



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



FACULTAD DE
INGENIERÍA

Aplicación de Physics-Informed Neural Networks (PINNs) para la predicción de caudal diario en la cuenca del Santa Lucía

Informe de Proyecto de Grado presentado por

Leandro Pereira, Tomás Spoturno

en cumplimiento parcial de los requerimientos para la graduación de la carrera
de Ingeniería en Computación de Facultad de Ingeniería de la Universidad de
la República

Supervisores

Alberto Castro
Angela Gorgoglione, Federico Vilaseca

Montevideo, 28 de abril de 2026



Aplicación de Physics-Informed Neural Networks (PINNs) para la predicción de caudal diario en la cuenca del Santa Lucía por Leandro Pereira, Tomás Spoturno tiene licencia [CC Atribución - No Comercial - Sin Derivadas 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Agradecimientos

Queremos comenzar agradeciendo a la Facultad de Ingeniería y a la Universidad de la República por brindarnos una formación de calidad, que nos preparó para ejercer como ingenieros con responsabilidad y solidez técnica. También agradecemos los recursos académicos y humanos que nos permitieron llevar adelante este proyecto.

Agradecemos especialmente a nuestros tutores, Alberto Castro, Angela Gorgolione y Federico Vilaseca, por acompañarnos, guiarnos y brindarnos confianza durante todo este proceso. Su orientación, motivación y seguimiento fue fundamental para concretar este trabajo.

Expresamos nuestro más profundo agradecimiento a nuestras familias, por haber sido nuestro pilar durante todos estos años de carrera. Gracias por priorizar siempre nuestra formación, por brindarnos el tiempo, el espacio y el apoyo necesarios para estudiar, y por acompañarnos en cada etapa con interés genuino.

Su aliento constante fue fundamental para llegar hasta esta instancia. Finalmente, agradecemos al tribunal por tomarse el tiempo de leer este informe y evaluar nuestro trabajo.

Resumen

Este trabajo aborda la predicción de caudal diario en la subcuenca Florida — Puente Ruta 5 del sistema hidrográfico del río Santa Lucía (Uruguay), mediante la comparación e integración de modelos hidrológicos conceptuales y técnicas de aprendizaje automático. El objetivo principal fue evaluar si la incorporación de conocimiento físico dentro de arquitecturas neuronales mejora el desempeño predictivo respecto a enfoques puramente conceptuales o puramente basados en datos.

Se implementaron y compararon tres familias de modelos: (i) el modelo hidrológico conceptual determinista GR4J, basado en ecuaciones que representan almacenamiento y transferencia de agua en la cuenca; (ii) modelos neuronales puramente basados en datos, incluyendo un perceptrón multicapa y una red neuronal recurrente para series temporales; y (iii) modelos híbridos basados en Redes Neuronales Informadas por la Física (*Physics-Informed Neural Networks*), donde una red neuronal estima variables hidrometeorológicas que luego se integran dinámicamente en el modelo GR4J. En particular, se evaluaron variantes donde la red predice la precipitación, la evapotranspiración o ambas.

El desarrollo incluyó preprocesamiento de datos hidrometeorológicos (consolidación temporal, imputación de faltantes y normalización), diseño de representaciones temporales mediante retardos explícitos y ventanas deslizantes, optimización automática de hiperparámetros utilizando la herramienta Optuna, y análisis de relevancia de variables mediante la técnica de atribución *Integrated Gradients*. Los modelos fueron implementados en el lenguaje de programación Python, utilizando la biblioteca de aprendizaje profundo PyTorch para la construcción, entrenamiento y evaluación de las redes neuronales, junto con herramientas de análisis numérico para la calibración y evaluación del modelo conceptual.

La evaluación se realizó sobre un conjunto de prueba independiente, empleando métricas tradicionales de desempeño hidrológico, como la eficiencia de Nash–Sutcliffe y el porcentaje de sesgo (PBIAS), complementadas con métricas generales de error ampliamente utilizadas en modelado numérico, como el error absoluto medio y la raíz del error cuadrático medio. La configuración híbrida en la que una red recurrente estima la evapotranspiración integrada en GR4J obtuvo el mejor desempeño global, mejorando la eficiencia predictiva y reduciendo el sesgo volumétrico respecto al modelo conceptual calibrado.

El análisis mostró que las mejoras se deben a una corrección dinámica del

balance hídrico interno del modelo físico. No obstante, la ausencia de restricciones explícitas permitió la generación de valores físicamente inconsistentes en la evapotranspiración estimada, lo que evidencia un compromiso entre precisión predictiva y coherencia física. Asimismo, el enfoque híbrido presentó mayor costo computacional de entrenamiento.

En conjunto, el trabajo demuestra que la integración estructurada entre modelos físicos y aprendizaje profundo puede mejorar la predicción hidrológica, aunque requiere mecanismos adicionales para garantizar consistencia física y escalabilidad computacional.

Palabras clave: Redes Neuronales Informadas por la Física (PINNs), Predicción de Caudal, Modelación Hidrológica, GR4J, Perceptrón Multicapa (MLP), Redes Neuronales Recurrentes (RNN), LSTM, GRU, Aprendizaje Automático, Cuenca del Santa Lucía

Índice general

1. Introducción	1
1.1. Motivación y contexto	1
1.2. Planteo del problema	2
1.3. Objetivos	3
1.3.1. Objetivo general	3
1.3.2. Objetivos específicos	3
1.4. Enfoque metodológico y resultados esperados	4
1.5. Organización del documento	5
2. Revisión de antecedentes	7
2.1. Hidrología	8
2.1.1. Ciclo Hidrológico	8
2.1.2. Agregación espacial de la precipitación	8
2.1.3. Modelo Conceptual GR4J	11
2.2. Aprendizaje automático y redes neuronales en modelación hidrológica	15
2.2.1. Perceptrón multicapa (MLP)	15
2.2.2. Redes neuronales recurrentes (RNN)	16
2.3. <i>Physics-Informed Neural Networks</i> (PINNs)	17
2.3.1. Formulación general	17
2.3.2. Aplicaciones en hidrología y recursos hídricos	18
3. Caso de estudio	21
3.1. Descripción de la cuenca de estudio	21
3.2. Información disponible	23
3.2.1. Estación hidrométrica	23
3.2.2. Estaciones pluviométricas	24
3.2.3. Estación meteorológica	25
3.3. Relaciones espaciales entre estaciones	26
3.3.1. Distancia entre la estación hidrométrica y las estaciones pluviométricas	26
3.3.2. Delimitación de áreas de influencia: polígonos de Thiessen	27
3.4. Síntesis	29

4. Análisis de datos	31
4.1. Consolidación inicial del conjunto de datos	31
4.2. Imputación de datos	32
4.3. Análisis exploratorio de datos	33
4.3.1. Análisis descriptivo	33
4.3.2. Análisis correlativo	33
4.4. Normalización de Datos	34
5. Desarrollo y evaluación de modelos	35
5.1. Estrategia de entrenamiento y evaluación	36
5.1.1. Partición de datos	36
5.1.2. Métricas de evaluación del desempeño	36
5.1.3. Evaluación del rendimiento computacional	37
5.1.4. Análisis cualitativo del desempeño	37
5.1.5. Tratamiento de valores faltantes	38
5.2. Interpretabilidad basada en atribución	38
5.3. Optimización y AutoML	38
5.3.1. Calibración del modelo mediante NSGA-II	38
5.3.2. Optimización de hiperparámetros con Optuna	39
5.4. Infraestructura computacional	39
5.5. GR4J	39
5.6. MLP	41
5.6.1. Desarrollo del modelo	42
5.6.2. Incorporación de variables	42
5.6.3. Selección de variables	45
5.6.4. Optimización de hiperparámetros	47
5.7. RNN	49
5.7.1. Desarrollo del modelo	49
5.7.2. Incorporación de variables	51
5.7.3. Selección de variables	52
5.7.4. Optimización de hiperparámetros	53
5.8. PINNs	55
5.8.1. Formulación general	56
5.8.2. Variantes evaluadas	58
5.8.3. Arquitectura de red	59
5.8.4. Incorporación y selección de variables	60
5.8.5. Selección de variables	62
5.8.6. Análisis del balance entre términos de la función de pérdida	63
5.8.7. Optimización de hiperparámetros	64
5.8.8. Consideraciones metodológicas específicas de la PINN	65
6. Resultados	69
6.1. Aclaración de entorno de ejecución	69
6.1.1. Experimentos con GR4J, MLP y RNN	69
6.1.2. Experimentos con PINNs	69
6.2. Análisis correlativo de datos	70

6.2.1.	Estadísticas descriptivas del caudal	70
6.2.2.	Matriz de correlación	71
6.3.	GR4J	72
6.3.1.	Introducción	72
6.3.2.	Optimización de parámetros	72
6.3.3.	Evaluación del desempeño hidrológico	73
6.4.	MLP	75
6.4.1.	Desempeño del modelo base	76
6.4.2.	Síntesis del proceso de optimización	76
6.4.3.	Resultados del modelo optimizado	77
6.5.	RNN	80
6.5.1.	Desempeño del modelo base	80
6.5.2.	Síntesis del proceso de optimización	80
6.5.3.	Resultados del modelo optimizado	81
6.6.	PINN-P MLP	84
6.6.1.	Desempeño del modelo base	85
6.6.2.	Síntesis del proceso de optimización	85
6.6.3.	Resultados del modelo optimizado	86
6.7.	PINN-P RNN	89
6.7.1.	Desempeño del modelo base	89
6.7.2.	Síntesis del proceso de optimización	89
6.7.3.	Resultados del modelo optimizado	90
6.8.	PINN-E MLP	94
6.8.1.	Desempeño del modelo base	94
6.8.2.	Síntesis del proceso de optimización	94
6.8.3.	Resultados del modelo optimizado	95
6.9.	PINN-E RNN	98
6.9.1.	Desempeño del modelo base	98
6.9.2.	Síntesis del proceso de optimización	98
6.9.3.	Resultados del modelo optimizado	99
6.10.	PINN-PE MLP	102
6.10.1.	Desempeño del modelo base	102
6.10.2.	Síntesis del proceso de optimización	103
6.10.3.	Resultados del modelo optimizado	103
6.11.	PINN-PE RNN	105
6.11.1.	Desempeño del modelo base	107
6.11.2.	Síntesis del proceso de optimización	107
6.11.3.	Resultados del modelo optimizado	108
6.12.	Comparación	110
7.	Discusión	123
7.1.	Desempeño del modelo PINN E RNN frente a GR4J	123
7.2.	Análisis de la evapotranspiración predicha por la PINN	127

8. Ingeniería de Software	129
8.1. Planificación general	129
8.2. Diagrama de Gantt	129
8.3. Descripción de las etapas	129
8.4. Conclusión	131
9. Conclusiones y Trabajo Futuro	133
9.1. Conclusiones	133
9.2. Trabajo futuro	135
Referencias	137
A. Análisis e imputación de datos	141
A.1. Análisis de valores faltantes	141
A.2. Estrategia de imputación	143
A.2.1. Imputación de la precipitación	143
A.2.2. Imputación de variables meteorológicas (INIA Las Brujas)	146
A.2.3. Imputación del caudal	147
A.2.4. Evaluación de métodos alternativos de imputación	148
A.3. Resultados de la imputación	150
B. Optimización detallada de modelos	153
B.1. MLP	153
B.1.1. Modelo base	154
B.1.2. Evaluación de la incorporación de variables	155
B.1.3. Depuración inicial por correlación	160
B.1.4. Análisis de relevancia mediante <i>Integrated Gradients</i>	162
B.1.5. Optimización de hiperparámetros (Optuna)	166
B.1.6. Evaluación extendida del modelo optimizado	168
B.2. RNN	172
B.2.1. Modelo base	173
B.2.2. Evaluación de la incorporación de variables	174
B.2.3. Depuración inicial por correlación	177
B.2.4. Análisis de relevancia mediante <i>Integrated Gradients</i>	178
B.2.5. Optimización de hiperparámetros (Optuna)	182
B.2.6. Evaluación extendida del modelo optimizado	184
B.3. PINN-P MLP	188
B.3.1. Modelo base	189
B.3.2. Evaluación de la incorporación de variables	191
B.3.3. Depuración inicial por correlación	193
B.3.4. Análisis de relevancia mediante <i>Integrated Gradients</i>	195
B.3.5. Análisis del balance entre términos de la función de pérdida	197
B.3.6. Optimización de hiperparámetros (Optuna)	199
B.3.7. Evaluación extendida del modelo optimizado	200
B.4. PINN-P RNN	206
B.4.1. Modelo base	206

B.4.2.	Evaluación de la incorporación de variables	208
B.4.3.	Selección de la ventana temporal (<i>sequence length</i>)	209
B.4.4.	Depuración inicial por correlación	210
B.4.5.	Análisis de relevancia mediante <i>Integrated Gradients</i>	212
B.4.6.	Análisis del balance entre términos de la función de pérdida	214
B.4.7.	Optimización de hiperparámetros (Optuna)	216
B.4.8.	Evaluación extendida del modelo optimizado	219
B.5.	PINN-E MLP	224
B.5.1.	Modelo base	224
B.5.2.	Evaluación de la incorporación de variables	226
B.5.3.	Análisis de relevancia mediante <i>Integrated Gradients</i>	231
B.5.4.	Análisis del balance entre términos de la función de pérdida	235
B.5.5.	Optimización de hiperparámetros (Optuna)	236
B.5.6.	Evaluación extendida del modelo optimizado	238
B.6.	PINN-E RNN	243
B.6.1.	Modelo base	243
B.6.2.	Evaluación de la incorporación de variables	245
B.6.3.	Selección de la ventana temporal (<i>sequence length</i>)	247
B.6.4.	Depuración inicial por correlación	248
B.6.5.	Análisis de relevancia mediante <i>Integrated Gradients</i>	250
B.6.6.	Análisis del balance entre términos de la función de pérdida	252
B.6.7.	Optimización de hiperparámetros (Optuna)	253
B.6.8.	Evaluación extendida del modelo optimizado	256
B.7.	PINN-PE MLP	261
B.7.1.	Modelo base	261
B.7.2.	Evaluación de la incorporación de variables	263
B.7.3.	Depuración inicial por correlación	266
B.7.4.	Análisis de relevancia mediante <i>Integrated Gradients</i>	267
B.7.5.	Análisis del balance entre términos de la función de pérdida	270
B.7.6.	Optimización de hiperparámetros (Optuna)	271
B.7.7.	Evaluación extendida del modelo optimizado	274
B.8.	PINN-PE RNN	278
B.8.1.	Modelo base	279
B.8.2.	Evaluación de la incorporación de variables	280
B.8.3.	Depuración inicial por correlación	281
B.8.4.	Análisis de relevancia mediante <i>Integrated Gradients</i>	282
B.8.5.	Análisis del balance entre términos de la función de pérdida	283
B.8.6.	Optimización de hiperparámetros (Optuna)	284
B.8.7.	Evaluación extendida del modelo optimizado	285

Capítulo 1

Introducción

1.1. Motivación y contexto

La predicción de caudales en ríos y arroyos es un componente central para la gestión de recursos hídricos, ya que contribuye a la toma de decisiones en ámbitos como la planificación del abastecimiento de agua, la operación de infraestructuras hidráulicas, la gestión de eventos extremos (crecidas y sequías) y la evaluación de impactos asociados a la variabilidad climática. En este marco, disponer de modelos capaces de reproducir y anticipar la respuesta hidrológica de una cuenca a partir de información hidrometeorológica resulta de particular relevancia.

El sistema hidrográfico del río Santa Lucía constituye un caso de interés en Uruguay por su relevancia ambiental y de gestión. En particular, se ha estudiado la relación entre el uso/cobertura del suelo y variables asociadas a la dinámica del sistema fluvial, por ejemplo en términos de calidad de agua ([Gorgoglione y cols., 2020](#)). Este tipo de antecedentes refuerza la pertinencia de contar con herramientas de modelación que permitan comprender y predecir el comportamiento hidrológico del sistema bajo distintas condiciones.

Desde el punto de vista metodológico, la modelación hidrológica ha utilizado tradicionalmente modelos conceptuales basados en ecuaciones que representan de forma agregada procesos como almacenamiento, infiltración, generación de escorrentía y propagación ([Chow y cols., 1988](#)). Un ejemplo ampliamente empleado en simulación lluvia-caudal es el modelo GR4J, caracterizado por su parsimonia y desempeño competitivo en distintos contextos hidrológicos ([Perrin y cols., 2003](#)). Este tipo de modelos ofrece coherencia física e interpretabilidad, aunque puede verse limitado por simplificaciones estructurales, sensibilidad a la calibración y dificultades para representar relaciones no lineales complejas presentes en los datos.

En paralelo, el desarrollo de técnicas de aprendizaje automático impulsó el uso de redes neuronales para tareas hidrológicas. Trabajos pioneros mostraron el potencial de las redes neuronales artificiales para modelar el proceso lluvia-

caudal (Hsu y cols., 1995), y posteriormente se sistematizaron conceptos y aplicaciones en hidrología (ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000a, 2000b). Más recientemente, el *deep learning* se consolidó como una alternativa para capturar no linealidades y dependencias temporales complejas en disciplinas de hidrología y recursos hídricos (Tripathy y Mishra, 2024). En particular, las arquitecturas recurrentes y sus variantes (p. ej., LSTM) han sido utilizadas exitosamente en problemas de modelación lluvia-caudal (Elman, 1990; Bengio y cols., 1994; Hochreiter y Schmidhuber, 1997; Kratzert y cols., 2018).

Sin embargo, los enfoques puramente *data-driven* suelen entrenarse sin restricciones físicas explícitas, lo que puede derivar en predicciones físicamente inconsistentes, menor robustez fuera del rango observado y dificultades de interpretación del mecanismo hidrológico representado. En respuesta a estas limitaciones, las *Physics-Informed Neural Networks* (PINNs) proponen incorporar conocimiento físico durante el entrenamiento, integrando ecuaciones o regularizaciones basadas en principios físicos en la función de pérdida (Raissi y cols., 2019; Karniadakis y cols., 2021). Si bien su formulación se difundió inicialmente en problemas gobernados por ecuaciones diferenciales, existen antecedentes recientes de aplicación en problemáticas de hidrología e hidráulica (p. ej., flujo en suelos no saturados y modelos de onda difusiva) (Bandai y Ghezzehei, 2022; Hou y cols., 2024). En este escenario, resulta pertinente estudiar enfoques híbridos que combinen la flexibilidad de las redes neuronales con estructuras conceptuales hidrológicas establecidas.

1.2. Planteo del problema

Esta tesis aborda el problema de la **predicción de caudal** en un punto de control a partir de información hidrometeorológica disponible, evaluando comparativamente tres enfoques con distinto grado de incorporación de física:

- un modelo conceptual determinista (GR4J) (Perrin y cols., 2003);
- modelos neuronales puramente *data-driven* (MLP y RNN) (Hsu y cols., 1995; ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000a; Kratzert y cols., 2018);
- modelos híbridos basados en PINNs, que integran restricciones físicas durante el entrenamiento (Raissi y cols., 2019; Karniadakis y cols., 2021).

El análisis se focaliza en la subcuenca **Florida – Puente Ruta 5**, delimitada dentro del sistema hidrográfico del río Santa Lucía. Este recorte espacial responde a un criterio metodológico: estudiar una subcuenca delimitada permite reducir la complejidad del problema hidrológico, concentrando el análisis en un único punto de salida de caudal, correspondiente a la estación **DINAGUA Florida – Puente Ruta 5**. Este enfoque facilita la validación inicial de modelos PINN, ya que reduce fuentes adicionales de incertidumbre asociadas a la

heterogeneidad espacial de la cuenca completa, permitiendo evaluar el desempeño en un dominio acotado y relativamente bien caracterizado, con potencial de escalamiento posterior a ámbitos de mayor complejidad.

En términos operativos, la investigación busca responder las siguientes preguntas:

- **¿En qué medida un modelo PINN basado en GR4J puede mejorar (o al menos igualar) el desempeño predictivo** frente a un enfoque puramente determinista (GR4J) y frente a modelos puramente neuronales (MLP/RNN)?
- **¿Qué compromisos (*trade-offs*) introduce el enfoque PINN** en términos de precisión, estabilidad de entrenamiento, interpretabilidad y costo computacional?

La evaluación del desempeño se apoya en métricas ampliamente utilizadas en hidrología para simulación y predicción de caudales, como la eficiencia de Nash–Sutcliffe (NSE) (Nash y Sutcliffe, 1970), considerando además lineamientos de buenas prácticas para la evaluación de modelos en simulaciones de cuencas (Moriassi y cols., 2007).

1.3. Objetivos

1.3.1. Objetivo general

Evaluar el desempeño de modelos *Physics-Informed Neural Networks* (PINNs) que incorporan la formulación física del modelo GR4J para la predicción del caudal diario en la subcuenca Florida – Puente Ruta 5, y compararlo con el de un modelo conceptual determinista (GR4J) y modelos puramente *data-driven* (MLP y RNN).

1.3.2. Objetivos específicos

- Cuantificar el desempeño predictivo del modelo hidrológico GR4J calibrado como línea base para la estimación de caudales en la subcuenca de estudio.
- Cuantificar el desempeño de modelos de aprendizaje automático tipo **MLP** y **RNN** en la predicción directa de caudal a partir de variables hidrometeorológicas de entrada.
- Evaluar el impacto de la incorporación de información física mediante arquitecturas **PINN** basadas en GR4J sobre la precisión en la predicción de caudales.
- Comparar distintos esquemas de acoplamiento en las arquitecturas PINN, donde la red neuronal estime (i) precipitación, (ii) evapotranspiración o (iii) ambas variables, en términos de desempeño hidrológico.

- Analizar la influencia de la representación temporal y de la incorporación de variables temporales en la capacidad predictiva de los modelos desarrollados.
- Determinar la relevancia de las variables de entrada mediante técnicas de interpretabilidad (p. ej., *Integrated Gradients*) (Sundararajan y cols., 2017), con el fin de sustentar decisiones de selección o descarte de *features*.
- Evaluar la sensibilidad del desempeño de las redes neuronales frente a la selección de hiperparámetros.
- Comparar los modelos propuestos en términos de desempeño predictivo y costo computacional utilizando un conjunto de prueba independiente, mantenido oculto durante la etapa de desarrollo.

1.4. Enfoque metodológico y resultados esperados

La tesis adopta un enfoque comparativo entre tres familias de modelos:

1. **Modelo conceptual determinista (GR4J):** utilizado como referencia basada en ecuaciones y parámetros calibrables.
2. **Modelos neuronales *data-driven* (MLP y RNN):** entrenados para predecir el caudal directamente, capturando relaciones no lineales y dependencias temporales.
3. **Modelos híbridos PINN: *data-driven* + GR4J:** variantes donde una red neuronal estima variables de forzante (precipitación y/o evapotranspiración) y dichas estimaciones se propagan a través de GR4J para obtener el caudal final. En los casos en que se predice solo una de las variables, la restante se toma desde el valor observado.

La Figura 1.1 presenta un esquema general de la metodología adoptada.

Para garantizar comparabilidad, se diseña un conjunto de pruebas replicable por arquitectura, que contempla decisiones de representación temporal (*lags* explícitos en MLP y ventanas temporales en RNN), incorporación de variables temporales, variantes de construcción de entrada (p. ej., precipitación agregada vs. por estación), análisis de relevancia/interpretabilidad por atribución (Sundararajan y cols., 2017) y optimización de hiperparámetros mediante búsqueda automatizada (Akiba y cols., 2019). Finalmente, el desempeño se evalúa en un conjunto de test separado y preservado durante el desarrollo.

En cuanto a **resultados esperados**, se plantea que la incorporación de estructura física mediante PINNs podría:

- mejorar la consistencia física de las predicciones y/o estabilizar el entrenamiento respecto de modelos puramente *data-driven*;

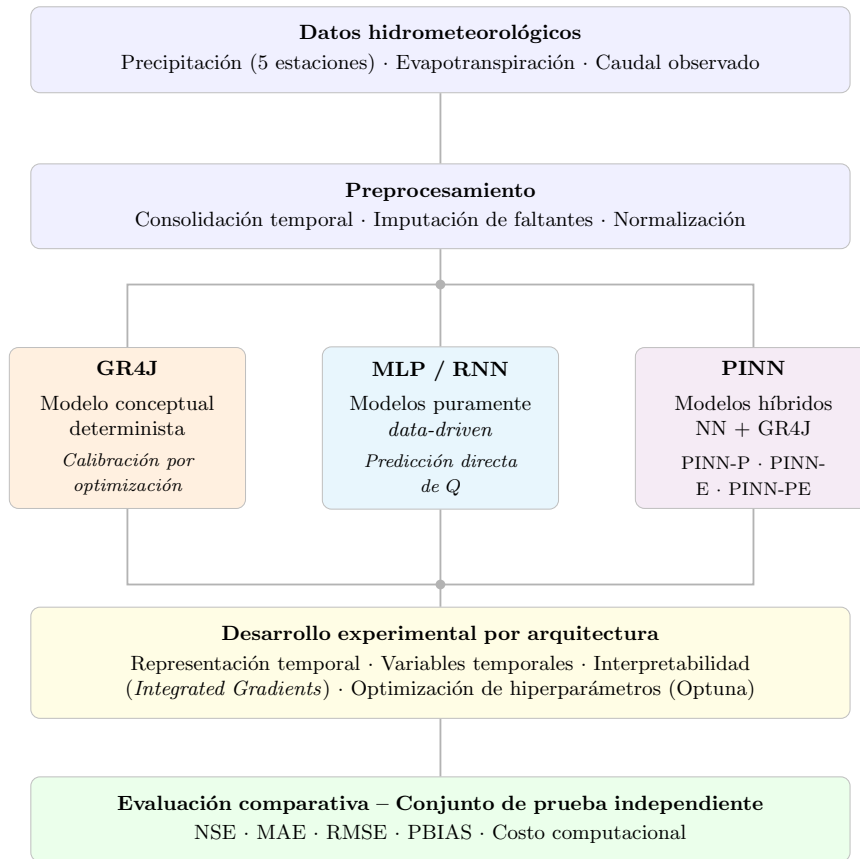


Figura 1.1: Esquema general de la metodología adoptada.

- aportar mayor robustez ante cambios de régimen hidrológico (dentro de los límites del caso de estudio);
- ofrecer un compromiso entre interpretabilidad (vía GR4J) y flexibilidad (vía red neuronal).

Al mismo tiempo, se espera identificar limitaciones y *trade-offs* del enfoque PINN, incluyendo complejidad metodológica, sensibilidad al diseño de la función de pérdida informada por la física y potencial incremento del costo computacional.

1.5. Organización del documento

El resto del documento se estructura de la siguiente manera:

- El **Capítulo 2** presenta la revisión de antecedentes, incluyendo funda-

mentos hidrológicos, el modelo conceptual GR4J, enfoques de aprendizaje automático para modelación hidrológica (MLP y RNN) y el marco de *Physics-Informed Neural Networks* (PINNs).

- El **Capítulo 3** describe el caso de estudio, detallando la cuenca analizada, las estaciones disponibles y las relaciones espaciales entre ellas.
- El **Capítulo 4** aborda el análisis y preprocesamiento de datos, incluyendo consolidación, imputación, análisis exploratorio y normalización.
- El **Capítulo 5** presenta el desarrollo y evaluación de los modelos, incluyendo la estrategia de entrenamiento, métricas de desempeño, optimización y configuraciones evaluadas.
- El **Capítulo 6** expone los resultados experimentales obtenidos para los distintos enfoques y sus respectivas variantes.
- El **Capítulo 7** desarrolla la discusión de los resultados, incluyendo el análisis comparativo entre modelos y la interpretación hidrológica de los hallazgos.
- El **Capítulo 8** resume los aspectos de ingeniería de *software* asociados al proyecto.
- Finalmente, el **Capítulo 9** presenta las conclusiones del trabajo y las líneas de investigación futura.

Capítulo 2

Revisión de antecedentes

En este capítulo se presentan los antecedentes teóricos que enmarcan el desarrollo del presente proyecto de grado, articulando conceptos de la hidrología con enfoques de modelación lluvia-escorrentía y técnicas modernas de aprendizaje automático. El objetivo es establecer el marco conceptual necesario para comprender tanto los procesos físicos que gobiernan la respuesta hidrológica de una cuenca como las estrategias de representación y predicción utilizadas en el trabajo.

En primer lugar, se describe la base hidrológica del estudio a partir del ciclo hidrológico y la noción de cuenca como sistema, introduciendo las principales entradas, salidas y procesos de almacenamiento. Dado que el análisis se aborda a escala de cuenca con modelos *lumped*, se incluye además una revisión de métodos clásicos de agregación espacial de la precipitación orientados a construir series representativas de precipitación media areal.

A continuación, se presenta el modelo conceptual GR4J como referencia consolidada para la simulación continua de caudales diarios, detallando su estructura general y su interpretación hidrológica. En paralelo, se revisan enfoques *data-driven* basados en redes neuronales, introduciendo arquitecturas *feed-forward* (MLP) y recurrentes (RNN), con énfasis en su capacidad para capturar relaciones no lineales y dependencias temporales en series hidrometeorológicas.

Finalmente, se introduce el marco de las *Physics-Informed Neural Networks* (PINNs) como una línea de trabajo que busca integrar información física dentro del proceso de aprendizaje, discutiendo su formulación general y algunas aplicaciones recientes en hidrología y recursos hídricos. En conjunto, estos antecedentes permiten contextualizar las decisiones metodológicas del proyecto y fundamentar el enfoque adoptado para el problema de modelación hidrológica abordado.

2.1. Hidrología

2.1.1. Ciclo Hidrológico

El ciclo hidrológico comprende el conjunto de fenómenos mediante los cuales el agua circula entre la atmósfera, la superficie terrestre y el subsuelo. Según [Chow y cols. \(1988\)](#), la hidrología se ocupa de la ocurrencia, circulación y distribución del agua sobre y bajo la superficie de la Tierra, así como de sus propiedades físicas y químicas y su relación con el medio ambiente. Este proceso constituye la base del **ciclo hidrológico**, el cual describe el movimiento continuo del agua dentro del sistema Tierra-atmósfera.

El ciclo hidrológico incluye la evaporación desde océanos y superficies continentales, el transporte del vapor de agua por la atmósfera, la condensación y formación de nubes, la precipitación, la infiltración en el suelo, el escurrimiento superficial y subterráneo, y finalmente el retorno del agua al océano. Este ciclo no tiene un punto inicial ni final definidos, sino que constituye una circulación continua de agua y energía. A escala planetaria puede considerarse un sistema cerrado, en el cual la cantidad total de agua permanece prácticamente constante; sin embargo, a escala local o de cuenca se comporta como un sistema abierto, con intercambios de masa y energía con su entorno.

De acuerdo con [Chow y cols. \(1988\)](#), el ciclo hidrológico puede interpretarse como un **sistema de flujo** en el que el agua es el elemento que se mueve bajo la influencia de la energía solar y de la gravedad. En este sistema, la *precipitación* actúa como la principal entrada de agua, mientras que la *escorrentía*, la *evapotranspiración* y los cambios de almacenamiento en el suelo y en los acuíferos constituyen sus salidas y transformaciones internas. El autor introduce el concepto de **sistema hidrológico** para representar de manera abstracta el comportamiento de una cuenca hidrográfica. La cuenca se concibe como una unidad con límites definidos que recibe una entrada de agua en forma de precipitación y genera una salida en forma de escorrentía, al tiempo que almacena y transfiere agua a través de procesos intermedios. Este enfoque permite analizar la respuesta de la cuenca ante un estímulo externo, como un evento de lluvia, y constituye la base conceptual para la modelación hidrológica.

2.1.2. Agregación espacial de la precipitación

En estudios hidrológicos a escala de cuenca, la precipitación suele medirse en estaciones puntuales distribuidas de manera irregular en el territorio. Sin embargo, modelos conceptuales *lumped* como GR4J requieren una única serie temporal representativa de la precipitación sobre toda el área de la cuenca. Por ello, es necesario estimar la **precipitación media areal** (PMA) a partir de las observaciones puntuales disponibles ([Chow y cols., 1988](#)).

Sea un conjunto de N estaciones con precipitación medida $P_i(t)$ en el instante (o día) t . De manera general, la PMA puede expresarse como una combinación ponderada:

$$\bar{P}(t) = \sum_{i=1}^N w_i P_i(t), \quad \sum_{i=1}^N w_i = 1, \quad w_i \geq 0,$$

donde los pesos w_i dependen del método de agregación espacial empleado. A continuación se describen tres métodos clásicos utilizados en hidrología: el promedio aritmético simple, los polígonos de Thiessen y la ponderación por distancia inversa.

Promedio aritmético simple

El método más sencillo consiste en calcular el **promedio aritmético** de las estaciones disponibles (Chow y cols., 1988). En este caso, se asigna el mismo peso a cada estación:

$$w_i = \frac{1}{N}, \quad i = 1, \dots, N,$$

por lo que la precipitación media areal se obtiene como:

$$\bar{P}(t) = \frac{1}{N} \sum_{i=1}^N P_i(t).$$

Este método supone implícitamente que la red de estaciones está distribuida de manera aproximadamente uniforme y que la variabilidad espacial de la precipitación dentro de la cuenca es reducida. Su principal ventaja es la simplicidad y el bajo costo computacional. Sin embargo, cuando la distribución de estaciones es irregular puede introducir sesgos al sobrerrepresentar zonas con mayor densidad de pluviómetros.

Polígonos de Thiessen

El método de **polígonos de Thiessen** fue propuesto por Thiessen (Thiessen, 1911) para estimar promedios de precipitación sobre grandes áreas a partir de observaciones puntuales. El procedimiento consiste en subdividir el área de estudio en regiones de influencia asociadas a cada estación, de modo que cualquier punto dentro de una región se encuentre más próximo a su estación correspondiente que a cualquier otra.

Geométricamente, los polígonos se construyen trazando las mediatrices de los segmentos que unen estaciones vecinas. La intersección de estas mediatrices define polígonos que delimitan el área de influencia de cada pluviómetro. Una vez recortados por el límite de la cuenca, se obtiene un área efectiva A_i asociada a la estación i .

Denotando por A el área total de la cuenca, los pesos se definen como:

$$w_i = \frac{A_i}{A},$$

y la precipitación media areal resulta:

$$\bar{P}(t) = \sum_{i=1}^N \left(\frac{A_i}{A} \right) P_i(t).$$

El método asigna pesos fijos basados únicamente en la geometría y no requiere supuestos adicionales sobre la estructura espacial de la precipitación. No obstante, asume que la precipitación es uniforme dentro de cada polígono y presenta discontinuidades en los límites entre regiones de influencia.

Ponderación por distancia inversa (IDW)

La **ponderación por distancia inversa** (*Inverse Distance Weighting*, IDW), también conocida como método de Shepard, fue propuesta por Shepard (Shepard, 1968). Se trata de un método determinístico de interpolación espacial basado en el supuesto de que observaciones más cercanas tienen mayor influencia que observaciones más lejanas.

Para estimar la precipitación en una ubicación \mathbf{x} dentro de la cuenca, se define:

$$\hat{P}(\mathbf{x}, t) = \frac{\sum_{i=1}^N w_i(\mathbf{x}) P_i(t)}{\sum_{i=1}^N w_i(\mathbf{x})}, \quad w_i(\mathbf{x}) = \frac{1}{d(\mathbf{x}, \mathbf{x}_i)^p},$$

donde \mathbf{x}_i es la ubicación de la estación i , $d(\cdot, \cdot)$ representa una distancia (típicamente euclídea) y $p > 0$ es un parámetro que controla la tasa de decaimiento de la influencia con la distancia.

Para obtener la precipitación media areal, puede interpolarse $\hat{P}(\mathbf{x}, t)$ en una grilla de M puntos dentro de la cuenca y luego promediar espacialmente:

$$\bar{P}(t) \approx \frac{1}{M} \sum_{j=1}^M \hat{P}(\mathbf{x}_j, t).$$

A diferencia del método de Thiessen, IDW produce un campo interpolado continuo y permite controlar el grado de suavizado mediante el parámetro p . Sin embargo, su desempeño depende de la densidad y distribución de estaciones, así como de la elección de sus parámetros.

Consideraciones comparativas

El promedio aritmético simple constituye el enfoque más básico y puede ser adecuado en cuencas pequeñas o con redes densas y homogéneas. El método de Thiessen introduce una corrección geométrica basada en el área de influencia de cada estación, mientras que IDW incorpora explícitamente la distancia como criterio de ponderación y genera una representación espacial continua.

La elección del método depende de la densidad de estaciones, la extensión de la cuenca y la variabilidad espacial esperada de la precipitación. En modelos *lumped*, la estimación de la precipitación media areal constituye un paso fundamental para asegurar coherencia entre las entradas meteorológicas y la representación conceptual del sistema hidrológico.

2.1.3. Modelo Conceptual GR4J

El modelo GR4J (*Génie Rural à 4 paramètres Journalier*) es un modelo conceptual de transformación lluvia–escorrentía desarrollado por Perrin, Michel y Andréassian (Perrin y cols., 2003), orientado a la simulación continua de caudales diarios en cuencas hidrográficas. Su diseño persigue un equilibrio entre simplicidad estructural y capacidad de representación hidrológica, utilizando únicamente cuatro parámetros calibrables que permiten capturar los procesos dominantes del ciclo hidrológico a escala de cuenca.

GR4J es un modelo *lumped*, por lo que asume que las variables de entrada (precipitación y evapotranspiración potencial) son espacialmente uniformes sobre la cuenca. A partir de ellas, el modelo representa el funcionamiento hidrológico mediante reservorios conceptuales, funciones de producción y enrutamiento, manteniendo a su vez un esquema de balance hídrico interno (Perrin y cols., 2003).

Parámetros del modelo

El modelo se define a partir de cuatro parámetros principales (Perrin y cols., 2003):

- x_1 : Capacidad del almacén de producción (mm).
- x_2 : Parámetro de intercambio subterráneo (mm).
- x_3 : Capacidad del almacén de enrutamiento (mm).
- x_4 : Tiempo de enrutamiento (días).

Estos parámetros controlan la retención de agua en el suelo, los procesos de intercambio con el entorno, y la dinámica del flujo retardado y directo hacia la salida de la cuenca.

Estructura general del modelo

El modelo puede dividirse conceptualmente en dos componentes principales (Perrin y cols., 2003), tal como se muestra en la Figura 2.1:

1. **Módulo de producción:** Representa el almacenamiento de agua en el suelo, la evapotranspiración y la generación del agua efectiva disponible para escorrentía.
2. **Módulo de enrutamiento:** Controla la transformación del agua generada en caudal mediante reservorios y funciones de propagación.

Ambos módulos se encuentran acoplados mediante el balance hídrico diario.

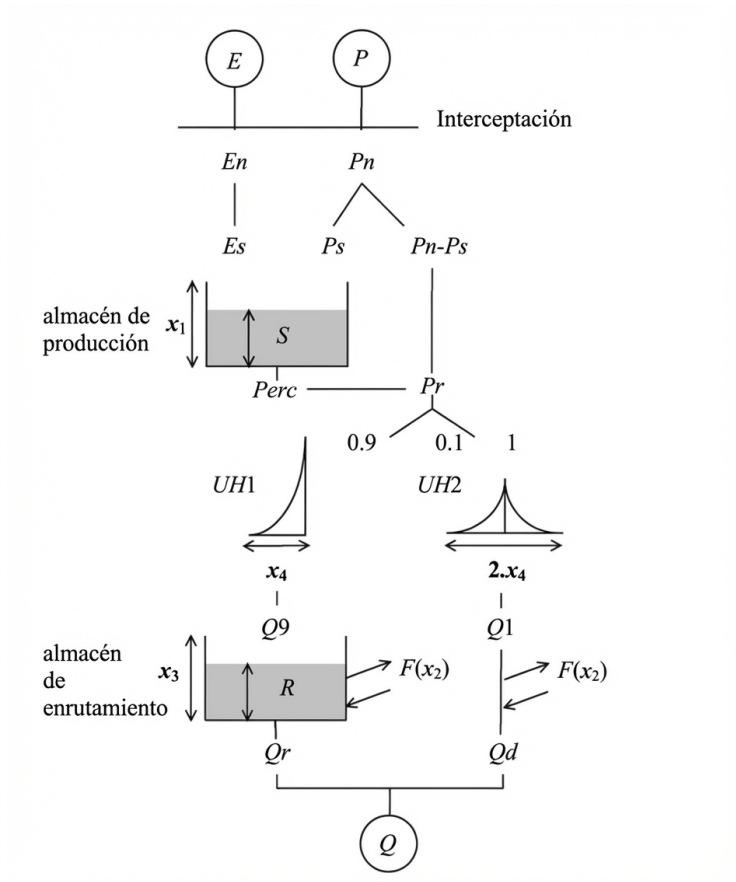


Figura 2.1: Diagrama del modelo GR4J

Módulo de producción

Sea P la precipitación diaria y E la evapotranspiración potencial, representadas en la parte superior del esquema (Figura 2.1), el modelo actúa sobre un almacén de capacidad x_1 cuyo estado en el día t se denota como S_t , correspondiente al almacén de producción S en la figura (Perrin y cols., 2003).

Cuando $P > E$, existe precipitación neta efectiva:

$$P_n = P - E$$

tal como se indica en el esquema de la Figura 2.1. Una fracción de esta precipitación se almacena en el reservorio de suelo (P_s en la figura), mientras que el excedente contribuye a la generación de escorrentía. La fracción que ingresa al almacén se determina mediante una función de saturación no lineal:

$$P_s = \frac{x_1 \left(1 - \left(\frac{S_t}{x_1} \right)^2 \right) \tanh\left(\frac{P_n}{x_1}\right)}{1 + \frac{S_t}{x_1} \tanh\left(\frac{P_n}{x_1}\right)}$$

y el almacenamiento se actualiza como:

$$S_{t+1} = S_t + P_s$$

Adicionalmente, el almacén de producción genera un flujo de percolación ($Perc$ en la Figura 2.1), que representa la transferencia de agua hacia el módulo de enrutamiento.

La precipitación efectiva generada $P_n - P_s$, junto con el flujo de percolación ($Perc$), conforma el flujo Pr que alimenta el módulo de enrutamiento, tal como se muestra en la Figura 2.1.

En caso contrario, cuando $E > P$, el modelo retira agua del almacén hasta donde sea posible:

$$E_n = E - P$$

La evapotranspiración real (E_s en la Figura 2.1) se determina mediante una expresión análoga:

$$E_s = \frac{S_t \left(2 - \frac{S_t}{x_1} \right) \tanh\left(\frac{E_n}{x_1}\right)}{1 + \left(1 - \frac{S_t}{x_1} \right) \tanh\left(\frac{E_n}{x_1}\right)}$$

y el almacenamiento actualizado resulta:

$$S_{t+1} = S_t - E_s$$

Módulo de enrutamiento y generación de caudal

El volumen de agua producido ($P_n - P_s$ en la Figura 2.1) se divide en dos componentes: una fracción que atraviesa un reservorio de enrutamiento y otra que se transmite más rápidamente, representando flujos rápidos asociados a eventos de lluvia intensa, tal como se observa en el esquema.

GR4J utiliza dos hidrogramas unitarios discretizados de diferente escala temporal (denominados UH1 y UH2 en la Figura 2.1), lo cual permite representar simultáneamente la respuesta rápida y retardada de la cuenca (Perrin y cols., 2003).

Una fracción del volumen es enrutada mediante un reservorio de capacidad x_3 , cuyo estado (R en la Figura 2.1) se actualiza mediante:

$$R_{t+1} = R_t + P_n - P_s - Q_r$$

donde Q_r representa el caudal filtrado en el reservorio, correspondiente al flujo de salida del almacén de enrutamiento mostrado en la figura.

Adicionalmente, el término de intercambio subterráneo (representado como $F(x_2)$ en la Figura 2.1) se modela mediante:

$$Ex = x_2 \left(\frac{R_t}{x_3} \right)^{\frac{7}{2}}$$

donde Ex puede ser positivo (aporte externo) o negativo (pérdida hacia el subsuelo o cuencas vecinas) (Perrin y cols., 2003).

Finalmente, el caudal total simulado (Q en la Figura 2.1) es resultado de la suma del flujo proveniente del reservorio de enrutamiento y del flujo rápido, ambos posteriormente ajustados por el término de intercambio.

Propósito y aplicaciones

El modelo GR4J ha sido ampliamente empleado en hidrología tanto operativa como científica debido a su robustez, eficiencia computacional y estructura conceptual clara (Perrin y cols., 2003). Ha sido aplicado para la simulación continua de caudales, análisis de respuesta hidrológica bajo diferentes regímenes climáticos, estudios de calibración y transferencia de parámetros, así como para comparaciones frente a modelos más complejos y frente a enfoques basados en aprendizaje automático.

En síntesis, GR4J constituye un modelo conceptual compacto y eficiente que ha demostrado desempeño satisfactorio en múltiples regiones y condiciones, convirtiéndose en un referente para la modelación lluvia–escorrentía a escala diaria (Perrin y cols., 2003).

2.2. Aprendizaje automático y redes neuronales en modelación hidrológica

En paralelo al desarrollo de modelos conceptuales y físicamente basados, la hidrología ha incorporado enfoques *data-driven* para representar relaciones complejas entre variables hidrometeorológicas y respuestas hidrológicas observadas. En este marco, las redes neuronales artificiales han sido utilizadas desde la década de 1990 como aproximadores no lineales con aplicaciones en el pronóstico de caudales y la modelación hidrológica ([ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000a, 2000b](#)). Un ejemplo temprano es el trabajo de [Hsu y cols. \(1995\)](#), donde se entrena una red neuronal para modelar el proceso lluvia–escorrentía a partir de series temporales diarias de precipitación y caudal en la salida de la cuenca, incluyendo valores rezagados de ambas variables para capturar la memoria hidrológica del sistema.

El incremento reciente en la disponibilidad de datos y recursos computacionales ha impulsado el uso de arquitecturas más profundas y de redes con memoria, así como el desarrollo de enfoques híbridos que integran restricciones físicas en el aprendizaje ([Tripathy y Mishra, 2024](#)). A continuación, se presentan las arquitecturas neuronales consideradas en este trabajo y su vínculo con aplicaciones hidrológicas.

2.2.1. Perceptrón multicapa (MLP)

El **perceptrón multicapa** (*Multi-Layer Perceptron*, MLP) es una red neuronal *feed-forward* compuesta por capas totalmente conectadas. Su entrenamiento se realiza típicamente mediante el algoritmo de *retropropagación del error* (*backpropagation*) ([Rumelhart y cols., 1986](#)). Además, se ha demostrado que redes *feed-forward* con al menos una capa oculta pueden aproximar funciones con alta generalidad bajo condiciones amplias, lo cual fundamenta su capacidad de representación ([Hornik y cols., 1989](#)).

Sea $\mathbf{x} \in \mathbb{R}^d$ el vector de entrada. Un MLP con L capas puede escribirse como:

$$\mathbf{a}^{(0)} = \mathbf{x}, \quad \mathbf{a}^{(\ell)} = \sigma^{(\ell)}\left(\mathbf{W}^{(\ell)}\mathbf{a}^{(\ell-1)} + \mathbf{b}^{(\ell)}\right), \quad \ell = 1, \dots, L,$$

donde $\mathbf{W}^{(\ell)}$ y $\mathbf{b}^{(\ell)}$ son los parámetros entrenables, $\sigma^{(\ell)}(\cdot)$ es una función de activación no lineal, y la salida del modelo es $\hat{y} = \mathbf{a}^{(L)}$.

En hidrología, el MLP se ha utilizado como aproximador de la transformación lluvia–escorrentía, aprendiendo una relación no lineal entre entradas meteorológicas (p. ej., precipitación y evapotranspiración potencial) y el caudal ([Hsu y cols., 1995](#)). Dado que un MLP no incorpora memoria temporal de forma explícita, es común construir vectores de entrada con rezagos o ventanas temporales (por ejemplo, $P_{t-k:t}$ y $E_{t-k:t}$) para capturar efectos de almacenamiento y respuesta retardada ([Besaw y cols., 2010](#)).

2.2.2. Redes neuronales recurrentes (RNN)

Las **redes neuronales recurrentes** (*Recurrent Neural Networks*, RNN) introducen realimentación en su arquitectura, permitiendo modelar dependencias temporales mediante un estado oculto que se actualiza secuencialmente (Elman, 1990). Para una serie de entradas $\{\mathbf{x}_t\}$, una formulación básica es:

$$\mathbf{h}_t = \phi(\mathbf{W}_x \mathbf{x}_t + \mathbf{W}_h \mathbf{h}_{t-1} + \mathbf{b}_h), \quad \hat{\mathbf{y}}_t = \psi(\mathbf{W}_y \mathbf{h}_t + \mathbf{b}_y),$$

donde \mathbf{h}_t es el estado oculto en el instante t , $\hat{\mathbf{y}}_t$ es la salida estimada del modelo, \mathbf{W}_x , \mathbf{W}_h y \mathbf{W}_y son matrices de pesos, \mathbf{b}_h y \mathbf{b}_y son vectores de sesgo, y $\phi(\cdot)$, $\psi(\cdot)$ son funciones no lineales.

Un problema clásico en el entrenamiento de RNN es la dificultad para aprender dependencias de largo plazo debido a gradientes que se desvanecen o explotan durante la retropropagación a través del tiempo (Bengio y cols., 1994). Para mitigar esta limitación, Hochreiter y Schmidhuber (1997) proponen la arquitectura **Long Short-Term Memory** (LSTM), que incorpora una memoria interna y compuertas que regulan qué información se retiene, se actualiza y se expone como salida. Esta estructura permite preservar información relevante durante horizontes temporales extensos, facilitando el modelado de dinámicas con memoria prolongada.

Como alternativa a LSTM, Cho y cols. (2014) introducen la arquitectura **Gated Recurrent Unit** (GRU), que simplifica el diseño al combinar las compuertas de entrada y olvido en una única compuerta de actualización, y prescindir de una memoria interna explícita separada del estado oculto. Su formulación básica puede expresarse como:

$$\begin{aligned} \mathbf{z}_t &= \sigma(\mathbf{W}_z \mathbf{x}_t + \mathbf{U}_z \mathbf{h}_{t-1}), & \mathbf{r}_t &= \sigma(\mathbf{W}_r \mathbf{x}_t + \mathbf{U}_r \mathbf{h}_{t-1}), \\ \tilde{\mathbf{h}}_t &= \tanh(\mathbf{W}_h \mathbf{x}_t + \mathbf{U}_h (\mathbf{r}_t \odot \mathbf{h}_{t-1})), \\ \mathbf{h}_t &= (1 - \mathbf{z}_t) \odot \mathbf{h}_{t-1} + \mathbf{z}_t \odot \tilde{\mathbf{h}}_t, \end{aligned}$$

En forma análoga a las RNN, la salida del modelo puede obtenerse a partir del estado oculto como $\hat{\mathbf{y}}_t = \psi(\mathbf{W}_y \mathbf{h}_t + \mathbf{b}_y)$.

donde \mathbf{z}_t y \mathbf{r}_t corresponden a las compuertas de actualización y reinicio, respectivamente; \mathbf{W}_z , \mathbf{W}_r , \mathbf{W}_h , \mathbf{U}_z , \mathbf{U}_r y \mathbf{U}_h son matrices de pesos, y \odot denota el producto elemento a elemento. Las GRU mantienen la capacidad de modelar dependencias temporales de mediano y largo plazo, con una arquitectura más parsimoniosa en términos de parámetros, lo que puede traducirse en menor costo computacional y mayor estabilidad en el entrenamiento.

En el contexto hidrológico, las RNN —y en particular LSTM y GRU— han mostrado capacidad para capturar la *memoria hidrológica* asociada a procesos de almacenamiento y liberación sin requerir la definición explícita de estados conceptuales. Para lluvia–escorrentía a escala diaria, Kratzert y cols. (2018) presentan el uso de LSTM en múltiples cuencas y comparan sus resultados con un modelo hidrológico establecido, destacando el potencial de este enfoque para modelación hidrológica regional.

2.3. *Physics-Informed Neural Networks* (PINNs)

Un desafío recurrente en modelos puramente *data-driven* es garantizar que las predicciones respeten restricciones físicas fundamentales (por ejemplo, leyes de conservación, coherencia dimensional o cotas de estabilidad). Si bien estos modelos pueden ajustarse adecuadamente a los datos disponibles, sus predicciones pueden resultar físicamente inconsistentes o implausibles fuera del dominio de entrenamiento (Karniadakis y cols., 2021).

En respuesta a esta limitación, se ha desarrollado el marco de *physics-informed learning*, definido como la incorporación explícita de conocimiento físico previo, derivado de leyes empíricas, principios matemáticos o ecuaciones diferenciales gobernantes, dentro del proceso de entrenamiento de algoritmos de aprendizaje automático (Karniadakis y cols., 2021). Dentro de este enfoque, las *Physics-Informed Neural Networks* (PINNs) constituyen una clase particular de modelos en los que las ecuaciones diferenciales que describen el sistema se integran directamente en la función de pérdida, actuando como una forma de regularización física del aprendizaje (Raissi y cols., 2019; Karniadakis y cols., 2021).

2.3.1. Formulación general

Considérese una ecuación diferencial que gobierna un proceso hidrológico, expresada abstractamente como:

$$\mathcal{N}[u(\mathbf{x}, t)] = 0, \quad (\mathbf{x}, t) \in \Omega \times [0, T],$$

junto con condiciones iniciales y/o de borde. En una PINN, se aproxima la solución mediante una red neuronal $u_\theta(\mathbf{x}, t)$ parametrizada por θ . La idea central es definir una pérdida que combine (i) el ajuste a datos observados y (ii) el residuo de la ecuación física evaluado en el dominio del problema (Raissi y cols., 2019). Para este último término, se emplean típicamente *puntos de colocación*, que pueden incluir tanto puntos con observaciones como puntos adicionales generados dentro del dominio sin datos asociados. En estos puntos se evalúa el residuo $\mathcal{N}[u_\theta]$ mediante diferenciación automática, penalizando el incumplimiento de la ecuación gobernante. La incorporación de puntos adicionales permite reforzar el cumplimiento de las restricciones físicas y densificar el entrenamiento incluso en escenarios con disponibilidad limitada de datos experimentales. Este mecanismo distingue a las PINNs de los enfoques puramente supervisados, al permitir que la información estructural del sistema complemente explícitamente la información contenida en los datos.

Un esquema típico de entrenamiento se escribe como:

$$\mathcal{L}(\theta) = \lambda_{\text{data}} \mathcal{L}_{\text{data}}(\theta) + \lambda_{\text{phys}} \mathcal{L}_{\text{phys}}(\theta) + \lambda_{\text{bc}} \mathcal{L}_{\text{bc}}(\theta),$$

donde $\mathcal{L}_{\text{data}}$ representa el error respecto a observaciones, $\mathcal{L}_{\text{phys}}$ penaliza el residuo $\mathcal{N}[u_\theta]$, y \mathcal{L}_{bc} incorpora condiciones iniciales y/o de borde. El balance entre términos se controla mediante los pesos λ .

Este marco permite abordar problemas **directos** e **inversos** dentro de una formulación unificada, incluyendo la identificación de parámetros o forzantes desconocidos (Raissi y cols., 2019; Karniadakis y cols., 2021). A la vez, la literatura reporta desafíos prácticos, como sensibilidad a inicialización, balanceo de pérdidas y mayores costos computacionales en procesos altamente no lineales (Karniadakis y cols., 2021).

2.3.2. Aplicaciones en hidrología y recursos hídricos

Las *Physics-Informed Neural Networks* (PINNs) han sido exploradas en hidrología principalmente en problemas donde existen ecuaciones gobernantes bien establecidas, tales como flujos subsuperficiales y procesos hidrodinámicos. Un ejemplo representativo es el trabajo de Bandai y Ghezzehei (2022), en el que se aplican PINNs a la ecuación de Richards para flujo no saturado, incorporando *domain decomposition* para manejar suelos estratificados con discontinuidades en la conductividad hidráulica.

Más recientemente, se han propuesto extensiones y variantes que mejoran la estabilidad y el desempeño de las PINNs en contextos hidrológicos complejos. En particular, Y. Li y cols. (2025) presentan un enfoque híbrido que acopla PINNs con simulaciones generadas mediante Hydrus-1D, mostrando mejoras en la resolución de la ecuación de Richards frente a formulaciones puramente informadas por la física. En el ámbito hidrológico-hidráulico, Tian y cols. (2025) utilizan PINNs para resolver las ecuaciones bidimensionales de aguas someras incorporando topografía y términos fuente de lluvia, lo que resulta relevante para aplicaciones en inundaciones y escorrentía superficial.

Asimismo, las PINNs han sido empleadas para abordar problemas inversos y de estimación de parámetros en sistemas hidráulicos con observaciones limitadas. Por ejemplo, Z. Li y cols. (2025) desarrollan una PINN para la predicción de flujos en canales abiertos con monitoreo escaso, demostrando que la incorporación explícita de las ecuaciones físicas permite reconstruir la dinámica espacio-temporal y estimar parámetros hidráulicos a partir de datos ruidosos.

En conjunto, estos trabajos recientes evidencian el potencial de las PINNs para integrar datos observacionales y conocimiento físico en sistemas hidrológicos gobernados por ecuaciones diferenciales, ofreciendo una alternativa prometedora frente a enfoques puramente basados en datos (Raissi y cols., 2019; Karniadakis y cols., 2021). A diferencia de los modelos de aprendizaje profundo convencionales, las PINNs incorporan las leyes físicas directamente en la función de pérdida, lo que les permite obtener soluciones físicamente consistentes incluso cuando la disponibilidad de datos observados es reducida (Raissi y cols., 2019).

Como antecedentes nacionales, aunque no basados en PINNs, cabe destacar dos trabajos relevantes. Por un lado, Duque y cols. (2022) aplican redes neuronales para la predicción empírica de niveles hidrométricos en la cuenca del Río Negro (Uruguay). Por otro lado, Vilaseca y cols. (2023) emplean modelos de *Random Forest* para la simulación de caudal diario en subcuencas del río Santa Lucía, identificando variables predictivas clave como la precipitación acumulada previa y los caudales rezagados. Ambos trabajos ponen de manifiesto el interés

creciente en la aplicación de técnicas de aprendizaje automático para problemas hidrológicos en el contexto uruguayo.

Capítulo 3

Caso de estudio

3.1. Descripción de la cuenca de estudio

La predicción del **caudal en sistemas fluviales** constituye una herramienta fundamental para la gestión sostenible de los recursos hídricos. Conocer la respuesta hidrológica de una cuenca frente a las condiciones meteorológicas permite anticipar situaciones de crecida o déficit hídrico, optimizar la planificación del uso del agua y contribuir a la gestión del riesgo asociado a eventos extremos. En este contexto, el desarrollo de modelos de estimación de caudal basados en información hidrometeorológica aporta herramientas para la **predicción, el manejo del recurso y la toma de decisiones en la gestión hídrica**.

En el contexto uruguayo, la **cuenca del río Santa Lucía** posee una relevancia estratégica, dado que constituye la principal fuente de abastecimiento de agua potable para el área metropolitana de Montevideo. En particular, el embalse de Paso Severino, ubicado sobre el río Santa Lucía Chico, concentra gran parte del sistema de captación destinado al consumo humano. Esta importancia justifica el interés en comprender y modelar con precisión la dinámica hidrológica de sus principales subcuencas aportantes.

El presente trabajo se focaliza en la **subcuenca del río Santa Lucía Chico** delimitada en la estación **DINAGUA Florida – Puente Ruta 5 (código 53.1)**, en adelante denominada *subcuenca Florida – Puente Ruta 5*. Esta estación constituye un punto de control hidrológico clave, ya que registra el **aporte natural principal al embalse de Paso Severino**.

De acuerdo con **Vilaseca y cols. (2023)**, la subcuenca del **Santa Lucía Chico** abarca aproximadamente **2478 km²** y presenta un uso del suelo dominado por **pasturas (82.4%)** y **cultivos (9.4%)**, con un clima templado y precipitaciones anuales entre **1000 y 1500 mm**. Las temperaturas varían típicamente entre **3 °C y 30 °C**, en concordancia con el régimen térmico del centro-sur del Uruguay.

Desde el punto de vista hidrológico, se trata de una **cuenca de cabecera** que concentra los escurrimientos generados en el tramo alto del sistema, presen-

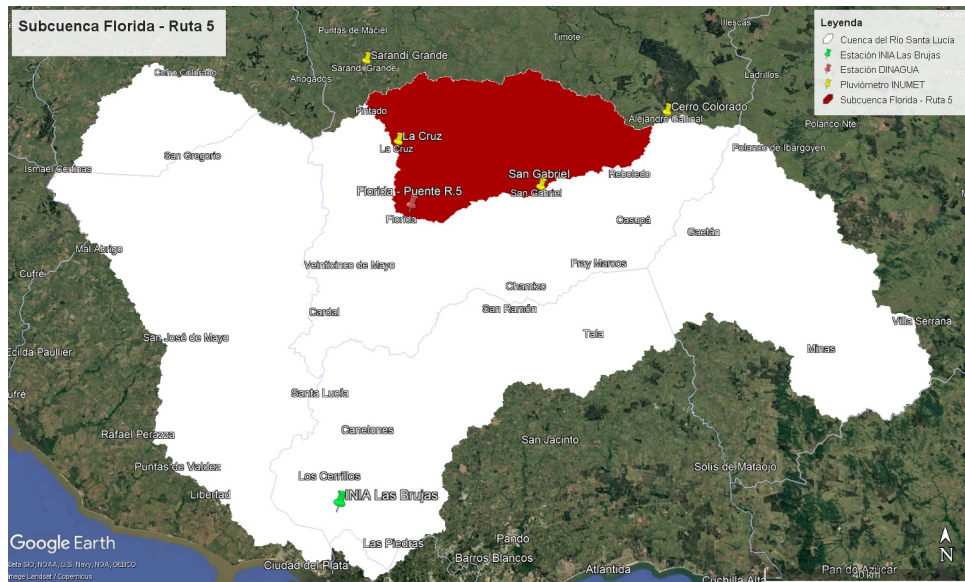


Figura 3.1: Subcuenca Florida – Puente Ruta 5 y estaciones hidrometeorológicas utilizadas en el estudio, clasificadas por organismo (DINAGUA, INUMET e INIA).

tando un comportamiento predominantemente **lótico**. Aguas abajo, el sistema adquiere un carácter más **léntico** debido a la presencia del embalse de Paso Severino, lo que genera diferencias en los tiempos de respuesta hidrológica entre las estaciones de Florida y Paso Severino (Vilaseca y cols., 2023).

Estudios previos han analizado la calidad del agua en este sistema. En particular, Gorgoglione y cols. (2020) demostraron que las actividades agrícolas, ganaderas y urbanas inciden en los procesos de eutrofización del embalse de Paso Severino, identificando correlaciones significativas entre el fósforo total y el uso agrícola del suelo, así como entre el nitrógeno y las áreas urbanizadas. Estos resultados evidencian la interacción entre procesos hidrológicos y factores antrópicos, reforzando la importancia de modelar adecuadamente la dinámica de caudales en esta subcuenca.

El recorte espacial adoptado responde a un criterio metodológico: analizar una subcuenca delimitada permite reducir la complejidad del problema hidrológico, concentrando la evaluación en un único punto de salida de caudal correspondiente a la estación Florida – Puente Ruta 5. Este enfoque facilita la validación inicial de los modelos de *Physics-Informed Neural Networks* (PINNs), al trabajar en un sistema relativamente bien caracterizado y representativo, que posteriormente podría escalarse a dominios espaciales de mayor complejidad.

Cabe destacar que esta subcuenca coincide con una de las zonas analizadas por Vilaseca y cols. (2023) para la simulación de caudal diario mediante modelos de *Random Forest*. En dicho trabajo se emplearon las mismas fuen-

tes de información hidrometeorológica utilizadas en la presente investigación (**caudales de DINAGUA**, **precipitaciones de INUMET** —La Cruz, San Gabriel, Sarandí Grande y Cerro Colorado— y **variables meteorológicas de INIA Las Brujas**) para el período 1989–2016. Los autores demostraron que el acumulado de precipitación de los siete días previos (P_{accum}) constituye un indicador eficiente del estado de humedad antecedente del suelo, mientras que los **caudales rezagados** (Q_{t-1} , Q_{t-2}) aportan la mayor capacidad predictiva para el caudal diario.

3.2. Información disponible

El estudio de la subcuenca requiere identificar y caracterizar las fuentes de información hidrometeorológica disponibles en la región. Estas provienen de distintas instituciones nacionales con funciones complementarias en la gestión del recurso hídrico: la Dirección Nacional de Aguas (DINAGUA), el Instituto Uruguayo de Meteorología (INUMET) y el Instituto Nacional de Investigación Agropecuaria (INIA).

La Figura 3.1 presenta las estaciones seleccionadas para el análisis en la cuenca del Santa Lucía y su entorno inmediato. Los criterios de selección y la justificación de las variables consideradas se detallan en los apartados siguientes.

3.2.1. Estación hidrométrica

La **Dirección Nacional de Aguas (DINAGUA)**, dependiente del Ministerio de Ambiente (MA), opera una extensa red de estaciones hidrométricas a nivel nacional. Los registros de esta institución comprenden series temporales de *caudal* (Q) [m^3/s], disponibles públicamente a través de su portal de datos hidrológicos abiertos ([Dirección Nacional de Aguas \(DINAGUA\), 2026](#)) y abarcando el período comprendido entre el **1 de enero de 1980** y el **4 de julio de 2023**. La frecuencia temporal de registro es **diaria**.

En la cuenca del río Santa Lucía, DINAGUA mantiene diversas estaciones de monitoreo distribuidas a lo largo del cauce principal y sus afluentes (véase Tabla 3.1).

Tabla 3.1: Estaciones de monitoreo y variables recopiladas por DINAGUA.

Variable	Estación
Caudal (Q) [m^3/s]	Fray Marcos (DINAGUA)
Caudal (Q) [m^3/s]	Florida (Puente Ruta 5)
Caudal (Q) [m^3/s]	Paso Pache (Ruta 5 nueva)
Caudal (Q) [m^3/s]	Picada de Varela
Caudal (Q) [m^3/s]	Paso Roldán
Caudal (Q) [m^3/s]	Paso de los Troncos
Caudal (Q) [m^3/s]	Santa Lucía (Ruta 11)

De este conjunto, la estación *Florida – Puente Ruta 5* fue seleccionada como punto de referencia para el presente estudio, dado que constituye el **punto de cierre hidrológico** de la subcuenca analizada. Su elección responde a que:

- delimita naturalmente la subcuenca de análisis,
- representa el punto de salida del flujo superficial acumulado en el área, y
- permite comparar las predicciones de los modelos con los valores observados por DINAGUA.

En este trabajo se considera el **caudal (Q)** como variable principal, dado que constituye la *variable de salida característica de los modelos hidrológicos*, y permite evaluar el desempeño del modelo en términos del flujo superficial observado en la estación de referencia.

El porcentaje de valores faltantes en la serie de caudal es de **7.52 %**.

Por tanto, la estación Florida – Puente Ruta 5 define el **marco de validación del modelo hidrológico desarrollado**, al permitir contrastar las simulaciones con los registros oficiales de DINAGUA.

3.2.2. Estaciones pluviométricas

El **Instituto Uruguayo de Meteorología (INUMET)** dispone de una amplia red de estaciones meteorológicas y pluviométricas distribuidas en todo el territorio nacional. Si bien INUMET registra diversas variables atmosféricas, en el presente trabajo se utilizaron únicamente los datos de precipitación acumulada diaria, que constituyen una fuente fundamental para caracterizar el régimen de lluvias y su relación con los procesos de escurrimiento superficial en la cuenca del río Santa Lucía.

Con fines académicos, se solicitó formalmente a INUMET el acceso a sus registros históricos (**Instituto Uruguayo de Meteorología (INUMET), 2023**). El conjunto de datos proporcionado abarca el período comprendido entre el **1 de enero de 1980** y el **27 de junio de 2023**, con una **frecuencia diaria** de registro, e incluye series de *precipitación (P)* expresadas en milímetros [mm], registradas en un conjunto amplio de estaciones pluviométricas distribuidas en la región de estudio.

No obstante, no todas las estaciones resultan relevantes para el área de estudio. Dado que el análisis se centra en la subcuenca asociada a la estación hidrométrica *Florida (Puente Ruta 5)*, se seleccionaron únicamente aquellas estaciones pluviométricas ubicadas en la parte alta y media de dicha subcuenca (véase Figura 3.1). Estas representan las áreas donde se generan los escurrimientos que posteriormente confluyen hacia el punto de control de caudal.

En función de criterios de representatividad espacial y disponibilidad de datos, el subconjunto final considerado en el modelado incluye las siguientes estaciones:

- Florida — 21.54 % de datos faltantes

- La Cruz — 2.44 % de datos faltantes
- San Gabriel — 2.32 % de datos faltantes
- Sarandí Grande — 1.37 % de datos faltantes
- Cerro Colorado — 0.19 % de datos faltantes

La inclusión de múltiples estaciones distribuidas espacialmente permite capturar la variabilidad espacial de las precipitaciones dentro de la cuenca, reduciendo la incertidumbre asociada a la estimación de la precipitación media areal y mejorando la representación de los aportes *aguas arriba* del punto de control.

3.2.3. Estación meteorológica

El **Instituto Nacional de Investigación Agropecuaria (INIA)** opera una red de estaciones meteorológicas distribuidas en distintas regiones del país, destinadas al monitoreo de variables atmosféricas y energéticas relevantes para la actividad agropecuaria y los estudios climáticos.

En este trabajo se incorporaron los registros provenientes de la estación *INIA Las Brujas* (**Instituto Nacional de Investigación Agropecuaria (INIA), 2026**), ubicada en el departamento de Canelones, al suroeste de la cuenca del río Santa Lucía (véase Figura 3.1). Si bien esta estación se encuentra fuera del límite inmediato de la subcuenca de interés, su inclusión se justifica debido a que las variables atmosféricas que mide presentan una variabilidad espacial relativamente baja a escala regional, especialmente en el Uruguay centro-sur. Por tanto, pueden considerarse representativas de las condiciones climáticas que influyen sobre la subcuenca analizada.

El conjunto de datos de INIA abarca el período comprendido entre el **1 de enero de 1980** y el **4 de julio de 2023**, con una **frecuencia diaria** de registro, e incluye las variables listadas en la Tabla 3.2.

Tabla 3.2: Variables meteorológicas registradas en la estación INIA Las Brujas.

Variable
Evapotranspiración Penman (ET) [mm]
Heliofanía (Hel) [hs]
Humedad relativa media (HR) [%]
Radiación solar (RS) [cal/cm ² /día]
Temperatura del aire media (TA) [°C (24hs)]
Temperatura del aire máxima (TA) [°C]
Temperatura del aire mínima (TA) [°C]
Velocidad del viento (VV) [2m/km/24h]

En términos de completitud, las series meteorológicas presentan porcentajes bajos de valores faltantes:

- Velocidad del viento (VV): 0.11 %

- Temperatura del aire media (TA): 0.10 %
- Evapotranspiración Penman (ET): 0.09 %
- Humedad relativa media (HR): 0.09 %
- Heliofanía (Hel): 0.0 %
- Radiación solar (RS): 0.0 %
- Temperatura del aire máxima (TA): 0.0 %
- Temperatura del aire mínima (TA): 0.0 %

Las variables meteorológicas registradas —tales como *temperatura*, *radiación solar*, *humedad relativa* y *velocidad del viento*— complementan la información pluviométrica de INUMET, aportando una descripción más integral del balance atmosférico y energético de la región.

De esta manera, la información de INIA Las Brujas contribuye a contextualizar los procesos hidrológicos que ocurren dentro de la subcuenca y a mejorar la interpretación de las relaciones entre precipitación, evapotranspiración y caudal.

3.3. Relaciones espaciales entre estaciones

La integración espacial de la información hidrometeorológica requiere considerar explícitamente la disposición geográfica relativa entre las estaciones de monitoreo y el punto de control de caudal. En esta sección se describen las métricas de distancia utilizadas y el procedimiento de delimitación de áreas de influencia mediante polígonos de Thiessen, los cuales serán empleados posteriormente en distintos enfoques de modelado.

3.3.1. Distancia entre la estación hidrométrica y las estaciones pluviométricas

Con el objetivo de cuantificar la proximidad espacial entre el punto de salida de la subcuenca (DINAGUA Florida – Puente Ruta 5) y las estaciones pluviométricas seleccionadas, se calcularon distancias lineales aproximadas entre ellas. Estas distancias serán utilizadas posteriormente en esquemas de ponderación por proximidad (por ejemplo, ponderación inversa a la distancia), como alternativa para integrar la precipitación espacialmente distribuida en una única serie representativa.

Las distancias consideradas (en kilómetros) fueron:

Florida:	0.1
La Cruz:	17
San Gabriel:	30
Sarandí Grande:	41
Cerro Colorado:	65

En el caso de la estación Florida, se asignó una distancia de 0.1 km a efectos computacionales, con el fin de evitar divisiones por cero en expresiones de ponderación inversa.

3.3.2. Delimitación de áreas de influencia: polígonos de Thiessen

Para estimar una precipitación media areal representativa de la subcuenca Florida – Puente Ruta 5 a partir de múltiples estaciones pluviométricas, se aplicó el método de los **polígonos de Thiessen**. Este procedimiento geométrico asigna a cada estación un área de influencia definida por el conjunto de puntos más cercanos a dicha estación que a cualquier otra, en términos de distancia euclidiana.

El procedimiento consistió en:

- construir los polígonos de Thiessen a partir de la ubicación geográfica de las estaciones Cerro Colorado, La Cruz, Florida, Sarandí Grande y San Gabriel;
- recortar dichos polígonos al contorno de la subcuenca de interés;
- calcular el porcentaje de área de cada polígono respecto al área total de la subcuenca.

Los pesos relativos obtenidos fueron:

Cerro Colorado:	0.1446
La Cruz:	0.3045
Florida:	0.1081
Sarandí Grande:	0.0490
San Gabriel:	0.3938

Estos coeficientes representan la fracción del área total de la subcuenca asociada a cada estación. La precipitación media areal diaria se calcula entonces como un promedio ponderado:

$$P_{\text{areal}}(t) = \sum_{i=1}^n w_i P_i(t)$$

donde w_i corresponde al peso del polígono de Thiessen de la estación i , y $P_i(t)$ es la precipitación diaria registrada en dicha estación.

La Figura 3.2 muestra la delimitación final de los polígonos recortados al contorno de la subcuenca, junto con la ubicación de las estaciones consideradas.

Este procedimiento será utilizado en capítulos posteriores para construir series integradas de precipitación representativas de la subcuenca, y para comparar enfoques alternativos de integración espacial basados en proximidad (ponderación por distancia) o en áreas de influencia (Thiessen).

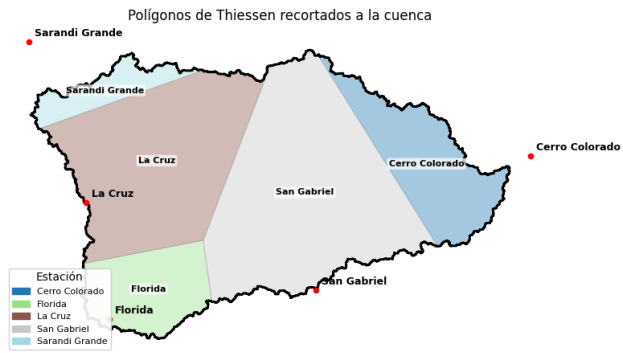


Figura 3.2: Polígonos de Thiessen recortados a la subcuenca Florida – Puente Ruta 5 y estaciones pluviométricas utilizadas.

3.4. Síntesis

En resumen, la selección de estaciones y el tratamiento espacial de la información responden a criterios **hidrológicos, espaciales y de disponibilidad de datos**:

- **DINAGUA Florida – Puente Ruta 5** define el *punto de salida hidrológico* de la subcuenca y la *variable objetivo (caudal)*, permitiendo contrastar simulaciones y observaciones.
- Las estaciones **INUMET** de **La Cruz, San Gabriel, Sarandí Grande, Cerro Colorado** y **Florida** aportan las *entradas pluviométricas* más representativas del área de drenaje y permiten capturar la variabilidad espacial de la precipitación.
- La estación **INIA Las Brujas** provee variables *meteorológicas de contexto* (temperatura, radiación, humedad, viento y evapotranspiración), cuya homogeneidad espacial regional justifica su utilización aun estando fuera del límite inmediato de la subcuenca.
- Para integrar la precipitación registrada en múltiples estaciones en una única serie representativa, se definieron dos elementos espaciales que serán utilizados como referencia metodológica en capítulos posteriores:
 - **Distancia geográfica** entre el punto de control de caudal y las estaciones pluviométricas, lo que permite implementar y evaluar esquemas de ponderación por proximidad (p. ej., ponderación inversa a la distancia).
 - **Polígonos de Thiessen** recortados a la subcuenca, a partir de los cuales se obtuvieron *pesos areales* para calcular precipitación media diaria mediante promedio ponderado.

Este conjunto de estaciones, junto con la caracterización espacial de sus relaciones, permite disponer de un marco de datos consistente para el desarrollo, comparación y validación de los modelos hidrológicos propuestos en los capítulos siguientes. En la Tabla 3.3 se presenta un resumen de la información utilizada para este estudio.

Tabla 3.3: Resumen de estaciones y variables utilizadas en el estudio.

Organismo	Estación	Variable	Valores faltantes [%]
DINAGUA	Florida – Puente Ruta 5	Caudal (Q) [m ³ /s]	7.52
INUMET	La Cruz	Precipitación (P) [mm]	2.44
	San Gabriel	Precipitación (P) [mm]	2.32
	Sarandí Grande	Precipitación (P) [mm]	1.37
	Cerro Colorado	Precipitación (P) [mm]	0.19
	Florida	Precipitación (P) [mm]	21.54
INIA	Las Brujas	Evapotranspiración Penman (ET) [mm]	0.09
		Heliofanía (Hel) [hs]	0.00
		Humedad relativa media (HR) [%]	0.09
		Radiación solar (RS) [cal/cm ² /día]	0.00
		Temperatura media (TA) [°C]	0.10
		Temperatura máxima (TA) [°C]	0.00
		Temperatura mínima (TA) [°C]	0.00
Velocidad del viento (VV) [2m/km/24h]	0.11		

Nota: Todas las series poseen frecuencia diaria. Los períodos de registro son 01/01/1980–04/07/2023 para DINAGUA e INIA, y 01/01/1980–27/06/2023 para INUMET.

Capítulo 4

Análisis de datos

En este capítulo se abordan las etapas de procesamiento y análisis preliminar de los datos utilizados en el desarrollo de los modelos predictivos. Una vez definidas las fuentes y variables relevantes en el capítulo anterior, el siguiente paso consiste en consolidar la información proveniente de las distintas instituciones, garantizando su coherencia temporal y espacial.

El proceso incluye la integración de las series hidrométricas, pluviométricas y meteorológicas en una base de datos unificada, la detección y corrección de valores faltantes y la aplicación de métodos de imputación para completar los registros incompletos. Posteriormente, se realiza un análisis correlativo de las variables para observar su relacionamiento, identificando posibles asociaciones significativas que permitan comprender mejor la influencia de unas sobre otras y su impacto en los resultados del estudio.

De esta manera, el capítulo establece las bases metodológicas para el entrenamiento de los modelos de predicción presentados en los apartados siguientes, asegurando que los datos empleados sean representativos, consistentes y adecuados para el análisis hidrológico.

4.1. Consolidación inicial del conjunto de datos

Una vez obtenidos los registros de las distintas fuentes (DINAGUA, INUMET e INIA), fue necesario integrarlos en un único conjunto de datos que sirviera de base para el modelado. Este proceso implicó varios pasos clave:

- **Unificación temporal:** todos los registros fueron alineados en torno a una misma escala temporal, utilizando como referencia la fecha. Esto permitió que, para cada día del período analizado, se dispusiera de un vector completo con las variables hidrológicas y meteorológicas correspondientes.
- **Integración de fuentes:** los distintos archivos provenientes de cada organismo fueron combinados empleando la fecha como índice común. De

esta forma se evitó la duplicidad y se garantizó la coherencia entre las distintas variables medidas.

- **Selección de estaciones relevantes:** del total de estaciones disponibles, únicamente se conservaron aquellas vinculadas a la subcuenca de estudio (Florida, La Cruz, San Gabriel, Sarandí Grande, Cerro Colorado e INIA Las Brujas), conforme al criterio metodológico descrito previamente.
- **Variables incluidas:** se mantuvieron todas las variables meteorológicas provistas por los conjuntos de datos, incorporando de manera íntegra la información disponible.
- **Variable objetivo:** además de las variables meteorológicas, se incorporó como variable objetivo el caudal registrado en la estación *Florida (Puente Ruta 5)*, correspondiente a la subcuenca en análisis.

El resultado de este proceso es un conjunto de datos consolidado, estructurado de manera que cada fila representa un día dentro del período de estudio y cada columna corresponde a una variable hidrológica o meteorológica de interés. Este conjunto de datos unificado constituye la base a partir de la cual se desarrollan la imputación de datos, el análisis exploratorio y los modelos predictivos presentados en los siguientes apartados.

4.2. Imputación de datos

Antes del modelado se analizó la presencia y distribución temporal de valores faltantes. La principal discontinuidad correspondió a la precipitación en la estación de Florida, asociada a la ausencia de registros en el período 1980–1989, lo que motivó el recorte del intervalo inicial para garantizar una ventana temporal homogénea. Tras esta depuración, todas las variables presentaron menos del 6.5% de datos faltantes. La imputación se realizó únicamente en las variables forzantes, es decir, en las variables de entrada del modelo (precipitación y variables meteorológicas). La precipitación se completó mediante interpolación espacial por distancia inversa ponderada (IDW, $p=2$), mientras que las variables meteorológicas de INIA Las Brujas, con menos del 0.15% de datos faltantes y comportamiento temporal suave, se interpolaron linealmente en el tiempo. El caudal no fue imputado para evitar introducir información artificial en la variable objetivo; sin embargo, las series forzantes imputadas se propagaron a través del modelo hidrológico para preservar la memoria del sistema. La selección del método de imputación para precipitación se basó en una validación cruzada específica, en la que el IDW mostró el mejor compromiso entre error, preservación de la estructura espacial y retención del conjunto de datos, evitando la pérdida del 43% de los registros que implicaría trabajar únicamente con casos completos. La metodología aplicada y los resultados detallados se presentan en el Anexo A.

4.3. Análisis exploratorio de datos

El objetivo de este apartado es caracterizar el comportamiento de las variables del conjunto de datos consolidado (véanse las secciones anteriores) y, en particular, analizar tanto sus propiedades estadísticas básicas como las correlaciones existentes entre ellas. Este diagnóstico orienta el proceso de *feature engineering*, permite detectar redundancias entre los predictores y aporta hipótesis sobre la dinámica hidrológica de la subcuenca.

4.3.1. Análisis descriptivo

En primera instancia, se realizará un análisis estadístico descriptivo de la variable caudal, con el fin de caracterizar su distribución y magnitud. Las medidas estadísticas serán calculados excluyendo los valores faltantes.

Se estimarán medidas de tendencia central (media y mediana), medidas de dispersión (desviación estándar) y valores extremos (mínimo y máximo). Estas métricas permitirán evaluar la estructura de la distribución, su grado de asimetría, la variabilidad de la serie temporal y la presencia de eventos extremos, aspectos característicos de las series hidrológicas.

Asimismo, estas medidas estadísticas proporcionarán un marco de referencia para la interpretación de las métricas de desempeño utilizadas en la etapa de modelado, particularmente el error absoluto medio (MAE) y la raíz del error cuadrático medio (RMSE). En efecto, la magnitud de dichos errores solo puede evaluarse adecuadamente en relación con la escala y la variabilidad propia de la variable objetivo. Por lo tanto, la caracterización estadística previa resulta necesaria para contextualizar la calidad de las estimaciones obtenidas.

4.3.2. Análisis correlativo

El análisis se realizará sobre el *dataset* obtenido tras aplicar el recorte temporal y las imputaciones descritas en secciones anteriores. Para estimar las dependencias lineales entre variables, se empleará el coeficiente de correlación de Pearson, calculado a escala diaria sobre pares de series temporales con datos disponibles.

A partir de la matriz de correlaciones resultante, se identificarán posibles relaciones fuertes entre variables explicativas, con el objetivo de detectar redundancias o fenómenos de multicolinealidad. En caso de observarse correlaciones elevadas entre predictores, se evaluará la conveniencia de eliminar o consolidar algunas variables.

Este procedimiento permitirá reducir la dimensionalidad efectiva del conjunto de datos, simplificar la estructura del modelo y potencialmente mejorar su capacidad de generalización.

4.4. Normalización de Datos

Con el objetivo de asegurar la comparabilidad entre variables y mejorar el desempeño de los modelos utilizados en los análisis posteriores, se aplicó un proceso de normalización a las variables consideradas en este estudio. Salvo que se indique explícitamente lo contrario en capítulos o secciones posteriores, todas las variables fueron transformadas mediante el *Standard Scaler* en su versión estándar.

El método *Standard Scaler* consiste en reescalar cada variable para que presente media cero y desviación estándar igual a uno. Formalmente, para cada observación x de una variable X , la transformación aplicada es:

$$x' = \frac{x - \mu}{\sigma}$$

donde μ es la media muestral de la variable y σ su desviación estándar. Esta normalización es adecuada cuando se trabaja con modelos sensibles a la escala de las variables, ya que garantiza que todas contribuyan de manera comparable evitando que aquellas con mayor rango o varianza dominen el proceso de ajuste o entrenamiento.

Es importante remarcar que el proceso de ajuste del normalizador (*fit*) se realizó **únicamente sobre el conjunto de entrenamiento**, con el fin de evitar filtración de información (*data leakage*). Posteriormente, la transformación correspondiente (*transform*) se aplicó tanto al conjunto de entrenamiento como a los conjuntos de validación y prueba. De esta manera, todos los subconjuntos mantienen la misma escala sin incorporar información externa al proceso de entrenamiento del modelo.

Este procedimiento se aplicó de manera consistente en todo el análisis, salvo en aquellos casos donde, por razones metodológicas o interpretativas descritas oportunamente, se requirió utilizar una transformación alternativa o preservar los valores originales.

Capítulo 5

Desarrollo y evaluación de modelos

En este capítulo se describe la metodología adoptada para el desarrollo, calibración y evaluación comparativa de los modelos considerados en este trabajo, con especial énfasis en la implementación y análisis de *Physics-Informed Neural Networks* (PINNs) como propuesta central de la tesis.

En primer lugar, se presenta la estrategia de partición cronológica de los datos, garantizando independencia temporal entre los conjuntos de entrenamiento, validación y prueba, con el objetivo de obtener una estimación realista del desempeño predictivo. A continuación, se introducen las métricas utilizadas para evaluar el comportamiento hidrológico de los modelos, junto con criterios de análisis cualitativo, evaluación del rendimiento computacional y tratamiento de valores faltantes.

Posteriormente, se incorpora un enfoque de interpretabilidad basado en atribución mediante *Integrated Gradients*, que permite analizar la contribución de las variables de entrada en los modelos neuronales y examinar la coherencia hidrológica de las predicciones.

Seguidamente, se describen las estrategias de calibración y optimización empleadas, incluyendo técnicas evolutivas y métodos de optimización bayesiana para la selección de hiperparámetros. Se presenta además la infraestructura de cómputo utilizada para la ejecución de los experimentos, necesaria para la realización de entrenamientos intensivos y búsquedas extensas en el espacio de configuraciones.

Finalmente, se detallan las metodologías específicas aplicadas a cada modelo. El modelo conceptual GR4J se adopta como línea base de referencia, mientras que arquitecturas neuronales convencionales (MLP y RNN) se implementan como puntos de comparación dentro del marco *data-driven*. Sobre esta base, se desarrolla la formulación *physics-informed*, integrando restricciones físicas en la función de pérdida y evaluando su impacto en la capacidad predictiva y en la consistencia hidrológica de los resultados.

5.1. Estrategia de entrenamiento y evaluación

5.1.1. Partición de datos

El conjunto de datos se dividió siguiendo una práctica habitual en el entrenamiento de modelos de aprendizaje automático. Se reserva un **80 %** de los registros totales para el proceso de entrenamiento, mientras que el **20 %** restante se destina exclusivamente a la evaluación final del desempeño (*conjunto de prueba*).

Dentro del subconjunto de entrenamiento, se aplica una subdivisión interna en proporción **70/30**, asignando un **70 %** de los datos al entrenamiento propiamente dicho y un **30 %** al conjunto de validación. Esta última partición permite ajustar los hiperparámetros y monitorear la capacidad de generalización del modelo durante el proceso de optimización, evitando el sobreajuste.

La partición se realiza de forma **cronológica**, manteniendo el orden temporal de las observaciones, de modo que los conjuntos de validación y prueba corresponden siempre a períodos posteriores a los utilizados en el entrenamiento.

5.1.2. Métricas de evaluación del desempeño

En todos los casos se calculan métricas de desempeño sobre los conjuntos de validación y prueba. En las ecuaciones presentadas, $Q_{\text{obs}}^{(i)}$ representa el caudal observado en el instante i , $Q_{\text{pred}}^{(i)}$ el caudal estimado por el modelo, \bar{Q}_{obs} la media de los caudales observados, $e_i = Q_{\text{obs}}^{(i)} - Q_{\text{pred}}^{(i)}$ el error en el instante i , y N el número total de observaciones consideradas.

