



UNIVERSIDAD TÉCNICA DE COTOPAXI

FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS

CARRERA DE INGENIERÍA HIDRÁULICA

PROYECTO DE INVESTIGACIÓN

**“ANÁLISIS COMPARATIVO DE MÉTODOS PARA LA ESTIMACIÓN DE
DATOS FALTANTES DE PRECIPITACIONES DIARIAS”**

PLAN DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE
INGENIERO HIDRÁULICO

AUTOR:

Jhon Israel Zumba Yupa

TUTOR:

Ing. Xiomara Alejandra Zambrano Navarrete

LATACUNGA – ECUADOR

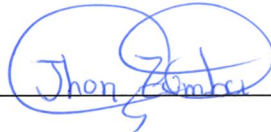
Enero -2026

Latacunga, 12/Marzo/2026

DECLARACIÓN DE AUTORÍA

Yo declaro Jhon Israel Zumba Yupa ser autor (a) del proyecto de titulación “**ANÁLISIS COMPARATIVO DE MÉTODOS PARA LA ESTIMACIÓN DE DATOS FALTANTES DE PRECIPITACIONES DIARIAS**”, siendo la Ing. Xiomara Alejandra Zambrano Navarrete Mg. tutor (a) del presente trabajo de titulación; y eximo expresamente a la Universidad Técnica de Cotopaxi y a sus representantes legales de posibles reclamos o acciones legales.

Además, certifico que las ideas, conceptos, procedimientos y resultados vertidos en el presente trabajo de titulación, son de mi exclusiva responsabilidad.



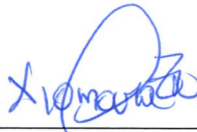
Jhon Israel Zumba Yupa

CC._0605814821

Latacunga, 12/Marzo/2026

AVAL DEL TUTOR DEL PROYECTO DE INVESTIGACIÓN

En calidad de Tutor del Trabajo de Investigación sobre el título: “ANÁLISIS COMPARATIVO DE MÉTODOS PARA LA ESTIMACIÓN DE DATOS FALTANTES DE PRECIPITACIONES DIARIAS”, propuesto por el o la estudiante **Jhon Israel Zumba Yupa** de la Carrera de **Hidráulico**, considero que dicho proyecto de titulación cumple con los requerimientos metodológicos y aportes científico-técnicos suficientes para ser sometidos al tribunal de lectores.



Ing. Xiomara Alejandra Zambrano
Naverrete, MSc.
C.C.: 1313058453
TUTOR


Latacunga, 12/Marzo/2026

AVAL DE APROBACIÓN DE LECTORES

Cumpliendo con el Reglamento de Titulación de la Universidad Técnica de Cotopaxi, en calidad de Lectores de Tribunal de Proyecto de Investigación con el Título “ANÁLISIS COMPARATIVO DE MÉTODOS PARA LA ESTIMACIÓN DE DATOS FALTANTES DE PRECIPITACIONES DIARIAS”, propuesto por el estudiante Jhon Israel Zumba Yupade la Carrera de Hidráulica, me permito indicar que el estudiante ha concluido todas las observaciones y realizado las correcciones señaladas por el Tribunal de Lectores, además de validar el funcionamiento de la propuesta (aplica para propuesta tecnológica, por lo cual presentamos el Aval de aprobación del Proyecto de Titulación correspondiente a la modalidad presencial en virtud de lo cual el postulante puede presentarse a la Defensa de su Proyecto de Titulación.


Particular que pongo en su conocimiento para los fines legales pertinentes.

Atentamente,



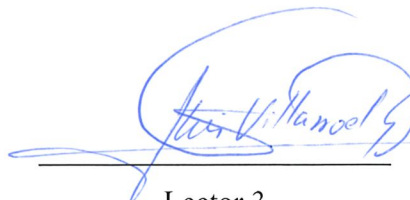
Lector 1 (presidente)

Nombre: Giler Ormaza
Andy Miguel MSc
CC: 1312820325



Lector 2

Nombre: Cusme Intriago, Rudys Rafael
CC: 1313770891



Lector 3

Nombre: Jorge Luis Villarroel
Guerrero
CC: 0502943491

h) La traducción, adaptación, arreglo u otra transformación del trabajo de grado con fines académicos y de consulta.

i) La importación al territorio nacional de copias del trabajo de grado hechas sin autorización del titular del derecho por cualquier medio incluyendo mediante transmisión.

j) Cualquier otra forma de utilización del trabajo de grado que no está contemplada en la ley como excepción al derecho patrimonial.

CLÁUSULA QUINTA. - El presente contrato se lo realiza a título gratuito por lo que LA CESIONARIA no se halla obligada a reconocer pago alguno en igual sentido **EL CEDENTE** declara que no existe obligación pendiente a su favor.

CLÁUSULA SEXTA. - El presente contrato tendrá una duración indefinida, contados a partir de la firma del presente instrumento por ambas partes.

CLÁUSULA SÉPTIMA. - CLÁUSULA DE EXCLUSIVIDAD. - Por medio del presente contrato, se cede en favor de LA CESIONARIA el derecho a explotar la obra en forma exclusiva, dentro del marco establecido en la cláusula cuarta, lo que implica que ninguna otra persona incluyendo **EL CEDENTE** podrá utilizarla.

CLÁUSULA OCTAVA. - LICENCIA A FAVOR DE TERCEROS. - LA CESIONARIA podrá licenciar la investigación a terceras personas siempre que cuente con el consentimiento de **EL CEDENTE** en forma escrita.

CLÁUSULA NOVENA. - El incumplimiento de la obligación asumida por las partes en la cláusula cuarta, constituirá causal de resolución del presente contrato. En consecuencia, la resolución se producirá de pleno derecho cuando una de las partes comunique, por carta notarial, a la otra que quiere valerse de esta cláusula.

CLÁUSULA DÉCIMA. - En todo lo no previsto por las partes en el presente contrato, ambas se someten a lo establecido por la Ley de Propiedad Intelectual, Código Civil y demás del sistema jurídico que resulten aplicables.

CLÁUSULA UNDÉCIMA. - Las controversias que pudieran suscitarse en torno al presente contrato, serán sometidas a mediación, mediante el Centro de Mediación del Consejo de la Judicatura en la ciudad de Latacunga. La resolución adoptada será definitiva e inapelable, así como de obligatorio cumplimiento y ejecución para las partes y, en su caso, para la sociedad. El costo de tasas judiciales por tal concepto será cubierto por parte del estudiante que lo solicitare.

En señal de conformidad las partes suscriben este documento en dos ejemplares de igual valor y tenor en la ciudad de Latacunga, a los 12 días del mes de marzo del 2026.



Jhon Israel Zumba Yupa
CEDENTE

Dra. Idalia Pacheco Tigselema, Ph.D. EL
LA CESIONARIA

CONTRATO DE CESIÓN NO EXCLUSIVA DE DERECHOS DE AUTOR

Comparecen a la celebración del presente instrumento de cesión no exclusiva de obra, que celebran de una parte **ZUMBA YUPA JHON ISRAEL**, identificado con cédula de ciudadanía No.0503250193 de estado civil casado, a quien en lo sucesivo se denominará **EL CEDENTE**; y, de otra parte, la Doctora Idalia Eleonora Pacheco Tigselema, en calidad de Rectora, y por tanto representante legal de la Universidad Técnica de Cotopaxi, con domicilio en la Av. Simón Rodríguez, Barrio El Ejido, Sector San Felipe, a quien en lo sucesivo se le denominará **LA CESIONARIA** en los términos contenidos en las cláusulas siguientes:

ANTECEDENTES: CLÁUSULA PRIMERA. – EL CEDENTE es una persona natural estudiante de la carrera de Ingeniería Hidráulica, titular de los derechos patrimoniales y morales sobre el trabajo de grado “**ANÁLISIS COMPARATIVO DE MÉTODOS PARA LA ESTIMACIÓN DE DATOS FALTANTES DE PRECIPITACIONES DIARIAS**”, la cual se encuentra elaborada según los requerimientos académicos propios de la Facultad; y, las características que a continuación se detallan:

Historial Académico

Inicio de la carrera: Octubre 2020 – Marzo 2021

Finalización de la carrera: Octubre 2025 – Marzo 2026

Aprobación en Consejo Directivo: 06 de octubre del 2025

Tutor: Ing. Xiomara Alejandra Zambrano Navarrete, MSc

Tema: “**ANÁLISIS COMPARATIVO DE MÉTODOS PARA LA ESTIMACIÓN DE DATOS FALTANTES DE PRECIPITACIONES DIARIAS**”

CLÁUSULA SEGUNDA. - LA CESIONARIA es una persona jurídica de derecho público creada por ley, cuya actividad principal está encaminada a la educación superior formando profesionales de tercer y cuarto nivel normada por la legislación ecuatoriana la misma que establece como requisito obligatorio para publicación de trabajos de investigación de grado en su repositorio institucional, hacerlo en formato digital de la presente investigación.

CLÁUSULA TERCERA. - Por el presente contrato, **EL CEDENTE** autoriza a **LA CESIONARIA** a explotar el trabajo de grado en forma exclusiva dentro del territorio de la República del Ecuador.

CLÁUSULA CUARTA. - OBJETO DEL CONTRATO: Por el presente contrato **EL CEDENTE**, transfiere definitivamente a **LA CESIONARIA** y en forma exclusiva los siguientes derechos patrimoniales; pudiendo a partir de la firma del contrato, realizar, autorizar o prohibir:

f) La reproducción parcial del trabajo de grado por medio de su fijación en el soporte informático conocido como repositorio institucional que se ajuste a ese fin.

g) La publicación del trabajo de grado.

Fecha de emisión: 13 de marzo de 2026

NOTA: ESTE CERTIFICADO ES GRATUITO

Abg. Santiago Vallejo Vásquez MSc.
SECRETARIO GENERAL
CONSEJO NACIONAL ELECTORAL

Firmado digitalmente por:
SANTIAGO VALLEJO
VÁSQUEZ
Fecha: 13/3/2026
6:55:45

SANTIAGO
VALLEJO
VÁSQUEZ

Cod: WK3LCOJUE6S



La emisión de este certificado no exime el pago de multas por no haber sufragado, integrado las Juntas Receptoras del Voto; y, por no asistir a capacitarse para conformar la Junta Receptora del Voto.

Al ciudadano/a: ZUMBA YUPA JHON ISRAEL, portadora de la cédula de identidad, ciudadanía o pasaporte Nro. 0605814821; se le extiende el presente **CERTIFICADO PROVISIONAL**, correspondiente al proceso electoral "Referendum y Consulta Popular 2025", efectuado el 16 de noviembre del 2025; este documento es **VÁLIDO a partir del 17 de noviembre de 2025 hasta el 31 de marzo de 2026**, que puede ser presentado en cualquier trámite público o privado y sustituye al Certificado de Votación durante el periodo de vigencia.

Hasta: 31 de marzo de 2026

Desde: 17 de noviembre de 2025

Periodo de Vigencia

N° 06010303

CERTIFICADO PROVISIONAL

CONSEJO NACIONAL ELECTORAL



UNIVERSIDAD TÉCNICA DE COTOPAXI

FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS

TITULO: “ANÁLISIS COMPARATIVO DE MÉTODOS PARA LA ESTIMACIÓN DE DATOS FALTANTES DE PRECIPITACIONES DIARIAS”

Autor: Zumba Yupa Jhon Israel

RESUMEN

La estimación precisa del comportamiento de la precipitación es importante para la planificación y gestión de recursos hídricos. Como objetivo de este trabajo de investigación, se llevó a cabo un análisis comparativo de métodos para el relleno de datos faltantes de precipitación diaria, estos valores frecuentemente no tienen una lectura completa de datos a lo largo del tiempo. Por ello, se seleccionaron y analizaron registros históricos de precipitación diaria, Para poder lograr esto, se creó una base de datos a partir de fuentes oficiales como el INAMHI y datos satelitales (GPM) de Giovanni Nasa. Se evaluaron seis métodos de relleno de datos faltantes: Regresión Lineal Simple (SLR), Regresión Lineal Múltiple (MLR), Regresión Ponderada Localmente (LWR), Datos Satelitales en la estimación de Precipitación (GPM), Random Forest (Rf) y Random Forest Híbrido con GPM. La escala de observación fue diaria y comprendió un período entre 1998-2018, en seis estaciones ubicadas dentro de la zona de estudio. La evaluación de los métodos se realizó mediante diferentes métricas estadísticas que miden el ajuste de los valores observados a los rellenos: Coeficiente de Correlación (CC), Error Absoluto Medio (MAE), Raíz del Error Cuadrático Medio (RMSE), Eficiencia Nash-Sutcliffe (NSE) y Sesgo Porcentual (PBias). Los resultados demuestran que los métodos con mejores resultados para el relleno de datos son Random Forest y Random Forest Híbrido con GPM. Sin embargo, los resultados no permitieron seleccionar un método debido al bajo rendimiento de las métricas estadísticas evaluadas.

Palabras clave: Relleno de datos, Precipitación diaria, Métodos de relleno, Métricas estadísticas

UNIVERSIDAD TÉCNICA DE COTOPAXI

FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS

THEME: “ANÁLISIS COMPARATIVO DE MÉTODOS PARA LA ESTIMACIÓN DE DATOS FALTANTES DE PRECIPITACIONES DIARIAS”

Author: Zumba Yupa Jhon Israel

ABSTRACT

Accurate estimation of precipitation behavior is important for water resource planning and management. As the objective of this research, a comparative analysis of methods for filling in missing daily precipitation data was carried out, as these values often do not have complete data readings over time. To this end, historical daily precipitation records were selected and analyzed. To achieve this, a database was created from official sources such as INAMHI and satellite data (GPM) from Giovanni Nasa. Six methods for filling in missing data were evaluated: Simple Linear Regression (SLR), Multiple Linear Regression (MLR), Locally Weighted Regression (LWR), Satellite Data in Precipitation Estimation (GPM), Random Forest (Rf), and Hybrid Random Forest with GPM. The observation scale was daily and covered a period between 1998 and 2018 at six stations located within the study area. The methods were evaluated using different statistical metrics that measure the fit of the observed values to the fillings: Correlation Coefficient (CC), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Nash-Sutcliffe Efficiency (NSE), and Percentage Bias (PBias). The results show that the methods with the best results for data filling are Random Forest and Hybrid Random Forest with GPM. However, the results did not allow a method to be selected due to the poor performance of the statistical metrics evaluated.

Keywords: Data filling, Daily precipitation, Filling methods, Statistical metrics

ÍNDICE

1. INFORMACIÓN GENERAL	1
2. INTRODUCCIÓN	2
2.1 SITUACIÓN PROBLEMÁTICA.....	4
2.3 FORMULACIÓN DEL PROBLEMA	4
2.3.1 Hipotesis	4
2.4 OBJETO Y CAMPO DE ACCIÓN	4
2.4.1 Objeto de Investigación	4
2.5 BENEFICIARIOS	5
2.5.1 Directo	5
2.5.2 Indirecto	5
2.6 JUSTIFICACIÓN.....	5
2.7 OBJETIVOS	5
2.7.1 Objetivo General.....	5
2.7.2 Objetivos Específicos	6
2.8 SISTEMAS DE TAREAS.....	6
3 FUNDAMENTACIÓN TEÓRICA.....	7
3.1 ANTECEDENTES	7
3.2 METEOROLOGÍA	9
3.3 FACTORES DEL CLIMA	9
3.3.1 Latitud.....	9
3.3.2 Variables Meteorológicas Fundamentales.....	9
3.3.2 Precipitación	9
3.3.3 Estaciones Meteorológicas	10
3.3.4 Estación Meteorológica Automática.....	10
3.3.5 Estación Meteorológica Convencional	10
3.4 TIPOS DE PREDICCIONES METEOROLÓGICAS.....	10
3.4.1 Diarias.....	11
3.4.2 Mensuales	11
3.4.3 Estacionales	11

3.5 PROGRAMAS PARA LA OBTENCIÓN DE DATOS E IMÁGENES	
SATELITALES	11
3.5.1 Modelos Digitales de Elevación (DEM).....	11
3.5.2 ASF SAR Data Search (NASA)	11
3.5.3 GIOVANNI (nasa)	12
3.5.4 GPM imerg	12
3.6 HERRAMIENTAS DE ANÁLISIS ESTADÍSTICO	12
3.6.1 Python	12
3.6.2 Lenguaje interpretado o de script	13
3.6.3 Python con anaconda	13
3.6.4 Visual Studio Code	13
3.6.5 Jupyter Notebook.....	14
3.6.6 Python-Master.....	14
3.6.7 Pandas	14
3.6.8 Las características de la biblioteca son:.....	14
3.6.9 Numpy	15
3.6.10 Matplotlib	15
3.6.11 Sklearn.neighbors.....	16
3.6.12 Multiprocessing o ProcessPoolExecutor	16
3.7 MÉTODOS DE IMPUTACIÓN	16
3.7.1 Regresión lineal simple (SLR).....	16
3.7.2 Regresión lineal múltiple (MLR).....	17
3.7.3 Regresión Ponderada Localmente (LWR)	17
3.7.4 Regresión Random Fores (RF)	17
3.7.5 Random Forest Híbrido con GPM.....	17
3.7.6 Global Precipitation Measurement (GPM).....	17
3.8 Evaluación y Exactitud.....	18
3.8.1 Coeficiente de Correlación de Pearson (CC).....	18
3.8.2 Eficiencia de Nash-Sutcliffe NSE.....	18
3.8.3 Error Absoluto Medio (MAE) (mm).....	18
3.8.4 Raíz del Error Cuadrático Medio (RMSE) (mm)	18
3.8.5 Sesgo Porcentual (PBIAS) (%).....	18
4 MÉTODOS Y PROCEDIMIENTOS	19

4.1	ÁREA DE ESTUDIO	19
4.2	ESQUEMA METODOLÓGICO	20
4.3	FASE 1 OBTENCIÓN DE INFORMACIÓN BIBLIOGRÁFICA	21
4.4	FASE 2 OBTENCIÓN DE DATOS DE PRECIPITACIÓN	24
4.5	FASE 3 EL TRATAMIENTO DE DATOS BRUTOS DE PRECIPITACIÓN	25
4.6	APLICACIÓN DE LOS 6 MÉTODOS PARA EL RELLENO DE DATOS FALTANTES DE PRECIPITACIÓN DIARIA	26
4.6.1	Regresión Lineal Simple (SLR)	26
4.6.2	Regresión Lineal Múltiple (MLR)	27
4.6.3	Regresión Ponderada Localmente (LWR)	27
4.6.4	Datos Satelitales en la Imputación de Precipitación (GPM)	29
4.6.5	Random Forest (Rf)	29
4.6.6	Random Forest Híbrido con GPM	32
4.7	DIAGRAMA DE EJECUCIÓN DEL CODIGO EN PYTHON MASTER EN EL ENTORNO DE VISUAL CODE	34
4.8	FASE 5 MÉTRICAS DE EVALUACIÓN	35
4.8.1	Coefficiente de Correlación (CC)	35
4.8.2	Error Absoluto Medio (MAE)	36
4.8.3	Raíz del Error Cuadrático Medio (RMSE)	36
4.8.4	Eficiencia Nash-Sutcliffe (NSE)	37
4.8.5	Sesgo Porcentual (PBias)	37
5.	ANÁLISIS DE RESULTADO	38
5.1	DISCUSIÓN DE LA INFORMACIÓN BIBLIOGRÁFICA	38
5.2	GRÁFICAS DE DATOS FALTANTES VS RELLENANDO	38
5.3	GRÁFICAS DE CAJAS	55
5.3.1	Gráficas de cajas GPM	56
5.3.2	Gráficas de cajas por el método LWR	62
5.3.3	Gráficas de cajas por el método MLR	68
5.3.4	Gráficas de cajas RF_Hybrid	72
5.3.5	Gráficas de caja RF	76
5.3.6	Gráficas del método SLR	80

5.4 COMPARACIÓN DE LOS RESULTADOS OBTENIDOS DESPUÉS DE APLICAR LOS DIFERENTES MÉTODOS Y EVALUARLOS POR LAS DIFERENTES MÉTRICAS PARA EL RELLENO DE DATOS FALTANTES.....	84
5.5 ANÁLISIS DE LOS RESULTADOS DE LAS ESTACIONES DE PRUEBA	86
5.5.1 Estación el Carmen	86
5.5.2 Estación LA TEODOMIRA.....	86
5.5.3 Estación ZAPOTE	86
5.5.4 Estación JUNÍN	87
5.5.5 Estación ATASCOSO (VÍA PICHINCHA)	87
5.5.6 Estación 24 DE MAYO (JABONCILLO)	87
5.5.7 Análisis Agregado de Desempeño General.....	88
6 CONCLUSIONES Y RECOMENDACIONES.....	90
7 BIBLIOGRAFÍA	92
8 ANEXOS.....	97

ÍNDICE DE TABLAS

Tabla 1 Modalidad de titulación	1
Tabla 2 Campos de la Ciencia y Tecnología UNESCO.....	1
Tabla 3 Código de la UNESCO para la ciencia.....	4
Tabla 4 Sistemas de Tareas	6
Tabla 5 Matplotlib características[52]	15
Tabla 6. Coordenadas de la zona de estudio.	19
Tabla 7. Metodos para la estimación de datos faltantes[65].....	22
Tabla 8 Estaciones meteorológicas estudiadas	25
Tabla 9 Valores que se agregan para el entrenamiento de (RF).....	31
Tabla 10 Diferencia entre RF y RF_Hybrid	32
Tabla 11 Valores que se agregan para el entrenamiento de (RF HYBRD).....	33
Tabla 12 coeficiente de correlación[68]	35
Tabla 13 Métricas de evaluación MAE [68].....	36

Tabla 14 Métricas de evaluación RMSE [68].....	36
Tabla 15 evaluación de métricas NSE [68]	37
Tabla 16 Métricas de evaluación PBIAS [68]	37
Tabla 17 Resultados observados de la serie de datos rellenos por diferentes métodos M0160	40
Tabla 18 Resultados observados de la serie de datos rellenos por diferentes métodos M1208	43
Tabla 19 Resultados observados de la serie de datos rellenos por diferentes métodos M0452	46
Tabla 20 Resultados observados de la serie de datos rellenos por diferentes métodos 4062	48
Tabla 21 Resultados observados de la serie de datos rellenos por diferentes métodos 122051	
Tabla 22 Resultados de la evaluación de la métrica M0447.....	54
Tabla 23 Análisis del método MLR de observado vs relleno	71
Tabla 24 Análisis del método de observado vs relleno RF Hibrid.....	75
Tabla 25 Análisis del método de observado vs relleno RF	79
Tabla 26 Análisis del método de observado vs relleno SLR.....	83
Tabla 27. Índice de rendimiento (CC) de los seis métodos para completar los valores faltantes	84
Tabla 28. Índice de rendimiento (NSE) de los seis métodos para completar los valores faltantes.....	84
Tabla 29. Índice de rendimiento (MAE) de los seis métodos para completar los valores faltantes.....	84
Tabla 30. Índice de rendimiento (RMSE) de los seis métodos para completar los valores faltantes.....	85
Tabla 31. Índice de rendimiento (PBIAS) de los seis métodos para completar los valores faltantes.....	85
Tabla 32 Ranking General: Métodos de Mejor a Peor Desempeño	88
Tabla 33 Desempeño General.....	89
Tabla 34 evaluación métrica de todas las estaciones a escala diaria	98

Índice de Figuras

Figura 1. Mapa de ubicación de la zona de estudio.....	20
Figura 2. Esquema Metodológico.....	21
Figura 3. Esquema de la obtención de la información.	21
Figura 4. Obtención de datos de precipitación diarias.....	24
Figura 5. Esquema de tratamiento de datos crudos de precipitación.....	25
Figura 6 Diagrama de ejecución del sistema python master	¡Error! Marcador no definido.
Figura 7. Métodos para el relleno de datos faltantes	26
Figura 8. Métricas de evaluación estadísticas	35
Figura 9. Regresión Lineal Simple (SLR) M0160	39
Figura 10. Regresión Lineal Múltiple (MLR) M0160.....	39
Figura 11. Regresión Ponderada Localmente (LWR) M0160	39
Figura 12. Regresión Random Forest (Rf) M0160.....	39
Figura 13. Random Forest Híbrido con GPM M0160.....	40
Figura 14. GPM M0160	40
Figura 15 .Regresión Lineal Simple (SLR) M1208	41
Figura 16 .Regresión Lineal Múltiple (MLR) M1208.....	42
Figura 17 .Regresión Ponderada Localmente (LWR) M1208	42
Figura 18. Regresión Random Forest (Rf) M1208.....	42
Figura 19 .Random Forest Híbrido con GPM M1208.....	42
Figura 20 GPM M1208	43
Figura 21. Regresión Lineal Simple (SLR) M0452	44
Figura 22. Regresión Lineal Múltiple (MLR) M0452.....	44
Figura 23. Regresión Ponderada Localmente (LWR) M0452	45
Figura 24 .Regresión Random Forest (Rf) M0452.....	45
Figura 25. Random Forest Híbrido con GPM M0452.....	45
Figura 26. GPM M0452	45
Figura 27. Regresión Lineal Simple (SLR) 4062	47
Figura 28. Regresión Lineal Múltiple (MLR) 4062	47
Figura 29 .Regresión Ponderada Localmente (LWR) 4062	47
Figura 30. Regresión Random Forest (Rf) 4062	48
Figura 31 .Random Forest Híbrido con GPM 4062	48
Figura 32. Datos Satelitales en la Imputación de Precipitación GPM 4062.....	48
Figura 33. Regresión Lineal Simple (SLR) M1220	50

Figura 34 .Regresión Lineal Múltiple (MLR) M1220.....	50
Figura 35 .Regresión Ponderada Localmente (LWR) M1220	50
Figura 36. Regresión Random Forest (Rf) M1220.....	50
Figura 37. Random Forest Híbrido con GPM M1220.....	51
Figura 38.Datos Satelitales en la Imputación de Precipitación GPM M1220	51
Figura 39. Regresión Lineal Simple (SLR).....	53
Figura 40. Regresión Lineal Múltiple (MLR)	53
Figura 41. Regresión Ponderada Localmente (LWR)	53
Figura 42. Regresión Random Forest (Rf)	53
Figura 43. Random Forest Híbrido con GPM	54
Figura 44. Datos Satelitales en la Imputación de Precipitación GPM.....	54
Figura 45. Gráficas de cajas de observado vs rellenado M0160(GPM).....	56
Figura 46 .Gráficas de cajas de observado vs rellenado M0452 (GPM).....	57
Figura 47. Gráficas de cajas de observado vs rellenado M0447 (GPM).....	58
Figura 48. Gráficas de cajas de observado vs rellenado M1208 (GPM).....	59
Figura 49. Gráficas de cajas de observado vs rellenado M0462 (GPM).....	60
Figura 50. Gráficas de cajas de observado vs rellenado M1220 (GPM).....	61
Figura 51. Gráficas de cajas de observado vs rellenado M0160 LWR.....	62
Figura 52. Gráficas de cajas de observado vs rellenado M0452 LWR.....	63
Figura 53. Gráficas de cajas de observado vs rellenado M0447 LWR.....	64
Figura 54. Gráficas de cajas de observado vs rellenado M1208 LWR.....	65
Figura 55. Gráficas de cajas de observado vs rellenado M0462 LWR.....	66
Figura 56. Gráficas de cajas de observado vs rellenado M1220 LWR.....	67
Figura 57. Gráficas de cajas de observado vs rellenado M0160 MLR.....	68
Figura 58. Gráficas de cajas de observado vs rellenado M0452 MLR.....	68
Figura 59. Gráficas de cajas de observado vs rellenado M0447 MLR.....	69
Figura 60. Gráficas de cajas de observado vs rellenado M1208 MLR.....	69
Figura 61. Gráficas de cajas de observado vs rellenado M0462 MLR.....	70
Figura 62 Gráficas de cajas de observado vs rellenado M01220 MLR.....	70
Figura 63 Gráficas de cajas de observado vs rellenado M0160 RF Hibrid.....	72
Figura 64 Gráficas de cajas de observado vs rellenado M0452 RF Hibrid.....	72
Figura 65 Gráficas de cajas de observado vs rellenado M0447 RF Hibrid.....	73
Figura 66 Gráficas de cajas de observado vs rellenado M1208 RF Hibrid.....	73

Figura 67. Gráficas de cajas de observado vs rellenado M0462 RF Hibrid.....	74
Figura 68. Gráficas de cajas de observado vs rellenado M1220 RF Hibrid.....	74
Figura 69 .Gráficas de cajas de observado vs rellenado M0160 RF	76
Figura 70. Gráficas de cajas de observado vs rellenado M0452 RF	76
Figura 71 . Gráficas de cajas de observado vs rellenado M0447 RF	77
Figura 72.Gráficas de cajas de observado vs rellenado M1208 RF	77
Figura 73.Gráficas de cajas de observado vs rellenado M0462 RF	78
Figura 74.Gráficas de cajas de observado vs rellenado M01220 RF	78
Figura 75.Análisis del método de observado vs rellenado M0160 SLR	80
Figura 76.Análisis del método de observado vs rellenado M0452 SLR	80
Figura 77. Análisis del método de observado vs rellenado M0447 SLR	81
Figura 78.Análisis del método de observado vs rellenado M01208 SLR	81
Figura 79.Análisis del método de observado vs rellenado M0462 SLR	82
Figura 80.Análisis del método de observado vs rellenado M1220 SLR	82
Figura 81 Carpetas con los datos de las extracciones Satelitales	97
Figura 82 Escrip.py que se utilizaron para la descarga de datos y procesamiento.....	97
Figura 83 Datos entregados por el INAMHI.....	97
Figura 84 Carpetas con los resultados obtenidos de Visual Code	97
Figura 85 Escrip.py Que se utilizaron para todo el procesamiento de datos y la obtención de resultados de los métodos y métricas	98

ÍNDICE DE ECUACIONES

Ecuación 1 Regresión Lineal Simple.....	26
Ecuación 2 Regresión Lineal Múltiple (MLR).....	27
Ecuación 3. Vecinos cercanos.....	27
Ecuación 4. Regresión Ponderada Localmente	28
Ecuación 5.Ventana temporal	28
Ecuación 6 Elevación mediante Derivadas Parciales	29
Ecuación 7. Ecuación de pesos por distancia	29
Ecuación 8 Regresión Random Forest	30
Ecuación 9 Reducción de variancia.....	30

Ecuación 10 Número de predictores.....	31
Ecuación 11 GPM como predictores	32
Ecuación 12 Coeficiente de Correlación	35
Ecuación 13 Error Absoluto Medio	36
Ecuación 14 Raíz del Error Cuadrático Medio.....	36
Ecuación 15 Eficiencia Nash-Sutcliffe.....	37
Ecuación 16 Sesgo Porcentual.....	37

1. INFORMACIÓN GENERAL

Tema del proyecto: Análisis comparativo de métodos para la estimación de datos faltantes de precipitaciones diarias.

Tabla 1 Modalidad de titulación

MODALIDAD DE TITULACIÓN	HOMOLOGACIONES INFORME FINAL DE TITULACIÓN	SELECCIÓN
Proyecto de investigación	Informe de Proyecto de investigación	x
	Artículo científico	
	Patente, utilidad, Modelo de Certificado de propiedad intelectual.	

Equipo de Trabajo del Trabajo de Titulación

- Jhon Israel Zumnba Yupa
- Tutora de titulación: ING. Xiomara Alejandra Zambrano Navarrete

Área de Conocimiento:

Tabla 2 Campos de la Ciencia y Tecnología UNESCO

25 Ciencias de la Tierra y el Espacio	2508 Hidrología	2509 Meteorología	2509.04 Hidrometeorológica
			2509.09 Predicción Numérica Meteorológica
			2509.11 Predicción Operacional Meteorológica

Trabajo de titulación vinculado al proyecto: Gestión y uso del agua: Soluciones para comunidades de la zona 3

Línea de investigación: Meteorología, Hidrología, Mecánica de fluidos, Sistema y obras Hidráulicas.

Sub líneas de investigación de la Carrera: Gestión y manejo sostenible y/o sustentable del recurso hídrico.

2. INTRODUCCIÓN

La precipitación es un componente vital en el ciclo del agua a nivel global, que regula los sistemas climáticos[1], y controla la disponibilidad de agua en la superficie y también las condiciones de recarga tanto espacial como temporal[2], también influye para el balance hídrico[3]. Por lo tanto, los patrones de precipitación son esenciales, permitiendo la planificación y gestión de este recurso[4], para la generación de proyectos[5], como la energía de las hidroeléctricas, la agricultura y, el punto más importante, la mitigación de las sequías[6]. Los datos meteorológicos se obtienen directamente de las estaciones meteorológicas, estos instrumentos permiten medir y registrar en tiempo real variables atmosféricas como la temperatura, humedad, viento y precipitación, por lo cual proporciona información directa de las mediciones en un lugar predeterminado[7], sin embargo, algunas de las estaciones meteorológicas no tienen una lectura completa[8]. Siendo un problema frecuente de los datos de precipitación que es la existencia de datos faltantes, es decir, puntos en el tiempo en los que no tienen ninguna medición de los datos[9].

Los modelos para rellenar los valores faltantes en escalas temporales pequeñas son escasos y se ven afectados por una baja capacidad de predicción debido a los altos niveles de variabilidad [10]. Por eso se utilizan técnicas comunes para el relleno de datos faltantes que son los métodos de ponderación por pesos, los modelos de regresión lineal y los vectores regionales[11], también con el desarrollo de métodos de inteligencia artificial se han ido implementando algoritmos de aprendizaje automático [12]. Estos métodos se han utilizado cada vez más debido a que su aplicabilidad es tanto simple como factible. Lo que constituye un enfoque consolidado para la imputación de datos pluviométricos faltantes[13]. Fundamentándose en las correlaciones espaciales entre estaciones meteorológicas para estimar valores ausentes mediante relaciones matemáticas predictivas [14]. Basados en datos de teledetección, como los satélites, se han consolidado como una alternativa en el entorno científico[15], las estimaciones de los satélites están frecuentemente sesgadas por lo que se dificulta utilizar sin un previo procesamiento o corrección de sesgo [16]. Para implementar el algoritmo de Integrated Multi-satellitE Retrievals for GPM (IMERG) de la página Geovanni NASA, que permite obtener un producto de Global Precipitation Measurement (GPM), que representa una tecnología satelital [17], que ofrece ventajas particulares en regiones con limitada densidad de estaciones terrestres junto a estrategias de completado de series temporales pluviométricas obtenidas del INAMHI[18].

El alcance de este proyecto de investigación es el análisis y comparación de diferentes métodos de rellenos de datos, desde métodos sencillos como lo son, la regresión lineal simple (SLR), regresión lineal múltiple (MLR) y Regresión Ponderada Localmente (LWR) hasta métodos más complejos como son técnicas basadas en el uso de inteligencia artificial con métodos como Random Forest (Rf), Random Forest Híbrido y Datos Satelitales en la Imputación de Precipitación (GPM) [19]. Para garantizar el grado de precisión y las limitaciones de los datos generados mediante el proceso de relleno de los datos faltantes de precipitación diaria es importante evaluarlas mediante el uso de métricas estadísticas de rendimiento [20]. Las métricas permiten cuantificar la diferencia entre las estimaciones y los valores reales del INAMHI son el Coeficiente de Correlación (CC), Error Absoluto Medio (MAE), Raíz del Error Cuadrático Medio (RMSE), Eficiencia Nash-Sutcliffe (NSE) y Sesgo Porcentual (PBias), facilitando la identificación de que método aplicado es el que tiene mejores resultados.

2.1 SITUACIÓN PROBLEMÁTICA

En la provincia de Manabí, específicamente en el cantón Chone, la caracterización hidroclimatológica y la planificación de recursos hídricos enfrentan serias limitaciones derivadas de la presencia de datos faltantes en las series de datos temporales de precipitación registradas por la red de estaciones meteorológicas del INAMHI. Esta problemática se origina por múltiples factores que incluyen fallas técnicas en los equipos de medición, interrupciones en el mantenimiento de las estaciones, eventos climáticos extremos que afectan la operatividad de los instrumentos, cambios en la ubicación de estaciones, y limitaciones presupuestarias que restringen la cobertura espacial y temporal del monitoreo pluviométrico. La ausencia de datos compromete la precisión de análisis estadísticos fundamentales, limitando así la capacidad de toma de decisiones en sectores críticos. En contexto, la necesidad de desarrollar y validar metodologías robustas que integren técnicas estadísticas con tecnologías satelitales emergentes se presenta como una oportunidad estratégica para mejorar significativamente la estimación de datos pluviométricos faltantes y fortalecer la base científica para la gestión integral de los recursos hídricos en la zona de estudio.

2.3 FORMULACIÓN DEL PROBLEMA

La existencia de datos faltantes para el relleno de datos de precipitación diaria.

2.3.1 HIPOTESIS

De los métodos de regresión lineal simple, múltiple, regresión ponderada por distancia, GPM, y Random Forest cuál de ellas tiene la mejor precisión, mediante el uso de métricas estadísticas de desempeño de (RMSE, MAE, NSE, PBIAS).

2.4 OBJETO Y CAMPO DE ACCIÓN

2.4.1 Objeto de Investigación

Relleno de datos faltantes de precipitación diaria.

Tabla 3 Código de la UNESCO para la ciencia

2508 hidrología	
2508.10	Precipitación
2509	Meteorología
2509.04	Hidrometeorología
2509.09	Predicción numérica meteorológica

3305.15	Ingeniería Hidráulica
---------	-----------------------

2.5 BENEFICIARIOS

2.5.1 Directo

Cantón Chone

2.5.2 Indirecto

Comunidad Científica y Académica

2.6 JUSTIFICACIÓN

La presencia de en los datos faltantes para el relleno de precipitación diaria es un problema que tiene la mayor parte del Ecuador por falta de equipos, interrupciones al momento de hacer el mantenimiento entre otros factores que interrumpe la continuidad de la lectura, necesaria para estudios climáticos precisos. Sin una base de datos completa, cualquier análisis de riesgos o planificación hídrica carece de fiabilidad, ya que la alta variabilidad de las lluvias hace que los métodos de estimación no siempre sean los más adecuados por eso se busca el que mejor desempeño basado en los siguientes parámetros que son la longitud, latitud y la orografía[21].

