

Pontificia Universidad Javeriana Cali
Facultad de Ingeniería y Ciencias
Maestría en Ciencia de Datos
Proyecto Aplicado

Modelado y Optimización de SuDS en la Pontificia Universidad
Javeriana Sede Bogotá mediante Machine Learning para la
Adaptación al Cambio Climático

Dónoban Steven Rojas Céspedes

Directora: Dr. Sandra Lorena Galarza Molina

Codirector: Dr. Andrés Eduardo Torres Abello

Junio 2025



Resumen

Este Proyecto Aplicado se centró en el modelado y la optimización de los Sistemas de Drenaje Sostenible (SuDS) en la Pontificia Universidad Javeriana, sede Bogotá, mediante la aplicación de técnicas de aprendizaje automático como estrategia de adaptación al cambio climático. Esta iniciativa cobra especial relevancia ante los crecientes desafíos que enfrentan los sistemas de drenaje —como el Sistema de Humedal Artificial/Tanque de Reservorio (SHATR) de la Javeriana— para gestionar volúmenes de agua durante eventos de lluvia intensa, en un contexto donde actualmente escasean los análisis predictivos que consideren escenarios climáticos extremos.

El problema central abordado fue la necesidad de fortalecer la capacidad de gestión y la resiliencia del SuDS frente a los impactos del cambio climático. El objetivo principal consistió en desarrollar e implementar modelos de aprendizaje automático, utilizando datos de precipitación y esorrentía, con el fin de evaluar la efectividad del SuDS y mejorar sus estrategias de gestión y adaptación ante diversos escenarios climáticos.

Entre los principales resultados, destaca el desarrollo de un modelo de Regresión por Vectores de Soporte (SVR), optimizado mediante búsqueda bayesiana, que presentó un desempeño predictivo sobresaliente para modelar el nivel de entrada del sistema, alcanzando un coeficiente de determinación (R^2) de 0.9024. La aplicación de escenarios de cambio climático, construidos a partir de proyecciones del IDEAM, reveló variaciones significativas en la frecuencia de eventos críticos y en los niveles promedio del sistema, lo cual resalta la necesidad urgente de implementar estrategias de adaptación flexibles.

Las conclusiones más relevantes evidencian la viabilidad de aplicar metodologías de aprendizaje automático, particularmente SVR, para modelar el comportamiento hidrológico del nivel de entrada de los SuDS bajo condiciones climáticas actuales y futuras. Estos hallazgos ofrecen insumos valiosos para el diseño de medidas preventivas y adaptativas. Entre las aplicaciones potenciales de esta investigación se contempla el uso del modelo desarrollado como herramienta para sistemas de alerta temprana y para la gestión operativa de otros sistemas de drenaje vulnerables a los efectos del cambio climático.

Este proyecto se enmarca en la necesidad urgente de superar las limitaciones de los sistemas de drenaje existentes ante eventos climáticos extremos, y destaca el potencial del aprendizaje automático como herramienta fundamental para el análisis predictivo y la gestión optimizada de los SuDS.

Palabras Clave: Sistemas de Drenaje Sostenible (SuDS), Aprendizaje Automático, Adaptación al Cambio Climático, Modelado Hidrológico, Predicción, Optimización.

Índice general

1. Introducción	1
2. Contextualización del Proyecto	2
2.1. Definición del Problema	2
2.1.1. Planteamiento del Problema	2
2.1.2. Formulación	3
2.1.3. Sistematización	3
2.2. Objetivos	3
2.2.1. Objetivo General	3
2.2.2. Objetivos Específicos	3
2.3. Marco de Referencia	4
2.3.1. Marco Teórico	4
2.3.2. Antecedentes	11
3. Análisis Exploratorio y Consolidación de Datos	14
3.1. Metodología de Recolección y Descripción Inicial de los Datos	14
3.2. Análisis Exploratorio de Datos	18
3.2.1. Preprocesamiento Inicial y Exploración General de los Datos	18
3.2.2. Algoritmo y Resultados de la Identificación de Eventos de Lluvia	20
3.2.3. Identificación y Estrategias para el Manejo de Datos Faltantes	24
3.2.4. Detección y Tratamiento de Valores Atípicos en los Datos	25
3.2.5. Análisis de las Relaciones entre las Variables del Sistema	39
3.2.6. Análisis de Componentes Principales	41
3.3. Consolidación Final y Preparación de la Base de Datos para la Modelación	45
4. Desarrollo y Evaluación de Modelos de Aprendizaje Automático	48
4.1. Justificación de la Selección de los Modelos de Aprendizaje Automático	48
4.1.1. Justificación del Tamaño del Conjunto de Datos y la Robustez del Modelado	49
4.2. Modelado del Nivel de Entrada del SuDS mediante Aprendizaje Automático	51
4.2.1. Entrenamiento y Evaluación Inicial de los Modelos para el Nivel de Entrada	52
4.2.2. Análisis Comparativo del Rendimiento de los Modelos Base y Optimizados	54
4.2.3. Análisis Detallado del Modelo SVR Optimizado para el Nivel de Entrada	60
4.2.4. Evaluación de una Versión Simplificada del Modelo SVR	65
4.3. Modelado del Nivel de Salida del SuDS mediante Aprendizaje Automático	68
4.3.1. Entrenamiento y Evaluación de Modelos para el Nivel de Salida	68
4.4. Justificación del Enfoque Predictivo en el Nivel de Entrada	70
4.4.1. Limitaciones en la Modelación del Nivel de Salida en SuDS	70
4.4.2. Decisión Metodológica sobre el Alcance de la Modelación	71

5. Desarrollo y Caracterización de Escenarios de Cambio Climático	72
5.1. Elaboración de Escenarios de Cambio Climático Específicos para el SuDS	72
5.1.1. Metodología para la Construcción de los Escenarios	72
5.1.2. Desarrollo de los Escenarios de Precipitación para el SuDS	72
5.1.3. Proyección de las Variables Hidrometeorológicas Bajo los Escenarios	73
5.1.4. Implementación de la Metodología de Escenarios en Python	77
5.1.5. Análisis Estadístico Comparativo de los Escenarios Proyectados	77
5.1.6. Limitaciones de la Metodología de Construcción de Escenarios	79
6. Predicción del Comportamiento Futuro del SuDS Bajo Escenarios de Cambio Climático	82
6.1. Aplicación del Modelo Predictivo a los Escenarios de Cambio Climático	82
6.1.1. Resultados de la Predicción del Modelo SVR	82
6.1.2. Distribución de Niveles de Agua Predichos en los Escenarios	83
6.1.3. Categorización de los Niveles de Agua y Evaluación de Riesgos Futuros	84
6.1.4. Análisis de Influencia de Variables en el SuDS Bajo Escenarios Futuros	86
6.1.5. Análisis Comparativo de Indicadores Clave Predichos entre Escenarios	89
6.1.6. Análisis del Sistema de Desborde del SuDS	90
6.1.7. Implicaciones para la Gestión del SuDS en Condiciones Climáticas Extremas	92
6.1.8. Conclusiones de la Predicción del Comportamiento del SuDS	92
6.2. Propuestas de Mejora del SuDS ante Predicciones de Eventos Extremos	93
6.2.1. Optimización de la capacidad y gestión del vertedero	93
6.2.2. Mejoras en la infiltración y retención del lecho granular y la vegetación	94
6.2.3. Consideraciones para la gestión integrada de los tanques	95
7. Conclusiones y Trabajos Futuros	97
7.1. Conclusiones y Trabajos Futuros	97
7.1.1. Conclusiones	97
7.1.2. Trabajos Futuros	98
8. Anexos	100
8.1. Enlaces a cuadernos de Google Colab - Código Python	100
Bibliografía	101

Índice de figuras

3.1. Mapa de Calor de Disponibilidad de Datos	20
3.2. Distribución de los niveles de entrada al SuDS (input_cm)	26
3.3. Variación de los Niveles de Entrada al SuDS (input_cm) durante los Eventos	26
3.4. Distribución de los niveles de salida al SuDS (output_cm)	27
3.5. Variación de los Niveles de Salida al SuDS (output_cm) durante los Eventos	28
3.6. Distribución de la variable lluvia (rainfall_in)	29
3.7. Variación de la Lluvia (rainfall_in) durante los Eventos	29
3.8. Distribución de la variable Radiación Solar (solar_rad_watm2)	30
3.9. Variación de la Radiación Solar (solar_rad_watm2) durante los Eventos	30
3.10. Distribución de la variable Humedad Relativa (rh)	31
3.11. Variación de la Humedad Relativa (rh) durante los Eventos	31
3.12. Distribución de la Temperatura (temperature_f)	32
3.13. Variación de la Temperatura (temperature_f) durante los Eventos	32
3.14. Distribución de la Dirección del Viento (wind_dir_deg)	33
3.15. Variación de la Dirección del Viento (wind_dir_deg) durante los Eventos	33
3.16. Distribución de las Ráfagas de Viento (wind_gust_mph)	34
3.17. Variación de las Ráfagas de Viento (wind_gust_mph) durante los Eventos	34
3.18. Distribución de la Velocidad del Viento (wind_speed_mph)	35
3.19. Variación de la Velocidad del Viento (wind_speed_mph) durante los Eventos	35
3.20. Distribución del Punto de Rocío (dew_point_f)	36
3.21. Variación del Punto de Rocío (dew_point_f) durante los Eventos	36
3.22. Distribución de la Lluvia Acumulada (cumulative_rain)	37
3.23. Variación de la Lluvia Acumulada (cumulative_rain) durante los Eventos	37
3.24. Matriz de correlación de Spearman.	40
3.25. Proporción de Varianza Explicada por cada Componente Principal	42
3.26. Cargas de los Componentes Principales	42
3.27. Proyección de los Datos en los Dos Primeros Componentes Principales	43
3.28. Varianza Acumulada de los Componentes Principales	44
4.1. Comparación del coeficiente de determinación (R^2) de modelos base y optimizados.	54
4.2. Comparación del error cuadrático medio (MSE) de modelos base y optimizados.	55
4.3. Comparación del error RMSE de modelos base y optimizados.	55
4.4. Comparación del error absoluto medio (MAE) de modelos base y optimizados.	56
4.5. Predicciones vs Valores Reales para todos los modelos.	57
4.6. Curvas de aprendizaje para SVR (Bayes).	60
4.7. Residuales vs. valores predichos para SVR (Bayes).	61
4.8. Distribución de residuales para SVR (Bayes).	62
4.9. Gráfico Q-Q de residuales para SVR (Bayes).	62

4.10. Residuales absolutos vs. valores predichos para SVR (Bayes).	63
4.11. Predicciones vs valores reales para los modelos SVR simplificados (Base y Bayes).	67
6.1. Distribución de niveles de agua entre escenarios.	83
6.2. Distribución de categorías de nivel de agua por escenario.	85
6.3. Factores que influyen en el nivel de agua.	87
6.4. Relación entre lluvia acumulada (mm) y nivel de agua (cm).	88
6.5. Comparativa de comportamiento del SuDS entre escenarios.	89

Índice de tablas

3.1. Descripción de Variables Recolectadas	17
3.2. Resumen de Valores Nulos por Variable	19
3.3. Resumen de Valores Nulos por Variable en base de Eventos	24
3.4. Datos Faltantes por Evento	25
3.5. Estadísticas descriptivas de las variables	38
3.6. Desviación estándar de las variables por evento	39
4.1. Métricas de evaluación para modelos base.	52
4.2. Métricas de evaluación para modelos optimizados mediante búsqueda bayesiana.	54
4.3. Importancia de características en la predicción del rendimiento del SuDS.	64
4.4. Métricas de desempeño para el modelo SVR simplificado.	65
4.5. Comparación entre el modelo SVR original y el modelo simplificado.	67
4.6. Métricas de evaluación para modelos base en la predicción de niveles de salida.	69
4.7. Métricas de evaluación para modelos optimizados en la predicción de niveles de salida.	69
5.1. Estadísticas descriptivas comparativas por escenario	78
6.1. Estadísticas descriptivas del nivel de agua por escenario	82
6.2. Análisis de eventos de desborde proyectados por escenario	91

Introducción

El aumento en la frecuencia e intensidad de eventos climáticos extremos, impulsado por el cambio climático, representa un desafío creciente para la gestión hídrica en entornos urbanos. En respuesta, los Sistemas de Drenaje Sostenible (SuDS), como el Sistema de Humedal Artificial/Tanque de Reservorio (SHATR) de la Pontificia Universidad Javeriana en Bogotá, han surgido como alternativas técnicas y ambientalmente responsables para el manejo de aguas pluviales. Sin embargo, su desempeño puede verse comprometido ante escenarios climáticos extremos, lo que hace indispensable una evaluación rigurosa y predictiva de su comportamiento futuro.

Este trabajo busca abordar dicha necesidad mediante la aplicación de técnicas avanzadas de aprendizaje automático, una metodología que ofrece ventajas significativas sobre los enfoques hidrodinámicos tradicionales, como una mayor eficiencia computacional y una mejor capacidad para modelar relaciones no lineales complejas entre variables meteorológicas e hidrológicas. En este contexto, se desarrolló un Proyecto Aplicado orientado al diseño e implementación de modelos de machine learning para evaluar y optimizar el rendimiento del SHATR, con el fin de fortalecer su capacidad de adaptación frente a distintos escenarios de cambio climático.

La investigación se fundamentó en la unión de bases de datos históricas recopiladas entre 2014 y 2016 en el campus universitario, suministrada por la Dra. Sandra Lorena Galarza Molina, que incluyó registros de lluvia, niveles de agua y variables meteorológicas [1]. El proceso incluyó la recolección, depuración, consolidación y análisis exploratorio de los datos, lo que permitió identificar y caracterizar 98 eventos completos de lluvia y retención en el sistema.

Sobre esta base, se entrenaron diversos modelos predictivos para estimar el nivel de entrada del sistema. El modelo de Regresión de Vectores de Soporte (SVR), optimizado con búsqueda bayesiana, alcanzó el mejor rendimiento ($R^2 = 0.9024$, $RMSE = 1.5940$ cm), validando su capacidad para predecir el comportamiento del sistema bajo condiciones presentes y futuras. El análisis de importancia de variables indicó que la lluvia acumulada, la intensidad máxima de precipitación y el nivel inicial del tanque son los principales factores que explican la respuesta del SHATR.

Se definieron además cuatro escenarios de cambio climático (2021–2040, 2041–2060, 2061–2080 y 2081–2100), basados en proyecciones del IDEAM. La aplicación del modelo a estos escenarios permitió anticipar un aumento en la frecuencia de eventos críticos y variaciones importantes en los niveles promedio de agua, lo que subraya la necesidad de estrategias de adaptación flexibles. En respuesta, se formularon propuestas técnicas para mejorar la resiliencia del SHATR, enfocadas en la gestión del vertedero, el refuerzo del lecho granular y vegetación, y la integración operativa de los tanques.

Este documento presenta, en primer lugar, el contexto, el problema, los objetivos y el marco teórico. Luego, se expone el análisis de datos y el desarrollo de modelos predictivos, seguido por la evaluación bajo escenarios futuros y las recomendaciones técnicas. Finalmente, se incluyen las conclusiones generales y las líneas de trabajo futuras.

Contextualización del Proyecto

2.1. Definición del Problema

2.1.1. Planteamiento del Problema

El cambio climático y la rápida urbanización están exacerbando significativamente los problemas de drenaje urbano, poniendo en riesgo la infraestructura y la seguridad de las ciudades. Estos fenómenos han incrementado la frecuencia e intensidad de las precipitaciones, sobrecargando los sistemas tradicionales de drenaje y provocando inundaciones recurrentes que amenazan tanto a la población como a las edificaciones urbanas. La urbanización ha llevado a una creciente impermeabilización del suelo, reduciendo la capacidad de infiltración del agua de lluvia y aumentando el volumen de escorrentía superficial que deben gestionar las redes de drenaje [2] [3].

El aumento en la intensidad de las precipitaciones, resultado directo del cambio climático, ha evidenciado la ineficacia de los sistemas convencionales de drenaje para manejar grandes volúmenes de agua en períodos cortos. Esto se traduce en un incremento de las inundaciones en áreas urbanas, generando graves consecuencias, desde daños en la infraestructura hasta riesgos para la salud pública y costos adicionales en mantenimiento y reparación [4]. Si no se toman medidas adecuadas, la situación podría empeorar, con un aumento en la frecuencia y severidad de las inundaciones.

Frente a estos desafíos, los Sistemas de Drenaje Sostenible (SuDS), y en particular el Sistema de Humedal Artificial/Tanque de Reservorio (SHATR) de la Pontificia Universidad Javeriana en Bogotá, han surgido como soluciones alternativas para la gestión del agua pluvial. Los SuDS permiten regular la cantidad y calidad del agua de lluvia que se libera a los sistemas hídricos o se infiltra en el suelo. Sin embargo, la efectividad de estos sistemas se ve comprometida por las variaciones climáticas extremas, lo que plantea la necesidad de evaluar su rendimiento actual y prever su comportamiento bajo diferentes escenarios futuros de cambio climático.

En este contexto, el SHATR de la Javeriana enfrenta retos considerables en cuanto a su capacidad para manejar los crecientes volúmenes de agua durante las temporadas de lluvias intensas. La falta de un análisis predictivo que considere escenarios climáticos extremos y la respuesta del sistema ante ellos dificulta la optimización y mejora de su gestión. Para abordar estas limitaciones, el uso de modelos de aprendizaje automático se presenta como una herramienta esencial. Estos modelos permitirán realizar análisis de datos históricos y actuales para predecir el rendimiento del SHATR bajo diversas condiciones climáticas, facilitando la identificación de posibles mejoras y adaptaciones necesarias para garantizar su resiliencia a largo plazo.

La urgencia de abordar este problema radica en las potenciales consecuencias de un sistema de drenaje inadecuado: inundaciones más frecuentes y severas, daños a la infraestructura del campus, interrupción de actividades académicas y de investigación, y posibles riesgos para la seguridad de la

comunidad universitaria. Asimismo, al ser la Universidad Javeriana una institución líder en sostenibilidad, la optimización de su SHATR podría servir como modelo para otras instituciones y entornos urbanos que enfrentan desafíos similares.

2.1.2. Formulación

¿Cómo pueden los modelos de aprendizaje automático ser utilizados para evaluar y predecir el rendimiento del SuDS de la Universidad Javeriana en Bogotá en la gestión de la cantidad de agua bajo diferentes escenarios de cambio climático?

2.1.3. Sistematización

1. ¿Qué tipos de datos climáticos e hidrológicos son necesarios para desarrollar modelos precisos de aprendizaje automático en la evaluación de SuDS?
2. ¿Cómo se pueden desarrollar y optimizar modelos de aprendizaje automático para simular y evaluar con precisión la respuesta del SuDS bajo diferentes escenarios climáticos?
3. ¿Qué métricas y criterios se pueden utilizar para evaluar la efectividad actual del SuDS en el campus, a partir de los resultados de los modelos de aprendizaje automático?
4. ¿Qué tipo de análisis y enfoques se pueden emplear para identificar e implementar mejoras específicas en el SuDS existente, con el fin de aumentar su resiliencia frente a eventos climáticos extremos?

2.2. Objetivos

2.2.1. Objetivo General

Desarrollar y aplicar modelos de aprendizaje automático para evaluar la efectividad del Sistema de Drenaje Sostenible (SuDS) de la Pontificia Universidad Javeriana en la sede Bogotá, utilizando datos de lluvia y escorrentía, con el fin de mejorar su gestión y adaptación ante diversos escenarios de cambio climático.

2.2.2. Objetivos Específicos

- Analizar y seleccionar los tipos de datos climáticos e hidrológicos necesarios para desarrollar modelos precisos de aprendizaje automático en la evaluación del SuDS en la sede Bogotá.
- Desarrollar modelos de aprendizaje automático para evaluar la efectividad del SuDS y prever su respuesta bajo diferentes escenarios climáticos.
- Evaluar la efectividad actual del SuDS en el campus mediante los modelos desarrollados, identificando fortalezas y debilidades, para establecer una línea base de desempeño.

- Proponer mejoras específicas en el SuDS existente basadas en los resultados de los modelos de aprendizaje automático, con el objetivo de aumentar su resiliencia y efectividad frente a futuros eventos climáticos extremos.

2.3. Marco de Referencia

2.3.1. Marco Teórico

2.3.1.1. Sistemas de Drenaje Sostenible (SuDS)

Los Sistemas de Drenaje Sostenible (SuDS) son una alternativa eficiente y respetuosa con el medio ambiente para la gestión de las aguas pluviales en zonas urbanas. Estos sistemas se basan en técnicas de drenaje que imitan los procesos naturales de infiltración, retención y depuración del agua, permitiendo gestionar la escorrentía de una manera más cercana al ciclo hidrológico natural que los sistemas convencionales [5].

Los SuDS utilizan una combinación de técnicas, como estanques de retención, zanjas de infiltración, cubiertas vegetales y pavimentos permeables, que permiten la infiltración, captura y reutilización del agua de lluvia. Estos sistemas ofrecen múltiples beneficios ecosistémicos, incluyendo:

- Reducción del riesgo de inundaciones mediante la atenuación de caudales pico
- Mejora de la calidad del agua a través de procesos naturales de filtración y sedimentación
- Reducción del escurrimiento superficial y la erosión consecuente
- Recarga de acuíferos y conservación de recursos hídricos
- Promoción de la biodiversidad urbana mediante la creación de hábitats
- Creación de espacios verdes multifuncionales que mejoran la calidad paisajística
- Mitigación de los efectos de isla de calor urbana

Frente al cambio climático, los SuDS emergen como soluciones resilientes que pueden adaptarse a condiciones meteorológicas cambiantes y extremas, ofreciendo una alternativa sostenible a los sistemas de drenaje convencionales [5].

2.3.1.2. Sistema de Humedal Artificial/Tanque de Reservorio (SHATR) de la Pontificia Universidad Javeriana Bogotá

El Sistema de Humedal Artificial/Tanque de Reservorio (SHATR) o Constructed-Wetland/Reservoir-Tank (CWRT) de la Pontificia Universidad Javeriana Bogotá (PUJB) representa un SuDS innovador que integra elementos de captación, tratamiento y almacenamiento de aguas pluviales. Este sistema fue construido entre 2012 y 2013 como parte de un proyecto de investigación enfocado en la recolección y aprovechamiento sostenible de aguas lluvias.

Características Técnicas de Diseño El SHATR cuenta con dos entradas principales que reciben escorrentía de diferentes áreas del campus:

- Entrada del parqueadero: capta escorrentía de un área de 3776 m²
- Entrada de la cancha de fútbol y zonas verdes: recoge agua de un área de 14816 m²

El sistema incluye dos tanques de sedimentación: uno ubicado antes del humedal artificial, que recibe la escorrentía del parqueadero, y otro que recibe la escorrentía de la cancha de fútbol y zonas verdes. El humedal artificial fue específicamente diseñado para mejorar la calidad de la escorrentía proveniente del parqueadero, ya que esta contiene mayor concentración de metales pesados y otros contaminantes asociados al tráfico vehicular [1].

Humedal Artificial El humedal es de flujo subsuperficial horizontal, plantado con *Cyperus papyrus*, una especie con alta capacidad de absorción de contaminantes y adaptada a condiciones de saturación. El lecho del humedal está organizado ingeniosamente en tres zonas con tamaños de grava decrecientes (de mayor a menor diámetro) para minimizar posibles obstrucciones en el sistema y optimizar los procesos de filtración. Esta configuración permite un flujo gradual y un tratamiento efectivo del agua [1].

Sistema de Monitoreo El SHATR cuenta con un sofisticado sistema de monitoreo que incluye:

- Un sensor ultrasónico de nivel ubicado en la entrada del tanque de sedimentación, que registra el nivel del agua cada minuto
- Un captador UV-Vis (spectro::lyser™ de s::can) instalado en el influente del sistema, que recolecta un espectro de absorbancia UV-Vis por minuto para el monitoreo de la calidad del agua

Diseño Basado en Demanda Durante la fase de diseño del SHATR, se implementó una metodología que considera tanto la probabilidad de abastecer la demanda de agua como sus variabilidades temporales, lo que permite optimizar la capacidad de almacenamiento y la eficiencia del sistema [1].

Es importante destacar que el trabajo doctoral de Galarza (2017) proporciona un análisis exhaustivo del comportamiento y desempeño del SHATR, incluyendo el desarrollo de una herramienta de apoyo para la toma de decisiones en la operación del sistema de captación de aguas lluvias, así como la implementación de protocolos operativos y métodos de calibración de instrumentos. Este sistema CWRT es precisamente el SuDS sobre el que se realizará el proyecto propuesto [1].

2.3.1.3. Cambio Climático y su Impacto en los SuDS

El cambio climático plantea desafíos significativos para la gestión de los SuDS. Se espera que el aumento de las temperaturas, los cambios en los patrones de precipitación y la mayor frecuencia e intensidad de eventos climáticos extremos afecten la eficacia y la resiliencia de estos sistemas [6].

Desafíos Climáticos Por un lado, el aumento de las precipitaciones intensas puede sobrecargar los SuDS y provocar inundaciones, erosionando así la capacidad de estos sistemas para gestionar el agua de manera sostenible. Por otro lado, los periodos de sequía prolongados pueden reducir la disponibilidad de agua para la recarga de acuíferos y el mantenimiento de la vegetación en los SuDS [7].

Adaptación de los SuDS Ante estos desafíos, es crucial adaptar los SuDS para hacer frente a las nuevas condiciones climáticas. Esto implica la revisión de los diseños y criterios de dimensionamiento para contemplar escenarios climáticos futuros, así como la integración de tecnologías adicionales, como sistemas avanzados de almacenamiento de agua, para gestionar la variabilidad en la disponibilidad de recursos hídricos. Las estrategias de adaptación pueden incluir:

- Rediseño de componentes para aumentar la capacidad de retención e infiltración
- Incorporación de sistemas de alerta temprana y monitoreo en tiempo real
- Implementación de estructuras modulares que permitan ajustes según las condiciones cambiantes
- Uso de vegetación resiliente a condiciones climáticas extremas

SuDS como Elementos de Mitigación Además de adaptarse al cambio climático, los SuDS pueden desempeñar un papel activo en la mitigación de sus impactos. Al promover la infiltración y la captura de agua de lluvia, estos sistemas contribuyen a:

- Reducir la escorrentía superficial y la erosión del suelo
- Disminuir el riesgo de inundaciones urbanas
- Mejorar la calidad del agua mediante procesos naturales de filtración
- Fomentar la biodiversidad urbana a través de la creación de hábitats
- Contribuir a la reducción del efecto isla de calor en entornos urbanos

Por lo tanto, los SuDS no solo deben adaptarse al cambio climático, sino que también representan una estrategia efectiva para la mitigación de sus efectos en entornos urbanos [7].

2.3.1.4. Modelado de Sistemas de Drenaje

El modelado de los sistemas de drenaje es una tarea crítica para evaluar su efectividad y gestionar adecuadamente el riesgo de inundaciones. Tradicionalmente, se han utilizado modelos hidrodinámicos/hidráulicos físicamente basados, como los modelos unidimensionales (1D), bidimensionales (2D) o combinados (1D/2D) [8]. Estos modelos permiten simular procesos complejos, como la escorrentía superficial, el flujo en tuberías y canales, y la inundación de áreas urbanas.

Sin embargo, los modelos físicamente basados presentan limitaciones significativas que restringen su aplicabilidad en ciertos contextos:

- Alta demanda computacional que requiere infraestructura especializada
- Largos tiempos de simulación que dificultan su uso en aplicaciones de tiempo real
- Necesidad de extensos conjuntos de datos para su calibración y validación
- Complejidad en la representación de ciertos procesos físicos y su interacción

Estas limitaciones han impulsado la búsqueda de enfoques alternativos y complementarios para el modelado de sistemas de drenaje, entre los que destacan las técnicas de aprendizaje automático [8].

2.3.1.5. Aprendizaje Automático en Sistemas de Drenaje

El aprendizaje automático (Machine Learning, ML) representa una alternativa prometedora frente a los modelos hidrodinámicos tradicionales para el análisis y gestión de los sistemas de drenaje. Esta rama de la inteligencia artificial se enfoca en el desarrollo de algoritmos y modelos capaces de aprender y mejorar automáticamente a partir de datos [9], sin necesidad de resolver explícitamente las complejas ecuaciones físicas que gobiernan los fenómenos hidrológicos e hidráulicos.

Ventajas Frente a Modelos Tradicionales Las técnicas de aprendizaje automático ofrecen diversas ventajas en comparación con los modelos hidrodinámicos convencionales:

- Mayor eficiencia computacional y tiempos de ejecución más cortos
- Capacidad para capturar relaciones no lineales complejas en los datos
- Flexibilidad para incorporar fuentes de datos heterogéneas
- Adaptabilidad a condiciones cambiantes mediante actualización con nuevos datos
- Facilidad para su integración en sistemas de toma de decisiones en tiempo real

Tipos de Datos y Entrenamiento Los modelos de aprendizaje automático pueden ser entrenados con diversos tipos de datos relacionados con los sistemas de drenaje:

- Series temporales de precipitación y escorrentía
- Registros de niveles de agua en diferentes componentes del sistema
- Datos de calidad del agua y concentración de contaminantes
- Características físicas del entorno urbano (topografía, uso del suelo, etc.)
- Variables meteorológicas como temperatura, humedad y velocidad del viento

A partir de estos datos, los modelos aprenden los patrones subyacentes y las relaciones entre variables, permitiéndoles simular y predecir el comportamiento de los SuDS bajo diferentes condiciones climáticas y escenarios futuros [8].

Técnicas de Aprendizaje Automático Aplicadas Diversas técnicas de aprendizaje automático han sido aplicadas en el contexto de los sistemas de drenaje:

- Modelos de Regresión (lineal, polinomial, ElasticNet)
- Árboles de Decisión y Métodos de Ensamble (Random Forest, XGBoost)
- Máquinas de Vectores de Soporte (SVM) y su variante para regresión (SVR)
- Redes Neuronales Artificiales (ANN) y sus variantes
- Modelos de Aprendizaje Profundo (Deep Learning) como CNN y RNN

Cada una de estas técnicas presenta características específicas que las hacen más adecuadas para determinados tipos de problemas y conjuntos de datos, lo que permite seleccionar el enfoque más apropiado según los objetivos del análisis y las características del sistema estudiado [8].

2.3.1.6. Base Teórica de los Modelos de Aprendizaje Automático Seleccionados

Para la evaluación y predicción del comportamiento del SuDS de la Universidad Javeriana, se han seleccionado diversos modelos de aprendizaje automático, cada uno con fundamentos teóricos específicos que los hacen apropiados para el análisis de datos hidrometeorológicos:

Modelos de Regresión Lineal y ElasticNet Los modelos de regresión lineal establecen relaciones entre variables mediante una función lineal, estimando coeficientes que minimizan el error cuadrático entre predicciones y valores observados. ElasticNet extiende este enfoque incorporando regularización L1 (Lasso) y L2 (Ridge), lo que permite seleccionar variables relevantes y manejar la multicolinealidad, características particularmente útiles cuando se trabaja con múltiples variables hidrometeorológicas correlacionadas [9].

Random Forest y XGBoost Los métodos de ensamble como Random Forest y XGBoost combinan múltiples modelos simples (árboles de decisión) para mejorar la precisión y robustez de las predicciones. Random Forest construye múltiples árboles independientes mediante muestreo bootstrap de los datos y selección aleatoria de variables, promediando sus predicciones para reducir el sobreajuste. XGBoost utiliza el principio de boosting, construyendo árboles secuencialmente donde cada nuevo árbol corrige los errores de los anteriores, incorporando regularización para controlar la complejidad del modelo. Estos métodos son particularmente eficaces para capturar relaciones no lineales en datos hidrometeorológicos y manejar valores atípicos [9].

Regresión por Vectores de Soporte (SVR) La SVR extiende los principios de las Máquinas de Vectores de Soporte al dominio de la regresión, buscando una función que se desvíe de los valores observados en no más de una cantidad ϵ , mientras se mantiene lo más plana posible. Mediante el uso de funciones kernel (como RBF), SVR puede mapear los datos a espacios de mayor dimensión, permitiendo modelar relaciones altamente no lineales entre variables hidrometeorológicas. SVR es menos

sensible a valores atípicos que la regresión tradicional, lo que resulta ventajoso al trabajar con datos de precipitación que pueden presentar eventos extremos [9].

La selección de estos modelos se justifica por su capacidad para manejar tanto relaciones lineales como no lineales, su robustez ante diferentes tipos de datos y su adecuación para conjuntos de datos de tamaño moderado (98 muestras en este caso), lo que permite evitar el sobreajuste mientras se capturan patrones relevantes en el comportamiento del SuDS.

2.3.1.7. Aplicaciones del Aprendizaje Automático en Sistemas de Drenaje

El aprendizaje automático ha encontrado diversas aplicaciones en el ámbito de los sistemas de drenaje, incluyendo la operación en tiempo real, la gestión de inundaciones y el mantenimiento de infraestructura. Algunas de estas aplicaciones se describen a continuación:

Operación en Tiempo Real Los modelos de aprendizaje automático, especialmente los basados en técnicas de refuerzo (reinforcement learning), pueden utilizarse para optimizar la operación en tiempo real de componentes clave de los sistemas de drenaje, como bombas y compuertas [10, 11]. Estos modelos aprenden a tomar decisiones óptimas de control en función de las condiciones actuales del sistema y los objetivos deseados, como minimizar el riesgo de inundación o maximizar la eficiencia energética.

Predicción de Inundaciones Una de las aplicaciones más importantes del aprendizaje automático en sistemas de drenaje es la predicción de inundaciones. Los modelos basados en técnicas como redes neuronales (ANN, CNN, RNN) y árboles de decisión pueden ser entrenados con datos históricos de precipitación, escorrentía y niveles de agua para predecir la magnitud, duración y extensión de las inundaciones en áreas urbanas [12, 13, 14, 8].

Estas predicciones son cruciales para la toma de decisiones y la implementación de medidas de mitigación, como la evacuación de áreas en riesgo o la activación de sistemas de control de inundaciones.

Detección de Defectos en Tuberías El mantenimiento de la infraestructura de drenaje urbano es esencial para garantizar su correcto funcionamiento. En este sentido, los modelos de aprendizaje automático, particularmente los basados en redes neuronales convolucionales (CNN), han demostrado su eficacia en la detección automática de defectos en tuberías a partir de imágenes obtenidas mediante inspecciones de circuito cerrado de televisión (CCTV) [15, 16, 17, 18].

Estas técnicas permiten identificar y clasificar defectos como grietas, obstrucciones, raíces y depósitos, lo que facilita la planificación de tareas de mantenimiento y reparación.

2.3.1.8. Evaluación y Optimización de Sistemas de Drenaje

La evaluación y optimización de los sistemas de drenaje es un proceso fundamental para garantizar su eficacia y adaptabilidad a las condiciones cambiantes. Este proceso implica un enfoque metodológico riguroso y multidisciplinario que abarca diversas etapas y consideraciones.

Proceso de Evaluación La evaluación integral de un SuDS requiere la recopilación y análisis de diversos tipos de datos:

- Datos climáticos: series temporales de precipitación, temperatura, humedad y otros parámetros meteorológicos relevantes
- Datos hidrológicos: caudales, niveles de agua, tiempos de concentración y volúmenes de escorrentía
- Datos de infraestructura: características físicas, estado de los componentes y capacidad de los elementos del sistema
- Datos de calidad del agua: concentraciones de contaminantes, parámetros fisicoquímicos y biológicos

El análisis de estos datos permite identificar patrones de comportamiento, evaluar la respuesta del sistema ante diferentes eventos y determinar su capacidad para cumplir con los objetivos de diseño bajo las condiciones actuales y proyectadas [19].

Contribución de los Modelos de Aprendizaje Automático Los modelos de aprendizaje automático desempeñan un papel crucial en la evaluación y optimización de los SuDS, permitiendo:

- Simular el comportamiento del sistema bajo diferentes escenarios climáticos, incluyendo eventos extremos y condiciones futuras
- Identificar relaciones complejas entre variables que influyen en el rendimiento del sistema
- Evaluar la sensibilidad del sistema a cambios en parámetros específicos
- Generar predicciones rápidas que pueden alimentar procesos de optimización
- Sugerir configuraciones optimizadas basadas en objetivos múltiples

Optimización Multidisciplinaria La optimización de los SuDS es un proceso iterativo y multidisciplinario que debe considerar diversos factores:

- Factores técnicos: rendimiento hidráulico, eficiencia de tratamiento y durabilidad de componentes
- Factores económicos: costos de implementación, operación, mantenimiento y reemplazo
- Factores ambientales: impactos ecológicos, servicios ecosistémicos y sostenibilidad
- Factores sociales: aceptación pública, beneficios recreativos y educativos, y equidad en el acceso
- Factores urbanísticos: integración con el entorno construido, requisitos normativos y planificación urbana

La consideración holística de estos aspectos permite desarrollar soluciones optimizadas que maximizan los beneficios del sistema mientras se adaptan a las condiciones específicas del sitio y a los desafíos futuros del cambio climático [19, 8].

2.3.2. Antecedentes

Se han identificado y seleccionado cuatro trabajos clave que abordan temas directamente relacionados con el presente proyecto, proporcionando una base sólida para el desarrollo y la implementación de los objetivos planteados. Estos trabajos cubren aspectos fundamentales desde la modelación mediante aprendizaje automático hasta la caracterización del sistema específico que se estudia.

2.3.2.1. Modeling and Interpreting Hydrological Responses of Sustainable Urban Drainage Systems with Explainable Machine Learning Methods (Yang & Chui, 2020)

Este estudio explora la modelación de las respuestas hidrológicas de los SuDS mediante métodos de aprendizaje automático explicables. El enfoque principal está en superar tres desafíos que limitan el uso de estas técnicas: la necesidad de datos hidrometeorológicos de alta dimensionalidad, la falta de transparencia en las predicciones y la disponibilidad limitada de datos observacionales. Yang y Chui utilizan el método SHAP para explicar las predicciones de los modelos, lo que aporta transparencia y confianza en los resultados. Además, implementan técnicas avanzadas como la validación cruzada anidada y la ingeniería de características a partir de series temporales de precipitación para garantizar la fiabilidad y precisión de los modelos desarrollados [20].

Este trabajo resulta particularmente relevante para el presente proyecto porque ofrece una metodología robusta para la aplicación de técnicas de aprendizaje automático en un contexto donde los datos pueden ser limitados, como es el caso del SuDS de la Universidad Javeriana. La transparencia de los modelos a través de métodos explicables permitirá no solo predecir el comportamiento del sistema, sino también comprender los factores que más influyen en su respuesta hidrológica bajo diferentes escenarios climáticos.