- **Error cuadrático medio (RMSE)**

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N e_i^2}$$

El RMSE penaliza en mayor medida los errores grandes debido al término cuadrático. Se expresa en las mismas unidades que el caudal (m^3/s), su rango es $[0, +\infty)$ y su valor óptimo es 0.

- **Error absoluto medio (MAE)**

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |e_i|$$

El MAE mide la magnitud promedio del error sin penalización cuadrática. Se expresa en m^3/s y su valor óptimo es 0.

- **Eficiencia de Nash–Sutcliffe (NSE)**

$$\text{NSE} = 1 - \frac{\sum_{i=1}^N e_i^2}{\sum_{i=1}^N (Q_{\text{obs}}^{(i)} - \bar{Q}_{\text{obs}})^2}$$

Propuesta por [Nash y Sutcliffe \(1970\)](#), la NSE es una métrica adimensional cuyo rango teórico es $(-\infty, 1]$. Valores iguales a 1 indican ajuste perfecto; valores iguales a 0 implican desempeño equivalente al uso de la media observada; valores negativos reflejan desempeño inferior a dicha referencia.

- **Sesgo porcentual (PBIAS)**

$$\text{PBIAS} = 100 \times \frac{\sum_{i=1}^N e_i}{\sum_{i=1}^N Q_{\text{obs}}^{(i)}}$$

El PBIAS cuantifica la tendencia promedio del modelo a sobreestimar o subestimar el caudal observado ([Moriassi y cols., 2007](#)). Su valor óptimo es 0. Valores positivos indican subestimación y valores negativos sobreestimación.

La interpretación del desempeño según NSE y PBIAS se realiza conforme a [Moriassi y cols. \(2015\)](#), como se muestra en la Tabla 5.1. Durante la optimización se busca maximizar NSE y minimizar $|\text{PBIAS}|$ en validación, mientras que la comparación final entre modelos se realiza exclusivamente sobre el conjunto de prueba.

Tabla 5.1: Clasificación del desempeño para simulaciones de caudal según NSE y PBIAS (adaptado de [Moriassi y cols. \(2015\)](#)).

Clasificación	NSE	$ \text{PBIAS} $
Muy bueno	$0.75 < \text{NSE} \leq 1.00$	$< 10\%$
Bueno	$0.65 < \text{NSE} \leq 0.75$	$10\% - 15\%$
Satisfactorio	$0.50 < \text{NSE} \leq 0.65$	$15\% - 25\%$
Insatisfactorio	$\text{NSE} \leq 0.50$	$> 25\%$

5.1.3. Evaluación del rendimiento computacional

Además del desempeño predictivo, se registra el tiempo requerido para el entrenamiento y la evaluación en cada conjunto. Este análisis permite comparar la eficiencia computacional entre arquitecturas, considerando que modelos con desempeño hidrológico similar pueden diferir en complejidad y costo temporal.

5.1.4. Análisis cualitativo del desempeño

Se realiza una inspección visual de las series temporales de caudal observado y simulado, evaluando la capacidad del modelo para reproducir tendencias generales, períodos secos y húmedos, eventos extremos y fases de recesión.

5.1.5. Tratamiento de valores faltantes

Las instancias sin observaciones disponibles no se utilizan ni para entrenamiento ni para evaluación. En consecuencia, todas las métricas se calculan únicamente sobre registros válidos.

En el caso de modelos PINN, se incorporan términos adicionales en la función de pérdida que permiten integrar información física en puntos sin observación directa.

5.2. Interpretabilidad basada en atribución

Además del desempeño agregado, resulta relevante comprender qué variables impulsan una predicción, con el fin de evaluar consistencia hidrológica y detectar posibles sesgos.

El método *Integrated Gradients* (IG), propuesto por [Sundararajan y cols. \(2017\)](#), asigna a cada variable de entrada una contribución asociada a una predicción específica:

$$\text{IG}_i(\mathbf{x}) = (x_i - x'_i) \int_0^1 \frac{\partial F(\mathbf{x}' + \alpha(\mathbf{x} - \mathbf{x}'))}{\partial x_i} d\alpha$$

donde $F(\cdot)$ representa la predicción del modelo, \mathbf{x}' es un punto de referencia (*baseline*) respecto al cual se calcula la atribución, y $\alpha \in [0, 1]$ es una variable de integración que define el camino entre \mathbf{x}' y \mathbf{x} . En aplicaciones hidrológicas, IG permite analizar la sensibilidad local del modelo a las variables de entrada, tales como la precipitación y otras variables meteorológicas, así como a la información contenida en estados previos de la serie temporal.

En el caso de los modelos recurrentes, la entrada corresponde a una secuencia $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_T\}$. Si bien el método permite calcular atribuciones para cada variable en cada instante temporal de la secuencia, en esta implementación las *Integrated Gradients* se evaluaron únicamente respecto al primer paso temporal de la ventana de entrada, el cual corresponde al instante más reciente disponible para la predicción, manteniendo constantes las restantes componentes de la secuencia.

En consecuencia, las contribuciones obtenidas mediante IG reflejan la sensibilidad de la predicción respecto a las variables en el instante más reciente de la ventana, sin descomponer explícitamente las contribuciones asociadas a los pasos temporales anteriores dentro de la secuencia.

5.3. Optimización y AutoML

5.3.1. Calibración del modelo mediante NSGA-II

La calibración del modelo hidrológico se formuló como un problema de optimización monobjetivo, consistente en la minimización de una función de desempeño definida sobre el conjunto de validación.

Para la exploración del espacio paramétrico se empleó el algoritmo evolutivo NSGA-II (Deb y cols., 2002). Si bien NSGA-II fue originalmente propuesto para problemas de optimización multiobjetivo, en este trabajo se utilizó en una configuración de objetivo único.

La métrica específica utilizada como función objetivo, así como los parámetros considerados en el proceso de calibración, se detallan en los capítulos correspondientes a cada experimento.

5.3.2. Optimización de hiperparámetros con Optuna

La optimización de hiperparámetros se planteó como la minimización de la función de pérdida (*loss*) evaluada sobre el conjunto de validación.

Se utilizó la librería Optuna (Akiba y cols., 2019), empleando el algoritmo `TPESampler` como estrategia de muestreo. Este método, basado en estimación de densidades mediante árboles de Parzen, modela de forma adaptativa la distribución de configuraciones prometedoras y guía la búsqueda hacia regiones con mayor probabilidad de mejora.

Asimismo, se utilizaron los mecanismos de *pruning* incorporados en Optuna para descartar tempranamente configuraciones con bajo desempeño intermedio, reduciendo el costo computacional del proceso de optimización.

Los hiperparámetros optimizados y sus respectivos rangos de búsqueda se especifican en las secciones metodológicas particulares de cada modelo.

5.4. Infraestructura computacional

Los experimentos se ejecutaron en el **Cluster-UY**, infraestructura nacional de cómputo científico de alto desempeño en Uruguay (Nesmachnow y Iturriaga, 2019), adecuada para entrenamientos intensivos y campañas extensas de optimización.

5.5. GR4J

El modelo hidrológico conceptual **GR4J** se adopta como modelo base para la representación del balance hídrico diario en la cuenca de estudio. Este modelo, propuesto por Perrin y cols. (2003), describe el ciclo hidrológico a través de cuatro parámetros principales $\{x_1, x_2, x_3, x_4\}$ que controlan los procesos de almacenamiento, intercambio subterráneo y enrutamiento del agua en la cuenca.

En esta investigación, el modelo GR4J cumple un doble propósito: (i) servir como **línea base** para la evaluación del desempeño hidrológico, y (ii) proveer la función de simulación de caudal que será utilizada en las secciones posteriores del trabajo en el contexto de **PINNs**.

El modelo fue implementado siguiendo estrictamente las ecuaciones y el esquema conceptual presentados en Perrin y cols. (2003), incluyendo el cálculo

diario de la precipitación y evapotranspiración efectivas, la contabilidad del almacenamiento de humedad del suelo, la percolación, la generación de escorrentía rápida y lenta, y el enrutamiento del caudal mediante hidrogramas unitarios.

Evolución interna y condiciones de simulación El modelo GR4J posee un **sistema interno de estados** que se actualiza de forma secuencial en cada paso temporal. En cada iteración, además del caudal simulado, el modelo actualiza variables internas asociadas al contenido de humedad del suelo, los almacenamientos de enrutamiento y los flujos de intercambio subterráneo.

Para garantizar la coherencia física de estas variables internas, es necesario cumplir dos condiciones durante la simulación:

1. **Continuidad temporal:** la serie de entrada debe mantenerse continua y sin huecos. Por esta razón, aunque los días con caudal observado faltante no se utilicen para el cálculo de las métricas de evaluación, el modelo se ejecuta igualmente sobre dichos registros. De esta manera, al disponer de los valores de precipitación y evapotranspiración, el sistema interno puede evolucionar de forma consistente a lo largo de todo el período de simulación.
2. **Período de calentamiento (*warm-up*):** dado que las variables internas del modelo no son conocidas al inicio de la simulación, se requiere un período inicial de estabilización para que los almacenamientos y flujos alcancen un régimen representativo. Con este fin, se reservaron los primeros **365 días** del conjunto de datos como período de *warm-up*. Dicho intervalo se incluye dentro del 80 % de los datos utilizados para la calibración, asegurando que el ajuste de parámetros se realice sobre un sistema hidrológicamente estabilizado.

Configuración y ajuste de parámetros

Para la definición de los parámetros $\{x_1, x_2, x_3, x_4\}$ del modelo GR4J se consideraron dos configuraciones complementarias:

1. **Configuración referencial:** se utilizaron los valores reportados por [Narbondo y cols. \(2020\)](#) correspondientes a la subcuenca del río Santa Lucía Chico, cuyo punto de cierre coincide con el de la subcuenca considerada en este trabajo. En dicho estudio, los autores presentan una calibración individual del modelo GR4J cuyos parámetros son:

$$x_1 = 101 \text{ mm}, \quad x_2 = -2.5 \text{ mm}, \quad x_3 = 49 \text{ mm}, \quad x_4 = 2.3 \text{ días.}$$

2. **Configuración optimizada:** se realizó un proceso de optimización de parámetros empleando la librería `Optuna` con el algoritmo evolutivo NSGA-II. Este proceso consistió en un total de **2000 trials**, utilizando como punto de partida los valores de [Narbondo y cols. \(2020\)](#), y definiendo rangos de búsqueda del $\pm 50 \%$ alrededor de cada uno de los parámetros

reportados. Si bien los rangos de búsqueda exceden ligeramente los intervalos de confianza al 80 % reportados en la Tabla 4 de [Narbondo y cols. \(2020\)](#) — $x_1 \in [100, 1200]$ mm, $x_2 \in [-5, 3]$ mm/d, $x_3 \in [20, 300]$ mm y $x_4 \in [1.1, 2.9]$ d— en los parámetros x_1 (límite inferior: 50.5 vs. 100) y x_4 (límite superior: 3.45 vs. 2.9), se verificó que los **valores óptimos obtenidos** se encuentran dentro de dichos intervalos de confianza. La función objetivo de la optimización fue maximizar el coeficiente de eficiencia de Nash–Sutcliffe (NSE) en el conjunto de calibración.

Ambas configuraciones fueron evaluadas en términos de su capacidad para reproducir el caudal observado, comparando el desempeño mediante métricas hidrológicas estándar. La configuración que muestre el mejor desempeño en el conjunto de validación será adoptada como **baseline definitivo**, y sus parámetros se utilizarán en la formulación del modelo GR4J dentro de las secciones de PINNs desarrolladas posteriormente. Recordemos que el conjunto de validación, al igual que el de entrenamiento y el de prueba, es el mismo que en el resto de modelos (definido en la sección [5.1.1](#)).

Integración de los datos de precipitación y evapotranspiración

El modelo GR4J requiere como entrada una única serie de precipitación representativa de toda la subcuenca. Dado que en este estudio se dispone de registros provenientes de múltiples estaciones pluviométricas, fue necesario integrar espacialmente la información disponible.

La construcción de la precipitación media areal se realizó siguiendo el procedimiento descrito en la Sección [3.3](#) (Relaciones espaciales entre estaciones), donde se definieron los pesos areales mediante el método de los **polígonos de Thiessen** recortados al contorno de la subcuenca. En consecuencia, la serie diaria utilizada como entrada en GR4J corresponde al promedio ponderado de las estaciones consideradas, según los coeficientes allí establecidos.

Por otro lado, el modelo GR4J requiere como segunda entrada una serie de **evapotranspiración potencial (ETP)**. En este caso, no fue necesario realizar ningún procedimiento adicional de integración espacial, dado que se dispone de una única serie diaria correspondiente a la estación **INIA Las Brujas**, la cual se consideró representativa de las condiciones atmosféricas regionales para el período de estudio.

5.6. MLP

En esta sección se describe el procedimiento seguido para el desarrollo y ajuste del modelo de predicción de caudales mediante redes neuronales de tipo MLP. El objetivo es presentar el flujo metodológico aplicado, sin entrar en los resultados obtenidos, los cuales serán abordados en un capítulo posterior.

5.6.1. Desarrollo del modelo

El desarrollo metodológico se inició empleando como insumo todas las variables descriptas previamente en el Capítulo 4, a fin de contar con un punto de partida amplio sobre el cual aplicar las técnicas posteriores de selección, generación y transformación de variables.

Para la arquitectura de la red se optó, tras pruebas exploratorias con configuraciones básicas, por una estructura fija que resultara suficientemente flexible y con un desempeño inicial aceptable. La finalidad de esta decisión fue delimitar un modelo estable sobre el cual concentrar los esfuerzos en la etapa de *feature engineering*, evitando que las modificaciones en la arquitectura interfieran con la evaluación del impacto de las variables de entrada.

La arquitectura adoptada corresponde a un perceptrón multicapa (MLP) con tres capas ocultas de 256, 128 y 64 neuronas, respectivamente, cada una con función de activación ReLU. Se incorporaron mecanismos de regularización mediante *dropout*, con tasas del 10 % y 5 % en las dos primeras capas ocultas, respectivamente. La capa de salida se definió sin activación, para producir directamente una predicción escalar del caudal. Se emplearon lotes de tamaño 32 y un máximo de 200 épocas, con criterio de *early stopping* y paciencia de 20 épocas.

El entrenamiento se realizó con el optimizador Adam, utilizando una tasa de aprendizaje inicial de 10^{-4} y tomando como función de pérdida el error cuadrático medio (MSE), estándar en problemas de regresión.

Cabe destacar que esta arquitectura se mantuvo invariante durante las fases de incorporación y selección de variables. La optimización de la arquitectura en sí misma, junto con otros hiperparámetros, se abordó posteriormente mediante un procedimiento de AutoML con la librería Optuna, descrito en la Sección 5.6.4.

5.6.2. Incorporación de variables

Variables temporales El conjunto de datos utilizado presenta cada observación indexada por su fecha correspondiente; sin embargo, dicha información no se encontraba representada explícitamente como una variable numérica interpretable por el modelo. Para incorporar información temporal de manera efectiva, se descompuso la fecha en sus componentes básicos (día, mes y año) y se generaron variables derivadas a partir de ellos.

En el caso del *año*, se decidió no incorporarlo como variable explicativa debido a que se trata de un valor estrictamente creciente que únicamente refleja el orden cronológico de las observaciones. Desde la perspectiva de una red neuronal, este número no posee significado estructural intrínseco más allá de su magnitud relativa; es decir, el modelo lo interpretaría simplemente como un escalar creciente sin información semántica adicional. En consecuencia, su inclusión podría inducir al aprendizaje de relaciones asociadas al paso del tiempo como tendencia numérica, más que a patrones sustantivos presentes en los datos.

Por otro lado, las variables *día* y *mes* presentan patrones estacionales relevantes. Sin embargo, representarlas como valores numéricos directos (1–31 y

1–12) introduce una falsa linealidad: por ejemplo, diciembre (12) queda numéricamente distante de enero (1), a pesar de ser meses consecutivos; de manera análoga, el día 1 y el día 31 no son interpretados como contiguos. Para evitar esta inconsistencia y preservar la naturaleza periódica de estas variables, se empleó una codificación trigonométrica basada en funciones seno y coseno, estrategia comúnmente utilizada para representar características cíclicas en modelos de aprendizaje automático (Mahajan y cols., 2021).

Para un componente genérico x que toma valores enteros en un ciclo de longitud T , se definen dos nuevas variables como:

$$x_{\sin} = \sin\left(2\pi \frac{x}{T}\right), \quad x_{\cos} = \cos\left(2\pi \frac{x}{T}\right).$$

En este trabajo, el día del mes se normaliza utilizando $T = 31$ como período de referencia. Si bien la duración real de los meses varía entre 28 y 31 días, la adopción de un período fijo permite mantener una representación estable del ciclo intramensual, introduciendo una distorsión mínima que no afecta la continuidad ni la periodicidad necesarias para capturar patrones estacionales. De este modo, por cada componente temporal considerado (día y mes), se generan dos variables adicionales correspondientes a las transformaciones seno y coseno.

Adicionalmente, con el objetivo de evaluar si la incorporación de estas características temporales mejora efectivamente el desempeño predictivo, se llevarán a cabo pruebas experimentales sobre el conjunto de validación. En particular, se compararán cuatro configuraciones: (i) no incluir variables temporales; (ii) incluir únicamente las transformaciones cíclicas del día; (iii) incluir únicamente las del mes; y (iv) incluir ambas variables simultáneamente. Cada configuración será evaluada mediante las métricas seleccionadas, lo que permitirá determinar empíricamente el aporte real de estas variables y discutir posteriormente cuál estrategia resulta más adecuada.

En síntesis, las variables temporales utilizadas en este trabajo se limitan a las transformaciones cíclicas del día y del mes, excluyendo el año por su falta de variabilidad predictiva. Asimismo, se evaluará de forma sistemática su impacto en el rendimiento del modelo para fundamentar su inclusión o descarte en la versión final.

Representación espacial de la precipitación Un segundo aspecto considerado fue la representación espacial de la precipitación. Dado que el *dataset* incluye registros de múltiples estaciones meteorológicas distribuidas en la región de estudio, es posible que distintas estaciones presenten valores similares, aportando información parcialmente redundante que podría introducir ruido o colinealidad en el modelo.

Para abordar esta situación, se definió un análisis comparativo entre dos enfoques: (i) mantener las precipitaciones de cada estación como variables independientes, y (ii) aplicar métodos de agregación espacial.

En este segundo enfoque se evaluaron tres alternativas específicas:

1. **Promedio simple** entre estaciones;

2. **Promedio ponderado por distancia** al punto de control de caudal, utilizando las distancias definidas en la Sección 3.3;
3. **Agregación mediante polígonos de Thiessen**, empleando los pesos areales obtenidos en la Sección 3.3.

De esta forma, se buscó establecer si una representación agregada de la precipitación proporciona una descripción más parsimoniosa (es decir, con menor redundancia de variables) y robusta del forzante hidrológico en comparación con el uso de estaciones individuales.

Inclusión de lags Finalmente, se consideró la incorporación de variables que representaran valores pasados de las series, comúnmente denominadas *lags*. En el contexto de modelos de series temporales, los *lags* constituyen una forma estándar de enriquecer el conjunto de variables, ya que permiten que el modelo utilice información de observaciones anteriores para mejorar la predicción del valor actual.

En este trabajo se generaron *lags* únicamente para la precipitación, con el objetivo de capturar la influencia que las lluvias de días previos pueden ejercer sobre el caudal observado en el presente. La incorporación de estos valores retrasados permite al modelo representar de manera más realista los procesos de respuesta de la cuenca frente a los eventos de precipitación y la persistencia temporal de sus efectos, ya que estos *lags* actúan como un *proxy* (aproximación indirecta) del estado de humedad del suelo, una variable clave que condiciona la generación de escorrentía.

Si bien inicialmente se consideró incluir también *lags* del caudal, esta opción fue descartada debido a la presencia de valores faltantes en dicha serie, tal como se analizó en el Capítulo 4. Dado que el caudal constituye la variable objetivo, se decidió no imputar los valores faltantes para evitar introducir sesgos o inconsistencias. En consecuencia, no es posible construir *lags* de forma consistente, ya que el valor inmediatamente anterior al observado podría corresponder, según el caso, al día anterior o incluso a fechas mucho más alejadas en el tiempo. Esta discontinuidad generaría una estructura temporal heterogénea y dificultaría la interpretación del modelo.

Por ello, se optó por restringir el análisis de *lags* a la variable de precipitación, cuya serie se encuentra completa y permite evaluar de manera más controlada el impacto de la memoria temporal sobre el desempeño predictivo del modelo.

En primer lugar, se planteó un experimento preliminar con el objetivo de evaluar si la incorporación de *lags* de precipitación contribuye efectivamente a mejorar el desempeño del modelo. Para ello, se fijó un horizonte arbitrario de **7 días de lag**, considerando que este valor es suficientemente amplio para capturar la respuesta retardada de la cuenca, en línea con el horizonte de acumulación empleado por Vilaseca y cols. (2023) como *proxy* de la condición de humedad antecedente del suelo.

Este experimento inicial se aplicó bajo distintos esquemas de representación espacial de la precipitación, a fin de determinar cuál de ellos resulta más ade-

cuado cuando se incluyen variables retardadas. Se compararon las siguientes configuraciones:

1. **Estaciones individuales:** se calcularon *lags* para cada una de las estaciones de precipitación de manera independiente.
2. **Promedio simple:** se calculó la precipitación media diaria entre todas las estaciones antes de generar los *lags*.
3. **Promedio ponderado por distancia:** se asignaron pesos inversamente proporcionales a la distancia al punto de control, según lo definido en la Sección 3.3.
4. **Polígonos de Thiessen:** se ponderó la precipitación según los pesos areales derivados de los polígonos recortados a la subcuenca (Sección 3.3).

El propósito de esta primera etapa fue determinar si la inclusión de *lags* resulta beneficiosa para el modelo y, en caso afirmativo, qué esquema de agregación espacial ofrece el mejor desempeño.

Una vez identificado el esquema espacial más adecuado, se procedió a una segunda etapa en la que se evaluó el **número óptimo de días de lag**. En esta fase, se mantuvo el esquema de agregación seleccionado previamente y se realizaron pruebas con distintos horizontes de retardo (de 1 a 14 días), con el objetivo de cuantificar en qué medida la memoria temporal de la precipitación mejora las predicciones del caudal y hasta qué punto la inclusión de más días aporta información relevante o introduce redundancia temporal.

5.6.3. Selección de variables

Si bien las variables definidas en las secciones previas son plausibles desde el punto de vista del dominio y de las series temporales, su efecto neto sobre la capacidad predictiva del modelo no es evidente *a priori*. Por este motivo, se estableció un procedimiento explícito para estimar la contribución y relevancia de cada variable, evitando depender exclusivamente de comparaciones directas de desempeño (p. ej., “mejora/no mejora” en una métrica), dado que tales comparaciones pueden resultar poco robustas frente a la variabilidad muestral o a interacciones entre características.

Protocolo general. La selección se estructuró en dos etapas complementarias y sucesivas: primero se evaluó la redundancia estadística entre variables, y luego se analizó su impacto efectivo sobre la predicción mediante *Integrated Gradients* (IG).

1. **Análisis de correlación.** En primer lugar, se aplicó el análisis de correlación desarrollado en la Sección 4.3, con el objetivo de identificar pares de variables altamente correlacionadas. En esta etapa se propusieron remociones de variables redundantes, priorizando mantener aquellas con mayor

respaldo conceptual o disponibilidad de datos. Cada eliminación fue seguida por una reevaluación del modelo para verificar que la supresión no produjera una degradación significativa en las métricas de validación.

2. **Análisis de relevancia mediante *Integrated Gradients*.** Una vez depurado el conjunto inicial por correlación, se estimaron las atribuciones de relevancia mediante *Integrated Gradients* (IG), un método de interpretabilidad basada en gradientes que asigna a cada entrada una contribución proporcional al cambio en la predicción al recorrer una trayectoria desde una entrada de referencia hasta la entrada observada. IG permite estimar la importancia relativa de cada variable para la predicción del modelo, preservando la estructura de la red y evitando modificaciones *ad hoc* del conjunto de características.

Sobre un modelo entrenado con la arquitectura base (Sección 5.6.1), se calcularon las atribuciones de IG para cada variable y se obtuvo un *ranking* de relevancia. Las atribuciones se interpretaron en términos de magnitud total (impacto absoluto sobre la predicción), dado que el signo únicamente indica la dirección de influencia (positiva o negativa) y no la calidad de la contribución.

Se implementó un esquema de depuración iterativa en el que se identificaron y eliminaron variables con bajo impacto medio (atribuciones de pequeña magnitud) o con atribuciones inestables entre los conjuntos de entrenamiento y validación. En cada iteración, el modelo se reentrenó y se recalcularon las atribuciones de IG, hasta alcanzar estabilidad en las relevancias o en el desempeño de validación.

El análisis de IG se realizó de manera separada sobre los subconjuntos de entrenamiento y validación, correspondientes respectivamente al 56 % (70 % del 80 %) y al 24 % (30 % del 80 %) de los datos originales. El conjunto de prueba (20 %) no fue utilizado en esta etapa, a fin de evitar *data leakage* en la selección de variables. Esta evaluación diferenciada permitió examinar la consistencia de las atribuciones: si las variables más relevantes en entrenamiento también muestran alta relevancia en validación, ello sugiere que el modelo está capturando relaciones genuinas del proceso físico; por el contrario, discrepancias marcadas podrían indicar sobreajuste u otros sesgos en el aprendizaje.

Criterios de decisión. La inclusión o exclusión de variables se basó en:

1. La **redundancia por correlación** entre variables (según el análisis de la Sección 4.3),
2. La **magnitud media de las atribuciones de IG** (impacto total sobre la predicción),
3. La **consistencia de las atribuciones** entre los conjuntos de entrenamiento y validación.
4. La **estabilidad del desempeño del modelo**: las exclusiones se mantuvieron únicamente cuando no provocaron una degradación apreciable en las métricas de evaluación.

Se favorecieron conjuntos de características más compactos siempre que no implicaran una pérdida sustancial de desempeño ni una reducción en la capacidad explicativa del modelo. El proceso iterativo de depuración se detuvo cuando las atribuciones de IG se estabilizaron entre iteraciones o cuando la eliminación de una variable provocó una disminución en las métricas de validación.

5.6.4. Optimización de hiperparámetros

Con las variables seleccionadas y tomando la arquitectura como parte del espacio de búsqueda, se realizó una optimización de hiperparámetros con *Optuna*, siguiendo un esquema de búsqueda automatizada orientado a minimizar la pérdida de validación (MSE). La optimización se llevó a cabo exclusivamente sobre el 80% de los datos destinado al ajuste del modelo (con una división interna de 70% para entrenamiento y 30% para validación). El 20% restante se mantuvo completamente reservado y no intervino en este proceso, de manera que pudiera utilizarse posteriormente como *hold-out* final para evaluar el desempeño del modelo seleccionado en condiciones de independencia total. En todos los casos se preservó la secuencia temporal de la serie, es decir, sin aplicar *shuffle*. La función objetivo entrenó el modelo con *early stopping* y registró la mínima pérdida de validación observada en cada corrida.

Se llevó a cabo **un estudio** con `n_trials = 1000`. En cada *trial*, se muestreó una configuración dentro de los rangos mostrados en la Tabla 5.2, se entrenó con el optimizador Adam y pérdida MSE, y se devolvió la mínima `val_loss`. Se fijó una semilla global para asegurar reproducibilidad del proceso completo. Además, se estableció un criterio de *early stopping global* de 100 *trials*: si no se observaba mejora en la pérdida de validación durante 100 iteraciones consecutivas, el proceso de optimización se detenía de manera anticipada.

Tabla 5.2: Espacio de búsqueda y tipo de muestreo en Optuna.

Bloque	Hiperparámetro	Rango / Opciones	Muestreo
Arquitectura	<code>n_layers</code>	[1, 5]	entero
	<code>width_l{i}</code>	[32, 1024]	entero (escala log)
	<code>activation</code>	{ReLU, Leaky ReLU, Tanh, SiLU, GELU}	categorico
	<code>dropout_base</code>	[0.0, 0.5]	continuo
Entrenamiento	<code>lr</code>	$[10^{-5}, 10^{-2}]$	continuo (escala log)
	<code>weight_decay</code>	$[10^{-8}, 10^{-2}]$	continuo (escala log)
	<code>batch_size</code>	{16, 32, 64, 128, 256}	categorico
	<code>epochs</code>	200	fijo
	<code>patience</code>	20	fijo
Validación	<i>split</i> temporal	56/24/20 (sin <i>shuffle</i>)	fijo

Detalles de implementación. La arquitectura candidata se construyó dinámicamente a partir de: número de capas ocultas, ancho por capa (en escala logarítmica) y elección de activación. El término de *dropout* se aplicó con un

esquema escalonado $p_i = \text{dropout_base} \cdot 0.75^{(i-1)}$, decreciendo a medida que avanza la profundidad. El optimizador Adam se inicializó con la tasa de aprendizaje muestreada y se inyectó *weight decay* directamente en los *param groups* del optimizador. La función de pérdida utilizada fue MSE y la métrica reportada para selección fue la mínima `val_loss` a lo largo del entrenamiento con *early stopping* (criterio de paciencia muestreado).

Adicionalmente, se empleó un *pruner* del tipo `MedianPruner`, configurado con `n_startup_trials=10`, `n_warmup_steps=20` e `interval_steps=5`. Este mecanismo permitió descartar de forma temprana aquellas configuraciones con desempeño por debajo de la mediana acumulada, reduciendo así el tiempo total de optimización al enfocar los recursos de cómputo en configuraciones más prometedoras.

Justificación de los rangos.

- **Profundidad y ancho** ($n_layers \in [1, 5]$, $\text{width_li} \in [32, 1024]$ en escala log): para datos tabulares, MLPs poco profundos a moderados suelen ser suficientes; el muestreo logarítmico del ancho favorece explorar órdenes de magnitud (de modelos compactos a medianos) sin sesgar hacia valores grandes.
- **Activaciones** (ReLU, Leaky ReLU, Tanh, SiLU, GELU): cubren funciones estándar con buen respaldo empírico en tabular; ReLU/Leaky ReLU por su estabilidad y eficiencia, Tanh por saturación controlada, y SiLU/GELU por suavidad y buenos resultados prácticos en redes modernas.
- **Dropout** [0.0; 0.5]: rango clásico en regularización para MLPs; incluye el caso sin *dropout* y valores moderados donde suele observarse mejor generalización.
- **Learning rate** [10^{-5} , 10^{-2}] (log): intervalo típico para Adam en regresión; la escala log facilita encontrar órdenes de magnitud adecuados.
- **Weight decay** [10^{-8} , 10^{-2}] (log): abarca desde regularización casi nula hasta niveles moderados; útil para controlar complejidad efectiva.
- **Batch size** {16, 32, 64, 128, 256}: potencias de dos habituales, equilibrando estabilidad de gradiente, tiempo por *epoch* y uso de memoria.

Procedimiento. Para cada *trial* se siguieron los siguientes pasos: (i) muestreo de una configuración de hiperparámetros, (ii) construcción de la arquitectura candidata, (iii) entrenamiento real sobre el 70 % del conjunto de entrenamiento reservado (80 % del conjunto total definido), y el restante 30 % de validación, (iv) monitoreo de la función de pérdida en validación (`val_loss`) con *early stopping* y criterio de paciencia muestreado, y (v) registro de la mínima `val_loss` alcanzada durante el entrenamiento.

La selección final de hiperparámetros se realizó escogiendo la configuración que minimizó la pérdida en el conjunto de validación. El bloque del 20 % reservado se utilizó exclusivamente en etapas posteriores como referencia independiente de desempeño, sin intervenir en la búsqueda de hiperparámetros.

5.7. RNN

En esta sección se describe el procedimiento seguido para el desarrollo y ajuste del modelo recurrente (RNN) aplicado a la predicción de caudales. Al igual que en el caso MLP (Sección 5.6.1), la estrategia metodológica buscó aislar el impacto de las decisiones de *feature engineering* respecto de las de arquitectura. Para ello, el trabajo se organizó en cuatro etapas secuenciales y acotadas: (i) fijar una arquitectura base de RNN con desempeño inicial aceptable; (ii) definir y congelar un conjunto de variables “base” para el entrenamiento; (iii) incorporar, de forma controlada, variables adicionales y evaluar la contribución global de todas las variables (tanto las base como las incorporadas) mediante análisis de correlación de las variables e *Integrated Gradients* (IG); y (iv) recién al final, optimizar la arquitectura y los hiperparámetros sobre el espacio de búsqueda correspondiente. Este flujo permitió comparar de manera consistente los efectos de las variables de entrada, reduciendo la confusión entre cambios en los insumos y cambios en la arquitectura.

5.7.1. Desarrollo del modelo

El modelo se desarrolló bajo un enfoque estructural híbrido **recurrente–denso (RNN → MLP)**, mantenido a lo largo de todo el trabajo. En este esquema, las capas recurrentes se encargan de capturar la dinámica temporal de las variables de entrada, mientras que las capas densas actúan como bloque final de proyección no lineal hacia la salida escalar. Este orden (primero la codificación temporal, luego la proyección densa) se preservó en todas las variantes y optimizaciones posteriores del modelo.

Bloque recurrente (codificador temporal). El bloque recurrente estuvo constituido por una red del tipo LSTM unidireccional, adecuada para predicción causal en tiempo real (la red sólo accede a información pasada). Este bloque se configuró con dos capas apiladas y 64 unidades ocultas por capa, aplicando *dropout* del 20 % entre capas para mejorar la generalización. Formalmente, cada secuencia de entrada $\mathbf{X} \in \mathbb{R}^{(\text{batch}, \text{seq_len}, \text{input_size})}$ se procesa paso a paso, y sólo la salida correspondiente al último instante temporal se conserva como vector representativo del estado global de la secuencia:

$$\mathbf{z} = \text{LSTM}(\mathbf{X})_{\text{last}} \in \mathbb{R}^h,$$

donde h denota el tamaño oculto del bloque recurrente. Esta representación condensada sirve como entrada directa para la etapa densa. En los casos en que

las entradas se presentan como vectores sin dimensión temporal explícita, el modelo las expande automáticamente a secuencias de longitud uno, garantizando compatibilidad con la interfaz recurrente.

Longitud de ventana temporal. Inicialmente, el modelo se configuró con secuencias de longitud uno (ventana temporal igual a 1), lo que implica que cada predicción se basa únicamente en la información del instante más reciente. Posteriormente, y de manera análoga a la incorporación explícita de rezagos (*lags*) en el modelo MLP, se evaluó la utilización de ventanas temporales de mayor longitud, permitiendo que la red recurrente accediera a múltiples pasos anteriores dentro de cada muestra.

Se analizaron distintas longitudes de secuencia como hiperparámetro, con el objetivo de determinar si la incorporación de memoria temporal explícita mejoraba el desempeño predictivo. Esta exploración permitió comparar el efecto de ampliar el horizonte de información histórica manteniendo constante la estructura arquitectónica general del modelo.

Bloque denso (cabeza MLP). A partir del vector recurrente \mathbf{z} se aplica un bloque denso compuesto por tres capas lineales con activación ReLU y mecanismos de *dropout* decreciente. La estructura utilizada fue:

$$h \rightarrow 128 \xrightarrow{\text{ReLU+Dropout } 0.10} 64 \xrightarrow{\text{ReLU+Dropout } 0.05} 32 \xrightarrow{\text{ReLU}} 1,$$

donde la última capa es lineal y genera la predicción escalar del caudal. Este bloque cumple un rol análogo al del MLP utilizado en la arquitectura previa, actuando como proyector flexible sobre el espacio de salida.

Entrenamiento y configuración. El modelo completo (RNN + MLP) se entrenó con el optimizador Adam, tasa de aprendizaje inicial 10^{-4} y función de pérdida MSE. Se emplearon *batches* de tamaño 32 y un máximo de 200 épocas, con criterio de *early stopping* y paciencia de 20 épocas.

Razonamiento de diseño. La decisión de estructurar la red en dos bloques secuenciales (recurrente y denso) respondió a una motivación tanto conceptual como práctica. Desde el punto de vista conceptual, la RNN actúa como un codificador temporal que resume patrones dinámicos, dependencias y memorias de corto y mediano plazo, mientras que la MLP cumple la función de transformador no lineal que traduce esa representación latente en una predicción escalar interpretable. Desde el punto de vista metodológico, mantener fija esta arquitectura general permitió reproducir las comparaciones entre configuraciones de variables sin introducir cambios estructurales que alterarían la naturaleza del aprendizaje temporal. El *dropout* se utilizó de forma análoga a la empleada en la MLP (tasas del 10% y 5% en las dos primeras capas densas), buscando un equilibrio entre capacidad de generalización y estabilidad numérica.

Evolución del enfoque. Durante las fases posteriores de optimización (Sección 5.7.4), se exploraron distintas combinaciones de número de capas recurrentes, tamaño del estado oculto, tasas de *dropout* y tipo de celda (LSTM, GRU o RNN simple). Sin embargo, la estructura recurrente–densa se mantuvo invariable, constituyendo el núcleo arquitectónico del modelo en todas sus variantes. Este enfoque permitió conservar una base coherente para la comparación entre configuraciones y garantizar que los cambios de desempeño observados se debieran efectivamente a las modificaciones en la configuración interna y no a alteraciones en el flujo general de información dentro de la red.

5.7.2. Incorporación de variables

Variables temporales El conjunto de datos conserva la estructura temporal descrita previamente, con cada observación asociada a su fecha correspondiente. De manera análoga a lo implementado para el modelo MLP (Sección 5.6.2), se incorporaron únicamente las variables temporales *día* y *mes*, representadas mediante sus transformaciones cíclicas seno y coseno. Esta codificación permite reflejar adecuadamente la naturaleza periódica de estos componentes temporales sin introducir la falsa linealidad que produciría utilizarlos como valores enteros.

En esta etapa se evaluará el impacto de incluir dichas representaciones cíclicas (día, mes y la combinación de ambas) con el fin de determinar empíricamente si aportan información relevante para el desempeño del modelo recurrente.

Representación espacial de la precipitación En cuanto a la representación de la precipitación, se retomaron las mismas alternativas analizadas en la etapa MLP (precipitaciones individuales, promedio simple, promedio ponderado por distancia y polígonos de Thiessen) con el propósito de contrastar los resultados bajo un marco experimental equivalente. El objetivo no fue asumir que la mejor representación obtenida en el MLP conservaría su superioridad en el contexto de una red recurrente, sino verificar si la estructura secuencial del modelo modifica la sensibilidad frente a la forma de representar la información espacial de la lluvia.

Tratamiento de valores pasados (*lags*) A diferencia del enfoque adoptado en el modelo MLP, en el caso de la red neuronal recurrente (RNN) no se incorporan *lags* explícitos de las variables de entrada. Esta decisión se fundamenta en la propia naturaleza de la arquitectura recurrente, la cual dispone de un estado interno que le permite modelar dependencias temporales a partir de secuencias de datos, sin necesidad de incluir manualmente valores pasados como variables independientes.

En este contexto, el análisis de la información temporal se aborda mediante la definición de una **ventana temporal de entrada**, que determina la longitud de la secuencia utilizada por la RNN en cada paso de entrenamiento. Como configuración inicial, se considera una ventana temporal de longitud **1**, equivalente a un esquema sin recurrencia efectiva, lo que permite establecer una referencia base del desempeño del modelo.

Posteriormente, se evalúa el impacto de incrementar progresivamente la longitud de la ventana temporal, de forma análoga al análisis de *lags* realizado en el MLP, pero manteniendo la coherencia con la estructura recurrente del modelo. En particular, se experimenta con ventanas temporales de tamaño comprendido entre **1 y 14 días**, con el objetivo de identificar el horizonte temporal óptimo que maximiza el desempeño predictivo de la RNN.

Este procedimiento permite analizar hasta qué punto el modelo se beneficia de secuencias temporales más largas y determinar el compromiso entre la cantidad de información temporal disponible y la complejidad efectiva del aprendizaje. De este modo, la modelación de dependencias temporales se realiza exclusivamente a través del mecanismo recurrente, evitando la introducción de *lags* explícitos que podrían resultar redundantes con la memoria interna de la red.

Todas las configuraciones se entrenan manteniendo constantes la arquitectura recurrente–densa y las condiciones de entrenamiento, de manera que las diferencias observadas en el desempeño puedan atribuirse únicamente a la longitud de la ventana temporal utilizada.

5.7.3. Selección de variables

Si bien las variables definidas en las secciones previas constituyen un conjunto coherente desde el punto de vista del dominio y de la dinámica temporal, su aporte efectivo al desempeño del modelo recurrente no es necesariamente equivalente al observado en el MLP. Por este motivo, se implementó un procedimiento explícito de evaluación y selección de variables, con el objetivo de estimar su contribución individual y reducir redundancias, siguiendo el mismo enfoque general aplicado en el modelo MLP (Sección 5.6.3), pero aplicado en este caso a la arquitectura recurrente.

Protocolo general. El proceso se estructuró en dos etapas complementarias: primero se evaluó la redundancia estadística entre variables, y luego se estimó su relevancia mediante *Integrated Gradients* (IG).

1. **Análisis de correlación.** Se aplicó el mismo análisis de correlación descrito en la Sección 4.3, con el fin de identificar variables altamente correlacionadas o redundantes. En cada caso se priorizó conservar aquellas con mayor respaldo conceptual o disponibilidad de datos. Las remociones fueron seguidas por un reentrenamiento de la red y una verificación de que no se produjeran degradaciones significativas en las métricas de validación.
2. **Análisis de relevancia mediante *Integrated Gradients*.** Una vez depurado el conjunto inicial, se estimaron las atribuciones de relevancia utilizando el método de *Integrated Gradients* (IG), que permite cuantificar la contribución de cada variable a la predicción del modelo manteniendo la estructura de la red intacta. Sobre un modelo entrenado con la arquitectura base definida en la Sección 5.7.1, se calcularon las atribuciones de IG

en los conjuntos de entrenamiento y validación, y se construyó un *ranking* de relevancia a partir de las magnitudes promedio de dichas atribuciones.

Las magnitudes de IG se interpretaron como medida del impacto total de cada variable sobre la predicción, mientras que el signo se consideró únicamente como indicador de la dirección de influencia. Se implementó un proceso de depuración iterativo en el que se eliminaron variables con baja relevancia media o con atribuciones inestables entre los conjuntos de entrenamiento y validación. Tras cada eliminación, el modelo se reentrenó y se recalcularon las atribuciones, repitiendo el procedimiento hasta alcanzar estabilidad tanto en las métricas de validación como en las relevancias estimadas por IG.

Al igual que en el MLP, el análisis de variables se realizó únicamente sobre los subconjuntos de entrenamiento y validación (80 % del total de datos), reservando el 20 % restante como conjunto de prueba para la evaluación final, evitando así *data leakage* durante la selección de variables.

Criterios de decisión. La inclusión o exclusión de variables se determinó considerando los siguientes criterios:

1. La **redundancia por correlación** entre variables,
2. La **magnitud promedio de las atribuciones de IG** (impacto total sobre la predicción),
3. La **consistencia de las atribuciones** entre entrenamiento y validación.
4. La **estabilidad del desempeño del modelo**: las exclusiones se mantuvieron únicamente cuando no provocaron una degradación apreciable en las métricas de evaluación.

Se favorecieron configuraciones más compactas siempre que no implicaran una pérdida sustancial de desempeño ni redujeran la capacidad del modelo para representar la dinámica temporal subyacente. El proceso se dio por concluido cuando las atribuciones de IG se estabilizaron o cuando la eliminación de una variable produjo una degradación apreciable en las métricas de validación.

5.7.4. Optimización de hiperparámetros

Con las variables seleccionadas y tomando la arquitectura recurrente–densa como parte del espacio de búsqueda, se llevó a cabo una optimización de hiperparámetros mediante la librería *Optuna*. El objetivo fue minimizar la pérdida de validación (MSE) sobre el conjunto de ajuste, siguiendo un esquema de búsqueda automatizada.

La optimización se realizó exclusivamente sobre el 80 % inicial de los datos destinado al ajuste (manteniendo la división temporal 70 % entrenamiento / 30 % validación sin *shuffle*), reservando el 20 % restante como *hold-out* independiente para la evaluación final. En cada *trial*, el modelo se entrenó con *early stopping* y se registró como métrica de desempeño la mínima pérdida de validación alcanzada durante el entrenamiento.

Espacio de búsqueda. El proceso exploró simultáneamente hiperparámetros asociados a tres bloques: (i) la arquitectura recurrente, (ii) la arquitectura del bloque denso (MLP), y (iii) el entrenamiento. La Tabla 5.3 resume el rango y tipo de muestreo utilizado en cada caso.

Tabla 5.3: Espacio de búsqueda y tipo de muestreo en Optuna para la RNN.

Bloque	Hiperparámetro	Rango / Opciones	Muestreo
Arquitectura RNN	<code>rnn_type</code>	{LSTM, GRU, RNN}	categórico
	<code>rnn_hidden_size</code>	[32, 256]	entero (escala log)
	<code>rnn_num_layers</code>	[1, 4]	entero
	<code>dropout_rnn</code>	[0.0, 0.5] (si <code>rnn_num_layers > 1</code>)	continuo
Bloque MLP	<code>n_mlp_layers</code>	[1, 5]	entero
	<code>mlp_width_l{i}</code>	[16, 512]	entero (escala log)
	<code>mlp_activation</code>	{ReLU, Leaky ReLU, Tanh, SiLU, GELU}	categórico
	<code>mlp_dropout_base</code>	[0.0, 0.5]	continuo
Entrenamiento	<code>lr</code>	[10^{-5} , 10^{-2}]	continuo (escala log)
	<code>weight_decay</code>	[10^{-8} , 10^{-2}]	continuo (escala log)
	<code>batch_size</code>	{16, 32, 64, 128, 256}	categórico
	<code>epochs</code>	200	fijo
	<code>patience</code>	20	fijo
Validación	<code>split</code> temporal	56/24/20 (sin <i>shuffle</i>)	fijo

Detalles de implementación. En cada *trial*, se muestreó una combinación de hiperparámetros dentro del espacio definido y se construyó dinámicamente una arquitectura RNN+MLP. La RNN se generó con el tipo de celda seleccionado (LSTM, GRU o RNN simple), número de capas y tamaño oculto, aplicando *dropout* entre capas cuando correspondía. La salida del bloque recurrente se conectó a una MLP cuya profundidad, ancho por capa, función de activación y tasa de *dropout* decreciente se determinaron también mediante muestreo aleatorio controlado. El optimizador Adam se inicializó con la tasa de aprendizaje muestreada y se ajustó el parámetro de *weight decay* directamente sobre los *param groups*.

Cada configuración se entrenó hasta un máximo de 200 épocas, utilizando *early stopping* con la paciencia seleccionada para evitar sobreajuste. La función objetivo de *Optuna* devolvió la mínima pérdida de validación registrada (`val_loss`) como métrica de evaluación.

Justificación de los rangos. Los intervalos definidos reflejan valores típicos en tareas de regresión temporal con redes recurrentes:

- **Tipo de celda recurrente:** Se consideraron tres variantes recurrentes (LSTM, GRU y RNN simple), todas evaluadas bajo el mismo protocolo de optimización. LSTM y GRU son arquitecturas con mecanismos de

compuertas que favorecen el aprendizaje de dependencias temporales prolongadas, mientras que la RNN simple representa una formulación más elemental.

- **Tamaño oculto y número de capas:** el rango [32, 256] y hasta cuatro capas permiten explorar desde arquitecturas compactas hasta configuraciones moderadamente profundas, manteniendo la estabilidad numérica del entrenamiento.
- **Dropout:** los rangos [0.0, 0.5] permiten controlar la regularización.
- **Hiperparámetros de entrenamiento:** los intervalos de `lr`, `weight_decay` y `batch_size` son equivalentes a los utilizados en la optimización de la MLP, garantizando consistencia entre experimentos.

Procedimiento. Cada ejecución del estudio siguió la siguiente secuencia: (i) muestreo de hiperparámetros, (ii) construcción de la arquitectura RNN+MLP, (iii) entrenamiento sobre el 70 % de los datos y validación sobre el 30 %, (iv) monitoreo de la pérdida en validación con *early stopping*, y (v) registro de la mínima `val_loss` alcanzada.

Al finalizar el estudio, se seleccionó la combinación de hiperparámetros que minimizó la pérdida de validación. El conjunto de *test* del 20 % se mantuvo completamente aislado y se utilizó únicamente para la evaluación final del modelo optimizado, garantizando independencia total respecto al proceso de búsqueda.

Adicionalmente, se empleó un mecanismo de *pruning* equivalente al utilizado en la optimización de la MLP (Sección 5.6.4), mediante un `MedianPruner`. Este se configuró con `n_startup_trials=10`, `n_warmup_steps=20` e `interval_steps=5`, permitiendo descartar tempranamente aquellas configuraciones cuyo desempeño intermedio resultaba inferior a la mediana acumulada. De este modo, se redujo el costo computacional del estudio concentrando los recursos en arquitecturas más prometedoras.

5.8. PINNs

En esta sección se detalla la metodología seguida para construir y entrenar redes neuronales informadas por la física (PINNs) acopladas al modelo hidrológico conceptual GR4J. El objetivo es imponer, durante el entrenamiento, la explicabilidad física entre las variables hidrológicas y el caudal mediante una función de pérdida que combina términos de ajuste a datos y de consistencia física.

En los modelos de aprendizaje profundo tradicionales aplicados a hidrología, la red neuronal se entrena para predecir directamente el caudal Q a partir de variables meteorológicas (por ejemplo, precipitación y evapotranspiración potencial). Este enfoque puramente empírico puede alcanzar buenas métricas de ajuste, pero carece de interpretabilidad física y puede llegar a generalizar pobremente fuera del rango de datos observados.

En el presente trabajo se busca incorporar las relaciones hidrológicas fundamentales dentro del proceso de aprendizaje mediante el acoplamiento explícito con el modelo conceptual GR4J. Dado que los parámetros estructurales de GR4J (x_1, x_2, x_3, x_4) son característicos de la cuenca y se mantienen fijos una vez calibrados, no resulta conveniente que la red los aprenda dinámicamente. En cambio, los verdaderos grados de libertad que permiten capturar la variabilidad temporal del sistema son los forzantes meteorológicos de entrada: la precipitación P y la evapotranspiración E .

Por ello, el enfoque propuesto consiste en que la red neuronal f_ϕ no prediga el caudal directamente, sino las series de P y/o E , las cuales se utilizan como entrada al modelo GR4J calibrado \mathcal{M}_θ . De esta forma, la predicción final del caudal emerge de la simulación física del modelo hidrológico, preservando la coherencia estructural del proceso lluvia–escorrentía.

En el entrenamiento, la función de pérdida combina dos componentes:

1. un término de datos, que mide el error entre las variables meteorológicas predichas y observadas, y
2. un término físico, que penaliza la discrepancia entre el caudal observado y el generado por GR4J al propagar esas predicciones.

Este esquema permite que la red aprenda representaciones consistentes con la física subyacente, mejorando la interpretabilidad y robustez del modelo resultante.

5.8.1. Formulación general

Sea \mathcal{M}_θ el modelo hidrológico GR4J con parámetros calibrados θ y estado interno \mathbf{s} . Sea f_ϕ una red neuronal parametrizada por ϕ , que recibe como entrada un lote de datos \mathbf{X} y produce estimaciones de variables meteorológicas requeridas por GR4J.

Definimos una función de pérdida compuesta como

$$\mathcal{L}(\phi) = w_{\text{data}} \mathcal{L}_{\text{data}}(\phi; \mathbf{X}) + w_{\text{phys}} \mathcal{L}_{\text{phys}}(\phi; \mathbf{X}),$$

donde w_{data} y w_{phys} ponderan el compromiso entre fidelidad a los datos observados y coherencia física impuesta por el modelo hidrológico.

Pérdida de datos. La red neuronal produce estimaciones

$$\hat{P} = f_\phi(\mathbf{X}) \quad \text{o} \quad \hat{E} = f_\phi(\mathbf{X}),$$

según la variante considerada. La pérdida de datos evalúa la capacidad predictiva directa de la red mediante el error cuadrático medio (MSE):

$$\mathcal{L}_{\text{data}} = \begin{cases} \text{MSE}(f_\phi(\mathbf{X}), P_{\text{obs}}), & \text{si se predice precipitación,} \\ \text{MSE}(f_\phi(\mathbf{X}), E_{\text{obs}}), & \text{si se predice evapotranspiración.} \end{cases}$$

En variantes multisalida, la pérdida se extiende sumando los términos MSE correspondientes a cada variable predicha.

Pérdida física. La pérdida física se construye propagando a través de GR4J las entradas requeridas (precipitación P y evapotranspiración E), que pueden ser observadas o estimadas por la red según la configuración adoptada. El modelo hidrológico genera entonces un caudal simulado

$$Q_{\text{GR4J}} = \mathcal{M}_\theta(P, E; \mathbf{s}),$$

y la pérdida física se define como

$$\mathcal{L}_{\text{phys}} = \text{MSE}(Q_{\text{GR4J}}, Q_{\text{obs}}).$$

Nótese que $\mathcal{L}_{\text{phys}}$ depende de ϕ de forma implícita cuando P y/o E provienen de $f_\phi(\mathbf{X})$.

Interpretación del sistema compuesto. En consecuencia, la red neuronal no predice directamente el caudal, sino que produce estimaciones de P y/o E , las cuales son utilizadas por el modelo hidrológico \mathcal{M}_θ para generar la predicción final Q_{GR4J} . De esta forma, el sistema compuesto mantiene la estructura física de GR4J mientras aprende relaciones empíricas entre las variables de entrada y el comportamiento hidrológico observado.

Sobre la no utilización de puntos de colocación. En las formulaciones clásicas de redes neuronales informadas por la física (PINNs), es habitual incorporar *puntos de colocación* adicionales —es decir, puntos sintéticos en el dominio de entrada— para imponer la ecuación diferencial gobernante aun en ausencia de datos observados. Sin embargo, en el presente trabajo no se emplean puntos de colocación artificiales.

La razón principal radica en la naturaleza *stateful* del modelo GR4J. A diferencia de muchas PINNs basadas en ecuaciones diferenciales locales, donde el residuo físico puede evaluarse en puntos independientes del dominio, GR4J es un modelo dinámico con memoria interna: el estado hidrológico en cada instante depende explícitamente de la evolución previa de los almacenamientos internos. Cada par de entradas (P_t, E_t) modifica los estados (S, R) , condicionando todas las salidas futuras.

La generación de puntos sintéticos implicaría introducir secuencias ficticias de precipitación y evapotranspiración que, al ser propagadas por el modelo, alterarían el estado interno de GR4J. Esto no sólo afectaría el valor instantáneo del caudal simulado, sino también toda la trayectoria posterior del sistema. En consecuencia, no es posible evaluar el residuo físico en puntos aislados sin alterar la dinámica global del modelo.

Por este motivo, la coherencia física se impone exclusivamente sobre las trayectorias observadas, preservando la continuidad temporal real del sistema hidrológico y evitando introducir estados internos artificiales que podrían distorsionar el comportamiento dinámico del modelo.

Consideraciones hidrológicas y configuración de GR4J. El modelo GR4J utilizado en esta etapa corresponde al mismo esquema conceptual descrito en la Sección 5.5, empleando los parámetros $\{x_1, x_2, x_3, x_4\}$ obtenidos en la configuración que mostró el mejor desempeño durante la calibración (ver Sección 5.5). Estos parámetros se mantienen fijos a lo largo del entrenamiento de las PINNs, dado que representan propiedades intrínsecas de la cuenca y no constituyen variables de predicción.

En cuanto a los forzantes meteorológicos, se aplican las mismas consideraciones de preprocesamiento presentadas previamente: la **precipitación diaria** se agrega espacialmente a partir de las estaciones disponibles mediante polígonos de Thiessen, de acuerdo con los pesos derivados en la Sección 5.5. El valor resultante, $P_{\text{Thiessen,obs}}(t)$, se utiliza de forma diferente según la variante del modelo:

- Cuando la red neuronal predice la precipitación, $P_{\text{Thiessen,obs}}$ actúa como referencia observada en el término de pérdida de datos ($\mathcal{L}_{\text{data}}$).
- Cuando la red predice únicamente la evapotranspiración, $P_{\text{Thiessen,obs}}$ se emplea como entrada fija al modelo GR4J, junto con la evapotranspiración predicha E_{pred} , para generar el caudal simulado Q_{GR4J} .

De esta forma, la precipitación agregada cumple un doble propósito: servir como dato de ajuste en las variantes que la predicen y como forzante exógeno en aquellas que sólo estiman la evapotranspiración.

Inicialización del estado (*warm-up*). De igual forma, se adopta el mismo procedimiento de *warm-up* de 365 días descrito en la Sección 5.5, utilizando series observadas de P y E para estabilizar los almacenamientos internos del modelo antes del inicio del entrenamiento. Durante el proceso de optimización, se mantiene la continuidad del estado entre *batches* para conservar la coherencia hidrológica de las simulaciones.

5.8.2. Variantes evaluadas

Se exploraron distintas variantes del esquema propuesto, que difieren en el grado de acoplamiento explícito con la física del modelo GR4J. En todos los casos, la red neuronal f_ϕ genera estimaciones de una o más variables meteorológicas que luego son utilizadas por el modelo hidrológico \mathcal{M}_θ para simular el caudal. Las pérdidas de datos y físicas se formulan según la variable (o combinación de variables) predicha.

PINN-E (predicción de evapotranspiración)

La red predice \hat{E} . La pérdida de datos compara \hat{E} con E_{obs} , y la pérdida física evalúa el caudal simulado al forzar GR4J con \hat{E} y la precipitación agregada observada P_{agg} :

$$\mathcal{L}_{\text{data}} = \text{MSE}(\hat{E}, E_{\text{obs}}), \quad \mathcal{L}_{\text{phys}} = \text{MSE}(Q_{\text{GR4J}}(P_{\text{agg}}, \hat{E}; \theta), Q_{\text{obs}}).$$

Esta formulación transfiere información física hacia la variable de salida \hat{E} y puede ser ventajosa cuando la serie de E observada es ruidosa.

PINN-P (predicción de precipitación agregada)

En esta variante, la red neuronal predice únicamente la serie diaria de precipitación agregada \hat{P}_{agg} , mientras que la evapotranspiración se mantiene fija a partir de los valores observados E_{obs} . De este modo, la red aprende a estimar la precipitación efectiva que, al ser introducida en el modelo GR4J junto con E_{obs} , reproduce de forma coherente el caudal observado.

La función de pérdida se define como:

$$\mathcal{L}_{\text{data}} = \text{MSE}(\hat{P}_{\text{agg}}, P_{\text{agg}}), \quad \mathcal{L}_{\text{phys}} = \text{MSE}(Q_{\text{GR4J}}(\hat{P}_{\text{agg}}, E_{\text{obs}}; \theta), Q_{\text{obs}}).$$

De esta manera, la componente física del entrenamiento retropropaga información hidrológica hacia la estimación de la precipitación, permitiendo que la red ajuste sus predicciones no sólo según las observaciones directas, sino también de acuerdo con la coherencia entre la lluvia estimada, la evapotranspiración observada y el caudal resultante. Esta formulación puede ser particularmente útil en escenarios donde la precipitación observada presenta incertidumbres o lagunas temporales.

PINN-PE (predicción conjunta de P agregado y E)

La red produce dos salidas $[\hat{P}_{\text{agg}}, \hat{E}]$ de manera análoga que las variantes anteriores. Ambas salidas son luego introducidas en el modelo GR4J para generar la simulación de caudal.

La pérdida de datos suma los términos correspondientes a la precipitación agregada y a la evapotranspiración, mientras que la pérdida física impone coherencia sobre el caudal simulado:

$$\mathcal{L}_{\text{data}} = \text{MSE}(\hat{P}_{\text{agg}}, P_{\text{agg}}) + \text{MSE}(\hat{E}, E_{\text{obs}}), \quad \mathcal{L}_{\text{phys}} = \text{MSE}(Q_{\text{GR4J}}(\hat{P}_{\text{agg}}, \hat{E}; \theta), Q_{\text{obs}}).$$

Esta variante fuerza a la red a aprender forzantes meteorológicos físicamente consistentes que, al ser introducidos en GR4J, reproduzcan el caudal observado.

5.8.3. Arquitectura de red

Las arquitecturas de red utilizadas en las variantes PINN siguen la misma estructura base que en los modelos previamente evaluados, correspondientes a redes neuronales de tipo perceptrón multicapa (MLP) y redes recurrentes (RNN). En este trabajo, la incorporación del conocimiento físico no implica una modificación estructural en la arquitectura de la red, sino un cambio en la formulación de la función de pérdida, tal como se detalló en la Sección 5.8.1.

De este modo, las redes MLP y RNN conservan su diseño original —número de capas, neuronas, funciones de activación y esquema de entrenamiento—, pero son optimizadas bajo una pérdida compuesta que combina los términos de ajuste

a datos y de coherencia física. Esto permite evaluar el impacto del acoplamiento físico sin introducir factores adicionales de complejidad arquitectónica, aislando así el efecto de la penalización física sobre el proceso de aprendizaje.

La metodología general de entrenamiento y análisis se mantiene consistente con la descrita para los modelos MLP y RNN en los capítulos anteriores. En particular, se aplican los mismos procedimientos de incorporación de variables temporales, agregación de precipitación e inclusión de rezagos (*lags*), así como los mismos criterios de selección de variables mediante análisis de correlación e *Integrated Gradients*.

Asimismo, se conservan los esquemas de búsqueda de hiperparámetros implementados con `Optuna`, junto con los mismos criterios de validación y evaluación del desempeño predictivo.

La única diferencia estructural respecto a los modelos puramente *data-driven* radica en la inclusión del término de coherencia física en la función de pérdida, que permite acoplar las predicciones de la red con el modelo hidrológico GR4J.

En esta etapa se analizan en paralelo ambas arquitecturas (PINN-MLP y PINN-RNN) sobre las tres variantes consideradas (P, E y PE), con el objetivo de identificar aquella que presenta un mejor equilibrio entre capacidad predictiva y consistencia física.

Las particularidades de implementación, los ajustes específicos y las consideraciones metodológicas propias de cada variante se describen en las subsecciones siguientes.

5.8.4. Incorporación y selección de variables

La selección e incorporación de variables de entrada sigue la misma metodología utilizada en los modelos MLP y RNN, basada en la evaluación sistemática de configuraciones de entrada y en el análisis de relevancia mediante correlación estadística e *Integrated Gradients*. No obstante, en las variantes PINN se introducen ajustes específicos vinculados a la naturaleza físico-informada del problema de predicción.

Los modelos previos (MLP y RNN) estaban orientados a predecir directamente el caudal, por lo que se incluían como entradas las variables meteorológicas y sus respectivos *lags* de precipitación. En la formulación PINN-PE, se cambio, la red no predice el caudal de forma directa, sino que estima las variables meteorológicas (precipitación agregada y evapotranspiración) que luego son utilizadas por el modelo GR4J para simular el caudal.

Por este motivo, la inclusión de variables retardadas (*lags*) se define de forma **condicionada a la variable objetivo de cada variante**. En las configuraciones donde la red predice evapotranspiración (**PINN-E** y **PINN-PE**), se incorporan ***lags* de evapotranspiración** (E_{t-1}, E_{t-2}, \dots) para capturar su dinámica temporal y mejorar la estimación de dicha magnitud. En cambio, en la variante **PINN-P** (predicción de precipitación), estos *lags* de E no se incluyen, ya que E se toma como forzante observado y no constituye variable de predicción.

De manera análoga, los **lags de precipitación** se consideran únicamente en aquellas variantes donde la red predice precipitación (**PINN-P** y **PINN-PE**). En la variante **PINN-E**, la precipitación se utiliza como entrada exógena observada y, por consistencia metodológica, no se expanden *lags* adicionales de P .

Adicionalmente, por consistencia con el planteo del problema, cuando una variable meteorológica es **objetivo de predicción** (por ejemplo, evapotranspiración en PINN-E o precipitación en PINN-P), dicha variable no se incorpora como entrada directa en su forma observada agregada. Esto evita *data leakage* —esto es, proporcionar al modelo la misma magnitud que debe estimar— y mantiene la coherencia entre el objetivo supervisado y el conjunto de variables explicativas.

En el caso de la evapotranspiración (PINN-E y PINN-PE), la serie observada de E se excluye completamente del conjunto de entradas.

En el caso de la precipitación (PINN-P y PINN-PE), no se utiliza como entrada la precipitación previamente agregada mediante esquemas espaciales (p. ej., Thiessen). No obstante, sí se incluyen como entradas las precipitaciones medidas en estaciones individuales, permitiendo que la red aprenda de forma implícita una combinación espacial efectiva. La precipitación agregada se emplea únicamente como referencia observada en el término de datos ($\mathcal{L}_{\text{data}}$), actuando como señal supervisada pero sin imponer *a priori* un esquema de integración espacial.

Variables temporales Al igual que en el modelo MLP y RNN, se incorporaron únicamente las variables temporales *día* y *mes*, representadas mediante sus transformaciones cíclicas seno y coseno. El objetivo es determinar empíricamente qué conjunto de variables temporales contribuye de manera más significativa al desempeño del modelo.

Representación espacial de la precipitación En las variantes donde la precipitación **no** es predicha por la red (p. ej., **PINN-E**), se mantiene el análisis comparativo entre distintos esquemas de agregación espacial (estaciones individuales, promedio simple, ponderación por distancia y polígonos de Thiessen), en línea con lo desarrollado para los modelos puramente basados en datos. Este análisis permite identificar la representación más adecuada del forzante P cuando éste se considera exógeno.

En las variantes donde la red **sí predice precipitación** (**PINN-P** y **PINN-PE**), la agregación espacial no se fija previamente. En su lugar, se entregan a la red las precipitaciones individuales de las estaciones disponibles, de modo que el modelo pueda aprender internamente una combinación espacial adecuada. La precipitación agregada por Thiessen se utiliza únicamente como referencia observada en el término de datos ($\mathcal{L}_{\text{data}}$).

Inclusión de lags En todos los casos, los *lags* se construyen únicamente para aquellas variables que constituyen **salida** de la PINN (precipitación en PINN-P

y PINN-PE; evapotranspiración en PINN-E y PINN-PE).

El tratamiento de los valores pasados en el marco de las redes neuronales informadas por la física (PINNs) depende de la arquitectura base empleada para la aproximación de la solución.

En los casos en que la PINN se implementa sobre una arquitectura de tipo MLP, se adopta el mismo procedimiento descrito previamente para dicho modelo. En primer lugar, se evalúa de manera preliminar si la inclusión de variables retrasadas (*lags*) aporta mejoras en el desempeño, fijando un horizonte temporal inicial de **7 días** y considerando los distintos esquemas de agregación espacial. Posteriormente, una vez identificado el esquema más adecuado, se determina el **número óptimo de días de retardo** mediante pruebas sistemáticas con horizontes comprendidos entre **1 y 14 días**.

Por el contrario, cuando la PINN se construye a partir de una arquitectura recurrente (RNN), no se incorporan *lags* explícitos de las variables de entrada. En este caso, la modelación de las dependencias temporales se realiza a través de la propia estructura recurrente, mediante la definición de una **ventana temporal de entrada**, siguiendo el mismo criterio metodológico descrito para las RNN convencionales. Se consideran ventanas temporales de longitud variable, comprendidas entre **1 y 14 días**, con el objetivo de analizar el impacto del horizonte temporal sobre el desempeño del modelo e identificar la configuración óptima.

De este modo, el tratamiento de la información temporal en las PINNs se adapta de forma coherente a la arquitectura subyacente, evitando redundancias y asegurando la consistencia metodológica entre los distintos enfoques considerados.

5.8.5. Selección de variables

Una vez definida la configuración inicial de variables (temporales, agregación espacial de precipitación y *lags*), se aplica el mismo procedimiento de selección de variables descrito para el modelo MLP, basado en el análisis de correlación y en la aplicación del método *Integrated Gradients* (IG).

Protocolo general. La selección se estructura en dos etapas complementarias y sucesivas:

1. **Análisis de correlación.** Se aplica el mismo análisis de correlación desarrollado en la Sección 4.3, identificando pares de variables altamente correlacionadas. Se proponen eliminaciones de variables redundantes, priorizando aquellas con mayor respaldo conceptual o disponibilidad de datos. Cada eliminación es seguida por una reevaluación del modelo para verificar que no produzca degradación significativa en las métricas de validación.
2. **Análisis de relevancia mediante *Integrated Gradients*.** Una vez depurado el conjunto inicial por correlación, se calculan las atribuciones de IG sobre el modelo entrenado con la arquitectura base. Se implementa un

esquema de depuración iterativa eliminando variables con bajo impacto medio o con atribuciones inestables entre los conjuntos de entrenamiento y validación. En cada iteración, el modelo se reentrena y se recalculan las atribuciones de IG, hasta alcanzar estabilidad en las relevancias o en el desempeño de validación.

Durante estas etapas, los **pesos de la función de pérdida se mantienen fijos** con valores $w_{\text{data}} = w_{\text{phys}} = 1$, de modo de otorgar igual ponderación al término de ajuste a datos y al término de coherencia física. Esta decisión busca aislar el efecto de la selección de variables sin introducir variaciones adicionales asociadas al balance entre componentes de la pérdida. El análisis específico del equilibrio óptimo entre términos se aborda en una etapa posterior, una vez determinado el conjunto final de entradas.

El análisis de IG se realiza de manera separada sobre los subconjuntos de entrenamiento y validación, correspondientes respectivamente al 56 % y al 24 % de los datos originales. El conjunto de prueba (20 %) no se utiliza en esta etapa, a fin de evitar *data leakage*.

Criterios de decisión. La inclusión o exclusión de variables se basa en:

1. La **redundancia por correlación** entre variables,
2. La **magnitud media de las atribuciones de IG** (impacto total sobre la predicción),
3. La **consistencia de las atribuciones** entre los conjuntos de entrenamiento y validación,
4. La **estabilidad del desempeño del modelo**: las exclusiones se mantienen únicamente cuando no provocan degradación apreciable en las métricas de evaluación.

5.8.6. Análisis del balance entre términos de la función de pérdida

Un aspecto distintivo de las redes neuronales informadas por la física es la presencia de múltiples términos en la función de pérdida, cada uno con un peso asociado que regula su contribución durante el entrenamiento. En las variantes de PINN implementadas, la función de pérdida se define como:

$$\mathcal{L}(\phi) = w_{\text{data}} \mathcal{L}_{\text{data}} + w_{\text{phys}} \mathcal{L}_{\text{phys}}$$

donde w_{data} y w_{phys} controlan el balance entre la fidelidad a las observaciones meteorológicas y la coherencia física del caudal simulado.

En lugar de optimizar automáticamente estos pesos mediante un algoritmo de búsqueda de hiperparámetros, en este trabajo se realiza un análisis sistemático de distintas combinaciones predefinidas, con el objetivo de evaluar cómo diferentes relaciones entre ambos términos afectan el desempeño del modelo.

Este análisis se lleva a cabo una vez definidas las variables de entrada, los esquemas de agregación espacial y los *lags* (o ventanas temporales), manteniendo fija una configuración arquitectónica base. De este modo, el efecto observado puede atribuirse exclusivamente al balance entre los términos de la función de pérdida, sin interferencias derivadas de la optimización posterior de hiperparámetros.

No se exploran todas las combinaciones posibles de pesos absolutos, sino únicamente distintos *ratios* entre la pérdida de datos y la pérdida física. Esto se debe a que combinaciones proporcionalmente equivalentes (por ejemplo, 0.25–0.25 y 1–1) preservan exactamente la misma relación entre términos y, por tanto, inducen el mismo equilibrio relativo durante el entrenamiento, aun cuando el valor absoluto de la función de pérdida sea diferente.

Las configuraciones evaluadas abarcan relaciones comprendidas hasta un máximo de 1 : 4 (o 4 : 1), es decir, no se consideran desbalances más extremos como 1 : 5 o 1 : 6, con el fin de evitar escenarios excesivamente dominados por uno de los términos que puedan comprometer la estabilidad o la interpretabilidad del modelo.

Asimismo, no se evalúa la configuración $w_{\text{data}} = 1$ y $w_{\text{phys}} = 0$, ya que ello implicaría prescindir completamente de la restricción física, desnaturalizando el enfoque de red neuronal informada por la física y reduciendo el modelo a una arquitectura puramente basada en datos.

Las configuraciones restantes incluyen casos cercanos a los extremos, con fuerte predominancia de uno de los términos, así como el caso balanceado 1 : 1, permitiendo caracterizar de manera explícita el compromiso entre ajuste empírico y consistencia física e identificar el *ratio* que produce el mejor desempeño predictivo en el sistema acoplado.

5.8.7. Optimización de hiperparámetros

Con las variables seleccionadas, los pesos de pérdida definidos, y tomando la arquitectura como parte del espacio de búsqueda, se realiza una optimización de hiperparámetros con *Optuna*, siguiendo un esquema de búsqueda automatizada orientado a minimizar la pérdida de validación (MSE).

En el caso de las **PINN-MLP**, la optimización arquitectónica replica exactamente el enfoque adoptado para la MLP pura (Sección 5.6.4), manteniendo idéntico espacio de búsqueda, criterios de evaluación y esquema de entrenamiento.

De manera análoga, en las **PINN-RNN** se emplea el mismo procedimiento definido para la RNN convencional (Sección 5.7.4), incluyendo la exploración del tipo de celda recurrente, tamaño oculto, número de capas, regularización y parámetros de entrenamiento.

En ambos casos, la única diferencia respecto a los modelos puramente basados en datos radica en la formulación de la función de pérdida, que incorpora el término de coherencia física, mientras que el esquema de optimización arquitectónica permanece inalterado.

5.8.8. Consideraciones metodológicas específicas de la PINN

Manejo del estado interno de GR4J. En las variantes PINN, el manejo del estado interno de GR4J adquiere particular relevancia, ya que es fundamental para garantizar la coherencia física entre las predicciones de la red y la evolución hidrológica simulada por el modelo conceptual. Antes del entrenamiento, se ejecuta un período de calentamiento (*warmup*) de 365 días con series observadas de precipitación P y evapotranspiración E , excluyendo dicho período del conjunto de entrenamiento, a fin de inicializar los almacenamientos (S, R) en un estado hidrológicamente realista.

Durante el entrenamiento, al inicio de cada época el estado de GR4J se reinicia al estado resultante del *warmup*. Dentro de una misma época, el estado se propaga continuamente entre *batches*, lo que permite que la memoria hidrológica del modelo evolucione de forma coherente a través de la serie temporal.

Para el cálculo del caudal con GR4J en los conjuntos de validación y prueba, se mantiene la consistencia temporal y de estado del modelo hidrológico respecto al entrenamiento. Durante la validación, el estado inicial de GR4J se obtiene continuando la simulación a partir del estado final alcanzado al concluir el conjunto de entrenamiento, garantizando así la coherencia hidrológica entre ambos períodos. De manera análoga, para el conjunto de prueba se utiliza como condición inicial el estado resultante de propagar consecutivamente los conjuntos de entrenamiento y validación. Este esquema asegura una evolución continua y físicamente consistente de los almacenamientos internos (S, R) a lo largo de toda la serie temporal, evitando reinicios arbitrarios que podrían alterar la dinámica del sistema.

Implicancias computacionales de la naturaleza *stateful*. La naturaleza dinámica y dependiente del estado de GR4J tiene además una consecuencia directa en el esquema de entrenamiento de la PINN. Dado que los almacenamientos internos del modelo se actualizan secuencialmente en el tiempo, las entradas deben procesarse estrictamente en orden cronológico. No es posible evaluar el modelo hidrológico en instantes temporales independientes ni reordenar los datos sin alterar la evolución del estado.

En términos computacionales, esto introduce una restricción relevante: la evaluación del término físico de la función de pérdida no puede paralelizarse completamente en el eje temporal. Aunque la red neuronal puede ejecutarse de manera vectorizada, el cálculo de Q_{GR4J} requiere la propagación paso a paso de la serie temporal, ya que cada salida depende del estado resultante del paso anterior.

En consecuencia, durante el entrenamiento se produce un cuello de botella en el cálculo de la pérdida física, asociado a la simulación secuencial del modelo hidrológico. Este comportamiento no constituye una limitación accidental de implementación, sino una consecuencia estructural del acoplamiento con un modelo dinámico con memoria. Por diseño, se espera entonces un mayor costo computacional respecto a modelos puramente basados en datos o PINNs formuladas sobre ecuaciones evaluables punto a punto.

Tratamiento de valores faltantes de caudal durante el entrenamiento.

El manejo de registros con caudal faltante sigue el mismo principio descrito previamente (Sección 5.5), donde se explicó la necesidad de preservar la continuidad temporal del modelo conceptual aun en presencia de observaciones nulas. Durante el entrenamiento, estos valores no se emplean para el cálculo de las métricas de ajuste ni de la función de pérdida, pero las entradas correspondientes se procesan igualmente por el modelo GR4J para asegurar la correcta actualización de sus estados internos. En la práctica, se aplica un **enmascaramiento selectivo** sobre la pérdida, de modo que sólo las observaciones de caudal disponibles contribuyen al cálculo de error, mientras que las restantes mantienen la continuidad hidrológica del sistema.

De esta forma, se garantiza una simulación físicamente consistente, preservando la evolución de los almacenamientos (S, R) y evitando discontinuidades artificiales en la dinámica del modelo.

Enmascaramiento diferenciado en la función de pérdida. La función de pérdida propuesta para la PINN combina dos términos complementarios: una pérdida de datos (*data loss*) que supervisa directamente las predicciones de precipitación y evapotranspiración frente a sus observaciones, y una pérdida física (*physics loss*) que impone consistencia hidrológica mediante el modelo GR4J, comparando el caudal simulado con el observado.

En presencia de series temporales con observaciones parciales de caudal, se implementa una estrategia de **enmascaramiento diferenciado**. La pérdida física se calcula únicamente sobre los registros donde existe caudal observado, mientras que la pérdida de datos se evalúa sobre todos los registros disponibles. Esto significa que, aunque un registro carezca de observación de caudal, no se descarta completamente del entrenamiento: si existen valores observados de precipitación o evapotranspiración, estos se utilizan para calcular la pérdida correspondiente, manteniendo el aprendizaje supervisado sobre las variables meteorológicas.

En la implementación práctica, este esquema se materializa asignando un valor faltante a la pérdida física ($\mathcal{L}_{\text{phys}} = 0$) cuando no se dispone de caudal observado en un registro dado. De este modo, durante el *backpropagation*, el gradiente correspondiente a dicho término se anula, y las actualizaciones de los parámetros del modelo se orientan exclusivamente según la pérdida de datos ($\mathcal{L}_{\text{data}}$). Esta formulación garantiza que la red continúe aprendiendo de las observaciones meteorológicas sin introducir gradientes erróneos derivados de comparaciones físicas inexistentes.

Somos conscientes de que este mecanismo produce el mismo efecto sobre el gradiente que el caso en el que $\mathcal{L}_{\text{phys}} = 0$ por un ajuste perfecto entre Q_{GR4J} y Q_{obs} ; en ambos escenarios, la componente física no aporta gradiente durante la retropropagación. Sin embargo, esta coincidencia no genera interferencias ni comportamientos indeseados: en ambos casos, la ausencia de gradiente físico es deseable, ya que permite que la optimización se dirija exclusivamente hacia la mejora del término de datos. De esta manera, el modelo mantiene la coherencia

entre las dos partes de la función de pérdida, aprovechando la información física solo cuando está disponible y concentrando el aprendizaje en los componentes observacionales en los restantes casos.

Capítulo 6

Resultados

6.1. Aclaración de entorno de ejecución

Todos los experimentos presentados en este trabajo fueron ejecutados en el sistema de cómputo de alto desempeño ClusterUY. Los recursos fueron solicitados al gestor de tareas del sistema, definiendo memoria, número de tareas y utilización de aceleradores GPU según el tipo de experimento.

Las ejecuciones se realizaron en nodos de cómputo que disponen de 40 núcleos de CPU por nodo, arquitectura típica de los nodos basados en procesadores de clase servidor orientados a cómputo científico. En todos los casos se asignaron 32 GB de memoria RAM por trabajo.

6.1.1. Experimentos con GR4J, MLP y RNN

Para los experimentos correspondientes a los modelos GR4J, MLP y RNN pura, cada ejecución se realizó utilizando:

- Un nodo con 40 núcleos de CPU disponibles,
- 32 GB de memoria RAM,
- 1 GPU dedicada.

Si bien el nodo dispone de 40 núcleos, los experimentos se ejecutaron como procesos individuales (sin paralelización explícita a nivel de CPU), utilizando principalmente un flujo de ejecución secuencial junto con aceleración por GPU para las operaciones matriciales intensivas.

6.1.2. Experimentos con PINNs

En el caso de los experimentos con *Physics-Informed Neural Networks* (PINNs), se utilizó la misma configuración en términos de nodo y memoria:

- Un nodo con 40 núcleos de CPU disponibles,

- 32 GB de memoria RAM,

pero sin utilización de GPU, por lo que las ejecuciones se realizaron exclusivamente sobre CPU.

La decisión de no emplear aceleración por GPU en las PINNs se fundamenta en su naturaleza predominantemente secuencial, tal como se explicó en la sección metodológica. Esta característica limita la posibilidad de aprovechar el paralelismo masivo que ofrecen las GPUs. En la práctica, se observó que el uso de GPU introducía sobrecostos asociados a la transferencia de datos y cambios de contexto entre CPU y GPU, resultando en mayores tiempos de ejecución que utilizando únicamente CPU.

6.2. Análisis correlativo de datos

6.2.1. Estadísticas descriptivas del caudal

Como punto de partida del análisis exploratorio, se calcularon las medidas estadísticas descriptivas básicas de la serie de caudal en la estación *Florida (Puente Ruta 5)*, excluyendo los valores faltantes. Estas medidas estadísticas permiten caracterizar la magnitud típica del flujo, su variabilidad y la presencia de eventos extremos.

Tabla 6.1: Medidas estadísticas descriptivas del caudal diario en Florida (Puente Ruta 5).

Medida estadística	Valor (m^3/s)
Media	24.46
Mediana	3.77
Desviación estándar	107.42
Mínimo	0.00
Máximo	3276.88

Al analizar los resultados (Tabla 6.1), se observa una diferencia marcada entre la media y la mediana, lo que evidencia una distribución fuertemente asimétrica hacia la derecha. Mientras que la mediana indica que la mayor parte del tiempo el caudal se mantiene en valores relativamente bajos, la media se ve considerablemente incrementada por la ocurrencia de eventos extremos de crecida.

La elevada desviación estándar, considerablemente mayor que la media, refuerza la existencia de una alta variabilidad y de una distribución con colas largas. El amplio rango observado confirma la coexistencia de situaciones de estiaje y crecidas severas dentro del período analizado.

Desde el punto de vista del modelado, esta estructura estadística implica que métricas como el MAE y el RMSE deben interpretarse en relación con esta alta dispersión y asimetría, siendo el RMSE particularmente sensible a errores en eventos extremos.

6.2.2. Matriz de correlación

El análisis de correlaciones constituye un paso fundamental para explorar la estructura interna del conjunto de datos y detectar patrones de dependencia entre variables hidrometeorológicas. A través de la matriz de correlación, se busca reconocer cuáles variables presentan comportamientos similares o complementarios y evaluar su potencial utilidad como predictores del caudal.

La Figura 6.1 muestra el mapa de calor de correlaciones entre el caudal objetivo en *Florida (Puente Ruta 5)*, las precipitaciones en las estaciones seleccionadas (*Florida, La Cruz, San Gabriel, Sarandí Grande, Cerro Colorado*) y las variables meteorológicas de *INIA Las Brujas* (Evapotranspiración, heliofanía, humedad relativa, radiación solar, temperaturas media/máxima/mínima y velocidad del viento).

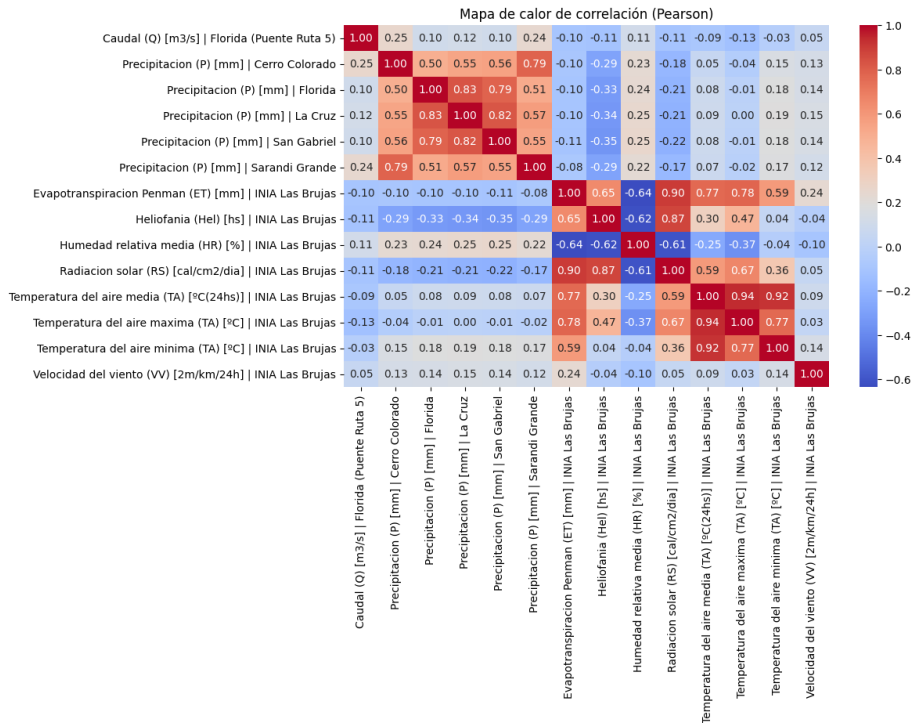


Figura 6.1: Mapa de calor de correlación (Pearson) entre variables hidrológicas y meteorológicas. Los valores van de -1 (azul) a $+1$ (rojo).

Los principales hallazgos de este análisis permiten identificar las relaciones más significativas entre las variables consideradas:

- **Caudal vs. precipitaciones (relación esperada pero moderada a escala diaria).** El caudal muestra correlaciones positivas con las precipitaciones de las estaciones cercanas (p. ej., Florida, San Gabriel, Sarandí

Grande y Cerro Colorado), aunque de magnitud moderada/baja. Esto es consistente con una respuesta hidrológica con *retardos* y no linealidades: a escala diaria, el pico de caudal suele ocurrir horas o días después de la lluvia y depende de la humedad antecedente.

- **Coherencia espacial de la lluvia.** Las precipitaciones entre estaciones presentan correlaciones elevadas (varias por encima de 0.7), indicando que los eventos pluviométricos que afectan la subcuenca tienen extensión regional y se registran de forma consistente en múltiples pluviómetros.
- **Estructura del balance energético.** Las variables atmosféricas de INIA están fuertemente interrelacionadas: radiación, temperatura y evapotranspiración correlacionan positivamente entre sí (valores altos), mientras que la humedad relativa muestra correlaciones negativas con estas (magnitud moderada/alta). La heliofanía acompaña a la radiación y a la evapotranspiración, reforzando la coherencia física del conjunto.
- **Baja conexión lineal del viento.** La velocidad del viento presenta correlaciones bajas con la mayoría de las variables y prácticamente nula con el caudal, sugiriendo que su aporte explicativo directo a escala diaria podría ser limitado.

6.3. GR4J

6.3.1. Introducción

En esta sección se presentan los resultados obtenidos mediante la implementación y calibración del modelo hidrológico conceptual **GR4J**. El objetivo es evaluar su capacidad para reproducir el comportamiento hidrológico de la cuenca bajo estudio y establecer una **línea base de desempeño** que servirá como referencia para los modelos basados en aprendizaje automático presentados en los apartados posteriores.

6.3.2. Optimización de parámetros

Dado que el modelo GR4J es de naturaleza determinista y no involucra un proceso de entrenamiento estadístico, los valores de desempeño se calcularon sobre la totalidad del conjunto de calibración (80%) bajo distintas combinaciones de parámetros generadas en cada *trial*. El objetivo de optimización fue maximizar el coeficiente de eficiencia de Nash–Sutcliffe (NSE). Tras un total de **2000 trials**, el proceso alcanzó un Nash–Sutcliffe Efficiency (NSE) = **0.65** (Mejor compromiso de desempeño sobre el conjunto de calibración del 80%).

Los valores de los parámetros del modelo GR4J se comparan en la Tabla 6.2, donde se presentan tanto los valores de referencia como los optimizados mediante Optuna.

Tabla 6.2: Comparación de parámetros del modelo GR4J: valores de referencia vs. optimizados

Parámetro	Referencia	Optimizado
x_1 (mm)	101.00	118.79
x_2 (mm)	-2.50	-1.26
x_3 (mm)	49.00	27.55
x_4 (días)	2.30	2.22

Comparando con los valores de referencia definidos por [Narbondo y cols. \(2020\)](#), se observa un leve incremento en la capacidad del depósito de producción (x_1) y una reducción más marcada en la capacidad del depósito de enrutamiento (x_3), lo que sugiere una mayor capacidad de almacenamiento en el suelo y una respuesta del caudal más rápida en el componente de enrutamiento. Los parámetros x_2 y x_4 presentan variaciones menores respecto a los valores de referencia, manteniendo una coherencia física con el comportamiento hidrológico esperado para la cuenca.