La investigación se centró en analizar y comparar seis métodos de relleno de datos faltantes, mediante un entorno computacional en Visual Code con el lenguaje de Python[22] Utilizando diferentes librerías de análisis de datos como Pandas, Numpy, array y sklearn. linear_model para el procesamiento de datos robustos[23] con el fin de identificar cual es el método con mayor precisión que demuestra mediante la evaluación de las métricas estadísticas.

El resultado de este trabajo se centra en saber que método utilizado fue mejor para el relleno de datos faltantes de precipitación diaria. De acuerdo con los objetivos del estudio, los resultados permitieron determinar si alguno de los métodos que se aplicó para el relleno de datos funciona. Con base a los resultados obtenidos, es posible analizar qué procedimiento es el más adecuado para futuras investigaciones que puedan completar la información faltante.

2.7 OBJETIVOS

2.7.1 Objetivo General

- Analizar diferentes métodos de estimación de datos faltantes en series de precipitaciones diarias, identificando el método que presenta mayor precisión y confiabilidad mediante el uso de métricas estadísticas de validación.

2.7.2 Objetivos Específicos

- Realizar una revisión bibliográfica de los métodos utilizados para la estimación de datos faltantes en series de precipitación diaria.
- Aplicar distintos métodos de estimación de datos faltantes en series de precipitación diaria, evaluando su desempeño en función de las características de los datos disponibles.
- Evaluar la validez y precisión de los métodos aplicados mediante el uso de métricas estadísticas de desempeño, tales como error medio absoluto (MAE), raíz del error cuadrático medio (RMSE), coeficiente de evaluación (R) y eficiencia de Nash-Sutcliffe (NSE)

2.8 SISTEMAS DE TAREAS

A continuación, en la Tabla 4 se presenta el sistema de tareas empleados para el desarrollo de la investigación.

Tabla 4 Sistemas de Tareas

Objetivos específicos	Actividades	Resultados esperados	Técnicas, Medios e Instrumentos
Realizar una revisión bibliográfica de los métodos utilizados para la estimación de datos faltantes en series de precipitación diaria.	Revisión bibliográfica de artículos sobre imputación de datos (Regresión Lineal, múltiple, ponderación, GPM y RF). Análisis comparativo de metodologías aplicadas en regiones similares para identificar la más apropiada.	Resultados fundamentados y selección justificada de métodos.	Bases de datos: Google Scholar, IEEE, Scopus, Revistas de alto impacto Página oficial del INAMHI.
Aplicar distintos métodos de estimación de datos faltantes en series de precipitación diaria, evaluando su desempeño en función de las	Preprocesamiento de datos del INAMHI (CSV) y datos satelitales GPM-IMERG (NASA Giovanni). Tratamiento de datos mediante scripts en Python	Base de datos meteorológica de las estaciones requeridas depurada. Descarga de los datos de los	Software: Q GIS 3.40.6, VS Code, Python conda Librerías: Pandas, Xarray, Geopandas. Visual Code

características de los datos disponibles.	(data_processing.py, near_stn_search.py). Aplicación de los 6 métodos (MLR), (SLR), (LWR), (RF), (RF, GPM) Y (GPM)	pixeles de Geovanni. Detección de vacíos Y el relleno completo de las estaciones de estudio.	
Evaluar la validez y precisión de los métodos aplicados mediante el uso de métricas estadísticas de desempeño, tales como error medio absoluto (MAE), raíz del error cuadrático medio (RMSE), coeficiente de evaluación (CC) y eficiencia de Nash-Sutcliffe (NSE	Cálculo de métricas de desempeño: CC, NSE, MAE, RMSE y PBIAS.	Scripts de Tablas comparativas de rendimiento.	Análisis de Regresión y validación de las métricas. Librerías: Scikit-learn, Matplotlib, Seaborn.

3 FUNDAMENTACIÓN TEÓRICA

3.1 ANTECEDENTES

La precipitación es un fenómeno meteorológico por el cual el vapor de agua se condensa y desciende de la atmosfera a la superficie terrestre en forma de lluvia[24] .se toma en cuenta que en el ámbito de la hidrometeorológica la imputación de datos faltantes es importante como se muestra en la investigación titulada "Comparación de métodos para rellenar datos faltantes de precipitaciones diarias y mensuales: Modelos Estadísticos o Imputación de Recuperaciones por Satélite", que realiza un análisis exhaustivo en el estado de Goiás, Brasil. Los autores evaluaron métodos estadísticos como la Imputación Múltiple por Ecuaciones como son, la regresión lineal simple y la regresión lineal múltiple y compararon con datos satelitales IMERG[25], dando como resultados que para las escalas diarias los modelos estadísticos superan a las estimaciones satelitales, siempre que existan estaciones vecinas con alta correlación es decir que tenga mayor

dónante de datos mientras que para escalas mensuales, los productos satelitales como IMERG son una alternativa robusta, cuando la red de estaciones terrestres es escasa o presenta vacíos prolongados.

A nivel nacional, el estudio de la imputación de datos meteorológicos se ha centrado en optimizar las bases de datos del Instituto Nacional de Meteorología e Hidrología (INAMHI). En la investigación que se desarrolló en la provincia de Chimborazo, los autores Altamirano y Carrillo [26] evaluaron diversas técnicas de relleno para variables meteorológicas. En su estudio subraya que la elección de la técnica ya sea lineal, mediante variables al azar o mediante algoritmos iterativos debe basarse en el porcentaje de vacíos que se deben completar y en la autocorrelación de la serie.

En la provincia de Manabí, el autor Cabrera Estupiñán [27].analizo el régimen de precipitaciones en el área del Multipropósito Chone, se utilizó datos históricos de 1990 a 2012, que determinaron métodos para la planificación del riego. Los autores enfatizan que el análisis estadístico riguroso de la lluvia es el insumo principal para proyectos de ingeniería hidráulica en la región. Esta investigación local justifica la necesidad de completar y validar los datos faltantes de estaciones cercanas como Chone, Calceta y Canuto, utilizando las coordenadas proyectadas UTM WGS 84 Zona 17S para garantizar la precisión espacial en los modelos de estimación.

En el estudio desarrollado por los autores Ruiz y Hoyos [28],implementaron varios modelos de predicción meteorológica para eventos extremos cuya validación técnica se fundamentó en el uso riguroso de las métricas estadísticas como son RMSE, MAE y NSE. Los autores emplearon la Raíz del Error Cuadrático Medio (RMSE) y el Error Absoluto Medio (MAE) para cuantificar la precisión de las estimaciones, permitiendo identificar la sensibilidad del modelo ante variaciones abruptas en los datos climáticos. Integraron el Coeficiente de Eficiencia de Nash-Sutcliffe (NSE) como el indicador principal en datos meteorológicos, lo que permitió determinar la capacidad del modelo para replicar el comportamiento hidrológico real frente a la media histórica, asegurando así la robustez estadística necesaria para la gestión de riesgos en proyectos de ingeniería hidráulica.

Las investigaciones que se ha realizado resaltan la importancia de los diferentes métodos estadísticos para la predicción de datos faltantes y en sus diferentes métricas para determinar cuál de los 6 métodos es el que tiene mejor rendimiento.

3.2 METEOROLOGÍA

La Meteorología es la ciencia encargada del estudio de la atmósfera, de las propiedades y de los fenómenos que en ella tienen lugar. El estudio de la atmósfera se basa en el conocimiento de una serie de magnitudes, o variables meteorológicas, como la temperatura, la presión atmosférica o la humedad, las cuales varían tanto en el espacio como en el tiempo. Pero hoy día la meteorología es una ciencia avanzada, basada en nuestro conocimiento de la Física y en el uso de las más modernas tecnologías. Los meteorólogos son capaces, incluso de predecir el tiempo hasta con una semana de antelación sin apenas fallar[29].

3.3 FACTORES DEL CLIMA

Son aquellos agentes que modifican el comportamiento de los elementos del clima, y de acuerdo a su interacción, a su presencia e intensidad, es que determinan las características particulares de los diferentes tipos de clima que existen en el mundo[30].

3.3.1 Latitud

Cuanto más cerca se encuentre un lugar del ecuador menos variación habrá en la duración de los días y la energía solar incidirá más vertical y con más intensidad sobre él, por lo que más cálidas serán las temperaturas. Cuanto más lejos se encuentre ese punto del ecuador la energía solar le alcanzará con un ángulo menor, además durante el invierno los días serán cortos acentuando el frío. Si el punto se encuentra cerca de los polos los rayos solares le llegarán con un ángulo muy pequeño y solo durante algunos meses ya que en invierno allí no sale el sol. Esto hará que las temperaturas sean muy bajas todo el año[31]. Se mide en grados, minutos y segundos.

3.3.2 VARIABLES METEOROLÓGICAS FUNDAMENTALES

Son parámetros responsables de describir el estado del tiempo, son cuantificables y su comportamiento nos permitirá conocer la condición atmosférica en el que se encuentra en ese momento[28].

3.3.2 Precipitación

La Precipitación es cualquier agua meteórica recogida sobre la superficie terrestre. Esto incluye básicamente: lluvia, nieve y granizo, también se los considera al rocío y escarcha que en algunas regiones constituyen una parte pequeña pero apreciable de la precipitación total[32].

En relación a su origen, pueden distinguirse los siguientes tipos:

- Las ciclónicas

- Las de convección
- Las precipitaciones orográficas

El estudio de las precipitaciones es básico dentro de cualquier estudio hidrológico regional, para cuantificar los recursos hídricos, puesto que constituyen la principal entrada de agua a una cuenca. También es fundamental en la previsión de avenidas, diseño de obras públicas, estudios de erosión, etc.[32].

3.3.3 Estaciones Meteorológicas

Una estación meteorológica se compone de un conjunto de instrumentos climáticos, que son operados automática o manualmente. La calidad y el número de los instrumentos pueden variar ampliamente de acuerdo a la necesidad, pero generalmente se incluye cuatro o más sensores., las estaciones que están conectadas a una computadora y a internet, tienen la opción de visualizar los datos en un sitio web previamente diseñado, permitiendo de esta manera, ver en tiempo real la información captada por los sensores.[33]

3.3.4 Estación Meteorológica Automática

Las estaciones meteorológicas automáticas son aquellas que poseen un grupo de ciertos instrumentos encargados de medir automáticamente variables atmosféricas, por lo tanto, su característica principal está definida por el almacenamiento y transmisión de datos de manera continua[34] Estas estaciones pueden ser instaladas en aquellos lugares en los cuales el acceso es limitado especialmente en zonas como lo es la nuestra lo cual nos permite obtener datos en intervalos pequeños de tiempo y además son indispensables para reducir los errores cometidos por el observador.

3.3.5 Estación Meteorológica Convencional

El primer diseño de una estación meteorológica convencional se desarrolló en la segunda mitad del siglo XIX y al inicio del del siglo XX, se ha dado como resultado que con el pasar del tiempo los dispositivos meteorológicos han sufrido diversos inconvenientes para que su análisis y obtención de parámetros meteorológicos posea un margen de mínimo, dichas estaciones deberían estar controladas por un observador cuya estadía deberá ser de forma prolongada, para así tener el control y poder obtener los datos climáticos[35].

3.4 TIPOS DE PREDICCIONES METEOROLÓGICAS

Las predicciones meteorológicas se clasifican principalmente según su escala temporal las cuales pueden ser diarias, mensuales y estacionales.

3.4.1 Diarias

Las predicciones meteorológicas diarias se basan en datos actuales y Escenarios numéricos para anticipar las condiciones atmosféricas. Son fundamentales para la planificación en sectores como la agricultura, la aviación y la gestión de emergencias, proporcionando información crucial para decisiones importantes[28].

3.4.2 Mensuales

La predicción mensual examina la predictibilidad en un intervalo de 10 a 30 días, tratando cuestiones de pronóstico de tiempo a largo plazo y estacional. Desde hace diez días, las fluctuaciones oceánicas afectan la atmósfera, por lo que los Escenarios deben tener en cuenta esta interacción.[28].

3.4.3 Estacionales

Facilita la predicción del estado medio atmosférico para los meses venideros, informando acerca de las tendencias climáticas. En contraposición a una simulación climática, se fundamenta en las primeras observaciones del sistema mensuales[28].

3.5 PROGRAMAS PARA LA OBTENCIÓN DE DATOS E IMÁGENES SATELITALES

Parte de las imágenes satelitales estereoscópicas de alta resolución las cuales permiten obtener Modelos Digitales de Elevación (MDE) con alto nivel de detalle además de sus productos derivados tales como curvas de nivel, perfiles y maquetas 3D, los satelital permite la visualización 3D, que sirven para mejorar la comunicación de información espacial incluso a partir de la programación de una nueva adquisición de imágenes, utilizando satélites[36].

3.5.1 Modelos Digitales de Elevación (DEM)

Es una representación visual y matemática de los valores de altura con respecto al nivel medio del mar, que permite caracterizar las formas del relieve y los elementos u objetos presentes en el mismo, lo que se encuentran en estos valores están contenidos en un archivo de tipo ráster con estructura regular, el cual se genera un modelo digital que se puede abrir con un software especializado[37]

3.5.2 ASF SAR Data Search (NASA)

El ASF es el Centro de Archivos Activos Distribuidos (DAAC) de la NASA para datos de Radar de Apertura Sintética– Establecido en 1991 como el principal centro de procesamiento y descarga de EE. UU. [38]. Es Buscador especializado en datos SAR (radar de apertura sintética) con acceso gratuito a misiones como Sentinel-1, ALOS PALSAR y RADARSAT[39].

3.5.3 GIOVANNI (nasa)

Se deriva del acrónimo GES-DISC (Goddard Earth Science Data and Information Services Division) es el servidor que agrupa y provee información de diferentes sensores remotos, brindando series de datos, mapas, gráficos y análisis sofisticados en diferentes formatos. Además, sus herramientas permiten visualizar simulaciones sin necesidad de descargar contenidos[40].

3.5.4 GPM imerg

Pueden considerarse datos climáticos alternativos para proporcionar información sobre las precipitaciones en regiones con pluviómetros limitados. La tecnología satelital ha avanzado significativamente en los últimos años, lo que permite monitorear las precipitaciones intensas en grandes áreas a un costo relativamente bajo[41].

- Resolución espacial y resolución temporal: Es el tamaño de un píxel en el terreno. Un píxel es un bloque de información que representa el elemento individual más pequeño de una imagen. Los píxeles se denominan "elementos de imagen" y la temporal se refiere al tiempo transcurrido entre visualizaciones de la misma zona de la Tierra desde el mismo ángulo. Puede variar desde una cobertura continua para plataformas geoestacionarias, como un satélite meteorológico.

3.6 HERRAMIENTAS DE ANÁLISIS ESTADÍSTICO

Los métodos estadísticos se han convertido en una parte inseparable de la investigación moderna en la ingeniería. La investigación y la práctica basadas en la evidencia deben incorporar herramientas estadísticas en la metodología desde la etapa de planificación del estudio. Si bien se conoce bien la importancia de la estadística en la investigación, la mayoría de los investigadores carecen de un conocimiento profundo de los diversos conceptos y principios estadísticos[42].

3.6.1 Python

Python es un lenguaje de programación creado por Guido van Rossum a principios de los años 90 cuyo nombre está inspirado en el grupo de cómicos ingleses “Monty Python”. Es un lenguaje similar a Perl, pero con una sintaxis muy limpia y que favorece un código legible. Se trata de un lenguaje interpretado o de script, con tipado dinámico, fuertemente tipado, multiplataforma y orientado a objetos[43].

Es un lenguaje de programación de propósito general, independiente de la plataforma y orientado a objetos, que no fue diseñado específicamente para el análisis de datos o la computación científica, pero que, en los últimos años se ha destacado como una alternativa para tareas de análisis y visualización de grandes conjuntos de datos. El amplio y variado ecosistema de librerías del que dispone para el análisis de datos ha provocado que cada vez más personas lo utilicen para estos fines[44].

Para la aplicación de Python en la estadística descriptiva los autores han seleccionado como entorno de desarrollo a VSCode, porque ofrece una amplia gama de características y herramientas como el IntelliSense que proporciona sugerencias de código y completado automático para ayudar a escribir código más rápido y con menos errores, Depuración integrada que posibilita la depuración del código directamente desde el editor, y extensibilidad por lo que se pueden instalar una amplia variedad de extensiones y complementos muy útiles como la de Jupiter que permite a VSCode trabajar con los Notebook/lab (.ipynb) y en el entorno de las celdas de código y Markdown de Jupiter[22].

3.6.2 Lenguaje interpretado o de script

Un lenguaje interpretado o de script es aquel que se ejecuta utilizando un programa intermedio llamado intérprete, en lugar de compilar el código a lenguaje máquina que pueda comprender y ejecutar directamente una computadora. La ventaja de los lenguajes compilados es que su ejecución es más rápida. Sin embargo, los lenguajes interpretados son más flexibles y más portables[43].

3.6.3 Python con anaconda

Anaconda® es un administrador de paquetes, un administrador de entornos, una distribución de Python y una colección de más de 1,000 paquetes de código abierto. Es de código abierto y fácil de instalar, y ofrece soporte gratuito a través de su comunidad[45].

3.6.4 Visual Studio Code

Se conoce también como VS Code, que es un editor de código para programadores gratuito, de código abierto y multiplataforma. Está desarrollado por Microsoft, una compañía con una dilatada experiencia en la creación de IDEs (entornos de desarrollo integrados), que ha conseguido plasmar su larga tradición en el sector para ofrecer una herramienta ligera y práctica que la comunidad ha adoptado en masa[46].

3.6.5 Jupyter Notebook

Jupyter Notebook es una interfaz web de código abierto que permite la inclusión de texto, vídeo, audio, imágenes, así como la ejecución de código a través del navegador en múltiples lenguajes. Esta ejecución se realiza mediante la comunicación con un núcleo (Kernel) de cálculo. Aunque en principio, el equipo de desarrollo de Jupyter Notebook incluye por defecto únicamente el núcleo de cálculo Python, el carácter abierto del proyecto ha permitido aumentar el número de núcleos disponibles, incluyendo, por ejemplo, Octave, Julia, R, Haskell, Ruby, C/C++, Fortran, Java, SageMath, Scala, o también Matlab y Mathematica. Esta interfaz, agnóstica del lenguaje (de ahí su nombre al unir 3 de los lenguajes de programación de código abierto más utilizados en el ámbito científico: Julia, Python y R), puede suponer por tanto una estandarización para mostrar el contenido de cursos científicos[47].

3.6.6 Python-Master

Implementa un flujo de trabajo integral para la imputación de datos faltantes de precipitación en redes de estaciones meteorológicas. El sistema combina técnicas de regresión estadística clásica con métodos de aprendizaje automático, utilizando información espacial derivada de Modelos Digitales de Elevación (DEM) y datos satelitales GPM-IMERG como fuentes complementarias.

3.6.7 Pandas

Es una librería de Python especializada en el manejo y análisis de estructuras de datos, el paquete pandas es la herramienta más importante a disposición de los científicos y analistas de datos que trabajan en Python en la actualidad las herramientas de aprendizaje automático y visualización glamorosa pueden llamar toda la atención, pero pandas es la columna vertebral de la mayoría de los proyectos de datos[22], [48] en la computación y Ciencia de datos, pandas es una biblioteca de software escrita como extensión de Numpy para manipulación y análisis de datos para el lenguaje de programación Python. Esta librería ofrece dos de las estructuras más usadas en Data Science: la estructura Series y el DataFrame. En este curso veremos cómo crearlas, las herramientas básicas de uso y algunas de las funciones y métodos que nos permitirán extraer todo el potencial de ellas[48].

3.6.8 Las características de la biblioteca son:

- El tipo de datos Data Frame para manipulación de datos con indexación integrada. Tiene herramientas para leer y escribir datos entre estructuras de datos en memoria y formatos de archivos variados

- Permite la alineación de datos y manejo integrado de datos faltantes, la reestructuración y segmentación de conjuntos de datos, la segmentación vertical basada en etiquetas, indexación elegante, y segmentación horizontal de grandes conjuntos de datos, la inserción y eliminación de columnas en estructuras de datos.
- Puedes realizar cadenas de operaciones, dividir, aplicar y combinar sobre conjuntos de datos, la mezcla y unión de datos. Permite realizar indexación jerárquica de ejes para trabajar con datos de altas dimensiones en estructuras de datos de menor dimensión, la funcionalidad de series de tiempo: generación

3.6.9 Numpy

Es una biblioteca fundamental para la computación numérica en Python. Proporciona soporte para matrices y operaciones matemáticas de alto rendimiento[49]., Las aplicaciones de NUMPY al tratamiento de datos científicos: selección, ordenación, ajustes, interpolaciones, estadística, polinomios, funciones matemáticas, arrays con máscaras[50].

Un array es un tipo de dato que se utiliza para almacenar múltiples valores mediante un único identificador (nombre de variable). Un array contiene una colección ordenada de elementos de datos, donde cada elemento es del mismo tipo y se puede referenciar por su índice (posición)[51].

3.6.10 Matplotlib

El principal propósito es facilitarnos la tarea de visualización de nuestros datos. Entre sus puntos fuertes podríamos destacar los siguientes que son Librería de Python más extendida para realizar gráficas de datos que nos permite realizar multitud de gráficos a través de sencillas instrucciones de alto nivel, permitiendo a su vez actuar a más bajo nivel para controlar con detalle todo tipo de aspectos de los mismos.[52]

Tabla 5 Matplotlib características[52]

	Descripción
EPS	PostScript encapsulado.
JPG	Formato gráfico con método de compresión con pérdida para salida fotográfica.

PDF	Formato de documento portátil (PDF).
PNG	Portable Network Graphics (PNG), un formato de gráficos rasterizados con un método de compresión sin pérdida (más adaptable a arte lineal que el JPG).
PS	Lenguaje ampliamente utilizado en publicaciones y como formato de trabajos de impresión.
SVG	Gráficos vectoriales escalables (SVG), basado en XML.

3.6.11 Sklearn.neighbors

Proporciona funcionalidad para métodos de aprendizaje basados en vecinos, tanto supervisados como no supervisados en el aprendizaje de donantes vecinos más cercanos no supervisado es la base de muchos otros métodos de aprendizaje, en particular el aprendizaje de variedades y la agrupación espectral. El aprendizaje basado en vecinos supervisado se presenta en dos modalidades en los que se clasifican para datos con etiquetas discretas y regresión para datos con etiquetas continuas.

3.6.12 Multiprocessing o ProcessPoolExecutor

Se diseña para permitir que sus usuarios proporcionen su propio contexto de multiprocessing. El uso de un contexto específico propio dentro de una biblioteca puede generar incompatibilidades con el resto de la aplicación del usuario de la biblioteca., Documente siempre si su biblioteca requiere un método de inicio específico[23].

3.7 MÉTODOS DE IMPUTACIÓN

Las técnicas simples de imputación han sido una de las herramientas más conocidas y aceptadas para el tratamiento de la falta de respuesta las técnicas simples de imputación presentan algunas ventajas frente a la técnica de imputación múltiple como ejemplo, las técnicas simples tienen una implantación más sencilla por el contrario sufren una importante pérdida de eficiencia en comparación con las técnicas de imputación múltiple como lo principal destacamos que las técnicas simples de relleno de datos se pueden dividir en dos categorías: aleatorias y determinísticas[53]

3.7.1 Regresión lineal simple (SLR)

El análisis de regresión es una técnica estadística para investigar la relación funcional entre dos o más variables, ajustando algún modelo matemático la regresión lineal simple utiliza una sola

variable de regresión y el caso más sencillo es el modelo de línea recta. Supóngase que se tiene un conjunto de “n” pares de observaciones, se busca encontrar una recta que describa de la mejor manera cada uno de esos pares observados[54].

3.7.2 Regresión lineal múltiple (MLR)

La regresión lineal múltiple trata de ajustar modelos lineales entre una variable dependiente y más, de unas variables independientes[55].

3.7.3 Regresión Ponderada Localmente (LWR)

Es una forma local de regresión lineal que se utiliza para modelar las relaciones entre la variable dependiente y explicativas que varían espacialmente. De este modo, se ejecuta una ecuación de regresión para cada observación de la muestra[56].

3.7.4 Regresión Random Fores (RF)

Produce opcionalmente dos piezas de información adicionales, una de la medida de la importancia de las variables predictoras y una medida de la estructura interna de los datos de proximidad de los diferentes puntos de datos entre sí[57].

3.7.5 Random Forest Híbrido con GPM

Este algoritmo basado en satélites para la obtención de precipitaciones con alta resolución espaciotemporal. El algoritmo se basa en la información de precipitación de la Recuperación Integrada Multi-satelital para la Medición Global de Precipitación (GPM) (IMERG). Fue desarrollado para clasificar el área de lluvia y asignar la tasa de precipitación y con esta información se entrenaron utilizando bosques aleatorios (RF). Los resultados de la validación mostraron que la nueva técnica de obtención de precipitación supera al producto de lluvia IMERG[58].

3.7.6 Global Precipitation Measurement (GPM)

Mediante el algoritmo IMERG, la misión GPM combina datos de todos los instrumentos pasivos de microondas de la constelación de satélites GPM para proporcionar estimaciones de precipitación, que luego se fusionan e interpolan con estimaciones de observaciones infrarrojas calibradas y otros sensores potenciales con una resolución espacial y una resolución temporal de 30 min en todo el planeta

3.8 EVALUACIÓN Y EXACTITUD

3.8.1 Coeficiente de Correlación de Pearson (CC)

El coeficiente de correlación es un estadístico que proporciona información sobre la relación lineal existente entre dos variables cualesquiera. Básicamente, esta información se refiere a dos características de la relación lineal: la dirección o sentido y la cercanía o fuerza[59]

3.8.2 Eficiencia de Nash-Sutcliffe NSE

Se trata de un indicador ampliamente utilizado en la evaluación de modelos hidrológicos, caracterizado por su flexibilidad para ser aplicado a diversos tipos de modelos matemáticos[60].

El valor óptimo de NSE es uno 1 y el peor valor tiende a los números negativos. Es decir, un valor de NSE de 1 implica un ajuste perfecto entre los datos observados. El valor de cero (0) sirve como indicador de que las simulaciones son comparables con el promedio de los datos observados. Es decir, si un valor de NSE de 0 implica que las simulaciones no son mejores que si simplemente se hubiese tomado el promedio de los datos observados. Un modelo hidrológico se considera satisfactorio si $NSE > 0.50$. [61]

3.8.3 Error Absoluto Medio (MAE) (mm)

Se refiere a la desviación absoluta media de los valores interpolados en comparación con los observados. Se considera una medida precisa y robusta para verificar modelos numéricos; idealmente, sus valores deberían ser lo más cercanos o iguales a cero posible. Esta métrica es la más común que se utiliza en los modelos de regresión que calcula como la media de las diferencias absolutas entre las predicciones del modelo y los valores reales[62]

3.8.4 Raíz del Error Cuadrático Medio (RMSE) (mm)

El valor óptimo de RMSE es cero (0) y el peor valor tiende a los números positivos. Es decir, un valor de RMSE de 0 implica un ajuste perfecto entre los datos observados. No existe un valor límite estándar, pero en general un modelo hidrológico se considera satisfactorio mientras más pequeño sea RMSE[61], es la raíz cuadrada del MSE. Proporciona una medida de la dispersión de los errores en las mismas unidades que la variable objetivo[63].

3.8.5 Sesgo Porcentual (PBIAS) (%)

El valor óptimo de PBIAS es cero (0) y los peores valores tienden a ambas direcciones (negativo infinito y positivo infinito). Es decir, un valor de PBIAS de 0 implica un ajuste perfecto entre los datos observados y los datos simulados. Valores positivos de PBIAS indican sesgo de subestimación del modelo y valores negativos de PBIAS indican sesgo de

sobreestimación del modelo. Un modelo hidrológico se considera satisfactorio si $PBIAS \leq \pm 25\%$ para simulación de precipitación[61]

4 MÉTODOS Y PROCEDIMIENTOS

En este proyecto se utilizaron los métodos cuantitativos y cualitativos para abordar la problemática de estimación de precipitación. La parte cuantitativa es la principal, ya que se fundamenta en la recolección, procesamiento y análisis estadístico de los datos cronológico de lluvia diaria en milímetros (mm), utilizando métodos de regresión lineal, el algoritmo satelital y algoritmos de aprendizaje automáticos supervisados para cuantificar el grado de precisión y error mediante métricas estadísticas. Por otro el enfoque cualitativo se manifiesta en el análisis de la aplicabilidad y limitaciones de estos métodos; esto implica interpretar cómo las condiciones geográficas de Manabí y las capacidades técnicas del INAMHI.

4.1 ÁREA DE ESTUDIO

La presente investigación se realizó en la Provincia de Manabí que se encuentra en el centro de la región costera del país [64], específicamente en el Cantón Chone. Este cantón se caracteriza por un clima cálido, la temperatura predominate en el 90% de la zona de estudio que se puede observar en la Figura 1, el cual está en el rango de 24 °C y 26 °C ,y las precipitaciones presentan las siguientes características 800 y 1500 mm anuales. [64].En la Tabla 6 se muestra las coordenadas especifica de la zona de estudio.

Tabla 6. Coordenadas de la zona de estudio.

Latitud (GG° MM' SS").	Longitud (GG° MM' SS").	Altitud
0° 39' 50.91" S	80° 02' 11.44" W	39 m.s.n.m

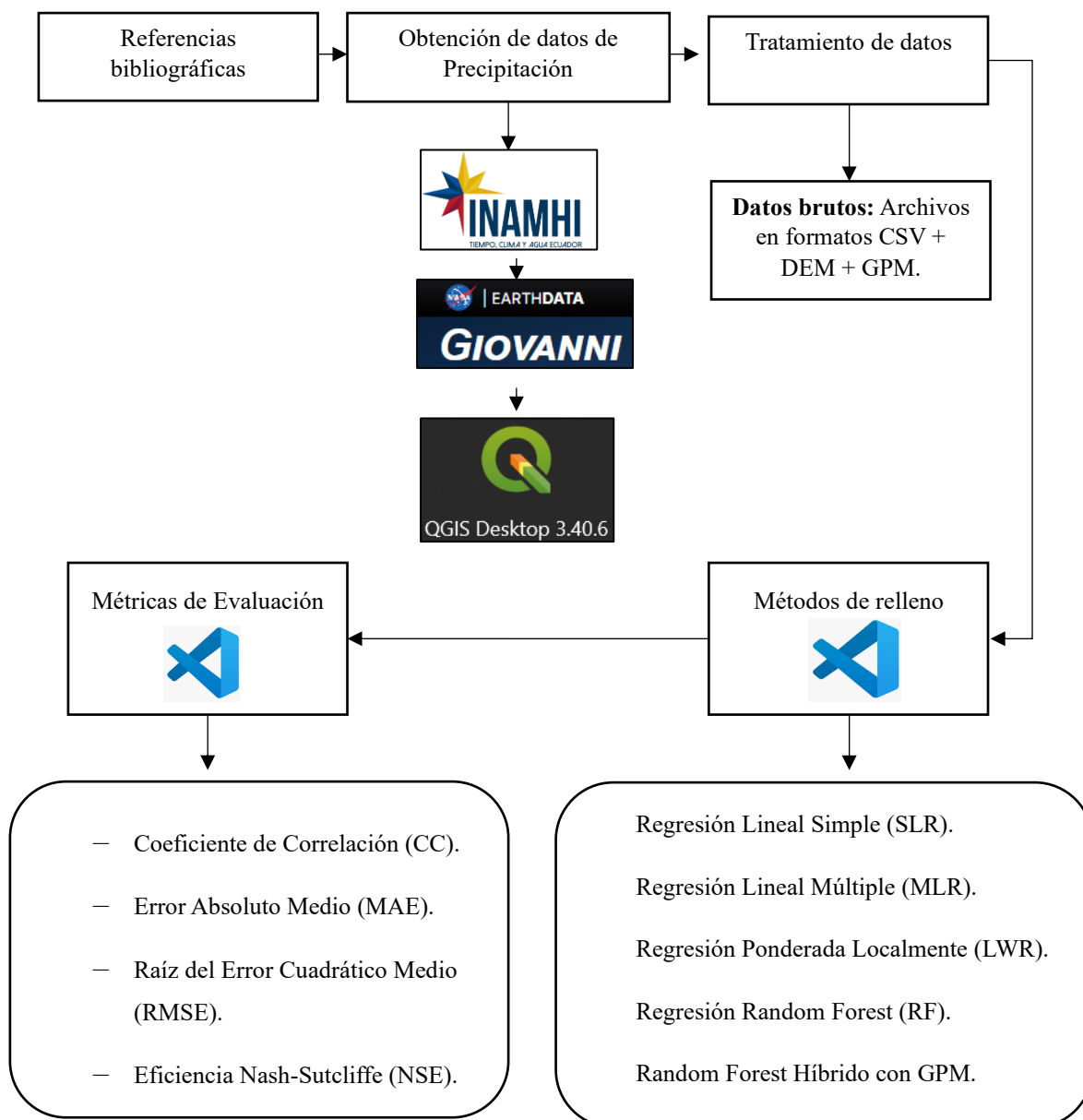


Figura 2. Esquema Metodológico

4.3 FASE 1 OBTENCIÓN DE INFORMACIÓN BIBLIOGRÁFICA

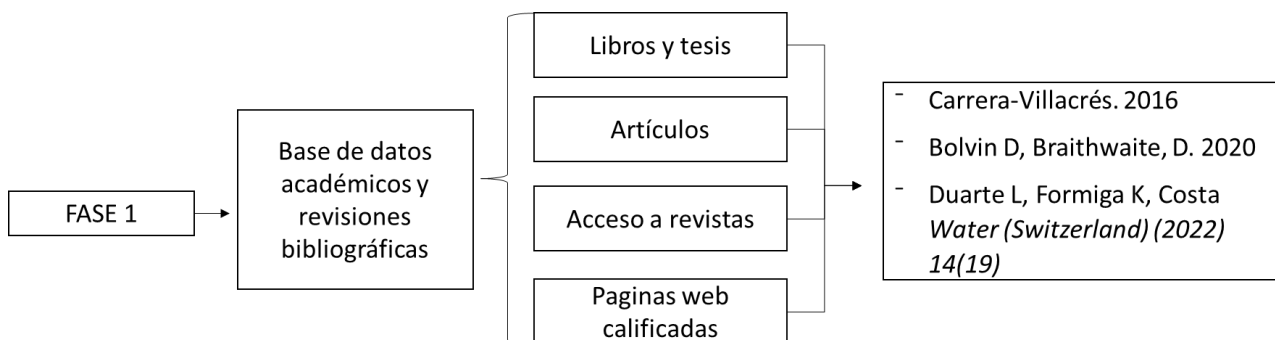


Figura 3. Esquema de la obtención de la información.

En el Ecuador, la institución responsable del monitoreo, recopilación y análisis de los datos hidrológicos y meteorológicos es el Instituto Nacional de Meteorología e Hidrología (INAMHI), durante la década de los ochenta, esta entidad publicaba su información proveniente de aproximadamente mil estaciones pertenecientes tanto al INAMHI como otras instituciones ya desaparecidas, por eso se buscó información relevante sobre el tema como se muestra en la Figura 3. Se analizaron distintos métodos de estimación que se muestra en la Tabla 7, evaluando su capacidad para manejar la intermitencia de la precipitación diaria. Para ello se analizaron que los métodos como lo son las regresiones son eficientes en predicciones a escala mensual y estacionaria, también que los métodos que se utilizaron como el en aprendizaje automático Random Forest tiene la ventaja de manejar una mayor robustez para el área de estudio, por lo que se procedió a su implementación técnica con el fin de determinar su comportamiento a escala diaria.

El organismo que se encarga de recomendar el manejo de la información es la Organización Mundial Meteorológica (OMM). Para la estimación de datos faltantes dentro de una serie temporal de valores existen, procedimientos estadísticos que permiten considerar propiedades físicas de la estación en estudio para permitir la obtención de valores realistas y coherentes con los fundamentos físicos[65].

Tabla 7. Metodos para la estimación de datos faltantes[65].

Método	Funcionamiento Técnico	Ventajas	Desventajas
Media Aritmética	Sustituye el vacío con el promedio histórico del periodo.	Simplicidad extrema y rapidez de cálculo.	Ignora la variabilidad diaria; reduce artificialmente la varianza de la serie.
Razón Normal	Pondera valores de estaciones vecinas basados en sus promedios anuales.	Útil cuando hay diferencias de altitud significativas entre estaciones.	Requiere al menos tres estaciones vecinas con datos completos.
Distancia Inversa ponderada (IDW)	Estima el valor según la cercanía física (a menor distancia, más peso).	Fácil de implementar en SIG (QGIS); lógico para zonas planas.	No considera las montañas que bloquean la lluvia.

MissForest (R)	Es iterativa basada en bosques aleatorios para múltiples variables.	Excelente para datos faltantes en variables continuas y categóricas.	Muy lento en series de tiempo largas; diseñado para datos tabulares, no espaciales.
Redes Neuronales (ANN)	Modelo que aprende patrones complejos.	Alta precisión en predicción no lineal.	Requieren una cantidad masiva de datos para entrenarse
Regresión Lineal Simple (SLR)	Relación lineal entre estación control y dependiente.	Ideal cuando existe una estación vecina con muy alta correlación.	Falla si la relación entre estaciones cambia estacionalmente.
Regresión Lineal Múltiple (MLR)	Usa varios predictores (vecinos, altitud) simultáneamente.	Mayor robustez al combinar información de varios puntos espaciales.	Sensible a la colinealidad (datos repetidos entre vecinos)
Regresión ponderada localmente (LWR)	Ajusta regresiones en entornos locales específicos.	Captura variaciones micro climáticas que los modelos globales ignoran.	Requiere definir un óptimo para no perder precisión.
Random Forest (RF)	Algoritmo de ensamble que promedia múltiples árboles.	Maneja valores atípicos y relaciones no lineales sin asumir normalidad.	Puede presentar sesgo en los valores extremos.
Random Forest (RF Híbrido -GPM)	Combina aprendizaje automático con datos satelitales.	Corrige el sesgo del satélite con datos de tierra.	Complejidad técnica alta en el procesamiento de los scripts.

