



UNIVERSIDAD DE MÁLAGA



Graduado en Ingeniería de Computadores

**Aplicación de modelos predictivos basados en inteligencia artificial para la gestión sostenible del agua en embalses de Andalucía**

*Transfer Learning para la adaptación eficiente de modelos predictivos entre embalses con características similares*

**Application of artificial intelligence-based predictive models for sustainable water management of Andalusian reservoirs**

*Transfer Learning for efficient adaptation of predictive models between reservoirs with similar characteristics*

Realizado por  
Rodrigo Hernández Barba

Tutorizado por  
Jamal Toutouh El Alamin  
Christian Cintrano López

Departamento  
Lenguajes y Ciencias de la Computación

MÁLAGA, SEPTIEMBRE 2025

Fecha defensa: septiembre de 2025



UNIVERSIDAD  
DE MÁLAGA



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA INFORMÁTICA  
GRADUADO EN INGENIERÍA DE COMPUTADORES

# **Aplicación de modelos predictivos basados en inteligencia artificial para la gestión sostenible del agua en embalses de Andalucía**

*Transfer Learning para la adaptación eficiente de modelos predictivos entre embalses con características similares*

## **Application of artificial intelligence-based predictive models for sustainable water management of Andalusian reservoirs**

*Transfer Learning for efficient adaptation of predictive models between reservoirs with similar characteristics*

Realizado por  
**Rodrigo Hernández Barba**

Tutorizado por  
**Jamal Toutouh El Alamin**  
**Christian Cintrano López**

Departamento  
**Lenguajes y Ciencias de la Computación**

UNIVERSIDAD DE MÁLAGA  
MÁLAGA, SEPTIEMBRE 2025

Fecha defensa: septiembre de 2025

# Abstract

This project presents a practical solution to support reservoir management in Andalusia through transfer learning. A reproducible ETL pipeline merges historical hydrological and meteorological series, harmonizes names and timestamps, and generates derived features and aggregates.

On this foundation, GRU models with long input windows and a 14-day operational horizon are trained and benchmarked against persistence and climatology. To scale modeling across sites, reservoirs are segmented into hydro-climatic clusters; per-cluster master models are adapted to related targets, reducing computational cost and convergence time with little loss of accuracy. The resulting system combines a consistent time-series repository (InfluxDB), models ready for inference, and a web interface for querying series and indicators.

Future work focuses on incorporating operational variables and quantifying predictive uncertainty.

**Keywords:** water management; reservoirs; Andalusia; time series; GRU; transfer learning; clustering; deep learning.

# Resumen

Este Trabajo Fin de Grado presenta una solución práctica para apoyar la gestión de embalses en Andalucía mediante aprendizaje por transferencia (*transfer learning*). Se construye un *pipeline ETL* reproducible que integra series históricas hidrológicas y meteorológicas, normaliza nombres y tiempos y genera variables derivadas.

Con esta base se entrenan modelos GRU con ventanas largas y horizonte operativo de 14 días, comparados frente a líneas base (persistencia y climatología). Para escalar el modelado entre ubicaciones, los embalses se segmentan en clústeres hidroclimáticos y se seleccionan modelos máster por clúster para su adaptación a destinos afines, reduciendo coste computacional y tiempo de convergencia sin pérdida apreciable de precisión. El sistema combina un repositorio temporal consistente (InfluxDB), modelos listos para inferencia y una interfaz web para consulta de series e indicadores.

Las líneas futuras incluyen variables de operación y cuantificación de la incertidumbre.

**Palabras clave:** gestión del agua; embalses; Andalucía; series temporales; GRU; *transfer learning*; *clustering*; aprendizaje profundo.



# Índice

<b>1. Introducción</b>	<b>9</b>
1.1. Motivación . . . . .	9
1.2. Objetivos . . . . .	10
1.3. Estructura del documento . . . . .	12
<b>2. Estado del arte</b>	<b>15</b>
2.1. Modelos computacionales aplicados al análisis y gestión del agua . . . . .	15
2.2. Modelos analizados . . . . .	17
2.3. Contexto del estudio . . . . .	21
2.3.1. Fenómeno de la sequía . . . . .	22
2.3.2. Política de desembalse y factores humanos . . . . .	23
2.3.3. Organización de los embalses andaluces . . . . .	25
2.3.4. Red de estaciones agroclimáticas . . . . .	27
2.4. Métricas . . . . .	28
2.4.1. Evaluación de modelos predictivos . . . . .	29
2.4.2. Validación del agrupamiento de embalses . . . . .	30
2.4.3. Selección de características y análisis . . . . .	32
2.4.4. Criterios de selección de máster de clúster . . . . .	32
2.4.5. Geolocalización y distancia . . . . .	36
2.5. Tecnologías empleadas . . . . .	37
2.5.1. Python . . . . .	37
2.5.2. InfluxDB (Influx DataBase) . . . . .	38
2.5.3. JavaScript . . . . .	39
2.5.4. VisualStudio Code . . . . .	39
2.5.5. Google Colaboratory . . . . .	40
2.5.6. Git + Github . . . . .	40
2.5.7. Docker . . . . .	40
2.5.8. Node.js . . . . .	41

2.5.9.	Bootstrap	41
2.5.10.	Express	42
2.5.11.	HTML (HyperText Markup Language)	43
2.5.12.	CSS (Cascade Style Sheets)	43
2.5.13.	Algunas librerías destacables	44
<b>3.</b>	<b>Analisis experimental</b>	<b>47</b>
3.1.	Procesado de datos para el modelado	47
3.1.1.	Selección de características	49
3.2.	Detalles	51
3.2.1.	<i>Clustering</i> y segmentación de embalses	53
3.2.2.	Entrenamiento de modelos predictivos	55
3.3.	Resultados	59
3.3.1.	Selección de características	59
3.3.2.	Resultados clustering	65
3.3.3.	Resultados entrenamiento de modelos predictivos	76
3.4.	Discusión general	128
3.4.1.	Selección de características	128
3.4.2.	<i>Clustering</i>	128
3.4.3.	Redes neuronales	130
3.4.4.	Entrenamiento de modelos predictivos	131
<b>4.</b>	<b>Desarrollo de la aplicación</b>	<b>135</b>
4.1.	Ingeniería del software	135
4.1.1.	Plataforma web	135
4.1.2.	Estructura del proyecto	136
4.1.3.	Despliegue de servicios	137
4.1.4.	Funcionamiento de la aplicación web	138
4.1.5.	Problemas encontrados y soluciones	138
4.2.	Procesos <i>ETL</i>	139
4.2.1.	Fase de extracción	139
4.2.2.	Fase de transformación	141

4.2.3. Fase de carga . . . . .	142
4.2.4. Incidencias y soluciones . . . . .	142
4.3. Metodología empleada . . . . .	143
<b>5. Conclusiones y trabajos futuros</b>	<b>145</b>
5.1. Conclusiones . . . . .	145
5.2. Trabajos futuros . . . . .	147
<b>Bibliografía</b>	<b>151</b>
<b>Apéndice A. Manual de Usuario</b>	<b>157</b>
<b>Apéndice B. Contexto del estudio</b>	<b>169</b>
B.1. Red de embalses . . . . .	169
B.2. Red de estaciones agroclimáticas de Andalucía . . . . .	177



# 1

# Introducción

## 1.1. Motivación

La gestión de los recursos hídricos en regiones mediterráneas se ha convertido en un desafío prioritario. En primer lugar, el cambio climático agrava la sequía, haciéndola más frecuente y prolongada, y vuelve irregulares los episodios de lluvia. En segundo lugar, el crecimiento demográfico y la mayor demanda de los sectores agrícola, ganadero, turístico e industrial presionan aún más el recurso. Estos factores ejercen una fuerte presión sobre los embalses, la principal fuente de suministro de agua dulce de Andalucía. Asegurar el abastecimiento de agua y optimizar su uso adquiere una importante relevancia, tanto para la preservación del medio ambiente como el crecimiento socioeconómico.

La Inteligencia Artificial es un campo de la informática que se orienta al desarrollo de sistemas en los que normalmente la inteligencia humana es un factor indispensable, como el aprendizaje y el razonamiento, entre otros. En la actualidad, el uso de la Inteligencia Artificial engloba multitud de campos muy diferentes, siendo especialmente útil en aquellas tareas en las que se deben procesar una gran cantidad de datos. Esto incluye áreas como la predicción de fenómenos complejos, visión por computador o el procesamiento del lenguaje natural, donde la Inteligencia Artificial proporciona soluciones precisas y robustas para la toma de decisiones.

La necesidad de este Trabajo Fin de Grado nace de la urgencia por encontrar una solución efectiva que garantice el suministro de agua a largo plazo mediante una correcta y precisa gestión del agua en los embalses andaluces. Con este fin, el trabajo consta de una parte común y dos líneas de trabajo. La primera, centrada en el desarrollo de modelos predictivos para analizar escenarios “*What-If*”. La segunda, enfocada en la aplicación de aprendizaje por trans-

ferencia (en adelante, *transfer learning*) para la reutilización de modelos entre embalses con características similares.

Mediante la aplicación de técnicas de *machine learning* (aprendizaje automático) y *deep learning* (aprendizaje profundo), se pretende desarrollar un modelo predictivo que incorpore diversas variables climáticas e hidrológicas, a partir de datos proporcionados por la Red de Información Agroclimática de Andalucía (RIA) [1] y la Red de Información Ambiental de Andalucía (REDIAM) [2], respectivamente, albergadas en una base de datos de series temporales. Asimismo, se propone incorporar estrategias de apoyo a la toma de decisiones para los responsables de la gestión hídrica. Se integrará el *transfer learning* para reutilizar modelos entrenados en un conjunto de datos sobre otros conjuntos relacionados. Con ello se reducen los recursos computacionales y se mejora la precisión de los modelos. De esta forma, se busca aprovechar el potencial de la Inteligencia Artificial para asentar las bases de un uso sostenible del agua de los embalses, garantizando el bienestar social, económico y medioambiental de Andalucía.

## 1.2. Objetivos

La gestión eficiente del agua exige herramientas predictivas fiables y operativas, especialmente en regiones con estrés hídrico como Andalucía. Este trabajo desarrolla un flujo de trabajo completo, desde la captura de datos hasta el despliegue de una aplicación web. Combina aprendizaje profundo y *transfer learning* para prever la evolución del volumen almacenado en los embalses y apoyar la toma de decisiones. El proyecto se alinea con los **Objetivos de Desarrollo Sostenible: ODS 6 (Agua limpia y saneamiento), ODS 11 (Ciudades y comunidades sostenibles), ODS 12 (Producción y consumo responsables) y ODS 13 (Acción por el clima)**.

**Objetivo general:** Diseñar, implementar y evaluar un sistema predictivo reproducible para los embalses de Andalucía, basado en redes neuronales recurrentes (GRU) y estrategias de *transfer learning* intraclúster, apoyado en un **flujo de procesamiento (pipeline) ETL** de datos públicos y una **aplicación web** para la exploración y comunicación de resultados.

**Objetivos específicos:**

1. **Extracción e integración de datos públicos.** Desarrollar un proceso automatizado de *web scraping* para recopilar series diarias hidrológicas (embalses) y meteorológicas (estaciones RIA), almacenarlas en **InfluxDB** y consolidarlas en **CSV normalizados por embalse**, garantizando trazabilidad de origen y versiones.
2. **Preprocesado y enriquecimiento para modelado.** Construir **dataframes enriquecidos** por embalse con transformaciones causales (*lags*, ventanas móviles, codificación estacional) y estandarización **z-score** solo con estadísticas de entrenamiento, evitando fugas de información y asegurando paridad de variables (*features*) entre el embalse origen (máster) y destino (*target*).
3. **Asociación espacial embalse–estaciones y continuidad temporal.** Implementar un algoritmo determinista que vincule a cada embalse hasta **tres estaciones meteorológicas cercanas** (distancia Haversine) y conmute de forma diaria según disponibilidad, preservando la representatividad espacial y la continuidad de la serie 2000–2024.
4. **Clustering hidro-climático y selección de masters.** Agrupar embalses en **clústeres homogéneos**, detectar **embalses atípicos** (en adelante, *outliers*) y, en cada clúster, seleccionar un **embalse máster** representativo y estable como donante de pesos para **transfer learning**; definir además un *target* por clúster para la evaluación de la transferencia.
5. **Modelado secuencial y transferencia.** Diseñar y entrenar **modelos GRU** con ventanas largas (120–180 pasos) y horizonte **H=14** días, comparando:
  - **entrenamiento desde cero** frente a **transfer learning** (versiones básica y extendida de *features*),
  - modos **directo** y **residual**, con protocolos de validación temporal adecuados.
6. **Evaluación rigurosa y análisis comparativo.** Establecer un marco de evaluación con **líneas base fuertes** (persistencia, climatología) y métricas estándar, aplicadas de forma homogénea en *holdout* anual y validación *rolling*, para cuantificar tanto la **precisión predictiva** como la **eficiencia computacional** (tiempo de entrenamiento) y detectar posibles casos de **transferencia negativa** (p. ej., embalses con comportamiento fuertemente operativo).

7. **Aplicación web para consulta y análisis.** Desarrollar una **plataforma web** (*frontend* React + Chart.js, *backend* Node/Express) que permita explorar tablas y gráficas históricas, visualizar un **mapa** de embalses y estaciones, filtrar por provincia/embalse/fechas y mostrar resultados de los modelos de forma accesible a distintos perfiles de usuario.
8. **Despliegue, reproducibilidad y documentación.** Preparar contenedores de los diferentes componentes (Docker), automatizar tareas recurrentes y documentar el **pipeline ETL**, la configuración de experimentos y el **manual de usuario**, asegurando **reproducibilidad** y mantenibilidad del sistema.

En conjunto, estos objetivos articulan un itinerario que va de los datos abiertos a la toma de decisiones, evaluando de forma crítica el valor añadido del **transfer learning** intraclúster frente al entrenamiento independiente y proporcionando una base tecnológica transferible a otros ámbitos o regiones.

### 1.3. Estructura del documento

El documento se organiza para guiar al lector desde el contexto y la motivación hasta los resultados experimentales y su despliegue software, poniendo el foco en el **modelado secuencial** y el **transfer learning intraclúster** aplicado a embalses andaluces. A grandes rasgos, se compone de **seis secciones y dos apéndices**:

- **Sección 1: Introducción** Presenta el problema y la motivación, los **objetivos** (incluida la doble línea de trabajo: modelos predictivos y transferencia entre embalses afines) y esta descripción de la estructura. Sirve de hoja de ruta del proyecto completo, del ETL a la aplicación web.
- **Sección 2: Estado del arte.** Revisa la literatura sobre predicción hidrológica y series temporales (RNN, LSTM/GRU) y contextualiza el caso andaluz (sequía, política de desembalse, organización de embalses y red agroclimática). Además, consolida en un único lugar las **métricas y criterios** usados después (evaluación de modelos, validación del *clustering*, selección de variables, estabilidad hidrológica y distancia geodésica) y describe las **tecnologías empleadas**.

- **Sección 3: Análisis experimental** Es el núcleo metodológico y empírico. Se divide en:
  1. **Detalles:** define el **procesado de datos para modelado** (dataframes diarios, *lags* y ventanas, estacionalidad, estandarización), la **selección de características**, el **clustering** y **segmentación** de embalses (incluida la elección de “másters” por clúster) y el **entrenamiento de modelos predictivos** (ventanas 120–180, horizonte H=14 y comparativa directo vs. residual con y sin transferencia).
  2. **Resultados:** reporta comparativas y figuras para la selección de características (feature selection), **agrupamiento** (K), **másters/outliers** y **transfer learning** (parejas donante-destino), incluyendo casos de éxito y límites.
  3. **Discusión General:** integra lecturas, implicaciones operativas y lecciones metodológicas (p. ej., cuándo preferir *transfer learning* frente a entrenamiento desde cero).
  
- **Sección 4: Aplicación:** Detalla **aspectos del modelado de la aplicación** (arquitectura *frontend-backend-BD*), la **plataforma web** para explorar datos y resultados, y el **pipeline ETL** (extracción, transformación y carga) con incidencias resueltas y la **metodología de trabajo seguida**.
  
- **Sección ??: Conclusiones y trabajos futuros:** Resume los principales hallazgos (valor del **transfer learning intraclúster** y del horizonte operativo H=14) y propone líneas de mejora (dato operativo, arquitecturas y MLOps).
  
- **Sección A: Apéndice A** Guía de navegación de la aplicación (tablas, gráficas y mapa).
  
- **Sección B: Apéndice B** Inventario operativo de embalses y estaciones, útil como referencia cruzada.



# 2

## Estado del arte

La sección sintetiza la evolución desde los modelos físico-conceptuales de lluvia–escorrentía hacia enfoques basados en datos. Revisa el aprendizaje automático clásico y el aprendizaje profundo para series temporales, con énfasis en RNN (LSTM/GRU) por su capacidad para capturar dependencias largas y retardos. Presenta el aprendizaje por transferencia como vía para reutilizar modelos entre embalses con afinidad hidroclimática y establece buenas prácticas: segmentación por clústeres y comparación con líneas base (persistencia y climatología). Por último, sitúa el problema en el contexto andaluz (sequía y políticas de explotación) y fija las métricas y criterios de validación que se emplearán en el resto del trabajo.

### 2.1. Modelos computacionales aplicados al análisis y gestión del agua

La gestión de los recursos hídricos en regiones con estrés, como Andalucía, exige herramientas predictivas que anticipen los niveles embalsados y se adapten a escenarios cambiantes. Tradicionalmente, esta necesidad se ha cubierto con modelos físico-conceptuales (por ejemplo, de lluvia–escorrentía) que representan el ciclo hidrológico mediante ecuaciones y parámetros calibrables. Su gran virtud es la interpretabilidad. Sus límites aparecen cuando la instrumentación es escasa, la calibración resulta costosa o la operación humana introduce dinámicas difíciles de parametrizar. En ese contexto han ganado peso los enfoques complementarios basados en datos. [3][4]

En el **estado del arte**, podemos encontrar varias familias de modelos. Las referencias estadísticas (persistencia y climatología) aportan líneas base sólidas en horizontes cortos. Los métodos de *machine learning* clásico (árboles de decisión, *Random Forest*, SVM) capturan relaciones no lineales con un coste de entrenamiento contenido y, además, ayudan a priorizar variables. Sobre esta base científica, el *deep learning* ha ganado protagonismo en series hidro-

meteorológicas por su capacidad para representar dependencias largas y retardos: las redes neuronales recurrentes (RNN), en particular LSTM y GRU, incorporan “memoria” del pasado mediante estados ocultos que se actualizan secuencialmente (entrenados por *backpropagation through time*), lo que permite modelar respuestas acumuladas características de los embalses [5][6][7].

Las **LSTM** introducen una celda de memoria y compuertas de entrada, olvido y salida para mitigar la desaparición del gradiente, con resultados contrastados en predicción hidrológica multicuenca y capacidad de generalización entre regímenes climáticos cuando se entrena de forma conjunta [8][9]. Las **GRU** simplifican la arquitectura a dos compuertas (actualización y reseteo), reducen el número de parámetros y el coste de cómputo, y ofrecen prestaciones comparables en horizontes operativos, con menor propensión al sobreajuste en conjuntos con muchas exógenas [10]. En paralelo, la adopción de Inteligencia Artificial (IA) se ha extendido a toda la cadena del agua (optimización del riego, detección de fugas, monitorización remota), lo que evidencia su madurez tecnológica más allá del laboratorio [11][12][13].

Una línea especialmente relevante para sistemas con datos heterogéneos por emplazamiento es el **aprendizaje por transferencia** (*transfer learning*): reutilizar representaciones aprendidas en un dominio fuente para acelerar y estabilizar el aprendizaje en un dominio destino relacionado. En ciencias ambientales ha mostrado beneficios en predicción climática estacional, eventos extremos y sistemas energéticos; en hidrología, su aplicación (aún emergente) sugiere mejoras con menos datos locales y menor coste de entrenamiento cuando existe afinidad hidro-climática entre cuencas [14][15][16][17]. Este marco enfatiza dos prácticas clave:

- segmentar previamente los sistemas (agrupamiento) para favorecer transferencias “intra-grupo” más fiables, y
- comparar el rendimiento frente a *baselines* fuertes (persistencia, climatología), imprescindibles para valorar la utilidad operativa cuando la inercia del sistema impone un techo alto [5][6].

En definitiva, el estado del arte converge en un enfoque híbrido y pragmático: aprovechar

la interpretabilidad de los modelos físicos cuando se requiere explicación causal; emplear *machine learning/deep learning* (LSTM/GRU) para capturar no linealidades y retardos difíciles de parametrizar; y apoyarse en *transfer learning* y segmentación para escalar el aprendizaje entre embalses de dinámica comparable, manteniendo la evaluación anclada en comparativas con persistencia y climatología para asegurar relevancia operativa.

## 2.2. Modelos analizados

A continuación se presentan los **modelos más adecuados para el problema tratado**: predicción de series hidrometeorológicas con **memoria, retardos y no linealidad**. Partimos de las **redes neuronales recurrentes** (RNN) por su capacidad para incorporar dependencia temporal, y revisamos sus dos variantes más utilizadas en hidrología, **LSTM** y **GRU**, destacando su funcionamiento, ventajas y limitaciones. También se resume el **esquema de entrada–salida** más adecuado para previsiones multi-día (*Many-to-One* y *Many-to-Many*) y se introducen los **criterios prácticos** que guían la elección entre arquitecturas. Por último, se enmarca su uso dentro de la estrategia de **aprendizaje por transferencia**, clave para reutilizar modelos entre embalses con dinámica comparable y diferentes niveles de instrumentación.

### Redes neuronales recurrentes (RNN)

Las redes neuronales son modelos de aprendizaje automático capaces de aproximar funciones no lineales a partir de datos. En el caso de las series temporales, como los registros meteorológicos e hidrológicos, la familia más utilizada son las redes neuronales recurrentes (RNN), cuya principal característica es la incorporación de memoria. Es decir, el estado actual del modelo depende no solo de la entrada en ese instante, sino también de las entradas anteriores. Esto resulta fundamental en el contexto de este trabajo, ya que los embalses responden a condiciones meteorológicas y de operación acumuladas en el tiempo.

Las RNN tienen aplicaciones muy diversas: desde la predicción de caudales y niveles de embalses, hasta el procesamiento de lenguaje natural (traducción automática, análisis de sentimiento), el reconocimiento de voz o el subtítulo de imágenes. Herramientas ampliamente utilizadas como *Google Translate*, Siri o la búsqueda por voz en motores de búsqueda se basan en este tipo de arquitecturas [5].

## Funcionamiento básico de las RNN

Tal y como se puede observar en figura 1 A diferencia de una red *feedforward*, donde cada entrada se procesa de manera independiente, una RNN mantiene un **estado oculto** que se actualiza en cada instante y que resume la información previa. El entrenamiento se lleva a cabo mediante el algoritmo de **retropropagación a través del tiempo** (*Backpropagation Through Time, BPTT*), que ajusta los parámetros acumulando gradientes sobre secuencias enteras. No obstante, las RNN presentan dos limitaciones importantes:

- **Desaparición del gradiente:** cuando los gradientes tienden a cero en secuencias largas, impidiendo que el modelo aprenda dependencias de largo plazo.
- **Explosión del gradiente:** cuando los gradientes crecen sin control, produciendo inestabilidad en el entrenamiento.

Para paliar estos problemas, se han desarrollado variantes más avanzadas de RNN, como las LSTM y las GRU.

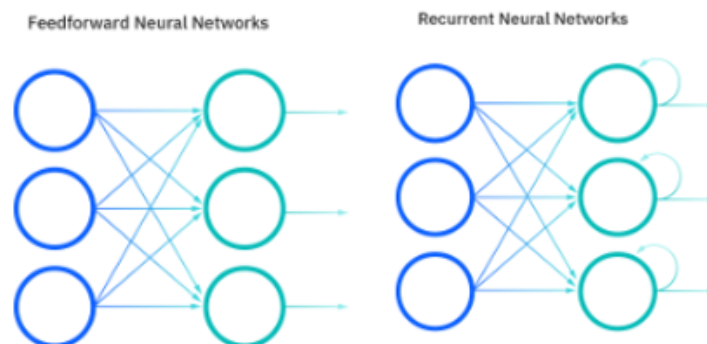


Figura 1: Redes neuronales prealimentadas contra redes neuronales recurrentes

## Tipologías de RNN

Existen distintos esquemas de entrada y salida:

- **One-to-One:** una entrada y una salida.
- **One-to-Many:** una entrada y múltiples salidas (ej. subtítulo de imágenes).
- **Many-to-One:** múltiples entradas y una salida (ej. clasificación de sentimientos o predicción de un valor futuro).

- **Many-to-Many**: múltiples entradas y múltiples salidas (ej. traducción automática).

A continuación, en la figura 2 se muestran los distintos tipos de redes neuronales recurrentes:

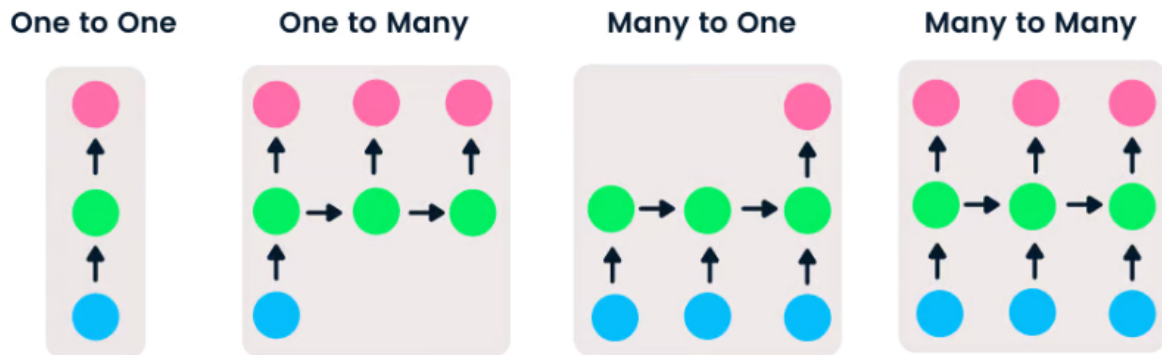
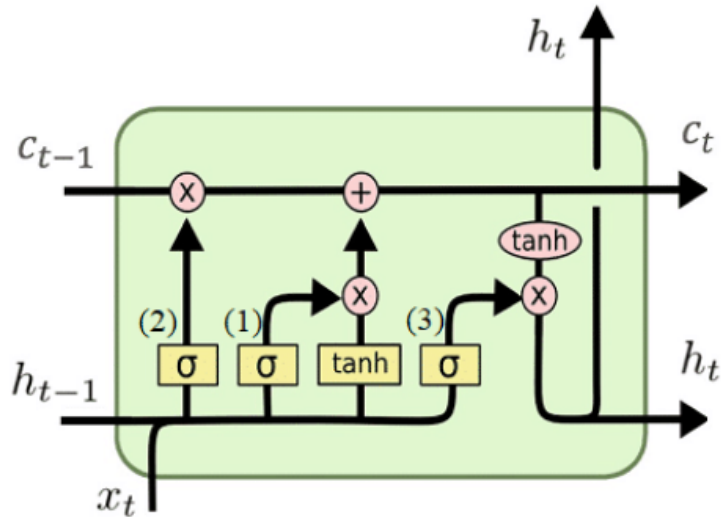


Figura 2: Tipos de redes neuronales recurrentes

En este trabajo se han empleado configuraciones *Many-to-One* y *Many-to-Many*, ya que permiten predecir horizontes de varios días en función de ventanas históricas.

### **Long Short-Term Memory (LSTM)**

Las **LSTM**, mostrada en figura 3, fueron introducidas por Hochreiter y Schmidhuber en 1997 [6] para resolver el problema del gradiente de desaparición. Incorporan una **celda de memoria** y **tres compuertas** (entrada, olvido y salida) que permiten regular el flujo de información, reteniendo aquella relevante y descartando la obsoleta. Gracias a esta arquitectura, las LSTM son capaces de aprender dependencias a largo plazo en secuencias temporales.



**LSTM**  
**(Long-Short Term Memory)**  
*Diagrama de una célula de memoria LSTM*

Figura 3: Arquitectura general de *LSTM*

Sus principales aplicaciones incluyen la **predicción de series temporales**, la **traducción automática**, la **generación de texto** y el **reconocimiento de voz**. No obstante, presentan algunas desventajas: mayor complejidad de entrenamiento, necesidad de más datos y recursos computacionales, y mayor riesgo de sobreajuste.

### ***Gated Recurrent Units (GRU)***

Las **GRU**, propuestas por Cho et al. (2014) [10], simplifican la arquitectura de las LSTM al reducir el número de compuertas a dos: **actualización** y **reseteo**. Estas controlan qué información se conserva del pasado y cuál se descarta. De este modo, las GRU combinan simplicidad y eficiencia, requiriendo menos parámetros y menor coste computacional, lo que se traduce en **entrenamientos más rápidos** y una menor probabilidad de sobreajuste.

La arquitectura general de una *GRU* se muestra en figura 4.

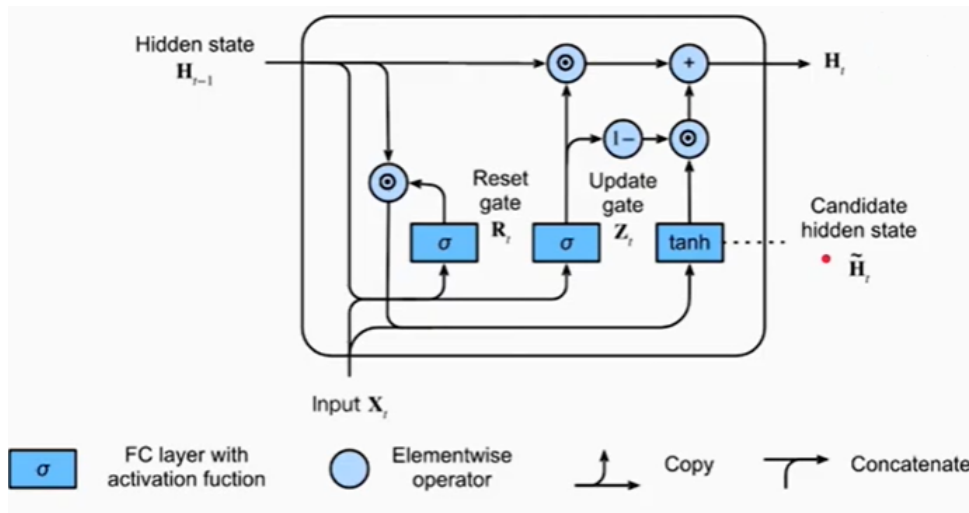


Figura 4: Arquitectura general de GRU

Entre sus aplicaciones destacan el **procesamiento de lenguaje natural**, el **reconocimiento de voz** y la **predicción de series temporales**. En el contexto de este trabajo, las GRU resultan especialmente adecuadas porque proporcionan métricas similares a las LSTM, pero con un 30–40 % menos de parámetros y menor coste de cómputo, lo que facilita el reentrenamiento periódico con datos de los embalses andaluces.

En conclusión, las redes neuronales recurrentes y sus variantes constituyen una herramienta fundamental para la predicción hidrológica en este trabajo. Su capacidad de manejar dependencias temporales y capturar relaciones no lineales las convierte en un pilar metodológico clave, especialmente en combinación con la estrategia de *transfer learning*, que permite transferir modelos entrenados en embalses con mayor disponibilidad de datos hacia otros con menor instrumentación.

### 2.3. Contexto del estudio

Esta sección enmarca el problema desde el punto de vista ambiental, institucional y operativo en Andalucía. Primero, se describe la **sequía** como rasgo estructural del régimen hidrológico y se distingue entre **sequía prolongada** y **escasez**, cuyo seguimiento mediante indicadores y umbrales permite activar medidas graduales de gestión. Después, se presenta el **marco de gobernanza**: la CHG en la cuenca del Guadalquivir (Comisión de Desembalse) y la Junta de Andalucía en las cuencas intracomunitarias (PES y Plan de Actuación), subra-

yando cómo las **decisiones humanas** modulan los desembalses. A continuación, se resume la **organización geográfica de los embalses**, condicionada por gradientes altitudinales y climáticos que afectan aportaciones, demanda y operación, y se precisan las variables de referencia (reserva en  $\text{hm}^3$  y porcentaje de reserva). Por último, se presenta la **red de estaciones agroclimáticas** y su integración con las series hidrológicas, base de los datos diarios utilizados en el modelado.

### 2.3.1. Fenómeno de la sequía

La sequía constituye un rasgo estructural del régimen hidroclimático andaluz y afecta, de forma recurrente, a la disponibilidad de agua para abastecimiento urbano, regadío, usos industriales y producción hidroeléctrica. En la demarcación del Guadalquivir (referencia hidrológica principal para Andalucía) el seguimiento operativo distingue entre **sequía prolongada** (déficit meteorológico sostenido) y **escasez** (insuficiencia coyuntural de recurso para atender demandas), evaluadas mediante indicadores normalizados por Unidades Territoriales (UTE/UTS) y umbrales que definen escenarios de normalidad, prealerta, alerta y emergencia. Esta arquitectura permite activar medidas graduales de gestión y priorización de usos en función del estado diagnosticado.

#### **Evolución reciente (2024–2025)**

Los informes mensuales de la Confederación Hidrográfica del Guadalquivir (CHG) muestran una mejora de las reservas embalsadas entre agosto de 2024 y agosto de 2025. A 31/08/2024, el volumen total almacenado en la demarcación alcanzaba  $2.653 \text{ hm}^3$  (33,0 % de la capacidad), con valores especialmente contenidos en el sistema de regulación general; a 31/08/2025, el total ascendía a  $3.655 \text{ hm}^3$  (45,5 %), reflejando una recuperación parcial tras los episodios secos previos. En paralelo, el **índice global de sequía prolongada** de la demarcación se situó en **0,674 en agosto de 2024** y **0,839 en agosto de 2025**, ambos compatibles con ausencia de sequía prolongada en términos meteorológicos, si bien coexistieron situaciones de **escasez** en varias UTE con estados de prealerta o alerta, por ejemplo, en Hoya de Guadix (2024) o Guardal (2025) [18][19][20].

**Impactos sectoriales.** La combinación de déficits pluviométricos, altas temperaturas y elevada evaporación reduce aportaciones y acelera descensos de almacenamiento, con repercusiones directas en:

- **abastecimientos urbanos** (modulaciones de demanda, búsqueda de recursos complementarios),
- **agricultura de regadío** (reducciones de dotación y ajustes de campaña) y
- **energía hidroeléctrica** (limitaciones de turbinado). La documentación operativa de CHG ilustra, además, el apoyo en **recursos no convencionales** en ámbitos extrapeninsulares de la demarcación (Ceuta/Melilla) [19], donde la operatividad de las plantas de **desalinización** forma parte del marco de garantía.

**Medidas de mitigación y gestión.** Los **Planes Especiales de Sequía** y los informes operativos asociados establecen un catálogo de actuaciones escalonadas, que incluyen campañas de ahorro, reasignaciones y priorización de usos, explotación coordinada de embalses, refuerzos de abastecimiento con fuentes alternativas (desalinización, subterráneas) y obras de emergencia en conexión y transporte. La activación se vincula a los escenarios UTE/UTS mencionados y se actualiza con la evolución de los indicadores (escasez y SPI para sequía prolongada).

En definitiva, el periodo reciente confirma la **variabilidad intraanual** propia del ámbito mediterráneo: episodios secos prolongados con impacto en almacenamiento y disponibilidad conviven con fases de recuperación parcial cuando las precipitaciones repuntan. La gestión adaptativa, basada en indicadores y medidas graduales, resulta clave para atenuar los efectos sobre abastecimiento y regadío y para sostener la resiliencia del sistema en escenarios de presión climática creciente.

### 2.3.2. Política de desembalse y factores humanos

#### Marco institucional y jurídico

En Andalucía conviven dos niveles de gestión:

## 1. Cuenca intercomunitaria del Guadalquivir (estatal)

La **Confederación Hidrográfica del Guadalquivir (CHG)** celebra cada campaña la Comisión de Desembalse, órgano donde se acuerdan volúmenes a liberar por sistemas (riego, abastecimiento, caudales ecológicos y otros usos). Los acuerdos fijan dotaciones, periodos y condicionantes operativos a partir de la situación hidrológica y de los planes vigentes. La propia CHG publica los “**Acuerdos de la Comisión de Desembalse**” [21], que articulan estas decisiones para cada año hidrológico.

## 2. Cuencas intracomunitarias (autonómicas)

En **Cuencas Mediterráneas Andaluzas, Guadalete-Barbate y Tinto-Odiel-Piedras**, la Junta de Andalucía gestiona la sequía mediante los **Planes Especiales de Sequía (PES)** y, desde 2022, a través del **Plan de Actuación frente a la Sequía**, que faculta una **Comisión para la Gestión de la Sequía** a adoptar medidas excepcionales: límites de consumo urbano, reducciones de dotaciones de riego, priorización de usos esenciales y otras actuaciones de ahorro. Este marco quedó activado por acuerdo del Consejo de Gobierno (5/09/2022) y regula cómo se aplican restricciones según los estados normalidad-prealerta-alerta-emergencia definidos en los PES [22].