6.3.3. Evaluación del desempeño hidrológico

Análisis de métricas de desempeño

Las métricas de desempeño utilizadas para la evaluación del modelo GR4J fueron las mismas presentadas previamente en el **Capítulo 5 - Desarrollo de Modelos**, donde se describieron sus formulaciones y fundamentos teóricos.

En particular, las métricas reportadas a continuación corresponden exclusivamente al **20 % del conjunto de datos reservado para prueba**, utilizado para analizar la capacidad del modelo de generalizar su comportamiento fuera del período de calibración. De esta manera, los valores presentados reflejan el desempeño del modelo en condiciones completamente independientes del ajuste de parámetros.

La Tabla 6.3 presenta el desempeño del modelo GR4J bajo las dos configuraciones de parámetros consideradas: la **referencial**, basada en los valores de [Narbondo y cols. \(2020\)](#), y la **optimizada**, obtenida mediante el proceso de búsqueda con *Optuna*.

Tabla 6.3: Desempeño hidrológico del modelo GR4J bajo ambas configuraciones de parámetros en el conjunto de prueba.

Configuración	NSE	PBIAS [%]	MAE [m^3/s]	RMSE [m^3/s]
Referencial	0.67	-17.34	19.01	74.63
Optimizada	0.69	-16.08	17.59	72.41

Los resultados muestran una **mejora sistemática del desempeño** en todas las métricas luego de la optimización de parámetros. El valor de NSE au-

mentó de 0.67 a 0.69, lo que representa una mejora relativa de aproximadamente **2.9 %** en la eficiencia de simulación. De forma consistente, tanto el RMSE como el MAE se reducen, indicando un ajuste más preciso del caudal simulado frente al observado, con una disminución más notoria en el error medio absoluto.

Asimismo, el sesgo (PBIAS) muestra una corrección adicional, pasando de -17.34% a -16.08% , lo que implica una reducción en la tendencia a la subestimación del caudal. Si bien la magnitud de la mejora no es drástica, el avance resulta **hidrológicamente coherente** y evidencia que el proceso de optimización permitió refinar el comportamiento del modelo sin comprometer su estabilidad física.

En conjunto, estos resultados confirman que la configuración optimizada ofrece un **mejor equilibrio entre precisión y consistencia hidrológica**, consolidando su adopción como configuración **baseline definitiva** para las comparaciones con modelos posteriores.

Análisis temporal de las simulaciones

La Figura 6.2 muestra la comparación temporal entre los caudales observados y los simulados por el modelo GR4J bajo las dos configuraciones analizadas (referencial y optimizada) durante el período de prueba correspondiente al 20 % final del registro.

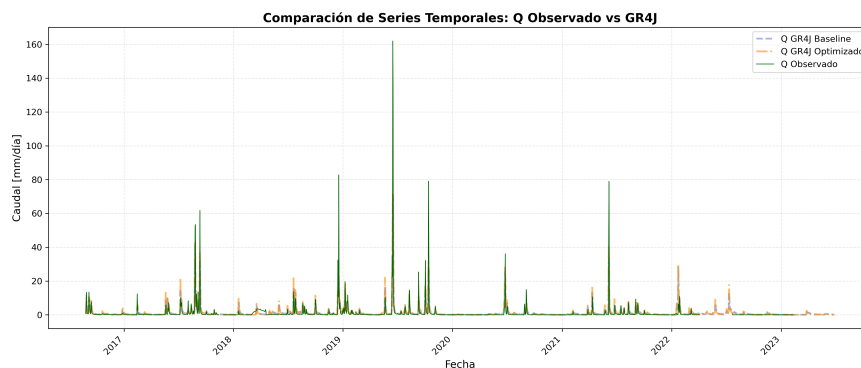


Figura 6.2: Comparación temporal entre los caudales observados y los simulados por el modelo GR4J (configuración base y optimizada).

Se observa que ambas configuraciones del modelo reproducen adecuadamente la dinámica general del caudal en la estación *Florida (Puente Ruta 5)*. Los períodos de estiaje y los principales eventos de crecida son correctamente capturados, lo que refleja la capacidad del modelo para representar los procesos de almacenamiento y liberación del agua en la cuenca.

Con el fin de profundizar en el análisis comparativo, la Figura 6.3 presenta un acercamiento al año 2021, período en el que se registraron varios eventos de crecida de distinta magnitud.

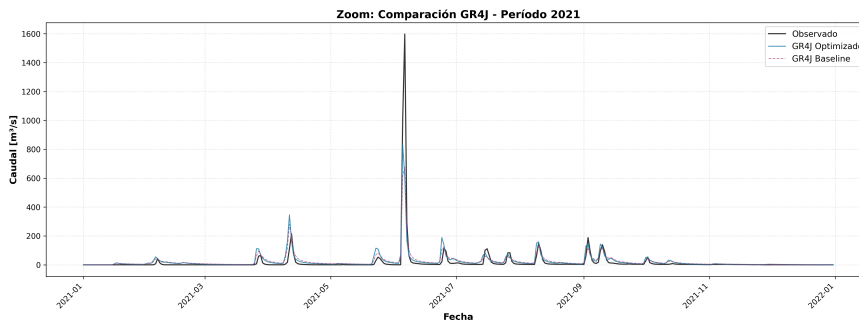


Figura 6.3: Detalle temporal del año 2021 para la comparación entre caudales observados y simulados por el modelo GR4J (configuración base y optimizada).

El análisis detallado del año 2021 evidencia que la configuración optimizada presenta una mejor representación de los eventos de crecida de magnitud media, reproduciendo con mayor precisión las pendientes de ascenso y recesión, así como una leve mejora en la magnitud de los picos simulados.

No obstante, durante el evento extremo registrado a mediados de año, ambas configuraciones subestiman significativamente el caudal máximo observado, mostrando una respuesta suavizada característica del modelo. Si bien la versión optimizada reduce parcialmente esta subestimación respecto a la configuración base, persiste una limitación estructural en la simulación de crecidas de alta intensidad.

Costo computacional

La simulación correspondiente a todo el período del conjunto de datos, utilizando los parámetros optimizados obtenidos en la etapa de calibración, demandó un tiempo total de **0.09 segundos** en el entorno de ejecución empleado. Este tiempo incluye la generación completa de la serie de caudales simulados.

6.4. MLP

En esta sección se presentan los resultados del modelo de predicción de caudales basado en redes neuronales *Multilayer Perceptron* (MLP). Se reporta el desempeño del modelo base, se resume el proceso de optimización aplicado (según la metodología del Capítulo correspondiente) y se presenta la evaluación final del modelo optimizado sobre el conjunto independiente de prueba. El detalle completo de los experimentos iterativos (incorporación/selección de variables, análisis de correlación, horizonte de *lags*, *Integrated Gradients* y Optuna) se incluye en el Anexo B.1.

6.4.1. Desempeño del modelo base

El primer experimento consistió en entrenar el MLP con la arquitectura fija descrita en la Sección 5.6.1, utilizando todas las variables iniciales sin selección ni optimización de hiperparámetros.

Tabla 6.4: Desempeño del modelo MLP base sobre el conjunto de validación

NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
0.06	14.91	32.84	103.61

Los resultados del modelo base (Tabla 6.4) muestran un desempeño limitado: el $NSE = 0.06$ indica que la red apenas mejora respecto a un predictor ingenuo basado en la media, mientras que el sesgo positivo ($PBIAS = 14.91\%$) sugiere una tendencia sistemática a sobreestimar caudales.

Este comportamiento es consistente con un modelo que no está capturando adecuadamente la dinámica hidrológica subyacente. En particular, una limitación esperable de esta configuración es la falta de memoria explícita: el caudal diario responde a precipitación acumulada y condiciones previas de la cuenca, por lo que la inclusión de *lags* es crítica. Esto motiva el proceso de optimización posterior, que se detalla en el Anexo B.1.

6.4.2. Síntesis del proceso de optimización

Siguiendo el procedimiento metodológico definido previamente, se realizaron experimentos controlados para mejorar el desempeño del MLP en validación, priorizando además la coherencia hidrológica y el control del sesgo. En síntesis, las decisiones que resultaron determinantes fueron:

- **Incorporación de memoria hidrológica:** la inclusión de *lags* de precipitación produjo la mejora más significativa del desempeño. La evaluación del horizonte temporal mostró que el modelo se beneficia claramente al incorporar varios días de precipitación previa, alcanzando una meseta de rendimiento alrededor de 8 días (ver Anexo B.1.2).
- **Representación de *lags* con precipitación areal:** se adoptó la precipitación ponderada mediante polígonos de Thiessen para construir los *lags*, al presentar el mejor compromiso entre precisión y estabilidad (Anexo B.1.2).
- **Selección y parsimonia de predictores:** se redujo redundancia en variables térmicas y radiativas altamente correlacionadas, manteniendo un conjunto parsimonioso sin deterioro relevante del ajuste, y con mejoras en sesgo (Anexo B.1.3).
- **Interpretabilidad (IG):** el análisis de *Integrated Gradients* confirmó que el modelo se apoya principalmente en la precipitación reciente (*lags* 1–2) y utiliza variables energéticas como moduladores secundarios (Anexo B.1.4).

- **Optimización de hiperparámetros (Optuna):** la búsqueda bayesiana identificó una arquitectura parsimoniosa de dos capas ocultas (Anexo B.1.5).

6.4.3. Resultados del modelo optimizado

Configuración final del modelo

El modelo final del MLP fue entrenado utilizando el conjunto de variables seleccionado tras el proceso de optimización detallado en el Anexo B.1. En particular, la configuración adoptada incluye:

- **Variables temporales:** codificación cíclica del mes (`month_sin`, `month_cos`).
- **Precipitación:** precipitación areal diaria ponderada mediante polígonos de Thiessen, junto con *lags* de 1 a 8 días.
- **Variables meteorológicas:** temperatura media del aire, radiación solar, heliofanía, humedad relativa media y velocidad del viento.

La arquitectura óptima, identificada mediante optimización bayesiana con Optuna, corresponde a una MLP con dos capas ocultas de 56 y 391 neuronas, activación `gelu`, tasa de aprendizaje 1.32×10^{-3} , *weight decay* de 4.1×10^{-5} , *dropout* base de 0.20 y *batch size* de 32. El entrenamiento se realizó con un máximo de 200 épocas y *early stopping* basado en la pérdida de validación.

Evaluación del modelo final optimizado

La Tabla 6.5 resume el desempeño del MLP optimizado. En el conjunto de prueba se obtiene $NSE = 0.67$, considerado *bueno* según Moriasi y cols. (2015), y un sesgo $PBIAS = 6.52\%$ dentro del rango *muy bueno*.

Tabla 6.5: Desempeño del modelo MLP optimizado en entrenamiento, validación y prueba.

Conjunto	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Entrenamiento	0.87	10.80	11.39	35.16
Validación	0.74	10.65	16.01	55.00
Prueba	0.67	6.52	15.96	73.57

Para enfatizar la mejora lograda respecto al modelo base, la Tabla 6.6 compara ambas configuraciones en el conjunto de validación.

En términos cuantitativos, el MLP optimizado reduce el MAE aproximadamente a la mitad en el conjunto de validación y aumenta de forma sustantiva la eficiencia (NSE), lo que evidencia el impacto de incorporar memoria pluviométrica y de optimizar la arquitectura.

Tabla 6.6: Comparación en conjunto de validación: MLP base vs. MLP optimizado.

Modelo	NSE	PBIAS [%]	MAE [m^3/s]	RMSE [m^3/s]
Base	0.06	14.91	32.84	103.61
Optimizado	0.74	10.65	16.01	55.00

Si bien el RMSE también disminuye de manera significativa, su magnitud relativa respecto al MAE sugiere que los errores más elevados continúan asociados a eventos de mayor magnitud, característicos de crecidas.

Desde un punto de vista hidrológico, la comparación temporal (Figura 6.4) muestra que el modelo reproduce adecuadamente la estacionalidad y la ocurrencia de la mayoría de los pulsos de caudal. Sin embargo, durante eventos extremos se observa una subestimación de los picos máximos, particularmente en el evento de mayor magnitud del período analizado, lo cual explica el aumento del RMSE.

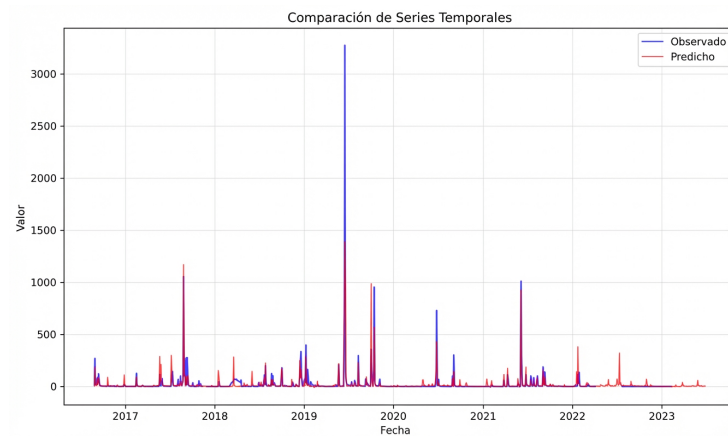


Figura 6.4: Serie temporal de caudales observados y predichos por el modelo MLP (conjunto de prueba).

El *scatterplot* en escala logarítmica (Figura 6.5) evidencia una buena alineación con la recta 1:1 en rangos medios, mientras que para caudales altos aumenta la dispersión y predominan puntos por debajo de la diagonal, confirmando la tendencia a subestimar crecidas extremas.

Finalmente, el análisis de atribuciones en prueba (Figura 6.6) confirma que la señal dominante proviene de los **lags de precipitación** (especialmente 1–2 días), mientras que variables energéticas como la radiación solar actúan como moduladores secundarios. Este patrón es consistente con la interpretación hidrológica: el caudal diario está fuertemente condicionado por la precipitación reciente, con efectos secundarios asociados al balance energético.

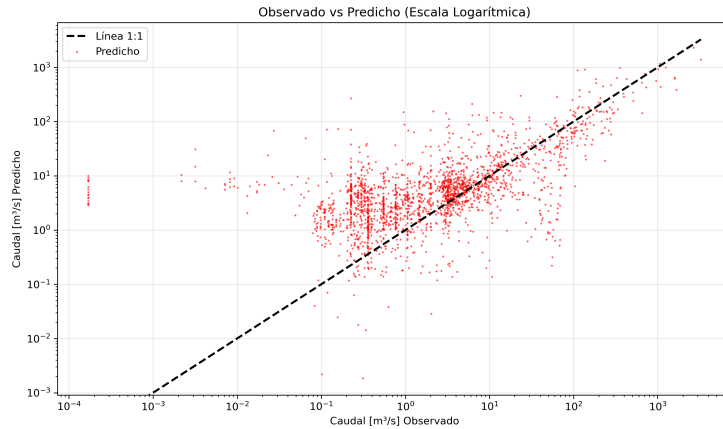


Figura 6.5: *Scatterplot* (escala logarítmica) para el modelo MLP optimizado.

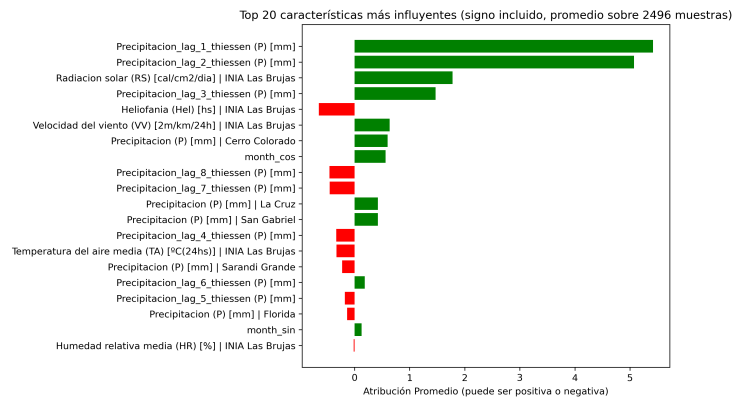


Figura 6.6: Top 20 características más influyentes según *Integrated Gradients* en el conjunto de prueba.

Costo computacional El entrenamiento completo del modelo MLP optimizado demandó 45.49 segundos en un total de 101 épocas en el entorno de cómputo utilizado. El tiempo de inferencia sobre los tres conjuntos fue despreciable a la precisión reportada por el sistema.

6.5. RNN

En esta sección se presentan los resultados del modelo de predicción de caudales basado en una arquitectura recurrente (RNN→MLP). Se reporta el desempeño del modelo base, se resume el proceso de optimización aplicado (según la metodología del Capítulo correspondiente) y se presenta la evaluación final del modelo optimizado sobre el conjunto independiente de prueba. El detalle completo de los experimentos iterativos (incorporación/selección de variables, análisis de correlación, horizonte de ventana temporal, *Integrated Gradients* y Optuna) se incluye en el Anexo B.2.

6.5.1. Desempeño del modelo base

El primer experimento consistió en entrenar la arquitectura fija definida en la Sección 5.7.1, utilizando todas las variables iniciales sin selección ni optimización de hiperparámetros.

Tabla 6.7: Desempeño del modelo RNN base sobre el conjunto de validación.

NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
0.04	14.17	33.35	104.85

Los resultados del modelo base (Tabla 6.7) muestran un desempeño limitado: el $NSE = 0.04$ indica que la red apenas mejora respecto a un predictor ingenuo basado en la media, mientras que el sesgo positivo ($PBIAS = 14.17\%$) sugiere una tendencia sistemática a subestimar caudales.

Este comportamiento es consistente con un modelo que, al procesar una secuencia unitaria (ventana de 1 día), no logra capturar las dependencias temporales del sistema hidrológico. Esto motiva el proceso de optimización posterior, que se detalla en el Anexo B.2.

6.5.2. Síntesis del proceso de optimización

Siguiendo el procedimiento metodológico definido previamente, se realizaron experimentos controlados para mejorar el desempeño del RNN en validación, priorizando además la coherencia hidrológica y el control del sesgo. En síntesis, las decisiones que resultaron determinantes fueron:

- **Variables temporales:** se conservó únicamente la codificación cíclica del mes (`month_sin`, `month_cos`), cuya inclusión ofreció la mayor estabilidad

entre las configuraciones evaluadas y contribuyó a reducir parcialmente el sesgo (Anexo B.2).

- **Representación espacial de la precipitación:** se mantuvo el enfoque de **estaciones individuales** (sin agregación espacial), al igual que lo adoptado en el modelo MLP. Si bien los métodos de agregación reducían ligeramente los errores medios, introducían un sesgo sistemático considerablemente mayor ($PBIAS > 23\%$); las estaciones individuales preservan la variabilidad espacial con el menor sesgo ($PBIAS = 12.94\%$) (Anexo B.2).
- **Incorporación de memoria temporal:** la selección de la ventana temporal de entrada produjo la mejora más significativa del desempeño. Se adoptó una **ventana de 9 días**, que alcanzó $NSE = 0.61$ con el menor $RMSE$ del grupo de ventanas largas y un sesgo relativamente controlado (Anexo B.2).
- **Selección y parsimonia de predictores:** tanto la depuración por correlación como el análisis de *Integrated Gradients* confirmaron que el conjunto completo de variables ofrece el mejor desempeño. Incluso la eliminación de las variables temporales (con las atribuciones más bajas según IG) produjo una degradación apreciable (NSE de 0.61 a 0.55) (Anexo B.2).
- **Interpretabilidad (IG):** el análisis de *Integrated Gradients* mostró que la velocidad del viento y la radiación solar son los predictores dominantes, seguidos por la humedad relativa y múltiples estaciones de precipitación (Anexo B.2).
- **Optimización de hiperparámetros (Optuna):** la búsqueda bayesiana (209 *trials*, 130 completados, 79 podados) identificó una celda RNN simple con 4 capas y 44 unidades ocultas, seguida de un MLP de 4 capas con activación `leaky_relu` y *batch size* de 16 (Anexo B.2).

6.5.3. Resultados del modelo optimizado

Configuración final del modelo

El modelo final del RNN fue entrenado utilizando el conjunto de variables seleccionado tras el proceso de optimización detallado en el Anexo B.2. En particular, la configuración adoptada incluye:

- **Variables temporales:** codificación cíclica del mes (`month_sin`, `month_cos`).
- **Precipitación:** precipitación diaria de todas las estaciones individuales, sin agregación espacial.
- **Variables meteorológicas:** radiación solar, heliofanía, temperatura media/máxima/mínima, humedad relativa media, velocidad del viento y evapotranspiración de Penman.
- **Ventana temporal:** 9 días.

La arquitectura óptima, identificada mediante optimización bayesiana con Optuna, corresponde a una red RNN de 4 capas con 44 unidades ocultas (*dropout* de 0.14), seguida de un MLP de 4 capas [323, 34, 32, 277] con activación `leaky_relu`, tasa de aprendizaje 1.29×10^{-4} , *dropout* base de 0.11, *weight decay* de 1.19×10^{-8} y *batch size* de 16.

Evaluación del modelo final optimizado

La Tabla 6.8 resume el desempeño del RNN optimizado. En el conjunto de prueba se obtiene $NSE = 0.59$, clasificado como *satisfactorio* según Moriasi y cols. (2015), y un sesgo $PBIAS = -15.99\%$ dentro del rango *satisfactorio*, que indica una tendencia a la sobreestimación del caudal.

Tabla 6.8: Desempeño del modelo RNN optimizado en entrenamiento, validación y prueba.

Conjunto	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Entrenamiento	0.69	-29.15	15.03	55.20
Validación	0.69	0.47	17.46	59.56
Prueba	0.59	-15.99	18.99	82.81

Para enfatizar la mejora lograda respecto al modelo base, la Tabla 6.9 compara ambas configuraciones en el conjunto de validación.

Tabla 6.9: Comparación en conjunto de validación: RNN base vs. RNN optimizado.

Modelo	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Base	0.04	14.17	33.35	104.85
Optimizado	0.69	0.47	17.46	59.56

En términos cuantitativos, el RNN optimizado reduce el MAE aproximadamente a la mitad y aumenta de forma sustantiva la eficiencia (NSE de 0.04 a 0.69 en validación), confirmando el impacto de incorporar la ventana temporal y optimizar la arquitectura. No obstante, en el conjunto de prueba el desempeño disminuye a $NSE = 0.59$, y el $PBIAS = -15.99\%$ revela una tendencia sistemática a la sobreestimación del caudal.

Desde un punto de vista hidrológico, la comparación temporal (Figura 6.7) muestra que el modelo reproduce adecuadamente la dinámica general y la ocurrencia de eventos, pero presenta una compresión del rango dinámico: las crecidas de mayor magnitud tienden a ser subestimadas, mientras que algunos eventos intermedios aparecen sobredimensionados.

El *scatterplot* en escala logarítmica (Figura 6.8) muestra una dispersión considerable en el rango medio y alto de caudales. Para valores elevados predomina la subestimación (puntos por debajo de la línea 1:1), mientras que en el rango

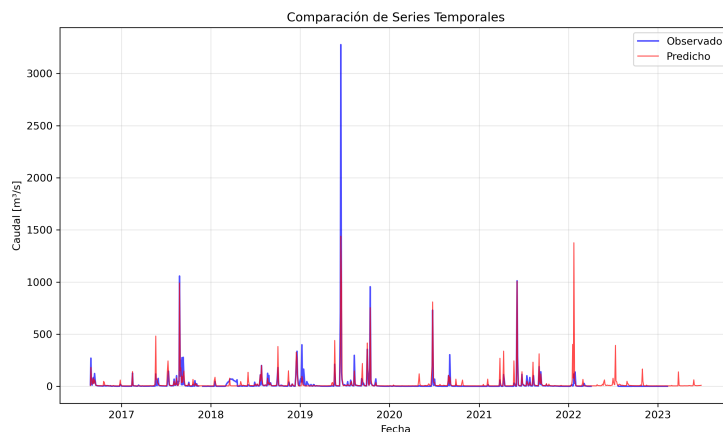


Figura 6.7: Serie temporal de caudales observados y predichos por el modelo RNN (conjunto completo).

intermedio se observan casos de sobreestimación, confirmando la distorsión del rango dinámico.

Además, se observa un patrón particular: una concentración de predicciones en un rango casi constante de caudal (aproximadamente $4\text{--}6\text{ m}^3/\text{s}$), al cual se asocian múltiples valores observados muy diferentes. Este comportamiento sugiere una tendencia del modelo a “colapsar” hacia valores medios ante determinadas condiciones, perdiendo sensibilidad frente a variaciones del caudal observado. Este efecto no se evidenciaba en el modelo MLP, y podría indicar limitaciones del RNN para capturar adecuadamente ciertos estados hidrológicos o una fuerte regresión hacia la media inducida por la función de pérdida.

Este tipo de patrón puede interpretarse como una forma de saturación del modelo o una restricción efectiva del espacio de salida, lo cual reduce la capacidad de representar extremos y contribuye tanto a la subestimación de crecidas como a la concentración de errores en determinados rangos.

Finalmente, el análisis de atribuciones en prueba (Figura 6.9) indica que la radiación solar presenta la mayor contribución positiva promedio, seguida por variables térmicas y de humedad. A diferencia del MLP, la señal de precipitación no domina claramente las atribuciones globales, lo que sugiere que el modelo recurrente distribuye la relevancia a lo largo de múltiples variables atmosféricas.

No obstante, debe considerarse que, en esta implementación, las *Integrated Gradients* se evaluaron únicamente sobre el primer paso temporal de la ventana de entrada. Dado que las redes recurrentes integran información a lo largo de toda la secuencia mediante su estado oculto, es posible que parte de la contribución asociada a la precipitación se encuentre distribuida en pasos temporales anteriores dentro de la ventana, los cuales no fueron descompuestos explícitamente en el análisis de atribuciones. En consecuencia, la interpretación presentada debe entenderse como una aproximación parcial a la dinámica completa de relevancia

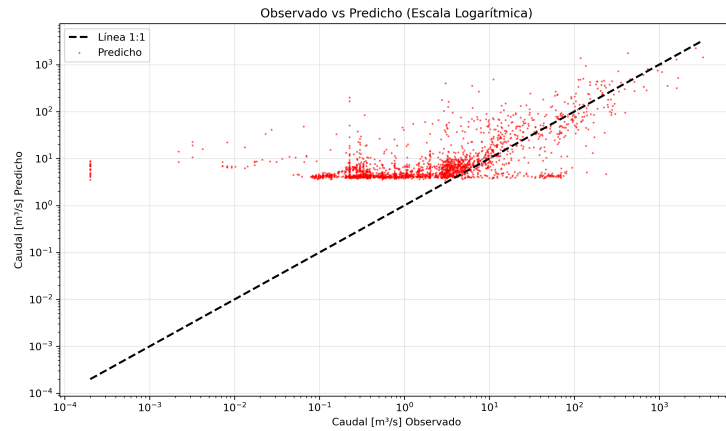


Figura 6.8: *Scatterplot* (escala logarítmica) para el modelo RNN optimizado.

temporal del modelo.

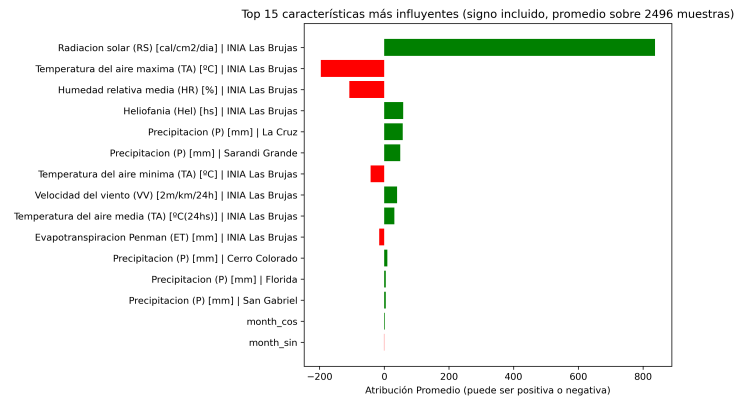


Figura 6.9: Top 15 características más influyentes según *Integrated Gradients* en el conjunto de prueba.

Costo computacional El entrenamiento completo del modelo RNN optimizado demandó 99.21 segundos en un total de 69 épocas en el entorno de cómputo utilizado. El tiempo de inferencia sobre los tres conjuntos fue de 0.01 segundos.

6.6. PINN-P MLP

En esta sección se presentan los resultados del modelo PINN-P con arquitectura MLP, donde la red neuronal predice exclusivamente la precipitación agregada diaria \hat{P}_{agg} , mientras que la evapotranspiración se mantiene fija a partir

de los valores observados. Se reporta el desempeño del modelo base, se resume el proceso de optimización aplicado y se presenta la evaluación final del modelo optimizado sobre el conjunto independiente de prueba. El detalle completo de los experimentos iterativos se incluye en el Anexo B.3.

6.6.1. Desempeño del modelo base

El primer experimento consistió en entrenar el modelo PINN-P (MLP) con la arquitectura base, utilizando todas las variables iniciales sin selección ni optimización de hiperparámetros. La predicción final de caudal emerge del acoplamiento físico con GR4J, evaluando $Q_{\text{GR4J}}(\hat{P}_{\text{agg}}, E_{\text{obs}}; \theta)$.

Tabla 6.10: Desempeño del modelo base PINN-P (MLP) sobre el conjunto de validación.

NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
0.66	13.30	15.41	62.35

El modelo base (Tabla 6.10) presenta un desempeño razonable ($NSE = 0.66$), con un sesgo positivo ($PBIAS = 13.30\%$) que indica una tendencia a la subestimación del volumen simulado.

6.6.2. Síntesis del proceso de optimización

Siguiendo el procedimiento metodológico definido previamente, se realizaron experimentos controlados para mejorar el desempeño del modelo. Las decisiones determinantes fueron:

- **Variables temporales:** se conservó la codificación cíclica del **mes** (`month_sin`, `month_cos`), cuya inclusión produjo la mejora más consistente en desempeño (NSE de 0.66 a 0.67) y redujo significativamente el sesgo ($PBIAS$ de 13.30% a 9.41%) (Anexo B.3).
- **Lags de precipitación:** la incorporación de *lags* de precipitación bajo distintos esquemas de representación espacial produjo un **deterioro sistemático** del desempeño, con incrementos del $PBIAS$ hasta valores superiores al 21%. Se decidió no incluir memoria temporal explícita en la formulación PINN-P (Anexo B.3).
- **Selección de variables (correlación e IG):** el análisis de correlación no justificó la eliminación de variables, y el análisis de *Integrated Gradients* mostró que radiación solar y velocidad del viento concentran la mayor atribución. La eliminación de variables temporales (atribuciones bajas según IG) deterioró el desempeño, por lo que se mantuvo el conjunto completo de variables (Anexo B.3).

- **Balance de la función de pérdida:** el desempeño resultó notablemente estable frente a variaciones en la relación $w_{\text{data}} : w_{\text{phys}}$ ($NSE \approx 0.67\text{--}0.68$ en todas las configuraciones). Se mantuvo la configuración balanceada $w_{\text{data}} = 1$, $w_{\text{phys}} = 1$ (Anexo B.3).
- **Optimización de hiperparámetros (Optuna):** la búsqueda bayesiana (202 *trials* completados, 26 podados) identificó una arquitectura de 3 capas ocultas [61, 35, 53] con activación `relu`, *dropout* de 0.31 y *batch size* de 32 (Anexo B.3).

6.6.3. Resultados del modelo optimizado

Configuración final del modelo

El modelo final fue entrenado con el conjunto de variables seleccionado tras el proceso de optimización. La configuración adoptada incluye:

- **Precipitación:** precipitaciones contemporáneas por estación (sin *lags*).
- **Variables meteorológicas:** radiación solar, heliofanía, evapotranspiración, temperatura media/máxima/mínima, humedad relativa y velocidad del viento.
- **Variables temporales:** codificación cíclica del mes (`month_sin`, `month_cos`).

La arquitectura óptima, identificada mediante Optuna, corresponde a una MLP de 3 capas ocultas [61, 35, 53] con activación `relu`, tasa de aprendizaje 4.89×10^{-3} , *dropout* base de 0.31, *weight decay* de 3.56×10^{-5} y *batch size* de 32. El balance de la función de pérdida se mantuvo en $w_{\text{data}} = 1$, $w_{\text{phys}} = 1$.

Evaluación del modelo final optimizado

La Tabla 6.11 resume el desempeño del modelo PINN-P (MLP) optimizado, mientras que la Tabla 6.12 compara los resultados frente al modelo base inicial. En el conjunto de prueba se obtiene $NSE = 0.74$, considerado *bueno* según Moriasi y cols. (2015) y un $PBIAS = -35.55\%$ considerado *insatisfactorio* y mostrando una tendencia a la sobreestimación del caudal.

Tabla 6.11: Desempeño del modelo PINN-P (MLP) optimizado en entrenamiento, validación y prueba.

Conjunto	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Entrenamiento	0.51	-68.14	22.79	71.08
Validación	0.75	-16.73	15.76	53.80
Prueba	0.74	-35.55	18.17	65.98

Tabla 6.12: Comparación en conjunto de validación: PINN-P MLP base vs. optimizado.

Modelo	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Base	0.66	13.30	15.41	62.35
Optimizado	0.75	-16.73	15.76	53.80

El modelo optimizado mejora sustancialmente la eficiencia respecto al base (NSE de 0.66 a 0.74 en prueba) y presenta una adecuada capacidad de generalización (NSE de 0.75 en validación a 0.74 en prueba). No obstante, presenta un sesgo volumétrico negativo significativo ($PBIAS = -35.55\%$ en prueba), que indica una tendencia sistemática a la sobreestimación del volumen total de caudal. Este sesgo contrasta con el comportamiento del modelo base ($PBIAS = 13.30\%$ positivo), sugiriendo que la optimización arquitectónica introduce un compromiso entre eficiencia estadística y conservación del volumen.

En el conjunto de entrenamiento, las métricas resultan inferiores a las de validación ($NSE = 0.51$ vs 0.75), comportamiento que se explica por la naturaleza compuesta de la función de pérdida en las PINNs: el mínimo de validación no necesariamente coincide con un mínimo del error de caudal en entrenamiento.

La comparación temporal (Figura 6.10) muestra que el modelo reproduce adecuadamente la dinámica general de la serie, capturando la ocurrencia de eventos de crecida y la estacionalidad, aunque tiende a suavizar algunos picos máximos y a sobreestimar varios de los picos pequeños y medianos, lo que explica el PBIAS negativo.

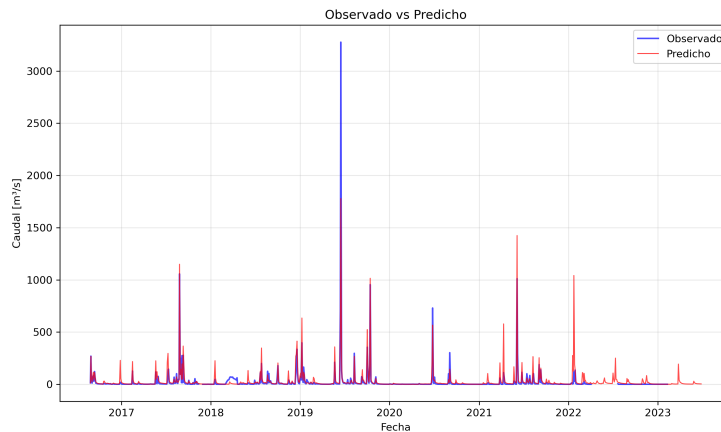


Figura 6.10: Serie temporal de caudales observados y predichos por el modelo PINN-P (MLP) optimizado.

El *scatterplot* en escala logarítmica (Figura 6.11) muestra buena alineación con la línea 1:1 en el rango medio de caudales, aunque con una tendencia a la

sobreestimación, mientras que para valores extremos se observa mayor dispersión y tendencia a la subestimación.

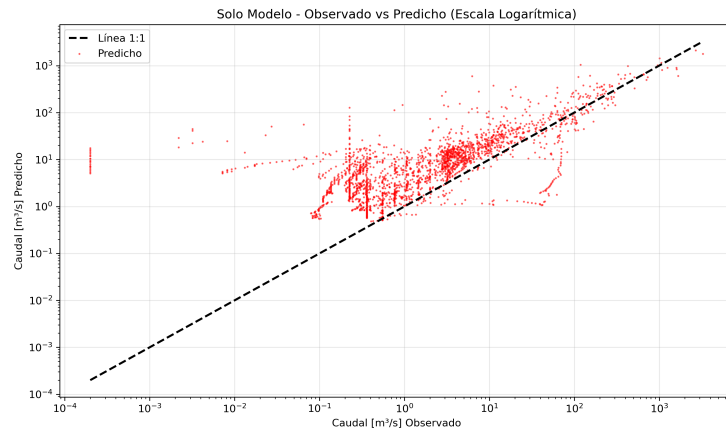


Figura 6.11: *Scatterplot* (escala logarítmica) para el modelo PINN-P (MLP) optimizado.

El análisis de atribuciones en prueba (Figura 6.12) muestra un cambio significativo respecto a las etapas exploratorias: las precipitaciones contemporáneas por estación concentran la mayor atribución positiva, mientras que las variables energéticas pasan a un rol secundario. Este patrón, alineado con la interpretación hidrológica clásica, resulta coherente con el hecho de que el modelo predice P , por lo que es esperable que las mayores atribuciones se concentren en las variables de precipitación correspondientes a las distintas estaciones.

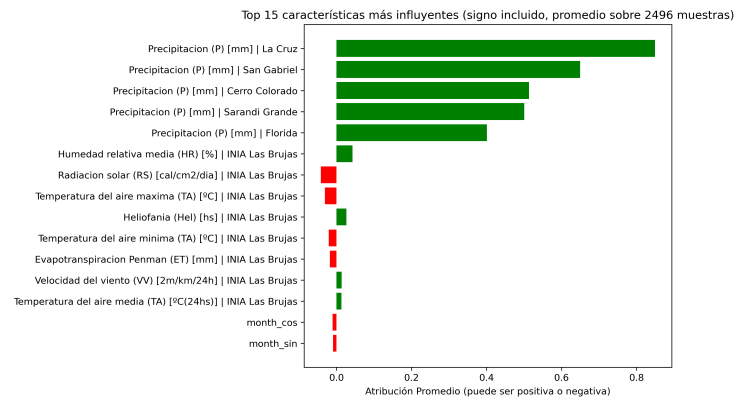


Figura 6.12: Top 15 características más influyentes según *Integrated Gradients* en el conjunto de prueba (PINN-P MLP).

Costo computacional El entrenamiento completo del modelo optimizado demandó 265.83 segundos en un total de 23 épocas. El tiempo de inferencia sobre todo el conjunto fue de 5.22 segundos.

6.7. PINN-P RNN

En esta sección se presentan los resultados del modelo PINN-P con arquitectura recurrente, donde la red neuronal predice exclusivamente la precipitación agregada diaria \hat{P}_{agg} , mientras que la evapotranspiración se mantiene fija a partir de los valores observados. Se reporta el desempeño del modelo base, se resume el proceso de optimización aplicado y se presenta la evaluación final del modelo optimizado sobre el conjunto independiente de prueba. El detalle completo de los experimentos iterativos se incluye en el Anexo B.4.

6.7.1. Desempeño del modelo base

El primer experimento consistió en entrenar el modelo PINN-P con arquitectura RNN base, utilizando todas las variables iniciales sin selección ni optimización de hiperparámetros. La configuración base empleó una secuencia de longitud 1, dos capas LSTM con 64 unidades ocultas y un módulo MLP posterior de tres capas (128→64→32) con activación ReLU.

Tabla 6.13: Desempeño del modelo base PINN-P sobre el conjunto de validación.

NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
0.71	0.13	15.89	57.25

El modelo base (Tabla 6.13) presenta un desempeño superior al de la PINN-P MLP ($NSE = 0.71$ vs 0.66), con un sesgo prácticamente nulo ($PBIAS = 0.13\%$) que contrasta con el 13.30% de la versión MLP.

6.7.2. Síntesis del proceso de optimización

Siguiendo el procedimiento metodológico definido previamente, se realizaron experimentos controlados para mejorar el desempeño del modelo. Las decisiones determinantes fueron:

- **Variables temporales:** a diferencia de la versión MLP (donde el mes mejoraba el desempeño), la incorporación de variables temporales **degradó consistentemente** todas las métricas en la RNN, incluyendo un aumento significativo del sesgo. Se decidió no incluir variables temporales (Anexo B.4).
- **Ventana temporal:** se adoptó una **ventana de 4 días**, que ofrece el mejor equilibrio entre eficiencia ($NSE = 0.70$) y control del sesgo

($PBIAS = -7.96\%$). La ventana unitaria alcanzaba mayor NSE (0.71) pero anula la capacidad recurrente del modelo (Anexo B.4).

- **Selección de variables (correlación e IG):** el análisis de correlación y de *Integrated Gradients* confirmaron que el conjunto completo de 12 variables ofrece el mejor desempeño. La radiación solar y velocidad del viento concentran la mayor atribución, con un patrón altamente consistente con la versión MLP (Anexo B.4).
- **Balance de la función de pérdida:** a diferencia de la MLP (donde el balance 1:1 era óptimo), la LSTM se beneficia de una **predominancia del término de datos**: $w_{\text{data}} = 1$, $w_{\text{phys}} = 0.25$, que alcanza $NSE = 0.71$ con sesgo prácticamente nulo ($PBIAS = -0.36\%$) (Anexo B.4).
- **Optimización de hiperparámetros (Optuna):** la búsqueda bayesiana (114 *trials* completados, 16 podados) identificó una celda LSTM con 1 capa y 70 unidades ocultas, seguida de un MLP de 1 capa con 390 neuronas, activación `gelu` y *batch size* de 16 (Anexo B.4).

6.7.3. Resultados del modelo optimizado

Configuración final del modelo

El modelo final fue entrenado con el conjunto de variables seleccionado tras el proceso de optimización. La configuración adoptada incluye:

- **Precipitación:** precipitaciones contemporáneas por estación (sin *lags*).
- **Variables meteorológicas:** radiación solar, heliofanía, temperatura media/máxima/mínima, humedad relativa y velocidad del viento.
- **Sin variables temporales.**
- **Ventana temporal:** 4 días.

La arquitectura óptima, identificada mediante Optuna, corresponde a una red LSTM de 1 capa con 70 unidades ocultas, seguida de un MLP de 1 capa con 390 neuronas y activación `gelu`, tasa de aprendizaje 2.75×10^{-3} , *dropout* base de 0.44, *weight decay* de 6.98×10^{-4} y *batch size* de 16. El balance de la función de pérdida se fijó en $w_{\text{data}} = 1$, $w_{\text{phys}} = 0.25$.

Evaluación del modelo final optimizado

La Tabla 6.14 resume el desempeño del modelo PINN-P optimizado. En el conjunto de prueba se obtiene $NSE = 0.67$, considerado *bueno* según Moriasi y cols. (2015) y un $PBIAS = -22.64\%$ considerado *insatisfactorio*.

La Tabla 6.15 compara el modelo optimizado con la configuración base en el conjunto de validación.

Tabla 6.14: Desempeño del modelo PINN-P optimizado en entrenamiento, validación y prueba.

Conjunto	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Entrenamiento	0.53	-42.53	19.00	69.72
Validación	0.64	-6.93	18.24	64.23
Prueba	0.67	-22.64	17.85	74.33

Tabla 6.15: Comparación: PINN-P RNN base vs. optimizado en conjunto de validación.

Modelo	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Base	0.71	0.13	15.89	57.25
Optimizado	0.64	-6.93	18.24	64.23

A diferencia de lo esperado, la configuración optimizada no supera al modelo base en el conjunto de validación. Se observa una disminución del NSE (de 0.71 a 0.64) y un deterioro en las métricas de error (MAE y RMSE), junto con la introducción de un sesgo negativo. Esto sugiere que el proceso de optimización pudo haber inducido un sobreajuste a ciertas dinámicas del conjunto de entrenamiento, afectando la capacidad de generalización del modelo.

El modelo optimizado presenta un $NSE = 0.67$ en prueba con un sesgo negativo significativo ($PBIAS = -22.64\%$), lo que indica una tendencia sistemática a la sobreestimación del volumen total de caudal. Si bien en validación el sesgo es más acotado ($PBIAS = -6.93\%$), la generalización temporal reintroduce un desbalance, particularmente asociado a eventos de bajo caudal, donde se concentra la mayor parte de la sobreestimación volumétrica.

En comparación con la versión MLP del modelo PINN-P optimizado ($NSE = 0.74$, $PBIAS = -35.55\%$ en prueba), la versión RNN presenta una eficiencia inferior, aunque con un sesgo menos pronunciado. Esto sugiere un intercambio entre eficiencia y balance volumétrico, manteniéndose en ambos casos la tendencia a la sobreestimación en el conjunto de prueba.

Asimismo, el modelo PINN-P (RNN) **no supera al modelo GR4J directo** en el conjunto de prueba ($NSE = 0.67$ frente a 0.69), lo que indica que la incorporación de una capa intermedia de estimación de precipitación mediante capas recurrentes no se traduce en una mejora neta respecto al modelo conceptual forzado con precipitación observada. Este resultado sugiere que el error introducido en la etapa de estimación de P compensa o incluso supera las posibles ganancias derivadas de la mayor flexibilidad del esquema híbrido.

En el conjunto de entrenamiento, las métricas resultan inferiores a las de validación y prueba ($NSE = 0.53$, $PBIAS = -42.53\%$), comportamiento consistente con la naturaleza compuesta de la función de pérdida en las PINNs, donde la optimización simultánea de términos de ajuste a datos y restricciones

físicas puede generar compromisos que penalizan las métricas tradicionales de desempeño hidrológico.

La comparación temporal (Figura 6.13) muestra que el modelo reproduce adecuadamente la dinámica general de la serie, capturando los períodos de estiaje y la ocurrencia de eventos de crecida intermedios. Sin embargo, se observa una subestimación marcada en eventos extremos, con el evento de mediados de 2019 ($\sim 3\,300\text{ m}^3/\text{s}$) reproducido con un pico de apenas $\sim 1\,400\text{ m}^3/\text{s}$.

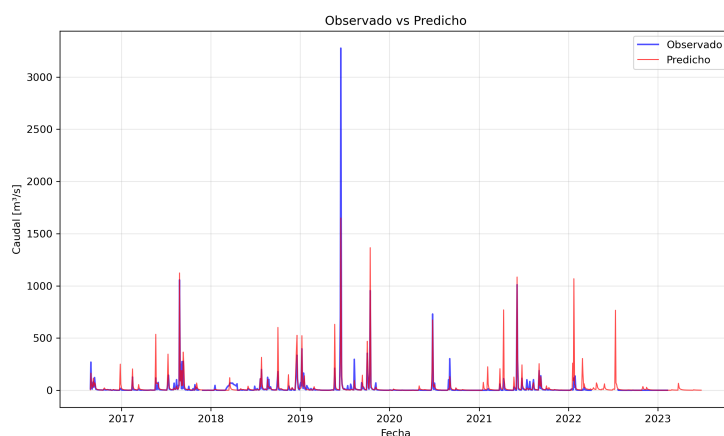


Figura 6.13: Serie temporal de caudales observados y predichos por el modelo PINN-P (RNN) optimizado.

El *scatterplot* en escala logarítmica (Figura 6.14) muestra alineación razonable con la línea 1:1 en el rango medio de caudales, con tendencia a la subestimación en caudales extremos y sobrepredicción en caudales bajos.

El análisis de atribuciones en prueba (Figura 6.15) muestra que la **radiación solar** concentra la mayor atribución positiva y la **velocidad del viento** la mayor atribución negativa, manteniendo el mismo patrón observado en las etapas exploratorias. A diferencia de la versión MLP optimizada (donde la precipitación pasaba a dominar las atribuciones), la RNN mantiene una dependencia predominante de variables energéticas y aerodinámicas.

Es importante destacar que, al igual que en otras arquitecturas recurrentes, el cálculo de *Integrated Gradients* se realizó sin considerar las $W - 1$ dimensiones adicionales de la ventana temporal (4 días). El contexto secuencial completo que utiliza el modelo no queda reflejado en las atribuciones calculadas, por lo que estos resultados deben interpretarse como una aproximación orientativa de la relevancia relativa de las variables.

Costo computacional El entrenamiento completo del modelo optimizado demandó 326.23 segundos en un total de 28 épocas. El tiempo de inferencia fue de 4.07 segundos.

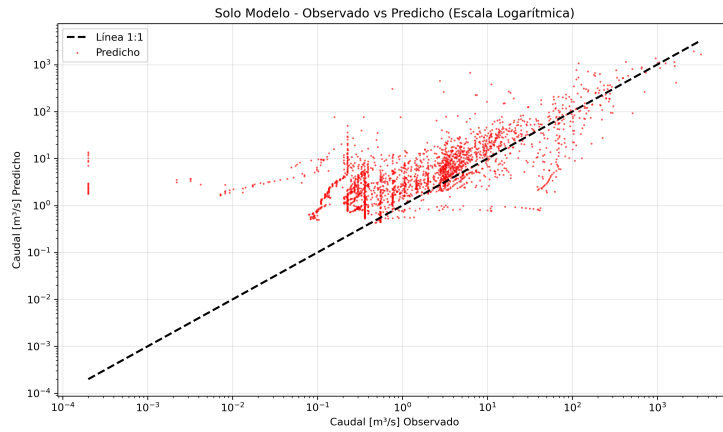


Figura 6.14: *Scatterplot* (escala logarítmica) para el modelo PINN-P (RNN) optimizado.

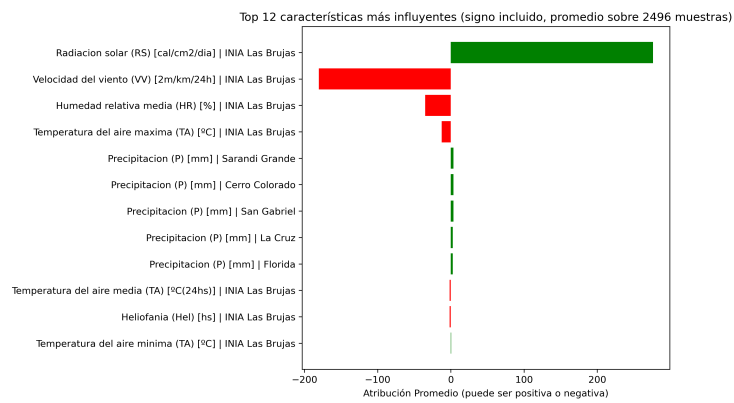


Figura 6.15: Características más influyentes según *Integrated Gradients* en el conjunto de prueba (PINN-P RNN).

6.8. PINN-E MLP

En esta sección se presentan los resultados del modelo PINN-E con arquitectura MLP, donde la red neuronal predice exclusivamente la evapotranspiración agregada diaria \hat{E}_{agg} , mientras que la precipitación se mantiene fija a partir de los valores observados. Se reporta el desempeño del modelo base, se resume el proceso de optimización aplicado y se presenta la evaluación final del modelo optimizado sobre el conjunto independiente de prueba. El detalle completo de los experimentos iterativos se incluye en el Anexo B.5.

6.8.1. Desempeño del modelo base

El primer experimento consistió en entrenar el modelo PINN-E (MLP) con la arquitectura base, utilizando todas las variables iniciales sin selección ni optimización de hiperparámetros. La predicción final de caudal emerge del acoplamiento físico con GR4J, evaluando $Q_{GR4J}(P_{Thiessen}, \hat{E}_{agg}; \theta)$.

Tabla 6.16: Desempeño del modelo base PINN-E (MLP) sobre el conjunto de validación.

NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
0.71	2.21	15.21	57.86

El modelo base (Tabla 6.16) presenta un desempeño razonable ($NSE = 0.71$), con un sesgo bajo ($PBIAS = 2.21\%$), lo que indica que el sistema acoplado es capaz de capturar una fracción importante de la dinámica de caudales.

6.8.2. Síntesis del proceso de optimización

Siguiendo el procedimiento metodológico definido previamente, se realizaron experimentos controlados para mejorar el desempeño del modelo. Las decisiones determinantes fueron:

- **Representación de la precipitación:** se adoptó el **promedio simple de estaciones** como forzante pluviométrico, al ofrecer el mejor compromiso entre eficiencia ($NSE = 0.72$) y parsimonia, reduciendo la dimensionalidad sin afectar el desempeño (Anexo B.5).
- **Memoria temporal:** se incorporaron **5 días de lags de evapotranspiración**, configuración que presentó el mejor control del sesgo volumétrico ($PBIAS \approx 0\%$) sin mejoras sustanciales más allá de ese horizonte (Anexo B.5).
- **Selección de variables (correlación e IG):** la evaluación por correlación no justificó eliminación de variables, pero el análisis de *Integrated*

Gradients mostró que las variables temporales (día y mes) tenían atribuciones cercanas a cero. Su eliminación mejoró consistentemente el desempeño (*NSE* de 0.72 a 0.74), por lo que fueron excluidas del conjunto final (Anexo B.5).

- **Balance de la función de pérdida:** el desempeño resultó estable frente a variaciones en la relación $w_{\text{data}} : w_{\text{phys}}$. Se mantuvo la configuración balanceada $w_{\text{data}} = 1$, $w_{\text{phys}} = 1$ (Anexo B.5).
- **Optimización de hiperparámetros (Optuna):** la búsqueda bayesiana (139 *trials* completados, 31 podados) identificó una arquitectura de 5 capas ocultas [81, 640, 979, 32, 83] con activación `relu` y *dropout* elevado (0.49) (Anexo B.5).

6.8.3. Resultados del modelo optimizado

Configuración final del modelo

El modelo final fue entrenado con el conjunto de variables seleccionado tras el proceso de optimización. La configuración adoptada incluye:

- **Precipitación:** promedio simple de estaciones.
- **Variables meteorológicas:** radiación solar, heliofanía, temperatura media/máxima/mínima, humedad relativa y velocidad del viento.
- **Memoria temporal:** 5 días de *lags* de evapotranspiración.
- **Sin variables temporales** (eliminadas tras análisis de IG).

La arquitectura óptima, identificada mediante Optuna, corresponde a una MLP de 5 capas ocultas [81, 640, 979, 32, 83] con activación `relu`, tasa de aprendizaje 1.26×10^{-3} , *dropout* base de 0.49, *weight decay* de 1.51×10^{-5} y *batch size* de 16. El balance de la función de pérdida se mantuvo en $w_{\text{data}} = 1$, $w_{\text{phys}} = 1$.

Evaluación del modelo final optimizado

La Tabla 6.17 resume el desempeño del modelo PINN-E optimizado, mientras que la Tabla 6.18 compara los resultados de dicho modelo con el base inicial. En el conjunto de prueba se obtiene $NSE = 0.71$, considerado *bueno* según Moriasi y cols. (2015) y un $PBIAS = -30.62\%$ considerado *insatisfactorio*.

El modelo optimizado mantiene la misma eficiencia que el base en prueba ($NSE = 0.71$) y mejora en validación ($NSE = 0.75$). No obstante, presenta un sesgo volumétrico negativo significativo ($PBIAS = -30.62\%$ en prueba), que indica una tendencia sistemática a la sobreestimación del volumen total de caudal. Este sesgo constituye una limitación relevante desde el punto de vista del balance hídrico.

Tabla 6.17: Desempeño del modelo PINN-E (MLP) optimizado en entrenamiento, validación y prueba.

Conjunto	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Entrenamiento	0.69	-46.06	19.37	56.73
Validación	0.75	-21.50	16.98	53.00
Prueba	0.71	-30.62	19.37	69.43

Tabla 6.18: Comparación en conjunto de validación: PINN-E MLP base vs. optimizado.

Modelo	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Base	0.71	2.21	15.21	57.86
Optimizado	0.75	-21.50	16.98	53.00

En el conjunto de entrenamiento, las métricas resultan inferiores a las de validación ($NSE = 0.69$ vs. 0.75), lo cual se explica por la naturaleza compuesta de la función de pérdida en las PINNs y el efecto regularizador del elevado *dropout* (0.49).

La comparación temporal (Figura 6.16) muestra que el modelo reproduce adecuadamente la dinámica general de la serie, capturando la mayoría de los eventos de crecida intermedia. Sin embargo, se observa una subestimación parcial en eventos extremos y una sobreestimación de los picos medianos y bajos.

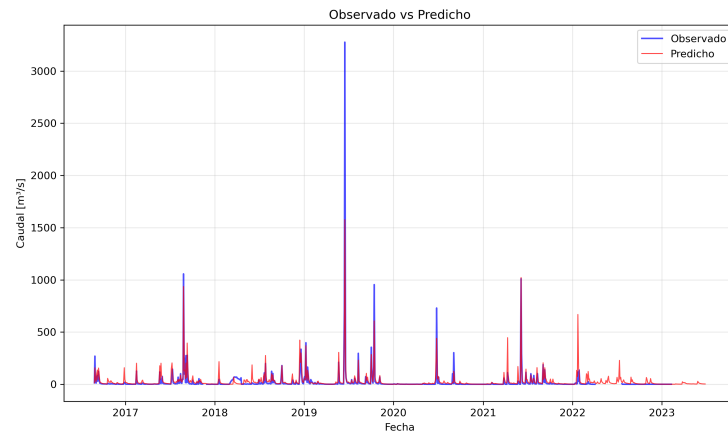


Figura 6.16: Serie temporal de caudales observados y predichos por el modelo PINN-E (MLP) optimizado.

El *scatterplot* en escala logarítmica (Figura 6.17) muestra que el modelo captura la tendencia general de los datos; sin embargo, la alineación con la línea

1:1 no es uniforme en todo el rango de caudales. Se observa una sobreestimación sistemática en valores bajos y medios, mientras que en caudales altos aparece una tendencia a la subestimación. Asimismo, la dispersión aumenta hacia los extremos, lo que sugiere la presencia de sesgos dependientes del régimen de caudal.

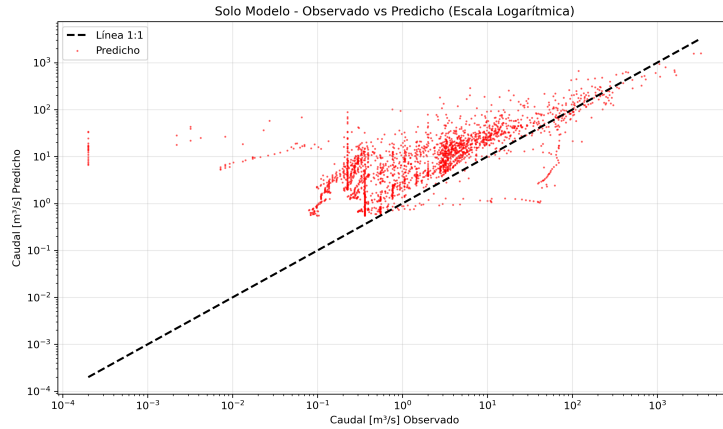


Figura 6.17: *Scatterplot* (escala logarítmica) para el modelo PINN-E (MLP) optimizado.

El análisis de atribuciones en prueba (Figura 6.18) indica que las variables térmicas y radiativas concentran la mayor atribución promedio, destacándose la temperatura máxima y la radiación solar. Este patrón es consistente con la formulación física del proceso de evapotranspiración.

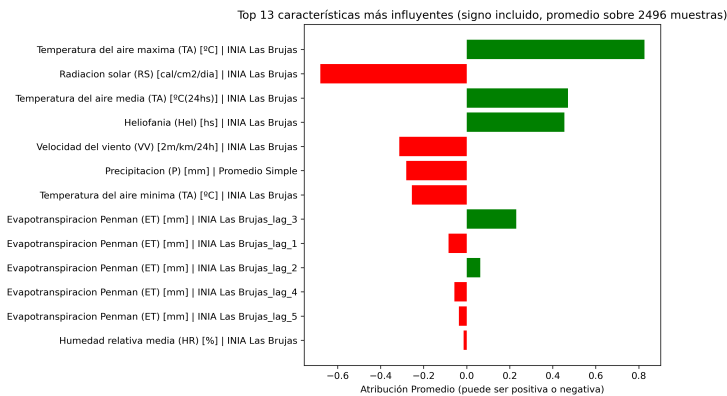


Figura 6.18: Top 13 características más influyentes según *Integrated Gradients* en el conjunto de prueba (PINN-E MLP).

Costo computacional El entrenamiento completo del modelo optimizado demandó 454.49 segundos en un total de 39 épocas. El tiempo de inferencia fue de 3.92 segundos.

6.9. PINN-E RNN

En esta sección se presentan los resultados del modelo PINN-E con arquitectura recurrente, donde la red neuronal predice exclusivamente la evapotranspiración agregada diaria \hat{E}_{agg} , mientras que la precipitación se mantiene fija a partir de los valores observados. Se reporta el desempeño del modelo base, se resume el proceso de optimización aplicado y se presenta la evaluación final del modelo optimizado sobre el conjunto independiente de prueba. El detalle completo de los experimentos iterativos se incluye en el Anexo B.6.

6.9.1. Desempeño del modelo base

El primer experimento consistió en entrenar el modelo PINN-E con arquitectura LSTM base, utilizando todas las variables iniciales sin selección ni optimización de hiperparámetros. La configuración base empleó una secuencia de longitud 1, dos capas LSTM con 64 unidades ocultas y un módulo MLP posterior de tres capas (128→64→32) con activación ReLU.

Tabla 6.19: Desempeño del modelo base PINN-E (RNN) sobre el conjunto de validación.

NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
0.73	-7.85	15.58	55.85

El modelo base (Tabla 6.19) presenta un desempeño razonable ($NSE = 0.73$), superior al observado en la configuración base de la PINN-E MLP ($NSE = 0.71$), con un sesgo moderado ($PBIAS = -7.85\%$) dentro del rango aceptable.

6.9.2. Síntesis del proceso de optimización

Siguiendo el procedimiento metodológico definido previamente, se realizaron experimentos controlados para mejorar el desempeño del modelo. Las decisiones determinantes fueron:

- **Variables temporales:** se conservó únicamente la codificación cíclica del mes (`month_sin`, `month_cos`), cuya inclusión produjo una leve mejora en $RMSE$, coherente con la estacionalidad marcada de la evapotranspiración. A diferencia de la PINN-E MLP (donde las variables temporales fueron eliminadas), en la RNN su remoción deterioraba significativamente el desempeño (Anexo B.6).

- **Representación de la precipitación:** se adoptó la **agregación mediante polígonos de Thiessen** como forzante pluviométrico, a diferencia de la PINN-E MLP donde el promedio simple resultaba óptimo. La capacidad de la capa recurrente para procesar secuencias temporales le permite aprovechar mejor esta representación espacial más sofisticada (Anexo B.6).
- **Ventana temporal:** se adoptó una **ventana de 8 días**, configuración que alcanzó el mejor desempeño global ($NSE = 0.77$, $PBIAS = 0.57\%$). (Anexo B.6).
- **Selección de variables (correlación e IG):** el análisis de correlación no justificó la eliminación de variables, y el análisis de *Integrated Gradients* confirmó que la velocidad del viento y las variables radiativas concentran la mayor atribución. Se mantuvo el conjunto completo de 10 variables (Anexo B.6).
- **Balance de la función de pérdida:** la configuración balanceada $w_{\text{data}} = 1$, $w_{\text{phys}} = 1$ dominó en todas las métricas, consistente con lo observado en la PINN-E MLP (Anexo B.6).
- **Optimización de hiperparámetros (Optuna):** la búsqueda bayesiana (141 *trials* completados, 70 podados) identificó una arquitectura basada en **celda GRU** (2 capas, 47 unidades ocultas) seguida de un MLP profundo de 5 capas, con activación *gelu* y *batch size* de 256 (Anexo B.6).

6.9.3. Resultados del modelo optimizado

Configuración final del modelo

El modelo final fue entrenado con el conjunto de variables seleccionado tras el proceso de optimización. La configuración adoptada incluye:

- **Precipitación:** agregación mediante polígonos de Thiessen.
- **Variables meteorológicas:** radiación solar, heliofanía, temperatura media/máxima/mínima, humedad relativa, velocidad del viento y evapotranspiración.
- **Variables temporales:** codificación cíclica del mes (*month_sin*, *month_cos*).
- **Ventana temporal:** 8 días.

La arquitectura óptima, identificada mediante Optuna, corresponde a una red GRU de 2 capas con 47 unidades ocultas (*dropout* recurrente de 0.16), seguida de un MLP de 5 capas [89, 150, 68, 38, 30] con activación *gelu*, tasa de aprendizaje 1.37×10^{-3} , *dropout* base de 0.20, *weight decay* de 1.89×10^{-6} y *batch size* de 256. El balance de la función de pérdida se mantuvo en $w_{\text{data}} = 1$, $w_{\text{phys}} = 1$.

Evaluación del modelo final optimizado

La Tabla 6.20 resume el desempeño del modelo PINN-E (RNN) optimizado. En el conjunto de prueba se obtiene $NSE = 0.77$, considerado *bueno* según Moriasi y cols. (2015), y un sesgo $PBIAS = 4.42\%$ dentro del rango *muy bueno*.

Tabla 6.20: Desempeño del modelo PINN-E (RNN) optimizado en entrenamiento, validación y prueba.

Conjunto	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Entrenamiento	0.81	-0.58	12.65	43.88
Validación	0.78	-2.47	14.38	49.93
Prueba	0.77	4.42	14.23	62.36

Para enfatizar la mejora lograda respecto al modelo base, la Tabla 6.21 compara ambas configuraciones en el conjunto de validación.

Tabla 6.21: Comparación en conjunto de validación: PINN-E RNN base vs. optimizado.

Modelo	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Base	0.73	-7.85	15.58	55.85
Optimizado	0.78	-2.47	14.38	49.93

El modelo optimizado mejora sustancialmente la eficiencia respecto al base (NSE de 0.73 a 0.78), con un sesgo volumétrico controlado ($PBIAS = -2.47\%$). La degradación gradual del NSE desde entrenamiento (0.81) hacia validación (0.78) y prueba (0.77) refleja una generalización temporal robusta.

En comparación con la versión MLP del modelo PINN-E optimizado ($NSE = 0.71$, $PBIAS = -30.62\%$ en prueba), la arquitectura recurrente presenta una mejora sustancial tanto en eficiencia (+6 puntos porcentuales de NSE) como en control del sesgo volumétrico, indicando que capitaliza de forma más efectiva las ventajas de la formulación PINN-E.

Un aspecto particularmente relevante es que el modelo PINN-E (RNN) **supera claramente al modelo GR4J directo** en el conjunto de prueba ($NSE = 0.77$ vs 0.69, mejora del 11.8 %), lo que indica que la estimación de evapotranspiración mediante la red neuronal, acoplada con GR4J, aporta una mejora neta significativa respecto al modelo conceptual alimentado directamente con forzantes observados.

La comparación temporal (Figura 6.19) muestra que el modelo reproduce adecuadamente la dinámica general de la serie, capturando tanto los períodos de estiaje como la mayoría de los eventos de crecida intermedia. Sin embargo, durante el evento extremo de mediados de 2019 ($\sim 3\,300\text{ m}^3/\text{s}$) se observa una subestimación parcial, lo cual explica el incremento del RMSE en prueba.

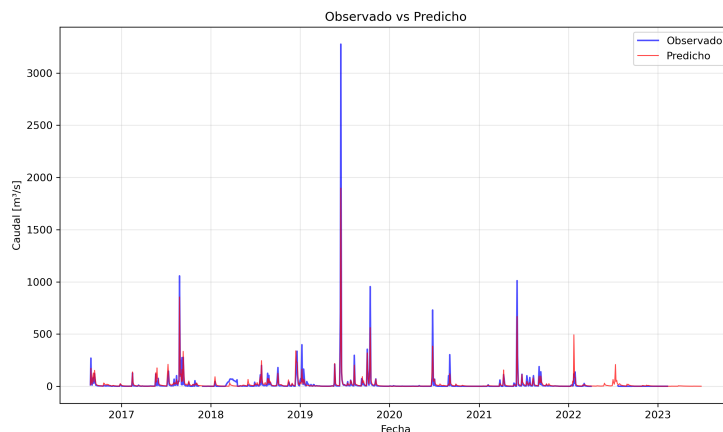


Figura 6.19: Serie temporal de caudales observados y predichos por el modelo PINN-E (RNN) optimizado.

El *scatterplot* en escala logarítmica (Figura 6.20) muestra buena alineación con la línea 1:1 en el rango medio-alto de caudales ($\sim 10\text{--}500\text{ m}^3/\text{s}$), con una tendencia leve a la sobrepredicción en caudales bajos y un ajuste razonable en eventos extremos, superior al observado en la formulación PINN-P.

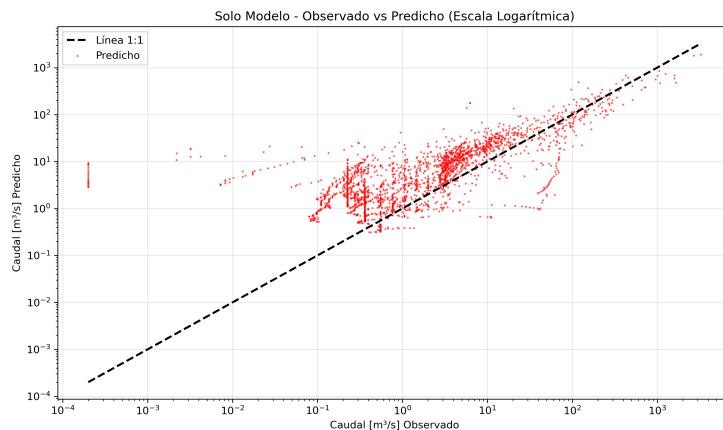


Figura 6.20: *Scatterplot* (escala logarítmica) para el modelo PINN-E (RNN) optimizado.

El análisis de atribuciones en prueba (Figura 6.21) muestra que la **radiación solar** concentra la mayor atribución positiva promedio, mientras que la **velocidad del viento** presenta la mayor atribución negativa. Este patrón es coherente con la dependencia física de la evapotranspiración respecto a la energía disponible, y es consistente con lo observado en la versión MLP de este modelo.

Sin embargo, es importante señalar una limitación metodológica inherente a la aplicación de *Integrated Gradients* a modelos recurrentes con ventana temporal: el cálculo se realizó evaluando cada muestra de forma independiente, sin considerar las $W - 1$ dimensiones adicionales que componen la secuencia de entrada (ventana de 8 días). Esta reducción implica que el contexto secuencial completo que utiliza realmente el modelo no queda reflejado en las atribuciones calculadas, lo que puede distorsionar la distribución de importancia entre variables. Por tanto, estos resultados constituyen una aproximación orientativa de la relevancia relativa de las variables y deben interpretarse con cautela.

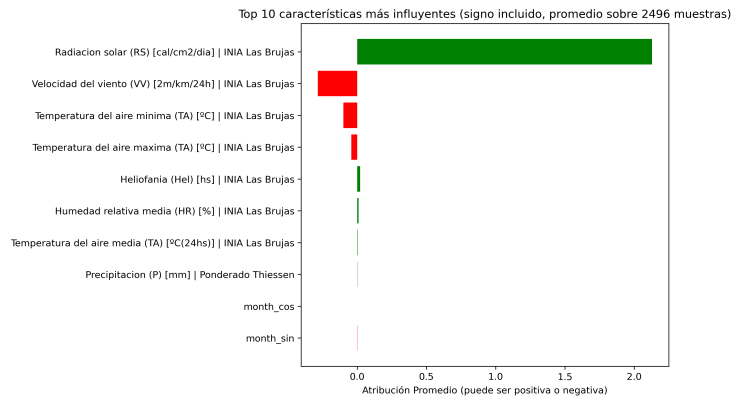


Figura 6.21: Top 10 características más influyentes según *Integrated Gradients* en el conjunto de prueba (PINN-E RNN).

Costo computacional El entrenamiento completo del modelo optimizado demandó 683.53 segundos en un total de 73 épocas. El tiempo de inferencia fue de 3.73 segundos.

6.10. PINN-PE MLP

En esta sección se presentan los resultados del modelo PINN-PE con arquitectura MLP, donde la red neuronal predice simultáneamente la precipitación agregada diaria \hat{P}_{agg} y la evapotranspiración agregada diaria \hat{E}_{agg} . Se reporta el desempeño del modelo base, se resume el proceso de optimización aplicado y se presenta la evaluación final del modelo optimizado sobre el conjunto independiente de prueba. El detalle completo de los experimentos iterativos se incluye en el Anexo B.7.

6.10.1. Desempeño del modelo base

El primer experimento consistió en entrenar el modelo PINN-PE (MLP) con la arquitectura base, utilizando todas las variables iniciales sin selección

ni optimización de hiperparámetros. La predicción final de caudal emerge del acoplamiento físico con GR4J, evaluando $Q_{\text{GR4J}}(\hat{P}_{\text{agg}}, \hat{E}_{\text{agg}}; \theta)$.

Tabla 6.22: Desempeño del modelo base PINN-PE (MLP) sobre el conjunto de validación.

NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
0.67	-0.60	16.49	61.30

El modelo base (Tabla 6.22) presenta un desempeño razonable ($NSE = 0.67$) con un sesgo prácticamente neutro ($PBIAS = -0.60\%$).

6.10.2. Síntesis del proceso de optimización

Siguiendo el procedimiento metodológico definido previamente, se realizaron experimentos controlados para mejorar el desempeño del modelo. Las decisiones determinantes fueron:

- **Variables temporales:** la incorporación de variables temporales no mejoró la eficiencia y deterioró el control del sesgo volumétrico ($PBIAS$ de -0.60% a $\sim 7-8\%$). Se decidió no incluir variables temporales (Anexo B.7).
- **Lags de precipitación y evapotranspiración:** la incorporación de *lags* produjo incrementos sustanciales del sesgo ($PBIAS > 7\%$) sin mejoras significativas en eficiencia. Se decidió no incluir memoria temporal explícita (Anexo B.7).
- **Selección de variables (correlación e IG):** ninguna configuración reducida superó al modelo base. Las eliminaciones que mejoraban levemente el NSE incrementaban el sesgo a valores superiores al 8% . Se mantuvo el conjunto completo de variables (Anexo B.7).
- **Balance de la función de pérdida:** el desempeño resultó altamente estable ($NSE \approx 0.67$ en todas las configuraciones). Se mantuvo la configuración balanceada $w_{\text{data}} = 1$, $w_{\text{phys}} = 1$ (Anexo B.7).
- **Optimización de hiperparámetros (Optuna):** la búsqueda bayesiana (142 *trials* completados, 29 podados) identificó una arquitectura de 4 capas ocultas [298, 970, 33, 87] con activación `gelu`, *dropout* de 0.10 y *batch size* de 16 (Anexo B.7).

6.10.3. Resultados del modelo optimizado

Configuración final del modelo

El modelo final fue entrenado con el conjunto de variables seleccionado tras el proceso de optimización. La configuración adoptada incluye:

- **Precipitación:** precipitaciones contemporáneas por estación (sin *lags*).
- **Variables meteorológicas:** radiación solar, heliofanía, temperatura media/máxima/mínima, humedad relativa y velocidad del viento.
- **Sin variables temporales ni *lags*.**

La arquitectura óptima, identificada mediante Optuna, corresponde a una MLP de 4 capas ocultas [298, 970, 33, 87] con activación `gelu`, tasa de aprendizaje 1.90×10^{-4} , *dropout* base de 0.10, *weight decay* de 2.63×10^{-6} y *batch size* de 16. El balance de la función de pérdida se mantuvo en $w_{\text{data}} = 1$, $w_{\text{phys}} = 1$.

Evaluación del modelo final optimizado

La Tabla 6.23 resume el desempeño del modelo PINN-PE (MLP) optimizado, mientras que la Tabla 6.24 permite comparar el desempeño frente al modelo base inicial. En el conjunto de prueba se obtiene $NSE = 0.72$, considerado *bueno* según Moriasi y cols. (2015) y un $PBIAS = -22.95\%$ considerado *insatisfactorio*.

Tabla 6.23: Desempeño del modelo PINN-PE (MLP) optimizado en entrenamiento, validación y prueba.

Conjunto	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Entrenamiento	0.61	-62.82	21.58	63.70
Validación	0.74	-14.42	16.44	54.08
Prueba	0.72	-22.95	17.29	68.70

Tabla 6.24: Comparación en conjunto de validación: PINN-PE MLP base vs. optimizado.

Modelo	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Base	0.67	-0.60	16.49	61.30
Optimizado	0.74	-14.42	16.44	54.08

El modelo optimizado mejora la eficiencia respecto al base en el conjunto de validación (NSE de 0.67 a 0.74). No obstante, presenta un sesgo volumétrico negativo significativo ($PBIAS = -22.95\%$ en prueba), que indica una tendencia sistemática a la sobreestimación del volumen total de caudal. Este sesgo contrasta con el comportamiento del modelo base ($PBIAS = -0.60\%$), sugiriendo que la optimización arquitectónica introduce un compromiso entre eficiencia estadística y conservación del volumen, patrón común en la mayoría de formulaciones PINN evaluadas.

En el conjunto de entrenamiento, las métricas resultan inferiores a las de validación ($NSE = 0.61$ vs 0.74), explicado por la naturaleza compuesta de la función de pérdida en las PINNs.

La comparación temporal (Figura 6.22) muestra que el modelo reproduce adecuadamente la dinámica general de la serie, capturando eventos de crecida intermedia y la estacionalidad, aunque con subestimación en eventos extremos y sobreestimación en picos intermedios y bajos.

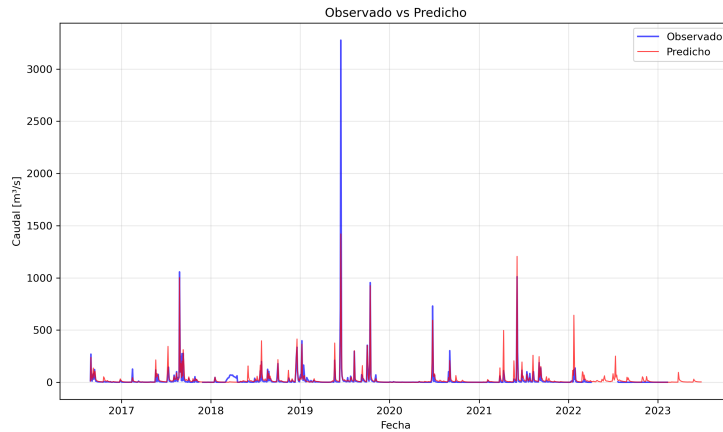


Figura 6.22: Serie temporal de caudales observados y predichos por el modelo PINN-PE (MLP) optimizado.

El *scatterplot* en escala logarítmica (Figura 6.23) muestra buena alineación con la línea 1:1 en el rango medio de caudales, aunque levemente por encima de la línea y subestimación para valores elevados.

El análisis de atribuciones en prueba (Figura 6.24) muestra que las precipitaciones por estación concentran las mayores contribuciones positivas, mientras que las variables térmicas y radiativas presentan atribuciones negativas de menor magnitud.

Costo computacional El entrenamiento completo del modelo optimizado demandó 215.09 segundos en un total de 22 épocas. El tiempo de inferencia fue de 3.69 segundos.

6.11. PINN-PE RNN

En esta sección se presentan los resultados del modelo PINN-PE con arquitectura recurrente, donde la red neuronal predice simultáneamente la precipitación agregada diaria \hat{P}_{agg} y la evapotranspiración \hat{E} . Se reporta el desempeño del modelo base, se resume el proceso de optimización aplicado y se presenta la evaluación final del modelo optimizado sobre el conjunto independiente

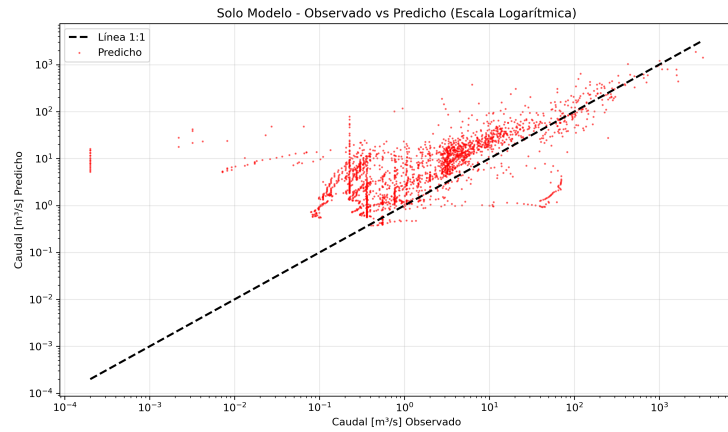


Figura 6.23: *Scatterplot* (escala logarítmica) para el modelo PINN-PE (MLP) optimizado.

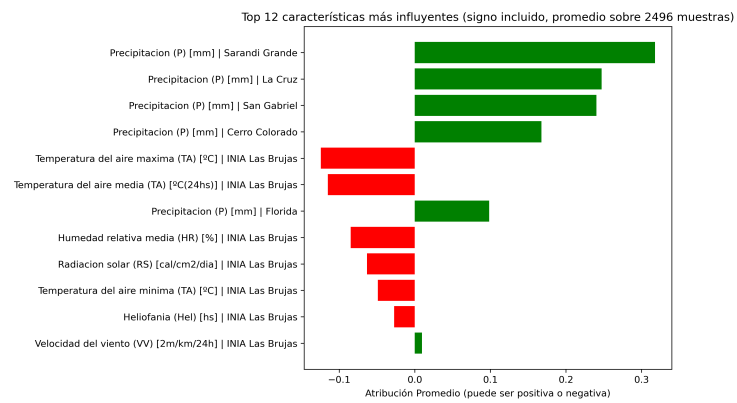


Figura 6.24: Top 12 características más influyentes según *Integrated Gradients* en el conjunto de prueba (PINN-PE MLP).

de prueba. El detalle completo de los experimentos iterativos se incluye en el Anexo B.8.

6.11.1. Desempeño del modelo base

El primer experimento consistió en entrenar el modelo PINN-PE con arquitectura RNN base, utilizando todas las variables iniciales sin selección ni optimización de hiperparámetros. La configuración base empleó una secuencia de longitud 1, dos capas LSTM con 64 unidades ocultas y un módulo MLP posterior de tres capas (128→64→32) con activación ReLU.

Tabla 6.25: Desempeño del modelo base PINN-PE sobre el conjunto de validación.

NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
0.69	-0.98	16.34	59.85

El modelo base (Tabla 6.25) presenta un $NSE = 0.69$ con un sesgo prácticamente nulo ($PBIAS = -0.98\%$). Este desempeño es inferior al de las configuraciones base de PINN-P RNN ($NSE = 0.71$) y PINN-E RNN ($NSE = 0.73$), lo cual es consistente con la mayor complejidad de la tarea dual.

6.11.2. Síntesis del proceso de optimización

Siguiendo el procedimiento metodológico definido previamente, se realizaron experimentos controlados para mejorar el desempeño del modelo. Las decisiones determinantes fueron:

- **Variables temporales:** la incorporación de variables temporales **degradó el desempeño**, especialmente la combinación día+mes ($NSE = 0.58$, $PBIAS = -51.47\%$). La configuración sin variables temporales alcanzó el mejor resultado ($NSE = 0.69$, $PBIAS = -0.98\%$). Se decidió no incluir variables temporales (Anexo B.8).
- **Ventana temporal:** se adoptó una **ventana de 5 días**, que ofrece el mejor equilibrio entre eficiencia ($NSE = 0.68$) y control del sesgo ($PBIAS = -1.14\%$) (Anexo B.8).
- **Selección de variables (correlación e IG):** el modelo base con 12 variables dominó en todas las métricas. Todas las exclusiones produjeron deterioros de 2-6 puntos de NSE y aumentos significativos del sesgo (Anexo B.8).
- **Balace de la función de pérdida:** se adoptó $w_{\text{data}} = 1$, $w_{\text{phys}} = 0.50$, que alcanza $NSE = 0.71$ con sesgo controlado ($PBIAS = -6.89\%$) (Anexo B.8).

- **Optimización de hiperparámetros (Optuna):** la búsqueda bayesiana (125 *trials* completados, 27 podados) identificó una celda LSTM con 1 capa y 171 unidades ocultas, seguida de un MLP de 3 capas [27, 31, 44], activación `gelu` y *batch size* de 32 (Anexo B.8).

6.11.3. Resultados del modelo optimizado

Configuración final del modelo

El modelo final fue entrenado con el conjunto de variables seleccionado tras el proceso de optimización. La configuración adoptada incluye:

- **Precipitación:** precipitaciones contemporáneas por estación.
- **Variables meteorológicas:** radiación solar, heliofanía, evapotranspiración, temperatura media/máxima/mínima, humedad relativa y velocidad del viento.
- **Sin variables temporales.**
- **Ventana temporal:** 5 días.

La arquitectura óptima, identificada mediante Optuna, corresponde a una red LSTM de 1 capa con 171 unidades ocultas, seguida de un MLP de 3 capas [27, 31, 44] con activación `gelu`, tasa de aprendizaje 1.12×10^{-3} , *dropout* base de 0.15, *weight decay* de 4.51×10^{-4} y *batch size* de 32. El balance de la función de pérdida se fijó en $w_{\text{data}} = 1$, $w_{\text{phys}} = 0.50$.

Evaluación del modelo final optimizado

La Tabla 6.26 resume el desempeño del modelo PINN-PE (RNN) optimizado. En el conjunto de prueba se obtiene $NSE = 0.71$, considerado *bueno* según Moriasi y cols. (2015), y un sesgo $PBIAS = -4.19\%$ dentro del rango *muy bueno*.

Tabla 6.26: Desempeño del modelo PINN-PE (RNN) optimizado en entrenamiento, validación y prueba.

Conjunto	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Entrenamiento	0.87	-33.76	13.07	37.27
Validación	0.69	1.08	15.90	59.28
Prueba	0.71	-4.19	15.29	69.51

Para comparar el desempeño de la optimización respecto al modelo base, la Tabla 6.27 muestra las métricas de resultado de ambas configuraciones en el conjunto de validación.

El modelo optimizado mantiene la misma eficiencia que el modelo base en el conjunto de validación ($NSE = 0.69$ en ambos casos), mostrando además una

Tabla 6.27: Comparación en conjunto de validación: PINN-PE RNN base vs. optimizado.

Modelo	NSE	PBIAS [%]	MAE [m^3/s]	RMSE [m^3/s]
Base	0.69	-0.98	16.34	59.85
Optimizado	0.69	1.08	15.90	59.28

ligera mejora en las métricas de error, con una reducción del MAE de 16.34 a 15.90 y del $RMSE$ de 59.85 a 59.28. En términos de sesgo volumétrico, el modelo optimizado presenta un $PBIAS = 1.08\%$, frente a -0.98% del modelo base, manteniéndose en ambos casos dentro de un rango de sesgo bajo y controlado.

En comparación con las formulaciones individuales en prueba, la PINN-PE RNN ($NSE = 0.71$) supera a la PINN-P RNN ($NSE = 0.67$) pero queda por debajo de la PINN-E RNN ($NSE = 0.77$).

La comparación temporal (Figura 6.25) muestra que el modelo reproduce adecuadamente la dinámica general, capturando períodos de estiaje y eventos de crecida intermedios. El evento extremo de mediados de 2019 ($\sim 3\,300\text{ m}^3/\text{s}$) es parcialmente reproducido con un pico de $\sim 1\,400\text{ m}^3/\text{s}$.

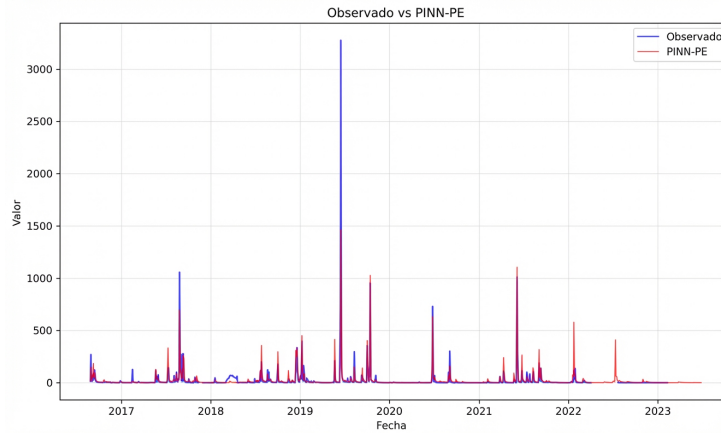


Figura 6.25: Serie temporal de caudales observados y predichos por el modelo PINN-PE (RNN) optimizado.

El *scatterplot* en escala logarítmica (Figura 6.26) muestra alineación razonable con la línea 1:1 en el rango medio-alto, con sobrepredicción en caudales bajos y ajuste aceptable en eventos extremos.

El análisis de atribuciones en prueba (Figura 6.27) muestra que la **humedad relativa media** concentra la mayor atribución positiva, mientras que la **velocidad del viento** presenta la mayor atribución negativa, de magnitud considerable. Las variables de precipitación de múltiples estaciones (La Cruz, Sarandí Grande, Cerro Colorado, San Gabriel, Florida) aparecen con atribuciones

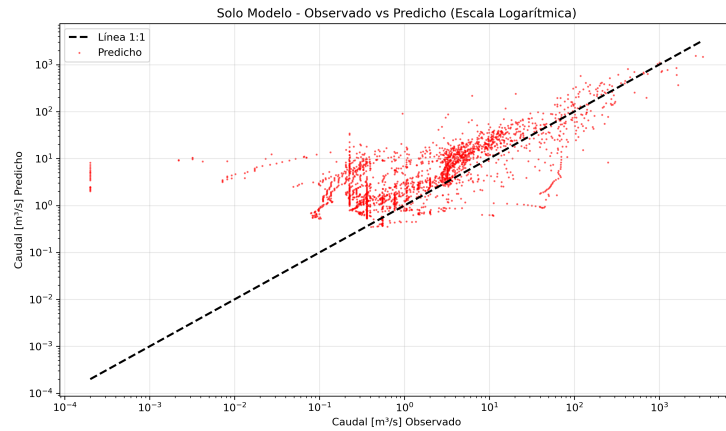


Figura 6.26: *Scatterplot* (escala logarítmica) para el modelo PINN-PE (RNN) optimizado.

positivas moderadas, lo que refleja la dependencia directa del caudal respecto al forzante pluviométrico. Las variables radiativas (heliofanía, radiación solar) y la temperatura máxima presentan atribuciones negativas de magnitud intermedia. Este patrón difiere del observado en las versiones de predicción individual (PINN-E, PINN-P), donde la radiación solar o la precipitación dominaban respectivamente, sugiriendo que la tarea dual redistribuye la relevancia relativa de los predictores.