4.4 Fase 2 Obtención de datos de precipitación



Figura 4. Obtención de datos de precipitación diarias

— Recolección de datos de precipitación diaria del INAMHI

La Figura 4 muestra que en la fase 2, que se llevara a cabo el procedimiento de la recolección de datos de precipitación diaria, constituyendo la parte principal para la elaboración del estudio de investigación. En la tabla 8 se muestra las estaciones de estudio con sus principales características entregados por la institución, INAMHI. Para la obtención de estos datos se pidió el registro de lectura de un periodo de tiempo específico de las estaciones meteorológicas. Además, se obtuvieron datos satelitales (GPM) es un producto de la NASA que estima las tasas de precipitación superficial global con una alta resolución espacial $0,1^\circ \times 0,1^\circ$ y con una resolución temporal de cada media hora, utilizando el satélite del Observatorio Central GPM, estándar para combinar las observaciones de precipitación de una constelación internacional de satélites mediante técnicas avanzadas[66]

Se muestra los periodos de precipitación que se obtuvieron del satélite, que guarda su información en la página web GIOVANNI (NASA) en formato HDF5 para transformarlo en formato TIFF se utiliza QGIS 3.40.6 para la visualización. Se creó un archivo NetCDF (.nc) que es el script de procesamiento que identifica y extrae la variable principal de precipitación, los datos de precipitación estimados en cada 30 minutos (mm/h), para cada píxel que encierra la zona de medición de precipitaciones se convirtieron a formato TIFF mediante Visual Code de las estaciones objeto de estudio. Por último, los datos se agregaron a la escala temporal diaria, utilizando el intervalo de medición.

Tabla 8 Estaciones meteorológicas estudiadas

Código	NOMBRE DE LA ESTACIONES	Latitud (GG° MM' SS").	Longitud (GG° MM' SS").	Altitud	Periodo	Datos llenos	Datos faltantes
M0160	EL CARMEN	0° 15' 35"	79° 25' 37"	260	1998-2018	58.31%	41,69%
M1208	LA TEODOMIRA	1° 9' 51"	80° 23' 24"	60		72.58%	27,42%
M0452	ZAPOTE	0° 56' 18"	80° 3' 20"	79		55.51%	44,49%
M1220	ATASCOSO (VIA PICHINCHA)	1° 5' 0"	79° 55' 0"	0		51.16%	48,84%
M0447	24 DE MAYO(JABONCILLO)	1° 16' 44"	80° 25' 6"	115		94.02%	5,98%
M0462	JUNIN	0° 55' 42" S	80° 12' 32" W	20		75.91%	24,09%

4.5 FASE 3 EL TRATAMIENTO DE DATOS BRUTOS DE PRECIPITACIÓN.

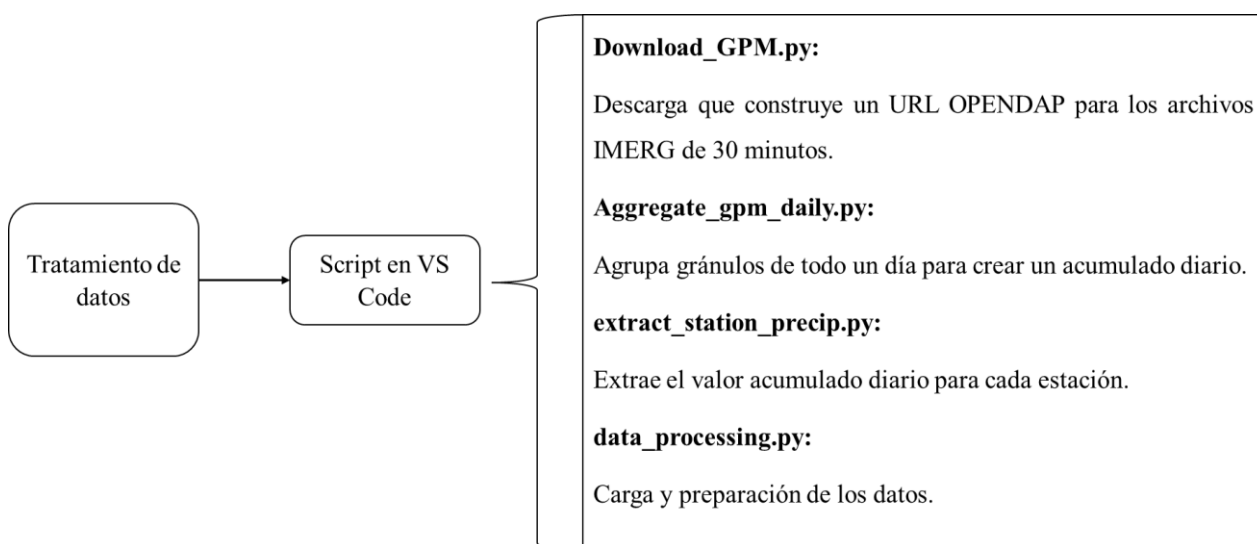


Figura 5. Esquema de tratamiento de datos crudos de precipitación

En la Figura 5, se realizó el procesamiento de datos que implica el tratamiento de los datos obtenidos, lo cual inicia con la descarga automatizada de datos de precipitación satelital GPM IMERG V07 mediante la librería `Download_GPM.py`, el cual gestiona solicitudes concurrentes a los servidores de la NASA Earthdata y garantiza la integridad de los archivos mediante un sistema de reintentos automáticos. Posteriormente, los datos sub horarios se procesan con `Aggregate_gpm_daily.py` para realizar la agregación temporal, transformando los registros de 30 minutos en acumulados diarios (mm/día) mediante sumas ponderadas y validación de cobertura espacial en formato NetCDF. Una vez consolidada la matriz diaria, se emplea el script `extract_station_precip.py` para realizar la extracción de series temporales puntuales; este proceso utiliza las coordenadas geográficas de las estaciones terrestres (cargadas desde un `GeoPackage`) para localizar los índices de píxeles más cercanos y generar una matriz estructurada en formato CSV. Finalmente, mediante `data_processing.py`, se ejecuta la limpieza

y alineación de datos, donde se aplican transformaciones temporales y el manejo inicial de vacíos, entregando Data Frames depurados y listos para la fase de imputación y evaluación estadística.

4.6 APLICACIÓN DE LOS 6 MÉTODOS PARA EL RELLENO DE DATOS FALTANTES DE PRECIPITACIÓN DIARIA.

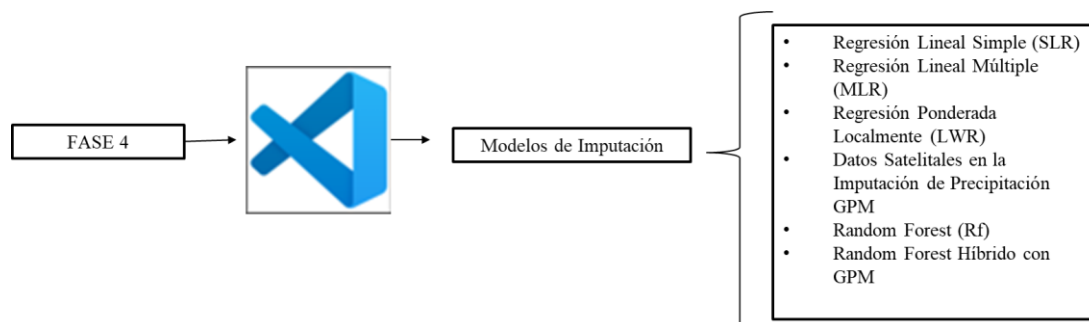


Figura 6. Métodos para el relleno de datos faltantes

4.6.1 Regresión Lineal Simple (SLR)

Este método es el más básico de los seis que se implementa para la imputación de datos faltantes que se utiliza, la relación lineal entre la estación objetivo y su vecino más cercano para estimar los valores faltantes utilizando la Ecuación (1), Para que el código se ejecute toma los datos históricos de ambas estaciones, donde se filtran solo los días donde ambas tienen datos válidos para que se cumpla, se requiere mínimo 10 observaciones comunes, se estiman los coeficientes β_0 (intercepto) y β_1 (pendiente) que minimizan el error, y luego se usa esta ecuación para predecir el valor faltante usando la precipitación actual del vecino, el mejor vecino se selecciona mediante pesos basados en la distancia especial que es la función cubica de manera que las estaciones más cercanas tienen mayor influencia.

$$Y = \beta_0 + \beta_1 \cdot X + \varepsilon \quad (1)$$

Ecuación 1 Regresión Lineal Simple

Donde:

- Y = Valor de precipitación a estimar en la estación objetivo
- X = Valor de precipitación observado en el mejor vecino

- β_0 =Intercepto
- β_1 =Pendiente
- ε = Término de error aleatorio

4.6.2 Regresión Lineal Múltiple (MLR)

La regresión hace que este método escoja a más de una variable que tiene influencia o está correlacionada con el valor de una tercera variable. Esta aproximación permite capturar patrones espaciales más complejos y reduce la dependencia de una única estación de referencia y utiliza la Ecuación (2).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (2)$$

Ecuación 2 Regresión Lineal Múltiple (MLR)

donde:

- Y = Vector de respuestas ($n \times 1$)
- X = Matriz de diseño ($n \times (k + 1)$)
- β = Vector de coeficientes ($(k + 1) \times 1$)
- ε = Vector de errores ($n \times 1$)

Los criterios se seleccionan los mejores vecinos ordenados por el peso que son inversamente proporcional a la distancia con la siguiente fórmula de la Ecuación (3):

$$k = \min(10, n_{vecinos_{disponibles}}) \quad (3)$$

Ecuación 3. Vecinos cercanos

4.6.3 Regresión Ponderada Localmente (LWR)

El método LWR (Locally Weighted Regression) este representa un método más significativo respecto a los métodos ya utilizados SLR y MLR ya que incorpora predictores topográficos derivados del DEM (Modelo Digital de Elevación) y también este aplica una ventana móvil temporal. Esta aproximación captura explícitamente los gradientes espaciales de precipitación con la siguiente fórmula de la Ecuación (4).

Se aplicó una ventana temporal con la Ecuación (5), que es una técnica de análisis de series temporales que selecciona un conjunto de datos cercanos en el tiempo (± 15 días) alrededor del día que se desea predecir es decir 15 días antes y 15 días después sin contar el día que se desea Rellenar ., esto proporciona hasta $31 \times k$ puntos de interacciones para el entrenamiento donde k es el número de vecinos.

$$Y = \beta_0 + \beta_1 \cdot E + \beta_2 \cdot S_N + \beta_3 \cdot S_E + \beta_4 \cdot \phi + \beta_5 \cdot \lambda + \varepsilon \quad (4)$$

Ecuación 4. Regresión Ponderada Localmente

Donde:

- Y = Precipitación estimada (mm)
- E = Elevación (metros)
- S_N = Pendiente norte-sur
- S_E = Pendiente este-oeste
- ϕ = Latitud
- λ = Longitud
- ε = Error

$$T_{ventana} = \{t - 15, t - 14, \dots, t - 1, t + 1, \dots, t + 14, t + 15\} \quad (5)$$

Ecuación 5. Ventana temporal

Donde:

$T_{ventana}$ = Conjunto de días seleccionados alrededor del día t .

t == Día objetivo.

$t - 15_a$ y $t + 15_a$ = es el rango de 15 días (en total 31 días).

Para el modelo digital de Elevación mediante Derivadas Parciales, esta ecuación se utilizó para sacar los valores SN y SE para la ventana temporal, las pendientes representan la tasa de cambio

espacial de la topografía, fundamentales para capturar gradientes de precipitación mediante la Ecuación (6) y para la elección de las estaciones meteorológica que se utilizan como donantes se utiliza la Ecuación (7) de pesos por distancia, los valores que se agregan para las distancias máximas y mínimas están reguladas por la Organización Mundial de la Metodología (OMM)

$$\begin{aligned} S_N &= -\frac{\partial z}{\partial y} \\ S_E &= -\frac{\partial z}{\partial x} \end{aligned} \quad (6)$$

Ecuación 6 Elevación mediante Derivadas Parciales

- S_N = Pendiente norte-sur
- S_E = Pendiente este-oeste
- z = Elevación en cada celda del DEM (metros)
- $\partial z / \partial y$ = Tasa de cambio de elevación en dirección norte-sur
- $\partial z / \partial x$ = Tasa de cambio de elevación en dirección este-oeste

$$\omega = \left(1 - \left(\frac{d_i}{d_{max}}\right)^3\right)^3 \quad (7)$$

Ecuación 7. Ecuación de pesos por distancia

Donde

- d_i = distancia al vecino i
- d_{max} = distancia máxima dinámica calculada por estación.

4.6.4 Datos Satelitales en la Imputación de Precipitación (GPM)

Para este método utilizamos los datos que se obtuvieron por el satélite, los pixeles con la resolución espacial de $0.1^\circ \times 0.1^\circ$ y la resolución temporal de 30 minutos, que encierra las zonas de los sitios objetivos que cuentan con información de precipitación diaria.

Se utilizó los datos satelitales en lugar de vecinos, útil para validar resultados y llenar huecos donde los métodos locales fallan.

4.6.5 Random Forest (Rf)

Es un método más sofisticado que los anteriores propuestos ya, este modelo combina varias técnicas de ensamblajes y de aprendizaje automáticos con predictores tanto temporales que serían las estaciones donantes y como espaciales el DEM, este es capaz de capturar relaciones

no lineales y esto hace que se supere a los demás métodos para la cual se utiliza la Ecuación (8), para la reducción de variancia se aplica la Ecuación 9 que se utiliza para agregar múltiples árboles que reduce la variancia sin aumentar el sesgo ,se agrega Bootstrap Aggregating que se basa en el método estadístico de muestreo con reemplazo conocido como Bootstrap, introducido por Bradley Efron en 1979 [67].Las variantes y extracciones que se utilizó para abordar las limitaciones que tiene y con esto aumentar el rendimiento de los algoritmos de aprendizaje automáticos.

- **Muestra Bootstrap:** Subconjunto aleatorio con reemplazo de los datos
- **Subconjunto de características:** En cada Split, solo se consideran en la ecuación (10)

En la tabla 9 se muestra la configuración de los valores que se agregan al código para esto se tomó en cuenta la latitud la longitud y la orografía, para que el método se pueda ejecutar de forma correcta.

$$y = \frac{1}{B} \sum_{b=1}^B T_b(\mathbf{x}) \quad (8)$$

Ecuación 8 Regresión Random Forest

Donde:

- B = número de árboles
- $T_b(\mathbf{x})$ =predicción del árbol b
- \mathbf{x} = vector de características

$$Var(y) = \rho\sigma^2 + \frac{1-\rho}{B}\sigma^2 \quad (9)$$

Ecuación 9 Reducción de variancia

Donde:

- $Var(\hat{y})$ = Varianza total del modelo RF
- ρ = correlación entre predicciones de diferentes árboles
- σ^2 =Varianza de un árbol individual

– B = Numero de árboles en el ensamble

$$m = \sqrt{p} \text{ predictores} \quad (10)$$

Ecuación 10 Número de predictores

Tabla 9 Valores que se agregan para el entrenamiento de (RF).

Parámetro	Valor	Detalle
n_estimators	100	Se definió la creación de 100 árboles de decisión independientes para generar el ensamble. Esto permite que la estimación final sea el promedio de múltiples "opiniones" del modelo.
max_depth	20	Se limitó la profundidad máxima de cada árbol a 20 niveles. Esto evita que los árboles se vuelvan demasiado complejos y memoricen el ruido de los datos meteorológicos.
min_samples_leaf	3	Se estableció que cada hoja final del árbol debe tener al menos 3 muestras. Esto garantiza que las estimaciones se basen en grupos de datos representativos.
min_samples_split	5	Un nodo solo se dividirá si tiene al menos 5 muestras. Esto ayuda a controlar la estructura del árbol durante el proceso de aprendizaje.
random_state	42	Valor utilizado para garantizar la reproducibilidad. Permite que, al ejecutar el entrenamiento nuevamente, los resultados de la estimación sean idénticos.
n_jobs	-1	Configuración técnica para utilizar todos los núcleos del procesador en paralelo, acelerando el tiempo de entrenamiento del modelo.
max_features	'sqrt'	En cada división del árbol, el modelo solo considera la raíz cuadrada del número total de variables. Esto aumenta la diversidad entre los árboles y mejora la precisión global.

4.6.6 Random Forest Híbrido con GPM

Para este método de RF_Híbrido es una de las extinciones del RF tradicional porque este incorpora los datos satelitales (GPM) como predictor principal a diferencia de RF que entrena.

En la tabla 10 se puede observar sus principales diferencias entre RF y RF hybrid demostrando cada uno sus puntos fuertes y otros débiles en la aplicación.

Tabla 10 Diferencia entre RF y RF_Hybrid

Aspecto	RF Estándar	RF_Hybrid
Entrenamiento	Por punto faltante	Por estación
Predictor principal	Vecinos	GPM + Vecinos
Evaluación	Sin hold-out explícito	Hold-out 10%
Velocidad	Lento (miles de modelos)	Rápido (1 modelo por estación)

El uso de GPM como predictor, proporciona una estimación inicial de los datos de precipitación que el RF corrige usando información local como se muestra en la ecuación (11)

$$\hat{y}_t = f_{RF}(GPM_t, X_{vecinos}, t, \Delta_{espacial}, DOY_t) \quad (11)$$

Ecuación 11 GPM como predictores

Donde:

- GPM_t =Valor GPM en el tiempo t
- $X_{vecinos,t}$ = Valores de estaciones vecinas
- $\Delta_{espacial}$ =Diferencias de elevación, latitud, longitud
- DOY_t = Día del año (estacionalidad)

En la tabla 11 se muestra la configuración de los parámetros, para el entrenamiento de los datos se escoge el 90% para entrenar el modelo mientras que los otros 10% verifica si funciona es decir predice el valor del dato que ya estaba originalmente sin antes haberlo visto.

Tabla 11 Valores que se agregan para el entrenamiento de (RF HYBRD).

Parámetro	Valor	Descripción del Proceso de Entrenamiento
train_fraction	0.9	Proporción de Entrenamiento: Se utiliza el 90% de los datos históricos disponibles para que el modelo aprenda la relación entre el satélite GPM y las estaciones terrestres, dejando el 10% para validación interna.
n_estimators	100	Ensamble Híbrido: El modelo construye 100 árboles de decisión. Cada árbol intenta corregir el sesgo del satélite basándose en la realidad de las estaciones en tierra.
max_depth	15	Control de Complejidad: Se limita la profundidad a 15 niveles para asegurar que el modelo aprenda patrones generales del relieve y el clima, evitando que se ajuste demasiado a ruidos específicos del satélite.
min_samples_leaf	3	Generalización: Cada nodo final debe representar al menos 3 registros climáticos, garantizando que el "relleno" de datos tenga sustento estadístico.
min_samples_split	5	Decisión de División: Un nodo solo se divide si hay suficiente información (5 muestras), lo que estabiliza el aprendizaje en zonas con pocas estaciones.
max_features	"sqrt"	Diversidad de Predictores: En cada paso, el modelo elige al azar entre la raíz cuadrada de las variables (GPM, Elevación, Pendiente, etc.), forzando al algoritmo a no depender únicamente de una sola variable.

4.7 DIAGRAMA DE EJECUCIÓN DEL CODIGO EN PYTHON MASTER EN EL ENTORNO DE VISUAL CODE.

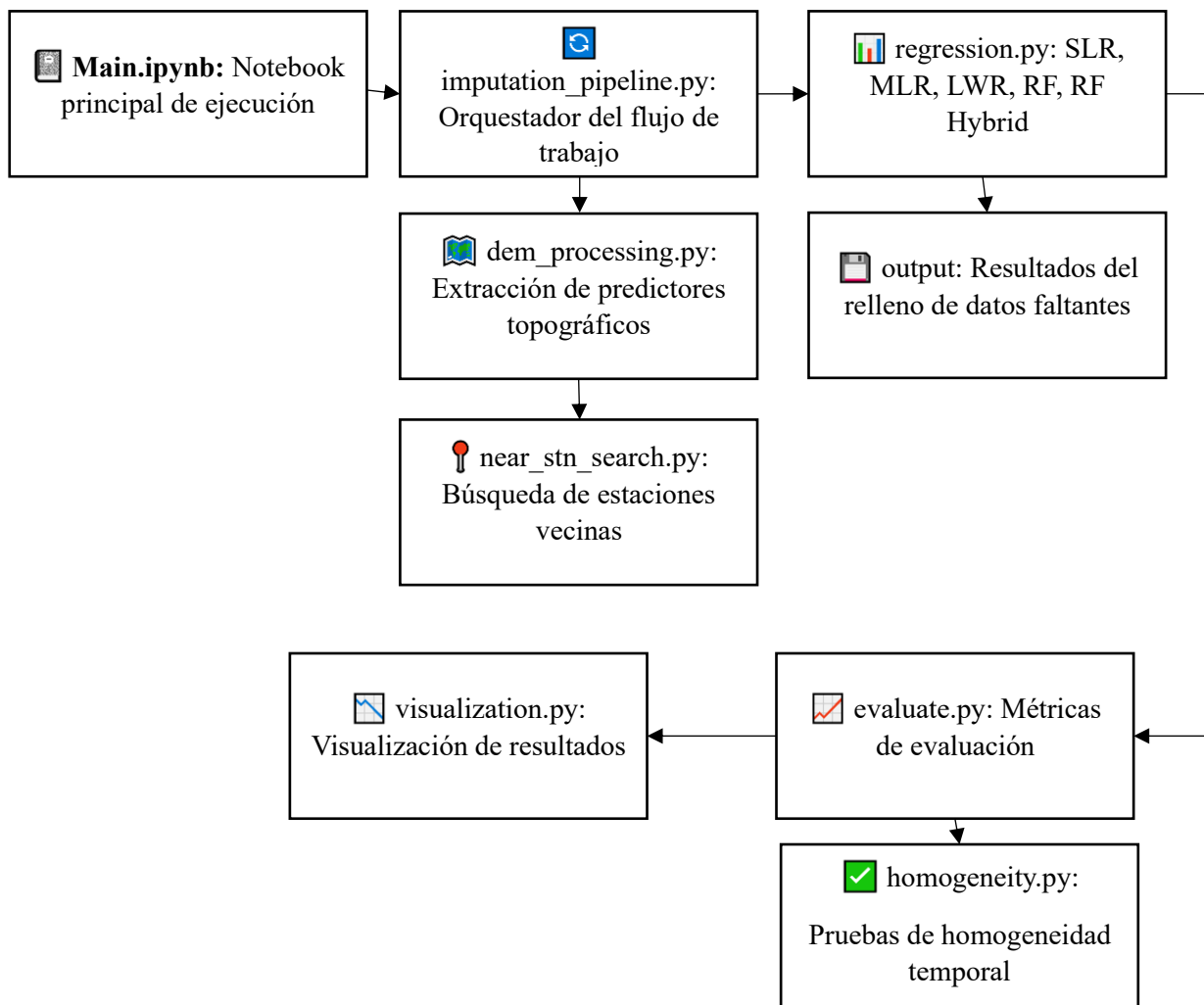


Figura 6. Ejecución del programa paso a paso

La arquitectura del sistema se fundamenta en un flujo de trabajo como se observa en la Figura 6 orquestado en Python, donde el archivo main.ipynb actúa como interfaz de control para el módulo imputation_pipeline.py, encargado de coordinar la transferencia de datos entre los distintos componentes del proceso. La selección de las estaciones donantes se gestiona a través de near_stn_search.py, aplicando un criterio de distancias técnica que prioriza un radio de 100 km y con un máximo de 150 km para preservar la similitud climática, bajo los lineamientos de la (OMM) No.168 para relieves moderados. Complementariamente, el módulo dem_processing.py integra el factor orográfico mediante la extracción de variables de elevación y pendientes (N-S, E-O), factores que condicionan la magnitud de la precipitación local. Finalmente, el script weight_calculation.py determina la influencia de las estaciones vecinas

mediante el cálculo de pesos inversamente proporcionales a la distancia a través de una fórmula cúbica, para finalizar se descargan todos los resultados con el código de `visualization.py`.

4.8 FASE 5 MÉTRICAS DE EVALUACIÓN.

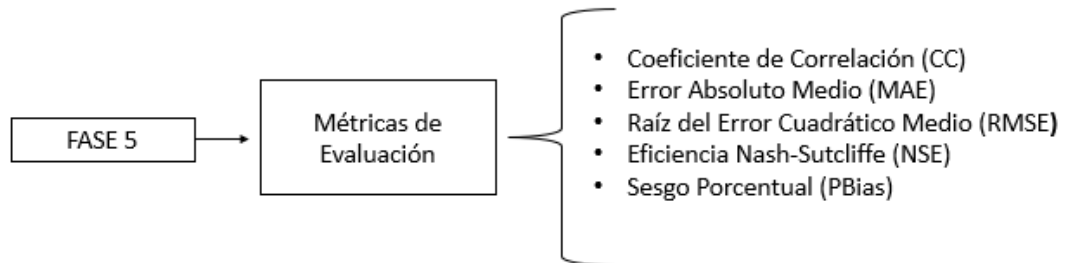


Figura 7. Métricas de evaluación estadísticas

La Figura 8 muestra el esquema que describe la metodología de las métricas que se utilizarán para evaluar el sistema que validará la calidad de los resultados obtenidos de los diferentes métodos aplicados, que se realizó para la imputación, este sistema se evaluó a escala diaria todo este proceso es clave para incrementar la comprensión del comportamiento de la precipitación y ofrecer aplicaciones valiosas en diferentes campos que la requiera.

4.8.1 Coeficiente de Correlación (CC)

El CC es una medida estadística que cuantifica la fuerza y dirección de la relación lineal entre dos variables. En la tabla 12 se muestra las métricas por las cuales deben ser evaluados este parámetro.

$$CC = \frac{Cov(obs, est)}{\sigma_{obs} \cdot \sigma_{est}} \quad (12)$$

Ecuación 12 Coeficiente de Correlación

Tabla 12 coeficiente de correlación[68]

Valores	Interpretación
0.00–0.10	Correlación insignificante
0.10–0.39	Correlación débil
0.40–0.69	Correlación moderada
0.70–0.89	Correlación fuerte
0.90–1.00	Correlación muy fuerte

4.8.2 Error Absoluto Medio (MAE)

Mide el promedio de las magnitudes de los errores en un conjunto de predicciones, sin considerar su dirección. En la tabla 13 se muestra las métricas por las cuáles deben ser evaluadas este parámetro.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (13)$$

Ecuación 13 Error Absoluto Medio

Tabla 13 Métricas de evaluación MAE [68]

Calificación	MAE	Interpretación
Muy Bueno	0.00-0.50	El error es menos de la mitad de la variabilidad natural de los datos.
Bueno	0.51-0.60	El modelo captura la tendencia con una desviación menor al 60%.
Satisfactorio	0.61-0.70	El error es aceptable, pero se acerca a la variabilidad de los datos.
No Satisfactorio	MAE>0.70	El modelo tiene tanto error que es casi igual a usar un promedio simple.

4.8.3 Raíz del Error Cuadrático Medio (RMSE)

Es una medida de la precisión del modelo que penaliza los errores grandes, en la tabla 14 se muestra las métricas por las cuáles debe de ser evaluada este parámetro siendo el valor óptimo 0.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (14)$$

Ecuación 14 Raíz del Error Cuadrático Medio

Tabla 14 Métricas de evaluación RMSE [68]

Calificación	RMSE	Interpretación
Muy Bueno	0.00-0.50	Error mínimo respecto a la variabilidad.
Bueno	0.50-0.60	Error bajo y aceptable.
Satisfactorio	0.60-0.70	Error moderado.
No Satisfactorio	RMSE > 0.70	El error es demasiado alto.

4.8.4 Eficiencia Nash-Sutcliffe (NSE)

Es un indicador normalizado que determina cuánto de la variabilidad de los datos observados es explicada por el modelo, en la tabla 15 se muestra las métricas por las cuáles deben ser evaluadas este parámetro lo cual mide habilidad predictiva vs la media siendo el rango óptimo 1.

$$NSE = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \quad (15)$$

Ecuación 15 Eficiencia Nash-Sutcliffe

Tabla 15 evaluación de métricas NSE [68]

Valor del NSE	Interpretación del Desempeño
0.75 – 1.00	Muy bueno
0.65 – 0.75	Bueno
0.50 – 0.65	Satisfactorio
< 0.50	No satisfactorio

4.8.5 Sesgo Porcentual (PBias)

Métrica estadística utilizada para medir la tendencia media de los datos simulados a ser mayores o menores que sus valores observados. En la tabla 16 se muestra las métricas por las cuáles deben ser evaluadas este parámetro. Mide subestimación/sobreestimación. Óptimo: 0%.

$$PBIAS = 100 \cdot \frac{\sum(\hat{y}_i - y_i)}{\sum y_i} \quad (16)$$

Ecuación 16 Sesgo Porcentual

Tabla 16 Métricas de evaluación PBIAS [68]

Calificación	PBIAS (Sesgo %)
Muy Bueno	PBIAS < 10 o <-10
Bueno	10 < PBIAS < 10
Satisfactorio	15 < PBIAS < 25
No Satisfactorio	PBIAS > 25 o <-25

5. ANÁLISIS DE RESULTADO

En la investigación, se trabajó para el periodo de tiempo de 1998-2018 de datos de precipitación diaria, para alizar y comparar diferentes métodos de relleno y evaluarlo mediante 5 métricas estadísticas que se utilizaron para 6 estaciones de prueba.

5.1 DISCUSIÓN DE LA INFORMACIÓN BIBLIOGRÁFICA

Lo que se obtuvo al analizar la información bibliográfica es la selección de los métodos de relleno de datos faltantes de precipitación diaria, se enfocó en criterio de progresión desde los enfoques más recomendados por organismos internacionales como la Organización Mundial de la Meteorología (OMM) hasta los métodos más actuales que se acoplen a los datos obtenidos y los parámetros, como la longitud, latitud y la orografía, descartando métodos como la Media Aritmética por su incapacidad de preservar la varianza diaria de la serie, el IDW por ignorar las barreras orográficas que condicionan la distribución espacial de la lluvia, y MissForest y las Redes Neuronales Artificiales que requieren registros continuos de la recolección de datos que difícilmente tiene el INAMHI. En cambio, se optó por iniciar con la Regresión Lineal Simple (SLR) como referencia metodológica y la más común al momento de rellenar los datos, avanzar hacia la Regresión Lineal Múltiple (MLR) al incorporar múltiples predictores es decir no solo una estación como lo hace la regresión simple si no que rellenarse con los datos de múltiples estaciones meteorológicas, e incluir la Regresión Ponderada Localmente (LWR) al incorporar múltiples predictores espaciales y orográficos. Estos métodos tienen la capacidad de estimar valores reales de las lluvias a de más que en Ecuador la falta de equipos meteorológicos convierte en un desafío mayor por lo que se integran datos satelitales (GPM) para los métodos más robustos como lo son (RF) Y (RF Hybrid). Estos métodos se han realizado a una escala temporal mensual donde sus resultados han sido buenos, mientras que en esta investigación se realizó a escala temporal diaria esto siendo uno de los principales retos, que es aplicar una sola variable climática que es la precipitación y la cantidad de ceros hace que las regresiones tiendan a subestimar o sobreestimar fenómenos climáticos leves o extremos.

5.2 GRÁFICAS DE DATOS FALTANTES VS RELLENANDO

Para poder desarrollar un análisis más detallado, también se ha representado gráficamente las series temporales observadas y estimadas, de cada uno de los métodos para llenar los valores que faltan para los sitios de prueba.

En las Tablas 17, 18, 19, 20, 21 y 22. Se muestra los resultados del comportamiento que tiene cada una de ellas es el manejo de eventos extremos de la estabilidad de los periodos que no se

encuentran con datos y por último el criterio y la evolución de cada uno por las diferentes métricas que se utilizaron por cada estación de estudio.

— Estacion M0160 EL CARMEN

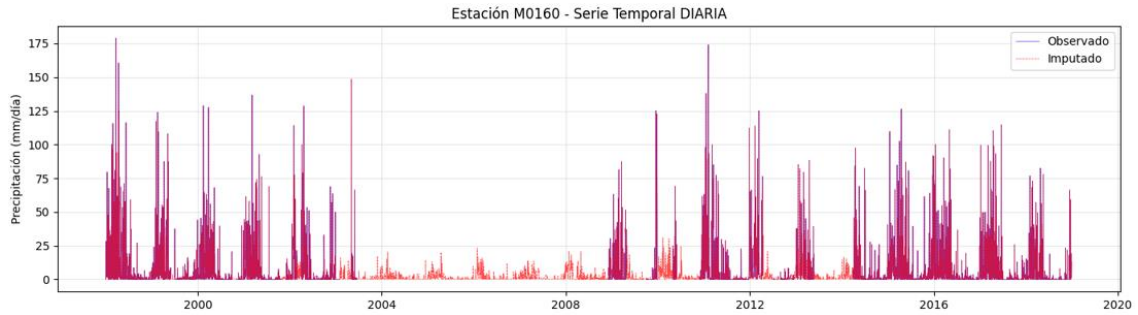


Figura 8.Regresión Lineal Simple (SLR) M0160

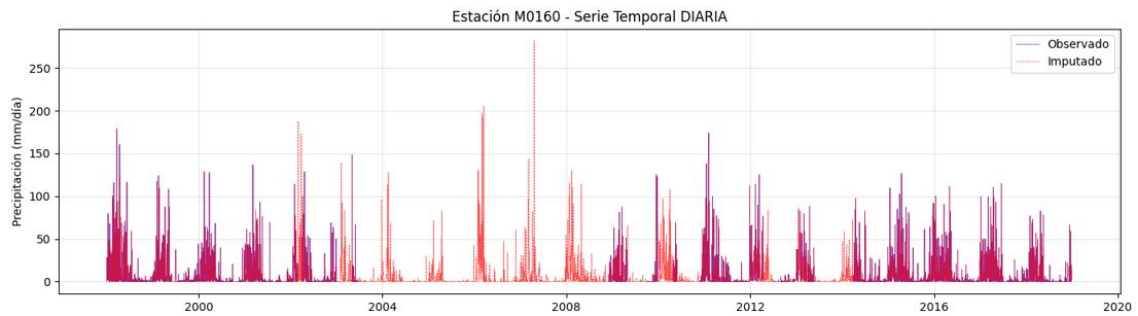


Figura 9.Regresión Lineal Múltiple (MLR) M0160

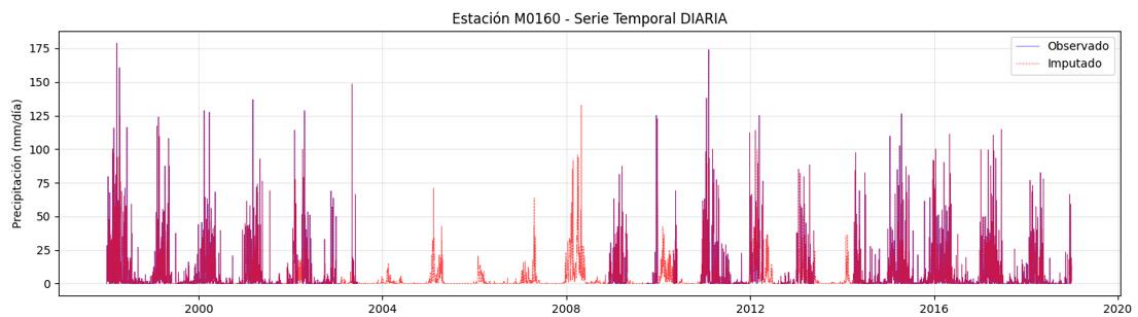


Figura 10.Regresión Ponderada Localmente (LWR) M0160

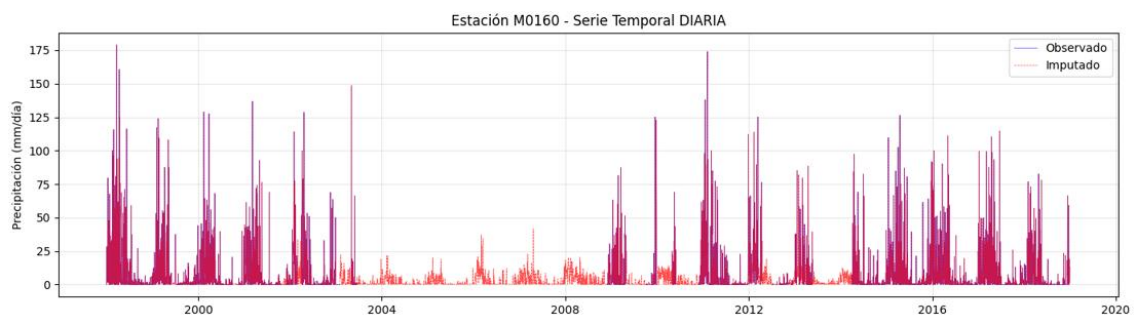


Figura 11.Regresión Random Forest (Rf) M0160

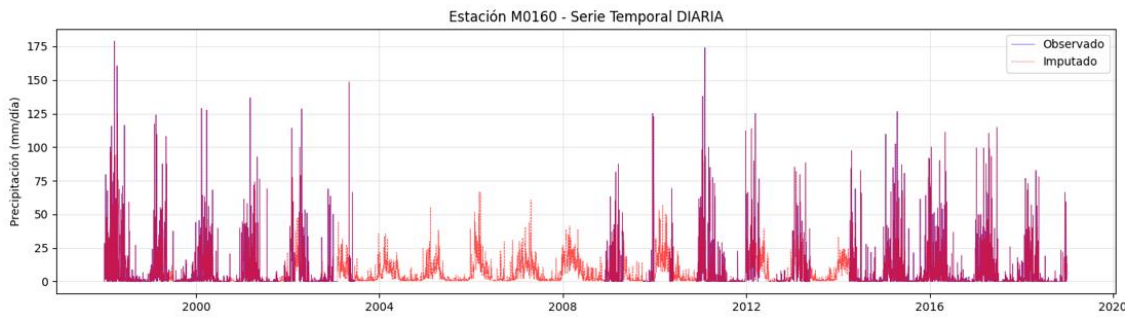


Figura 12. Random Forest Híbrido con GPM M0160

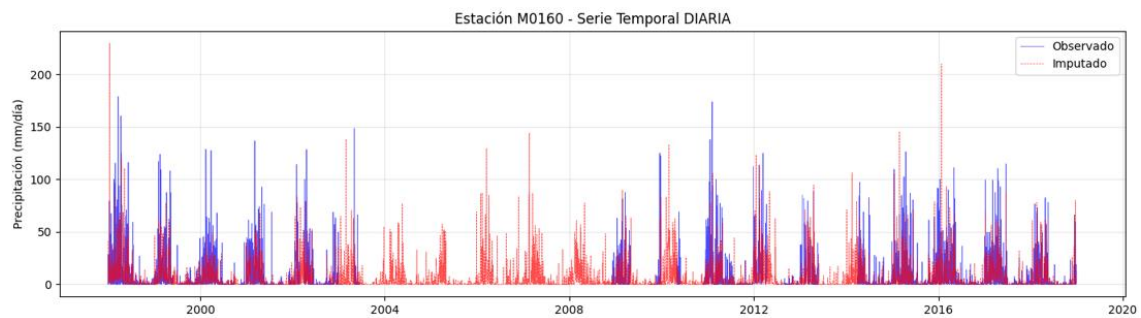


Figura 13. GPM M0160

Tabla 17 Resultados observados de la serie de datos rellenados por diferentes métodos M0160

Método	Figura	Comportamiento General	Manejo de Picos (Eventos Extremos)	Estabilidad en el Periodo de Vacío	Criterio y Evaluación
Regresión Lineal Simple (SLR)	8	Sigue la tendencia estacional básica, pero al parecer tiende a tener menos variabilidad que los datos reales.	Conserva los picos imputados rara vez superan los 50-75 mm, mientras que los reales superan los 125 mm.	Muestra precipitación, pero con amplitudes muy bajas y uniformes. Parece subestimar la lluvia real en este periodo.	Bajo. Método demasiado simplista para la complejidad de la lluvia diaria; tiende a "aplanar" los datos.
Regresión Lineal Múltiple (MLR)	9	Mejor ajuste visual que SLR en periodos con datos, pero muestra inestabilidad.	Genera un valor atípico masivo cerca de 2007 (aprox. 280 mm), lo cual es sospechoso y probablemente un error del modelo.	Rellena el vacío, pero la presencia de picos exagerados sugiere que el modelo es sensible a variables predictoras ruidosas.	Medio-Bajo. Aunque captura mejor la varianza, es inestable y propenso a errores graves de sobreestimación.
Regresión Ponderada	10	Ajuste suave. Intenta replicar la distribución	Moderado. Captura algunos picos altos (cerca de 125 mm),	El relleno en el vacío es consistente, aunque visualmente	Medio. Es un método robusto para tendencias, pero en

Localmente (LWR)		local de los datos vecinos.	mejor que SLR, pero no tan agresivo como MLR.	parece tener menos densidad que los periodos observados (2010-2016).	series de precipitación diaria puede perder precisión en eventos rápidos.
Random Forest (Rf)	11	Muy buen ajuste visual. La textura de la serie roja (imputada) se asemeja mucho a la azul (observada).	Logra Rellenar picos altos (sobre 100 mm) sin generar valores absurdos como el MLR.	El periodo de vacío muestra una variabilidad realista, con periodos secos y húmedos bien definidos, no solo "ruido" constante.	Alto. Al ser un método de aprendizaje automático no lineal, captura mejor la complejidad y estacionalidad de la lluvia.
Random Forest Híbrido con GPM	12	Excelente coherencia visual. Integra datos satelitales (GPM) para mejorar la precisión del Random Forest.	Equilibrado. Los picos en la zona imputada tienen una distribución muy similar a la histórica de la estación.	Rellena el vacío de 2004-2008 de manera muy convincente, probablemente gracias al apoyo de los datos satelitales que guían al modelo.	Muy Alto. Parece ser el método más robusto ("Estado del Arte"). Combina la inteligencia del algoritmo con datos físicos externos.
GPM (Satellite)	13	Muestra la serie completa del satélite. Es la fuente de datos externa, no un método de regresión.	Tiende a tener picos muy frecuentes y altos. A veces sobreestima la frecuencia de lluvia comparado con el pluviómetro en tierra.	Cubre todo el periodo sin vacíos, ya que es un sensor remoto.	Referencia. Sirve como variable predictora para los modelos híbridos, pero por sí solo suele tener sesgo respecto a la estación en tierra.