De forma práctica, estas comisiones autonómicas han ido modulando restricciones por demarcación. Por ejemplo, en **Guadalete-Barbate** se han ajustado en distintas convocatorias **límites de consumo urbano ( $\approx 200\text{--}225\text{ L/hab/día}$ ) y reducciones para regadío** en función de la evolución de reservas y aportaciones, con comunicación pública de cada acuerdo por la Junta [23].

Además, la Junta dispone de normativa y guías técnicas donde se establecen **umbrales de referencia** para consumo urbano en sequía (p. ej., 200 L/hab/día) y el **escalado de medidas** desde campañas informativas hasta restricciones y cortes programados en niveles severos, lo que enmarca la toma de decisiones en Cuencas Mediterráneas Andaluzas, Guadalete-Barbate y Tinto-Odiel-Piedras.

**Factores humanos que condicionan el desembalse** Aunque la hidrología y la meteorología

logía marcan la disponibilidad, el **perfil de demanda y las decisiones de gestión** terminan modulando los desembalses. En Andalucía destacan:

- **Abastecimiento urbano y turístico.** La estacionalidad del consumo (verano y picos turísticos, especialmente en **Costa del Sol y litoral mediterráneo**) obliga a programar liberaciones que garanticen continuidad del servicio dentro de los límites fijados por las comisiones (p. ej., 200–225 L/hab/día cuando hay sequía).
- **de regantes** solicitan dotaciones en función de cultivos (olivar, hortofrutícolas, arroz, etc.) y calendario agrícola. CHG y las comisiones autonómicas ajustan **dotaciones y ventanas de riego** y, si es preciso, recortan porcentajes para preservar el suministro esencial y la conservación de reservas. En el Guadalquivir estas decisiones se formalizan en la **Comisión de Desembalse** anual [21].
- **Exigencias ambientales y seguridad.** Los **caudales ecológicos** y otras condiciones ambientales de los planes hidrológicos condicionan caudales mínimos y curvas de explotación; la **seguridad de presas** y maniobras de mantenimiento también originan desembalses puntuales. (Estos condicionantes se integran en los PES y en la operativa ordinaria de los organismos de cuenca) [22].
- **Medidas excepcionales de sequía.** En fases de **alerta o emergencia**, las comisiones (Junta) pueden imponer **recortes a riego, restricciones urbanas, cambios de prioridad de usos** y medidas de ahorro; en Guadalete-Barbate se han anunciado varias actualizaciones de restricciones en los últimos años conforme cambiaba el escenario.

### 2.3.3. Organización de los embalses andaluces

La planificación hidrológica en Andalucía persigue dos grandes objetivos: conservar y proteger el dominio público hidráulico, y satisfacer las demandas de agua de una población diversa y en crecimiento. Esta comunidad autónoma presenta una marcada heterogeneidad climática e hidrológica, lo que se traduce en la existencia de más de 70 embalses que cumplen funciones como el control de avenidas, generación hidroeléctrica, abastecimiento urbano y riego agrícola.

La gestión de estos embalses, como se ha mencionado anteriormente, está dividida entre organismos estatales y autonómicos, en función de si la cuenca hidrográfica es intercomunitaria o intracomunitaria:

- **Confederaciones hidrográficas:** organismos estatales que gestionan las cuencas intercomunitarias. En Andalucía destaca la Confederación Hidrográfica del Guadalquivir (CHG), la más extensa de la región.
- **Demarcaciones hidrográficas:** organismos dependientes de la Junta de Andalucía encargados de las cuencas intracomunitarias, entre las que se encuentran las Cuencas Mediterráneas Andaluzas, Guadalete-Barbate y Tinto-Odiel-Piedras.

Este reparto competencial permite adaptar la gestión del agua a las particularidades territoriales, garantizando una mayor eficiencia en la administración de los recursos hídricos.

### **Organización geográfica**

La red de embalses andaluces se distribuye de manera desigual siguiendo la organización hidrográfica de la región. La mayor concentración se localiza en la Demarcación del Guadalquivir, que vertebra el interior peninsular desde Sierra Morena y las Subbéticas hacia el Atlántico; en el litoral atlántico (Guadalete-Barbate y Tinto-Odiel-Piedras) predominan presas de tamaño medio orientadas al abastecimiento y a la laminación de avenidas en cuencas cortas y de respuesta rápida; y a lo largo de las Cuencas Mediterráneas Andaluzas los embalses se encadenan en valles encajados de la Bética (Serranía de Ronda, Sierra Nevada y sierras prelitorales), con menores cuencas tributarias y marcada estacionalidad.

En el plano altitudinal, el sector oriental y bético concentra obras en cotas medias-altas (valles de montaña y piedemontes), mientras que en el occidental (campiñas y marismas) abundan infraestructuras en cotas bajas ligadas a amplias llanuras aluviales.

Esta gradación se alinea con el gradiente climático (desde condiciones semiáridas en el sureste hasta ambientes subhúmedos en el valle medio del Guadalquivir y húmedos en enclaves orográficos del suroccidente y Sierra Nevada), de modo que la localización condiciona el régimen de aportaciones, la presión de demanda y la estrategia operativa de cada sistema.

Cada uno de los embalses cuenta con dos variables que se almacenan en una base de datos de series temporales:

- **Reserva (hm3):** Es el volumen de agua almacenada en el embalse dedicado a abastecimiento o riego. Se mide en hectómetros cúbicos.
- **Porcentaje de reserva (%):** Es la relación entre la reserva y la capacidad total del embalse. Se mide en tantos por cien.

#### 2.3.4. Red de estaciones agroclimáticas

Una **estación agroclimática** es un conjunto de instrumentos y sensores destinados a registrar, con precisión y continuidad, las principales variables meteorológicas que condicionan los procesos hidrológicos y agronómicos (temperatura, humedad, precipitación, radiación, viento, evapotranspiración de referencia, entre otras). Su implantación exige un estudio previo de ubicación, exposición y configuración instrumental para asegurar mediciones representativas y comparables en el tiempo. Existen dos tipologías: **manuales** (instrumentación analógica con lectura periódica) y **automáticas** (sensórica digital con registro y envío telemático continuo), siendo estas últimas las que permiten cubrir zonas de difícil acceso y horarios no laborables con mayor resolución.

Las estaciones automáticas empleadas en este estudio comparten una arquitectura habitual:

- **mástil** con soportes para piranómetro y anemoveleta
- caja intemperie con registrador de datos (datalogger)
- regulador y módem GSM, alimentación autónoma (batería recargada por panel solar)
- conjunto de **sensores** (sonda combinada de temperatura–humedad protegida frente a radiación y condensación, pluviómetro, anemómetro y veleta, piranómetro)
- La **unidad central** procesa las señales y las transmite, garantizando la trazabilidad de cada observación

Las **variables** registradas diariamente incluyen:

- temperaturas máxima, mínima y media (°C)
- humedades relativa máxima, mínima y media (%)
- velocidad y racha máxima del viento (m/s)
- dirección del viento (°)
- radiación global (MJ/m<sup>2</sup>)
- precipitación (mm)
- evapotranspiración de referencia (ET<sub>o</sub>), estimada a partir de radiación, viento, temperatura y humedad

Estas series nutren el conjunto de datos meteorológicos del proyecto y se integran con las series hidrológicas de los embalses.

Para cada embalse del estudio se asoció, de forma dinámica, hasta **tres estaciones agroclimáticas cercanas** (priorizando siempre la más próxima operativa en cada fecha) con el fin de  **cubrir todo el periodo 2000–2024**  con datos diarios consistentes. Esta asociación se realizó mediante geolocalización (coordenadas de presa y estaciones) y un mapa de correspondencia embalse–estación, normalizando además la nomenclatura (homogeneización de acentos, artículos y espacios) para evitar ambigüedades en los procesos posteriores de *clustering* y entrenamiento de modelos.

## 2.4. Métricas

Este apartado consolida y formaliza todas las métricas usadas en la memoria, agrupadas por finalidad y con sus fórmulas. Donde procede, se indican las líneas base con las que se comparan los modelos predictivos.

### 2.4.1. Evaluación de modelos predictivos

En los experimentos de modelado (másters, *outliers* y *transfer learning*) se evalúa el ajuste sobre el **porcentaje de llenado a H=14 días**. Dado que la persistencia es un baseline muy competitivo a este horizonte, las métricas absolutas se acompañan siempre de una **comparación contra líneas base** (persistencia y climatología).

- **Mean Absolute Error (MAE)**[24]

Mide la desviación media en valor absoluto entre observaciones y predicciones; en este proyecto se interpreta en **puntos porcentuales (p.p.)**.

$$\mathbf{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - x_i|$$

- **Root Mean Squared Error (RMSE)**[25]

Esta métrica penaliza más los errores grandes, revelando picos de fallo.

$$\mathbf{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2}$$

Si  $\mathbf{RMSE} \gg \mathbf{MAE}$ , existen episodios de error alto (colas gruesas); si son similares, el error es homogéneo.

- **Mean Absolute Percentage Error (MAPE)**[26]

Error relativo medio en %.

$$\mathbf{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - x_i}{y_i} \right|$$

Intuitivo para comparar entre embalses con escalas distintas; evitar cuando  $y_i$  se acerca a 0 (inestabilidad del denominador).

- **Coficiente de determinación ( $R^2$ )**[27]

Métrica que mide la proporción de varianza del objetivo explicada por el modelo.

$$\mathbf{R}^2 = 1 - \frac{\sum_{i=1}^n (y_i - x_i)^2}{\sum_{i=1}^n (y_i - \mu_y)^2}$$

1 es perfecto; 0 equivale a predecir la media; **valores negativos** indican peor que la media. A H=14, acercarse o superar la **persistencia** es el listón relevante.

▪ **skill respecto a una línea base (climatología/persistencia)**

Mejora del modelo sobre una referencia (persistencia o climatología).

$$R^2 : Skill = R_{\text{modelo}}^2 - R_{\text{base}}^2$$

Interpretación de los resultados:

- **>0**: el modelo supera la referencia
- **≈ 0**: similar (puede ser aceptable si la base es muy fuerte;
- **<0**: peor que la base

**Significado de las variables empleadas:**

- $y_i$ : valor observado del objetivo en el instante (o fecha)  $i$ . En este trabajo: **porcentaje de llenado, en puntos porcentuales (p.p.)**.
- $\hat{y}_i$ : valor predicho por el modelo para ese mismo instante  $i$ , también en p.p
- $n$ : número total de observaciones sobre las que se evalúa la métrica (todas las fechas y, si aplica, todos los plazos del horizonte considerados en la evaluación).
- $R_{\text{modelo}}^2$  y  $R_{\text{base}}^2$ : coeficiente de determinación calculado en el **mismo conjunto de evaluación** para el **modelo** y para la **línea base** (persistencia o climatología), respectivamente.
- $\bar{y}_i$ : media aritmética de los valores observados en el conjunto de evaluación
- **Líneas base:**
  - **Persistencia:**  $\hat{y}_{t+h} = y_t$  (el valor futuro a  $t+h$  se toma igual al valor actual)
  - **Climatología:** media histórica por **día del año** del horizonte evaluado.

**2.4.2. Validación del agrupamiento de embalses**

Para justificar la elección de la partición en  $K$  clústeres se recurrirá a métricas internas de cohesión y separación. A continuación se **presentan y fundamentan** tres indicadores complementarios (**Silhouette**, **Calinski–Harabasz** y **Davies–Bouldin**) seleccionados porque no requieren etiquetas externas, aportan perspectivas ortogonales sobre la estructura de

los datos (calidad punto a punto, razón entre dispersión intra/intergrupo y similitud entre clústeres) y son computacionalmente eficientes y de uso extendido en la literatura. Todas las métricas deben recalcularse en cada experimento y para cada valor de K, pues sus magnitudes no son comparables entre ejecuciones distintas. En primer término, se expone el **método de Silhouette**.

- **Método de Silhouette**[28]

Calidad de la asignación por punto, comparando compactación interna y separación al clúster vecino.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

Siendo a(i) la distancia media a su clúster y b(i) la mínima distancia media a otros.

Rango [-1,1][1,1][1,1]. En datos **altamente dimensionales** con ruido, **0.2–0.4** puede ser un rango aceptable si hay coherencia espacial y temática; valores próximos a 0 indican fronteras difusas.

- **Índice Calinski-Harabasz**[29]

Es una métrica que se encarga de medir la razón entre dispersión **inter-clúster** y **intra-clúster**.

$$CH = \frac{\text{tr}(B_k)/(k - 1)}{\text{tr}(W_k)/(n - k)}$$

Mayores valores sugieren grupos más compactos y separados. **Comparativa por K:** no hay umbral universal.

- **Índice de Davies-Bouldin** [30]

Similitud media entre clústeres usando razón de dispersiones internas sobre distancia entre centros.

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{\substack{j=1 \\ j \neq i}}^k \frac{s_i + s_j}{M_{ij}}$$

En torno a 1 suele indicar particiones razonables; <1 muy buena separación; >2 sugiere solapamiento notable. **Recalcular para cada K.**

### 2.4.3. Selección de características y análisis

Para estudiar la señal de las variables y priorizar familias de predictores se combinan medidas de correlación y relevancia.

- **Correlación de Pearson**[31]

Mide la asociación lineal entre dos variables, en  $[-1,1][1,1][1,1]$ .

$$r = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

Signo positivo (relación directa) o negativo (inversa). En reservas **sin desfase** es normal observar  $|r|$  bajos (inercia del sistema); al introducir **lags** (0–60 días) emergen patrones físicamente plausibles (temperatura y radiación con signo negativo; precipitación/humedad con signo positivo). Uso en la elaboración de los *heatmaps* por embalse/provincia y barridos de *lags*.

- **Correlación de Spearman**[32]

Asociación monótona (de rangos); capta relaciones no lineales suaves. Su fórmula es la **Pearson aplicado a los rangos**. Apoya el análisis de **representatividad intra-clúster** y la robustez frente a *outliers* cuando Pearson es demasiado restrictiva.

- **Random Forest** [33]

- **Impureza**. Suma de reducciones de impureza (varianza) por variable. Es rápida, pero sesgada hacia variables muy correlacionadas.
- **Permutación**. Caída de rendimiento al **barajar** una columna manteniendo el resto; ofrece una estimación más **robusta** a colinealidad, a costa de mayor coste.

En esta línea de trabajo, ha sido empleada para contrastar hallazgos de correlación/LAGS y fijar una **parquedad por familias** (temperaturas/ humedades/ vientos/ radiación/ precipitación) evitando duplicidades.

### 2.4.4. Criterios de selección de máster de clúster

Además de la calidad de ajuste, la selección de **embalses máster** exige estabilidad y representatividad, para que el **transfer learning** sea fiable dentro de cada clúster.

▪ **Nash-Sutcliffe Efficiency (NSE)**

El **Nash-Sutcliffe Efficiency (NSE)** compara el error del modelo con la variabilidad natural de la serie respecto a su media:

$$NSE = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Toma **1** como ajuste perfecto; **0** equivale a **predecir siempre la media**; valores negativos implican peor que esa referencia.

Usar  $NSE \geq 0.50$  como umbral implica que el modelo explica al menos la mitad de la variación respecto a la media histórica. Es un estándar operativo razonable en hidrología diaria porque:

- discrimina modelos que **añaden valor real** frente a usar una media fija;
- es **sensible a errores grandes** (al ser cuadrático), por lo que penaliza picos de fallo;
- facilita comparar candidatos (másters/targets) con una **métrica bien conocida** en la comunidad.

Las opciones serían:

- $NSE \geq 0.50$ : desempeño **aceptable/satisfactorio**; hay ganancia clara frente a la media.
- $NSE \approx 0.65-0.75$ : **bueno-muy bueno** (orientativo).
- $0 \leq NSE < 0.50$ : mejor que la media, pero con valor añadido limitado; recomendable para uso exploratorio o con medidas de mejora/calibración antes de uso operativo.
- $NSE < 0$ : el modelo es **peor que una media constante** (debe revisarse).

El NSE puede verse afectado por **episodios extremos** y por la persistencia de la serie; por eso lo complementamos con otras métricas (MAE/RMSE,  $R^2$ , skill vs. persistencia/-climatología) y con el control de **sesgo** anterior. En conjunto,  $|\text{Bias}| \leq 5 \text{ p.p.} + NSE \geq 0.50$  define un modelo **centrado** y con capacidad explicativa suficiente para uso operativo y para ser elegido **máster/target** en el flujo de *transfer learning*.

■ **Kling-Gupta Efficiency (KGE)**

Mide el balance simultáneo de **correlación** ( $r$ ), **variabilidad relativa** y **sesgo**.

$$KGE = 1 - \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2}$$

por lo que 1 es perfecto (alta correlación, sin sesgo y con variabilidad bien reproducida) y pueden darse valores **negativos** cuando el ajuste es pobre. En la práctica de hidrología operativa, **KGE  $\geq$  0.45** se usa como **umbral orientativo**: indica que el modelo **capta razonablemente la señal**, mantiene **sesgo controlado** y reproduce **variabilidad** sin desviarse en exceso. En tu flujo, ese umbral marca **aceptación** (p. ej., para seleccionar “másters” o validar un modelo); por debajo de 0.45, el desempeño se considera insuficiente para uso operativo.

■ **Sesgo medio (Bias)**

El **bias** (sesgo) es el **error medio del modelo**: indica si, en promedio, **sobreestima** o **subestima** el nivel/porcentaje del embalse.

$$\text{Bias} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)$$

donde  $y_i$  es el valor observado y  $\hat{y}_i$  el predicho. Lo expresamos en **puntos porcentuales (p.p.)** cuando el objetivo es el **% de llenado**.

Las opciones son:

- **Bias  $\approx$  0**: el modelo no tiene tendencia sistemática.
- **Bias  $>$  0**: sobreestima de forma sostenida.
- **Bias  $<$  0**: subestima de forma sostenida.

Fijar **|Bias|  $\leq$  5 p.p.** exige que el **desplazamiento sistemático** del modelo (su “desalineación” media) sea **pequeño** a escala operativa. Es un umbral práctico:

- controla errores persistentes que podrían **acumular decisiones equivocadas** (por ejemplo, liberar/retener agua por una señal sesgada);
- deja margen para la **variabilidad diaria** sin penalizar al modelo por ruido inevitable;

- es comparable entre embalses porque está en p.p. del porcentaje de llenado.

Cumplir  $|Bias| \leq 5$  p.p. **no garantiza baja dispersión** (puede haber alta variabilidad alrededor de cero), pero sí que el **error promedio** está **bien centrado**.

#### ▪ Variabilidad mínima de la serie ( $\sigma$ )

Es la **desviación estándar** del objetivo (porcentaje de llenado) en puntos porcentuales:

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2}$$

$\sigma \geq 5$  p.p. Si la serie es casi plana ( $\sigma$  muy baja), **no hay señal dinámica suficiente** para evaluar de forma fiable un modelo:

- Métricas relativas como  $R^2$  o NSE se vuelven **inestables** (el denominador es muy pequeño).
- Errores absolutos (MAE/RMSE) pueden ser **artificialmente bajos** sin que el modelo haya aprendido nada sustantivo. Exigir **al menos 5 p.p.** de variabilidad garantiza que el embalse presenta cambios apreciables y que la evaluación del desempeño es **informativa y comparable**.

#### ▪ Test de Kolmogorov-Smirnov (KS)

Comparamos la **distribución** del objetivo en los **últimos 2 años** con la del **histórico** para detectar cambios de régimen (derivadas). Usamos la **prueba de Kolmogorov-Smirnov (KS) a dos muestras**:

- **Hipótesis nula ( $H_0$ )**: ambas muestras provienen de la **misma** distribución.
- **Estadístico D**: **máxima diferencia** entre las funciones de distribución empíricas de ambas muestras (cuanto mayor, más cambio).
- **Valor-p ( $p$ )**: probabilidad de observar un D tan grande (o mayor) si  $H_0$  fuera cierta.

**Regla práctica de aceptación:**

- Si  $p \geq 0.05$  **no rechazamos  $H_0$** : no hay evidencia de cambio; consideramos la **serie estable**.

- Complementamos con un **umbral geométrico**  $D \leq 0.05$  como chequeo adicional de **tamaño de efecto** (útil porque el  $p$ -valor depende del tamaño muestral). Cumplir cualquiera de los dos criterios nos da suficiente confianza de **estacionariedad operativa** para selección de másters/targets y para aplicar *transfer learning* sin arrastre de sesgos recientes.
- $p \geq 0.05$  o  $D \leq 0.05$  - **estable** (sin evidencia de deriva).
- $p < 0.05$  y  $D > 0.05$  - **posible cambio de régimen**; conviene revisar o excluir.

#### ■ Representatividad intra-clúster

Mide qué “típico” es un embalse dentro de su grupo.

Los indicadores son: **Distancia al centro** (en PCA/espacio original) y **correlación (Pearson/Spearman)** con el perfil medio del clúster. Se usa para priorizar **másters que representen al conjunto**, no solo que ajusten bien su propia serie.

#### 2.4.5. Geolocalización y distancia

##### ■ Fórmula de Haversine

Se usan las coordenadas de presa y estaciones; la proximidad se calcula con la **fórmula de Haversine**. Esta fórmula estima la distancia ortodrómica (de círculo máximo) entre dos puntos de la Tierra, asumiendo una esfera. A partir de latitudes y longitudes en radianes:

$$d = 2R \arcsin \left( \sqrt{\sin^2 \left( \frac{\varphi_2 - \varphi_1}{2} \right) + \cos \varphi_1 \cos \varphi_2 \sin^2 \left( \frac{\lambda_2 - \lambda_1}{2} \right)} \right)$$

Donde las variables utilizadas son las siguientes:

- $\varphi_1, \varphi_2$ : latitudes de los dos puntos.
- $\lambda_1, \lambda_2$ : longitudes de los dos puntos.
- $R$ : radio de la Tierra (usa 6371km para  $d$  en km, o 6371000m  $d$  en metros).
- $d$ : distancia ortodrómica (de círculo máximo) entre ambos puntos.

## 2.5. Tecnologías empleadas

En este punto se presenta el **ecosistema tecnológico** que sostiene el proyecto de extremo a extremo, desde la captura del dato hasta su visualización y despliegue. La **capa de datos y ETL** se articula en torno a **Python** (pandas, NumPy, Selenium) y una **base de series temporales (InfluxDB)** que permite ingesta, limpieza y consultas eficientes por tiempo. La **capa de modelado** emplea **TensorFlow/Keras** y utilidades de **scikit-learn** para el entrenamiento y evaluación de modelos secuenciales (GRU), con **Joblib** para serialización y **Google Colab** como entorno de cómputo en la nube cuando conviene. La **capa de servicio** expone los resultados mediante **Node.js** y **Express**, y se orquesta con **Docker** para garantizar portabilidad y reproducibilidad. Por último, la **capa de presentación** combina **HTML/CSS/Bootstrap** y **JavaScript/React** (con **Chart.js**) para ofrecer una interfaz interactiva. Todo el desarrollo se apoya en **Git + GitHub** y **Visual Studio Code**, priorizando software abierto, trazabilidad y despliegues consistentes. Esta combinación asegura **escalabilidad, mantenimiento y uso operativo** en un contexto de datos temporales de alta frecuencia.

### 2.5.1. Python

Python [34] ha sido el lenguaje de programación principal en el desarrollo de este trabajo debido a su versatilidad y al amplio ecosistema de librerías orientadas al análisis de datos, la modelización estadística y la construcción de modelos de aprendizaje automático. Su sintaxis sencilla y su gran comunidad de usuarios lo convierten en una herramienta idónea para proyectos de investigación aplicada. Además, Python ha facilitado la integración con bases de datos, procesos *ETL* y el despliegue de modelos en entornos de producción.

Las características más importantes de este lenguaje son:

- **Tipo dinámico:** no es necesario declarar el tipo de dato; se infiere en tiempo de ejecución.
- **Sintaxis legible:** facilita el desarrollo y la revisión de código por equipos.
- **Ecosistema científico:** librerías para datos y ciencia (NumPy, pandas, SciPy, matplotlib).

- **Aprendizaje automático:** soporte maduro para ML/DL (scikit-learn, TensorFlow, PyTorch).
- **Gestión de entornos:** aislamiento y dependencias con venv/conda/poetry.
- **Interoperabilidad:** integración sencilla con APIs, bases de datos y servicios externos.

### 2.5.2. InfluxDB (Influx DataBase)

InfluxDB [35] se ha utilizado como base de datos temporal para almacenar y consultar las series cronológicas procedentes de embalses y estaciones meteorológicas. Su arquitectura está optimizada para manejar grandes volúmenes de datos indexados por tiempo, lo que permite realizar consultas rápidas y eficientes. Esto ha resultado fundamental para la gestión de datos hidrológicos y meteorológicos, que presentan alta frecuencia temporal y requieren operaciones como agregaciones y filtrados dinámicos.

Algunas de las características más relevantes y que ha decantado la balanza a usar esta base de datos son:

- **Modelo de series temporales:** esquema *measurement-tag-field* optimizado por tiempo.
- **Ingesta eficiente:** protocolo de línea (*Line Protocol*) para altas tasas de escritura.
- **Motor orientado a tiempo:** almacenamiento TSM y compacción por ventanas temporales.
- **Lenguaje de consulta:** soporte de Flux/InfluxQL para agregaciones y *downsampling*.
- **Retención y continuas:** políticas de retención y *continuous queries* para resúmenes.
- **Rendimiento en consultas:** lecturas rápidas sobre rangos temporales y *group by time()*.

InfluxDB admite dos lenguajes de consulta:

- **InfluxQL (Influx Query Language):** es un lenguaje de consulta similar a SQL orientado a trabajar con datos y series temporales. Se usa para versiones 1.x de influxDB y en algunos entornos de InfluxDB 2.0.

- Flux: es el lenguaje de consultas para InfluxDB 2.0, diseñado para ser más flexible y potente que InfluxQL. A partir de la versión 1.8 de InfluxDB se integra con InfluxQL para uso de producción.

### 2.5.3. JavaScript

JavaScript [36] fue creado en 1995 por Brendan Eich en Netscape Communications como lenguaje de programación para páginas web en el navegador de la compañía, *Netscape Navigator*. Con el tiempo, el resto de navegadores gráficos lo adoptaron hasta convertirlo en el estándar de facto del desarrollo en el lado del cliente. En este proyecto se ha empleado principalmente en el *frontend* de la aplicación web, lo que ha permitido implementar funcionalidades interactivas y dinámicas: exploración de las predicciones de los modelos, visualización de gráficas en tiempo real y gestión de filtros personalizados para cada embalse o estación.

Algunas de las características más relevantes de este lenguaje son:

- Ejecución en navegador: interacción directa con el DOM sin *plugins*.
- **Asincronía nativa:** *Promises* y *async/await* para I/O no bloqueante.
- **Módulos ES:** organización del código y reutilización entre componentes.
- **Ecosistema npm:** amplio catálogo de paquetes para el *frontend*.
- **Integración con APIs:** consumo sencillo de *REST* y *WebSocket* en clientes web.
- **Portabilidad:** funciona en cualquier navegador moderno.

### 2.5.4. VisualStudio Code

VisualStudio Code [37] es un editor de código fuente con licencia de código abierto, desarrollada por *Microsoft* y lanzado al público en 2015. Esta aplicación se ha utilizado como entorno de desarrollo integrado (IDE) para el trabajo de programación tanto en Python como en JavaScript. Su compatibilidad con extensiones específicas (control de versiones, entornos virtuales, integración con Docker, entre otras) ha facilitado el flujo de trabajo, la depuración de código y la colaboración.

### 2.5.5. Google Colaboratory

Google Colaboratory [38] es una herramienta gratuita, desarrollado en 2017 por *Google Research* alojada en *Jupyter Notebook* y basada en la nube que permite al usuario escribir, ejecutar y compartir código Python a través de los llamados *notebooks* o cuadernos, sin necesidad de configuración ni instalación en su máquina local. Es una gran solución especialmente en áreas como la ciencia de datos y el aprendizaje automático ya que nos permite entrenar modelos de *machine learning* y *deep learning* en CPU, GPU y TPU.

### 2.5.6. Git + Github

Git [39] es un sistema de control de versiones distribuido y desarrollado originalmente por Linus Torvalds en 2005, que registra la evolución del código mediante instantáneas (*commits*), ramas y fusiones, lo que facilita el trabajo colaborativo, la recuperación histórica y la gestión paralela de cambios. En este proyecto, el control de versiones se ha gestionado con Git, permitiendo llevar un registro detallado de las modificaciones y garantizar la trazabilidad del proyecto. *GitHub* ha servido como repositorio remoto, favoreciendo la colaboración, el almacenamiento seguro y el despliegue continuo de la aplicación. Además, el uso de ramas (*branches*) ha permitido experimentar con diferentes configuraciones de modelos sin comprometer la estabilidad del proyecto principal.

### 2.5.7. Docker

Docker [40] es una plataforma de contenedores que utiliza virtualización a nivel de sistema operativo para empaquetar aplicaciones y sus dependencias en imágenes que se ejecutan como contenedores aislados y portables sobre cualquier *host* con *Docker Engine*. En este proyecto se ha utilizado para la containerización de la aplicación, facilitando la portabilidad y la reproducibilidad del entorno de ejecución. Mediante contenedores se encapsularon dependencias, configuraciones y servicios (como InfluxDB y el backend), lo que simplificó el despliegue en diferentes máquinas y garantizó la consistencia del sistema en todas las fases del desarrollo.

Algunas de las ventajas de Docker frente a otras plataformas como *Red Hat OpenShift* o *Podman* son:

- **Contenedores reproducibles:** mismo entorno en desarrollo y producción.

- **Imágenes declarativas:** construcción con *Dockerfile* y caché por capas.
- **Orquestación ligera:** *docker-compose* para varios servicios (p. ej., API + InfluxDB).
- **Aislamiento de dependencias:** evita conflictos de librerías y versiones.
- **Portabilidad:** despliegue consistente en distintas máquinas/OS.
- **Redes y volúmenes:** comunicación segura y persistencia de datos.

### 2.5.8. Node.js

Node.js [41] es un entorno de ejecución de JavaScript del lado del servidor, basado en el motor V8 de Chrome y en un modelo de E/S no bloqueante orientado a eventos, diseñado para aplicaciones de red escalables y de baja latencia. En este trabajo se empleó en el *backend* de la aplicación web para gestionar la comunicación entre servidor y cliente. Su capacidad para manejar múltiples conexiones concurrentes de forma eficiente lo convierte en una opción adecuada para casos con consultas en tiempo real, como la recuperación de datos de predicción o el filtrado dinámico de series temporales.

Algunas de las características más destacables son:

- **E/S no bloqueante:** *event loop* eficiente para muchas conexiones concurrentes.
- **Tiempo real:** soporte nativo de *WebSocket* y *streams*.
- **API sencilla:** creación rápida de servicios y *microservices*.
- **Integración natural:** encaje directo con *Express* y *stacks full-JS*.

### 2.5.9. Bootstrap

Bootstrap [42] es un *framework frontend* orientado a construir interfaces web adaptables a cualquier dispositivo. Proporciona un sistema de rejilla (*grid*) y un catálogo amplio de componentes (botones, formularios, cards, navbars) que aceleran la maquetación y garantizan el diseño responsive sin partir de cero. Se distribuye tanto en versión precompilada (CSS/JS listos para usar) como en código fuente para personalización avanzada.

En este proyecto se empleó para estructurar las páginas públicas (inicio, datos y mapa) y asegurar la correcta visualización en móviles y escritorio. Entre sus ventajas prácticas destacan: curva de aprendizaje moderada con abundante documentación y comunidad; rejilla fluida para organizar columnas que se adaptan al ancho de pantalla; compatibilidad multiplataforma; y utilidades de imagen/espaciado que simplifican la estética sin *CSS ad-hoc*.

Algunas de sus ventajas son: [**nubecollectiva-bootstrap**]

- **Sistema de rejilla:** grid responsivo con *breakpoints* móviles y escritorio.
- **Componentes listos:** botones, cards, formularios, *navbars* y modales reutilizables.
- **Estilos utilitarios:** clases para espaciado, color, tipografía y visibilidad.
- **Documentación amplia:** guías y ejemplos para acelerar la maquetación.
- **Compatibilidad:** soporte multiplataforma y navegadores principales.

#### 2.5.10. Express

Express [43] es el *framework* de referencia para desarrollar APIs con Node.js. Facilita el enrutado, la gestión de *middleware* (autenticación, compresión, *logging*) y la exposición de endpoints REST de forma ligera y escalable. Aquí se utilizó para implementar la API que sirve datos históricos y agregados a la interfaz React, encapsulando consultas hacia InfluxDB mediante controladores sencillos y *middlewares* de validación y caché.

Algunas de las ventajas son:

- **Enrutado minimalista:** definición clara de rutas y controladores.
- **Middleware flexible:** autenticación, compresión, logging y caché *plug-and-play*.
- **Endpoints REST:** exposición de APIs ligeras y escalables.
- **Ecosistema maduro:** miles de *middlewares* y utilidades disponibles.
- **Configuración rápida:** arranque de servicios con pocas líneas de código.

### 2.5.11. HTML (HyperText Markup Language)

HTML [44] es el lenguaje de marcado estándar de la Web: define la estructura y el significado del contenido que renderiza el navegador (títulos, párrafos, tablas, formularios, SVG, etc.). No es un lenguaje de programación, sino una especificación de etiquetas y atributos que describe la semántica del documento y permite su integración con hojas de estilo (CSS) y *scripts* (JavaScript).

En la aplicación se siguieron convenciones semánticas (uso de `<header>`, `<main>`, `<section>`, `<nav>`, `<table>`...) para favorecer la accesibilidad y el SEO, además de facilitar el estilo con *Bootstrap* y CSS y la interacción con componentes React.

Algunas de las características destacables son:

- **Lenguaje de marcado:** usa etiquetas (tags) para definir la forma en la que los contenidos se muestran en la web.
- **Independiente de la plataforma:** se puede ver en cualquier navegador web de cualquier sistema operativo.
- **Basado en texto plano:** se puede editar con cualquier editor de texto.
- **Integrable:** se integra con CSS, JavaScript, SVG, entre otros.

### 2.5.12. CSS (Cascade Style Sheets)

CSS [45] es el lenguaje de estilos que controla la presentación visual del documento HTML (colores, tipografías, márgenes, *layout*, media queries). La “cascada” determina cómo se combinan reglas procedentes de distintos orígenes y niveles de especificidad. En este trabajo, además de las hojas de estilo de *Bootstrap*, se incorporaron estilos propios para ajustar la identidad visual (paleta, espaciados, *hover states*) y realizar pequeñas adaptaciones responsivas. Separar contenido (HTML/JSX) de presentación (CSS) permitió reutilizar componentes y aplicar cambios globales con bajo esfuerzo, reduciendo duplicidades y mejorando los tiempos de carga.

Alguna de sus ventajas son:

- Los cambios visuales lo hacemos en el mismo lugar y se aplica a toda la página.
- No se duplican los estilos en distintos lugares.

- El contenido que se transmite es menor, lo que hace que las páginas se carguen más rápido.
- A la hora de llevarlo a otras plataformas, como móviles, tablets, es más fácil de adaptarlo y crear versiones diferentes.

### 2.5.13. Algunas librerías destacables

- **Tensorflow**

**TensorFlow** es el *framework* de cómputo numérico y *deep learning* utilizado para definir, entrenar y desplegar los modelos secuenciales a través de su API **Keras**. Se eligió por su madurez en producción, la integración con Keras, que acelera el prototipado, las opciones de distribución y un ecosistema bien documentado que facilita el empaquetado y la integración en contenedores.

#### ¿Por qué TensorFlow frente a PyTorch?

La elección de *TensorFlow* frente a *PyTorch* respondió a su ruta de producción más madura (*TF Serving/TF Lite*), la integración nativa con Keras (que acelera el prototipado sin sacrificar rendimiento), los mecanismos de distribución (*tf.distribute*) y una documentación y ecosistema orientados a despliegue empresarial, lo que facilitó el empaquetado y la integración en contenedores.

- **Pandas**

Herramienta principal para la manipulación de datos tabulares y series temporales. Se utilizó en la lectura/escritura de CSV, el ordenado y el parsing robusto de fechas, la tipificación numérica y la limpieza selectiva de nulos. En la ingeniería de variables permitió extraer features de estacionalidad (día del año y día de la semana, posteriormente transformados con NumPy), generar *lags* con *shift* (p. ej., 1/3/7/14/30 días), acumulados, medias exponenciales, deltas del objetivo y *lag* anual. También se empleó para construir la climatología por día del año y para materializar particiones temporales (p. ej., *holdout* anual) antes del entrenamiento.