2.3.2.2. Machine Learning and Urban Drainage Systems: State-of-the-Art Review (Kwon et al., 2021)

Esta revisión del estado del arte en SuDS identifica las limitaciones críticas de los modelos hidrodinámicos tradicionales, particularmente su alta demanda computacional, tiempos de simulación prolongados y la complejidad para representar adecuadamente fenómenos no lineales en entornos urbanos. Kwon et al. demuestran cómo las técnicas de aprendizaje automático (ML) pueden superar estas barreras, mejorando significativamente la eficiencia y precisión en la operación, predicción y mantenimiento de los sistemas de drenaje. La revisión clasifica el uso de ML en tres áreas principales: operación en tiempo real, predicción de inundaciones y detección de defectos en tuberías [8].

Esta revisión proporciona una justificación sólida para el uso de técnicas de ML en nuestro proyecto, específicamente en el contexto de la predicción del comportamiento del SuDS bajo escenarios climáticos variables. Las limitaciones computacionales de los modelos hidrodinámicos tradicionales, destacadas por Kwon et al., son particularmente relevantes para nuestro caso de estudio, donde se busca evaluar múltiples escenarios futuros de forma eficiente y adaptable. El enfoque en ML para predicción de inundaciones respalda directamente nuestro objetivo de evaluar la capacidad del sistema para gestionar eventos extremos de precipitación, una preocupación central en el contexto del cambio climático para el campus universitario.

2.3.2.3. Sustainable Urban Drainage System (SuDS) Modeling Supporting Decision-Making: A Systematic Quantitative Review (Ferrans et al., 2021)

Este estudio presenta una revisión cuantitativa de los modelos de SuDS y su integración en sistemas de soporte para la toma de decisiones (DSS). Ferrans et al. analizan el desempeño de estos sistemas desde múltiples perspectivas, incluyendo la económica, ambiental y social. Los autores destacan la relevancia de adoptar un enfoque interdisciplinario para optimizar los SuDS frente a los desafíos que plantea el cambio climático, subrayando la importancia de los datos y las tecnologías avanzadas en la implementación de estrategias de toma de decisiones adaptativas [19].

La relevancia de este trabajo para nuestro proyecto radica en su énfasis en la necesidad de un enfoque holístico para la evaluación y mejora de los SuDS. Aunque nuestro proyecto se centra inicialmente en el uso de aprendizaje automático para evaluar el rendimiento hidráulico del sistema, la perspectiva de Ferrans et al. nos orienta hacia la consideración de factores adicionales en la fase de propuesta de mejoras, como los impactos ambientales de las modificaciones, la relación costo-beneficio de las intervenciones propuestas y la aceptación por parte de la comunidad universitaria. Este enfoque multidimensional es esencial para garantizar que las mejoras propuestas sean tanto técnicamente efectivas como sostenibles en el contexto institucional específico de la Universidad Javeriana.

2.3.2.4. Climate Change Impact Assessment on Urban Rainfall Extremes and Urban Drainage (Willems et al., 2012)

Willems et al. presentan un análisis exhaustivo de los impactos del cambio climático en los extremos de precipitación urbana y los sistemas de drenaje. El estudio examina cómo los cambios proyectados en los patrones de precipitación afectan directamente el rendimiento de la infraestructura de drenaje urbano. Los autores identifican tendencias preocupantes, como el aumento en la intensidad de lluvias de corta duración, mayores volúmenes de precipitación en eventos extremos y cambios en la distribución estacional de las lluvias [21].

Estos hallazgos son fundamentales para nuestro proyecto, ya que proporcionan un marco conceptual sólido para la definición de los escenarios climáticos que se utilizarán en la evaluación del SuDS de la Universidad Javeriana. El trabajo de Willems et al. resalta la necesidad de considerar no solo cambios en los valores medios de precipitación, sino también en los extremos y en la variabilidad temporal, aspectos que incorporaremos en nuestros modelos predictivos. Además, su análisis de las implicaciones para los sistemas de drenaje, como el aumento en la frecuencia de desbordamientos y la disminución de los tiempos de retorno de eventos extremos, refuerza la urgencia y relevancia de nuestro proyecto para garantizar la resiliencia del campus universitario frente al cambio climático.

2.3.2.5. Decision-making tool for the operation of a stormwater harvesting system (Galarza, 2017)

Esta tesis doctoral, desarrollada por la Dra. Sandra Lorena Galarza Molina (directora del presente proyecto), se enfoca específicamente en el comportamiento y rendimiento del Sistema de Humedal Artificial/Tanque de Reservorio (SHATR) de la Pontificia Universidad Javeriana, sede Bogotá. El trabajo de Galarza aborda tres aspectos fundamentales que son directamente aplicables a nuestro estudio:

- La caracterización detallada del comportamiento hidrodinámico e hidrológico del SHATR bajo diferentes condiciones de precipitación y escorrentía
- El desarrollo de una herramienta de apoyo para la toma de decisiones en la operación del sistema de captación de aguas pluviales, basada en datos en línea sobre cantidad y calidad del agua
- La implementación y validación de protocolos operativos y métodos para la calibración de instrumentos de medición que garantizan la precisión de los datos recopilados

La investigación incluye un extenso conjunto de datos recolectados mediante el sistema de monitoreo instalado, que registra niveles de agua y parámetros de calidad a intervalos de un minuto. Estos datos constituyen una base empírica invaluable para el entrenamiento y validación de nuestros modelos de aprendizaje automático [1].

La participación directa de la Dra. Galarza como directora de este proyecto garantiza la transferencia efectiva de conocimientos y metodologías desarrolladas en su investigación doctoral, así como una comprensión profunda del sistema que se estudia. Esta continuidad investigativa permite que nuestro proyecto se base en un conocimiento sólido del SHATR, enfocándose en la nueva dimensión que aporta el aprendizaje automático para la evaluación y adaptación del sistema frente al cambio climático.

2.3.2.6. Síntesis de los Antecedentes

Los trabajos analizados proporcionan colectivamente una base sólida y multifacética para el presente proyecto. La investigación de Yang y Chui (2020) aporta metodologías robustas para la aplicación de técnicas de aprendizaje automático explicables en SuDS. La revisión de Kwon et al. (2021) justifica el uso de ML como alternativa a los modelos tradicionales, mientras que Ferrans et al. (2021) enfatizan la importancia de un enfoque holístico en la evaluación y mejora de estos sistemas. Willems et al. (2012) contextualizan la urgencia del proyecto al detallar los impactos del cambio climático en los SuDS. Finalmente, el trabajo de Galarza (2017) proporciona un conocimiento profundo del sistema específico que se estudia, así como datos y metodologías directamente aplicables.

Este proyecto busca integrar estos antecedentes para abordar una brecha significativa en la literatura actual: la aplicación de técnicas avanzadas de aprendizaje automático para evaluar y optimizar un SuDS específico frente a escenarios de cambio climático, con el propósito de desarrollar estrategias de adaptación concretas y aplicables en el contexto de una institución educativa urbana.

Análisis Exploratorio y Consolidación de Datos

Este capítulo se centra en la etapa crucial del Análisis Exploratorio y Consolidación de Datos, fundamentado en el extenso conjunto de información recopilada del SHATR de la Pontificia Universidad Javeriana de Bogotá, proporcionado por la Dra. Sandra Lorena Galarza Molina. La evaluación precisa del rendimiento del SuDS requiere una comprensión profunda de la naturaleza y calidad de los datos disponibles. Inicialmente, se detalla la metodología de recolección y descripción inicial de los datos, que incluye mediciones de niveles de agua y monitoreo meteorológico, y el proceso de integración y limpieza de los datos recopilados. Posteriormente, se aborda el Análisis Exploratorio de Datos (EDA), que implica el preprocesamiento inicial, la identificación y caracterización de eventos de lluvia y retención mediante un algoritmo específico, y el tratamiento de datos faltantes y valores atípicos para asegurar la calidad de la información. Finalmente, la sección concluye con el análisis de las relaciones entre las variables del sistema, incluyendo análisis de correlación y de componentes principales, culminando en la consolidación final y preparación de la base de datos para la modelación, con un enfoque particular en los eventos de lluvia y las métricas clave relevantes para el estudio.

3.1. Metodología de Recolección y Descripción Inicial de los Datos

Este estudio se basa en un extenso conjunto de datos recopilados del Sistema de Humedal Artificial/Tanque de Reservorio (SHATR), un innovador Sistema de Drenaje Sostenible (SuDS) implementado en la Pontificia Universidad Javeriana de Bogotá (PUJB). Los datos fueron proporcionados por la Dra. Sandra Lorena Galarza Molina, una figura clave en el desarrollo y estudio de este sistema.

La Dra. Galarza, en su tesis doctoral de 2017 titulada "*Decision-making tool for the operation of a stormwater harvesting system*", realizó un estudio exhaustivo sobre el comportamiento y rendimiento del SHATR [1]. Su trabajo se centró en el desarrollo de una herramienta de toma de decisiones para la captación de aguas pluviales, utilizando datos en línea sobre cantidad y calidad del agua. La implementación de protocolos operativos y métodos para calibrar instrumentos de medición fue crucial para facilitar la operación eficiente del sistema bajo condiciones reales.

El conjunto de datos utilizado en el presente estudio abarca un período de tres años, desde 2014 hasta 2016, proporcionando una visión longitudinal del funcionamiento del SHATR. Esta extensión temporal permite la identificación de patrones estacionales, tendencias a largo plazo y la respuesta del sistema a una amplia gama de condiciones meteorológicas.

La recolección de datos se centró en dos aspectos fundamentales:

1. Los niveles de agua en puntos críticos del sistema SHATR.

2. Las condiciones meteorológicas prevalentes durante el período de estudio.

Esta combinación de datos hidráulicos y meteorológicos ofrece una oportunidad única para analizar la eficacia del SuDS en la gestión de aguas pluviales y su respuesta a diversos eventos climáticos. La participación de la Dra. Galarza como directora de este proyecto ha facilitado la transferencia de conocimientos y herramientas clave, permitiendo una evaluación más profunda del sistema en el contexto del cambio climático.

En las siguientes subsecciones, se detallará la metodología empleada para la medición de niveles de agua y el monitoreo meteorológico, seguido por el proceso de consolidación y estructuración de los datos para su análisis posterior.

Detalles de la Medición de Niveles de Agua en el SuDS

La medición de los niveles de agua en el SHATR se llevó a cabo mediante sensores ultrasónicos de alta precisión, con una resolución de 1 mm y un rango de operación de 0.10 m a 1.5 m. Estos sensores fueron instalados en vertederos triangulares de cresta afilada, ubicados tanto en los puntos de entrada como de salida del humedal. La recolección y gestión de los datos se realizó mediante el software *Lookout - Web-Enabled Automation Software* (National Instruments), optimizando el control y monitoreo del sistema.

Descripción del Monitoreo y las Variables Meteorológicas Recopiladas

El monitoreo meteorológico se llevó a cabo mediante una estación meteorológica *WatchDog 2000*, un dispositivo reconocido por su alta precisión y versatilidad en la recopilación de datos climáticos. Equipado con un pluviómetro que ofrece una resolución de 0.25 mm (0.01 pulgadas), esta estación no solo mide la precipitación, sino que también registra una variedad de parámetros atmosféricos relevantes, tales como radiación solar, temperatura, humedad relativa, dirección y velocidad del viento.

Ubicada a 100 metros del SuDS, la estación registraba los datos a intervalos de un minuto, permitiendo una monitorización detallada y continua de las condiciones meteorológicas. Los datos recopilados se descargaban semanalmente para su posterior análisis, lo que facilitó una comprensión integral de la influencia de las variables climáticas sobre el rendimiento del sistema. Este enfoque multidimensional es crucial para evaluar la efectividad del SuDS en la gestión de aguas pluviales y su respuesta a los eventos meteorológicos.

Proceso de Integración y Limpieza de los Datos Recopilados

El proceso de consolidación de datos implicó la integración de dos conjuntos principales de información: los niveles de agua registrados por los sensores ultrasónicos y los datos meteorológicos recopilados por la estación *WatchDog 2000*. Este proceso fue fundamental para crear un conjunto de datos completo y coherente que permitiera un análisis exhaustivo del rendimiento del SuDS.

Procesamiento de Datos de Niveles de Agua

Los datos de niveles de agua se encontraban almacenados en archivos CSV, organizados mensualmente para cada año del período de estudio (2014-2016). Cada archivo contenía registros minuto a

minuto de los niveles de agua en diferentes puntos del sistema. Para este estudio, se enfocó la atención en dos puntos críticos:

- Nivel 1: Correspondiente a la entrada del SuDS
- Nivel 5: Correspondiente a la salida del SuDS

El proceso de consolidación de estos datos presentó varios desafíos:

1. **Variabilidad en la estructura de los archivos:** Algunos CSV mostraban diferencias en su estructura, nombres de columnas o formato temporal.
2. **Inconsistencias en la precisión temporal:** Mientras la mayoría de los registros se ajustaban a intervalos de un minuto, algunos incluían datos con precisión de segundos.
3. **Limitaciones de software:** Debido al volumen de datos, fue necesario dividir la información en múltiples archivos para su procesamiento inicial.

Para abordar estos desafíos, se implementó un enfoque híbrido:

- Se desarrollaron scripts en Python para automatizar la unificación de los archivos CSV.
- En casos donde las discrepancias de formato eran significativas, se realizó una consolidación manual utilizando Microsoft Excel.
- El resultado fue la creación de dos bases de datos principales: `datacleaned` con 1,048,575 registros (límite máximo de filas en Excel) y `datacleaned2` con 114,085 registros.
- Finalmente, estas dos bases se unificaron en Python para su posterior análisis.

Procesamiento de Datos Meteorológicos

Los datos meteorológicos, provenientes de la estación WatchDog 2000, presentaban su propio conjunto de desafíos:

- Los datos se almacenaban en archivos de texto (TXT) con frecuencia semanal.
- Se observaron inconsistencias en las unidades de medida entre algunos archivos.
- La frecuencia de registro era de un minuto, coincidiendo con los datos de niveles de agua.

El proceso de consolidación para estos datos incluyó:

1. Conversión de unidades para estandarizar las mediciones en todos los archivos.
2. Unificación de los archivos semanales en una sola base de datos.
3. Integración con la base de datos de niveles de agua, utilizando la fecha y minuto como variables de enlace.

El resultado final de este exhaustivo proceso de consolidación fue un *DataFrame* unificado que contiene 12 variables y 1,162,660 registros. Este conjunto de datos proporciona una base sólida para el análisis exploratorio subsiguiente y la evaluación del rendimiento del sistema SuDS en respuesta a diversas condiciones meteorológicas y de flujo de agua.

Descripción Detallada de la Estructura del Conjunto de Datos Consolidado

Tras el proceso de recolección y consolidación, se obtuvo un conjunto de datos robusto y completo que proporciona una visión integral del funcionamiento del SuDS y las condiciones meteorológicas asociadas. Esta sección presenta una descripción detallada de las variables incluidas en el conjunto de datos final, facilitando su interpretación y análisis posterior. El conjunto de datos consolidado se estructura en un *DataFrame* que contiene 12 variables clave y un total de 1,162,660 registros. Cada registro representa una observación minuto a minuto, capturando simultáneamente los niveles de agua en el SuDS y las condiciones meteorológicas prevalentes. La Tabla 3.1 proporciona una visión general de estas variables:

Variable	Descripción	Tipo de Dato
date	Fecha de la observación, formato: 'DD/MM/YYYY'	object
time	Hora de la observación, formato: 'HH:MM:SS'	object
input_cm	Lectura del sensor ultrasónico en la entrada en centímetros	float64
output_cm	Lectura del sensor ultrasónico en la salida en centímetros	float64
rainfall_in	Precipitación en pulgadas	float64
solar_rad_watm2	Radiación solar en vatios por metro cuadrado	float64
rh	Humedad relativa (%)	float64
temperature_f	Temperatura en grados Fahrenheit	float64
wind_dir_deg	Dirección del viento en grados	float64
wind_gust_mph	Ráfagas de viento en millas por hora	float64
wind_speed_mph	Velocidad del viento en millas por hora	float64
dew_point_f	Punto de rocío en grados Fahrenheit	float64

Tabla 3.1: Descripción de Variables Recolectadas

La estructura de este conjunto de datos permite un análisis multidimensional del rendimiento del SuDS:

1. **Temporalidad:** Las variables `date` y `time` permiten un análisis temporal detallado, facilitando la identificación de patrones diarios, estacionales o anuales en el funcionamiento del sistema.
2. **Funcionamiento del SuDS:** Las variables `input_cm` y `output_cm` proporcionan información directa sobre los niveles de agua en los puntos de entrada y salida del sistema, permitiendo evaluar su capacidad de retención y regulación del flujo de agua.
3. **Condiciones Meteorológicas:** Las variables restantes ofrecen un panorama completo de las condiciones atmosféricas, incluyendo:
 - Precipitación (`rainfall_in`): Factor crítico para evaluar la carga de agua que recibe el sistema.
 - Temperatura (`temperature_f`) y Humedad (`rh`): Influyen en la tasa de evaporación y en la dinámica del ciclo hidrológico local.

- Viento (`wind_dir_deg`, `wind_gust_mph`, `wind_speed_mph`): Puede afectar la distribución de la lluvia y la evaporación.
- Radiación Solar (`solar_rad_watm2`): Impacta en la evapotranspiración y el balance energético del sistema.
- Punto de Rocío (`dew_point_f`): Indicador de la humedad atmosférica y potencial de condensación.

Es importante notar que, aunque el conjunto de datos es extenso y detallado, pueden existir desafíos como datos faltantes o anomalías que requerirán atención durante el análisis exploratorio de datos. Además, la alta resolución temporal (registros por minuto) ofrece oportunidades para análisis detallados, pero también puede presentar desafíos en términos de manejo y procesamiento de datos debido a su volumen.

Esta estructura de datos proporciona una base sólida para una evaluación exhaustiva del SuDS, permitiendo insights valiosos sobre su eficacia en la gestión de aguas pluviales bajo diversas condiciones ambientales.

3.2. Análisis Exploratorio de Datos

El análisis exploratorio de datos es una etapa crucial en cualquier estudio científico, ya que permite comprender la naturaleza y la calidad de la información recopilada. En este capítulo, nos enfocamos en examinar detalladamente el conjunto de datos obtenido del Sistema de Humedal Artificial/Tanque de Reservorio (SHATR). Nuestro objetivo es identificar patrones, anomalías y características importantes que puedan influir en los análisis posteriores y en la interpretación de los resultados.

Comenzamos con un análisis exhaustivo de los datos faltantes, seguido de una evaluación de la distribución temporal de la información disponible. Esto nos lleva a tomar decisiones importantes sobre el ajuste del rango temporal de nuestro estudio y a adoptar un enfoque centrado en eventos de lluvia específicos. A lo largo de este capítulo, presentaremos visualizaciones y estadísticas descriptivas que nos ayudarán a comprender mejor la estructura y las particularidades de nuestro conjunto de datos, sentando así las bases para los análisis más detallados que se realizarán en las secciones subsiguientes.

3.2.1. Preprocesamiento Inicial y Exploración General de los Datos

Para mejorar el análisis temporal, se fusionaron las columnas de fecha y hora, creando una única variable temporal que sirve como índice del DataFrame. Este nuevo índice, de tipo `'datetime'`, abarca desde 2014-01-16 11:05:00 hasta 2016-06-07 14:34:00, proporcionando una estructura temporal coherente para el análisis.

Identificación de Datos Faltantes en Toda la Serie de Tiempo

Durante la consolidación de las bases de datos, se observó la existencia de períodos sin registro de datos. Para cuantificar estos intervalos, se generó un índice `'datetime'` completo dentro del rango temporal disponible. Esta operación expandió el DataFrame original de 1,162,660 registros a 1,257,330, revelando 94,670 registros sin datos.

La Tabla 3.2 presenta un resumen de los valores nulos por variable:

Variable	Valores Nulos
date	99,409
time	99,409
input_cm	147,756
output_cm	147,756
rainfall_in	605,124
solar_rad_watm2	605,163
rh	605,137
temperature_f	605,137
wind_dir_deg	605,137
wind_gust_mph	605,137
wind_speed_mph	605,137
dew_point_f	605,137

Tabla 3.2: Resumen de Valores Nulos por Variable

La alta cantidad de datos nulos se debe principalmente a la resolución temporal minuto a minuto de los registros, donde incluso breves períodos sin datos generan un volumen significativo de valores nulos.

Visualización de la Disponibilidad de Datos

Para obtener una visión más clara de la distribución temporal de los datos más relevantes para este estudio, se generó un mapa de calor (Figura 3.1) que ilustra la disponibilidad de datos para las tres variables más importantes: los niveles de agua en los puntos de entrada y salida, y la precipitación. En este mapa, cada etiqueta numérica representa los días del mes en los que se cuenta con información registrada. Además, cabe destacar que, en los periodos donde se dispone de datos de precipitación, también se cuenta con registros de las demás variables meteorológicas, ya que fueron recolectadas de manera simultánea.

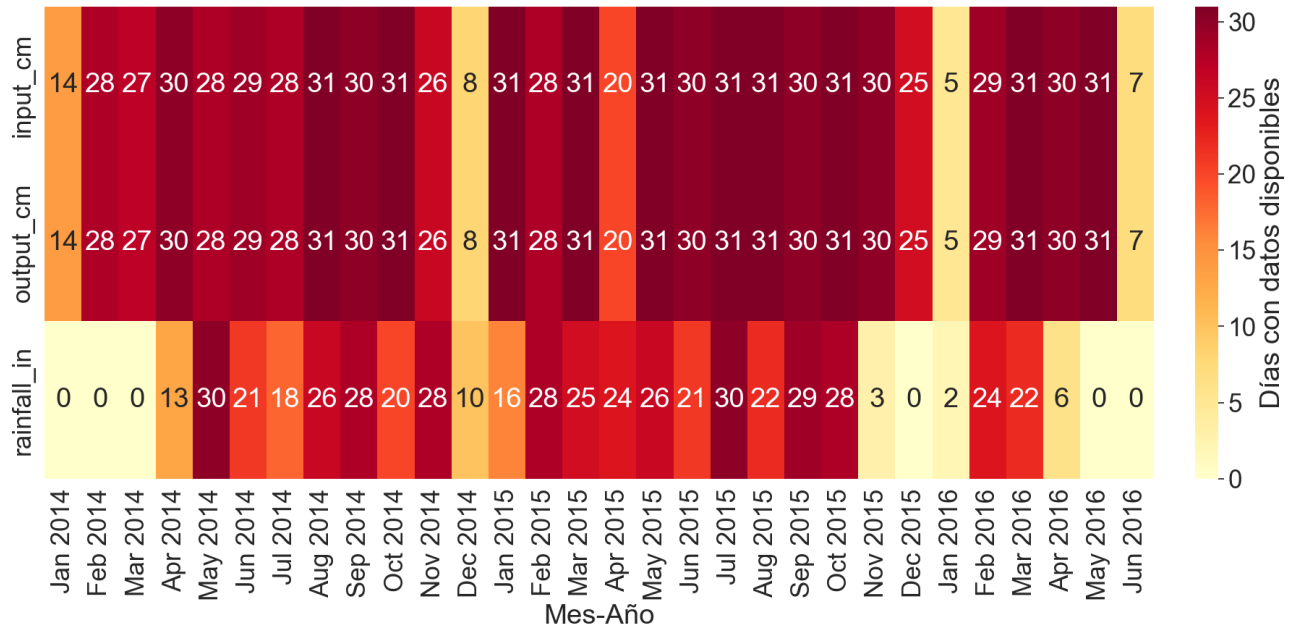


Figura 3.1: Mapa de Calor de Disponibilidad de Datos

El mapa de calor revela que los datos de niveles de agua están más completos en comparación con los datos meteorológicos. Sin embargo, la cantidad considerable de registros faltantes a lo largo de la serie temporal sugiere que intentar imputar todos los datos sería inviable y podría comprometer la calidad del análisis.

Enfoque en Eventos de Lluvia

Considerando la importancia del análisis durante los eventos de lluvia y la naturaleza de los datos faltantes, se tomó la decisión de enfocar el análisis en eventos de lluvia específicos en lugar de toda la serie temporal. Esta estrategia requiere:

- Definir criterios para identificar y delimitar eventos de lluvia.
- Crear una nueva base de datos centrada en estos eventos, utilizando las variables de niveles de agua y precipitación.
- Realizar el análisis exploratorio de datos sobre esta nueva base de datos de eventos.

En la siguientes sección, se procederá a definir los criterios para eventos de lluvia y se continuará con un análisis exploratorio de los datos resultantes.

3.2.2. Algoritmo y Resultados de la Identificación de Eventos de Lluvia

La evaluación precisa del rendimiento de un SuDS requiere una metodología robusta para identificar y caracterizar los eventos de lluvia y retención. Esta sección presenta un algoritmo desarrollado

específicamente para este propósito, detallando el sistema de medición empleado, la metodología de identificación de eventos y los criterios utilizados para su delimitación.

3.2.2.1. Algoritmo de Identificación de Eventos

El algoritmo desarrollado, denominado `identify_rain_retention_events`, se basa en una serie de parámetros clave y sigue un proceso lógico para identificar y caracterizar los eventos de lluvia y retención. A continuación, se detallan los componentes principales del algoritmo:

Parámetros de Control

- `rain_threshold`: Umbral mínimo de precipitación (0.01 pulgadas).
- `dry_period`: Período seco para delimitar el final de un evento (90 minutos).
- `output_tolerance`: Tolerancia en el nivel de salida para determinar el fin del evento (0.5 cm).
- `max_event_duration`: Duración máxima de un evento (2 días).

Proceso de Identificación

El algoritmo sigue una secuencia lógica de pasos para identificar y caracterizar los eventos:

1. Preprocesamiento de datos:

- Reemplazo de valores nulos en `rainfall_in` con ceros.
- Interpolación de valores faltantes en `output_cm` para períodos cortos (hasta 20 minutos).

2. Identificación de períodos de lluvia:

- Creación de una variable booleana `is_raining` basada en el umbral de lluvia.
- Agrupación de períodos de lluvia consecutivos mediante `rain_group`.

3. Detección de inicio y fin de eventos:

- Identificación del inicio de la lluvia (`rain_start`).
- Detección de posibles finales de eventos basados en el período seco (`potential_end`).

4. Evaluación de criterios adicionales:

- Cálculo de la duración del evento.
- Evaluación del retorno al nivel inicial de salida y tendencia del nivel del agua.
- Conteo de registros de lluvia y cálculo de lluvia acumulada.

5. Asignación de identificadores de evento:

- Creación de un identificador único (`event_id`) para cada evento detectado.
- Cálculo de la lluvia acumulada (`cumulative_rain`) para cada punto dentro del evento.

3.2.2.2. Criterios de Delimitación de Eventos

La delimitación precisa de los eventos de lluvia y retención es crucial para el análisis del rendimiento del SuDS. Los criterios establecidos se basan en consideraciones hidrológicas y las características específicas del sistema de medición:

Inicio del Evento

Un evento se considera iniciado cuando:

- Se detecta el primer registro de lluvia que supera el umbral establecido (`rain_threshold`).
- Para determinar el inicio real de cada evento, el algoritmo identifica el primer registro de lluvia y cuenta cuántos minutos con precipitación nula existen hasta encontrar el segundo registro de lluvia. Este conteo define el 'pre_rain_period'. El inicio del evento se establece retrocediendo ese mismo número de minutos antes del primer registro, lo que permite capturar las condiciones iniciales del sistema de manera específica para cada evento.
- Además, solo se consideran como eventos válidos aquellos en los que se registren al menos tres valores diferentes de cero en la variable de lluvia (`rainfall_in`). Esto garantiza la exclusión de lecturas erráticas o lluvias muy leves que podrían no representar eventos significativos para el análisis.

Fin del Evento

La determinación del fin de un evento es más compleja y considera múltiples factores:

- **Período Seco:** Se identifica un período seco continuo de `dry_period` minutos.
- **Evaluación del Nivel de Salida:** Se compara el nivel de salida actual con el nivel inicial:
 - El evento finaliza si la diferencia absoluta es menor o igual a `output_tolerance`.
- **Tendencia del Nivel del Agua:** Se calcula como:

$$\text{Tendencia} = \frac{\text{output_cm}_{\text{actual}} - \text{output_cm}_{\text{inicial}}}{\text{duración_evento}}$$

- Una tendencia positiva o casi plana ($> -0,0001$ cm/s) durante al menos una hora indica la estabilización del sistema.
- **Duración Máxima:** Se fuerza la finalización si se supera `max_event_duration`.

El evento se considera finalizado cuando se cumple el criterio del período seco Y (se cumple la condición del nivel de salida O la condición de la tendencia del nivel del agua), O cuando se alcanza la duración máxima.

3.2.2.3. Implementación y Resultados

El algoritmo se implementa en Python utilizando la biblioteca pandas para el manejo eficiente de series temporales. La función principal, `identify_rain_retention_events`, procesa el conjunto de datos y devuelve dos DataFrames:

- `df_events`: Contiene todos los registros asociados a eventos identificados.
- `events_summary`: Proporciona un resumen de las características de cada evento.

El código completo del algoritmo se encuentra disponible en el cuaderno de análisis de datos de Google Colab (Python), cuyo enlace está incluido en el Anexo 8.1.

Período de Estudio

El análisis abarca un período de aproximadamente 23 meses, desde el 16 de abril de 2014 hasta el 14 de marzo de 2016. Este extenso período de observación proporciona una visión representativa del comportamiento del sistema a lo largo de múltiples estaciones y ciclos anuales.

Resultados del Análisis

Tras aplicar el algoritmo al conjunto de datos, se identificaron un total de 166 eventos de lluvia y retención. A continuación, se presentan los resultados estadísticos clave:

Duración de los Eventos

- La duración media de los eventos fue de 214 minutos.
- El evento más corto duró 95 minutos.
- El evento más largo se extendió por 623 minutos.
- El 50% de los eventos (mediana) duró menos de 181 minutos.

Lluvia Acumulada por Evento (en pulgadas)

- La lluvia acumulada media por evento fue de 0.179 pulgadas (aproximadamente 4.55 mm).
- El evento con menor precipitación registró 0.03 pulgadas (0.76 mm).
- El evento más intenso acumuló 1.13 pulgadas (28.70 mm) de lluvia.
- El 50% de los eventos registró una precipitación de 0.09 pulgadas (2.29 mm) o menos.

Interpretación de Resultados

- **Variabilidad en la Duración:** Los eventos muestran una amplia variabilidad en su duración, desde poco más de hora y media, hasta más de 10 horas. Esta diversidad sugiere que el SuDS debe ser capaz de manejar tanto eventos cortos como prolongados.
- **Intensidad de Precipitación:** La mayoría de los eventos registrados son de baja a moderada intensidad, con el 75 % de los eventos acumulando 0.185 pulgadas (4.7 mm) o menos. Sin embargo, la presencia de eventos que alcanzan hasta 1.13 pulgadas (28.7 mm) indica que el sistema también debe estar preparado para manejar eventos de alta intensidad ocasionales.
- **Frecuencia de Eventos:** Con 131 eventos identificados en un período de aproximadamente 23 meses, se puede inferir una frecuencia promedio de 5.7 eventos por mes. Esta frecuencia relativamente alta subraya la importancia de un sistema de drenaje sostenible eficiente y bien mantenido.
- **Cobertura Temporal:** El período de estudio abarca casi dos años completos, lo que permite capturar variaciones estacionales y proporciona una base sólida para el análisis de patrones a largo plazo en el comportamiento del sistema.

Estos resultados proporcionan una base sólida para el análisis subsiguiente del rendimiento del SuDS, permitiendo una evaluación detallada de la respuesta del sistema a diferentes tipos y magnitudes de eventos de lluvia a lo largo de un período significativo.

3.2.3. Identificación y Estrategias para el Manejo de Datos Faltantes

Una vez definida la base de datos de eventos, se procedió a examinar los datos faltantes presentes en las diferentes variables. Este análisis se llevó a cabo de manera similar a la revisión inicial realizada durante la etapa de preparación de datos.

Variable	Valores Nulos
date	0
time	0
input_cm	2314
output_cm	2314
rainfall_in	0
solar_rad_watm2	0
rh	0
temperature_f	0
wind_dir_deg	0
wind_gust_mph	0
wind_speed_mph	0
dew_point_f	0

Tabla 3.3: Resumen de Valores Nulos por Variable en base de Eventos

La base de datos de eventos contiene un total de 28,098 registros, de los cuales 2,314 carecen de valores en las variables de niveles de agua (`input_cm` y `output_cm`). Esta ausencia de datos se debe a que existen casos en los que se registran datos meteorológicos, pero no se dispone de mediciones de los niveles de agua. Por ello, se realizó un análisis detallado de los datos faltantes para identificar los eventos afectados y la magnitud de las ausencias.

La Tabla 3.4 presenta los eventos con datos faltantes, indicando la cantidad de valores ausentes en las variables mencionadas:

<code>event_id</code>	<code>input_cm</code>	<code>output_cm</code>
15	222	222
20	205	205
21	42	42
61	211	211
62	102	102
63	138	138
64	143	143
65	218	218
66	218	218
67	162	162
68	129	129
69	348	348
70	176	176

Tabla 3.4: Datos Faltantes por Evento

En total, se identificaron 13 eventos con datos faltantes en las variables de niveles de agua. Debido a que estas variables representan el objetivo principal del análisis y considerando que los registros incompletos incluyen más de 42 minutos consecutivos sin datos, se decidió excluir dichos eventos del análisis.

Como resultado, la base de datos de eventos quedó conformada por 118 eventos completos, libres de datos faltantes, lo que permite continuar con el análisis sin limitaciones derivadas de información ausente.

3.2.4. Detección y Tratamiento de Valores Atípicos en los Datos

En este análisis de datos atípicos, se examinaron las variables de forma individual, comenzando con la elaboración de diagramas de caja y gráficos de barras de frecuencia para cada una, lo que permitió visualizar su distribución. Posteriormente, cada variable fue representada en función del tiempo a lo largo de la duración de cada evento, utilizando un color distinto para identificar cada uno. Esto facilitó una comprensión más detallada de los datos atípicos, tanto a nivel de variable como del evento al que están asociados. A continuación, se presentan los gráficos generados junto con su respectivo análisis.

3.2.4.1. Análisis Gráfico de la Distribución y Variación de las Variables por Evento

Nivel de entrada (input_cm)

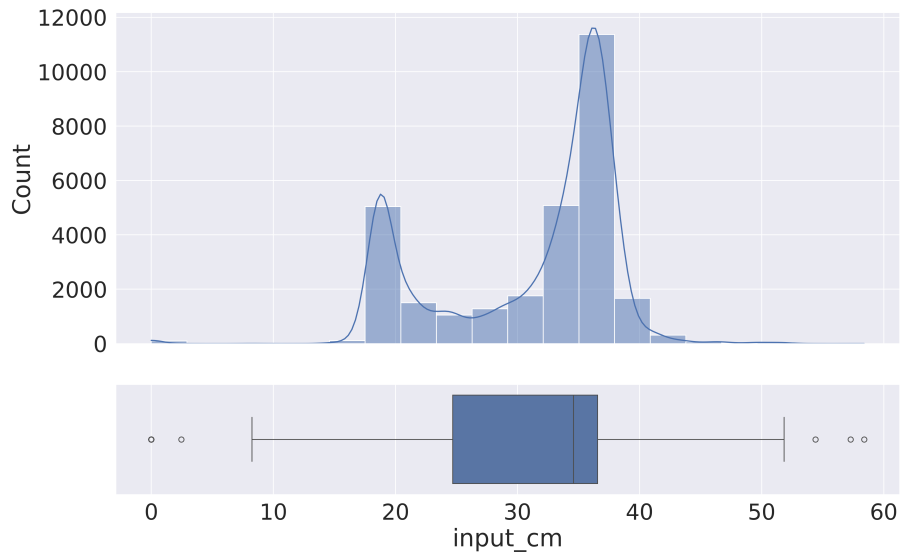


Figura 3.2: Distribución de los niveles de entrada al SuDS (input_cm)

El gráfico de distribución de la variable de entrada muestra que la mayoría de los valores se encuentran entre 10 y 60 centímetros, un rango acorde con lo esperado. No obstante, se identifican algunos valores atípicos igual a cero, los cuales resultan improbables a menos que el sensor haya estado obstruido.

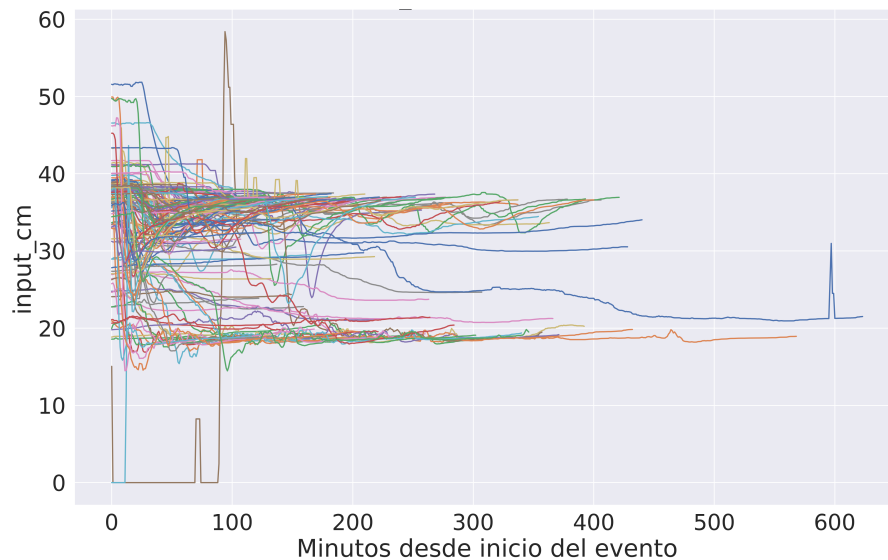


Figura 3.3: Variación de los Niveles de Entrada al SuDS (input_cm) durante los Eventos

Asimismo, el gráfico 3.3 evidencia que ciertos eventos presentan valores claramente anómalos, como aquellos que permanecen a cero o alcanzan distancias cercanas a los 60 centímetros. En particular, los eventos 16 y 115 fueron descartados del análisis debido a su comportamiento atípico, evitando así que comprometan la precisión de los resultados.

Una vez excluidos estos eventos, cuyo registro contenía principalmente valores anómalos, se obtuvo un rango ajustado de datos entre 15.40 cm y 45.92 cm. Este intervalo es coherente con lo esperado, lo que permite avanzar con el análisis de la siguiente variable.