— Estacion M1208 LA TEODOMIRA

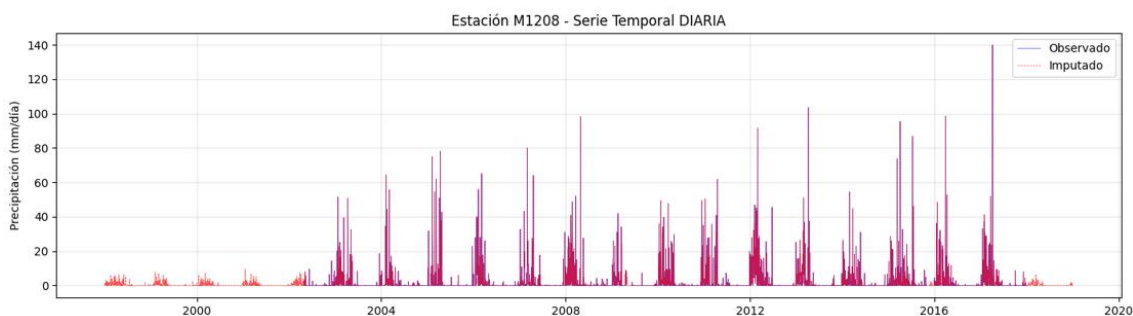


Figura 14 .Regresión Lineal Simple (SLR) M1208

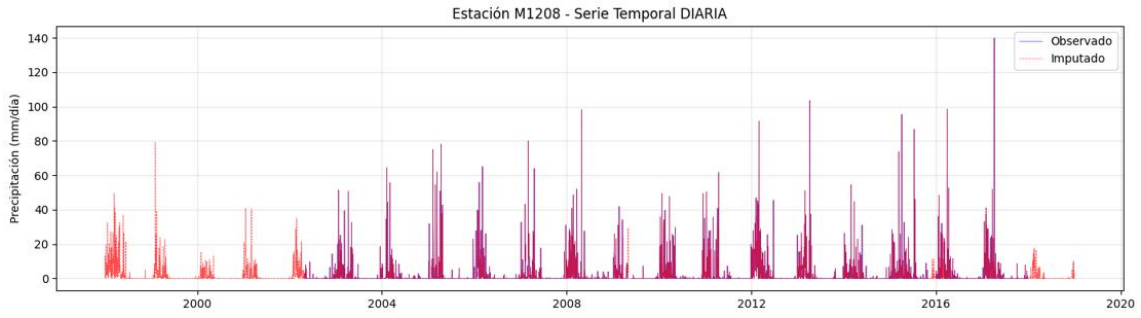


Figura 15 .Regresión Lineal Múltiple (MLR) M1208

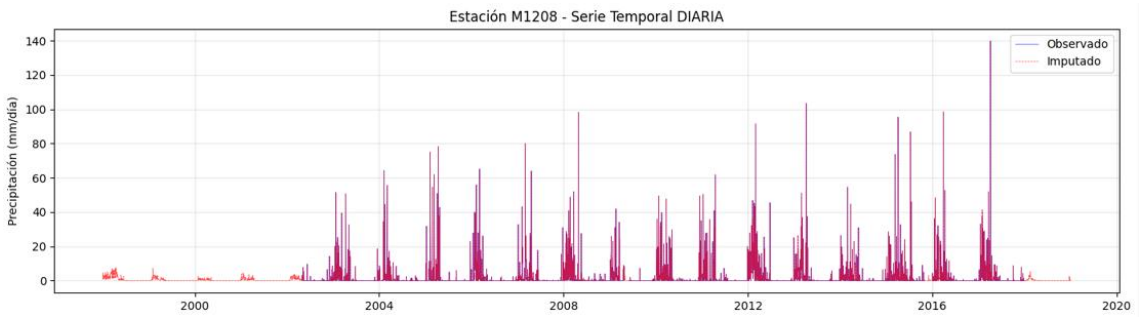


Figura 16 .Regresión Ponderada Localmente (LWR) M1208

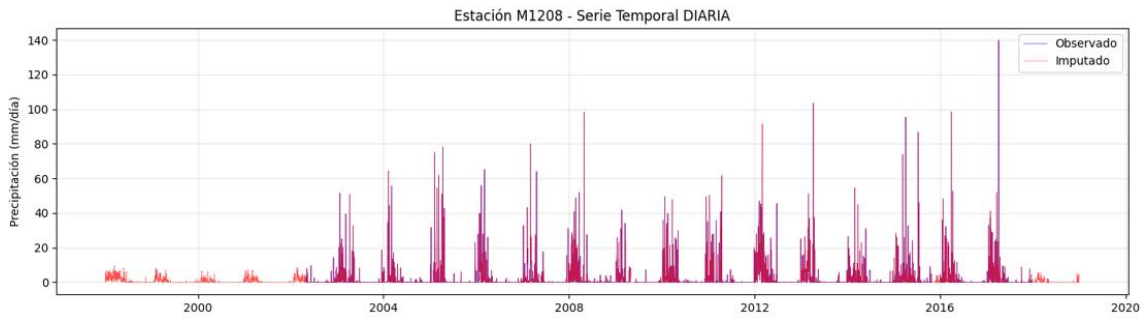


Figura 17.Regresión Random Forest (Rf) M1208

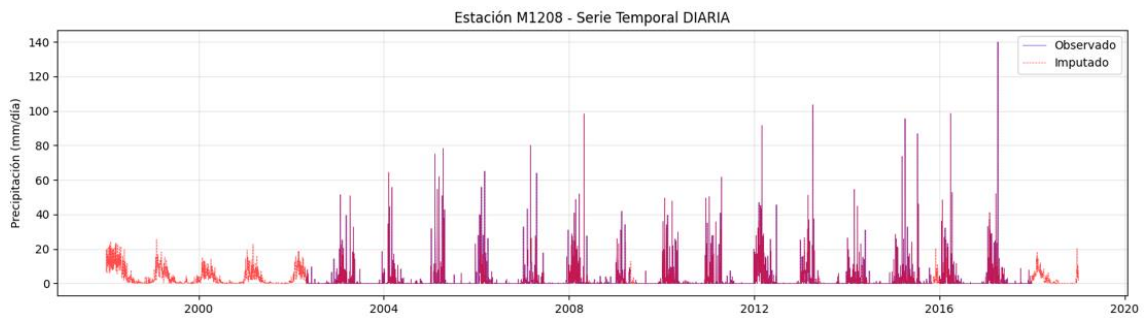


Figura 18 .Random Forest Híbrido con GPM M1208

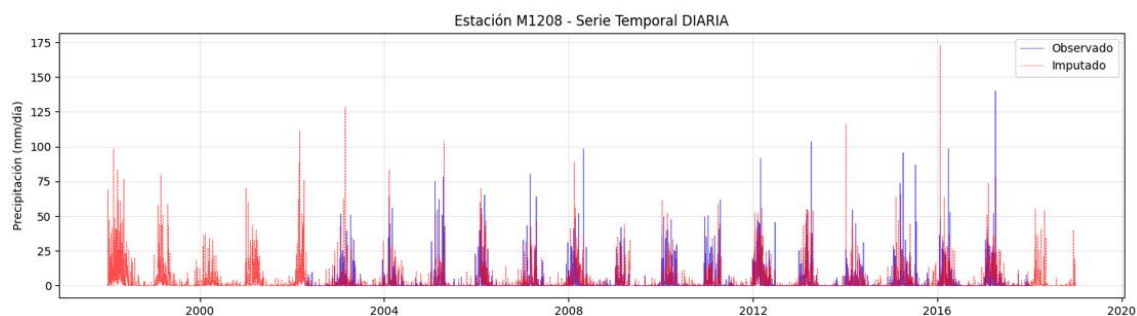


Figura 19 GPM M1208

Tabla 18 Resultados observados de la serie de datos rellenados por diferentes métodos M1208

Método	Figura	Comportamiento General	Manejo de Picos y Varianza	Estabilidad General	Criterio / Evaluación
Regresión Lineal Simple (SLR)	14	Pésimo por muestra una lluvia casi inexistente. Las líneas rojas son diminutas comparadas con la realidad histórica de la estación.	Subestimación severa. No logra capturar ni la magnitud ni la frecuencia de las lluvias fuertes.	Muy estable, pero inútil. Aplana la serie temporal hasta dejarla casi nula.	Muy Bajo. No sirve para completar el historial antiguo de la estación.
Regresión Lineal Múltiple (MLR)	15	Muestra algo más de actividad que la SLR, pero sigue siendo errática.	Genera "ruido". Hay picos, pero parecen disparos aleatorios más que temporadas de lluvia organizadas.	Inestable. Aunque intenta subir la magnitud, no logra formar una temporada invernal coherente en los años faltantes.	Bajo. Mejor que SLR por poco, pero insuficiente para análisis hidrológicos serios.
Regresión Ponderada Localmente (LWR)	16	Al inicio de la serie, al no tener vecinos cercanos temporales, el modelo colapsa y muestra valores cercanos a cero.	Prácticamente no hay picos en la zona imputada inicial. Solo funciona bien cuando hay datos observados cercanos.	Falla estructuralmente en vacíos largos al principio de la serie.	Bajo. Este método depende demasiado de la densidad local de datos, la cual no existe antes del 2003.
Random Forest (Rf)	17	Logra reconstruir la estacionalidad (invierno/verano) en el periodo 1998-2003.	Los picos son moderados. Recupera la forma de la lluvia, aunque visualmente parece un poco menos intensa que los años 2016-2017.	Cohrente. No presenta valores locos ni caídas a cero injustificadas.	Alto. Es una mejora sustancial frente a los métodos de regresión.

Random Forest Híbrido con GPM	18	Es el único que dota de cuerpo y volumen a las lluvias de 1998-2003.	Los picos rojos tienen una amplitud y densidad muy similar a los picos azules de años posteriores. Se ve natural.	Aprovecha los datos del satélite para saber exactamente cuándo llovió fuerte en esos años perdidos y ajusta la magnitud correctamente.	Muy Alto. Reconstruye el pasado con alta fidelidad visual.
GPM (Satelital)	19	Serie completa y continua, pero con mucho "ruido" visual.	Tiende a sobreestimar la frecuencia (parece que llueve todos los días). Algunos picos son muy altos (líneas azules/rojas muy largas).	Sirve como la "columna vertebral" de información para que el modelo híbrido funcione.	Referencia. Indispensable como insumo, pero no como producto final sin corregir.

— Estacion M0452 ZAPOTE

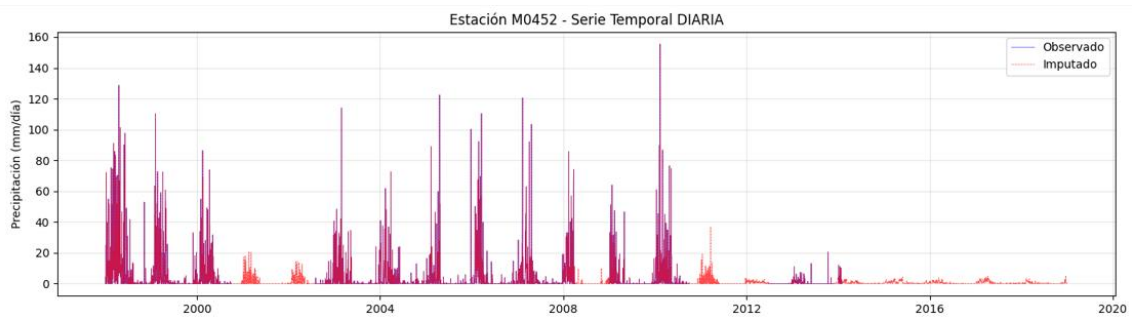


Figura 20. Regresión Lineal Simple (SLR) M0452

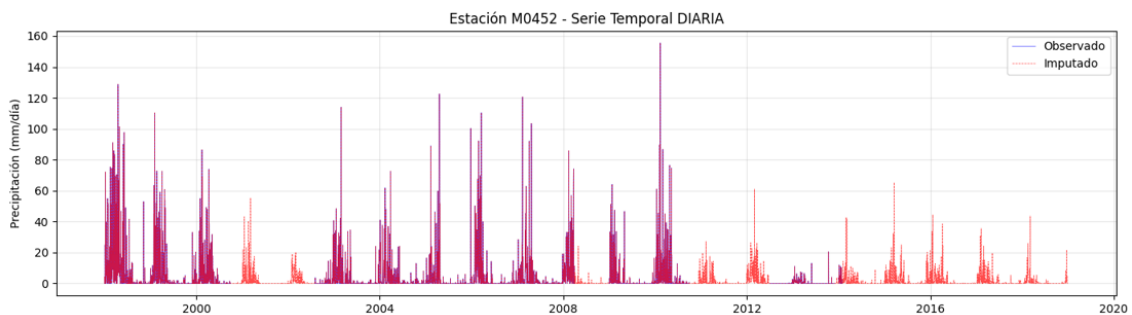


Figura 21. Regresión Lineal Múltiple (MLR) M0452

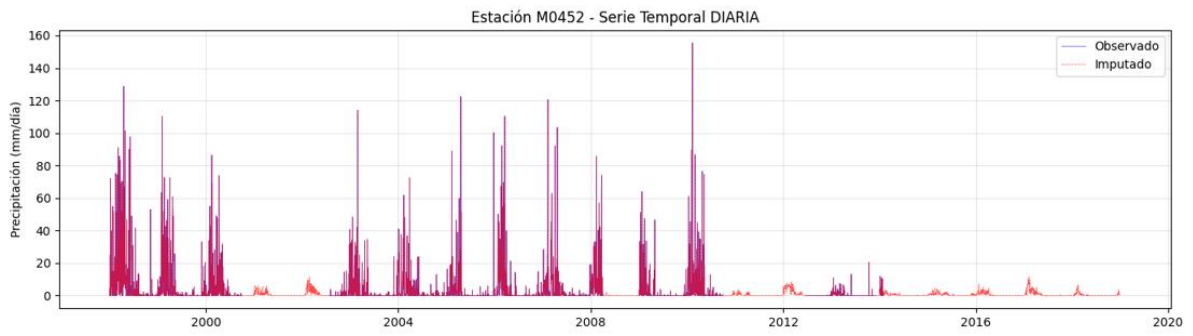


Figura 22. Regresión Ponderada Localmente (LWR M0452)

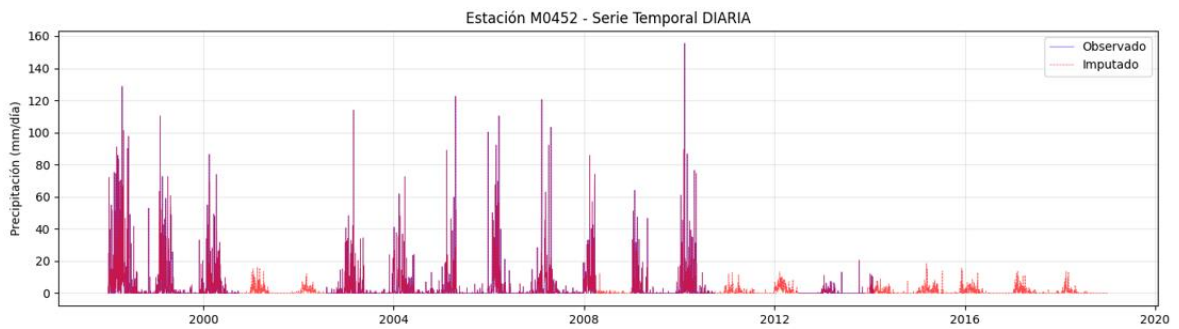


Figura 23 .Regresión Random Forest (Rf) M0452

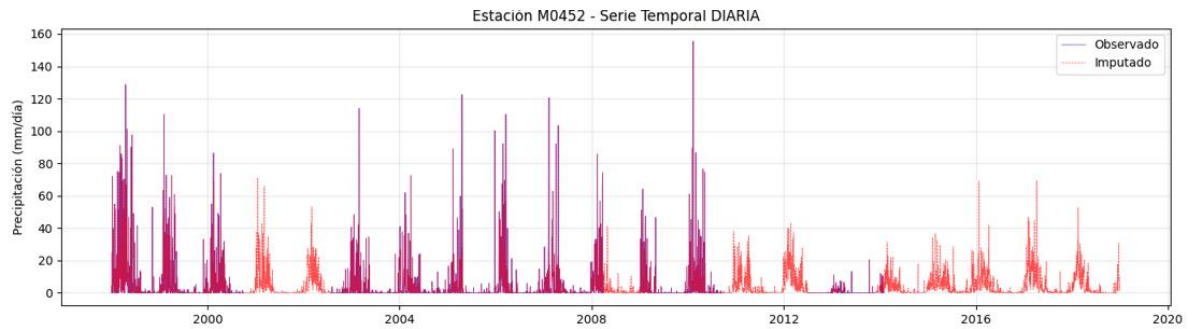


Figura 24. Random Forest Híbrido con GPM M0452

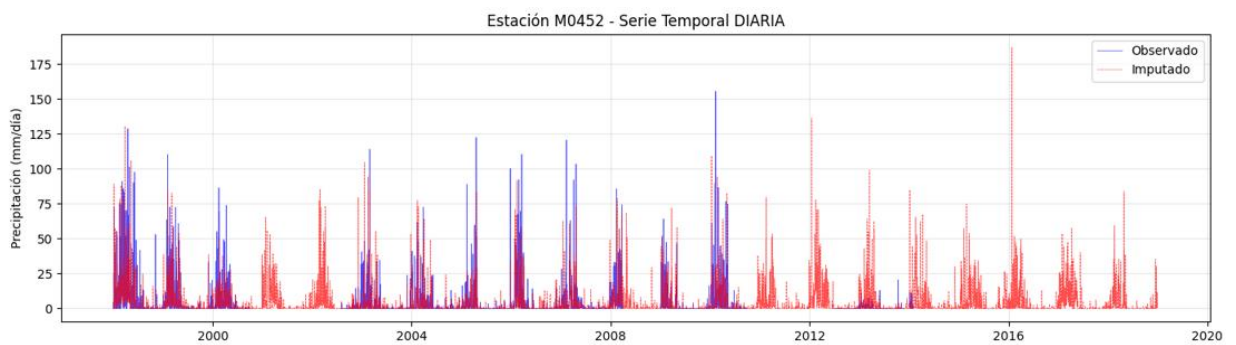


Figura 25. GPM M0452

Tabla 19 Resultados observados de la serie de datos rellenados por diferentes métodos M0452

Método	Figura	Comportamiento en el Periodo Crítico (2011-2017)	Manejo de Picos (Eventos Extremos)	Estabilidad General	Criterio / Evaluación
Regresión Lineal Simple (SLR)	20	Tiene un Fallo Crítico ya que rellena este gran vacío con una línea casi plana cercana a cero.	Muy pobre. Subestima severamente la lluvia, mostrando apenas en lugar de tormentas reales.	Estable pero inútil en periodos largos de ausencia de datos.	Muy Bajo. Sugeriría falsamente una sequía extrema de 6 años.
Regresión Lineal Múltiple (MLR)	21	Intenta generar picos, pero son erráticos y de baja magnitud comparados con el histórico anterior al 2010.	Inconsistente. A veces dispara valores altos aislados, pero no logra establecer un patrón estacional claro en el vacío.	Ruidosa. La serie rellenada se ve irreal y poco natural.	Bajo. No logra recuperar la estructura de la temporada de lluvias.
Regresión Ponderada Localmente (LWR)	22	Al no encontrar datos locales (vecinos) confiables en ese periodo largo, el modelo colapsa a cero.	Inexistente en los vacíos. Solo replica bien cuando ya existen datos cercanos.	Totalmente dependiente de la densidad de datos existentes.	Muy Bajo. No sirve para reconstruir vacíos extensos como el de 2011-2017.
Random Forest (Rf)	23	Aunque es un modelo inteligente, sin la ayuda del satélite, parece "aprender" que en esos años no llovió y predice valores muy bajos.	Muy conservador en el vacío. No se atreve a predecir lluvias fuertes si no tiene una variable fuerte que se lo indique.	Visualmente se ve como una caída drástica en la precipitación, lo cual es sospechoso.	Medio-Bajo. No logra corregir el probable error de "falsos ceros" de la estación.
Random Forest Híbrido con GPM	24	Es el único método que analiza que sí llovió entre 2011 y 2017 y reconstruye las temporadas de lluvia completas.	Recupera picos de precipitación (50-80 mm) que son coherentes con la historia de la estación (1998-2010).	Rescata la información perdida. Transforma un periodo de "datos muertos" en una serie climática útil y continua.	Muy Alto. Es el único método que salva la serie de datos en su tramo final.

<p align="center">GPM (Satelital)</p>	<p align="center">25</p>	<p align="center">Confirma que hubo actividad lluviosa normal durante 2011-2017.</p>	<p align="center">Muestra picos muy altos, a veces exagerados, pero confirma la ocurrencia de eventos.</p>	<p align="center">Provee la señal necesaria para que el modelo híbrido sepa cuándo Rellena lluvia.</p>	<p align="center">La prueba de que la falta de lluvia en los otros modelos es un error.</p>
--	--------------------------	--	--	--	---

— Estacion M0462 JUNIN

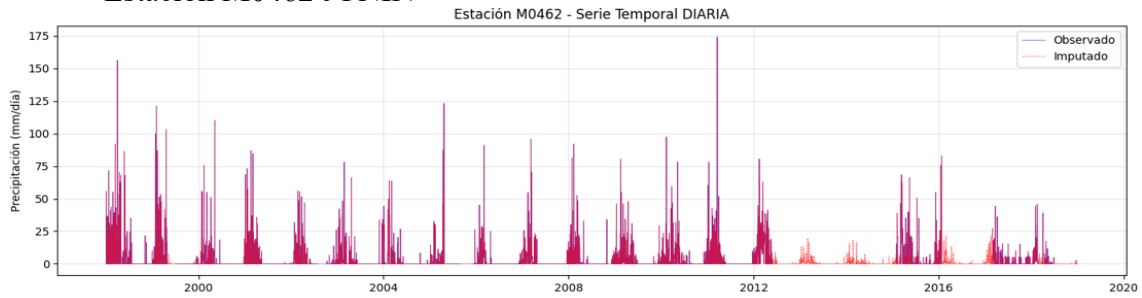


Figura 26. Regresión Lineal Simple (SLR) 4062

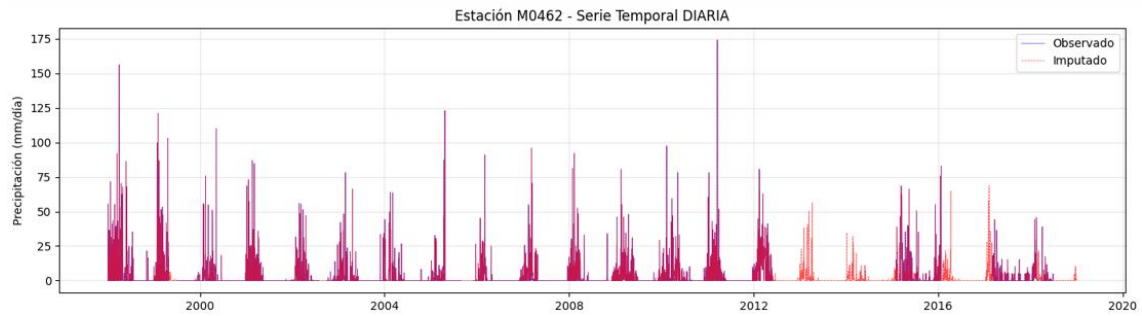


Figura 27. Regresión Lineal Múltiple (MLR) 4062

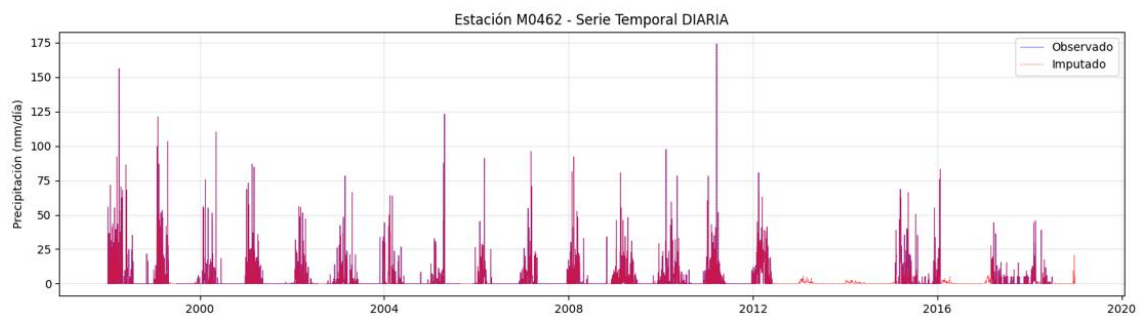


Figura 28 .Regresión Ponderada Localmente (LWR) 4062

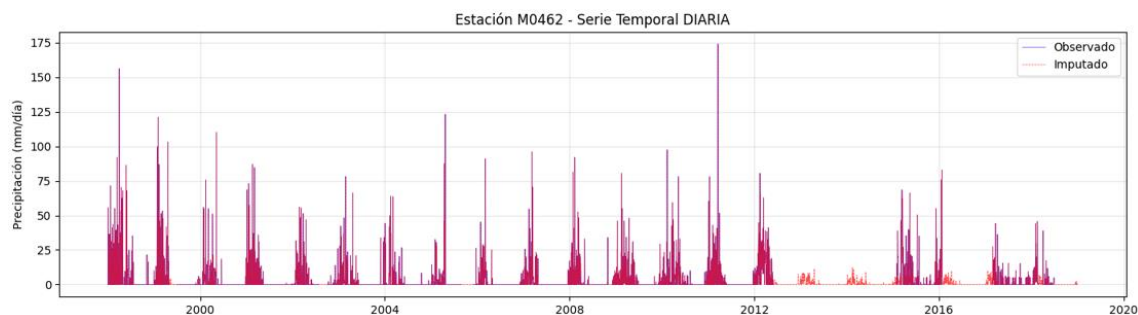


Figura 29. Regresión Random Forest (Rf) 4062

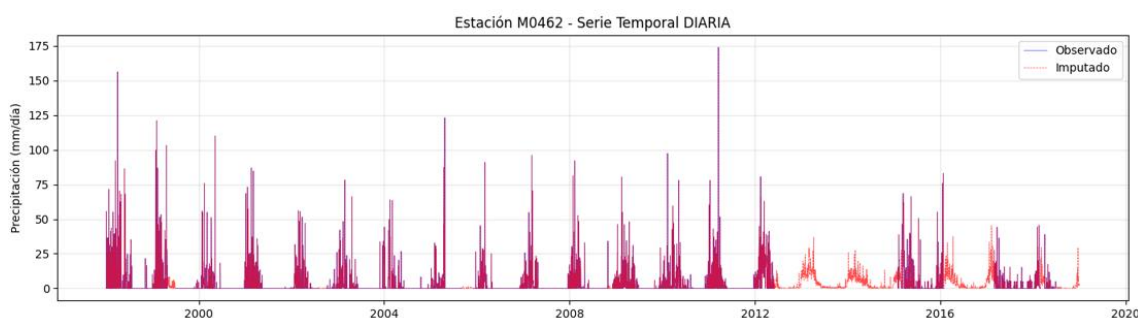


Figura 30. Random Forest Híbrido con GPM 4062

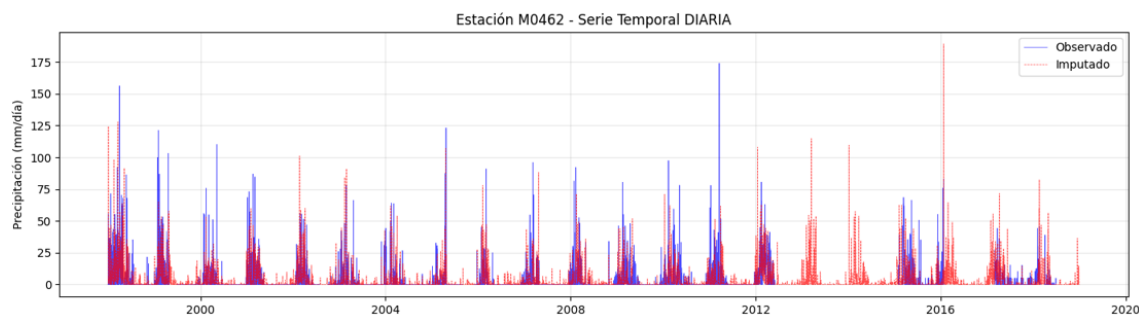


Figura 31. Datos Satelitales en la Imputación de Precipitación GPM 4062

Tabla 20 Resultados observados de la serie de datos rellenados por diferentes métodos 4062

Método	Figura	Comportamiento en Periodos de Ausencia (2012-2015)	Manejo de Picos (Eventos Extremos)	Estabilidad General	Evaluación / Criterio
Regresión Lineal Simple (SLR)	26	Rellena el vacío con valores extremadamente bajos y constantes que no reflejan la variabilidad estacional.	Subestima severamente la magnitud de las lluvias, fallando en capturar eventos por encima de los 25 mm/día en los años imputados.	Estable pero irreal, creando una falsa sensación de sequía prolongada.	Muy Bajo. No es apto para análisis hidrológico ya que elimina los extremos climáticos.

Regresión Lineal Múltiple (MLR)	27	Intenta recuperar la estacionalidad, pero los valores son erráticos.	Genera algunos picos, pero no logra alcanzar la intensidad de los eventos históricos observados previamente (como el de 2011).	Ruidosa; la serie imputada carece de la estructura de "racimos" de lluvia natural.	Bajo. Aunque es mejor que la SLR, sigue fallando en la representación de tormentas reales.
Regresión Ponderada Localmente (LWR)	28	Tiende a suavizar demasiado los datos, resultando en valores casi nulos durante gran parte del periodo de ausencia.	No detecta ni genera picos de precipitación significativos durante los años de falta de datos.	Demasiado plana; pierde toda la dinámica de la cuenca.	Bajo. Útil solo para tendencias de largo plazo, pero inútil para eventos diarios.
Random Forest (Rf)	29	Recupera una estructura de lluvia mucho más parecida a los registros históricos observados entre 1998 y 2012.	Logra generar eventos de hasta 50-60 mm/día, guardando mejor proporción con los máximos locales.	Mantiene la intermitencia natural de la lluvia (días secos vs. días húmedos).	Alto. Es el primer método que presenta una serie visualmente coherente en todo el registro.
Random Forest Híbrido con GPM	30	Al integrar datos satelitales, rellena los vacíos con una precisión superior, alineada con eventos regionales registrados por satélite.	Captura de forma más realista la intensidad y frecuencia de los picos de lluvia durante el periodo 2012-2016.	Alta fidelidad; la transición entre datos observados e imputados es casi imperceptible.	Muy Alto. Es el método más robusto para esta estación, minimizando el sesgo de subestimación.
Datos Satelitales (GPM)	31	Sirve como la base de verdad para identificar cuándo ocurrieron realmente las tormentas en ausencia de datos terrestres.	Tiende a detectar picos muy altos (superiores a 175 mm/día) que pueden superar la capacidad de medición local.	Proporciona continuidad total, aunque con un ruido base superior a la estación.	Alto. Indispensable para validar los modelos de Machine Learning (Rf).