- **Scikit-learn (sklearn)**

Conjunto de utilidades para preprocessing, validación y aprendizaje clásico. Se utilizó *StandardScaler* para estandarizar las features (ajuste en entrenamiento y transformación en validación), se calcularon métricas de evaluación (*MAE*, *RMSE a partir de MSE*, *MAPE* y  $R^2$ ) y se apoyó la selección de variables con *RandomForestRegressor* y *Permutation Importance* sobre *TimeSeriesSplit*, asegurando causalidad en las comparaciones.

- **Matplotlib**

Biblioteca de visualización científica con la que se generaron las figuras empleadas en la memoria: curvas de entrenamiento/validación, comparativas “Real vs. Predicción” en *holdout* y en periodos largos, y gráficos de apoyo para el análisis exploratorio (p. ej., *heatmaps* de correlación). Se configuró un *backend* sin GUI para ejecución en servidor y se homogenizó la estética (rejilla, layout ajustado y exportación a PNG con dpi=150).

- **Joblib**

Utilizada para la serialización eficiente de artefactos ligeros, en particular el *scaler* de entrada y transformaciones auxiliares, lo que permitió reutilizar el mismo preprocesado en evaluación y re-ploteo sin necesidad de reentrenar.

- **Selenium + Webdriver Manager**

Conjunto para automatización de navegación web y descarga de ficheros cuando las fuentes carecen de API pública. Se empleó para *web scraping* de REDIAM (embalses) y de la Red RIA (meteorología), orquestando sesiones, resolviendo dependencias del driver de navegador y alimentando el *pipeline* de ingesta con datos actualizados.

- **React**

Biblioteca de interfaz de usuario con la que se implementó el *frontend*, estructurado en componentes reutilizables (tablas, selectores, gráficos) que consumen la API del *backend* y permiten filtrar por embalse, estación, provincia y ventana temporal, manteniendo desacopladas la presentación y la lógica de acceso a datos.

- **Chart.js** Librería de gráficos integrada en el *frontend* (a través de *wrappers* para React) para representar series temporales y agregados en líneas, barras y áreas, con leyendas interactivas y *tooltips* que facilitan la exploración de resultados y su comunicación al usuario final.

- **NumPy** Biblioteca fundamental para cálculo numérico y operaciones vectorizadas. Sirvió como base para construir ventanas temporales y tensores de entrada a red (*np.array*, *np.stack*), para generar codificaciones estacionales mediante funciones trigonométricas (*np.sin*, *np.cos*) y para cálculos auxiliares en métricas (p. ej., *np.sqrt* al derivar RMSE). Se estableció *np.random.seed(42)* como semilla de reproducibilidad compartida.

# 3

## Analisis experimental

Este capítulo presenta el marco completo del análisis: primero se describe la **extracción y procesamiento común de los datos** (fuentes, depuración, construcción de series y *features*, normalización y esquemas de validación), y a continuación se detallan los **experimentos realizados** (*clustering*, selección de horizonte, entrenamiento de modelos y *transfer learning*, así como comparativas de formulaciones). El objetivo es ofrecer un discurso **progresivo y trazable**: de los datos a los resultados, pasando por decisiones de diseño que justifican la configuración final.

### 3.1. Procesado de datos para el modelado

El **procesado y la selección de características** descritos en esta sección **aplican a todos los experimentos** del capítulo. En consecuencia, el orden será: en primer lugar, **extracción y procesamiento de datos**; en segundo lugar, **construcción y selección de características**; y, por último, **experimentación**. De este modo, todos los resultados posteriores heredan exactamente las mismas reglas de depuración, ingeniería y selección de variables, normalización y validación, garantizando **comparabilidad y reproducibilidad** y evitando el *leakage*.

El punto de partida fue un repositorio diario y homogéneo para 73 embalses andaluces, construido a partir de dos fuentes públicas capturadas mediante *web scraping*: REDIAM - Embalses de Andalucía [1] (volumen embalsado y porcentaje de llenado, con metadatos espaciales) y la Red Agroclimática RIA de la Junta de Andalucía [2] (precipitación, temperatura, humedad, radiación y viento). Las series brutas se volcaron inicialmente en InfluxDB y, tras una depuración básica, se exportaron a un CSV por embalse (una fila por día), que actúa como

insumo común para *clustering* y modelado.

Para asegurar coherencia entre catálogos heterogéneos se aplicó una estandarización nominal (eliminación de tildes, uso de mayúsculas y supresión de artículos/espacios, p. ej., **La Breña - Breña** y **Jándula - JANDULA**) tanto en embalses como en estaciones; además, se realizó un saneamiento sistemático: parseo robusto de fechas (zona horaria neutral), eliminación de duplicados temporales, tipificación numérica y descarte de valores físicamente imposibles.

Cada embalse se vinculó a una serie meteorológica “compuesta” a partir de hasta tres estaciones candidatas, seleccionadas por proximidad geográfica y vigencia operativa en 2000–2024. La proximidad se determinó con la métrica geodésica de Haversine. Además, para cada día, se adoptó una conmutación determinista: se usa la estación activa más cercana; si aparece una más próxima operativa o la actual deja de estarlo, se conmuta desde ese día. En la práctica, ningún embalse necesitó más de tres estaciones. Todo el proceso queda trazado (códigos, distancias y calendario de conmutaciones), garantizando continuidad temporal y representatividad espacial incluso cuando la estación más próxima no cubre todo el intervalo.

Sobre el CSV base (fecha ISO-8601, reserva y capacidad en  $\text{hm}^3$ , porcentaje de llenado y variables meteorológicas compuestas) se construyó el *dataframe* de modelado con **transformaciones causales** (calculadas solo con información disponible hasta  $t$ ): ventanas secuenciales de longitud fija (*steps* 120 o 180) y formulación directa a un horizonte  $H=14$  días (el modelo estima de una vez  $t+1\dots t+14$ ); memoria explícita mediante *lags* cortos 1,3,7,14,30 sobre meteorología y objetivo; agregados móviles multiescala 3,7,14,30 (medias, desviaciones, mínimos y máximos) especialmente en precipitación, temperatura y humedad; y codificación estacional con senos/cosenos del día del año y del día de la semana. Cuando procede, se integraron atributos estáticos (capacidad, superficie, coordenadas) desde el inventario maestro para análisis y segmentación, manteniendo la compatibilidad con el resto del *pipeline*.

Antes del entrenamiento se aplicó estandarización **z-score** usando exclusivamente estadísticas del **conjunto de entrenamiento**; estos normalizadores se reutilizaron sin reajuste en

validación y prueba para evitar *leakage* (cuando información que no debería estar disponible durante la etapa de entrenamiento se filtra en el proceso).

En escenarios de *transfer learning* se preservó estrictamente la paridad de features (orden, tipos y escalado) y, cuando fue pertinente, se reutilizaron los escaladores del **máster** para reducir el covariate *shift* durante el *fine-tuning*. El tratamiento de faltantes se pospuso al final del flujo: tras generar *lags* y *rolling* (que introducen *NaN* iniciales), se seleccionó [**objetivo + features**] y se aplicó un **dropna** limitado a ese bloque, evitando pérdidas de información ajenas al modelo. Asimismo, se controló la colinealidad extrema dentro de familias muy correlacionadas para favorecer la parquedad del conjunto de variables.

La validación combinó un **hold-out cronológico anual** (cola reciente) con **validación rolling** sobre tramos próximos al presente, e incluyó líneas base de **persistencia** y **climatología** para contextualizar el desempeño. Las definiciones y criterios de lectura de las métricas empleadas se desarrollan en el apartado “Métricas Empleadas”. Finalmente, cada experimento almacenó artefactos de reproducibilidad (lista y orden de *features*, escaladores, índices temporales y parámetros), lo que asegura **trazabilidad** entre embalses y facilita la repetición de resultados tanto en entrenamiento desde cero como en *transfer learning*.

### 3.1.1. Selección de características

Con el fin de asegurar la comparabilidad y la validez de los resultados, se realizó un proceso de preparación de datos, estrategias de entrenamiento y métricas de evaluación. A continuación, se definen los conceptos utilizados y se ofrecen pautas explícitas para interpretar los valores que aparecen posteriormente en los resultados.

#### preparación de datos y estabilidad del entrenamiento

Para garantizar comparabilidad y reproducibilidad, **todos los experimentos comparten la siguiente configuración** fija:

- **Arquitectura (GRU):** dos capas recurrentes con 128 y 64 unidades, seguidas de bloques densos 64–32–1 y activación lineal en la salida; con *dropout* en las capas donde se indica.
- **Optimización:** Adam con *gradient clipping clipnorm = 1.0*.

- **Función de pérdida:** Huber como referencia (y MSE en comparativas específicas).
- **Callbacks:** *EarlyStopping*, *ReduceLROnPlateau* y *ModelCheckpoint*.
- **Formulación:** modo **directo** como configuración operativa; modo **residual** reservado a casos muy inerciales. En validación se emplea *teacher forcing*; en residual, la red predice el incremento ( $\Delta$ ) que se suma al último valor observado.
- **Esquema temporal:** ventanas de *steps* = 120/180 y horizonte  $H = 14$  como estándar del capítulo (otros horizontes se estudian cuando se indique).
- **Particionado y entrenamiento:** separación cronológica (*hold-out* anual + validación *rolling*), sin *shuffle*; normalización *z-score* ajustada solo en entrenamiento y aplicada a validación/prueba para evitar *leakage*.

### Normalización *z-score*

Todas las variables numéricas se estandarizan restando su media y dividiendo por su desviación típica, calculadas exclusivamente sobre el conjunto de entrenamiento. Esta transformación evita dominancias por escala y previene fugas de información temporal.

### Pérdida Huber

Función de coste robusta que combina un tramo cuadrático para errores pequeños y uno lineal para errores grandes. Frente al MSE, reduce la influencia de valores atípicos y favorece estimaciones estables en series con picos.

### *Early stopping*

Criterio de parada temprana: si la métrica de validación deja de mejorar durante un número prefijado de épocas, se detiene el entrenamiento y se conserva el mejor estado del modelo observado hasta ese momento, mitigando el sobreajuste.

### ***Learning-rate schedulers***

Mecanismos que reducen automáticamente la tasa de aprendizaje cuando la métrica de validación se estanca. Facilitan una convergencia más fina en las últimas fases del entrenamiento.

### ***Warm-up (en transfer learning)***

Fase inicial en la que se congelan las capas recurrentes preentrenadas y se reentrena únicamente la “cabeza” de salida. Tras esta estabilización, se procede al fine-tuning con descongelado progresivo.

### **Esquema de validación temporal**

#### ***Hold-out anual***

Separación cronológica en la que uno o varios años completos se reservan para evaluación final. Evita la contaminación temporal entre entrenamiento y prueba.

#### ***Validación rolling.***

Conjunto de particiones deslizantes (*train/valid*) que reproducen distintos cortes temporales. Permite comprobar la consistencia del modelo ante cambios de régimen y reduce la dependencia del resultado respecto a un único periodo.

## **3.2. Detalles**

Aquí se detallan los pasos que convierten los datos brutos en modelos operativos y reproducibles. En primer lugar, se presenta el ***clustering*** de embalses: detección de atípicos con HDBSCAN, agrupación aglomerativa (*Ward*) y elección de un “máster” por clúster según desempeño, estabilidad y representatividad. Por último, se justifican el **horizonte H=14**, la **formulación directa** y la arquitectura **GRU**, y se detalla el protocolo de **aprendizaje por transferencia** (congelación, *fine-tuning* y control de transferencias negativas). Todo el proceso queda trazado mediante artefactos de reproducibilidad, garantizando comparabilidad entre

embalses y replicación de resultados.

A continuación se describen los experimentos en el **orden exacto** en que se ejecutan y con sus dependencias explícitas:

### 1. **Clustering y segmentación de embalses**

Particionado de los embalses ( $K=3,4,5$ ) tras filtrar atípicos, evaluando cohesión/separación (Silhouette, Calinski-Harabasz, Davies-Bouldin) para elegir  $K$  y proponer másters por clúster.

### 2. **Selección de horizonte de predicción**

Comparación de  $H \in \{1,3,7,14,30\}$  con validación temporal y líneas base (persistencia, climatología) para fijar el horizonte operativo que mejor equilibra precisión y utilidad.

### 3. **Entrenamiento base** (másters y *outliers*)

Modelos entrenados desde cero con el pipeline común (GRU,  $H=14$ ; features 32 vs 82) para establecer referencias sólidas por clúster y en embalses singulares.

### 4. **Transfer learning intra-clúster**

Transferencia máster  $\rightarrow$  *target* con *warm-up* y *fine-tuning* controlado, comprobando ahorro de coste y mantenimiento de calidad dentro de clústeres afines.

### 5. **Entrenamiento directo vs. residual**

Comparativa metodológica por embalse para valorar estabilidad, dependencia de la persistencia y realismo temporal, definiendo el criterio operativo de elección.

### 6. **Caso de estudio:** Jándula (scratch vs. *transfer learning*)

Comparación controlada (misma ventana, horizonte de predicción, pérdida y *scheduler*) para cuantificar el *trade-off* precisión-coste y las ganancias del *transfer learning* dentro del clúster.

### 7. **Experimentos fallidos y límites del enfoque**

Documentación de ensayos que no superan líneas base, justificación de descartes (p. ej., horizonte diario) y lecciones para el diseño final.

### 3.2.1. *Clustering* y segmentación de embalses

El propósito de esta fase es organizar los embalses andaluces en **grupos hidro-climáticamente homogéneos** para favorecer **transferencias de conocimiento** fiables entre sistemas con dinámica comparable. Esta segmentación cumple una doble función en el diseño experimental: (i) **orientar la elección de un “máster” por clúster** (embalse donante de pesos en los escenarios de *transfer learning*) y (ii) **aislar embalses atípicos** cuya conducta singular desaconseja la transferencia directa.

El punto de partida es un **conjunto de datos diarios de alta dimensión** ( $\approx 100$  variables) que integra indicadores **hidrológicos** (porcentaje y/o reserva), **meteorológicos** (precipitación, temperatura, humedad, radiación y viento) y **rasgos físicos** del sistema (capacidad, superficie y *proxies* morfométricos). Las series dinámicas se enriquecen con **retardos y acumulados de 3–60 días** para capturar memoria hidrológica; se imputan huecos breves cuando procede y se aplica **estandarización z-score**, de modo que las comparaciones entre embalses reflejen **patrones de respuesta** y no diferencias de escala.

El flujo metodológico se articula en dos etapas. En primer lugar, se realiza una **detección de atípicos mediante HDBSCAN** con el fin de identificar embalses cuya “firma” tempoclimática no se integra en ningún grupo compacto. Estos casos se etiquetan como *outliers* y se excluyen del agrupamiento principal para evitar que **distorsionen las fronteras** entre clústeres. En segundo lugar, sobre el conjunto depurado se aplica un esquema de **Agglomerative Clustering** con **enlace de Ward** y **distancia euclídea**, **explorando diferentes particiones** ( $K \in \{3, 4, 5\}$ ). La calidad de las particiones se contrasta con **métricas internas** y **proyecciones** en espacios reducidos (PCA/UMAP/t-SNE), lo que permite seleccionar el número de clústeres que mejor **equilibra cohesión y separación** sin entrar en **sobre-segmentaciones**.

#### **Justificación del rango** ( $K \in \{3, 4, 5\}$ )

La elección de este rango no es arbitraria y responde a:

1. **Objetivo operativo e interpretabilidad:** se busca discriminar **grandes regímenes hidro-climáticos** con lectura física clara y asignar **un máster por clúster**; con  $K > 5$  proliferan micro-clústeres de **baja utilidad operativa**.

2. **Tamaño mínimo útil por clúster:** el diseño experimental requiere, por clúster, al menos un **máster** y un **destino** para validar transferencias; valores mayores de  $K$  generan grupos con **cardinalidad insuficiente**, lo que **compromete** la selección de máster y la evaluación de *transfer learning*.
3. **Parcimonia en alta dimensión:** con  $\sim 100$  rasgos, incrementar  $K$  tiende a **degradar la separación relativa** (concentración de distancias) y a **inestabilizar** las asignaciones; se aplica el principio de **parcimonia**: usar el **menor**  $K$  que aporte **mejoras sustantivas** en cohesión/separación.
4. **Evidencia empírica:** las **métricas internas** y las **proyecciones** (véase el apartado de resultados) muestran un **máximo claro en**  $K = 4$  y **deterioro** al pasar a  $K = 5$ ; por encima de 5 **no se observa ganancia** en separación y sí **sobre-segmentación**, por lo que **no se amplió** el barrido de  $K$ .
5. **Coherencia con el plan experimental:** acotar  $K$  a 3–5 **alineó** el número de grupos con el **catálogo de másters/targets** y con el **presupuesto computacional**, manteniendo comparabilidad entre configuraciones sin inflar innecesariamente el número de modelos a entrenar.

La **selección del máster** de cada clúster se resuelve con un procedimiento reproducible que combina tres dimensiones: **desempeño, estabilidad y representatividad**. En desempeño se consideran **indicadores operativos** calculados de forma homogénea; en estabilidad se compara el **comportamiento reciente** frente al histórico para descartar series con **cambios de régimen**; y en representatividad se priorizan los embalses más próximos al “**centro**” del **clúster** según correlaciones y distancias en el espacio de rasgos. La decisión final **favorece** al candidato que cumple los **umbrales definidos** y **maximiza** la combinación de representatividad y *skill* frente a referencias internas. Todo el proceso deja **trazabilidad** en ficheros tabulares (con el detalle de candidatos, criterios y elección) a fin de garantizar **reproducibilidad y auditoría**.

Con este enfoque, el *clustering* proporciona una **cartografía funcional** de los embalses andaluces que sustenta tanto el **diseño de los experimentos de *transfer learning*** como la **interpretación posterior** de resultados, manteniendo **separadas** las peculiaridades locales (*outliers*) y **consolidando** grupos donde la **transferencia de pesos** es más sólida.

### 3.2.2. Entrenamiento de modelos predictivos

#### Elección del horizonte temporal

La selección del horizonte se realizó de forma empírica sobre el embalse de Cala, evaluando  $H \in \{1, 3, 7, 14, 30\}$  días con ventana fija ( $steps=120$ ), formulación residual (la red aprende la corrección sobre la persistencia) y un conjunto secuencial de 82 variables (salvo en  $H=14$ , donde se contrastó además un conjunto enriquecido ( $\approx 290$  features)). Para cada  $H$  se inspeccionaron:

- un *hold-out* anual con validación *rolling* (series real vs. predicha),
- curvas de entrenamiento/validación con pérdida Huber y
- comparativas “real vs. predicción” en una ventana larga ( $\approx 15$  años).

El patrón observado fue el esperable: el desempeño se degrada conforme aumenta  $H$ , pero  **$H=14$**  mantuvo un rendimiento estable y un error medio bajo, ofreciendo a la vez **utilidad operativa**: dos semanas permiten programar explotación, bombeos y comunicación con antelación suficiente. Además, el uso del set enriquecido ( $\approx 290$  variables) reforzó la memoria hidrológica y los forzamientos climáticos necesarios para sostener la predicción quincenal. Por ello,  **$H=14$**  se adopta como **estándar** para el entrenamiento de másters y outliers y como base para las transferencias intra-clúster.

#### Entrenamiento de modelos predictivos

- **Alcance y diseño**

Por criterios de eficiencia y no redundancia, no se entrenaron modelos para los 73 embalses de forma individual. Se optó por un diseño escalable con 12 modelos:

- **4 outliers** (embalses singulares sin clúster, entrenados desde cero y resultado de la fase de detección de *outliers*; , Cala, Canales, Iznájar, San Rafael de Navallana),
- **4 másters** (uno por clúster, con histórico largo y estable, entrenados desde cero y resultado escogido tras los experimentos de *clustering* y elección de máster de clúster),

- **4 targets** (uno por clúster) que reciben *transfer learning* desde su máster para validar la transferencia dentro del clúster sin reentrenar todos los casos.

Este esquema equilibra representatividad, coste computacional y rigor comparativo.

- **Variables de entrada**

Se trabajó con dos familias de *features*:

- **Básicas ( $\approx 32$ ):** calendario con codificación estacional, precipitación, radiación, temperatura media; retardos, acumulados y variaciones del nivel/porcentaje.
- **Extendidas ( $\approx 82$ ):** añaden temperaturas máximas/mínimas, humedades (media/máx./mín.), viento (velocidad media/máx., dirección), radiación desagregada, más retardos/ventanas móviles; incluyen reserva ( $\text{hm}^3$ ) y porcentaje con sus *lags*. El objetivo es capturar con mayor fidelidad los forzamientos y la memoria del sistema.

- **Formulación y arquitectura**

La tarea consiste en **predecir 14 días de % de llenado** a partir de ventanas históricas de  $\text{steps} \in \{120, 180\}$ . Se adoptó el **modo directo** (la red mapea la ventana al vector  $t+1\dots t+14$ ) para evitar acumulación de error de enfoques recursivos y se comparó con el **modo residual** (corrección sobre la persistencia). Los modelos son **GRU** (y, en contraste, LSTM), con profundidad “*medium*” para  $\approx 82$  *features* y “**base**” para  $\approx 32$ , *dropout* y *weight decay*, **pérdida Huber**, *early stopping* y programadores de *learning rate*. Todas las entradas **se estandarizan (*z-score*)** con estadísticas de entrenamiento.

- **Evolución y líneas base**

Se aplicó validación temporal (***hold-out* anual** con *rolling* reciente) y se comparó frente a **climatología** y **persistencia** (muy competitiva a  $H=14$  por la inercia de los embalses). Además de las métricas cuantitativas, se revisaron curvas de pérdida y gráficos real-predicho para detectar sobreajuste y evaluar la capacidad de seguir picos, transiciones y retardos.

- ***Transfer learning***

El **transfer learning** traslada pesos y normalizadores desde un dominio fuente (máster) a un dominio destino (*target*) dentro del mismo **clúster hidro-climático**, donde la afinidad favorece la reutilización de patrones (inercia, estacionalidad, respuesta a forzamientos). El **pipeline** aplicado fue:

1. Entrenar el **máster** (misma ventana, H=14 y conjunto de *features*).
2. Exportar **pesos y normalizadores**.
3. **Calentamiento** (*warm-up*) en el destino con el bloque recurrente **congelado** y la cabeza **reinicializada**.
4. **Ajuste fino** con **descongelado gradual** y *learning rates* diferenciados.
5. Selección por **mínima pérdida de validación**.
6. Verificación frente a **persistencia y climatología**, registrando tiempos.

Se controlaron riesgos de **transferencia negativa**, *covariate shift* y *concept drift* con una elección cuidada del donante, pilotos cortos y regularización. Las parejas **máster-destino** fueron: **Andévalo-Guadalquivir, Guadalquivir-Jándula, Guadalquivir-Guadalquivir, Barbate-Casasola**. Los embalses sin donante adecuado (Cala, Canales, Iznájar y San Rafael de Navarra) **se entrenaron desde cero**.

**Residual vs. directo.** El modo **residual** puede ser muy competitivo en sistemas altamente inerciales, pero su dependencia de la persistencia lo hace más frágil en validaciones largas o en transferencias (cambian los residuales). El **directo** mostró mayor estabilidad temporal y se adopta como **referencia para explotación** y *transfer learning*, reservando el residual para casos muy inerciales y homogéneos.

- **Comparativa en Jándula: entrenamiento completo vs. transfer learning**

### **Comparativa en Jándula: entrenamiento completo vs. transfer learning**

**Alcance y justificación del ejemplo:** Antes de presentar el conjunto completo de resultados por embalse, se muestra **Jándula** como **estudio de caso representativo** por tres motivos:

- dispone de una *pareja máster*→*destino* bien definida (Guadalmena→Jándula), lo que permite evaluar con claridad la transferencia intra-clúster;
- cuenta con **series largas y estables** que hacen comparables las alternativas *scratch* vs. *transfer* bajo el **mismo protocolo** (H=14, steps=180, modo directo);
- sus resultados son **informativos del compromiso precisión–coste**, ni triviales (casos donde todo funciona) ni patológicos (casos incompatibles).

El resto de embalses se reporta a continuación en sus apartados específicos y en los cuadros comparativos correspondientes.

En Jándula se comparó un modelo independiente (desde cero) y un modelo con *transfer learning* desde Guadalmena, ambos con H=14, steps=180, modo directo y el mismo protocolo de entrenamiento.

- **Independiente (82 features):** alcanzó la máxima precisión y gran solidez de generalización, a costa de mayor tiempo de entrenamiento; recomendable cuando se prioriza exactitud y hay presupuesto computacional.
- **Transfer learning extendido (82 features):** calidad muy próxima al anterior con tiempos sensiblemente menores; opción preferente cuando se busca equilibrio precisión–coste y existe afinidad físico-climática.
- **Transfer learning básico (32 features):** evidenció que un set reducido puede ser insuficiente para la dinámica de Jándula; la transferencia exige señal exógena rica (precipitación, radiación, humedad, viento, acumulados y retardos) para adaptarse bien al destino.

*Implicación operativa.* Con másters bien entrenados y paridad de *features*, el *transfer learning* permite escalar a nuevos embalses con **coste reducido** y **calidad cercana** al entrenamiento independiente. En Jándula: *precisión máxima* con **independiente (82)**; *mejor balance precisión–coste* con **transfer learning extendido desde Guadalmena**.

### 3.3. Resultados

#### 3.3.1. Selección de características

- **Correlaciones entre variables meteorológicas**

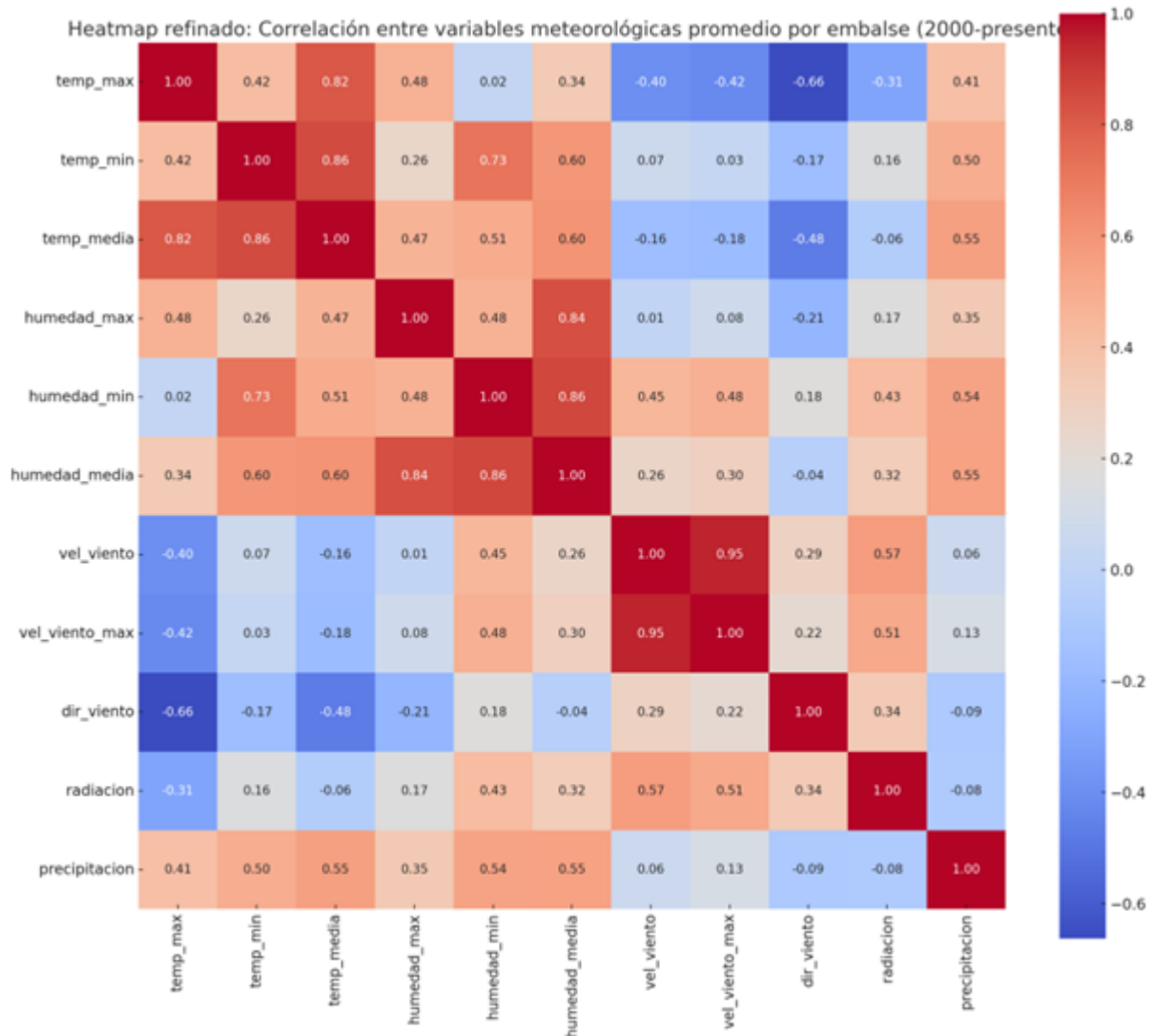


Figura 5: Heatmap agregado por embalse

El heatmap agregado por embalse (Figura 5) muestra correlaciones altas dentro de cada familia:

- **Temperaturas** (máx, media, mín) con  $r$  en torno a **0.82–0.92**.
- **Humedades** (media, máx, mín) con  $r$  en torno a **0.84–0.86**.

- **Viento: velocidad media y velocidad máxima con  $r \approx 0.95$ .**

Aparecen asociaciones **negativas** entre **temperatura** y **dirección del viento** ( $r$  alrededor de  $-0.66$ ) y relaciones **moderadas** entre **radiación** y **velocidad del viento** ( $r \approx 0.57$ ). La **precipitación** presenta correlaciones **bajas** a moderadas con otras magnitudes ( $|r| \leq 0.55$ ).