Es importante destacar que, al igual que en las demás arquitecturas recurrentes, el cálculo de *Integrated Gradients* se realizó sin considerar las $W - 1$ dimensiones adicionales de la ventana temporal (5 días). El contexto secuencial completo que integra el modelo no queda reflejado en las atribuciones calculadas, por lo que estos resultados deben interpretarse como una aproximación orientativa de la relevancia relativa de las variables.

Costo computacional El entrenamiento completo del modelo optimizado demandó 603.05 segundos en un total de 63 épocas. El tiempo de inferencia fue de 4.09 segundos.

6.12. Comparación

En esta sección se comparan los modelos desarrollados utilizando el 20% final de los datos como conjunto de prueba, el cual no fue empleado durante el entrenamiento ni en la calibración. El modelo conceptual GR4J se adopta como línea base para evaluar las mejoras introducidas por los modelos de aprendizaje automático y sus variantes con restricciones físicas (PINN).

La evaluación cuantitativa se realiza mediante las métricas NSE, PBIAS,

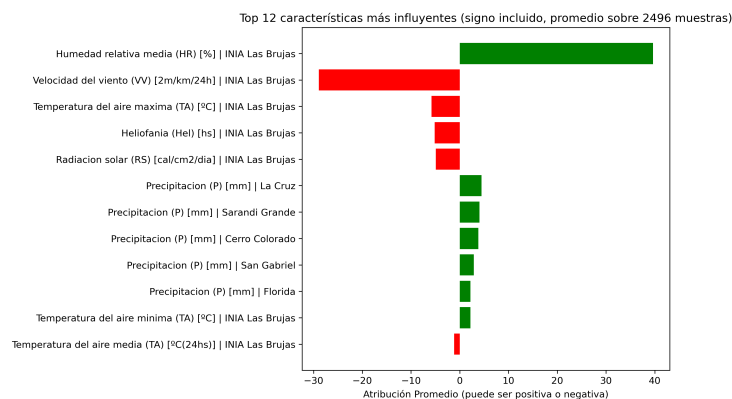


Figura 6.27: Top 12 características más influyentes según *Integrated Gradients* en el conjunto de prueba (PINN-PE RNN).

MAE y RMSE. Para la interpretación del NSE y del PBIAS se utiliza la clasificación propuesta por [Moriassi y cols. \(2015\)](#), presentada en la Tabla 5.1.

En la Tabla 6.28 se muestran los resultados obtenidos, donde los valores de NSE y PBIAS se presentan con codificación de color según su categoría de desempeño.

Tabla 6.28: Resultados finales sobre el conjunto de prueba (20 % final de los datos).

Modelo	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]	Tiempo de inferencia [s]
GR4J	0.69	-16.08	17.59	72.41	0.09
MLP	0.67	6.52	15.96	73.57	0.00
RNN	0.59	-15.99	18.99	82.81	0.01
PINN P MLP	0.74	-35.55	18.17	65.98	5.22
PINN P RNN	0.67	-22.64	17.85	74.33	4.07
PINN E MLP	0.71	-30.62	19.37	69.43	3.92
PINN E RNN	0.77	4.42	14.23	62.36	3.73
PINN PE MLP	0.72	-22.95	17.29	68.70	3.69
PINN PE RNN	0.71	-4.19	15.29	69.51	4.09

Codificación de colores (según [Moriassi y cols. \(2015\)](#)): **Muy bueno** (NSE > 0.75; |PBIAS| < 10%), **Bueno** (0.65 < NSE ≤ 0.75; 10 < |PBIAS| ≤ 15%), **Satisfactorio** (0.50 < NSE ≤ 0.65; 15 < |PBIAS| ≤ 25%), **Insatisfactorio** (NSE ≤ 0.50; |PBIAS| > 25%).

En términos de eficiencia predictiva global (NSE), el modelo GR4J presenta un desempeño *bueno* (0.69), constituyendo una línea base sólida. Entre los modelos de aprendizaje automático tradicionales, el MLP mantiene una clasificación *bueno* (0.67), mientras que la RNN desciende a la categoría *satisfactorio* (0.59), evidenciando mayores dificultades para generalizar en el conjunto de prueba final.

La incorporación de restricciones físicas mediante PINNs produce mejoras notables en determinados casos. El modelo PINN E RNN alcanza el mayor NSE (0.77), clasificándose como *muy bueno* y superando tanto al *baseline* como al resto de los modelos evaluados. Asimismo, varias configuraciones PINN presentan valores de NSE superiores a 0.70, consolidándose dentro de la categoría *buena*. Estos resultados sugieren que la inclusión explícita de información física contribuye a mejorar la representación dinámica del sistema hidrológico y su capacidad de generalización fuera del conjunto de entrenamiento.

En cuanto al sesgo volumétrico (PBIAS), el comportamiento es más heterogéneo. El GR4J presenta un desempeño *satisfactorio* (-16.08%), mientras que el MLP y el PINN E RNN alcanzan categoría *muy buena*, con valores absolutos menores al 10%. Sin embargo, algunas variantes PINN muestran sesgos superiores al 25%, clasificándose como *insatisfactorias*. Esto indica que una mejora en eficiencia global (NSE) no necesariamente implica una adecuada reproducción del balance volumétrico total, evidenciando la importancia de evaluar simultáneamente métricas de eficiencia y de sesgo.

Las métricas MAE y RMSE refuerzan estas conclusiones. El menor RMSE corresponde al PINN E RNN (62.36), lo que sugiere una mejor representación de la variabilidad temporal y de los eventos de mayor magnitud respecto al GR4J. En general, las configuraciones que combinan restricciones físicas con arquitectura recurrente tienden a reducir el error cuadrático medio, lo cual resulta consistente con una mejor captura de las dependencias temporales del proceso hidrológico.

En lo que respecta al tiempo de inferencia, las diferencias entre modelos resultan poco significativas en términos prácticos, dado que todos los valores se encuentran en el orden de los segundos o fracciones de segundo. En consecuencia, para aplicaciones operativas convencionales, el costo de evaluación individual no constituye un factor limitante. No obstante, se observa que las configuraciones PINN presentan tiempos de inferencia mayores que los modelos puramente basados en datos. Este comportamiento es consistente con lo discutido en capítulos anteriores, donde se señaló que la inclusión del modelo hidrológico GR4J dentro de la arquitectura implica una evaluación secuencial asociada a su naturaleza dependiente del estado. Por diseño, se espera entonces un mayor costo computacional relativo respecto a arquitecturas *feedforward* o recurrentes sin acoplamiento físico, aun cuando dicho incremento no resulte determinante en el contexto del conjunto de prueba analizado.

En síntesis, desde el punto de vista estrictamente predictivo, las configuraciones PINN, en particular la arquitectura PINN E RNN, evidencian mejoras consistentes respecto al modelo conceptual GR4J y a las redes puramente basadas en datos.

Análisis del costo computacional de entrenamiento

Si bien las diferencias en tiempo de inferencia no resultan determinantes en el contexto evaluado, el proceso de entrenamiento sí presenta contrastes significativos entre las distintas configuraciones. Con el objetivo de analizar este aspecto,

en la Tabla 6.29 se resumen el tiempo total de entrenamiento, la cantidad de épocas recorridas y el tiempo promedio por época para cada modelo.

Tabla 6.29: Costo computacional durante el entrenamiento de los modelos.

Modelo	Tiempo total (s)	Épocas	Tiempo promedio por época (s/época)
GR4J	-	-	-
MLP	45.49	101	0.45
RNN	99.21	69	1.44
PINN P MLP	265.83	23	11.56
PINN P RNN	326.23	28	11.65
PINN E MLP	454.49	39	11.65
PINN E RNN	683.53	73	9.36
PINN PE MLP	215.09	22	9.78
PINN PE RNN	603.05	63	9.57

El tiempo promedio por época se calcula como el cociente entre el tiempo total de entrenamiento y el número de épocas recorridas.

El modelo GR4J no se incluye en esta comparación, dado que no requiere un proceso de entrenamiento iterativo como las redes neuronales, sino únicamente una etapa de calibración paramétrica independiente.

Entre los modelos puramente basados en datos, se observa que el MLP presenta el menor costo computacional total (45.49 s) y el menor tiempo por época (0.45 s/época). La RNN, en cambio, prácticamente duplica el tiempo total (99.21 s) y triplica el costo por época (1.44 s/época). Esta diferencia resulta consistente con la naturaleza secuencial de la arquitectura recurrente, que mantiene estados internos dependientes del tiempo y limita la paralelización completa del proceso de entrenamiento, incrementando el costo computacional por iteración respecto a una red *feedforward*.

Las configuraciones PINN exhiben un incremento sustancial en el tiempo de entrenamiento respecto a las redes puramente basadas en datos. Mientras que MLP y RNN presentan tiempos por época del orden de 10^{-1} y 10^0 segundos, respectivamente, las PINNs muestran valores comprendidos aproximadamente entre 9 y 12 segundos por época. Esto implica un aumento de un orden de magnitud en el costo por iteración, lo cual confirma el impacto computacional asociado a la incorporación del término físico en la función de pérdida.

Este comportamiento era esperable desde el punto de vista metodológico. La inclusión del modelo hidrológico GR4J dentro del proceso de optimización introduce una evaluación secuencial paso a paso, derivada de su naturaleza dinámica y dependiente del estado. Dado que los almacenamientos internos deben actualizarse estrictamente en orden cronológico, la evaluación del término físico no puede paralelizarse completamente en el eje temporal, generando un cuello de botella computacional estructural durante cada época de entrenamiento.

Al analizar el tiempo total de entrenamiento, se observa que la variabilidad

entre configuraciones PINN no responde exclusivamente al costo por época, que resulta relativamente homogéneo dentro de cada familia, sino también al número de épocas necesarias para converger. Por ejemplo, el modelo PINN E RNN alcanza el mayor tiempo total (683.53 s), lo cual se explica tanto por un costo por época elevado (9.36 s/época) como por un número considerable de épocas (73). En contraste, configuraciones como PINN PE MLP presentan tiempos totales menores (215.09 s) principalmente debido a un número reducido de épocas (22), aun cuando el costo por iteración es del mismo orden.

Asimismo, se observa que las diferencias entre variantes con arquitectura MLP y RNN dentro de las PINNs no modifican sustancialmente el orden de magnitud del costo por época, lo que sugiere que el componente dominante del tiempo computacional está asociado a la simulación secuencial del modelo físico más que a la arquitectura neuronal en sí misma.

En conjunto, los resultados confirman que la incorporación de restricciones físicas mediante el acoplamiento con GR4J introduce un incremento significativo en el costo computacional de entrenamiento, aproximadamente un orden de magnitud superior al de los modelos puramente basados en datos. Sin embargo, parte de la variabilidad en el tiempo total se explica por diferencias en la dinámica de convergencia. De este modo, las mejoras observadas en desempeño predictivo se obtienen a costa de un mayor esfuerzo computacional, evidenciando un compromiso claro entre precisión y eficiencia en el proceso de entrenamiento.

Comparación visual

La comparación cuantitativa (Tabla 6.28) se complementa con un análisis visual orientado a identificar patrones de error estructurales que no siempre son evidentes en métricas agregadas. En particular, se discuten: (i) la dinámica temporal en un período representativo, (ii) el balance volumétrico acumulado, (iii) la distribución de caudales (FDC) y (iv) la reproducción de caudales pico. Este conjunto permite interpretar de manera integrada el desempeño relativo de los modelos, explicando por qué algunas configuraciones pueden alcanzar alta eficiencia (NSE) pero a la vez exhibir sesgos volumétricos relevantes (PBIAS) o discrepancias persistentes en caudales bajos.

Series temporales en un período representativo (escala lineal y log).

En la Figura 6.28 se presenta la comparación multi-modelo para el período 2021-07-01 a 2021-12-31 en escala lineal, donde se enfatizan los eventos de mayor magnitud. En términos generales, la mayoría de los modelos reproduce adecuadamente el *timing* de los principales pulsos hidrológicos: los ascensos y descensos ocurren en fechas consistentes con lo observado, lo que sugiere que las entradas meteorológicas permiten capturar razonablemente la ocurrencia de eventos. Sin embargo, se observan diferencias notorias en la *magnitud* de los picos y en la forma de las recesiones posteriores. En particular, algunas configuraciones PINN tienden a sobredimensionar picos específicos, mientras que otras los amortiguan, evidenciando que una correcta sincronización temporal no garantiza una reproducción adecuada de amplitudes.

La Figura 6.29 muestra el mismo período en escala logarítmica, donde las diferencias entre modelos se vuelven considerablemente más evidentes. A diferencia de la representación en escala lineal, dominada visualmente por los picos, la escala log permite analizar con mayor resolución el comportamiento en caudales medios y bajos, así como la forma de las recesiones.

En esta representación se observa una separación más clara entre las curvas, particularmente en el tramo de caudales pequeños. Varias configuraciones mantienen valores sistemáticamente superiores a los observados durante los períodos inter-evento, lo que se traduce en recesiones más lentas y un sostenimiento artificial del caudal de base. Esta discrepancia resulta mucho más notoria en escala log, donde la distancia vertical entre curvas refleja diferencias relativas significativas.

Este comportamiento es coherente con los valores de PBIAS negativos de mayor magnitud reportados en la Tabla 6.28, ya que la sobreestimación persistente en el rango bajo contribuye acumulativamente a un exceso volumétrico a lo largo del período de prueba.

En contraste, el modelo **PINN E RNN** muestra una convergencia progresiva hacia la serie observada durante las fases de recesión, reduciendo la separación sistemática en el tramo medio-bajo. Si bien no elimina completamente las discrepancias, su comportamiento en escala log evidencia una representación más realista del vaciamiento del sistema, lo que sugiere una mejor captura de la dinámica interna de almacenamiento y explica su mejor desempeño conjunto en NSE y PBIAS.

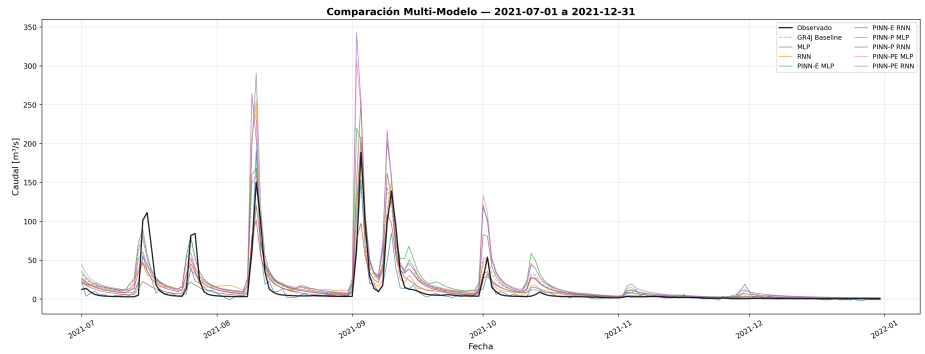


Figura 6.28: Comparación multi-modelo en el período 2021-07-01 a 2021-12-31 (escala lineal).

Un aspecto adicional que resulta evidente en las Figuras 6.28 y 6.29 es la similitud estructural entre las configuraciones PINN y el modelo conceptual GR4J. Esta coherencia no es casual, sino consecuencia directa de la arquitectura adoptada. En las configuraciones PINN, la red neuronal no reemplaza al modelo hidrológico, sino que predice una o más entradas forzantes (precipitación P , evapotranspiración E , o ambas), las cuales se incorporan al esquema dinámico de GR4J manteniendo los mismos parámetros x_1, x_2, x_3, x_4 . Las variables no

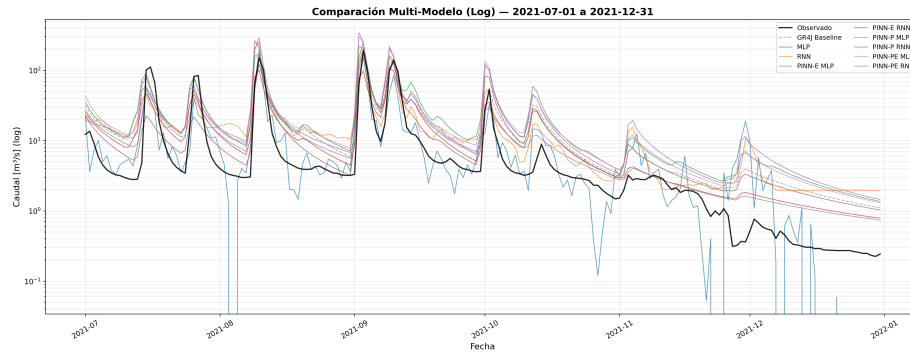


Figura 6.29: Comparación multi-modelo en el período 2021-07-01 a 2021-12-31 (escala logarítmica).

predichas por la red se utilizan con sus valores observados.

Como resultado, la estructura dinámica interna del sistema —almacenamientos, intercambio subterráneo y función de transferencia— permanece gobernada por GR4J. En consecuencia, las curvas simuladas por las PINNs tienden a conservar la forma general y la tendencia temporal del modelo conceptual, mostrando diferencias principalmente en magnitud y en la intensidad relativa de los eventos.

Este comportamiento explica por qué las configuraciones PINN presentan trayectorias visualmente similares entre sí y respecto a GR4J, pero con ajustes cuantitativos diferenciados. La red neuronal actúa esencialmente modulando la señal de entrada al modelo físico, introduciendo correcciones que pueden mejorar la eficiencia global o el balance volumétrico, sin alterar la estructura hidrológica subyacente. Desde esta perspectiva, las mejoras observadas en el **PINN E RNN** no corresponden a un cambio estructural del modelo conceptual, sino a una optimización informada de sus forzantes, lo que refuerza la coherencia física de la simulación.

Balance volumétrico y caudal acumulado. La Figura 6.30 presenta el volumen acumulado para el período completo de prueba. Este tipo de visualización es particularmente útil para interpretar el PBIAS: separaciones persistentes y crecientes entre curvas reflejan sesgos sistemáticos en el balance de masa.

Se observa que varias configuraciones PINN exhiben una deriva positiva respecto de lo observado (curvas acumuladas por encima), coherente con los valores de PBIAS negativos de mayor magnitud reportados en la Tabla 6.28. Esta sobreestimación volumétrica se origina, como se evidenció en la escala logarítmica de las series temporales (Figura 6.29), en la persistencia de caudales de base superiores a los observados durante las fases de recesión.

Entre todas las configuraciones evaluadas, se observa que únicamente el **MLP** y el **PINN E RNN** mantienen trayectorias acumuladas cercanas a la serie observada durante la mayor parte del período de prueba. Este compor-

tamiento es coherente con sus valores de PBIAS de baja magnitud absoluta (6.52 % para MLP y 4.42 % para PINN E RNN), clasificados como *muy buenos*. En ambos casos, la separación entre curvas no muestra una deriva progresiva marcada, lo que indica un adecuado balance volumétrico global.

No obstante, al contrastar este resultado con las series temporales (Figuras 6.28 y 6.29), emergen diferencias relevantes entre ambas configuraciones. El **MLP**, si bien logra un volumen acumulado consistente, presenta una señal más irregular, con mayor cantidad de picos y mínimos locales que no siempre corresponden con la dinámica observada. Esta mayor variabilidad introduce oscilaciones que, aunque se compensan en el balance integrado, reducen la coherencia temporal del modelo.

En cambio, el **PINN E RNN** combina una trayectoria acumulada cercana a la observada con una representación temporal más estable y estructuralmente consistente, especialmente en la forma de las recesiones y en el rango de caudales bajos. Es decir, no solo reproduce adecuadamente el balance global hídrico, sino que lo hace manteniendo una dinámica hidrológica más realista.

Esta diferencia resulta relevante: mientras que el MLP alcanza un buen desempeño volumétrico principalmente por compensación de errores locales, el **PINN E RNN** logra simultáneamente coherencia dinámica y consistencia acumulada, lo que refuerza su mejor desempeño global.

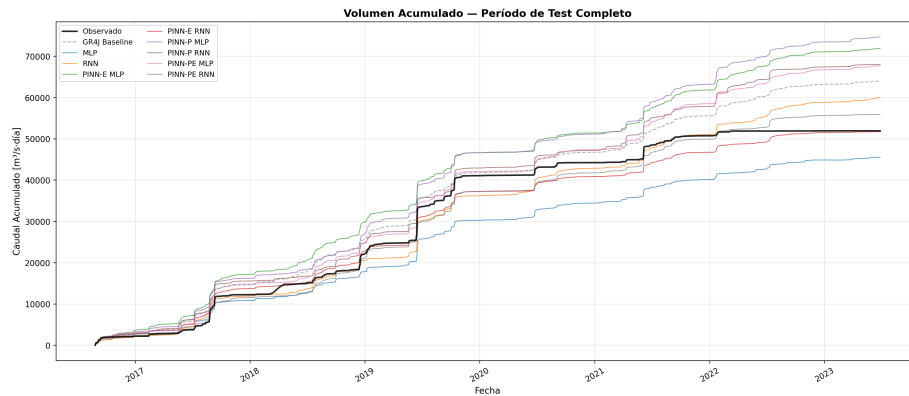


Figura 6.30: Volumen acumulado para el período completo de prueba. Diferencias persistentes entre curvas indican sesgos sistemáticos en el balance volumétrico.

Distribución de caudales: curvas de duración (FDC). Las curvas de duración de caudales (Figura 6.31) permiten evaluar el ajuste de la distribución simulada en todo el rango hidrológico. En escala lineal, las diferencias en el tramo de mayor magnitud (bajas probabilidades de excedencia) aparecen relativamente acotadas, mientras que la escala logarítmica evidencia con mayor claridad discrepancias en caudales medios y bajos. En particular, varios modelos tienden

a sobreestimar caudales para probabilidades de excedencia medias-altas, lo que implica una representación de caudales de base más elevada que la observada. Este patrón es coherente con (i) la deriva observada en el volumen acumulado para ciertos modelos y (ii) los sesgos PBIAS negativos de mayor magnitud. Dentro de este conjunto, el **PINN E RNN** presenta un ajuste comparativamente más cercano al observado en el tramo medio-bajo, aun cuando persisten diferencias en los extremos del rango (muy altos y muy bajos caudales).

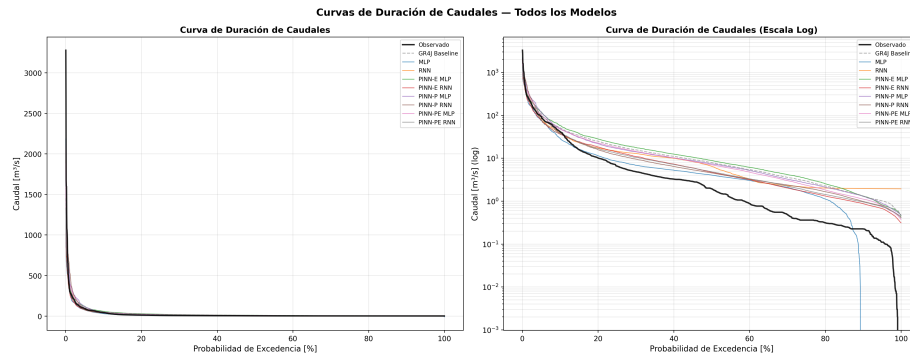


Figura 6.31: Curvas de duración de caudales (FDC) para todos los modelos, en escala lineal (izquierda) y logarítmica (derecha).

Reproducción de caudales pico (eventos extremos). La Figura 6.32 sintetiza la habilidad de los modelos para reproducir los 20 mayores caudales pico del período de prueba, comparando caudal observado versus predicho. En general, se aprecia una dificultad compartida para replicar los eventos más extremos: para los mayores valores observados, las predicciones tienden a ubicarse por debajo de la línea 1:1, indicando subestimación de picos muy altos. Asimismo, en picos de magnitud intermedia aparecen casos de sobreestimación puntual (puntos por encima de 1:1), lo que sugiere que distintos modelos compensan errores entre eventos moderados y extremos.

En comparación con el modelo conceptual **GR4J**, se observa que este último es, en numerosos casos, el que presenta mayor subestimación de los picos más intensos, ubicándose sistemáticamente más alejado de la línea 1:1 en el extremo superior del rango. En este sentido, los modelos generados, tanto las redes puramente basadas en datos como las configuraciones PINN, muestran una mejora relativa en la representación de eventos de alta magnitud, reduciendo parcialmente la brecha respecto a los valores observados. Esta mejora resulta particularmente relevante desde el punto de vista hidrológico, dado que los caudales pico suelen estar asociados a eventos de mayor impacto.

Si bien en el rango de caudales medios y bajos el comportamiento entre modelos tiende a ser más similar, en el dominio de los máximos se aprecia con mayor claridad la contribución de las configuraciones desarrolladas frente al *baseline* conceptual. En este contexto, el **PINN E RNN** muestra una dispersión

más contenida en rangos medios (consistente con sus menores MAE/RMSE), aunque la subestimación de los eventos más extremos permanece como una limitación estructural compartida.

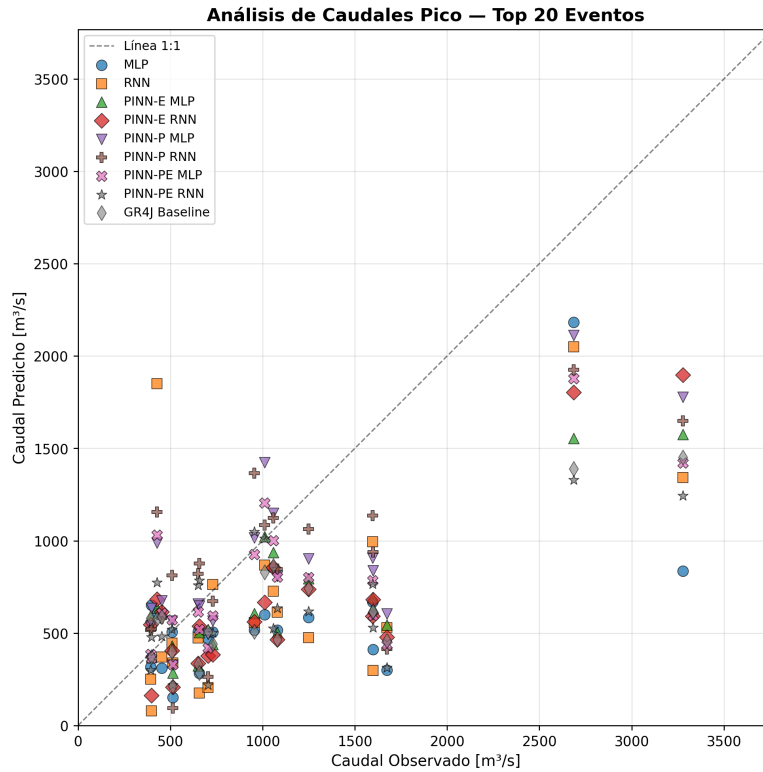


Figura 6.32: Análisis de caudales pico: comparación observado vs predicho para los 20 mayores eventos del período de prueba.

Síntesis del análisis visual. En conjunto, la evidencia gráfica converge con las métricas de la Tabla 6.28 y permite identificar tres hallazgos transversales. Primero, la principal fuente de sesgo volumétrico reside en la sobreestimación de caudales de base: las configuraciones con PBIAS negativo elevado sostienen recepciones más lentas que las observadas (Figuras 6.29 y 6.31), generando una deriva positiva persistente en el volumen acumulado (Figura 6.30). Segundo, la subestimación de los eventos más extremos constituye una limitación compartida, si bien los modelos desarrollados reducen la brecha respecto al *baseline* **GR4J** (Figura 6.32). Tercero, entre todas las configuraciones, el **PINN E RNN** logra el mejor equilibrio entre eficiencia global, control del sesgo volumétrico y coherencia dinámica: mantiene una trayectoria acumulada cercana a la observada sin la irregularidad del MLP, presenta un ajuste más consistente en el

tramo medio–bajo de la FDC y exhibe la menor dispersión en rangos medios de caudales pico.

Conclusión

La comparación final sobre el conjunto de prueba muestra que la incorporación de restricciones físicas mediante configuraciones PINN puede mejorar el desempeño predictivo, pero que dichas mejoras no son uniformes entre variantes y deben evaluarse en conjunto con el sesgo volumétrico y la coherencia dinámica del hidrograma.

El modelo conceptual **GR4J**, con un NSE de 0.69 (categoría *buena*) y un PBIAS de -16.08% (*satisfactorio*), constituye una línea base sólida pero con margen de mejora, particularmente en el balance volumétrico. Integrandolo la evidencia numérica (Tabla 6.28) y el análisis visual, se identifica al **PINN E RNN** como la mejor configuración: alcanza el mayor NSE (0.77, *muy bueno*), el menor MAE (14.23) y el menor RMSE (62.36), junto con un PBIAS de 4.42% (*muy bueno*), superando al *baseline* en todas las métricas evaluadas. Los gráficos confirman que esta configuración combina coherencia dinámica en el tramo medio–bajo (es decir, una representación más realista de las recesiones y los caudales de base en escala logarítmica), estabilidad en el volumen acumulado y menor dispersión en caudales pico respecto al resto de los modelos.

La mejora, sin embargo, no es gratuita: las configuraciones PINN presentan un costo computacional de entrenamiento aproximadamente un orden de magnitud superior al de los modelos puramente basados en datos. Asimismo, si bien la subestimación de los eventos más extremos persiste como una limitación compartida, los modelos desarrollados evidencian una mejora relativa frente al **GR4J** en la reproducción de caudales pico, particularmente en rangos intermedios.

Es importante señalar que estos resultados corresponden a una única subcuenca (Florida – Puente Ruta 5) y a un período temporal específico, por lo que su generalización a otras cuencas o condiciones hidrológicas debe ser evaluada en estudios complementarios.

En consecuencia, el capítulo siguiente corresponde a la *discusión* y se centrará exclusivamente en el modelo **PINN E RNN**. Allí se analizará en detalle su comportamiento frente al modelo conceptual **GR4J**, con énfasis en comprender *qué aspectos del sistema mejora* cuando se incorpora un componente basado en datos dentro de un esquema físico.

En particular, la configuración seleccionada no solo predice la evapotranspiración E , sino que lo hace bajo una formulación híbrida en la que la función de pérdida combina dos contribuciones: (i) el ajuste directo entre la evapotranspiración observada y la estimada, y (ii) el ajuste del caudal simulado por **GR4J** —utilizando la E predicha y la precipitación observada— respecto al caudal observado. De este modo, la red neuronal aprende una estimación de E que no solo reproduce su señal observada, sino que además resulta hidrológicamente consistente al integrarse dentro del modelo conceptual.

La discusión se orientará entonces a evaluar si esta estrategia, que combina aprendizaje basado en datos con restricciones físicas explícitas, permite mejorar el desempeño de un modelo puramente físico sin perder coherencia estructural. En particular, se analizará bajo qué condiciones hidrológicas (p. ej., períodos húmedos y secos, caudales bajos y eventos extremos) la predicción de E aporta ventajas prácticas respecto al GR4J, y en cuáles dichas mejoras podrían no resultar significativas.

Capítulo 7

Discusión

7.1. Desempeño del modelo PINN E RNN frente a GR4J

Los resultados cuantitativos mostraron que la configuración **PINN E RNN** supera al modelo conceptual **GR4J** en términos de NSE, PBIAS, RMSE y MAE. No obstante, para comprender la naturaleza de dichas mejoras es necesario analizar el comportamiento dinámico del modelo en distintos rangos hidrológicos.

Comportamiento en la serie temporal completa. La Figura 7.1 presenta la comparación entre el caudal observado, GR4J y PINN E RNN para todo el período de prueba.

En esta representación se observa que ambos modelos reproducen adecuadamente el *timing* de los eventos principales, lo que confirma que la estructura dinámica del sistema continúa gobernada por el esquema conceptual de GR4J. La forma general del hidrograma es similar en ambas simulaciones, lo cual era esperable dado que la arquitectura PINN no modifica los almacenamientos ni la función de transferencia del modelo físico.

Sin embargo, en los eventos de mayor magnitud del período completo se aprecia que la **PINN E RNN** reduce parcialmente la subestimación presente en GR4J, acercándose más a los valores máximos observados. Esta mejora en los picos extremos contribuye de manera significativa a la reducción del RMSE y al aumento del NSE.

Análisis en un período representativo de caudales medios. Para analizar con mayor detalle el comportamiento en un intervalo dominado por eventos de magnitud intermedia, la Figura 7.2 muestra el período julio–diciembre de 2021.

En este intervalo se observan diferencias relevantes:

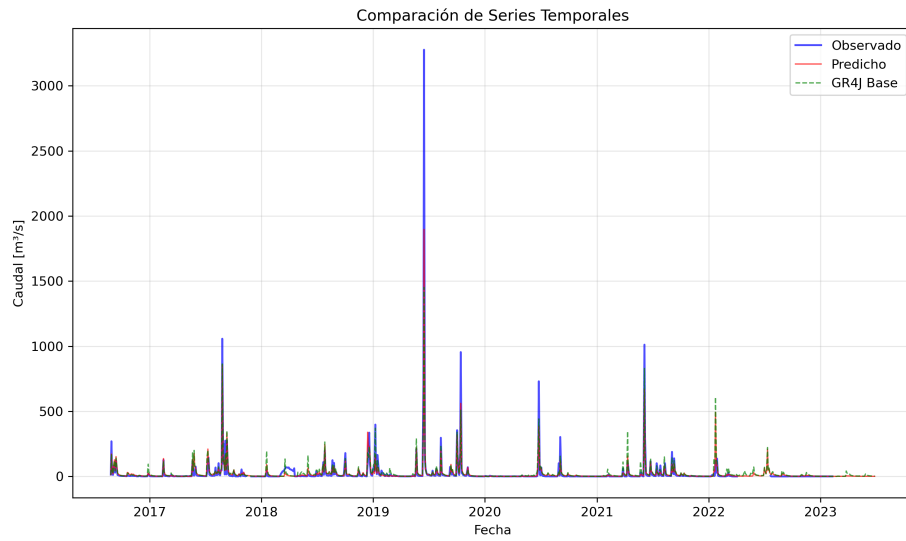


Figura 7.1: Serie temporal completa del período de prueba. Comparación entre observado, GR4J y PINN E RNN.

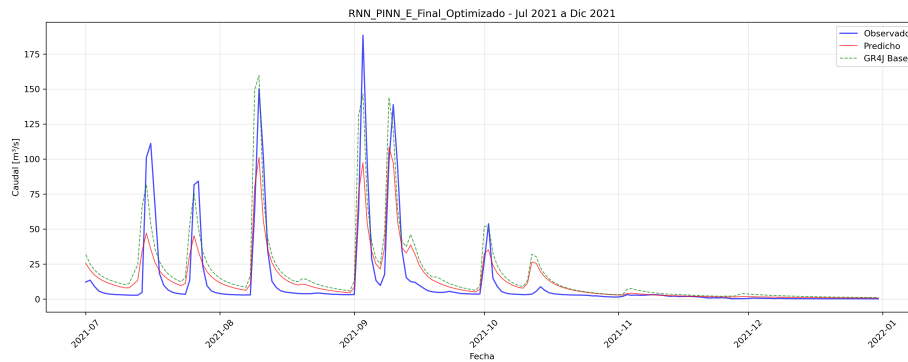


Figura 7.2: Zoom del período julio–diciembre de 2021.

- En varios picos de magnitud intermedia, el modelo **GR4J** reproduce adecuadamente la amplitud de los máximos, mostrando un ajuste comparable o incluso ligeramente superior al de la PINN en términos puntuales.
- En contraste, durante las fases de recesión y en los caudales bajos posteriores a los eventos, la **PINN E RNN** presenta una dinámica más cercana a la observada, con descensos más pronunciados y menor sobreestimación sistemática.

Este comportamiento sugiere que las mejoras introducidas por la PINN no son uniformes en todos los rangos de magnitud. Mientras que en picos extremadamente altos la corrección es favorable, en eventos intermedios el modelo conceptual mantiene un desempeño competitivo.

Distribución global de caudales. La curva de permanencia en escala logarítmica (Figura 7.3) permite analizar el comportamiento en todo el rango hidrológico.

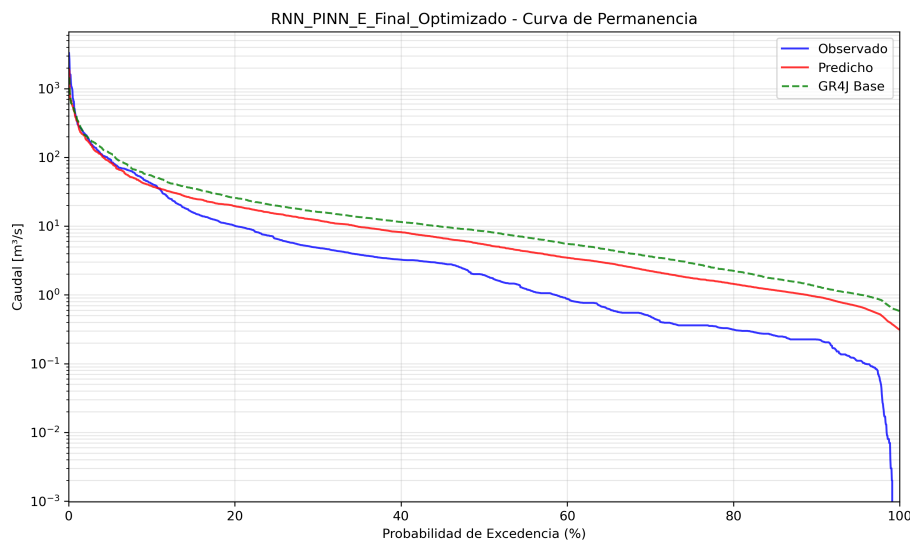


Figura 7.3: Curva de permanencia (escala logarítmica) para observado, GR4J y PINN E RNN.

En el tramo de altas probabilidades de excedencia (caudales bajos), GR4J tiende a sostener valores superiores a los observados, coherente con su PBIAS negativo (-16 %), que indica una sobreestimación volumétrica global.

La **PINN E RNN** reduce esta discrepancia en el rango medio-bajo, acercando su distribución a la observada y disminuyendo la separación sistemática. Este comportamiento explica su mejora sustancial en PBIAS (4.42 %) y confirma que la reducción del sesgo volumétrico no se debe a compensaciones aleatorias, sino a un ajuste consistente del balance hídrico.

En el extremo superior de la distribución (bajas probabilidades de excedencia), ambos modelos presentan cierta subestimación de los máximos más extremos, aunque la PINN reduce parcialmente la brecha respecto a GR4J.

Limitaciones estructurales impuestas por GR4J. Un aspecto central es que la arquitectura híbrida mantiene intacta la estructura conceptual de GR4J. La red neuronal únicamente modifica la señal de evapotranspiración utilizada como forzante, sin alterar los almacenamientos internos ni la función de transferencia.

En consecuencia, la forma general del hidrograma permanece condicionada por las ecuaciones del modelo físico. Las mejoras observadas se manifiestan principalmente en magnitud y balance volumétrico, más que en una transformación estructural de la respuesta hidrológica.

Costo computacional y compromiso precisión–eficiencia. El incremento en desempeño predictivo se obtiene a costa de un mayor costo computacional durante el entrenamiento. La configuración PINN E RNN presentó el mayor tiempo total entre los modelos evaluados, asociado tanto al número de épocas como al procesamiento secuencial requerido por la integración dinámica de GR4J dentro de la función de pérdida.

Este comportamiento refleja un compromiso claro entre precisión predictiva y eficiencia computacional, inherente a la incorporación explícita de restricciones físicas dependientes del estado. No obstante, en el contexto del presente estudio, los tiempos de entrenamiento obtenidos se mantuvieron del orden de minutos, por lo que no constituyeron una limitación operativa significativa.

Sin embargo, este aspecto debería considerarse en escenarios con conjuntos de datos de mayor extensión temporal, resolución más fina o en aplicaciones que requieran reentrenamientos frecuentes. En tales casos, el costo computacional podría incrementarse sustancialmente, dado que la evaluación secuencial del modelo hidrológico limita la paralelización completa del proceso de optimización. Por lo tanto, si bien el enfoque híbrido demuestra ventajas claras en términos de desempeño predictivo, su escalabilidad computacional constituye un factor relevante a tener en cuenta en aplicaciones a mayor escala.

Síntesis parcial. Integrando la evidencia cuantitativa y visual, puede afirmarse que la configuración **PINN E RNN** mejora el desempeño del modelo conceptual principalmente en:

- la representación de picos de gran magnitud,
- la dinámica de recesión y los caudales bajos,
- y la reducción del sesgo volumétrico global.

En el rango de eventos intermedios, el modelo GR4J mantiene un desempeño competitivo y en algunos casos comparable o superior en términos de amplitud puntual.

Estos resultados indican que la mejora introducida por la PINN no proviene de una modificación estructural del modelo hidrológico, sino de una corrección informada de sus forzantes. La siguiente sección analizará en detalle la señal de evapotranspiración E predicha por la red y su impacto en el balance hídrico interno de GR4J.

7.2. Análisis de la evapotranspiración predicha por la PINN

Con el objetivo de comprender el origen de las mejoras observadas en la simulación de caudales, se analizó la señal de evapotranspiración E predicha por la configuración **PINN E RNN** en el conjunto de prueba, comparándola con la evapotranspiración observada utilizada por el modelo conceptual.

La Figura 7.4 muestra la serie temporal de E observada y la estimada por la PINN.

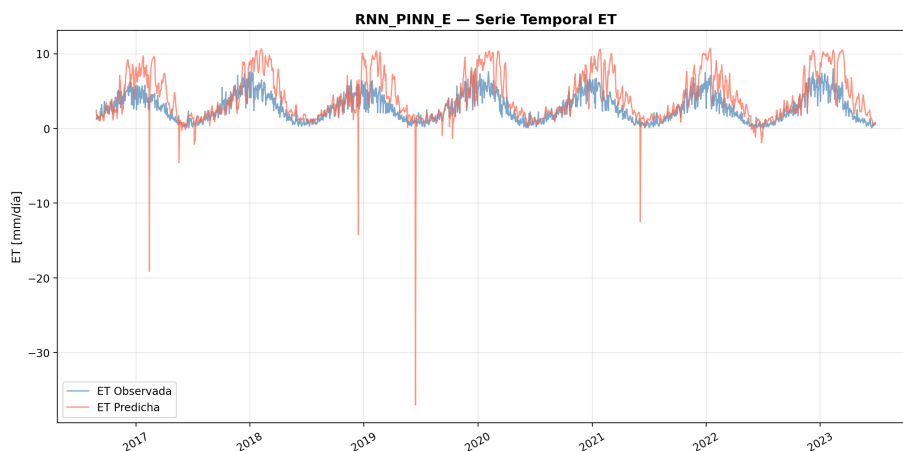


Figura 7.4: Serie temporal de evapotranspiración observada y predicha por la PINN E RNN en el conjunto de prueba.

En términos generales, la red neuronal logra capturar adecuadamente la periodicidad estacional de la evapotranspiración, reproduciendo los ciclos anuales de incremento y disminución asociados a la dinámica climática. Este comportamiento indica que la PINN no ignora la señal física subyacente, sino que la mantiene como componente estructural de la predicción.

Sin embargo, se observan diferencias significativas en magnitud respecto a los valores observados. En numerosos períodos la evapotranspiración predicha supera sistemáticamente a la observada, mientras que en otros se registran subestimaciones pronunciadas. De manera particularmente relevante, se detectan episodios puntuales donde la red genera valores negativos de E .

Desde un punto de vista físico, valores negativos de evapotranspiración carecen de interpretación directa en el contexto hidrológico convencional, ya que implicarían un flujo neto de agua hacia el sistema suelo-planta desde la atmósfera en términos de balance vertical, lo cual no corresponde al proceso modelado. La aparición de estos valores indica que la red no posee restricciones explícitas que impidan soluciones fuera del dominio físicamente plausible.

Este comportamiento puede entenderse a la luz de la función de pérdida utilizada. La PINN optimiza simultáneamente (i) el ajuste entre E observada y estimada, y (ii) el ajuste entre el caudal simulado por GR4J, utilizando la E predicha, y el caudal observado. En consecuencia, la red puede modificar E de manera que compense deficiencias estructurales del modelo conceptual en determinados rangos hidrológicos.

En particular, cuando GR4J tiende a subestimar picos de caudal elevados, la red puede reducir artificialmente la evapotranspiración (e incluso generar valores negativos) para incrementar el agua disponible en el balance hídrico y así amplificar la respuesta del modelo. De manera análoga, en períodos donde GR4J sobreestima caudales bajos, la PINN puede incrementar E por encima de la observada, reduciendo el volumen efectivo y corrigiendo la sobreestimación sistemática.

Este mecanismo explica coherentemente los patrones identificados en la sección anterior: la mejora en la representación de picos de gran magnitud y la reducción del sesgo en caudales bajos están asociadas a ajustes compensatorios en la señal de evapotranspiración.

No obstante, es importante destacar que dichas correcciones no siempre conservan consistencia física estricta. La ausencia de restricciones explícitas (por ejemplo, imponer $E \geq 0$) permite que la red explore soluciones matemáticamente óptimas para minimizar el error en caudal, aun cuando algunas de ellas carezcan de interpretación hidrológica directa.

Desde una perspectiva metodológica, este resultado revela simultáneamente una fortaleza y una limitación del enfoque híbrido. Por un lado, demuestra la capacidad de la PINN para identificar y corregir sesgos estructurales del modelo conceptual a través de sus forzantes. Por otro, evidencia la necesidad de incorporar restricciones explícitas que limiten el espacio de soluciones a valores físicamente plausibles, aspecto que se retoma en las líneas de trabajo futuro.

En síntesis, la mejora en desempeño del modelo híbrido no proviene de una modificación estructural de GR4J, sino de una reinterpretación adaptativa de la evapotranspiración como variable de ajuste dentro del balance hídrico. La red neuronal actúa como un corrector dinámico que redistribuye el agua disponible en el sistema para compensar deficiencias en determinados rangos hidrológicos.

Este análisis permite comprender el origen de las mejoras observadas en caudal: la PINN no reproduce simplemente la evapotranspiración observada, sino que aprende una señal funcional de E que optimiza la respuesta hidrológica global del modelo. Así, el aumento en eficiencia predictiva se explica por una corrección interna del balance de masa, que mejora la representación de picos elevados y reduce la sobreestimación en caudales bajos, aunque a costa de introducir episodios puntuales de inconsistencia física.

Capítulo 8

Ingeniería de Software

8.1. Planificación general

El desarrollo del presente trabajo de tesis se organizó en etapas de alto nivel, acordes al carácter exploratorio del problema y al costo computacional asociado a la experimentación.

El proyecto tuvo una duración inicialmente estimada de ocho meses. No obstante, durante su ejecución se solicitó una prórroga hasta completar doce meses, motivada principalmente por la necesidad de ejecutar un conjunto extenso de pruebas estandarizadas sobre múltiples arquitecturas.

En particular, los procesos de optimización de hiperparámetros, implementados mediante la librería *Optuna*, implicaron la evaluación de un gran número de configuraciones, lo que derivó en tiempos de cómputo acumulados del orden de días, a pesar de que cada entrenamiento individual se ejecuta en escalas de segundos o minutos.

8.2. Diagrama de Gantt

La Figura 8.1 presenta el diagrama de Gantt correspondiente a la planificación efectiva del proyecto, contemplando la extensión del plazo.

8.3. Descripción de las etapas

Durante los primeros meses se desarrolló una etapa de estudio e investigación, orientada a consolidar los fundamentos teóricos necesarios y a delimitar el alcance del trabajo.

Luego se abordó una etapa de implementación de *baselines*. Como primer paso, se desarrolló un modelo base GR4J, utilizado como línea de base conceptual para evaluar el desempeño de los modelos de aprendizaje automático frente a un enfoque hidrológico tradicional.

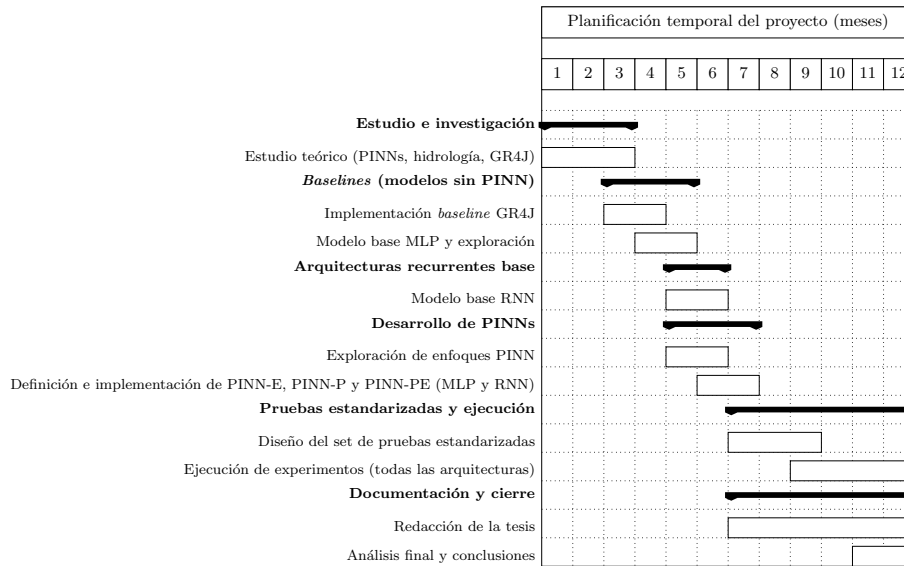


Figura 8.1: Diagrama de Gantt del desarrollo de la tesis.

Posteriormente, se implementaron modelos puramente *data-driven* sin información física incorporada, comenzando con un modelo base MLP y luego una versión equivalente basada en arquitecturas recurrentes (RNN). Estos modelos se utilizaron como referencia para analizar el impacto de la incorporación de conocimiento físico en etapas posteriores.

Sobre esta base se inició el desarrollo de modelos PINN, atravesando una fase exploratoria en la que se evaluaron distintos enfoques para la incorporación de restricciones físicas. Como resultado de este proceso se definieron tres arquitecturas finales: PINN-E, PINN-P y PINN-PE. Cada una de estas arquitecturas fue implementada tanto en versiones MLP como RNN, permitiendo un análisis comparativo consistente.

Hacia el final del período originalmente planificado se diseñó un conjunto de pruebas estandarizadas, basadas en las experiencias obtenidas durante las etapas exploratorias, con el objetivo de evaluar de manera homogénea todas las arquitecturas definidas (*baselines* y modelos PINN).

La ejecución de dichas pruebas se extendió durante gran parte del proyecto, incluyendo el período de prórroga, debido al elevado tiempo de cómputo requerido por algunos experimentos. Esta etapa constituyó el principal factor que motivó la extensión del plazo.

Finalmente, se analizaron los resultados obtenidos, se elaboraron las conclusiones y se completó la redacción final de la tesis.

8.4. Conclusión

La planificación adoptada permitió estructurar el desarrollo de un trabajo de carácter exploratorio, garantizando una evaluación rigurosa y homogénea de las arquitecturas propuestas, aun cuando fue necesario extender el plazo inicialmente estimado.

Capítulo 9

Conclusiones y Trabajo Futuro

9.1. Conclusiones

El presente trabajo tuvo como objetivo general evaluar el desempeño de modelos *Physics-Informed Neural Networks* (PINNs) que incorporan la formulación física del modelo GR4J para la predicción del caudal diario en la subcuenca Florida – Puente Ruta 5, comparándolo con el de un modelo conceptual determinista (GR4J) y modelos puramente *data-driven* (MLP y RNN).

A lo largo del desarrollo se diseñaron, implementaron y compararon múltiples configuraciones, modelos puramente basados en datos y variantes híbridas, evaluadas sobre un conjunto de prueba independiente correspondiente al 20 % final de los datos disponibles. El análisis integró métricas cuantitativas (NSE, PBIAS, MAE, RMSE), inspección visual de series temporales, curvas de duración y análisis de eventos extremos.

Principales resultados alcanzados

En relación con los objetivos específicos planteados, los resultados obtenidos permiten extraer las siguientes conclusiones principales:

- El modelo conceptual **GR4J** calibrado constituyó una línea base sólida (NSE 0.69, categoría *buena*), validando su utilidad como referencia para la comparación con los demás enfoques.
- Los modelos puramente *data-driven* mostraron desempeños diferenciados: el **MLP** alcanzó un balance volumétrico *muy bueno* (PBIAS 6.52 %) con eficiencia similar a GR4J, mientras que la **RNN** presentó menor capacidad de generalización en el conjunto de prueba (NSE 0.59).
- La incorporación de restricciones físicas mediante la arquitectura **PINN E RNN** permitió mejorar el desempeño predictivo global respecto al modelo

conceptual GR4J, alcanzando el mayor NSE (0.77, *muy bueno*) y el menor RMSE (62.36) entre las configuraciones evaluadas.

- La comparación de distintos esquemas de acoplamiento (predicción de P , E o ambas) mostró que la variante que predice únicamente la evapotranspiración con arquitectura recurrente (PINN E RNN) ofrece el mejor compromiso entre eficiencia global y consistencia volumétrica.
- La mejora no fue uniforme en todos los rangos hidrológicos. Se observaron avances claros en la representación de picos de gran magnitud y en la dinámica de recesión y caudales bajos, mientras que en eventos de magnitud intermedia el modelo conceptual mantuvo un desempeño competitivo.
- El análisis de la evapotranspiración predicha mostró que la red neuronal actúa como un corrector dinámico del balance hídrico interno de GR4J. Las mejoras en caudal se explican por ajustes compensatorios en la señal de E , que redistribuyen la disponibilidad de agua dentro del modelo.
- No obstante, la ausencia de restricciones físicas explícitas permitió que la red genere valores puntuales de evapotranspiración físicamente inconsistentes (por ejemplo, valores negativos), evidenciando una limitación estructural del enfoque implementado.
- El costo computacional de entrenamiento de las configuraciones PINN resultó significativamente mayor que el de los modelos puramente basados en datos. Si bien en el contexto del presente trabajo los tiempos fueron manejables (del orden de minutos), la escalabilidad computacional constituye un aspecto relevante en escenarios de mayor complejidad.

Aprendizajes y limitaciones del trabajo

Durante el desarrollo surgieron diversos aprendizajes metodológicos y se identificaron limitaciones que permiten realizar una autocrítica constructiva del proceso:

- **Alcance geográfico:** todos los resultados corresponden a una única subcuenca (Florida – Puente Ruta 5) del sistema hidrográfico del río Santa Lucía. La transferibilidad de las conclusiones a cuencas con características hidrológicas, climáticas o de cobertura del suelo diferentes no ha sido evaluada y constituye una limitación del estudio.
- La optimización de hiperparámetros mediante Optuna se realizó utilizando la función de pérdida de la red como función objetivo. Sin embargo, dado que la evaluación final se centró en métricas hidrológicas como NSE y PBIAS, podría haber sido más consistente optimizar directamente en función de dichas métricas. Esta decisión pudo haber influido en ciertas discrepancias observadas entre desempeño durante entrenamiento y evaluación final.

- El hecho de no provenir estrictamente del área hidrológica implicó que algunos criterios de ajuste y selección intermedia de modelos podrían no haber estado completamente alineados con prácticas habituales del dominio. Este aspecto influyó en decisiones metodológicas que podrían refinarse en futuros desarrollos.
- No se incorporaron restricciones explícitas que garantizaran consistencia física en la evapotranspiración predicha, lo cual permitió soluciones matemáticamente eficientes pero parcialmente inconsistentes desde el punto de vista hidrológico.
- La aplicación de *Integrated Gradients* a los modelos recurrentes (RNN, LSTM, GRU) se realizó evaluando cada muestra de forma independiente, sin considerar las $W - 1$ dimensiones adicionales de la ventana temporal. Esta limitación implica que el contexto secuencial completo que integra el modelo no queda reflejado en las atribuciones calculadas: las entradas de los pasos anteriores de la ventana no son procesadas por IG, lo que puede distorsionar la distribución de importancia entre variables. Los análisis de IG sobre arquitecturas recurrentes reportados en este trabajo constituyen por tanto aproximaciones orientativas de la relevancia relativa de las variables.

En conjunto, el trabajo logró cumplir el objetivo planteado: evaluar si la incorporación de aprendizaje basado en datos dentro de un esquema físico mejora el desempeño predictivo manteniendo coherencia estructural. Los resultados obtenidos indican que el enfoque híbrido puede aportar mejoras relevantes, aunque con compromisos asociados en términos de interpretabilidad física y costo computacional.

9.2. Trabajo futuro

A partir de los resultados obtenidos y las limitaciones identificadas, se proponen las siguientes líneas de trabajo futuro:

Mejoras en la optimización y criterios de entrenamiento

- Implementar procesos de optimización de hiperparámetros que utilicen directamente métricas hidrológicas como NSE y/o PBIAS como función objetivo, en lugar de la función de pérdida, alineando así el proceso de entrenamiento con los criterios finales de evaluación.

Restricciones físicas y consistencia hidrológica

- Incorporar restricciones explícitas que aseguren $E \geq 0$, por ejemplo mediante funciones de activación adecuadas o penalizaciones específicas en la función de pérdida.

- Evaluar esquemas de regularización física que limiten el rango de corrección permitido sobre la evapotranspiración.

Extensiones estructurales del modelo

- Analizar la posibilidad de utilizar un período de *warm-up* variable para GR4J, determinado dinámicamente en función de la estabilización de sus almacenamientos internos.
- Evaluar arquitecturas recurrentes con conexiones *skip* o mecanismos de memoria extendida para mejorar la captura de dependencias temporales de largo plazo.
- Explorar el uso de otros modelos hidrológicos conceptuales o semi-distribuidos (por ejemplo, SWAT) como estructura física base dentro del esquema híbrido.

Interpretabilidad y análisis de atribuciones

- Extender el cálculo de *Integrated Gradients* para considerar explícitamente la totalidad de la secuencia de entrada en modelos recurrentes, evaluando atribuciones conjuntas sobre las W dimensiones temporales.

Información adicional y generalización

- Estudiar el impacto de incorporar retardos (*lags*) del caudal como entradas adicionales en los modelos evaluados, utilizando distintas estrategias para el manejo de los valores faltantes.
- Validar el enfoque propuesto en otras subcuencas del sistema hidrográfico del río Santa Lucía, así como en cuencas con características hidrológicas contrastantes, para evaluar la transferibilidad y robustez de las conclusiones.

Cierre final

En conjunto, las líneas propuestas permiten vislumbrar una evolución natural del enfoque híbrido desarrollado en este trabajo. La integración entre modelos físicos y aprendizaje automático no solo representa una mejora incremental en desempeño predictivo, sino una dirección metodológica con potencial para redefinir el equilibrio entre interpretabilidad y capacidad de ajuste en modelado hidrológico, particularmente en contextos donde se dispone de estructura conceptual previa y datos observacionales suficientes.

El presente estudio constituye así una base sobre la cual pueden construirse desarrollos más robustos, físicamente consistentes y aplicables a un rango más amplio de escenarios hidrológicos.

Referencias

- Akiba, T., Sano, S., Yanase, T., Ohta, T., y Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. En *Proceedings of the 25th acm sigkdd international conference on knowledge discovery & data mining* (pp. 2623–2631). ACM. doi: 10.1145/3292500.3330701
- ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. (2000a). Artificial neural networks in hydrology. II: Hydrologic applications. *Journal of Hydrologic Engineering*, 5(2), 124–137. doi: 10.1061/(ASCE)1084-0699(2000)5:2(124)
- ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. (2000b). Artificial neural networks in hydrology. I: Preliminary concepts. *Journal of Hydrologic Engineering*, 5(2), 115–123. doi: 10.1061/(ASCE)1084-0699(2000)5:2(115)
- Bandai, T., y Ghezzehei, T. A. (2022). Forward and inverse modeling of water flow in unsaturated soils with discontinuous hydraulic conductivities using physics-informed neural networks with domain decomposition. *Hydrology and Earth System Sciences*, 26, 4469–4495. doi: 10.5194/hess-26-4469-2022
- Bengio, Y., Simard, P., y Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2), 157–166. doi: 10.1109/72.279181
- Besaw, L. E., Rizzo, D. M., Bierman, P. R., y Hackett, W. R. (2010). Advances in ungauged streamflow prediction using artificial neural networks. *Journal of Hydrology*, 385, 1–?? doi: 10.1016/j.jhydrol.2010.02.037
- Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., y Bengio, Y. (2014). Learning phrase representations using rnn encoder–decoder for statistical machine translation. En *Proceedings of the 2014 conference on empirical methods in natural language processing (emnlp)* (pp. 1724–1734).
- Chow, V. T., Maidment, D. R., y Mays, L. W. (1988). *Applied hydrology*. New York, NY: McGraw-Hill.
- Deb, K., Pratap, A., Agarwal, S., y Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, 6(2), 182–197. doi: 10.1109/4235.996017
- Dirección Nacional de Aguas (DINAGUA). (2026). *Sistema nacional de información hídrica (sih) – datos hidrológicos abiertos*. Descarga-

- do 2026-02-26, de <https://www.ambiente.gub.uy/SIH-JSF/paginas/visualizador/visualizador.xhtml> (Ministerio de Ambiente, Uruguay. Datos utilizados correspondientes al período hasta 2023)
- Duque, J. S., do Santos, L., Santos, R., Arteaga, J. A., y Aubet, N. (2022). *Level river forecasting using empirical hydrological modeling for rio negro basin uruguay*. (Proceedings paper (SBMAC)) doi: 10.5540/03.2022.009.01.0269
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14(2), 179–211. doi: 10.1207/s15516709cog1402_1
- Gorgoglione, A., Gregorio, J., Ríos, A., Alonso, J., Chreties, C., y Fossati, M. (2020). Influence of land use/land cover on surface-water quality of santa lucía river, uruguay. *Sustainability*, 12(11), 4692. Descargado de <https://www.mdpi.com/2071-1050/12/11/4692> doi: 10.3390/su12114692
- Hochreiter, S., y Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. doi: 10.1162/neco.1997.9.8.1735
- Hornik, K., Stinchcombe, M., y White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359–366. doi: 10.1016/0893-6080(89)90020-8
- Hou, Q., Li, Y., Singh, V. P., y Sun, Z. (2024). Physics-informed neural network for diffusive wave model. *Journal of Hydrology*, 637, 131261. doi: 10.1016/j.jhydrol.2024.131261
- Hsu, K.-L., Gupta, H. V., y Sorooshian, S. (1995). Artificial neural network modeling of the rainfall–runoff process. *Water Resources Research*, 31(10), 2517–2530. doi: 10.1029/95WR01955
- Instituto Nacional de Investigación Agropecuaria (INIA). (2026). *Banco de datos agroclimático – unidad gras*. Descargado 2026-02-26, de <http://www.inia.uy/gras/Clima/Banco-datos-agroclimatico> (INIA, Uruguay. Datos utilizados correspondientes al período hasta 2023)
- Instituto Uruguayo de Meteorología (INUMET). (2023). *Registros históricos de precipitación diaria*. (Datos provistos mediante solicitud formal con fines académicos. INUMET, Uruguay)
- Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., y Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, 3(6), 422–440. doi: 10.1038/s42254-021-00314-5
- Kim, J.-K., y Park, S. (2021). Robust regression imputation. *Journal of the American Statistical Association*, 116(533), 400–412. doi: 10.1080/01621459.2020.1775744
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K., y Herrnegger, M. (2018). Rainfall–runoff modelling using long short-term memory (lstm) networks. *Hydrology and Earth System Sciences*, 22(11), 6005–6022. doi: 10.5194/hess-22-6005-2018
- Li, Y., Sun, Q., Fu, Y., y Wei, J. (2025). Solving the richards infiltration equation by coupling physics-informed neural networks with Hydrus-1D. *Scientific Reports*, 15, 18649. doi: 10.1038/s41598-025-02978-w
- Li, Z., Mu, T., Li, X., Li, P., Feng, J., Xu, H., . . . Qian, S. (2025). Physics-informed neural network for hydraulic prediction in open-channel water

- transfer projects with sparse monitoring data. *Water Research*, 287, 124507. doi: 10.1016/j.watres.2025.124507
- Mahajan, T., Singh, G., y Bruns, G. (2021, marzo). An experimental assessment of treatments for cyclical data. En *Proceedings of the 2021 computer science conference for csu undergraduates*. Descargado de <https://cscsu-conference.github.io/index.html> (Virtual conference)
- Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., y Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3), 885–900. doi: 10.13031/2013.23153
- Moriasi, D. N., Gitau, M. W., Pai, N., y Daggupati, P. (2015). Hydrologic and water quality models: Performance measures and evaluation criteria. *Transactions of the ASABE*, 58(6), 1763–1785. doi: 10.13031/trans.58.10715
- Narbondo, S., Gorgoglione, A., Crisci, M., y Chreties, C. (2020). Enhancing physical similarity approach to predict runoff in ungauged watersheds in sub-tropical regions. *Water*, 12(528), 1–19. Descargado de <https://www.mdpi.com/2073-4441/12/2/528>
- Nash, J. E., y Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part i—a discussion of principles. *Journal of Hydrology*, 10(3), 282–290. doi: 10.1016/0022-1694(70)90255-6
- Nesmachnow, S., y Iturriaga, S. (2019). Cluster-uy: Collaborative scientific high performance computing in uruguay. En M. Torres y J. Klapp (Eds.), *Supercomputing: Third international conference, isum 2019, guadalajara, mexico, march 18–22, 2019, revised selected papers* (Vol. 1151, pp. 188–202). Springer, Cham. doi: 10.1007/978-3-030-38043-4_16
- Perrin, C., Michel, C., y Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, 279(1–4), 275–289.
- Raissi, M., Perdikaris, P., y Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378, 686–707. doi: 10.1016/j.jcp.2018.10.045
- Rumelhart, D. E., Hinton, G. E., y Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536. doi: 10.1038/323533a0
- Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. En *Proceedings of the 1968 23rd acm national conference* (pp. 517–524). ACM. doi: 10.1145/800186.810616
- Sundararajan, M., Taly, A., y Yan, Q. (2017). Axiomatic attribution for deep networks. En *Proceedings of the 34th international conference on machine learning* (Vol. 70, pp. 3319–3328). PMLR. Descargado de <https://proceedings.mlr.press/v70/sundararajan17a.html>
- Teegavarapu, R. S. V., y Chandramouli, V. (2005). Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. *Journal of Hydrology*, 312(1–4), 191–206.

- Thiessen, A. H. (1911). Precipitation averages for large areas. *Monthly Weather Review*, 39(7), 1082–1089. doi: 10.1175/1520-0493(1911)39(1082b:PAFLA)2.0.CO;2
- Tian, Y., Ding, S., Huang, L., Su, G., y Chen, J. (2025). Physics-informed neural networks for solving the two-dimensional shallow water equations with terrain topography and rainfall source terms. *Water Resources Research*, 61(9). doi: 10.1029/2025WR040052
- Tripathy, K. P., y Mishra, A. K. (2024). Deep learning in hydrology and water resources disciplines: concepts, methods, applications, and research directions. *Journal of Hydrology*, 628, 130458. doi: 10.1016/j.jhydrol.2023.130458
- Vilaseca, F., Castro, A., Chreties, C., y Gorgoglione, A. (2023). Assessing influential rainfall–runoff variables to simulate daily streamflow using random forest. *Hydrological Sciences Journal*. Descargado de <https://doi.org/10.1080/02626667.2023.2232356> (Universidad de la República, Montevideo, Uruguay) doi: 10.1080/02626667.2023.2232356

Anexo A

Análisis e imputación de datos

Este anexo documenta en detalle el proceso de análisis de valores faltantes, la estrategia de imputación aplicada y los resultados de la validación cruzada que justifican la selección del método empleado.

A.1. Análisis de valores faltantes

Con el objetivo de evaluar la completitud de las series temporales y determinar la necesidad de aplicar técnicas de imputación, se analizó la proporción de valores faltantes en cada variable del conjunto de datos. La Tabla A.1 resume el número total de observaciones ausentes y su porcentaje relativo (que ya se había reportado en la descripción de variables del capítulo anterior) respecto al total de registros disponibles (15.884 filas).

Tabla A.1: Cantidad y porcentaje de valores faltantes por variable.

Variable	Valores faltantes	Porcentaje [%]
Precipitación (P) [mm] — Florida	3421	21.54
Caudal (Q) [m ³ /s] — Florida (Puente Ruta 5)	1194	7.52
Precipitación (P) [mm] — La Cruz	388	2.44
Precipitación (P) [mm] — San Gabriel	368	2.32
Precipitación (P) [mm] — Sarandí Grande	217	1.37
Precipitación (P) [mm] — Cerro Colorado	30	0.19
Velocidad del viento (VV) [2m/km/24h] — INIA Las Brujas	17	0.11
Temperatura del aire media (TA) [°C (24hs)] — INIA Las Brujas	16	0.10
Evapotranspiración Penman (ET) [mm] — INIA Las Brujas	15	0.09
Humedad relativa media (HR) [%] — INIA Las Brujas	15	0.09

Si bien la tabla permite cuantificar la magnitud de los valores ausentes, resulta necesario analizar también su distribución temporal. Para ello se elaboró un mapa binario de valores faltantes, mostrado en la Figura A.1, donde el color rojo representa la presencia de datos faltantes y el verde los registros válidos.

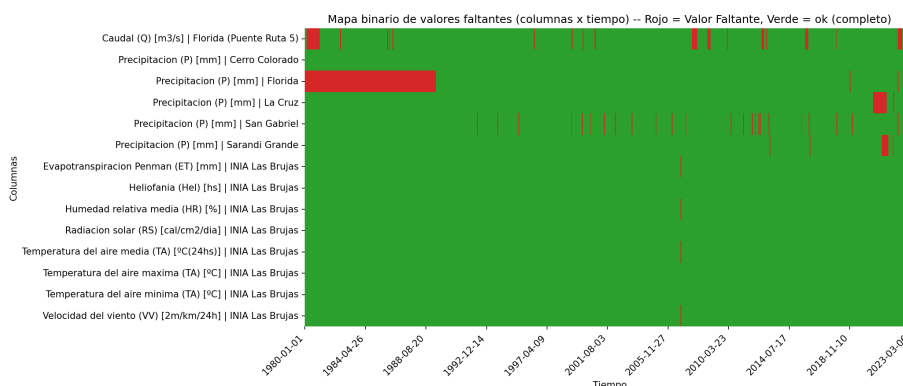


Figura A.1: Mapa binario de valores faltantes por variable a lo largo del tiempo. En rojo se indican los valores ausentes y en verde los registros completos.

La Figura A.1 permite visualizar claramente los patrones de ausencia a lo largo del período de estudio. La mayor concentración de valores faltantes corresponde a la precipitación registrada en la estación *Florida*, donde se observa un tramo inicial sin datos que se extiende durante desde el 1 de Enero de 1980 hasta el 1 de Mayo de 1989. Este comportamiento sugiere que el pluviómetro pudo no haber estado operativo o instalado durante los primeros años de la serie.

En contraste, las restantes estaciones de precipitación (*La Cruz*, *San Gabriel*, *Sarandí Grande* y *Cerro Colorado*) exhiben una cobertura mucho más completa, con ausencias esporádicas y de menor duración. La ventana temporal más prolongada de datos faltantes entre estas corresponde a la estación *La Cruz*, con una interrupción de 365 días entre el 1 de julio de 2020 y el 30 de junio de 2021. En *Sarandí Grande* se observa una brecha de 166 días (del 16 de febrero al 31 de julio de 2021), mientras que *San Gabriel* y *Cerro Colorado* presentan interrupciones breves de solo unos pocos días.

Por su parte, la serie de caudal en *Florida* (*Puente Ruta 5*) presenta vacíos del 7.52 %, concentrados principalmente en el tramo inicial y en algunos períodos aislados posteriores, posiblemente asociados a fallas temporales en los sensores o tareas de mantenimiento en la estación de aforo.

Finalmente, las variables meteorológicas medidas por el *INIA Las Brujas* muestran una continuidad casi total, sin intervalos prolongados de valores faltantes, lo que confirma la estabilidad y calidad de las mediciones provenientes de dicha fuente.

A partir de estos resultados, se identificaron las variables y períodos más afectados por la falta de información, lo que permitió definir el tratamiento

adecuado en cada caso. En particular, la ausencia prolongada de registros en la estación *Florida* motivó el recorte del período inicial (desde el 1 de Enero de 1980 hasta el 1 de Mayo de 1989) del conjunto de datos, de modo que el análisis y la imputación se desarrollaran sobre un intervalo temporal continuo y homogéneo, libre de lagunas simultáneas entre variables.

La Tabla A.2 presenta la cantidad y el porcentaje de valores faltantes remanentes tras la depuración inicial, evidenciando una reducción significativa en comparación con el estado previo.

Tabla A.2: Cantidad y porcentaje de valores faltantes por variable tras el recorte inicial.

Variable	Valores faltantes	Porcentaje [%]
Caudal (Q) [m ³ /s] — Florida (Puente Ruta 5)	796	6.38
Precipitación (P) [mm] — La Cruz	387	3.10
Precipitación (P) [mm] — San Gabriel	367	2.94
Precipitación (P) [mm] — Sarandí Grande	216	1.73
Precipitación (P) [mm] — Cerro Colorado	29	0.23
Velocidad del viento (VV) [2m/km/24h] — INIA Las Brujas	17	0.14
Temperatura del aire media (TA) [°C (24hs)] — INIA Las Brujas	16	0.13
Evapotranspiración Penman (ET) [mm] — INIA Las Brujas	15	0.12
Humedad relativa media (HR) [%] — INIA Las Brujas	15	0.12
Precipitación (P) [mm] — Florida	13	0.10

La Figura A.2 muestra la nueva distribución de los valores faltantes luego del recorte temporal aplicado. En comparación con el mapa anterior, se observa una reducción considerable en la cantidad de huecos, particularmente en la estación *Florida*, donde desaparece el extenso tramo inicial sin registros. En general, las ausencias remanentes se presentan de forma más dispersa y acotada, lo que indica una mejora sustancial en la consistencia del conjunto de datos.

Este procedimiento permitió establecer una base temporal uniforme y reducir el riesgo de sesgos derivados de lagunas prolongadas en los datos históricos. A partir de esta versión depurada del *dataset*, se implementaron diferentes estrategias de imputación según el tipo y la naturaleza de las variables consideradas, con el objetivo de completar los valores faltantes manteniendo la coherencia física y estadística de las series.

A.2. Estrategia de imputación

A.2.1. Imputación de la precipitación

La precipitación constituye una variable con alta variabilidad, lo que hace que su imputación represente un desafío particularmente complejo. A diferencia de variables atmosféricas como la temperatura o la humedad, que presentan

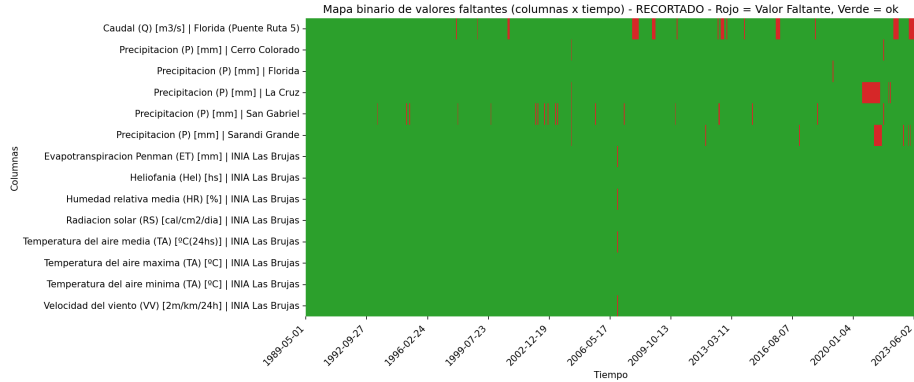


Figura A.2: Mapa binario de valores faltantes tras el recorte inicial del período de análisis.

cambios suaves y pueden considerarse uniformes a escala regional, la precipitación muestra patrones localizados muy influenciados por la topografía, y otros procesos. Esto implica que el valor faltante en una estación no puede estimarse simplemente promediando las estaciones vecinas sin considerar su localización geográfica relativa.

Para abordar este problema se implementó el método de **Interpolación por Distancia Inversa Ponderada (IDW, por sus siglas en inglés *Inverse Distance Weighting*)**, muy utilizado en hidrología para la estimación espacial de precipitaciones a partir de observaciones puntuales dispersas (Chow y cols., 1988). Este método se fundamenta en el principio de que estaciones más cercanas geográficamente presentan mayor similitud en sus registros pluviométricos que aquellas más alejadas, y por tanto deben tener mayor peso en la estimación del valor faltante.

Fundamento teórico del método IDW El método IDW estima el valor faltante de precipitación en una estación objetivo a partir de una combinación ponderada de las observaciones disponibles en las estaciones vecinas. La ponderación se determina en función de la distancia euclidiana entre las estaciones, asignando mayor peso a aquellas más próximas y reduciendo progresivamente la influencia de las más alejadas.

Formalmente, para una estación objetivo i con valor faltante en el instante t , el valor imputado $\hat{P}_i(t)$ se calcula como:

$$\hat{P}_i(t) = \frac{\sum_{j \neq i} w_{ij} \cdot P_j(t)}{\sum_{j \neq i} w_{ij}} \quad (\text{A.1})$$

donde:

- $P_j(t)$ es la precipitación observada en la estación j en el instante t

- w_{ij} es el peso asignado a la estación j para estimar el valor en la estación i
- La sumatoria se realiza sobre todas las estaciones j que disponen de datos válidos en el instante t , excluyendo la estación objetivo i

Los pesos w_{ij} se definen como la inversa de la distancia elevada a una potencia p :

$$w_{ij} = \frac{1}{d_{ij}^p} \quad (\text{A.2})$$

donde:

- d_{ij} es la distancia euclidiana entre las estaciones i y j , calculada en coordenadas geográficas (latitud, longitud)
- p es el exponente de distancia, que controla la tasa de decaimiento del peso con la distancia

En la literatura hidrológica, el valor más utilizado para el exponente es $p = 2$ (Shepard, 1968; Teegavarapu y Chandramouli, 2005), lo que corresponde a la inversa del cuadrado de la distancia. Este valor ha demostrado proporcionar un balance adecuado entre la influencia de estaciones cercanas y lejanas, evitando tanto una concentración excesiva del peso en la estación más próxima como una distribución demasiado uniforme que ignore la estructura espacial de los datos. Para la implementación en este trabajo se adoptó este valor estándar.

Distancia euclidiana entre estaciones La distancia d_{ij} entre dos estaciones i y j se calcula a partir de sus coordenadas geográficas mediante la métrica euclidiana:

$$d_{ij} = \sqrt{(\text{lat}_i - \text{lat}_j)^2 + (\text{lon}_i - \text{lon}_j)^2} \quad (\text{A.3})$$

donde lat_i y lon_i representan la latitud y longitud de la estación i , respectivamente. Si bien esta aproximación no considera la curvatura terrestre —lo que sería más preciso con distancias geodésicas—, resulta suficientemente exacta para la escala espacial de la cuenca del Santa Lucía, donde las diferencias entre ambas métricas son despreciables.

Criterios de aplicación y tratamiento de casos especiales Durante la implementación del método IDW se aplicaron los siguientes criterios para garantizar la robustez y coherencia física de las imputaciones:

- **Número mínimo de estaciones:** Se requiere al menos una estación vecina con datos no faltantes para realizar la imputación. Si ninguna estación dispone de datos en el instante considerado, el valor faltante se trata mediante interpolación temporal lineal como método de respaldo (descrito más adelante).

- **Umbral de no precipitación:** En casos donde la mayoría de las estaciones vecinas registran precipitación nula (valor cero), se asume que el evento pluviométrico no afectó la región y se imputa un valor de cero en la estación objetivo, evitando la estimación de precipitaciones artificiales pequeñas producto del promedio ponderado.

Interpolación temporal como método de respaldo En situaciones excepcionales donde no es posible aplicar el método IDW —por ejemplo, cuando todas las estaciones vecinas presentan simultáneamente valores faltantes en un mismo instante—, se recurre a la **interpolación temporal lineal** como estrategia de respaldo. Esta técnica estima el valor faltante a partir de los valores válidos inmediatamente anteriores y posteriores en la misma estación:

$$\hat{P}_i(t) = P_i(t - \Delta t) + \frac{P_i(t + \Delta t) - P_i(t - \Delta t)}{2\Delta t} \cdot \Delta t \quad (\text{A.4})$$

donde Δt representa el intervalo temporal hasta la observación válida más cercana (en días). La interpolación temporal se aplica únicamente a **lagunas cortas**, típicamente de hasta 7 días consecutivos, dado que períodos más extensos requieren aproximaciones más sofisticadas que consideren patrones estacionales o climatológicos.

Este método de respaldo garantiza que, incluso en condiciones adversas de disponibilidad de datos, el conjunto de datos puede completarse manteniendo cierta coherencia temporal sin introducir discontinuidades en las series.

A.2.2. Imputación de variables meteorológicas (INIA Las Brujas)

A diferencia de la precipitación, que presenta alta variabilidad espacial y requiere observaciones de múltiples estaciones distribuidas en la cuenca, las variables meteorológicas registradas en la estación INIA Las Brujas (evapotranspiración, temperatura del aire, velocidad del viento y humedad relativa) exhiben gradientes espaciales suaves y menor variabilidad local. Estas características físicas permiten que una única estación represente de manera adecuada las condiciones atmosféricas de toda la región de estudio, siempre que se encuentre ubicada dentro o en las proximidades de la cuenca.

Como se indicó en la sección de análisis de valores faltantes (Tabla A.2), estas variables presentan una proporción muy baja de datos ausentes tras el recorte temporal inicial: menos del 0.15 % en todos los casos. Además, los valores faltantes se distribuyen de forma dispersa a lo largo de la serie temporal, sin presentar lagunas prolongadas ni períodos continuos sin registro.

Formulación matemática Para una variable meteorológica X medida en la estación INIA Las Brujas, el valor imputado en el instante t con dato faltante se calcula mediante interpolación lineal entre las observaciones válidas inmediatamente anteriores y posteriores:

$$\hat{X}(t) = X(t_{\text{ant}}) + \frac{X(t_{\text{post}}) - X(t_{\text{ant}})}{t_{\text{post}} - t_{\text{ant}}} \cdot (t - t_{\text{ant}}) \quad (\text{A.5})$$

donde:

- $X(t_{\text{ant}})$ es el último valor válido observado antes del instante t
- $X(t_{\text{post}})$ es el primer valor válido observado después del instante t
- t_{ant} y t_{post} son los instantes temporales correspondientes a dichas observaciones

En el caso de lagunas en los extremos de la serie temporal (valores faltantes al inicio o al final), se aplica **extrapolación hacia adelante** (*forward fill*) o **hacia atrás** (*backward fill*), respectivamente, utilizando el valor válido más cercano. Este criterio garantiza la completitud del conjunto de datos sin introducir discontinuidades artificiales en los bordes temporales.

VARIABLES TRATADAS CON INTERPOLACIÓN TEMPORAL LINEAL Las variables meteorológicas de INIA Las Brujas imputadas mediante este método son:

- **Evapotranspiración Penman (ET)** [mm]
- **Temperatura del aire media (TA)** [°C]
- **Velocidad del viento (VV)** [2m/km/24h]
- **Humedad relativa media (HR)** [%]

A.2.3. Imputación del caudal

A diferencia de las variables meteorológicas (precipitación, evapotranspiración, temperatura), el caudal constituye la **variable objetivo** del presente estudio. Esta distinción es fundamental: imputar valores faltantes de caudal introduciría información artificial justamente en la variable que se pretende predecir, sesgando la validación del modelo y comprometiendo la evaluación de su desempeño real.

Por este motivo, se adoptó la siguiente estrategia:

- **No imputación del caudal:** Los valores faltantes de caudal **no se imputan** bajo ninguna circunstancia. Los registros sin observación de caudal se mantienen en el conjunto de datos, pero se utilizan exclusivamente como soporte temporal para las variables forzantes.
- **Preservación de la continuidad temporal:** En las variantes que incorporan el modelo hidrológico conceptual GR4J, las variables forzantes (precipitación P y evapotranspiración E) imputadas se propagan a través del modelo para actualizar su estado interno, incluso en días sin observación de caudal. De este modo se preserva la memoria hidrológica de la cuenca sin introducir valores sintéticos de caudal.

- **Uso selectivo en el entrenamiento:** Durante el entrenamiento y la evaluación de los modelos, únicamente los registros con caudal observado contribuyen al cálculo de las funciones de pérdida y de las métricas de desempeño.

Esto preserva la integridad de la variable objetivo y garantiza que las métricas de desempeño reportadas se basen exclusivamente en datos de caudal observados, lo que mantiene la coherencia física del estado hidrológico interno del modelo.

A.2.4. Evaluación de métodos alternativos de imputación

Con el objetivo de validar la elección del método IDW para la imputación de precipitaciones y evaluar si existen alternativas que pudieran ofrecer mejor desempeño, se realizó un estudio comparativo exhaustivo entre diferentes estrategias de imputación. La comparación incluyó tanto métodos tradicionales de uso general como técnicas especializadas para datos hidrológicos y meteorológicos.

Metodología de evaluación Para evitar sesgos derivados del uso del conjunto de testing en la selección de métodos de preprocesamiento, la evaluación se realizó mediante una técnica de **validación cruzada de imputación**. Este enfoque evalúa la calidad intrínseca de cada método de imputación sin necesidad de entrenar el modelo predictivo final.

La metodología consiste en los siguientes pasos:

1. Se seleccionan aleatoriamente valores **conocidos** del conjunto de datos (aproximadamente 10 % de los valores disponibles en cada iteración).
2. Los valores seleccionados se eliminan temporalmente, simulando valores faltantes artificiales.
3. Se aplica el método de imputación candidato para estimar estos valores artificialmente faltantes.
4. Se comparan los valores imputados con los valores reales previamente guardados.
5. Este proceso se repite durante 5 iteraciones con diferentes muestras aleatorias para obtener estimaciones robustas del desempeño.

Esta técnica de validación cruzada es análoga a la validación cruzada utilizada en el entrenamiento de modelos, pero aplicada específicamente a la imputación de datos. Permite evaluar objetivamente la capacidad de cada método para recuperar valores faltantes sin comprometer la integridad de la validación.

Sea y_i el valor real observado y \hat{y}_i el valor imputado correspondiente, para $i = 1, \dots, n$, donde n representa el número total de observaciones consideradas en

la evaluación. Se denotan por \bar{y} y \tilde{y} las medias de los valores reales e imputados, respectivamente.

Con esta notación, las métricas de evaluación consideradas se formulan como sigue:

- **Error Absoluto Medio (MAE)**: medida de la magnitud promedio del error de imputación, expresada en las unidades originales de cada variable:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

- **Raíz del Error Cuadrático Medio (RMSE)**: métrica que penaliza más fuertemente los errores grandes de imputación:

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

- **Correlación de Pearson (r)**: medida de la similitud lineal entre los valores reales e imputados:

$$r = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \tilde{y})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\hat{y}_i - \tilde{y})^2}}$$

- **Sesgo (Bias)**: diferencia promedio entre valores imputados y reales, que identifica si el método tiende a sobreestimar o subestimar sistemáticamente:

$$\text{Bias} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)$$

- **Preservación de datos (%)**: proporción de registros temporales conservados tras la imputación, relativa al conjunto original:

$$\text{Preservación} = \frac{N_{\text{válidos tras imputación}}}{N_{\text{original}}} \times 100$$

- **Tiempo de ejecución**: se definió como el tiempo total requerido por el algoritmo para completar el proceso de imputación, medido en segundos.

Métodos evaluados Se implementaron y evaluaron los siguientes métodos de imputación:

1. **Interpolación IDW (*Inverse Distance Weighted*)**: Método seleccionado para este estudio, basado en promedios ponderados por la inversa del cuadrado de la distancia entre estaciones. Fundamentado en la metodología de Chow para imputación de precipitaciones (Chow y cols., 1988).
2. **Interpolación por Kriging simplificado**: Variante simplificada del método geoestadístico de Kriging, que utiliza la matriz de correlación entre estaciones como ponderación espacial en lugar de realizar un modelado completo del variograma.
3. **RRRI (*Robust Random Regression Imputation*)**: Método basado en regresión robusta con estimador MM, que preserva la varianza de los datos originales mediante la adición de términos de error aleatorios (Kim y Park, 2021).
4. ***K-Nearest Neighbors* (KNN)**: Imputa valores faltantes usando el promedio de los k vecinos más similares en el espacio multivariado de características ($k = 5$).
5. **Interpolación lineal temporal**: Interpolación temporal simple entre observaciones válidas consecutivas, sin considerar información espacial.
6. **Media estacional**: Imputa valores faltantes utilizando la media mensual histórica correspondiente al mes del valor ausente.
7. ***Forward/Backward Fill***: Propagación hacia adelante del último valor válido, complementada con propagación hacia atrás para valores faltantes al inicio de la serie.
8. **Eliminación de filas (*Drop Nulls*)**: Método de referencia que simplemente elimina todas las filas con al menos un valor faltante. Representa el límite superior de desempeño posible con datos completos, pero al costo de pérdida significativa de información.