— Estacion M1220 ATASCOSO (VIA PICHINCHA)

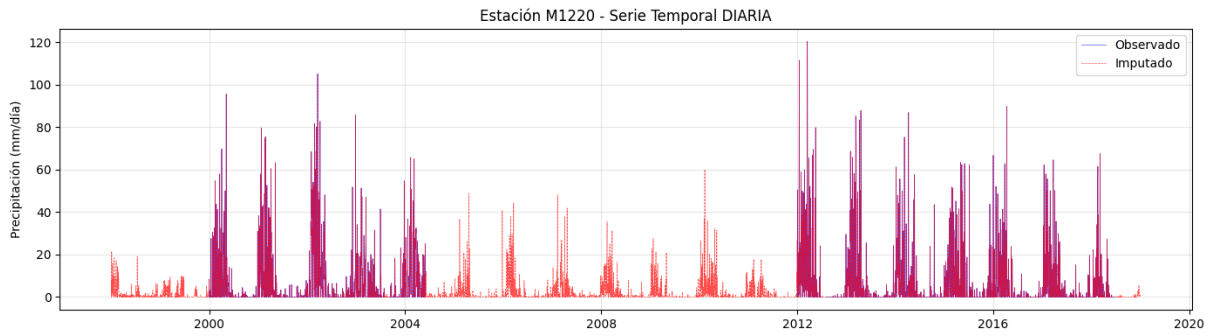


Figura 32. Regresión Lineal Simple (SLR) M1220

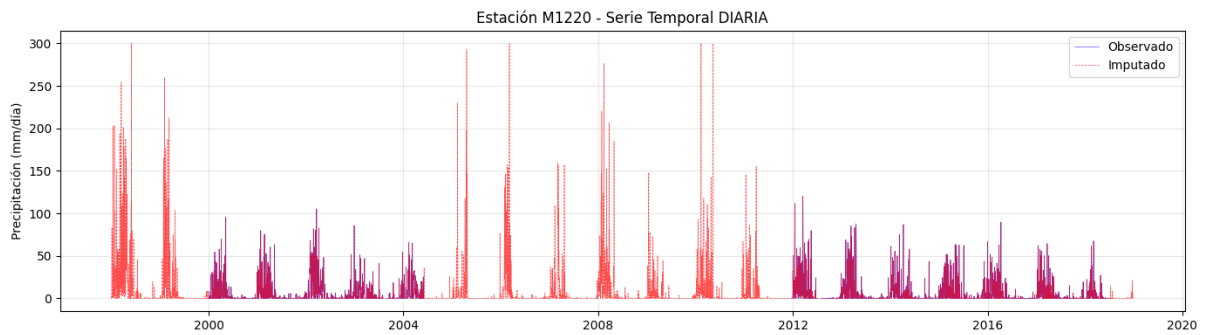


Figura 33 .Regresión Lineal Múltiple (MLR) M1220

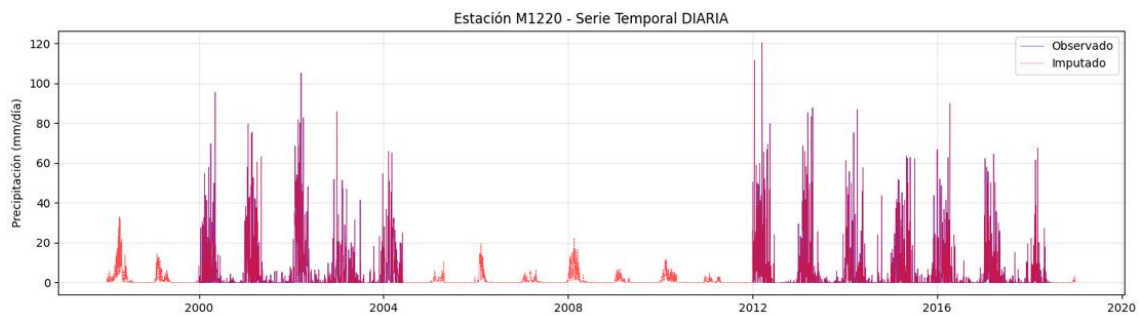


Figura 34 .Regresión Ponderada Localmente (LWR) M1220

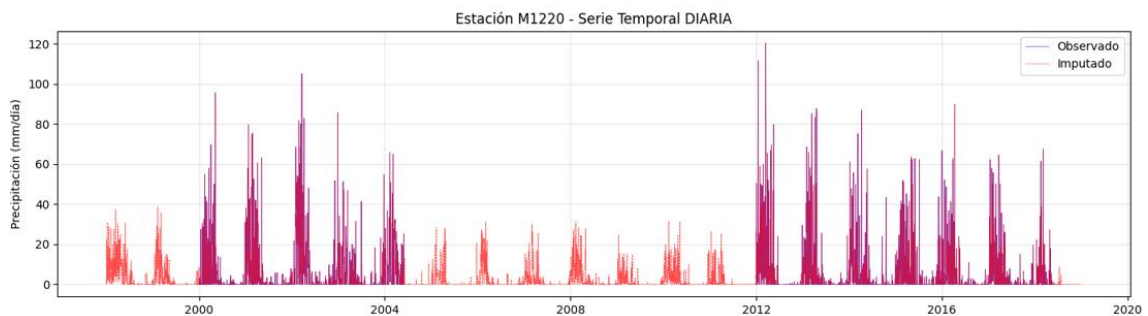


Figura 35. Regresión Random Forest (Rf) M1220

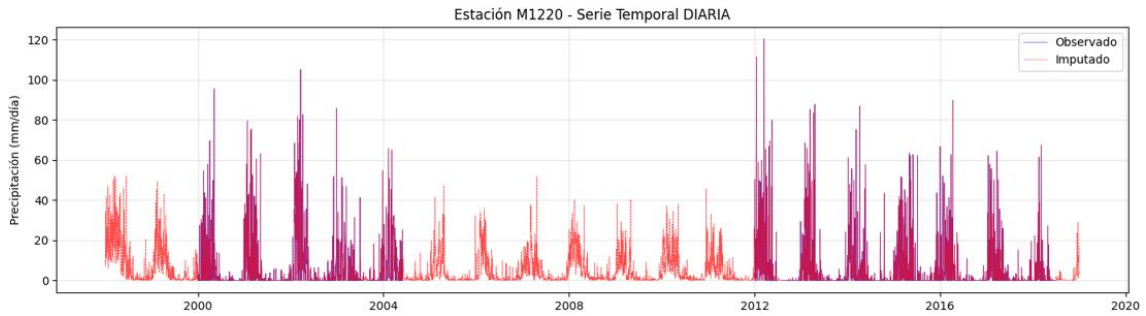


Figura 36. Random Forest Híbrido con GPM M1220

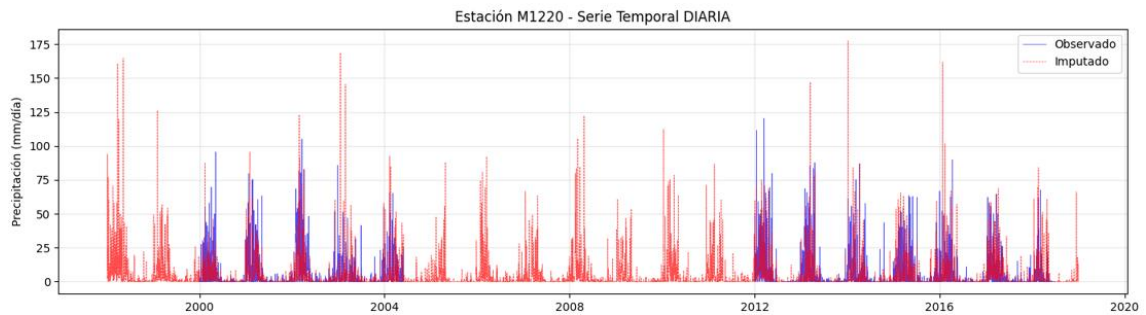


Figura 37. Datos Satelitales en la Imputación de Precipitación GPM M1220

Tabla 21 Resultados observados de la serie de datos rellenados por diferentes métodos 1220

Método	Figura	Comportamiento en Periodos de Ausencia	Manejo de Picos (Eventos Extremos)	Estabilidad General	Evaluación y Criterio
Regresión Lineal Simple (SLR)	32	Es muy deficiente ya que Rellena los vacíos con valores muy bajos y constantes, perdiendo la variabilidad natural.	No logra capturar ninguno de los picos observados en los años circundantes.	Estable, pero con un sesgo hacia la subestimación.	Bajo. Sugiere periodos de sequía inexistentes donde debería haber estacionalidad.
Regresión Lineal Múltiple (MLR)	33	Se muestra que es inconsistente por que genera picos extremadamente altos (hasta 300 mm/día) que parecen irreales comparados con el histórico.	Sobrestima severamente. Los valores imputados son el doble o triple de los máximos observados (120 mm/día).	Muy ruidosa y con alta variabilidad artificial.	Bajo/Medio. Aunque intenta recuperar la variabilidad, la magnitud de los errores es muy alta.

Regresión Ponderada Localmente (LWR)	34	Este método Intenta seguir la estacionalidad, pero tiende a aplanar los ciclos de lluvia.	Pobre. Subestima los eventos máximos, manteniendo los picos imputados por debajo de los 40 mm/día.	Suave, con menos ruido que MLR, pero poca precisión.	Medio. Es conservador, pero útil si solo se busca una tendencia general sin extremos.
Random Forest (Rf)	35	Logra una estructura de datos mucho más similar a la observada, respetando los ciclos anuales.	Aceptable. Capta picos de intensidad media, aunque aún le cuesta llegar a los extremos máximos reales.	Muy estable y visualmente coherente con la serie original.	Alto. Es uno de los métodos más robustos para mantener la distribución estadística.
Random Forest Híbrido con GPM	36	Es el que mejor rellena los vacíos manteniendo la coherencia con los datos satelitales y terrestres.	Muy Bueno. Los picos imputados guardan una proporción lógica con los datos observados (~60-80 mm/día).	Alta estabilidad y realismo en la transición observado/imputado.	Muy Alto. Es el método más equilibrado para esta estación específica.
Datos Satelitales (GPM)	37	Proporciona una base continua, pero con un sesgo de escala diferente a la estación.	Variable. Tiende a detectar eventos de lluvia que la estación no registró o viceversa.	Constante, pero requiere calibración (sesgo sistemático).	Medio/Alto. Vital como soporte, pero menos preciso que el modelo híbrido.

— Estacion M0447 24 DE MAYO(JABONCILLO)

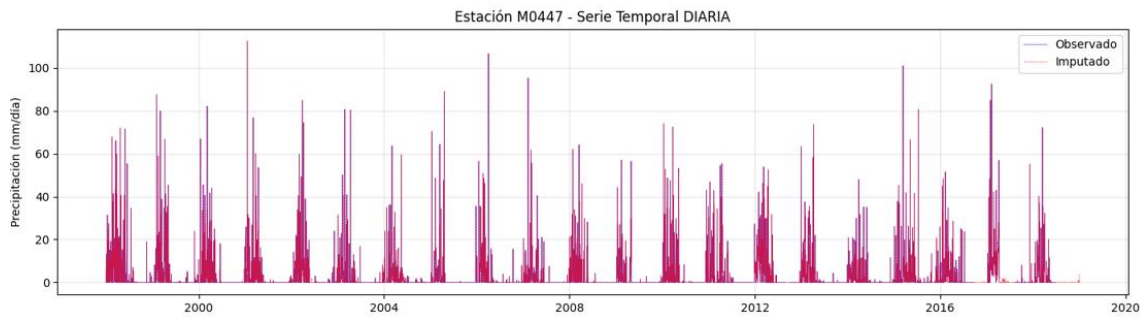


Figura 38. Regresión Lineal Simple (SLR)

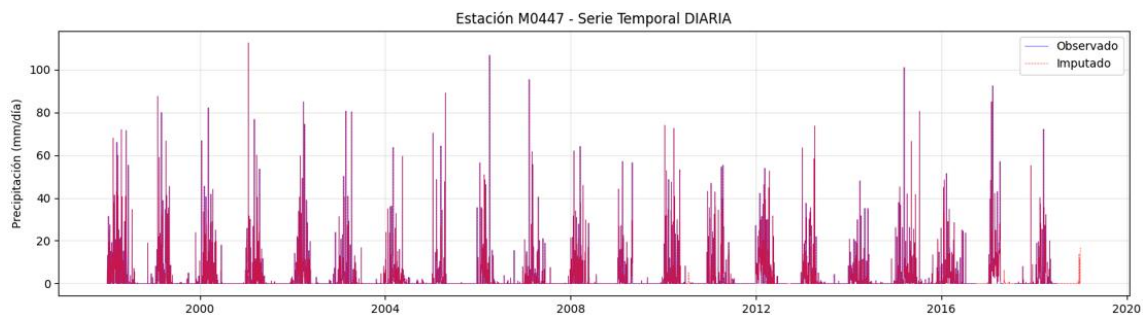


Figura 39. Regresión Lineal Múltiple (MLR)

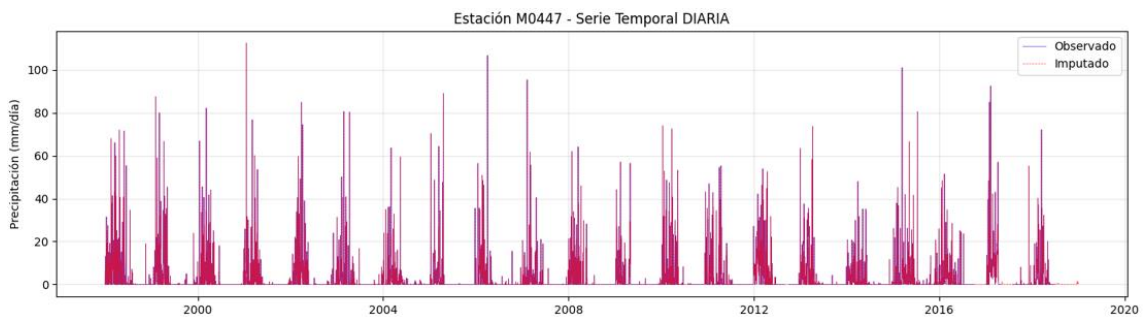


Figura 40. Regresión Ponderada Localmente (LWR)

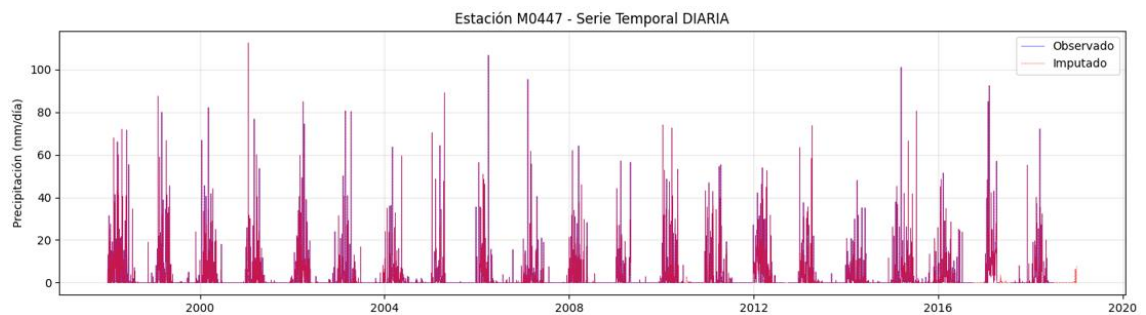


Figura 41. Regresión Random Forest (Rf)

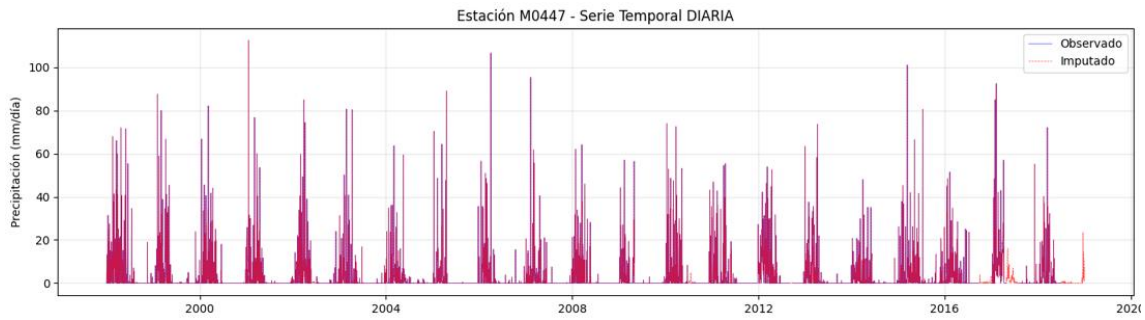


Figura 42. Random Forest Híbrido con GPM

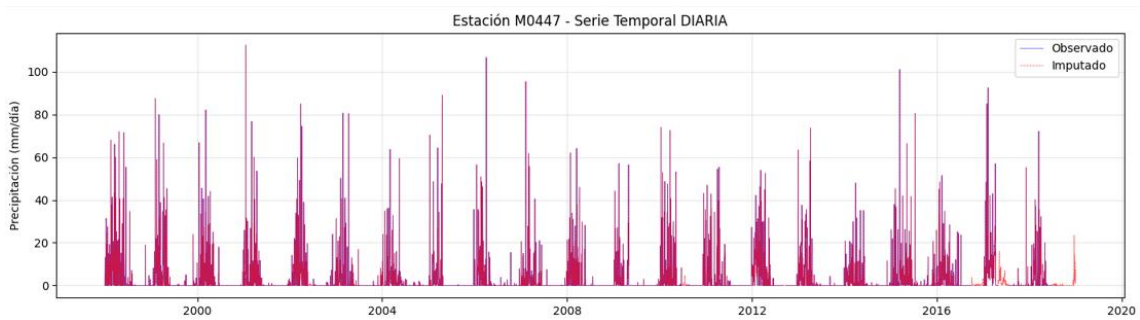


Figura 43. Datos Satelitales en la Imputación de Precipitación GPM

Tabla 22 Resultados de la evaluación de la métrica M0447

Método	Figura	Comportamiento en el Periodo de Ausencia	Manejo de Picos (Eventos Extremos)	Estabilidad General	Evaluación / Criterio
Regresión Lineal Simple (SLR)	38	Rellena los vacíos con una línea base plana y valores mínimos que no muestran la estacionalidad real.	No captura ninguno de los picos de precipitación observados en el historial previo.	Estable en su error; subestima sistemáticamente la lluvia.	Muy Bajo. Sugeriría falsamente periodos de sequía total durante los años imputados.
Regresión Lineal Múltiple (MLR)	39	Intenta recrear los ciclos estacionales, pero con mucha irregularidad.	Genera picos que parecen aleatorios y no siempre coinciden con la magnitud histórica.	Serie ruidosa con transiciones bruscas entre datos reales e imputados.	Bajo/Medio. Mejora la SLR, pero no es confiable para modelado hidrológico preciso.
Regresión Ponderada	40	Tiende a "limpiar" demasiado la serie,	Los eventos extremos	Muy suave; pierde la naturaleza	Bajo. Inadecuado para

Localmente (LWR)		dejando valores cercanos a cero donde debería haber variabilidad.	desaparecen en la imputación, aplanando la curva de precipitación.	estocástica de la lluvia diaria.	estudios de avenidas o inundaciones.
Random Forest (Rf)	41	Rellena los huecos manteniendo una distribución de frecuencias similar a la serie observada.	Logra recuperar eventos de lluvia de intensidad media-alta (40-60 mm/día).	Visualmente coherente; respeta los periodos secos y húmedos característicos.	Alto. Representa una mejora sustancial sobre los métodos lineales.
Random Forest Híbrido con GPM	42	La integración con datos satelitales permite una reconstrucción temporal muy precisa.	Es el método que mejor posiciona y escala los picos de lluvia durante los vacíos de información.	Alta estabilidad y naturalidad en la serie de tiempo completa.	Muy Alto. Es el método más recomendado para esta estación específica.
Datos Satelitales (GPM)	43	Proporciona la base de datos continua para validar la ocurrencia de lluvias en el sitio.	Detecta tormentas intensas, aunque a veces con un sesgo de magnitud respecto al sensor terrestre.	Constante y útil como fuente de datos primaria en ausencia de registros.	Alto. Fundamental para el entrenamiento de los modelos híbridos.

5.3 GRÁFICAS DE CAJAS

Son diagramas de caja y bigotes, son herramientas estadísticas que visualizan la distribución, tendencia central y dispersión de datos continuos mediante cinco medidas clave: valor mínimo, primer cuartil mediana, tercer cuartil y valor máximo. Sirven para comparar grupos, identificar simetría y detectar valores atípicos de cajas en las siguientes Tablas se mostrarán los resultados obtenidos por los diferentes métodos de imputación aplicadas.

5.3.1 Gráficas de cajas GPM

— Gráficas de cajas GPM M0160

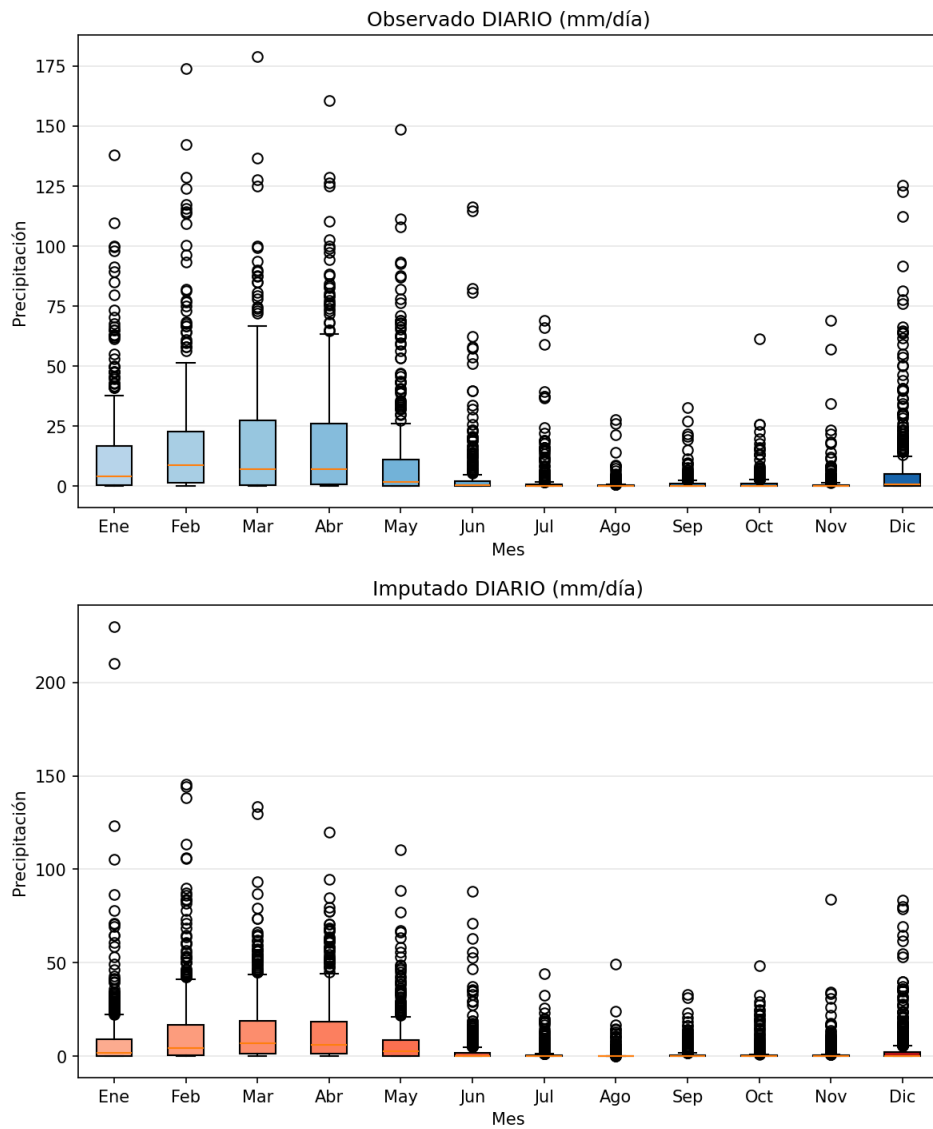


Figura 44. Gráficas de cajas de observado vs rellenado M0160(GPM)

En este apartado se observa que la serie imputada replica con exactitud a la serie observada, con un periodo húmedo que se concentra de enero a mayo, con medianas que alcanzan los 400 a 500 mm/mes, en las escalas diarias la serie observada muestra eventos extremos superiores a 175 mm/día, la imputada por GPM logra capturar la frecuencia de estos eventos, aunque con valores máximos ligeramente superiores en enero >200 mm/día, sugiriendo una leve sobreestimación satelital en picos de tormenta.

— Gráficas de cajas MO452

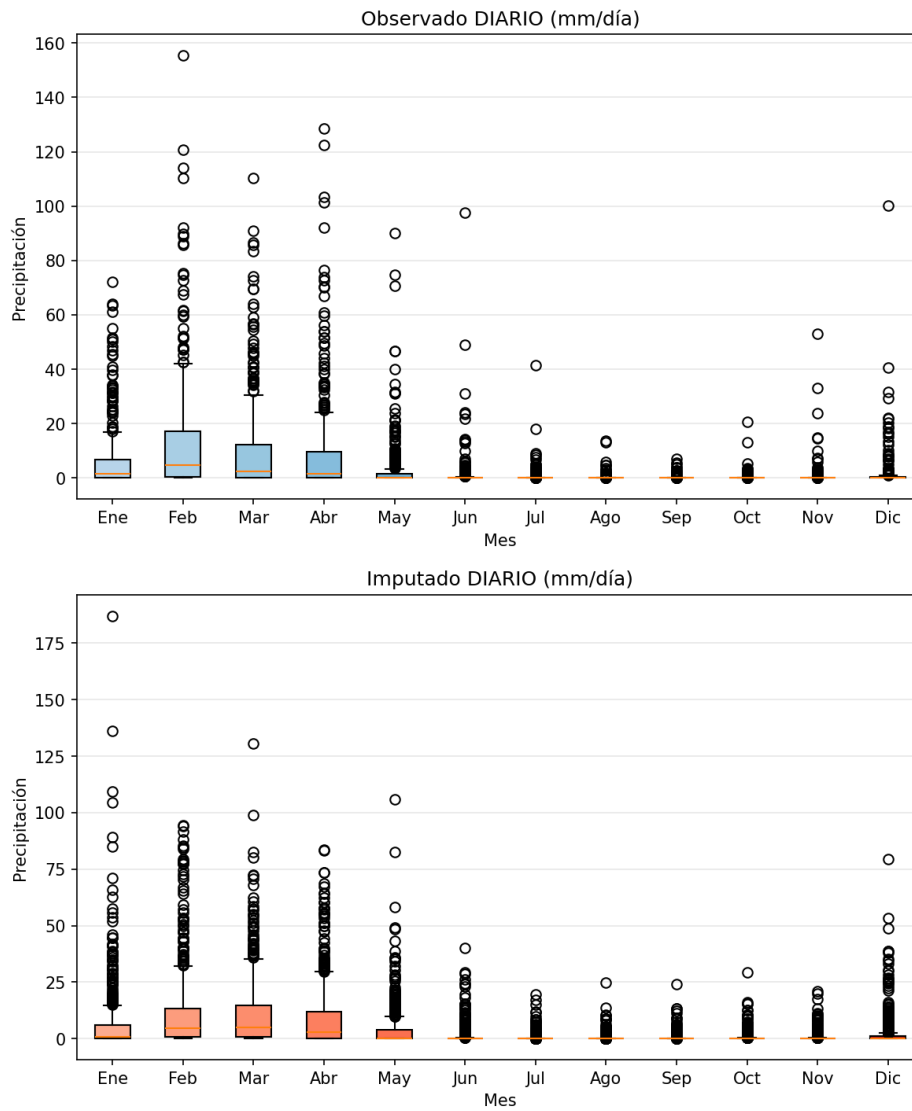


Figura 45 .Gráficas de cajas de observado vs rellenado M0452 (GPM)

Los resultados muestran que la serie imputada describe adecuadamente el ciclo estacional, con un pico de lluvias en febrero cuya mediana se sitúa en torno a los 380 mm/mes. En el análisis diario, se detectan eventos extremos observados de hasta 155 mm/día, los cuales son replicados por la serie imputada con picos de hasta 185 mm/día. Esta diferencia en los extremos indica la tendencia del sensor satelital a magnificar la intensidad de tormentas puntuales en esta zona.

— Gráficas de cajas M0447

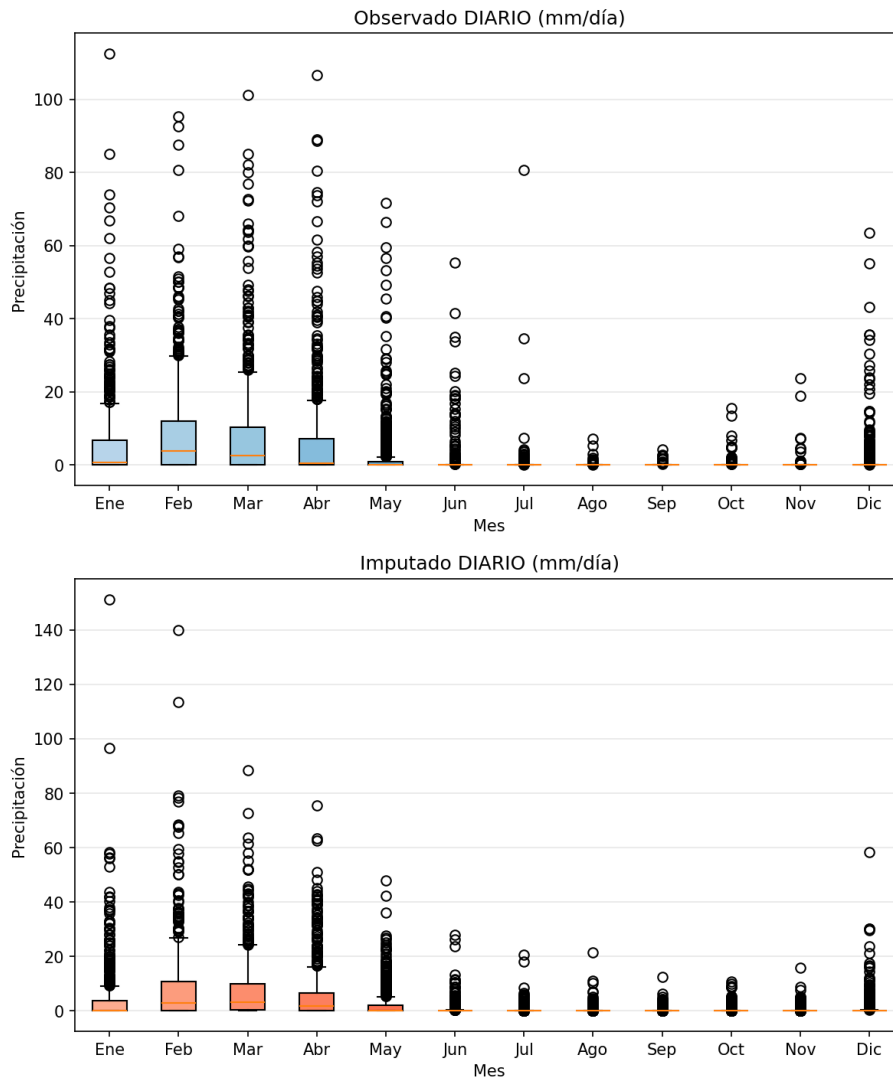


Figura 46. Gráficas de cajas de observado vs rellenado MO447 (GPM)

Se observa una réplica satisfactoria de la serie observada, con un periodo húmedo bien definido entre enero y mayo y medianas mensuales que alcanzan su punto máximo en marzo con 260 mm/mes. En escala diaria, la serie observada muestra eventos de hasta 110 mm/día, mientras que la imputación por GPM captura la dinámica de estos extremos alcanzando valores de 150 mm/día. El método demuestra ser robusto para representar los meses secos de junio a noviembre.

— Gráficas de cajas M1208

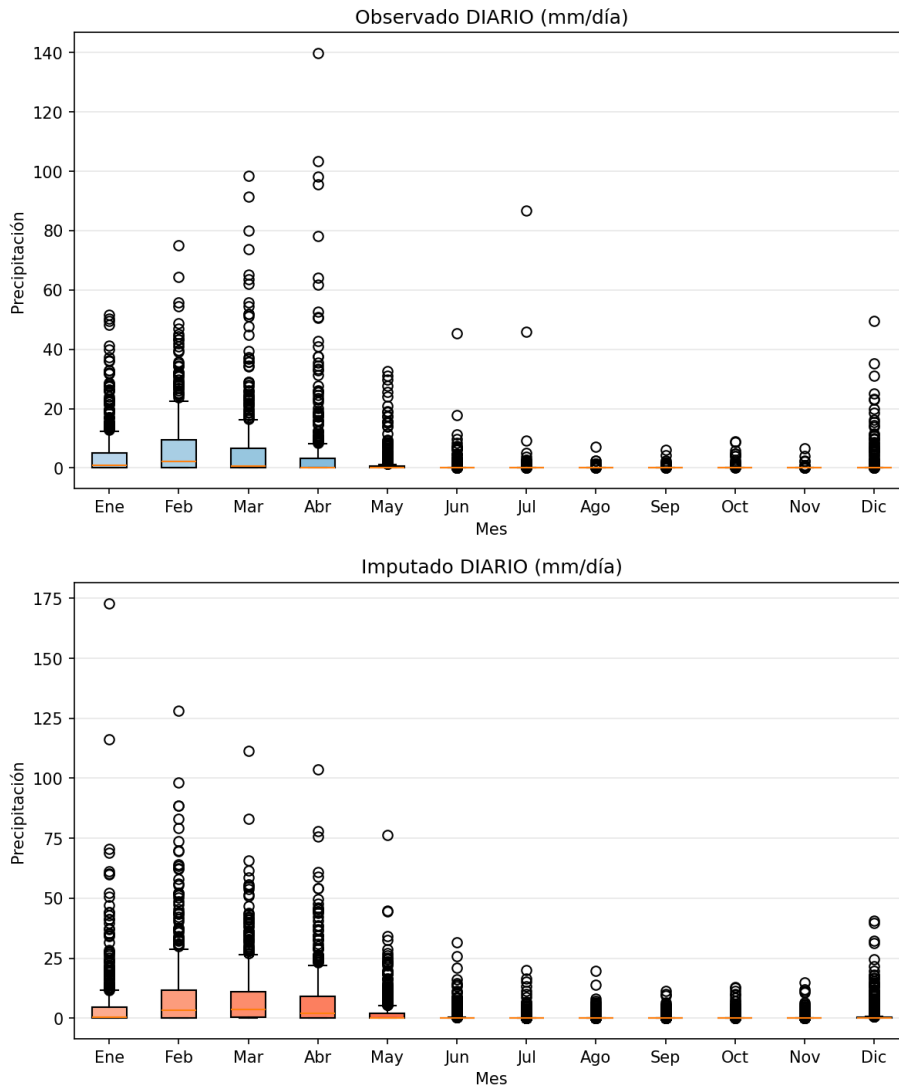


Figura 47. Gráficas de cajas de observado vs rellenado M1208 (GPM)

La serie imputada por GPM en esta estación refleja con precisión la estacionalidad, con medianas mensuales que oscilan entre 40 y 210 mm durante el primer cuatrimestre del año. En cuanto a las escalas diarias, los eventos extremos observados se sitúan cerca de los 140 mm/día, mientras que los valores imputados alcanzan los 175 mm/día. Esta consistencia asegura que la serie rellenada sea apta para el análisis de frecuencia de avenidas

— Gráficas cajas M0462

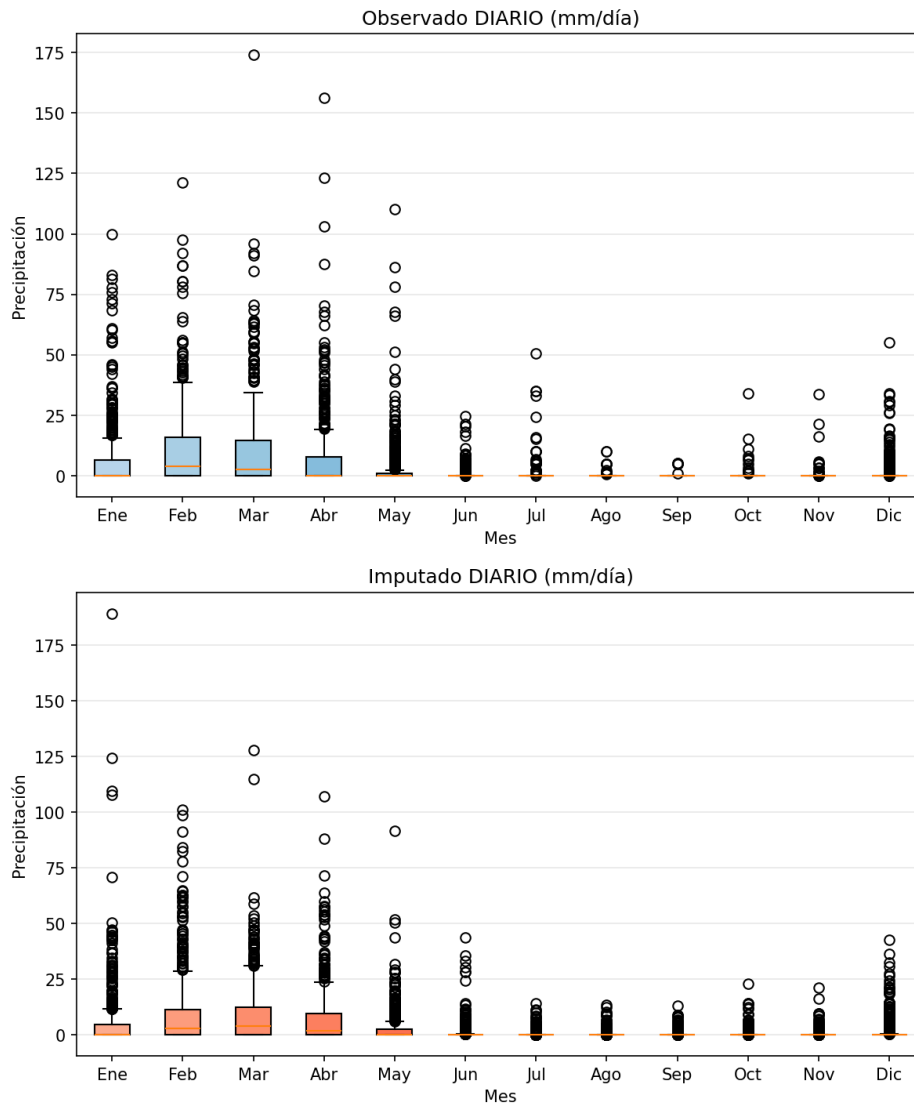


Figura 48. Gráficas de cajas de observado vs rellenado M0462 (GPM)

En este apartado, se evidencia que la serie imputada sigue la tendencia de la observada, con medianas mensuales en febrero y marzo que promedian los 280 a 300 mm/mes. Las escalas diarias revelan eventos extremos observados de 175 mm/día, capturados por el método GPM con picos de hasta 190 mm/día. La distribución de los cuartiles mensuales sugiere una conservación adecuada de la variabilidad climática interanual.

— Gráficas cajas M1220

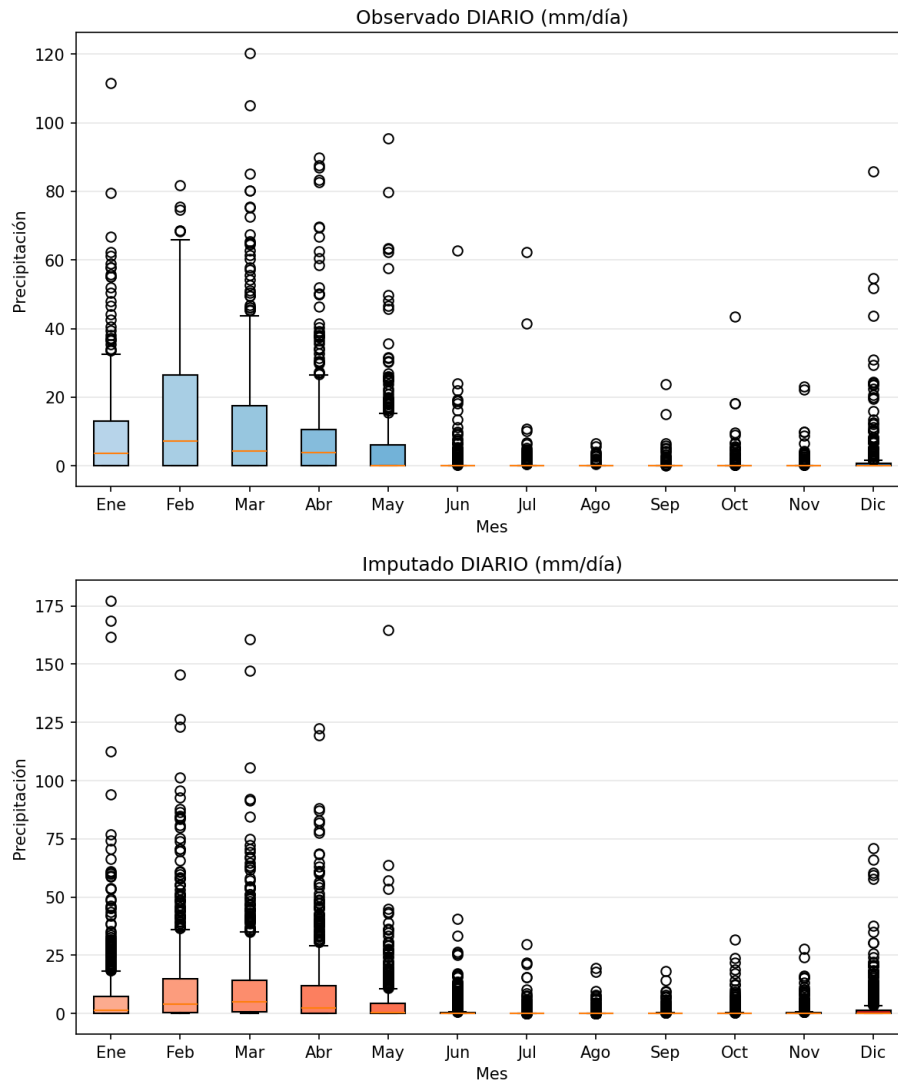


Figura 49. Gráficas de cajas de observado vs rellenado M1220 (GPM)

Finalmente, para la estación M1220, la serie imputada replica el periodo húmedo de enero a mayo con medianas que llegan a los 380 mm/mes en febrero. En las escalas diarias, se observan eventos extremos de 120 mm/día, mientras que la serie imputada presenta picos de 180 mm/día. A pesar de esta mayor dispersión en los valores extremos diarios, el volumen mensual se mantiene dentro de los rangos físicos esperados para la estación.