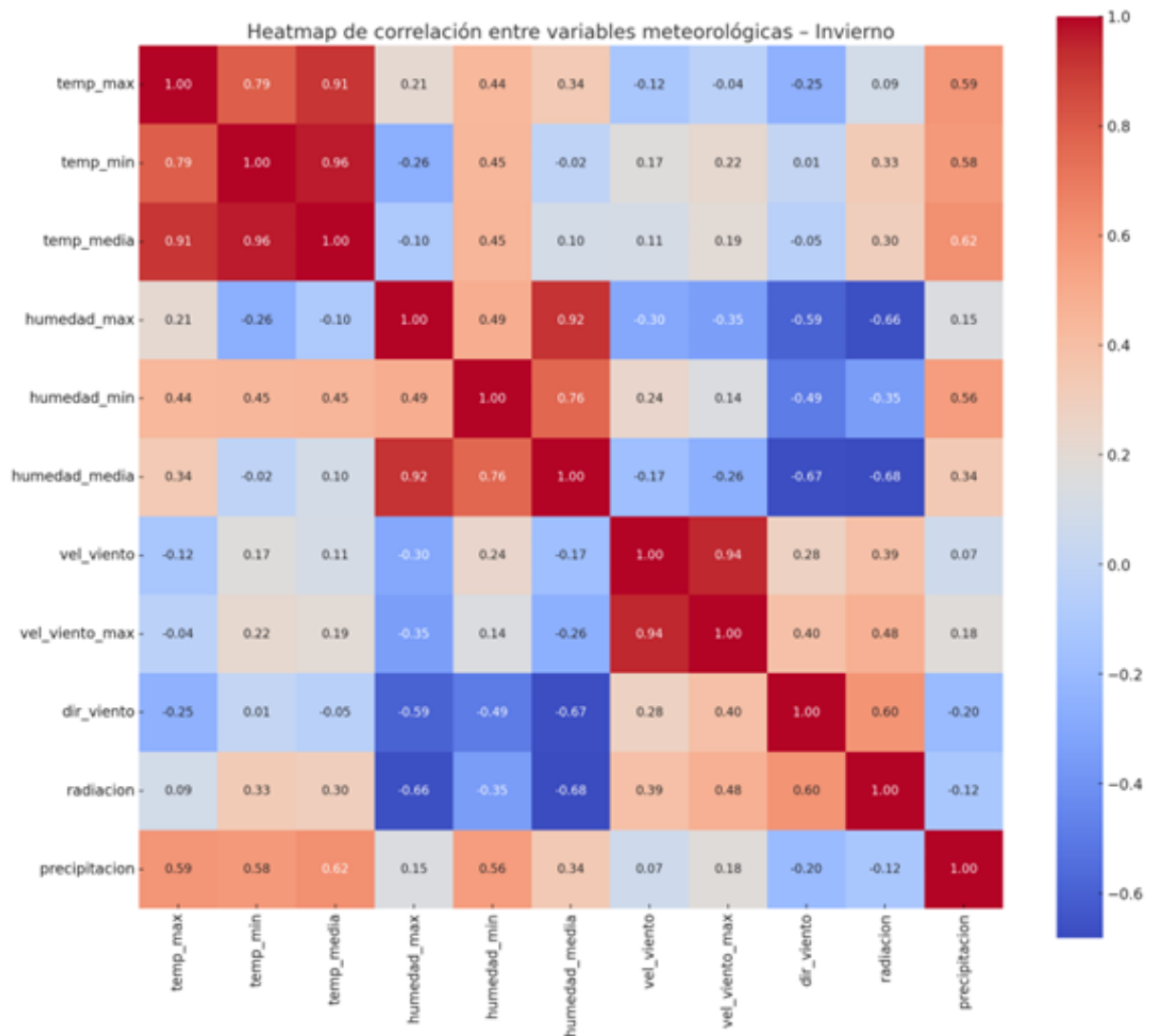


Figura 6: *Heatmap* estacional - Invierno

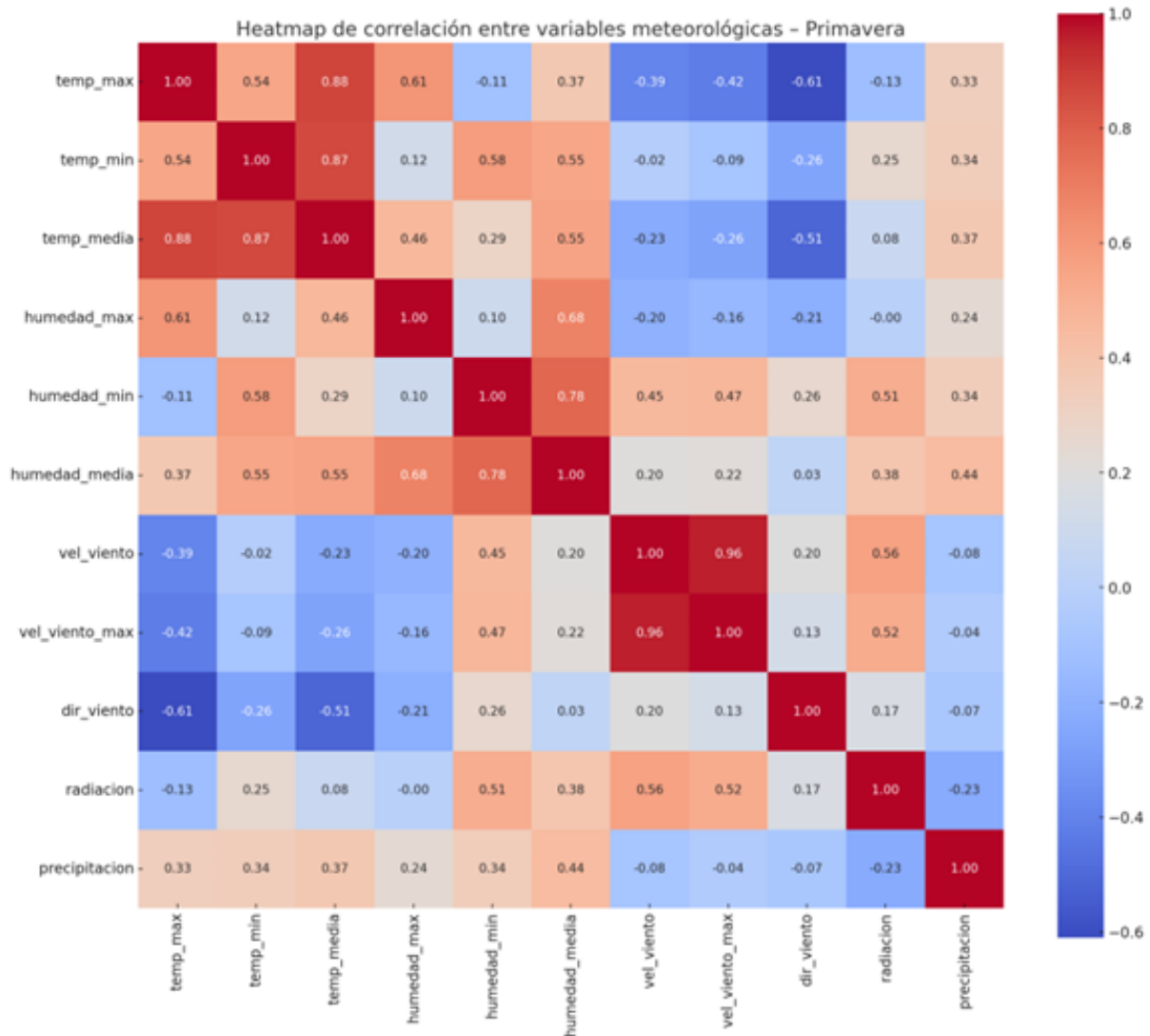


Figura 7: *Heatmap* estacional - Primavera

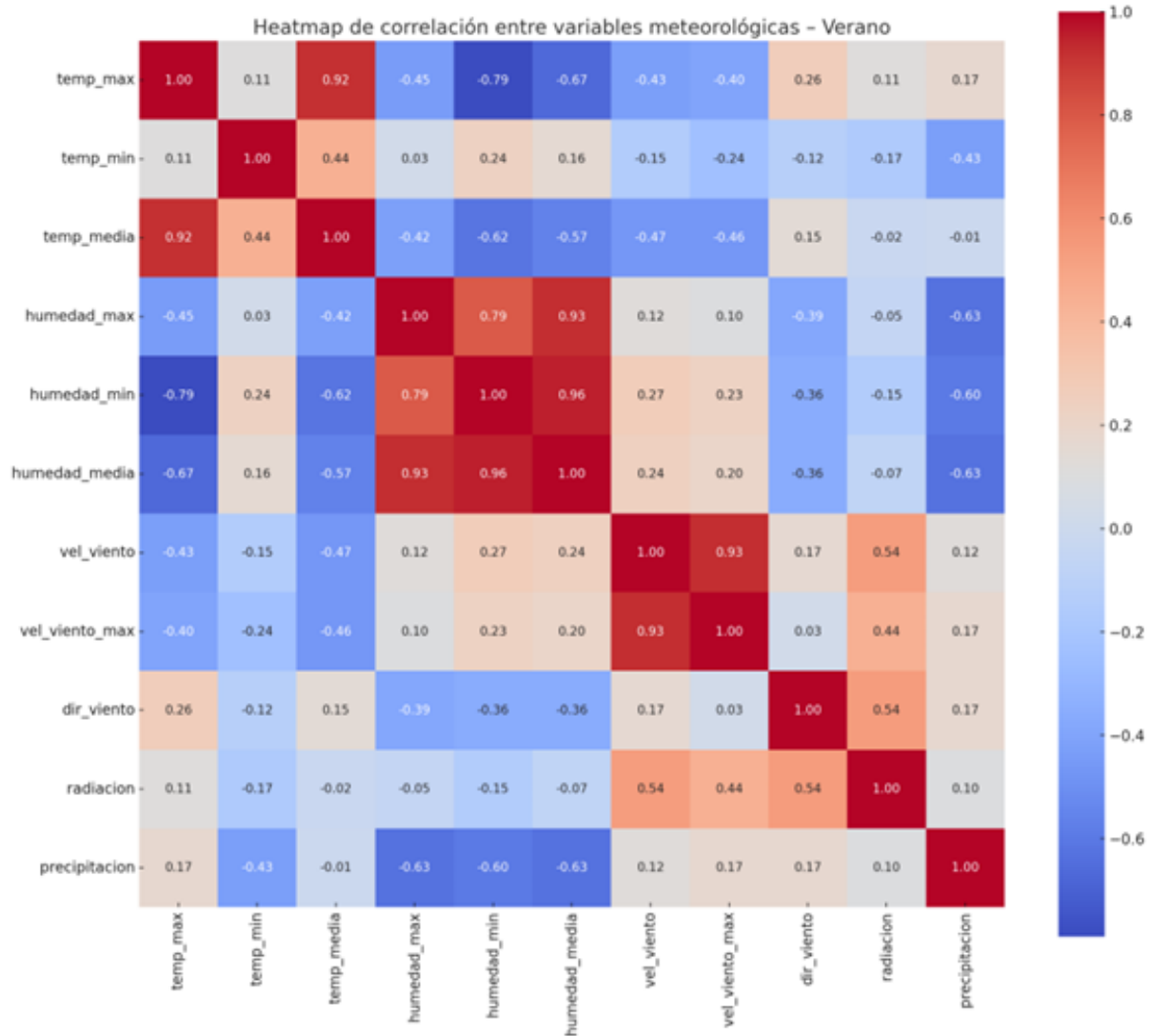


Figura 8: *Heatmap* estacional - Verano

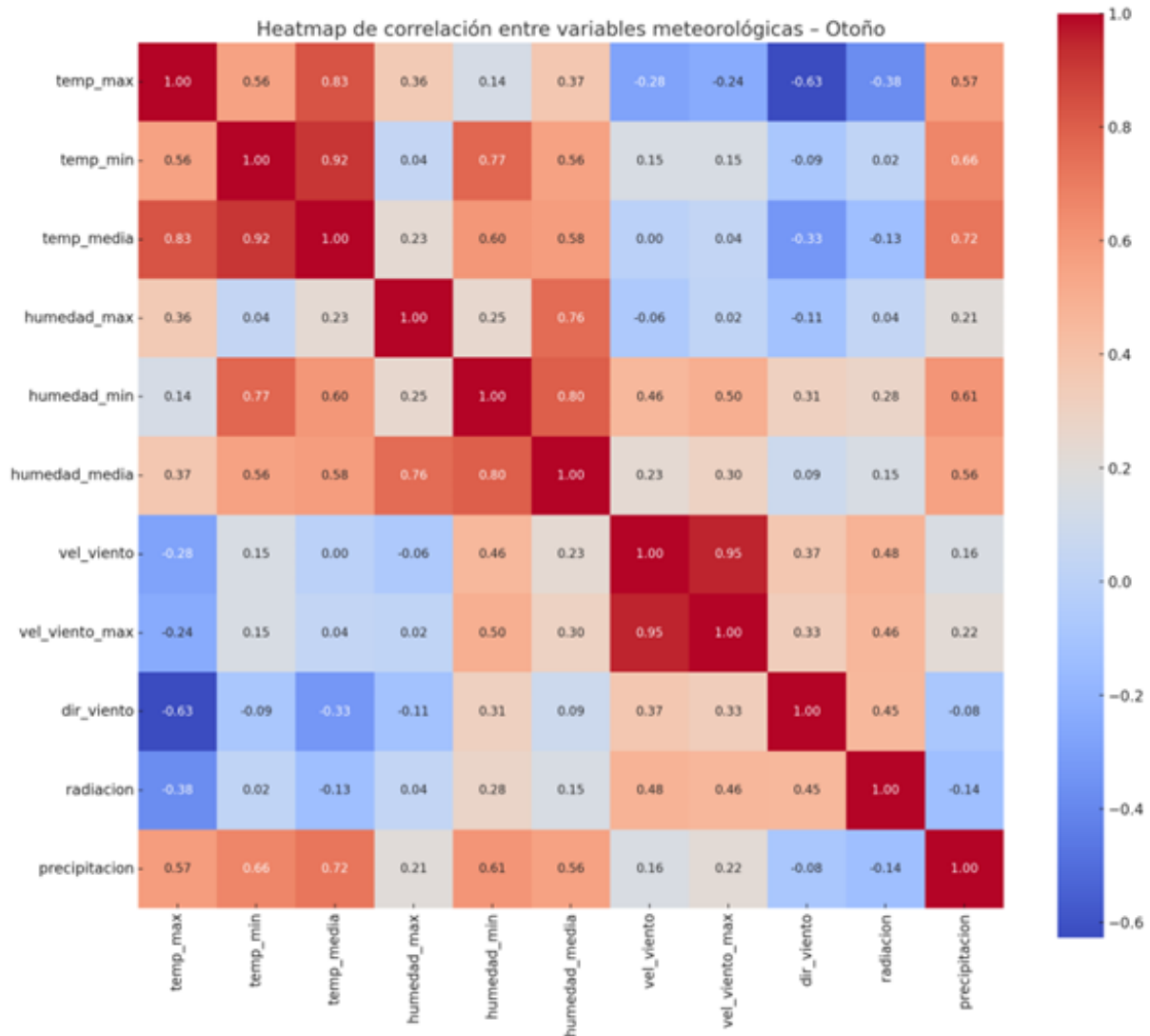


Figura 9: *Heatmap* estacional - Otoño

Los **heatmaps** estacionales (Figs.6,7,8,9) mantienen ese patrón intra-clúster y muestran variaciones de intensidad por estación: las tres variables de **temperatura** continúan altamente correlacionadas entre sí, **humedad** conserva relación positiva dentro de su familia y **velocidad** y **velocidad máxima de viento** siguen estrechamente asociadas; la **radiación** se correlaciona positivamente con las temperaturas y negativamente con las humedades en todas las estaciones, con magnitudes que fluctúan según el periodo.

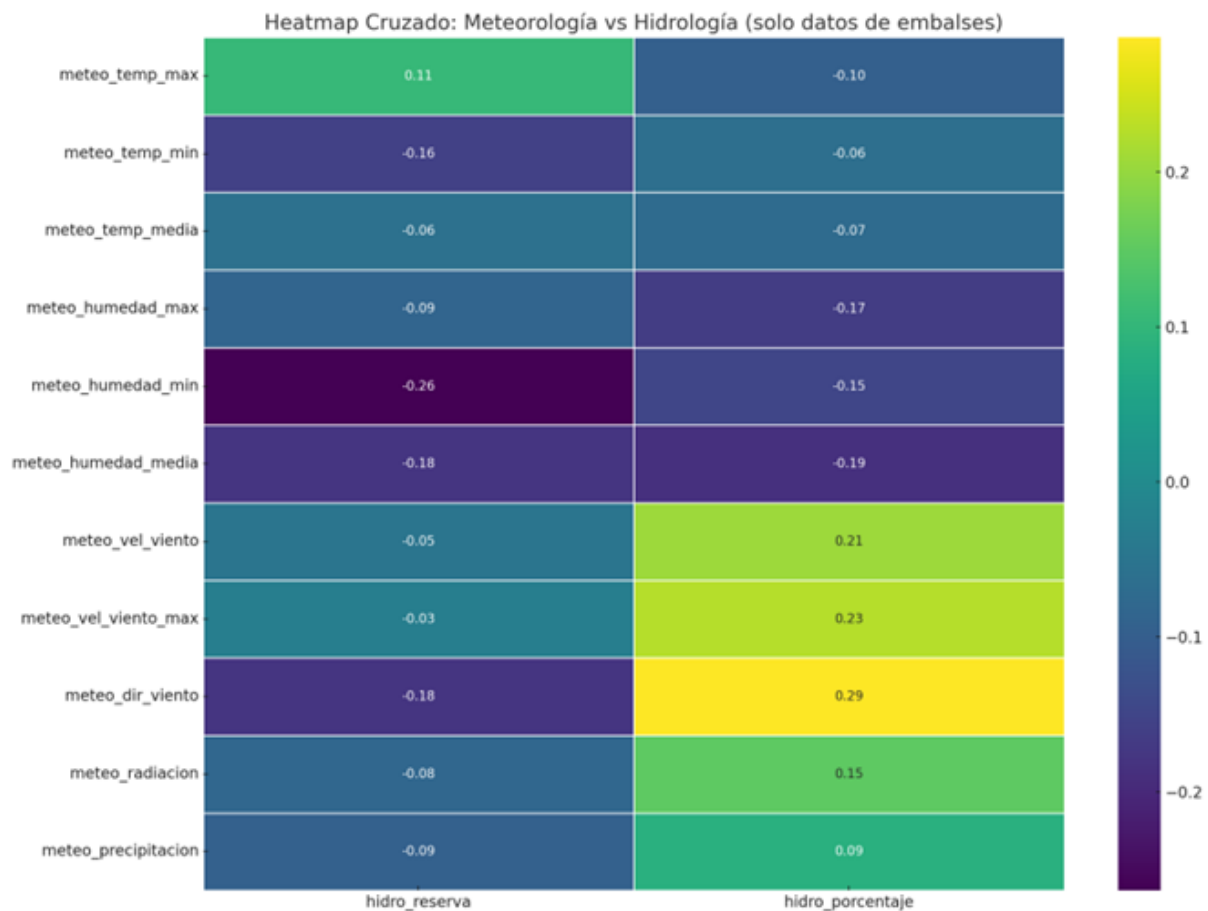


Figura 10: *Heatmap* cruzado - meteorología vs. hidrología

En el *heatmap* cruzado (Figura 10), las correlaciones a **desfase cero** se sitúan en **rangos bajos**. Para reserva, los coeficientes oscilan aproximadamente entre **-0.26 y +0.11**; para porcentaje, entre **-0.19 y +0.29**. Se observan valores positivos moderados con algunas variables de **viento** y **radiación** respecto al **porcentaje**, y coeficientes cercanos a cero o ligeramente negativos con la mayoría de las magnitudes respecto a **reserva**.

- **Heatmap** correlación por provincia

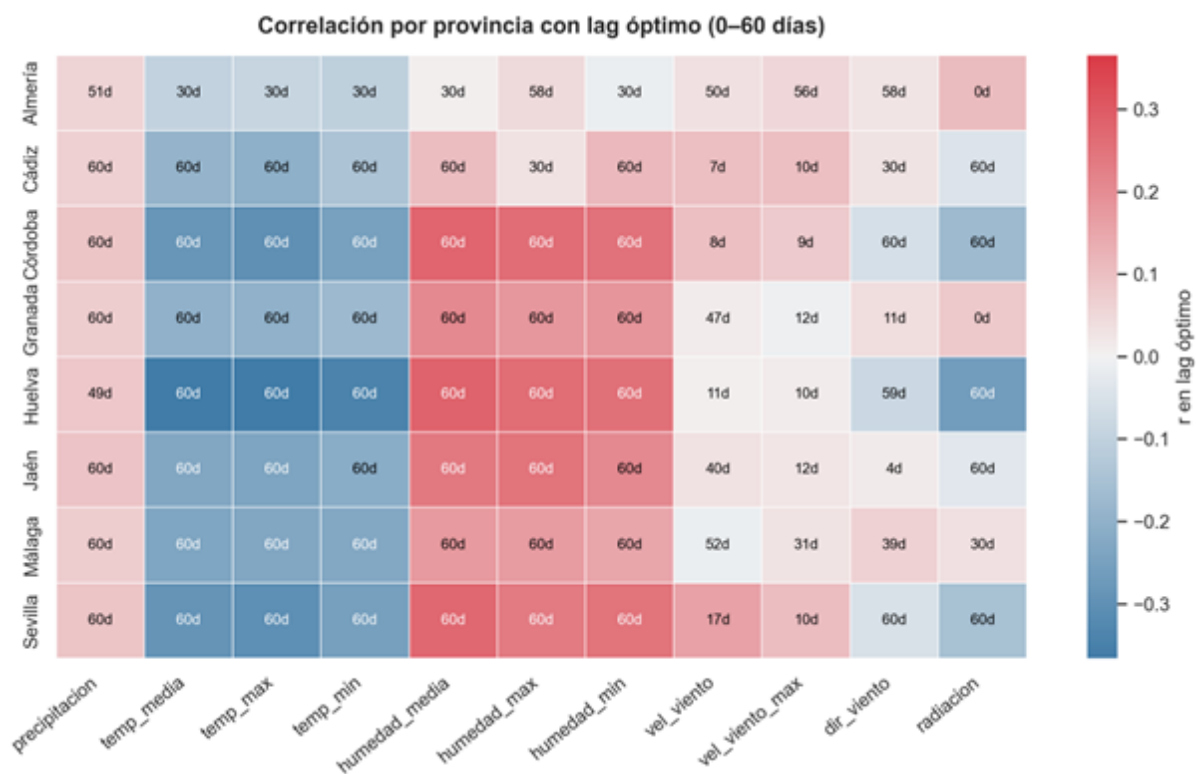


Figura 11: *Heatmap* correlación por provincias con *lag* óptimo

El mapa de **lag óptimo** (Figura 11) resume, para cada **provincia** y **variable**, el desfase (0–60 días) en el que se obtuvo el **máximo**  $|r|$ . En general:

- Las **temperaturas** y las **humedades** alcanzan su máximo en **desfases largos**, con frecuencia **47–60 días**.
- La **precipitación** registra *lags* óptimos entre **49 y 60 días** en varias provincias.
- Las variables de **viento** muestran **lags más variados**, con valores óptimos que van de **7 a 59 días**, según provincia y métrica.

### 3.3.2. Resultados clustering

El **análisis de agrupación** identificó **cuatro outliers** (**San Rafael de Navallana, Cala, Canales e Iznájar**), cuyos patrones hidro-meteorológicos y/o geométricos no presentan vecindad compacta con el resto. Estos casos se excluyeron del particionado principal y se reportan como “no asignables” a clúster por comportamiento singular.

■ Experimento 1 (K=3)

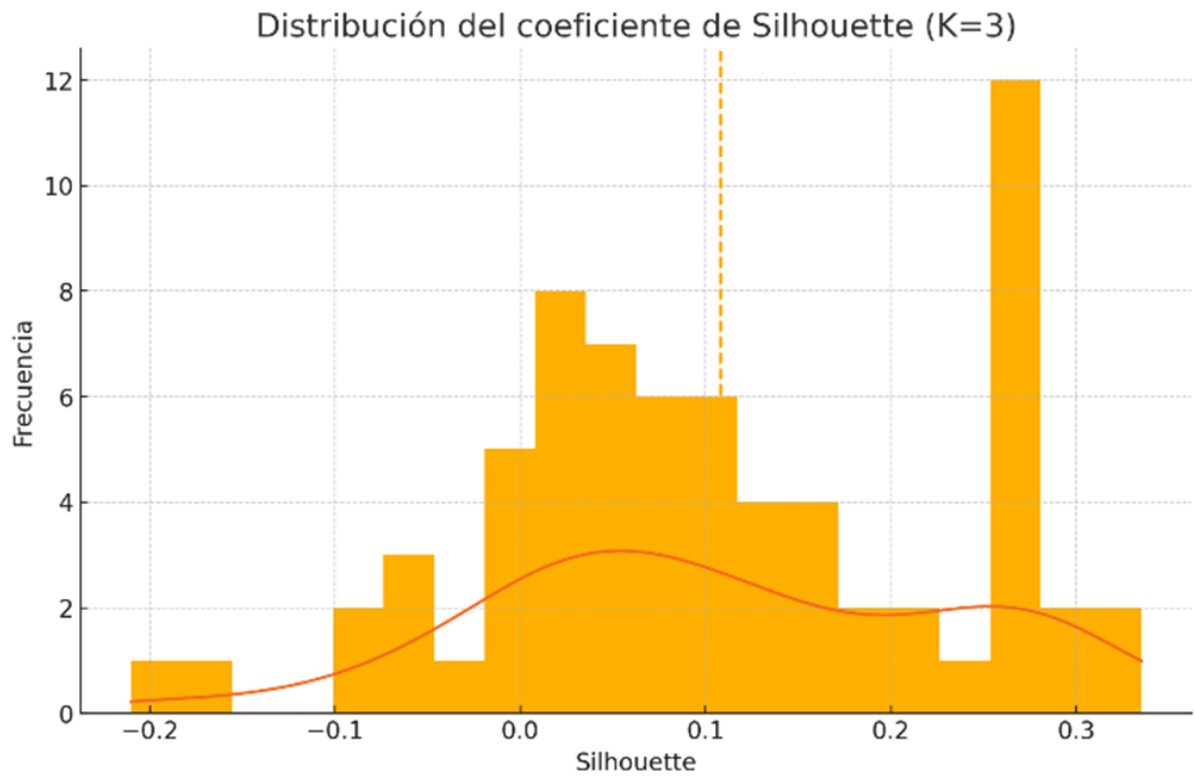


Figura 12: Distribución Coeficiente de Silhouette para K = 3

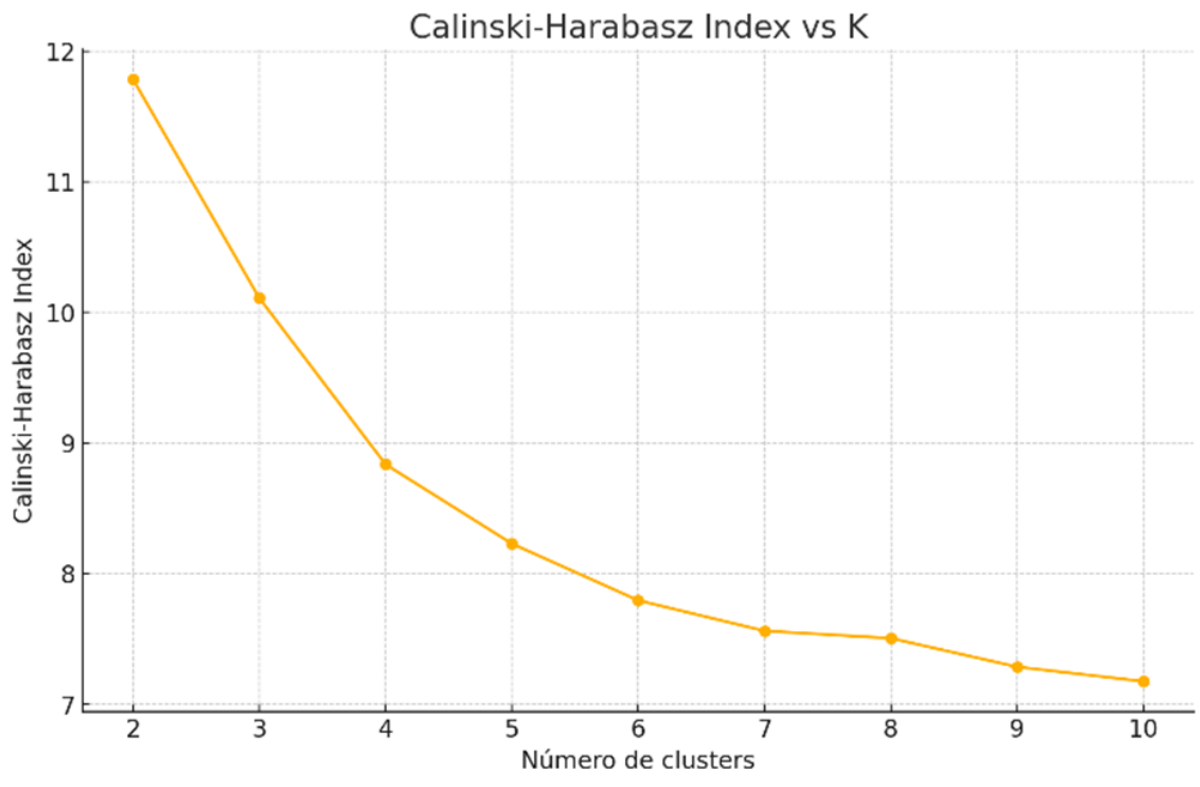


Figura 13: Índices Calinski-Harabasz

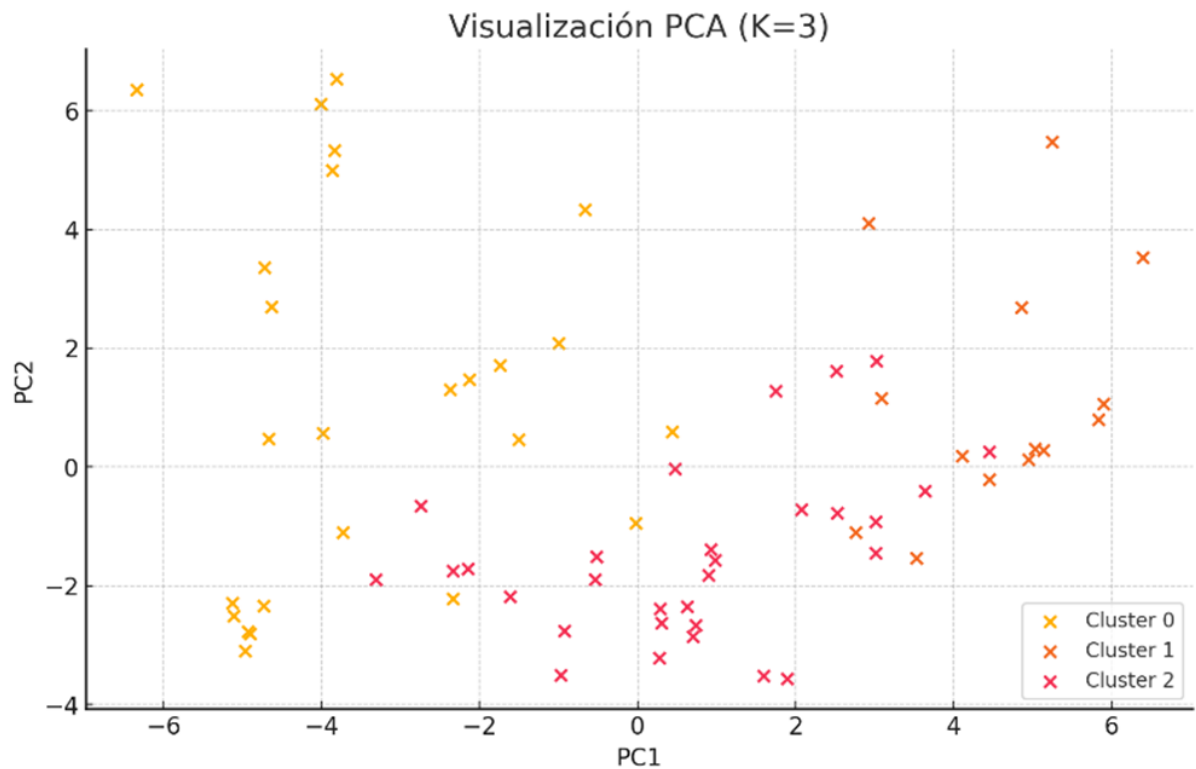


Figura 14: Visualización PCA para K = 3

Con  $K=3$ , las métricas internas fueron **Silhouette = 0.108**, **Calinski–Harabasz (CH) = 10.11** e **índice Davies–Bouldin (DB) = 2.213** (esta última, mostrada en Figura.13). El **histograma de Silhouette** (Figura 12) se concentró en torno a 0, y las **proyecciones PCA**, representadas en Figura 14, mostraron **superposición entre grupos**, reflejando una separación limitada.

- **Experimento 2 (K = 4)**

En la segunda configuración, con cuatro clústeres, los resultados mejoraron sustancialmente. El **índice de Silhouette subió a 0.253**, el **CH alcanzó 29.3** y el **DB descendió a 1.171**. Estos indicadores apuntan a un equilibrio adecuado entre cohesión interna y separación externa. El histograma de Silhouette mostró una media próxima a 0.24, con una proporción mayor de embalses bien integrados en sus grupos y menos casos en valores negativos. Además, las proyecciones PCA y t-SNE reflejaron clústeres más definidos, aunque con cierta proximidad entre las fronteras de algunos embalses.

En cuanto a la composición, el **clúster 0** incluyó **34 embalses** de gran capacidad y comportamiento relativamente estable; el **clúster 1**, con **17 embalses**, agrupó cuencas de menor tamaño y elevada variabilidad climática; el **clúster 2**, con **14 embalses**, correspondió a zonas áridas o semiáridas, caracterizadas por temperaturas y radiación altas, baja precipitación y niveles medios bajos; finalmente, el **clúster 3**, con **solo 4 embalses**, destacó por un elevado porcentaje medio de llenado y una dinámica asociada a precipitaciones acumuladas.

Respecto a la selección de másters, la propuesta inicial fue **Guadalupe (Clúster 0)**, **Andévalo (Clúster 1)**, **Guadalupe (Cluster 2)** y **Breña (Clúster 3)**. Sin embargo, la evaluación de estabilidad mostró que **Breña no cumplía los criterios**, por lo que se sustituyó por **Barbate como máster del clúster 3**. El conjunto final de másters quedó formado por **Guadalupe, Andévalo, Guadalupe y Barbate**.

- **Silhouette Score (0.253)**: Indica una separación moderada entre clusters, coherente con la alta dimensionalidad del conjunto de datos (103 variables hidrológicas, meteorológicas y categóricas). En *clustering* geográfico o ambiental, valores entre 0.2 y 0.4 pueden ser aceptables si los datos son ruidosos o si hay muchas variables correlacionadas.

El siguiente gráfico (Figura 15) muestra un histograma correspondiente a la **distribución de los coeficientes de Silhouette**.

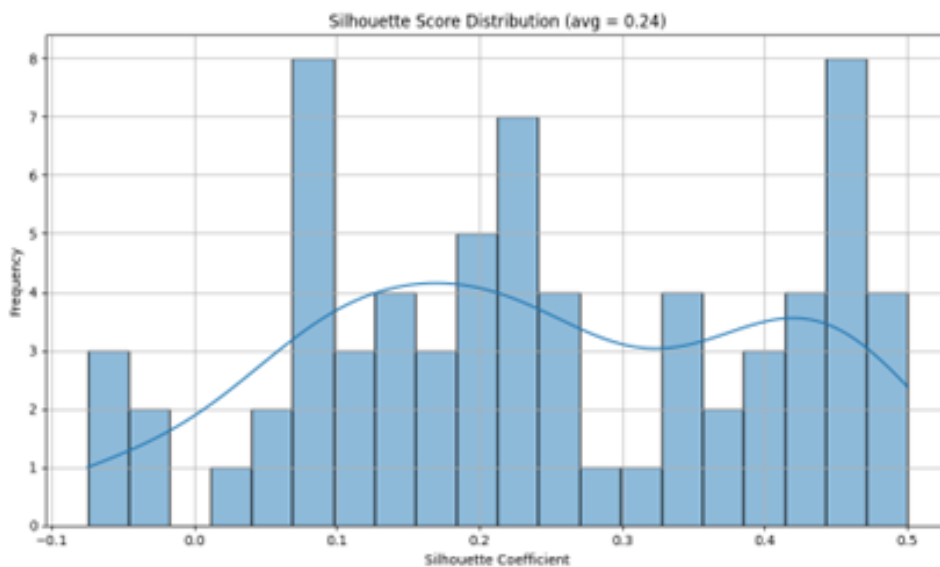


Figura 15: Distribución Coeficiente de Silhouette para K = 4

- **Bins (intervalos en el eje X).** El rango total de Silhouette (-0.1 a 0.5) se divide en 19 subintervalos de igual amplitud. Cada barra cubre un trozo de ese rango.
- **Frecuencia (eje Y).** La altura de cada barra muestra el número de embalses cuyos coeficientes caen dentro de ese *bin*. Sumando todas esas barras recuperamos los 69 embalses analizados.
- **Curva de densidad (KDE).** La línea suave aproxima la distribución continua de los valores, ayudando a identificar superpoblaciones (picos de concentración)
- **Silhouette medio.** La media aritmética de todos los  $s(i)$  es aproximadamente 0.24. Un valor medio entre 0.2 y 0.4 se considera aceptable para datos de alta dimensionalidad y ruido natural, como ocurre con series hidrológicas con más de 100 variables.
  - Hay un núcleo central que agrupa la mayoría de los embalses (aproximadamente el 40 %) en valores de silhouette entre 0.1 y 0.3. estop refleja una separación moderada: los clústers son internamente compactos pero existe cierta proximidad entre grupos vecinos.
  - Un subconjunto destacado (alrededor del 15 %) alcanza valores elevados (0.4-0.5), indicando embalses cuya dinámica temporal encaja de forma muy clara en su cluster asignado.
  - La asimetría de distribución (cola más larga hacia la derecha) sugiere que, aunque muchos embalses queden en la zona media, existe un grupo de “ejemplares” muy bien diferenciados.
- **Índice Calinski-Harabasz (29.3):** Tal y como se muestra en Figura 16, se observa una dispersión inter-clúster razonable frente a la intra-clúster. Esto indica que los clústers están bien separados y son internamente compactos

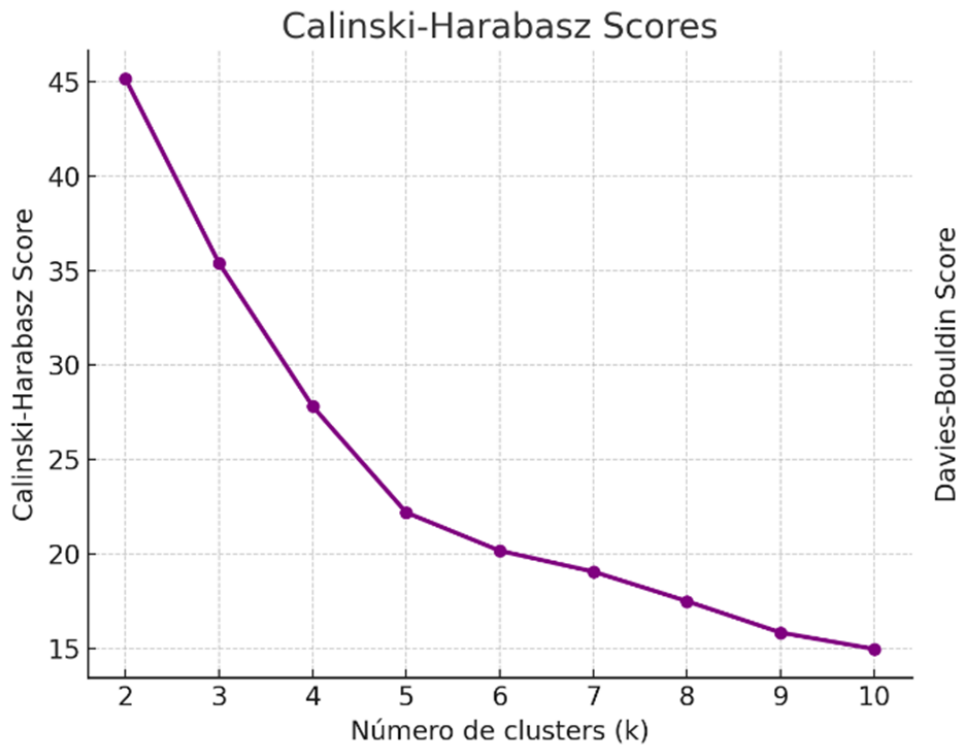


Figura 16: Índices Calinski-Harabasz para K = 4

- **Índice Davies-Bouldin (1.171):** Confirma una similitud moderada entre grupos, aceptable en contextos sin etiqueta previas (*ground thuth*) Representado en Figura 17.

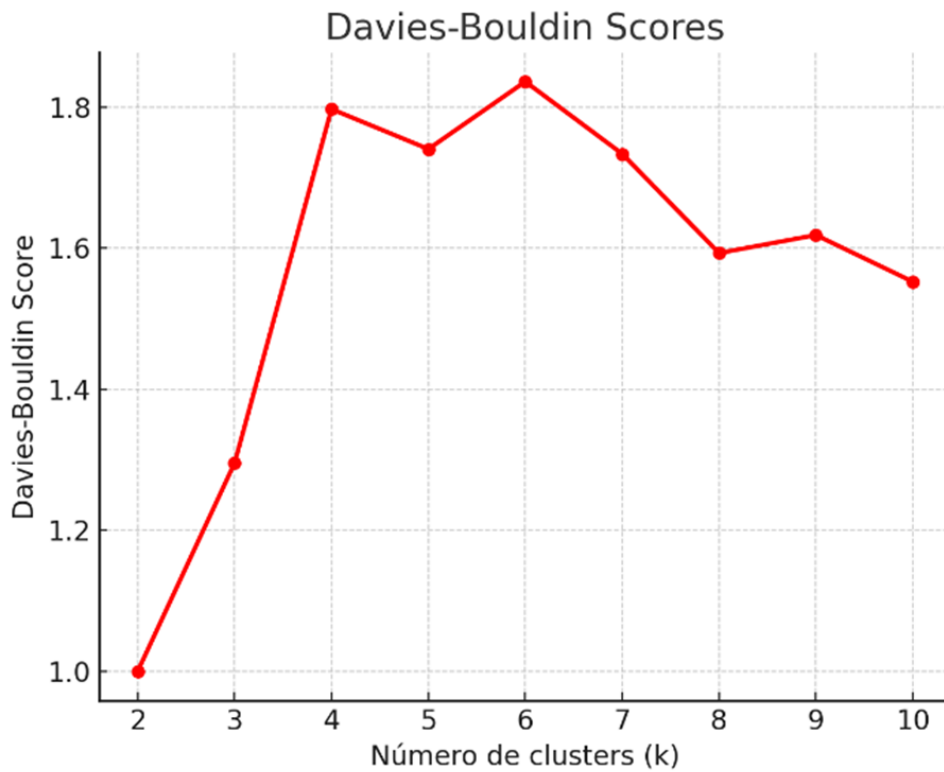


Figura 17: Índices Davies-Bouldin para K = 4

Las técnicas de reducción de dimensionalidad **PCA** y **t-SNE** (Figura 19 y Figura 18, respectivamente) revelaron solapamientos parciales, pero destacaron la coherencia espacial de los clusters y la clara diferenciación de los *outliers*.