A.3. Resultados de la imputación

Análisis de resultados y selección del método Los resultados de la validación cruzada revelan que el método IDW alcanza el mejor balance entre calidad de imputación, preservación de datos y eficiencia computacional. La Tabla A.3 presenta las métricas de desempeño obtenidas para las estaciones de precipitación mediante la técnica de validación cruzada descrita anteriormente.

Como se observa en la Tabla A.3, el método IDW logra correlaciones superiores a 0.80 en las estaciones de Florida, La Cruz y San Gabriel, indicando una excelente capacidad para preservar los patrones espaciales de precipitación. Los

Tabla A.3: Resultados de validación cruzada del método IDW para estaciones de precipitación (5 iteraciones, 10 % de datos removidos en cada iteración).

Estación	MAE [mm]	RMSE [mm]	Correlación	Sesgo [mm]
Florida	1.88 ± 0.11	5.94 ± 0.31	0.82 ± 0.03	0.40 ± 0.17
La Cruz	2.05 ± 0.28	6.57 ± 0.84	0.88 ± 0.02	-0.50 ± 0.14
San Gabriel	2.05 ± 0.09	6.10 ± 0.44	0.83 ± 0.02	0.16 ± 0.24
Sarandí Grande	2.86 ± 0.18	8.91 ± 0.90	0.66 ± 0.04	0.24 ± 0.28
Cerro Colorado	2.81 ± 0.19	8.55 ± 0.88	0.67 ± 0.08	-0.04 ± 0.31

errores de imputación resultan bajos en términos absolutos, con errores absolutos medios (MAE) entre 1.9 y 2.9 mm y errores cuadráticos medios (RMSE) entre 5.9 y 8.9 mm. Estos valores son reducidos considerando que eventos de precipitación frecuentemente superan los 20–30 mm diarios. Los sesgos observados son en general pequeños (menores a 0.5 mm en magnitud), esto significa que el método es preciso y no tiende a dar resultados más altos o más bajos de lo que realmente son.

Las estaciones de Sarandí Grande y Cerro Colorado muestran correlaciones ligeramente menores (alrededor de 0.66–0.67) y errores mayores. Esto puede deberse a su ubicación geográfica relativa o a características locales de precipitación que presentan mayor variabilidad.

En cuanto a la preservación de datos, el método IDW mantiene el **99.99 %** de los registros originales, completando valores faltantes sin descartar información histórica. En contraste, la alternativa de **eliminación directa de filas con valores faltantes** descartaría el 43.26 % del conjunto de datos (reduciendo de 15.519 a 8.806 registros). Si bien trabajar únicamente con datos completos evita cualquier error de imputación, la pérdida masiva de información compromete la capacidad del modelo para capturar la variabilidad hidrológica a largo plazo y eventos extremos que podrían estar presentes en los períodos eliminados.

Justificación de la elección final Con base en los resultados presentados, se seleccionó el método IDW como estrategia de imputación para las series de precipitación por las siguientes razones:

- **Excelente calidad de imputación:** Las métricas de la Tabla A.3 demuestran que el método preserva adecuadamente los patrones espaciotemporales de precipitación, con correlaciones mayoritariamente superiores a 0.80 y errores relativamente bajos.
- **Preservación prácticamente completa del conjunto de datos:** Con 99.99 % de preservación, se maximiza la información disponible para entrenamiento y validación de modelos, en contraste con la pérdida del 43 % que implicaría eliminar filas con valores faltantes.
- **Eficiencia computacional:** El tiempo de procesamiento reducido (0.27 segundos) permite su aplicación eficiente a conjuntos de datos grandes.

- **Fundamentación física sólida:** El método IDW se basa en el principio de que las estaciones más cercanas geográficamente presentan patrones pluviométricos más similares, coherente con los procesos meteorológicos regionales.
- **Validación independiente del modelo final:** La selección se realizó mediante validación cruzada de imputación, evaluando la capacidad del método para recuperar valores conocidos sin utilizar el conjunto de testing del modelo predictivo. Esto evita sesgos de selección y garantiza que la elección es independiente del desempeño del modelo final.
- **Amplia aceptación en hidrología:** IDW es un método estándar para la imputación espacial de precipitaciones, ampliamente utilizado y validado en estudios hidrológicos ([Chow y cols., 1988](#)).

Anexo B

Optimización detallada de modelos

El presente anexo documenta en detalle el proceso iterativo de construcción, evaluación y optimización de los modelos desarrollados en esta tesis. Mientras que el Capítulo 6 presenta una síntesis de los resultados finales y los principales hallazgos, aquí se incluyen:

- Experimentos intermedios de incorporación y selección de variables.
- Evaluaciones comparativas de configuraciones alternativas.
- Análisis de correlación y reducción de dimensionalidad.
- Interpretabilidad mediante *Integrated Gradients*.
- Búsqueda y selección de hiperparámetros con *Optuna*.

El objetivo es garantizar trazabilidad metodológica y transparencia en las decisiones adoptadas, sin sobrecargar el cuerpo principal del documento.

Cada modelo se presenta en una subsección independiente, manteniendo una estructura homogénea que facilita la comparación entre enfoques.

B.1. MLP

Esta sección presenta el desarrollo detallado del modelo *Multilayer Perceptron* (MLP), incluyendo todos los experimentos realizados durante el proceso de optimización.

El cuerpo principal (Sección 6.4) resume únicamente los resultados del modelo base y del modelo final optimizado. Aquí se documentan las etapas intermedias que condujeron a la configuración seleccionada.

B.1.1. Modelo base

El primer experimento consistió en entrenar el modelo MLP con la arquitectura fija descrita en la Sección 5.6.1, utilizando todas las variables iniciales sin selección ni optimización de hiperparámetros.

Tabla B.1: Desempeño del modelo base sobre conjunto de validación.

NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
0.06	14.91	32.84	103.61

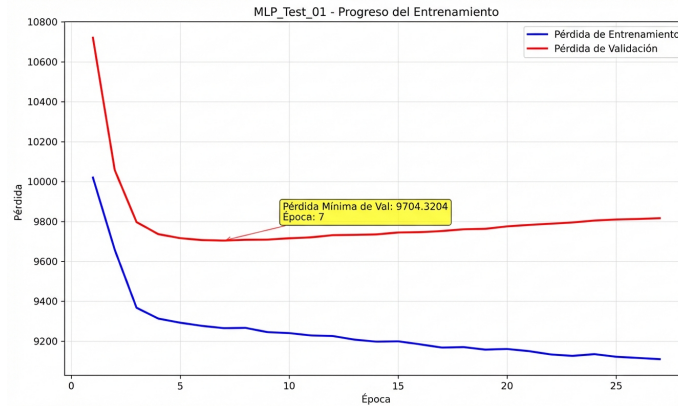


Figura B.1: Evolución de la función de pérdida (MSE) en entrenamiento y validación para el modelo base.

La Figura B.1 muestra la evolución de la pérdida durante el entrenamiento del modelo base. Se observa que la función de pérdida en entrenamiento desciende de manera continua hasta estabilizarse, mientras que la pérdida en validación presenta una reducción inicial pero luego tiende a incrementarse levemente a partir de la época 7, punto en el cual se alcanza la mínima pérdida de validación (*early stopping*).

Este comportamiento indica la aparición temprana de un leve *overfitting*, dado que el modelo continúa ajustándose a los datos de entrenamiento sin lograr una mejora equivalente en el conjunto de validación. En consecuencia, la capacidad de generalización del modelo se ve limitada, lo que sugiere que la arquitectura base, en su configuración actual, no logra representar de manera adecuada las relaciones subyacentes entre las variables predictoras y el caudal.

Los resultados cuantitativos (Tabla B.1) evidencian un desempeño general limitado. El valor de $NSE = 0.06$ indica que el modelo apenas logra explicar una fracción mínima de la variabilidad observada en los caudales, mientras que el $PBIAS$ positivo (14.91 %) sugiere una tendencia sistemática a la subestimación de los valores observados. Asimismo, el error cuadrático medio ($RMSE = 103.61$)

resulta elevado en relación con la magnitud típica de los caudales, lo que refuerza la conclusión de que el modelo base presenta una capacidad predictiva reducida.

En conjunto, estos resultados reflejan que la configuración inicial del MLP no captura adecuadamente las relaciones subyacentes entre las variables meteorológicas y el caudal.

B.1.2. Evaluación de la incorporación de variables

En esta etapa se evaluó el impacto de incorporar variables temporales representadas de forma cíclica mediante las transformaciones seno y coseno. Tal como se describió en la Sección 5.6.2, el objetivo fue determinar si la inclusión del día y/o del mes mejora el desempeño predictivo del modelo MLP. Para ello se realizó un análisis comparativo entre cuatro configuraciones: (i) sin variables temporales, (ii) incorporación únicamente del día, (iii) incorporación únicamente del mes, y (iv) incorporación conjunta del día y del mes.

Cada configuración se entrenó utilizando la misma arquitectura y los mismos hiperparámetros, de modo que las diferencias observadas reflejen únicamente el aporte de las variables temporales. El desempeño se evaluó sobre el conjunto de validación mediante las métricas MAE, RMSE, NSE y PBIAS. Los resultados obtenidos se presentan en la Tabla B.2.

Variables temporales

Tabla B.2: Comparación de desempeño para distintas combinaciones de variables temporales cíclicas sobre el conjunto de validación.

Configuración	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Sin variables	0.06	14.91	32.84	103.61
día	0.07	11.67	33.36	103.33
mes	0.07	11.89	33.23	103.25
día, mes	0.06	13.91	33.16	103.91

Los resultados muestran que las diferencias entre configuraciones son relativamente acotadas. En términos de error absoluto medio (MAE), la opción sin variables presenta el valor más bajo, aunque las configuraciones que incorporan exclusivamente el mes o la combinación día–mes presentan valores comparables. En contraste, la incorporación del día de forma aislada exhibe el mayor MAE, lo que sugiere que la variabilidad intramensual aporta información limitada al desempeño del modelo.

En cuanto al NSE, las configuraciones que incorporan variables temporales muestran valores levemente superiores respecto a la opción sin variables. La mayor mejora se obtiene al incluir únicamente el mes o el día (NSE = 0.07), mientras que la inclusión conjunta de día y mes no genera beneficios adicionales y presenta un valor levemente inferior (NSE = 0.06).

Respecto al sesgo, las configuraciones que incorporan variables temporales presentan una reducción moderada de la sobreestimación observada en la opción sin variables, reflejada en menores valores de PBIAS. Si bien el sesgo sigue siendo positivo y de magnitud considerable, esta disminución sugiere que las variables temporales aportan información adicional para ajustar la escala de las predicciones, aun cuando el impacto sobre las métricas de error (MAE y RMSE) resulta limitado.

Desde una perspectiva hidrológica, la ligera superioridad de la variable *mes* en MAE y RMSE puede interpretarse por su capacidad para capturar patrones estacionales característicos del régimen de caudales. Procesos como la acumulación de agua, los períodos de precipitaciones intensas o los ciclos de evapotranspiración operan predominantemente a escalas mensuales o estacionales, más que a nivel del día específico del mes. En este sentido, el mes constituye un descriptor temporal más estable y estrechamente vinculado a los mecanismos físicos que modulan el caudal, mientras que el día del mes rara vez se asocia a fenómenos hidrológicos sistemáticos. Esta consideración explica por qué la inclusión del mes resulta, en este caso, más informativa que la del día.

En síntesis, aunque las mejoras cuantitativas son moderadas, la incorporación del mes como variable temporal cíclica aporta señales estacionales relevantes, reduce el sesgo y presenta el mayor valor de NSE (0.07) entre las configuraciones evaluadas. En consecuencia, y considerando tanto la evidencia empírica como la coherencia hidrológica, se decidió conservar únicamente la variable *mes* para las etapas posteriores del modelado.

Representación espacial de la precipitación

Tabla B.3: Desempeño comparativo según esquema de agregación de precipitación.

Método de agregación	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Estaciones individuales	0.07	11.89	33.23	103.25
Promedio simple	0.07	25.21	30.62	103.24
Promedio ponderado por distancia	0.04	24.23	31.71	105.00
Polígonos de Thiessen	0.05	25.19	31.04	104.16

La tabla B.3 resume los resultados obtenidos al utilizar distintos esquemas de agregación de precipitación. Dado que las diferencias globales entre los métodos de agregación son reducidas en términos de *RMSE* y *NSE*, no se observa una ventaja clara de un esquema sobre otro en cuanto al desempeño global del

modelo. Si bien el método de promedio simple presenta el menor valor de *MAE*, esta mejora no se traduce en incrementos apreciables en la eficiencia del modelo ni en una reducción consistente del error cuadrático medio.

No obstante, el análisis del *PBIAS* revela diferencias relevantes entre los enfoques evaluados. El esquema basado en *estaciones individuales* presenta un sesgo considerablemente menor que los métodos de agregación espacial, mientras que las alternativas —promedio simple, promedio ponderado por distancia y polígonos de Thiessen— exhiben sobreestimaciones sistemáticas significativamente más elevadas.

La menor magnitud del *PBIAS* en el enfoque de estaciones individuales indica que este método reproduce de manera más equilibrada el volumen total de caudal, evitando el sesgo positivo introducido por los esquemas de agregación espacial. Aunque esta ventaja no se refleja en mejoras sustanciales en las métricas de error, sí representa un aspecto relevante desde el punto de vista de la coherencia hidrológica y la conservación de la señal física del proceso.

Por este motivo, se decidió **mantener el enfoque basado en estaciones individuales**, ya que ofrece un compromiso más adecuado entre desempeño global y control del sesgo, a la vez que preserva la variabilidad espacial original de la precipitación sin introducir procesos de agregación que no aportan beneficios consistentes en términos predictivos.

Inclusión de *lags*

Con el objetivo de evaluar el efecto de la memoria temporal en el desempeño del modelo, se realizaron distintos experimentos orientados a analizar la utilidad de incorporar valores retardados (*lags*) de la precipitación como variables explicativas.

Etapla 1: Evaluación preliminar de la inclusión de *lags* En primer lugar, se llevó a cabo una prueba preliminar para determinar si la incorporación de *lags* de precipitación mejora efectivamente la capacidad predictiva del modelo. Para este propósito, se fijó un horizonte arbitrario de **7 días de lag**, considerando que este rango es lo suficientemente amplio para capturar la respuesta retardada de la cuenca.

El análisis se realizó utilizando cuatro esquemas de agregación espacial de la precipitación: mantener las estaciones independientes, promedio simple, promedio ponderado por distancia y polígonos de Thiessen. La Tabla B.4 resume los resultados obtenidos en esta etapa.

Los resultados presentados en la Tabla B.4 muestran de manera clara el impacto positivo que tiene la incorporación de *lags* de precipitación sobre el desempeño del modelo. En comparación con el caso base (*sin lags*), la introducción de siete días de memoria temporal mejora sustancialmente todas las métricas, incrementando el *NSE* desde valores cercanos a 0.07 hasta un rango comprendido entre 0.54 y 0.66, y reduciendo de forma significativa tanto el *MAE* como el *RMSE*. Estos resultados confirman que la respuesta retardada de la cuenca contiene información esencial para la predicción del caudal.

Tabla B.4: Desempeño comparativo según esquema de agregación de precipitación.

Método de agregación	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Sin lags	0.07	11.89	33.23	103.25
Polígonos de Thiessen	0.66	10.05	17.28	62.52
Promedio simple	0.62	12.38	18.08	66.08
Estaciones individuales	0.61	12.15	18.67	66.56
Promedio ponderado por distancia	0.54	9.69	20.01	72.36

En relación con las diferencias entre los esquemas de agregación espacial, se observa que el método de **polígonos de Thiessen** alcanza el mejor desempeño global. Este enfoque presenta los menores errores ($MAE = 17.28$ y $RMSE = 62.52$ m³/s) y el mayor valor de NSE (0.66), lo que evidencia una mayor capacidad para reproducir la variabilidad observada del caudal. Asimismo, su $PBIAS$ (10.05 %) se mantiene dentro de un rango moderado, indicando un balance razonable en la estimación del volumen total.

Por el contrario, el esquema de *promedio ponderado por distancia* exhibe el peor desempeño de la comparación, con el $RMSE$ más elevado (72.36 m³/s) y el menor valor de NSE (0.54), lo que sugiere que la ponderación por distancia puede introducir una pérdida de información relevante en esta cuenca. El enfoque de *estaciones independientes* y el de *promedio simple* presentan resultados intermedios, con valores similares de error y eficiencia, aunque ambos mantienen un sesgo positivo algo mayor que el observado para los métodos ponderados.

En síntesis, aunque todos los esquemas muestran mejoras sustanciales al incorporar *lags* de precipitación, el método basado en **polígonos de Thiessen** se destaca por ofrecer la combinación más favorable de precisión, eficiencia y estabilidad. Por estas razones, este enfoque será adoptado para las siguientes etapas del análisis, en las cuales se determinará el número óptimo de días de *lag* a considerar en el modelo.

Etapas 2: Determinación del horizonte óptimo de lags Luego de comprobar el efecto positivo de la incorporación de *lags* y de identificar al esquema de agregación mediante **polígonos de Thiessen** como el más adecuado, se llevó a cabo un segundo experimento destinado a determinar el número óptimo de días de retardo que maximiza el desempeño del modelo.

En esta etapa, se mantuvo constante el esquema espacial de Thiessen y se variaron los horizontes temporales de *lag* desde 1 hasta 14 días. El objetivo fue analizar en qué medida la memoria temporal de la precipitación contribuye a

mejorar las predicciones del caudal y hasta qué punto la incorporación de valores demasiado antiguos introduce redundancia o pérdida de generalización.

La Tabla B.5 presenta los resultados obtenidos para cada configuración de horizonte temporal, utilizando los mismos indicadores de desempeño que en la etapa anterior.

Tabla B.5: Evaluación del horizonte de *lags* de precipitación (agregación mediante polígonos de Thiessen).

Horizonte de lag (días)	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
1 día	0.39	21.01	24.86	83.30
2 días	0.63	10.57	20.16	64.68
3 días	0.63	12.11	18.58	65.04
4 días	0.64	7.29	18.09	63.93
5 días	0.67	6.97	17.72	61.13
6 días	0.66	7.87	17.79	62.13
7 días	0.66	10.05	17.28	62.52
8 días	0.68	5.47	17.15	60.74
9 días	0.66	7.91	17.63	62.08
10 días	0.67	6.14	17.18	61.27
11 días	0.66	5.45	17.81	62.32
12 días	0.66	9.73	17.24	61.94
13 días	0.69	7.35	16.89	59.88
14 días	0.67	8.14	17.02	61.41

Los resultados de la Tabla B.5 muestran una mejora clara en el desempeño del modelo al incrementar el horizonte temporal de *lag* desde 1 hasta aproximadamente 4–5 días. La disminución progresiva de las métricas de error (*MAE* y *RMSE*), junto con el aumento del coeficiente de eficiencia de Nash–Sutcliffe (*NSE*), evidencia la relevancia de incorporar memoria temporal en las entradas del modelo, dado que la precipitación reciente condiciona directamente la respuesta del caudal.

A partir del quinto día, las mejoras comienzan a atenuarse y el desempeño entra en una región de rendimientos decrecientes. Los valores de *NSE* se estabilizan mayormente en el rango 0.66–0.68, mientras que las reducciones adicionales en *MAE* y *RMSE* resultan moderadas y no siempre consistentes entre horizontes consecutivos. Este comportamiento sugiere que el modelo alcanza una meseta en la información efectiva aportada por los *lags*, más allá de la cual el beneficio marginal de incorporar memoria adicional es limitado.

Dentro de este régimen, el horizonte de **8 días** se destaca por presentar un desempeño particularmente equilibrado. Este caso combina uno de los valores más bajos de *RMSE* (60.74 m³/s), un *NSE* elevado (0.68) y un *PBIAS* reducido (5.47 %), indicando una mejora simultánea en precisión, eficiencia y control del sesgo. Si bien el horizonte de 13 días alcanza el mayor valor de *NSE* (0.69), la

diferencia respecto a 8 días es marginal (del orden de 0.01) y se acompaña de un aumento en el *PBIAS* y de una mayor complejidad estructural del modelo.

Desde una perspectiva de parsimonia, la selección de horizontes más cortos resulta preferible. Cada día adicional de *lag* incrementa el número de variables de entrada, aumentando la complejidad del modelo, el costo computacional y el riesgo de sobreajuste. En este contexto, el horizonte de 8 días ofrece un compromiso más favorable al capturar la memoria temporal relevante del sistema sin introducir redundancia innecesaria en las variables explicativas.

En conjunto, los resultados indican que extender el horizonte de *lags* más allá de 8 días no produce mejoras consistentes que justifiquen el aumento en complejidad del modelo. Considerando simultáneamente el desempeño predictivo, el control del sesgo y el principio de parsimonia, se selecciona un horizonte de **ocho días de retardo** como la configuración más adecuada para las etapas posteriores del modelado.

B.1.3. Depuración inicial por correlación

La matriz descrita en la sección 6.2.2 presenta la correlación de Pearson entre las principales variables hidrometeorológicas del conjunto de datos. Se observa que las precipitaciones registradas en las distintas estaciones muestran correlaciones altas entre sí ($r > 0.75$), lo cual era esperable dada la proximidad espacial de los pluviómetros y la ocurrencia simultánea de eventos de lluvia a escala regional. Sin embargo, estas variables no fueron modificadas en esta etapa, ya que en la Sección B.1.2 se evaluaron distintos esquemas de agregación espacial (promedio simple, ponderado por distancia, polígonos de Thiessen) y se concluyó que mantener las precipitaciones por estación ofrecía el mejor desempeño.

Más allá de las precipitaciones, la matriz permite identificar varios grupos de variables con correlaciones fuertes, particularmente entre las variables meteorológicas medidas en la estación INIA Las Brujas. Los principales conjuntos correlacionados son los siguientes:

- **Temperatura del aire máxima, media y mínima:** muestran correlaciones entre 0.77 y 0.94, lo que sugiere una redundancia significativa entre estas mediciones.
- **Radiación solar (RS), heliofanía (Hel) y evapotranspiración de Penman (ET):** presentan correlaciones positivas elevadas ($r = 0.87$ entre RS y Hel, y $r = 0.90$ entre RS y ET), mientras que la correlación entre Hel y ET es moderada ($r = 0.67$). Estas tres variables están asociadas a la disponibilidad energética y al balance radiativo diario.

Otras variables, como la velocidad del viento (VV), mostraron correlaciones débiles o moderadas con el resto del conjunto, por lo que fueron retenidas sin modificaciones.

El análisis anterior sugiere que las temperaturas extremas (máxima y mínima) y las variables radiativas derivadas (heliofanía y evapotranspiración) podrían

aportar información redundante, dado su alto grado de correlación con otras variables del mismo grupo físico. No obstante, la simple correlación estadística no es suficiente para decidir su exclusión, ya que cada variable podría contribuir de forma diferente al desempeño global del modelo hidrológico.

Por esta razón, se planteó una evaluación empírica orientada a cuantificar el impacto de eliminar estas variables sobre las métricas de desempeño. En lugar de analizar por separado cada grupo, se optó por realizar una **evaluación combinada**, que permita capturar posibles interacciones entre las variables térmicas y radiativas, y determinar un subconjunto reducido que mantenga la capacidad predictiva del modelo.

Evaluación combinada de la reducción de variables correlacionadas

A partir de los valores de correlación observados, se diseñaron distintos escenarios de eliminación de variables para evaluar la sensibilidad del modelo a la reducción de dimensionalidad. Las configuraciones se definieron en función de las variables removidas respecto al modelo base, el cual incluye la totalidad de las variables disponibles. En todos los casos, el modelo se reentrenó completamente y se compararon las métricas de validación (MAE, RMSE, NSE y PBIAS) con respecto al modelo base.

Las configuraciones evaluadas incluyen:

- Modelo base (sin eliminación de variables).
- Eliminación de las temperaturas extremas ($T_{\text{máx}}$, $T_{\text{mín}}$).
- Eliminación conjunta o individual de las variables radiativas derivadas (Hel y ET).
- Combinaciones de eliminación entre variables térmicas y radiativas.

Los resultados presentados en la Tabla B.6 muestran que la reducción del conjunto de variables mediante la eliminación de aquellas fuertemente correlacionadas no conduce a mejoras significativas en el desempeño global del modelo, aunque sí introduce diferencias relevantes en términos de sesgo y complejidad estructural.

El modelo base, que conserva todas las variables originales, presenta el mayor valor de NSE (0.68) y el menor $RMSE$ ($60.74 \text{ m}^3/\text{s}$), confirmando que la inclusión completa de la información disponible permite capturar con mayor fidelidad la variabilidad del caudal. Sin embargo, algunas configuraciones reducidas alcanzan desempeños muy cercanos, con diferencias marginales en NSE y errores, lo que sugiere que parte de la información contenida en las variables correlacionadas es redundante desde el punto de vista predictivo.

Entre las configuraciones evaluadas, la eliminación de $T_{\text{máx}}$, $T_{\text{mín}}$ y ET (manteniendo temperatura media, radiación solar y heliofanía) se destaca por presentar el menor valor de $PBIAS$ (2.32 %), lo que indica una mejora sustancial en el control del sesgo respecto al modelo base. Esta reducción del sesgo se logra

Tabla B.6: Resultados de desempeño para configuraciones con variables removidas.

Variables re-movidas	NSE	PBIAS [%]	MAE [m³/s]	RMSE [m³/s]
Ninguna (modelo base)	0.68	5.47	17.15	60.74
$T_{\text{máx}}$, $T_{\text{mín}}$	0.66	6.56	17.16	62.23
Hel, ET	0.64	10.95	17.38	63.83
ET	0.64	8.81	17.66	63.85
Hel	0.64	7.63	17.96	64.09
$T_{\text{máx}}$, $T_{\text{mín}}$, Hel, ET	0.63	8.01	18.27	65.04
$T_{\text{máx}}$, $T_{\text{mín}}$, Hel	0.67	3.26	17.44	61.78
$T_{\text{máx}}$, $T_{\text{mín}}$, ET	0.67	2.32	17.47	61.27

a costa de una disminución muy moderada del desempeño global, con un *NSE* de 0.67 y un *RMSE* apenas superior al del modelo completo. Configuraciones alternativas con mayor reducción de variables presentan deterioros más marcados en las métricas de error y eficiencia, sin beneficios adicionales en términos de sesgo.

Desde una perspectiva de parsimonia, la eliminación de variables altamente correlacionadas resulta deseable siempre que no implique una pérdida sustancial de capacidad predictiva. En este sentido, la configuración con eliminación de $T_{\text{máx}}$, $T_{\text{mín}}$ y ET ofrece un equilibrio particularmente favorable, al reducir el número total de variables de entrada y mejorar el sesgo, manteniendo un nivel de desempeño comparable al del modelo base.

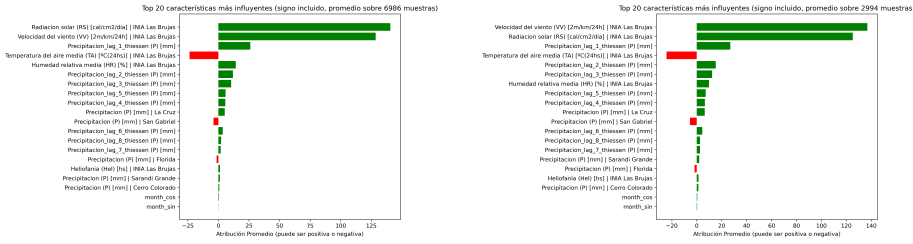
En conjunto, estos resultados indican que es posible simplificar el modelo sin comprometer de manera significativa su desempeño. Considerando simultáneamente la precisión, la eficiencia, el control del sesgo y el principio de parsimonia, se selecciona la configuración con eliminación de $T_{\text{máx}}$, $T_{\text{mín}}$ y ET como el conjunto de variables más adecuado para las etapas posteriores del modelado.

B.1.4. Análisis de relevancia mediante *Integrated Gradients*

Una vez definido el conjunto de variables depurado a partir del análisis de correlación, se evaluó la relevancia individual de cada característica mediante el método de *Integrated Gradients* (IG). Este enfoque permite cuantificar la contribución de cada variable a la predicción del modelo, asignando una atribución proporcional al cambio en la salida de la red al recorrer una trayectoria entre una entrada de referencia y la observación real. De esta manera, IG proporciona una medida interpretativa coherente con la estructura interna del modelo, sin requerir modificaciones en la arquitectura ni en el proceso de entrenamiento.

El cálculo se realizó por separado sobre los subconjuntos de entrenamiento y

validación, correspondientes respectivamente al 56 % (70 % del 80 %) y al 24 % (30 % del 80 %) de las observaciones totales. El conjunto de prueba (20 %) se reservó para la evaluación final, evitando *data leakage* durante la selección de variables.



(a) Subconjunto de entrenamiento.

(b) Subconjunto de validación.

Figura B.2: Importancia relativa de variables según *Integrated Gradients* (IG) en los conjuntos de entrenamiento y validación.

Los resultados obtenidos mediante IG mostraron una notable consistencia entre los conjuntos de entrenamiento (Figura B.2a) y validación (Figura B.2b), indicando que el modelo capturó patrones estables y no dependientes del subconjunto utilizado.

Un hallazgo central del análisis fue que varias variables meteorológicas medidas en la estación INIA Las Brujas —principalmente la **velocidad del viento (VV)**, la **radiación solar (RS)** y la **temperatura del aire media (TA)**— aparecen como las características más influyentes, con atribuciones promedio significativamente superiores al resto. Estas variables presentan efectos positivos o negativos consistentes, reflejando su impacto directo sobre la estimación del caudal.

Por otra parte, los **rezagos de precipitación** mostraron contribuciones relevantes y consistentemente positivas, aunque de menor magnitud que las variables meteorológicas principales. Se observa el patrón esperado de decaimiento: los rezagos 1 y 2 son los más influyentes, y la importancia disminuye progresivamente con rezagos más largos.

En contraste, las formas contemporáneas de **precipitación diaria** presentaron atribuciones promedio muy cercanas a cero, indicando un aporte marginal al modelo. Un comportamiento similar se observó en las variables temporales (*month_sin* y *month_cos*) y en la **humedad relativa media**, cuyas contribuciones resultaron mínimas.

Finalmente, variables como la **evapotranspiración Penman** y la **heliografía** mostraron atribuciones pequeñas pero estables, sugiriendo un rol secundario o indirecto en la dinámica del caudal.

Estos resultados motivaron la aplicación de un **procedimiento iterativo de depuración** guiado por IG, orientado a evaluar empíricamente si la eliminación de variables con atribuciones cercanas a cero podía mejorar la estabilidad y el desempeño del modelo sin pérdida de información relevante.

Procedimiento iterativo de depuración mediante IG. El proceso de depuración se estructuró en una serie de pasos repetitivos orientados a identificar y eliminar variables con influencia marginal sobre la predicción. Los pasos fueron los siguientes: (i) cálculo de las atribuciones de IG en los conjuntos de entrenamiento y validación; (ii) identificación de la variable con menor atribución promedio en validación; (iii) eliminación de dicha variable del conjunto de entrada y reentrenamiento completo del modelo; (iv) comparación de las métricas de desempeño (MAE, RMSE, NSE, PBIAS) respecto al modelo base; y (v) nuevo análisis de IG para verificar la estabilidad del patrón de relevancias. El ciclo se repitió hasta que no se observaron mejoras significativas o cambios estructurales en la distribución de atribuciones.

En la primera iteración, las variables con la atribución media más baja fueron *month_sin* y *month_cos*, ambas con contribuciones prácticamente nulas tanto en entrenamiento como en validación. Dado que estas dos características representan una codificación trigonométrica complementaria del mes del año, se decidió eliminarlas en conjunto para evitar una representación incompleta del ciclo anual. A partir de ello se evaluó una nueva versión del modelo *sin* estas dos variables, con el fin de determinar si su exclusión impactaba en la estabilidad o el desempeño general del sistema predictivo:

Tabla B.7: Comparación del desempeño del modelo al eliminar las variables codificadas de mes.

Configuración	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Modelo con variables de mes	0.67	2.32	17.47	61.27
Modelo sin variables de mes	0.69	9.66	17.07	59.94

En la primera iteración del procedimiento, las variables con la menor atribución media según IG fueron *month_sin* y *month_cos*, ambas con contribuciones prácticamente nulas tanto en el conjunto de entrenamiento como en el de validación. Dado que estas dos variables representan una codificación trigonométrica complementaria del ciclo anual, se decidió eliminarlas en conjunto para evaluar empíricamente el impacto de su exclusión sobre el desempeño del modelo.

La Tabla B.7 presenta la comparación entre el modelo con variables de mes y el modelo sin dichas características. Los resultados muestran que la eliminación de *month_sin* y *month_cos* produce mejoras moderadas en algunas métricas de ajuste, reflejadas en una disminución del *RMSE* (de 61.27 a 59.94 m³/s) y un incremento del coeficiente de eficiencia de Nash–Sutcliffe (de 0.67 a 0.69).

No obstante, esta mejora viene acompañada de un aumento significativo del sesgo porcentual. El *PBIAS* se incrementa de 2.32 % a 9.66 %, lo que indica una mayor desviación sistemática en la estimación del volumen total de caudal. Desde una perspectiva hidrológica, este incremento del sesgo resulta relevante,

ya que afecta directamente la capacidad del modelo para reproducir balances hídricos de forma adecuada, aun cuando las métricas de error puntual mejoren levemente.

Estos resultados sugieren que, si bien las variables temporales de mes presentan una contribución directa baja según IG, su inclusión cumple un rol regulador en la estabilidad estadística del modelo, particularmente en términos de control del sesgo. La información estacional que aportan parece actuar como un anclaje temporal que limita desviaciones sistemáticas en la magnitud de las predicciones.

En consecuencia, y priorizando la coherencia hidrológica y el control del sesgo por sobre mejoras marginales en métricas de ajuste, se decidió **mantener las variables *month_sin* y *month_cos*** en el conjunto final de entrada. El procedimiento iterativo de depuración se dio por finalizado en esta etapa, al no identificarse otras variables con atribuciones marginales cuya eliminación mejorara de forma consistente el desempeño global del modelo.

Conjunto final de variables

El conjunto final de variables fue definido a partir de un proceso integrado de análisis por correlación, evaluación empírica del desempeño y análisis de relevancia mediante *Integrated Gradients* (IG), priorizando la coherencia hidrológica, el control del sesgo y la parsimonia del modelo por sobre mejoras marginales en métricas de ajuste.

A lo largo del procedimiento se evaluó la eliminación progresiva de variables correlacionadas o con baja atribución promedio según IG. Si bien algunas exclusiones mostraron reducciones leves en métricas como *RMSE* o incrementos marginales del *NSE*, dichas mejoras se vieron acompañadas, en varios casos, por aumentos significativos del sesgo porcentual (*PBIAS*) o por pérdidas de estabilidad del modelo. En consecuencia, la selección final se basó en un criterio de equilibrio entre precisión, estabilidad estadística y representación adecuada de los procesos físicos subyacentes.

Las variables incluidas en el conjunto final son las siguientes:

- **Precipitación:** se conservaron las precipitaciones contemporáneas de las estaciones, junto con los **rezagos de 1 a 8 días** agregados mediante polígonos de Thiessen. Este conjunto captura de manera adecuada la memoria hidrológica del sistema, reflejando la respuesta retardada del caudal ante la precipitación acumulada, tal como fue evidenciado por el análisis de desempeño y las atribuciones obtenidas con IG.
- **Radiación y energía:** se conservaron la **radiación solar (RS)** y la **heliofanía (Hel)**, ambas asociadas al balance energético en superficie. La **evapotranspiración de Penman (ET)** fue eliminada del conjunto final debido a su alta correlación con estas variables y a que su exclusión no produjo un deterioro apreciable en las métricas de desempeño. En particular, las configuraciones que prescinden de ET mantuvieron valores de *NSE*,

RMSE y *MAE* comparables al modelo base, a la vez que permitieron reducir la dimensionalidad del modelo. De este modo, se evitó la inclusión de información redundante, favoreciendo una representación energética más parsimoniosa sin comprometer la capacidad predictiva.

- **Temperatura del aire:** se conservó únicamente la **temperatura del aire media**, dado que la eliminación de las temperaturas extrema máxima y mínima no implicó pérdidas relevantes de información y permitió una reducción del número total de variables sin deteriorar significativamente el desempeño del modelo.
- **Humedad relativa (HR) y velocidad del viento (VV):** ambas variables presentaron atribuciones consistentes según IG y baja colinealidad con el resto de los predictores, por lo que se mantuvieron en el conjunto final.
- **Variables temporales:** se conservaron las codificaciones trigonométricas *month_sin* y *month_cos*. Si bien estas variables presentan atribuciones directas bajas según IG, su eliminación generó un incremento significativo del sesgo (*PBIAS*), indicando que capturan información estacional relevante para el control de desviaciones sistemáticas en la estimación del caudal.

En síntesis, el conjunto final de variables representa una versión depurada y coherente del conjunto original, que preserva la información hidrometeorológica esencial para la predicción del caudal, evita redundancias innecesarias y mantiene un adecuado control del sesgo. Esta configuración ofrece un equilibrio robusto entre capacidad predictiva, estabilidad estadística y parsimonia estructural, y fue la utilizada en las etapas posteriores de optimización y validación del modelo MLP.

B.1.5. Optimización de hiperparámetros (Optuna)

Para determinar la configuración más adecuada de la arquitectura MLP se utilizó el *framework* de optimización bayesiana *Optuna*. El espacio de búsqueda considerado, detallado en la Sección 5.6.4, incluyó hiperparámetros estructurales (número de capas ocultas y cantidad de neuronas por capa), así como parámetros asociados al entrenamiento (tasa de aprendizaje, *weight decay*, *dropout* y *batch size*).

El proceso de optimización constó de un total de 175 *trials*, de los cuales 89 finalizaron correctamente y 86 fueron podados de forma anticipada mediante el algoritmo de *median pruning*. Esta estrategia permitió descartar configuraciones poco prometedoras en etapas tempranas del entrenamiento, reduciendo significativamente el costo computacional del proceso.

Durante las primeras iteraciones se observó una elevada variabilidad en el valor de la función objetivo, con algunos *trials* que presentaron pérdidas de validación muy altas, lo que evidencia una fase inicial de exploración amplia del

espacio de búsqueda. A medida que avanzó la optimización, Optuna concentró el muestreo en regiones más prometedoras, logrando una disminución progresiva del mejor valor alcanzado. La Figura B.3 muestra la evolución del valor objetivo para cada *trial*, junto con la trayectoria del mejor resultado acumulado.

El uso conjunto de *early stopping* y *pruning* permitió acelerar la convergencia del proceso. En el *trial* óptimo, la detención temprana se produjo una vez alcanzado un mínimo estable de la pérdida de validación, con un valor $val_loss = 2753.44$, que se mantuvo como el mejor resultado registrado durante toda la optimización.

La configuración seleccionada corresponde al *trial* número 74 y se resume en la Tabla B.8. A diferencia de configuraciones más complejas evaluadas durante la búsqueda, el modelo óptimo presenta una arquitectura relativamente simple, compuesta por **dos capas ocultas**. Esta solución sugiere que, dado el conjunto de variables seleccionado previamente, una profundidad moderada resulta suficiente para capturar las relaciones no lineales entre los predictores y el caudal.

El uso de la función de activación `gelu`, junto con una tasa de aprendizaje del orden de 10^{-3} , un *batch size* intermedio y una regularización suave mediante *weight decay* y *dropout*, favoreció un entrenamiento estable y una buena capacidad de generalización, evitando configuraciones excesivamente complejas que no aportaron mejoras sustanciales en el desempeño.

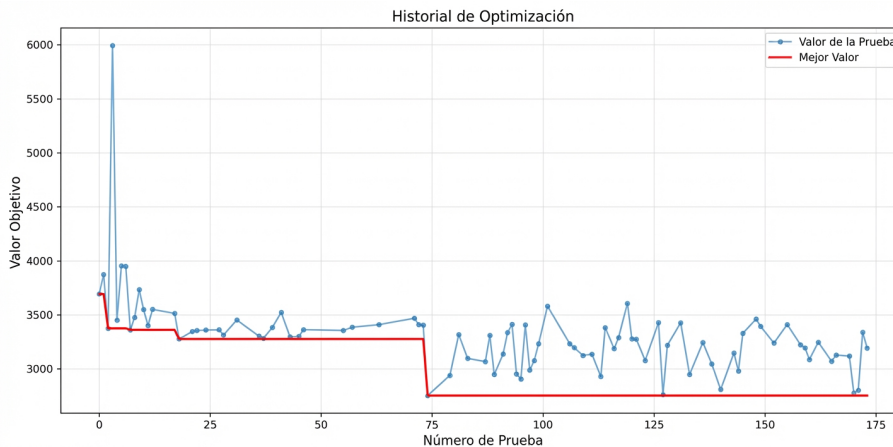


Figura B.3: Historial del proceso de optimización realizado con Optuna. La línea azul muestra el valor de la función objetivo para cada *trial*, mientras que la línea roja indica el mejor valor obtenido hasta ese punto.

En síntesis, la optimización mediante Optuna permitió identificar una configuración parsimoniosa y robusta del modelo MLP, que logra un buen compromiso entre capacidad representacional, estabilidad del entrenamiento y desempeño en validación. Estos resultados refuerzan la importancia de una búsqueda sistemática de hiperparámetros, ya que arquitecturas más profundas o complejas

Tabla B.8: Configuración óptima seleccionada por Optuna.

Hiperparámetro	Valor óptimo
Número de capas ocultas	2
Neuronas por capa	[56, 391]
Función de activación	<code>gelu</code>
Tasa de aprendizaje	0.001315
<i>Dropout</i> base	0.204
<i>Batch size</i>	32
<i>Weight decay</i>	4.09×10^{-5}

no garantizaron mejoras adicionales en el ajuste del modelo.

B.1.6. Evaluación extendida del modelo optimizado

El modelo final se entrenó utilizando el conjunto de datos integrado descrito en capítulos anteriores, construido a partir de la combinación de las series hidrométricas de DINAGUA y las series meteorológicas de INUMET e INIA para la cuenca del río Santa Lucía. A partir de los conjuntos separados de datos de cada institución, se unificaron las bases mediante la fecha como clave común, conservando únicamente:

- El caudal en Florida (Puente Ruta 5), utilizado como variable objetivo.
- Las variables meteorológicas de las estaciones Florida, La Cruz, San Gabriel, Sarandí Grande, Cerro Colorado e INIA Las Brujas: precipitación, evapotranspiración de Penman, heliofanía, humedad relativa media, radiación solar, temperatura media, máxima y mínima del aire, y velocidad del viento.

Los datos faltantes en las series meteorológicas se imputaron mediante interpolación por distancia inversa (IDW), mientras que los valores faltantes de caudal no se imputaron, conservándose explícitamente como faltantes para ser excluidos del cálculo de métricas.

Para capturar la memoria hidrológica de la cuenca, se construyeron *lags* de precipitación de 1 a 8 días a partir de la precipitación areal diaria ponderada mediante polígonos de Thiessen. Asimismo, se incorporaron variables temporales mediante codificación cíclica del mes (*month_sin*, *month_cos*).

La arquitectura final, optimizada mediante Optuna, corresponde a una MLP con dos capas ocultas de 56 y 391 neuronas, activación `gelu`, tasa de aprendizaje 1.32×10^{-3} , *weight decay* de 4.1×10^{-5} , *dropout* base de 0.20 y *batch size* de 32. El entrenamiento se realizó durante un máximo de 200 épocas con *early stopping* basado en la pérdida de validación.

Desempeño cuantitativo

La Tabla B.9 presenta el desempeño del modelo en entrenamiento, validación y prueba.

Tabla B.9: Desempeño del modelo MLP optimizado en los distintos conjuntos.

Conjunto	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Entrenamiento	0.87	10.80	11.39	35.16
Validación	0.74	10.65	16.01	55.00
Prueba	0.67	6.52	15.96	73.57

En el conjunto de entrenamiento el modelo alcanza una eficiencia elevada ($NSE = 0.87$), lo que indica una alta capacidad de ajuste a los datos históricos. Sin embargo, la disminución del NSE en validación (0.74) y prueba (0.67) evidencia un comportamiento esperable de pérdida de desempeño al generalizar temporalmente.

En el conjunto de prueba, el valor $NSE = 0.67$ se clasifica como *bueno* según los criterios de (Moriasi y cols., 2015). El MAE de 15.96 m³/s indica una adecuada precisión en condiciones medias, mientras que el incremento del RMSE hasta 73.57 m³/s refleja la influencia de errores asociados a eventos extremos de crecida.

El PBIAS en prueba (6.52%) indica una leve subestimación sistemática, dentro del rango *muy bueno* ($|\text{PBIAS}| < 10\%$). Resulta destacable que el sesgo se reduce respecto a entrenamiento y validación, lo que sugiere que la generalización temporal no introduce desbalances volumétricos significativos.

Evolución del entrenamiento

La Figura B.4 muestra la evolución de la pérdida de entrenamiento y validación. Se observa una rápida disminución inicial seguida de una estabilización progresiva. La mínima pérdida de validación se alcanza en la época 81, momento en el cual se activa el criterio de *early stopping*.

La separación creciente entre las curvas de entrenamiento y validación hacia el final del proceso indica la aparición de cierto sobreajuste leve, aunque contenido, coherente con la diferencia moderada entre NSE de entrenamiento y validación.

Análisis gráfico del ajuste

La Figura B.5 muestra la comparación temporal en el conjunto completo. El modelo reproduce adecuadamente la estacionalidad y la ocurrencia de la mayoría de los pulsos de caudal.

Sin embargo, durante eventos extremos se observa una subestimación de los picos máximos, particularmente en el evento de mayor magnitud del período analizado, lo que explica el aumento del RMSE respecto al MAE.

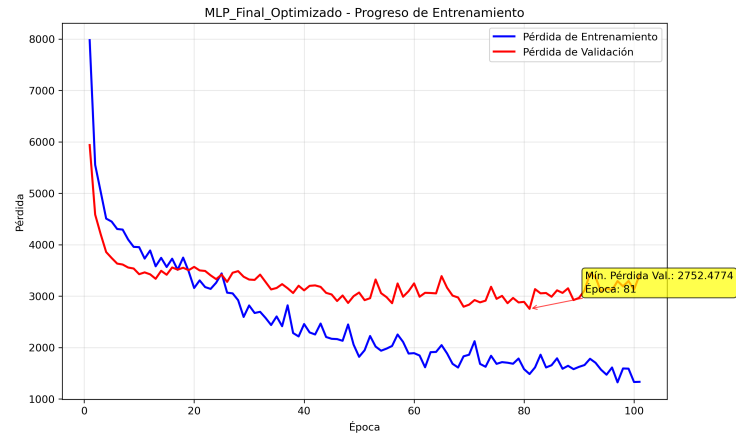


Figura B.4: Evolución de la pérdida de entrenamiento y validación del modelo MLP optimizado.

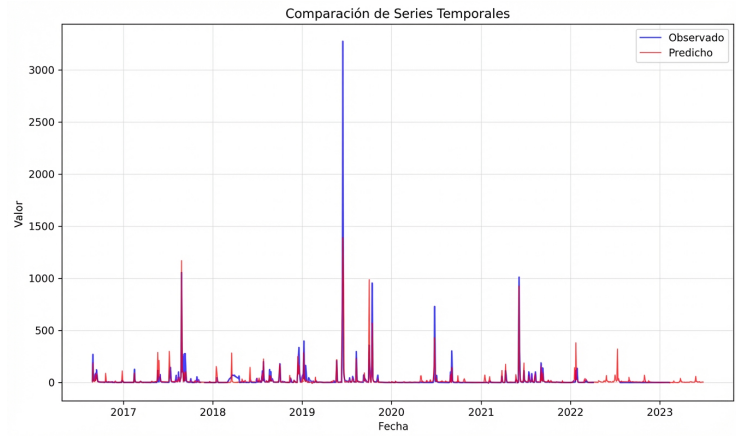


Figura B.5: Serie temporal de caudales observados y predichos por el modelo MLP.

El *scatterplot* en escala logarítmica (Figura B.6) muestra una adecuada alineación con la línea 1:1 en rangos medios de caudal. Para valores elevados se observa dispersión creciente y predominio de puntos por debajo de la diagonal, confirmando la tendencia a subestimar crecidas extremas.

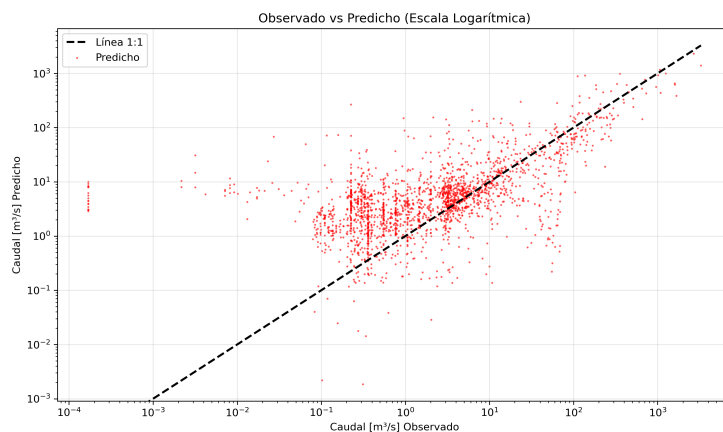


Figura B.6: Diagrama observado vs predicho en escala logarítmica.

El análisis de residuos (Figura B.7) muestra una concentración alrededor de cero sin patrón temporal sistemático evidente. Los mayores residuos negativos coinciden con eventos extremos, mientras que el resto de la serie presenta errores acotados.

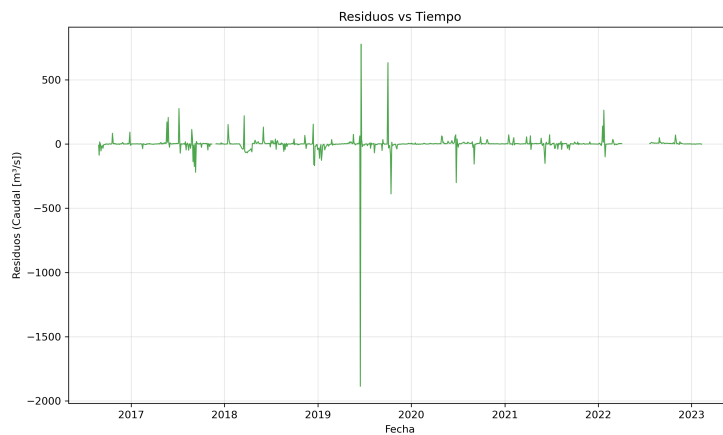


Figura B.7: Residuos del modelo MLP en función del tiempo.

Análisis de relevancia mediante *Integrated Gradients*

El análisis de atribución sobre el conjunto de prueba mostrado en la figura B.8 indica que las mayores contribuciones positivas corresponden a los *lags* de precipitación ponderados por Thiessen (especialmente los *lags* 1 y 2 días), lo que confirma la dominancia de la señal pluviométrica reciente en la predicción del caudal.

La radiación solar también presenta contribución positiva significativa, mientras que variables como heliofanía y algunos *lags* más largos muestran contribuciones negativas. Las variables cíclicas de mes presentan una influencia moderada, contribuyendo principalmente a la estabilización del sesgo global.

Este patrón es coherente con la interpretación hidrológica: el modelo basa su capacidad predictiva en la memoria corta de la precipitación y utiliza las variables energéticas como moduladores secundarios.

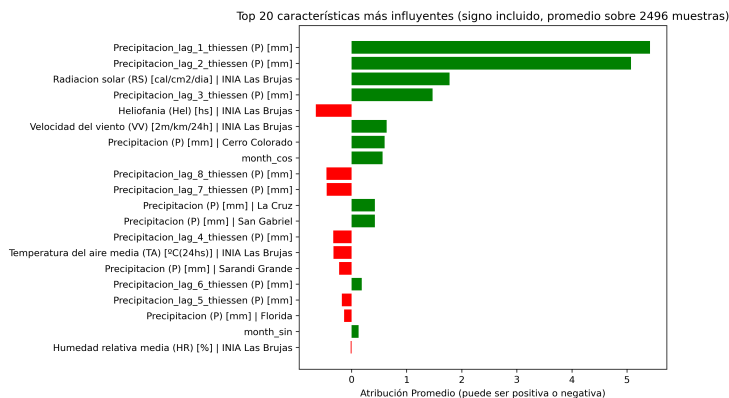


Figura B.8: Top 20 características más influyentes según *Integrated Gradients* en el conjunto de prueba.

Costo computacional

El entrenamiento completo del modelo MLP optimizado demandó 45.49 segundos en un total de 101 épocas en el entorno de cómputo utilizado. El tiempo de inferencia sobre los tres conjuntos fue despreciable a la precisión reportada por el sistema.

B.2. RNN

Esta sección presenta el desarrollo detallado del modelo RNN, incluyendo todos los experimentos realizados durante el proceso de optimización.

El cuerpo principal (Sección 6.5) resume únicamente los resultados del modelo base y del modelo final optimizado. Aquí se documentan las etapas intermedias que condujeron a la configuración seleccionada.

B.2.1. Modelo base

El primer experimento consistió en entrenar la arquitectura fija definida en 5.7.1 con el conjunto de variables inicial (sin selección ni optimización de hiperparámetros).

Tabla B.10: Desempeño del modelo RNN base sobre conjunto de validación.

NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
0.04	14.17	33.35	104.85

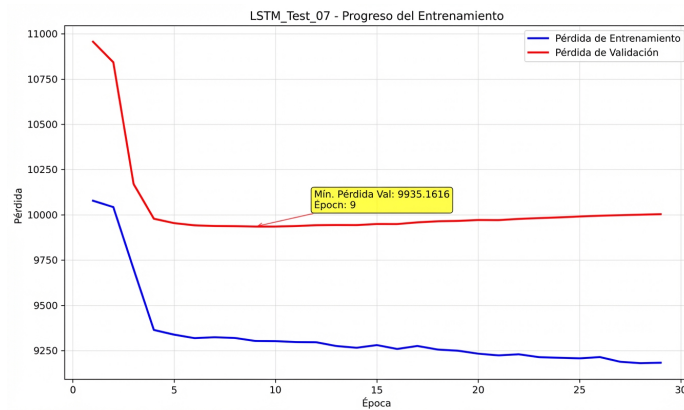


Figura B.9: Evolución de la pérdida (MSE) en entrenamiento y validación para el modelo RNN base.

En la Figura B.9 se observa la evolución de la pérdida durante el entrenamiento del modelo base. La **pérdida de entrenamiento** (línea azul) desciende de forma continua a lo largo de las épocas, mientras que la **pérdida de validación** (línea roja) disminuye rápidamente en las primeras iteraciones y luego se estabiliza a partir de la época 4, alcanzando su valor mínimo en la época 9 ($val_loss = 9935.16$). Este comportamiento sugiere la aparición de un leve *overfitting*, dado que el modelo continúa mejorando su desempeño sobre el conjunto de entrenamiento sin lograr una reducción adicional en la pérdida de validación.

Los resultados del modelo base (Tabla B.10) muestran un desempeño limitado, con $RMSE = 104.85 \text{ m}^3/\text{s}$, $MAE = 33.35 \text{ m}^3/\text{s}$ y $NSE = 0.04$, lo que evidencia una capacidad predictiva reducida. El sesgo global es moderadamente alto ($PBIAS = 14.17\%$), indicando una tendencia sistemática del modelo a subestimar los caudales observados. En conjunto, estos resultados reflejan que la configuración inicial del modelo recurrente no logra capturar de forma adecuada las dependencias temporales del sistema hidrológico.

B.2.2. Evaluación de la incorporación de variables

El análisis se planteó de forma **iterativa** y acumulativa (ver Sección 5.7.2), conservando las configuraciones que mostraron mejoras o estabilidad.

Variables temporales

Los resultados se presentan en la Tabla B.11.

Tabla B.11: Comparación de desempeño para distintas combinaciones de variables temporales cíclicas sobre el conjunto de validación.

Configuración	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Sin variables	0.04	14.17	33.35	104.85
día	0.04	13.85	33.29	104.64
mes	0.04	12.94	33.41	104.57
día, mes	0.04	12.65	33.51	104.75

De forma similar a lo observado en el modelo *MLP* (ver Sección B.1.2), los resultados del RNN indican que la incorporación de variables temporales produce únicamente **mejoras marginales** respecto a la configuración base. Entre las alternativas evaluadas, la inclusión individual del *mes* presenta el desempeño más estable, con una ligera reducción del *RMSE* y una mejora más consistente del *PBIAS*, mientras que la variable *día* aporta beneficios limitados y la combinación conjunta no introduce mejoras adicionales en términos de error o eficiencia.

Al igual que en el análisis previo, estas diferencias sugieren que el modelo recurrente presenta una **sensibilidad acotada** frente a la desagregación temporal diaria, pero responde de manera algo más estable a patrones de variabilidad estacional. Desde una perspectiva hidrológica, este comportamiento es consistente con la interpretación desarrollada para el MLP: el *mes* captura los ciclos estacionales que gobiernan la dinámica de caudales, mientras que el *día* raramente se asocia a procesos hidrológicos sistemáticos. En consecuencia, la señal mensual constituye un descriptor temporal más robusto y relevante que la variabilidad intramensual.

En síntesis, aunque las diferencias cuantitativas son moderadas, la incorporación del *mes* ofrece la mayor estabilidad entre las configuraciones evaluadas y contribuye a una reducción parcial del sesgo del modelo. Considerando la evidencia empírica tanto del *MLP* como del *RNN*, así como su coherencia con los mecanismos hidrológicos subyacentes, se decidió **conservar únicamente la variable temporal *mes*** para las etapas posteriores del modelado.

Representación espacial de la precipitación

Los resultados de la Tabla B.12 muestran que los distintos métodos de agregación de la precipitación presentan diferencias moderadas en las métricas de

Tabla B.12: Desempeño comparativo según esquema de agregación de precipitación.

Método de agregación	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Estaciones individuales	0.04	12.94	33.41	104.57
Promedio simple	0.07	23.51	30.99	103.25
Promedio ponderado por distancia	0.04	24.84	31.36	104.82
Polígonos de Thiessen	0.05	23.63	31.30	104.04

desempeño. Si bien los esquemas basados en agregación espacial (*promedio simple*, *promedio ponderado por distancia* y *polígonos de Thiessen*) logran reducciones en los errores (*MAE* y *RMSE*) y ligeras mejoras en la eficiencia (*NSE*) respecto al uso de estaciones individuales, estas mejoras no resultan sustanciales desde el punto de vista predictivo.

Sin embargo, al analizar el *PBIAS*, se observa una diferencia marcada entre los enfoques. El esquema basado en **estaciones individuales** presenta un sesgo significativamente menor (12.94%), mientras que todos los métodos de agregación espacial exhiben valores superiores al 23%, indicando una tendencia mucho más pronunciada a la subestimación de los caudales. En este contexto, el *PBIAS* se convierte en un criterio clave para la evaluación comparativa de los esquemas de representación de la precipitación.

Por este motivo, se decidió **mantener el enfoque basado en estaciones individuales**. Si bien los métodos de agregación espacial permiten reducir ligeramente los errores medios, lo hacen a costa de introducir un sesgo sistemático considerablemente mayor. El uso de estaciones individuales no solo minimiza este sesgo, sino que además preserva la información espacial original de cada estación, lo cual resulta consistente con la naturaleza altamente heterogénea de la precipitación. En conjunto, el menor *PBIAS* y la conservación de la resolución espacial justifican la elección de este esquema por sobre las alternativas basadas en agregación.

Análisis de la ventana temporal de entrada

Los resultados de la Tabla B.13 muestran que la longitud de la ventana temporal de entrada constituye un factor determinante en el desempeño del modelo RNN. La comparación con la configuración base, correspondiente a una ventana de un día, evidencia una mejora sustancial al incorporar secuencias temporales más largas, con reducciones marcadas tanto en el *MAE* como en el *RMSE*, así como incrementos significativos en la eficiencia del modelo (*NSE*).

Tabla B.13: Desempeño del modelo RNN para distintas longitudes de ventana temporal.

Ventana (días)	NSE	PBIAS [%]	MAE [m^3/s]	RMSE [m^3/s]
1	0.04	12.94	33.41	104.57
2	0.45	13.95	24.36	79.27
3	0.58	25.88	19.40	69.58
4	0.55	25.14	19.25	71.89
5	0.61	24.64	17.53	67.16
6	0.60	26.58	18.07	67.30
7	0.58	21.18	18.26	69.36
8	0.58	19.89	18.13	69.27
9	0.61	17.15	17.78	66.94
10	0.60	16.83	17.77	67.34
11	0.60	17.72	17.90	67.90
12	0.59	15.96	17.94	68.49
13	0.58	16.35	18.15	69.19
14	0.58	19.53	18.06	68.98

En particular, el desempeño mejora de manera pronunciada al incrementar la ventana temporal desde 1 hasta aproximadamente 5 días, alcanzando valores de NSE cercanos a 0.60 y reducciones de error superiores al 35 % respecto a la configuración inicial. A partir de este punto, el beneficio marginal de aumentar la longitud de la ventana se vuelve más limitado, observándose un comportamiento relativamente estable de las métricas para ventanas comprendidas entre 5 y 12 días.

Al analizar el comportamiento conjunto de las métricas, se observa que las ventanas de 9 y 10 días alcanzan los valores más altos de eficiencia, con $NSE = 0.61$, junto con errores moderados y valores de $PBIAS$ sensiblemente inferiores a los observados para ventanas más cortas. En particular, la ventana de 9 días presenta el menor $RMSE$ ($66.94 \text{ m}^3/\text{s}$) dentro de este grupo y un $PBIAS$ relativamente contenido (17.15%), lo que indica un adecuado equilibrio entre precisión, eficiencia y control del sesgo.

En base a estos resultados, se seleccionó una **ventana temporal de 9 días** como configuración óptima para las etapas posteriores del modelado. Esta elección se fundamenta en que dicha ventana alcanza uno de los mayores valores de eficiencia del modelo, comparable al máximo observado, sin introducir incrementos significativos en el sesgo, y ofrece un compromiso favorable entre desempeño predictivo y estabilidad general del modelo.

Este resultado pone de manifiesto que, aun en arquitecturas recurrentes con mecanismos de memoria interna, la definición de la longitud de la secuencia de entrada constituye un hiperparámetro clave. En este caso, una ventana temporal moderadamente extensa permite capturar dependencias relevantes del sistema hidrológico sin degradar el balance global del modelo, evidenciando que la ade-

cuada selección del horizonte temporal resulta determinante para optimizar el desempeño de la RNN.

B.2.3. Depuración inicial por correlación

Con el fin de reducir posibles redundancias en el conjunto de variables de entrada y evaluar su impacto sobre el desempeño del modelo recurrente, se aplicó un procedimiento de depuración inicial basado en el análisis de correlación, siguiendo el mismo enfoque utilizado previamente para el modelo MLP (Sección B.1.3). Para cada configuración reducida se reentrenó completamente la arquitectura definida en la Sección 5.7.3, manteniendo constantes el resto de los hiperparámetros y la ventana temporal seleccionada, y se compararon las métricas de validación (MAE, RMSE, NSE y PBIAS) respecto del modelo base que utiliza todas las variables disponibles.

La Tabla B.14 resume los resultados obtenidos para las diferentes configuraciones evaluadas, incluyendo variantes con eliminación de variables de temperatura, exclusión de variables radiativas derivadas y combinaciones reducidas de ambos grupos.

Tabla B.14: Resultados de desempeño para configuraciones con variables removidas (modelo RNN).

Variables re-movidas	NSE	PBIAS [%]	MAE [m³/s]	RMSE [m³/s]
Ninguna (modelo base)	0.61	17.15	17.78	66.94
T _{máx} , T _{mín} , ET	0.59	22.37	18.25	68.70
T _{máx} , T _{mín} , Hel, ET	0.58	17.04	18.73	69.53
Hel	0.57	14.40	18.66	69.79
T _{máx} , T _{mín} , Hel	0.57	22.80	18.57	69.96
T _{máx} , T _{mín}	0.57	24.44	18.39	70.42
ET	0.56	9.48	19.28	70.74
Hel, ET	0.52	22.63	19.57	74.42

Los resultados muestran un patrón claro y consistente: todas las configuraciones reducidas presentan un desempeño inferior al modelo base. La configuración completa alcanza el mejor ajuste global, con un valor de NSE = 0.61, el menor error cuadrático medio (RMSE = 66.94) y uno de los menores errores absolutos (MAE = 17.78). Asimismo, exhibe un sesgo moderado (PBIAS = 17.15 %), que se mantiene dentro de un rango comparable o inferior al de la mayoría de las configuraciones alternativas.

La reducción de variables, ya sea mediante la eliminación de componentes térmicas o de variables radiativas derivadas, genera un deterioro sistemático del desempeño del modelo. En particular, los modelos simplificados presentan

incrementos del RMSE de entre 2 y 11 % y reducciones del NSE hacia valores comprendidos entre 0.52 y 0.59. Las configuraciones más reducidas, como aquellas que conservan únicamente variables radiativas, exhiben el peor desempeño global, evidenciando la pérdida de información relevante para la predicción del caudal.

En función de estos resultados se concluye que la mejor alternativa es conservar la **configuración base con todas las variables**. Esta opción garantiza el mayor nivel de eficiencia predictiva, evita incrementos innecesarios del error y preserva la información necesaria para representar de forma adecuada los procesos hidrológicos en el marco de la arquitectura recurrente.

B.2.4. Análisis de relevancia mediante *Integrated Gradients*

Posterior al análisis de correlación, se procedió a estimar la relevancia individual de cada predictor utilizando el método de *Integrated Gradients* (IG), aplicado sobre el modelo recurrente entrenado con la arquitectura base y la ventana temporal seleccionada. Este enfoque permite cuantificar la contribución marginal de cada variable a la predicción, preservando tanto la estructura secuencial del modelo como las interacciones temporales aprendidas durante el entrenamiento. Las Figuras B.10a y B.10b muestran los valores medios de atribución para los conjuntos de entrenamiento y validación, respectivamente.

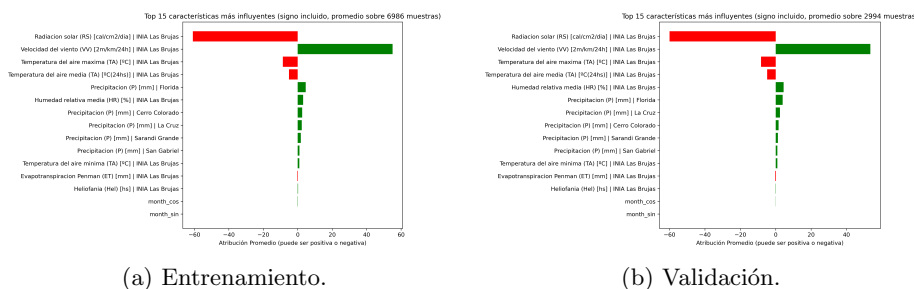


Figura B.10: Importancia relativa de variables según *Integrated Gradients* en el modelo RNN.

Los resultados revelan un patrón de relevancia altamente consistente entre entrenamiento y validación. La variable con mayor contribución positiva es, de forma destacada, la **velocidad del viento (VV)** medida en la estación INIA Las Brujas, que presenta atribuciones promedio muy superiores al resto y se posiciona como la señal dominante dentro de la secuencia de entrada. Le siguen en importancia la **radiación solar (RS)** y, en menor medida, la **humedad relativa media (HR)**, también registradas en la misma estación. Este conjunto de variables conforma un primer bloque de predictores altamente influyentes, coherente con su rol en los procesos energéticos y atmosféricos que condicionan la dinámica hidrológica.

Un segundo bloque de relevancia está compuesto por múltiples variables de

precipitación medidas en distintas estaciones, lo que pone de manifiesto la importancia de la memoria hidrológica espacialmente distribuida y la capacidad del modelo recurrente para integrar información secuencial proveniente de diferentes puntos de la cuenca. Las variables de temperatura (media, máxima y mínima) presentan contribuciones más moderadas y con signos mixtos, lo que sugiere un rol complementario dentro del modelo, pero no dominante.

En contraste, las variables temporales codificadas mediante funciones trigonométricas (*month_sin* y *month_cos*) exhiben atribuciones promedio cercanas a cero en ambos conjuntos, ubicándose sistemáticamente en la cola del *ranking* de relevancia. Este resultado indica que, una vez incorporada explícitamente la información temporal a través de la ventana deslizante, la contribución marginal de los codificadores estacionales resulta limitada dentro del modelo recurrente.

En conjunto, los valores de IG sugieren la existencia de un subconjunto de variables cuya relevancia es marcadamente inferior al resto, y cuya eliminación podría potencialmente simplificar el modelo. De acuerdo con la metodología definida en la Sección 5.7.3, se adoptó un **procedimiento iterativo de poda guiada por IG**, en el cual se elimina en cada iteración la variable (o grupo de variables) con menor atribución promedio y se reentrena el modelo completo, evaluando las métricas de validación en cada paso.

Procedimiento iterativo de eliminación guiada por IG

A partir del *ranking* de relevancia obtenido mediante *Integrated Gradients*, se implementó un procedimiento iterativo de depuración orientado a identificar un subconjunto óptimo de predictores para el modelo RNN. En cada iteración se eliminan las variables con menor atribución promedio y se reentrena completamente la arquitectura, manteniendo constantes la ventana temporal, la arquitectura y las condiciones de entrenamiento, de modo que las diferencias observadas en el desempeño puedan atribuirse exclusivamente a la eliminación de variables.

Dado que las variables *month_sin* y *month_cos* presentaron las atribuciones promedio más bajas y cercanas a cero en el conjunto de validación, se decidió iniciar el proceso de depuración eliminando conjuntamente ambos codificadores temporales. Esta decisión se fundamenta en que dichas variables representan una misma fuente de información estacional y su impacto marginal dentro del modelo recurrente resultó mínimo según el análisis de IG.

Resultados de la primera iteración: eliminación de variables temporales

En la primera iteración del procedimiento iterativo, se eliminaron las variables *month_sin* y *month_cos* del conjunto de entrada. El modelo recurrente fue reentrenado completamente con el conjunto reducido, manteniendo fija la arquitectura base y la ventana temporal seleccionada. La Tabla B.15 presenta la comparación directa entre el desempeño del modelo completo y el obtenido tras eliminar las variables temporales.

Tabla B.15: Desempeño del modelo RNN antes y después de eliminar las variables temporales.

Configuración	NSE	PBIAS [%]	MAE [m^3/s]	RMSE [m^3/s]
Modelo base (todas las variables)	0.61	17.15	17.78	66.94
Sin <i>month_sin</i> , <i>month_cos</i>	0.55	20.15	19.31	71.98

Los resultados muestran un **deterioro consistente del desempeño** tras la eliminación de las variables temporales. El *MAE* aumenta aproximadamente un 9%, mientras que el *RMSE* se incrementa en torno a un 7%. Asimismo, la eficiencia del modelo cae de $\text{NSE} = 0.61$ a 0.55, y el sesgo global (PBIAS) aumenta hasta el 20.15%, indicando una mayor tendencia a la sobreestimación de los caudales.

Si bien el análisis de *Integrated Gradients* indicaba que los codificadores estacionales presentaban una contribución marginal reducida, la degradación observada tras su eliminación evidencia que dichas variables aportan información complementaria que el modelo recurrente utiliza en combinación con la dinámica secuencial aprendida. Este resultado confirma que una baja atribución individual no implica necesariamente prescindibilidad dentro del proceso de entrenamiento, especialmente en modelos recurrentes donde las interacciones temporales juegan un rol central.

Dado que la eliminación del primer conjunto de variables ya produce una degradación apreciable del desempeño, se concluye que no es conveniente continuar con nuevas iteraciones de eliminación guiada por IG. En consecuencia, se mantiene como configuración final el **modelo completo con todas las variables**, que ofrece el mejor equilibrio entre precisión, estabilidad y capacidad de generalización.

Conjunto final de variables

Integrando los resultados de las evaluaciones acumulativas realizadas a lo largo del proceso de modelado —incluyendo la incorporación progresiva de variables, la representación espacial de la precipitación, la selección de la ventana temporal de entrada, la depuración por correlación y el análisis de relevancia mediante *Integrated Gradients*— se determinó el conjunto final de variables que maximiza el desempeño del modelo RNN.

En todos los casos evaluados, tanto la eliminación de variables basada en correlaciones como la poda guiada por IG condujeron a un deterioro sistemático de las métricas de validación. En particular, incluso la remoción de predictores con baja atribución promedio según IG produjo pérdidas apreciables de eficiencia y aumentos del error. En consecuencia, el conjunto óptimo de variables coincide con la **configuración completa empleada en la mejor versión del**

modelo, utilizando una ventana temporal de entrada de **9 días**.

El conjunto final de predictores seleccionados para la arquitectura RNN se detalla a continuación:

- **Precipitación (P)**. Se conservan las precipitaciones diarias medidas en todas las estaciones individuales incluidas en el conjunto original, sin aplicar esquemas de agregación espacial. Esta representación presentó el menor sesgo global (*PBIAS*) entre las alternativas evaluadas y permite preservar la variabilidad espacial fina asociada a la distribución de la lluvia en la cuenca. La información temporal de la precipitación se incorpora exclusivamente a través de una **ventana deslizante de 9 días**, coherente con la naturaleza secuencial del modelo RNN, lo que permite capturar de forma explícita la memoria hidrológica sin recurrir a *lags* manuales.
- **Variables energéticas y radiativas**. Se incluyen todas las variables atmosféricas medidas en la estación INIA Las Brujas:
 - **Velocidad del viento (VV)**,
 - **Radiación solar (RS)**,
 - **Humedad relativa media (HR)**,
 - **Heliofanía (Hel)**,
 - **Evapotranspiración de Penman (ET)**.

El análisis mediante IG identificó a la velocidad del viento y la radiación solar como los predictores más influyentes del modelo, seguidos por la humedad relativa. Si bien la heliofanía y la evapotranspiración presentaron atribuciones promedio más reducidas, su eliminación generó un deterioro claro del desempeño, lo que indica que aportan información complementaria relevante dentro del flujo secuencial del modelo recurrente.

- **Temperatura del aire**. Se conservan las temperaturas diarias **mínima**, **media** y **máxima** registradas en INIA Las Brujas. Estas variables mostraron contribuciones moderadas y con signos mixtos en el análisis de IG, pero su remoción produjo incrementos del error y reducciones de la eficiencia, por lo que ninguna fue considerada prescindible.
- **Variables temporales**. Se mantienen las variables estacionales codificadas mediante funciones trigonométricas (*month_sin* y *month_cos*). Aunque estas variables presentaron atribuciones promedio bajas en el análisis de IG, su eliminación produjo una degradación significativa del desempeño del modelo, evidenciando que aportan información estacional complementaria que resulta relevante cuando se combina con la ventana temporal de entrada.

En conjunto, el modelo RNN final aprovecha de manera efectiva la combinación de información espacial detallada de la precipitación, señales atmosféricas

energéticas, condiciones térmicas locales y una formulación explícita de memoria temporal mediante una ventana deslizante. La evidencia empírica demuestra que cualquier reducción del conjunto completo de variables conduce a un deterioro del desempeño, por lo que la configuración final corresponde al **conjunto completo de predictores optimizado a través del proceso secuencial de evaluación y selección descrito**.

B.2.5. Optimización de hiperparámetros (Optuna)

Para determinar la configuración más adecuada de la arquitectura RNN→MLP se utilizó el *framework* de optimización bayesiana *Optuna*. El espacio de búsqueda considerado, detallado en la Sección 5.7.4, incluyó hiperparámetros estructurales de la red recurrente (tipo de celda, tamaño del estado oculto, número de capas y *dropout*), así como parámetros asociados a la red densa posterior y al proceso de entrenamiento (tasa de aprendizaje, *weight decay*, *dropout* del MLP y *batch size*).

El proceso de optimización constó de un total de **209 trials**, de los cuales **130 finalizaron correctamente** y **79 fueron podados de forma anticipada** mediante el algoritmo de *median pruning*. No se registraron *trials* fallidos. Esta estrategia permitió descartar configuraciones poco prometedoras en etapas tempranas del entrenamiento, reduciendo significativamente el costo computacional del proceso.

Durante las primeras iteraciones se observó una elevada variabilidad en el valor de la función objetivo, con algunos *trials* que presentaron pérdidas de validación muy altas, lo que evidencia una fase inicial de exploración amplia del espacio de búsqueda. A medida que avanzó la optimización, Optuna concentró el muestreo en regiones más prometedoras, logrando una disminución progresiva del mejor valor alcanzado. La Figura B.11 muestra la evolución del valor objetivo para cada *trial*, junto con la trayectoria del mejor resultado acumulado.

El uso conjunto de *early stopping* y *pruning* permitió acelerar la convergencia del proceso. En el *trial* óptimo, identificado como el número **108**, la detención temprana se produjo una vez alcanzado un mínimo estable de la pérdida de validación, con un valor final de $val_loss = 3185.67$, que se mantuvo como el mejor resultado registrado durante toda la optimización.

La Figura B.12 ilustra el comportamiento del mecanismo de poda, donde se observa que la mayoría de los *trials* descartados presentan valores intermedios de la función objetivo considerablemente superiores a los alcanzados por las configuraciones completas, validando la efectividad del criterio de *pruning* adoptado.

La configuración óptima seleccionada corresponde a una arquitectura recurrente de complejidad moderada, combinada con una red densa posterior relativamente profunda. La Tabla B.16 resume los hiperparámetros seleccionados por Optuna para el mejor *trial*.

La Figura B.13 presenta las curvas de aprendizaje correspondientes a los cinco mejores *trials*. En todos los casos se observa una reducción rápida de la pérdida durante las primeras épocas, seguida de una estabilización progresi-

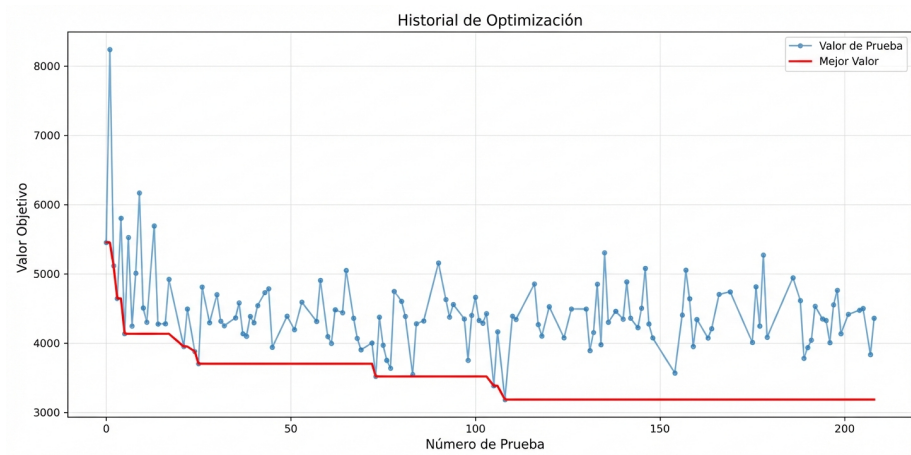


Figura B.11: Historial del proceso de optimización con Optuna para el modelo RNN.

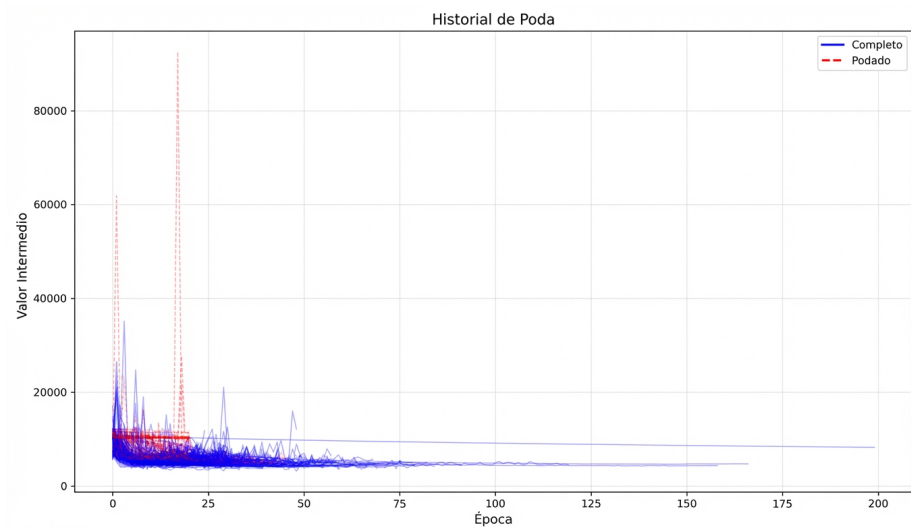


Figura B.12: Historial del proceso de *pruning* durante la optimización con Optuna.

Tabla B.16: Configuración óptima seleccionada por Optuna (RNN→MLP).

Hiperparámetro	Valor óptimo
Tipo de celda RNN	RNN
Tamaño oculto RNN	44
Número de capas RNN	4
<i>Dropout</i> RNN	0.14
Capas MLP (número y anchos)	4 capas: [323, 34, 32, 277]
Activación MLP	<code>leaky_relu</code>
<i>Dropout</i> base MLP	0.11
Tasa de aprendizaje	1.29×10^{-4}
<i>Batch size</i>	16
<i>Weight decay</i>	1.19×10^{-8}

va, lo que indica un entrenamiento estable y consistente entre configuraciones competitivas.

En síntesis, la optimización mediante Optuna permitió identificar una configuración robusta del modelo RNN→MLP, capaz de equilibrar adecuadamente capacidad representacional y estabilidad del entrenamiento. A diferencia de arquitecturas recurrentes más simples evaluadas durante la búsqueda, la combinación de múltiples capas RNN con una red densa posterior profunda permitió capturar de forma más efectiva las dependencias temporales y no lineales del sistema. Estos resultados refuerzan la importancia de una búsqueda sistemática de hiperparámetros, ya que configuraciones más complejas no garantizaron mejoras adicionales sin una regularización adecuada.

B.2.6. Evaluación extendida del modelo optimizado

El modelo RNN final se entrenó utilizando el conjunto de datos integrado descrito en secciones previas, construido a partir de la combinación de las series hidrométricas de DINAGUA y las series meteorológicas de INUMET e INIA para la cuenca del río Santa Lucía. Al igual que en el caso del modelo MLP, las bases se unificaron mediante la fecha como clave común, conservando el caudal diario en Florida (Puente Ruta 5) como variable objetivo y las variables meteorológicas seleccionadas como predictores.

La matriz de entrada incluyó las precipitaciones diarias de todas las estaciones consideradas, las variables atmosféricas medidas en INIA Las Brujas (radiación solar, velocidad del viento, humedad relativa, heliofanía, evapotranspiración de Penman y temperaturas mínima, media y máxima), así como las variables temporales codificadas de forma cíclica. A diferencia del MLP, la información temporal se incorporó exclusivamente mediante una **ventana deslizante de 9 días**, explotando la naturaleza recurrente del modelo.

El particionamiento temporal reservó el 20 % final del período como conjunto

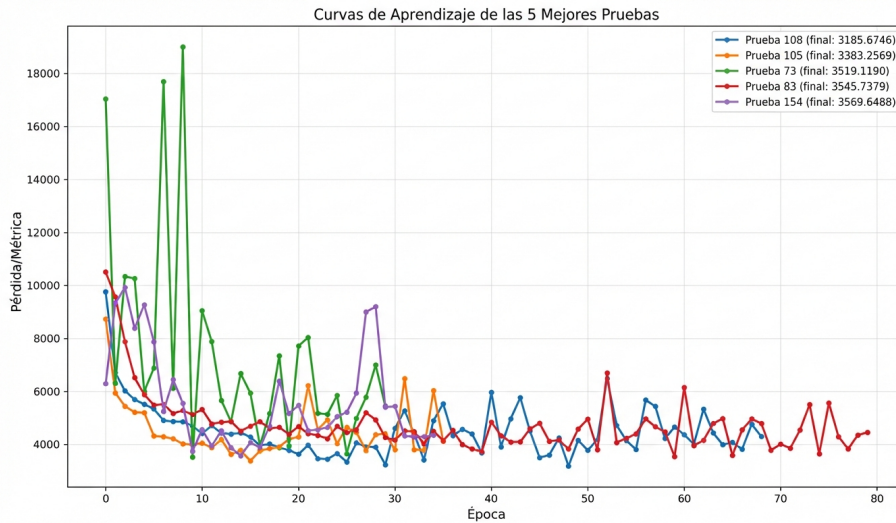


Figura B.13: Curvas de aprendizaje de los cinco mejores *trials* evaluados por Optuna para el modelo RNN.

de prueba independiente, mientras que el 80 % restante se utilizó para entrenamiento y validación interna. Las métricas se calcularon únicamente sobre días con caudales observados válidos.

Desempeño cuantitativo

La Tabla B.17 presenta el desempeño del modelo en entrenamiento, validación y prueba.

Tabla B.17: Desempeño del modelo RNN optimizado en los distintos conjuntos.

Conjunto	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Entrenamiento	0.69	-29.15	15.03	55.20
Validación	0.69	0.47	17.46	59.56
Prueba	0.59	-15.99	18.99	82.81

En el conjunto de entrenamiento el modelo alcanza un $NSE = 0.69$, valor que se mantiene prácticamente constante en validación (0.69), lo que indica ausencia de sobreajuste significativo durante el proceso de optimización. No obstante, en el conjunto independiente de prueba el desempeño disminuye hasta $NSE = 0.59$, clasificándose como *satisfactorio* según los criterios de [Moriassi y cols. \(2015\)](#).

El incremento del RMSE en prueba (82.81 m³/s) respecto a entrenamiento y validación evidencia la influencia de errores asociados a eventos de crecida. El

MAE (18.99 m³/s) confirma una capacidad razonable de ajuste en condiciones medias, aunque inferior a la alcanzada por el modelo MLP.

El PBIAS en prueba (-15.99%) revela una tendencia sistemática a la **sobreestimación** del caudal. Si bien en validación el sesgo es prácticamente nulo (0.47%), la generalización temporal introduce nuevamente un desbalance.

Evolución del entrenamiento

La Figura B.14 muestra la evolución de la pérdida. Se observa una disminución pronunciada en las primeras épocas, seguida de oscilaciones significativas en la curva de validación. La mínima pérdida de validación se alcanza en la época 49, momento en el cual se activa el criterio de *early stopping*.

Las oscilaciones persistentes de la pérdida de validación sugieren mayor inestabilidad en el proceso de optimización respecto al MLP, posiblemente asociada a la dinámica interna de la arquitectura recurrente y a la propagación de gradientes a lo largo de la secuencia temporal.

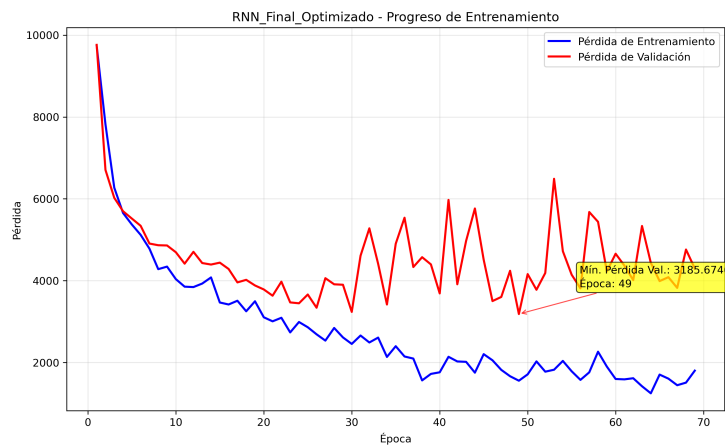


Figura B.14: Evolución de la pérdida de entrenamiento y validación del modelo RNN optimizado.

Análisis gráfico del ajuste

La Figura B.15 muestra la comparación temporal en el conjunto completo. El modelo reproduce adecuadamente la dinámica general y la ocurrencia de eventos, pero presenta una alteración sistemática en la amplitud de los pulsos de caudal.

En particular, las crecidas de mayor magnitud tienden a ser subestimadas, mientras que algunos eventos de magnitud intermedia aparecen sobredimensionados respecto a los valores observados. Este comportamiento sugiere una compresión del rango dinámico de la señal predicha, con una tendencia del modelo a concentrar las estimaciones en un intervalo intermedio de caudales.

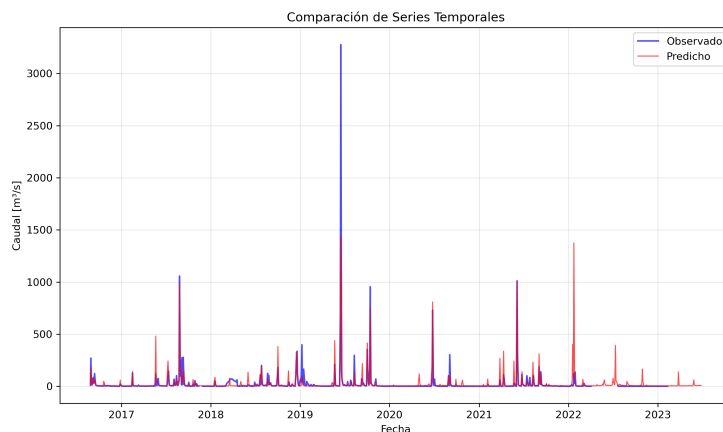


Figura B.15: Serie temporal de caudales observados y predichos por el modelo RNN.