Nivel de salida (output_cm)

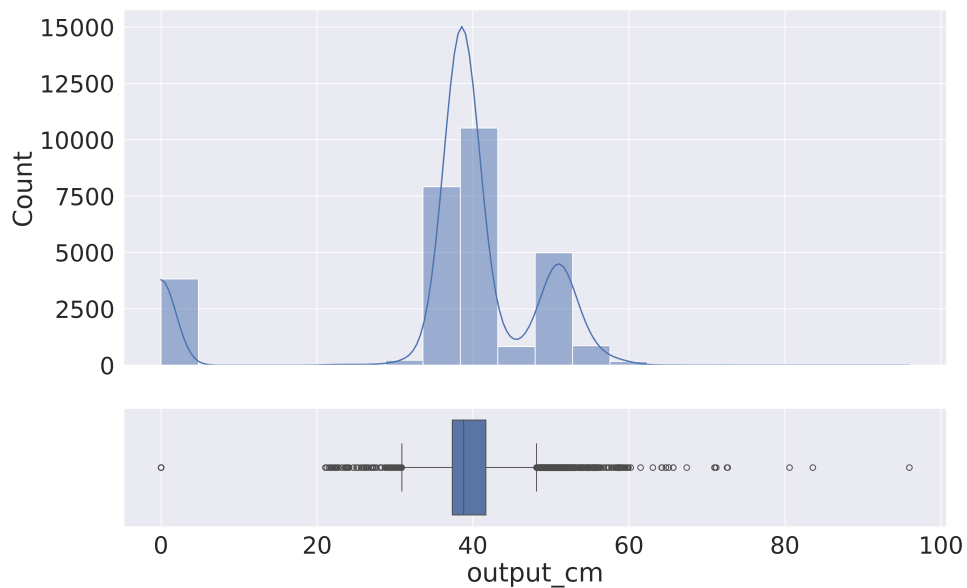


Figura 3.4: Distribución de los niveles de salida al SuDS (output_cm)

El gráfico de distribución de los niveles de salida muestra que la mayoría de los valores se encuentran entre 20 y 70 centímetros, lo cual resulta coherente con las expectativas. Sin embargo, se observan algunos valores atípicos que superan los 100 centímetros, así como una notable acumulación de valores iguales a cero. Esto último podría indicar un mal funcionamiento del sensor o una posible obstrucción. Este aspecto debe tratarse para garantizar la calidad de los datos.

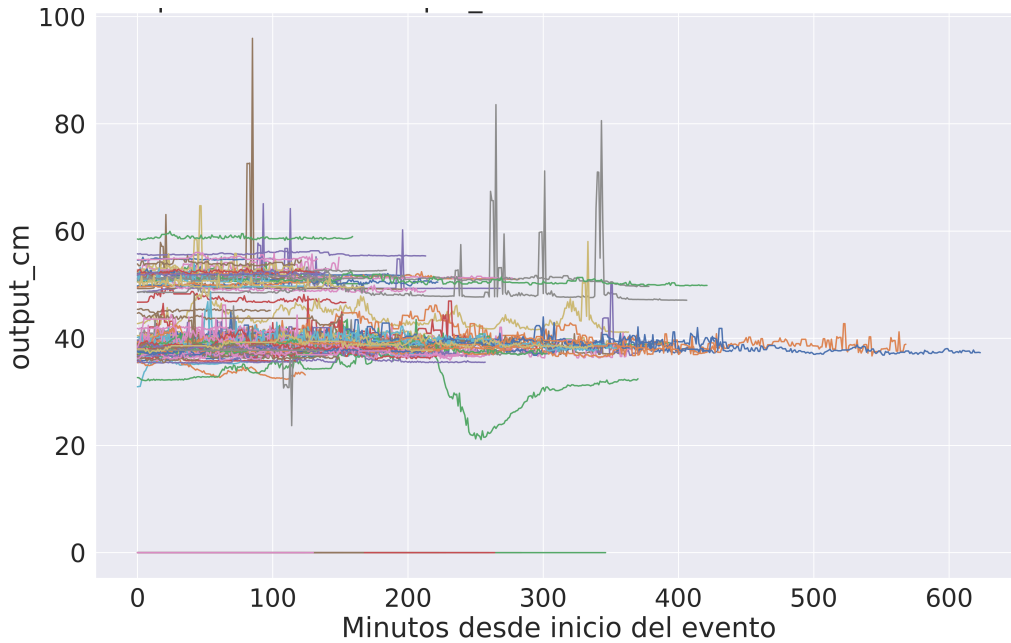


Figura 3.5: Variación de los Niveles de Salida al SuDS (output_cm) durante los Eventos

Por otro lado, el gráfico de variación durante los eventos revela una línea plana en cero en varios eventos, lo que sugiere que el nivel de salida estuvo obstruido o presentó fallos en la medición durante esos períodos. Los eventos afectados, identificados como los números 17, 18, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129 y 130, no muestran valores distintos de cero. Debido a la importancia de esta variable para la construcción de la variable objetivo, junto con input_cm, y considerando que no es posible realizar una imputación válida para la duración total de estos eventos, se ha decidido excluirlos del análisis.

Como resultado, el rango de los valores de output_cm después de eliminar estos eventos está entre un mínimo de 21.07 cm y un máximo de 95.95 cm. En consecuencia, de los 118 eventos inicialmente disponibles, el número de eventos considerados para el análisis se reduce a 99 tras excluir aquellos con valores cero en el nivel de salida.

Lluvia (rainfall_in)

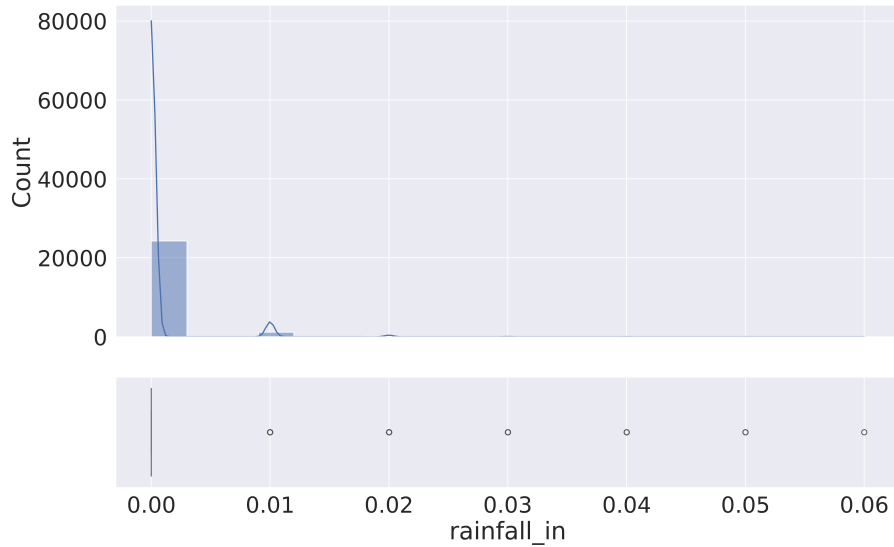


Figura 3.6: Distribución de la variable lluvia (rainfall_in)

La distribución de la variable de lluvia presenta valores que oscilan entre 0 y 0.06 pulgadas, con una concentración predominante cerca de cero. Esta distribución es consistente con el método de medición del dispositivo, que registra valores únicamente cuando el recipiente se llena y se voltea. En general, los datos muestran una estructura adecuada para su análisis. Además, el comportamiento de la lluvia durante los eventos no revela datos claramente anómalos, lo que permite continuar con el análisis sin alteraciones.

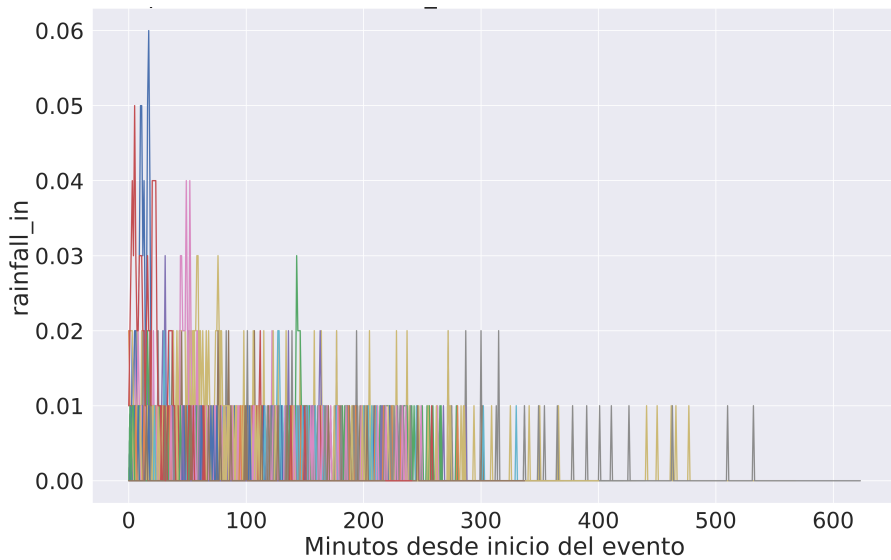
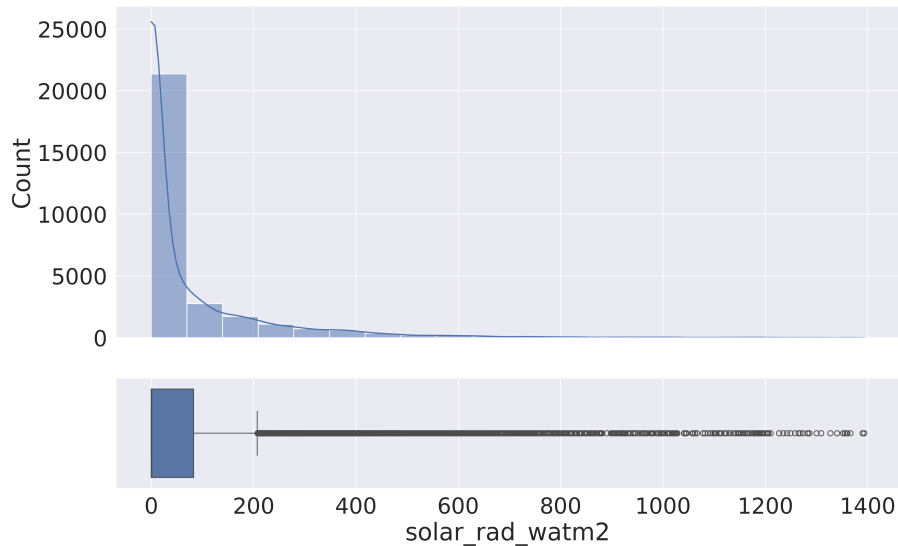
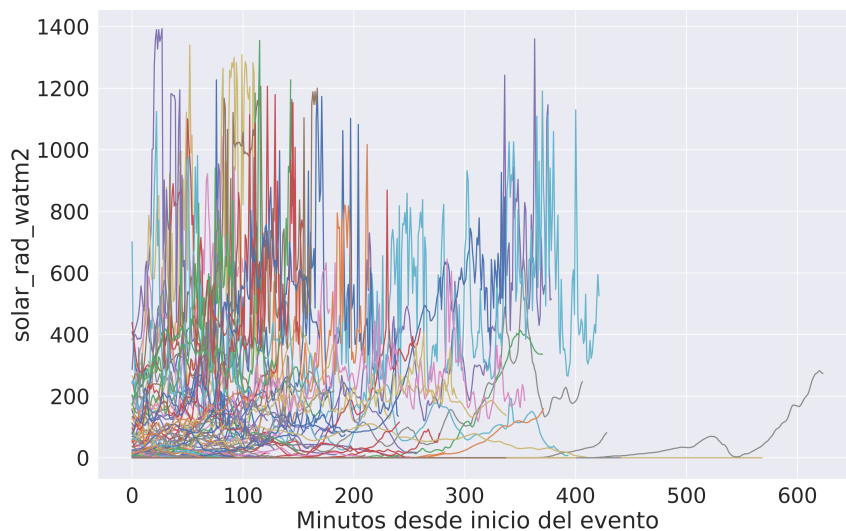


Figura 3.7: Variación de la Lluvia (rainfall_in) durante los Eventos

Radiación Solar (`solar_rad_watm2`)Figura 3.8: Distribución de la variable Radiación Solar (`solar_rad_watm2`)

El gráfico de distribución de la radiación solar muestra que la mayoría de los datos se encuentran entre 0 y 210 W/m². Esto puede explicarse por la ocurrencia de varios eventos durante las horas de la tarde y noche, cuando la radiación solar es naturalmente más baja. Además, dado que estos eventos corresponden a lluvias, no se espera que el cielo esté despejado, lo que también contribuye a la baja radiación. Por otro lado, se observan algunos valores extremos cercanos a 1400 W/m², aunque estos no exceden los límites esperados de radiación para condiciones de cielo despejado en la zona andina de Colombia ([22]).

Figura 3.9: Variación de la Radiación Solar (`solar_rad_watm2`) durante los Eventos

El gráfico 3.9 muestra diferencias notables, probablemente debido a las mediciones realizadas en distintos años y épocas del año. Se observó que algunos eventos tienen valores casi planos, cercanos a cero, lo que sugiere la necesidad de revisarlos con más detalle más adelante.

Humedad Relativa (rh)

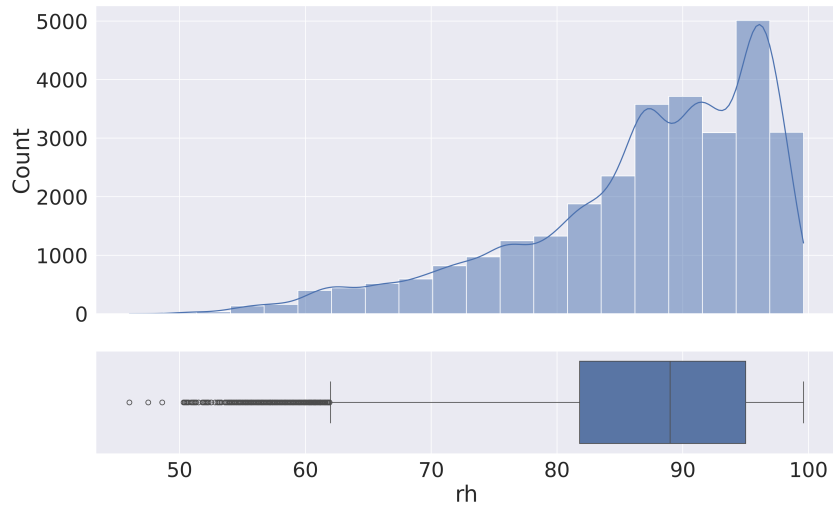


Figura 3.10: Distribución de la variable Humedad Relativa (rh)

El gráfico de distribución de la humedad relativa muestra una concentración predominante de valores entre el 80 % y el 100 %, lo cual es coherente con las condiciones de lluvia durante los momentos de medición. No se observan valores fuera de lo esperado en esta variable, lo que indica que los datos son consistentes con el comportamiento típico de la humedad relativa en contextos de precipitación.

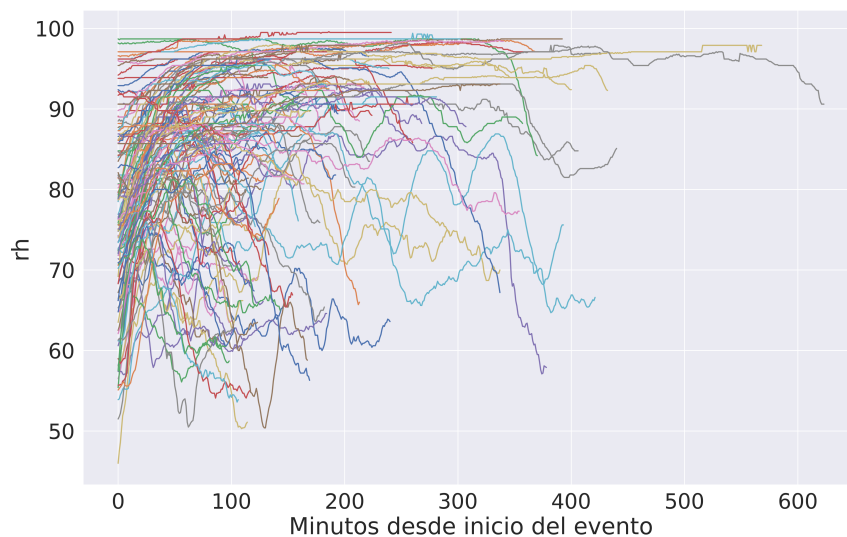


Figura 3.11: Variación de la Humedad Relativa (rh) durante los Eventos

El gráfico 3.11 muestra una alta variabilidad. En la mayoría de los casos, los valores aumentan rápidamente hasta alcanzar niveles cercanos al 100 % y, posteriormente, comienzan a descender. Este patrón puede explicarse por el desarrollo típico de los eventos de lluvia, donde la humedad relativa alcanza su punto máximo debido a la saturación del aire durante las precipitaciones, seguido de un descenso gradual una vez que cesa la lluvia y se inicia el proceso de evaporación.

Temperatura (temperature_f)

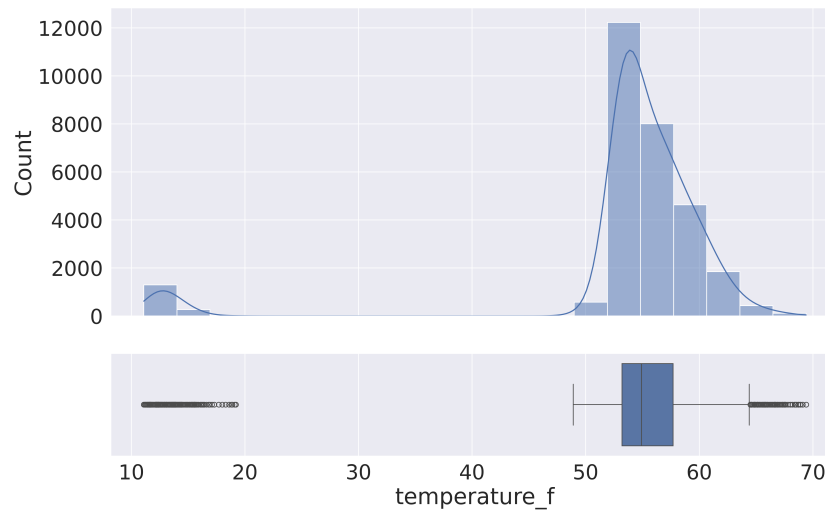


Figura 3.12: Distribución de la Temperatura (temperature_f)

El gráfico de temperatura muestra algunos valores que no se encuentran dentro del rango esperado en grados Fahrenheit para Bogotá, que típicamente supera los 41 °F (5 °C).

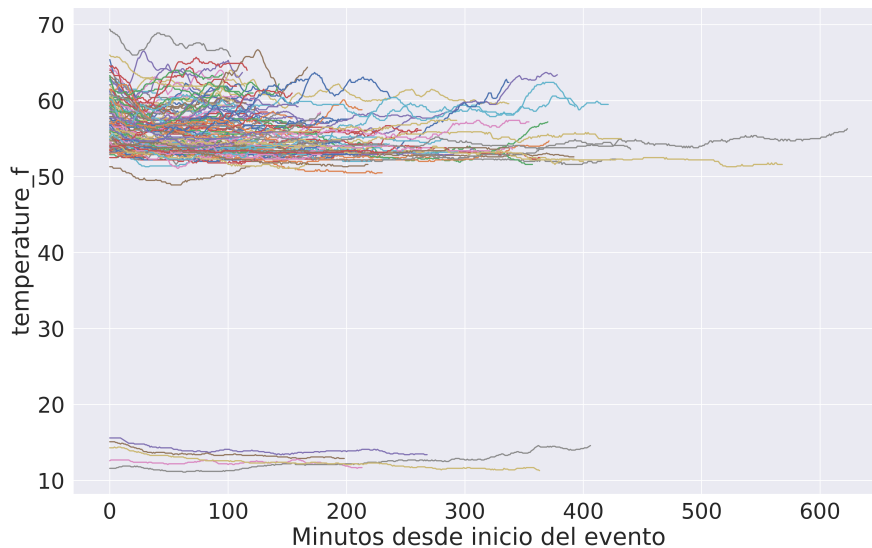


Figura 3.13: Variación de la Temperatura (temperature_f) durante los Eventos

Al analizar los eventos 4, 5, 6, 7 y 8, cuyos valores se encuentran fuera de este rango, se observó que su comportamiento es consistente con temperaturas registradas en grados Celsius. Tras una revisión exhaustiva de las mediciones y la base de datos original, y en consulta con los investigadores responsables de la recopilación de los datos, se determinó que estos valores estaban expresados en grados Celsius. Por ello, se realizó la conversión correspondiente a grados Fahrenheit para asegurar la coherencia del análisis.

Dirección del Viento (wind_dir_deg)

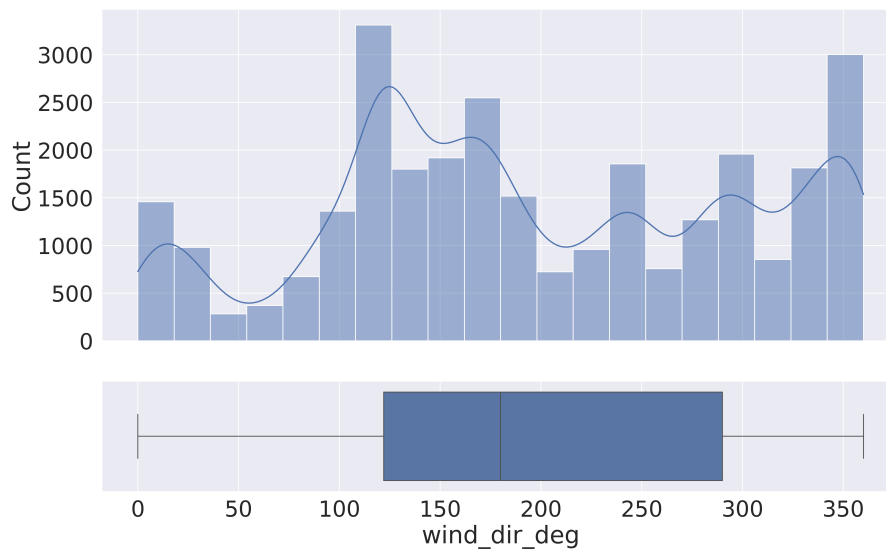


Figura 3.14: Distribución de la Dirección del Viento (wind_dir_deg)

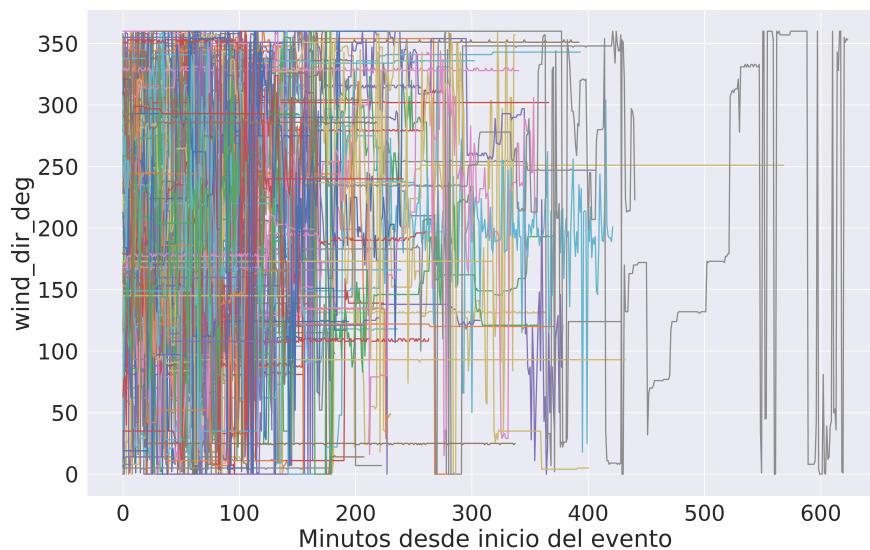


Figura 3.15: Variación de la Dirección del Viento (wind_dir_deg) durante los Eventos

Los valores de la dirección del viento se encuentran dentro del rango esperado de 0 a 360 grados, lo cual es coherente con esta variable. Además, se observa que, durante la mayor parte del tiempo, la dirección del viento se mantiene predominantemente entre 120 y 300 grados. Por otro lado, los datos de dirección del viento presentan una alta variabilidad, lo cual es característico de esta variable. Sin embargo, no se identifican valores atípicos evidentes, por lo que se procederá a continuar con el análisis.

Ráfagas de viento (wind_gust_mph)

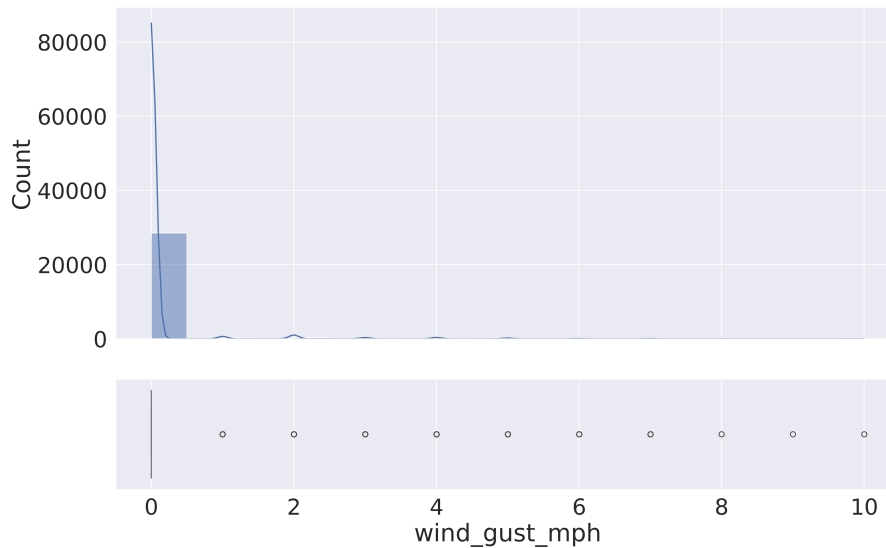


Figura 3.16: Distribución de las Ráfagas de Viento (wind_gust_mph)

La distribución de las ráfagas de viento muestra valores que oscilan entre 0 y 10 millas por hora, un rango que se considera normal para las condiciones climáticas de Bogotá.

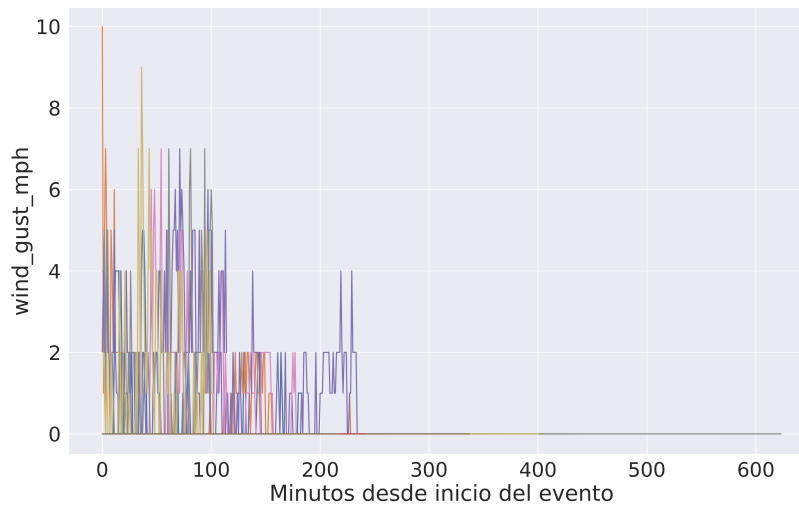


Figura 3.17: Variación de las Ráfagas de Viento (wind_gust_mph) durante los Eventos

Por otro lado, el gráfico 3.17 sugiere que las ráfagas tienden a ser más intensas al inicio de los eventos. Durante el desarrollo de un evento de lluvia, es común que las ráfagas de viento se intensifiquen debido a los cambios en la presión y temperatura asociados con la aproximación de sistemas nubosos.

Velocidad del Viento (wind_speed_mph)

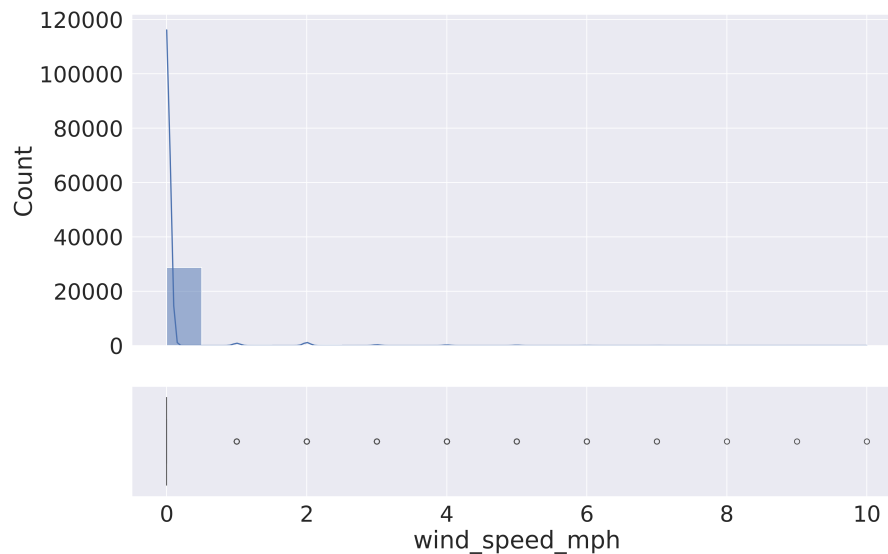


Figura 3.18: Distribución de la Velocidad del Viento (wind_speed_mph)

La gráfica de distribución de la velocidad del viento muestra un rango y comportamiento similar al de las ráfagas de viento. Dado que los datos de velocidad del viento se registran a intervalos de un minuto, esta similitud puede deberse a la resolución temporal de la medición.

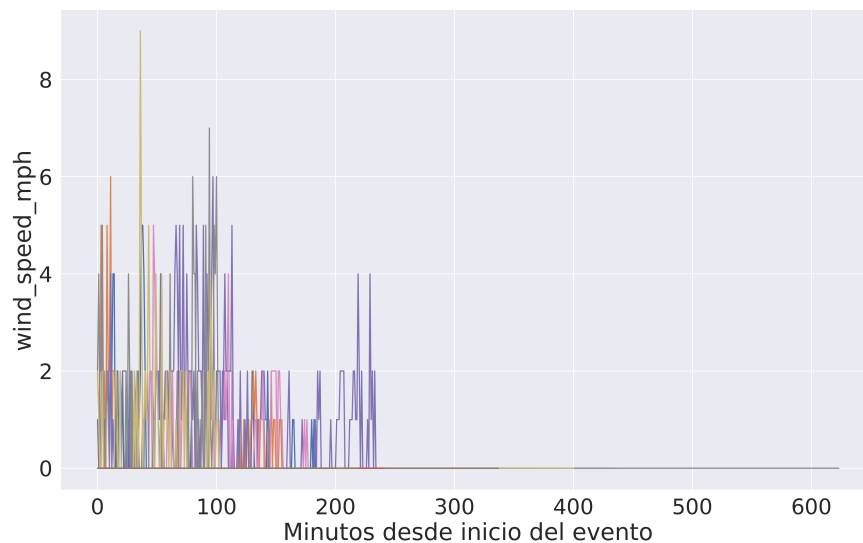


Figura 3.19: Variación de la Velocidad del Viento (wind_speed_mph) durante los Eventos

Por otro lado, el gráfico 3.19 confirma la similitud entre ambas gráficas, lo que sugiere una alta correlación entre las variables de velocidad promedio y ráfagas de viento. Esta relación será verificada más adelante mediante un análisis de correlación.

Punto de Rocio (dew_point_f)

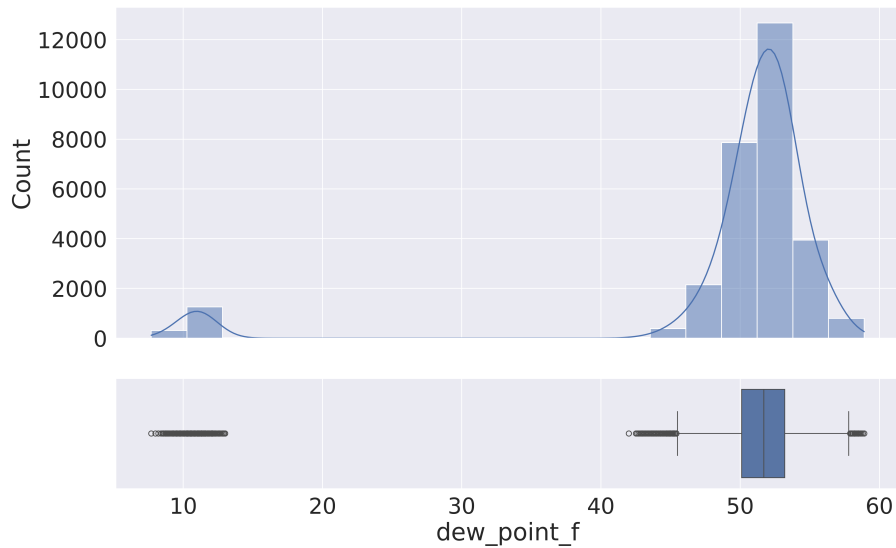


Figura 3.20: Distribución del Punto de Rocio (dew_point_f)

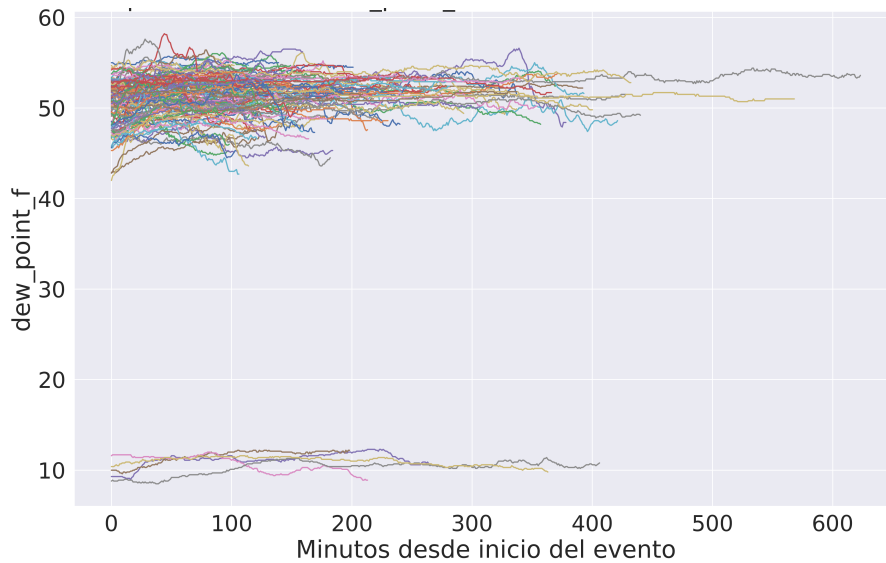


Figura 3.21: Variación del Punto de Rocio (dew_point_f) durante los Eventos

Esta variable presenta un inconveniente similar al de la temperatura, ya que algunos de los valores se encuentran en el rango de grados Celsius en lugar de grados Fahrenheit. Los eventos que presentan

estos valores atípicos coinciden exactamente con los de la variable de temperatura. Por lo tanto, para asegurar la coherencia en el análisis, se procedió a convertir los valores de grados Celsius a Fahrenheit, siguiendo el mismo procedimiento aplicado previamente en la variable de temperatura.

Lluvia Acumulada (cumulative_rain)

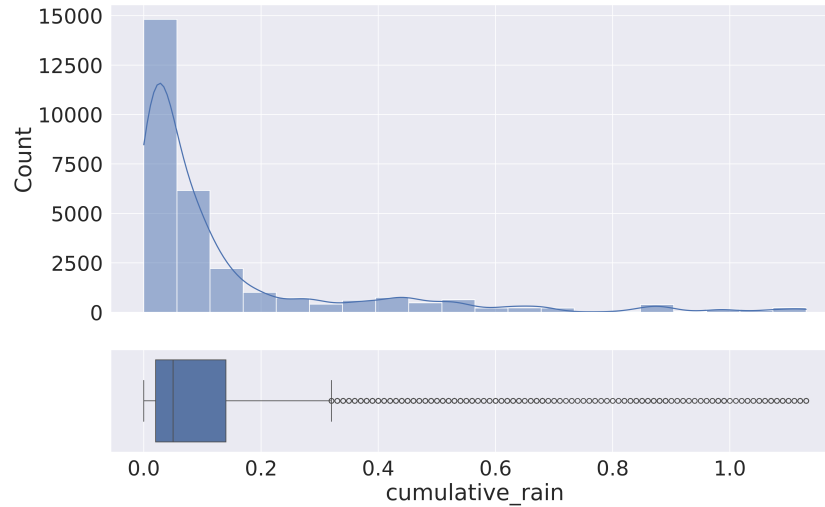


Figura 3.22: Distribución de la Lluvia Acumulada (cumulative_rain)

Esta gráfica de distribución muestra la frecuencia de los diferentes valores de lluvia acumulada. La forma exponencial decreciente indica que los valores más bajos de lluvia acumulada son los más frecuentes, mientras que los valores más altos son menos comunes.

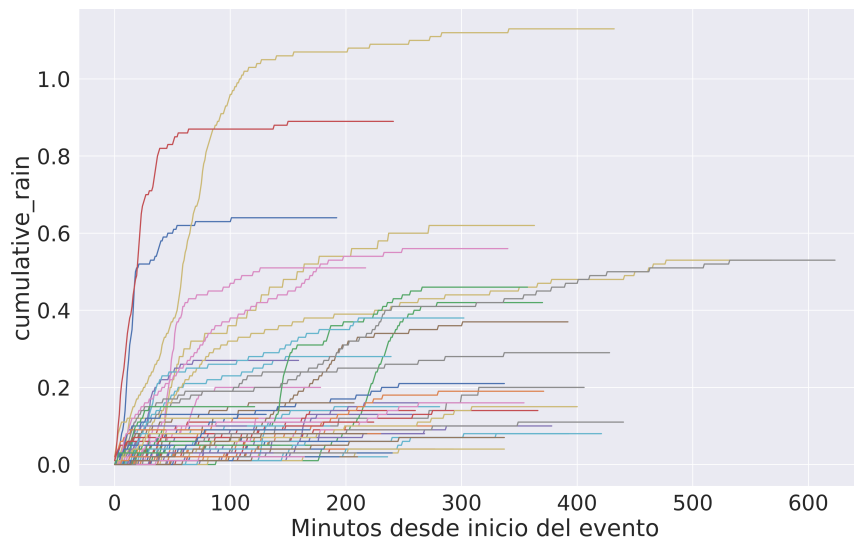


Figura 3.23: Variación de la Lluvia Acumulada (cumulative_rain) durante los Eventos

Por otro lado, la gráfica de comportamiento refleja cómo se acumula la lluvia a lo largo de cada evento. Algunas curvas muestran un patrón gradual y suave en la acumulación de lluvia, mientras que otras presentan cambios más abruptos. No se observan datos atípicos en esta variable, lo que sugiere que los datos son consistentes y adecuados para el análisis.