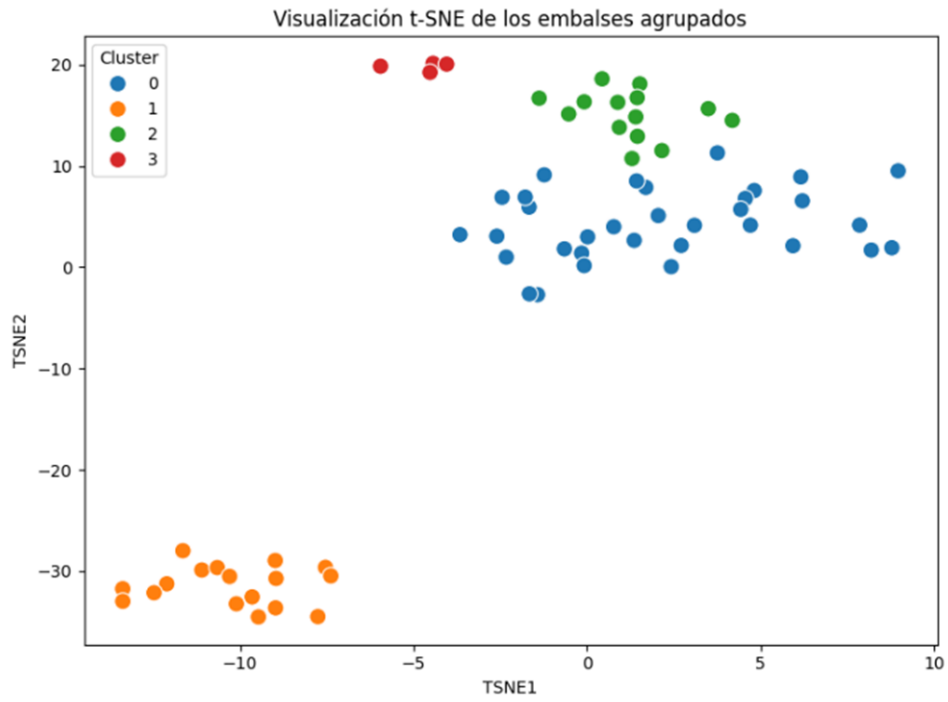


Figura 18: Visualización t-SNE para  $K = 4$

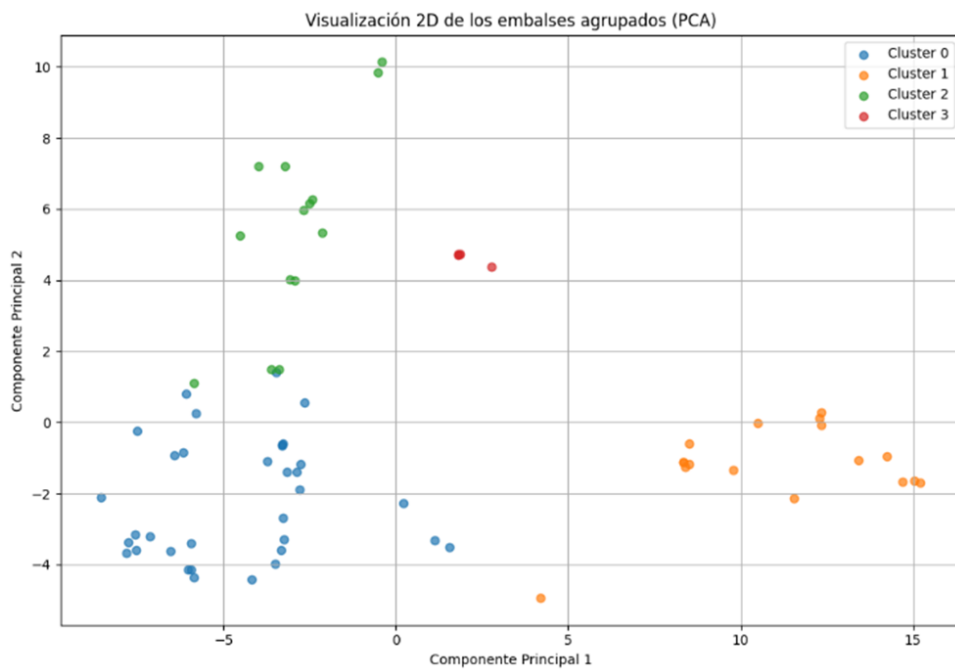


Figura 19: Visualización PCA para  $K = 4$

■ Experimento 3 ( $K = 5$ )

El experimento con cinco clústeres no ofreció resultados favorables. **El índice de Silhouette descendió a 0.133, el CH bajó a 8.26 y el DB aumentó a 1.826** (ambos mostrados en 21), lo que indica pérdida de cohesión y mayor solapamiento entre grupos. El histograma de Silhouette (Figura 20) mostró que la mayoría de embalses se situaban en torno a valores próximos a cero, y las proyecciones PCA, representadas en Figura 22 revelaron clústeres pequeños y poco diferenciados. En la práctica, la partición generó micro-clústeres con escasa representatividad, lo que dificultaba tanto la interpretación como la transferencia de modelos.

En todas las configuraciones se mantuvo estable la lista de **outliers**, formada por los embalses **San Rafael de Navallana, Cala, Canales e Iznájar**, cuyo comportamiento singular justificó su exclusión del proceso de agrupamiento.

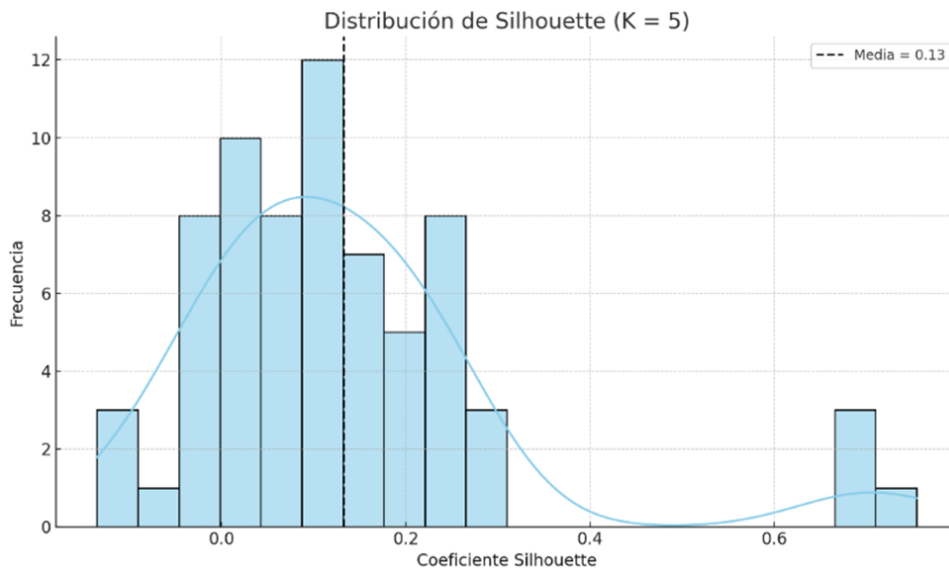


Figura 20: Distribución Coeficiente de Silhouette para K = 5

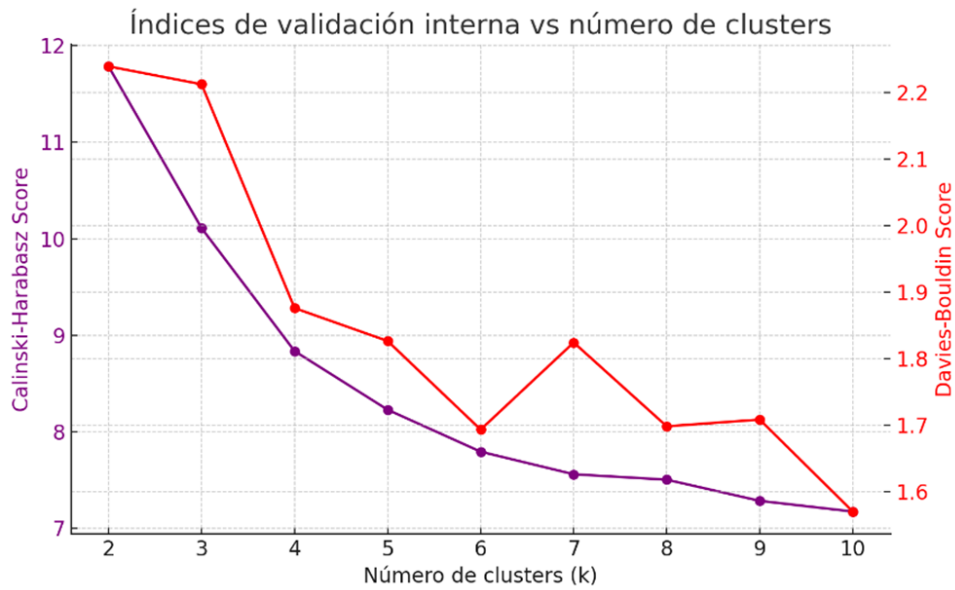


Figura 21: Índices de validación interna para K = 5

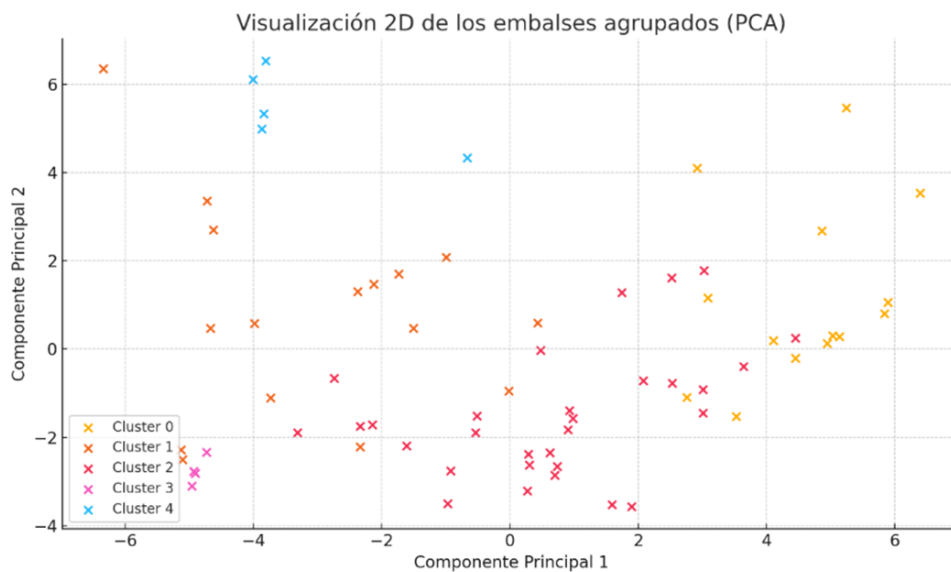


Figura 22: Visualización PCA para K = 5

En resumen, las tres configuraciones evaluadas mostraron diferencias notables en las métricas internas. Mientras que con **K=3 los grupos eran poco compactos y muy solapados**, y con **K=5 aparecía sobresegmentación en microclústeres poco representativos**, la opción de **K=4 se consolidó como la más equilibrada**.

En este trabajo **recalculamos Calinski–Harabasz y Davies–Bouldin en cada experimento**: aunque las gráficas muestren los índices para cada valor de K, **cada cambio de K (y,**

**por tanto, de partición)** exige volver a computarlos bajo el mismo *dataframe* y preprocesado para que la comparación sea válida.

Para mayor claridad, en la tabla 3 se recogen de manera comparativa los principales indicadores de calidad obtenidos en cada experimento.

Tabla 1: Resultados de indicadores de calidad del agrupamiento por experimento

Configuración	Silhouette	Calinski–Harabasz	Davies–Bouldin
K = 3	0,108	10,11	2,213
K = 4	0,253	29,30	1,171
K = 5	0,133	8,26	1,826

#### Observaciones principales.

- **K = 3:** grupos poco definidos, alta superposición.
- **K = 4:** mejor equilibrio; clústeres compactos y separados, histograma de silhouette con media  $\approx 0,24$ .
- **K = 5:** sobre-segmentación: micro-clústeres y mayor solapamiento.

#### 3.3.3. Resultados entrenamiento de modelos predictivos

##### Elección del horizonte temporal

La comparación de horizontes en **Cala** (*steps* = 120, formulación **residual** y 82 features; en **H=14** se probó además un set **enriquecido**) se sustentó en:

- **hold-out anual** con validación *rolling*, representadas en las figs.24,27,30,33,36
- **curvas de pérdida** (Huber), mostradas en las figs.25,28,31,34 y 37) y
- **real vs. predicción en 15 años** (ver figs. 23,26,29,32 y 35), complementado con **radar por horizonte** ( Figura 40), **MAE/RMSE** (Figura 38) vs. **H** y **skill vs. climatología** (Figura 39).

- Experimento 1 (H = 1 día)

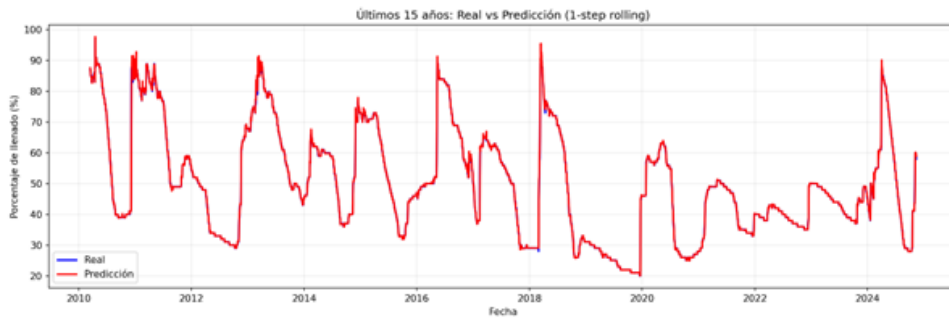


Figura 23: Evolución del porcentaje de llenado en los últimos 15 años con H = 1

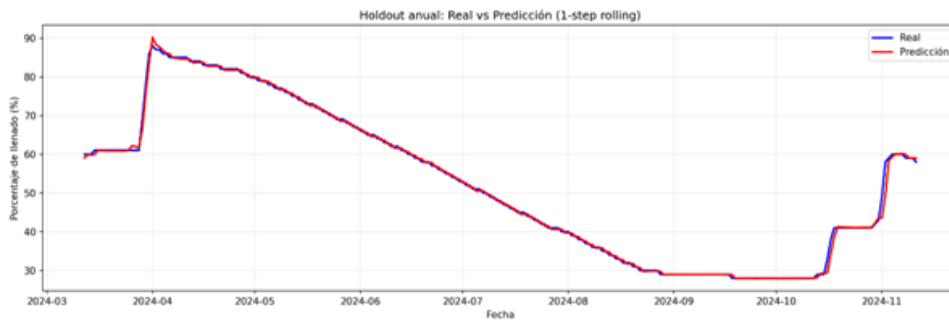


Figura 24: Evolución del porcentaje de llenado en un año con H = 1



Figura 25: Curva de entrenamiento del modelo predictivo con H = 1

MAE 0.472 p.p., RMSE 0.910 p.p. 0.89 %,  $R^2$  0.998; bases:  $R_{\text{clima}}^2$  0.560,  $R_{\text{persistente}}^2$  0.995, skill +0.437;  $\approx$  2251 s. Reproduce fielmente ciclo anual e histórico en rolling 1-paso; la validación es estable con ligera oscilación final.

- Experimento 2 (H = 3 días)



Figura 26: Evolución del porcentaje de llenado en los últimos 15 años con  $H = 3$

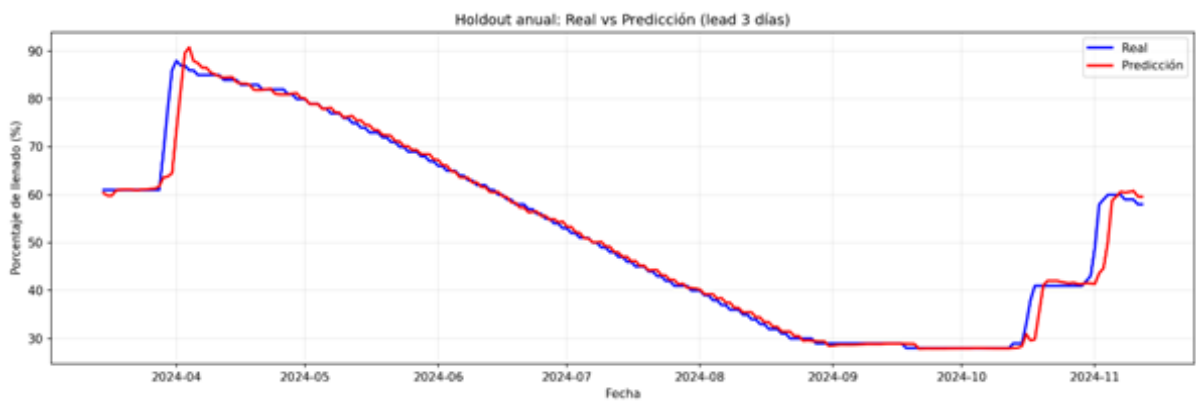


Figura 27: Evolución del porcentaje de llenado en un año con  $H = 3$

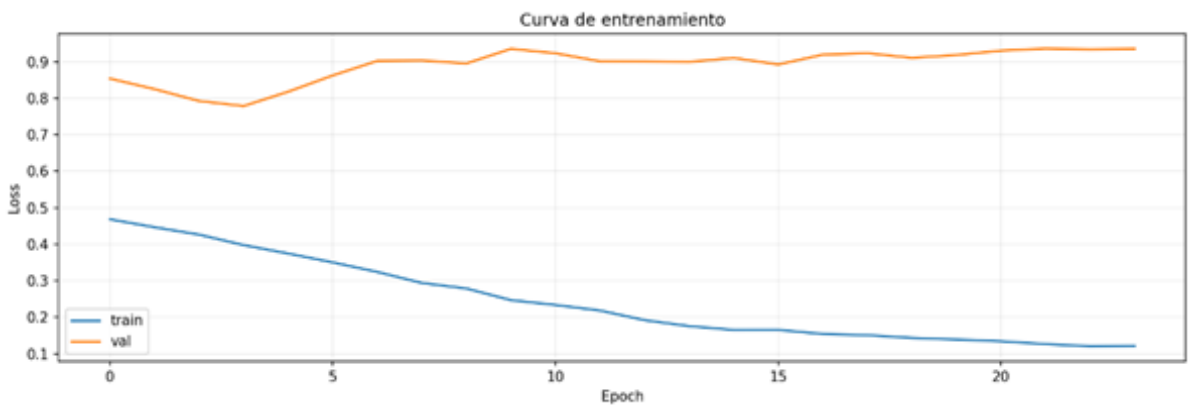


Figura 28: Curva de entrenamiento del modelo predictivo con  $H = 3$

MAE 1.086, RMSE 2.785, MAPE 1.99 %,  $R^2$  0.979; clima 0.558, persistente 0.970, skill +0.422;  $\approx$  1160 s. Ajuste estrecho con leves desfases en picos; la validación sugiere moderado sobreajuste.

■ Experimento 3 (H = 7 días)

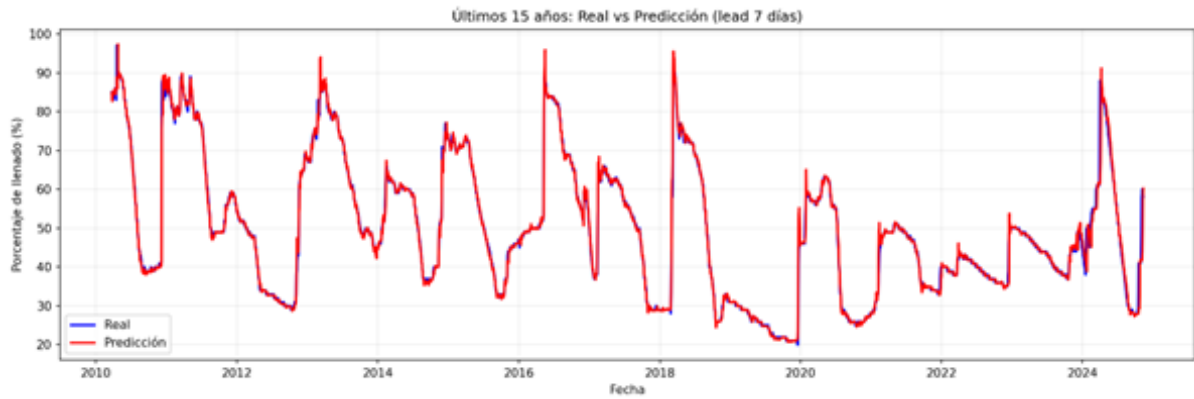


Figura 29: Evolución del porcentaje de llenado en los últimos 15 años con H = 7

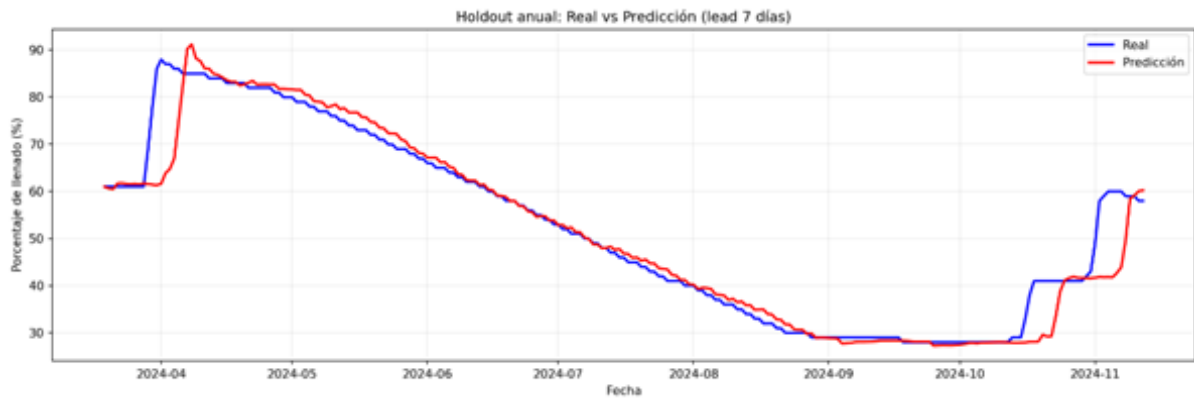


Figura 30: Evolución del porcentaje de llenado en un año con H = 7

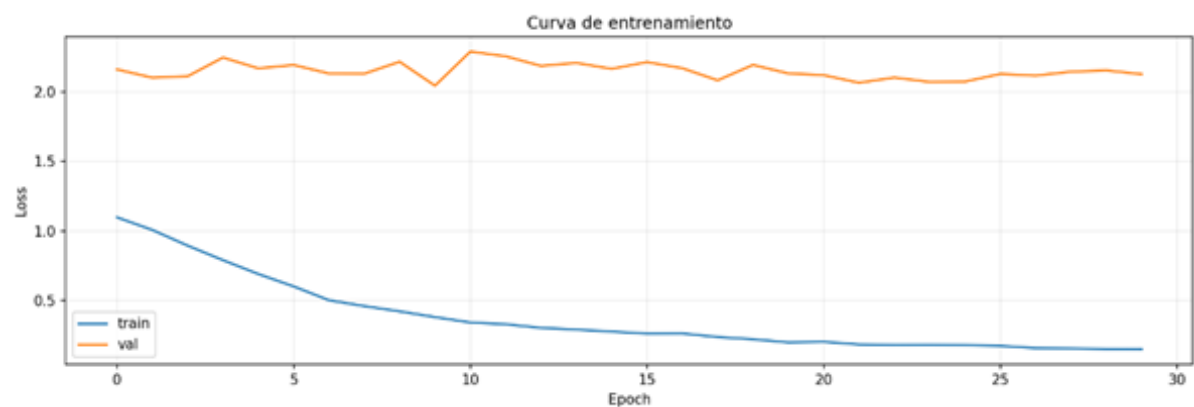


Figura 31: Curva de entrenamiento del modelo predictivo con H = 7

MAE 2.441, RMSE 5.183, MAPE 4.39 %,  $R^2$  0.929; clima 0.556, persistente 0.908, skill +0.373;  $\approx$  1658 s. Sigue bien tendencia y descensos prolongados; la validación se estabiliza alta al final.

■ Experimento 4 (H = 14 días)



Figura 32: Evolución del porcentaje de llenado en los últimos 15 años con H = 14

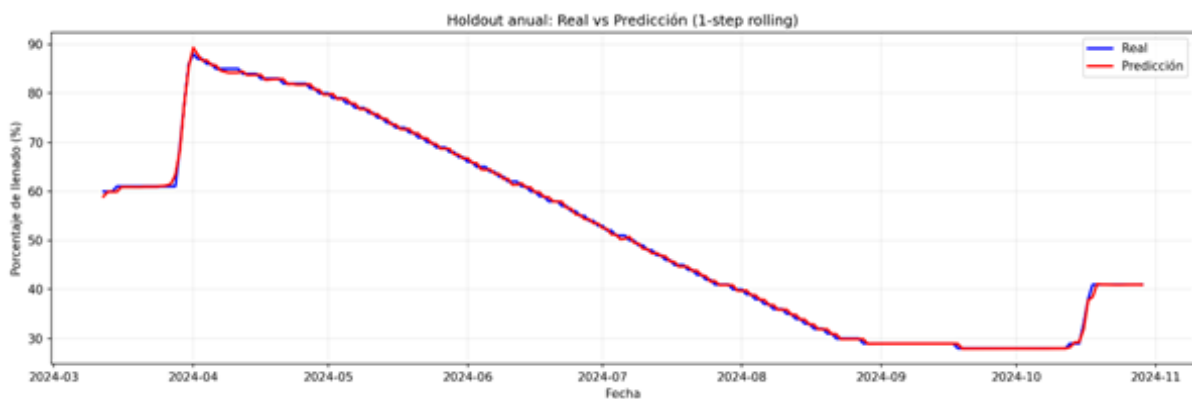


Figura 33: Evolución del porcentaje de llenado en un año con H = 14

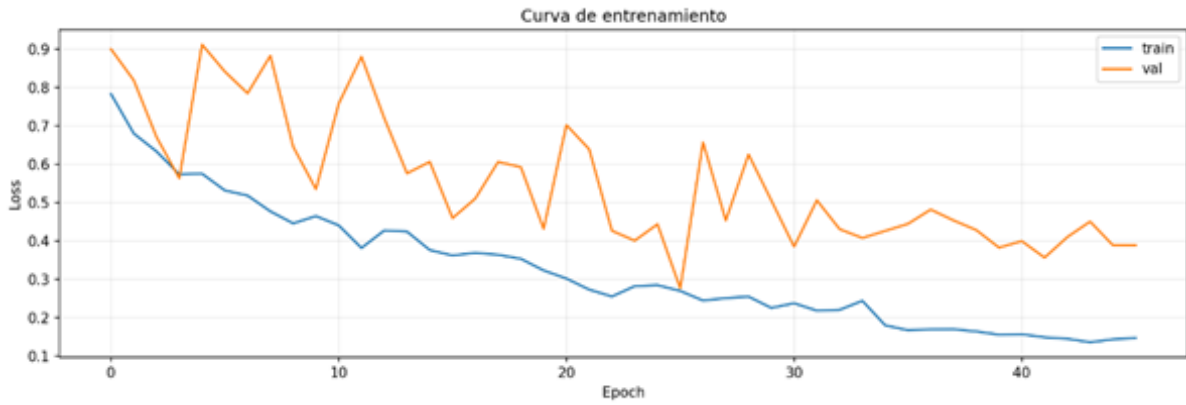


Figura 34: Curva de entrenamiento del modelo predictivo con H = 14

MAE 0.375, RMSE 0.526, MAPE 0.73 %,  $R^2$  0.999; clima 0.584, persistente 0.996, skill +0.415;  $\approx$  3470 s. Solape casi completo real-predicho; el residual corrige sistemáticamente fallos de persistencia en picos/cambios.

■ Experimento 5 (H = 30 días)



Figura 35: Evolución del porcentaje de llenado en los últimos 15 años con H = 30

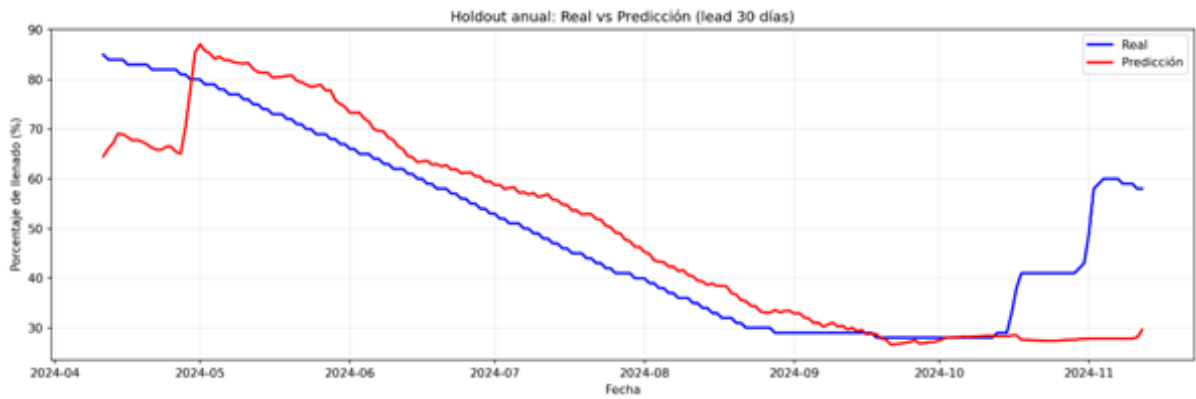


Figura 36: Evolución del porcentaje de llenado en un año con H = 30

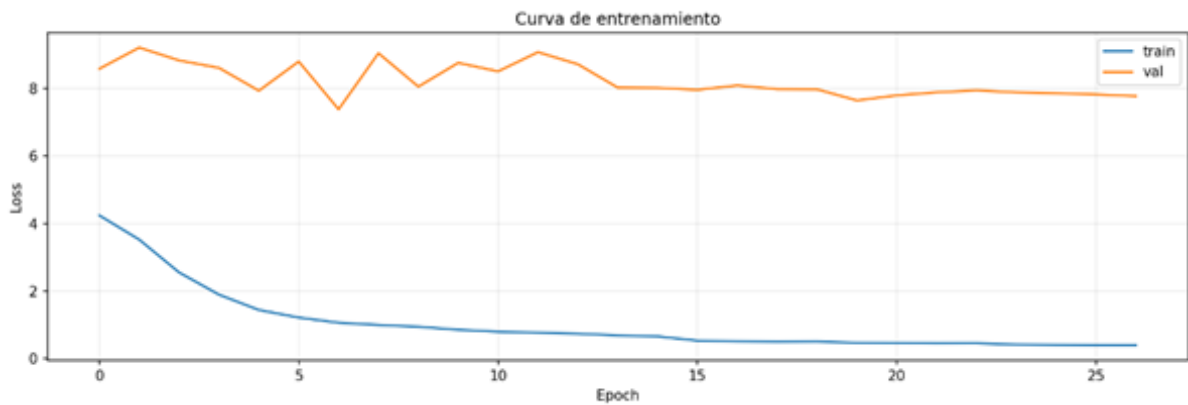


Figura 37: Curva de entrenamiento del modelo predictivo con H = 30

MAE 7.855, RMSE 10.523, MAPE 14.90 %,  $R^2$  0.682; clima 0.564, persistente 0.443, skill +0.119;  $\approx$  1553 s. Mantiene forma global, pero acumula errores de fase y amplitud; validación elevada.

Tabla 2: Resumen métricas evaluativas para los distintos horizontes de predicción

Horizonte	MAE (p.p.)	RMSE (p.p.)	MAPE (%)	$R^2$	$R^2_{\text{persistente}}$	$R^2_{\text{climatología}}$	skill climatología	Tiempo (s)
1	0,472	0,910	0,89	0,998	0,995	0,560	0,437	2251
3	1,086	2,785	1,99	0,979	0,970	0,558	0,422	1160
7	2,441	5,183	4,39	0,929	0,908	0,556	0,373	1658
<b>14</b>	<b>0,375</b>	<b>0,526</b>	<b>0,73</b>	<b>0,999</b>	<b>0,996</b>	<b>0,584</b>	<b>0,415</b>	<b>3470</b>
30	7,855	10,523	14,90	0,682	0,443	0,564	0,119	1553

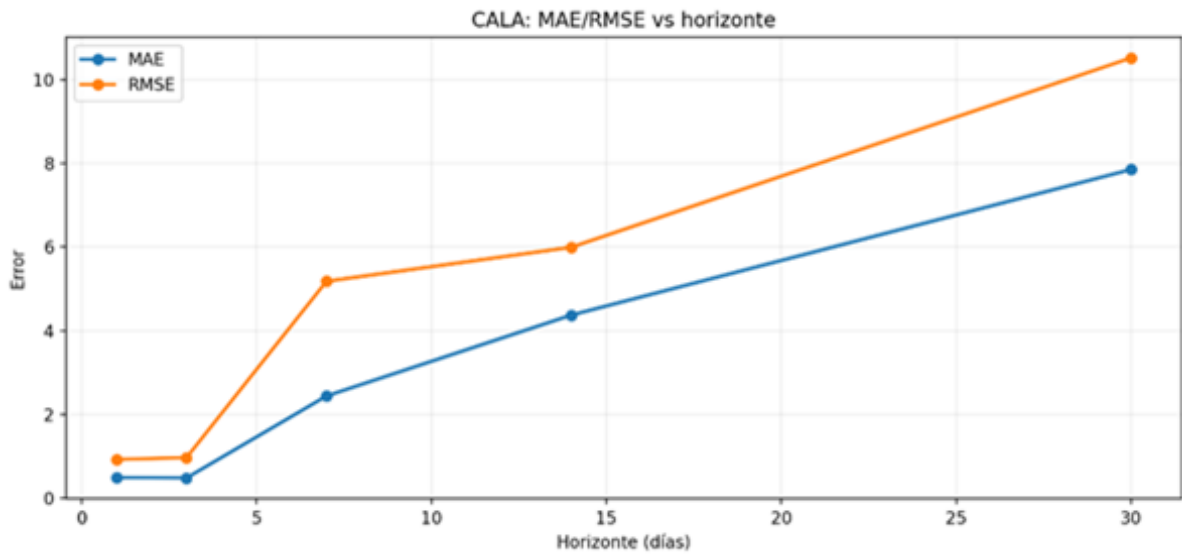


Figura 38: Evolución de los valores de error por horizonte de predicción

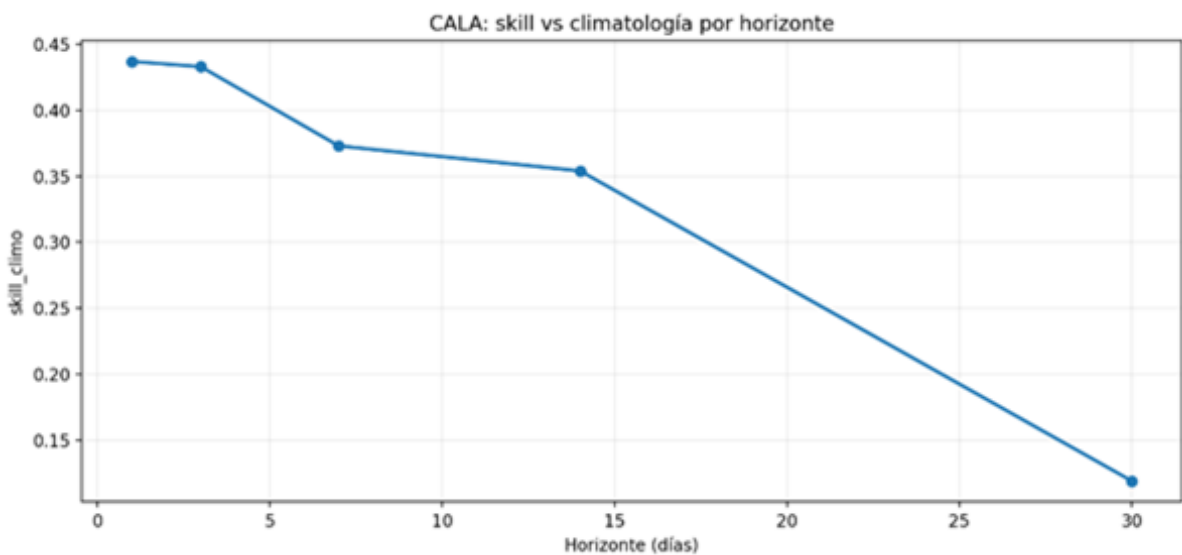


Figura 39: Robustez frente a valores climatológicos según horizonte de predicción

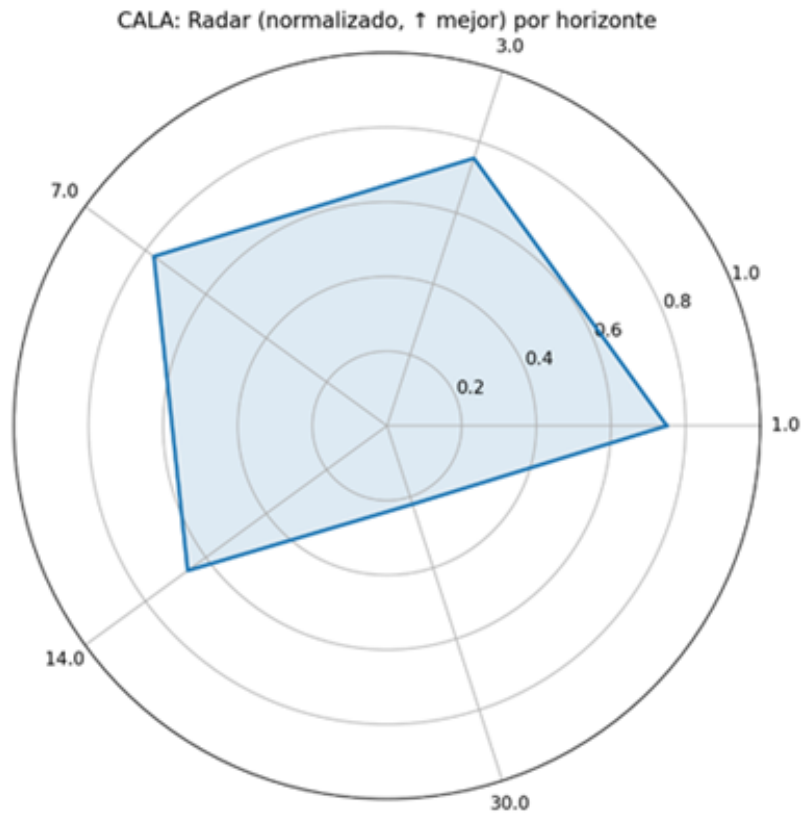


Figura 40: Radar normalizado de desempeño por horizonte

### Entrenamiento de modelos predictivos - *Másters* y *Outliers*

La **evaluación** se realizó con *hold-out* anual y validación *rolling*, reportando MAE, RMSE, MAPE y  $R^2$ , Skill frente a climatología y persistencia, además del **tiempo de entrenamiento** en segundos.