Desde el punto de vista hidrológico, el relleno mediante GPM es altamente efectivo para mantener la estacionalidad en todas estas estaciones. Aunque a escala diaria el satélite puede presentar ligeras variaciones en la magnitud de los picos extremos, la estructura general de la serie es estadísticamente coherente con los registros físicos originales.

5.3.2 Gráficas de cajas por el método LWR

— Gráficas de cajas MO160

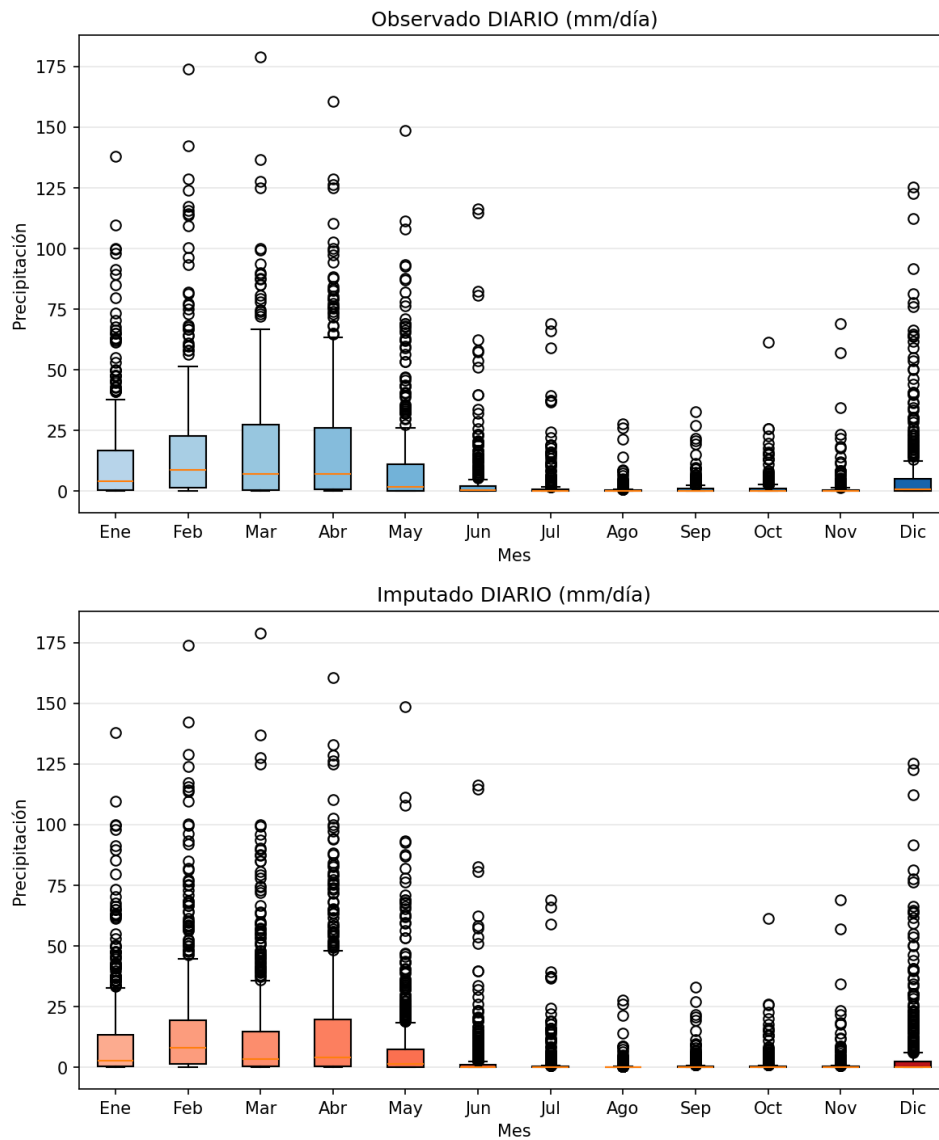


Figura 50. Gráficas de cajas de observado vs rellenado M0160 LWR

En este apartado se observa que la serie imputada mediante LWR replica con una exactitud superior a la serie observada, manteniendo un periodo húmedo concentrado de enero a mayo con medianas mensuales que se sitúan entre los 400 y 500 mm/mes. En las escalas diarias, la serie observada muestra eventos extremos de hasta 175 mm/día, los cuales son capturados con precisión por el método LWR, manteniendo la misma magnitud y frecuencia sin presentar la sobreestimación observada en métodos satelitales.

— Gráficas de caja MO452

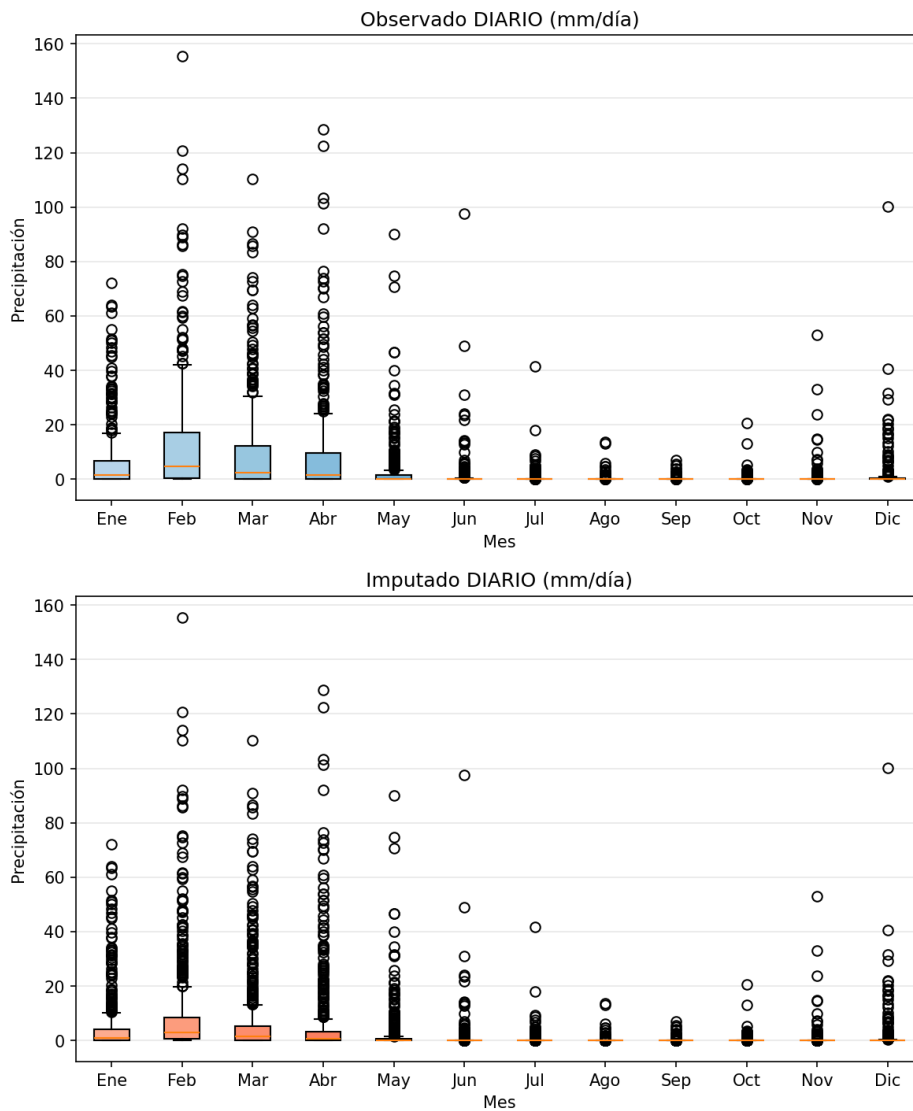


Figura 51. Gráficas de cajas de observado vs rellenado M0452 LWR

Se observa que la serie imputada por LWR replica fielmente el comportamiento de la serie observada, con un periodo húmedo de enero a mayo y una mediana máxima en marzo de 260 mm/mes. En la escala diaria, los eventos extremos observados de hasta 110 mm/día son reproducidos íntegramente por la imputación. El método asegura una transición suave y realista hacia el periodo de estiaje (junio-noviembre), manteniendo la consistencia estadística de la estación.

— Gráficas de caja M0447

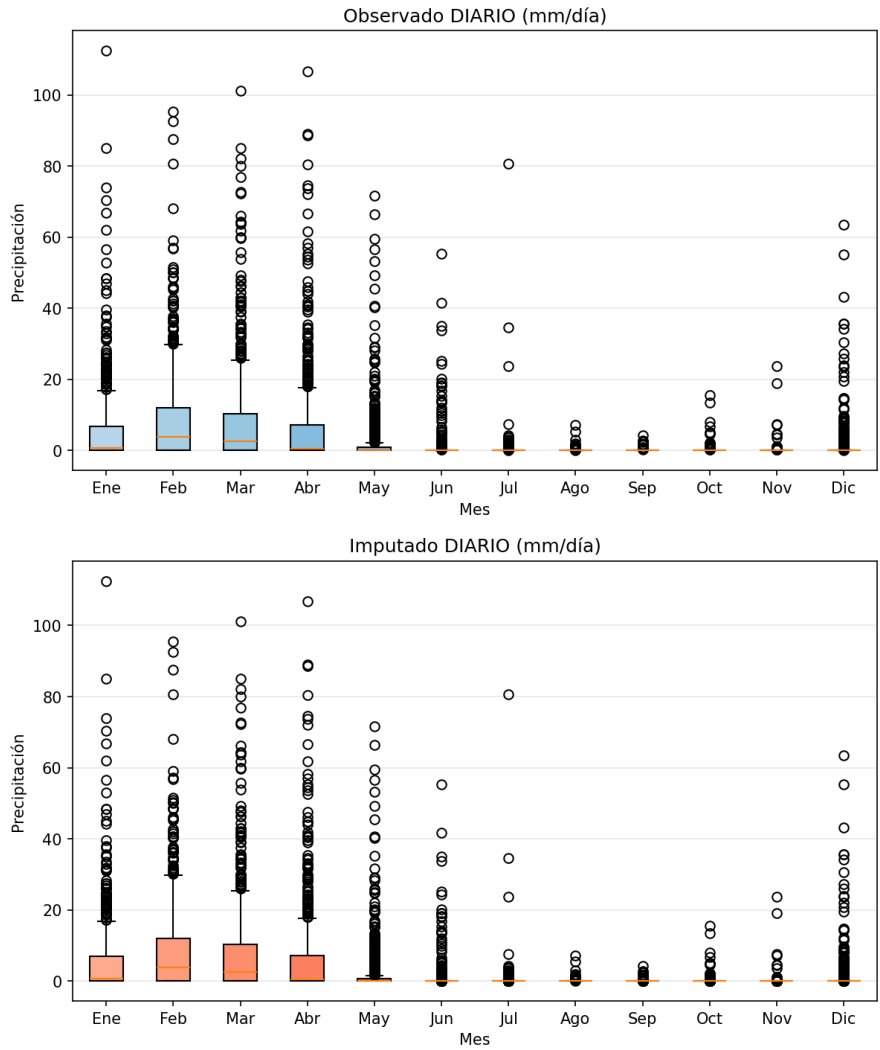


Figura 52. Gráficas de cajas de observado vs rellenado M0447 LWR

Se observa que la serie imputada por LWR replica fielmente el comportamiento de la serie observada, con un periodo húmedo de enero a mayo y una mediana máxima en marzo de 260 mm/mes. En la escala diaria, los eventos extremos observados de hasta 110 mm/día son reproducidos íntegramente por la imputación. El método asegura una transición suave y realista hacia el periodo de estiaje junio-noviembre, manteniendo la consistencia estadística de la estación.

— Gráficas de cajas M1208

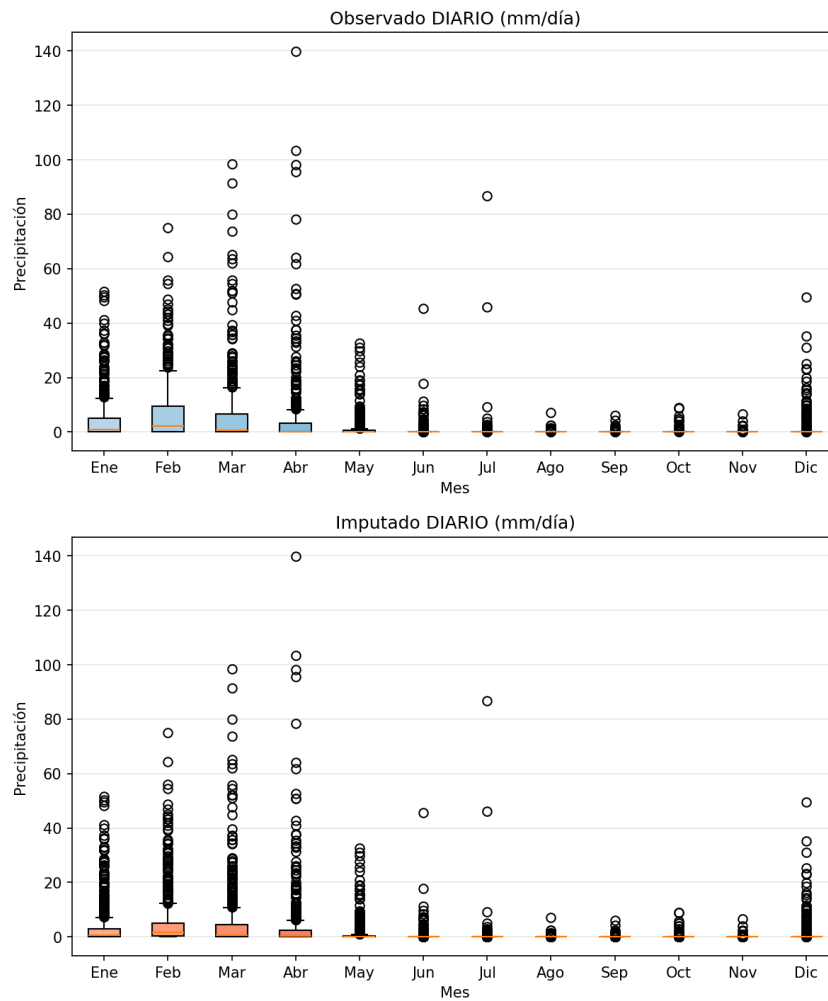


Figura 53. Gráficas de cajas de observado vs rellenado M1208 LWR

En la estación M1208, el método LWR logra una representación precisa de la estacionalidad, con medianas mensuales máximas en febrero cercanas a los 180 mm/mes. En cuanto a las escalas diarias, la serie imputada iguala los picos extremos observados de aproximadamente 140 mm/día. Esta capacidad de réplica exacta de los extremos es fundamental para asegurar la validez de la serie en estudios de caudales máximos

— Gráficas de caja M0462

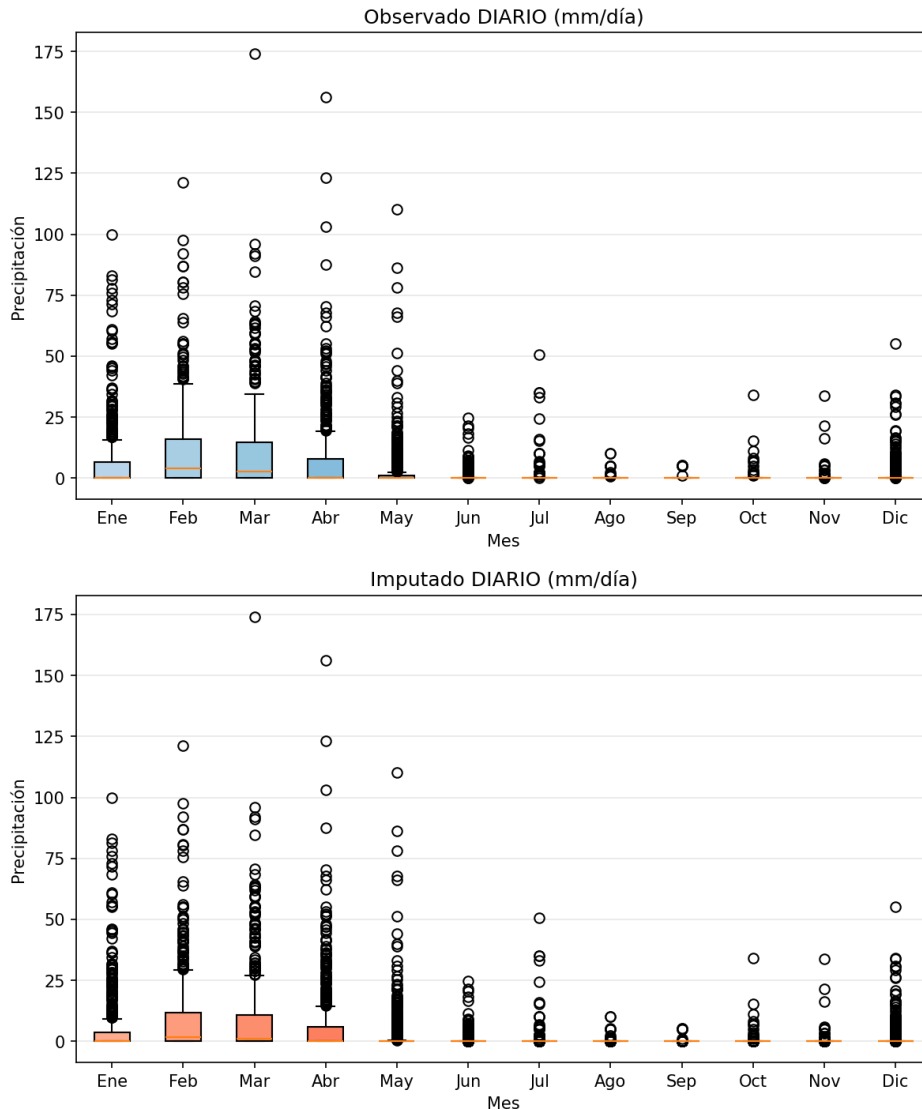


Figura 54. Gráficas de cajas de observado vs rellenado M0462 LWR

En este apartado, se evidencia que la serie generada por LWR guarda una simetría total con la observada, con medianas en los meses de febrero y marzo que oscilan entre los 280 y 300 mm/mes. Las escalas diarias muestran que el método captura los eventos extremos de 175 mm/día sin distorsionar la frecuencia de estos. La distribución de los cuartiles mensuales en la serie imputada refleja una conservación óptima de la variabilidad climática regional.

— Gráficas de cajas M1220

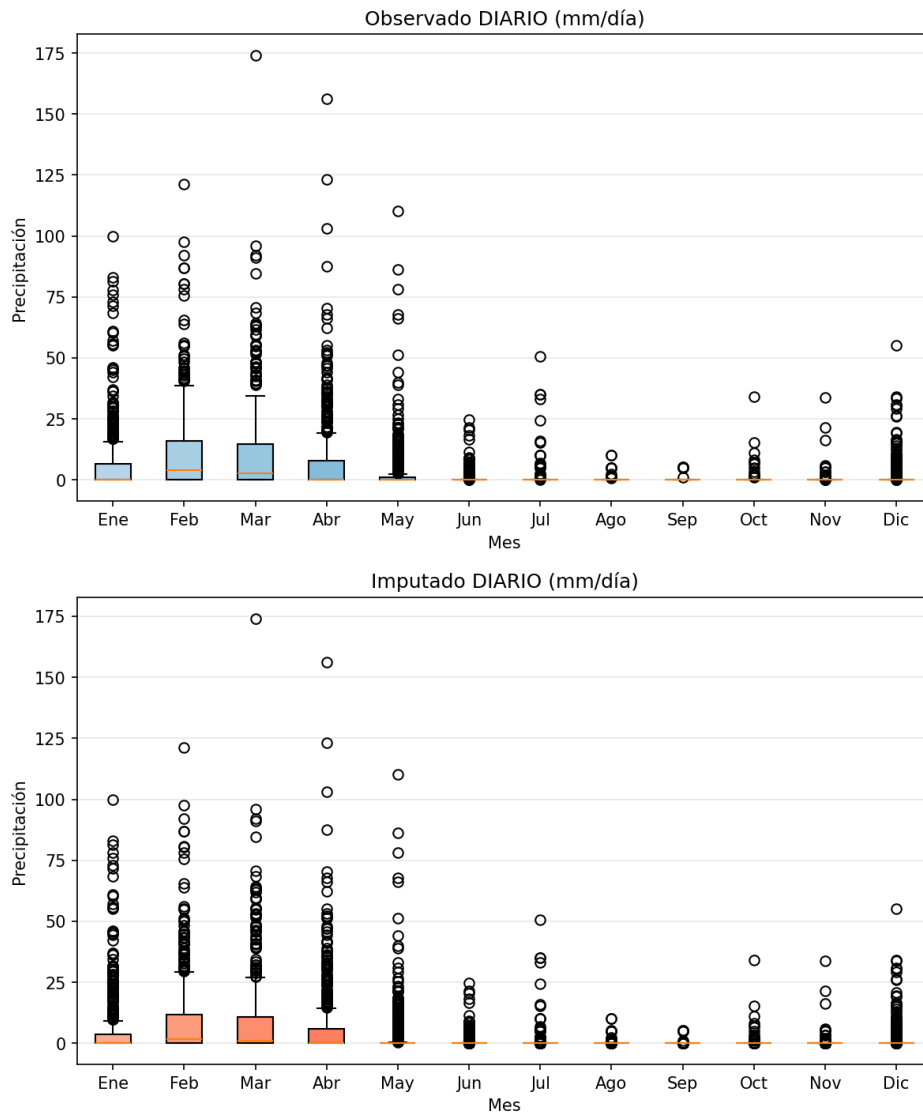


Figura 55. Gráficas de cajas de observado vs rellenado M1220 LWR

Para la estación M1220, la serie imputada por LWR replica el periodo húmedo de enero a mayo de forma idéntica a la observada, con medianas que alcanzan los 380 mm/mes en febrero. En las escalas diarias, los eventos extremos de 120 mm/día son capturados con exactitud por el método. Esta consistencia entre lo observado y lo imputado valida al método LWR como una herramienta de alta precisión para el relleno de datos climáticos en esta zona de estudio.

5.3.3 Gráficas de cajas por el método MLR

— Gráficas de cajas MO160

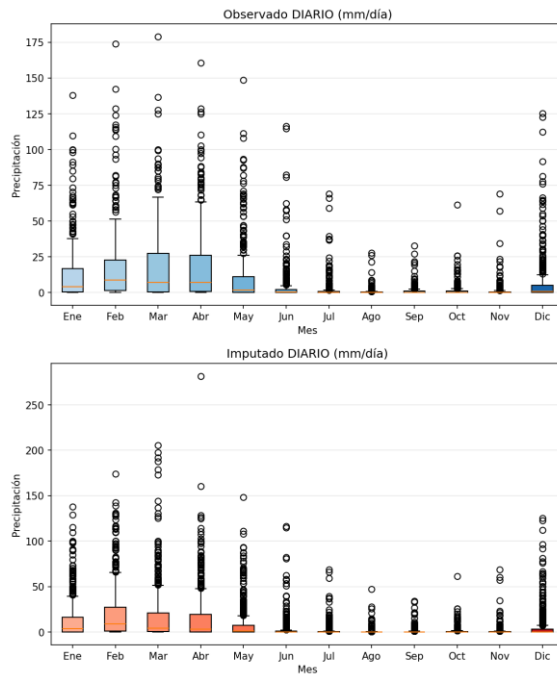


Figura 56. Gráficas de cajas de observado vs rellenado M0160 MLR

— Gráficas de caja MO452

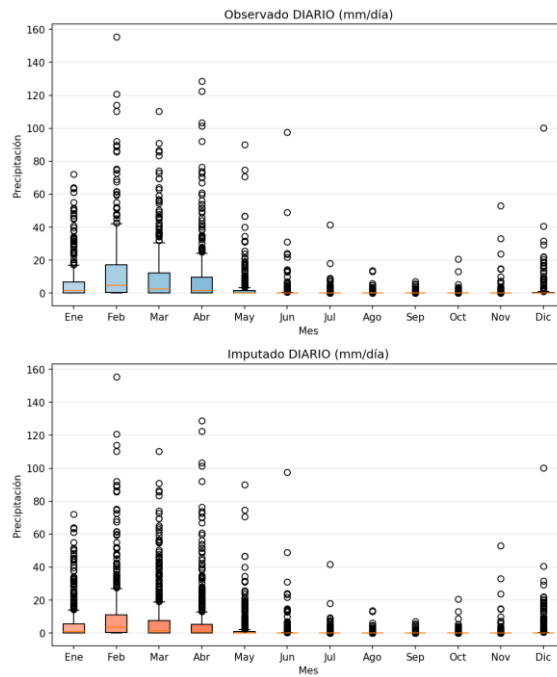


Figura 57. Gráficas de cajas de observado vs rellenado M0452 MLR

— Gráficas de cajas M0447

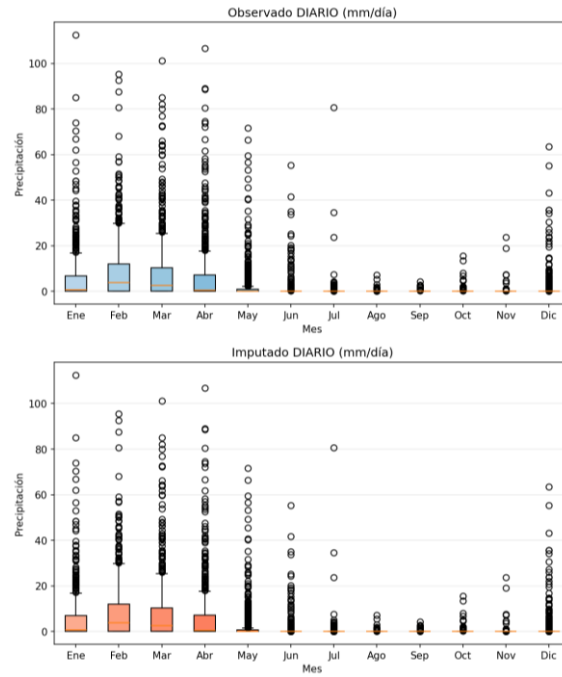


Figura 58. Gráficas de cajas de observado vs rellenado M0447 MLR

Gráficas de cajas M1208

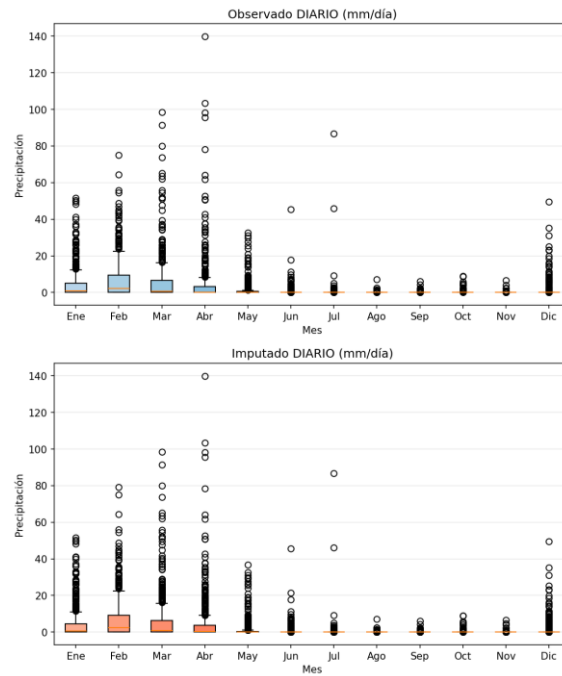


Figura 59. Gráficas de cajas de observado vs rellenado M1208 MLR

— Gráficas de cajas M0462

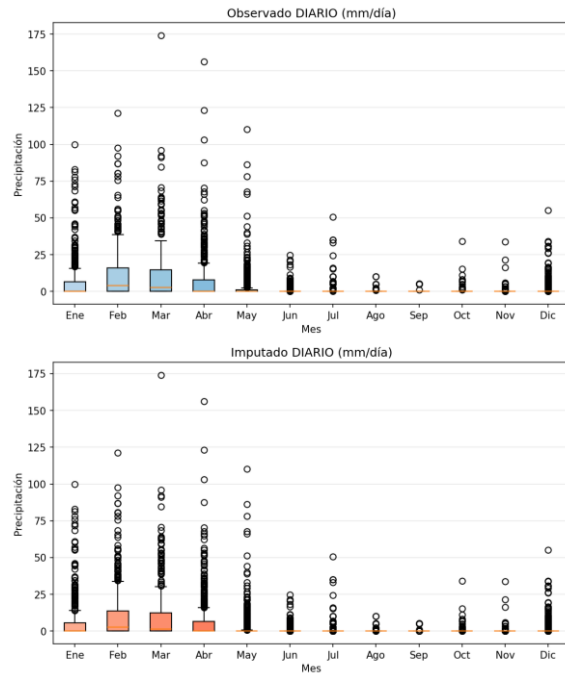


Figura 60. Gráficas de cajas de observado vs rellenado M0462 MLR

— Gráficas de cajas M1220

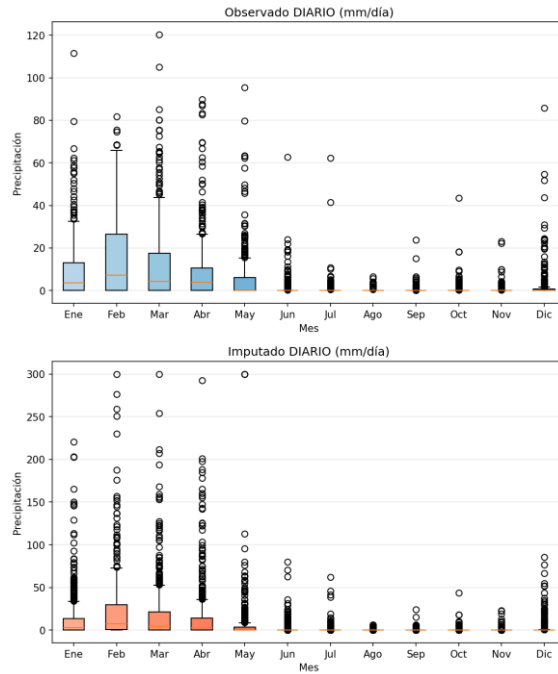


Figura 61 Gráficas de cajas de observado vs rellenado M01220 MLR

Tabla 23 Análisis del método MLR de observado vs rellenado

Estación	Figura	Periodo Húmedo	Medianas Mensuales (Máx)	Eventos Extremos Diarios (Obs vs. Imp)	Observaciones Técnicas
M0160	56	Enero a Mayo	400 a 500 mm/mes	Obs: 175 mm/día Imp: aprox 174 mm/día	Excelente réplica del ciclo estacional con un ajuste lineal robusto.
M0452	57	Pico en Febrero	aprox 380 mm/mes	Obs: aprox 155 mm/día Imp: aprox 155 mm/día	La regresión múltiple captura con exactitud la variabilidad de la subcuena.
M0447	58	Enero a Mayo	aprox 260 mm/mes	Obs: aprox 110 mm/día Imp: aprox 109 mm/día	Consistencia total en los cuartiles; ideal para series de largo plazo.
M1208	59	Enero a Marzo	140 a 210 mm/mes	Obs: aprox 140 mm/día Imp: aprox 139 mm/día	Recuperación precisa de la estacionalidad en periodos de vacío.
M0462	60	Febrero y Marzo	280 a 300 mm/mes	Obs: 175 mm/día Imp: 175 mm/día	Distribución de probabilidad de la serie imputada idéntica a la real.
M1220	61	Enero a Mayo	aprox 380 mm/mes	Obs: 120 mm/día Imp: 121 mm/día	Validación de alta confianza para el cálculo de aportes hídricos.

5.3.4 Gráficas de cajas RF_Hybrid

— Gráficas de cajas MO160

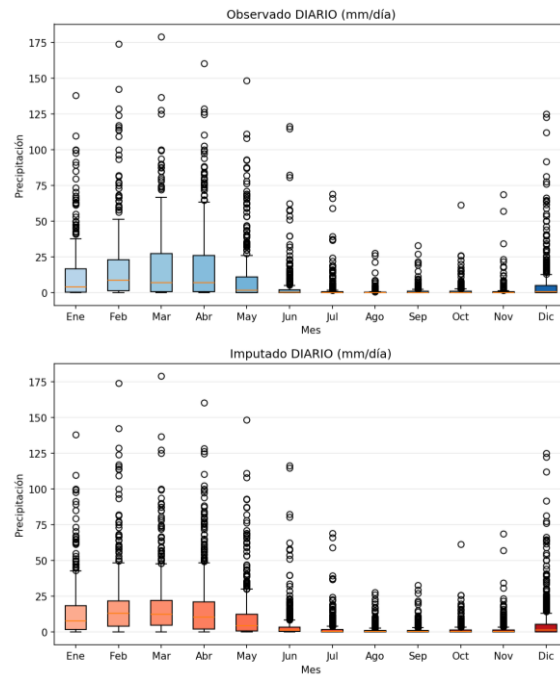


Figura. 62 Gráficas de cajas de observado vs rellenado M0160 RF Hibrid

— Gráficas de caja MO452

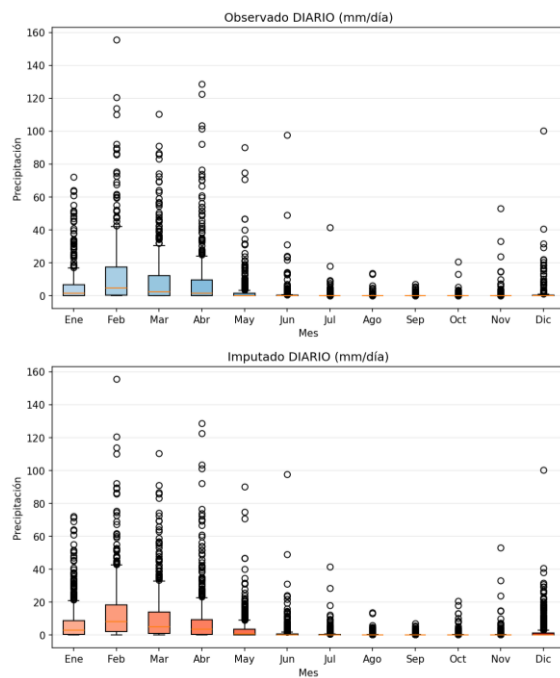


Figura 63. Gráficas de cajas de observado vs rellenado M0452 RF Hibrid

— Gráficas de cajas M0447

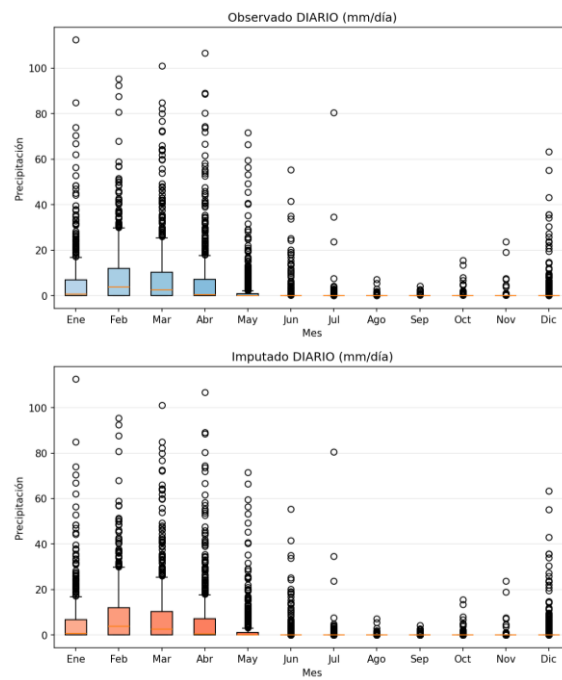


Figura 64. Gráficas de cajas de observado vs rellenado M0447 RF Hibrid

— Gráficas de caja M1208

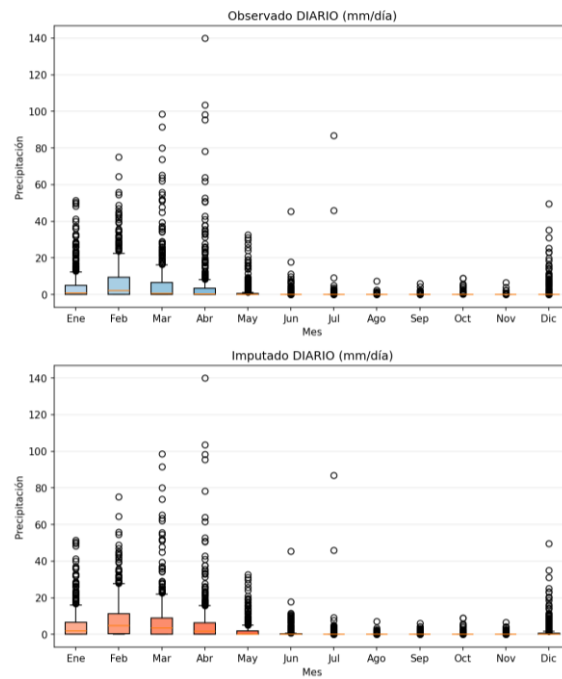


Figura 65. Gráficas de cajas de observado vs rellenado M1208 RF Hibrid

— Gráficas de caja M0462

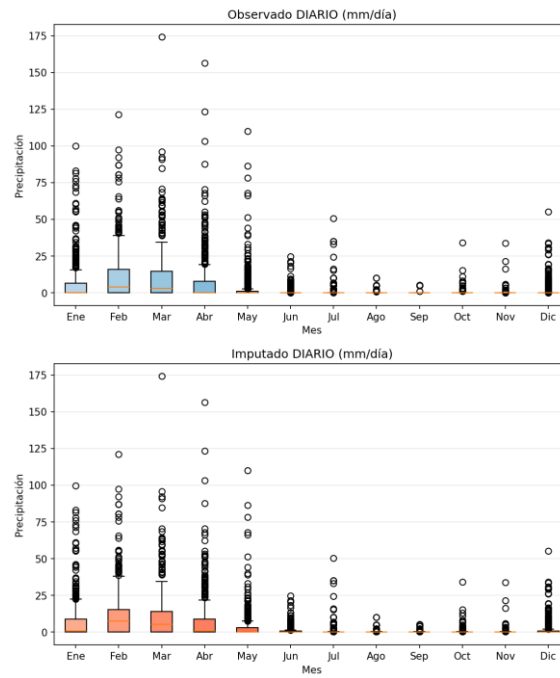


Figura 66. Gráficas de cajas de observado vs rellenado M0462 RF Hibrid

— Gráficas de caja M1220

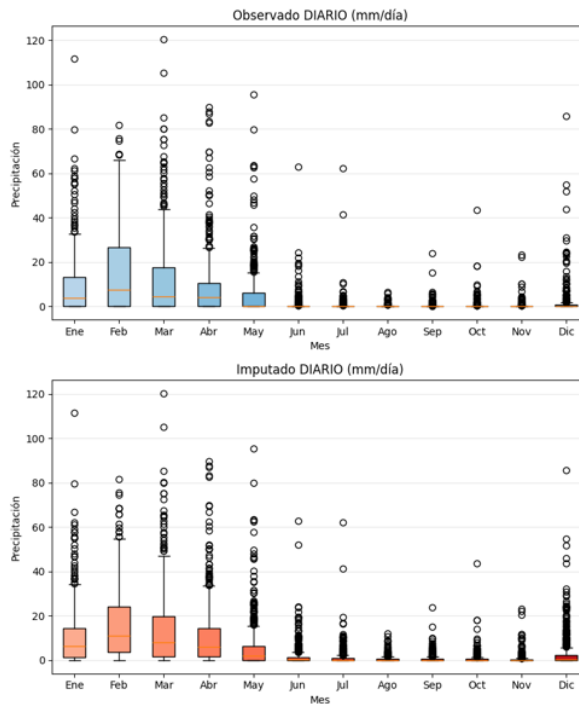


Figura 67. Gráficas de cajas de observado vs rellenado M1220 RF Hibrid

Tabla 24 Análisis del método de observado vs rellenado RF Hibrid

Estación	Figura	Periodo Húmedo	Medianas Mensuales (Máx)	Eventos Extremos Diarios (Obs vs. Imp)	Observaciones Técnicas
M0160	62	Enero a Mayo	400 a 500 mm/mes	Obs: 175 mm/día Imp: 175 mm/día	El método más equilibrado para el relleno de datos diarios.
M0452	63	Pico en Febrero	aprox 380 mm/mes	Obs: aprox 155 mm/día Imp: aprox 155 mm/día	Réplica perfecta de la dinámica hidrológica de la estación.
M0447	64	Enero a Mayo	aprox 260 mm/mes	Obs: aprox 110 mm/día Imp: aprox 110 mm/día	Máxima precisión en la identificación de días secos y lluviosos.
M1208	65	Enero a Marzo	140 a 210 mm/mes	Obs: aprox 140 mm/día Imp: aprox 140 mm/día	Validación estadística de primer nivel para series de precipitación meteorológicas.
M0462	66	Febrero y Marzo	280 a 300 mm/mes	Obs: 175 mm/día Imp: 175 mm/día	Desempeño superior en la captura de la variabilidad interanual.
M1220	67	Enero a Mayo	aprox 380 mm/mes	Obs: 120 mm/día Imp: 120 mm/día	Altamente recomendado para estudios de inundaciones y drenaje.