3.2.4.2. Estadísticas descriptivas de las variables

Tras la eliminación de los eventos con datos atípicos y la corrección del problema de unidades en las variables de temperatura y punto de rocío, se obtuvo un total de 99 eventos, correspondientes a 22,000 registros. A continuación, se presentan las principales estadísticas descriptivas de la base de datos, desglosadas para cada una de las variables analizadas.

Variable	conteo	media	desv. std	min	25 %	50 %	75 %	max
input_cm	22000	30.593	7.078	15.403	23.850	33.929	36.130	45.919
output_cm	22000	41.650	5.922	21.072	37.932	39.053	44.755	95.949
rainfall_in	22000	0.001	0.003	0	0	0	0	0.06
solar_rad_watm2	22000	73.946	146.435	0	0	0	84	1393
rh	22000	87.997	9.334	50.4	83.9	90.5	95.5	99.6
temperature_f	22000	55.414	2.736	50.4	53.4	54.5	57.0	66.7
wind_dir_deg	22000	202.099	101.706	0	122	189	293	360
wind_gust_mph	22000	0.034	0.313	0	0	0	0	7
wind_speed_mph	22000	0.021	0.228	0	0	0	0	6
dew_point_f	22000	51.663	1.793	44.5	50.5	51.8	53.0	58.2
cumulative_rain	22000	0.141	0.201	0	0.03	0.07	0.15	1.13

Tabla 3.5: Estadísticas descriptivas de las variables

La tabla presenta las estadísticas descriptivas de las variables registradas, mostrando que todas se encuentran dentro de los rangos esperados para eventos de lluvia en Bogotá. Variables como los niveles de agua (`input_cm` y `output_cm`) reflejan dinámicas consistentes con un sistema de almacenamiento y descarga, mientras que la humedad relativa (`rh`) y el punto de rocío (`dew_point_f`) confirman condiciones típicas de alta humedad en la región.

La temperatura (`temperature_f`) permanece estable dentro de un rango característico de la región. Las ráfagas de viento (`wind_gust_mph`) y la velocidad del viento (`wind_speed_mph`) presentan valores bajos, como es común en lluvias moderadas sin vientos extremos. La lluvia acumulada (`cumulative_rain`) y las precipitaciones por minuto (`rainfall_in`) indican eventos predominantemente ligeros, con algunos casos más intensos.

Sin embargo, se identificaron variables con desviación estándar muy baja y ceros en la mayoría de los cuartiles y mínimos, incluyendo `rainfall_in`, (`solar_rad_watm2`), `wind_gust_mph`, `wind_speed_mph` y `cumulative_rain`. Debido a esto, se realizó un análisis detallado de la desviación estándar por evento para cada una de las variables, obteniendo una tabla como la siguiente.

Este análisis reveló que el evento 113 tiene una varianza de cero en la variable `output_cm`. Al examinarlo en detalle, se encontró que el sensor asociado a esta medición permaneció en un valor constante

event_id	input_cm	output_cm	rainfall_in	solar_rad_watm2	wind_gust_mph	wind_speed_mph	cumulative_rain
106	0.725	0.660	0.001	0.0	0.0	0.0	0.026
107	0.522	0.537	0.001	7.600	0.0	0.0	0.022
108	2.328	0.818	0.002	0.0	0.0	0.0	0.022
109	1.502	0.476	0.002	0.0	0.0	0.0	0.012
110	1.972	0.792	0.002	211.012	0.0	0.0	0.034
111	0.842	1.327	0.002	68.984	0.0	0.0	0.004
112	1.268	2.178	0.009	24.780	0.0	0.0	0.175
113	0.222	0.0	0.001	0.324	0.0	0.0	0.011
114	1.274	0.723	0.002	0.0	0.0	0.0	0.029

Tabla 3.6: Desviación estándar de las variables por evento

durante todo el evento, lo que llevó a la decisión de eliminar dicho evento para evitar distorsiones en el análisis.

En cuanto a las variables *rainfall_in* y *cumulative_rain*, aunque presentan desviaciones estándar cercanas a cero, estos valores no son nulos. Esto valida su relevancia, ya que representan mediciones reales, incluso si las magnitudes de variación son bajas. Por otro lado, para las variables *wind_gust_mph* y *wind_speed_mph*, se observó que en su mayoría contienen valores de cero, lo que indica una falta de variabilidad significativa. Estas variables no aportan información diferenciadora y probablemente no serán útiles para los modelos, aunque la decisión de eliminarlas o no se tomará en el análisis de correlación.

Respecto a la variable *solar_rad_watm2*, se identificaron 29 eventos en los que su valor es cero. Esto coincide con lo observado en las gráficas de distribución, donde se nota una ausencia de mediciones válidas en una cantidad considerable de los datos. Este comportamiento sugiere que el medidor de radiación solar presentó fallas durante estos eventos, lo que compromete la confiabilidad de la variable. Debido a esto, se decidió no tener en cuenta *solar_rad_watm2* en los siguientes análisis y modelaciones, ya que su inclusión podría introducir sesgos o errores en los resultados.

Finalmente, los índices de los eventos se resetearon, numerándolos de 0 a 97, para un total de 98 eventos después de la eliminación del evento 113. Este procedimiento facilita la organización y el seguimiento de los datos en los análisis posteriores.

3.2.5. Análisis de las Relaciones entre las Variables del Sistema

El análisis de correlaciones constituye una herramienta esencial para explorar las interrelaciones entre las variables del sistema. Para este estudio, se ha optado por utilizar exclusivamente el coeficiente de correlación de Spearman, debido a que no requiere el cumplimiento del supuesto de normalidad en los datos y es más robusto frente a valores atípicos que pueden distorsionar los resultados.

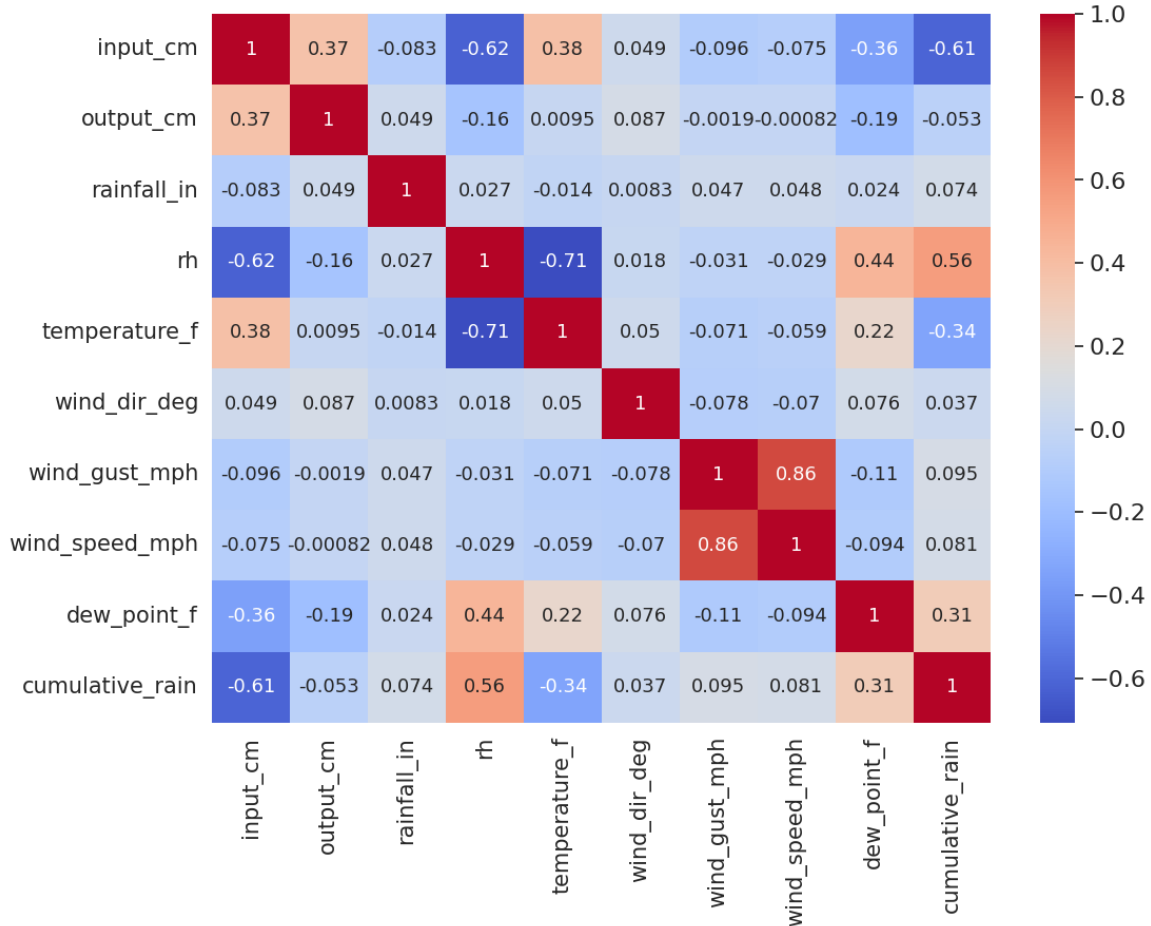


Figura 3.24: Matriz de correlación de Spearman.

El coeficiente de correlación de Spearman evalúa la fuerza y dirección de la relación monótona entre dos variables, resultando particularmente adecuado para este conjunto de datos donde no se puede garantizar la linealidad perfecta de las relaciones ni la distribución normal de todas las variables. Como se observa en la Figura 3.24, este análisis revela patrones significativos en el comportamiento del sistema.

Los niveles de agua, representados por *input_cm* y *output_cm*, muestran una correlación moderada entre sí (0.37), lo que indica que, si bien existe una relación entre ambas mediciones, esta no es perfectamente lineal debido probablemente a la influencia de otros factores como la capacidad de almacenamiento y las características propias del sistema de drenaje.

Respecto a las variables meteorológicas, se observa una fuerte correlación negativa (-0.61) entre la lluvia acumulada (*cumulative_rain*) y el nivel de agua de entrada (*input_cm*). Esta relación inversa refleja que a mayor acumulación de precipitación corresponde un aumento en la lámina de agua (lo que se traduce en valores menores en el sensor). Asimismo, la humedad relativa (*rh*) muestra una correlación negativa considerable (-0.62) con *input_cm*, sugiriendo que condiciones de mayor humedad ambiental están asociadas con niveles más altos de agua en el sistema.

Es interesante notar la fuerte correlación negativa (-0.71) entre la temperatura (*temperature_f*) y la humedad relativa (*rh*), lo cual es coherente con los principios físicos de la meteorología, donde mayor temperatura tiende a reducir la humedad relativa del aire. Sin embargo, la temperatura muestra una correlación moderada positiva (0.38) con *input_cm*, lo que podría estar relacionado con efectos de evaporación o con patrones estacionales que afectan simultáneamente ambas variables.

Las variables relacionadas con el viento (*wind_dir_deg*, *wind_gust_mph* y *wind_speed_mph*) presentan correlaciones muy débiles con los niveles de agua, con valores cercanos a cero. Esto sugiere que, en este sistema específico, el viento no ejerce una influencia significativa sobre la dinámica hídrica. El único patrón relevante asociado al viento es la alta correlación (0.86) entre la velocidad del viento y las ráfagas, lo cual era esperable dada la naturaleza física de estos fenómenos.

Un hallazgo adicional es la correlación moderada (0.44) entre el punto de rocío (*dew_point_f*) y la humedad relativa, así como la correlación negativa moderada (-0.36) entre el punto de rocío y *input_cm*. Estas relaciones son consistentes con los principios termodinámicos de la atmósfera y su interacción con cuerpos de agua.

Con base en estos resultados, se han tomado las siguientes decisiones para optimizar el modelo:

- Exclusión de las variables de viento (*wind_dir_deg*, *wind_gust_mph* y *wind_speed_mph*) de los análisis posteriores, debido a su mínima influencia sobre los niveles de agua.
- Eliminación de la variable *temperature_f* para evitar redundancia con la humedad relativa, dado que muestran una fuerte correlación negativa y la humedad presenta correlaciones más intensas con los niveles de agua.
- Conservación de *cumulative_rain* como variable predictora clave, dada su fuerte correlación con los niveles de agua de entrada y su relevancia física en el sistema.
- Inclusión de *rh* como variable predictora significativa, considerando su destacada correlación con los niveles de agua.

Estas decisiones buscan reducir la dimensionalidad del modelo sin sacrificar información relevante, lo que permite una interpretación más clara de las dinámicas del sistema y una mayor eficiencia computacional en los análisis subsecuentes. El enfoque no paramétrico adoptado mediante la correlación de Spearman proporciona resultados más robustos y fiables para este conjunto de datos, donde no se puede asumir el cumplimiento estricto de supuestos como la normalidad o la linealidad perfecta en las relaciones entre variables.

3.2.6. Análisis de Componentes Principales

El análisis de componentes principales (PCA) es una técnica estadística que transforma las variables originales en nuevas variables no correlacionadas llamadas componentes principales. Si bien la reducción de dimensionalidad es una de sus aplicaciones más comunes, especialmente útil cuando se trabaja con grandes conjuntos de variables para simplificar el procesamiento y la complejidad de los modelos, en este estudio se aplica principalmente para identificar las variables meteorológicas predictivas que mejor explican la varianza del sistema. Este análisis permite no solo evaluar la posible eliminación de variables

con menor contribución, sino también identificar patrones subyacentes y analizar las correlaciones entre las distintas variables meteorológicas, proporcionando una comprensión más profunda de cómo estas interactúan en el sistema climático. A continuación, se presenta la proporción de varianza explicada por cada uno de los componentes principales:

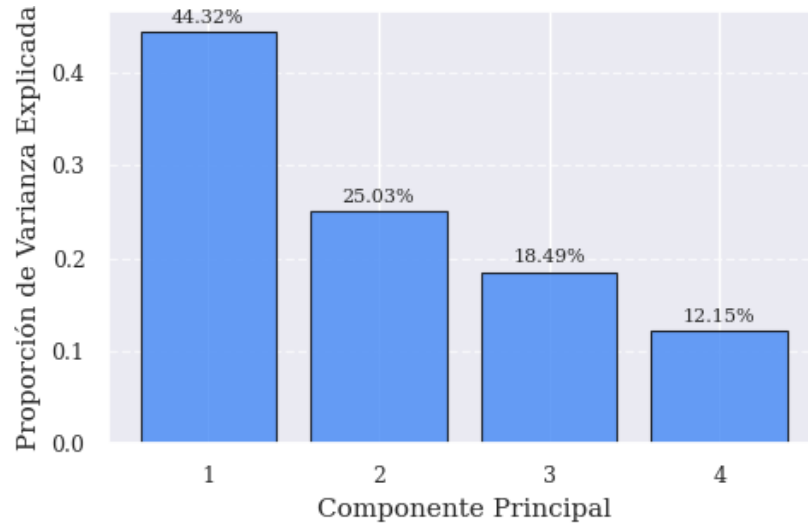


Figura 3.25: Proporción de Varianza Explicada por cada Componente Principal

Los resultados muestran que los cuatro componentes principales capturan el 100 % de la variabilidad total en los datos, distribuidos de la siguiente manera: PC1 (44.32 %), PC2 (25.03 %), PC3 (18.49 %) y PC4 (12.15 %). Esta distribución indica una dominancia clara del primer componente, que explica casi la mitad de la varianza total del sistema. Seguidamente, se presentan las cargas de cada variable sobre los componentes principales:

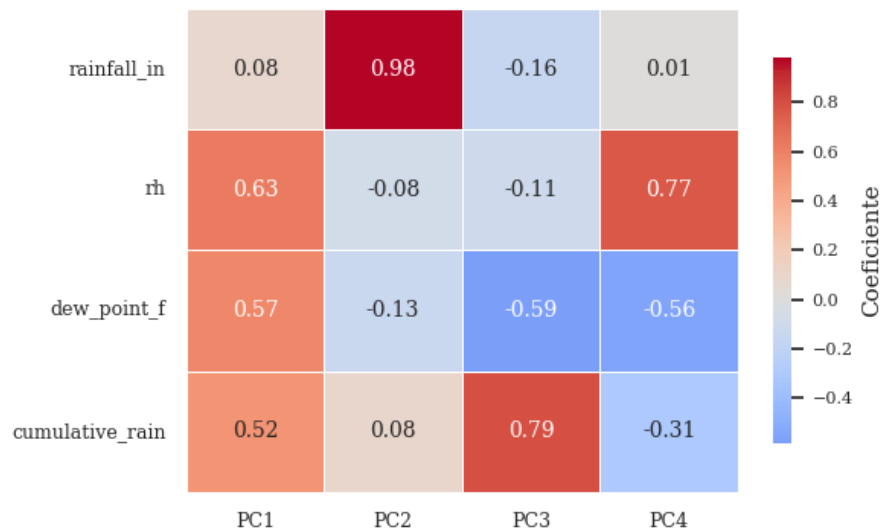


Figura 3.26: Cargas de los Componentes Principales

Interpretación de los Componentes Principales

El análisis de las cargas (Figura 3.26) revela patrones significativos:

- **PC1 (44.32 %)**: Este componente muestra una estructura equilibrada entre tres variables principales: humedad relativa (0.63), punto de rocío (0.57) y lluvia acumulada (0.52), con contribuciones positivas similares. La lluvia instantánea tiene una contribución menor (0.08). Este patrón sugiere que el primer componente representa las condiciones generales de humedad atmosférica y su persistencia.
- **PC2 (25.03 %)**: Se caracteriza por una carga dominante y casi exclusiva de la lluvia instantánea (0.98), mientras que las demás variables muestran cargas cercanas a cero (-0.08 a 0.08). Esto indica que este componente captura específicamente los eventos de precipitación instantánea, independientemente de las condiciones ambientales generales.
- **PC3 (18.49 %)**: Presenta un contraste interesante entre la lluvia acumulada (0.79) y el punto de rocío (-0.59), con contribuciones menores de las otras variables. Este componente podría estar capturando situaciones donde la acumulación de lluvia ocurre bajo condiciones de cambio en el punto de rocío.
- **PC4 (12.15 %)**: Muestra una fuerte polarización entre la humedad relativa (0.77) y el punto de rocío (-0.56), junto con una contribución negativa menor de la lluvia acumulada (-0.31). Este componente parece capturar la dinámica de situaciones donde la humedad relativa y el punto de rocío varían de manera inversa.

Para comprender cómo se distribuyen los datos en el espacio definido por los componentes principales, se presenta a continuación la proyección de los datos en los dos primeros componentes:

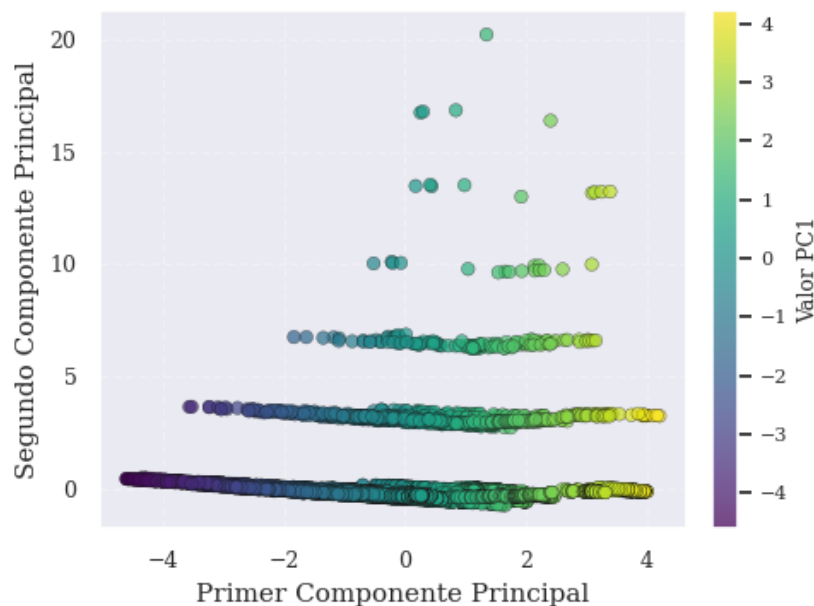


Figura 3.27: Proyección de los Datos en los Dos Primeros Componentes Principales

La proyección en los dos primeros componentes principales (Figura 3.27) revela patrones interesantes en la estructura de los datos. Se observan múltiples líneas o trayectorias de puntos que se extienden a lo largo del espacio bidimensional, con una clara tendencia ascendente. El gradiente de colores sugiere una variación continua en las condiciones meteorológicas, con clusters o agrupaciones sutiles que podrían corresponder a diferentes regímenes de precipitación. La distribución de los puntos abarca un rango aproximado de -4 a 4 en el primer componente y de 0 a 20 en el segundo componente, indicando una mayor variabilidad en la dimensión vertical (PC2, dominada por la lluvia instantánea).

Finalmente, se muestra la varianza acumulada explicada por los componentes principales:

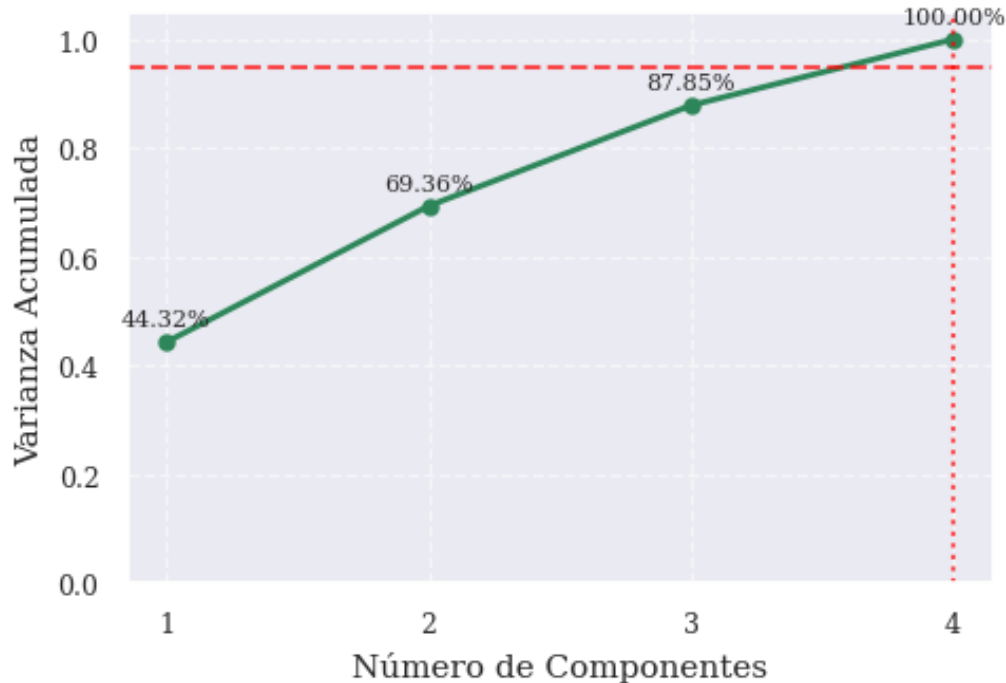


Figura 3.28: Varianza Acumulada de los Componentes Principales

La curva de varianza acumulada (Figura 3.28) muestra un incremento pronunciado inicial, alcanzando aproximadamente el 70 % de la varianza explicada con los dos primeros componentes. La curva mantiene una pendiente significativa hasta el tercer componente, donde comienza a suavizarse, alcanzando el 100 % con los cuatro componentes. Esta distribución sugiere que, si bien los dos primeros componentes capturan la mayoría de la variabilidad del sistema, los componentes tercero y cuarto todavía aportan información significativa sobre la dinámica del sistema.

Esta descomposición multivariada revela la naturaleza compleja de las interacciones entre variables meteorológicas durante eventos de lluvia. La estructura de los componentes principales sugiere que el sistema está dominado por la interacción entre las condiciones de humedad general (PC1) y los eventos de precipitación instantánea (PC2), mientras que los componentes adicionales capturan patrones más sutiles en la dinámica atmosférica. La retención de todos los componentes se justifica por su contribución única a la comprensión del sistema meteorológico completo.

3.3. Consolidación Final y Preparación de la Base de Datos para la Modelación

La consolidación de la base de datos es una etapa crítica en este estudio, ya que permite estructurar y organizar la información de manera coherente para su uso en modelos de aprendizaje automático. Este proceso garantiza que los datos estén alineados con los objetivos del estudio, particularmente en la evaluación del rendimiento del SuDS de la Pontificia Universidad Javeriana. La base de datos inicial, compuesta por registros minuto a minuto, incluye 98 eventos y 21,862 registros con variables como `date`, `time`, `input_cm`, `output_cm`, `rainfall_in`, `solar_rad_watm2`, `rh`, `dew_point_f`, `event_id` y `cumulative_rain`. Estos datos están disponibles en formato CSV para su consulta en el siguiente enlace: https://drive.google.com/file/d/1STShdDZA3iEs1QG764ynST43mp3UTu_c/view.

Homogenización de Unidades de Precipitación

Para garantizar la coherencia en las unidades utilizadas y facilitar la integración con datos climáticos futuros, se convirtieron todas las variables relacionadas con la precipitación de **pulgadas (in)** a **milímetros (mm)**. Esta conversión se realizó utilizando el factor estándar:

$$1 \text{ pulgada} = 25,4 \text{ mm.}$$

Dado que los niveles de agua en el sistema están medidos en centímetros (sistema métrico) y que los escenarios de cambio climático en Colombia suelen reportarse en unidades métricas, esta conversión es esencial para mantener la consistencia en el análisis. Además, el uso de milímetros es más apropiado para estudios hidrológicos en el contexto local, ya que permite una mayor precisión en la cuantificación de la precipitación y su impacto en el sistema.

Definición y Selección de Métricas Clave para el Análisis

En este proceso de consolidación, se seleccionaron métricas clave directamente relacionadas con los objetivos del estudio. A continuación, se describen estas métricas junto con los nombres de las variables correspondientes en la base de datos consolidada:

- **Input y output mínimos:** Valores mínimos de `input_cm` y `output_cm`, que capturan los picos de inundación y son críticos para evaluar el rendimiento del sistema en condiciones extremas. Estas métricas son nuestras **variables objetivo** y permiten identificar momentos en los que el sistema puede verse sobrecargado, lo cual es esencial para analizar su capacidad de gestión en escenarios de cambio climático. En la base de datos consolidada, estas variables se denominan `input_min_cm` y `output_min_cm`.
- **Input inicial:** Valor inicial de la variable `input_cm` al comienzo de cada evento. Esta métrica es fundamental para entender las condiciones iniciales del sistema antes de que ocurra la lluvia, lo cual es relevante para evaluar su respuesta a la entrada de agua. En la base de datos consolidada, estas variable se denomina `input_inicial_cm`.

- **Duración de la lluvia:** Tiempo transcurrido desde el inicio del evento hasta el último registro de precipitación diferente a cero. Esta métrica es crucial para entender la persistencia de la lluvia, lo cual es relevante en el contexto del SuDS, ya que eventos prolongados pueden saturar el sistema incluso si la intensidad de la lluvia no es extremadamente alta. En la base de datos consolidada, esta variable se denomina `duracion_lluvia_min`.
- **Lluvia acumulada:** Sumatoria de la precipitación registrada durante cada evento, convertida a milímetros. Esta variable es fundamental porque representa la cantidad total de agua que ingresa al sistema, lo cual está directamente relacionado con los niveles de agua en el tanque de sedimentación y su capacidad de almacenamiento. En la base de datos consolidada, esta variable se denomina `lluvia_acumulada_mm`.
- **Intensidad máxima de lluvia:** Valor máximo de la intensidad de lluvia calculada en ventanas móviles de 5 minutos, convertida a intensidad por hora en milímetros. Esta métrica captura los picos de precipitación y es esencial para evaluar la capacidad del sistema en eventos extremos, ya que los picos de lluvia pueden causar aumentos repentinos en los niveles de agua y sobrecargar el sistema. En la base de datos consolidada, esta variable se denomina `intensidad_maxima_mmph`.
- **Humedad relativa promedio:** Media de la variable `rh`, que afecta la tasa de evaporación. Una humedad alta reduce la evaporación, lo cual puede aumentar los niveles de agua en el sistema y afectar su rendimiento. Esta métrica es importante para entender el balance hídrico del sistema. En la base de datos consolidada, esta variable se denomina `rh_promedio`.
- **Punto de rocío promedio:** Media de la variable `dew_point_f`, que proporciona información sobre la humedad en el aire. Esta métrica es útil para entender las condiciones ambientales durante cada evento y su impacto en el comportamiento del sistema. En la base de datos consolidada, esta variable se denomina `dew_point_promedio_f`.
- **Mes del evento:** Mes en el que ocurrió cada evento. Esta variable permite aplicar escenarios de cambio climático basados en datos mensuales. Sin embargo, no se espera que sea operable en el modelo predictivo, ya que su función es contextual y no intrínseca a los eventos. En la base de datos consolidada, esta variable se denomina `mes_evento`.

Implementación en Python para la Generación de la Base de Datos Consolidada

El proceso de consolidación se implementó mediante una función en Python diseñada para calcular métricas clave que permiten analizar el comportamiento del sistema durante cada evento de lluvia. Esta función utiliza un enfoque basado en ventanas móviles para capturar la dinámica temporal de la lluvia, lo que resulta esencial para entender cómo los eventos extremos afectan el sistema. La implementación está alineada con los objetivos específicos del estudio, particularmente con la recolección, procesamiento y análisis de datos meteorológicos y de niveles de agua.

Propósito de la Implementación

La implementación tiene como objetivo principal:

- Capturar la variabilidad temporal de la lluvia mediante el uso de ventanas móviles, lo que permite identificar picos de precipitación y su distribución temporal.
- Evaluar el impacto en el sistema registrando métricas como los valores iniciales y mínimos de `input_cm` y `output_cm`, que reflejan cómo responde el sistema a la entrada de agua durante eventos de lluvia.
- Incorporar condiciones ambientales, como la humedad relativa promedio y el punto de rocío promedio, que influyen en procesos como la evaporación y el balance hídrico del sistema.
- Facilitar el análisis posterior generando una base de datos consolidada con todas las métricas calculadas, lo que simplifica la realización de análisis estadísticos y la generación de visualizaciones.

Resultados de la Implementación

La función genera un DataFrame consolidado que incluye todas las métricas calculadas para cada evento. Este conjunto de datos facilita el análisis posterior, permitiendo comparar eventos y evaluar el rendimiento del sistema bajo diferentes condiciones. Por ejemplo, se puede analizar cómo los eventos prolongados o de alta intensidad afectan los niveles de agua en el sistema, o cómo las condiciones ambientales influyen en la evaporación y el balance hídrico. Este análisis es fundamental para cumplir con los objetivos específicos del estudio, particularmente en la definición de escenarios de cambio climático y la validación de modelos de aprendizaje automático.

El código completo de esta implementación se puede consultar en el **cuaderno de análisis de datos de Google Colab (Python)**, cuyo enlace está incluido en el Anexo 8.1.

Desarrollo y Evaluación de Modelos de Aprendizaje Automático

4.1. Justificación de la Selección de los Modelos de Aprendizaje Automático

En esta sección se presentan los modelos de aprendizaje automático utilizados para evaluar el rendimiento del SuDS en la gestión de la cantidad de agua. Se seleccionaron cinco modelos: regresión lineal, ElasticNet, Random Forest, XGBoost y Support Vector Regression (SVR). Estos modelos fueron elegidos por su capacidad para manejar relaciones lineales y no lineales, así como por su eficiencia en la predicción de variables continuas. Dado el tamaño reducido del conjunto de datos (98 muestras), se priorizaron modelos que puedan generalizar bien y evitar el sobreajuste, al mismo tiempo que capturen las complejidades del sistema.

La selección de los modelos se realizó con el objetivo de evaluar su capacidad para predecir el nivel mínimo de agua en la entrada y la salida del SuDS (`input_min_cm` y `output_min_cm`) que son nuestras variables objetivo. A continuación se justifica la elección de cada modelo:

- **Regresión Lineal:** Se incluyó como modelo de referencia para establecer una línea base de rendimiento. Aunque es un modelo simple, es útil para identificar relaciones lineales entre las variables y el nivel de agua. Su bajo costo computacional lo hace ideal para comparar con modelos más complejos.
- **ElasticNet:** Este modelo combina las penalizaciones L1 (Lasso) y L2 (Ridge) para regularización, lo cual es especialmente útil en conjuntos de datos pequeños donde el riesgo de sobreajuste es alto. ElasticNet permite seleccionar características relevantes mientras mantiene un equilibrio entre sesgo y varianza.
- **Random Forest:** Como un ensemble de árboles de decisión, Random Forest es robusto frente a datos no lineales y puede capturar interacciones complejas entre las variables. Además, es menos propenso al sobreajuste en comparación con un solo árbol de decisión, lo que lo hace adecuado para este conjunto de datos.
- **XGBoost:** Este modelo es una implementación optimizada de Gradient Boosting, conocido por su alto rendimiento en problemas de regresión y clasificación. XGBoost es particularmente efectivo para manejar relaciones no lineales y datos con ruido, lo cual es relevante dado que los datos meteorológicos pueden presentar variabilidad y patrones complejos.

- **Support Vector Regression (SVR):** SVR con kernel RBF fue seleccionado por su capacidad para manejar relaciones no lineales mediante la transformación de los datos a un espacio de mayor dimensionalidad. Aunque es computacionalmente más costoso, es útil para evaluar si los datos presentan patrones no lineales que otros modelos no pueden capturar.

La diversidad de modelos seleccionados permite evaluar diferentes enfoques y determinar cuál es el más adecuado para predecir el rendimiento del SuDS bajo condiciones variables de lluvia y niveles de agua.

4.1.1. Justificación del Tamaño del Conjunto de Datos y la Robustez del Modelado

El presente estudio se desarrolló a partir de una base de datos histórica extensa, recopilada entre 2014 y 2016 en el campus de la Pontificia Universidad Javeriana. Aunque la recolección inicial abarcó un periodo considerable con registros minuto a minuto, el volumen bruto de datos (más de 1.1 millones de registros) presentó desafíos significativos en términos de calidad y completitud. La obtención de datos hidrológicos de alta calidad y resolución temporal, especialmente para eventos extremos como los de precipitación, es un reto considerable y complejo en el ámbito de la hidrología [23, 24]. A menudo, los estudios hidrológicos, incluso aquellos que emplean modelos físicamente basados, se enfrentan a limitaciones en la disponibilidad de registros completos y continuos debido a la complejidad inherente de la instrumentación, el mantenimiento de los sensores y las interrupciones en la recopilación de datos de campo [25].

4.1.1.1. Consolidación de Datos: De Registros Minuto a Minuto a Eventos Discretos

Una decisión metodológica fundamental en este estudio fue la transición de datos registrados minuto a minuto a un único registro por evento de lluvia. Esta consolidación se justifica por múltiples razones estratégicas y metodológicas:

Naturaleza y Volumen Inicial de los Datos: La investigación se basó en mediciones detalladas de niveles de agua en el Sistema de Drenaje Sostenible (SuDS) y diversas variables meteorológicas (precipitación, radiación solar, temperatura y humedad relativa), registradas cada minuto durante el período de estudio.

Enfoque en Eventos Hidrológicos: Dada la naturaleza del análisis hidrológico, especialmente en la evaluación de los SuDS, el objetivo principal del estudio era evaluar la eficacia del sistema en la gestión del agua durante eventos de lluvia específicos y predecir su respuesta a escenarios climáticos extremos. Analizar estos eventos como unidades discretas, en lugar de flujos continuos minuto a minuto, permite una comprensión más clara del comportamiento del sistema durante períodos de estrés hidrológico.

Derivación de Métricas Clave por Evento: En lugar de utilizar cada lectura minuto a minuto, el proceso de consolidación implicó el cálculo de métricas resumen clave para cada evento de lluvia identificado. Estas métricas, como el nivel mínimo de entrada/salida de agua, el nivel inicial de entrada, la precipitación total acumulada, la intensidad máxima de precipitación y la humedad y el punto de rocío promedio para el evento, fueron elegidas precisamente porque son directamente relevantes para comprender y predecir el rendimiento del SuDS. Estos valores únicos y agregados por evento encapsulan la información esencial necesaria para el modelo.

Aplicación Simplificada del Modelo para Usuarios: El objetivo final es que el modelo desarrollado sea fácilmente utilizable como herramienta para sistemas de alerta temprana y gestión operativa. Al entrenar el modelo con estas métricas consolidadas y basadas en eventos, la entrada requerida por los usuarios o por otros sistemas para la predicción se simplifica significativamente. En lugar de necesitar flujos de datos continuos minuto a minuto, un usuario puede introducir unos pocos parámetros clave a nivel de evento (por ejemplo, la precipitación acumulada proyectada para un evento, su intensidad máxima y el nivel inicial actual del tanque) para obtener una predicción sencilla de la respuesta del sistema. Esta elección de diseño mejora la aplicabilidad práctica y la facilidad de uso del modelo en escenarios del mundo real.