El *scatterplot* en escala logarítmica (Figura B.16) muestra una dispersión considerable en el rango medio y alto de caudales. Para valores elevados predomina la subestimación (puntos por debajo de la línea 1:1), mientras que en el rango intermedio se observan múltiples casos de sobreestimación.

Este patrón confirma que el modelo no solo atenúa los extremos superiores, sino que también amplifica algunos pulsos moderados, evidenciando una distorsión en la representación relativa de la amplitud de los eventos.

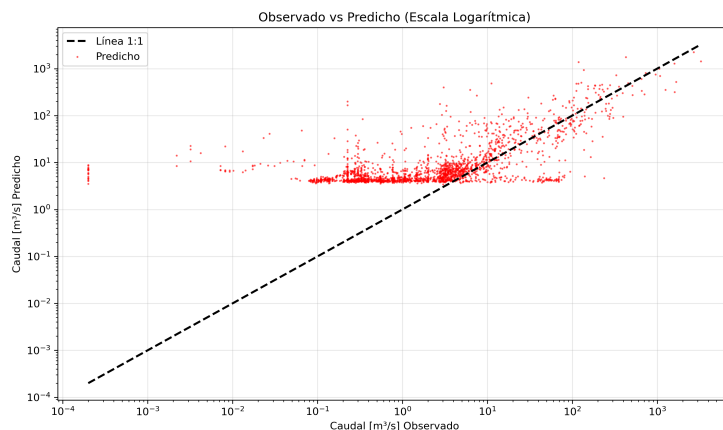


Figura B.16: Diagrama observado vs predicho en escala logarítmica.

El análisis de residuos (Figura B.17) muestra errores concentrados alrededor de cero para la mayor parte de la serie, aunque con valores extremos negativos asociados a crecidas importantes. No se identifican patrones temporales

sistemáticos, pero sí mayor amplitud de residuos respecto al MLP.

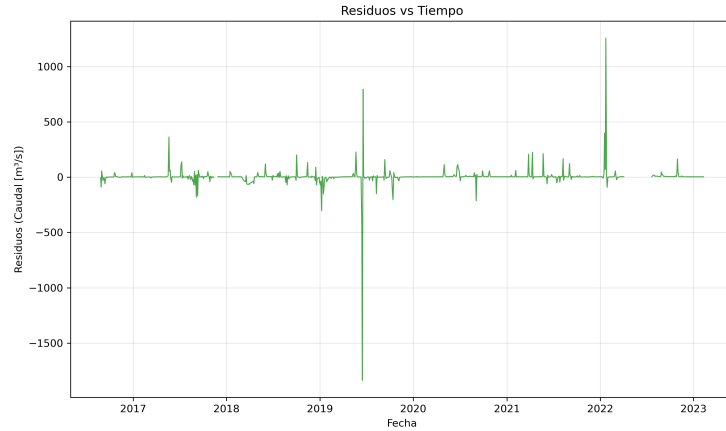


Figura B.17: Residuos del modelo RNN en función del tiempo.

Análisis de relevancia mediante *Integrated Gradients*

El análisis de atribución (Figura B.18) indica que la radiación solar presenta la mayor contribución positiva promedio, seguida por algunas variables térmicas y de humedad. A diferencia del MLP, la señal de precipitación no domina claramente las atribuciones globales.

No obstante, debe considerarse que, en esta implementación, las *Integrated Gradients* se evaluaron únicamente sobre el primer paso temporal de la ventana de entrada. Dado que las redes recurrentes integran información a lo largo de toda la secuencia mediante su estado oculto, es posible que parte de la contribución asociada a la precipitación se encuentre distribuida en pasos temporales anteriores dentro de la ventana, los cuales no fueron descompuestos explícitamente en el análisis de atribuciones. En consecuencia, la interpretación presentada debe entenderse como una aproximación parcial a la dinámica completa de relevancia temporal del modelo.

Costo computacional

El entrenamiento completo del modelo demandó 99.21 segundos en un total de 69 épocas en el entorno de cómputo utilizado. Por su parte, el tiempo de inferencia fue de 0.01 segundos, de acuerdo con los registros del sistema.

B.3. PINN-P MLP

Esta sección presenta el desarrollo detallado del modelo PINN-P con arquitectura MLP, incluyendo todos los experimentos realizados durante el proceso de optimización.

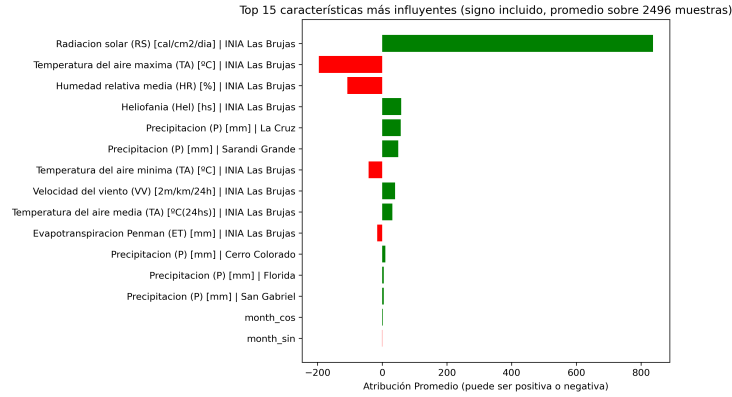


Figura B.18: Top 15 características más influyentes según *Integrated Gradients* en el conjunto de prueba.

El cuerpo principal (Sección 6.6) resume únicamente los resultados del modelo base y del modelo final optimizado. Aquí se documentan las etapas intermedias que condujeron a la configuración seleccionada.

B.3.1. Modelo base

El primer experimento en el marco del enfoque PINN consistió en entrenar una red MLP bajo la formulación **PINN-P**, es decir, haciendo que la red prediga exclusivamente la precipitación agregada diaria \hat{P}_{agg} , mientras que la evapotranspiración se mantiene fija a partir de los valores observados E_{obs} . La predicción final de caudal no es producida directamente por la red, sino que emerge del acoplamiento físico con GR4J, evaluando el caudal simulado $Q_{GR4J}(\hat{P}_{agg}, E_{obs}; \theta)$ y comparándolo con el caudal observado mediante el término físico de la función de pérdida.

Dado que el conjunto de prueba se reserva exclusivamente para la evaluación final del modelo, el desempeño en esta etapa se reporta sobre el **conjunto de validación**, utilizando MAE, RMSE, NSE y PBIAS calculadas sobre el caudal simulado por GR4J.

Tabla B.18: Desempeño del modelo base PINN-P (MLP) sobre el conjunto de validación.

NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
0.66	13.30	15.41	62.35

La Figura B.19 muestra la evolución de la pérdida durante el entrenamiento del modelo. Se observa una reducción pronunciada tanto en entrenamiento como en validación durante las primeras épocas, seguida por una estabilización

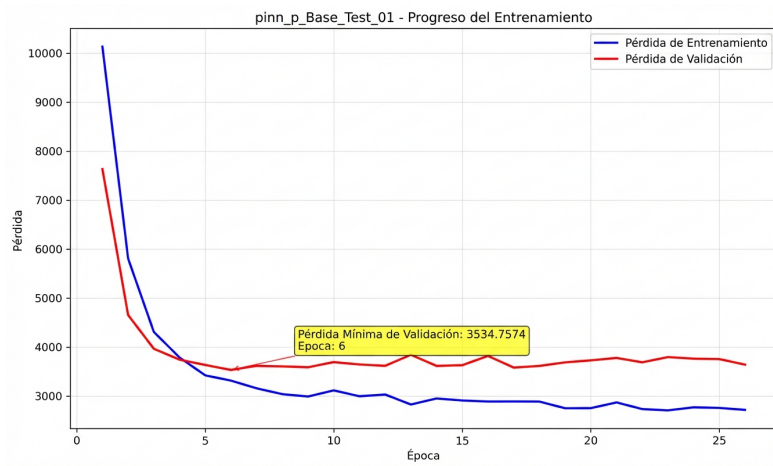


Figura B.19: Evolución de la función de pérdida (entrenamiento y validación) para el modelo base PINN-P (MLP).

temprana del desempeño en validación. En particular, la pérdida de validación alcanza su mínimo en la época 6, a partir de la cual se aprecia una separación progresiva entre ambas curvas: la pérdida de entrenamiento continúa disminuyendo, mientras que la pérdida de validación se mantiene aproximadamente constante o presenta oscilaciones leves.

Este patrón es consistente con la aparición temprana de *overfitting*, indicando que el modelo comienza a especializarse en el conjunto de entrenamiento sin traducir ese ajuste en mejoras sostenidas sobre datos no vistos. En el contexto PINN, este comportamiento resulta relevante porque sugiere que, aun con el acoplamiento físico con GR4J, el equilibrio inicial entre los términos de pérdida y la configuración base del modelo no alcanzan para sostener una mejora estable en generalización, lo cual justifica las etapas posteriores de ajuste (incorporación de variables, selección de *lags* y análisis del balance $w_{\text{data}} : w_{\text{phys}}$).

En términos cuantitativos, la tabla B.18 el modelo base presenta un desempeño razonable sobre validación, con $NSE = 0.66$, lo que indica que el sistema acoplado explica aproximadamente dos tercios de la variabilidad observada del caudal en este subconjunto. El error cuadrático medio ($RMSE = 62.35 \text{ m}^3/\text{s}$) se mantiene en un rango moderado, mientras que el sesgo porcentual positivo ($PBIAS = 13.30 \%$) evidencia una tendencia sistemática a la subestimación del volumen total simulado.

En conjunto, estos resultados muestran que la configuración base de la PINN-P es capaz de capturar una fracción importante de la dinámica de caudales a través del modelo GR4J forzado con \hat{P}_{agg} .

B.3.2. Evaluación de la incorporación de variables

En esta etapa se analiza el impacto de distintas configuraciones de variables de entrada sobre el desempeño del modelo PINN-P basado en arquitectura MLP. Tal como se detalló en la metodología, el objetivo es determinar qué combinación de variables mejora la capacidad del sistema acoplado (red neuronal + GR4J) para reproducir el caudal observado en el conjunto de validación.

A diferencia del modelo MLP tradicional, en la formulación PINN-P la red neuronal no predice directamente el caudal, sino la precipitación agregada diaria \hat{P}_{agg} , que luego es utilizada como forzante del modelo GR4J junto con la evapotranspiración observada. En consecuencia, cualquier modificación en las variables de entrada puede afectar tanto la calidad de la estimación de \hat{P}_{agg} como la coherencia hidrológica del caudal simulado.

Las métricas reportadas en esta sección se calculan sobre el conjunto de validación y corresponden al caudal generado por GR4J, evaluado mediante MAE, RMSE, NSE y PBIAS. En todos los experimentos se mantuvo constante la arquitectura base del modelo y los pesos de la función de pérdida, de modo que las diferencias observadas puedan atribuirse exclusivamente a la configuración de variables considerada.

Variables temporales

En primer lugar, se evaluó el efecto de incorporar variables temporales representadas de forma cíclica mediante transformaciones seno y coseno del *mes* y del *día*. La inclusión de estas variables busca capturar patrones estacionales que puedan influir en la estimación de la precipitación agregada y, en consecuencia, en la simulación del caudal a través de GR4J.

Se compararon cuatro configuraciones: (i) sin variables temporales, (ii) incorporación únicamente del mes, (iii) incorporación únicamente del día, y (iv) incorporación conjunta de día y mes. Los resultados obtenidos se presentan en la Tabla B.19.

Tabla B.19: Comparación de desempeño para distintas combinaciones de variables temporales sobre el conjunto de validación (PINN-P MLP).

Configuración	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Ninguna	0.66	13.30	15.41	62.35
día	0.66	13.82	15.46	62.32
mes	0.67	9.41	15.18	61.15
día, mes	0.66	16.37	15.24	62.30

Los resultados muestran que la incorporación de la variable *mes* produce la mejora más consistente en el desempeño del modelo. Esta configuración alcanza el mayor valor de eficiencia ($NSE = 0.67$) y los menores errores en términos de $RMSE$ (61.15 m³/s) y MAE (15.18 m³/s), además de reducir de forma

significativa el sesgo porcentual ($PBIAS = 9.41\%$) respecto a la configuración sin variables temporales.

En contraste, la incorporación exclusiva del *día* no genera mejoras apreciables respecto al caso base. Las métricas permanecen prácticamente inalteradas en términos de NSE y $RMSE$, mientras que el sesgo aumenta levemente. La inclusión conjunta de día y mes tampoco aporta beneficios adicionales; por el contrario, incrementa el $PBIAS$ hasta 16.37% , indicando una mayor subestimación sistemática del caudal.

Desde una perspectiva hidrológica, estos resultados son coherentes con la escala temporal dominante de los procesos que regulan la dinámica del caudal. Los patrones estacionales —asociados a ciclos de precipitación, radiación y evapotranspiración— operan principalmente a escala mensual, mientras que el día del mes no suele estar vinculado a mecanismos físicos recurrentes en el sistema hidrológico. En el contexto PINN-P, donde la red debe estimar una precipitación efectiva consistente con el caudal observado a través de GR4J, la información estacional proporcionada por el mes actúa como un regulador que mejora tanto la eficiencia como el control del sesgo.

En consecuencia, y considerando simultáneamente el desempeño cuantitativo y la coherencia hidrológica del sistema acoplado, se decidió **conservar únicamente la variable temporal correspondiente al mes** para las etapas posteriores del modelado.

Representación espacial de los *lags* de precipitación

En esta etapa se evaluó el efecto de incorporar valores retardados (*lags*) de la precipitación como variables de entrada, considerando distintos esquemas de representación espacial. A diferencia del modelo MLP tradicional, en la formulación PINN-P la red neuronal predice directamente la precipitación agregada diaria \hat{P}_{agg} . Por este motivo, el análisis se restringe a la inclusión de *lags* como variables explicativas auxiliares, manteniendo la precipitación contemporánea como variable objetivo exclusiva del modelo.

Se compararon cinco configuraciones: (i) sin *lags* de precipitación, (ii) *lags* ponderados por distancia, (iii) *lags* ponderados mediante polígonos de Thiessen, (iv) promedio simple de estaciones, y (v) *lags* individuales por estación. Las métricas se calcularon sobre el conjunto de validación, evaluando el caudal simulado por GR4J. Los resultados se presentan en la Tabla B.20.

Los resultados muestran que la incorporación de *lags* de precipitación no mejora el desempeño del modelo PINN-P; por el contrario, produce un deterioro sistemático en todas las métricas evaluadas. La configuración sin *lags* alcanza el mayor valor de eficiencia ($NSE = 0.67$), así como los menores valores de $RMSE$ ($61.15\text{ m}^3/\text{s}$) y MAE ($15.18\text{ m}^3/\text{s}$). Asimismo, presenta el menor sesgo porcentual ($PBIAS = 9.41\%$), considerablemente inferior al observado en las configuraciones con *lags*.

En particular, se observa un aumento marcado del $PBIAS$ al incorporar memoria temporal. Las configuraciones con agregación espacial de *lags* elevan el sesgo a valores comprendidos entre 17.24% y 21.89% , indicando una tendencia

Tabla B.20: Comparación de desempeño según esquema de representación espacial de los *lags* de precipitación (PINN-P MLP).

Estrategia	NSE	PBIAS [%]	MAE [m³/s]	RMSE [m³/s]
Sin <i>lags</i> de precipitación	0.67	9.41	15.18	61.15
<i>Lags</i> ponderados por distancia	0.66	17.24	15.53	62.27
<i>Lags</i> ponderados Thiessen	0.64	19.98	15.66	63.84
Promedio simple estaciones	0.64	21.66	15.78	64.40
<i>Lags</i> por estación	0.63	21.89	15.89	64.89

creciente a la subestimación del volumen total simulado. Este comportamiento sugiere que la inclusión de información retardada de precipitación introduce redundancias o interacciones que afectan negativamente la coherencia global del sistema acoplado.

Desde el punto de vista de la parsimonia, la incorporación de *lags* implica además un incremento sustancial en el número de variables de entrada, aumentando la complejidad estructural del modelo sin aportar mejoras en desempeño. Dado que la configuración sin *lags* no sólo resulta más simple, sino que además exhibe el mejor comportamiento en términos de eficiencia, error y control del sesgo, se decidió **no incorporar valores retardados de precipitación** en el modelo PINN-P.

En consecuencia, no se consideró necesario realizar un análisis adicional para determinar el horizonte óptimo de *lags*, ya que la evidencia empírica indica que su inclusión no resulta beneficiosa en esta formulación.

B.3.3. Depuración inicial por correlación

Siguiendo el mismo procedimiento aplicado al modelo MLP y RNN, se realizó un análisis de correlación entre las variables meteorológicas con el objetivo de identificar posibles redundancias y evaluar la factibilidad de reducir la dimensionalidad del modelo sin deteriorar su desempeño.

Tal como se observó previamente, las variables térmicas (temperatura máxima, media y mínima) presentan correlaciones elevadas entre sí, al igual que las variables asociadas al balance radiativo (radiación solar, heliofanía y evapotranspiración). Dado que la inclusión de variables altamente correlacionadas puede introducir redundancia e incrementar innecesariamente la complejidad del modelo, se diseñaron distintos escenarios de reducción progresiva.

En todos los casos, el modelo fue reentrenado completamente bajo la formulación PINN-P, manteniendo fija la arquitectura base y los pesos de la función de pérdida. Las métricas se calcularon sobre el conjunto de validación, evaluando

el caudal simulado por GR4J.

Evaluación combinada de reducción de variables correlacionadas

Las configuraciones evaluadas fueron las siguientes:

- Modelo base (todas las variables seleccionadas previamente).
- Sólo radiación solar y heliofanía (sin evapotranspiración).
- Sólo temperatura media (sin temperaturas extrema máxima y mínima).
- Radiación solar y evapotranspiración (sin heliofanía).
- Sólo radiación solar.
- Temperatura media + radiación solar + heliofanía.
- Temperatura media + radiación solar + evapotranspiración.
- Temperatura media + radiación solar.

Tabla B.21: Resultados de desempeño para configuraciones de reducción de variables correlacionadas (PINN-P MLP).

Variables re-movidas	NSE	PBIAS [%]	MAE [m³/s]	RMSE [m³/s]
Ninguna (modelo base)	0.67	9.41	15.18	61.15
ET	0.67	15.10	15.05	61.52
T _{máx} , T _{mín}	0.67	13.42	15.04	61.73
Hel	0.66	16.53	15.12	62.04
Hel, ET	0.66	16.38	15.17	62.35
T _{máx} , T _{mín} , ET	0.66	9.89	15.54	62.41
T _{máx} , T _{mín} , Hel	0.66	10.41	15.62	62.58
T _{máx} , T _{mín} , Hel, ET	0.65	16.16	15.38	63.26

Los resultados de la Tabla B.21 muestran que ninguna de las configuraciones reducidas logra superar el desempeño del modelo base de manera consistente. Si bien algunas alternativas presentan valores de MAE levemente inferiores, estas mejoras puntuales no se traducen en incrementos sostenidos del NSE ni en reducciones sistemáticas del RMSE.

El modelo base mantiene el mayor valor de eficiencia ($NSE = 0.67$) junto con el menor RMSE (61.15 m³/s), lo que indica que la inclusión completa del conjunto de variables permite una mejor reproducción de la variabilidad

del caudal en validación. Además, presenta el menor valor de sesgo porcentual ($PBIAS = 9.41\%$).

En particular, las configuraciones que eliminan simultáneamente variables térmicas y radiativas tienden a incrementar el sesgo, alcanzando valores superiores al 15 %, lo que indica una mayor subestimación sistemática del volumen total simulado. Este comportamiento sugiere que, en el contexto PINN-P, la interacción entre variables energéticas y térmicas contribuye de manera complementaria a la estimación de una precipitación efectiva coherente con la dinámica de GR4J.

Desde una perspectiva de parsimonia, si bien algunas configuraciones reducen el número de variables de entrada, esta simplificación no produce mejoras claras en desempeño que justifiquen la pérdida de información potencial. En consecuencia, se decidió **mantener el modelo base con el conjunto completo de variables**, dado que ofrece el mejor equilibrio entre eficiencia, precisión y control del sesgo.

B.3.4. Análisis de relevancia mediante *Integrated Gradients*

Una vez definido el conjunto de variables a partir del análisis de correlación, se evaluó la relevancia individual de cada característica mediante el método de *Integrated Gradients* (IG). Este enfoque permite cuantificar la contribución promedio de cada variable a la predicción del modelo, proporcionando una medida interpretativa consistente con la estructura interna de la red.

El cálculo se realizó por separado sobre los subconjuntos de entrenamiento y validación. La Figura B.20 presenta las 15 variables con mayor atribución promedio (incluyendo el signo) en ambos subconjuntos.

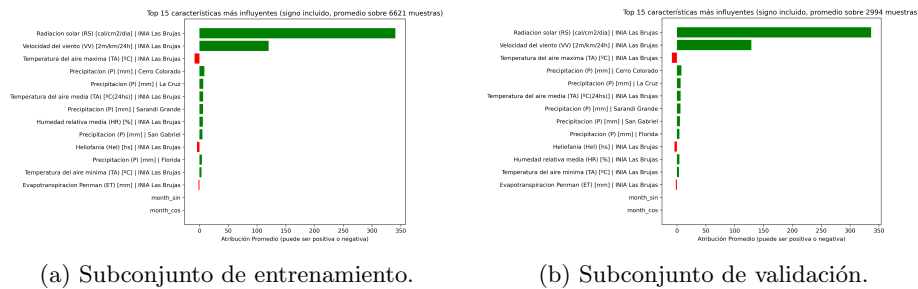


Figura B.20: Importancia relativa de variables según *Integrated Gradients* (IG) en los conjuntos de entrenamiento y validación (PINN-P MLP).

Los resultados muestran una alta consistencia entre entrenamiento y validación, indicando que el modelo aprende patrones estables y no dependientes del subconjunto utilizado. En ambos casos, las variables asociadas al balance energético y aerodinámico en INIA Las Brujas —en particular **radiación solar (RS)** y **velocidad del viento (VV)**— aparecen como las características más influyentes. En un segundo nivel de relevancia se observa la contribución

de variables térmicas y de humedad, así como de las precipitaciones medidas en distintas estaciones, con atribuciones positivas de menor magnitud.

En contraste, las variables temporales *month_sin* y *month_cos* presentan atribuciones promedio cercanas a cero en ambas particiones, sugiriendo un aporte directo reducido dentro del modelo. Sin embargo, al igual que en el caso del modelo MLP, una atribución baja según IG no implica necesariamente que la variable sea prescindible, ya que puede aportar información complementaria o actuar como regulador estadístico del modelo.

Evaluación empírica de eliminación de variables de mes. Dado que *month_sin* y *month_cos* presentan atribuciones marginales, se evaluó empíricamente el efecto de excluirlas en conjunto. La Tabla B.22 compara el desempeño del modelo con y sin variables de mes en el conjunto de validación.

Tabla B.22: Comparación de desempeño del modelo PINN-P (MLP) en el conjunto de validación al eliminar las variables temporales de mes.

Configuración	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Modelo con variables de mes	0.67	9.41	15.18	61.15
Modelo sin variables de mes	0.66	13.30	15.41	62.35

Los resultados muestran que la eliminación de las variables de mes produce un deterioro consistente del desempeño: el *NSE* disminuye de 0.67 a 0.66, el *RMSE* aumenta de 61.15 a 62.35 m³/s, y el sesgo porcentual se incrementa de 9.41 % a 13.30 %. En particular, el aumento del *PBIAS* indica una pérdida relevante en el control del sesgo global, aun cuando las variables de mes presenten una contribución directa baja según IG.

En consecuencia, y priorizando la estabilidad del modelo y el control del sesgo por sobre la depuración basada exclusivamente en atribuciones marginales, se decidió **mantener las variables *month_sin* y *month_cos*** en el conjunto final de entrada. Dado que no se identificaron otras variables con atribuciones bajas cuya exclusión mejorara de forma consistente el desempeño, el procedimiento iterativo de depuración se dio por finalizado en esta etapa.

Conjunto final de variables

El conjunto final de variables para el modelo PINN-P (MLP) fue definido a partir de un proceso secuencial que incluyó: (i) incorporación progresiva de variables temporales, (ii) evaluación de la posible inclusión de *lags* de precipitación, (iii) análisis de correlación entre variables meteorológicas, y (iv) análisis de relevancia mediante *Integrated Gradients* (IG).

El modelo base inicial no incluía variables temporales ni *lags*. En una primera etapa se evaluó la incorporación de información estacional, observándose que

la inclusión del mes mejoraba de forma consistente el desempeño y reducía el sesgo. Por este motivo, se decidió incorporar *month_sin* y *month_cos* en las configuraciones posteriores.

Posteriormente se evaluó la incorporación de *lags* de precipitación bajo distintos esquemas de representación espacial. Los resultados mostraron un deterioro sistemático en eficiencia y un incremento significativo del *PBIAS*, por lo que se concluyó que la inclusión de memoria temporal explícita en la red no resulta beneficiosa en la formulación PINN-P. En consecuencia, los *lags* no fueron incorporados en el modelo final.

El análisis de correlación permitió explorar configuraciones reducidas de variables térmicas y radiativas; sin embargo, ninguna alternativa logró superar de forma consistente al modelo con el conjunto completo de variables meteorológicas. Por su parte, el análisis de IG mostró patrones de relevancia estables entre entrenamiento y validación, sin identificar variables cuya eliminación produjera mejoras sostenidas en desempeño.

En función de estos resultados, el conjunto final de variables quedó conformado por:

- **Precipitaciones contemporáneas por estación** (sin *lags*).
- **Radiación solar (RS), heliofanía (Hel) y evapotranspiración (ET).**
- **Temperatura del aire** (media, máxima y mínima).
- **Humedad relativa (HR) y velocidad del viento (VV).**
- **Variables temporales** *month_sin* y *month_cos*.

Esta configuración representa el mejor equilibrio entre eficiencia predictiva, control del sesgo y complejidad estructural en el marco de la formulación PINN-P. Sobre este conjunto se desarrollaron las etapas posteriores de análisis del balance entre los términos de la función de pérdida y la optimización de hiperparámetros.

B.3.5. Análisis del balance entre términos de la función de pérdida

Uno de los aspectos distintivos de las redes neuronales informadas por la física es la presencia de múltiples términos en la función de pérdida, cuya ponderación relativa determina el equilibrio entre ajuste empírico y coherencia física. En el caso de PINN-P, la función de pérdida combina el término de datos (ajuste de \hat{P}_{agg}) y el término físico (ajuste del caudal simulado por GR4J), ponderados mediante los coeficientes w_{data} y w_{phys} .

Siguiendo la estrategia metodológica previamente descrita, se evaluaron distintas combinaciones de pesos manteniendo fijo el conjunto de variables seleccionado. Se analizaron relaciones comprendidas hasta un máximo de 1:4 en ambos sentidos, excluyendo el caso $w_{phys} = 0$, ya que ello implicaría eliminar completamente la restricción física y desnaturalizar la formulación PINN.

La Tabla B.23 presenta las métricas obtenidas en el conjunto de validación para cada combinación evaluada.

Tabla B.23: Desempeño del modelo PINN-P (MLP) para distintas combinaciones de pesos en la función de pérdida.

$w_{\text{data}} - w_{\text{phys}}$	NSE	PBIAS [%]	MAE [m^3/s]	RMSE [m^3/s]
0.00 – 1.00	0.67	11.32	15.08	61.26
0.25 – 1.00	0.67	8.69	15.20	61.07
0.50 – 1.00	0.67	8.95	15.19	61.10
0.75 – 1.00	0.67	8.74	15.19	60.98
1.00 – 1.00	0.67	9.41	15.18	61.15
1.00 – 0.75	0.67	10.17	15.15	61.27
1.00 – 0.50	0.67	9.77	15.17	61.21
1.00 – 0.25	0.68	9.68	15.13	60.93
0.75 – 0.50	0.68	8.66	15.18	60.86
0.75 – 0.25	0.67	9.74	15.17	61.12

Los resultados muestran que el desempeño del modelo es notablemente estable frente a variaciones moderadas en la relación entre los términos de pérdida. Las diferencias en NSE se mantienen dentro de un rango estrecho (aproximadamente 0.67–0.68), mientras que las variaciones en $RMSE$ y MAE resultan marginales. Asimismo, el $PBIAS$ oscila en un intervalo acotado, sin observarse desbalances extremos incluso en configuraciones con predominancia de uno de los términos.

Se observa que algunas combinaciones con mayor peso relativo del término de datos ($w_{\text{data}} > w_{\text{phys}}$) alcanzan valores levemente superiores de eficiencia ($NSE \approx 0.68$) y menores valores de $RMSE$. No obstante, estas diferencias son del orden de centésimas y no representan mejoras sustanciales en términos prácticos.

De manera similar, configuraciones con predominancia del término físico mantienen niveles de desempeño comparables, lo que sugiere que la restricción impuesta por GR4J actúa como un regulador robusto incluso cuando su peso relativo varía dentro del rango analizado.

En conjunto, los resultados indican que el modelo PINN-P presenta una baja sensibilidad al ajuste fino de los pesos dentro del intervalo considerado, evidenciando una formulación estable del problema de optimización. Dado que ninguna combinación sobresale de manera clara sobre las demás y considerando el principio de simplicidad metodológica, se decidió **mantener la configuración balanceada** $w_{\text{data}} = 1$ y $w_{\text{phys}} = 1$ para las etapas posteriores de optimización y evaluación final.

Esta elección preserva un equilibrio explícito entre supervisión empírica y coherencia física, sin privilegiar artificialmente ninguno de los dos componentes de la función de pérdida.

B.3.6. Optimización de hiperparámetros (Optuna)

Una vez definido el conjunto final de variables y fijado el balance de la función de pérdida en $w_{\text{data}} = 1$ y $w_{\text{phys}} = 1$, se procedió a la optimización de hiperparámetros mediante el *framework* *Optuna*, utilizando el mismo esquema metodológico aplicado en el modelo MLP.

El proceso de búsqueda incluyó hiperparámetros estructurales (número de capas ocultas y neuronas por capa), así como parámetros asociados al entrenamiento (función de activación, tasa de aprendizaje, *dropout*, *weight decay* y *batch size*). La función objetivo correspondió a la pérdida de validación (MSE), evaluada bajo la formulación compuesta propia de la PINN.

El estudio culminó con los siguientes resultados:

- **Mejor valor de pérdida de validación:** 2658.51
- ***Trials* completados:** 202
- ***Trials* podados:** 26

La Figura B.21 muestra la evolución del valor objetivo a lo largo de los *trials*. Durante las primeras iteraciones se observa una fase de exploración amplia, con valores elevados y alta variabilidad. A medida que avanzó el proceso, el algoritmo concentró la búsqueda en regiones más prometedoras del espacio de hiperparámetros, logrando una reducción progresiva del mejor valor alcanzado.

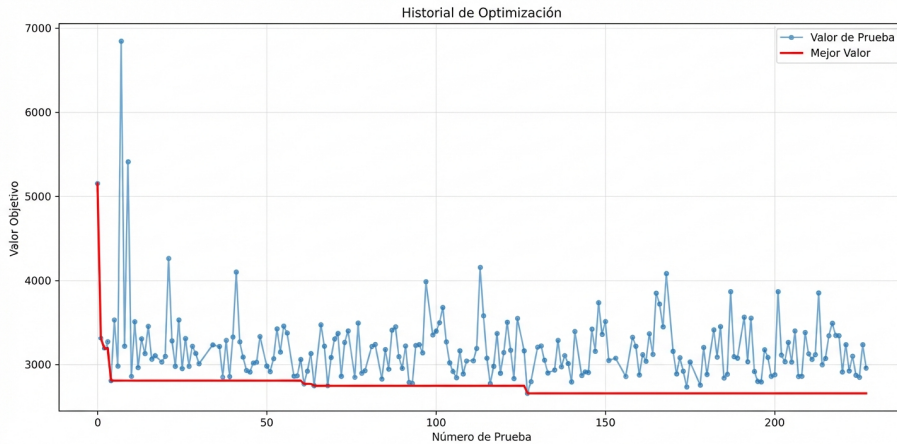


Figura B.21: Historial del proceso de optimización realizado con Optuna para PINN-P (MLP). La línea azul muestra el valor de la función objetivo para cada *trial*, mientras que la línea roja indica el mejor valor obtenido hasta ese punto.

El mejor resultado se obtuvo en el *trial* 127, cuya configuración arquitectónica se resume en la Tabla B.24. A diferencia del modelo MLP tradicional, donde la arquitectura óptima fue relativamente compacta, en el caso de PINN-P la

solución seleccionada presenta **tres capas ocultas**, lo que sugiere que la tarea de estimar precipitación bajo una restricción física puede requerir una mayor capacidad representacional.

Tabla B.24: Configuración óptima seleccionada por Optuna para PINN-P (MLP).

Hiperparámetro	Valor óptimo
Número de capas ocultas	3
Neuronas por capa	[61, 35, 53]
Función de activación	<code>relu</code>
Tasa de aprendizaje	0.00489
<i>Dropout</i> base	0.31
<i>Batch size</i>	32
<i>Weight decay</i>	3.56×10^{-5}

La función de activación seleccionada fue `relu`. Asimismo, la tasa de aprendizaje óptima ($\approx 4.9 \times 10^{-3}$) resulta elevada en comparación con configuraciones típicas de entrenamiento estable, mientras que el valor de *dropout* base (0.31) indica una regularización relativamente fuerte, consistente con la profundidad del modelo.

La Figura B.22 presenta las curvas de aprendizaje correspondientes a los cinco mejores *trials*. Se observa que, aunque todos convergen hacia valores similares de pérdida, existe variabilidad en la estabilidad de las trayectorias a lo largo de las épocas.

Finalmente, la Figura B.23 muestra el historial de *pruning*. El uso de poda permitió descartar configuraciones poco prometedoras en etapas tempranas, reduciendo el costo computacional total sin comprometer la calidad del resultado final.

En conjunto, la optimización mediante Optuna permitió identificar una arquitectura más profunda y con mayor regularización que la obtenida para el MLP tradicional, lo que sugiere que la formulación informada por la física introduce un problema de optimización con características diferentes. No obstante, el proceso mostró una convergencia estable y consistente, respaldando la robustez de la configuración seleccionada.

B.3.7. Evaluación extendida del modelo optimizado

El modelo final PINN-P (MLP) se entrenó utilizando la arquitectura óptima identificada mediante Optuna (Sección B.3.6), manteniendo el balance de pérdida $w_{\text{data}} = 1$ y $w_{\text{phys}} = 1$, y el conjunto de variables definido en las etapas previas.

La evaluación se realizó de forma independiente sobre los conjuntos de entrenamiento, validación y prueba, preservando la partición temporal original y aplicando el mismo criterio de enmascaramiento para los registros con caudal

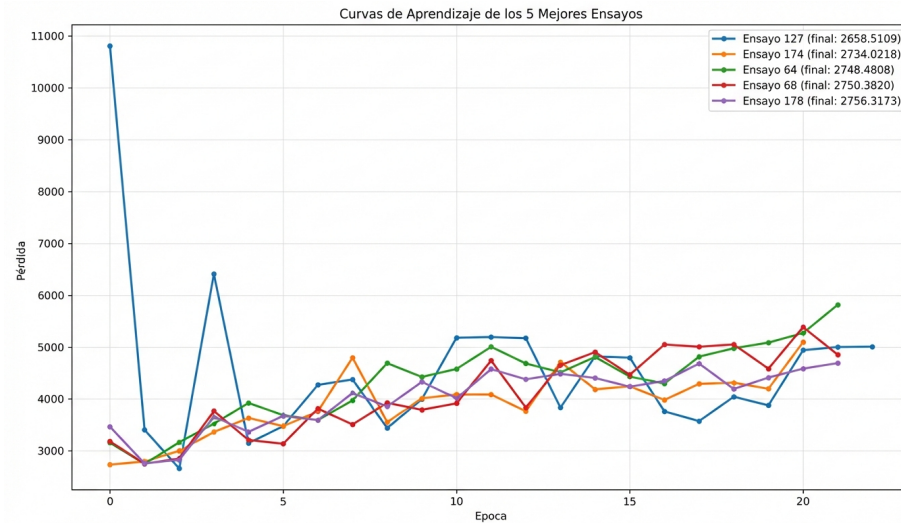


Figura B.22: Curvas de aprendizaje de los cinco mejores *trials* del proceso de optimización de Optuna (PINN-P MLP).

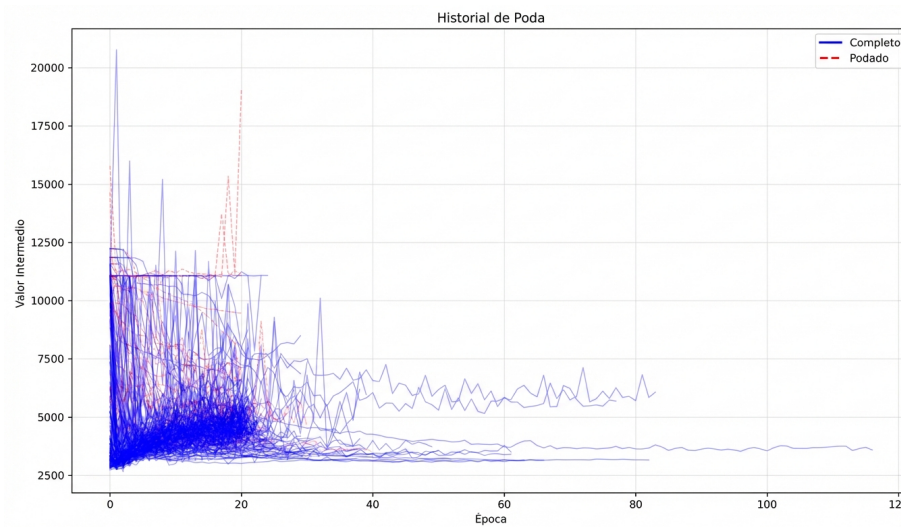


Figura B.23: Historial de *pruning* durante la optimización con Optuna (PINN-P MLP). Las trayectorias podadas se interrumpen tempranamente al no mostrar mejoras en la función objetivo.

faltante. Las métricas consideradas fueron MAE, RMSE, NSE y PBIAS y los resultados se encuentran en la tabla B.25.

Tabla B.25: Desempeño del modelo PINN-P optimizado en los distintos conjuntos.

Conjunto	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Entrenamiento	0.51	-68.14	22.79	71.08
Validación	0.75	-16.73	15.76	53.80
Prueba	0.74	-35.55	18.17	65.98

Análisis del desempeño

En el conjunto de validación se obtiene un desempeño elevado, con un $NSE = 0.75$ y un $RMSE = 53.80$ m³/s, superior al observado en etapas preliminares del modelo base. En el conjunto de prueba, el modelo mantiene una eficiencia comparable ($NSE = 0.74$), lo que indica una adecuada capacidad de generalización temporal.

No obstante, se observa un incremento significativo del sesgo porcentual negativo ($PBIAS = -35.55\%$ en el conjunto de prueba), indicando una tendencia sistemática a la sobreestimación del volumen total de caudal. Este comportamiento contrasta con las configuraciones preliminares evaluadas antes de la optimización, donde el sesgo se mantenía dentro de rangos más moderados.

En el conjunto de entrenamiento, las métricas resultan considerablemente más bajas ($NSE = 0.51$ y $PBIAS = -68.14\%$). Este comportamiento aparentemente contraintuitivo —desempeño inferior en entrenamiento que en validación— se explica por la naturaleza compuesta de la función de pérdida en las PINNs.

Dado que la optimización se realiza sobre una combinación ponderada de pérdida de datos y pérdida física, el mínimo de validación no necesariamente coincide con un mínimo estricto del error directo sobre caudal en entrenamiento. En particular, el modelo puede privilegiar configuraciones que mejoren la coherencia física global del sistema aunque ello implique un mayor error puntual sobre ciertos períodos del conjunto de entrenamiento.

La Figura B.24 muestra la evolución de la pérdida durante el entrenamiento. Se observa que la pérdida de validación alcanza su mínimo en una etapa temprana (época 3), mientras que la pérdida de entrenamiento permanece en valores relativamente elevados, lo que refuerza la idea de que el criterio de optimización no está alineado exclusivamente con el ajuste empírico directo.

Análisis gráfico del ajuste

La Figura B.25 presenta la comparación temporal entre caudal observado y predicho en el conjunto completo. El modelo reproduce adecuadamente la dinámica general de la serie, capturando la ocurrencia de eventos de crecida y la

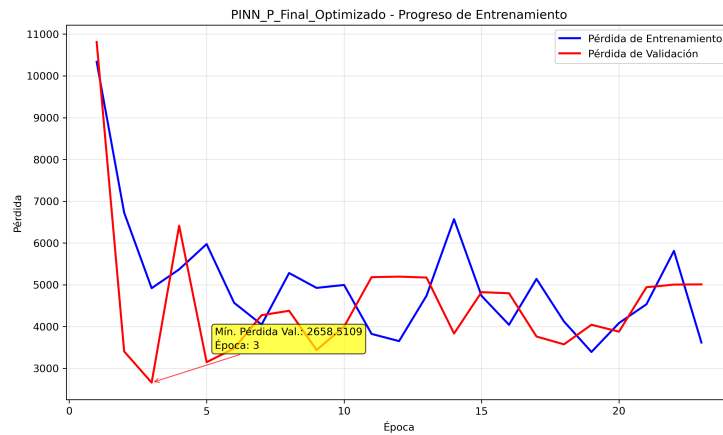


Figura B.24: Evolución de la pérdida de entrenamiento y validación del modelo PINN-P optimizado.

estacionalidad, aunque tiende a suavizar algunos picos máximos y sobreestimar picos medianos y bajos.

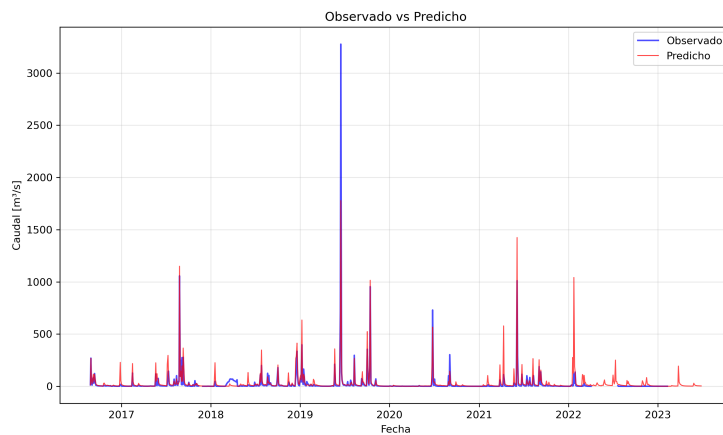


Figura B.25: Serie temporal de caudales observados y predichos por el modelo PINN-P optimizado.

El *scatterplot* en escala logarítmica (Figura B.26) muestra una buena alineación con la línea 1:1 en el rango medio de caudales, aunque tendiendo a la sobreestimación, mientras que para valores extremos se observa mayor dispersión y una tendencia a la subestimación en eventos de gran magnitud.

El análisis de residuos (Figura B.27) confirma la presencia de errores concentrados en eventos extremos, con algunos valores atípicos asociados a crecidas particulares, mientras que en condiciones normales de caudal los residuos se

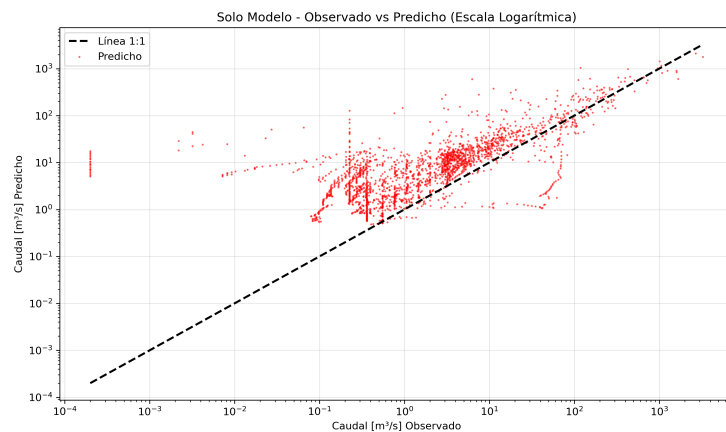


Figura B.26: Diagrama observado vs predicho en escala logarítmica para el modelo PINN-P optimizado.

mantienen acotados.

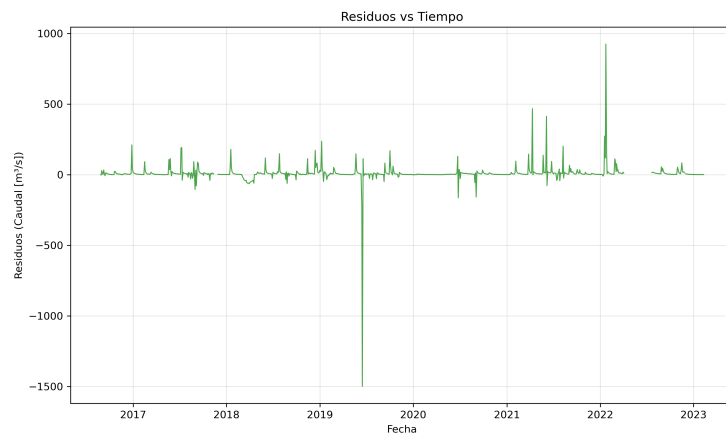


Figura B.27: Residuos del modelo PINN-P optimizado en función del tiempo.

Análisis de atribución final mediante *Integrated Gradients*

Con el objetivo de interpretar el comportamiento del modelo optimizado, se realizó un análisis de relevancia de variables utilizando el método *Integrated Gradients* (IG) sobre el conjunto de prueba.

La Figura B.28 muestra las 15 características con mayor atribución promedio (incluyendo el signo). A diferencia de los análisis exploratorios realizados en etapas previas —donde la radiación solar y la velocidad del viento presenta-

ban las mayores contribuciones—, en el modelo final optimizado se observa un cambio significativo en el patrón de relevancia.

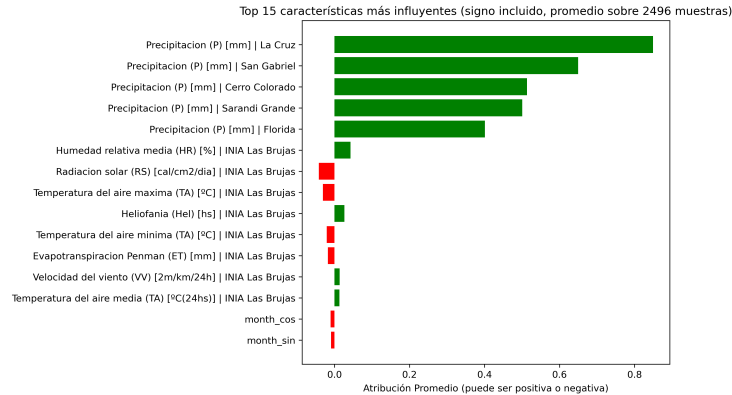


Figura B.28: Top 15 características más influyentes según *Integrated Gradients* en el conjunto de prueba (modelo PINN-P optimizado).

En particular, las precipitaciones contemporáneas por estación (La Cruz, San Gabriel, Cerro Colorado, Sarandí Grande y Florida) pasan a concentrar la mayor parte de la atribución positiva del modelo. Este resultado indica que, una vez optimizada la arquitectura y ajustados los hiperparámetros, la red prioriza la información pluviométrica directa como principal determinante del caudal simulado.

Las variables meteorológicas asociadas a procesos energéticos (radiación solar, heliofania, evapotranspiración) presentan contribuciones de menor magnitud, mientras que las variables térmicas y temporales (*month_sin*, *month_cos*) exhiben una influencia marginal.

Este cambio respecto a los análisis iniciales sugiere que el proceso de optimización arquitectónica modifica la forma en que la red distribuye la importancia entre predictores. Durante la fase exploratoria, la arquitectura menos especializada parecía apoyarse con mayor intensidad en variables radiativas y de viento. En cambio, el modelo final optimizado muestra un comportamiento más alineado con la interpretación hidrológica clásica, donde la precipitación de las estaciones constituye el principal forzante de la precipitación agregada predecida.

Costo computacional

El entrenamiento completo del modelo optimizado demandó 265.83 segundos en un total de 23 épocas en el entorno de cómputo utilizado. Por su parte, el tiempo de inferencia fue de 5.22 segundos, según los registros del sistema.

B.4. PINN-P RNN

Esta sección presenta el desarrollo detallado del modelo PINN-P con arquitectura RNN, incluyendo todos los experimentos realizados durante el proceso de optimización.

El cuerpo principal (Sección 6.7) resume únicamente los resultados del modelo base y del modelo final optimizado. Aquí se documentan las etapas intermedias que condujeron a la configuración seleccionada.

B.4.1. Modelo base

El primer experimento en el marco del enfoque PINN-P con arquitectura recurrente consistió en entrenar una red LSTM bajo la formulación **PINN-P**, es decir, haciendo que la red prediga exclusivamente la precipitación agregada diaria \hat{P}_{agg} , mientras que la evapotranspiración se mantiene fija a partir de los valores observados E_{obs} . La predicción final de caudal no es producida directamente por la red, sino que emerge del acoplamiento físico con GR4J, evaluando el caudal simulado $Q_{\text{GR4J}}(\hat{P}_{\text{agg}}, E_{\text{obs}}; \theta)$ y comparándolo con el caudal observado mediante el término físico de la función de pérdida.

A diferencia de la arquitectura MLP utilizada en la sección anterior, la red LSTM procesa secuencias temporales de longitud configurable (*sequence length*), lo que le permite capturar dependencias temporales de forma implícita a través de su estado oculto. La configuración base del modelo empleó una secuencia de longitud 1 (equivalente a una ventana unitaria), dos capas LSTM con 64 unidades ocultas, y un módulo MLP posterior de tres capas (128→64→32) con activación ReLU.

Dado que el conjunto de prueba se reserva exclusivamente para la evaluación final del modelo, el desempeño en esta etapa se reporta sobre el **conjunto de validación**, utilizando MAE, RMSE, NSE y PBIAS calculadas sobre el caudal simulado por GR4J (Tabla B.26).

Tabla B.26: Desempeño del modelo base PINN-P (RNN) sobre el conjunto de validación.

NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
0.71	0.13	15.89	57.25

La Figura B.29 muestra una reducción pronunciada de ambas curvas durante las primeras épocas: la pérdida de entrenamiento cae desde valores cercanos a 8000 hasta aproximadamente 4000 en apenas tres épocas, mientras que la pérdida de validación desciende desde ~4600 hasta su mínimo global de 3017.96 en la **época 3**. A partir de ese punto, la pérdida de entrenamiento continúa disminuyendo de forma gradual —alcanzando valores en torno a 2800–3100—, mientras que la pérdida de validación se incrementa y presenta oscilaciones

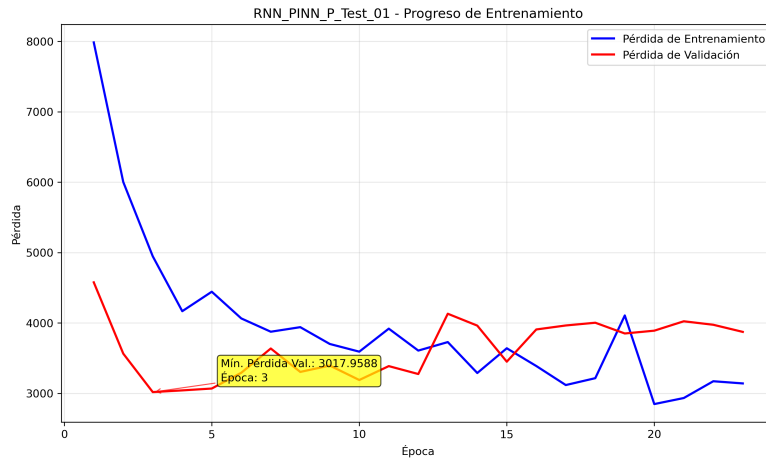


Figura B.29: Evolución de la función de pérdida (entrenamiento y validación) para el modelo base PINN-P (RNN).

marcadas en un rango de 3 000–4 100, sin lograr estabilizarse ni retornar a valores cercanos al mínimo alcanzado.

Este patrón es indicativo de una aparición temprana de *overfitting*, incluso más tempranamente que la observada en la versión MLP del modelo base (donde el mínimo de validación se alcanzaba en la época 6). La separación progresiva entre ambas curvas confirma que el modelo comienza a especializarse rápidamente en el conjunto de entrenamiento sin traducir ese ajuste en mejoras sostenidas sobre datos no vistos. Adicionalmente, las oscilaciones pronunciadas en la pérdida de validación sugieren una sensibilidad del modelo a las características específicas de los *batches* y a la configuración base de hiperparámetros (tasa de aprendizaje, *dropout*), aspectos que justifican la etapa posterior de optimización mediante Optuna.

En el contexto PINN, este comportamiento resulta relevante porque indica que, aun con el acoplamiento físico con GR4J, la configuración base de la LSTM no logra mantener una convergencia estable en validación, lo que refuerza la necesidad de ajustar tanto la arquitectura como el balance entre los términos de la función de pérdida en las etapas posteriores.

En términos cuantitativos, el modelo base RNN presenta un desempeño superior al observado en la configuración base del modelo MLP ($NSE = 0.66$, $RMSE = 62.35 \text{ m}^3/s$). Con $NSE = 0.71$, la red con LSTM explica aproximadamente el 71 % de la variabilidad observada del caudal en el conjunto de validación, lo que representa una mejora de 5 puntos porcentuales respecto a la versión MLP.

El error cuadrático medio ($RMSE = 57.25 \text{ m}^3/s$) se reduce en más de $5 \text{ m}^3/s$, mientras que el sesgo porcentual ($PBIAS = 0.13 \%$) resulta prácticamente nulo, indicando que el modelo no tiene una tendencia sistemática de

sobreestimación ni subestimación del volumen total simulado. Este valor contrasta notablemente con el sesgo de 13.30 % observado en el modelo base MLP, sugiriendo que la arquitectura recurrente logra un mejor balance volumétrico entre el caudal simulado y el observado.

En conjunto, estos resultados muestran que la configuración base de la PINN-P con arquitectura RNN resulta más efectiva que su contraparte MLP, tanto en eficiencia como en control del sesgo. No obstante, al igual que en el caso MLP, se procedió al desarrollo de las etapas posteriores de ajuste, orientadas a evaluar la incorporación de variables, la selección de la ventana temporal y el balance entre los términos de la función de pérdida.

B.4.2. Evaluación de la incorporación de variables

En esta etapa se analiza el impacto de distintas configuraciones de variables de entrada sobre el desempeño del modelo PINN-P basado en arquitectura RNN. Tal como se detalló en la metodología, el objetivo es determinar qué combinación de variables mejora la capacidad del sistema acoplado (red neuronal + GR4J) para reproducir el caudal observado en el conjunto de validación.

Al igual que en la versión MLP, en la formulación PINN-P la red neuronal no predice directamente el caudal, sino la precipitación agregada diaria \hat{P}_{agg} , que luego es utilizada como forzante del modelo GR4J junto con la evapotranspiración observada. En consecuencia, cualquier modificación en las variables de entrada puede afectar tanto la calidad de la estimación de \hat{P}_{agg} como la coherencia hidrológica del caudal simulado.

Las métricas reportadas en esta sección se calculan sobre el conjunto de validación y corresponden al caudal generado por GR4J, evaluado mediante MAE, RMSE, NSE y PBIAS.

Variables temporales

En primer lugar, se evaluó el efecto de incorporar variables temporales representadas de forma cíclica mediante transformaciones seno y coseno del *mes* y del *día*. La inclusión de estas variables busca capturar patrones estacionales que puedan influir en la estimación de la precipitación agregada y, en consecuencia, en la simulación del caudal a través de GR4J.

Se compararon cuatro configuraciones: (i) sin variables temporales, (ii) incorporación únicamente del día, (iii) incorporación únicamente del mes, y (iv) incorporación conjunta de día y mes. Los resultados obtenidos se presentan en la Tabla B.27.

Los resultados muestran que la incorporación de variables temporales no sólo no mejora el desempeño del modelo, sino que produce un deterioro consistente en todas las métricas evaluadas. La configuración sin variables temporales presenta el mayor valor de eficiencia ($NSE = 0.71$), el menor error ($RMSE = 57.25$ m³/s, $MAE = 15.89$ m³/s) y, notablemente, un sesgo volumétrico prácticamente nulo ($PBIAS = 0.13$ %).

Tabla B.27: Comparación de desempeño para distintas combinaciones de variables temporales sobre el conjunto de validación (PINN-P RNN).

Configuración	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Ninguna	0.71	0.13	15.89	57.25
día	0.69	-8.04	17.14	59.06
mes	0.69	-27.38	18.03	59.30
día, mes	0.67	-22.97	17.84	61.04

Desde una perspectiva hidrológica, la presencia de un sesgo marcado al incorporar variables de fecha sugiere que estas introducen relaciones artificiales que interfieren con la estimación de la precipitación agregada, degradando la coherencia física del sistema acoplado. Las variables hidrológicas y meteorológicas incluidas en las secuencias de entrada ya contienen implícitamente la información estacional relevante (ciclos de precipitación, temperatura, radiación), y su procesamiento secuencial por la LSTM es suficiente para capturar estos patrones.

En consecuencia, y considerando estos resultados y la coherencia del sistema acoplado, se decidió **no incorporar variables temporales** en las etapas posteriores del modelado con arquitectura RNN.

B.4.3. Selección de la ventana temporal (*sequence length*)

Un aspecto clave de la arquitectura LSTM es la longitud de la secuencia temporal de entrada (*sequence length*), que determina el horizonte de memoria explícita disponible para la red. A diferencia de los *lags* individuales evaluados en la versión MLP donde cada valor retardado se incorpora como una variable adicional, en la formulación recurrente la ventana temporal define el número de pasos consecutivos procesados por la LSTM, permitiendo capturar dependencias temporales.

Se evaluaron configuraciones con ventanas de 1 a 14 días, manteniendo constante el conjunto de variables y la arquitectura base del modelo (sin variables temporales, según la decisión adoptada en la etapa anterior). Las métricas se calcularon sobre el conjunto de validación, evaluando el caudal simulado por GR4J.

Los resultados evidenciados en la Tabla B.28 muestran que la ventana de 1 día alcanza el mejor desempeño global, con el mayor *NSE* (0.71), el menor *RMSE* (57.25 m³/s) y un sesgo prácticamente nulo (*PBIAS* = 0.13 %). A medida que se incrementa la longitud de la secuencia, el desempeño tiende a deteriorarse, con variaciones significativas tanto en eficiencia como en sesgo volumétrico.

Sin embargo, la selección de una ventana unitaria no resulta metodológicamente adecuada para una arquitectura recurrente: con una secuencia de longitud 1, la LSTM no dispone de pasos temporales previos sobre los cuales operar su

Tabla B.28: Comparación de desempeño según longitud de la ventana temporal (PINN-P RNN).

Ventana (días)	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
1	0.71	0.13	15.89	57.25
2	0.62	5.47	18.48	65.61
3	0.65	-17.78	18.16	63.30
4	0.70	-7.96	16.47	58.33
5	0.67	-20.55	18.05	61.09
6	0.68	8.15	16.48	60.62
7	0.67	-2.53	17.02	61.62
8	0.66	9.23	16.56	62.59
9	0.64	10.77	16.54	64.02
10	0.68	0.26	17.01	60.35
11	0.66	5.61	17.04	62.52
12	0.70	-14.20	17.00	58.50
13	0.67	4.63	16.43	61.11
14	0.67	12.25	16.13	61.76

mecanismo de memoria, por lo que su comportamiento se reduce al de una red *feedforward* con una transformación lineal adicional. En este escenario, el uso de una arquitectura RNN no se justifica frente a un modelo MLP convencional, ya que la capacidad de captura de dependencias temporales queda anulada.

Entre las ventanas de mayor longitud, la configuración de **4 días** presenta el mejor equilibrio entre eficiencia y control del sesgo, con $NSE = 0.70$, $RMSE = 58.33 \text{ m}^3/\text{s}$ y un sesgo moderado ($PBIAS = -7.96 \%$) que se mantiene dentro del rango aceptable ($\pm 10 \%$). La ventana de 12 días también alcanza un NSE comparable (0.70), pero con un sesgo considerable ($PBIAS = -14.20 \%$), lo que la descarta como opción preferente.

En consecuencia, considerando el equilibrio entre desempeño, control del sesgo y la capacidad de la red para aprovechar información temporal en las etapas posteriores de optimización, se decidió **adoptar una ventana temporal de 4 días** para las etapas siguientes del modelado.

B.4.4. Depuración inicial por correlación

Siguiendo el mismo procedimiento aplicado al modelo MLP, se realizó un análisis de correlación entre las variables meteorológicas con el objetivo de identificar posibles redundancias y evaluar la factibilidad de reducir la dimensionalidad del modelo sin deteriorar su desempeño.

Tal como se observó en las formulaciones previas, las variables térmicas (temperatura máxima, media y mínima) presentan correlaciones elevadas entre sí, al igual que las variables asociadas al balance radiativo (radiación solar, heliofanía y evapotranspiración). Se diseñaron distintos escenarios de reducción

progresiva para evaluar el impacto de eliminar variables redundantes.

En todos los casos, el modelo fue reentrenado completamente bajo la formulación PINN-P con arquitectura RNN, manteniendo fija la ventana temporal de 4 días y los pesos de la función de pérdida en $w_{\text{data}} = 1$ y $w_{\text{phys}} = 1$. Las métricas se calcularon sobre el conjunto de validación, evaluando el caudal simulado por GR4J.

Evaluación combinada de reducción de variables correlacionadas

Las configuraciones evaluadas fueron las siguientes:

- Modelo base (todas las variables, 12 *features*).
- Sólo temperatura media (sin temperaturas máxima y mínima, 10 *features*).
- Sin radiación solar (11 *features*).
- Sin heliofanía (11 *features*).
- Sólo temperatura media y sin radiación solar (9 *features*).

Tabla B.29: Resultados de desempeño para configuraciones de reducción de variables correlacionadas (PINN-P RNN).

Configuración	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Modelo base	0.70	-7.96	16.47	58.33
Sin temp. máx. y mín.	0.67	-8.55	17.07	61.55
Sin radiación solar	0.66	-13.23	18.26	62.04
Sin heliofanía	0.67	-6.19	17.49	61.42
Sin temp. máx./mín. ni RS	0.66	7.08	16.56	61.88

Los resultados (Tabla B.29) muestran que ninguna de las configuraciones reducidas logra igualar el desempeño del modelo base. Todas las exclusiones producen un deterioro del *NSE* del orden de 3 a 4 puntos porcentuales, junto con incrementos apreciables del *RMSE* (entre 3 y 4 m³/s).

El modelo base mantiene el mayor valor de eficiencia (*NSE* = 0.70) junto con el menor *RMSE* (58.33 m³/s), lo que indica que la inclusión completa del conjunto de variables permite una mejor reproducción de la variabilidad del caudal en validación. Si bien su sesgo porcentual (*PBIAS* = -7.96 %) no es el menor en valor absoluto, se mantiene dentro del umbral aceptable de ±10 %.

En particular, la eliminación de la radiación solar produce el mayor deterioro, tanto en eficiencia (*NSE* = 0.66) como en sesgo (*PBIAS* = -13.23 %),

lo que confirma la relevancia de esta variable en la estimación de la precipitación agregada dentro del sistema acoplado. La eliminación de las temperaturas extremas también degrada consistentemente el desempeño, lo que indica que la información térmica complementaria aporta información relevante para la red LSTM.

En consecuencia, y dado que ninguna exclusión produce mejoras en desempeño que justifiquen la pérdida de información potencial, se decidió **mantener el modelo base con el conjunto completo de 12 variables**, priorizando el desempeño global del sistema acoplado.

B.4.5. Análisis de relevancia mediante *Integrated Gradients*

Una vez definido el conjunto de variables a partir del análisis de correlación, se evaluó la relevancia individual de cada característica mediante el método de *Integrated Gradients* (IG). Este enfoque permite cuantificar la contribución promedio de cada variable a la predicción del modelo, proporcionando una medida interpretativa consistente con la estructura interna de la red.

El cálculo se realizó por separado sobre los subconjuntos de entrenamiento y validación. La Figura B.30 presenta las variables con mayor atribución promedio (incluyendo el signo) en ambos subconjuntos.

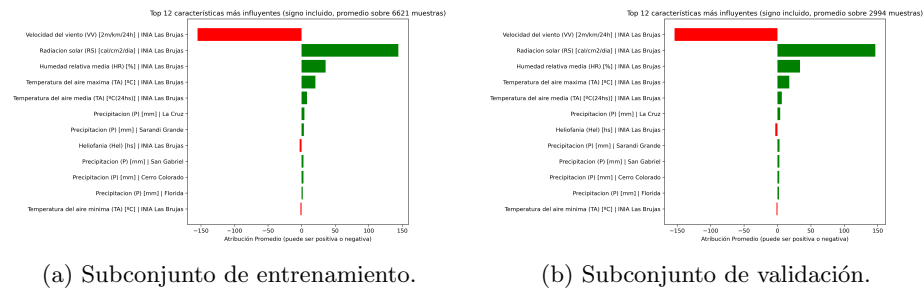


Figura B.30: Importancia relativa de variables según *Integrated Gradients* (IG) en los conjuntos de entrenamiento y validación (PINN-P RNN).

Los resultados muestran una estructura de relevancia altamente consistente entre entrenamiento y validación, lo que indica que el modelo aprende patrones estables y no dependientes del subconjunto utilizado. En ambos casos, la distribución de atribuciones presenta una polarización marcada entre dos variables dominantes: la **velocidad del viento (VV)** aparece como la característica con mayor atribución negativa (en torno a -150), mientras que la **radiación solar (RS)** concentra la mayor atribución positiva (aproximadamente $+150$). La magnitud de estas dos variables supera ampliamente al resto, evidenciando que el modelo LSTM asigna un peso desproporcionado a estas señales para la estimación de la precipitación agregada.

En un segundo nivel de relevancia se ubican la **humedad relativa (HR)**, la **temperatura máxima** y la **temperatura media**, todas con atribuciones

positivas moderadas. Las precipitaciones contemporáneas por estación (La Cruz, Sarandí Grande, San Gabriel, Cerro Colorado, Florida) presentan contribuciones positivas de menor magnitud, mientras que la heliofanía exhibe una atribución negativa leve. La **temperatura mínima** aparece como la variable con menor atribución promedio en ambas particiones.

Este patrón de relevancia es altamente consistente con el observado en la versión MLP del modelo base PINN-P, donde la radiación solar y la velocidad del viento también emergían como las características más influyentes. La similitud entre ambas arquitecturas sugiere que, en la formulación PINN-P con la configuración base, el sistema acoplado se apoya preferentemente en variables asociadas al balance energético y aerodinámico para estimar la precipitación agregada, independientemente de la estructura de la red.

Si bien la temperatura del aire mínima presenta la menor atribución promedio, como se verificó en el análisis de correlación (Sección B.4.4), la exclusión conjunta de las temperaturas máxima y mínima produjo un deterioro del desempeño del modelo (NSE de 0.70 a 0.67). Dado que el modelo opera con sólo 12 variables de entrada, no existe un problema de excesiva dimensionalidad, y la eliminación de una variable con baja atribución individual no garantiza una mejora en el rendimiento global.

En consecuencia, se decidió **mantener el conjunto completo de 12 variables sin modificaciones** para las etapas posteriores del modelado, dando por finalizado el proceso iterativo de depuración.

Conjunto final de variables

El conjunto final de variables para el modelo PINN-P RNN fue definido a partir de un proceso secuencial que incluyó: (i) evaluación de la incorporación de variables temporales, (ii) selección de la ventana temporal, (iii) análisis de correlación entre variables meteorológicas, y (iv) análisis de relevancia mediante *Integrated Gradients* (IG).

En una primera etapa se evaluó la incorporación de información estacional, observándose que —a diferencia de la versión MLP— la inclusión de variables temporales degradaba consistentemente el desempeño del modelo RNN. Este resultado se atribuye a la capacidad intrínseca de la arquitectura recurrente para capturar patrones estacionales a través del procesamiento secuencial, lo que hace redundante la incorporación explícita de señales temporales.

Posteriormente se evaluaron distintas longitudes de ventana temporal, seleccionándose una ventana de 4 días como la configuración que ofrece el mejor equilibrio entre eficiencia y control del sesgo dentro de un horizonte temporalmente coherente con los procesos hidrológicos de la cuenca.

El análisis de correlación permitió explorar configuraciones reducidas; sin embargo, todas las exclusiones produjeron deterioros del desempeño del orden de 3–4 puntos porcentuales de NSE , sin aportar mejoras compensatorias. El análisis de IG confirmó la estructura de relevancia observada, sin identificar variables cuya eliminación produjera mejoras sostenidas.

En función de estos resultados, el conjunto final de variables quedó conformado por:

- **Precipitaciones contemporáneas por estación** (sin *lags*).
- **Radiación solar (RS), heliofanía (Hel)**.
- **Temperatura del aire** (media, máxima y mínima).
- **Humedad relativa (HR) y velocidad del viento (VV)**.
- **Sin variables temporales**.

Esta configuración se procesa mediante una ventana temporal de 4 días, aprovechando la capacidad recurrente de la LSTM para capturar la dinámica temporal sin necesidad de incorporar variables temporales explícitas. Sobre este conjunto se desarrollaron las etapas posteriores de análisis del balance entre los términos de la función de pérdida y la optimización de hiperparámetros.

B.4.6. Análisis del balance entre términos de la función de pérdida

Uno de los aspectos distintivos de las redes neuronales informadas por la física es la presencia de múltiples términos en la función de pérdida, cuya ponderación relativa determina el equilibrio entre ajuste empírico y coherencia física. En el caso de PINN-P, la función de pérdida combina el término de datos (ajuste de \hat{P}_{agg}) y el término físico (ajuste del caudal simulado por GR4J), ponderados mediante los coeficientes w_{data} y w_{phys} .

Siguiendo la estrategia metodológica previamente descrita, se evaluaron distintas combinaciones de pesos manteniendo fijo el conjunto de variables y la ventana temporal seleccionados. Se analizaron 13 configuraciones que cubren relaciones comprendidas hasta un máximo de 1:4 en ambos sentidos, incluyendo los casos extremos $w_{data} = 0$ (sólo restricción física) y $w_{phys} = 0$ (sólo supervisión por datos).

La Tabla B.30 presenta las métricas obtenidas en el conjunto de validación para cada combinación evaluada.

A diferencia de lo observado en la versión MLP, donde el desempeño resultaba notablemente estable frente a variaciones en los pesos, en la arquitectura RNN se aprecia una mayor sensibilidad a la ponderación relativa de los términos de pérdida. Las diferencias en *NSE* abarcan un rango más amplio (0.65–0.72), y el sesgo varía considerablemente entre configuraciones.

El caso extremo sin restricción física ($w_{data} = 1$, $w_{phys} = 0$) alcanza el mayor *NSE* (0.72) y el menor *RMSE* (56.96 m³/s), lo que indica que la supervisión directa por datos resulta más efectiva para maximizar la eficiencia estadística. No obstante, este resultado carece de interés práctico en el marco PINN, ya que la eliminación completa del término físico desnaturaliza la formulación informada por la física.

Tabla B.30: Desempeño del modelo PINN-P (RNN) para distintas combinaciones de pesos en la función de pérdida.

$w_{\text{data}} - w_{\text{phys}}$	NSE	PBIAS [%]	MAE [m^3/s]	RMSE [m^3/s]
0.00 – 1.00	0.65	17.07	16.85	63.59
1.00 – 0.00	0.72	-9.79	16.14	56.96
0.25 – 1.00	0.68	15.94	16.14	60.51
0.25 – 0.75	0.67	8.91	16.72	61.14
0.50 – 1.00	0.70	9.36	15.59	58.50
0.50 – 0.75	0.70	9.69	15.92	58.88
0.75 – 1.00	0.71	1.42	16.09	57.64
1.00 – 1.00	0.70	-7.96	16.47	58.33
1.00 – 0.75	0.69	-7.37	16.95	59.42
1.00 – 0.50	0.69	3.87	16.15	59.56
0.75 – 0.50	0.70	-4.04	16.36	58.87
0.75 – 0.25	0.69	-11.09	17.35	59.31
1.00 – 0.25	0.71	-0.36	15.98	57.60

Entre las configuraciones que preservan la restricción física, se destaca la combinación $w_{\text{data}} = 1.00$, $w_{\text{phys}} = 0.25$, que alcanza un $NSE = 0.71$ y un $RMSE = 57.60 \text{ m}^3/\text{s}$ (valores prácticamente equivalentes al caso sin física) mientras que presenta un sesgo excepcionalmente bajo ($PBIAS = -0.36 \%$), el mejor entre todas las configuraciones evaluadas. Esta configuración representa un equilibrio óptimo: mantiene la coherencia física del sistema acoplado sin sacrificar la eficiencia predictiva.

La configuración balanceada $w_{\text{data}} = 1$, $w_{\text{phys}} = 1$ presenta un desempeño competitivo ($NSE = 0.70$), pero con un sesgo moderado ($PBIAS = -7.96 \%$). La combinación $w_{\text{data}} = 0.75$, $w_{\text{phys}} = 1.00$ alcanza un NSE ligeramente superior (0.71) con un sesgo muy bajo ($PBIAS = 1.42 \%$), pero con un MAE más elevado.

En conjunto, los resultados indican que el modelo PINN-P con arquitectura RNN se beneficia de una predominancia moderada del término de datos sobre el término físico. La restricción impuesta por GR4J contribuye de forma más sutil al control del sesgo volumétrico cuando opera como un regulador secundario, sin dominar la optimización.

En consecuencia, considerando simultáneamente la eficiencia estadística, el control del sesgo y la preservación de la coherencia física del sistema acoplado, se decidió **adoptar la configuración** $w_{\text{data}} = 1.00$ y $w_{\text{phys}} = 0.25$ para las etapas posteriores de optimización y evaluación final.

Esta elección se distingue de la adoptada en la versión MLP ($w_{\text{data}} = 1$, $w_{\text{phys}} = 1$), lo que sugiere que la interacción entre la arquitectura de la red y la ponderación de los términos de pérdida no es trivial. La mayor capacidad representacional de la LSTM parece requerir un menor peso de la restricción física para lograr un equilibrio óptimo.

B.4.7. Optimización de hiperparámetros (Optuna)

Una vez definido el conjunto final de variables, la ventana temporal de 4 días y el balance de función de pérdida en $w_{\text{data}} = 1.00$ y $w_{\text{phys}} = 0.25$, se procedió a la optimización de hiperparámetros mediante el *framework* *Optuna*.

A diferencia de la versión MLP, el espacio de búsqueda incluyó hiperparámetros específicos de la arquitectura recurrente: tipo de celda recurrente (LSTM, GRU o RNN), tamaño del estado oculto y número de capas recurrentes. Además, se optimizaron los parámetros del módulo MLP posterior (número de capas, ancho, función de activación y *dropout*) así como los parámetros de entrenamiento (tasa de aprendizaje, *weight decay* y *batch size*).

La función objetivo correspondió a la pérdida de validación (MSE), evaluada bajo la formulación compuesta propia de la PINN.

El estudio culminó con los siguientes resultados:

- **Mejor valor de pérdida de validación:** 709.51
- ***Trials* completados:** 114
- ***Trials* podados:** 16

La Figura B.31 muestra la evolución del valor objetivo a lo largo de los *trials*. Durante las primeras iteraciones se observa una fase de exploración amplia, con valores elevados y alta variabilidad. A medida que avanzó el proceso, el algoritmo concentró la búsqueda en regiones más prometedoras del espacio de hiperparámetros, logrando una reducción progresiva del mejor valor alcanzado.

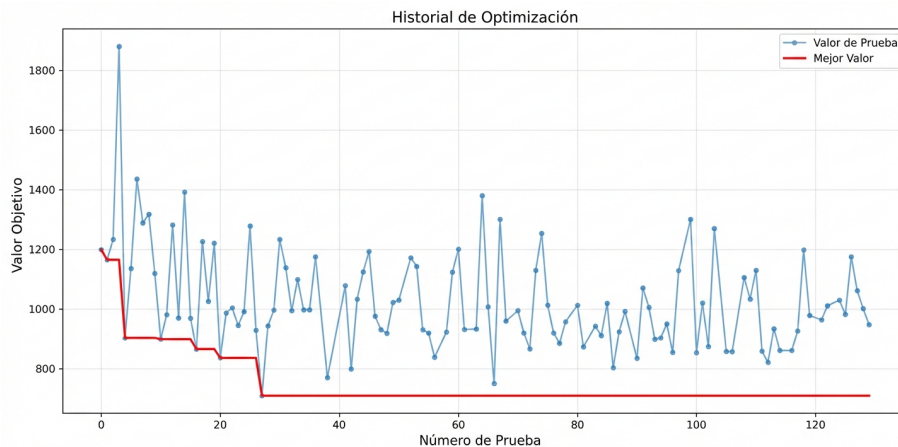


Figura B.31: Historial del proceso de optimización realizado con Optuna para PINN-P RNN. La línea azul muestra el valor de la función objetivo para cada *trial*, mientras que la línea roja indica el mejor valor obtenido hasta ese punto.

La configuración óptima se resume en la Tabla B.31.

Tabla B.31: Configuración óptima seleccionada por Optuna para PINN-P RNN.

Hiperparámetro	Valor óptimo
Tipo de celda recurrente	LSTM
Tamaño del estado oculto	70
Capas recurrentes	1
Capas MLP	1
Ancho capa MLP	390
Función de activación	gelu
Tasa de aprendizaje	0.00275
<i>Dropout</i> base	0.44
<i>Batch size</i>	16
<i>Weight decay</i>	6.98×10^{-4}

El proceso de optimización seleccionó una **celda LSTM** por sobre las alternativas GRU y RNN, lo que sugiere que el mecanismo de compuerta de olvido propio de la LSTM resulta más adecuado para la tarea de estimación de precipitación en este contexto. La arquitectura óptima presenta una estructura simplificada en el componente recurrente (**una sola capa** con 70 unidades ocultas), compensada por un módulo MLP de ancho considerable (390 unidades) con activación **gelu**.

Esta configuración contrasta con la arquitectura base utilizada en las etapas exploratorias (2 capas LSTM con 64 unidades y MLP de 3 capas), sugiriendo que una representación recurrente más compacta, seguida de una capa densa de mayor capacidad, resulta más efectiva para la tarea de estimación de precipitación dentro del sistema PINN.

La tasa de aprendizaje óptima ($\approx 2.75 \times 10^{-3}$) es considerablemente superior a la utilizada en la configuración base (10^{-4}), lo que indica que el modelo se beneficia de actualizaciones de gradiente más agresivas. El valor de *dropout* (0.44) indica una regularización fuerte, mientras que el *weight decay* (6.98×10^{-4}) aporta una penalización adicional sobre los pesos, contribuyendo conjuntamente al control del sobreajuste.

La Figura B.32 presenta las curvas de aprendizaje correspondientes a los cinco mejores *trials*. Se observa que, aunque todos convergen hacia valores similares de pérdida, existe variabilidad en la estabilidad de las trayectorias a lo largo de las épocas.

Finalmente, la Figura B.33 muestra el historial de *pruning*. El uso de poda permitió descartar configuraciones poco prometedoras en etapas tempranas, reduciendo el costo computacional total sin comprometer la calidad del resultado final.

En conjunto, la optimización mediante Optuna identificó una arquitectura que simplifica el componente recurrente y concentra la capacidad representacional en el módulo MLP posterior. Este resultado sugiere que, en la formulación PINN-P, la LSTM opera principalmente como un codificador temporal com-

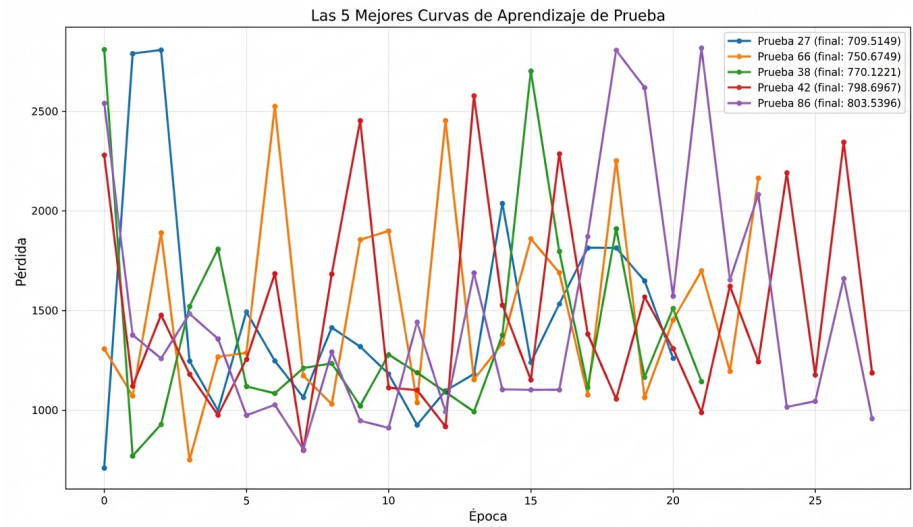


Figura B.32: Curvas de aprendizaje de los cinco mejores *trials* del proceso de optimización de Optuna (PINN-P RNN).

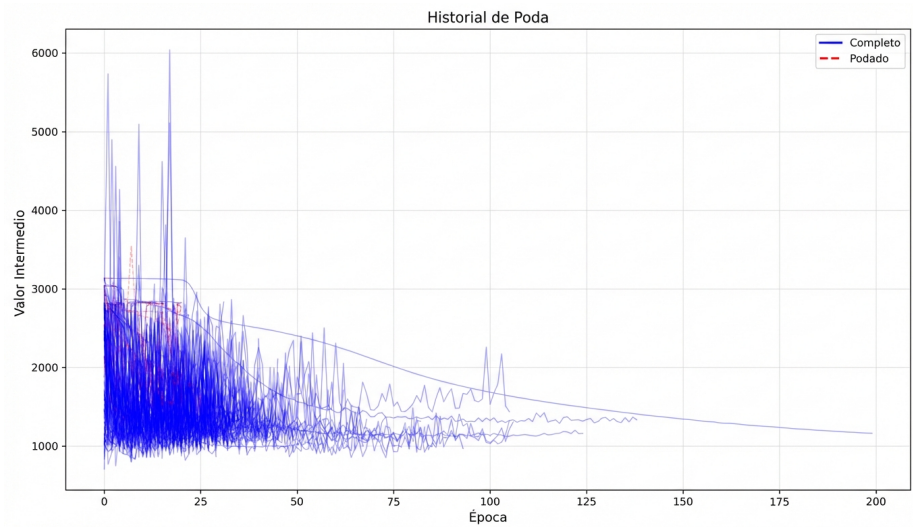


Figura B.33: Historial de *pruning* durante la optimización con Optuna (PINN-P RNN). Las trayectorias podadas se interrumpen tempranamente al no mostrar mejoras en la función objetivo.

pacto cuya salida es procesada por una capa densa grande para producir la estimación final de precipitación.

B.4.8. Evaluación extendida del modelo optimizado

El modelo final PINN-P RNN se entrenó utilizando la arquitectura óptima identificada mediante Optuna (Sección B.4.7), con el balance de pérdida $w_{\text{data}} = 1.00$ y $w_{\text{phys}} = 0.25$, una ventana temporal de 4 días, y el conjunto de variables definido en las etapas previas.

La evaluación se realizó de forma independiente sobre los conjuntos de entrenamiento, validación y prueba, preservando la partición temporal original y aplicando el mismo criterio de enmascaramiento para los registros con caudal faltante. Las métricas consideradas fueron MAE, RMSE, NSE y PBIAS. La Tabla B.32 presenta el desempeño del modelo en los distintos conjuntos.

Tabla B.32: Desempeño del modelo PINN-P optimizado RNN en los distintos conjuntos.

Conjunto	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Entrenamiento	0.53	-42.53	19.00	69.72
Validación	0.64	-6.93	18.24	64.23
Prueba	0.67	-22.64	17.85	74.33

A modo de referencia, el modelo GR4J ejecutado de forma directa sobre el mismo período de prueba obtiene $NSE = 0.69$, $RMSE = 72.36$ m³/s y $PBIAS = -16.08$ %.

Análisis del desempeño

En el conjunto de prueba, el modelo PINN-P RNN optimizado alcanza un $NSE = 0.67$ y un $RMSE = 74.33$ m³/s. Si bien la eficiencia estadística se sitúa en un rango adecuado, el modelo presenta un sesgo negativo considerable ($PBIAS = -22.64$ %), indicando una tendencia a la sobreestimación del volumen total de caudal.

En el conjunto de validación se obtiene un $NSE = 0.64$ con un $RMSE = 64.23$ m³/s y un sesgo moderado ($PBIAS = -6.93$ %), mientras que en entrenamiento el desempeño es marcadamente inferior ($NSE = 0.53$, $PBIAS = -42.53$ %). Este patrón de desempeño inferior en entrenamiento que en validación y prueba es consistente con el observado en la versión MLP del modelo PINN-P, y se explica por la composición de la función de pérdida donde la optimización sobre una combinación ponderada de pérdida de datos y pérdida física no garantiza un mínimo estricto del error directo sobre caudal en el conjunto de entrenamiento.

En comparación con la versión MLP del modelo PINN-P optimizado ($NSE = 0.74$, $PBIAS = -35.55$ % en prueba), la RNN presenta una eficiencia inferior

pero un sesgo más controlado. No obstante, ambas formulaciones comparten la tendencia a la sobreestimación en el conjunto de prueba, patrón que contrasta con los sesgos más moderados observados durante las etapas de validación.

Es relevante señalar que el modelo PINN-P RNN optimizado no supera el desempeño del modelo GR4J directo en el conjunto de prueba ($NSE = 0.69$ vs 0.67), lo que indica que, en este caso particular, la adición de la capa de estimación de precipitación mediante la red LSTM no aporta una mejora neta respecto al modelo conceptual alimentado directamente con la precipitación observada.

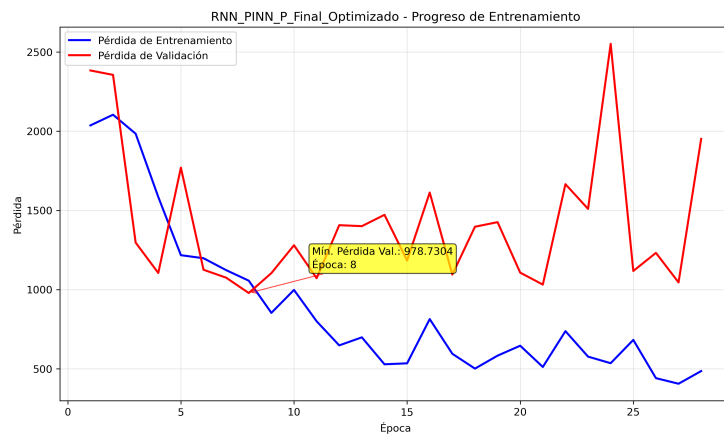


Figura B.34: Evolución de la pérdida de entrenamiento y validación del modelo PINN-P optimizado (RNN).

La Figura B.34 muestra la evolución de la pérdida durante el entrenamiento. Se observa que ambas curvas descienden de forma pronunciada durante las primeras épocas, con la pérdida de entrenamiento cayendo desde ~ 2100 hasta valores en torno a 1000 en las primeras 8 épocas. La pérdida de validación alcanza su mínimo global de 978.73 en la **época 8**, lo que representa una convergencia más tardía que la observada tanto en el modelo base LSTM (época 3) como en el modelo PINN-P MLP optimizado (época 3), lo que indica que la arquitectura optimizada logra un proceso de aprendizaje más gradual.

No obstante, a partir de la época 8 se observa una separación progresiva entre ambas curvas: la pérdida de entrenamiento continúa disminuyendo, mientras que la pérdida de validación presenta oscilaciones pronunciadas en un rango amplio (~ 1000 – 2500), con picos particularmente marcados alrededor de las épocas 17 y 23–24. Este patrón confirma la presencia de *overfitting*, aunque con una aparición más tardía respecto al modelo base, lo que indica que la optimización de hiperparámetros mediante Optuna logró extender levemente la ventana de aprendizaje.

Análisis gráfico del ajuste

La Figura B.35 presenta la comparación temporal entre caudal observado y predicho en el conjunto de prueba.

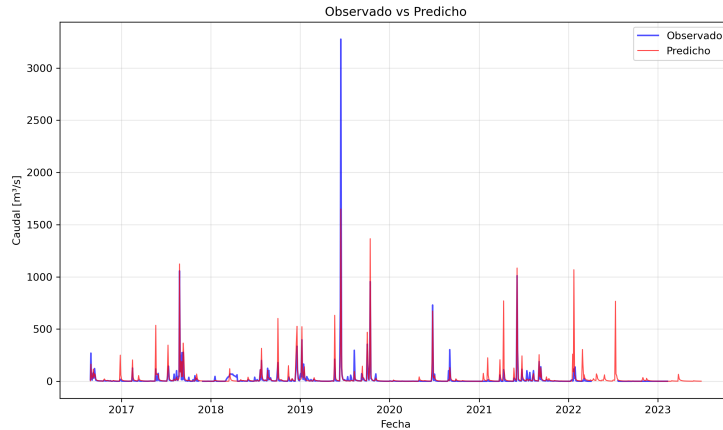


Figura B.35: Serie temporal de caudales observados y predichos por el modelo PINN-P optimizado (RNN) en el conjunto de prueba.