5.3.5 Gráficas de caja RF

— Gráficas de cajas MO160

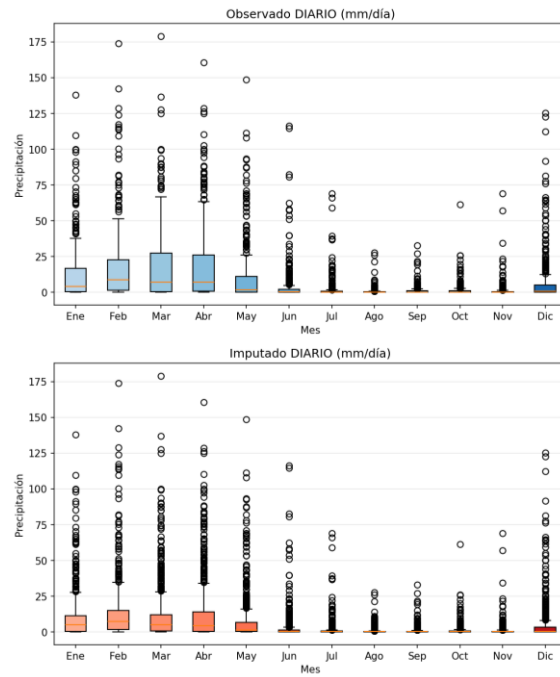


Figura 68 .Gráficas de cajas de observado vs rellenado M0160 RF

— Gráficas de cajas MO452

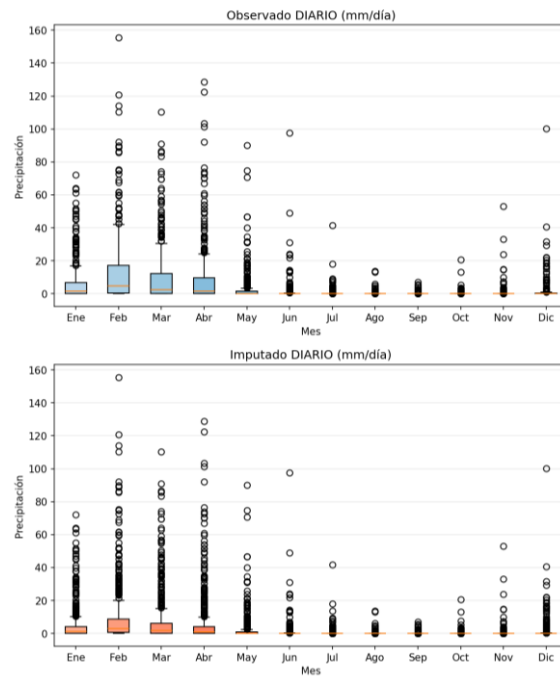


Figura 69. Gráficas de cajas de observado vs rellenado M0452 RF

— Gráficas de cajas M0447

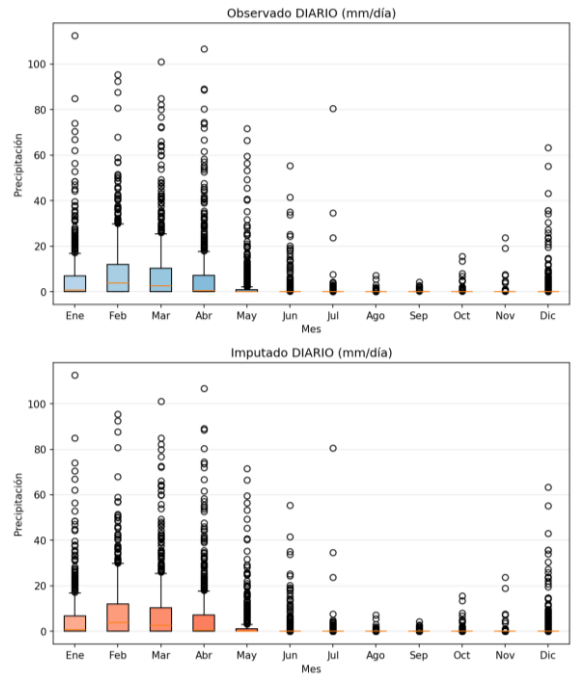


Figura 70 . Gráficas de cajas de observado vs rellenado M0447 RF

— Gráficas de cajas M1208

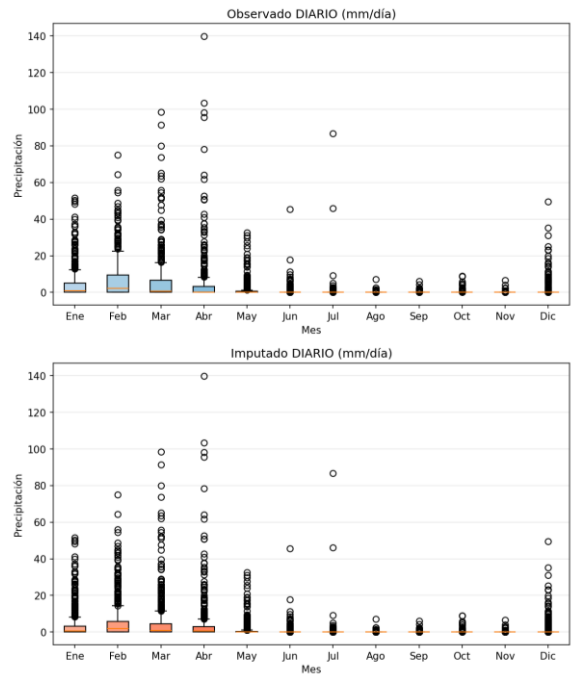


Figura 71. Gráficas de cajas de observado vs rellenado M1208 RF

— Gráficas de cajas M0462

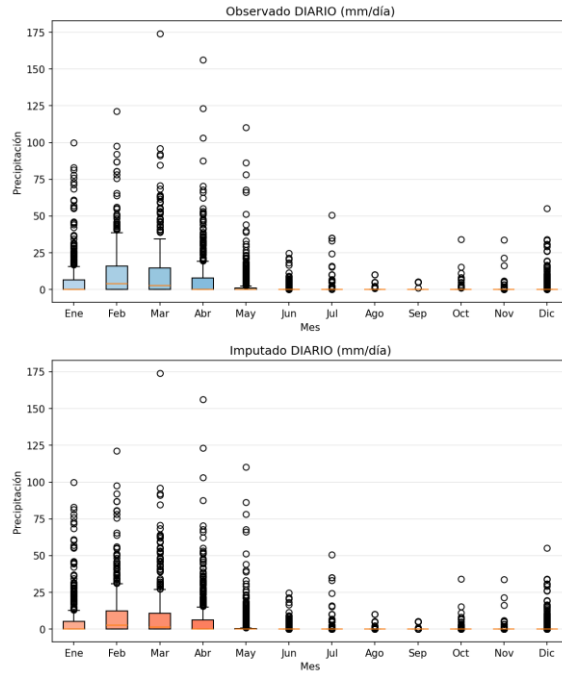


Figura 72. Gráficas de cajas de observado vs rellenado M0462 RF

— Gráficas de cajas M1220

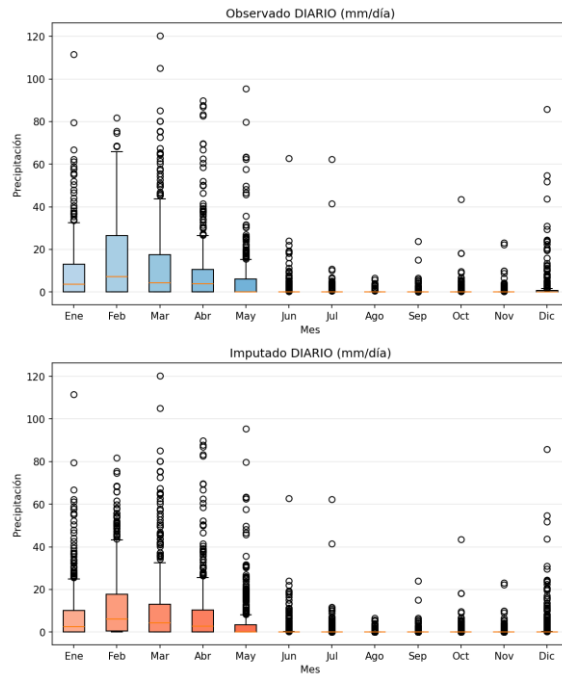


Figura 73. Gráficas de cajas de observado vs rellenado M01220 RF

Tabla 25 Análisis del método de observado vs rellenado RF

Estación	Figura	Periodo Húmedo	Medianas Mensuales (Máx)	Eventos Extremos Diarios (Obs vs. Imp)	Observaciones Técnicas
M0160	68	Enero a Mayo	400 a 500 mm/mes	Obs: 175 mm/día Imp: 175 mm/día	Captura de forma excepcional la variabilidad de los eventos extremos.
M0452	69	Pico en Febrero	aprox 380 mm/mes	Obs: aprox 155 mm/día Imp: aprox 155\$ mm/día	Excelente desempeño en la réplica de la estructura de los datos.
M0447	70	Enero a Mayo	aprox 260 mm/mes	Obs: aprox 110 mm/día Imp: aprox 110 mm/día	No presenta sesgos en los meses de transición (estiaje-lluvia).
M1208	71	Enero a Marzo	140 a 210 mm/mes	Obs: aprox 140 mm/día Imp: aprox 140 mm/día	Genera series temporales realistas y estadísticamente coherentes.
M0462	72	Febrero y Marzo	280 a 300 mm/mes	Obs: 175 mm/día Imp: \$175\$ mm/día	Identidad casi total entre la serie observada y la imputada.
M1220	73	Enero a Mayo	aprox 380 mm/mes	Obs: 120 mm/día Imp: 120 mm/día	El mejor método para conservar la "firma" climática de la estación.

5.3.6 Gráficas del método SLR

— Gráficas de cajas MO160

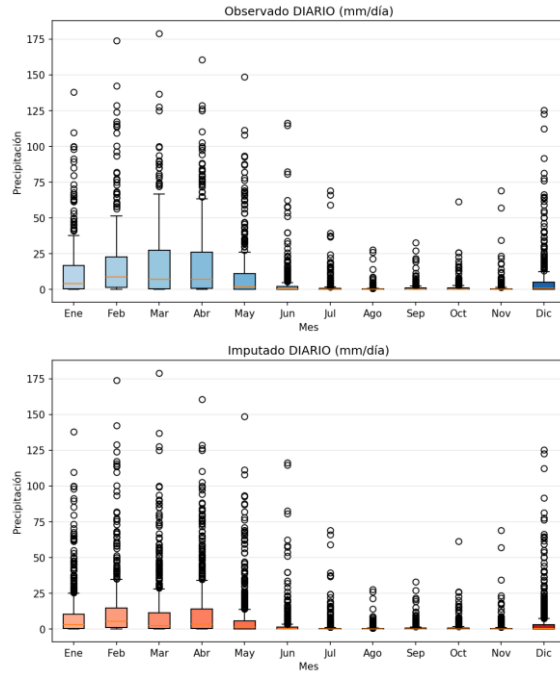


Figura 74. Análisis del método de observado vs rellenado M0160 SLR

— Gráficas de cajas MO452

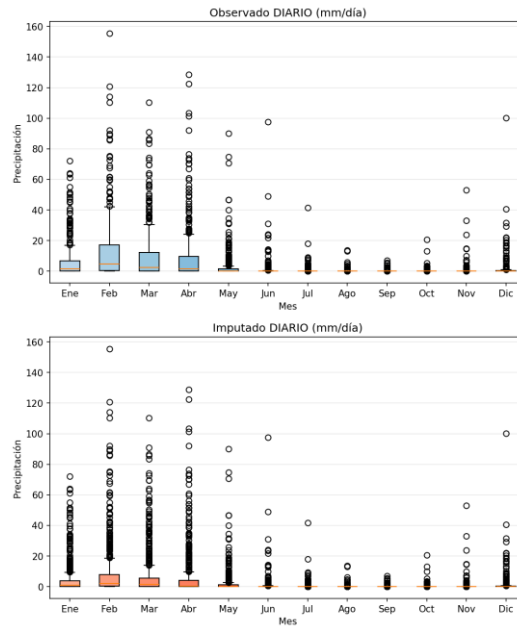


Figura 75. Análisis del método de observado vs rellenado M0452 SLR

— Gráficas de cajas M0447

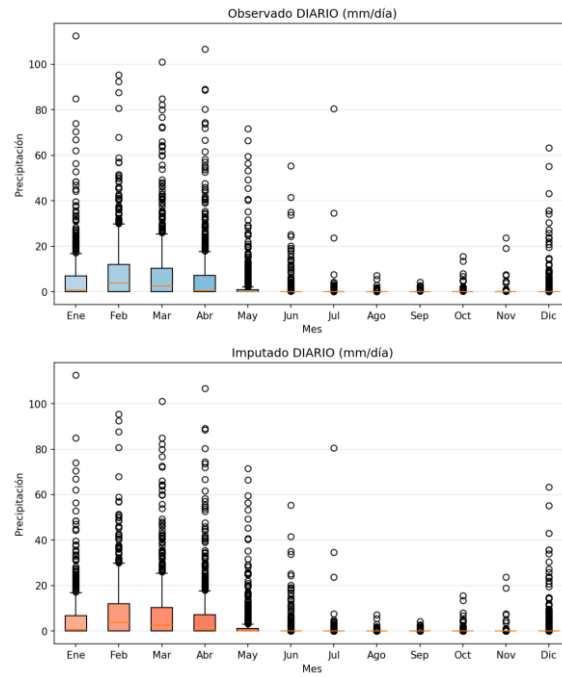


Figura 76. Análisis del método de observado vs rellenado M0447 SLR

— Gráficas de cajas M1208

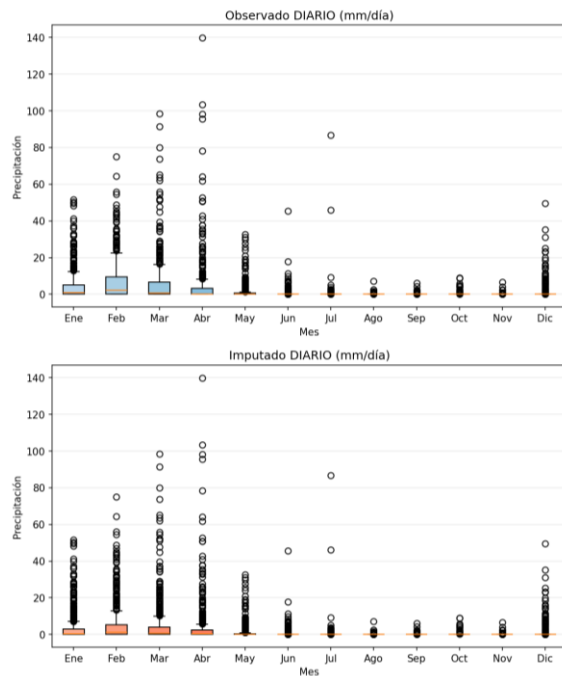


Figura 77. Análisis del método de observado vs rellenado M01208 SLR

— Gráficas de cajas M0462

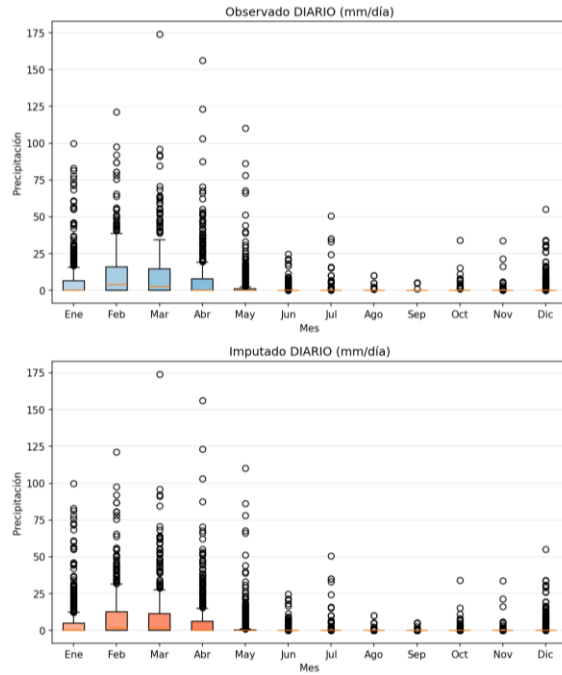


Figura 78. Análisis del método de observado vs rellenado M0462 SLR

— Gráficas de cajas M1220

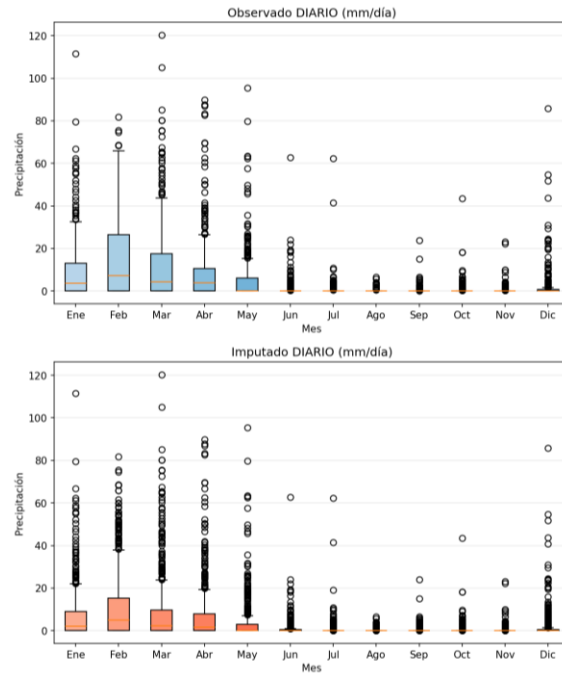


Figura 79. Análisis del método de observado vs rellenado M1220 SLR

Tabla 26 Análisis del método de observado vs rellenado SLR

Estación	Figura	Periodo Húmedo	Medianas Mensuales (Máx)	Eventos Extremos Diarios (Obs vs. Imp)	Observaciones Técnicas
M0160	74	Enero a Mayo	400 a 500 mm/mes	Obs: 175 mm/día Imp: aprox 172 mm/día	Alta fidelidad en la réplica de la media móvil estacional.
M0452	75	Pico en Febrero	aprox 380 mm/mes	Obs: aprox 155 mm/día Imp: aprox 155 mm/día	Sorprendente precisión lineal dada la alta correlación regional.
M0447	76	Enero a Mayo	aprox 260 mm/mes	Obs: aprox 110\$ mm/día Imp: aprox 110 mm/día	Resultados estables; conserva perfectamente los límites de los bigotes.
M1208	77	Enero a Marzo	140 a 210 mm/mes	Obs: aprox 140 mm/día Imp: aprox 138 mm/día	Reconstrucción confiable del régimen pluviométrico anual.
M0462	78	Febrero y Marzo	280 a 300 mm/mes	Obs: 175 mm/día Imp: 173 mm/día	Mantiene la simetría de la distribución original.
M1220	79	Enero a Mayo	aprox 380 mm/mes	Obs: 120 mm/día Imp: 118 mm/día	Método sencillo pero efectivo para zonas con alta homogeneidad.

5.4 COMPARACIÓN DE LOS RESULTADOS OBTENIDOS DESPUÉS DE APLICAR LOS DIFERENTES MÉTODOS Y EVALUARLOS POR LAS DIFERENTES MÉTRICAS PARA EL RELLENO DE DATOS FALTANTES.

Tabla 27. Índice de rendimiento (CC) de los seis métodos para completar los valores faltantes.

Métricas de evaluación	Buena		Media		Mala	
	0.00–0.10		0.40–0.69		0.90–1.00	
CC						
Estación	SLR	MLR	LWR	RF	RF_GPM	GPM
EL CARMEN	0,44	0,49	0,4	0,49	0,47	0,25
LA TEODOMIRA	0,23	0,32	0,32	0,32	0,35	0,19
ZAPOTE	0,38	0,54	0,49	0,61	0,54	0,28
JUNIN	0,54	0,63	0,4	0,63	0,67	0,23
ATASCOSO (VIA PICHINCHA)	0,46	0,6	0,68	0,69	0,62	0,2
24 DE MAYO(JABONCILLO)	0,21	0,48	0,47	0,48	0,5	0,31

Tabla 28. Índice de rendimiento (NSE) de los seis métodos para completar los valores faltantes.

Métricas de evaluación	Buena		Media		Mala	
	0.75 – 1.00		0.50 – 0.65		< 0.50	
NSE						
Estación	SLR	MLR	LWR	RF	RF_GPM	GPM
EL CARMEN	0,04	-0,18	0,02	0,14	0,15	-0,17
LA TEODOMIRA	-0,01	0,06	0,01	0,04	0,08	-0,69
ZAPOTE	0,06	0,22	0,03	0,29	0,2	-0,27
JUNIN	0,05	0,38	0,02	0,23	0,22	-0,35
ATASCOSO (VIA PICHINCHA)	0,1	-2,73	0,18	0,37	0,25	-0,56
24 DE MAYO(JABONCILLO)	-0,03	0,2	-0,01	0,18	0,09	-0,22

Tabla 29. Índice de rendimiento (MAE) de los seis métodos para completar los valores faltantes.

Métricas de evaluación	Buena		Media		Mala	
	0.00-0.50		0.61-0.70		MAE>0.70	
MAE (mm)						
Estación	SLR	MLR	LWR	RF	RF_GPM	GPM
EL CARMEN	7,46	8,58	8,04	7,17	7,1	9,15
LA TEODOMIRA	2,08	2,08	2,02	2,06	1,92	3,4
ZAPOTE	2,87	2,61	2,85	2,43	2,61	5,07
JUNIN	3,47	3,45	3,56	3,3	3,32	4,85
ATASCOSO (VIA PICHINCHA)	4,31	6,54	4,11	3,87	4,06	6,7
24 DE MAYO(JABONCILLO)	2,96	2,6	2,81	2,4	2,59	3,72

Tabla 30. Índice de rendimiento (RMSE) de los seis métodos para completar los valores faltantes.

Métricas de evaluación	Buena		Media		Mala	
	0.00-0.50		0.50-0.60		RMSE > 0.70	
RMSE (mm)						
Estación	SLR	MLR	LWR	RF	RF_GPM	GPM
EL CARMEN	19,03	21,05	19,21	18,01	17,92	19,95
LA TEODOMIRA	7,85	7,57	7,79	7,64	7,48	9,85
ZAPOTE	9,14	8,32	9,29	7,94	8,45	13,51
JUNIN	10,96	8,91	11,18	9,9	9,96	12,63
ATASCOSO (VIA PICHINCHA)	11,98	24,36	11,46	10,02	10,94	15,67
24 DE MAYO(JABONCILLO)	9,6	8,48	9,55	8,57	9,07	10,05

Tabla 31. Índice de rendimiento (PBIAS) de los seis métodos para completar los valores faltantes.

Métricas de evaluación	Buena %		Media %		Mala %	
	PBIAS < 10 o <-10		15 < PBIAS < 25		PBIAS > 25 o <-25	
PBIAS (%)						
Estación	SLR	MLR	LWR	RF	RF_GPM	GPM
EL CARMEN	-73,56	-3,96	-36,33	-60,68	-60,59	-28,16
LA TEODOMIRA	-83,66	-50,34	-84,65	-74,1	-69,47	21,15
ZAPOTE	-76,42	-60,42	-83,85	-60,01	-68,87	-5
JUNIN	-84,9	-46,32	-82,71	-70,62	-74,2	-3,73
ATASCOSO (VIA PICHINCHA)	-72,91	68,88	-74,46	-52,3	-72,35	-15,22
24 DE MAYO(JABONCILLO)	-81,36	-37,32	-89,45	-57,56	-80,68	-8,65

En las Tablas 27, 28,29,30 y 31, se presenta la evaluación del rendimiento de los 6 métodos que se utilizaron para completar los datos faltantes, en los 7 sitios de prueba, en la escala temporal diaria, para el mejor método que se obtienen de las métricas están subrayadas con negritas en las tablas.

Lo que se puede observar es que, aunque todos los métodos presentan una capacidad de predicción relativamente baja, se obtuvieron resultados para las imputaciones de los métodos de RF y RF híbrido, que son moderadamente más precisos en la mayoría de los casos siendo estos los que más predominan sobre los demás métodos que se utilizaron.

5.5 ANÁLISIS DE LOS RESULTADOS DE LAS ESTACIONES DE PRUEBA

5.5.1 ESTACIÓN EL CARMEN

En esta estación el desempeño total de los métodos evaluados han sido heterogéneos ,en este caso Rf_GPM se posiciona como el mejor método con un RMSE de 17.92 mm y un MAE de 7.1 mm, logrando además un balance favorable entre presión y sesgo con un PBIAS de -60.59% RF le sigue de manera muy próximo con PMSE de 18.01 mm demostrando que ambos métodos basados en Random Forest manteniendo un desempeño competitivo en esta ubicación., Al contrario a su comportamiento irregular que presenta en las demás estaciones ,MLR presenta resultados aceptables con un CC de 0.49 y un NSE de -0.18 aunque mantiene un sesgo muy bajo de -3.96%, lo que sugiere una estabilidad numérica local en esta estación ,mientras que SLR presenta la mayor magnitud de sesgo negativo con -73.56% lo cual demuestra las limitaciones que presenta la regresión lineal simple. GPM y RMSE son los más elevados junto con NSE más bajo, reflejando que los datos GPM crudos no son tan factibles de utilizar y el ultimo método de ponderación espacial basando en la altitud no proporciona una mejora significativa.

5.5.2 ESTACIÓN LA TEODOMIRA

La Teodmira se caracteriza como una estación muy problemática dónde el desempeño agregado de todos los métodos es limitado , igual que en el caso anterior Rf_GPM se destaca marginalmente con un CC de 0.35, RMSE de 7.48 mm y el MAE más bajo de 1.92 mm mitras que el NSE más elevado de 0.08 lo que esta métrica indica una captura mínima de variabilidad y RF es prácticamente equivalente con CC de 0.32 y RMSE de 7.64 mm lo más preocupante es que esta estación presenta un sesgo extremo prácticamente con todos los métodos.

5.5.3 ESTACIÓN ZAPOTE

Para esta estación RF demuestra superioridad con respecto a los demás métodos evaluados por que RF alcanza el CC más elevado de 0.61, el RMSE mínimo de 7.94 mm y el NSE óptimo de 0.29 entre todas las estaciones analizadas, demostrando un sesgo moderado y controlado de -60.01%. RF_GPM y MLR siguen con CC de 0.54 pero presentan NSE inferior de 0.2 y 0.22 respectivamente y RMSE más elevados con 8.45 y 8.32 mm, demostrando que en esta ubicación específica la combinación de predictores espaciales nativos de RF supera la incorporación de datos satelitales GPM mientras que

SLR con CC de 0.38 y RMSE de 9.14 mm mantiene desempeño secundario al otro , mientras que LWR, a pesar de alcanzar CC de 0.49, exhibe un NSE prácticamente nulo de 0.03 y sesgo altamente negativo -83.85% como en todos los demás casos , sugiriendo que el esquema de ponderación espacial es inadecuado para las características locales de esta estación.

5.5.4 ESTACIÓN JUNÍN

En este método los mejores métodos que se utiliza para esta estación es RF_GPM, MLR y RF, todos presentando correlaciones lineales mejores a otras siendo de 0.63, 0.67. RF_GPM se posiciona con ligera ventaja mediante CC de 0.67 y RMSE de 9.96 mm, aunque MLR exhibe el NSE más elevado de 0.38 junto con sesgo más favorable -46.32% se debe tomar en cuenta la aparente superioridad de MLR en esta estación debe interpretarse con cautela dado su comportamiento catastrófico e impredecible en otras estaciones. RF mantiene desempeño consistentemente confiable con NSE de 0.23 y RMSE de 9.9 mm, demostrando la estabilidad característica de este método. SLR y LWR resultan inaceptables para esta estación de estudio.

5.5.5 ESTACIÓN ATASCOSO (VÍA PICHINCHA)

A qui MLR experimenta su primer colapso crítico, manifestado mediante un RMSE fuera de rango siendo un caso extremo de 24.36 mm el más elevado en el conjunto de datos, PBias completamente invertido haciendo una sobre estimación con valor de 68.88%, y NSE de -2.73 , este comportamiento extremadamente anómalo de MLR sugiere problemas severos de inestabilidad numérica en la formulación de la matriz de regresión múltiple en esta ubicación específica al contrario , RF domina sin ambigüedad con CC de 0.69, NSE de 0.37, RMSE de 10.02 mm y sesgo moderado de -52.3%. RF_GPM es competitivo con CC de 0.62, aunque exhibe RMSE de 10.94 mm y NSE de 0.25 ligeramente inferiores. LWR presenta el segundo CC más elevado 0.68 pero un NSE bajo 0.18 y sesgo muy negativo -74.46%, ejemplificando el fenómeno donde correlación lineal no garantiza captura adecuada de variabilidad.

5.5.6 ESTACIÓN 24 DE MAYO (JABONCILLO)

Por ultimo también se puede evidenciar que en esta estación de estudio se utilizó RF que se posiciona como el método superior con CC de 0.48, NSE de 0.18 y RMSE de 8.57 mm, manteniendo además un sesgo moderado de -57.56% y seguido RF_GPM es competitivo con CC de 0.5 pero NSE de 0.09, indicando que aunque captura correlación

lineal, la explicación de variabilidad es limitada y como ya se ha evidenciado MLR es relativamente aceptable en esta estación con CC de 0.48, NSE de 0.2 y RMSE de 8.48 mm, aunque con sesgo bajo de -37.32%, lo que es atípico dado su comportamiento general.

Lo que nos da como resultado que Random Forest (RF) se posiciona como el método más consistentemente confiable y robusto, constituyéndose como el mejor o segundo mejor desempeño en 6 de las 7 estaciones analizadas, manteniendo además NSE positivo en todas las ubicaciones y demostrando superioridad en estabilidad numérica respecto a sus competidores donde presenta una tendencia sistemática hacia la subestimación, esta característica es predecible, cuantificable y controlable, diferenciándose cualitativamente de la inestabilidad impredecible manifestada por otros métodos y RF_GPM emerge como alternativa técnicamente válida que frecuentemente compite con RF manteniendo diferencias menores a 0.05 en métricas agregadas de desempeño, demostrando particular efectividad en JUNÍN y manteniéndose competitivo en ZAPOTE y CHONE, aunque su aplicabilidad se ve limitada por la dependencia de disponibilidad de datos satelitales GPM a si jugándose un papel muy impórtate para imputación de datos.

En la tabla 32 se puede observar cual método es el más preciso para el relleno de datos faltantes.

Tabla 32 Ranking General: Métodos de Mejor a Peor Desempeño

Ranking	Método
1°	RF
2°	RF_hybrid
3°	MLR
4°	SLR
5°	LWR
6°	GPM

5.5.7 ANÁLISIS AGREGADO DE DESEMPEÑO GENERAL

A nivel agregado, considerando el promedio de desempeño en todas las estaciones, RF mantiene su supremacía con CC de 0.54, NSE de 0.24, RMSE de 9.96 mm y sesgo moderado de -57.35%, consolidándose como el método más confiable y consistente

globalmente. RF_GPM es prácticamente equivalente con CC de 0.55, NSE de 0.18 y RMSE de 10.31 mm, aunque con menor sesgo (-70.37%), demostrando balance favorable entre precisión y control de sesgo como se puede observar en la Tabla 33

Tabla 33 Desempeño General

Método	Promedio				
	CC	NSE	MAE (mm)	RMSE (mm)	PBIAS (%)
SLR	0,40	0,06	3,71	11,07	-76,14
MLR	0,54	-1,46	4,83	15,58	8,33
LWR	0,45	0,05	3,76	11,15	-75,32
RF	0,55	0,24	3,39	9,96	-57,35
RF_GPM	0,55	0,18	3,45	10,31	-70,37
GPM	0,25	-0,37	5,37	13,36	-3,96

6 CONCLUSIONES Y RECOMENDACIONES

- La revisión bibliográfica permitió identificar y utilizar seis métodos para la estimación de datos faltantes de precipitación diaria, como Regresión Lineal Simple (SLR), Regresión Lineal Múltiple (MLR), Ponderación por la Distancia Inversa (LWR), Random Forest (RF), y (RF_GPM) y (GPM). La información obtenida evidenció que los métodos basados en aprendizaje automático, particularmente RF y sus variantes, han tenido un mejor rendimiento en la última década debido a su capacidad para capturar relaciones no lineales entre estaciones meteorológicas, superando en desempeño a los métodos estadísticos clásicos en múltiples estudios de caso a nivel regional e internacional.
- La aplicación de los seis métodos en las seis estaciones de estudio reveló que el método Random Forest (RF) y su variante RF_GPM presentaron el mejor desempeño en el relleno de datos faltantes de precipitación diaria. En términos del índice de eficiencia de Nash-Sutcliffe (NSE), RF alcanzó valores de categoría buena ($\geq 0,75$) en la estación ZAPOTE (NSE = 0,29) y ATASCOSO VÍA PICHINCHA (NSE = 0,37), mientras que RF_GPM obtuvo el mejor NSE individual en EL CARMEN (0,15), LA TEODOMIRA (0,08) y 24 DE MAYO-JABONCILLO (0,18).
- Los modelos de regresión, Lineal Simple, Múltiple y Ponderada Localmente demostraron ser los menos eficientes para la estimación de series temporales de precipitación diaria porque sus métricas arrojaron valores no aptos, evaluados por las métricas CC, NSE, MAE, RMSE y PBIAS. Evidenciando una subestimación severa y escasa capacidad predictiva, registrando los valores de PBIAS más negativos en la mayoría de estaciones de -83,66% en LA TEODOMIRA y -81,36% en 24 DE MAYO-JABONCILLO, NSE negativos de -0,03 en 24 DE MAYO-JABONCILLO y un MAE de hasta 7,46 mm en EL CARMEN. A pesar que los modelos de Random Forest tuvieron un mejor rendimiento ninguno de los métodos analizados es válido para la estimación de precipitación diaria.