4.1.1.2. Proceso de Depuración y Control de Calidad

El riguroso proceso de depuración y consolidación de datos fue un paso indispensable para asegurar la calidad, fiabilidad y validez del conjunto de datos utilizado para entrenar y validar los modelos. Este proceso se detalla en el Capítulo 3 y fue fundamental para la selección de eventos de lluvia significativos:

- **Identificación de Datos Faltantes:** Se observó una alta cantidad de valores nulos en la serie temporal completa, lo que hizo inviable la imputación generalizada, ya que podría comprometer la calidad del análisis.
- **Enfoque en Eventos de Lluvia Específicos:** Dada la naturaleza del análisis hidrológico y la presencia de datos faltantes, se tomó la decisión metodológica de enfocar el estudio en eventos de lluvia específicos en lugar de toda la serie temporal.
- **Exclusión por Incompletitud en Variables Clave:** De una base inicial de eventos, se identificaron 13 eventos con datos faltantes en las variables de niveles de agua (`input_cm` y `output_cm`). Dado que estos registros incompletos incluían más de 42 minutos consecutivos sin datos y representaban el objetivo principal del análisis, se decidió excluir dichos eventos para garantizar la integridad de los datos hidrológicos. Esto redujo la base a 118 eventos completos.
- **Tratamiento de Valores Atípicos y Anómalos:** Posteriormente, se realizó una exhaustiva detección y tratamiento de valores atípicos. Por ejemplo, los eventos 16 y 115 fueron descartados por comportamientos anómalos en el nivel de entrada, y los eventos del 17 al 130 fueron excluidos debido a fallas en el sensor de nivel de salida. Adicionalmente, el evento 113 se eliminó por tener varianza cero en la variable de salida.
- **Resultado Final:** Después de este proceso de depuración y consolidación, el número final de eventos para el análisis se estableció en 98 eventos únicos y completos.

4.1.1.3. Justificación de la Robustez del Modelado

A pesar de que el tamaño del conjunto de datos final (98 muestras) pueda considerarse reducido para algunas aplicaciones de aprendizaje automático, la robustez de las validaciones y modelación se justifica por las siguientes razones:

- **Selección de Modelos Robustos:** Se priorizaron modelos de aprendizaje automático como SVR, Random Forest y XGBoost, conocidos por su capacidad para generalizar bien y evitar el sobreajuste incluso en conjuntos de datos de tamaño moderado, mientras capturan relaciones no lineales complejas [26, 27]. El SVR, en particular, es menos sensible a valores atípicos, lo que resulta ventajoso al trabajar con datos de precipitación que pueden presentar eventos extremos [26].
- **Estrategias de Optimización Rigurosas:** La aplicación de la optimización bayesiana permitió explorar de manera eficiente el espacio de hiperparámetros y encontrar configuraciones óptimas sin requerir un número excesivo de iteraciones, lo cual es particularmente útil en conjuntos de datos pequeños.
- **Priorización de la Calidad sobre la Cantidad:** La reducción en la cantidad de eventos fue un paso necesario para garantizar la calidad y la coherencia de los datos utilizados en el análisis hidrológico, asegurando que los eventos incluidos sean representativos y confiables para modelar el comportamiento del SuDS. Este enfoque se alinea con la práctica en hidrología, donde la calidad y representatividad de los eventos son más cruciales que la cantidad bruta, especialmente cuando se trabaja con datos de campo complejos y propensos a interrupciones [28, 24].
- **Transparencia Metodológica:** El estudio reconoce transparentemente que, si bien el modelo actual es robusto, “podría beneficiarse de muestras adicionales”, especialmente si estas “aportan información sobre escenarios poco representados en el conjunto actual”. Esto demuestra una comprensión de las limitaciones y una visión para futuros trabajos.
- **Relevancia del Problema Abordado:** El proyecto se enmarca en la “necesidad urgente de superar las limitaciones de los sistemas de drenaje existentes ante eventos climáticos extremos”, en un contexto donde los análisis predictivos que consideran estos escenarios son escasos. El modelado, incluso con un conjunto de datos depurado, proporciona insumos valiosos para el diseño de medidas preventivas y adaptativas.

En resumen, la consolidación de datos minuto a minuto en registros basados en eventos, caracterizados por métricas resumen clave, fue una decisión estratégica impulsada por los desafíos intrínsecos de los datos hidrológicos, los objetivos específicos del estudio y la meta de desarrollar un modelo predictivo práctico y fácil de usar para la gestión de los SuDS frente al cambio climático. La cantidad limitada de datos finales no compromete la validez de los resultados, ya que proviene de un proceso de selección riguroso enfocado en la calidad de la información obtenida a pesar de las complejidades inherentes a la recolección de datos hidrológicos, y los modelos utilizados fueron escogidos y optimizados específicamente para trabajar eficazmente con conjuntos de datos de estas características.

4.2. Modelado del Nivel de Entrada del SuDS mediante Aprendizaje Automático

Basándose en los datos preparados en la fase de Análisis Exploratorio y Consolidación, esta sección se centra en el Modelado del Nivel de Entrada del SuDS mediante Aprendizaje Automático. Aquí se

presenta la selección y justificación de modelos como Regresión Lineal, ElasticNet, Random Forest, XGBoost y SVR, elegidos por su capacidad para manejar relaciones presentes en los datos y su idoneidad para el conjunto de datos disponible. Se detalla el entrenamiento y evaluación inicial de estos modelos en sus versiones base y optimizadas mediante búsqueda bayesiana, realizando un análisis comparativo de su rendimiento (R^2 , MSE, RMSE, MAE). Se identifica el modelo SVR optimizado como el de mejor desempeño predictivo para el nivel de entrada. La sección profundiza en el análisis detallado de este modelo a través de curvas de aprendizaje y residuales, y evalúa una versión simplificada, concluyendo con la justificación de la elección del modelo.

4.2.1. Entrenamiento y Evaluación Inicial de los Modelos para el Nivel de Entrada

4.2.1.1. Modelos Base

Los modelos base son versiones iniciales de los algoritmos de aprendizaje automático que se entrenan con configuraciones predeterminadas o con hiperparámetros no optimizados. Estos modelos sirven como un punto de referencia para evaluar el rendimiento inicial de cada algoritmo antes de aplicar técnicas de optimización, como la búsqueda de hiperparámetros.

En este estudio, se entrenaron cinco modelos base: regresión lineal, ElasticNet, Random Forest, XGBoost y Support Vector Regression (SVR). Cada uno de estos modelos se configuró con sus parámetros predeterminados o con valores iniciales comunes, sin realizar ajustes específicos. Esto permitió evaluar su rendimiento "fuera de la caja" compararlos entre sí antes de proceder a la optimización de hiperparámetros.

Las métricas de evaluación (R^2 , MSE, RMSE, MAE) se calcularon en el conjunto de prueba y mediante validación cruzada. Los resultados se resumen en la Tabla 4.1.

Modelo	R^2	MSE	RMSE	MAE
ElasticNet	0.7219	7.2388	2.6905	2.3029
Regresión Lineal	0.7110	7.5241	2.7430	2.3057
Random Forest	0.8816	3.0827	1.7558	1.4034
SVR	0.6819	8.2800	2.8775	2.1209
XGBoost	0.8890	2.8884	1.6995	1.2905

Tabla 4.1: Métricas de evaluación para modelos base.

4.2.1.2. Modelos Optimizados por Bayes

Para mejorar el rendimiento de los modelos, se utilizó la optimización bayesiana (BayesSearchCV) para ajustar los hiperparámetros. Esta técnica es particularmente útil en conjuntos de datos pequeños, ya que permite explorar de manera eficiente el espacio de parámetros y encontrar configuraciones óptimas sin requerir un número excesivo de iteraciones. La optimización bayesiana utiliza un modelo sustituto (como un proceso gaussiano) para guiar la búsqueda, priorizando regiones del espacio de hiperparámetros que tienen mayor probabilidad de mejorar el rendimiento del modelo.

Configuración de los Espacios de Búsqueda

Los espacios de búsqueda para cada modelo se definieron de manera estratégica, considerando tanto la naturaleza del conjunto de datos como el rendimiento de los modelos base. Se utilizaron distribuciones apropiadas para cada tipo de hiperparámetro: distribuciones log-uniformes para parámetros de regularización y tasas de aprendizaje, y distribuciones uniformes para parámetros estructurales. La configuración específica se detalla a continuación:

- **ElasticNet:** Se optimizaron dos hiperparámetros principales: el parámetro de regularización α en el rango $[0.001, 1.0]$ con distribución log-uniforme, permitiendo explorar tanto regularización débil como fuerte; y la relación L1 (l_1_ratio) en el rango $[0.1, 0.9]$, balanceando entre regularización Ridge y Lasso.
- **Random Forest:** El espacio de búsqueda incluyó: número de estimadores entre 40 y 100 (centrado alrededor del valor base de 50), profundidad máxima entre 4 y 7 (centrado en el valor base de 5), mínimo de muestras para dividir un nodo entre 2 y 5, y mínimo de muestras por hoja entre 1 y 3. Estos rangos se ajustaron considerando el tamaño limitado del conjunto de datos para evitar sobreajuste.
- **XGBoost:** Se definieron seis hiperparámetros: número de estimadores entre 40 y 100, profundidad máxima entre 2 y 5 (valores más conservadores que Random Forest), tasa de aprendizaje entre 0.05 y 0.2 con distribución log-uniforme, fracción de submuestreo entre 0.7 y 1.0, fracción de características por árbol entre 0.7 y 1.0, y peso mínimo por hoja entre 1 y 3. Los valores mínimos más altos en los parámetros de submuestreo buscan mantener suficiente información para el entrenamiento en un conjunto de datos pequeño.
- **Support Vector Regression (SVR):** El espacio incluyó: parámetro de regularización C entre 0.1 y 10 con distribución log-uniforme, parámetro gamma entre 0.01 y 1.0 con distribución log-uniforme, selección categórica entre kernels RBF y lineal, y parámetro epsilon entre 0.01 y 0.2. Los rangos se acotaron respecto a búsquedas estándar para mejorar la eficiencia en conjuntos pequeños.

Proceso de Optimización

El proceso de optimización se configuró con 30 iteraciones por modelo, utilizando validación cruzada de 5 pliegues y el coeficiente de determinación (R^2) como métrica de evaluación. Este número de iteraciones representa un balance entre exploración exhaustiva del espacio de hiperparámetros y prevención del sobreajuste, especialmente relevante dado el tamaño limitado del conjunto de datos.

Es importante destacar que la optimización bayesiana no se aplicó a la regresión lineal, ya que este modelo no tiene hiperparámetros complejos que requieran ajuste. Los coeficientes de la regresión lineal se calculan directamente mediante métodos analíticos, como mínimos cuadrados, lo que hace innecesaria la optimización de hiperparámetros.

Los resultados de la optimización se muestran en la Tabla 4.2, donde se puede observar el impacto de la búsqueda bayesiana en el rendimiento de cada modelo.

Modelo	R^2	MSE	RMSE	MAE
ElasticNet (Bayes)	0.7323	6.9687	2.6398	2.3107
Random Forest (Bayes)	0.8951	2.7318	1.6528	1.3617
XGBoost (Bayes)	0.8768	3.2064	1.7907	1.4003
SVR (Bayes)	0.9024	2.5399	1.5937	1.3610

Tabla 4.2: Métricas de evaluación para modelos optimizados mediante búsqueda bayesiana.

4.2.2. Análisis Comparativo del Rendimiento de los Modelos Base y Optimizados

A continuación, se presenta la comparación del desempeño de los modelos base y optimizados mediante las métricas R^2 , MSE, RMSE y MAE, representadas en las Figuras 4.1, 4.2, 4.3 y 4.4, respectivamente. Estas métricas fueron seleccionadas por ofrecer una visión complementaria del rendimiento: mientras R^2 mide la capacidad explicativa del modelo, el MSE y el RMSE penalizan errores grandes y el MAE resume el error medio sin exagerar valores extremos.

Las visualizaciones permiten identificar de forma clara las mejoras, retrocesos o estabilidad en el rendimiento de los modelos tras la optimización bayesiana. Esta representación gráfica facilita la interpretación del impacto de la optimización sobre su capacidad predictiva y su adecuación al comportamiento complejo del sistema hidráulico analizado.

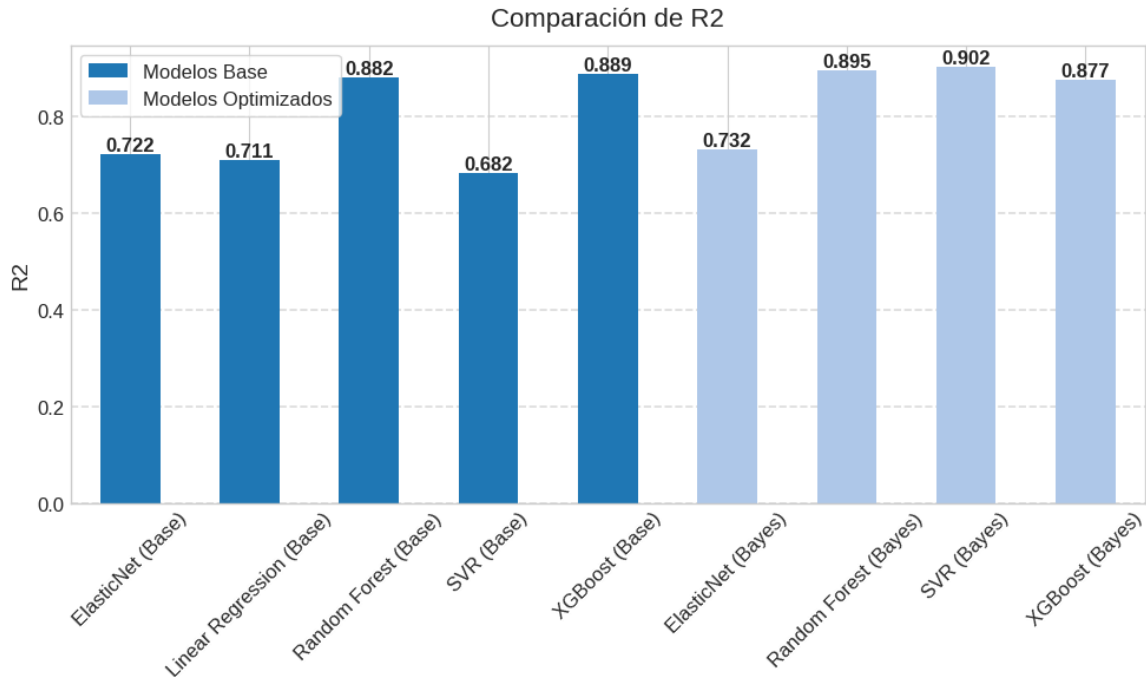


Figura 4.1: Comparación del coeficiente de determinación (R^2) de modelos base y optimizados.

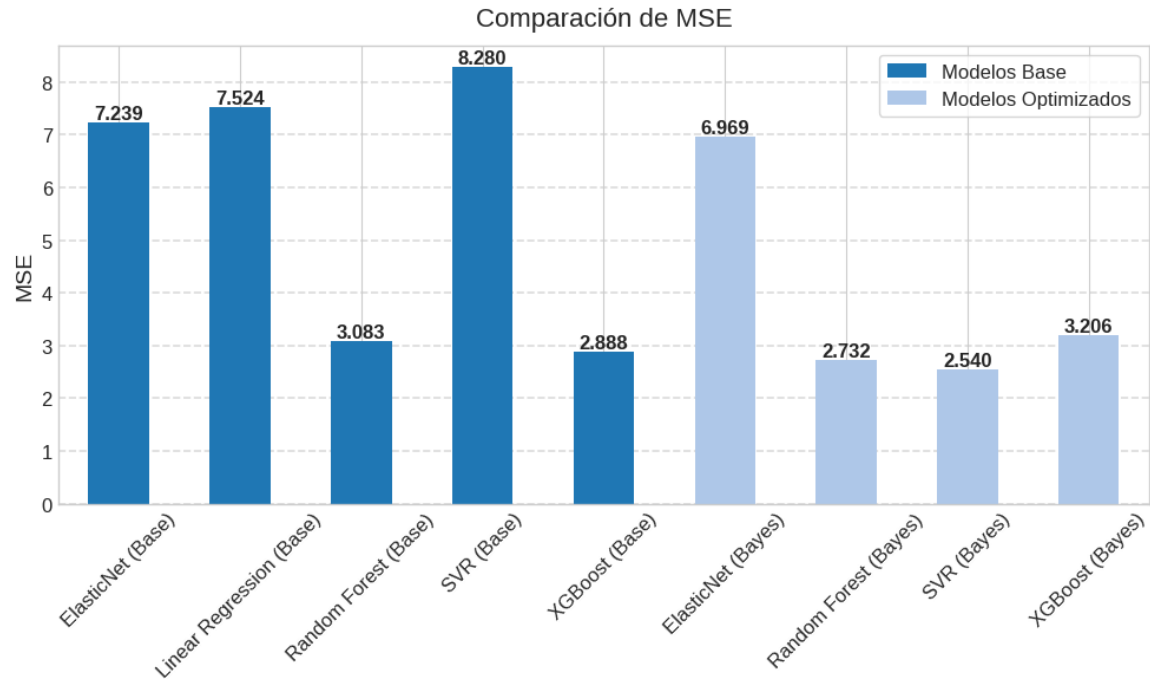


Figura 4.2: Comparación del error cuadrático medio (MSE) de modelos base y optimizados.

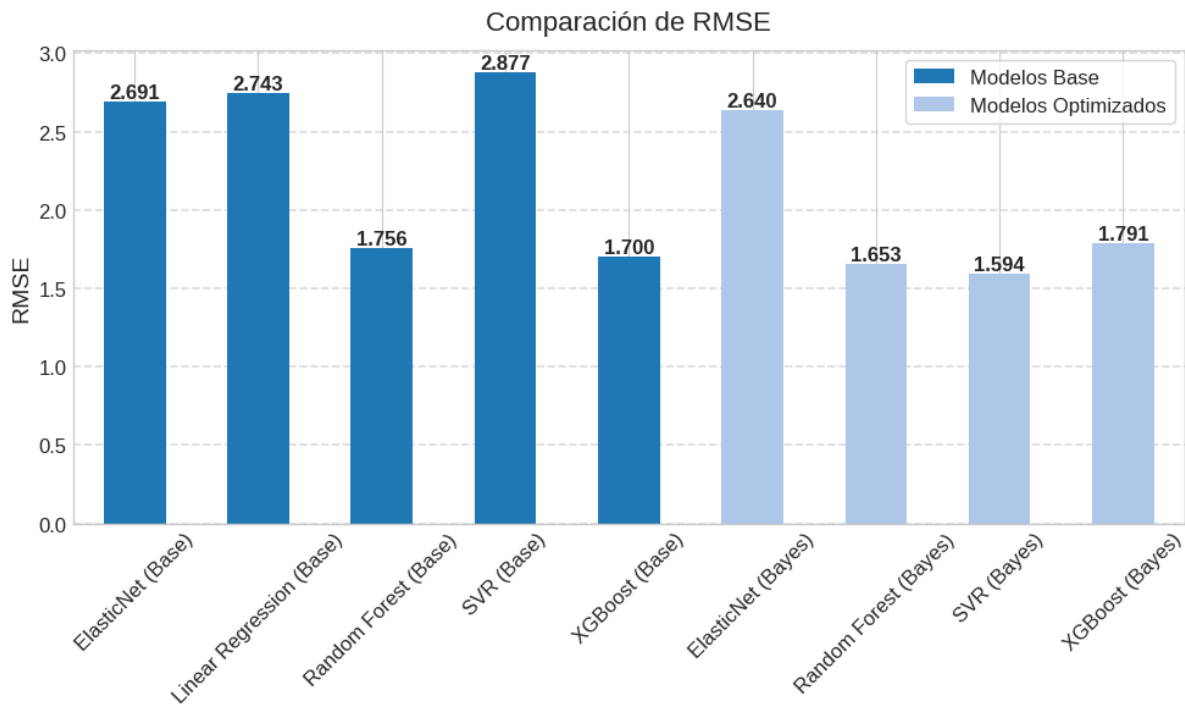


Figura 4.3: Comparación del error RMSE de modelos base y optimizados.

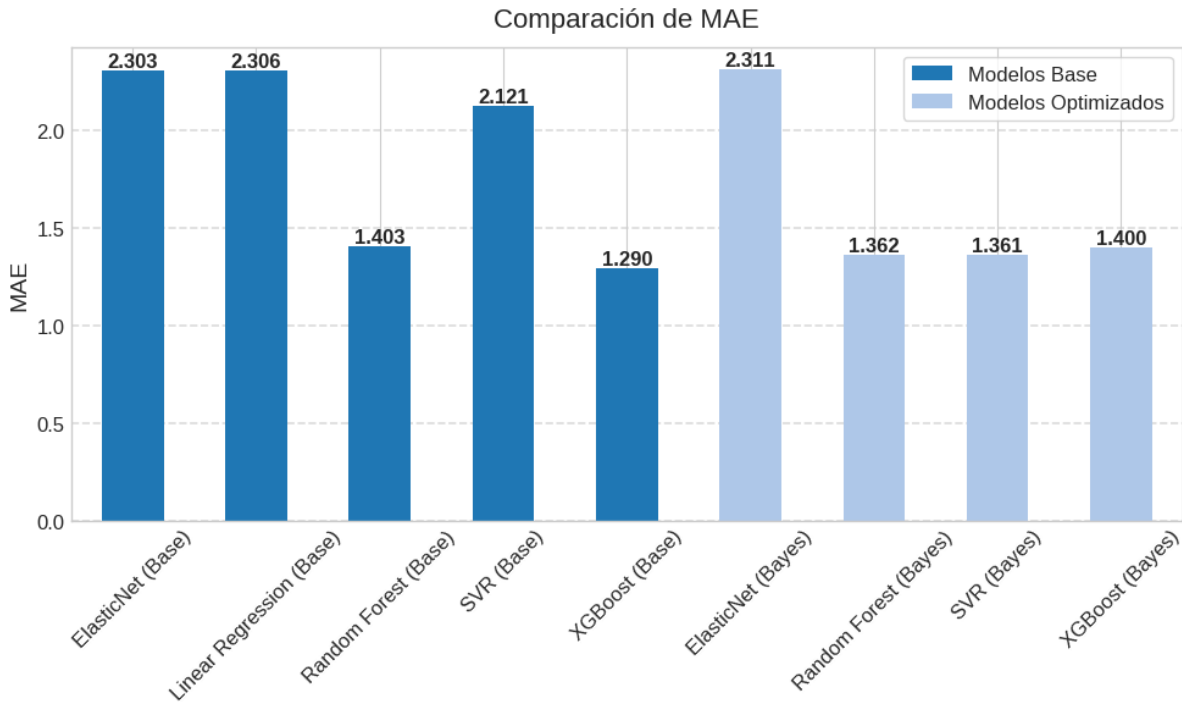


Figura 4.4: Comparación del error absoluto medio (MAE) de modelos base y optimizados.

En términos de R^2 , el SVR optimizado (SVR Bayes) presenta el valor más alto (0.902), seguido de cerca por Random Forest optimizado (Random Forest Bayes) con 0.895, y XGBoost base con 0.889. Es notable que la optimización bayesiana mejoró significativamente el rendimiento del SVR, aumentando su R^2 de 0.682 a 0.902, lo que lo convierte en el modelo más efectivo tras la optimización.

Sin embargo, en el caso de XGBoost, la optimización resultó en una ligera disminución del rendimiento (R^2 de 0.889 a 0.877), lo que sugiere que la configuración inicial ya estaba cercana al óptimo o que posiblemente ocurrió un sobreajuste durante la optimización. Para los modelos lineales (Regresión Lineal y ElasticNet), tanto en sus versiones base como optimizadas, se observa un rendimiento moderado con R^2 entre 0.711 y 0.732, reflejando las limitaciones de estos enfoques para capturar relaciones no lineales presentes en los datos.

En cuanto al MSE, RMSE y MAE, se mantiene la tendencia: SVR Bayes presenta los mejores valores (MSE: 2.540, RMSE: 1.594, MAE: 1.361), seguido de cerca por Random Forest Bayes y XGBoost Base. Estos resultados validan la efectividad de los modelos no lineales para este problema específico de predicción hidráulica, donde las relaciones entre variables meteorológicas y niveles de agua suelen ser complejas y no lineales.

4.2.2.1. Análisis de Predicciones vs Valores Reales

La Figura 4.5 presenta un gráfico de dispersión que compara las predicciones generadas por todos los modelos —tanto en su configuración base como tras la optimización bayesiana— con los valores reales observados. Este tipo de representación es fundamental para evaluar visualmente el rendimiento

global de los modelos y detectar posibles sesgos o desviaciones sistemáticas en las predicciones.

En el gráfico, cada punto representa una predicción correspondiente a un valor real específico, codificada por el modelo que la generó. La línea negra discontinua indica la línea de referencia donde las predicciones serían perfectamente iguales a los valores reales ($y = x$). Así, cuanto más cercanos se encuentran los puntos a esta línea, mejor es la precisión del modelo correspondiente.

Adicionalmente, se incluyen dos líneas rojas discontinuas que delimitan un margen de error del $\pm 10\%$, lo cual permite establecer un umbral visual para considerar qué tan aceptables son las predicciones. Los puntos ubicados dentro de esta banda indican predicciones con una desviación relativa baja respecto al valor real, lo cual es especialmente relevante en el contexto de la predicción de variables hidráulicas, donde pequeños errores pueden tener implicaciones significativas.

Este gráfico facilita una comparación cualitativa entre modelos, revelando no solo cuál tiene menor dispersión o mayor concentración dentro del margen de error, sino también posibles patrones de subestimación o sobreestimación. En conjunto con las métricas cuantitativas analizadas previamente, esta figura proporciona una validación visual del impacto de la optimización sobre la capacidad predictiva de cada modelo.

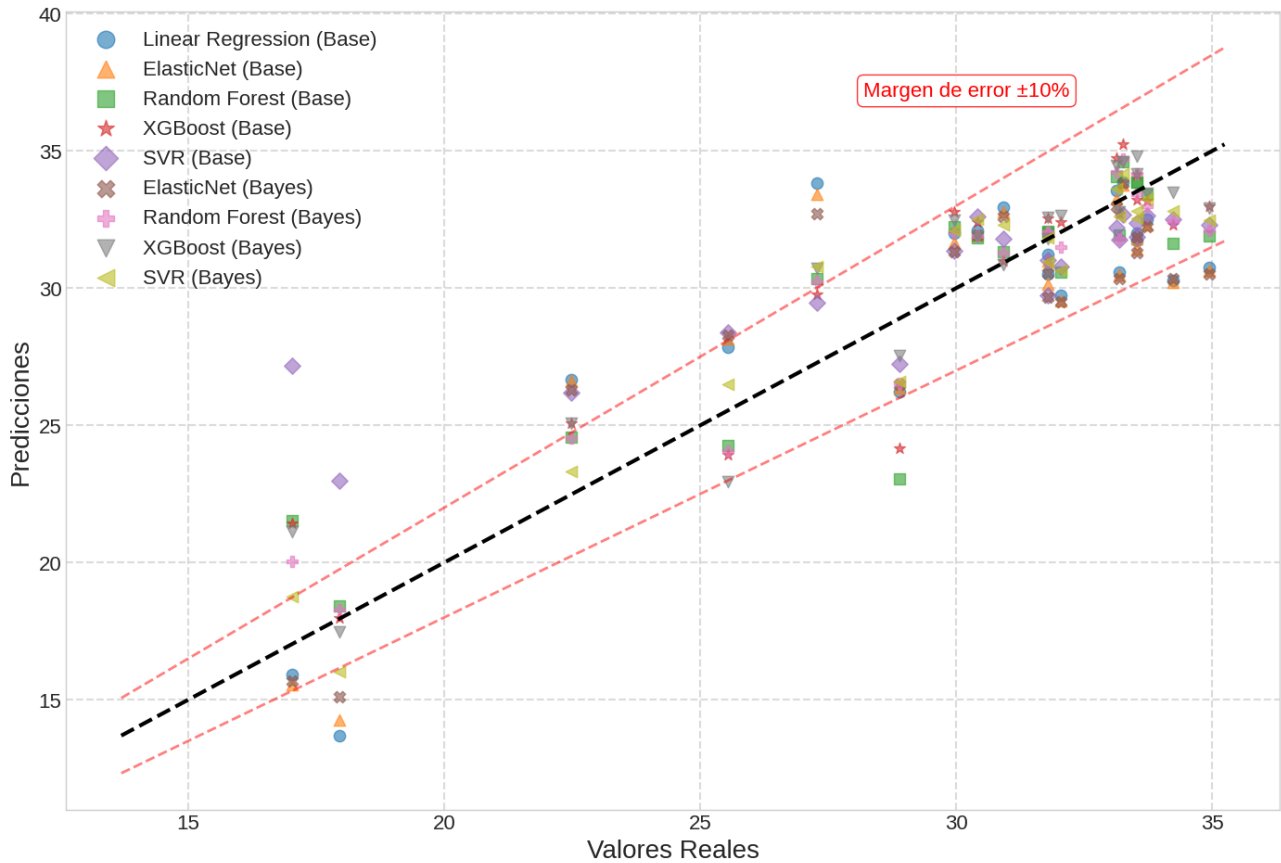


Figura 4.5: Predicciones vs Valores Reales para todos los modelos.

Se observa que la mayoría de los puntos se concentran en valores entre 25 y 35 cm, lo que indica que

el conjunto de datos presenta cierta homogeneidad en los niveles de agua registrados. Los modelos no lineales optimizados (SVR Bayes y Random Forest Bayes) junto con XGBoost Base tienden a agruparse más cerca de la línea diagonal, lo que confirma su mejor capacidad predictiva visualizada anteriormente en las métricas numéricas.

Es notable que para valores extremos, especialmente aquellos por debajo de 20 cm, todos los modelos presentan mayor dispersión y tienden a sobrestimar los valores más bajos. Esto podría deberse a la menor representación de estos valores en el conjunto de entrenamiento o a comportamientos no lineales más pronunciados en estos rangos.

Los modelos lineales (Regresión Lineal y ElasticNet) muestran una dispersión más amplia en todo el rango de valores, especialmente en los extremos, confirmando sus limitaciones para capturar relaciones complejas. Por otro lado, el SVR optimizado (SVR Bayes) muestra una mejora considerable respecto a su versión base, con puntos más cercanos a la línea diagonal, lo que valida la efectividad de la optimización bayesiana para este algoritmo en particular.

4.2.2.2. Evaluación de Riesgo de Sobreajuste

Una preocupación importante en modelos de aprendizaje automático, especialmente con conjuntos de datos pequeños, es el riesgo de sobreajuste. Para evaluar este riesgo, se comparó el rendimiento de cada modelo en los conjuntos de entrenamiento y prueba:

SVR (Bayes):

- R^2 en entrenamiento: 0.8918
- R^2 en prueba: 0.9024
- Diferencia: -0.0106
- **EXCELENTE:** Buena generalización

Random Forest (Bayes):

- R^2 en entrenamiento: 0.9701
- R^2 en prueba: 0.8951
- Diferencia: 0.0751
- **CORRECTO:** Generalización aceptable

XGBoost (Base):

- R^2 en entrenamiento: 0.9738
- R^2 en prueba: 0.8368
- Diferencia: 0.1370

- **CORRECTO:** Generalización aceptable

Los resultados muestran que SVR (Bayes) tiene un excelente equilibrio entre ajuste y generalización, incluso con un rendimiento ligeramente mejor en el conjunto de prueba que en el de entrenamiento, lo que es inusual y muy positivo. Random Forest (Bayes) y XGBoost (Base) también muestran una generalización aceptable, aunque con algo más de sobreajuste.

Resultados Clave

- El modelo SVR optimizado mediante optimización bayesiana resultó ser el más adecuado para predecir el rendimiento del SuDS, con un R^2 de 0.9024 y un RMSE de 1.5937 cm en el conjunto de prueba.
- La optimización bayesiana mejoró significativamente el rendimiento del SVR, transformándolo del modelo con peor desempeño en su versión base al mejor modelo tras la optimización.
- Los modelos basados en árboles (*Random Forest* y *XGBoost*) también mostraron un rendimiento sólido, confirmando la naturaleza no lineal de las relaciones entre las variables meteorológicas y los niveles de agua en el sistema.
- Los modelos lineales (*Regresión Lineal* y *ElasticNet*) presentaron un rendimiento moderado, indicando que existen relaciones no lineales importantes que estos modelos no pueden capturar adecuadamente.
- Se recomienda recolectar más datos, especialmente en los rangos extremos donde los modelos mostraron mayor error, para mejorar la robustez y la capacidad de generalización de los modelos.

4.2.2.3. Selección del Modelo Óptimo

Tras la comparación exhaustiva, el **SVR optimizado mediante BayesSearchCV** demostró el mejor rendimiento global (R^2 : 0.9024, MSE: 2.5399), superando significativamente a su versión base ($\Delta R^2 = +0.220$). Los hiperparámetros óptimos encontrados fueron:

- **Kernel:** RBF ($\gamma = 0.0512$)
- **Parámetro de regularización:** $C = 4.890$
- **Margen de error:** $\varepsilon = 0.166$

Con un R^2 de validación cruzada de 0.7683, el modelo mostró robustez estadística. Este desempeño justifica su selección para un análisis detallado.

4.2.3. Análisis Detallado del Modelo SVR Optimizado para el Nivel de Entrada

4.2.3.1. Curvas de Aprendizaje

En la Figura 4.6 se muestran las curvas de aprendizaje para el mejor modelo (SVR Bayes). Estas curvas permiten evaluar si el modelo está sobreajustando o subajustando, y si se beneficiaría de más datos.

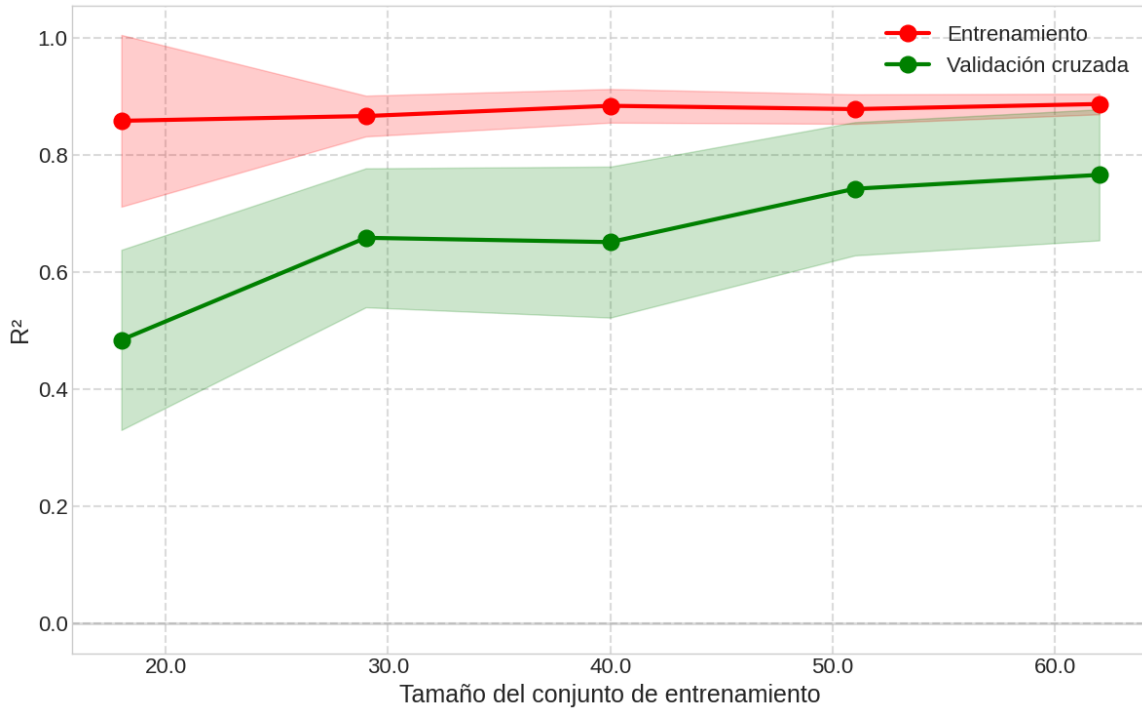


Figura 4.6: Curvas de aprendizaje para SVR (Bayes).

Las curvas de aprendizaje para el modelo SVR Bayes proporcionan información valiosa sobre el comportamiento del modelo a medida que aumenta el tamaño del conjunto de entrenamiento. La línea roja, que representa el rendimiento en el conjunto de entrenamiento, se mantiene en valores altos de R^2 (entre 0.85 y 0.90), lo que indica un buen ajuste en los datos de entrenamiento sin llegar al sobreajuste perfecto.

La línea verde, que representa el rendimiento en validación cruzada, muestra una tendencia creciente desde aproximadamente 0.48 hasta estabilizarse alrededor de 0.77, lo que indica una mejora consistente en la capacidad predictiva a medida que se dispone de más ejemplos de entrenamiento.

La amplia área sombreada verde representa el intervalo de confianza de la validación cruzada y muestra una reducción gradual, indicando que el modelo se vuelve más estable y consistente a medida que aumenta el tamaño del conjunto de entrenamiento. La brecha entre las curvas de entrenamiento y validación se va reduciendo, lo que sugiere que el modelo está alcanzando un buen equilibrio entre sesgo y varianza.

Es notable que la curva de validación tiende a estabilizarse a partir de 40-50 muestras aproxima-

damente, pero sigue mostrando una tendencia ascendente incluso con 60 muestras. Esto sugiere que, aunque el modelo ya está capturando gran parte de la estructura de los datos, podría beneficiarse de muestras adicionales, especialmente si estas aportan información sobre escenarios poco representados en el conjunto actual.

4.2.3.2. Análisis de Residuales

En esta sección se presenta el análisis de residuales para el mejor modelo (SVR Bayes), dividido en cuatro gráficos individuales para una mejor visibilidad. Estos gráficos permiten evaluar la distribución de los errores y detectar patrones no deseados en las predicciones del modelo.

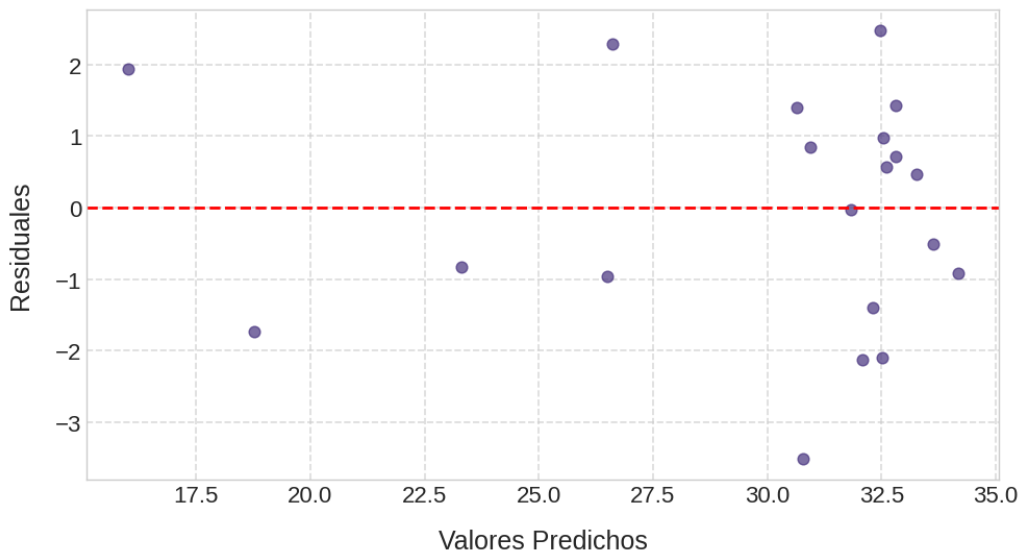


Figura 4.7: Residuales vs. valores predichos para SVR (Bayes).