El modelo reproduce adecuadamente la dinámica general de la serie, capturando la ocurrencia y el momento de los principales eventos de crecida así como los períodos de estiaje. Los caudales bajos y medios son bien representados a lo largo de todo el período de prueba (2016–2023).

Sin embargo, se observa una subestimación marcada en los eventos extremos. El evento de mayor magnitud, registrado a mediados de 2019 con un caudal observado cercano a $3\,300\text{ m}^3/\text{s}$, es reproducido por el modelo con un pico de apenas $\sim 1\,400\text{ m}^3/\text{s}$, lo que representa una subestimación superior al 55 %. Otros eventos de magnitud intermedia ($\sim 1\,000\text{--}1\,100\text{ m}^3/\text{s}$ en 2017, 2021 y 2022) son mejor capturados, aunque con una tendencia general a sobreestimar los picos máximos. Esta sobreestimación sistemática en picos intermedios y bajos es consistente con el sesgo volumétrico negativo reportado ($PBIAS = -22.64\%$) y resulta visible gráficamente en la diferencia entre las curvas observada y predicha durante los episodios de crecida.

El *scatterplot* en escala logarítmica (Figura B.36) permite evaluar la alineación entre caudales observados y predichos en distintos rangos de magnitud.

Se observa una alineación razonable con la línea 1:1 en el rango medio de caudales ($\sim 10\text{--}100\text{ m}^3/\text{s}$), donde la nube de puntos se distribuye de forma relativamente simétrica en torno a la diagonal. Para caudales bajos ($< 1\text{ m}^3/\text{s}$), la dispersión aumenta considerablemente, con una tendencia del modelo a sobrepredecir (puntos por encima de la línea 1:1). En el rango de caudales altos ($> 100\text{ m}^3/\text{s}$), los puntos se distribuyen tanto por encima como por debajo de la diagonal, mientras que para los valores observados más extremos ($> 1\,000\text{ m}^3/\text{s}$) se confirma la tendencia a la subestimación, con las predicciones ubicándose

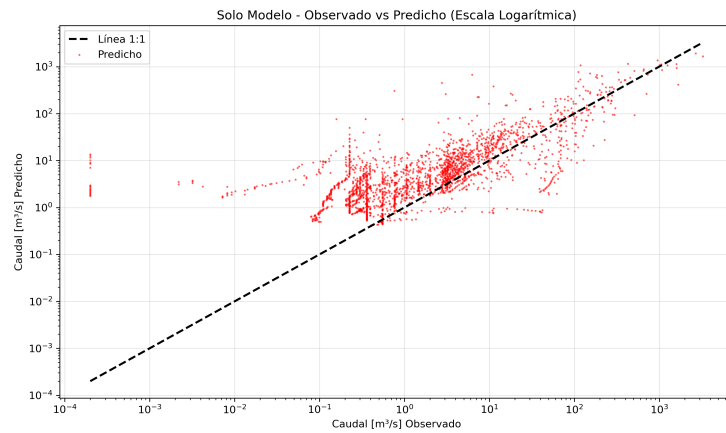


Figura B.36: Diagrama observado vs predicho en escala logarítmica para el modelo PINN-P optimizado (RNN).

por debajo de la línea 1:1. En conjunto, el diagrama refleja un ajuste aceptable en condiciones medias, pero con limitaciones en ambos extremos del rango de caudales.

El análisis de residuos (Figura B.37) permite evaluar la distribución temporal de los errores del modelo.

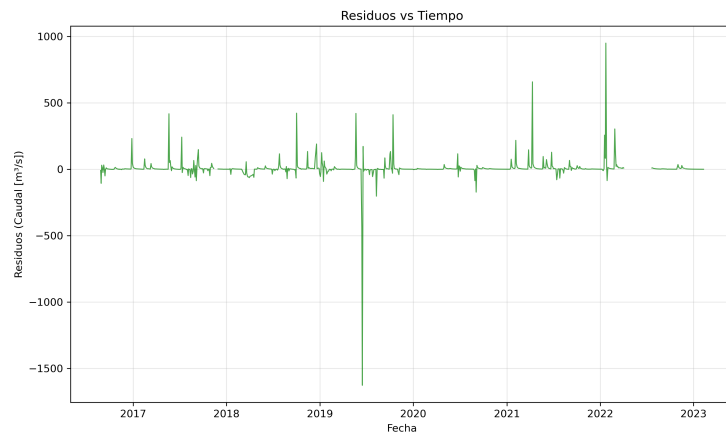


Figura B.37: Residuos del modelo PINN-P optimizado (RNN) en función del tiempo.

Los residuos se mantienen en general centrados en torno a cero durante la mayor parte del período, confirmando que en condiciones normales de caudal el modelo no presenta desviaciones sistemáticas. No obstante, se identifican valores atípicos de magnitud considerable asociados a eventos extremos. El residuo

negativo más pronunciado ($\sim -1\,600\text{ m}^3/\text{s}$) corresponde al evento de crecida de mediados de 2019, donde el modelo subestima drásticamente el pico observado. En contraste, se observan residuos positivos de menor magnitud pero recurrentes ($\sim 400\text{--}950\text{ m}^3/\text{s}$) distribuidos a lo largo del período, particularmente en 2017, 2018–2019, 2021 y 2022, asociados a eventos donde el modelo sobreestima parcialmente el caudal.

La distribución de residuos es asimétrica: el mayor valor atípico negativo supera en magnitud a los positivos. No se detectan patrones sistemáticos crecientes o decrecientes en el tiempo, lo que indica que el modelo no presenta deriva temporal en su desempeño.

Análisis de atribución final mediante *Integrated Gradients*

Con el objetivo de interpretar el comportamiento del modelo optimizado, se realizó un análisis de relevancia de variables utilizando el método *Integrated Gradients* (IG) sobre el conjunto de prueba.

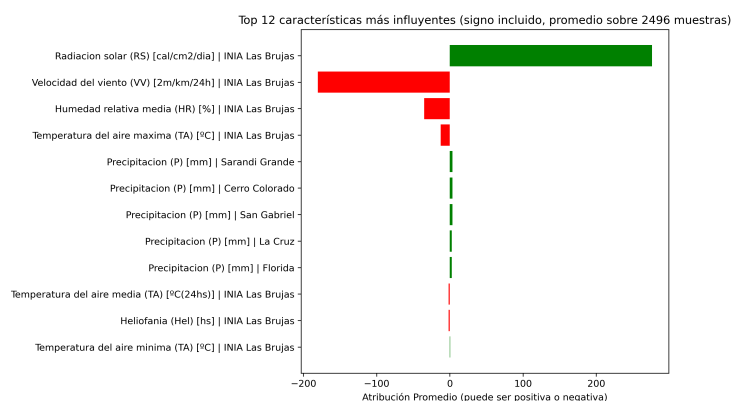


Figura B.38: Top características más influyentes según *Integrated Gradients* en el conjunto de prueba (modelo PINN-P optimizado RNN).

La Figura B.38 muestra que la **radiación solar (RS)** concentra la mayor atribución positiva ($\sim +275$), seguida por la **velocidad del viento (VV)** con la mayor atribución negativa (~ -200). En un tercer nivel, la **humedad relativa (HR)** presenta una atribución negativa moderada (~ -50), mientras que la **temperatura máxima** contribuye con una atribución positiva reducida. Las precipitaciones contemporáneas por estación (Sarandí Grande, Cerro Colorado, San Gabriel, La Cruz, Florida) presentan atribuciones positivas muy cercanas a cero, y las variables de temperatura media, heliofanía y temperatura mínima exhiben contribuciones prácticamente nulas.

Este patrón mantiene la misma estructura observada en las etapas exploratorias (Sección B.4.5), donde RS y VV ya dominaban ampliamente la distribución de relevancias, aunque con magnitudes absolutas ligeramente superiores en el modelo final (~ 275 vs ~ 150). A diferencia de lo observado en el modelo PINN-P

MLP optimizado —donde la optimización arquitectónica mediante Optuna producía un cambio significativo en el patrón de relevancia, con las precipitaciones contemporáneas pasando a concentrar la mayor atribución—, en la versión RNN el patrón permanece dominado por variables asociadas al balance energético y aerodinámico.

No obstante, debe considerarse que, en esta implementación, las *Integrated Gradients* se evaluaron únicamente sobre el primer paso temporal de la ventana de entrada. Dado que las redes recurrentes integran información a lo largo de toda la secuencia mediante su estado oculto, es posible que parte de la contribución asociada a la precipitación se encuentre distribuida en pasos temporales anteriores dentro de la ventana, los cuales no fueron descompuestos explícitamente en el análisis de atribuciones. En consecuencia, la interpretación presentada debe entenderse como una aproximación parcial a la dinámica completa de relevancia temporal del modelo.

Costo computacional

El entrenamiento completo del modelo optimizado demandó 326,23 segundos en un total de 28 épocas en el entorno de cómputo utilizado. Por su parte, el tiempo de inferencia fue de 4.07 segundos, según los registros del sistema.

B.5. PINN-E MLP

Esta sección presenta el desarrollo detallado del modelo PINN-E con arquitectura MLP, incluyendo todos los experimentos realizados durante el proceso de optimización.

El cuerpo principal (Sección 6.8) resume únicamente los resultados del modelo base y del modelo final optimizado. Aquí se documentan las etapas intermedias que condujeron a la configuración seleccionada.

B.5.1. Modelo base

A continuación se presenta la formulación **PINN-E** basada en arquitectura MLP, en la cual la red neuronal predice exclusivamente la evapotranspiración agregada diaria \hat{E}_{agg} , mientras que la precipitación se mantiene fija a partir de los valores observados P_{obs} . En particular, se dispone de registros de precipitación en distintas estaciones dentro de la cuenca, y la señal utilizada como forzante del modelo GR4J corresponde a la precipitación agregada espacialmente mediante el método de polígonos de Thiessen, tal como se describió en la metodología de trabajo.

La predicción final de caudal no es producida directamente por la red, sino que emerge del acoplamiento físico con GR4J, evaluando el caudal simulado $Q_{GR4J}(P_{Thiessen}, \hat{E}_{agg}; \theta)$ y comparándolo con el caudal observado mediante el término físico de la función de pérdida.

Dado que el conjunto de prueba se reserva exclusivamente para la evaluación final del modelo, el desempeño en esta etapa se reporta sobre el **conjunto de**

validación, utilizando MAE, RMSE, NSE y PBIAS calculadas sobre el caudal simulado por GR4J.

Tabla B.33: Desempeño del modelo base PINN-E (MLP) sobre el conjunto de validación.

NSE	PBIAS [%]	MAE [m^3/s]	RMSE [m^3/s]
0.71	2.21	15.21	57.86

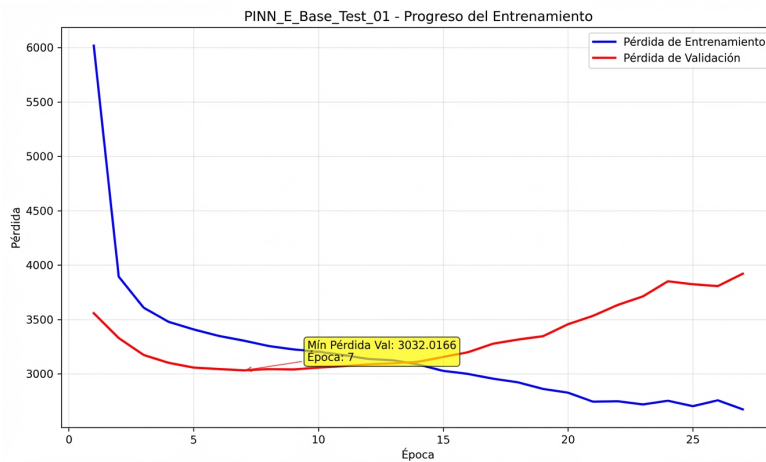


Figura B.39: Evolución de la función de pérdida (entrenamiento y validación) para el modelo base PINN-E (MLP).

La Figura B.39 muestra la evolución de la pérdida durante el entrenamiento del modelo. Se observa una reducción pronunciada tanto en entrenamiento como en validación durante las primeras épocas, seguida por una estabilización temprana del desempeño en validación. En particular, la pérdida de validación alcanza su mínimo en la época 7, a partir de la cual se aprecia una separación progresiva entre ambas curvas: la pérdida de entrenamiento continúa disminuyendo, mientras que la pérdida de validación comienza a incrementarse de forma sostenida.

Este patrón es consistente con la aparición temprana de *overfitting*, indicando que el modelo comienza a especializarse en el conjunto de entrenamiento sin traducir ese ajuste en mejoras sostenidas sobre datos no vistos. En el contexto PINN, este comportamiento resulta relevante porque sugiere que, aun con el acoplamiento físico con GR4J, el equilibrio inicial entre los términos de pérdida y la configuración base del modelo no alcanzan para sostener una mejora estable en generalización, lo cual justifica las etapas posteriores de ajuste (incorporación de variables, selección de *lags* y análisis del balance $w_{\text{data}} : w_{\text{phys}}$).

En términos cuantitativos, el modelo base presenta un desempeño razonable

sobre validación (Tabla B.33), con $NSE = 0.71$, lo que indica que el sistema acoplado explica aproximadamente el 71 % de la variabilidad observada del caudal en este subconjunto. El error cuadrático medio ($RMSE = 57.86 \text{ m}^3/\text{s}$) se mantiene en un rango moderado, mientras que el sesgo porcentual ($PBIAS = 2.21 \%$) evidencia una tendencia leve a la subestimación del volumen total simulado, considerablemente menor que la observada en configuraciones preliminares del enfoque PINN-P en su versión MLP.

En conjunto, estos resultados muestran que la configuración base de la PINN-E es capaz de capturar una fracción importante de la dinámica de caudales a través del modelo GR4J forzado con \hat{E}_{agg} .

B.5.2. Evaluación de la incorporación de variables

Variables temporales

En primer lugar, se evaluó el efecto de incorporar variables temporales representadas de forma cíclica mediante transformaciones seno y coseno del *mes* y del *día*. La inclusión de estas variables busca capturar patrones estacionales que puedan influir en la estimación de la evapotranspiración agregada y, en consecuencia, en la simulación del caudal a través de GR4J.

En la formulación PINN-E, la red neuronal predice \hat{E}_{agg} , mientras que la precipitación agregada espacialmente mediante polígonos de Thiessen se mantiene fija como forzante del modelo hidrológico. Por lo tanto, la incorporación de información temporal puede contribuir a mejorar la representación de la variabilidad estacional de la demanda evaporativa, proceso fuertemente controlado por la radiación, la temperatura y la duración del día.

Se compararon cuatro configuraciones: (i) sin variables temporales, (ii) incorporación únicamente del mes, (iii) incorporación únicamente del día, y (iv) incorporación conjunta de día y mes. Los resultados obtenidos se presentan en la Tabla B.34.

Tabla B.34: Comparación de desempeño para distintas combinaciones de variables temporales sobre el conjunto de validación (PINN-E MLP).

Configuración	NSE	PBIAS [%]	MAE [m^3/s]	RMSE [m^3/s]
Ninguna	0.71	2.21	15.21	57.86
día	0.71	3.61	15.11	57.78
mes	0.71	4.20	14.81	57.22
día, mes	0.71	1.33	15.14	57.40

Los resultados muestran que la incorporación de variables temporales produce mejoras leves pero consistentes respecto a la configuración sin información temporal explícita. Si bien todas las configuraciones presentan valores de eficiencia similares ($NSE \approx 0.71$), se observan diferencias en términos de error y sesgo.

La inclusión exclusiva del mes alcanza el menor $RMSE$ ($57.22 \text{ m}^3/\text{s}$) y el menor MAE ($14.81 \text{ m}^3/\text{s}$). No obstante, la incorporación conjunta de día y mes presenta el menor sesgo volumétrico ($PBIAS = 1.33 \%$), evidenciando un mejor control del balance total simulado.

Desde una perspectiva hidrológica, este comportamiento es coherente con la naturaleza del proceso modelado. A diferencia de la precipitación, la evapotranspiración presenta una marcada estacionalidad asociada a la disponibilidad energética y a la duración del día. En este contexto, tanto el mes como el día pueden actuar como reguladores temporales que estabilizan la estimación de \hat{E}_{agg} dentro del sistema acoplado.

Si bien las diferencias en eficiencia son marginales, la reducción observada en el sesgo bajo la configuración conjunta sugiere un comportamiento global más equilibrado del modelo. En consecuencia, se decidió **conservar ambas variables temporales (día y mes)** para las etapas posteriores del modelado.

Representación espacial de la precipitación observada

A diferencia de la formulación PINN-P —donde la red neuronal predice directamente la precipitación agregada y, por lo tanto, no resulta metodológicamente apropiado incorporar versiones agregadas adicionales como variables de entrada— en la formulación PINN-E la precipitación se mantiene fija y exclusivamente observada.

Esto permite evaluar distintos esquemas de agregación espacial de la precipitación medida en las estaciones de la cuenca, con el objetivo de analizar si la forma en que se construye el forzante pluviométrico afecta el desempeño del sistema acoplado (red neuronal + GR4J).

Se compararon cuatro configuraciones:

- Serie original por estación (una variable por estación).
- Promedio simple de estaciones.
- Promedio ponderado por distancia.
- Agregación mediante polígonos de Thiessen.

Las métricas se calcularon sobre el conjunto de validación, evaluando el caudal simulado por GR4J. Los resultados se presentan en la Tabla B.35.

Los resultados muestran que las diferencias entre los distintos métodos de agregación son relativamente acotadas. Todas las configuraciones presentan valores de eficiencia similares ($NSE \approx 0.71\text{--}0.72$), con variaciones moderadas en términos de error y sesgo.

El **promedio simple de estaciones** alcanza el mayor valor de eficiencia ($NSE = 0.72$) y el menor $RMSE$ ($56.62 \text{ m}^3/\text{s}$), además de reducir el MAE respecto a la configuración original por estación. Asimismo, presenta un sesgo volumétrico bajo y de signo negativo moderado ($PBIAS = -1.81 \%$).

Desde una perspectiva estructural, la configuración original incorpora una variable por estación, aumentando la dimensionalidad del modelo sin aportar

Tabla B.35: Comparación de desempeño según método de agregación espacial de la precipitación observada (PINN-E MLP).

Método	NSE	PBIAS [%]	MAE [m^3/s]	RMSE [m^3/s]
Original (por estación)	0.71	1.33	15.14	57.40
Promedio simple	0.72	-1.81	14.25	56.62
Ponderado	0.72	-2.03	15.32	56.77
Thiessen				
Ponderado por distancia	0.72	-1.25	15.27	56.88

mejoras sustanciales en desempeño. En contraste, el promedio simple resume la información pluviométrica en una única señal representativa de la cuenca, reduciendo la complejidad del modelo y favoreciendo la parsimonia.

El hecho de que los distintos métodos produzcan desempeños similares es coherente con la formulación PINN-E: dado que la red neuronal predice la evapotranspiración y no la precipitación, la sensibilidad del sistema acoplado a variaciones en la representación espacial del forzante pluviométrico resulta limitada. En otras palabras, la precipitación actúa como forzante externo fijo, mientras que el ajuste fino del modelo se produce principalmente a través de la estimación de \hat{E}_{agg} .

En consecuencia, considerando simultáneamente el desempeño cuantitativo y el principio de parsimonia estructural, se decidió **adoptar el promedio simple de estaciones como representación final de la precipitación observada** para las etapas posteriores del modelado.

Incorporación de *lags* de evapotranspiración

En esta etapa se evaluó el efecto de incorporar valores retardados (*lags*) de la evapotranspiración como variables de entrada del modelo PINN-E. Dado que la red neuronal predice \hat{E}_{agg} , la inclusión de memoria temporal explícita puede contribuir a capturar persistencias asociadas a procesos energéticos y condiciones atmosféricas que evolucionan de forma gradual en el tiempo.

Se analizaron configuraciones con horizontes de memoria comprendidos entre 0 y 14 días de *lags*. En todos los casos se mantuvo constante el conjunto de variables previamente seleccionado, así como la arquitectura base del modelo. Las métricas se calcularon sobre el conjunto de validación, evaluando el caudal simulado por GR4J.

Los resultados de la tabla B.36 muestran que la incorporación de memoria temporal produce variaciones marginales en la eficiencia del modelo, manteniéndose el *NSE* en un rango muy estrecho alrededor de 0.71–0.72 para la mayoría de las configuraciones evaluadas. Las diferencias en *RMSE* y *MAE* también resultan acotadas, sin observarse mejoras sustanciales asociadas a ho-

Tabla B.36: Comparación de desempeño según número de *lags* de evapotranspiración incorporados (PINN-E MLP).

<i>Lags</i>	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
0	0.72	-1.80	15.25	56.62
1	0.72	-2.17	15.24	56.52
2	0.72	-3.52	15.28	56.43
3	0.72	-2.48	15.22	56.54
4	0.72	-1.70	15.24	56.43
5	0.72	0.03	15.21	56.85
6	0.72	-0.38	15.15	56.56
7	0.72	-1.11	15.26	56.74
8	0.72	-1.11	15.23	56.45
9	0.71	0.50	15.22	57.09
10	0.72	1.18	15.16	56.98
11	0.72	-0.53	15.23	56.67
12	0.71	1.56	15.21	57.19
13	0.72	-1.64	15.28	56.69
14	0.72	0.46	15.20	56.98

rizontes de memoria extensos.

En términos de sesgo volumétrico, se observa mayor variabilidad. La configuración con 5 días de *lags* presenta el valor de *PBIAS* más cercano a cero (0.03%), indicando un balance prácticamente neutro entre el volumen simulado y el observado. En contraste, configuraciones con mayor número de *lags* no evidencian mejoras sistemáticas y, en algunos casos, introducen desviaciones adicionales.

Desde una perspectiva estructural, el incremento en el número de *lags* implica un aumento directo en la dimensionalidad del modelo y, por lo tanto, en su complejidad paramétrica. Dado que las mejoras en eficiencia son mínimas y no justifican la incorporación de un número elevado de variables adicionales, resulta razonable privilegiar configuraciones más parsimoniosas.

En consecuencia, considerando simultáneamente el control del sesgo volumétrico y el principio de parsimonia, se decidió **incorporar 5 días de *lags* de evapotranspiración** en el modelo PINN-E para las etapas posteriores del análisis.

Evaluación combinada de reducción de variables correlacionadas

Las configuraciones evaluadas fueron las siguientes:

- Modelo base (todas las variables seleccionadas previamente).
- Sólo temperatura media (sin temperaturas máxima y mínima).
- Sólo radiación solar.

- Sólo heliofanía.
- Temperatura media + heliofanía.
- Temperatura media + radiación solar.

En todos los casos se mantuvieron fijas la precipitación agregada mediante promedio simple, las variables temporales (día y mes) y los 5 días de *lags* de evapotranspiración definidos en la etapa anterior.

Tabla B.37: Resultados de desempeño para configuraciones de reducción de variables correlacionadas (PINN-E MLP).

Configuración	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Modelo base (todas las variables)	0.72	0.03	15.21	56.85
Sólo temperatura media	0.72	0.24	15.19	56.73
Sólo heliofanía	0.72	0.51	15.23	56.80
Sólo radiación solar	0.72	1.17	15.37	56.96
Temp. media + heliofanía	0.72	0.24	15.27	56.96
Temp. media + radiación solar	0.71	0.60	15.30	57.39

Los resultados, evidenciados en la tabla B.37 muestran que todas las configuraciones presentan valores de eficiencia prácticamente equivalentes ($NSE = 0.72$), lo que indica que la reducción de variables correlacionadas no produce cambios sustanciales en la capacidad explicativa del modelo.

Las configuraciones que eliminan simultáneamente variables térmicas y radiativas tienden a mostrar un leve deterioro en el desempeño, particularmente en el caso de temperatura media + radiación solar, donde el NSE desciende a 0.71 y el $RMSE$ aumenta a 57.39 m³/s.

Desde una perspectiva física, estos resultados son coherentes con la naturaleza del proceso modelado. En la formulación PINN-E, la red neuronal estima la evapotranspiración, variable fuertemente controlada por la energía disponible y la temperatura del aire. En consecuencia, la eliminación de variables asociadas al balance radiativo puede afectar la capacidad del modelo para capturar adecuadamente la dinámica de \hat{E}_{agg} , aunque el impacto observado sea moderado.

Si bien algunas configuraciones reducidas muestran desempeños similares al modelo base, las mejoras no son lo suficientemente consistentes como para justificar la eliminación permanente de variables. Además, el modelo base mantiene el mejor control del sesgo volumétrico, aspecto particularmente relevante en la simulación del balance hídrico.

En consecuencia, se decidió **mantener el modelo base con el conjunto completo de variables**, privilegiando la estabilidad global del sistema acoplado por sobre reducciones marginales en la dimensionalidad.

B.5.3. Análisis de relevancia mediante *Integrated Gradients*

Una vez definido el conjunto de variables a partir del análisis de correlación, se evaluó la relevancia individual de cada característica mediante el método de *Integrated Gradients* (IG). Este enfoque permite cuantificar la contribución promedio de cada variable a la predicción del modelo, proporcionando una medida interpretativa consistente con la estructura interna de la red.

El cálculo se realizó por separado sobre los subconjuntos de entrenamiento y validación. Las Figuras B.40a y B.40b presentan las características con mayor atribución promedio (incluyendo el signo) en ambos subconjuntos.

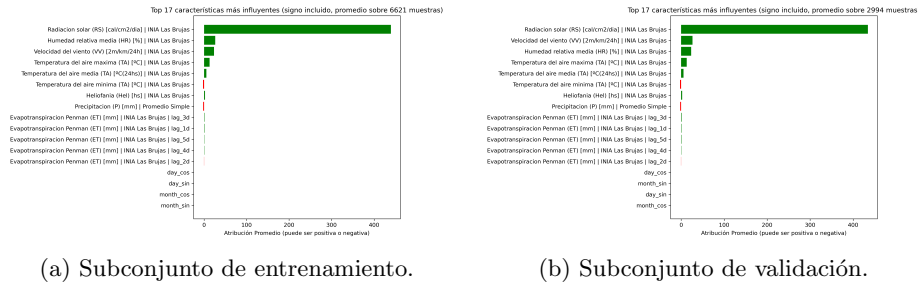


Figura B.40: Importancia relativa de variables según *Integrated Gradients* (IG) en los conjuntos de entrenamiento y validación para el modelo PINN-E MLP.

Los resultados muestran una estructura de relevancia altamente consistente entre entrenamiento y validación. En ambos casos, la **radiación solar (RS)** aparece como la variable dominante, con una magnitud de atribución considerablemente superior al resto. En un segundo nivel de importancia se ubican la velocidad del viento, la humedad relativa y las variables térmicas, todas coherentes con el control físico de la evapotranspiración.

En contraste, las variables temporales (`day_sin`, `day_cos`, `month_sin`, `month_cos`) presentan atribuciones promedio cercanas a cero en ambas particiones, lo que sugiere un aporte directo reducido dentro del modelo.

Evaluación empírica de eliminación de variables temporales.

Dado que las variables asociadas al día y al mes presentan atribuciones marginales según IG, se evaluó empíricamente el efecto de excluirlas bajo tres configuraciones:

- Eliminación únicamente de variables de mes.
- Eliminación únicamente de variables de día.

- Eliminación conjunta de día y mes.

Los resultados obtenidos sobre el conjunto de validación se presentan en la Tabla B.38.

Tabla B.38: Comparación de desempeño del modelo PINN-E (MLP) bajo eliminación de variables temporales.

Configuración	NSE	PBIAS [%]	MAE [m^3/s]	RMSE [m^3/s]
Modelo con día y mes	0.72	0.03	15.21	56.85
Sin mes	0.72	-2.03	15.45	56.82
Sin día	0.73	0.89	14.75	55.26
Sin día y mes	0.74	-1.38	15.19	55.03

Los resultados muestran que la eliminación exclusiva de las variables de mes no produce mejoras sustanciales en eficiencia y genera un aumento del sesgo volumétrico. En cambio, la eliminación de las variables asociadas al día conduce a un incremento apreciable del desempeño ($NSE = 0.73$) y a una reducción significativa del $RMSE$.

La eliminación conjunta de día y mes produce el mayor valor de eficiencia ($NSE = 0.74$) y el menor $RMSE$ ($55.03 \text{ m}^3/\text{s}$), manteniendo un sesgo volumétrico bajo ($PBIAS = -1.38\%$), dentro de rangos considerados adecuados para la simulación hidrológica.

Este comportamiento indica que, en el contexto de PINN-E, la información temporal no sólo presenta baja relevancia según IG, sino que además puede introducir redundancia respecto a las variables energéticas y térmicas que ya capturan de manera explícita la estacionalidad del sistema.

En consecuencia, considerando simultáneamente:

(i) la baja atribución observada en el análisis de IG, (ii) la mejora consistente en eficiencia al eliminar las variables temporales, (iii) y el principio de parsimonia estructural,

se decidió **excluir las variables de día y mes del conjunto final de entrada** en el modelo PINN-E.

Análisis de IG luego de eliminar variables temporales.

Dado que la eliminación conjunta de las variables de día y mes produjo una mejora en eficiencia, se recalculó el análisis de *Integrated Gradients* sobre el modelo actualizado, con el objetivo de evaluar la nueva distribución de relevancias.

Las Figuras B.41 y B.42 presentan las atribuciones promedio en los conjuntos de entrenamiento y validación, respectivamente.

En ambos subconjuntos se mantiene una estructura de relevancia consistente, dominada ampliamente por la radiación solar, seguida por variables asociadas al balance energético y aerodinámico.

Las menores atribuciones promedio corresponden principalmente a los *lags* de evapotranspiración. Sin embargo, estos forman parte de una estructura de

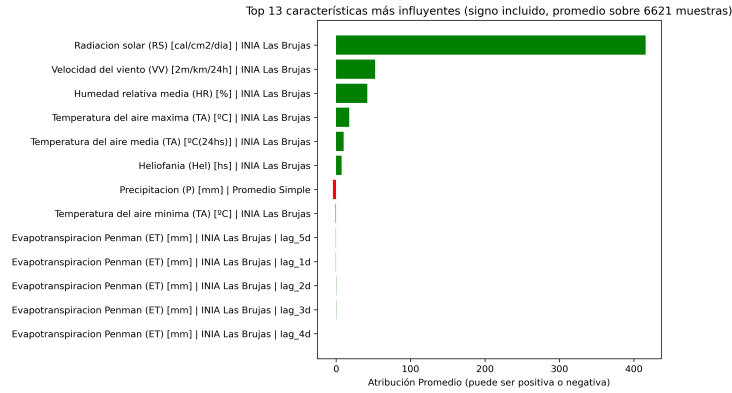


Figura B.41: Importancia relativa de variables según *Integrated Gradients* (IG) en el conjunto de entrenamiento, luego de eliminar día y mes (PINN-E MLP).

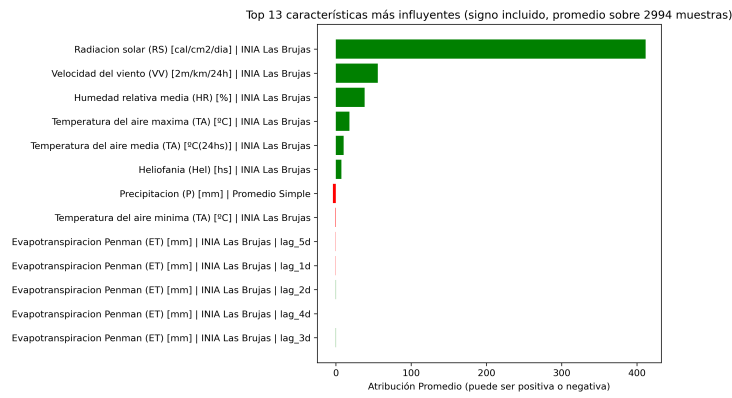


Figura B.42: Importancia relativa de variables según *Integrated Gradients* (IG) en el conjunto de validación, luego de eliminar día y mes (PINN-E MLP).

memoria temporal definida previamente a partir del análisis empírico de desempeño. Si bien algunos presentan contribuciones individuales reducidas, operan conjuntamente como un bloque coherente que captura persistencias de corto plazo en la dinámica de la evapotranspiración.

La eliminación aislada de un único *lag* alteraría la consistencia temporal del esquema adoptado, mientras que la eliminación completa del componente de memoria implicaría regresar a configuraciones previamente evaluadas que mostraron un desempeño inferior en términos de eficiencia y control del sesgo.

En consecuencia, se decidió mantener el conjunto completo de 5 días de *lags*, dando por finalizado el proceso iterativo de depuración en esta etapa.

Conjunto final de variables

El conjunto final de variables para el modelo PINN-E (MLP) fue definido a partir de un proceso secuencial que incluyó:

1. Incorporación de variables temporales.
2. Evaluación de distintos métodos de agregación espacial de la precipitación observada.
3. Análisis del impacto de *lags* de evapotranspiración.
4. Depuración de variables correlacionadas.
5. Análisis de relevancia mediante *Integrated Gradients* (IG).

En una primera etapa se evaluó la incorporación de información temporal (día y mes), observándose inicialmente mejoras marginales en eficiencia respecto a la configuración sin variables temporales explícitas. En consecuencia, ambas variables fueron incorporadas en las configuraciones posteriores.

Posteriormente se analizaron distintos métodos de agregación espacial de la precipitación observada. Si bien las diferencias en desempeño fueron acotadas, el promedio simple de estaciones alcanzó el mejor equilibrio entre eficiencia y parsimonia estructural, reduciendo la dimensionalidad respecto a la representación original por estación. Por este motivo, se adoptó el promedio simple como representación definitiva del forzante pluviométrico.

En una tercera etapa se evaluó la incorporación de memoria temporal mediante *lags* de evapotranspiración. Las variaciones en eficiencia fueron moderadas para distintos horizontes de memoria; no obstante, la configuración con 5 días de *lags* permitió mantener un adecuado control del sesgo volumétrico sin incrementar excesivamente la complejidad del modelo. En consecuencia, se incorporaron 5 días de memoria en la formulación base.

El análisis de correlación entre variables térmicas y radiativas permitió explorar configuraciones reducidas; sin embargo, las mejoras observadas no fueron suficientemente consistentes como para justificar la eliminación permanente de variables. Se decidió conservar el conjunto completo de variables meteorológicas energéticas y térmicas.

Finalmente, el análisis de relevancia mediante *Integrated Gradients* mostró que las variables temporales presentaban atribuciones promedio cercanas a cero. La evaluación empírica de su eliminación evidenció una mejora consistente en el desempeño, alcanzando un $NSE = 0.74$ tras excluir conjuntamente día y mes. En consecuencia, las variables temporales fueron removidas del conjunto final.

En función de estas decisiones, el conjunto final de variables del modelo PINN-E quedó conformado por:

- **Precipitación agregada mediante promedio simple de estaciones.**
- **Radiación solar (RS) y heliofanía (Hel).**
- **Temperatura del aire** (media, máxima y mínima).
- **Humedad relativa (HR) y velocidad del viento (VV).**
- **5 días de lags de evapotranspiración.**

Esta configuración representa el mejor equilibrio entre eficiencia predictiva, estabilidad del entrenamiento y parsimonia estructural dentro de la formulación PINN-E. Sobre este conjunto se desarrollaron las etapas posteriores de análisis del balance de la función de pérdida y optimización de hiperparámetros.

B.5.4. Análisis del balance entre términos de la función de pérdida

Al igual que en la formulación PINN-P, la función de pérdida en PINN-E combina un término de datos (asociado al ajuste de \hat{E}_{agg}) y un término físico (asociado al ajuste del caudal simulado por GR4J), ponderados mediante los coeficientes w_{data} y w_{phys} .

Con el conjunto final de variables ya definido, se evaluaron distintas combinaciones de pesos, manteniendo fija la arquitectura base del modelo. Se analizaron relaciones comprendidas hasta un máximo de 1:4 en ambos sentidos, excluyendo el caso $w_{phys} = 0$, ya que ello implicaría eliminar completamente la restricción física y desnaturalizar la formulación PINN.

La Tabla B.39 presenta las métricas obtenidas sobre el conjunto de validación para cada combinación evaluada.

Los resultados muestran que el desempeño del modelo es altamente estable frente a variaciones moderadas en la relación entre los términos de pérdida. Las diferencias en NSE se mantienen dentro de un rango muy estrecho (0.73–0.74), mientras que las variaciones en $RMSE$ y MAE son marginales.

En términos de sesgo volumétrico, tampoco se observan desbalances significativos. Todas las configuraciones presentan valores de $PBIAS$ cercanos a cero, lo que indica que el modelo mantiene un adecuado control del balance hídrico independientemente del peso relativo asignado a cada término.

Las diferencias observadas entre configuraciones son mínimas y no evidencian ventajas consistentes asociadas a una relación particular entre los términos

Tabla B.39: Desempeño del modelo PINN-E (MLP) para distintas combinaciones de pesos en la función de pérdida.

$w_{\text{data}} - w_{\text{phys}}$	NSE	PBIAS [%]	MAE [m^3/s]	RMSE [m^3/s]
0.00 – 1.00	0.73	-0.74	15.19	55.04
0.25 – 1.00	0.73	-0.33	15.16	55.05
0.25 – 0.75	0.73	-0.61	15.17	55.06
0.50 – 1.00	0.73	-0.75	15.18	55.08
0.50 – 0.75	0.73	-0.01	15.15	55.07
0.75 – 1.00	0.74	-0.67	15.17	55.03
1.00 – 1.00	0.74	-1.38	15.19	55.03
1.00 – 0.75	0.73	-0.66	15.18	55.05
1.00 – 0.50	0.73	-1.28	15.19	55.07
0.75 – 0.50	0.73	-0.53	15.18	55.07
0.75 – 0.25	0.74	-1.42	15.19	55.01
1.00 – 0.25	0.74	-0.91	15.15	54.99

de pérdida. En términos prácticos, todas las combinaciones evaluadas pueden considerarse equivalentes dentro del nivel de precisión reportado.

En consecuencia, dado que no se identificaron mejoras consistentes asociadas a una combinación específica de pesos y considerando el principio de simplicidad metodológica, se decidió **mantener la configuración balanceada** $w_{\text{data}} = 1$ y $w_{\text{phys}} = 1$.

B.5.5. Optimización de hiperparámetros (Optuna)

Una vez definido el conjunto final de variables y fijado el balance de la función de pérdida en $w_{\text{data}} = 1$ y $w_{\text{phys}} = 1$, se procedió a la optimización de hiperparámetros mediante el *framework* *Optuna*, utilizando el mismo esquema metodológico aplicado en los modelos anteriores.

El proceso de búsqueda incluyó hiperparámetros estructurales (número de capas ocultas y neuronas por capa), así como parámetros asociados al entrenamiento (función de activación, tasa de aprendizaje, *dropout*, *weight decay* y *batch size*). La función objetivo correspondió a la pérdida de validación (MSE), evaluada bajo la formulación compuesta propia de la PINN-E.

El estudio culminó con los siguientes resultados:

- **Mejor valor de pérdida de validación:** 2533.30
- ***Trials* completados:** 139
- ***Trials* podados:** 31

La Figura B.43 muestra la evolución del valor objetivo a lo largo de los *trials*. Durante las primeras iteraciones se observa una fase de exploración amplia, con elevada variabilidad en los valores de la función objetivo. A medida que avanzó

el proceso, el algoritmo concentró la búsqueda en regiones más prometedoras del espacio de hiperparámetros, logrando una reducción progresiva del mejor valor alcanzado.

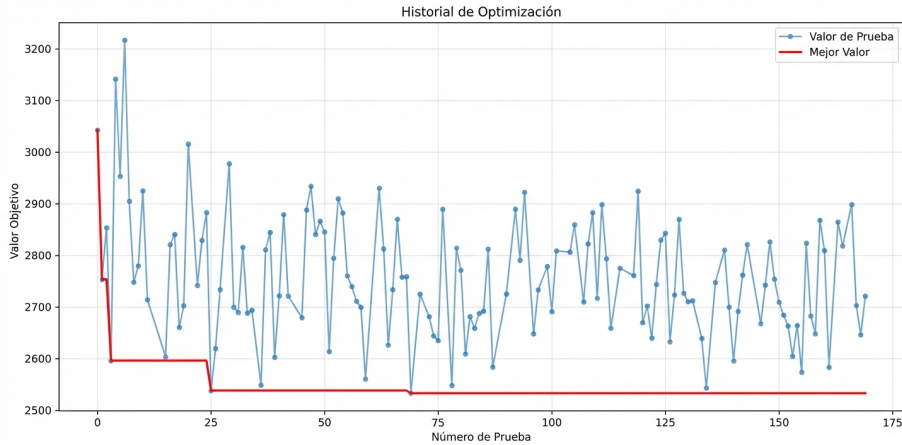


Figura B.43: Historial del proceso de optimización realizado con Optuna para PINN-E (MLP).

El mejor resultado se obtuvo en el *trial* 69. La configuración arquitectónica óptima se resume en la Tabla B.40.

Tabla B.40: Configuración óptima seleccionada por Optuna para PINN-E (MLP).

Hiperparámetro	Valor óptimo
Número de capas ocultas	5
Neuronas por capa	[81, 640, 979, 32, 83]
Función de activación	relu
Tasa de aprendizaje	0.00126
<i>Dropout</i> base	0.49
<i>Batch size</i>	16
<i>Weight decay</i>	1.51×10^{-5}

La arquitectura seleccionada presenta cinco capas ocultas con una configuración no monótona en el número de neuronas, incluyendo dos capas de gran ancho (640 y 979 unidades). Asimismo, el valor elevado de *dropout* (0.49) indica un nivel significativo de regularización, consistente con la profundidad y capacidad representacional del modelo.

La tasa de aprendizaje óptima (1.26×10^{-3}) resulta más conservadora que la obtenida en PINN-P en su misma versión (MLP), lo que sugiere un proceso de optimización más estable para esta formulación.

La Figura B.44 presenta las curvas de aprendizaje correspondientes a los cinco mejores *trials*. Se observa que, si bien todos convergen hacia valores similares de pérdida, existen diferencias en estabilidad y oscilaciones intermedias durante el entrenamiento.

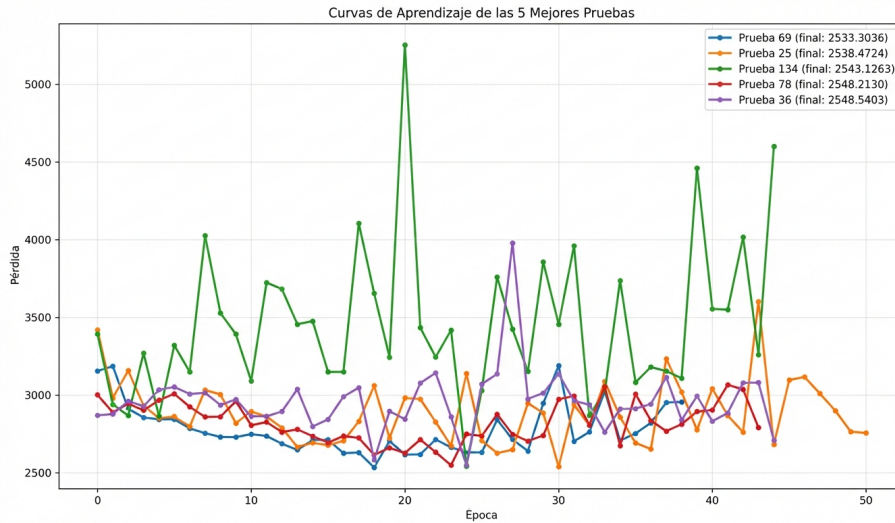


Figura B.44: Curvas de aprendizaje de los cinco mejores *trials* del proceso de optimización de Optuna (PINN-E MLP).

Finalmente, la Figura B.45 muestra el historial de *pruning*. La poda permitió descartar configuraciones poco prometedoras en etapas tempranas, reduciendo el costo computacional total sin afectar la calidad del resultado final.

En conjunto, la optimización mediante Optuna permitió identificar una arquitectura significativamente más profunda y con mayor capacidad representacional que la configuración base. Este resultado sugiere que la tarea de estimar evapotranspiración bajo una restricción física impuesta por GR4J requiere una mayor flexibilidad estructural para capturar las interacciones no lineales entre variables meteorológicas y memoria temporal.

B.5.6. Evaluación extendida del modelo optimizado

El modelo final PINN-E (MLP) se entrenó utilizando la arquitectura óptima identificada mediante Optuna, manteniendo el balance de pérdida $w_{\text{data}} = 1$ y $w_{\text{phys}} = 1$, y el conjunto final de variables definido en las etapas previas de selección e interpretación.

La evaluación se realizó de forma independiente sobre los conjuntos de entrenamiento, validación y prueba, preservando la partición temporal original y aplicando el mismo criterio de enmascaramiento para los registros con caudal faltante. Las métricas consideradas fueron MAE, RMSE, NSE y PBIAS. La Tabla B.41 presenta el desempeño del modelo en los distintos conjuntos.

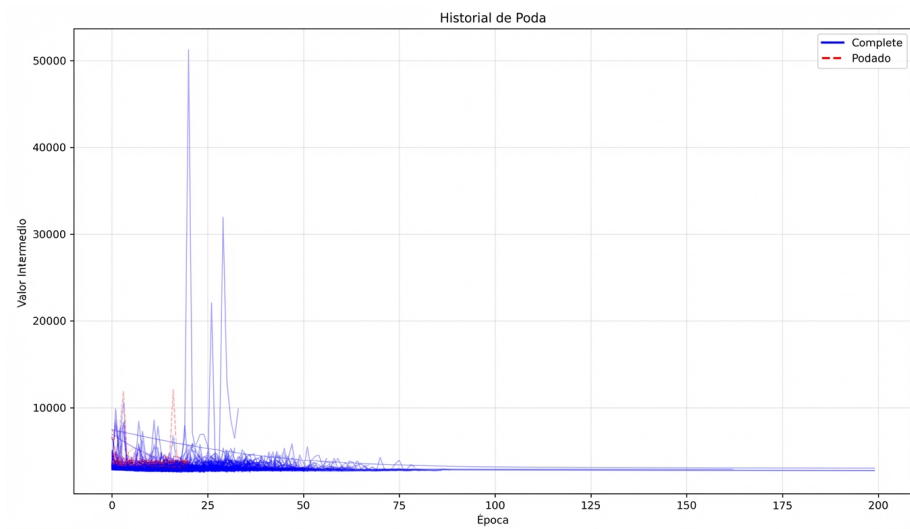


Figura B.45: Historial de *pruning* durante la optimización con Optuna (PINN-E MLP).

Tabla B.41: Desempeño del modelo PINN-E optimizado en los distintos conjuntos.

Conjunto	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Entrenamiento	0.69	-46.06	19.37	56.73
Validación	0.75	-21.50	16.98	53.00
Prueba	0.71	-30.62	19.37	69.43

Análisis del desempeño

En el conjunto de validación se obtiene un desempeño elevado, con un $NSE = 0.75$ y un $RMSE = 53.00 \text{ m}^3/\text{s}$, en línea con las mejores configuraciones intermedias evaluadas durante el desarrollo del modelo. En el conjunto de prueba, el modelo mantiene una eficiencia adecuada ($NSE = 0.71$), lo que indica una capacidad razonable de generalización temporal.

No obstante, se observa un sesgo porcentual negativo en todos los conjuntos, particularmente en entrenamiento ($PBIAS = -46.06\%$) y prueba ($PBIAS = -30.62\%$), lo que evidencia una tendencia sistemática a la sobreestimación del volumen total de caudal. Si bien el sesgo es menor que el observado en el modelo PINN-P optimizado, continúa representando una limitación relevante desde el punto de vista del balance hídrico.

En el conjunto de entrenamiento, las métricas resultan inferiores a las de validación ($NSE = 0.69$ frente a 0.75). Este comportamiento, también observado en PINN-P, puede explicarse por la naturaleza compuesta de la función de pérdida en las PINNs.

Dado que la optimización se realiza sobre una combinación ponderada de pérdida de datos y pérdida física, el mínimo de validación no necesariamente coincide con un mínimo estricto del error directo sobre caudal en entrenamiento. Asimismo, el valor elevado de *dropout* utilizado en la arquitectura seleccionada introduce un efecto regularizador que puede penalizar el ajuste en entrenamiento, resultando en pérdidas mayores respecto a validación.

La Figura B.46 muestra la evolución de la pérdida durante el entrenamiento. Se observa que la pérdida de validación alcanza su mínimo en la época 19, mientras que la pérdida de entrenamiento permanece sistemáticamente por encima, reforzando la idea de que el criterio de optimización no está alineado exclusivamente con el ajuste empírico directo.

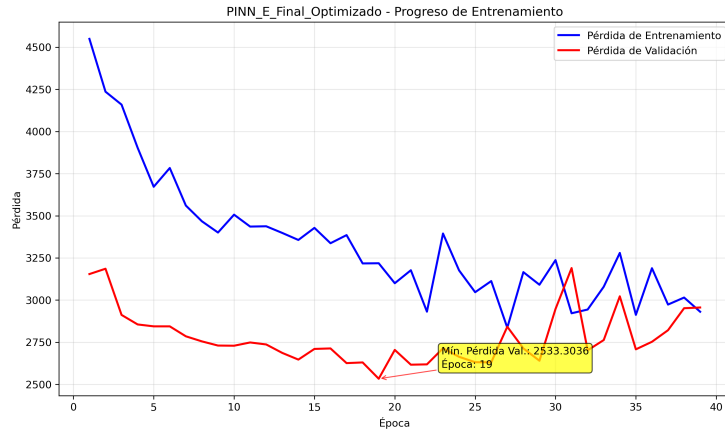


Figura B.46: Evolución de la pérdida de entrenamiento y validación del modelo PINN-E optimizado.

Análisis gráfico del ajuste

La Figura B.47 presenta la comparación temporal entre caudal observado y predicho en el conjunto de test final. El modelo reproduce adecuadamente la dinámica general de la serie, capturando la mayoría de los eventos de crecida intermedia y la estructura estacional.

Sin embargo, se observa una subestimación parcial en eventos extremos, particularmente en el evento de mayor magnitud registrado alrededor de 2019, donde el pico máximo no es completamente reproducido.

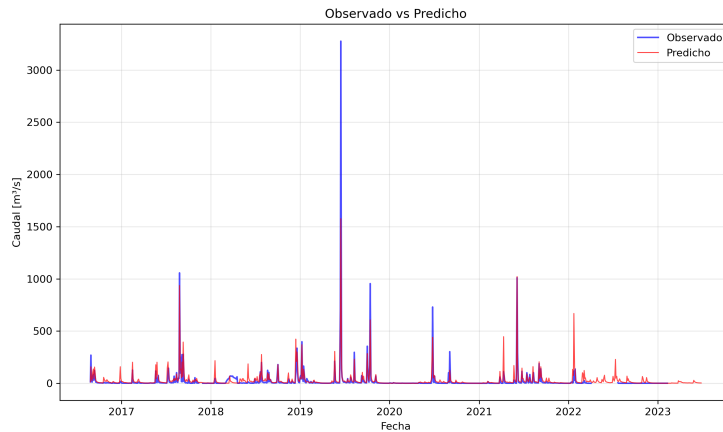


Figura B.47: Serie temporal de caudales observados y predichos por el modelo PINN-E optimizado.

El *scatterplot* en escala logarítmica (Figura B.48) muestra una buena alineación con la línea 1:1 en el rango medio de caudales, con una tendencia a sobreestimación, mientras que para valores extremos se observa mayor dispersión y una tendencia general a la subestimación en eventos de gran magnitud.

El análisis de residuos (Figura B.49) confirma que la mayoría de los errores se concentran en torno a cero, mientras que los valores más extremos se asocian a eventos de crecida de gran magnitud. No se observa un patrón sistemático creciente o decreciente en el tiempo, aunque el sesgo negativo global indica una subestimación acumulativa del volumen simulado.

Análisis de atribución final mediante *Integrated Gradients*

Con el objetivo de interpretar el comportamiento del modelo optimizado, se realizó un análisis de relevancia de variables utilizando el método *Integrated Gradients* (IG) sobre el conjunto de prueba.

La Figura B.50 muestra las 13 características con mayor atribución promedio (incluyendo el signo).

Se observa que las variables térmicas y radiativas concentran la mayor atribución promedio, destacándose particularmente la temperatura máxima y la

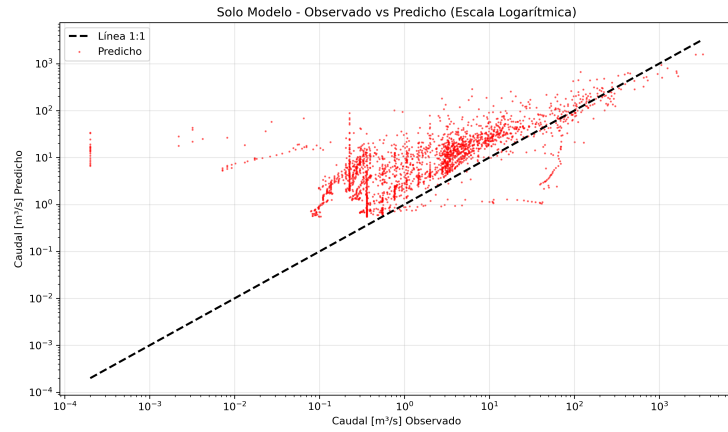


Figura B.48: Diagrama observado vs predicho en escala logarítmica para el modelo PINN-E optimizado.

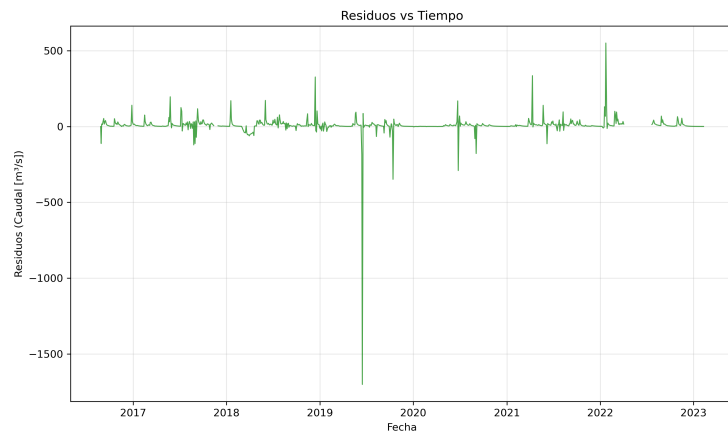


Figura B.49: Residuos del modelo PINN-E optimizado en función del tiempo.

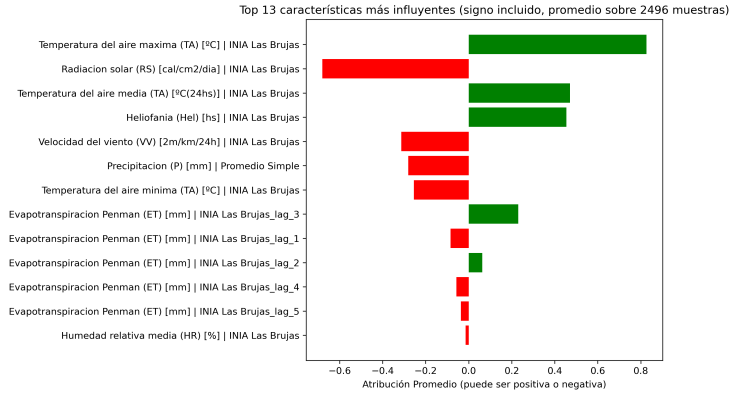


Figura B.50: Top 13 características más influyentes según *Integrated Gradients* en el conjunto de prueba (modelo PINN-E optimizado).

radiación solar. Los *lags* de evapotranspiración presentan contribuciones de menor magnitud relativa.

El patrón de atribuciones obtenido es consistente con la formulación física del proceso de evapotranspiración, donde las variables energéticas y térmicas desempeñan un rol central en la determinación de la demanda evaporativa.

Costo computacional

El entrenamiento completo del modelo optimizado demandó 454.49 segundos en un total de 39 épocas en el entorno de cómputo utilizado. Por su parte, el tiempo de inferencia fue de 3.92 segundos, según los registros del sistema.

B.6. PINN-E RNN

Esta sección presenta el desarrollo detallado del modelo PINN-E con arquitectura RNN, incluyendo todos los experimentos realizados durante el proceso de optimización.

El cuerpo principal (Sección 6.9) resume únicamente los resultados del modelo base y del modelo final optimizado. Aquí se documentan las etapas intermedias que condujeron a la configuración seleccionada.

B.6.1. Modelo base

El primer experimento en el marco del enfoque PINN-E con arquitectura recurrente consistió en entrenar una red LSTM bajo la formulación **PINN-E**, en la cual la red neuronal predice exclusivamente la evapotranspiración agregada diaria \hat{E}_{agg} , mientras que la precipitación se mantiene fija a partir de los valores observados P_{obs} . La predicción final de caudal no es producida directamente

por la red, sino que emerge del acoplamiento físico con GR4J, evaluando el caudal simulado $Q_{GR4J}(P_{obs}, \hat{E}_{agg}; \theta)$ y comparándolo con el caudal observado mediante el término físico de la función de pérdida.

A diferencia de la arquitectura MLP utilizada en la sección anterior, la red LSTM procesa secuencias temporales de longitud configurable (*sequence length*), lo que le permite capturar dependencias temporales de forma implícita a través de su estado oculto. La configuración base del modelo empleó una secuencia de longitud 1 (equivalente a una ventana unitaria), dos capas LSTM con 64 unidades ocultas, y un módulo MLP posterior de tres capas (128→64→32) con activación ReLU.

Dado que el conjunto de prueba se reserva exclusivamente para la evaluación final del modelo, el desempeño en esta etapa se reporta sobre el **conjunto de validación**, utilizando MAE, RMSE, NSE y PBIAS calculadas sobre el caudal simulado por GR4J.

Tabla B.42: Desempeño del modelo base PINN-E RNN sobre el conjunto de validación.

NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
0.73	-7.85	15.58	55.85

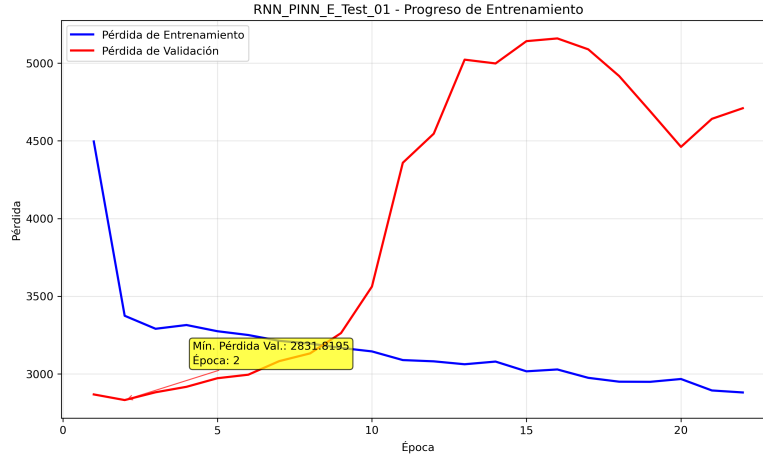


Figura B.51: Evolución de la función de pérdida (entrenamiento y validación) para el modelo base PINN-E (RNN).

La Figura B.51 muestra que la pérdida de validación alcanza su mínimo global de 2831.82 en la **época 2**, a partir de la cual se incrementa de forma sostenida hasta valores superiores a 5000, mientras que la pérdida de entrenamiento continúa disminuyendo gradualmente desde ~4500 hasta ~2900. Este patrón indica una aparición muy temprana de *overfitting*, incluso más precoz

que la observada en el modelo base PINN-E MLP (época 7). La separación marcada entre ambas curvas a partir de la época 8 confirma la necesidad de optimizar la arquitectura y los hiperparámetros en las etapas posteriores.

En términos cuantitativos, el modelo base recurrente presenta un desempeño superior al observado en la configuración base del modelo MLP ($NSE = 0.71$, $RMSE = 57.86 \text{ m}^3/\text{s}$). Con $NSE = 0.73$ (Tabla B.42), la red LSTM explica aproximadamente el 73 % de la variabilidad observada del caudal en el conjunto de validación, lo que representa una mejora de 2 puntos porcentuales respecto a la versión MLP.

El error cuadrático medio ($RMSE = 55.85 \text{ m}^3/\text{s}$) se reduce en aproximadamente $2 \text{ m}^3/\text{s}$, mientras que el sesgo porcentual ($PBIAS = -7.85 \%$) indica una tendencia moderada a la sobreestimación del volumen total simulado. Este valor es algo más elevado en magnitud que el observado en el modelo base MLP ($PBIAS = 2.21 \%$), aunque se mantiene dentro del rango aceptable ($\pm 10 \%$).

Es relevante señalar que el modelo PINN-E (RNN) base supera al modelo GR4J directo ($NSE = 0.72$, $RMSE = 56.95 \text{ m}^3/\text{s}$, $PBIAS = -9.86 \%$) en términos de eficiencia y error, lo que indica que la estimación de evapotranspiración mediante la red recurrente, acoplada con GR4J, ya es capaz de mejorar la simulación hidrológica respecto al modelo conceptual alimentado directamente con precipitación observada.

En conjunto, estos resultados muestran que la configuración base de la PINN-E con arquitectura RNN resulta más efectiva que su contraparte MLP en eficiencia, aunque con un sesgo ligeramente mayor en magnitud. No obstante, al igual que en los modelos anteriores, se procedió al desarrollo de las etapas posteriores de ajuste.

B.6.2. Evaluación de la incorporación de variables

En esta etapa se analiza el impacto de distintas configuraciones de variables de entrada sobre el desempeño del modelo PINN-E basado en arquitectura RNN. Al igual que en la versión MLP, en la formulación PINN-E la red neuronal predice la evapotranspiración agregada diaria \hat{E}_{agg} , que luego es utilizada como forzante del modelo GR4J junto con la precipitación observada.

Las métricas reportadas en esta sección se calculan sobre el conjunto de validación y corresponden al caudal generado por GR4J, evaluado mediante MAE, RMSE, NSE y PBIAS.

Variables temporales

En primer lugar, se evaluó el efecto de incorporar variables temporales representadas de forma cíclica mediante transformaciones seno y coseno del *mes* y del *día*. La inclusión de estas variables busca capturar patrones estacionales que puedan influir en la estimación de la evapotranspiración agregada y, en consecuencia, en la simulación del caudal a través de GR4J.

Se compararon cuatro configuraciones: (i) sin variables temporales, (ii) incorporación únicamente del día, (iii) incorporación únicamente del mes, y (iv)

incorporación conjunta de día y mes. Los resultados obtenidos se presentan en la Tabla B.43.

Tabla B.43: Comparación de desempeño para distintas combinaciones de variables temporales sobre el conjunto de validación (PINN-E RNN).

Configuración	NSE	PBIAS [%]	MAE [m^3/s]	RMSE [m^3/s]
Ninguna	0.7270	-7.85	15.58	55.85
día	0.7250	-7.71	15.68	56.06
mes	0.7297	-9.23	15.86	55.57
día, mes	0.7265	-6.07	15.36	55.91

Los resultados muestran que las diferencias entre configuraciones son marginales en términos de eficiencia ($NSE \approx 0.73$ en todos los casos). No obstante, se observan variaciones en las métricas secundarias que permiten distinguir las configuraciones.

La incorporación exclusiva del *mes* alcanza el menor $RMSE$ ($55.57 \text{ m}^3/\text{s}$), indicando una leve mejora en la precisión global de la simulación. Si bien el sesgo porcentual aumenta ligeramente en magnitud ($PBIAS = -9.23 \%$), este incremento es moderado y se mantiene dentro del rango aceptable.

Este comportamiento difiere del observado en la formulación PINN-P con RNN, donde la incorporación de variables temporales degradaba consistentemente el desempeño. La diferencia puede explicarse por el proceso modelado: la evapotranspiración presenta una estacionalidad marcada, por lo que la información mensual puede actuar como un complemento útil incluso para una arquitectura recurrente. En contraste, la precipitación presenta mayor variabilidad entre estaciones, lo que hace que la variable temporal resulte menos informativa o incluso perjudicial.

En consecuencia, considerando la mejora en $RMSE$ y la coherencia con la estacionalidad del proceso de evapotranspiración, se decidió **conservar únicamente la variable temporal correspondiente al mes** para las etapas posteriores del modelado.

Representación espacial de la precipitación observada

Se compararon cuatro configuraciones:

- Serie original por estación (una variable por estación).
- Promedio simple de estaciones.
- Promedio ponderado por distancia.
- Agregación mediante polígonos de Thiessen.

Las métricas se calcularon sobre el conjunto de validación, evaluando el caudal simulado por GR4J.

Tabla B.44: Comparación de desempeño según método de agregación espacial de la precipitación observada (PINN-E RNN).

Método	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Original (por estación)	0.73	-9.23	15.86	55.57
Promedio simple	0.74	-9.37	15.36	54.89
Ponderado por distancia	0.72	-9.41	15.87	56.29
Ponderado Thiessen	0.74	-6.01	15.23	54.16

Los resultados muestran que la **agregación mediante polígonos de Thiessen** alcanza el mejor desempeño global, con el mayor NSE (0.74), el menor $RMSE$ (54.16 m³/s) y el menor MAE (15.23 m³/s). Además, presenta el menor sesgo volumétrico en valor absoluto ($PBIAS = -6.01$ %, considerablemente inferior al observado en las demás configuraciones).

Este resultado difiere del observado en la versión MLP de la PINN-E, donde el promedio simple alcanzaba el mejor desempeño. La diferencia sugiere que la capacidad de la LSTM para procesar secuencias temporales le permite aprovechar mejor una representación espacial más sofisticada de la precipitación.

Desde una perspectiva hidrológica, la agregación por polígonos de Thiessen proporciona una representación espacial ponderada por la proximidad geográfica de las estaciones al punto de interés, lo que resulta más coherente con la distribución real de la precipitación sobre la cuenca. Además, esta configuración reduce la dimensionalidad respecto a la representación original por estación, lo que favorece la parsimonia estructural.

En consecuencia, se decidió **adoptar la agregación mediante polígonos de Thiessen** como representación definitiva de la precipitación observada para las etapas posteriores del modelado.

B.6.3. Selección de la ventana temporal (*sequence length*)

Un aspecto clave de la arquitectura RNN es la longitud de la secuencia temporal de entrada. Se evaluaron configuraciones con ventanas de 1 a 14 días, manteniendo constante el conjunto de variables (con mes y precipitación Thiessen) y la arquitectura base del modelo. Los resultados se presentan en la Tabla B.45.

Los resultados muestran una tendencia general de mejora del desempeño al incrementar la longitud de la ventana temporal, a diferencia de lo observado en la formulación PINN-P con RNN, donde la ventana unitaria ofrecía el mejor resultado puntual. Este comportamiento es coherente con la naturaleza del proceso modelado: la evapotranspiración presenta una dinámica más gradual y persistente que la precipitación, por lo que la incorporación de memoria temporal adicional aporta información relevante para su estimación.

Tabla B.45: Comparación de desempeño según longitud de la ventana temporal (PINN-E RNN).

Ventana (días)	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
1	0.74	-6.01	15.23	54.16
2	0.73	-12.26	15.83	55.33
3	0.73	-6.74	15.55	55.30
4	0.75	1.56	14.58	53.94
5	0.76	-16.86	15.74	52.16
6	0.74	-11.72	15.70	54.19
7	0.76	-7.74	15.10	52.29
8	0.77	0.57	14.72	51.21
9	0.75	-19.76	16.88	53.61
10	0.76	-4.30	14.78	52.38
11	0.76	-4.95	15.00	52.61
12	0.75	-4.58	15.35	53.02
13	0.75	-19.47	16.21	53.46
14	0.74	-17.43	16.48	54.17

La configuración de **8 días** alcanza el mejor desempeño global, con el mayor *NSE* (0.77), el menor *RMSE* (51.21 m³/s) y un sesgo volumétrico prácticamente nulo (*PBIAS* = 0.57 %). Ventanas más largas (10–14 días) mantienen eficiencias comparables (*NSE* ≈ 0.74–0.76), pero con mayor variabilidad en el sesgo y, en general, sin superar el desempeño de la ventana de 8 días. Desde una perspectiva hidrológica, la ventana de 8 días resulta coherente con la escala temporal de los procesos que regulan la evapotranspiración, que incluyen la evolución de las condiciones radiativas, térmicas y de humedad del suelo a lo largo de varios días. Esta longitud de secuencia permite a la red LSTM capturar la persistencia de las condiciones atmosféricas sin introducir excesiva complejidad.

En consecuencia, se decidió **adoptar una ventana temporal de 8 días** para las etapas siguientes del modelado.

B.6.4. Depuración inicial por correlación

Siguiendo el mismo procedimiento aplicado en las formulaciones anteriores, se realizó un análisis de correlación entre las variables meteorológicas con el objetivo de identificar posibles redundancias y evaluar la factibilidad de reducir la dimensionalidad del modelo.

En todos los casos, el modelo fue reentrenado completamente bajo la formulación PINN-E con arquitectura RNN, manteniendo fija la ventana temporal de 8 días y los pesos de la función de pérdida en $w_{\text{data}} = 1$ y $w_{\text{phys}} = 1$. Las métricas se calcularon sobre el conjunto de validación, evaluando el caudal simulado por GR4J.

Evaluación combinada de reducción de variables correlacionadas

Las configuraciones evaluadas fueron las siguientes:

- Modelo base (todas las variables, 10 *features*).
- Sólo temperatura media (sin temperaturas máxima y mínima).
- Sin radiación solar.
- Sin heliofanía.
- Sólo temperatura media y sin radiación solar.

Los resultados se presentan en la Tabla B.46.

Tabla B.46: Resultados de desempeño para configuraciones de reducción de variables correlacionadas (PINN-E RNN).

Configuración	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Modelo base	0.77	0.57	14.72	51.21
Sin temp. máx. y mín.	0.75	-1.79	14.71	53.61
Sin radiación solar	0.74	-18.33	16.37	53.99
Sin heliofanía	0.76	-7.90	15.28	52.86
Sin temp. máx./mín. ni RS	0.75	-10.22	15.36	53.93

Los resultados muestran que el modelo base con el conjunto completo domina consistentemente en todas las métricas relevantes. Presenta el mayor *NSE* (0.77), el menor *RMSE* (51.21 m³/s) y, notablemente, el sesgo volumétrico más bajo en valor absoluto (*PBIAS* = 0.57 %).

Todas las exclusiones producen un deterioro del desempeño, con caídas del *NSE* de 1 a 3 puntos porcentuales. La eliminación de la radiación solar genera el mayor deterioro, tanto en eficiencia (*NSE* = 0.74) como en sesgo (*PBIAS* = -18.33 %), lo que confirma la relevancia central de esta variable en la estimación de la evapotranspiración dentro del sistema acoplado, resultado coherente con la física del proceso.

La eliminación de la heliofanía produce un impacto más moderado (*NSE* = 0.76), mientras que la exclusión de las temperaturas extremas reduce el *NSE* a 0.75 sin afectar significativamente el MAE.

En consecuencia, y dado que ninguna exclusión produce mejoras en desempeño, se decidió **mantener el modelo base con el conjunto completo**, priorizando tanto el desempeño global como el control del sesgo volumétrico.

B.6.5. Análisis de relevancia mediante *Integrated Gradients*

Una vez definido el conjunto de variables a partir del análisis de correlación, se evaluó la relevancia individual de cada característica mediante el método de *Integrated Gradients* (IG). El cálculo se realizó por separado sobre los subconjuntos de entrenamiento y validación. La Figura B.52 presenta las variables con mayor atribución promedio en ambos subconjuntos.

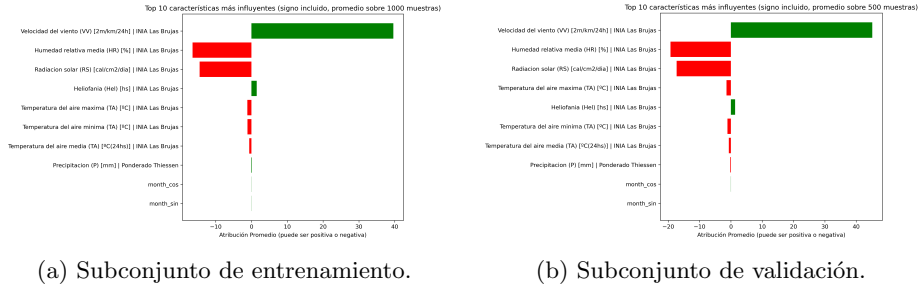


Figura B.52: Importancia relativa de variables según *Integrated Gradients* (IG) en los conjuntos de entrenamiento y validación (PINN-E RNN).

Los resultados muestran una estructura de relevancia altamente consistente entre entrenamiento y validación. En ambos casos, la **velocidad del viento (VV)** aparece como la variable con mayor atribución positiva ($\sim +40$), seguida por la **humedad relativa (HR)** y la **radiación solar (RS)** con atribuciones negativas de magnitud comparable (~ -10 a -20). La heliofanía presenta una atribución positiva leve, mientras que las variables térmicas (máxima, mínima y media) contribuyen con atribuciones negativas moderadas. La precipitación Thiessen y las variables temporales (`month_sin`, `month_cos`) exhiben atribuciones cercanas a cero.

Este patrón es coherente con la física de la evapotranspiración, donde las variables aerodinámicas y energéticas desempeñan un rol central. A diferencia de la versión MLP de la PINN-E —donde la radiación solar dominaba ampliamente—, en la LSTM la velocidad del viento emerge como la variable más influyente, lo que sugiere que la arquitectura recurrente prioriza la componente aerodinámica del proceso evaporativo.

Evaluación empírica de eliminación de variables temporales.

Dado que en análisis previos las variables temporales han presentado atribuciones marginales, se evaluó empíricamente el efecto de excluir la variable `month_sin`. Los resultados se presentan en la Tabla B.47.

Los resultados muestran que la eliminación de la variable de mes produce un deterioro consistente del desempeño: el *NSE* disminuye de 0.77 a 0.75, el *RMSE* aumenta de 51.21 a 53.75 m³/s, y el sesgo porcentual se incrementa de 0.57 % a 7.35 %. La pérdida de control del sesgo resulta particularmente relevante, ya que el sesgo casi nulo del modelo con mes representaba una de sus fortalezas principales.

Tabla B.47: Comparación de desempeño del modelo PINN-E RNN al eliminar la variable temporal de mes.

Configuración	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Modelo con mes	0.77	0.57	14.72	51.21
Modelo sin mes	0.75	7.35	15.05	53.75

Este resultado contrasta con la formulación PINN-E MLP, donde la eliminación de las variables temporales producía una mejora del desempeño (NSE de 0.72 a 0.74). La diferencia sugiere que, en la arquitectura RNN procesando secuencias de 8 días, la información mensual aporta una señal complementaria que la red no logra capturar completamente a través del procesamiento secuencial de las variables meteorológicas.

En consecuencia, se decidió **mantener la variable `month_sin` y `month_cos`** en el conjunto final de entrada, dando por finalizado el proceso iterativo de depuración.

Conjunto final de variables

El conjunto final de variables para el modelo PINN-E RNN fue definido a partir de un proceso secuencial que incluyó: (i) evaluación de la incorporación de variables temporales, (ii) selección del método de agregación espacial de la precipitación, (iii) selección de la ventana temporal, (iv) análisis de correlación entre variables meteorológicas, y (v) análisis de relevancia mediante *Integrated Gradients* (IG).

En una primera etapa se evaluó la incorporación de información estacional, observándose que a diferencia de la formulación PINN-P con RNN, la inclusión del mes producía una leve mejora en el $RMSE$, resultado coherente con la estacionalidad marcada de la evapotranspiración. Se decidió incorporar las variables `month_sin` y `month_cos`.

Posteriormente se evaluaron distintos métodos de agregación espacial de la precipitación observada, seleccionándose la agregación mediante polígonos de Thiessen por alcanzar el mejor desempeño global ($NSE = 0.74$, $PBIAS = -6.01\%$).

La selección de la ventana temporal reveló una tendencia de mejora con secuencias más largas, seleccionándose una ventana de 8 días ($NSE = 0.77$, $PBIAS = 0.57\%$) como la configuración óptima.

El análisis de correlación confirmó que el conjunto completo de variables ofrece el mejor desempeño, y la evaluación mediante IG descartó la eliminación de variables temporales, cuya remoción producía un deterioro significativo.

En función de estos resultados, el conjunto final de variables quedó conformado por:

- **Precipitación agregada mediante polígonos de Thiessen.**

- **Radiación solar (RS)** y **heliofanía (Hel)**.
- **Temperatura del aire** (media, máxima y mínima).
- **Humedad relativa (HR)** y **velocidad del viento (VV)**.
- **Variables temporales** `month_sin` y `month_cos`.

Esta configuración se procesa mediante una ventana temporal de 8 días. Sobre este conjunto se desarrollaron las etapas posteriores de análisis del balance entre los términos de la función de pérdida y la optimización de hiperparámetros.

B.6.6. Análisis del balance entre términos de la función de pérdida

Al igual que en las formulaciones anteriores, la función de pérdida en PINN-E combina un término de datos (asociado al ajuste de \hat{E}_{agg}) y un término físico (asociado al ajuste del caudal simulado por GR4J), ponderados mediante los coeficientes w_{data} y w_{phys} .

Se evaluaron 13 configuraciones que cubren relaciones comprendidas hasta un máximo de 1:4 en ambos sentidos, incluyendo los casos extremos. La Tabla B.48 presenta las métricas obtenidas en el conjunto de validación.

Tabla B.48: Desempeño del modelo PINN-E RNN para distintas combinaciones de pesos en la función de pérdida.

$w_{data} - w_{phys}$	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
0.00 – 1.00	0.76	-13.41	15.62	52.19
1.00 – 0.00	0.72	-9.90	16.26	57.05
0.25 – 1.00	0.75	-13.21	15.81	53.76
0.25 – 0.75	0.75	-20.93	16.64	53.66
0.50 – 1.00	0.76	-4.41	15.23	52.71
0.50 – 0.75	0.75	-25.19	17.15	53.44
0.75 – 1.00	0.75	-5.28	15.10	53.31
1.00 – 1.00	0.77	0.57	14.72	51.21
1.00 – 0.75	0.76	-4.51	14.88	52.51
1.00 – 0.50	0.77	-10.43	15.62	51.36
0.75 – 0.50	0.76	-2.67	15.08	52.46
0.75 – 0.25	0.77	-6.05	15.26	51.64
1.00 – 0.25	0.75	-2.23	14.91	53.45

Los resultados muestran que la configuración balanceada $w_{data} = 1.00$, $w_{phys} = 1.00$ alcanza el mejor desempeño global, con el mayor *NSE* (0.77), el menor *RMSE* (51.21 m³/s), el menor *MAE* (14.72 m³/s) y el sesgo volumétrico más bajo en valor absoluto (*PBIAS* = 0.57 %). Esta configuración domina simultáneamente en las cinco métricas evaluadas, lo que constituye un resultado poco frecuente en este tipo de análisis.

A diferencia de lo observado en la formulación PINN-P con RNN donde un peso reducido del término físico ($w_{\text{phys}} = 0.25$) ofrecía el mejor equilibrio, en la PINN-E el balance simétrico resulta óptimo. Este resultado es consistente con lo observado en la versión MLP de la PINN-E, donde también se adoptó la configuración balanceada.

En consecuencia, se decidió **mantener la configuración balanceada** $w_{\text{data}} = 1$ y $w_{\text{phys}} = 1$ para las etapas posteriores de optimización y evaluación final.

B.6.7. Optimización de hiperparámetros (Optuna)

Una vez definido el conjunto final de variables, la ventana temporal de 8 días y el balance de función de pérdida en $w_{\text{data}} = 1.00$ y $w_{\text{phys}} = 1.00$, se procedió a la optimización de hiperparámetros mediante el *framework Optuna*.

El espacio de búsqueda incluyó hiperparámetros específicos de la arquitectura recurrente (tipo de celda, tamaño del estado oculto, número de capas, *dropout* recurrente), así como los parámetros del módulo MLP posterior y los parámetros de entrenamiento. La función objetivo correspondió a la pérdida de validación (MSE), evaluada bajo la formulación compuesta propia de la PINN-E.

El estudio culminó con los siguientes resultados:

- **Mejor valor de pérdida de validación:** 2334.95
- ***Trials* completados:** 141
- ***Trials* podados:** 70

La Figura B.53 muestra la evolución del valor objetivo a lo largo de los *trials*.

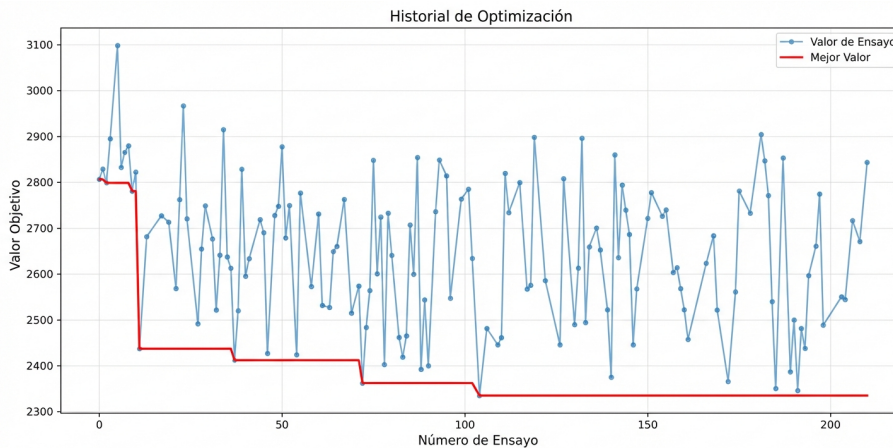


Figura B.53: Historial del proceso de optimización realizado con Optuna para PINN-E RNN.

Tabla B.49: Configuración óptima seleccionada por Optuna para PINN-E RNN.

Hiperparámetro	Valor óptimo
Tipo de celda recurrente	GRU
Tamaño del estado oculto	47
Capas recurrentes	2
<i>Dropout</i> recurrente	0.16
Capas MLP	5
Neuronas por capa MLP	[89, 150, 68, 38, 30]
Función de activación	gelu
Tasa de aprendizaje	0.00137
<i>Dropout</i> base MLP	0.20
<i>Batch size</i>	256
<i>Weight decay</i>	1.89×10^{-6}

La configuración óptima se resume en la Tabla B.49.

A diferencia de lo observado en las formulaciones PINN-P (donde Optuna seleccionaba una celda LSTM), en la PINN-E el proceso de optimización identificó una **celda GRU** como la más adecuada. La GRU, al disponer de un mecanismo de compuertas más simple (sin celda de estado separada), puede resultar más eficiente para capturar la dinámica relativamente gradual de la evapotranspiración, reduciendo el número de parámetros sin sacrificar capacidad de representación.

La arquitectura seleccionada presenta un componente recurrente compacto (2 capas GRU con 47 unidades ocultas) seguido de un módulo MLP considerablemente más profundo (**5 capas**) con una configuración no monótona en el número de neuronas (89→150→68→38→30). Esta estructura profunda del componente *feedforward* sugiere que la estimación de evapotranspiración requiere una transformación no lineal más compleja que la estimación de precipitación.

El *batch size* seleccionado (256) es significativamente mayor que el utilizado en las formulaciones PINN-P (16–32), lo que puede favorecer la estabilidad del gradiente durante el entrenamiento y es consistente con la mayor cantidad de datos procesados en secuencias de 8 días. El *weight decay* resulta muy bajo (1.89×10^{-6}), indicando no ser necesario una regularización L2.

La Figura B.54 presenta las curvas de aprendizaje correspondientes a los cinco mejores *trials*.

La Figura B.55 muestra el historial de *pruning*.

Es notable que la tasa de poda resultó significativamente más alta que en las formulaciones PINN-P (70 de 141 *trials* podados, $\approx 50\%$), lo que sugiere que el espacio de hiperparámetros de la PINN-E contiene regiones más amplias de bajo desempeño. No obstante, la convergencia del proceso fue adecuada, permitiendo identificar una configuración arquitectónica robusta.

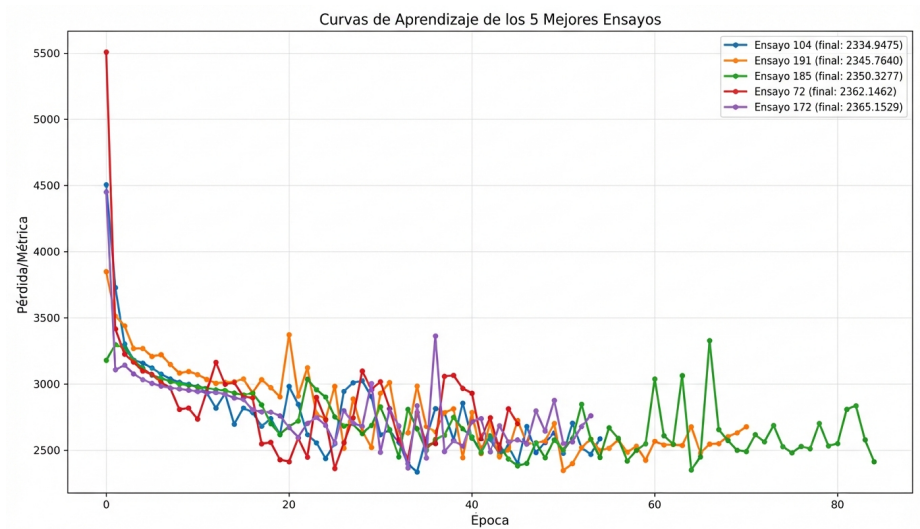


Figura B.54: Curvas de aprendizaje de los cinco mejores *trials* del proceso de optimización de Optuna (PINN-E RNN).

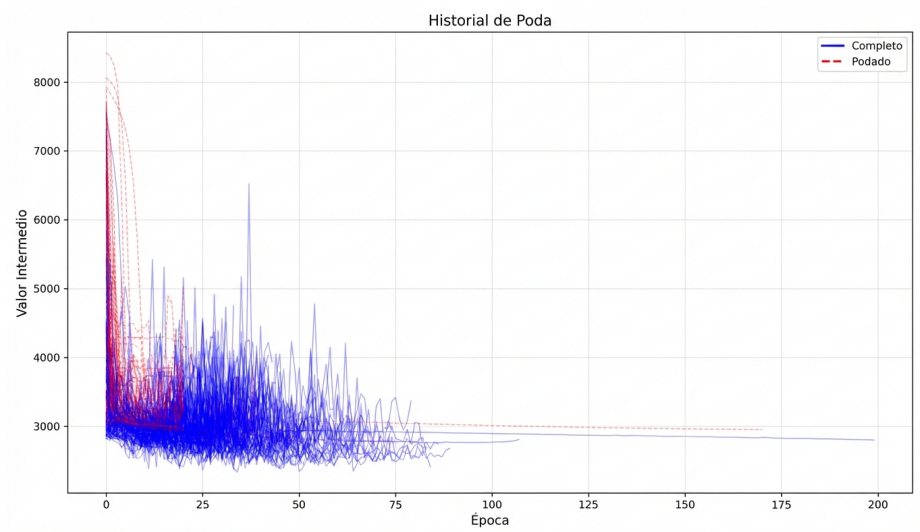


Figura B.55: Historial de *pruning* durante la optimización con Optuna (PINN-E RNN).

B.6.8. Evaluación extendida del modelo optimizado

El modelo final PINN-E RNN se entrenó utilizando la arquitectura óptima identificada mediante Optuna (Sección B.6.7), con el balance de pérdida $w_{\text{data}} = 1.00$ y $w_{\text{phys}} = 1.00$, una ventana temporal de 8 días, y el conjunto de variables definido en las etapas previas.

La evaluación se realizó de forma independiente sobre los conjuntos de entrenamiento, validación y prueba, preservando la partición temporal original y aplicando el mismo criterio de enmascaramiento para los registros con caudal faltante. Las métricas consideradas fueron MAE, RMSE, NSE y PBIAS. La Tabla B.50 presenta el desempeño del modelo en los distintos conjuntos.

Tabla B.50: Desempeño del modelo PINN-E optimizado (RNN) en los distintos conjuntos.

Conjunto	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Entrenamiento	0.81	-0.58	12.65	43.88
Validación	0.78	-2.47	14.38	49.93
Prueba	0.77	4.42	14.23	62.36

A modo de referencia, el modelo GR4J ejecutado de forma directa sobre el mismo período de prueba obtiene $NSE = 0.69$, $RMSE = 72.36$ m³/s y $PBIAS = -16.08$ %.

Análisis del desempeño

En el conjunto de entrenamiento, el modelo alcanza un $NSE = 0.81$ con $RMSE = 43.88$ m³/s y un sesgo prácticamente nulo ($PBIAS = -0.58$ %), lo que indica un ajuste muy preciso a los datos de entrenamiento tanto en variabilidad como en volumen total. En validación, el desempeño se mantiene elevado ($NSE = 0.78$, $RMSE = 49.93$ m³/s), con un sesgo leve ($PBIAS = -2.47$ %), lo que evidencia una buena capacidad de generalización y ausencia de sobreajuste significativo.

En el conjunto de prueba, el modelo PINN-E RNN optimizado alcanza un $NSE = 0.77$ y un $RMSE = 62.36$ m³/s, con un sesgo volumétrico positivo moderado ($PBIAS = 4.42$ %) que indica una leve tendencia a la subestimación del volumen total de caudal. La degradación gradual del NSE desde entrenamiento (0.81) hacia validación (0.78) y prueba (0.77) resulta moderada y consistente, lo que refleja una generalización temporal robusta. El incremento del $RMSE$ en prueba (62.36 m³/s frente a 49.93 m³/s en validación) sugiere la influencia de eventos extremos propios del período de evaluación.