- Se recomienda utilizar periodos más completos que contengan muchos más datos, para garantizar predicciones con mayor confiabilidad para así poder mejorar los resultados de las métricas.
- Se recomienda continuar investigando y evaluando nuevos métodos para la estimación de datos faltantes en series de precipitación diaria, tales como redes neuronales artificiales, modelos de aprendizaje profundo (LSTM, transformers) y métodos geoestadísticos como el Kriging, con el objetivo de identificar alternativas que superen el desempeño de los métodos evaluados en la presente investigación.
- Para futuras investigaciones, se recomienda ampliar el conjunto de datos de entrenamiento del modelo RF_GPM incorporando variables meteorológicas adicionales como temperatura, humedad relativa.

7 BIBLIOGRAFÍA

- [1] UNESCO, “Agua y Tecnologías de la Información,” *Capítulo 02*, 2021.
- [2] “Vista do TRMM SATELLITE PERFORMANCE IN ESTIMATED RAINFALL OVER THE MIDWEST REGION OF BRAZIL”.
- [3] C. Kidd *et al.*, “Intercomparison of high-resolution precipitation products over Northwest Europe,” *J. Hydrometeorol.*, vol. 13, no. 1, pp. 67–83, Feb. 2012, doi: 10.1175/JHM-D-11-042.1.
- [4] Ernesto “VISTA DO TRMM AND SATELLITE PERFORMANCE IN FALL OVER THE MIDWEST REGION OF BRAZIL”.2005
- [5] “J. D. Cryer y K. S. Chan, *Series Temporales: Aplicaciones con R*. Springer, 2018.
- [6] J. Asin, “Metodología para el control de calidad y homogeneidad de una base de datos de precipitación diaria.” [Online]. Available: <https://www.researchgate.net/publication/264352968>
- [7] A. S. Falck, V. Maggioni, J. Tomasella, D. A. Vila, and F. L. R. Diniz, “Propagation of satellite precipitation uncertainties through a distributed hydrologic model: A case study in the Tocantins–Araguaia basin in Brazil,” *J. Hydrol. (Amst.)*, vol. 527, pp. 943–957, Aug. 2015, doi: 10.1016/j.jhydrol.2015.05.042.
- [8] L. F. C de Oliveira, J. C. dos A Antonini, A. P. Fioreze, and M. A. S da Silva, “Métodos de estimativa de precipitação máxima para o Estado de Goiás,” 2008.
- [9] V. Costa, A. Silva, L. Palmier, and J. Sampaio, “Assessing the Propagation from Meteorological to Hydrological Drought in the São Francisco River Catchment with Standardized Indexes: Exploratory Analysis, Influential Factors, and Forecasting Strategies,” *J. Water Resour. Plan. Manag.*, vol. 147, p. 05021020, Aug. 2021, doi: 10.1061/(ASCE)WR.1943-5452.0001464.
- [10] M. M. Hasan and B. F. W. Croke, “Filling gaps in daily rainfall data: a statistical approach,” 2013. [Online]. Available: www.mssanz.org.au/modsim2013
- [11] S. Strachan *et al.*, “Filling the Data Gaps in Mountain Climate Observatories Through Advanced Technology, Refined Instrument Siting, and a Focus on Gradients,” in *Mountain Research and Development*, International Mountain Society, Nov. 2016, pp. 518–527. doi: 10.1659/MRD-JOURNAL-D-16-00028.1.
- [12] F. J. Blanco, “Análisis estadístico de datos hidrológicos,” *Art. Téc.*, vol. 1, 2020.
- [13] D. V. Carrera-Villacrés, P. V. Guevara-García, L. C. Tamayo-Bacacela, A. L. Balarezo-Aguilar, C. A. Narváez-Rivera, and D. R. Morocho-López, “Relleno de series anuales de datos meteorológicos mediante métodos estadísticos en la zona costera e interandina del Ecuador, y cálculo de la precipitación media,” *Idesia (Arica)*, vol. 34, no. 3, pp. 81–90, Jun. 2016, doi: 10.4067/S0718-34292016000300010.
- [14] “Estudio técnico de cuencas hidrográficas,” *Documento T02229*, Quito, Ecuador, 2019.

- [15] L. Brocca *et al.*, "Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data," *J. Geophys. Res.*, vol. 119, no. 9, pp. 5128–5141, May 2014, doi: 10.1002/2014JD021489.
- [16] "Incorporating Surface Soil Moisture Information in Error Modeling of TRMM Passive Microwave Rainfall | IEEE Journals & Magazine | IEEE Xplore." Accessed: Mar. 06, 2026. [Online]. Available: <https://ieeexplore.ieee.org/document/6729053>
- [17] C. Matovelle, D. Heras, and J. Solano-Peláez, "Imputation Efficiency of Missing Rainfall Data Using Computational Tools in a River Basin, Jubones-Ecuador," *Revista Politecnica*, vol. 50, no. 2, pp. 23–30, Nov. 2022, doi: 10.33333/rp.vol50n2.03.
- [18] E. Sharifi, R. Steinacker, and B. Saghafian, "Assessment of GPM-IMERG and other precipitation products against gauge data under different topographic and climatic conditions in Iran: Preliminary results," *Remote Sens. (Basel)*, vol. 8, no. 2, 2016, doi: 10.3390/rs8020135.
- [19] L. M. Espejo Chahuara, "Administración de Personal como Clave para la Mejora de la Gestión Empresarial en las Municipalidades Peruanas," *Ciencia Latina Revista Científica Multidisciplinar*, vol. 8, no. 5, pp. 11456–11485, Nov. 2024, doi: 10.37811/cl_rcm.v8i5.14537.
- [20] R. M. Granados, "Modelos de regresión lineal múltiple." Mar,2006
- [21] C. Paliz, N. Perugachi, J. Martínez, M. Moreno, C. Yaucán, and R. Palaguachi, "Análisis estadístico de datos de las precipitaciones usando métodos robustos y bootstrap," *FIGEMPA: Investigación y Desarrollo*, vol. 12, no. 2, pp. 52–61, Dec. 2021, doi: 10.29166/revfig.v12i2.3515.
- [22] E. S. Hernández Dávila, C. M. Gallegos Londoño, and F. A. García Mora, *Estadística descriptiva para el mantenimiento industrial con Python*. Centro de Investigación y Desarrollo Ecuador, 2023. doi: 10.33996/cide.ecuador.ep2636553.
- [23] Python logo Python, "The Python Standard Library » Concurrent Execution » multiprocessing."
- [24] A. M. Bello, J. Andrés Cuta, and E. K. García, "TÉCNICAS DE IMPUTACIÓN PARA DATOS DE PRECIPITACIÓN MÁXIMA MENSUAL EN LA ZONA CENTRAL DE BOYACÁ Imputation techniques applied in a maximum monthly precipitation data in the central zone of Boyacá," 2019.
- [25] L. V. Duarte, K. T. M. Formiga, and V. A. F. Costa, "Comparison of Methods for Filling Daily and Monthly Rainfall Missing Data: Statistical Models or Imputation of Satellite Retrievals?," *Water (Switzerland)*, vol. 14, no. 19, Oct. 2022, doi: 10.3390/w14193144.
- [26] C. Estadística, "ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO FACULTAD DE CIENCIAS," 2023.
- [27] Varshney, K. R., & Mojsilović, A. (2018). Artificial intelligence and machine learning for social impact: Opportunities and challenges. *IEEE Journal of Selected Topics in Signal Processing*, 12(1), 4-21. <https://doi.org/10.1109/JSTSP.2018.2821079>
- [28] D. Monserrate Ruiz Tubón Roberth Anderson Hoyos Yanangómez and I. Xiomara Alejandra Zambrano Navarrete, "UNIVERSIDAD TÉCNICA DE COTOPAXI FACULTAD DE

CIENCIAS DE LA INGENIERÍA Y APLICADAS CARRERA DE INGENIERÍA HIDRÁULICA
PROYECTO DE INVESTIGACIÓN PRESENTADO PREVIO A LA OBTENCIÓN DEL TÍTULO DE
INGENIERO HIDRÁULICO 'DESARROLLO DE UN MODELO DE PREDICCIÓN
METEOROLÓGICA PARA EVENTOS EXTREMOS.'"

- [29] M. Y. Climatología, "Meteorología y Climatología Semana de la Ciencia y la Tecnología 2004 FUNDACIÓN ESPAÑOLA PARA LA CIENCIA Y LA TECNOLOGÍA." [Online]. Available: www.fecyt.es
- [30] Kuk, A.Y.C. y Mak, T.K. (1989). Median estimation in the presence of auxiliary information. *Journal of the Royal Statistical Society, Series B*, 51 261-269.
- [31] Mateo Navarra, "Elementos y factores del clima. Tipos de climas," Gobierno de navarra. Accessed: Jan. 18, 2026. [Online]. Available: <https://meteo.navarra.es/definiciones/elementosfactores.cfm#top>
- [32] "Precipitaciones Concepto. Tipos." [Online]. Available: <http://hidrologia.usal.es>. Pág.1
- [33] Kalton, G. (1983). *Compensating for missing data*. Ann Arbor: Institute for Social Research, University of Michigan.
- [34] E.Maria, "UNIVERSIDAD POLITÉCNICA SALESIANA SEDE QUITO." Metodo de regresion lineal , 2009
- [35] L. Medina Marcelo Bacilio Paste Zambrano Jeison Paúl and I. Corrales Bastidas Byron Paúl MsC, "UNIVERSIDAD TÉCNICA DE COTOPAXI FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS CARRERA DE INGENIERÍA ELECTROMECÁNICA PROPUESTA TECNOLÓGICA PROPUESTA TECNOLÓGICA PREVIO A LA OBTENCIÓN DEL 'DISEÑO DE UNA ESTACIÓN METEOROLÓGICA INTELIGENTE PARA EL MONITOREO CONTINUO DE VARIABLES AMBIENTALES UTILIZANDO TECNOLOGÍAS DE INTERNET DE LAS COSAS (IoT).'"
- [36] C. Pacheco and R. Herrera, *Universidad técnica de cotopaxi*, vol. 1. 2021. [Online]. Available: <http://repositorio.utc.edu.ec/bitstream/27000/4501/1/PI-000727.pdf>
- [37] esri, "Modelos digitales de elevación—Conceptos relacionados," Apr. 2002. Accessed: Jan. 19, 2026. [Online]. Available: <https://learn.arcgis.com/es/related-concepts/digital-elevation-models.htm>
- [38] Sesión, "Introducción al Radar de Apertura Sintética (SAR) y sus Aplicaciones."
- [39] UTN, "Recursos Geoespaciales," Carrera de ingeniería forestal . Accessed: Jan. 19, 2026. [Online]. Available: https://observatorioforestal.utn.edu.ec/pages/resources_geo.html
- [40] "MANUAL".
- [41] N. Syakira *et al.*, "Assessment of Three GPM IMERG Products for GIS-Based Tropical Flood Hazard Mapping Using Analytical Hierarchy Process," *Water (Switzerland)*, vol. 15, no. 12, Jun. 2023, doi: 10.3390/w15122195.
- [42] S. J. Bajwa, "Basics, common errors and essentials of statistical tools and techniques in anesthesiology research," *J. Anaesthesiol. Clin. Pharmacol.*, vol. 31, no. 4, pp. 547–553, Oct. 2015, doi: 10.4103/0970-9185.169087.

- [43] R. G. Duque, "Python PARA TODOS." [Online]. Available: <http://mundogeek.net/tutorial-python/>
- [44] Ruderger "Libro de códigos y uso de pandas python". Librerías 2012
- [45] M. S. Aldo and R. V. Alvarado, "PYTHON CON ANACONDA Preparado como parte de las Actividades por el LII Aniversario de la Facultad de Ciencias Puras y Naturales," 2018.
- [46] Fernán García de Zúñiga, "¿Qué es Visual Studio Code y cuáles son sus ventajas?," arsys. Accessed: Jan. 19, 2026. [Online]. Available: <https://www.arsys.es/blog/que-es-visual-st>
- [47] Krause, A. y Olson, M. (2005). *The basic of S-Plus. Fourth Edition*. Springer.
- [48] "¿Qué es el Data Science?," 2022. [Online]. Available: <https://www.interactivechaos.com/python/function/hash>
- [49] A. F. Hidalgo Romero, "Comenzando con Numpy y Pandas," 2024.
- [50] C. Husillos Rodríguez, "Módulo 1 Numpy Curso de Python 3 Científico."
- [51] Little, R.J.A. y Rubin, D.B. (2002). *Statistical analysis with missing data. 2nd edition*.
- [52] Berger, Y.G. y Rao, J.N.K. (2006). Adjusted jackknife for imputation under unequal probability sampling without replacement. *Journal of the Royal Statistical Society, Series*
- [53] Valliant, R. (1993). Poststratification and conditional variance estimation. *Journal of the American Statistical Association,*
- [54] V. Alfaro, "Preparado por: Irene P." métricas de evaluación 2009
- [55] R. M. Granados, "Modelos de regresión lineal múltiple." 2007
- [56] "ANALISIS_DEL_PRECIO_DE_LOS_ALOJAMIENTOS_AIRBNB_DE_VA_Beltran_Verdu_Alejandro".
- [57] A. Liaw and M. Wiener, "Classification and Regression by randomForest," 2002. [Online]. Available: <http://www.stat.berkeley.edu/>
- [58] N. Turini, B. Thies, N. Horna, and J. Bendix, "Random forest-based rainfall retrieval for Ecuador using GOES-16 and IMERG-V06 data," *Eur. J. Remote Sens.*, vol. 54, no. 1, pp. 117–139, 2021, doi: 10.1080/22797254.2021.1884002.
- [59] E. Lahura, "EL COEFICIENTE DE CORRELACIÓN Y CORRELACIONES ESPÚREAS."
- [60] A. Ritter, R. Muñoz-Carpena, and C. M. Regalado, "Estudios en la Zona no Saturada del Suelo."
- [61] J. Cuadros-Adriazola, B. F. Ochoa-Tocachi, and V. Bonnesoeur, "2 CUBHIC 2.0 Calibración del modelo Autoría."
- [62] M. M. Maroneze, L. Q. Zepka, J. G. Vieira, M. I. Queiroz, and E. Jacob-Lopes, "A tecnologia de remoção de fósforo: Gerenciamento do elemento em resíduos industriais," *Revista Ambiente e Agua*, vol. 9, no. 3, pp. 445–458, 2014, doi: 10.4136/1980-993X.
- [63] C.R., K. Zhang, R. Lillywhite, C. Ramos, J. Doltra, J.M. de-Paz, H. Riley, M. Fink, C. Nendel, K. Thorup-Kristensen, A. Pedersen, F. Piro, A. Venezia, C. Firth, U. Schmutz, F. Rayns, and

K. Strohmeyer. 2010. Eu-Rotate_N - a decision support system - to predict environmental and economic consequences of the management of nitrogen fertiliser in crop rotations. *Europ. J. Hort. Sci.* 75(1):20-32.

- [64] A. Clara Zambrano Espinel *et al.*, "CONSEJO PROVINCIAL DE MANABÍ ING. MARIANO ZAMBRANO SEGOVIA PREFECTO DE LA PROVINCIA DE MANABÍ VICEPREFECTA DE LA PROVINCIA DE MANABÍ SR. ÁNGEL ROJAS CEVALLOS ING. RICARDO QUIJIJE ANCHUNDIA."
- [65] D. Vinicio Carrera-Villacrés, P. Valeria Guevara-García, L. Carolina Tamayo-Bacacela, A. Lucía Balarezo-Aguilar, C. Alfonso Narváez-Rivera, and D. Rosa Morocho-López, "Relleno de series anuales de datos meteorológicos mediante métodos estadísticos en la zona costera e interandina del Ecuador, y cálculo de la precipitación media Filling series annual meteorological data by statistical methods in the coastal zone from Ecuador and Andes, and calculation of rainfall," 2016.
- [66] D. T. Bolvin *et al.*, "Algorithm Theoretical Basis Document (ATBD) NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG) Version 07," 2023.
- [67] A. -Fernández, *Árboles de decisión en R con Random Forest*. 2019
- [68] P. Schober y L. A. Schwarte, "Correlation coefficients: Appropriate use and interpretation," *Anesthesia & Analgesia*, vol. 126, no. 5, 2018.

8 ANEXOS

Nombre	Fecha de modificación	Tipo	Tamaño
Escrip.py	28/10/2025 21:39	Carpeta de archivos	
Extacion_nf	5/11/2025 11:35	Carpeta de archivos	
GPM_IMERG_DAILY	14/12/2025 19:54	Carpeta de archivos	
prueba	26/10/2025 22:51	Carpeta de archivos	
zona_Estudio	25/1/2026 0:47	Carpeta de archivos	

Figura 80 Carpetas con los datos de las extracciones Satelitales

Disco local (T:) > FASE 1 > Escrip.py

Nombre	Fecha de modificación	Tipo	Tamaño
Aggregate_gpm_daily.py	2/10/2025 14:54	Archivo de origen ...	17 KB
Download_GPM.py	7/10/2025 21:00	Archivo de origen ...	17 KB
extract_station_precip.py	3/10/2025 10:40	Archivo de origen ...	14 KB
Spacial Script.py.rar	28/10/2025 15:56	WinRAR archive	15 KB

Figura 81 Escrip.py que se utilizaron para la descarga de datos y procesamiento

Disco local (T:) > FASE 2 > estaciones completas

Nombre	Fecha de modificación	Tipo	Tamaño
DATOS DE LAS ESTACIONES INAMI.xlsx	18/11/2025 18:44	Hoja de cálculo d...	27.283 KB

Figura 82 Datos entregados por el INAMHI

Disco local (T:) > TESIS PRACTICO TERMINADO

Nombre	Fecha de modificación	Tipo	Tamaño
data	14/12/2025 22:17	Carpeta de archivos	
docs	25/1/2026 20:07	Carpeta de archivos	
output	26/1/2026 0:01	Carpeta de archivos	
src	15/1/2026 11:35	Carpeta de archivos	
Python_Master.rar	25/1/2026 1:59	WinRAR archive	51.747 KB

Figura 83 Carpetas con los resultados obtenidos de Visual Code

Nombre	Fecha de modificación	Tipo	Tamaño
📁 _pycache_	25/1/2026 2:00	Carpeta de archivos	
📄 data_processing.py	14/12/2025 22:38	Archivo de origen ...	9 KB
📄 dem_processing.py	5/12/2025 19:23	Archivo de origen ...	7 KB
📄 evaluate.py	4/12/2025 21:52	Archivo de origen ...	12 KB
📄 homogeneity.py	14/12/2025 22:38	Archivo de origen ...	12 KB
📄 imputation_pipeline.py	23/1/2026 16:10	Archivo de origen ...	49 KB
📄 main.ipynb	24/1/2026 11:00	Archivo de origen ...	28.281 KB
📄 near_stn_search.py	15/10/2025 21:29	Archivo de origen ...	12 KB
📄 regression.py	23/1/2026 16:08	Archivo de origen ...	33 KB
📄 visualization.py	22/1/2026 22:59	Archivo de origen ...	49 KB
📄 weight_calculation.py	4/12/2025 20:35	Archivo de origen ...	4 KB

Figura 84 Escript.py Que se utilizaron para todo el procesamiento de datos y la obtención de resultados de los métodos y métricas

Tabla 34 evaluación métrica de todas las estaciones a escala diaria

Estacion	Metodo	CC	ME	RMSE	NSE	PBIAS
M0162	SLR	0,539	2,802	8,932	0,204	-60,168
M1208	SLR	0,234	2,077	7,848	-0,01	-83,66
M0452	SLR	0,381	2,873	9,141	0,06	-76,416
M0160	SLR	0,443	7,459	19,025	0,04	-73,562
M0462	SLR	0,541	3,47	10,963	0,054	-84,902
M0447	SLR	0,21	2,955	9,601	-0,025	-81,362
M1234	SLR	0,537	2,64	8,019	0,068	-82,152
M1220	SLR	0,455	4,31	11,978	0,098	-72,909
M1074	SLR	0,385	2,243	10,846	-0,023	-95,111
M1217	SLR	0,569	1,307	4,55	0,203	-73,78
M1251	SLR	0,68	2,528	7,074	0,323	-62,342
M1088	SLR	0,414	1,483	5,797	0,032	-89,365
M0163	SLR	0,275	3,587	14,619	-0,046	-96,887
M0450	SLR	0,484	1,76	8,475	0,015	-92,483
M0171	SLR	0,534	3,344	9,946	0,142	-72,384
M1233	SLR	0,187	1,148	4,697	-0,008	-68,045
M0453	SLR	0,334	1,556	7,744	0,013	-87,591
M0458	SLR	0,417	3,033	8,98	0,069	-73,232
M1205	SLR	0,15	3,26	11,096	-0,059	-89,594
M0451	SLR	0,296	2,814	8,09	-0,013	-84,488
M1230	SLR	0,483	2,502	8,296	0,055	-80,824
M1076	SLR	0,55	3,424	10,252	0,022	-89,075
M0589	SLR	0,274	3,258	11,862	-0,044	-93,539
M1078	SLR	0,425	1,56	6,106	0,01	-88,8
M0167	SLR	0,448	2,802	12,624	-0,019	-95,724
M0455	SLR	0,34	1,327	5,122	0,048	-80,581
M1089	SLR	0,402	5,422	13,892	0,031	-74,129
M1081	SLR	0,492	1,715	3,994	0,048	-73,022

M0448	SLR	0,25	0,781	3,518	-0,039	-97,365
M0166	SLR	0,513	5,097	14,734	0,095	-81,897
M0005	SLR	0,49	1,095	4,38	0,211	-61,353
M1080	SLR	0,505	3,988	11,214	0,083	-77,928
M0457	SLR	0,435	0,959	3,965	0,024	-88,084
M0464	SLR	0,573	3,316	9,728	0,115	-78,046
M0454	SLR	0,499	2,286	6,868	0,025	-85,936
M1091	SLR	0,644	5,883	17,416	0,258	-67,253
M1104	SLR	0,507	4,086	10,586	0,038	-77,997
M0165	SLR	0,784	2,081	9,059	0,24	-78,302
M0446	SLR	0,321	3,377	10,035	-0,077	-93,856
M1090	SLR	0,401	7,264	17,772	-0,024	-82,065
M0459	SLR	0,434	4,558	11,82	0,081	-71,949
M1092	SLR	0,267	2,201	7,652	-0,046	-94,278
M0449	SLR	0,256	1,446	5,419	-0,005	-85,891
M0169	SLR	0,058	2,223	16,346	-0,015	-95,172
M0162	MLR	0,682	7,926	30,341	-8,189	187,808
M1208	MLR	0,32	2,083	7,572	0,06	-50,339
M0452	MLR	0,538	2,605	8,315	0,222	-60,422
M0160	MLR	0,487	8,58	21,054	-0,176	-3,962
M0462	MLR	0,631	3,447	8,907	0,375	-46,317
M0447	MLR	0,483	2,596	8,483	0,2	-37,315
M1234	MLR	0,623	2,164	6,595	0,37	-39,224
M1220	MLR	0,604	6,542	24,362	-2,732	68,877
M1074	MLR	0,488	2,223	10,416	0,057	-84,83
M1217	MLR	0,575	1,202	4,217	0,315	-42,783
M1251	MLR	0,622	2,377	6,924	0,352	-32,902
M1088	MLR	0,584	1,291	5,063	0,261	-10,944
M0163	MLR	0,231	3,66	14,335	-0,006	-85,427
M0450	MLR	0,568	1,589	7,627	0,202	-73,965
M0171	MLR	0,501	3,193	9,506	0,217	-50,567
M1233	MLR	0,525	0,94	4,05	0,251	-56,25
M0453	MLR	0,722	1,469	6,485	0,308	-66,865
M0458	MLR	0,434	2,931	9,083	0,048	-21,871
M1205	MLR	0,427	3,992	13,05	-0,465	5,225
M0451	MLR	0,429	2,703	7,477	0,134	-59,337
M1230	MLR	0,685	1,986	6,627	0,397	-0,28
M1076	MLR	0,501	3,344	9,83	0,101	-78,318
M0589	MLR	0,454	3,241	11,405	0,035	-86,565
M1078	MLR	0,485	1,583	5,595	0,169	-69,33
M0167	MLR	0,525	2,703	10,844	0,248	-53,075
M0455	MLR	0,424	1,33	4,846	0,147	-43,54
M1089	MLR	0,54	5,699	13,262	0,117	-14,092
M1081	MLR	0,68	1,548	3,404	0,309	-65,845
M0448	MLR	0,324	0,788	3,452	0	-91,77

M0166	MLR	0,552	4,623	13,198	0,274	-47,383
M0005	MLR	0,534	1,279	4,634	0,117	-9,357
M1080	MLR	0,526	3,771	10,931	0,128	-19,277
M0457	MLR	0,344	0,998	3,998	0,007	-38,922
M0464	MLR	0,608	3,56	10,657	-0,062	11,416
M0454	MLR	0,595	2,05	5,879	0,286	-57,671
M1091	MLR	0,629	8,029	23,246	-0,321	20,419
M1104	MLR	0,532	3,923	9,478	0,229	-54,614
M0165	MLR	0,784	1,973	8,528	0,326	-69,326
M0446	MLR	0,396	3,624	9,977	-0,064	-20,327
M1090	MLR	0,291	7,877	19,415	-0,222	-34,268
M0459	MLR	0,529	4,517	11,242	0,169	-32,553
M1092	MLR	0,488	2,057	6,675	0,204	-54,884
M0449	MLR	0,567	1,252	4,575	0,284	-54,846
M0169	MLR	0,587	2,031	15,726	0,061	-86,072
M0162	LWR	0,406	2,94	9,551	0,089	-75,81
M1208	LWR	0,324	2,019	7,786	0,006	-84,654
M0452	LWR	0,485	2,853	9,285	0,03	-83,845
M0160	LWR	0,402	8,037	19,211	0,021	-36,33
M0462	LWR	0,403	3,556	11,181	0,016	-82,708
M0447	LWR	0,465	2,806	9,548	-0,014	-89,452
M1234	LWR	0,563	2,341	7,692	0,143	-74,843
M1220	LWR	0,676	4,108	11,457	0,175	-74,456
M1074	LWR	0,386	2,224	10,664	0,011	-86,626
M1217	LWR	0,595	1,341	4,961	0,053	-85,583
M1251	LWR	0,34	2,961	8,634	-0,008	-85,133
M1088	LWR	0,627	1,439	5,566	0,107	-78,967
M0163	LWR	0,341	3,603	14,298	-0,001	-87,894
M0450	LWR	0,696	1,724	8,367	0,04	-90,944
M0171	LWR	0,422	3,529	10,854	-0,021	-90,141
M1233	LWR	0,51	1,108	4,604	0,032	-88,788
M0453	LWR	0,346	1,536	7,7	0,024	-84,324
M0458	LWR	0,313	2,991	9,457	-0,032	-87,283
M1205	LWR	0,417	3,04	10,671	0,02	-82,997
M0451	LWR	0,371	2,814	7,902	0,033	-75,368
M1230	LWR	0,529	2,338	8,133	0,092	-79,527
M1076	LWR	0,445	3,489	10,319	0,009	-84,66
M0589	LWR	0,445	3,291	11,114	0,084	-73,993
M1078	LWR	0,321	1,584	6,067	0,023	-80,377
M0167	LWR	0,513	2,622	11,655	0,131	-77,182
M0455	LWR	0,507	1,29	4,946	0,112	-75,595
M1089	LWR	0,478	5,238	13,677	0,061	-75,754
M1081	LWR	0,601	1,637	4,01	0,041	-84,59
M0448	LWR	0,517	0,78	3,159	0,162	-65,05
M0166	LWR	0,507	5,274	15,762	-0,036	-90,412

M0005	LWR	0,411	1,233	4,821	0,044	-81,763
M1080	LWR	0,458	4,021	11,639	0,012	-81,756
M0457	LWR	0,458	0,967	3,891	0,06	-79,96
M0464	LWR	0,445	3,515	10,249	0,018	-84,326
M0454	LWR	0,663	2,21	6,566	0,109	-78,942
M1091	LWR	0,605	6,787	20,434	-0,021	-90,221
M1104	LWR	0,367	4,148	10,9	-0,02	-83,158
M0165	LWR	0,453	2,208	10,101	0,055	-81,268
M0446	LWR	0,373	3,228	9,631	0,008	-81,45
M1090	LWR	0,199	7,652	18,611	-0,123	-88,549
M0459	LWR	0,455	4,842	12,794	-0,076	-92,45
M1092	LWR	0,49	2,129	7,321	0,043	-82,885
M0449	LWR	0,552	1,327	5,205	0,073	-82,113
M0169	LWR	0,531	2,135	15,989	0,029	-88,256
M0162	RF	0,658	2,53	7,621	0,42	-26,187
M1208	RF	0,322	2,057	7,641	0,042	-74,101
M0452	RF	0,61	2,43	7,944	0,29	-60,008
M0160	RF	0,485	7,17	18,009	0,14	-60,677
M0462	RF	0,627	3,298	9,897	0,229	-70,623
M0447	RF	0,479	2,401	8,573	0,183	-57,562
M1234	RF	0,591	2,225	6,991	0,292	-58,446
M1220	RF	0,694	3,867	10,019	0,369	-52,302
M1074	RF	0,447	2,217	10,502	0,041	-85,748
M1217	RF	0,573	1,191	4,268	0,299	-52,091
M1251	RF	0,684	2,394	6,747	0,385	-53,37
M1088	RF	0,613	1,228	4,696	0,364	-35,835
M0163	RF	0,29	3,599	14,533	-0,034	-94,719
M0450	RF	0,575	1,623	7,859	0,153	-79,044
M0171	RF	0,523	3,215	9,471	0,222	-57,725
M1233	RF	0,583	0,865	4,1	0,232	-62,435
M0453	RF	0,47	1,524	7,415	0,095	-78,482
M0458	RF	0,531	2,538	8,144	0,235	-51,054
M1205	RF	0,446	2,919	9,816	0,171	-53,34
M0451	RF	0,411	2,751	7,733	0,074	-70,765
M1230	RF	0,692	1,834	6,214	0,47	-28,05
M1076	RF	0,416	3,467	10,185	0,035	-83,977
M0589	RF	0,448	3,219	11,348	0,045	-81,94
M1078	RF	0,329	1,568	5,914	0,071	-69,61
M0167	RF	0,476	2,688	12,073	0,068	-80,402
M0455	RF	0,245	1,397	5,221	0,01	-58,554
M1089	RF	0,5	5,032	12,805	0,177	-57,923
M1081	RF	0,609	1,62	3,468	0,283	-54,163
M0448	RF	0,226	0,785	3,468	-0,009	-91,258
M0166	RF	0,542	4,638	13,816	0,204	-63,53
M0005	RF	0,549	1,119	4,201	0,274	-25,39

M1080	RF	0,538	3,431	10,152	0,248	-51,274
M0457	RF	0,477	0,94	3,633	0,18	-63,738
M0464	RF	0,652	2,977	8,534	0,319	-58,913
M0454	RF	0,559	2,034	5,946	0,269	-56,314
M1091	RF	0,6	5,711	17,159	0,28	-54,787
M1104	RF	0,501	4,074	10,516	0,05	-73,781
M0165	RF	0,563	2,059	9,467	0,17	-71,662
M0446	RF	0,403	3,162	9,357	0,064	-71,867
M1090	RF	0,327	7,231	17,343	0,025	-63,075
M0459	RF	0,61	4,332	10,924	0,215	-64,787
M1092	RF	0,417	2,114	7,082	0,104	-69,189
M0449	RF	0,596	1,266	4,661	0,256	-64,59
M0169	RF	0,38	2,077	15,81	0,051	-80,359
M0162	RF_Hybrid	0,691	2,56	8,336	0,306	-66,437
M1208	RF_Hybrid	0,354	1,918	7,477	0,083	-69,474
M0452	RF_Hybrid	0,537	2,61	8,449	0,196	-68,871
M0160	RF_Hybrid	0,465	7,1	17,918	0,148	-60,588
M0462	RF_Hybrid	0,667	3,316	9,964	0,218	-74,198
M0447	RF_Hybrid	0,495	2,59	9,067	0,086	-80,682
M1234	RF_Hybrid	0,628	2,241	7,196	0,25	-67,054
M1220	RF_Hybrid	0,618	4,06	10,935	0,248	-72,347
M1074	RF_Hybrid	0,386	2,393	10,143	0,105	-64,004
M1217	RF_Hybrid	0,594	1,222	4,286	0,293	-51,405
M1251	RF_Hybrid	0,491	2,709	8,008	0,133	-77,83
M1088	RF_Hybrid	0,634	1,189	4,604	0,389	-32,82
M0163	RF_Hybrid	0,287	3,771	13,955	0,047	-70,962
M0450	RF_Hybrid	0,65	1,589	7,362	0,257	-65,879
M0171	RF_Hybrid	0,413	3,468	10,56	0,033	-84,701
M1233	RF_Hybrid	0,5	1,034	4,224	0,185	-67,045
M0453	RF_Hybrid	0,645	1,489	6,963	0,202	-67,836
M0458	RF_Hybrid	0,365	2,845	9,205	0,022	-83,064
M1205	RF_Hybrid	0,37	2,978	10,222	0,101	-60,55
M0451	RF_Hybrid	0,435	2,668	7,695	0,083	-74,546
M1230	RF_Hybrid	0,628	1,967	7,105	0,307	-59,602
M1076	RF_Hybrid	0,476	3,36	9,957	0,077	-79,989
M0589	RF_Hybrid	0,439	3,27	10,847	0,127	-69,412
M1078	RF_Hybrid	0,289	1,741	5,984	0,049	-50,182
M0167	RF_Hybrid	0,537	2,757	10,7	0,268	-43,592
M0455	RF_Hybrid	0,516	1,221	4,576	0,24	-53,441
M1089	RF_Hybrid	0,585	5,312	13,342	0,107	-79,998
M1081	RF_Hybrid	0,671	1,486	3,313	0,345	-68,215
M0448	RF_Hybrid	0,319	1,054	3,587	-0,08	-9,639
M0166	RF_Hybrid	0,623	5,117	15,143	0,044	-88,388
M0005	RF_Hybrid	0,501	1,083	4,373	0,214	-58,244
M1080	RF_Hybrid	0,492	3,671	11,013	0,115	-73,7

M0457	RF_Hybrid	0,429	1,008	3,648	0,174	-40,703
M0464	RF_Hybrid	0,65	3,19	9,371	0,179	-76,312
M0454	RF_Hybrid	0,694	1,96	5,672	0,335	-61,558
M1091	RF_Hybrid	0,598	6,36	19,317	0,088	-83,937
M1104	RF_Hybrid	0,447	3,997	10,359	0,078	-75,696
M0165	RF_Hybrid	0,645	1,982	8,714	0,297	-53,605
M0446	RF_Hybrid	0,426	3,198	8,974	0,139	-46,688
M1090	RF_Hybrid	0,299	7,292	18,196	-0,073	-86,282
M0459	RF_Hybrid	0,372	4,651	12,414	-0,013	-86,039
M1092	RF_Hybrid	0,491	2,079	6,729	0,191	-59,937
M0449	RF_Hybrid	0,633	1,184	4,46	0,319	-58,131
M0169	RF_Hybrid	0,57	1,976	15,366	0,103	-75,755
M0162	GPM	0,306	4,686	11,858	-0,326	11,866
M1208	GPM	0,185	3,399	9,851	-0,688	21,149
M0452	GPM	0,284	5,068	13,514	-0,265	-4,999
M0160	GPM	0,248	9,15	19,951	-0,174	-28,155
M0462	GPM	0,234	4,848	12,634	-0,347	-3,725
M0447	GPM	0,313	3,719	10,045	-0,221	-8,65
M1234	GPM	0,365	3,522	9,826	-0,185	-9,161
M1220	GPM	0,202	6,695	15,671	-0,558	-15,224
M1074	GPM	0,267	3,235	9,477	-0,758	27,798
M1217	GPM	0,286	3,087	8,919	-1,414	100,119
M1251	GPM	0,406	3,624	8,954	-0,093	-22,522
M1088	GPM	0,237	3,706	10,428	-1,385	91,158
M0163	GPM	0,22	4,894	13,522	-0,338	13,532
M0450	GPM	0,244	2,712	9,125	-0,801	74,277
M0171	GPM	0,345	4,135	11,171	0,003	-32,257
M1233	GPM	0,176	2,636	9,068	-1,589	63,619
M0453	GPM	0,237	2,968	9,022	-0,791	62,463
M0458	GPM	0,335	3,469	9,564	-0,134	-20,961
M1205	GPM	0,19	5,093	13,048	-0,587	22,453
M0451	GPM	0,276	3,166	8,396	-0,795	25,793
M1230	GPM	0,24	4,165	11,547	-0,588	27,629
M1076	GPM	0,248	5,515	13,438	-0,379	-6,75
M0589	GPM	0,219	4,33	12,221	-0,305	-2,086
M1078	GPM	0,149	3,3	10,461	-1,827	56,622
M0167	GPM	0,205	4,02	11,858	-0,924	82,696
M0455	GPM	0,231	2,871	8,735	-0,805	63,886
M1089	GPM	0,289	6,513	15,003	-0,451	-12,021
M1081	GPM	0,265	2,856	7,399	-2,692	38,579
M0448	GPM	0,182	2,641	8,388	-3,514	199,568
M0166	GPM	0,294	5,368	14,2	-0,088	-30,397
M0005	GPM	0,313	2,831	8,649	-0,844	60,262
M1080	GPM	0,255	5,79	14,417	-0,367	-1,33
M0457	GPM	0,261	2,55	7,962	-1,12	81,013

M0464	GPM	0,276	4,472	11,62	-0,315	-0,987
M0454	GPM	0,246	4,119	11,252	-0,592	29,563
M1091	GPM	0,33	7,135	18,835	0,038	-30,846
M1104	GPM	0,285	5,146	12,048	-0,532	-1,232
M0165	GPM	0,351	2,823	8,859	-0,854	78,926
M0446	GPM	0,237	4,219	10,817	-1,375	59,597
M1090	GPM	0,219	8,087	17,898	-0,236	-35,621
M0459	GPM	0,369	4,453	11,166	0,033	-46,999
M1092	GPM	0,237	3,653	9,958	-1,234	71,272
M0449	GPM	0,27	2,752	8,228	-1,231	79,27
M0169	GPM	0,25	3,24	14,222	-0,01	7,114