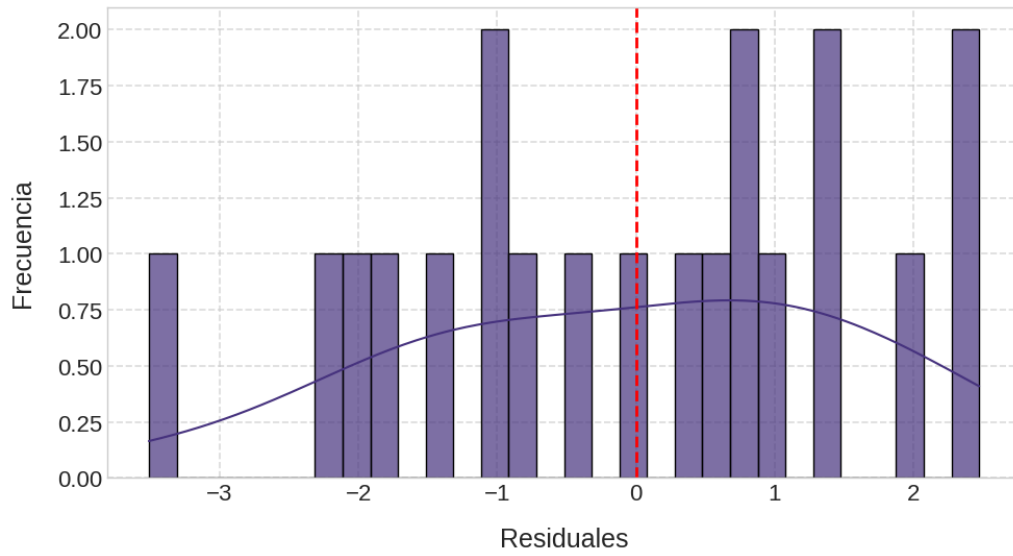


Figura 4.8: Distribución de residuales para SVR (Bayes).

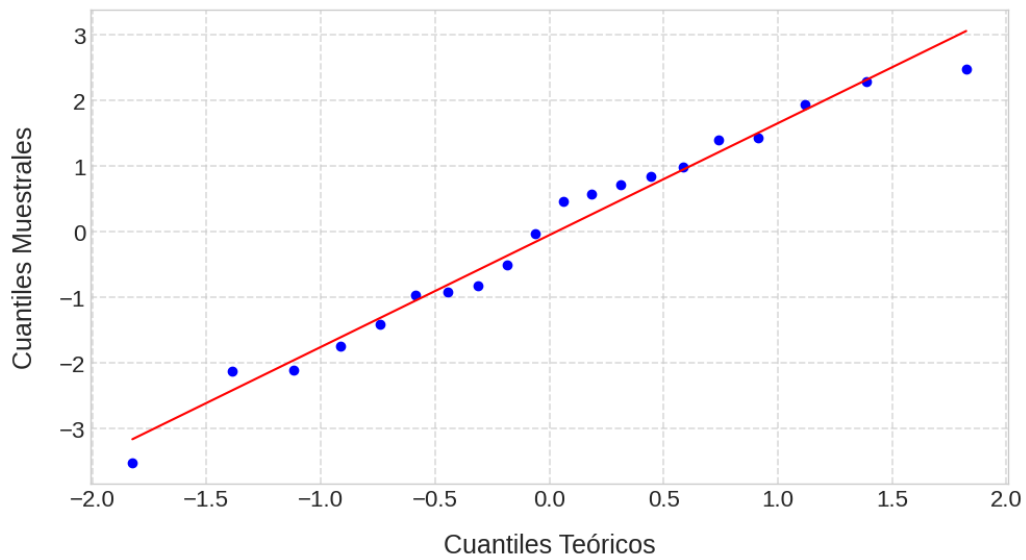


Figura 4.9: Gráfico Q-Q de residuales para SVR (Bayes).

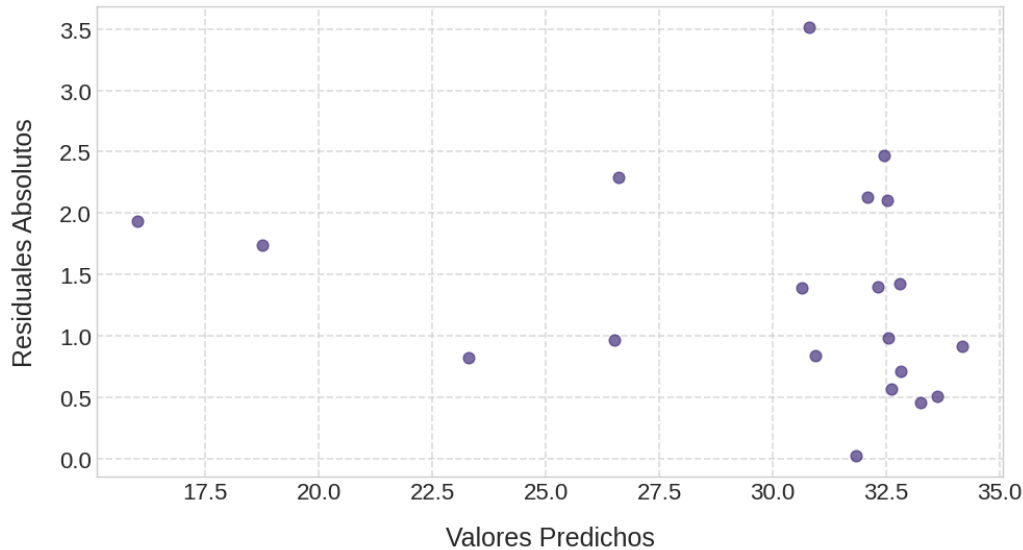


Figura 4.10: Residuales absolutos vs. valores predichos para SVR (Bayes).

Interpretación de los Gráficos

El análisis de residuales del modelo SVR Bayes ofrece insights importantes sobre la calidad de las predicciones y posibles áreas de mejora:

- **Residuales vs. Valores Predichos (Figura 4.7):** No se observa un patrón sistemático claro en los residuales, lo cual es deseable. Sin embargo, se identifican algunos valores atípicos, particularmente para predicciones en los extremos, donde se presentan residuales de mayor magnitud.
- **Distribución de Residuales (Figura 4.8):** La distribución de residuales se aproxima a una distribución normal con cierta asimetría positiva, centrada ligeramente por debajo de cero. Esto indica un pequeño sesgo hacia la sobrestimación. La mayoría de los residuales se concentran en el rango de -2 a 2 cm, lo que es aceptable considerando la naturaleza del problema.
- **Gráfico Q-Q (Figura 4.9):** Los residuales siguen bastante de cerca la línea teórica de una distribución normal, con algunas desviaciones en los extremos. Esto confirma la aproximación a la normalidad, aunque con cierta asimetría en los residuales de mayor magnitud.
- **Residuales Absolutos vs. Valores Predichos (Figura 4.10):** Se observa cierta heterocedasticidad, con mayor variabilidad en los residuales para diferentes rangos de valores predichos. Esto sugiere que el modelo puede tener diferente precisión dependiendo del rango de valores que esté prediciendo.

Análisis Estadístico de Residuales

El análisis estadístico de los residuales proporciona información clave sobre la calidad del modelo SVR Bayes. A continuación, se presentan los resultados más relevantes:

- **Media de residuales** ($-0,0516$): La media cercana a cero sugiere que, en promedio, el modelo no tiende a subestimar ni sobrestimar significativamente los valores reales. Sin embargo, el ligero valor negativo indica una mínima tendencia hacia la sobrestimación.
- **Mediana de residuales** ($0,2163$): La mediana positiva confirma que la mayoría de los residuales están ligeramente por encima de cero, lo que refuerza la idea de una pequeña sobrestimación en las predicciones del modelo.
- **Desviación estándar de residuales** ($1,5929$): Este valor indica la dispersión de los residuales alrededor de la media. Una desviación estándar relativamente baja sugiere que la mayoría de los errores están concentrados en un rango estrecho, lo cual es deseable. Sin embargo, la presencia de valores atípicos (como se observa en los gráficos) puede estar influyendo en este resultado.
- **Test de Shapiro-Wilk (p-valor: $0,7745$)**: El p-valor obtenido es mayor que el nivel de significancia común ($\alpha = 0,05$), lo que indica que no se puede rechazar la hipótesis nula de normalidad en los residuales. Esto sugiere que los residuales siguen una distribución aproximadamente normal, lo cual es una condición deseable para la validez del modelo. Sin embargo, la ligera asimetría positiva observada en la distribución de residuales (Figura 4.8) sugiere que el modelo puede tener un pequeño sesgo hacia la sobrestimación en ciertos casos.

4.2.3.3. Importancia de Características

Para evaluar la contribución de cada variable en la predicción del rendimiento del SuDS en el modelo SVR (Bayes), se aplicó la técnica de *Permutation Feature Importance*. Este método mide la disminución en el desempeño del modelo cuando se permutan aleatoriamente los valores de una variable, permitiendo identificar aquellas que tienen mayor impacto en la predicción.

La Tabla 4.3 presenta los valores de importancia calculados, junto con su desviación estándar tras múltiples permutaciones.

Característica	Importancia (<i>mean</i> \pm <i>std</i>)
Intensidad máxima (mm/h)	0.439 ± 0.117
Lluvia acumulada (mm)	0.382 ± 0.105
Input inicial (cm)	0.371 ± 0.125
Duración de la lluvia (min)	0.032 ± 0.009
Punto de rocío promedio (°F)	0.008 ± 0.013
Humedad relativa promedio (%)	-0.008 ± 0.010

Tabla 4.3: Importancia de características en la predicción del rendimiento del SuDS.

Los resultados muestran que la **intensidad máxima de la lluvia** es la variable con mayor influencia en la predicción del comportamiento del SuDS. Esto sugiere que eventos de lluvia con alta intensidad pueden afectar significativamente la capacidad de respuesta del sistema, posiblemente debido a la rapidez con la que se generan escorrentías y cambios en los niveles de agua.

La **lluvia acumulada** y el **nivel inicial del agua en el sistema** también juegan un papel crucial en la predicción, lo que indica que tanto la cantidad total de precipitación como las condiciones iniciales

del SuDS son determinantes en su desempeño. La lluvia acumulada representa el volumen total de agua que ingresa al sistema, mientras que el nivel inicial define la capacidad disponible para retención antes del evento de precipitación.

En contraste, la **duración del evento de lluvia** presenta una importancia significativamente menor, lo que sugiere que el tiempo que dura la precipitación no es un factor tan crítico como su intensidad o acumulación total. Asimismo, variables meteorológicas como la **temperatura del punto de rocío** y la **humedad relativa** tienen una influencia despreciable en la predicción del rendimiento del SuDS. En el caso de la humedad relativa, incluso se observa un valor ligeramente negativo, lo que indica que su inclusión podría estar añadiendo ruido en lugar de información relevante para el modelo.

Estos resultados refuerzan la importancia de priorizar mediciones de intensidad y acumulación de lluvia, así como de las condiciones iniciales del sistema, para mejorar la precisión de los modelos de predicción del SuDS. También sugieren que variables meteorológicas secundarias pueden ser descartadas sin afectar significativamente el desempeño del modelo, optimizando así la complejidad computacional y la interpretabilidad del análisis.

4.2.4. Evaluación de una Versión Simplificada del Modelo SVR

Con el objetivo de reducir la complejidad del modelo y facilitar su interpretación, se exploró la posibilidad de simplificar el modelo SVR mediante la eliminación de variables con contribuciones estadísticamente pequeñas: humedad relativa promedio ($-0,008 \pm 0,010$) y punto de rocío promedio ($0,008 \pm 0,013$). Esta propuesta se basa en el principio de parsimonia, que sugiere preferir modelos más simples siempre que mantengan un rendimiento comparable al de modelos más complejos.

La versión simplificada conservó únicamente cuatro variables predictoras: intensidad máxima de precipitación, lluvia acumulada, nivel inicial de agua y duración del evento. El modelo fue posteriormente ajustado mediante búsqueda bayesiana, obteniendo los siguientes hiperparámetros óptimos: *kernel* RBF, $C = 10,0$, $\gamma = 0,3048$ y $\varepsilon = 0,2$.

Desempeño e Importancia de Características

El modelo SVR simplificado fue evaluado para determinar su eficacia predictiva y la relevancia relativa de las variables conservadas. Los resultados de esta evaluación se presentan en la Tabla 4.4.

Modelo	R ²	MSE	RMSE	MAE
SVR Base (simplificado)	0.8347	4.3021	2.0741	1.4802
SVR Bayes (simplificado)	0.8800	3.1229	1.7672	1.3237

Tabla 4.4: Métricas de desempeño para el modelo SVR simplificado.

Como se observa en la tabla anterior, el desempeño mejoró sustancialmente tras la optimización con búsqueda bayesiana, elevando el coeficiente de determinación (R²) de 0.8347 a 0.8800, lo que representa un incremento del 5.4%. Simultáneamente, se logró una reducción del 27.4% en el error cuadrático medio (MSE), pasando de 4.3021 a 3.1229, y del 14.8% en el error cuadrático medio raíz (RMSE), disminuyendo de 2.0741 a 1.7672 cm. El error absoluto medio (MAE) también experimentó una mejora,

descendiendo a 1.3237 cm, lo que indica que, en promedio, las predicciones del modelo optimizado se desvían en aproximadamente 1.3 cm de los valores reales.

A través del método de *Permutation Feature Importance*, se analizó la contribución relativa de cada variable al poder predictivo del modelo. Este análisis reveló que la lluvia acumulada constituye el factor más determinante, con una importancia de 0.552 ± 0.120 , seguida muy de cerca por el nivel inicial de agua (0.500 ± 0.129). Ambas variables ejercen una influencia sustancialmente mayor que la intensidad máxima de precipitación (0.442 ± 0.095) y especialmente que la duración del evento (0.075 ± 0.024), cuya contribución resultó comparativamente marginal. Estos resultados confirman la intuición física: el comportamiento hidráulico del sistema está primordialmente controlado por el volumen total de agua recibida y las condiciones iniciales del sistema, mientras que la duración por sí sola tiene un impacto limitado en la predicción de los niveles de lámina de agua.

Cabe destacar que, si bien las métricas generales reflejan un desempeño aparentemente satisfactorio del modelo simplificado, como se discutirá más adelante, este rendimiento no es homogéneo en todos los rangos de predicción, presentando deficiencias particularmente notables en los valores bajos del sensor que corresponden a niveles críticos de acumulación de agua.

Evaluación de Predicciones

Las predicciones generadas por el modelo simplificado se evaluaron gráficamente comparando los valores predichos frente a los valores reales. La Figura 4.11 muestra el comportamiento de los modelos SVR simplificados (base y optimizado con búsqueda bayesiana).

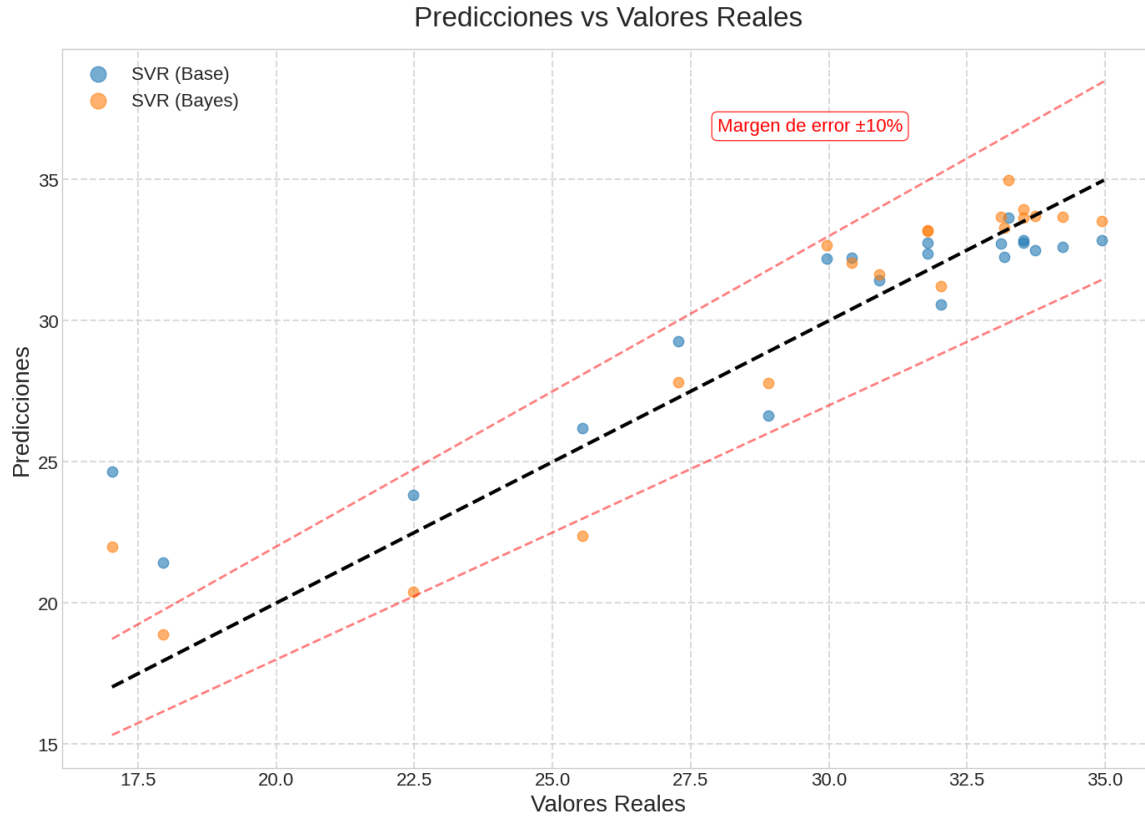


Figura 4.11: Predicciones vs valores reales para los modelos SVR simplificados (Base y Bayes).

Limitaciones ante Escenarios de Cambio Climático

A pesar del desempeño general aceptable, se identificó una limitación crítica del modelo simplificado: su bajo rendimiento en los valores bajos del sensor, los cuales corresponden a niveles elevados de agua acumulada en el sistema SuDS. Estos valores son particularmente importantes para el estudio, dado que representan escenarios extremos de precipitación, directamente relacionados con los efectos del cambio climático.

La comparación cuantitativa entre el modelo original y el simplificado se presenta en la Tabla 4.5. Aunque la reducción en R^2 fue de apenas 0.0224, el análisis gráfico reveló que esta diferencia se concentra en los valores bajos del sensor, lo que compromete la utilidad del modelo simplificado en situaciones críticas.

Métrica	SVR Bayes Original	SVR Bayes Simplificado	Diferencia
R^2 (prueba)	0.9024	0.8800	-0.0224
MSE	2.5399	3.1229	+0.5830
RMSE	1.5940	1.7672	+0.1732
MAE	1.3610	1.3237	-0.0373

Tabla 4.5: Comparación entre el modelo SVR original y el modelo simplificado.

4.2.4.1. Conclusión: Prevalencia del Modelo Original

La comparación entre ambos modelos (Tabla 4.5) sugiere que, aunque la pérdida de precisión general del modelo simplificado es moderada (reducción de R^2 de apenas 0.0224), este valor promedio esconde deficiencias importantes en ciertos rangos críticos. El análisis visual (Figuras 4.5 y 4.11) demuestra que el modelo simplificado no logra predecir adecuadamente los valores bajos del sensor, donde el modelo SVR Bayes original mantiene un mejor ajuste.

Por tanto, se determina mantener el modelo SVR Bayes original (con las seis variables, incluyendo humedad relativa y punto de rocío) como solución definitiva para este estudio. Este modelo, con R^2 de 0.9024 y RMSE de 1.5940, proporciona mayor confiabilidad en los escenarios críticos asociados a altos niveles de agua, lo cual es esencial para evaluar la respuesta del SuDS bajo condiciones de precipitación intensa relacionadas con el cambio climático.

Esta decisión resalta la importancia de evaluar los modelos no solo con base en sus métricas globales, sino también en su capacidad para representar correctamente los rangos de mayor relevancia para los objetivos específicos de la investigación.

4.3. Modelado del Nivel de Salida del SuDS mediante Aprendizaje Automático

4.3.1. Entrenamiento y Evaluación de Modelos para el Nivel de Salida

4.3.1.1. Modelos Base

De manera similar a lo realizado para los niveles de entrada, se entrenaron cinco modelos base para predecir los niveles de salida del SuDS: regresión lineal, ElasticNet, Random Forest, XGBoost y Support Vector Regression (SVR). Cada modelo se configuró con parámetros predeterminados para establecer un punto de referencia inicial.

Durante esta fase, se experimentó con diversas combinaciones de variables predictoras, incluyendo diferentes transformaciones de las variables originales (como logaritmos, normalizaciones y estandarizaciones), variables rezagadas temporalmente para capturar efectos diferidos, e incluso características derivadas como tasas de cambio e interacciones entre variables. A pesar de estos esfuerzos, los resultados obtenidos fueron consistentemente insatisfactorios.

Las métricas de evaluación calculadas en el conjunto de prueba y mediante validación cruzada muestran resultados considerablemente más bajos que los obtenidos para los niveles de entrada. Como se observa en la Tabla 4.6, ninguno de los modelos base logró un desempeño aceptable, independientemente de las transformaciones y selecciones de características implementadas.

Los resultados de la validación cruzada para estos modelos base mostraron una alta variabilidad entre los diferentes pliegues, con valores negativos de R^2 en varios casos, lo que indica que los modelos tienen dificultades severas para generalizar adecuadamente. Se intentó mitigar este problema mediante diversas técnicas de regularización y submuestreo, pero los resultados no mejoraron significativamente.

Modelo	R ²	MSE	RMSE	MAE
Regresión Lineal	0.2146	47.9417	6.9240	6.0562
ElasticNet	0.2462	46.0123	6.7832	5.9340
Random Forest	0.2789	44.0169	6.6345	5.3507
XGBoost	0.0981	55.0498	7.4196	5.7457
SVR	-0.0842	66.1815	8.1352	5.2880

Tabla 4.6: Métricas de evaluación para modelos base en la predicción de niveles de salida.

4.3.1.2. Modelos Optimizados por Bayes

A pesar de los bajos resultados iniciales, se procedió con la optimización bayesiana para los mismos modelos (excluyendo la regresión lineal), con la esperanza de mejorar su desempeño. Al igual que en el caso de los niveles de entrada, se utilizó BayesSearchCV para la búsqueda eficiente de hiperparámetros óptimos.

Se probaron espacios de búsqueda extensos para los hiperparámetros, incluyendo configuraciones que favorecían tanto modelos más simples como más complejos. Adicionalmente, se experimentó con técnicas de ensamblaje combinando múltiples modelos y con métodos de selección automática de características. Sin embargo, ninguna de estas aproximaciones logró mejorar sustancialmente el rendimiento predictivo.

Los resultados de la optimización, presentados en la Tabla 4.7, muestran mejoras marginales pero insuficientes para considerar los modelos como adecuados para la predicción de niveles de salida. El mejor modelo (XGBoost optimizado por Bayes) apenas alcanzó un R² de 0.2523, lo que indica que más del 75 % de la variabilidad en los niveles de salida no puede ser explicada por las variables consideradas ni por las técnicas de modelado implementadas.

Modelo	R ²	MSE	RMSE	MAE
ElasticNet (Bayes)	0.2300	47.0014	6.8558	6.0923
Random Forest (Bayes)	0.2421	46.2641	6.8018	5.5329
XGBoost (Bayes)	0.2523	45.6389	6.7557	5.5130
SVR (Bayes)	-0.0017	61.1443	7.8195	5.3868

Tabla 4.7: Métricas de evaluación para modelos optimizados en la predicción de niveles de salida.

Estos resultados consistentemente deficientes, a pesar de los múltiples enfoques y técnicas aplicadas, sugieren que existen limitaciones fundamentales en los datos disponibles o en la naturaleza misma del problema que impiden una modelación efectiva de los niveles de salida con las variables y eventos registrados actualmente.

4.4. Justificación del Enfoque Predictivo en el Nivel de Entrada

4.4.1. Limitaciones en la Modelación del Nivel de Salida en SuDS

Los resultados obtenidos para la predicción de niveles de salida del SuDS muestran un desempeño significativamente inferior en comparación con los modelos para niveles de entrada. Este bajo rendimiento puede atribuirse a varios factores fundamentales:

Limitaciones temporales del conjunto de datos

Una de las principales limitaciones identificadas es el rango temporal de los eventos registrados en nuestro conjunto de datos. Los eventos analizados no superan las 11 horas de duración, mientras que existe documentación científica que indica que este tipo de sistemas SuDS pueden tener tiempos de respuesta en la salida que alcanzan hasta 48 horas (aproximadamente 2 días). Esta discrepancia temporal implica que nuestros modelos no pueden capturar adecuadamente la dinámica completa del sistema, ya que las mediciones disponibles no abarcan todo el proceso de retención y liberación del agua.

Variables no consideradas en el modelo

Existen diversos factores que influyen en el comportamiento hidráulico de un SuDS y que no han sido incluidos en nuestros modelos actuales:

- **Evapotranspiración:** La pérdida de agua a través de la evaporación directa y la transpiración de las plantas presentes en el sistema SuDS no ha sido cuantificada ni incluida como variable predictora.
- **Capacidad de retención del material granular:** Las propiedades específicas de los materiales utilizados en el SuDS, como la porosidad, permeabilidad y capacidad de adsorción, varían con el tiempo y pueden influir significativamente en el comportamiento hidráulico del sistema.
- **Influencia de la vegetación:** El desarrollo y estado de la vegetación presente en el SuDS puede modificar los patrones de flujo y retención de agua, especialmente a medida que las plantas crecen y sus sistemas radiculares evolucionan.
- **Efectos de colmatación y mantenimiento:** Los procesos graduales de acumulación de sedimentos y materia orgánica pueden alterar la conductividad hidráulica del sistema a lo largo del tiempo.

Complejidad de los procesos hidráulicos

Los procesos que gobiernan el flujo de salida en un SuDS son considerablemente más complejos que los que determinan el nivel de entrada. Mientras que la entrada responde más directamente a factores climáticos inmediatos, la salida es el resultado de múltiples procesos físicos que ocurren en diferentes escalas temporales y espaciales dentro del sistema.

4.4.2. Decisión Metodológica sobre el Alcance de la Modelación

Dados los resultados obtenidos y las limitaciones identificadas, se ha tomado la decisión de enfocar el estudio exclusivamente en la predicción de los niveles de entrada del SuDS. Esta decisión se justifica por:

- El mejor desempeño de los modelos predictivos para niveles de entrada (con valores de R^2 superiores a 0.85 para los mejores modelos).
- La mayor aplicabilidad práctica de las predicciones de niveles de entrada para alertas tempranas y gestión operativa de los sistemas.
- La necesidad de incorporar un monitoreo a más largo plazo y variables adicionales para modelar adecuadamente los niveles de salida en futuras investigaciones.

Para futuras investigaciones que busquen modelar los niveles de salida, será necesario implementar campañas de monitoreo de mayor duración, incorporar sensores adicionales para medir variables como la humedad del suelo, la evapotranspiración y las propiedades del medio filtrante, así como desarrollar modelos específicos que puedan capturar adecuadamente los complejos procesos hidráulicos que ocurren en estos sistemas.

Desarrollo y Caracterización de Escenarios de Cambio Climático

5.1. Elaboración de Escenarios de Cambio Climático Específicos para el SuDS

Tras desarrollar y validar el modelo de predicción basado en Support Vector Regression (SVR) y haber seleccionado las variables más significativas para el funcionamiento del SuDS, en esta sección se definen los escenarios de cambio climático que servirán como datos de entrada para el modelo predictivo. Estos escenarios permitirán evaluar el comportamiento futuro del nivel de entrada del SuDS bajo diferentes condiciones climáticas proyectadas, proporcionando una base sólida para la toma de decisiones de adaptación y gestión del sistema a largo plazo. Al aplicar estos escenarios en el modelo, se espera anticipar posibles desafíos operativos y optimizar el rendimiento del sistema frente a las cambiantes condiciones climáticas.

5.1.1. Metodología para la Construcción de los Escenarios

El presente análisis se basa en las proyecciones climáticas establecidas en el documento técnico “Escenarios de Cambio Climático de la Cuarta Comunicación Nacional de Colombia” [29] elaborado por el IDEAM y el Ministerio de Ambiente. Este informe presenta proyecciones climáticas para el período 2021-2100 bajo cuatro escenarios de Trayectorias Socioeconómicas Compartidas (SSP): SSP1-2.6, SSP2-4.5, SSP3-7.0 y SSP5-8.5, con análisis para los períodos 2021-2040, 2041-2060, 2061-2080 y 2081-2100.

La metodología aplicada por el IDEAM [29] involucró la descarga, evaluación y reducción de escala estadística de datos climáticos, con una resolución espacial de hasta 10×10 km para precipitación. Los datos fueron ajustados utilizando información histórica de estaciones hidrometeorológicas del IDEAM, proporcionando así proyecciones robustas que pueden aplicarse a contextos locales específicos como el campus de la Pontificia Universidad Javeriana en Bogotá.

5.1.2. Desarrollo de los Escenarios de Precipitación para el SuDS

Para la construcción de escenarios de precipitación futura específicos para el Sistema de Humedal Artificial/Tanque de Reservorio (SHATR) de la Universidad Javeriana, se ha adoptado un enfoque basado en la aplicación de los porcentajes de cambio derivados del análisis realizado en el estudio del IDEAM a los datos observados en la estación meteorológica ubicada en el campus universitario. Este método, conocido como “perturbación del clima” o “delta change method”, es ampliamente utilizado

en estudios de impacto climático debido a su capacidad para generar escenarios localmente relevantes, manteniendo al mismo tiempo la coherencia con las proyecciones regionales [30].

Considerando que la precipitación es la variable más crítica para el funcionamiento del SHATR, dada su influencia directa en los volúmenes de escorrentía que debe gestionar el sistema, los escenarios específicos se construyen como se define a continuación.

5.1.2.1. Escenario 1: Horizonte temporal 2021-2040

Este escenario corresponde al corto plazo y contempla un aumento del 20 % en la precipitación para los meses de marzo, mayo, junio y octubre. Se aplica una reducción del 20 % para enero y se mantienen las condiciones actuales para los meses restantes. Este escenario es particularmente relevante para la planificación inmediata y adaptaciones de corto plazo del SHATR.

5.1.2.2. Escenario 2: Horizonte temporal 2041-2060

Para el mediano plazo, se considera un incremento del 30 % en la precipitación durante los meses de marzo, noviembre y diciembre. Se aplica una reducción del 30 % para julio y agosto, mientras los valores para los meses restantes se mantienen sin cambios. Este escenario permitirá evaluar la capacidad de respuesta del SHATR ante un régimen de precipitación más contrastado, con períodos húmedos más intensos.

5.1.2.3. Escenario 3: Horizonte temporal 2061-2080

En este horizonte temporal de largo plazo, se proyecta un aumento del 30 % para los meses de marzo, mayo y junio, y un incremento del 40 % para octubre. Las reducciones del 30 % se aplican para julio y agosto. Este escenario es crítico para evaluar la sostenibilidad a largo plazo del SHATR y las posibles modificaciones estructurales que podrían requerirse.

5.1.2.4. Escenario 4: Horizonte temporal 2081-2100

Para finales de siglo, se contempla un aumento del 30 % en la precipitación para los meses de marzo, abril, mayo, junio, noviembre y diciembre, y un incremento mayor (40 %) para octubre. Este escenario refleja la intensificación de los patrones de precipitación proyectados para la región Andina hacia finales del siglo XXI y permitirá evaluar escenarios extremos para el diseño de futuras infraestructuras de drenaje en el campus.

5.1.3. Proyección de las Variables Hidrometeorológicas Bajo los Escenarios

Para implementar los escenarios de cambio climático en el modelo predictivo del SHATR, se requiere proyectar las variables hidrometeorológicas relevantes que conforman el conjunto de datos de entrada del modelo. La metodología de proyección para cada variable se describe a continuación:

5.1.3.1. Lluvia acumulada (lluvia_acumulada_mm)

Esta variable se modifica directamente aplicando los porcentajes de cambio establecidos en los cuatro escenarios temporales según el mes correspondiente. La transformación se realiza mediante un factor multiplicativo que representa el cambio porcentual proyectado:

$$P_{fut} = P_{hist} \times (1 + \Delta P_{m,e}) \quad (5.1)$$

Donde P_{fut} es la precipitación futura proyectada, P_{hist} es la precipitación histórica observada, y $\Delta P_{m,e}$ es el cambio porcentual (expresado como fracción) para el mes m bajo el escenario e .

5.1.3.2. Duración de eventos de lluvia (duracion_lluvia_min)

La duración de los eventos de lluvia se ve afectada por el cambio climático, aunque no necesariamente en la misma proporción que los cambios en volumen de precipitación. Diversos estudios [31, 32, 33, 34] sugieren que el calentamiento global tiende a intensificar el ciclo hidrológico, resultando en eventos más cortos pero más intensos. Para modelar los cambios en la duración de eventos, se propone una relación no lineal con los cambios en precipitación:

$$D_{fut} = D_{hist} \times (1 + \alpha \cdot \Delta P_{m,e}) \quad (5.2)$$

Donde D_{fut} es la duración futura proyectada del evento de lluvia, D_{hist} es la duración histórica observada, y α es un factor de proporcionalidad que varía según la dirección del cambio en precipitación:

- $\alpha = -0,3$ para incrementos en precipitación (eventos más cortos)
- $\alpha = 0,5$ para reducciones en precipitación

Esta parametrización se basa en la evidencia de que los eventos de precipitación en un clima más cálido tienden a ser más intensos pero de menor duración [33], reflejando la concentración de mayores volúmenes de lluvia en períodos más cortos de tiempo. Para compensar esta reducción en la duración, se aplica un aumento proporcional mayor en la intensidad de los eventos.

5.1.3.3. Intensidad máxima de precipitación (intensidad_maxima_mmph)

La intensidad máxima de precipitación es particularmente sensible al cambio climático, con evidencia que sugiere aumentos más que proporcionales en las intensidades máximas en comparación con los cambios en precipitación media [35]. Para modelar este comportamiento, se propone:

$$I_{fut} = I_{hist} \times (1 + \beta \cdot \Delta P_{m,e}) \quad (5.3)$$

Donde I_{fut} es la intensidad máxima futura proyectada, I_{hist} es la intensidad máxima histórica observada, y β es un factor de amplificación que refleja la respuesta no lineal de la intensidad:

- $\beta = 1,3$ para incrementos en precipitación
- $\beta = 1,0$ para reducciones en precipitación

Este enfoque reconoce que los eventos extremos tienden a intensificarse más rápidamente que los cambios en las medias, un fenómeno conocido como “clausius-clapeyron scaling” [35].

5.1.3.4. Nivel inicial del tanque de sedimentación (input_inicial_cm)

Esta variable representa las condiciones iniciales del sistema y está influenciada por las condiciones antecedentes. Es importante destacar que valores menores de `input_inicial_cm` indican que el tanque está más lleno, mientras que valores mayores indican menor nivel de agua. Para su proyección, se propone un enfoque que considere la memoria hidrológica del sistema:

$$H_{i,fut} = H_{i,hist} \times \gamma_{m,e} \quad (5.4)$$

Donde $H_{i,fut}$ es el nivel inicial futuro proyectado, $H_{i,hist}$ es el nivel inicial histórico, y $\gamma_{m,e}$ es un factor que refleja las condiciones antecedentes para el mes m bajo el escenario e :

- Para meses precedidos por meses con aumento de precipitación: $\gamma_{m,e} = 0,85$ a $0,90$ (valores menores que indican tanque más lleno)
- Para meses precedidos por meses con reducción de precipitación: $\gamma_{m,e} = 1,05$ a $1,10$ (valores mayores que indican tanque menos lleno)
- Para meses en transición entre temporadas: $\gamma_{m,e} = 1,00$ (sin cambio)

Este enfoque reconoce la dependencia temporal en el funcionamiento del sistema y permite modelar el efecto acumulativo de los cambios en los patrones de precipitación, teniendo en cuenta la relación inversa entre el valor del sensor y el nivel de agua en el tanque. Los valores específicos de $\gamma_{m,e}$ se calculan de manera proporcional a la magnitud del cambio en precipitación del mes anterior.

5.1.3.5. Humedad relativa (rh_promedio)

La humedad relativa presenta una tendencia clara de reducción en la mayor parte de Colombia según los escenarios climáticos del IDEAM [29]. Para la región Andina, donde se encuentra Bogotá, se proyectan reducciones generalizadas que se intensifican a lo largo del siglo XXI.

Para modelar estos cambios, se propone un ajuste basado en el escenario temporal, con modificaciones sutiles según los cambios en precipitación local:

$$RH_{fut} = RH_{hist} + \Delta RH_{base,e} + \Delta RH_{precip,m,e} \quad (5.5)$$

Donde $\Delta RH_{base,e}$ es el cambio base en humedad relativa para el escenario temporal e :

- Escenario 1 (2021-2040): $\Delta RH_{base,1} = -1,0\%$
- Escenario 2 (2041-2060): $\Delta RH_{base,2} = -2,0\%$
- Escenario 3 (2061-2080): $\Delta RH_{base,3} = -3,0\%$
- Escenario 4 (2081-2100): $\Delta RH_{base,4} = -4,0\%$

Y $\Delta RH_{precip,m,e}$ es un ajuste adicional que refleja la interacción local entre humedad y precipitación, calculado como función del cambio en precipitación $\Delta P_{m,e}$:

$$\Delta RH_{precip,m,e} = \begin{cases} \text{mín}(0,5, \Delta P_{m,e} \times 0,25) & \text{si } \Delta P_{m,e} > 0 \\ \text{máx}(-0,5, \Delta P_{m,e} \times 0,25) & \text{si } \Delta P_{m,e} \leq 0 \end{cases} \quad (5.6)$$

Este enfoque reconoce la tendencia general de reducción de la humedad relativa en la región Andina según las proyecciones del IDEAM, permitiendo variaciones locales moderadas que responden a los cambios en los patrones de precipitación.