En  $H=14$  la **persistencia** es una línea base **muy competitiva**, por lo que todas las comparaciones se informan explícitamente respecto a ella. Para cada modelo se registran **tiempos de entrenamiento, curvas de pérdida y series real vs. predicción**, garantizando comparabilidad entre embalses, conjuntos de *features* y estrategias (entrenamiento completo vs. *transfer learning*).

- **Modelo embalse de Cala**

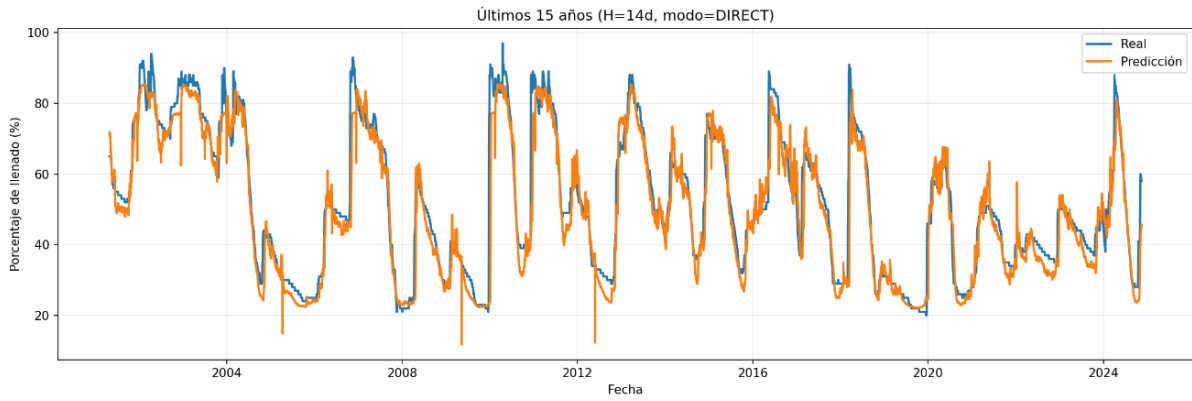


Figura 41: Evolución del porcentaje de llenado en los últimos 15 años para el embalse de Cala

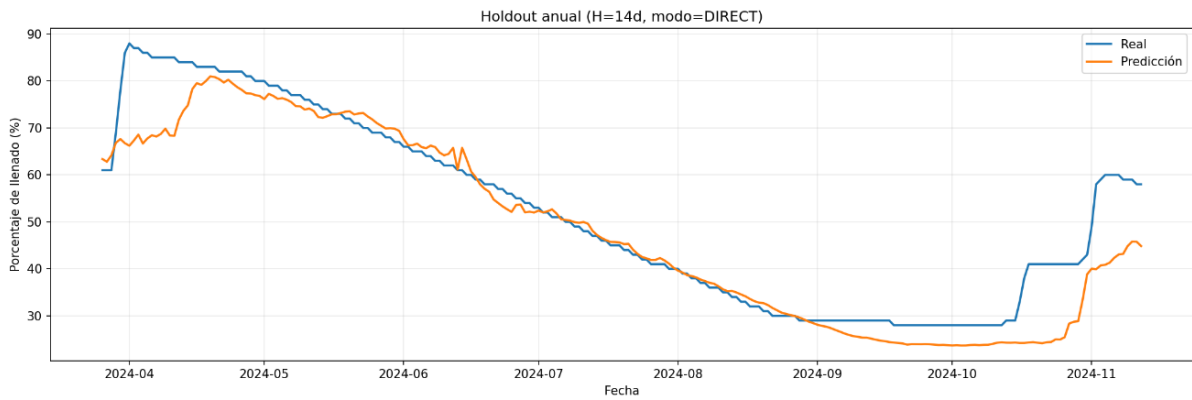


Figura 42: Evolución del porcentaje de llenado en un año para el embalse de Cala

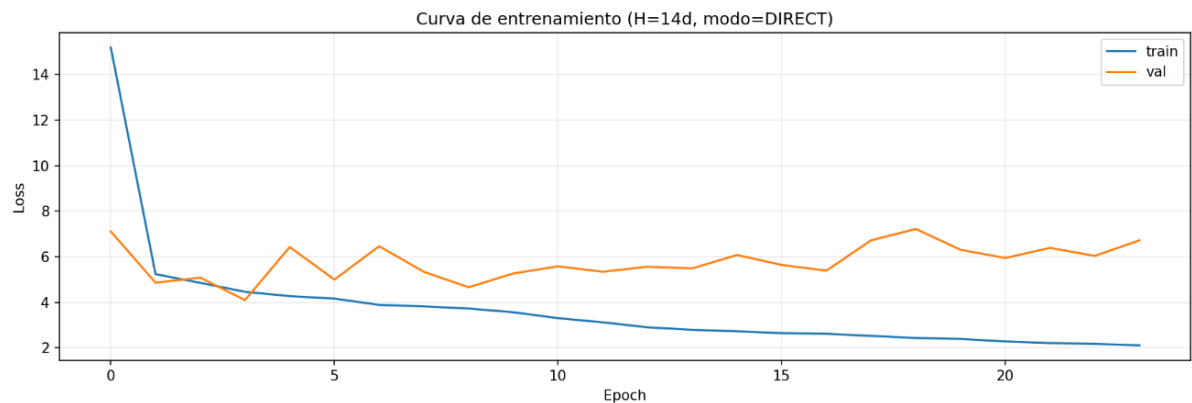


Figura 43: Curva de entrenamiento del modelo predictivo para el embalse de Cala

El modelo directo con 82 variables presenta **MAE = 4.554 p.p.**, **RMSE = 7.127 p.p.** y **MAPE = 9.19 %**, con  **$R^2 = 0.869$** . Frente a las referencias, la persistencia alcanza  **$R^2 = 0.995$**  y

la climatología  $R^2 = 0.554$ , de modo que el **skill** reportado es **+0.316**. El tiempo de entrenamiento fue 1241.8 s.

■ **Modelo embalse de Canales**

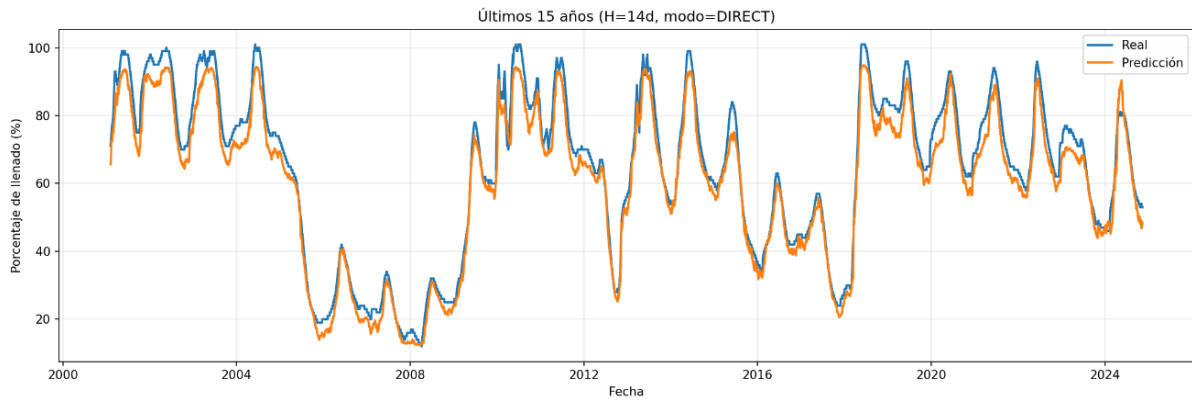


Figura 44: Evolución del porcentaje de llenado en los últimos 15 años para el embalse de Canales

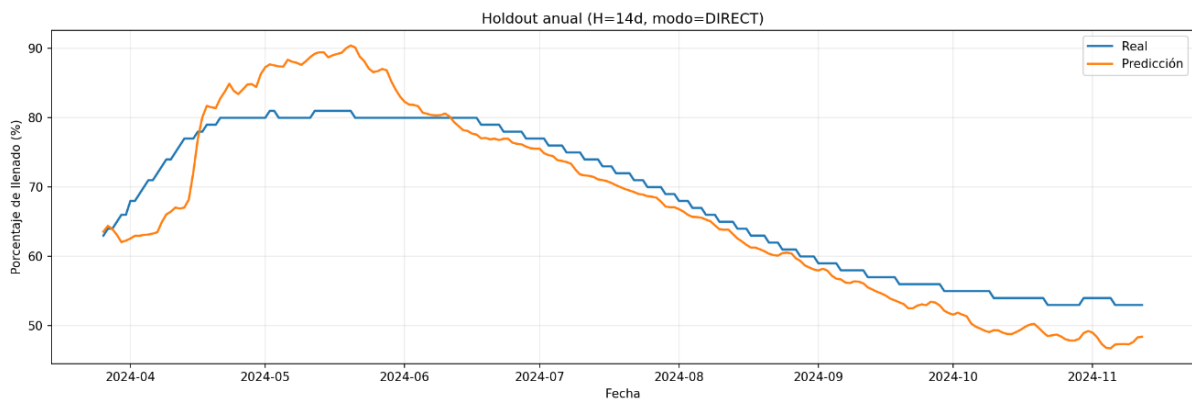


Figura 45: Evolución del porcentaje de llenado en un año para el embalse de Canales

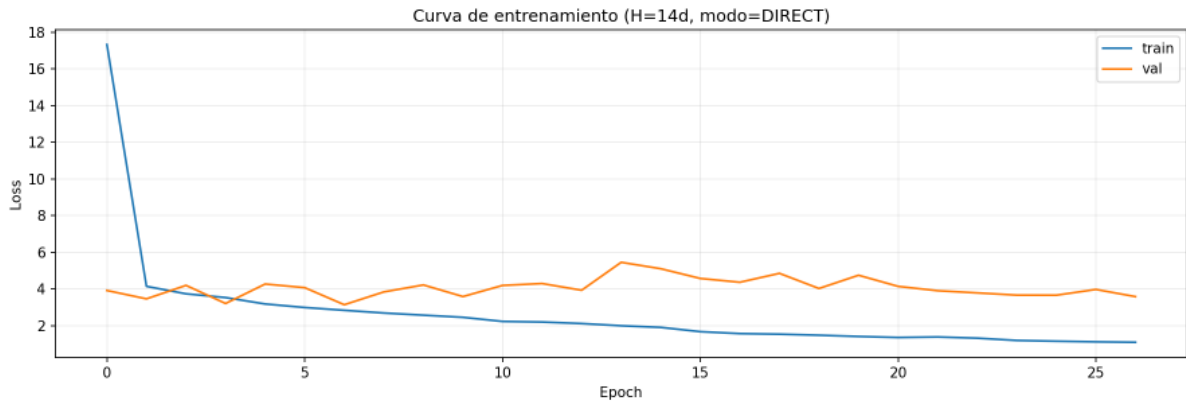


Figura 46: Curva de entrenamiento del modelo predictivo para el embalse de Canales

El modelo directo con 82 variables presenta **MAE = 3.623 p.p.**, **RMSE = 4.395 p.p.** y **MAPE = 5.40 %**, con  **$R^2 = 0.821$** . Frente a las referencias, la **persistencia alcanza  $R^2 = 0.998$**  y la **climatología  $R^2 = 0.876$** , de modo que el **skill** reportado es **-0.055**. El tiempo de entrenamiento fue **1334.6 s**.

■ **Modelo embalse de Iznájar**

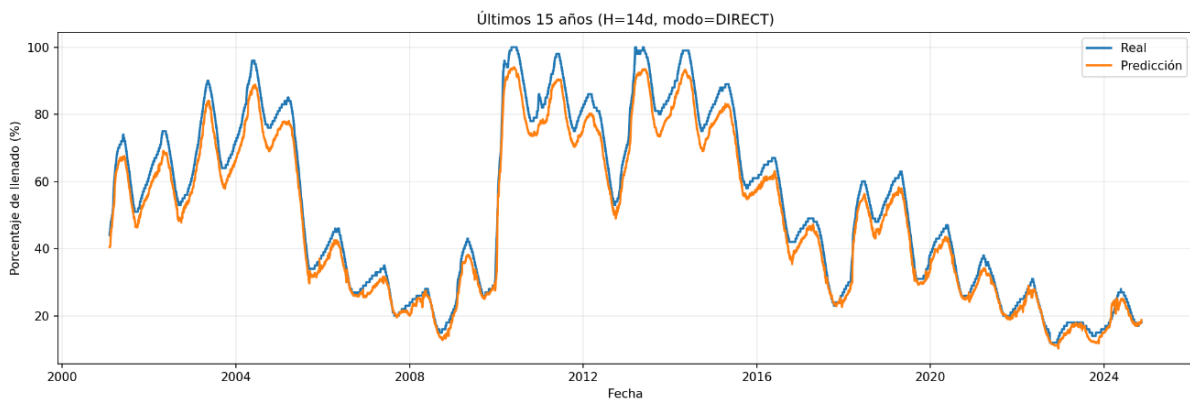


Figura 47: Evolución del porcentaje de llenado en los últimos 15 años para el embalse de Iznájar

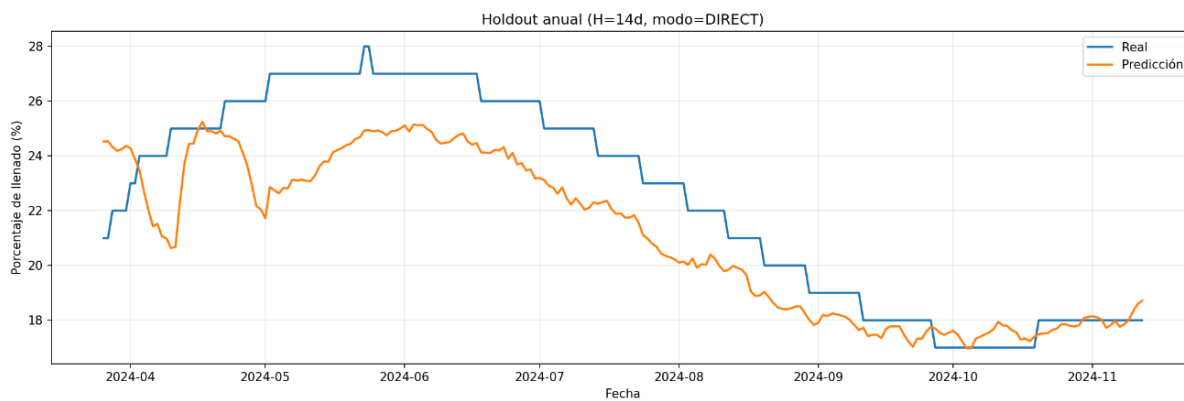


Figura 48: Evolución del porcentaje de llenado en un año para el embalse de Iznájar

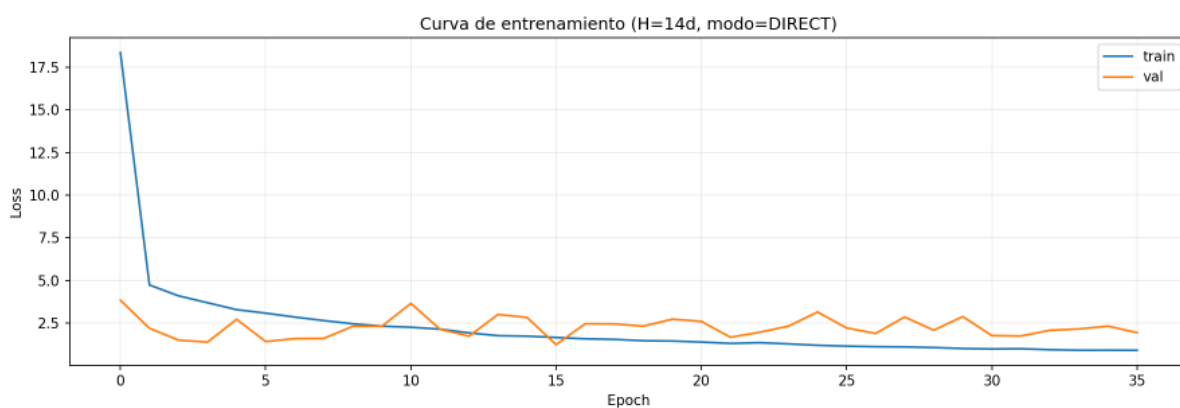


Figura 49: Curva de entrenamiento del modelo predictivo para el embalse de Iznájar

Los errores se sitúan en **MAE = 1.677 p.p.** y **RMSE = 2.021 p.p.** con **MAPE = 7.00 %** y **R<sup>2</sup> = 0.704**. Las referencias muestran **persistencia R<sup>2</sup> = 0.994** y **climatología R<sup>2</sup> = -72.770**; el **skill** registrado es **73.474**. **Tiempo de entrenamiento: 1800.7 s.**

- **Modelo embalse de San Rafael de Navallana**

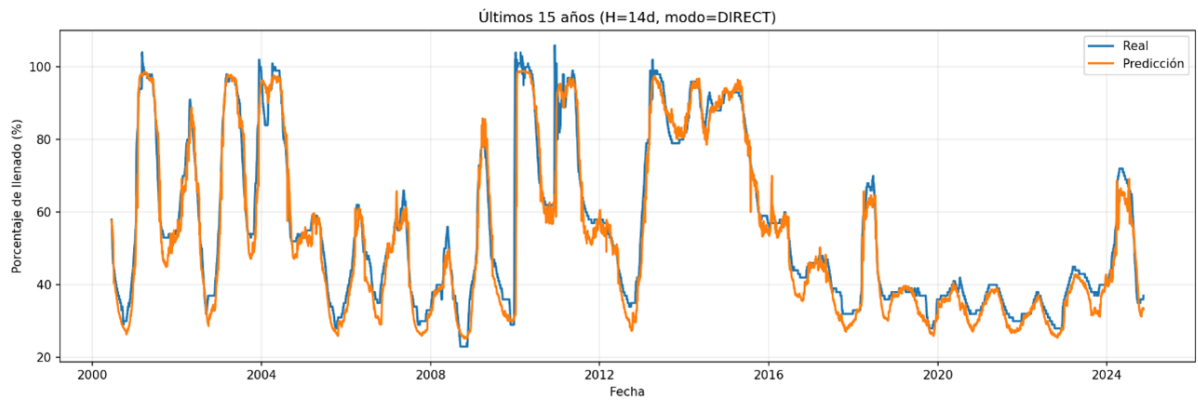


Figura 50: Evolución del porcentaje de llenado en los últimos 15 años para el embalse de San Rafael de Navallana

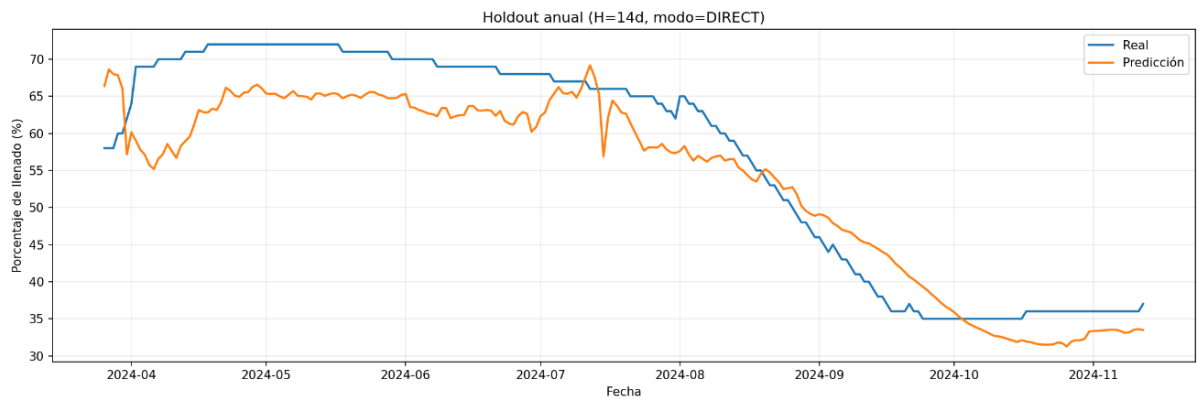


Figura 51: Evolución del porcentaje de llenado en un año para el embalse de San Rafael de Navallana

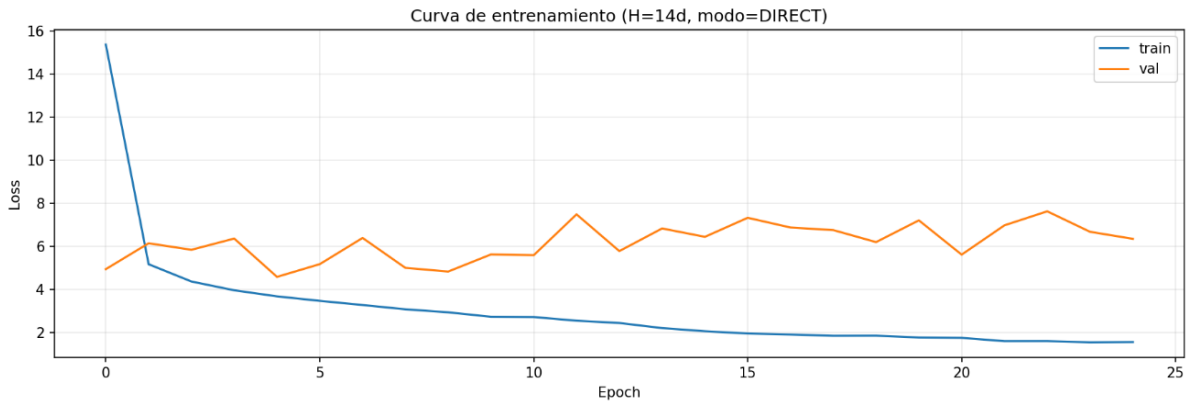


Figura 52: Curva de entrenamiento del modelo predictivo para el embalse de San Rafael de Navallana

El ajuste obtenido es **MAE = 5.094 p.p.**, **RMSE = 5.796 p.p.**, **MAPE = 8.87 %**, **con  $R^2 = 0.844$** . La persistencia alcanza  **$R^2 = 0.998$**  y la climatología  **$R^2 = 0.737$** ; el skill informado es **0.107**. Duración del entrenamiento: 1279.9 s.

■ **Modelo embalse de Barbate**

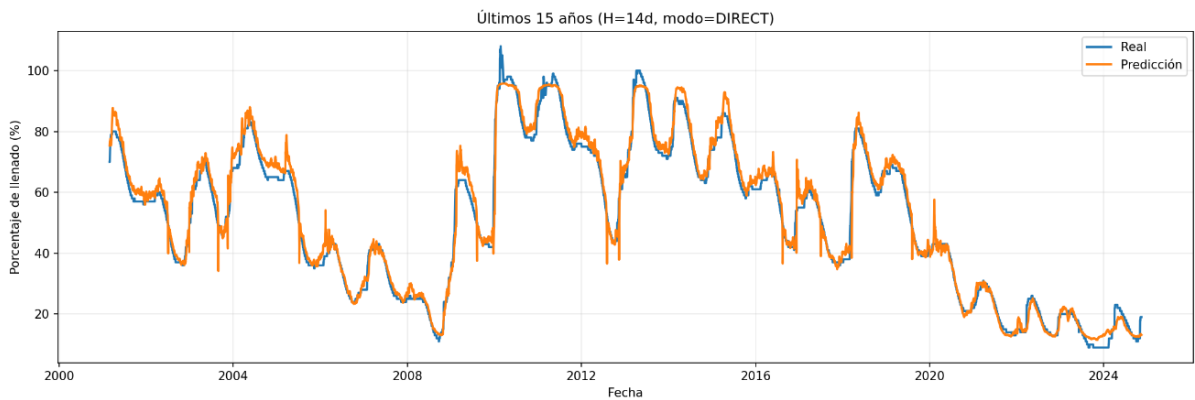


Figura 53: Evolución del porcentaje de llenado en los últimos 15 años para el embalse de Barbate

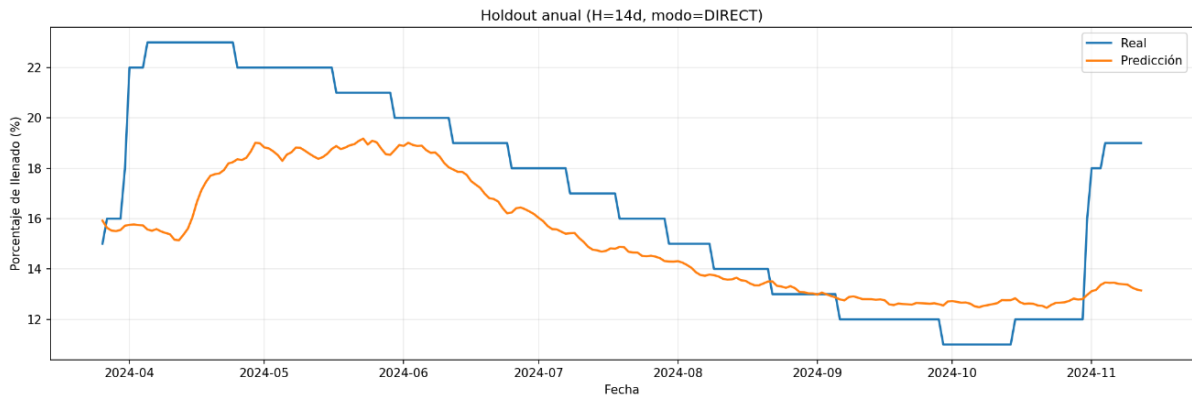


Figura 54: Evolución del porcentaje de llenado en un año para el embalse de Barbate

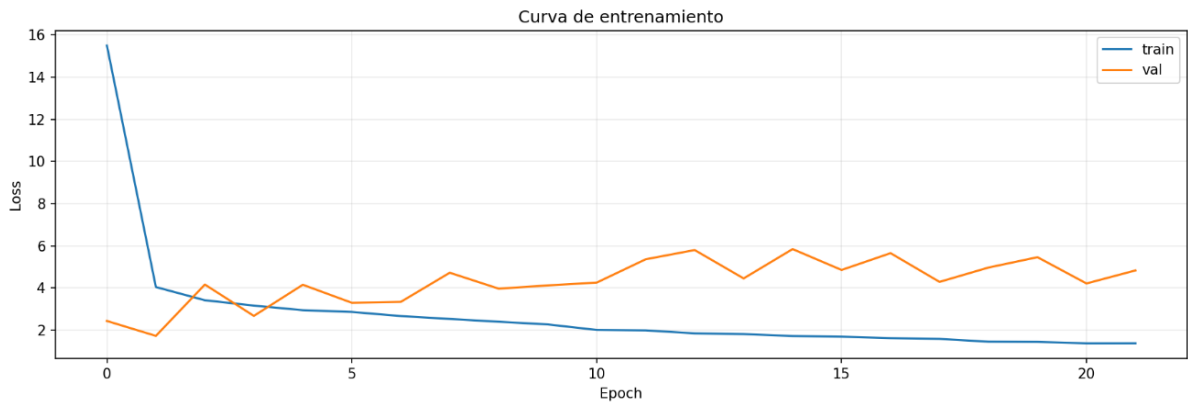


Figura 55: Curva de entrenamiento del modelo predictivo para el embalse de Barbate

Con 82 variables, el modelo reporta **MAE = 2.186 p.p.**, **RMSE = 2.930 p.p.**, **MAPE = 11.73 %** y  **$R^2 = 0.471$** . Las referencias son **persistencia  $R^2 = 0.985$**  y **climatología  $R^2 = -79.183$** , con **skill = 79.654**. **Tiempo: 1938.2 s.**

- **Modelo embalse de Guadalmena**

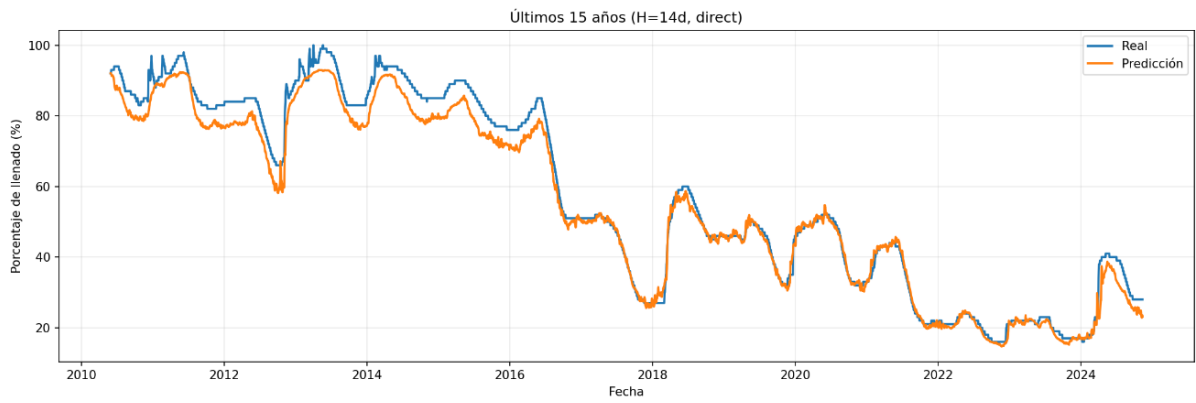


Figura 56: Evolución del porcentaje de llenado en los últimos 15 años para el embalse de Guadalmena

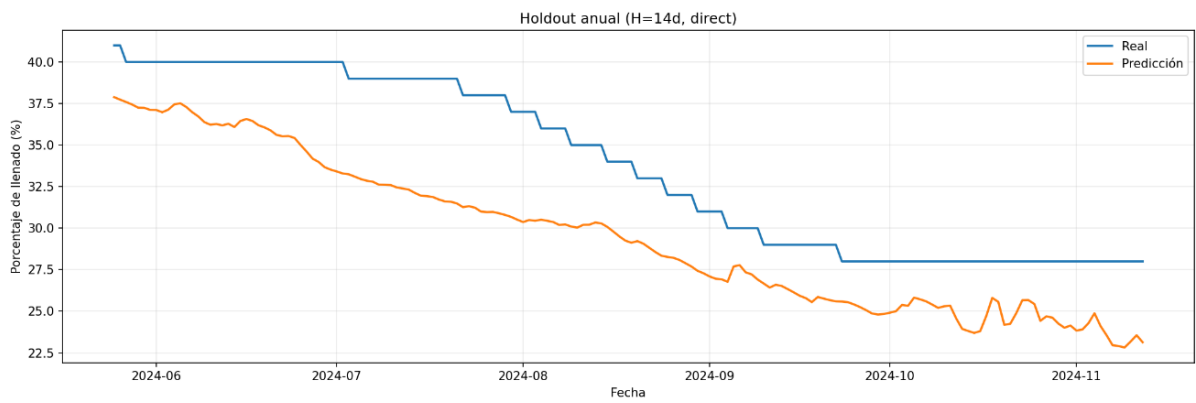


Figura 57: Evolución del porcentaje de llenado en un año para el embalse de Guadalmena

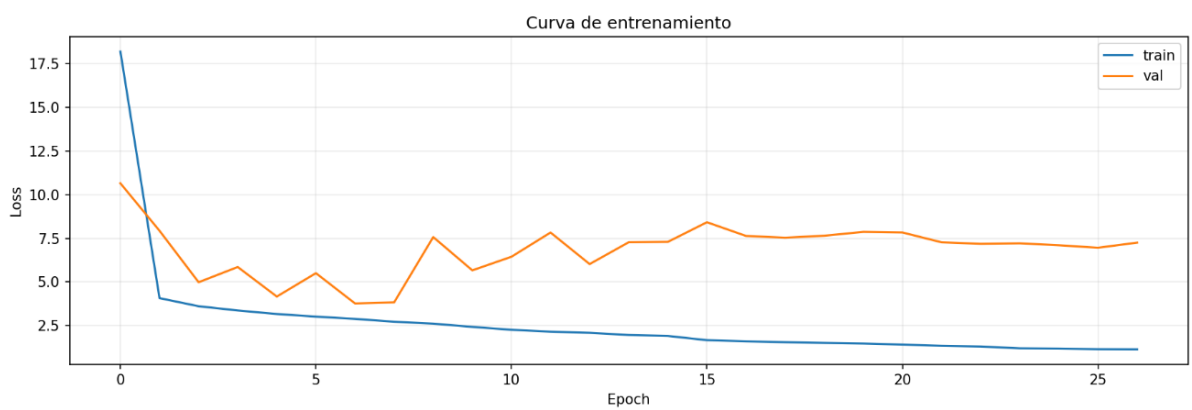


Figura 58: Curva de entrenamiento del modelo predictivo para el embalse de Guadalmena

Para 82 variables, las métricas son **MAE = 4.273 p.p.**, **RMSE = 4.532 p.p.**, **MAPE = 12.55 %**

y  $R^2 = 0.202$ . Las referencias indican **persistencia**  $R^2 = 0.997$  y **climatología**  $R^2 = -17.722$ ; **skill** = 17.924. **Tiempo: 2125.0 s.**

■ **Modelo embalse de Andévalo**

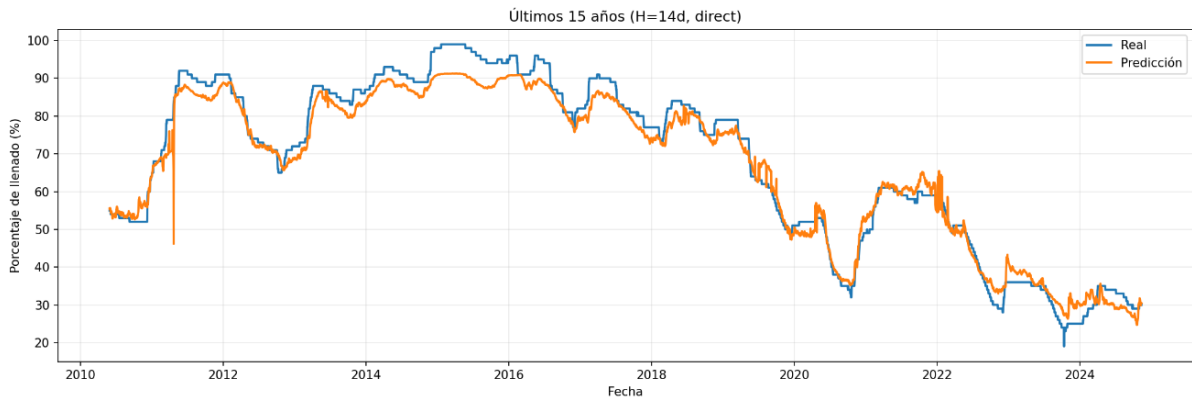


Figura 59: Evolución del porcentaje de llenado en los últimos 15 años para el embalse de Andévalo