El sesgo resulta considerablemente menor en magnitud que el observado en las formulaciones PINN-P, tanto con MLP ($PBIAS = -35.55$ %) como con RNN ($PBIAS = -22.64$ %), y de signo opuesto.

En comparación con la versión MLP del modelo PINN-E optimizado ($NSE = 0.71$, $PBIAS = -30.62$ % en prueba), la RNN presenta una mejora sustancial

tanto en eficiencia (+6 puntos porcentuales de NSE) como en control del sesgo volumétrico. Este resultado indica que la arquitectura recurrente logra capitalizar de forma más efectiva las ventajas de la formulación PINN-E.

Un aspecto particularmente relevante es que el modelo PINN-E RNN **supera claramente al modelo GR4J directo** en el conjunto de prueba ($NSE = 0.77$ vs 0.69, mejora del 11.8 %), lo que indica que la estimación de evapotranspiración mediante la red recurrente, acoplada con GR4J, aporta una mejora neta significativa respecto al modelo conceptual alimentado directamente con forzantes observados. Este contraste es notable frente a la formulación PINN-P, donde ninguna de las arquitecturas lograba superar al GR4J directo.

La Figura B.56 muestra la evolución de la pérdida durante el entrenamiento.

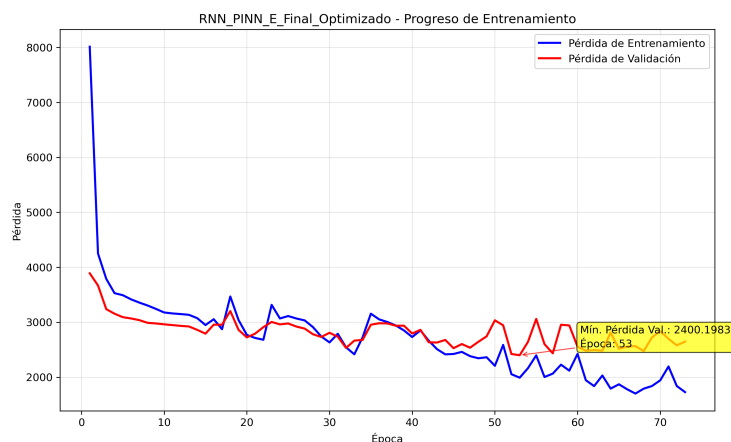


Figura B.56: Evolución de la pérdida de entrenamiento y validación del modelo PINN-E optimizado (RNN).

La Figura B.56 muestra un proceso de convergencia significativamente más estable que el observado en el modelo base. Ambas curvas descienden de forma conjunta durante las primeras 15 épocas, manteniéndose próximas entre sí hasta aproximadamente la época 45. La pérdida de validación alcanza su mínimo global de 2400.20 en la **época 53**, valor considerablemente inferior al del modelo base (2831.82 en época 2), lo que confirma la efectividad de la optimización arquitectónica.

A partir de la época 53, la pérdida de entrenamiento continúa disminuyendo (~ 1700 – 2000), mientras que la de validación se estabiliza en torno a 2400–2600. La separación entre curvas es mucho más contenida que en el modelo base y que en las formulaciones PINN-P, indicando un mejor control del sobreajuste gracias a la regularización moderada ($dropout$ 0.20) y al $batch$ size elevado (256).

Análisis gráfico del ajuste

La Figura B.57 presenta la comparación temporal entre caudal observado y predicho en el conjunto de prueba.

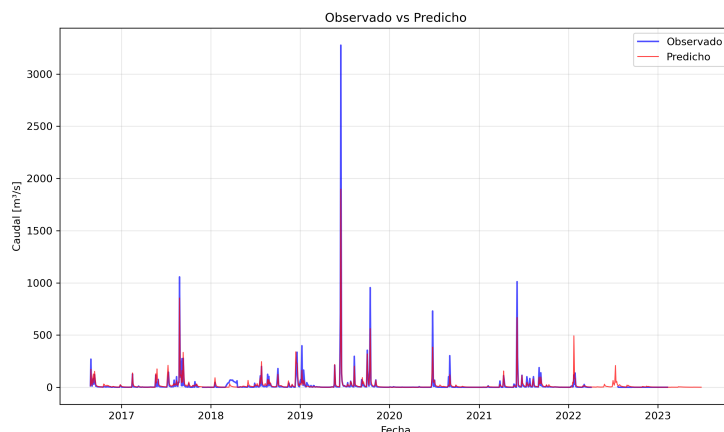


Figura B.57: Serie temporal de caudales observados y predichos por el modelo PINN-E optimizado (RNN) en el conjunto de prueba.

El modelo reproduce adecuadamente la dinámica general de la serie, capturando tanto los períodos de estiaje como la mayoría de los eventos de crecida intermedia a lo largo del período 2016–2023. Los caudales medios y bajos son bien representados, y los eventos de magnitud intermedia ($\sim 300\text{--}500\text{ m}^3/\text{s}$) son razonablemente capturados en su ocurrencia y magnitud.

El evento extremo de mediados de 2019 ($\sim 3300\text{ m}^3/\text{s}$) es parcialmente reproducido con un pico de $\sim 1400\text{ m}^3/\text{s}$, evidenciando una subestimación en eventos de gran magnitud. No obstante, otros eventos significativos (2017, 2019 previos, 2021) son mejor capturados que en la formulación PINN-P. La leve sobreestimación global ($PBIAS = 4.42\%$) no resulta visualmente evidente, lo que indica que se distribuye de forma difusa a lo largo de la serie.

El *scatterplot* en escala logarítmica (Figura B.58) permite evaluar la alineación entre caudales observados y predichos en distintos rangos de magnitud.

Se observa una buena alineación con la línea 1:1 en el rango medio-alto de caudales ($\sim 10\text{--}500\text{ m}^3/\text{s}$), con la nube de puntos distribuida de forma relativamente simétrica en torno a la diagonal. Para caudales bajos ($< 1\text{ m}^3/\text{s}$) se aprecia una tendencia a la sobrepredicción, con puntos ubicados por encima de la línea 1:1. En el rango de caudales extremos ($> 1000\text{ m}^3/\text{s}$), los puntos se mantienen cercanos a la diagonal, indicando un mejor ajuste en eventos de gran magnitud respecto a la formulación PINN-P.

El análisis de residuos (Figura B.59) permite evaluar la distribución temporal de los errores del modelo.

Los residuos se mantienen centrados en torno a cero durante la mayor parte del período, sin patrones sistemáticos crecientes o decrecientes en el tiempo. El

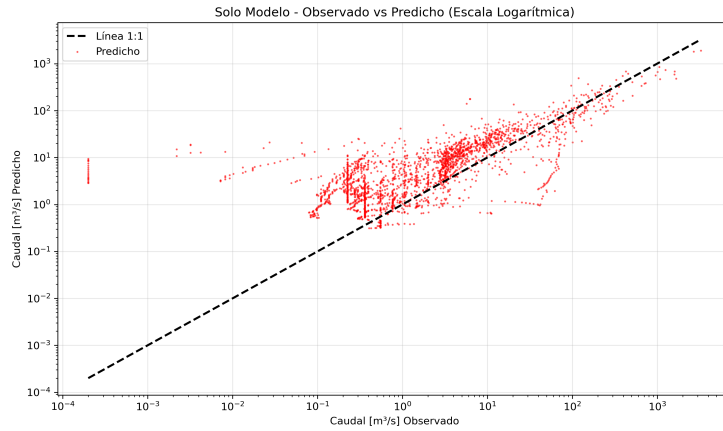


Figura B.58: Diagrama observado vs predicho en escala logarítmica para el modelo PINN-E optimizado (RNN).

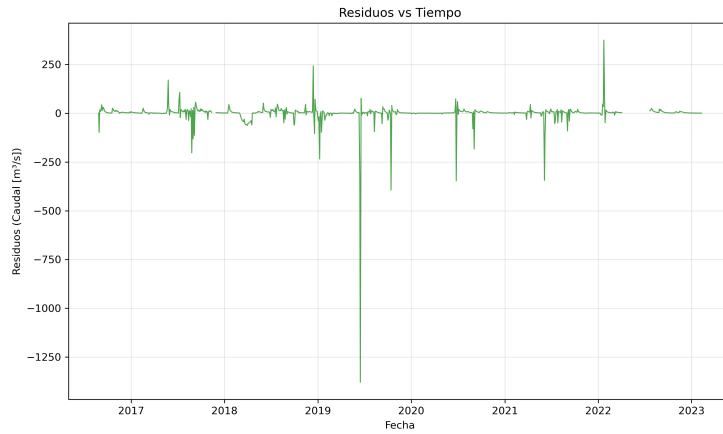


Figura B.59: Residuos del modelo PINN-E optimizado (RNN) en función del tiempo.

residuo negativo más pronunciado ($\sim -1\,300\text{ m}^3/\text{s}$) corresponde al evento extremo de mediados de 2019. Se observan residuos positivos de magnitud moderada ($\sim 150\text{--}350\text{ m}^3/\text{s}$) distribuidos a lo largo del período, particularmente en 2017, 2019 y 2022. En general, la amplitud de los residuos es menor que la observada en la formulación PINN-P RNN ($\pm 1\,600\text{ m}^3/\text{s}$ vs $\pm 1\,300\text{ m}^3/\text{s}$), lo que es consistente con el mejor desempeño global del modelo.

Análisis de atribución final mediante *Integrated Gradients*

Con el objetivo de interpretar el comportamiento del modelo optimizado, se realizó un análisis de relevancia de variables utilizando el método *Integrated Gradients* (IG) sobre el conjunto de prueba.

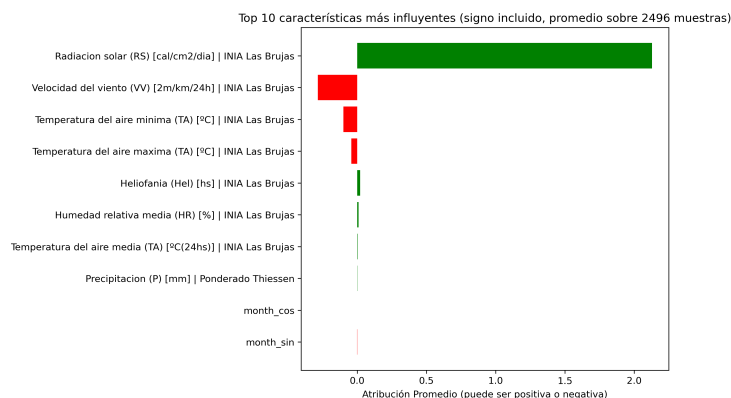


Figura B.60: Top características más influyentes según *Integrated Gradients* en el conjunto de prueba (modelo PINN-E optimizado RNN).

La Figura B.60 muestra que la **radiación solar (RS)** concentra la mayor atribución positiva promedio, mientras que la **velocidad del viento (VV)** presenta la mayor atribución negativa. Este patrón es coherente con la dependencia física de la evapotranspiración respecto a la energía disponible y es consistente con lo observado en la versión MLP de este modelo, donde la radiación solar también dominaba la distribución de relevancias.

Es interesante notar que el patrón de atribuciones del modelo final optimizado difiere del observado durante las etapas exploratorias (Sección B.6.5), donde la velocidad del viento presentaba la mayor atribución positiva y la radiación solar una atribución negativa. Este cambio puede atribuirse a la optimización arquitectónica realizada mediante Optuna, que modificó tanto el tipo de celda recurrente (de LSTM a GRU) como la profundidad del módulo MLP posterior, alterando la forma en que el modelo pondera las distintas variables de entrada.

No obstante, debe considerarse que, en esta implementación, las *Integrated Gradients* se evaluaron únicamente sobre el primer paso temporal de la ventana de entrada. Dado que las redes recurrentes integran información a lo largo de

toda la secuencia mediante su estado oculto, es posible que parte de la contribución asociada a las variables meteorológicas se encuentre distribuida en pasos temporales anteriores dentro de la ventana, los cuales no fueron descompuestos explícitamente en el análisis de atribuciones. En consecuencia, la interpretación presentada debe entenderse como una aproximación parcial a la dinámica completa de relevancia temporal del modelo.

Costo computacional

El entrenamiento completo del modelo optimizado demandó 683.53 segundos en un total de 73 épocas en el entorno de cómputo utilizado. Por su parte, el tiempo de inferencia fue de 3.73 segundos, de acuerdo con los registros del sistema.

B.7. PINN-PE MLP

Esta sección presenta el desarrollo detallado del modelo PINN-PE con arquitectura MLP, incluyendo todos los experimentos realizados durante el proceso de optimización.

El cuerpo principal (Sección 6.10) resume únicamente los resultados del modelo base y del modelo final optimizado. Aquí se documentan las etapas intermedias que condujeron a la configuración seleccionada.

B.7.1. Modelo base

El primer experimento para la formulación **PINN-PE** consistió en entrenar una red MLP haciendo que la red prediga simultáneamente la precipitación agregada diaria \hat{P}_{agg} y la evapotranspiración agregada diaria \hat{E}_{agg} . A diferencia de las variantes PINN-P y PINN-E, en este caso ninguna de las dos variables forzantes se mantiene fija a partir de observaciones, sino que ambas emergen como salidas del modelo neuronal.

La predicción final de caudal no es producida directamente por la red, sino que se obtiene a partir del acoplamiento físico con GR4J, evaluando el caudal simulado $Q_{GR4J}(\hat{P}_{agg}, \hat{E}_{agg}; \theta)$ y comparándolo con el caudal observado mediante el término físico de la función de pérdida.

Dado que el conjunto de prueba se reserva exclusivamente para la evaluación final del modelo, el desempeño en esta etapa se reporta sobre el **conjunto de validación**, utilizando MAE, RMSE, NSE y PBIAS calculadas sobre el caudal simulado por GR4J.

La Figura B.61 muestra la evolución de la pérdida durante el entrenamiento del modelo. Se observa una reducción pronunciada tanto en entrenamiento como en validación durante las primeras épocas, seguida por una estabilización temprana del desempeño en validación. En particular, la pérdida de validación alcanza su mínimo en la época 4, a partir de la cual se aprecia una separación

Tabla B.51: Desempeño del modelo base PINN-PE (MLP) sobre el conjunto de validación.

NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
0.67	-0.60	16.49	61.30

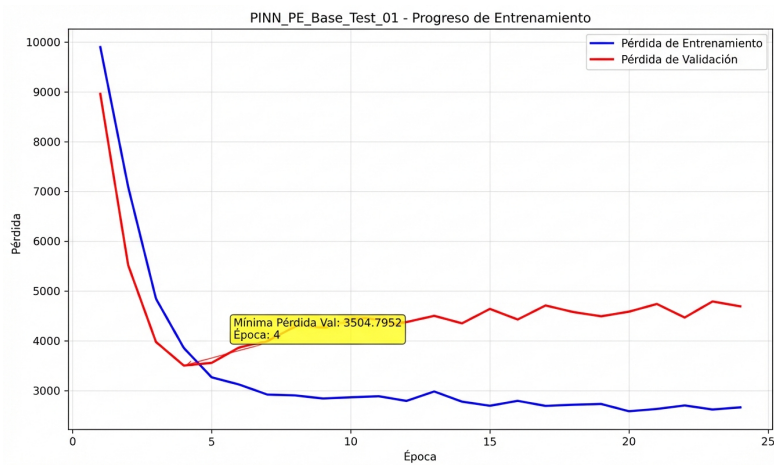


Figura B.61: Evolución de la función de pérdida (entrenamiento y validación) para el modelo base PINN-PE (MLP).

progresiva entre ambas curvas: la pérdida de entrenamiento continúa disminuyendo, mientras que la pérdida de validación comienza a incrementarse y a presentar oscilaciones moderadas.

Este patrón es consistente con la aparición temprana de *overfitting*, indicando que el modelo comienza a especializarse en el conjunto de entrenamiento sin traducir ese ajuste en mejoras sostenidas sobre datos no vistos. En el contexto PINN, este comportamiento resulta relevante porque sugiere que, aun con el acoplamiento físico con GR4J, el equilibrio inicial entre los términos de pérdida y la configuración base del modelo no alcanza para sostener una mejora estable en generalización, lo cual justifica las etapas posteriores de ajuste (incorporación de variables, selección de *lags* y análisis del balance entre términos de pérdida).

En términos cuantitativos (Tabla B.51), el modelo base presenta un desempeño razonable sobre validación, con $NSE = 0.67$, lo que indica que el sistema acoplado explica aproximadamente dos tercios de la variabilidad observada del caudal en este subconjunto. El error cuadrático medio ($RMSE = 61.30 \text{ m}^3/\text{s}$) se mantiene en un rango moderado, mientras que el sesgo porcentual resulta prácticamente neutro ($PBIAS = -0.60\%$).

En conjunto, estos resultados muestran que la configuración base de la PINN-PE es capaz de capturar una fracción importante de la dinámica de caudales a través del modelo GR4J forzado con \hat{P}_{agg} y \hat{E}_{agg} .

B.7.2. Evaluación de la incorporación de variables

En esta etapa se analiza el impacto de distintas configuraciones de variables de entrada sobre el desempeño del modelo PINN-PE basado en arquitectura MLP. Tal como se detalló en la metodología, el objetivo es determinar qué combinación de variables mejora la capacidad del sistema acoplado (red neuronal + GR4J) para reproducir el caudal observado en el conjunto de validación.

En la formulación PINN-PE la red neuronal predice simultáneamente la precipitación agregada diaria \hat{P}_{agg} y la evapotranspiración agregada diaria \hat{E}_{agg} , las cuales son utilizadas como forzantes del modelo GR4J. En consecuencia, cualquier modificación en las variables de entrada puede afectar tanto la calidad de ambas estimaciones como la coherencia hidrológica del caudal simulado.

Las métricas reportadas en esta sección se calculan sobre el conjunto de validación y corresponden al caudal generado por GR4J, evaluado mediante MAE, RMSE, NSE y PBIAS. En todos los experimentos se mantuvo constante la arquitectura base del modelo y los pesos de la función de pérdida, de modo que las diferencias observadas puedan atribuirse exclusivamente a la configuración de variables considerada.

Variables temporales

En primer lugar, se evaluó el efecto de incorporar variables temporales representadas de forma cíclica mediante transformaciones seno y coseno del *mes* y del *día*. La inclusión de estas variables busca capturar patrones estacionales

que puedan influir en la estimación conjunta de la precipitación y la evapotranspiración agregadas, y en consecuencia en la simulación del caudal a través de GR4J.

Se compararon cuatro configuraciones: (i) sin variables temporales, (ii) incorporación únicamente del mes, (iii) incorporación únicamente del día, y (iv) incorporación conjunta de día y mes. Los resultados obtenidos se presentan en la Tabla B.52.

Tabla B.52: Comparación de desempeño para distintas combinaciones de variables temporales sobre el conjunto de validación (PINN-PE MLP).

Configuración	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Ninguna	0.67	-0.60	16.49	61.30
día	0.66	9.53	15.69	62.37
mes	0.67	7.69	15.51	61.26
día, mes	0.67	7.34	15.58	61.57

Los resultados muestran que las configuraciones sin variables temporales, con *mes* y con *día, mes* presentan valores de eficiencia prácticamente equivalentes ($NSE = 0.67$). En términos de $RMSE$, las diferencias son marginales y no evidencian mejoras sustanciales asociadas a la incorporación de información temporal explícita.

Sin embargo, sí se observan diferencias relevantes en el sesgo volumétrico. La configuración sin variables temporales presenta un $PBIAS$ prácticamente neutro (-0.60%), mientras que la incorporación de *mes* o de *día, mes* incrementa el sesgo positivo hasta valores cercanos al 7–8 %. La inclusión exclusiva de *día* produce además un deterioro leve en eficiencia ($NSE = 0.66$) y el mayor sesgo entre las configuraciones evaluadas.

Desde una perspectiva hidrológica, este comportamiento sugiere que la información estacional explícita no aporta mejoras claras en eficiencia cuando la red ya dispone de variables meteorológicas energéticas y térmicas que capturan indirectamente la variabilidad anual. En el caso particular de PINN-PE, donde ambas forzantes (\hat{P}_{agg} y \hat{E}_{agg}) son estimadas por la red, la incorporación de variables temporales puede alterar el equilibrio conjunto entre precipitación y evapotranspiración, afectando el balance volumétrico del sistema acoplado.

Considerando simultáneamente la estabilidad en eficiencia y el mejor control del sesgo volumétrico, se decidió **no incorporar variables temporales** en el modelo PINN-PE para las etapas posteriores del análisis.

Incorporación de *lags* de precipitación y evapotranspiración

En esta etapa se evaluó la incorporación de memoria temporal explícita mediante valores retardados (*lags*) de precipitación y evapotranspiración como variables de entrada del modelo PINN-PE.

A diferencia de la formulación PINN-E, en este caso la red neuronal predice

simultáneamente \hat{P}_{agg} y \hat{E}_{agg} . Por este motivo, no resulta metodológicamente apropiado fijar una única forma de agregación espacial de la precipitación desde el inicio. En consecuencia, se realizó una prueba preliminar combinando distintos esquemas de *lags* y diferentes formas de agregación espacial de la precipitación, con el objetivo de identificar la configuración más prometedora para una exploración posterior más detallada.

En todos los casos se incorporaron inicialmente 7 días de *lags*. Se evaluaron las siguientes configuraciones:

- *Lags* de precipitación únicamente, bajo distintos esquemas de agregación espacial.
- *Lags* de evapotranspiración únicamente.
- Incorporación conjunta de *lags* de precipitación y evapotranspiración.
- Configuración sin *lags* (modelo base).

En la configuración conjunta, los *lags* de precipitación se construyeron utilizando la agregación espacial mediante polígonos de Thiessen, manteniendo coherencia con la representación hidrológicamente consistente del forzante pluviométrico.

Las métricas se calcularon sobre el conjunto de validación, evaluando el caudal simulado por GR4J. Los resultados se presentan en la Tabla B.53.

Tabla B.53: Comparación preliminar de configuraciones con *lags* de precipitación y evapotranspiración (PINN-PE MLP).

Configuración	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Sin <i>lags</i>	0.67	-0.60	16.49	61.30
P (7) – ponderado distancia	0.68	8.06	15.90	60.77
P (7) – Thiessen	0.67	8.70	15.90	61.49
P (7) – promedio	0.66	10.00	15.99	62.09
P (7) – por estación	0.65	7.18	15.99	63.28
E (7)	0.65	10.72	16.11	63.22
P (7) + E (7) – Thiessen	0.66	7.03	16.41	62.36

En términos de eficiencia, la configuración con 7 *lags* de precipitación ponderados por distancia alcanza el mayor valor de *NSE* (0.68) y el menor *RMSE* (60.77 m³/s). No obstante, esta mejora relativa en eficiencia se acompaña de un incremento sustancial del sesgo volumétrico (*PBIAS* = 8.06%), indicando una tendencia marcada a la subestimación del volumen total simulado.

De manera similar, las restantes configuraciones que incorporan *lags* presentan valores de *NSE* comprendidos entre 0.65 y 0.67, pero en todos los casos el

PBIAS se incrementa de forma significativa respecto al modelo base, alcanzando valores entre 7 % y 11 %. La incorporación exclusiva de *lags* de evapotranspiración conduce además a un deterioro adicional en eficiencia ($NSE = 0.65$) y a uno de los mayores sesgos observados.

En contraste, la configuración sin *lags* mantiene un *PBIAS* prácticamente neutro (-0.60%), lo que indica un adecuado control del balance hídrico global del sistema acoplado. Si bien algunas configuraciones con memoria temporal muestran mejoras marginales en *NSE* o reducciones leves en *RMSE*, dichas variaciones resultan cuantitativamente pequeñas frente al incremento considerable del sesgo volumétrico.

Desde una perspectiva hidrológica, el control del volumen total simulado constituye un aspecto crítico en la evaluación del modelo. En este contexto, el deterioro sistemático del *PBIAS* asociado a la incorporación de *lags* no se ve compensado por las leves variaciones observadas en las demás métricas.

En consecuencia, considerando prioritariamente el control del sesgo volumétrico y el equilibrio global del sistema, se decidió **no incorporar *lags* de precipitación ni de evapotranspiración en el modelo PINN-PE**, manteniendo la configuración sin memoria temporal explícita para las etapas posteriores del análisis.

B.7.3. Depuración inicial por correlación

Siguiendo el mismo procedimiento aplicado en las formulaciones anteriores, se realizó un análisis de correlación entre las variables meteorológicas con el objetivo de identificar posibles redundancias y evaluar la factibilidad de reducir la dimensionalidad del modelo sin deteriorar su desempeño.

Al igual que en PINN-E, la variable evapotranspiración observada no forma parte del conjunto de entrada, ya que en la formulación PINN-PE la red neuronal predice directamente \hat{E}_{agg} . En consecuencia, las combinaciones evaluadas se restringen a variables térmicas y radiativas, sin considerar escenarios que involucren la eliminación de evapotranspiración como predictor.

En todos los casos se mantuvo la configuración sin *lags* y sin variables temporales definida en las etapas previas. Las métricas se calcularon sobre el conjunto de validación, evaluando el caudal simulado por GR4J. Los resultados se presentan en la Tabla B.54.

Las configuraciones que eliminan heliofanía o radiación solar alcanzan valores de eficiencia levemente superiores ($NSE = 0.68$). Sin embargo, estas mejoras se acompañan de incrementos significativos en el sesgo volumétrico, con *PBIAS* superiores al 8 %, indicando una subestimación sistemática del volumen total simulado.

En contraste, el modelo base mantiene un *PBIAS* cercano a cero (-0.60%), lo que evidencia un mejor control del balance hídrico global. Si bien el escenario con sólo temperatura media logra un sesgo también reducido (-1.01%), no presenta mejoras en eficiencia respecto al modelo base y mantiene errores ligeramente superiores en términos de *RMSE* y *MAE*.

Tabla B.54: Resultados de desempeño para configuraciones de reducción de variables correlacionadas (PINN-PE MLP).

Configuración	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Modelo base	0.67	-0.60	16.49	61.30
Remove heliofanía	0.68	10.08	15.53	60.75
Remove radiación solar	0.68	8.81	15.54	60.94
Sólo temperatura media	0.67	-1.01	16.50	61.60
Temp. media + Remove Heliofanía	0.67	2.72	16.14	61.78
Temp. media + Remove Radiación Sol.	0.66	3.08	16.21	62.35

Las configuraciones con mayor reducción estructural (eliminación simultánea de variables radiativas y simplificación térmica) no muestran mejoras consistentes en ninguna de las métricas evaluadas y, en algunos casos, deterioran tanto eficiencia como control del sesgo.

Desde una perspectiva hidrológica, la leve mejora en *NSE* observada al eliminar variables radiativas no compensa el incremento sustancial en el sesgo volumétrico. Dado que el control del volumen total constituye un aspecto central en la evaluación del sistema acoplado, se prioriza la estabilidad del balance hídrico por sobre variaciones marginales en eficiencia relativa.

En consecuencia, se decidió **mantener el modelo base con el conjunto completo de variables meteorológicas**, preservando el equilibrio entre eficiencia predictiva y control del sesgo volumétrico en la formulación PINN-PE.

B.7.4. Análisis de relevancia mediante *Integrated Gradients*

Una vez definido el conjunto de variables a partir del análisis de correlación, se evaluó la relevancia individual de cada característica mediante el método de *Integrated Gradients* (IG). Este enfoque permite cuantificar la contribución promedio de cada variable a la predicción del modelo, proporcionando una medida interpretativa consistente con la estructura interna de la red.

El cálculo se realizó por separado sobre los subconjuntos de entrenamiento y validación. Las Figuras B.62a y B.62b presentan las características con mayor atribución promedio (incluyendo el signo) en ambos subconjuntos.

Los resultados muestran una estructura de relevancia altamente consistente entre entrenamiento y validación. En ambos casos, la **radiación solar (RS)** aparece como la variable dominante en magnitud, seguida por la **humedad**

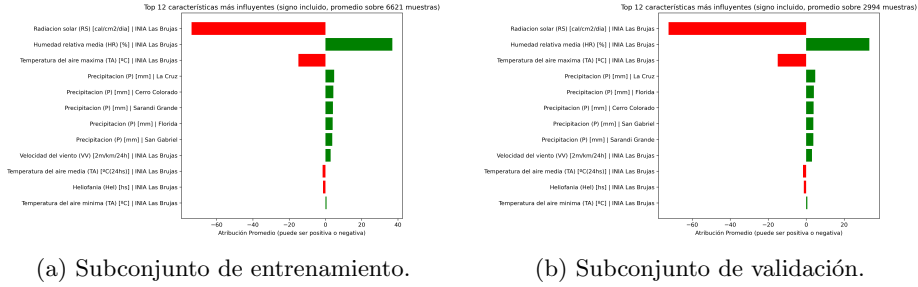


Figura B.62: Importancia relativa de variables según *Integrated Gradients* (IG) en los conjuntos de entrenamiento y validación para el modelo PINN-PE MLP.

relativa media (HR). En un segundo nivel de importancia se ubican variables térmicas y aerodinámicas (temperatura máxima y velocidad del viento), junto con las precipitaciones medidas por estación, con atribuciones positivas de menor magnitud relativa.

Este patrón resulta coherente con la formulación PINN-PE. Dado que la red debe estimar simultáneamente \hat{P}_{agg} y \hat{E}_{agg} , es esperable que variables energéticas y atmosféricas (RS, HR, temperatura, viento) tengan un rol central en la estimación de la componente evaporativa, mientras que la información pluviométrica por estación contribuya a la construcción de una precipitación efectiva consistente con la dinámica de GR4J.

Evaluación empírica de eliminación de variables con baja atribución.

A partir del análisis de IG se identificaron variables con atribución promedio reducida dentro del conjunto de entrada. En particular, se evaluó empíricamente el efecto de eliminar la **temperatura del aire mínima**, con el objetivo de verificar si su remoción podía simplificar el modelo sin deteriorar el desempeño.

La Tabla B.55 compara el desempeño del modelo base frente a la configuración con la variable eliminada, sobre el conjunto de validación.

Tabla B.55: Comparación de desempeño del modelo PINN-PE (MLP) al eliminar una variable de baja atribución según IG.

Configuración	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Modelo base	0.67	-0.60	16.49	61.30
Sin temperatura mínima	0.66	16.17	15.43	61.94

Los resultados muestran que, si bien la eliminación de la temperatura mínima reduce el *MAE*, produce un deterioro en eficiencia (*NSE* desciende de 0.67 a 0.66) y un incremento marcado del sesgo volumétrico, con *PBIAS* = 16.17%. Este aumento indica una subestimación sistemática del volumen total simulado, sustancialmente mayor que la observada en el modelo base.

Desde una perspectiva hidrológica, este resultado es particularmente relevante: aun cuando una variable presente baja atribución promedio en IG, su contribución puede operar como regulador del balance conjunto entre precipitación y evapotranspiración dentro del sistema acoplado con GR4J. En este contexto, el deterioro del *PBIAS* no se ve compensado por mejoras parciales en error medio absoluto.

En consecuencia, priorizando el control del sesgo volumétrico y la estabilidad del balance hídrico global, se decidió **mantener la temperatura mínima dentro del conjunto final de variables**, dando por finalizado el proceso iterativo de depuración en esta etapa.

Conjunto final de variables

El conjunto final de variables para el modelo PINN-PE (MLP) fue definido a partir de un proceso secuencial que incluyó:

1. Evaluación de variables temporales.
2. Análisis preliminar de incorporación de *lags* de precipitación y evapotranspiración.
3. Depuración de variables correlacionadas.
4. Análisis de relevancia mediante *Integrated Gradients* (IG).

En una primera etapa se evaluó la incorporación de información temporal (día y mes). Los resultados mostraron valores de eficiencia similares ($NSE = 0.67$) respecto a la configuración sin variables temporales explícitas, pero con un incremento considerable del sesgo volumétrico en las configuraciones que incorporaban dichas variables. Dado que el modelo base presentaba un *PBIAS* prácticamente neutro, se decidió no incluir variables temporales en la formulación final.

Posteriormente se analizó la incorporación de memoria temporal mediante *lags* de precipitación y evapotranspiración, considerando distintos esquemas de agregación espacial en el caso de la precipitación. Si bien una configuración particular alcanzó un $NSE = 0.68$, esta mejora fue acompañada por incrementos sustanciales en el sesgo volumétrico (valores superiores al 8 %). En contraste, el modelo sin *lags* mantuvo un *PBIAS* cercano a cero, evidenciando un mejor control del balance hídrico global. En consecuencia, se decidió no incorporar memoria temporal explícita en el modelo PINN-PE.

El análisis de correlación permitió evaluar configuraciones reducidas eliminando variables radiativas y térmicas potencialmente redundantes. Algunas combinaciones mostraron leves incrementos en eficiencia ($NSE = 0.68$), pero nuevamente asociados a incrementos significativos del sesgo volumétrico. El modelo base mantuvo el mejor equilibrio entre eficiencia y control del volumen total simulado, por lo que se preservó el conjunto completo de variables meteorológicas.

Finalmente, el análisis de relevancia mediante *Integrated Gradients* identificó variables con atribución promedio reducida. Se evaluó empíricamente la eliminación de la temperatura mínima; sin embargo, esta modificación produjo un deterioro en el sesgo volumétrico ($PBIAS = 16.17\%$), aun cuando el MAE mostrara una leve mejora. Dado que el control del balance hídrico constituye un criterio prioritario en la evaluación del sistema acoplado, se decidió mantener dicha variable en el modelo.

En función de estas decisiones, el conjunto final de variables del modelo PINN-PE quedó conformado por:

- **Precipitaciones contemporáneas por estación** (sin *lags* ni agregaciones adicionales).
- **Radiación solar (RS) y heliofanía (Hel).**
- **Temperatura del aire** (media, máxima y mínima).
- **Humedad relativa media (HR).**
- **Velocidad del viento (VV).**

Esta configuración representa el mejor equilibrio entre eficiencia predictiva, estabilidad del entrenamiento y control del sesgo volumétrico dentro de la formulación PINN-PE. Sobre este conjunto se desarrollaron las etapas posteriores de análisis del balance entre los términos de la función de pérdida y la optimización de hiperparámetros.

B.7.5. Análisis del balance entre términos de la función de pérdida

En la formulación PINN-PE, la función de pérdida combina un término de datos, asociado al ajuste conjunto de \hat{P}_{agg} y \hat{E}_{agg} , y un término físico, correspondiente al ajuste del caudal simulado por GR4J, ponderados mediante los coeficientes w_{data} y w_{phys} .

Con el conjunto final de variables ya definido, se evaluaron distintas combinaciones de pesos manteniendo fija la arquitectura base del modelo. Se analizaron relaciones comprendidas hasta un máximo de 1:4 en ambos sentidos, excluyendo el caso $w_{phys} = 0$, ya que ello implicaría eliminar completamente la restricción física y desnaturalizar la formulación PINN.

La Tabla B.56 presenta las métricas obtenidas sobre el conjunto de validación para cada combinación evaluada.

Los resultados muestran que el desempeño del modelo es altamente estable frente a variaciones moderadas en la relación entre los términos de pérdida. La eficiencia se mantiene prácticamente constante ($NSE = 0.66-0.67$) en todas las configuraciones evaluadas, mientras que las diferencias en $RMSE$ y MAE resultan marginales.

En términos de sesgo volumétrico, tampoco se observan desbalances extremos: los valores de $PBIAS$ se mantienen en un rango relativamente acotado,

Tabla B.56: Desempeño del modelo PINN-PE (MLP) para distintas combinaciones de pesos en la función de pérdida.

$w_{\text{data}} - w_{\text{phys}}$	NSE	PBIAS [%]	MAE [m^3/s]	RMSE [m^3/s]
0.00 – 1.00	0.67	0.19	16.48	61.64
0.25 – 1.00	0.67	0.96	16.46	61.84
0.25 – 0.75	0.67	-0.11	16.53	61.62
0.50 – 1.00	0.66	1.72	16.41	61.96
0.50 – 0.75	0.67	-0.68	16.54	61.43
0.75 – 1.00	0.67	0.69	16.41	61.57
1.00 – 1.00	0.67	-0.60	16.49	61.30
1.00 – 0.75	0.67	-0.77	16.53	61.29
1.00 – 0.50	0.67	1.42	16.30	61.40
0.75 – 0.50	0.67	-0.60	16.49	61.29
0.75 – 0.25	0.67	2.72	16.12	61.21
1.00 – 0.25	0.67	1.64	16.25	61.16

aunque algunas configuraciones con predominancia del término de datos presentan incrementos hacia valores positivos superiores al 2 %.

Si bien ciertas combinaciones muestran reducciones leves en *RMSE* o *MAE*, estas diferencias no se traducen en mejoras sustanciales en eficiencia ni en un control consistentemente superior del sesgo volumétrico. En términos prácticos, todas las configuraciones evaluadas pueden considerarse equivalentes dentro del nivel de precisión reportado.

En consecuencia, dado que no se identificaron ventajas claras asociadas a una combinación específica de pesos y considerando el principio de simplicidad metodológica, se decidió **mantener la configuración balanceada** $w_{\text{data}} = 1$ y $w_{\text{phys}} = 1$ para las etapas posteriores de optimización.

Esta elección preserva un equilibrio explícito entre ajuste empírico y coherencia física, sin privilegiar artificialmente ninguno de los dos componentes de la función de pérdida.

B.7.6. Optimización de hiperparámetros (Optuna)

Una vez definido el conjunto final de variables y fijado el balance de la función de pérdida en $w_{\text{data}} = 1$ y $w_{\text{phys}} = 1$, se procedió a la optimización de hiperparámetros mediante el *framework* *Optuna*, siguiendo el mismo esquema metodológico aplicado en los modelos PINN-P y PINN-E.

El proceso de búsqueda incluyó hiperparámetros estructurales (número de capas ocultas y neuronas por capa), así como parámetros asociados al entrenamiento (función de activación, tasa de aprendizaje, *dropout*, *weight decay* y *batch size*). La función objetivo correspondió a la pérdida de validación (MSE), evaluada bajo la formulación compuesta propia de la PINN-PE.

El estudio culminó con los siguientes resultados:

- **Mejor valor de pérdida de validación: 2693.20**
- ***Trials* completados: 142**
- ***Trials* podados: 29**

La Figura B.63 muestra la evolución del valor objetivo a lo largo de los *trials*. Durante las primeras iteraciones se observa una fase de exploración amplia, con valores elevados y alta variabilidad. A medida que avanzó el proceso, el algoritmo fue concentrando la búsqueda en regiones más prometedoras del espacio de hiperparámetros, logrando reducciones progresivas en el mejor valor alcanzado hasta estabilizarse en torno a 2693.

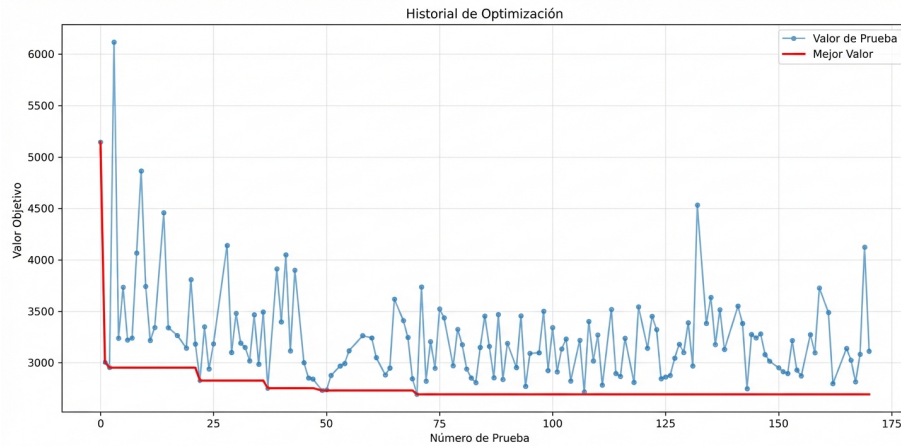


Figura B.63: Historial del proceso de optimización realizado con Optuna para PINN-PE (MLP). La línea azul representa el valor de la función objetivo en cada *trial*, mientras que la línea roja indica el mejor valor alcanzado hasta ese punto.

El mejor resultado se obtuvo en el *trial* 70. La configuración arquitectónica óptima se resume en la Tabla B.57.

La arquitectura seleccionada presenta cuatro capas ocultas con una estructura no monótona en el número de neuronas, incluyendo una segunda capa de gran ancho (970 unidades), lo que sugiere una elevada capacidad representacional intermedia. A diferencia de PINN-E, donde se observó una fuerte regularización mediante valores altos de *dropout*, en PINN-PE el valor óptimo de *dropout* resulta moderado (0.10), indicando un equilibrio distinto entre capacidad y regularización.

Asimismo, la función de activación seleccionada fue *gelu*, a diferencia de las formulaciones previas donde predominó *relu*. Este resultado sugiere que la estimación conjunta de precipitación y evapotranspiración bajo restricción física puede beneficiarse de funciones de activación más suaves, capaces de capturar transiciones no lineales de forma más estable.

Tabla B.57: Configuración óptima seleccionada por Optuna para PINN-PE (MLP).

Hiperparámetro	Valor óptimo
Número de capas ocultas	4
Neuronas por capa	[298, 970, 33, 87]
Función de activación	gelu
Tasa de aprendizaje	1.90×10^{-4}
Dropout base	0.10
Batch size	16
Weight decay	2.63×10^{-6}

La tasa de aprendizaje óptima (1.90×10^{-4}) es considerablemente más baja que la obtenida en PINN-P y también inferior a la de PINN-E, lo que indica un proceso de optimización más conservador, posiblemente asociado a la mayor complejidad de la formulación conjunta.

La Figura B.64 presenta las curvas de aprendizaje correspondientes a los cinco mejores *trials*. Se observa que, aunque convergen hacia valores similares de pérdida, existe variabilidad en la estabilidad de las trayectorias, con algunos ensayos mostrando oscilaciones intermedias más pronunciadas.

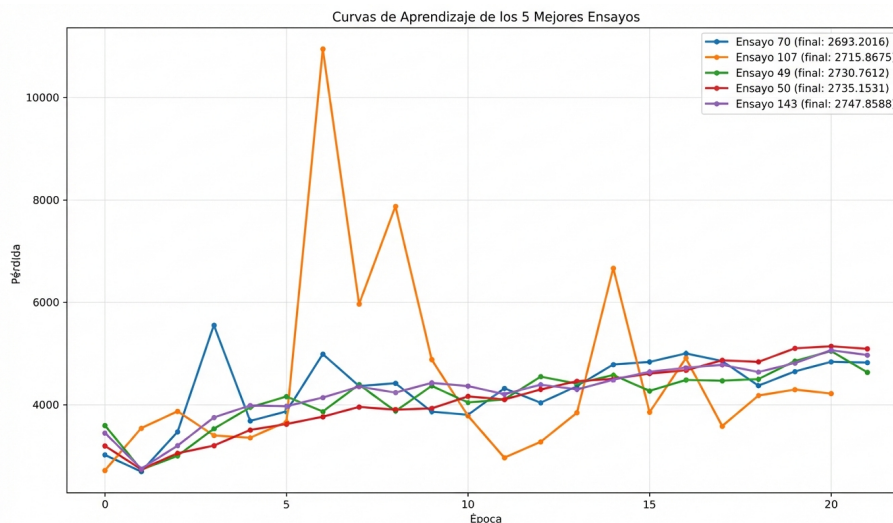


Figura B.64: Curvas de aprendizaje de los cinco mejores *trials* del proceso de optimización de Optuna (PINN-PE MLP).

Finalmente, la Figura B.65 muestra el historial de *pruning*. La poda permitió descartar configuraciones poco prometedoras en etapas tempranas, reduciendo el costo computacional total sin comprometer la calidad del resultado final.

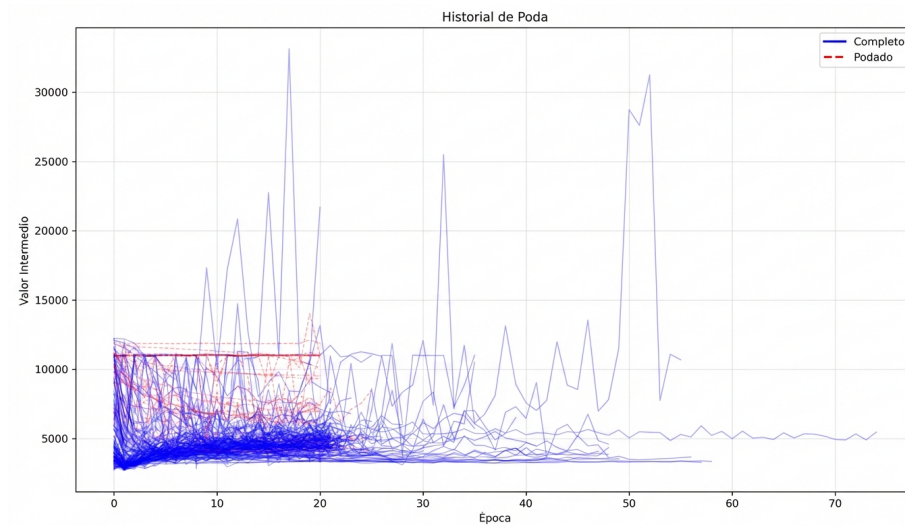


Figura B.65: Historial de *pruning* durante la optimización con Optuna (PINN-PE MLP).

En conjunto, la optimización mediante Optuna permitió identificar una arquitectura profunda y de alta capacidad intermedia, con una estrategia de entrenamiento conservadora y regularización moderada. Estos resultados sugieren que la formulación PINN-PE —al estimar simultáneamente precipitación y evapotranspiración— introduce un problema de optimización de mayor complejidad estructural que las formulaciones univariadas, requiriendo una arquitectura flexible pero cuidadosamente regularizada para mantener estabilidad en el entrenamiento.

B.7.7. Evaluación extendida del modelo optimizado

El modelo final PINN-PE (MLP) se entrenó utilizando la arquitectura óptima identificada mediante Optuna (Sección B.7.6), manteniendo el balance de pérdida $w_{\text{data}} = 1$ y $w_{\text{phys}} = 1$, y el conjunto final de variables definido en las etapas previas.

La evaluación se realizó de forma independiente sobre los conjuntos de entrenamiento, validación y prueba, preservando la partición temporal original y aplicando el mismo criterio de enmascaramiento para los registros con caudal faltante. (Tabla B.58) Las métricas consideradas fueron MAE, RMSE, NSE y PBIAS.

Análisis del desempeño

En el conjunto de validación el modelo alcanza un desempeño elevado, con un $NSE = 0.74$ y un $RMSE = 54.08 \text{ m}^3/\text{s}$, en línea con las mejores configuraciones

Tabla B.58: Desempeño del modelo PINN-PE optimizado en los distintos conjuntos.

Conjunto	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Entrenamiento	0.61	-62.82	21.58	63.70
Validación	0.74	-14.42	16.44	54.08
Prueba	0.72	-22.95	17.29	68.70

intermedias observadas durante el desarrollo del modelo. En el conjunto de prueba se mantiene una eficiencia adecuada ($NSE = 0.72$), lo que indica una capacidad razonable de generalización temporal.

No obstante, se observa un sesgo porcentual negativo en todos los conjuntos, particularmente pronunciado en entrenamiento ($PBIAS = -62.82\%$) y también significativo en prueba ($PBIAS = -22.95\%$). Este resultado evidencia una tendencia sistemática a la sobreestimación del volumen total de caudal, más marcada que en las formulaciones PINN-P y PINN-E evaluadas previamente.

Al igual que en los otros enfoques PINN, el desempeño en entrenamiento resulta inferior al de validación ($NSE = 0.61$ frente a 0.74). Este comportamiento, aparentemente contraintuitivo, se explica por la naturaleza compuesta de la función de pérdida: el proceso de optimización no minimiza exclusivamente el error directo sobre caudal, sino una combinación ponderada entre el término de datos (precipitación y evapotranspiración estimadas) y el término físico (caudal simulado por GR4J).

La Figura B.66 muestra la evolución de la pérdida durante el entrenamiento. Se observa que la pérdida de validación alcanza su mínimo en una etapa muy temprana (época 2), mientras que la pérdida de entrenamiento permanece sistemáticamente por encima. Este patrón refuerza la idea de que el criterio de optimización prioriza configuraciones que mejoran el equilibrio global del sistema acoplado aunque no necesariamente reduzcan el error puntual sobre el conjunto de entrenamiento.

Análisis gráfico del ajuste

La Figura B.67 presenta la comparación temporal entre caudal observado y predicho en el conjunto completo. El modelo reproduce adecuadamente la dinámica general de la serie, capturando la ocurrencia de eventos de crecida intermedia y la estacionalidad. Sin embargo, se observa una subestimación en varios eventos extremos, y una sobreestimación en los picos medianos y bajos, lo cual contribuye al sesgo volumétrico negativo observado en las métricas globales.

El *scatterplot* en escala logarítmica (Figura B.68) muestra una buena alineación con la línea 1:1 en el rango medio de caudales, aunque por encima de la línea, marcando una sobreestimación, mientras que para valores elevados se observa dispersión creciente y predominio de puntos por debajo de la diagonal, indicando subestimación en eventos de gran magnitud.

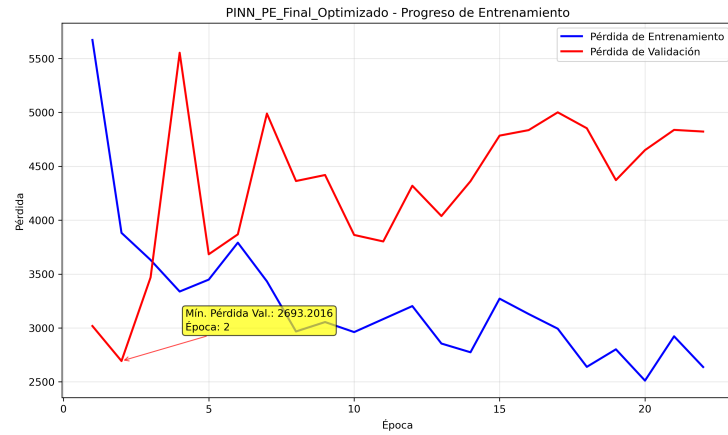


Figura B.66: Evolución de la pérdida de entrenamiento y validación del modelo PINN-PE optimizado.

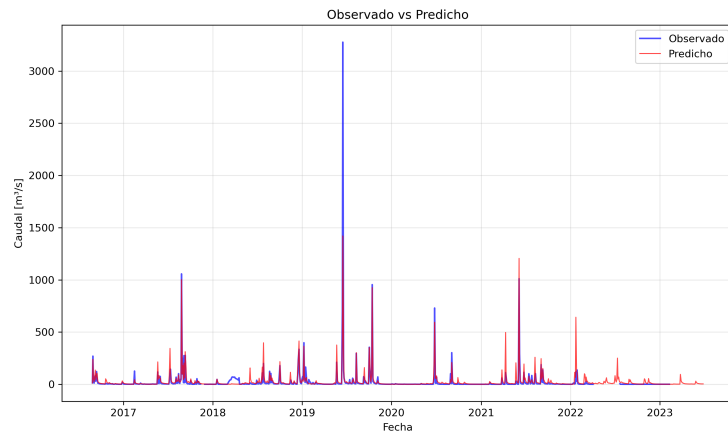


Figura B.67: Serie temporal de caudales observados y predichos por el modelo PINN-PE optimizado.

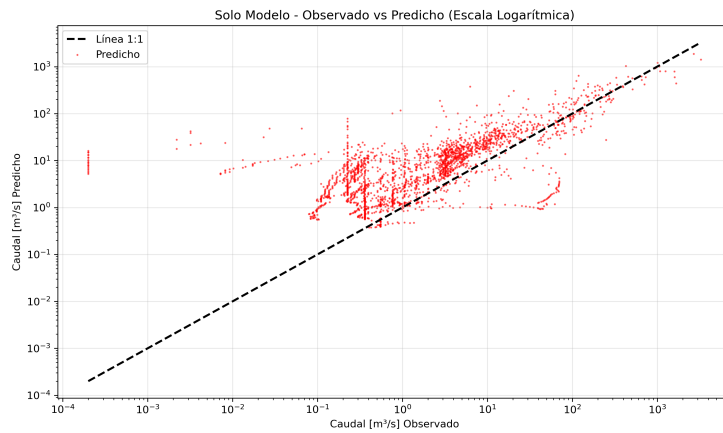


Figura B.68: Diagrama observado vs predicho en escala logarítmica para el modelo PINN-PE optimizado.

El análisis de residuos (Figura B.69) confirma que la mayoría de los errores se concentran en torno a cero, aunque se observan valores extremos asociados a eventos particulares de crecida. No se identifica un patrón temporal sistemático, pero el sesgo negativo global indica una sobreestimación acumulativa del volumen simulado.

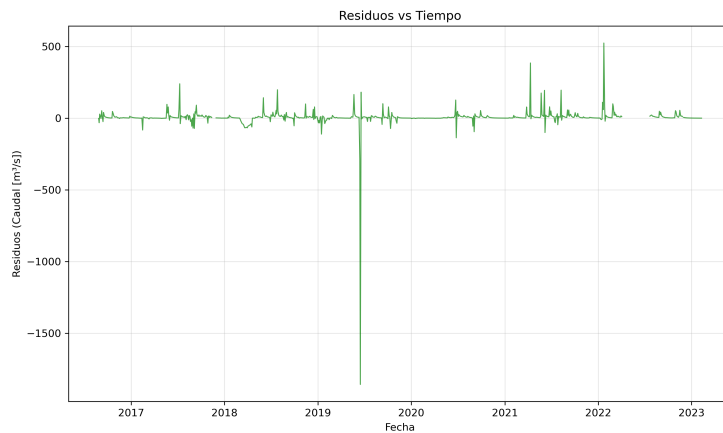


Figura B.69: Residuos del modelo PINN-PE optimizado en función del tiempo.

Análisis de atribución final mediante *Integrated Gradients*

Con el objetivo de interpretar el comportamiento del modelo optimizado, se realizó un análisis de relevancia de variables utilizando *Integrated Gradients* (IG) sobre el conjunto de prueba.

La Figura B.70 muestra las 12 características con mayor atribución promedio (incluyendo el signo). Se observa que las precipitaciones por estación concentran las mayores contribuciones positivas, destacándose Sarandí Grande, La Cruz, San Gabriel y Cerro Colorado.

En contraste, las variables térmicas y radiativas presentan atribuciones negativas de menor magnitud, mientras que la velocidad del viento muestra una contribución prácticamente nula.

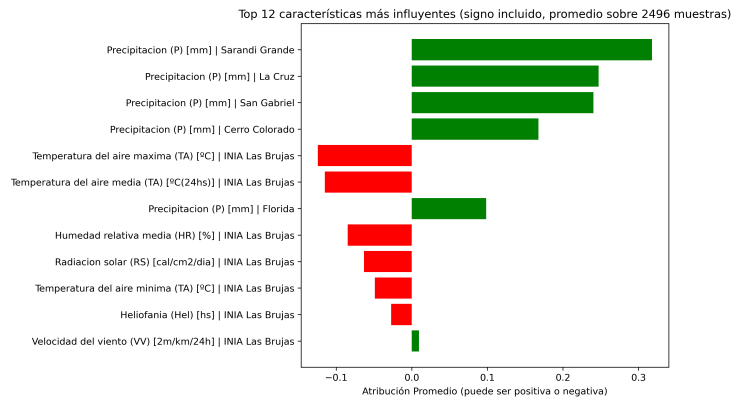


Figura B.70: Top 12 características más influyentes según Integrated Gradients en el conjunto de prueba (modelo PINN-PE optimizado).

Este patrón sugiere que, tras la optimización arquitectónica, el modelo PINN-PE tiende a apoyarse fuertemente en la señal pluviométrica directa, mientras que las variables asociadas al balance energético actúan como moduladores secundarios. Esta priorización de la precipitación puede contribuir a la adecuada reproducción de la dinámica relativa (reflejada en el NSE), pero también podría estar asociada al deterioro del balance volumétrico observado en el *PBIAS*.

Costo computacional

El entrenamiento completo del modelo optimizado demandó 215.09 segundos en 22 épocas en el entorno de cómputo utilizado. Por su parte, el tiempo de inferencia fue de 3.69 segundos, según los registros del sistema.

B.8. PINN-PE RNN

Esta sección presenta el desarrollo detallado del modelo PINN-PE con arquitectura recurrente, incluyendo todos los experimentos realizados durante el proceso de optimización.

El cuerpo principal (Sección 6.11) resume únicamente los resultados del modelo base y del modelo final optimizado. Aquí se documentan las etapas intermedias que condujeron a la configuración seleccionada.

B.8.1. Modelo base

El primer experimento en el marco del enfoque PINN-PE con arquitectura recurrente consistió en entrenar una red LSTM bajo la formulación **PINN-PE**, en la cual la red neuronal predice simultáneamente la precipitación agregada diaria \hat{P}_{agg} y la evapotranspiración \hat{E} . La predicción final de caudal emerge del acoplamiento físico con GR4J, evaluando $Q_{GR4J}(\hat{P}_{agg}, \hat{E}; \theta)$.

La configuración base empleó una secuencia de longitud 1, dos capas LSTM con 64 unidades ocultas, y un módulo MLP posterior de tres capas (128→64→32) con activación ReLU.

Tabla B.59: Desempeño del modelo base PINN-PE RNN sobre el conjunto de validación.

NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
0.69	-0.98	16.34	59.85

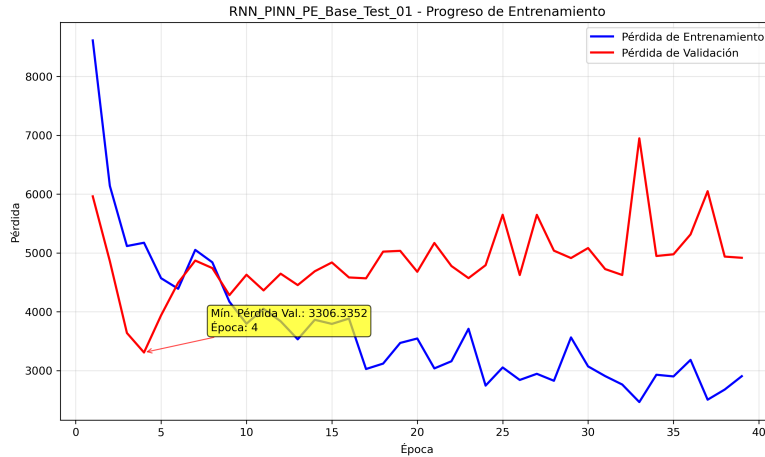


Figura B.71: Evolución de la función de pérdida (entrenamiento y validación) para el modelo base PINN-PE RNN.

La Figura B.71 muestra que la pérdida de validación alcanza su mínimo de 3306.34 en la **época 4**, a partir de la cual se incrementa de forma sostenida hasta valores en torno a 5000–5600. La pérdida de entrenamiento, en contraste, continúa disminuyendo gradualmente desde ~8600 hasta ~2500. Este patrón indica una aparición temprana de *overfitting*, con una separación progresiva entre ambas curvas que confirma la necesidad de optimizar la arquitectura en las etapas posteriores.

En términos cuantitativos (Tabla B.59), el modelo base RNN presenta un $NSE = 0.69$ con un sesgo prácticamente nulo ($PBIAS = -0.98\%$). Este

desempeño es inferior al observado en las configuraciones base de PINN-P RNN ($NSE = 0.71$) y PINN-E RNN ($NSE = 0.73$), lo que es consistente con la mayor complejidad de la tarea: la red debe estimar simultáneamente dos forzantes del modelo hidrológico.

No obstante, al igual que en los modelos anteriores, se procedió al desarrollo de las etapas posteriores de ajuste.

B.8.2. Evaluación de la incorporación de variables

En esta etapa se analiza el impacto de distintas configuraciones de variables de entrada sobre el desempeño del modelo PINN-PE basado en arquitectura RNN.

Variables temporales

Se evaluó el efecto de incorporar variables temporales representadas de forma cíclica. Se compararon cuatro configuraciones: (i) sin variables temporales, (ii) incorporación del día, (iii) incorporación del mes, y (iv) incorporación conjunta de día y mes. (Tabla B.60)

Tabla B.60: Comparación de desempeño para distintas combinaciones de variables temporales sobre el conjunto de validación (PINN-PE RNN).

Configuración	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Ninguna	0.69	-0.98	16.34	59.85
día	0.62	-0.40	18.51	66.14
mes	0.68	5.83	15.93	60.63
día, mes	0.58	-51.47	22.87	68.95

Los resultados muestran que la configuración sin variables temporales alcanza el mejor desempeño global ($NSE = 0.69$, $RMSE = 59.85$ m³/s) con un sesgo prácticamente nulo ($PBIAS = -0.98$ %). La incorporación del día degrada el desempeño ($NSE = 0.62$), mientras que la inclusión conjunta de día y mes produce un deterioro severo ($NSE = 0.58$, $PBIAS = -51.47$ %).

Este patrón es consistente con el observado en la formulación PINN-P con RNN, donde las variables temporales también resultaban perjudiciales. La arquitectura recurrente captura implícitamente los patrones estacionales a través del procesamiento secuencial, haciendo redundante la información temporal explícita.

En consecuencia, se decidió **no incorporar variables temporales** en las etapas posteriores del modelado.

Selección de la ventana temporal (*sequence length*)

Se evaluaron configuraciones con ventanas de 1 a 14 días. (Tabla B.61)

Tabla B.61: Comparación de desempeño según longitud de la ventana temporal (PINN-PE RNN).

Ventana (días)	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
1	0.69	-0.98	16.34	59.85
2	0.60	-33.83	20.99	67.86
3	0.67	-12.80	18.03	61.63
4	0.71	-29.23	18.60	57.61
5	0.68	-1.14	16.74	60.07
6	0.62	1.15	18.07	65.72
7	0.68	-28.06	18.91	60.73
8	0.66	-22.65	19.13	61.90
9	0.70	10.31	15.94	58.99
10	0.66	-14.88	19.27	62.73
11	0.68	-4.81	16.76	60.60
12	0.65	-8.81	17.49	62.88
13	0.65	3.24	17.22	62.82
14	0.68	-5.39	16.63	60.72

La ventana de 4 días alcanza el mayor NSE (0.71) pero con un sesgo considerable ($PBIAS = -29.23\%$). Al igual que en la formulación PINN-P con RNN, la selección de una ventana unitaria no se justifica metodológicamente para una arquitectura recurrente, ya que anula la capacidad de captura de dependencias temporales.

La configuración de **5 días** presenta el mejor equilibrio entre eficiencia ($NSE = 0.68$, $RMSE = 60.07 \text{ m}^3/\text{s}$) y control del sesgo ($PBIAS = -1.14\%$), manteniéndose dentro del rango aceptable. Esta longitud de secuencia es coherente con los tiempos de respuesta de la cuenca y permite a la red aprovechar efectivamente su capacidad recurrente.

En consecuencia, se decidió **adoptar una ventana temporal de 5 días**.

B.8.3. Depuración inicial por correlación

Se evaluó el impacto de eliminar variables correlacionadas, manteniendo fija la ventana temporal de 5 días y los pesos $w_{\text{data}} = 1$, $w_{\text{phys}} = 1$.

Observando la tabla B.62, el modelo base domina en todas las métricas relevantes, con el mayor NSE (0.68) y el menor sesgo en valor absoluto ($PBIAS = -1.14\%$). Todas las exclusiones producen deterioros del desempeño de 2 a 6 puntos porcentuales de NSE y aumentos significativos del sesgo.

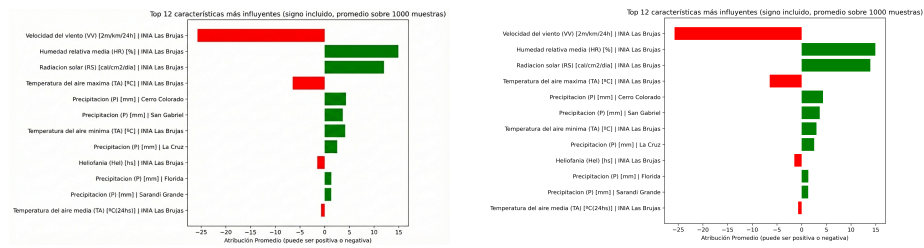
En consecuencia, se decidió **mantener el modelo base**.

Tabla B.62: Resultados de desempeño para configuraciones de reducción de variables correlacionadas (PINN-PE RNN).

Configuración	NSE	PBIAS [%]	MAE [m^3/s]	RMSE [m^3/s]
Modelo base	0.68	-1.14	16.74	60.07
Sin temp. máx. y mín.	0.66	-16.10	18.40	62.31
Solo RS (sin ET)	0.67	-27.60	18.77	61.19
Solo ET (sin RS)	0.62	-7.95	18.31	65.59
Temp. media + RS + ET	0.66	-24.23	19.24	62.35

B.8.4. Análisis de relevancia mediante *Integrated Gradients*

Se evaluó la relevancia individual de cada variable mediante *Integrated Gradients* (IG) sobre los subconjuntos de entrenamiento y validación.



(a) Subconjunto de entrenamiento.

(b) Subconjunto de validación.

Figura B.72: Importancia relativa de variables según *Integrated Gradients* (IG) en los conjuntos de entrenamiento y validación (PINN-PE RNN).

La Figura B.72 muestra que la **velocidad del viento (VV)** concentra la mayor atribución negativa (~ -25), mientras que la **humedad relativa (HR)** y la **radiación solar (RS)** presentan las mayores atribuciones positivas ($\sim +15$). La temperatura máxima contribuye con una atribución negativa moderada (~ -5). Las precipitaciones por estación (Cerro Colorado, San Gabriel, La Cruz, Florida, Sarandí Grande) y la temperatura mínima presentan atribuciones positivas de menor magnitud ($\sim +2-4$), mientras que la heliofanía y la temperatura media exhiben contribuciones marginales. Este patrón combina elementos de las formulaciones PINN-P y PINN-E, reflejando la naturaleza dual de la tarea.

El análisis confirmó que todas las variables contribuyen de forma relevante al modelo, sin identificar candidatas claras para eliminación. Dado que el análisis de correlación ya había demostrado que la exclusión de cualquier variable degrada el desempeño, se decidió **mantener el conjunto completo de 12 variables sin modificaciones**.

Conjunto final de variables

El conjunto final de variables para el modelo PINN-PE RNN quedó conformado por:

- **Precipitaciones contemporáneas por estación**
- **Radiación solar (RS), heliofanía (Hel)**
- **Temperatura del aire** (media, máxima y mínima).
- **Humedad relativa (HR) y velocidad del viento (VV).**
- **Sin variables temporales.**

Esta configuración se procesa mediante una ventana temporal de 5 días.

B.8.5. Análisis del balance entre términos de la función de pérdida

Se evaluaron 13 configuraciones de pesos. (Tabla B.63)

Tabla B.63: Desempeño del modelo PINN-PE RNN para distintas combinaciones de pesos en la función de pérdida.

$w_{\text{data}} - w_{\text{phys}}$	<i>NSE</i>	<i>PBIAS</i> [%]	<i>MAE</i> [m ³ /s]	<i>RMSE</i> [m ³ /s]
0.00 – 1.00	0.71	7.05	15.50	57.52
1.00 – 0.00	0.56	-2.54	18.51	70.56
0.25 – 1.00	0.65	4.42	16.70	63.18
0.25 – 0.75	0.67	5.06	16.76	61.68
0.50 – 1.00	0.69	5.15	16.38	59.82
0.50 – 0.75	0.69	-16.72	17.76	59.49
0.75 – 1.00	0.67	-0.43	16.93	61.09
1.00 – 1.00	0.68	-1.14	16.74	60.07
1.00 – 0.75	0.72	-18.12	17.54	56.57
1.00 – 0.50	0.71	-6.89	16.15	58.03
0.75 – 0.50	0.70	-9.24	17.02	58.20
0.75 – 0.25	0.68	-6.91	16.78	60.25
1.00 – 0.25	0.70	-18.59	17.31	58.94

A diferencia de las formulaciones PINN-P y PINN-E, donde el desempeño era relativamente estable o dominado por una configuración clara, en la PINN-PE se observa una mayor variabilidad entre las combinaciones evaluadas (*NSE* entre 0.56 y 0.72).

La configuración $w_{\text{data}} = 1.00$, $w_{\text{phys}} = 0.75$ alcanza el mayor *NSE* (0.72) y el menor *RMSE* (56.57 m³/s), pero con un sesgo elevado (*PBIAS* = -18.12 %). El caso extremo sin término de datos ($w_{\text{data}} = 0$, $w_{\text{phys}} = 1$) presenta un buen

NSE (0.71) con sesgo positivo moderado ($PBIAS = 7.05\%$), lo que sugiere que la restricción física mediante GR4J es particularmente informativa en esta formulación.

La combinación $w_{\text{data}} = 1.00$, $w_{\text{phys}} = 0.50$ ofrece un buen equilibrio: $NSE = 0.71$, $RMSE = 58.03 \text{ m}^3/\text{s}$ y un sesgo controlado ($PBIAS = -6.89\%$) dentro del rango aceptable.

En consecuencia, se decidió **adoptar la configuración** $w_{\text{data}} = 1.00$ y $w_{\text{phys}} = 0.50$.

B.8.6. Optimización de hiperparámetros (Optuna)

Una vez definido el conjunto final de variables, la ventana temporal de 5 días y el balance $w_{\text{data}} = 1.00$, $w_{\text{phys}} = 0.50$, se procedió a la optimización mediante *Optuna*.

- **Mejor valor de pérdida de validación:** 1556.52
- ***Trials* completados:** 125
- ***Trials* podados:** 27

La Figura B.73 muestra la evolución del valor objetivo a lo largo de los *trials*.

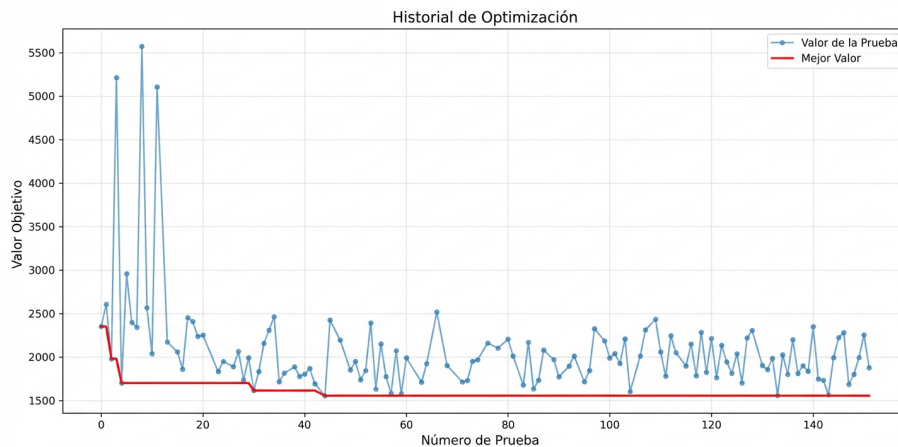


Figura B.73: Historial del proceso de optimización realizado con Optuna para PINN-PE RNN.

La configuración óptima se resume en la Tabla B.64.

A diferencia de la PINN-E —donde Optuna seleccionó una celda GRU—, en la PINN-PE se seleccionó una **celda LSTM** con un estado oculto considerablemente mayor (**171 unidades**), lo que sugiere que la tarea de estimar simultáneamente precipitación y evapotranspiración requiere una mayor capacidad de representación temporal. El módulo MLP posterior presenta 3 capas

Tabla B.64: Configuración óptima seleccionada por Optuna para PINN-PE RNN.

Hiperparámetro	Valor óptimo
Tipo de celda recurrente	LSTM
Tamaño del estado oculto	171
Capas recurrentes	1
Capas MLP	3
Neuronas por capa MLP	[27, 31, 44]
Función de activación	gelu
Tasa de aprendizaje	0.00112
<i>Dropout</i> base MLP	0.15
<i>Batch size</i>	32
<i>Weight decay</i>	4.51×10^{-4}

con un ancho reducido (27→31→44), configuración opuesta a la observada en PINN-P (1 capa de 390 unidades), indicando que la complejidad se concentra en el codificador recurrente.

En la Figura B.74 se aprecia que incluso los cinco mejores *trials* presentan curvas de aprendizaje marcadamente ruidosas, con oscilaciones abruptas y picos pronunciados a lo largo de buena parte del entrenamiento. Este comportamiento sugiere una dinámica de optimización inestable en términos locales, posiblemente asociada a la interacción entre el término físico y el término de datos en la función de pérdida. No obstante, a pesar de esta alta variabilidad inter-época, todos los ensayos convergen finalmente hacia un rango de pérdidas similar (en torno a 1550–1600), lo que indica que el proceso de optimización logra estabilizarse globalmente y alcanzar soluciones de desempeño comparable.

Por su parte, la Figura B.75 evidencia que el mecanismo de *pruning* descarta de forma temprana aquellos ensayos con pérdidas intermedias significativamente superiores, concentrando el presupuesto computacional en configuraciones prometedoras. La mayoría de los *trials* podados se interrumpen en etapas tempranas del entrenamiento, lo que confirma la eficacia del criterio de poda para acelerar la búsqueda sin comprometer la calidad de la solución final.

B.8.7. Evaluación extendida del modelo optimizado

El modelo final se entrenó utilizando la arquitectura óptima de Optuna, con $w_{\text{data}} = 1.00$, $w_{\text{phys}} = 0.50$, y una ventana temporal de 5 días.

La evaluación se realizó de forma independiente sobre los conjuntos de entrenamiento, validación y prueba, preservando la partición temporal original y aplicando el mismo criterio de enmascaramiento para los registros con caudal faltante.

Las métricas consideradas fueron MAE, RMSE, NSE y PBIAS. (Tabla B.65)
A modo de referencia, el modelo GR4J ejecutado de forma directa sobre

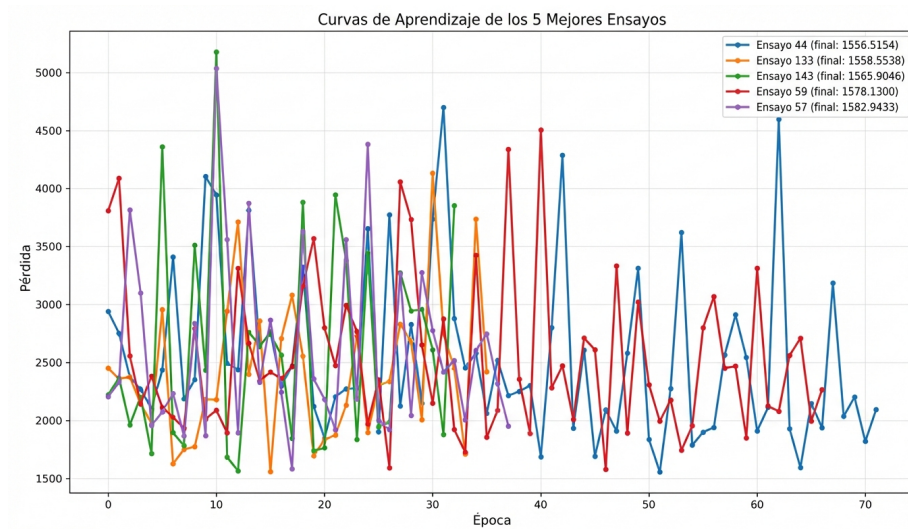


Figura B.74: Curvas de aprendizaje de los cinco mejores *trials* del proceso de optimización de Optuna (PINN-PE RNN).

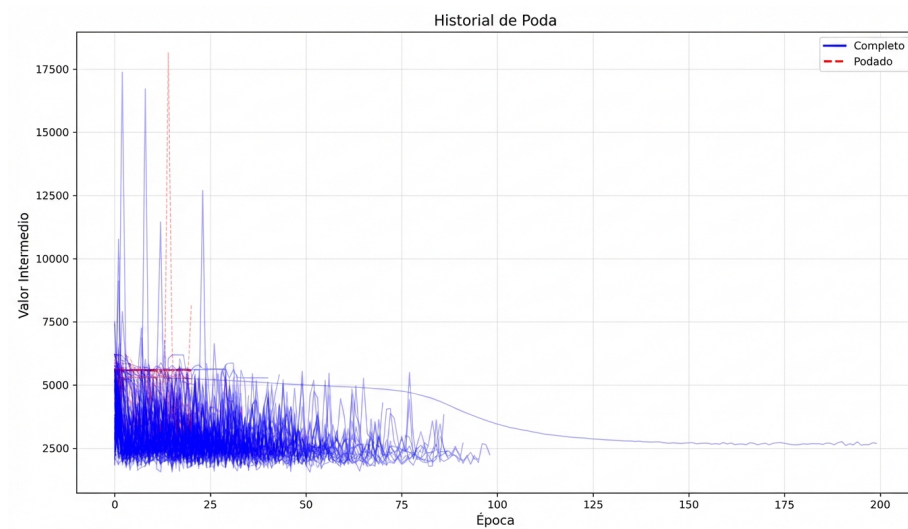


Figura B.75: Historial de *pruning* durante la optimización con Optuna (PINN-PE RNN).

Tabla B.65: Desempeño del modelo PINN-PE optimizado (RNN) en los distintos conjuntos.

Conjunto	NSE	PBIAS [%]	MAE [m ³ /s]	RMSE [m ³ /s]
Entrenamiento	0.87	-33.76	13.07	37.27
Validación	0.69	1.08	15.90	59.28
Prueba	0.71	-4.19	15.29	69.51

el mismo período de prueba obtiene $NSE = 0.59$, $RMSE = 82.77$ m³/s y $PBIAS = -5.05$ %.

Análisis del desempeño

En el conjunto de entrenamiento, el modelo alcanza un $NSE = 0.87$ con $RMSE = 37.27$ m³/s, lo que refleja un ajuste elevado a los datos de entrenamiento. Sin embargo, el sesgo en entrenamiento resulta considerable ($PBIAS = -33.76$ %), indicando una tendencia marcada a la sobreestimación del volumen de caudal durante la fase de optimización. Este patrón, también observado en las demás formulaciones PINN, se explica por la naturaleza de la función de pérdida, donde el término físico introduce restricciones que pueden penalizar el ajuste directo sobre caudal.

En validación, el desempeño disminuye a $NSE = 0.69$ con $RMSE = 59.28$ m³/s, pero el sesgo se reduce drásticamente hasta un valor prácticamente nulo ($PBIAS = 1.08$ %), lo que indica un equilibrio adecuado entre ajuste y generalización.

En el conjunto de prueba, el modelo PINN-PE RNN optimizado alcanza un $NSE = 0.71$ con un sesgo controlado ($PBIAS = -4.19$ %). La mejora del NSE respecto a validación (0.71 vs 0.69) sugiere que el período de prueba contiene patrones que el modelo logra capturar de forma efectiva. Este resultado representa una **mejora sustancial respecto al modelo GR4J directo** (+12 puntos porcentuales de NSE), lo que indica que la estimación conjunta de precipitación y evapotranspiración mediante la red LSTM aporta un valor agregado significativo al sistema acoplado.

En comparación con las formulaciones individuales en el conjunto de prueba, la PINN-PE RNN ($NSE = 0.71$) supera a la PINN-P RNN ($NSE = 0.67$) pero queda por debajo de la PINN-E RNN ($NSE = 0.77$). No obstante, su sesgo ($PBIAS = -4.19$ %) es el más controlado entre todas las formulaciones PINN con RNN.

La Figura B.76 muestra que la pérdida de entrenamiento desciende de forma sostenida desde ~ 4500 hasta valores en torno a 600–700, mientras que la pérdida de validación presenta oscilaciones pronunciadas en un rango amplio (~ 1700 – 4400), alcanzando su mínimo de 1668.45 en la **época 43**. La marcada separación entre ambas curvas y la elevada variabilidad de la pérdida de validación indican un proceso de entrenamiento relativamente inestable, lo que sugiere

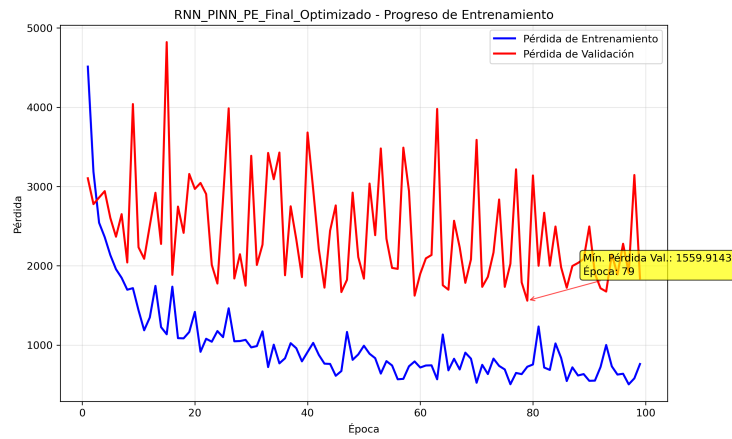


Figura B.76: Evolución de la pérdida de entrenamiento y validación del modelo PINN-PE optimizado RNN.

que la tarea dual de estimar precipitación y evapotranspiración simultáneamente introduce una mayor dificultad de optimización. No obstante, el mecanismo de *early stopping* con paciencia de 20 épocas permitió seleccionar adecuadamente el punto de mejor generalización.

Análisis gráfico del ajuste

La Figura B.77 presenta la comparación temporal entre caudal observado y predicho.

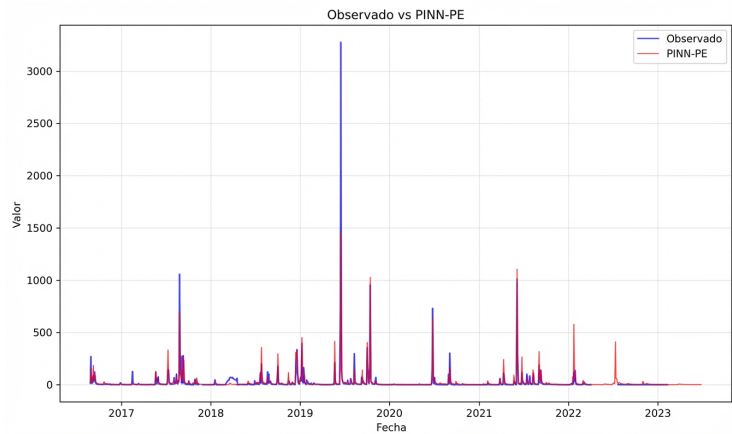


Figura B.77: Serie temporal de caudales observados y predichos por el modelo PINN-PE optimizado RNN en el conjunto de prueba.

El modelo reproduce adecuadamente la dinámica general de caudales, capturando los períodos de estiaje y la mayoría de los eventos de crecida intermedia. El evento extremo de mediados de 2019 ($\sim 3300 \text{ m}^3/\text{s}$) es parcialmente reproducido con un pico de $\sim 1400 \text{ m}^3/\text{s}$, evidenciando subestimación en eventos de gran magnitud. Otros eventos significativos (2017, 2019 previos, 2021) son razonablemente capturados, incluso siendo sobreestimados en algunos casos.

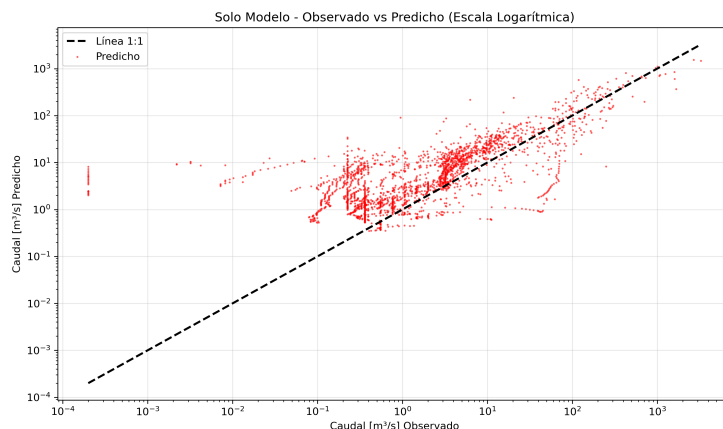


Figura B.78: Diagrama observado vs predicho en escala logarítmica para el modelo PINN-PE optimizado RNN.

En la Figura B.78 se observa una alineación razonable con la línea 1:1 en el rango medio-alto de caudales ($\sim 10\text{--}500 \text{ m}^3/\text{s}$). Para caudales bajos ($< 1 \text{ m}^3/\text{s}$), el modelo tiende a sobrepredecir (puntos por encima de la diagonal). En el rango de caudales extremos ($> 1000 \text{ m}^3/\text{s}$), los puntos se mantienen cercanos a la diagonal, indicando un ajuste aceptable en eventos de gran magnitud.

Los residuos visualizados en la Figura B.79 se mantienen centrados en torno a cero durante la mayor parte del período, sin patrones sistemáticos en el tiempo. El residuo negativo más pronunciado ($\sim -1800 \text{ m}^3/\text{s}$) corresponde al evento extremo de mediados de 2019. Se observan residuos positivos moderados ($\sim 100\text{--}500 \text{ m}^3/\text{s}$) distribuidos a lo largo del período, particularmente en 2017 y 2022. La amplitud general de los residuos es comparable a la observada en la PINN-E RNN.

Análisis de atribución final mediante *Integrated Gradients*

Con el objetivo de interpretar el comportamiento del modelo optimizado, se realizó un análisis de relevancia de variables utilizando el método *Integrated Gradients* (IG) sobre el conjunto de prueba.

La Figura B.80 muestra que la **humedad relativa media (HR)** concentra la mayor atribución positiva, mientras que la **velocidad del viento (VV)** presenta la mayor atribución negativa, de magnitud considerable. Las variables

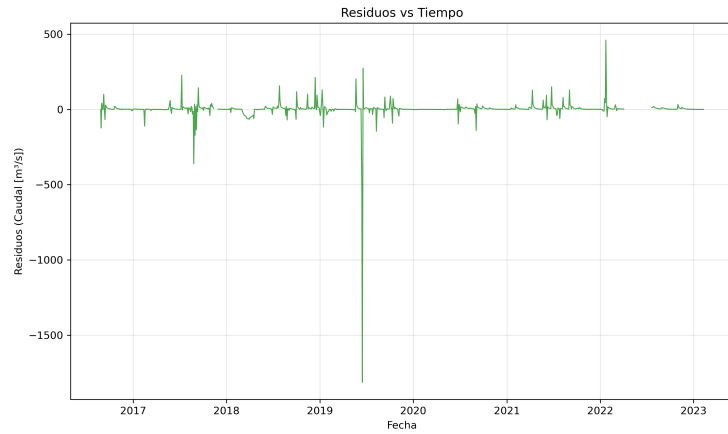


Figura B.79: Residuos del modelo PINN-PE optimizado (RNN) en función del tiempo.

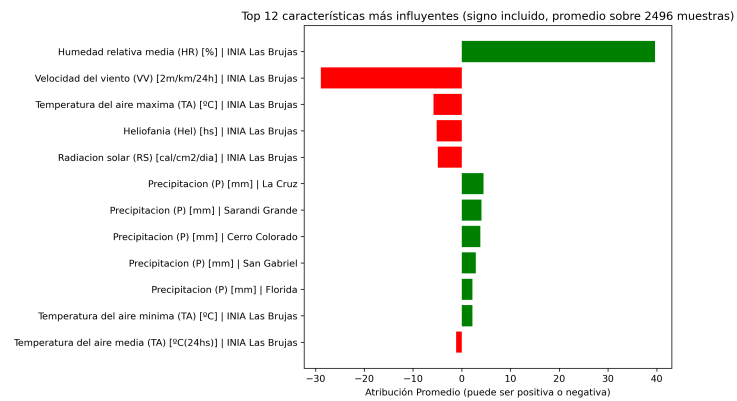


Figura B.80: Top características más influyentes según Integrated Gradients en el conjunto de prueba (modelo PINN-PE optimizado RNN).

de precipitación de múltiples estaciones (La Cruz, Sarandí Grande, Cerro Colorado, San Gabriel, Florida) aparecen con atribuciones positivas moderadas, lo que refleja la dependencia directa del caudal respecto al forzante pluviométrico. Las variables radiativas (heliofanía, radiación solar) y la temperatura máxima presentan atribuciones negativas de magnitud intermedia.

Este patrón difiere parcialmente del observado durante las etapas exploratorias (Sección B.8.4), donde la radiación solar presentaba una atribución positiva significativa y la humedad relativa una contribución moderada. En el modelo final optimizado, la humedad relativa asume el rol dominante positivo, mientras que la radiación solar pasa a exhibir una atribución negativa. Este cambio puede atribuirse a la optimización arquitectónica mediante Optuna, que redistribuye la forma en que el modelo pondera las distintas variables de entrada.

Asimismo, el patrón difiere del observado en las formulaciones de predicción individual (PINN-E, PINN-P), donde la radiación solar o la precipitación dominaban respectivamente, sugiriendo que la tarea dual redistribuye la relevancia relativa de los predictores.

No obstante, debe considerarse que, al igual que en las demás arquitecturas recurrentes, el cálculo de *Integrated Gradients* se realizó sin considerar las $W - 1$ dimensiones adicionales de la ventana temporal (5 días). El contexto secuencial completo que integra el modelo no queda reflejado en las atribuciones calculadas, por lo que estos resultados deben interpretarse como una aproximación orientativa de la relevancia relativa de las variables.

Costo computacional

El entrenamiento completo del modelo optimizado demandó 603.05 segundos en un total de 63 épocas en el entorno de cómputo utilizado. Por su parte, el tiempo de inferencia fue de 4.09 segundos, según los registros del sistema.