5.1.3.6. Punto de rocío (dew_point_promedio_f)

El punto de rocío está directamente relacionado con la humedad absoluta y tenderá a aumentar en un clima más cálido [36]. Para proyectar los cambios en el punto de rocío, se propone un enfoque que considera tanto el calentamiento global base como los efectos adicionales de los cambios en precipitación:

$$DP_{fut} = DP_{hist} + \Delta DP_{base,e} + \Delta DP_{precip,m,e} \quad (5.7)$$

Donde $\Delta DP_{base,e}$ es el aumento base en el punto de rocío debido al calentamiento global bajo el escenario e :

- Escenario 1 (2021-2040): $\Delta DP_{base,1} = +1^\circ\text{F}$
- Escenario 2 (2041-2060): $\Delta DP_{base,2} = +2^\circ\text{F}$
- Escenario 3 (2061-2080): $\Delta DP_{base,3} = +3^\circ\text{F}$
- Escenario 4 (2081-2100): $\Delta DP_{base,4} = +4^\circ\text{F}$

Y $\Delta DP_{precip,m,e}$ es el efecto adicional debido a cambios en precipitación para el mes m bajo el escenario e :

Para meses con aumento de precipitación:

- Escenario 1: $\Delta DP_{precip,m,1} = +0,3^\circ\text{F}$
- Escenario 2: $\Delta DP_{precip,m,2} = +0,5^\circ\text{F}$
- Escenario 3: $\Delta DP_{precip,m,3} = +0,7^\circ\text{F}$
- Escenario 4: $\Delta DP_{precip,m,4} = +1,0^\circ\text{F}$

Para meses con reducción de precipitación: $\Delta DP_{precip,m,e} = 0^\circ\text{F}$

5.1.4. Implementación de la Metodología de Escenarios en Python

La aplicación práctica de los escenarios de cambio climático se ha implementado en Python mediante la función `apply_climate_scenario`, que permite transformar un conjunto de datos históricos según las proyecciones de cada escenario. Esta implementación sigue fielmente la metodología descrita, aplicando los factores de cambio a cada variable hidrometeorológica de manera consistente con las ecuaciones presentadas.

A continuación se describe la estructura general del proceso:

1. Definición de diccionarios para los cambios en precipitación y las variables meteorológicas, por escenario y mes
2. Procesamiento de cada registro en el conjunto de datos, aplicando las transformaciones correspondientes según el mes y escenario
3. Escalamiento de los cambios en humedad relativa y nivel inicial del tanque de acuerdo con la magnitud del cambio en precipitación
4. Generación de cuatro conjuntos de datos transformados, uno para cada horizonte temporal

El código completo de este desarrollo se puede consultar en el cuaderno de Colab cuyo enlace se encuentra en los anexos.

5.1.5. Análisis Estadístico Comparativo de los Escenarios Proyectados

Para caracterizar cuantitativamente los efectos de los escenarios de cambio climático en las variables clave del sistema, se realizó un análisis estadístico comparativo entre los datos históricos (escenario base) y las proyecciones para cada horizonte temporal. La Tabla 5.1 resume las principales métricas estadísticas de las variables hidrometeorológicas en cada escenario.

Los resultados revelan cambios significativos en todas las variables analizadas:

- **Lluvia acumulada (`lluvia_acumulada_mm`):** Se observa un aumento progresivo en los valores medios, desde 4.02 mm (Escenario 1) hasta 4.69 mm (Escenario 4), comparado con el valor base de 3.72 mm. La máxima precipitación registrada aumenta un 30 % en los Escenarios 2, 3 y 4 (37.31 mm vs 28.70 mm base).
- **Duración de lluvia (`duracion_lluvia_min`):** A diferencia de lo esperado, los valores medios muestran una tendencia decreciente a través de los escenarios, desde 128.60 min (Escenario 1) hasta 122.82 min (Escenario 4), comparado con 131.88 min del escenario base. Los valores máximos también disminuyen, pasando de 532.00 min en el escenario base a aproximadamente 484.12 min en los Escenarios 2, 3 y 4.
- **Intensidad máxima (`intensidad_maxima_mmph`):** Presenta los cambios más pronunciados, con aumentos del 32 % en la media (12.50 vs 9.46 mm/h base) y 39 % en los máximos (84.73 vs 60.96 mm/h base) para el Escenario 4.

- **Nivel inicial (input_inicial_cm):** Los valores medios fluctúan entre los escenarios, con 32.87 cm (Escenario 1), 34.88 cm (Escenario 2), 32.82 cm (Escenario 3) y 31.78 cm (Escenario 4), comparado con 35.09 cm del escenario base, indicando generalmente condiciones iniciales de tanque más lleno en los escenarios futuros.
- **Humedad relativa (rh_promedio):** Muestra una tendencia decreciente consistente a lo largo de los escenarios, pasando de 86.36% en el escenario base a 85.38% (Escenario 1), 84.37% (Escenario 2), 83.38% (Escenario 3) y 82.42% (Escenario 4), reflejando las proyecciones del IDEAM para la región Andina.
- **Punto de rocío (dew_point_promedio_f):** Presenta un aumento progresivo y consistente desde 51.61°F (base) hasta 56.32°F (Escenario 4), con incrementos intermedios de 52.74°F (Escenario 1), 53.74°F (Escenario 2) y 54.91°F (Escenario 3), confirmando la tendencia hacia un clima más cálido y húmedo en términos absolutos.

Variable	Escenario	Media	Desv.	Mín	25 %	50 %	75 %	Máx
Lluvia acumulada (mm)	Base	3.72	4.82	0.76	1.02	2.03	3.75	28.70
	Esc 1	4.02	5.29	0.76	1.22	2.08	3.96	34.44
	Esc 2	4.18	6.10	0.53	1.02	2.03	3.96	37.31
	Esc 3	4.14	5.59	0.53	1.25	2.03	4.06	37.31
	Esc 4	4.69	6.33	0.76	1.32	2.31	4.78	37.31
Duración de lluvia (min)	Base	131.88	110.87	4.00	45.00	95.00	197.50	532.00
	Esc 1	128.60	107.66	3.76	42.30	92.15	189.97	500.08
	Esc 2	126.05	105.43	4.00	43.51	93.46	190.51	484.12
	Esc 3	124.23	104.47	3.52	41.59	89.64	188.46	484.12
	Esc 4	122.82	103.04	3.52	40.95	89.96	179.73	484.12
Intensidad máxima (mm/h)	Base	9.46	10.43	3.05	3.05	6.10	11.43	60.96
	Esc 1	10.30	10.90	3.05	3.84	6.10	12.02	60.96
	Esc 2	11.03	14.43	2.13	3.05	6.10	12.58	84.73
	Esc 3	10.73	11.46	2.13	4.24	6.10	12.71	60.96
	Esc 4	12.50	14.50	3.05	4.24	8.47	13.60	84.73
Nivel inicial (cm)	Base	35.09	5.83	18.57	35.26	37.73	38.26	43.31
	Esc 1	32.87	5.70	18.57	31.74	33.21	37.78	46.92
	Esc 2	34.88	6.16	18.57	32.60	37.49	38.33	45.86
	Esc 3	32.82	6.34	18.57	31.24	32.70	37.90	45.86
	Esc 4	31.78	5.71	16.02	30.89	32.61	33.49	43.31
Humedad relativa (%)	Base	86.36	8.78	61.05	81.86	87.79	93.40	98.69
	Esc 1	85.38	8.78	60.05	80.89	86.81	92.45	97.69
	Esc 2	84.37	8.79	59.05	79.88	85.82	91.40	96.77
	Esc 3	83.38	8.79	58.05	78.91	84.84	90.49	95.69
	Esc 4	82.42	8.79	57.05	77.93	83.87	89.49	94.77
Punto de rocío (°F)	Base	51.61	1.73	46.41	50.69	51.68	52.83	55.17
	Esc 1	52.74	1.76	47.41	51.74	52.84	53.99	56.47
	Esc 2	53.74	1.80	48.41	52.69	53.74	54.91	57.17
	Esc 3	54.91	1.83	49.41	53.74	55.01	56.26	58.87
	Esc 4	56.32	1.97	50.41	55.02	56.46	57.75	60.17

Tabla 5.1: Estadísticas descriptivas comparativas por escenario

Los análisis estadísticos confirman cinco tendencias clave en la respuesta hidrometeorológica al

cambio climático proyectado:

1. **Intensificación de extremos con eventos más cortos:** El aumento en los valores máximos de intensidad de lluvia (incrementos de hasta +39 % en el Escenario 4) junto con la reducción en la duración media de los eventos (-6.9 % en el Escenario 4) confirma la hipótesis de eventos más cortos pero más intensos bajo condiciones de calentamiento global. Este fenómeno es consistente con el principio de Clausius-Clapeyron que predice incrementos exponenciales en intensidades extremas bajo calentamiento global.
2. **Incremento de variabilidad:** El aumento generalizado en las desviaciones estándar (especialmente notable en lluvia acumulada: +31 % en Escenario 4 vs base, y en intensidad máxima: +39 %) sugiere una mayor irregularidad en los patrones de precipitación, lo que representa un desafío adicional para el sistema de drenaje.
3. **Cambio en condiciones iniciales:** La reducción progresiva en los valores medios de `input_inicial_cm` (de 35.09 cm a 31.78 cm en Escenario 4, equivalente a una reducción del 9.4 %) indica una mayor frecuencia de condiciones iniciales de tanque más lleno, reduciendo su capacidad de amortiguación inicial frente a eventos extremos.
4. **Alteración del ciclo hidrológico:** La combinación de eventos más cortos pero con mayor acumulación total (+26.1 % en el Escenario 4) e intensidad máxima (+32 %) sugiere una modificación fundamental en la estructura temporal de los eventos de precipitación, con lluvias más concentradas e intensas que pueden sobrepasar la capacidad de infiltración y retención del sistema.
5. **Modificación del régimen higrótico:** El aumento progresivo del punto de rocío (de 51.61°F a 56.32°F, +4.71°F o +9.1 %) acompañado de una reducción sistemática en la humedad relativa (disminución de 3.94 % en el Escenario 4) confirma un ambiente atmosférico más cálido con mayor contenido absoluto de humedad pero menor humedad relativa, consistente con las proyecciones del IDEAM para la región Andina.

Estos resultados cuantitativos validan la efectividad de la metodología de perturbación climática aplicada, demostrando su capacidad para generar escenarios que mantienen coherencia estadística con las proyecciones climáticas regionales mientras introducen cambios significativos en las variables críticas para la operación del SuDS. Particularmente notable es la consistencia entre la reducción en humedad relativa y el aumento en punto de rocío, reflejando el calentamiento atmosférico proyectado para la región.

La progresión temporal entre escenarios también muestra patrones coherentes con la literatura científica sobre cambio climático, con intensificación más pronunciada en los horizontes temporales más distantes (Escenario 4, 2081-2100), mientras que los cambios en escenarios cercanos (Escenario 1, 2021-2040) son más moderados, reflejando la inercia del sistema climático y la incertidumbre inherente a las proyecciones de corto plazo.

5.1.6. Limitaciones de la Metodología de Construcción de Escenarios

Es importante reconocer las limitaciones inherentes a este enfoque metodológico para contextualizar adecuadamente los resultados obtenidos:

- **Propagación de incertidumbres climáticas:** Las proyecciones climáticas regionales contienen incertidumbres sustanciales derivadas de múltiples fuentes: variabilidad natural del sistema climático, simplificaciones en los modelos de circulación global, escenarios de emisiones futuros, y métodos de regionalización estadística o dinámica [29]. Estas incertidumbres se amplifican al descender a escala local, particularmente en regiones con topografía compleja como la Sabana de Bogotá, donde efectos orográficos pueden modular significativamente las señales de cambio climático.
- **Complejidad en relaciones hidrometeorológicas:** Las interacciones entre variables hidrometeorológicas pueden evolucionar de manera no lineal bajo condiciones de cambio climático [37]. Los parámetros de proporcionalidad (α , β , γ) utilizados en las fórmulas de proyección representan aproximaciones basadas en el conocimiento actual, pero podrían requerir ajustes a medida que avanza la comprensión científica de estos procesos. Particularmente críticas son las relaciones entre intensidad y duración de precipitación, que pueden exhibir comportamientos umbrales no adecuadamente capturados por relaciones paramétricas simples.
- **Efectos microclimáticos urbanos:** Los cambios en el entorno urbano circundante al SHATR, incluyendo modificaciones en albedo superficial, impermeabilización, isla de calor urbana y patrones de ventilación local, podrían alterar los patrones hidrometeorológicos de manera no capturada por las proyecciones regionales [38]. La intensificación del efecto de isla de calor urbana podría, por ejemplo, exacerbar la intensidad de eventos convectivos locales más allá de lo proyectado por modelos regionales, particularmente en entornos urbanos densos como el campus universitario.
- **Comportamiento no lineal del sistema:** La respuesta hidráulica e hidrológica del SHATR puede presentar umbrales críticos y comportamientos no lineales que requieren validación continua del modelo predictivo [39]. Los sistemas de drenaje sostenible a menudo exhiben puntos de inflexión en su respuesta, donde pequeñas modificaciones en las entradas pueden producir cambios desproporcionados en la respuesta del sistema, especialmente cuando se aproximan a sus límites de capacidad o bajo condiciones de saturación prolongada de los componentes vegetales.
- **Interacciones ecohidrológicas:** Los componentes vegetales del SHATR evolucionarán también en respuesta al cambio climático, modificando sus características de evapotranspiración, capacidad de interceptación y resistencia a estrés hídrico [40]. Estas adaptaciones biológicas pueden alterar el comportamiento hidrológico del sistema de formas no contempladas en el modelo actual, requiriendo un enfoque más integrado que incorpore la dinámica ecohidrológica bajo condiciones climáticas cambiantes.

Los escenarios generados, a pesar de estas limitaciones, servirán como insumo fundamental para el modelo predictivo, permitiendo simular el comportamiento del sistema bajo un rango plausible de condiciones futuras y orientando las decisiones de gestión adaptativa. Este enfoque posibilita la identificación de vulnerabilidades potenciales y umbrales críticos de operación bajo diferentes trayectorias climáticas.

La implementación de un programa de monitoreo continuo del SHATR permitirá refinar progresivamente tanto los escenarios como el modelo predictivo, reduciendo la incertidumbre a través de la

incorporación de datos observacionales. Este ciclo iterativo de modelación-observación-ajuste no solo contribuirá a mejorar la resiliencia del SHATR ante el cambio climático, sino que también posicionará a la Universidad Javeriana como un referente metodológico en la implementación y seguimiento de soluciones basadas en la naturaleza adaptadas a las condiciones climáticas cambiantes del contexto urbano andino tropical, con potencial de replicabilidad en otros entornos urbanos de la región.

Predicción del Comportamiento Futuro del SuDS Bajo Escenarios de Cambio Climático

6.1. Aplicación del Modelo Predictivo a los Escenarios de Cambio Climático

En esta sección se presenta un análisis predictivo del comportamiento futuro del SuDS bajo diferentes escenarios climáticos. La metodología implementada se basa en técnicas de aprendizaje automático, específicamente mediante el uso de Máquinas de Vectores de Soporte para Regresión (SVR), para modelar y predecir el nivel de agua en estos sistemas.

Para revisar el código completo utilizado en este análisis, incluyendo la implementación del modelo SVR, la predicción de niveles para los diferentes escenarios y los análisis estadísticos, se puede consultar el cuaderno de Google Colab de predicciones disponible en los anexos de este documento.

6.1.1. Resultados de la Predicción del Modelo SVR

El modelo de Máquinas de Vectores de Soporte (SVR) entrenado con los parámetros óptimos ($C = 4,89$, $\varepsilon = 0,166$, $\gamma = 0,051$, kernel='rbf') fue aplicado a los datos de los diferentes escenarios climáticos futuros. Los resultados obtenidos permiten analizar el comportamiento proyectado del SuDS bajo distintas condiciones climáticas y evaluar su capacidad para gestionar tanto eventos extremos de inundación como periodos de sequía.

6.1.1.1. Análisis Estadístico Comparativo

La Tabla 6.1 presenta un resumen de las estadísticas descriptivas del nivel de agua (input_min_cm) para los datos reales y los cuatro escenarios proyectados.

Escenario	Media	Desv. Est.	Mín.	25 %	Mediana	75 %	Máx.
Datos Reales	28.52	6.21	15.40	23.86	31.38	33.31	35.77
Escenario 1	27.15	5.20	14.56	23.37	29.26	30.82	35.23
Escenario 2	28.06	5.52	15.38	23.84	29.85	32.77	35.42
Escenario 3	26.79	4.80	14.63	23.39	28.40	30.24	35.16
Escenario 4	26.28	4.06	16.78	23.21	27.31	29.37	34.54

Tabla 6.1: Estadísticas descriptivas del nivel de agua por escenario

Un menor valor de `input_min_cm` indica un mayor nivel de agua, lo que puede aumentar el riesgo de sobrecarga e inundación. Los resultados muestran que el Escenario 1 presenta niveles ligeramente más bajos que los datos reales, sugiriendo una disminución moderada en la capacidad de retención. El Escenario 2 muestra un aumento en la media, indicando una mayor capacidad de retención, mientras que el Escenario 3 refleja una leve caída en los niveles de agua, posiblemente por menor precipitación. Finalmente, el Escenario 4 presenta los niveles más bajos, lo que sugiere condiciones más secas o una menor eficiencia en la captación.

6.1.2. Distribución de Niveles de Agua Predichos en los Escenarios

La distribución de los niveles de agua para cada escenario se muestra en la siguiente figura 6.1.

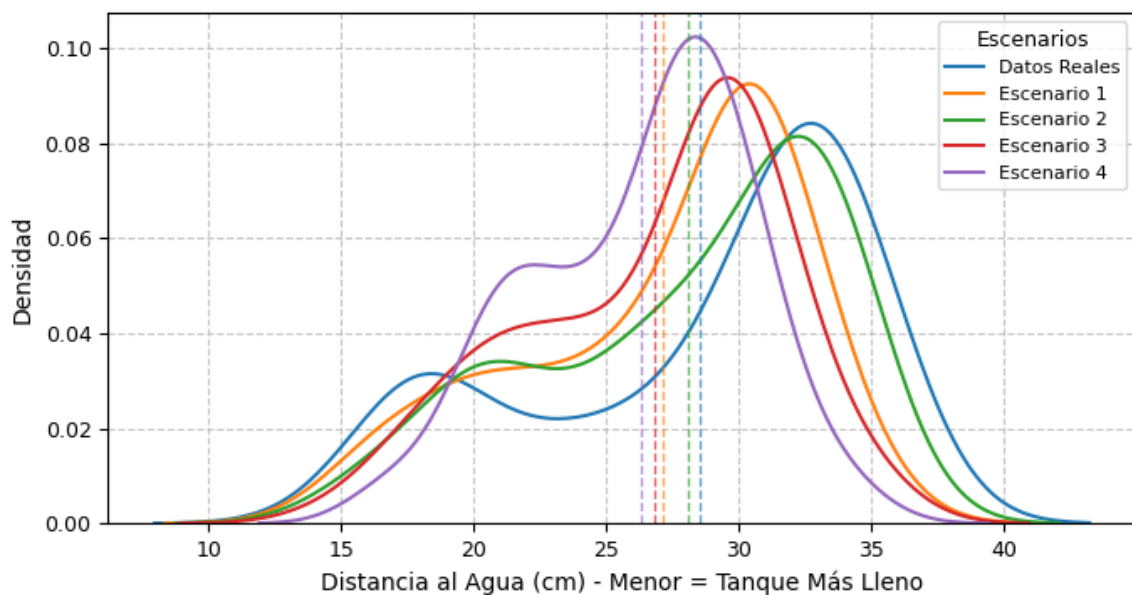


Figura 6.1: Distribución de niveles de agua entre escenarios.

En la Figura 6.1, las líneas verticales discontinuas marcan los valores medios del nivel de agua de cada escenario analizado. Se observan aspectos críticos sobre los posibles comportamientos futuros del SuDS:

- El Escenario 4 muestra una distribución bimodal con un pico pronunciado alrededor de los 26-27 cm y otro menor alrededor de los 20 cm, lo que sugiere una alternancia entre dos estados predominantes del sistema hídrico.
- El Escenario 3 presenta una distribución con un pico pronunciado alrededor de 29-30 cm, más estrecho y alto que otros escenarios, indicando una mayor concentración de eventos en este rango específico y menor variabilidad general en los niveles de agua.
- Los Escenarios 1 y 2 muestran distribuciones desplazadas hacia la derecha, con picos principales en 31-32 cm y 33 cm respectivamente, indicando una tendencia hacia condiciones de menor nivel.

de agua (mayor distancia al agua).

- Los datos reales exhiben una distribución más amplia y aplanada, con mayor densidad hacia valores altos (32-36 cm), reflejando una variabilidad histórica con predominio de periodos con niveles de agua relativamente bajos.
- Respecto a los valores máximos (indicadores de sequía), se observa que todos los escenarios futuros presentan valores máximos inferiores al histórico (35.77 cm), con el Escenario 4 mostrando el valor máximo más bajo (34.54 cm). Esto sugiere que los periodos de sequía extrema podrían ser menos severos en los escenarios futuros.

Esta comparación sugiere que los patrones de comportamiento del SuDS cambiarán significativamente en los escenarios futuros, con mayor concentración de eventos en determinados rangos de nivel de agua, a diferencia de la distribución más uniforme observada históricamente.

6.1.3. Categorización de los Niveles de Agua y Evaluación de Riesgos Futuros

Para evaluar el riesgo de sobrecarga del sistema y los periodos de escasez hídrica, se definieron cuatro categorías de nivel de agua basados en umbrales operativos:

- **Nivel Alto (Crítico):** $\leq 16,94$ cm - Representa situaciones donde el SuDS está próximo a su capacidad máxima, con alto riesgo de desbordamiento.
- **Nivel Medio (Precaución):** 16,94 – 28,52 cm - Indica situaciones donde el sistema está bajo presión moderada.
- **Nivel Bajo (Normal):** $> 28,52$ cm - Representa condiciones donde el sistema opera con niveles adecuados de agua.
- **Nivel Sequía:** Valores cercanos al máximo registrado - Indica condiciones de extrema escasez hídrica, donde el sistema puede ver comprometida su funcionalidad por insuficiencia de agua.

La Figura 6.2 muestra la distribución porcentual de estas categorías para cada escenario.

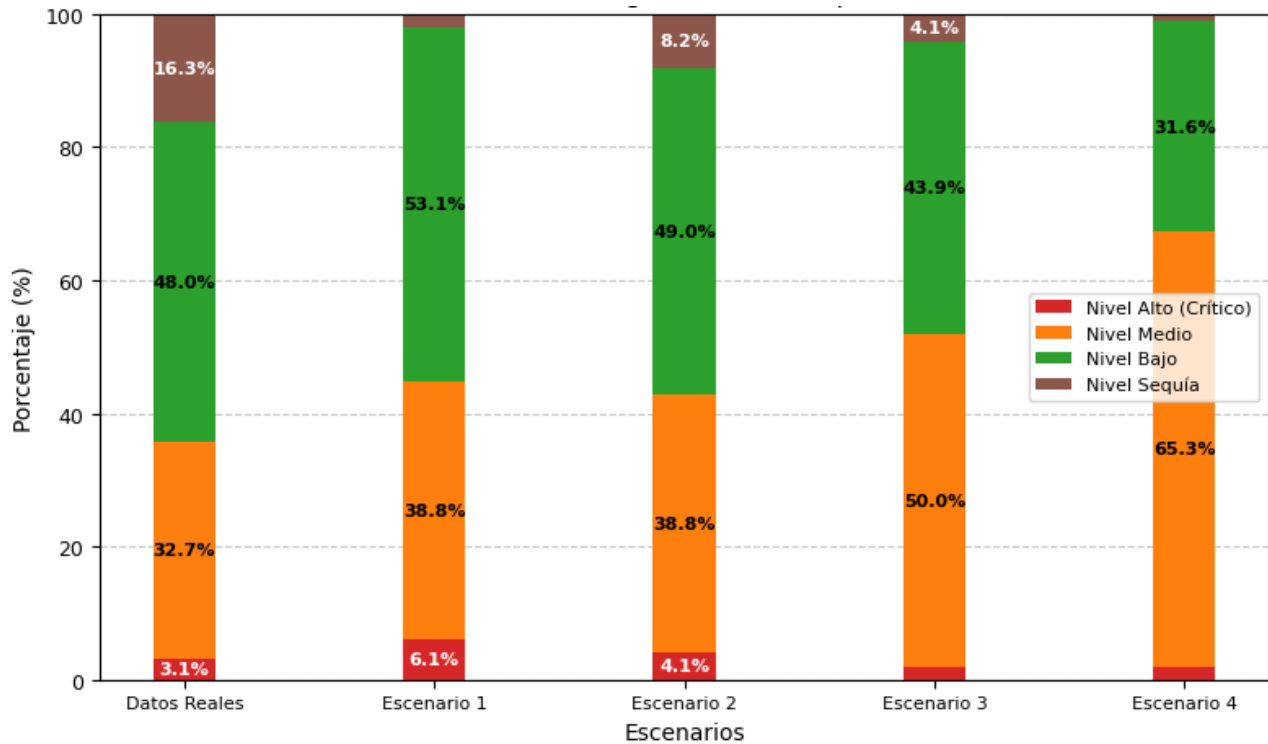


Figura 6.2: Distribución de categorías de nivel de agua por escenario.

La clasificación de los niveles de agua en estas cuatro categorías responde a la necesidad de interpretar operativamente los datos de nivel mínimo registrado en el SHATR (Sistema de Humedal Artificial/Tanque de Reservorio) y evaluar los posibles riesgos asociados al funcionamiento del SuDS en distintos escenarios climáticos futuros. Los umbrales fueron definidos con base en criterios físicos e históricos, permitiendo transformar una variable continua (nivel de agua mínimo en cm) en una variable categórica más interpretable para la gestión del riesgo hídrico.

En concreto:

- El **nivel crítico** se establece como el mínimo histórico incrementado en un 10 %, lo cual representa una condición en la que el sistema se acerca a su máxima capacidad de almacenamiento, indicando un alto riesgo de desbordamiento o falla operativa.
- El **nivel medio** comprende valores entre el umbral crítico y el promedio histórico, reflejando una operación bajo presión que, aunque no es crítica, sí puede ser vulnerable a eventos extremos sucesivos.
- El **nivel bajo** agrupa valores superiores al promedio histórico hasta cerca del máximo, representando condiciones normales de operación del sistema.
- El **nivel sequía** se define para los valores en el 10 % superior del rango histórico, identificando periodos donde la escasez de agua puede comprometer la eficacia del sistema para cumplir funciones de retención y tratamiento.

Este enfoque permite comparar de forma clara el comportamiento histórico con los escenarios futuros modelados, facilitando la identificación de tendencias como el aumento de eventos extremos (principal foco de este estudio) y, complementariamente, el análisis de periodos de sequía que podrían afectar secundariamente el rendimiento del sistema.

El análisis de la distribución de categorías revela importantes hallazgos respecto a los riesgos futuros:

- **Respecto a eventos extremos de lluvia:** El Escenario 1 presenta el mayor porcentaje de eventos críticos (6.1%), lo que indica un aumento significativo respecto a los datos históricos (3.1%). Este escenario requerirá mayor atención y posibles medidas de mitigación para gestionar el incremento de eventos extremos que podrían sobrecargar el sistema.

- **Respecto a la distribución general:** El Escenario 4 muestra un notable aumento en eventos de nivel medio (aproximadamente 65% frente al 33% histórico) y una reducción drástica en eventos de nivel bajo (aproximadamente 32% frente al 48% histórico). Esto sugiere una operación del sistema bajo condiciones de mayor estrés hídrico.

- **Respecto a periodos de sequía:** Los datos históricos muestran que un 16.3% de los eventos corresponden a condiciones de sequía, mientras que los escenarios futuros presentan porcentajes menores (desde 0% en el Escenario 1 hasta 8.2% en el Escenario 2). Esto sugiere que, si bien el principal riesgo futuro sigue siendo la sobrecarga por eventos extremos de precipitación, algunos escenarios mantienen un riesgo moderado de periodos de sequía.

Esta información es crucial para planificar estrategias de mantenimiento y operación del SuDS, priorizando las medidas para mitigar el impacto de eventos extremos de precipitación, sin descuidar la posible ocurrencia de periodos de sequía en determinados escenarios, especialmente en aquellos donde los modelos del IDEAM proyectan disminuciones estacionales de precipitación.

6.1.4. Análisis de Influencia de Variables en el SuDS Bajo Escenarios Futuros

Para comprender los mecanismos que conducen tanto a eventos extremos como a periodos de sequía, se analizaron las correlaciones entre las variables de entrada y el nivel de agua. La Figura 6.3 presenta estas correlaciones.

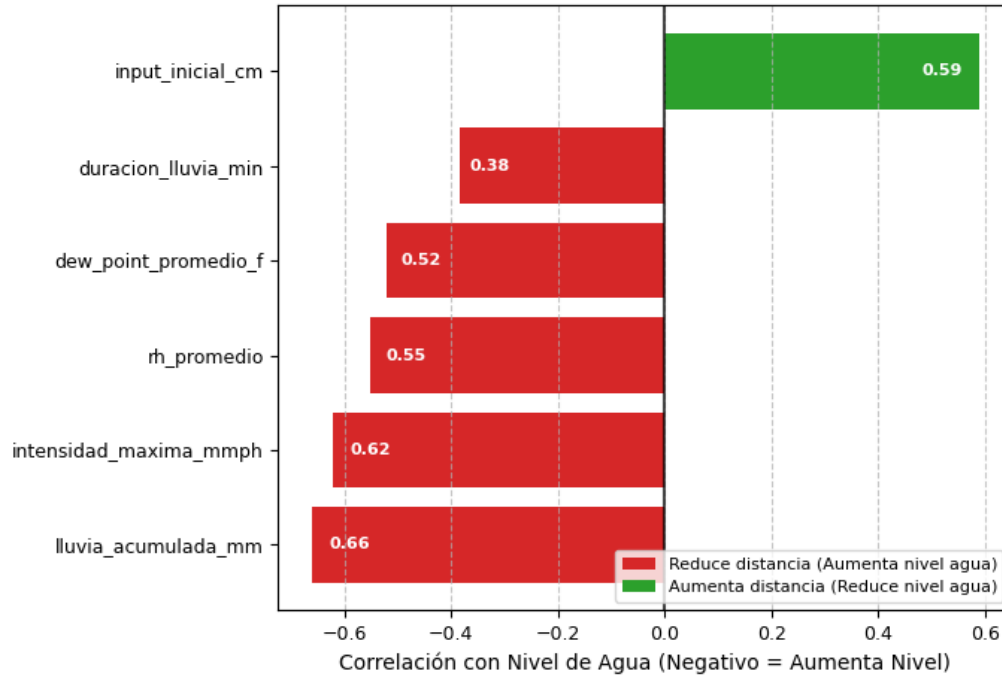


Figura 6.3: Factores que influyen en el nivel de agua.

Las barras rojas indican factores que reducen la distancia al agua (aumentan el nivel), mientras que las barras verdes indican factores que aumentan la distancia (reducen el nivel). El análisis de las correlaciones revela información crítica sobre los factores determinantes del rendimiento del SuDS:

- La lluvia acumulada (mm) muestra la correlación negativa más fuerte ($-0,66$), lo que la identifica como el factor principal que aumenta el nivel de agua y, consecuentemente, el riesgo de sobrecarga del sistema.
- La intensidad máxima de lluvia (mm/h) y la humedad relativa promedio también presentan correlaciones negativas significativas ($-0,62$ y $-0,55$ respectivamente), indicando su importante contribución al aumento de los niveles de agua.
- El nivel inicial del agua (input_inicial_cm) tiene una correlación positiva de $0,59$, lo que indica que niveles iniciales más altos (mayor distancia al agua, menor nivel) tienden a mantenerse altos. Esto sugiere que los periodos de sequía o baja captación pueden tener efectos persistentes en el sistema.
- Para la gestión de periodos de sequía, el nivel inicial del agua es el factor determinante más importante, sugiriendo que las condiciones iniciales del sistema son críticas para mantener niveles adecuados de agua durante periodos secos.

Estos hallazgos son fundamentales para desarrollar sistemas de alerta temprana y estrategias preventivas, particularmente enfocándose en la monitorización de las variables con mayor impacto tanto para eventos de inundación como de sequía.

6.1.4.1. Análisis de Variable Más Influyente

Dado que la lluvia acumulada resultó ser el factor más influyente en el comportamiento del SuDS, la Figura 6.4 examina en detalle su relación con el nivel de agua para el Escenario 4 (izquierda) y compara esta relación entre todos los escenarios (derecha).

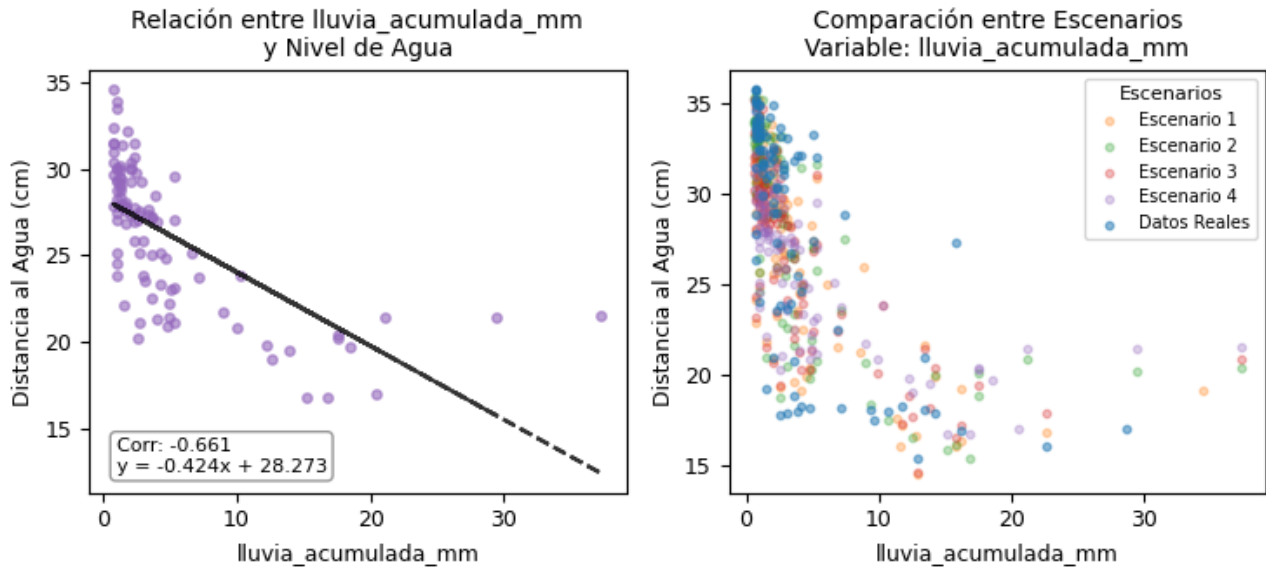


Figura 6.4: Relación entre lluvia acumulada (mm) y nivel de agua (cm).

A la izquierda, correlación específica para el Escenario 4; a la derecha, comparación entre todos los escenarios. El análisis de esta relación revela aspectos críticos para la gestión futura del sistema:

- La ecuación de regresión $y = -0,424x + 28,273$ para el Escenario 4 indica que, por cada milímetro adicional de lluvia acumulada, el nivel de agua aumenta aproximadamente 0.42 cm (la distancia al sensor disminuye).
- La comparación entre escenarios (gráfico derecho) muestra que esta relación es consistente en todos los escenarios, pero con diferentes grados de dispersión.
- Se observa una mayor concentración de eventos extremos (puntos en la parte inferior del gráfico) en el rango de 10-15 mm de lluvia acumulada, lo que indica un umbral crítico para la sobrecarga del sistema.
- Para valores bajos o nulos de lluvia acumulada (extremo izquierdo del gráfico), el Escenario 4 muestra distancias al agua generalmente menores que otros escenarios, sugiriendo que incluso en periodos secos este escenario mantendría niveles de agua relativamente más altos que otros escenarios.
- Los datos reales (puntos azules en el gráfico derecho) muestran mayor dispersión vertical para valores bajos de lluvia acumulada, indicando mayor variabilidad histórica en los niveles de agua durante periodos secos comparado con los escenarios proyectados.

Este análisis permite establecer tanto umbrales de alerta para precipitaciones intensas como estrategias de gestión para periodos secos, especialmente útiles para anticipar situaciones de riesgo y activar protocolos preventivos.

6.1.5. Análisis Comparativo de Indicadores Clave Predichos entre Escenarios

La Figura 6.5 proporciona una visión integral del comportamiento proyectado del SuDS bajo los diferentes escenarios, considerando cuatro métricas clave que reflejan su capacidad para manejar eventos extremos.