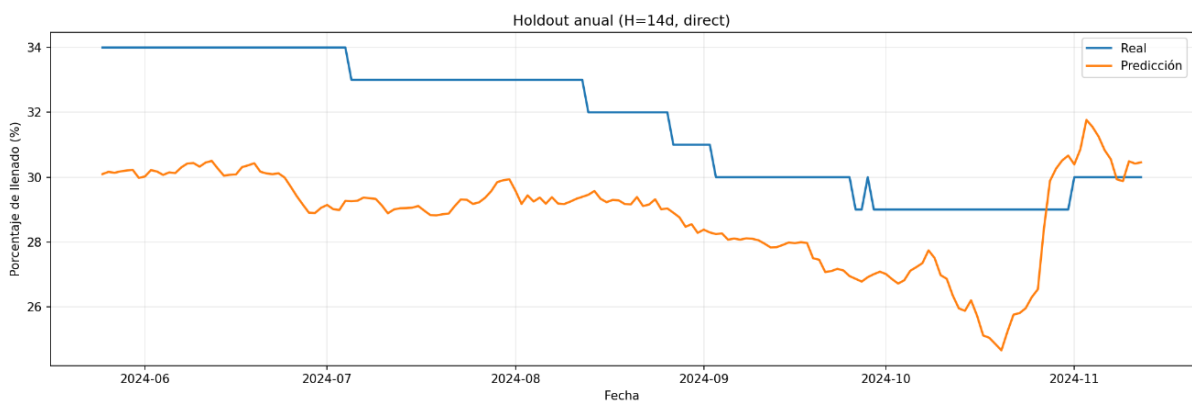


Figura 60: Evolución del porcentaje de llenado en un año para el embalse de Andévalo

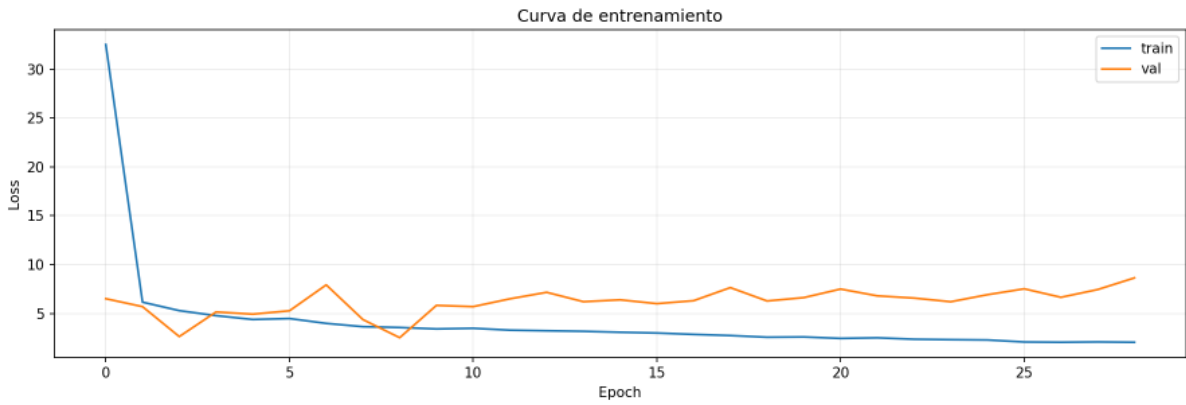


Figura 61: Curva de entrenamiento del modelo predictivo para el embalse de Andévalo

El modelo con 82 variables registra **MAE = 3.001 p.p.**, **RMSE = 3.196 p.p.**, **MAPE = 9.37 %** y **R<sup>2</sup> = -1.720**. Las referencias son **persistencia R<sup>2</sup> = 0.988** y **climatología R<sup>2</sup> = -253.739**; **skill = 252.019**. **Tiempo de entrenamiento: 1349.3 s.**

■ **Modelo embalse de Guadalteba**

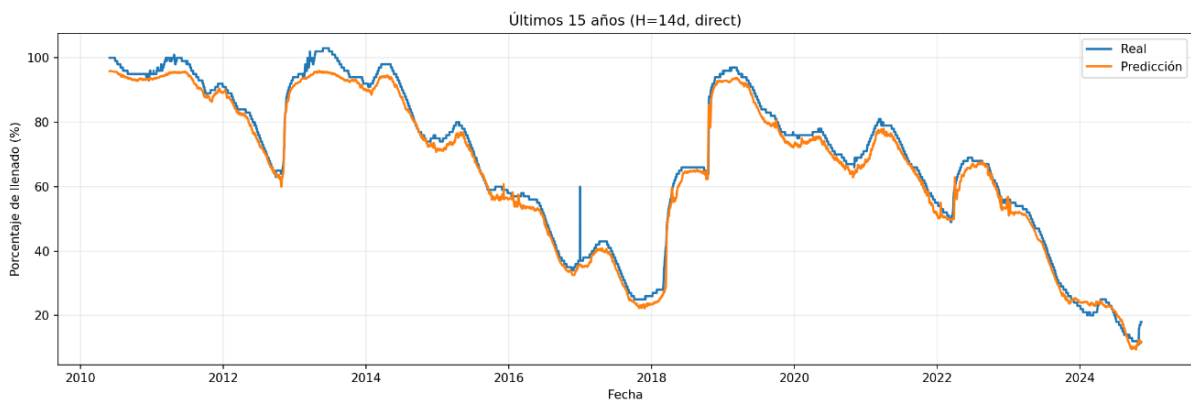


Figura 62: Evolución del porcentaje de llenado en los últimos 15 años para el embalse de Guadalteba

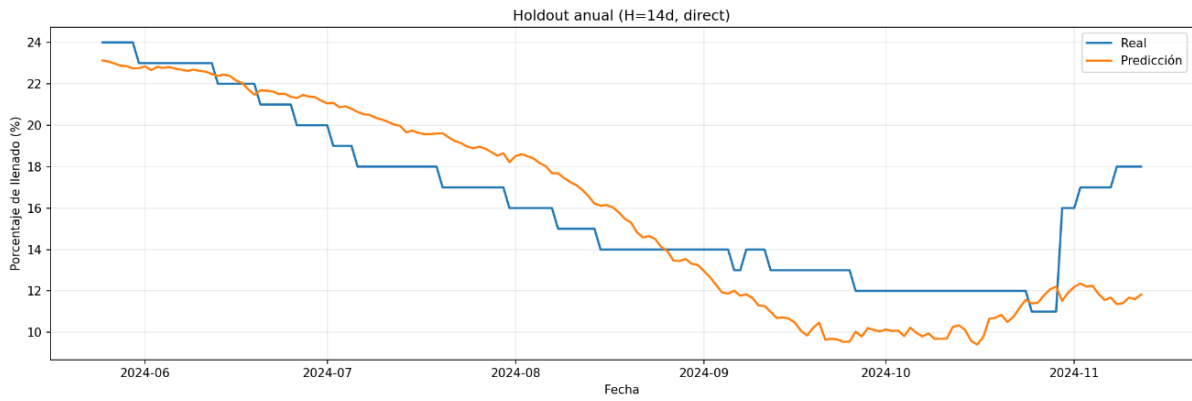


Figura 63: Evolución del porcentaje de llenado en un año para el embalse de Guadalteba

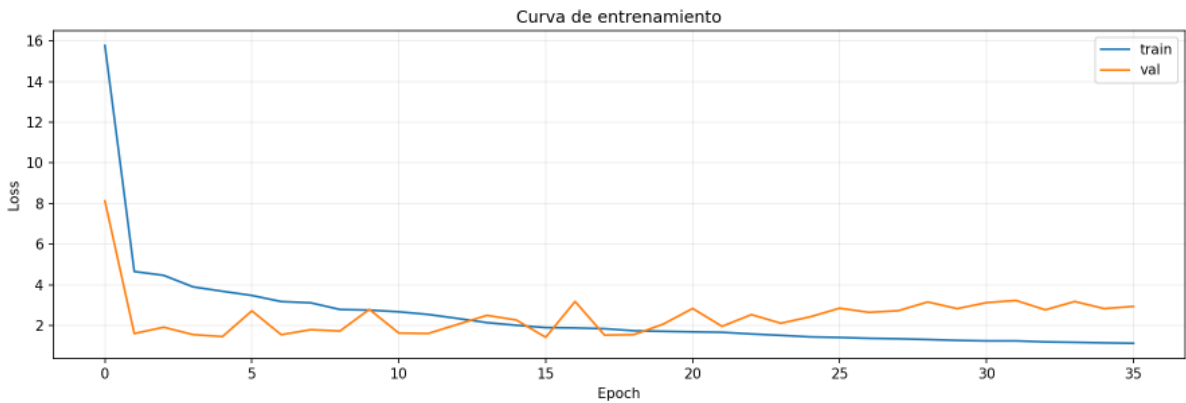


Figura 64: Curva de entrenamiento del modelo predictivo para el embalse de Guadalteba

Con 82 variables, el resultado es **MAE = 1.872 p.p.**, **RMSE = 2.300 p.p.**, **MAPE = 12.35 %** y  **$R^2 = 0.632$** . Las referencias reportan **persistencia  $R^2 = 0.983$**  y **climatología  $R^2 = -94.271$** ; **skill = 94.903**. El **entrenamiento tomó 2785.7 s**.

Tabla 3: Métricas de evaluación obtenidas en el entrenamiento de los modelos predictivos

Embalse	Steps	MAE (p.p.)	RMSE (p.p.)	MAPE (%)	$R^2$	$R^2_{\text{persistente}}$	$R^2_{\text{climatología}}$	skill climatología	Tiempo (s)
Cala (outlier)	120	4,554	7,127	9,19 %	0,869	0,995	0,55	0,316	1241,8
Canales (outlier)	120	3,623	4,395	5,40 %	0,821	0,998	0,87	-0,055	1334,6
Iznájar (outlier)	120	1,677	2,021	7,00 %	0,704	0,994	-72,77	73,474	1800,7
San Rafael de Navallana (outlier)	120	5,094	5,796	8,87 %	0,844	0,998	0,73	0,107	1279,9
Barbate (máster)	120	2,186	2,930	11,73 %	0,471	0,985	-79,18	79,654	1938,2
Guadalmena (máster)	180	4,273	4,532	12,55 %	0,202	0,997	-17,72	17,924	2125,0
Andévalo (máster)	180	3,001	3,196	9,37 %	-1,720	0,988	-253,7	252,019	1349,3
Guadalteba (máster)	180	1,872	2,300	12,35 %	0,632	0,983	-94,27	94,903	2785,7

Los valores de skill corresponden al diferencial  $R^2_{modelos} - R^2_{climatología}$  reportado en cada caso; los tiempos se indican en segundos e incluyen el early stopping cuando procede.

### Resultados modelos predictivos entrenados mediante *transfer learning*

#### ■ Modelo embalse de Casasola

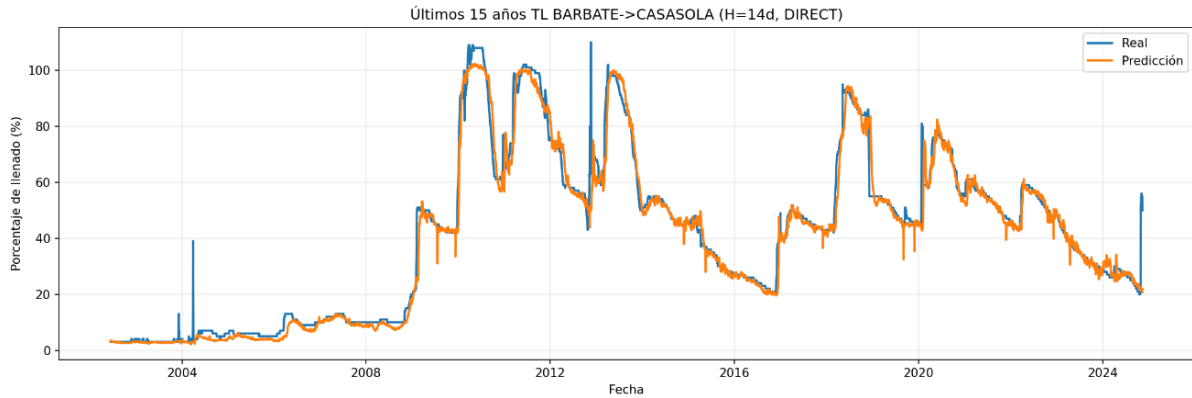


Figura 65: Evolución del porcentaje de llenado en los últimos 15 años para el embalse de Casasola

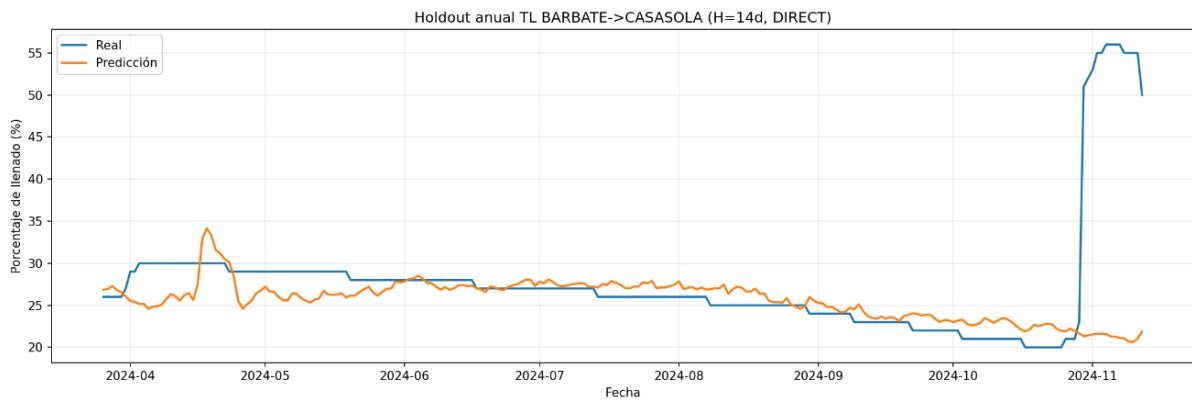


Figura 66: Evolución del porcentaje de llenado en un año para el embalse de Casasola

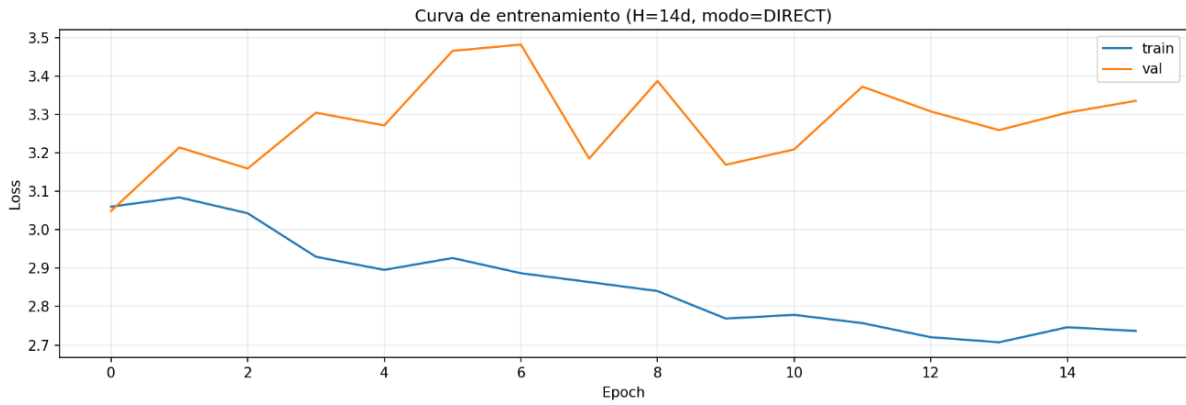


Figura 67: Curva de entrenamiento del modelo predictivo para el embalse de Casasola

Transfer learning 82: MAE 3.494, RMSE 8.348, MAPE 9.44 %,  $R^2$  -0.297; persist 0.933, clima -4.406, skill +4.109, 1773.1 s, 120 épocas.

■ **Modelo embalse de Guadalcacín**

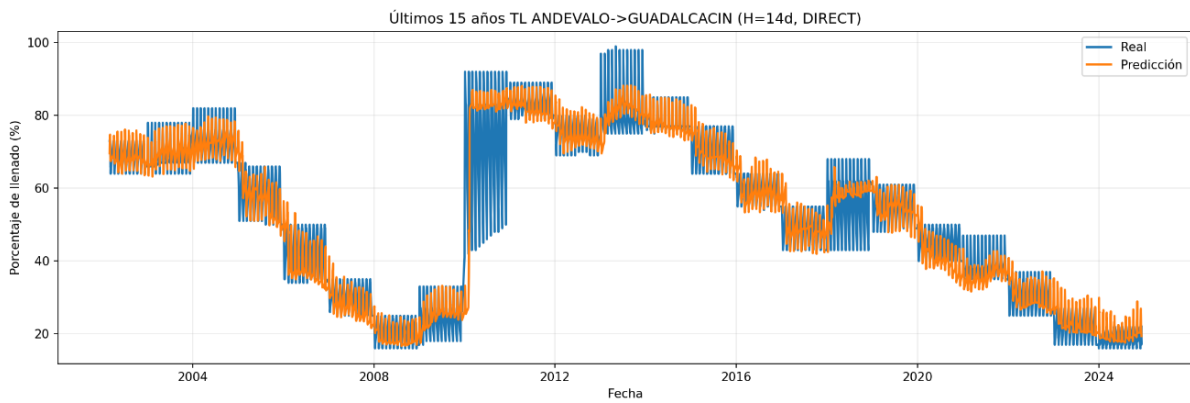


Figura 68: Evolución del porcentaje de llenado en los últimos 15 años para el embalse de Guadalcacín

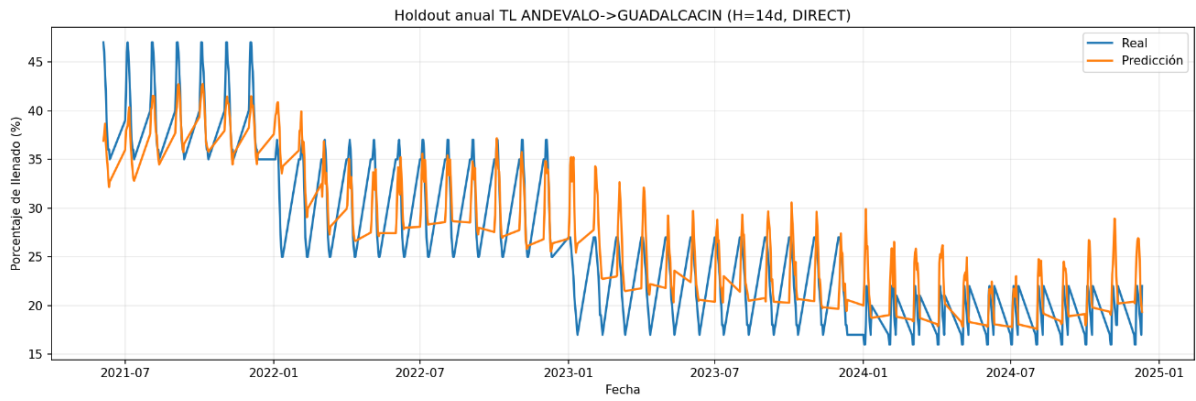


Figura 69: Evolución del porcentaje de llenado en un año para el embalse de Guadalcacín

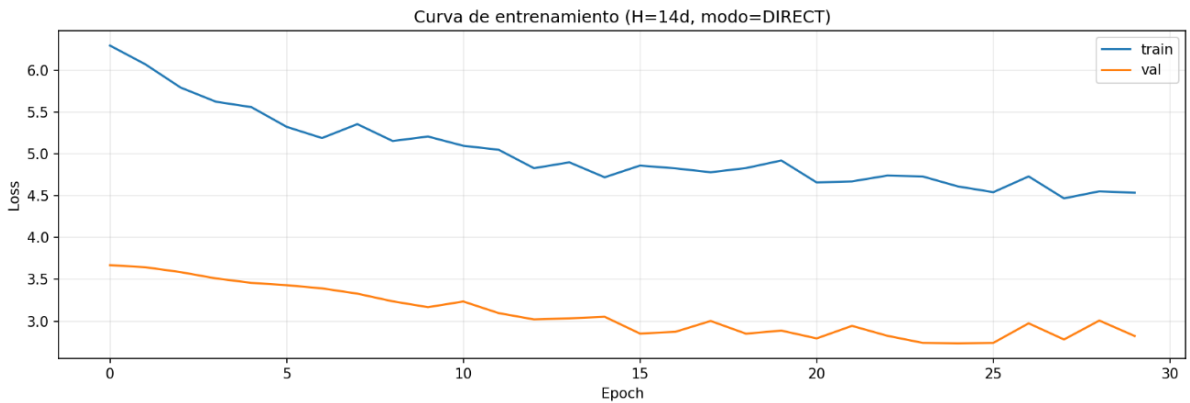


Figura 70: Curva de entrenamiento del modelo predictivo para el embalse de Guadalcacín

Transfer learning 82: MAE 3.212, RMSE 3.892, MAPE 13.06 %,  $R^2$  0.793; persist 0.887, clima -16.213, skill +17.007, 1058.2 s, 180 épocas.

- **Modelo embalse de Guadalhorce**

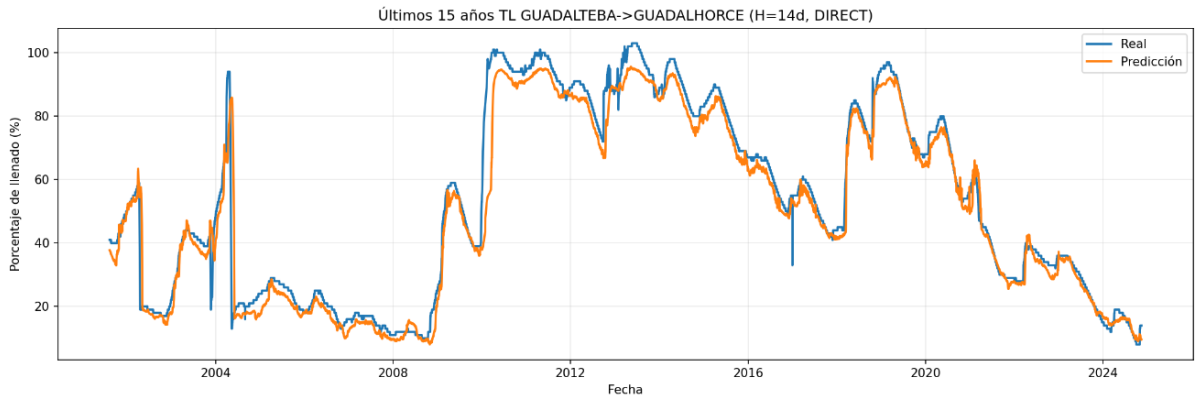


Figura 71: Evolución del porcentaje de llenado en los últimos 15 años para el embalse de Guadalhorce

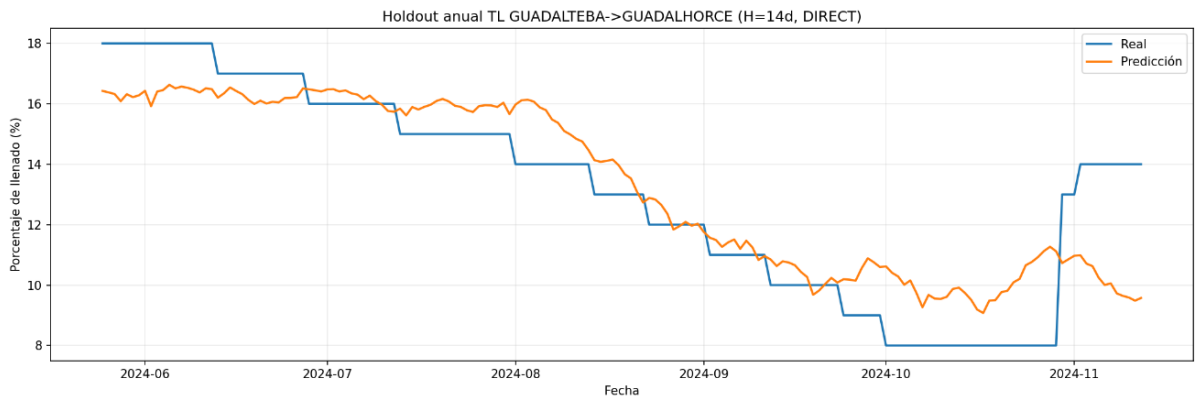


Figura 72: Evolución del porcentaje de llenado en un año para el embalse de Guadalhorce

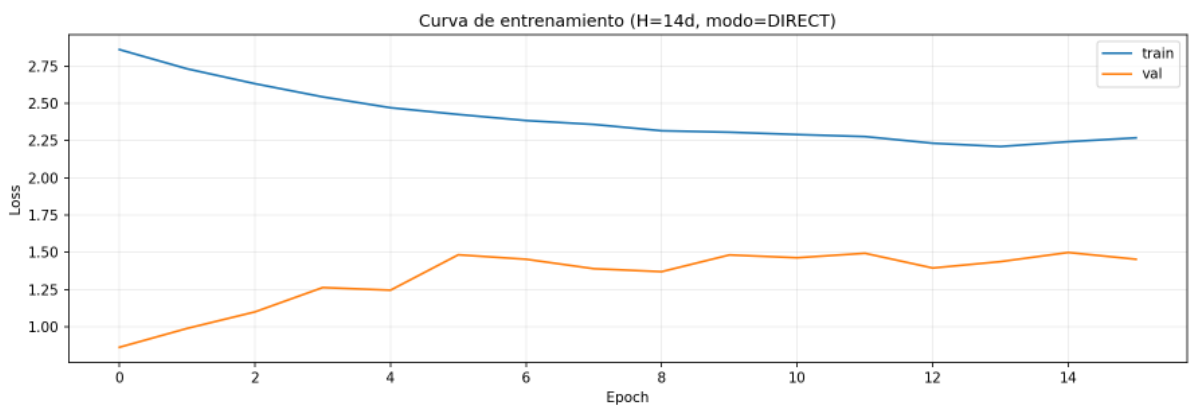


Figura 73: Curva de entrenamiento del modelo predictivo para el embalse de Guadalhorce

Transfer learning 82: MAE 1.286, RMSE 1.628, MAPE 11.01

## ■ Modelo embalse de Jándula

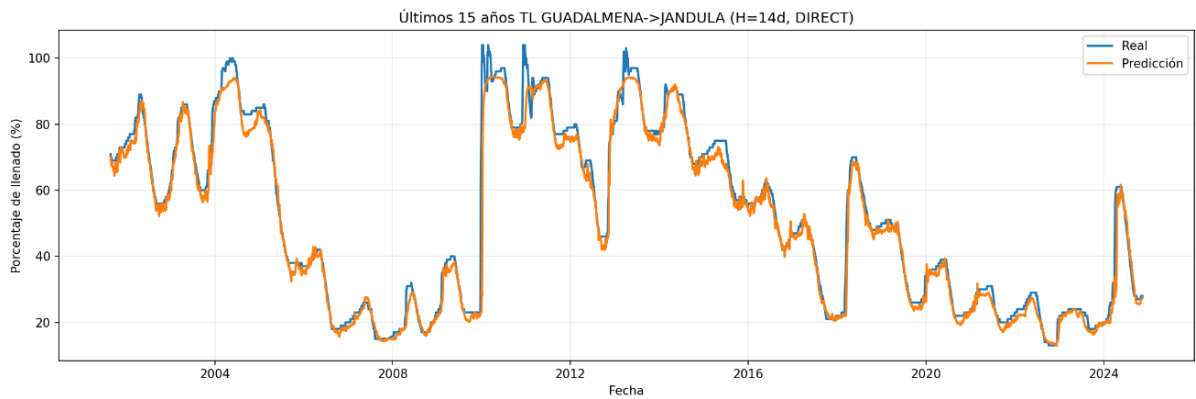


Figura 74: Evolución del porcentaje de llenado en los últimos 15 años para el embalse de Jándula

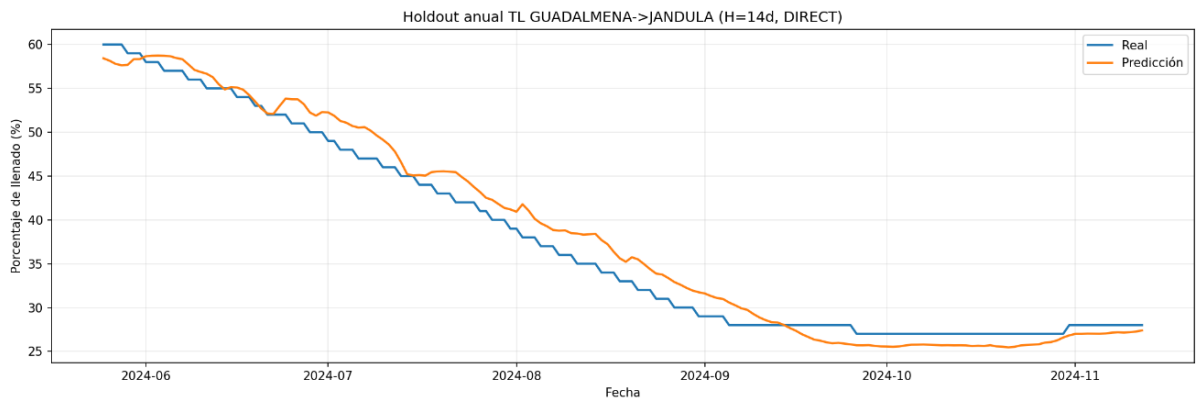


Figura 75: Evolución del porcentaje de llenado en un año para el embalse de Jándula

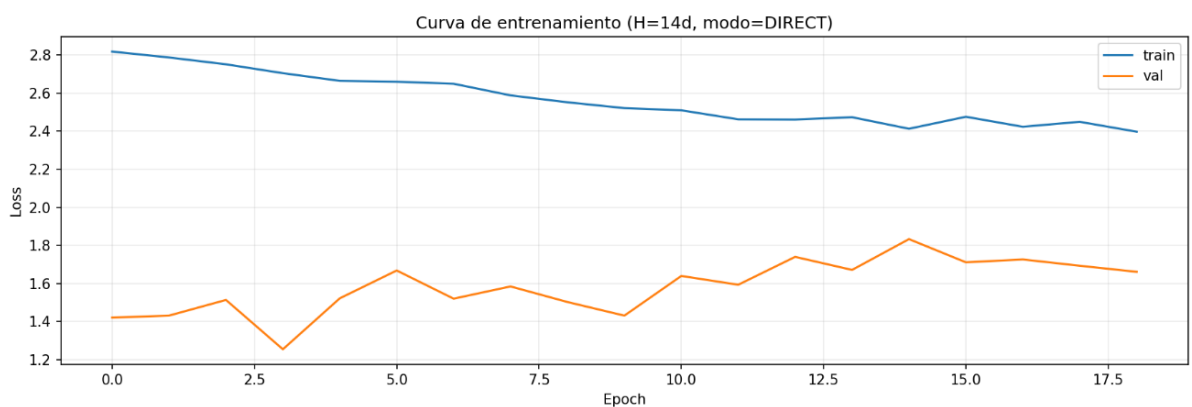


Figura 76: Curva de entrenamiento del modelo predictivo para el embalse de Jándula

Transfer learning 82: MAE 1.725, RMSE 1.946, MAPE 4.87 %,  $R^2$  0.970; persist 0.998, clima -0.260, skill +1.230, 2357.3 s, 180 épocas.

Tabla 4: Métricas de evaluación obtenidas en el entrenamiento de los modelos predictivos mediante *transfer learning*

Embalse	Steps	MAE (p.p.)	RMSE (p.p.)	MAPE (%)	$R^2$	$R^2_{\text{persistente}}$	$R^2_{\text{climatología}}$	Skill	Tiempo (s)
Jándula	180	1,725	1,946	4,87 %	0,970	0,998	-0,260	+1,230	2357,3
Guadalhorce	180	1,286	1,628	11,01 %	0,765	0,981	-132,239	+133,004	1907,6
Guadalcaçín	180	3,212	3,892	13,06 %	0,793	0,887	-16,213	+17,007	1058,2
Casasola	120	3,494	8,348	9,44 %	-0,297	0,933	-4,406	+4,109	1773,1

### Modo de entrenamiento Residual- Comparativa por embalse (H = 14)

A continuación, se muestran los resultados del entrenamiento, con modo residual, de todos los modelos predictivos (con H = 14).

#### ■ Modelo residual para embalse de Cala

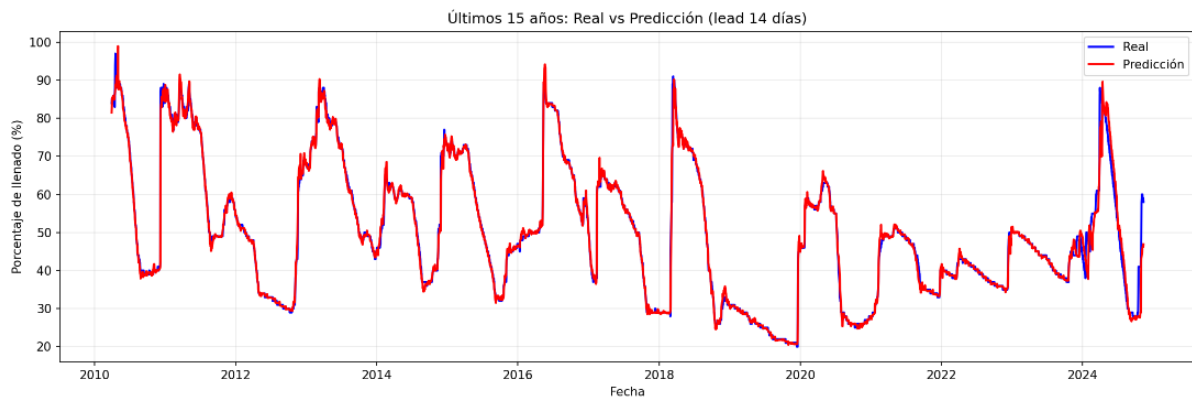


Figura 77: Evolución del porcentaje de llenado en los últimos 15 años en modo residual para el embalse de Cala

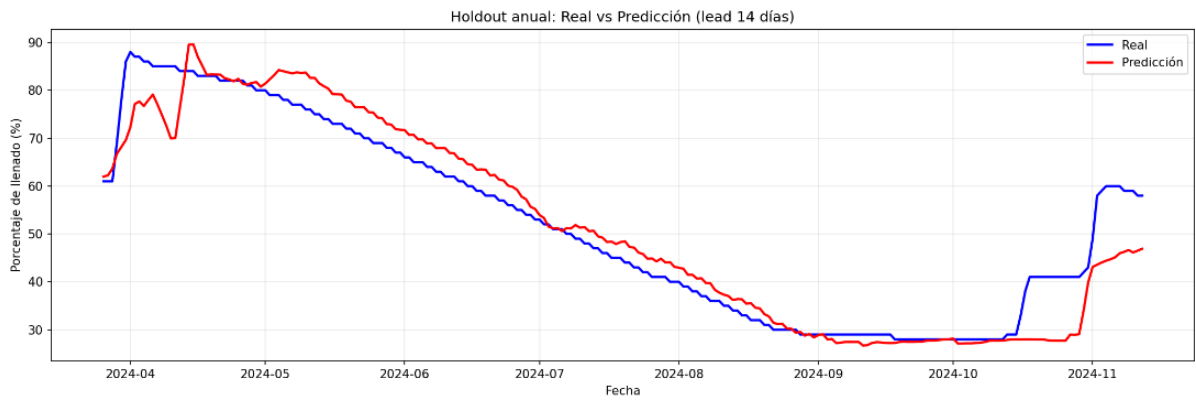


Figura 78: Evolución del porcentaje de llenado en un año en modo residual para el embalse de Cala

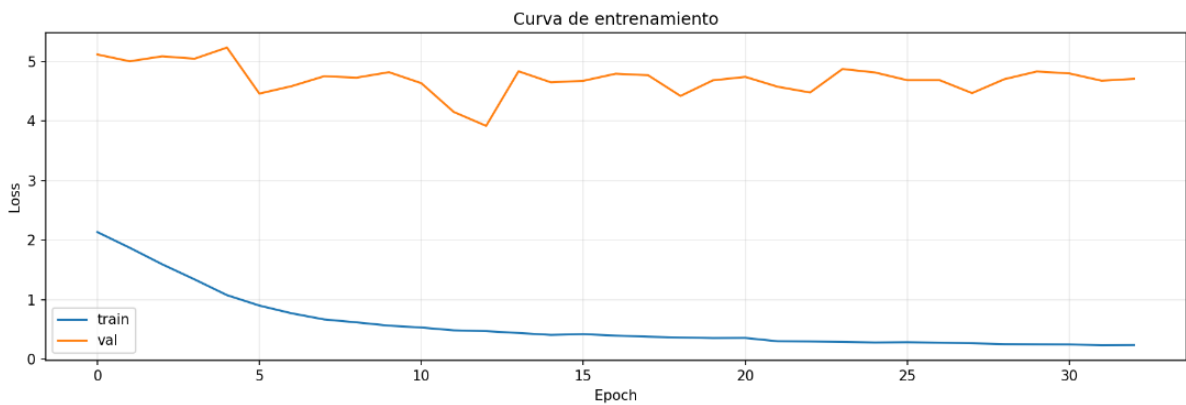


Figura 79: Curva de entrenamiento del modelo predictivo en modo residual para el embalse de Cala

Residual MAE 4.375 / RMSE 5.997 /  $R^2$  0.907; Directo MAE 4.554 / RMSE 7.127 /  $R^2$  0.869; persist 0.995.