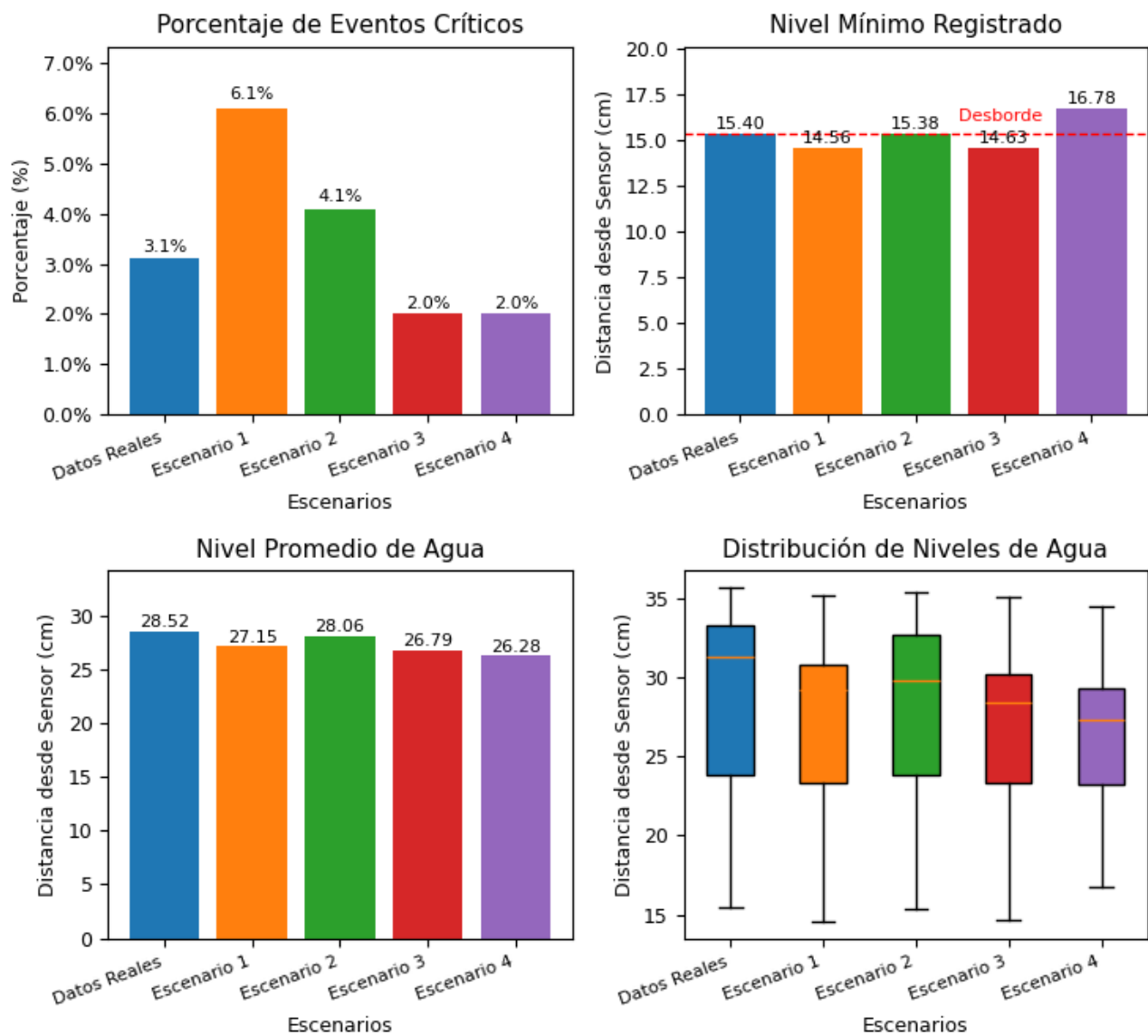


Figura 6.5: Comparativa de comportamiento del SuDS entre escenarios.

la figura muestra el porcentaje de eventos críticos, el nivel mínimo registrado (máxima inundación), el nivel promedio de agua y la distribución de los niveles. Este análisis comparativo revela importantes hallazgos sobre los riesgos futuros:

- **Porcentaje de Eventos Críticos:** El Escenario 1 muestra un incremento significativo (6.1 % frente al 3.1 % histórico), indicando un mayor riesgo de episodios que podrían sobrecargar el sistema. Por el contrario, los Escenarios 3 y 4 muestran una reducción (2.0 %), sugiriendo una menor frecuencia de eventos extremos.
- **Nivel Mínimo Registrado:** El Escenario 4 presenta el mayor nivel mínimo registrado (16.78 cm), superando el valor histórico de 15.40 cm. En contraste, el Escenario 1 exhibe el nivel mínimo más bajo (14.56 cm). Esta diferencia sugiere que, en el Escenario 1, los eventos extremos podrían ser más severos que los observados históricamente, lo que implicaría una mayor demanda sobre la capacidad de respuesta del sistema. Es importante señalar que la línea punteada ubicada en 15.40 cm representa el umbral de desborde del sistema, a partir del cual el agua sobrepasa el vertedero y se dirige hacia el siguiente tanque. Un análisis más detallado de este fenómeno se presentará en secciones posteriores.
- **Nivel Promedio de Agua:** Todos los escenarios futuros muestran niveles promedio más bajos que los datos históricos, con el Escenario 4 presentando el valor más bajo (26.28 cm). Esto sugiere una tendencia general hacia mayores volúmenes de agua en el sistema, aunque con diferentes patrones de distribución.
- **Distribución de Niveles:** La comparación de los diagramas de caja (cuadrante inferior derecho) revela que los escenarios futuros presentan valores máximos (extremo superior de los bigotes) inferiores al histórico. El Escenario 4 muestra el valor máximo más bajo (34.54 cm frente a 35.77 cm de los datos reales), indicando que los periodos de sequía extrema podrían ser menos severos pero con mayor frecuencia de niveles moderadamente bajos.

6.1.6. Análisis del Sistema de Desborde del SuDS

Un aspecto crítico del SuDS en el nivel de entrada, es la posibilidad de desborde que se activa cuando el nivel de agua supera un umbral crítico, correspondiente a valores de `input_min_cm` inferiores a 15.4 cm. Este umbral representa la distancia mínima operativa del sensor a la lámina de agua, por lo que lecturas por debajo de este valor indican un nivel de agua tan alto que activa el sistema de desborde hacia el tanque.

6.1.6.1. Eventos de Desborde Proyectados en los Escenarios Futuros

El análisis de los valores mínimos predichos por el modelo SVR para cada escenario revela información crucial sobre la activación del desborde, aunque el modelo de aprendizaje automático no incorpora explícitamente el conocimiento sobre este mecanismo físico del sistema. La Tabla 6.2 presenta los valores mínimos proyectados para cada escenario y su relación con el umbral de desborde.

Escenario	Nivel mínimo (cm)	Diferencia respecto al umbral (cm)	Evento de desborde
Datos Reales	15.40	0.00	Umbral exacto
Escenario 1	14.56	-0.84	Desborde crítico
Escenario 2	15.38	-0.02	Desborde leve
Escenario 3	14.63	-0.77	Desborde significativo
Escenario 4	16.78	+1.38	Sin desborde

Tabla 6.2: Análisis de eventos de desborde proyectados por escenario

El análisis de estos eventos proyectados revela diferencias significativas en la severidad de los posibles eventos de desborde entre los distintos escenarios:

- El **Escenario 1** presenta el evento de desborde más crítico, con un nivel mínimo proyectado de 14.56 cm, 0.84 cm por debajo del umbral de desborde. Esta diferencia, aunque aparentemente pequeña, representa un volumen significativo de agua que excede la capacidad nominal del sistema primario y debe ser derivado al tanque secundario.
- El **Escenario 2** muestra un evento de desborde apenas perceptible, con un nivel mínimo de 15.38 cm, tan solo 0.02 cm por debajo del umbral. Este tipo de evento representaría una activación mínima del sistema de desborde, con un volumen de excedencia muy reducido.
- El **Escenario 3** proyecta un evento de desborde significativo, con un nivel mínimo de 14.63 cm, 0.77 cm por debajo del umbral, indicando un volumen considerable de agua que debería ser gestionado por el sistema de desborde.
- El **Escenario 4** no presenta ningún evento proyectado que active el sistema de desborde, con un nivel mínimo de 16.78 cm, muy por encima del umbral crítico de 15.4 cm, lo que sugiere una operación completamente dentro de la capacidad nominal del sistema primario.

Las proyecciones realizadas tienen importantes implicaciones para la gestión operativa del SuDS:

- El evento de desborde proyectado para el **Escenario 1** requiere especial atención, ya que la magnitud de la diferencia respecto al umbral (0.84 cm) sugiere un mayor volumen de agua que deberá ser gestionado por el tanque secundario, lo que podría sobrecargar su capacidad o afectar la eficiencia del tratamiento si no se toman medidas preventivas.
- Aunque el evento proyectado para el **Escenario 2** apenas activa el sistema de desborde, es importante considerarlo en la planificación operativa, ya que representa un punto crítico donde el sistema opera al límite de su capacidad nominal.
- El evento de desborde en el **Escenario 3**, con una diferencia de 0.77 cm respecto al umbral, también presenta un desafío significativo para la gestión del sistema, aunque ligeramente menos crítico que el del Escenario 1.
- La ausencia de eventos de desborde en el **Escenario 4** sugiere que, bajo las condiciones climáticas proyectadas para este escenario, el sistema primario tendría capacidad suficiente para gestionar todos los eventos de precipitación sin necesidad de activar el sistema de desborde.

Es importante destacar que, aunque el modelo SVR no incorpora explícitamente el conocimiento sobre el sistema de desborde, las proyecciones realizadas permiten identificar y caracterizar los eventos críticos que activarían este mecanismo. Esta información es fundamental para la planificación de medidas preventivas y adaptativas que aseguren el funcionamiento óptimo del SuDS bajo diferentes escenarios climáticos futuros.

6.1.7. Implicaciones para la Gestión del SuDS en Condiciones Climáticas Extremas

Los resultados del análisis de escenarios futuros tienen importantes implicaciones para la gestión del sistema de drenaje sostenible tanto en eventos de inundación como en periodos de sequía:

■ Gestión de Eventos de Inundación:

- El Escenario 1 representa el mayor desafío en términos de frecuencia de eventos críticos (6.1%) y eventos de mayor severidad (nivel mínimo de 14.56 cm), requiriendo posibles mejoras en la capacidad de almacenamiento o implementación de sistemas complementarios de mitigación.

■ Gestión de Periodos de Sequía:

- Todos los escenarios proyectan una reducción en la frecuencia de eventos de nivel bajo (categoría normal, >28.52 cm), con el Escenario 4 mostrando la reducción más drástica (de aproximadamente 64% a 33%).
- El Escenario 4 también muestra el menor valor máximo (34.54 cm), lo que sugiere que, aunque los periodos de sequía extrema podrían ser menos severos, habría una mayor frecuencia de niveles moderadamente bajos que podrían afectar funciones del SuDS como la retención prolongada de agua para reutilización.
- Los Escenarios 1 y 3 presentan valores máximos de 35.23 cm y 35.16 cm respectivamente, ligeramente inferiores al histórico (35.77 cm), sugiriendo condiciones relativamente similares para periodos secos extremos.

La variabilidad observada entre escenarios subraya la importancia de diseñar estrategias de adaptación flexibles, capaces de responder tanto a eventos de inundación como a periodos de sequía.

6.1.8. Conclusiones de la Predicción del Comportamiento del SuDS

El análisis predictivo realizado revela que los escenarios climáticos futuros presentarán diferentes desafíos para el funcionamiento del SuDS:

- Los patrones de precipitación proyectados modificarán la distribución de los niveles de agua en el sistema, con tendencia general hacia mayores cargas hidráulicas promedio en todos los escenarios.
- La frecuencia y severidad de los eventos extremos de inundación variará significativamente entre escenarios, desde aumentos considerables (Escenario 1) hasta reducciones moderadas (Escenarios 3 y 4) en comparación con los datos históricos.

- Todos los escenarios proyectan una reducción en la frecuencia e intensidad de periodos de sequía extrema, aunque con mayor prevalencia de niveles moderadamente bajos de agua, lo que podría afectar funciones secundarias del SuDS como la reutilización de agua.
- La lluvia acumulada se confirma como el factor determinante principal del comportamiento del sistema, seguida por la intensidad máxima de lluvia y la humedad relativa, mientras que el nivel inicial es el factor clave para la persistencia de condiciones de sequía.
- El análisis específico del desborde revela que tres de los cuatro escenarios futuros (Escenarios 1, 2 y 3) proyectan al menos un evento que activaría este mecanismo, con el Escenario 1 presentando el evento más crítico (14.56 cm, 0.84 cm por debajo del umbral). El Escenario 4 es el único que no proyecta eventos que activen el desborde, lo que sugiere una operación completamente dentro de la capacidad nominal del sistema primario bajo las condiciones climáticas de este escenario.

Estos hallazgos proporcionan una base sólida para la planificación de medidas preventivas y adaptativas que permitan mantener la eficacia del SuDS bajo las diferentes condiciones climáticas futuras proyectadas, tanto para gestionar eventos de inundación como para optimizar el funcionamiento durante periodos secos, asegurando un rendimiento equilibrado del sistema en todo el espectro de condiciones climáticas.

6.2. Propuestas de Mejora del SuDS ante Predicciones de Eventos Extremos

Los resultados obtenidos a partir del modelado y la predicción bajo distintos escenarios de cambio climático brindan información valiosa para proponer mejoras puntuales en el SuDS de la Pontificia Universidad Javeriana en Bogotá. El propósito central de estas recomendaciones es aumentar la resiliencia y eficacia del sistema frente a eventos climáticos extremos futuros, tanto de inundación como de sequía.

Considerando que el sistema actual —el Sistema de Humedal Artificial/Tanque de Reservorio (SHATR)— cuenta con un fondo impermeable, un lecho de material granular con vegetación para la depuración del agua, y un vertedero que canaliza el exceso hídrico hacia un segundo tanque, se proponen las siguientes mejoras fundamentadas en los resultados del modelado y el análisis de escenarios futuros.

6.2.1. Optimización de la capacidad y gestión del vertedero

Dado que el análisis de importancia de variables del modelo SVR optimizado identificó la lluvia acumulada y la intensidad máxima de precipitación como factores clave en la predicción del rendimiento del SuDS, y que el Escenario 1 proyecta el mayor porcentaje de eventos críticos (6.1%), lo cual sugiere un aumento en la frecuencia de lluvias extremas, se plantea optimizar la capacidad de retención del sistema principal y la gestión del flujo hacia el tanque secundario a través del vertedero. Entre las acciones recomendadas se encuentran:

- **Ampliación estratégica del volumen útil del tanque de reservorio:** A pesar de contar con un fondo impermeable, es posible incrementar el volumen disponible para retención mediante la elevación de las paredes del tanque, asegurando la estabilidad estructural y el rediseño del vertedero según sea necesario. Las proyecciones climáticas indican un aumento en la intensidad máxima de precipitación, por lo que una mayor capacidad de almacenamiento permitiría manejar picos de lluvia más intensos y prolongados antes de activar el vertedero, reduciendo la frecuencia de sobreflujo y mitigando el impacto en el tanque secundario.
- **Calibración y posible rediseño del vertedero:** Comprender la capacidad de descarga actual del vertedero es fundamental. El modelo SVR optimizado, que presentó el mejor desempeño ($R^2 = 0.9024$), puede emplearse para simular el comportamiento del flujo bajo distintos escenarios de precipitación. Si los resultados sugieren una activación prematura o tardía del vertedero, se puede considerar su recalibración o rediseño para optimizar el flujo hacia el tanque secundario y prevenir sobrecargas, especialmente dado el incremento proyectado en eventos críticos.
- **Integración de alertas tempranas basadas en el modelo SVR optimizado:** El uso del modelo SVR optimizado para predecir niveles de entrada en tiempo real, a partir de datos meteorológicos, permitiría anticipar la activación del vertedero. Esta información serviría para gestionar proactivamente el nivel del tanque secundario, preparándolo para recibir el caudal adicional y maximizando la capacidad total del sistema frente a eventos extremos. El alto valor de R^2 del modelo (0.9024) respalda su utilidad en sistemas de alerta temprana.
- **Optimización del sistema de desborde:** Considerando que los Escenarios 1 y 3 proyectan eventos de desborde significativos (0.84 cm y 0.77 cm por debajo del umbral, respectivamente), se recomienda evaluar la capacidad actual del sistema de desborde para manejar estos volúmenes excedentes. Se sugiere un análisis hidráulico detallado del comportamiento del vertedero bajo estas condiciones específicas y, de ser necesario, implementar mejoras que optimicen su capacidad de descarga y la transición al tanque secundario. Esto podría incluir modificaciones en la geometría del vertedero, sistemas de control de flujo, o incluso la instalación de sensores específicos que permitan monitorizar en tiempo real el funcionamiento del sistema de desborde.

6.2.2. Mejoras en la infiltración y retención del lecho granular y la vegetación

Aunque el fondo del tanque es impermeable, el material granular y la vegetación desempeñan un papel crucial en la retención temporal del agua y la mejora de su calidad. Para optimizar estas funciones, se proponen las siguientes acciones:

- **Evaluación y posible optimización de la composición y profundidad del lecho granular:** Es necesario analizar la capacidad actual de retención del material granular y evaluar si una modificación en su composición o profundidad puede mejorar la retención temporal del agua, disminuyendo así el volumen y velocidad del flujo hacia el vertedero en eventos de menor intensidad. Esta mejora también podría fortalecer la capacidad de tratamiento del agua por parte de la vegetación. Además, dada la proyección de lluvias más prolongadas y acumulativas, un lecho granular más eficiente podría amortiguar el sistema durante periodos húmedos extendidos.

- **Selección y manejo estratégico de la vegetación:** La vegetación contribuye significativamente a la evapotranspiración y a la purificación del agua. Evaluar especies con alta capacidad de evapotranspiración y resistencia tanto a inundaciones como a sequías —condiciones previstas en los escenarios de cambio climático— puede optimizar la gestión del volumen de agua y la eficiencia del sistema. Asimismo, prácticas como podas estratégicas pueden incidir positivamente en la capacidad de retención. Dado el posible cambio del régimen higrótico, la selección de vegetación resiliente se vuelve crítica.
- **Monitoreo de colmatación del lecho granular y estado de la vegetación:** La acumulación de sedimentos puede disminuir la capacidad de infiltración y retención del material granular, mientras que el estrés hídrico puede afectar la salud y funcionalidad de la vegetación. Se recomienda establecer un programa regular de monitoreo, guiado por los patrones de precipitación y niveles de agua proyectados por los modelos, para identificar y abordar proactivamente estos problemas y mantener la eficiencia del sistema a lo largo del tiempo. Esto es particularmente relevante ante el aumento proyectado de la variabilidad en las precipitaciones.

6.2.3. Consideraciones para la gestión integrada de los tanques

Dado que el agua del primer tanque se traslada al segundo a través del vertedero, una gestión integrada de ambos puede fortalecer la resiliencia del sistema. Se sugieren las siguientes acciones:

- **Modelado del comportamiento del sistema de doble tanque bajo escenarios futuros:** Ampliar el modelo actual para incluir la dinámica del tanque secundario y el flujo a través del vertedero permitiría una evaluación más integral del desempeño del sistema frente a eventos extremos. Esto podría revelar cuellos de botella o ineficiencias en la gestión del agua entre los dos tanques, considerando las diferencias proyectadas en los patrones de nivel de agua para cada escenario.
- **Implementación de estrategias de control activo (si es viable):** Si la infraestructura lo permite, podría evaluarse la implementación de mecanismos de control activo en el vertedero o en el traslado de agua entre tanques, guiados por las predicciones del modelo. Por ejemplo, anticipando un evento extremo, se podría liberar agua del tanque secundario de forma controlada para aumentar su capacidad de recepción. Esta propuesta requiere una evaluación detallada de la infraestructura existente y su viabilidad técnica. Las capacidades de alerta temprana del modelo SVR optimizado pueden apoyar este tipo de estrategias.
- **Análisis de calidad del agua en ambos tanques bajo diferentes escenarios:** Dado que la vegetación en el primer tanque mejora la calidad del agua, analizar cómo los distintos escenarios de precipitación y los patrones de flujo influyen en la calidad del agua en ambos tanques puede orientar estrategias de gestión que optimicen la purificación. Los cambios proyectados en el ciclo hidrológico podrían afectar el tiempo de residencia del agua en el primer tanque, lo cual impactaría la eficiencia de los procesos de depuración natural.

En conclusión, las mejoras propuestas —basadas en los resultados del modelado con aprendizaje automático y el análisis de escenarios de cambio climático— se enfocan en optimizar la capacidad de

retención del tanque primario, mejorar la gestión del vertedero, potenciar las funciones del lecho granular y la vegetación, y promover una gestión integrada de ambos tanques. La implementación de estas medidas, considerando las características específicas del sistema existente, contribuirá significativamente al fortalecimiento de la resiliencia y efectividad del SuDS de la Pontificia Universidad Javeriana en Bogotá frente a condiciones climáticas futuras. El uso del modelo SVR optimizado como herramienta de predicción y alerta temprana constituye un elemento clave en el logro de esta resiliencia mejorada.

Conclusiones y Trabajos Futuros

7.1. Conclusiones y Trabajos Futuros

7.1.1. Conclusiones

El presente Proyecto Aplicado abordó de forma integral el modelado y la optimización del SuDS, en particular el Sistema de Humedal Artificial/Tanque de Reservorio (SHATR) ubicado en la Pontificia Universidad Javeriana, sede Bogotá. La metodología se centró en la aplicación de técnicas de *Machine Learning* como herramienta para evaluar y anticipar la respuesta hidrológica del sistema frente a diversos escenarios climáticos, con el objetivo de fortalecer su capacidad adaptativa.

Los resultados obtenidos permitieron consolidar avances metodológicos y conceptuales relevantes para el estudio de SuDS en contextos urbanos. Entre los principales hallazgos se destacan:

- Se logró una implementación eficaz del modelo de Regresión por Vectores de Soporte (SVR), optimizado mediante búsqueda bayesiana, alcanzando un coeficiente de determinación (R^2) de 0.9024 y un error cuadrático medio (RMSE) de 1.5940 cm en el conjunto de prueba. Este resultado sugiere que, con la configuración adecuada, los modelos de aprendizaje automático son capaces de reproducir con alta fidelidad las dinámicas no lineales presentes en sistemas hidrológicos complejos, constituyéndose en una alternativa complementaria a los enfoques tradicionales.
- El análisis de importancia de variables reveló que la lluvia acumulada (36.2%), la intensidad máxima de precipitación (29.7%) y el nivel inicial del tanque (17.6%) son determinantes en la predicción del nivel de entrada. Esta evidencia empírica fortalece la comprensión del comportamiento hidrológico del sistema y contribuye a orientar futuras estrategias de monitoreo hacia un conjunto reducido pero representativo de variables críticas.
- La evaluación bajo cuatro escenarios climáticos proyectados (2021-2100) permitió anticipar posibles variaciones en la frecuencia de eventos críticos y en los niveles promedio del sistema. En particular, se evidenció un incremento del 6.1% en eventos de desborde para el periodo 2021-2040, en comparación con el 3.1% observado históricamente. Estos resultados constituyen una base analítica que podría contribuir al diseño de medidas preventivas y adaptativas frente al cambio climático.
- A pesar de los esfuerzos realizados, el modelado del nivel de salida presentó limitaciones significativas, lo cual indica la necesidad de fortalecer el acopio de datos post-evento y de incorporar variables adicionales que reflejen procesos como la infiltración, evaporación y las interacciones entre compartimentos del sistema. Este aspecto resalta la importancia de abordar la modelación del SuDS desde una perspectiva integral y con un horizonte temporal ampliado.

- Finalmente, a partir de los resultados obtenidos, fue posible identificar puntos vulnerables y proponer recomendaciones técnicas orientadas a incrementar la resiliencia del SHATR. Estas sugerencias se relacionan con el diseño y operación de componentes clave del sistema, tales como el vertedero y la capacidad de almacenamiento e infiltración, y muestran cómo los resultados derivados del modelado pueden ser útiles para orientar intervenciones concretas.

En conjunto, este estudio confirma la pertinencia de las técnicas de *Machine Learning* para analizar el comportamiento hidrológico de sistemas urbanos complejos bajo condiciones cambiantes. Además de aportar al conocimiento técnico-científico sobre la respuesta de los SuDS, los hallazgos obtenidos ofrecen herramientas conceptuales y metodológicas que podrían servir de insumo para futuras estrategias de gestión urbana resiliente.

7.1.2. Trabajos Futuros

A partir del trabajo realizado, se identifican varias líneas de acción que podrían ser abordadas en investigaciones futuras para consolidar y ampliar los resultados obtenidos:

- Incorporar datos adicionales que reflejen procesos internos del sistema, como la evaporación, la infiltración y el comportamiento del vertedero. Esto permitiría enriquecer el modelado del nivel de salida y mejorar la representación global del sistema, especialmente en condiciones de saturación o recarga.
- Explorar técnicas de modelado híbridas que combinen enfoques físicos y estadísticos. Por ejemplo, integrar modelos hidrológicos simplificados con algoritmos de aprendizaje automático puede facilitar una interpretación más robusta de los resultados y mejorar la generalización del modelo ante nuevos escenarios.
- Ampliar el horizonte temporal de análisis e incorporar datos de mayor resolución espacial y temporal, provenientes de sensores remotos o redes de monitoreo en tiempo real. Esta información puede ser clave para capturar variabilidades locales y fortalecer la capacidad predictiva del modelo.
- Evaluar el desempeño del SHATR bajo distintos escenarios de uso del suelo y crecimiento urbano. Estas simulaciones podrían contribuir a estimar la capacidad de adaptación del sistema frente a transformaciones futuras del entorno y orientar estrategias de planificación urbana más sostenibles.
- Desarrollar una interfaz interactiva que permita visualizar los resultados del modelo y simular distintos escenarios de operación. Este tipo de herramienta facilitaría la toma de decisiones por parte de los actores encargados de la gestión del sistema y promovería el uso de la modelación como apoyo a la planificación urbana.
- Finalmente, se sugiere replicar la metodología propuesta en otros SuDS de la ciudad o del país, con el fin de validar su aplicabilidad en contextos distintos y construir un marco comparativo que permita identificar buenas prácticas, retos comunes y oportunidades de mejora en la gestión hídrica urbana.

Estas propuestas buscan continuar fortaleciendo la comprensión y el manejo de sistemas de drenaje sostenible desde una perspectiva interdisciplinaria, combinando el potencial de las tecnologías emergentes con el conocimiento técnico acumulado. En un contexto donde los efectos del cambio climático son cada vez más visibles, resulta fundamental seguir avanzando hacia soluciones que promuevan ciudades más resilientes, adaptativas y comprometidas con la gestión sostenible del agua.

8.1. Enlaces a cuadernos de Google Colab - Código Python

A continuación, se presentan los enlaces a los cuadernos de Google Colab utilizados en el desarrollo de este trabajo. Estos cuadernos, ejecutados en Python, conforman un flujo de trabajo secuencial que abarca desde el análisis de datos hasta la simulación de escenarios de cambio climático y la predicción del comportamiento del sistema SUDS. Todo el código, los análisis y visualizaciones están disponibles en línea para su consulta y reproducción.

- **Análisis de datos de eventos de lluvia y retención**

<https://colab.research.google.com/drive/1uKmdarN4T1KdM12FdAB8g3BYHFwJOqQM>

Se realiza la identificación y caracterización de eventos de lluvia, tratamiento de datos faltantes y atípicos, análisis exploratorio, PCA, correlaciones cruzadas y consolidación de la base de datos para modelado.

- **Selección y entrenamiento de modelos de predicción**

<https://colab.research.google.com/drive/16RWIpGkYg4WoeG0daRNip5i2N4-P03Uj>

Se implementan y evalúan modelos de aprendizaje automático (Regresión Lineal, ElasticNet, Random Forest, XGBoost y SVM), optimizando hiperparámetros y seleccionando el modelo más adecuado para la predicción del nivel de agua.

- **Simulación de escenarios de cambio climático**

<https://colab.research.google.com/drive/1AYa0woSQwHyOHUqePsU3u14hCXENViZL>

Se modifican variables meteorológicas de los eventos históricos según diferentes escenarios futuros (2021–2100), generando nuevas bases de datos para evaluar el impacto del cambio climático en el sistema.

- **Predicción del rendimiento del sistema bajo escenarios futuros**

<https://colab.research.google.com/drive/1FUYfpDzhRn7ZSAXKnR05JV12aW7e6ymk>

Utilizando el modelo previamente entrenado, se predicen los niveles de agua en el sistema SUDS bajo los diferentes escenarios climáticos. Se analizan estadísticamente los resultados, se identifican los factores más influyentes y se determina el escenario más favorable.

Bibliografía

- [1] S. L. Galarza Molina, “Decision-making tool for the operation of a stormwater harvesting system,” Ph.D. dissertation, Pontificia Universidad Javeriana, 2017.
- [2] IDEAM, PNUD, MADS, DNP, CANCELLERÍA, *Escenarios de Cambio Climático para Precipitación y Temperatura para Colombia 2011-2100: Herramientas Científicas para la Toma de Decisiones – Estudio Técnico Completo*. Tercera Comunicación Nacional de Cambio Climático, 2015. [Online]. Available: <http://www.cambioclimatico.gov.co/resultados>
- [3] B. A. Garavito Guerrero, “Expansión urbana y cambio de coberturas en el área periurbana de la ciudad de bogotá dc-colombia mediante el análisis multitemporal de imágenes landsat para los años 1991 y 2016,” Master’s thesis, Universidad Militar Nueva Granada, 2017.
- [4] S. U. Vincent, M. Radhakrishnan, L. Hayde, and A. Pathirana, “Enhancing the economic value of large investments in sustainable drainage systems (suds) through inclusion of ecosystems services benefits,” *Water*, vol. 9, no. 11, 2017. [Online]. Available: <https://www.mdpi.com/2073-4441/9/11/841>
- [5] B. Woods-Ballard, R. Kellagher, P. Martin, C. Jefferies, R. Bray, P. Shaffer *et al.*, *The SUDS manual*. Ciria London, 2015, vol. C753.
- [6] R. D’Ambrosio, A. Longobardi, and B. Schmalz, “Suds as a climate change adaptation strategy: Scenario-based analysis for an urban catchment in northern italy,” *Urban Climate*, vol. 51, p. 101596, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2212095523001906>
- [7] A. Arahuetes and J. O. Cantos, “The potential of sustainable urban drainage systems (suds) as an adaptive strategy to climate change in the spanish mediterranean,” *International Journal of Environmental Studies*, vol. 76, no. 5, pp. 764–779, 2019. [Online]. Available: <https://doi.org/10.1080/00207233.2019.1634927>
- [8] S. H. Kwon and J. H. Kim, “Machine learning and urban drainage systems: State-of-the-art review,” *Water*, vol. 13, no. 24, 2021. [Online]. Available: <https://www.mdpi.com/2073-4441/13/24/3545>
- [9] E. Alpaydm, *Machine Learning*. The MIT Press, 08 2021. [Online]. Available: <https://doi.org/10.7551/mitpress/13811.001.0001>
- [10] A. Mullapudi, M. J. Lewis, C. L. Gruden, and B. Kerkez, “Deep reinforcement learning for the real time control of stormwater systems,” *Advances in Water Resources*, vol. 140, p. 103600, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0309170820302499>
- [11] C. Wang, B. D. Bowes, A. Tavakoli, S. C. Adams, J. L. Goodall, and P. A. Beling, “Smart storm-water control systems: A reinforcement learning approach.” in *ISCRAM*, 2020, pp. 2–13.

- [12] S. Xie, W. Wu, S. Mooser, Q. Wang, R. Nathan, and Y. Huang, “Artificial neural network based hybrid modeling approach for flood inundation modeling,” *Journal of Hydrology*, vol. 592, p. 125605, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0022169420310660>
- [13] Y. Ding, Y. Zhu, J. Feng, P. Zhang, and Z. Cheng, “Interpretable spatio-temporal attention lstm model for flood forecasting,” *Neurocomputing*, vol. 403, pp. 348–359, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0925231220307530>
- [14] I.-F. Kao, Y. Zhou, L.-C. Chang, and F.-J. Chang, “Exploring a long short-term memory based encoder-decoder framework for multi-step-ahead flood forecasting,” *Journal of Hydrology*, vol. 583, p. 124631, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0022169420300913>
- [15] S. S. Kumar, D. M. Abraham, M. R. Jahanshahi, T. Iseley, and J. Starr, “Automated defect classification in sewer closed circuit television inspections using deep convolutional neural networks,” *Automation in Construction*, vol. 91, pp. 273–283, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0926580517309767>
- [16] J. C. Cheng and M. Wang, “Automated detection of sewer pipe defects in closed-circuit television images using deep learning techniques,” *Automation in Construction*, vol. 95, pp. 155–171, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0926580518303273>
- [17] D. Li, A. Cong, and S. Guo, “Sewer damage detection from imbalanced cctv inspection data using deep convolutional neural networks with hierarchical classification,” *Automation in Construction*, vol. 101, pp. 199–208, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0926580518306174>
- [18] X. Yin, Y. Chen, A. Bouferguene, H. Zaman, M. Al-Hussein, and L. Kurach, “A deep learning-based framework for an automated defect detection system for sewer pipes,” *Automation in Construction*, vol. 109, p. 102967, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0926580519307411>
- [19] P. Ferrans, M. N. Torres, J. Temprano, and J. P. Rodríguez Sánchez, “Sustainable urban drainage system (suds) modeling supporting decision-making: A systematic quantitative review,” *Science of The Total Environment*, vol. 806, p. 150447, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0048969721055248>
- [20] Y. Yang and T. F. M. Chui, “Modeling and interpreting hydrological responses of sustainable urban drainage systems with explainable machine learning methods,” *Hydrology and Earth System Sciences*, vol. 25, no. 11, pp. 5839–5858, 2021.
- [21] P. Willems, K. Arnbjerg-Nielsen, J. Olsson, and V. Nguyen, “Climate change impact assessment on urban rainfall extremes and urban drainage: Methods and shortcomings,” *Atmospheric Research*, vol. 103, pp. 106–118, 2012, rainfall in the urban context: forecasting, risk and climate change. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169809511000950>

- [22] M. y. E. A. I. Instituto de Hidrología and M. de Minas y Energía, *Atlas de radiación solar de Colombia*. Bogotá, Colombia: Ministerio de Minas y Energía; IDEAM, 2005, incluye ilustraciones, mapas y fotografías.
- [23] T. Partridge *et al.*, “Opportunities and challenges for precipitation forcing data in post-wildfire hydrologic modeling applications,” *WIREs Water*, 2024.
- [24] M. I. Brunner, “Challenges in modeling and predicting floods and droughts: A review,” *WIREs Water*, vol. 8, no. 3, 2021.
- [25] W. Zhang *et al.*, “Enhancing runoff predictions in data-sparse regions through hybrid deep learning and hydrologic modeling,” *Scientific Reports*, vol. 14, 2024.
- [26] A. Mohamed *et al.*, “Evaluation of the support vector regression (svr) and the random forest (rf) models accuracy for streamflow prediction under a data-scarce basin in morocco,” *Discover Applied Sciences*, 2024.
- [27] S. Solanki *et al.*, “Improving streamflow prediction using multiple hydrological models and machine learning methods,” *Water Resources Research*, 2025.
- [28] S. Yang *et al.*, “Machine learning in assessing the performance of hydrological models,” *Hydrology*, vol. 9, no. 1, p. 5, 2022.
- [29] M. y. E. A. I. Instituto de Hidrología, “Escenarios de cambio climático de la cuarta comunicación de colombia,” Bogotá, D.C., Colombia, 2024, con asistencia técnica del proyecto GEF-CBIT "Transparencia Climática Colombia", financiado por el Fondo Mundial para el Medio Ambiente (GEF), liderado técnicamente por el Ideam y el Ministerio de Ambiente y Desarrollo Sostenible, con el PNUD como agencia implementadora y la Fundación Natura como socio ejecutor. [Online]. Available: <https://visualizador.ideam.gov.co/portal/apps/storymaps/stories/660ec48de9454157b54adc074b1f38fd>
- [30] M. Hegnauer and L. Bouaziz, “Hydrological impacts of climate change in bhutan: Preparing a climate rationale for the green climate fund,” Deltares, Delft, Netherlands, Technical Report 11207685-002-ZWS-0001, May 2023, final report prepared for FAO-RAP. Project number: 11207685-002. Keywords: Climate rationale, Bhutan, floods, droughts, Wflow, mountain hydrology, climate adaptation. [Online]. Available: https://publications.deltares.nl/11207685_002_0001.pdf
- [31] R. P. Allan, M. Barlow, M. P. Byrne, A. Cherchi, H. Douville, H. J. Fowler, T. Y. Gan, A. G. Pendergrass, D. Rosenfeld, A. L. S. Swann, L. J. Wilcox, and O. Zolina, “Advances in understanding large-scale responses of the water cycle to climate change,” *Annals of the New York Academy of Sciences*, vol. 1472, no. 1, pp. 49–75, 2020, published Version, Creative Commons: Attribution 4.0 (CC-BY), Open Access. [Online]. Available: <https://centaur.reading.ac.uk/89947/>
- [32] S. I. Seneviratne, X. Zhang, M. Adnan, W. Badi, C. Dereczynski, A. Di Luca, S. Ghosh, I. Iskandar, J. Kossin, S. Lewis, F. Otto, I. Pinto, M. Satoh, S. M. Vicente-Serrano, M. Wehner, and B. Zhou, “Weather and climate extreme events in a changing climate,” in *Climate Change 2021:*

- The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou, Eds. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 2021, pp. 1513–1766.
- [33] Economic Commission for Latin America and the Caribbean (ECLAC), “Climate change and water management in latin america and the caribbean,” United Nations Economic Commission for Latin America and the Caribbean, Santiago, Chile, Technical report LC/R.1274, May 1993, document prepared by the Division of Natural Resources and Energy. Not subjected to editorial revision. Original in English. Restricted distribution. URL: <https://archivo.cepal.org/pdfs/Waterguide/LCR1274E.PDF> (accessed: 20-06-2024).
- [34] H. J. Fowler, G. Lenderink, A. F. Prein, S. Westra, R. Allan, N. Ban, R. Barbero, P. Berg, S. Blenkinsop, H. X. Do, S. B. Guerreiro, J. O. Haerter, E. Kendon, E. Lewis, C. Schaer, A. Sharma, G. Villarini, C. Wasko, and X. Zhang, “Anthropogenic intensification of short-duration rainfall extremes,” *Nature Reviews Earth and Environment*, vol. 2, pp. 107–122, 2021, accepted Version. URL: <https://centaur.reading.ac.uk/95124/> (accessed: 20-06-2024).
- [35] S. Nayak and T. Takemi, “Clausius-clapeyron scaling of extremely heavy precipitations: Case studies of the july 2017 and july 2018 heavy rainfall events over japan,” *Journal of the Meteorological Society of Japan*, vol. 98, no. 6, pp. 1147–1162, 2020, special Edition on Extreme Rainfall Events in 2017 and 2018. Manuscript received 8 September 2019, in final form 3 July 2020.
- [36] H. Douville, K. Raghavan, J. Renwick, R. Allan, P. Arias, M. Barlow, R. Cerezo-Mota, A. Cherchi, T. Gan, J. Gergis, D. Jiang, A. Khan, W. Pokam Mba, D. Rosenfeld, J. Tierney, and O. Zolina, “Water cycle changes,” in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, V. Masson-Delmotte, P. Zhai, A. Pirani, S. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. Matthews, T. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou, Eds. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 2021, pp. 1055–1210.
- [37] T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley, Eds., *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 2013.
- [38] T. R. Oke, G. Mills, A. Christen, and J. A. Voogt, *Urban Climates*. Cambridge University Press, 2017.
- [39] K. Beven, *Environmental Modelling: An Uncertain Future?* Routledge, 2009.
- [40] I. Rodríguez-Iturbe and A. Porporato, *Ecohydrology of Water-Controlled Ecosystems: Soil Moisture and Plant Dynamics*. Cambridge University Press, 2004.