- **Modelo residual para embalse de Canales**

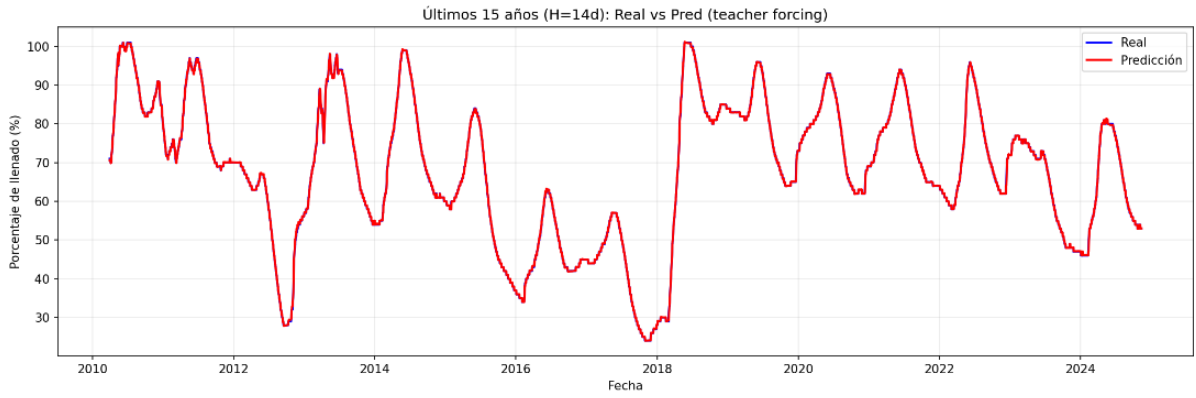


Figura 80: Evolución del porcentaje de llenado en los últimos 15 años en modo residual para el embalse de Canales

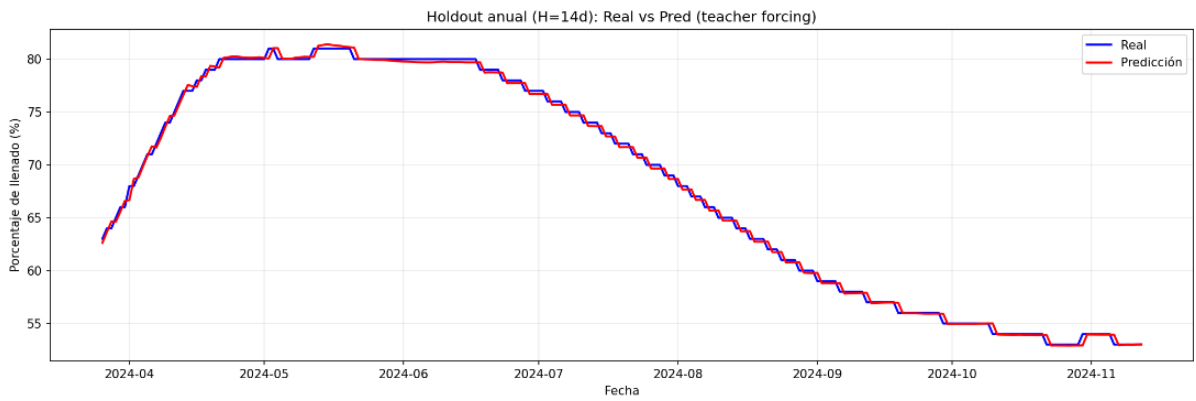


Figura 81: Evolución del porcentaje de llenado en un año en modo residual para el embalse de Canales

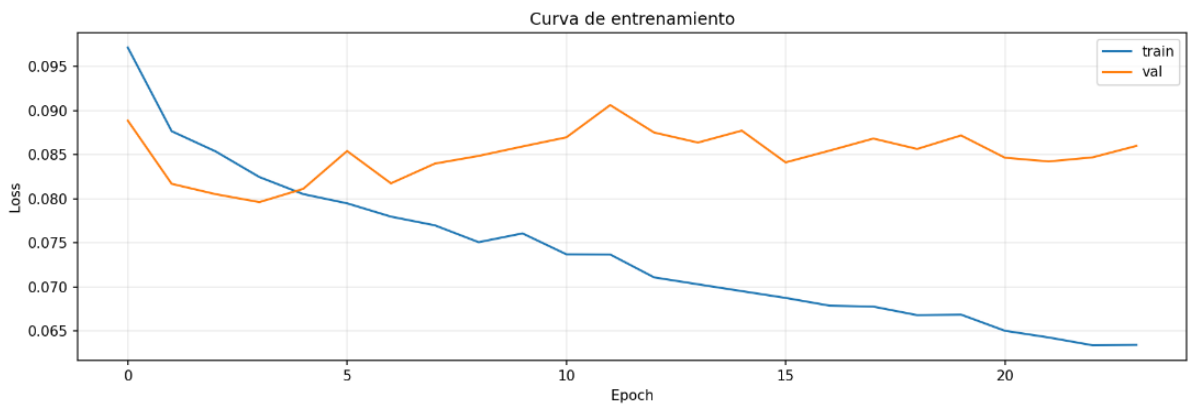


Figura 82: Curva de entrenamiento del modelo predictivo en modo residual para el embalse de Canales

Residual MAE 0.304 / RMSE 0.400 / MAPE 0.44 % /  $R^2$  0.999; Directo  $R^2$  0.821 (inferior a climatología); persist 0.998.

■ **Modelo residual para embalse de Iznájar**

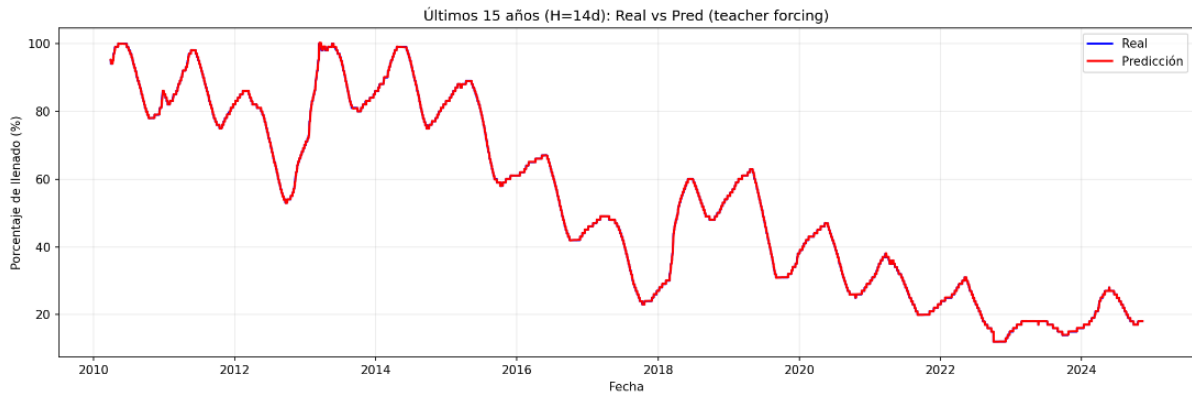


Figura 83: Evolución del porcentaje de llenado en los últimos 15 años en modo residual para el embalse de Iznájar

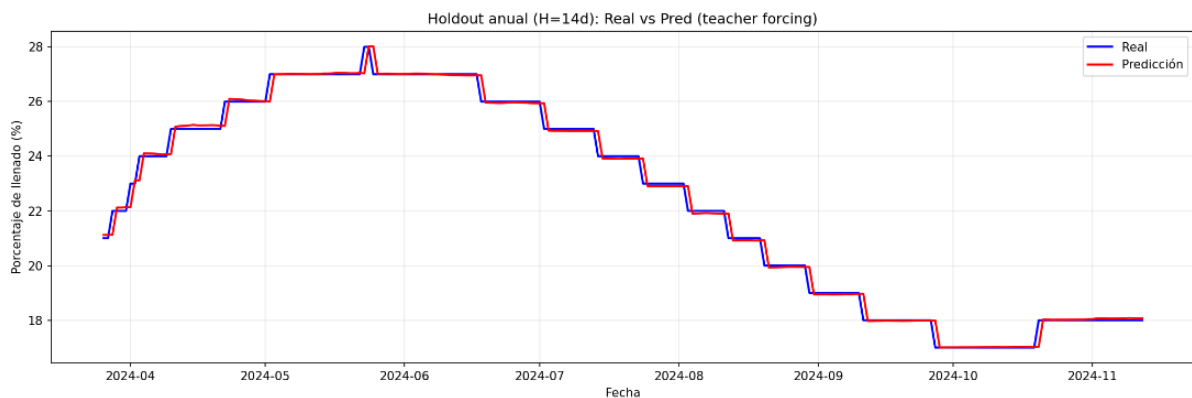


Figura 84: Evolución del porcentaje de llenado en un año en modo residual para el embalse de Iznájar

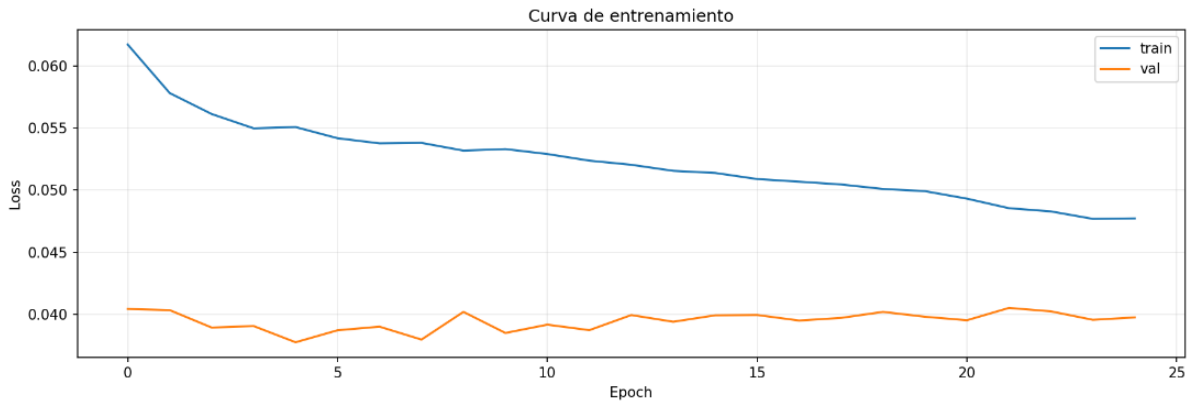


Figura 85: Curva de entrenamiento del modelo predictivo en modo residual para el embalse de Iznájar

Residual MAE 0.125 / RMSE 0.275 / MAPE 0.56 % /  $R^2$  0.995; Directo MAPE 7.00 % /  $R^2$  0.704; persist 0.994; skill residual +73.8.

■ **Modelo residual para embalse de San Rafael de Navallana**

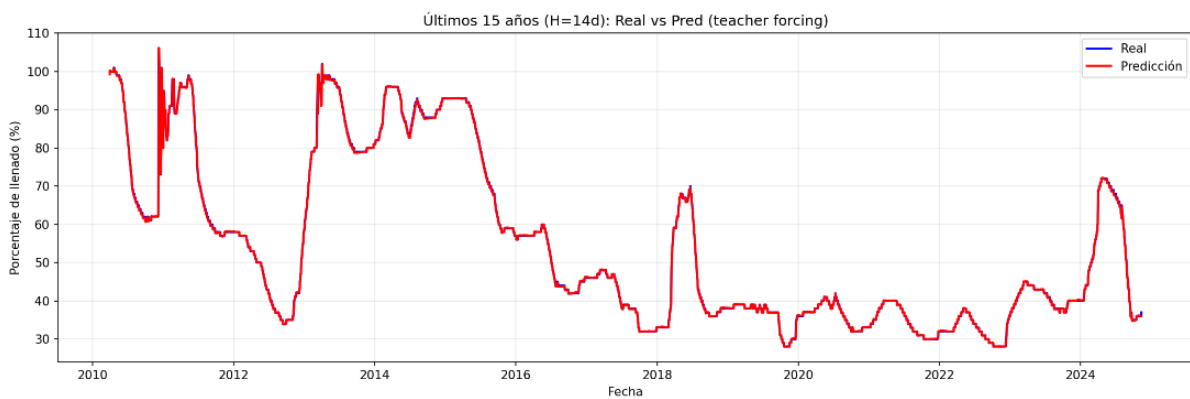


Figura 86: Evolución del porcentaje de llenado en los últimos 15 años en modo residual para el embalse de San Rafael de Navallana

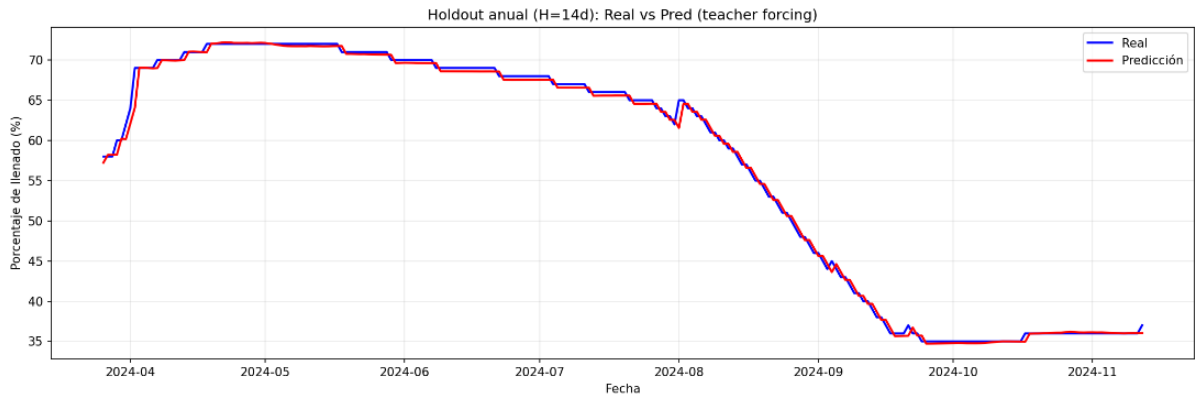


Figura 87: Evolución del porcentaje de llenado en un año en modo residual para el embalse de San Rafael de Navallana

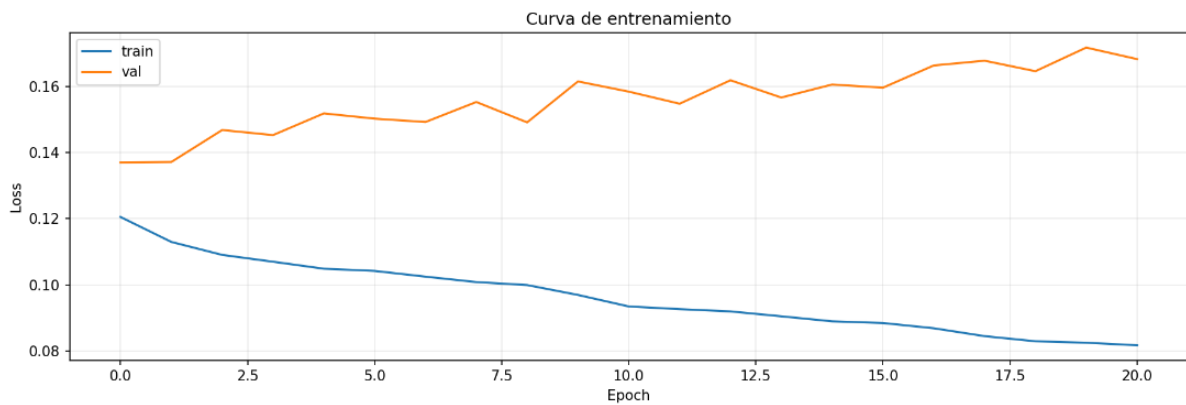


Figura 88: Curva de entrenamiento del modelo predictivo en modo residual para el embalse de San Rafael de Navallana

Residual MAE 0.403 / MAPE 0.73 % /  $R^2$  0.998; Directo MAE 5.094 /  $R^2$  0.844; persist 0.998.

- **Modelo residual para embalse de Andévalo**

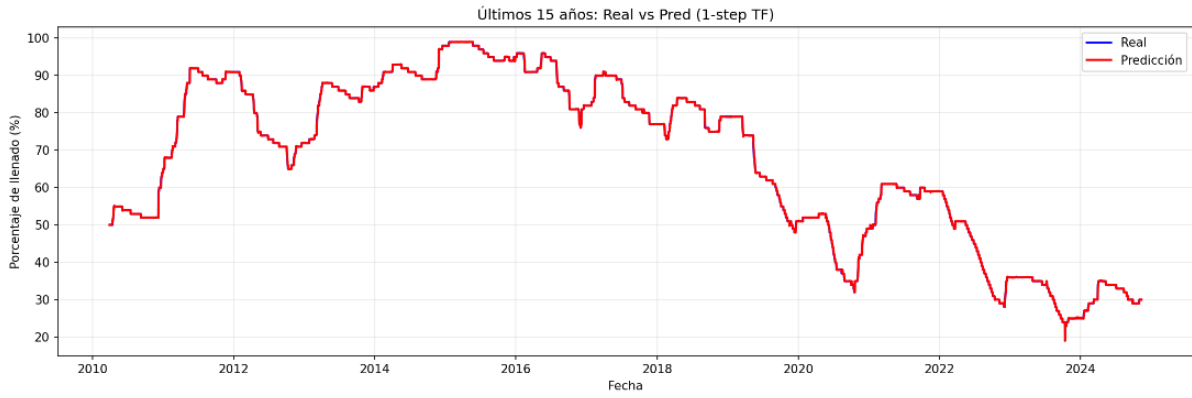


Figura 89: Evolución del porcentaje de llenado en los últimos 15 años en modo residual para el embalse de Andévalo

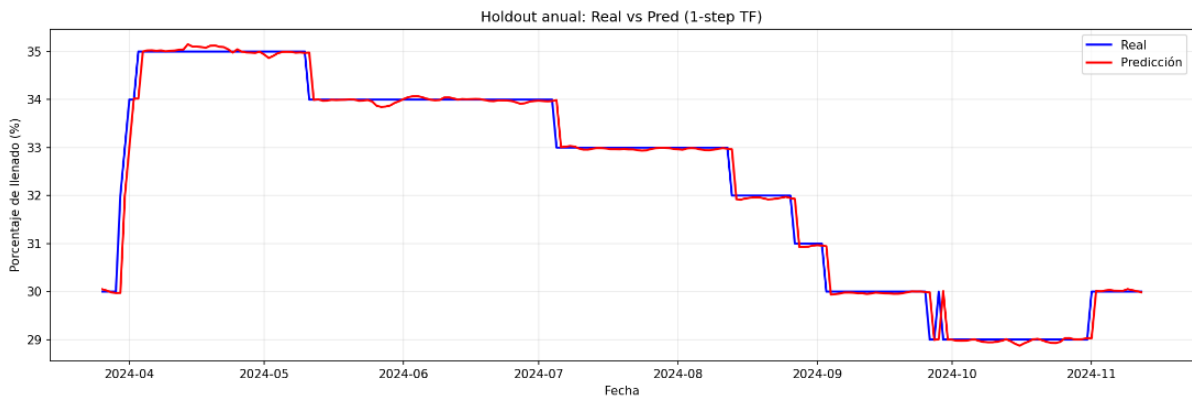


Figura 90: Evolución del porcentaje de llenado en un año en modo residual para el embalse de Andévalo

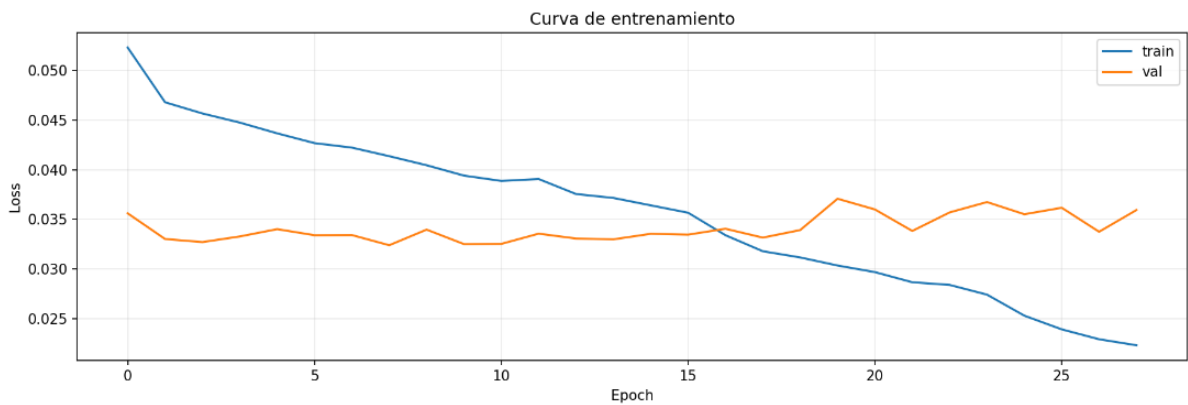


Figura 91: Curva de entrenamiento del modelo predictivo en modo residual para el embalse de Andévalo

Residual MAE 0.094 / RMSE 0.263 /  $R^2$  0.985; Directo  $R^2$  -1.720; persist 0.988.

■ **Modelo residual para embalse de Barbate**

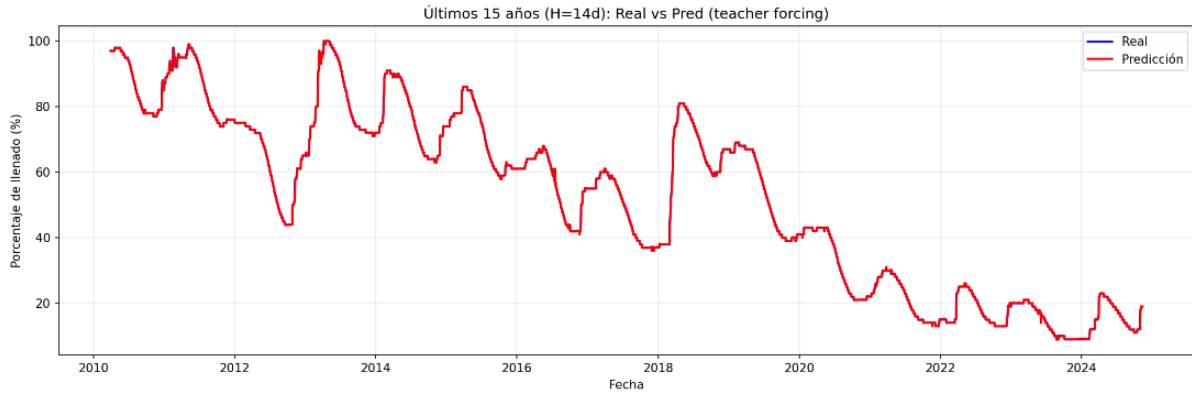


Figura 92: Evolución del porcentaje de llenado en los últimos 15 años en modo residual para el embalse de Barbate

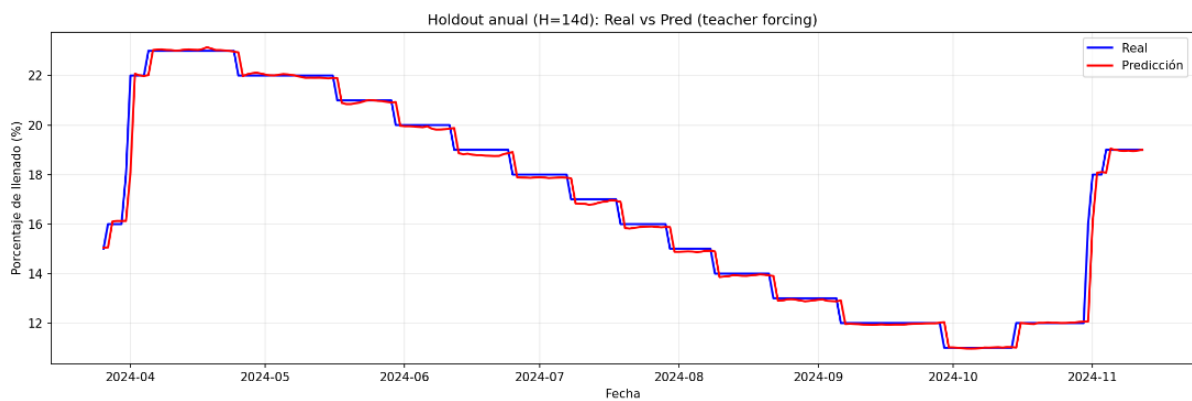


Figura 93: Evolución del porcentaje de llenado en en un año en modo residual para el embalse de Barbate

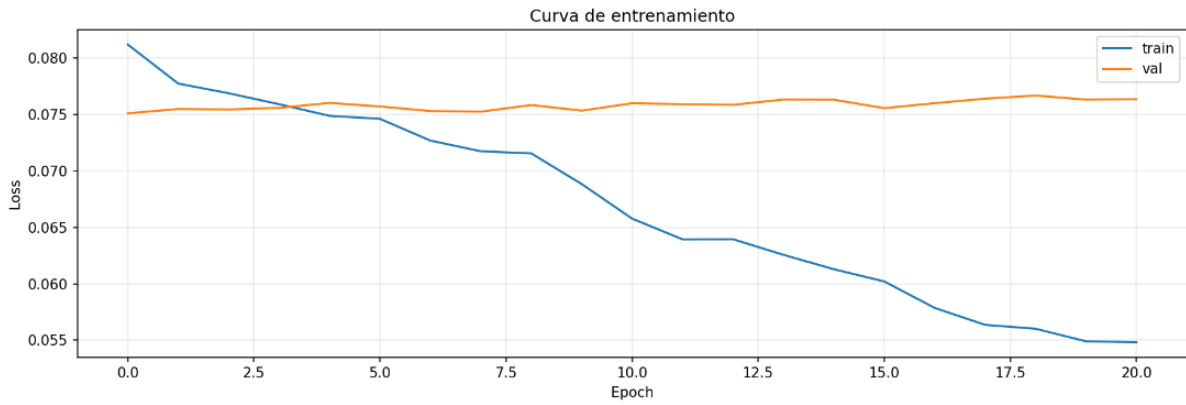


Figura 94: Curva de entrenamiento del modelo predictivo en modo residual para el embalse de Barbate

Residual MAE 0.184 / RMSE 0.481 /  $R^2$  0.986; Directo  $R^2$  0.471; persist 0.985.

■ **Modelo residual para embalse de Guadalteba**

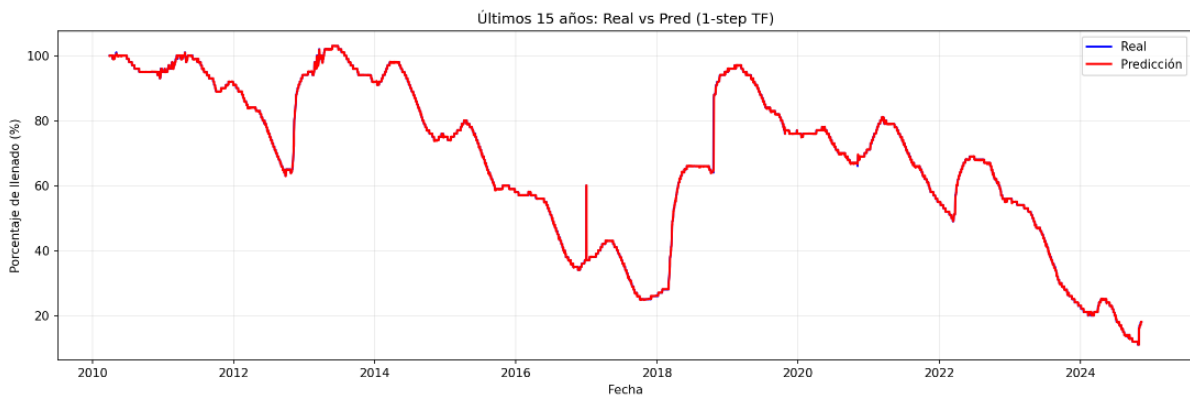


Figura 95: Evolución del porcentaje de llenado en los últimos 15 años en modo residual para el embalse de Guadalteba

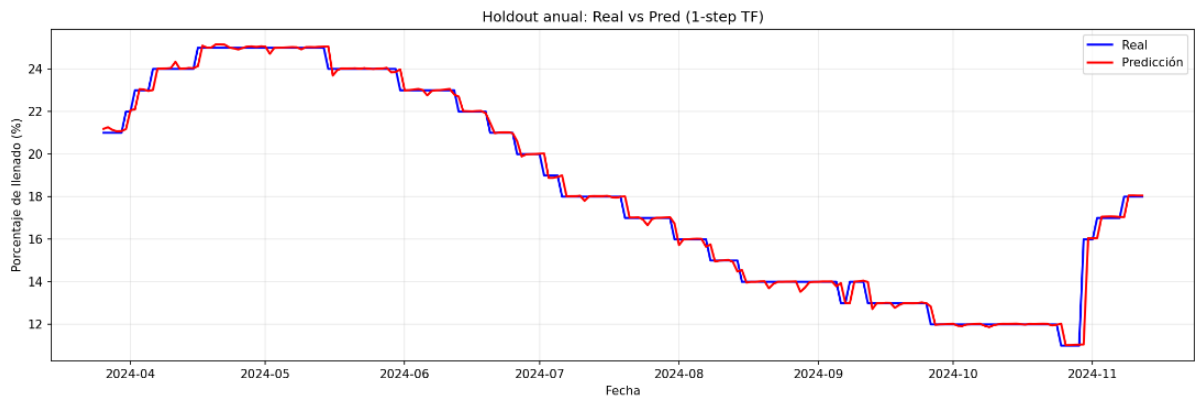


Figura 96: Evolución del porcentaje de llenado en un año en modo residual para el embalse de Guadalteba

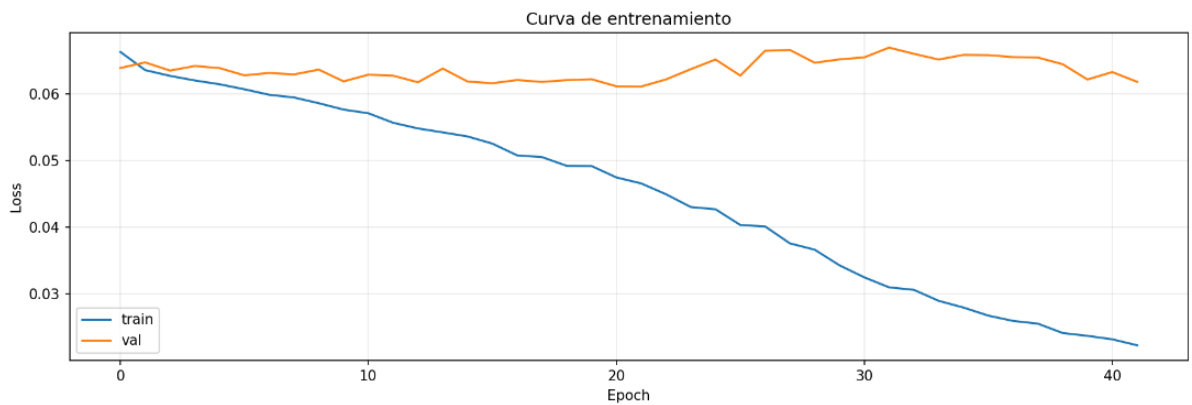


Figura 97: Curva de entrenamiento del modelo predictivo en modo residual para el embalse de Guadalteba

Residual MAE 0.155 / MAPE 0.91 % /  $R^2$  0.992; Directo MAPE 12.35 % /  $R^2$  0.632; persist 0.983.

- **Modelo residual para embalse de Guadalmena**

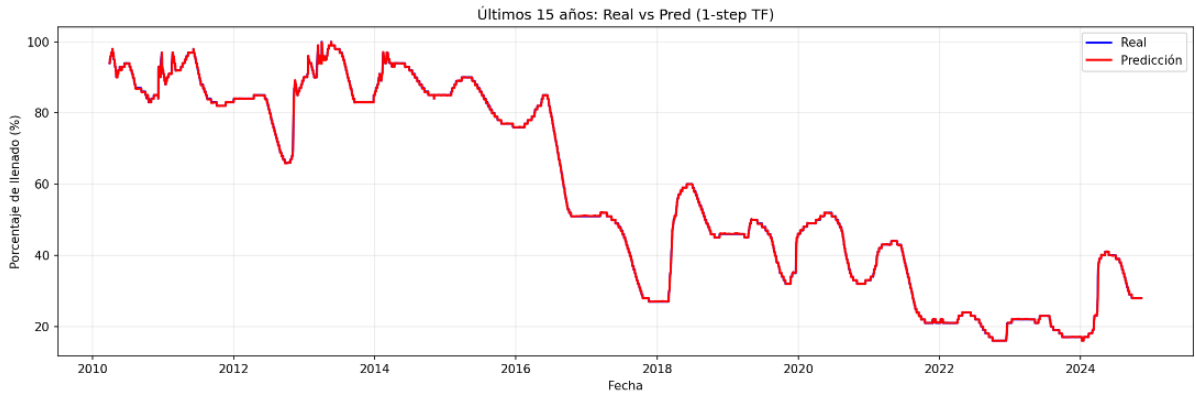


Figura 98: Evolución del porcentaje de llenado en los últimos 15 años en modo residual para el embalse de Guadalmena

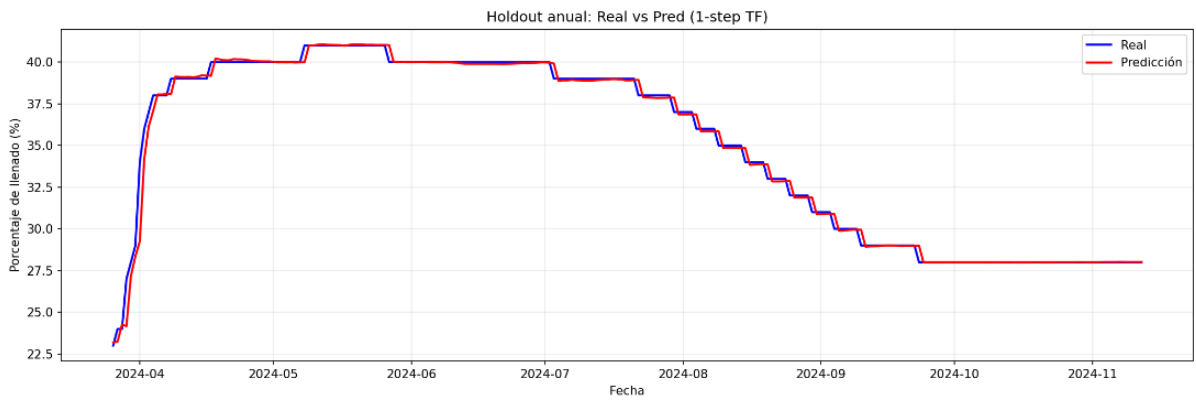


Figura 99: Evolución del porcentaje de llenado en un año en modo residual para el embalse de Guadalmena

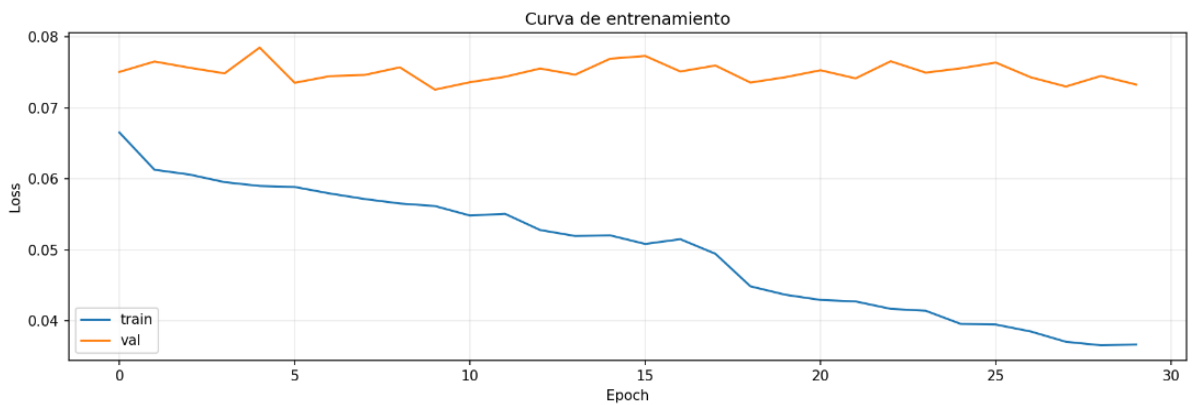


Figura 100: Curva de entrenamiento del modelo predictivo en modo residual para el embalse de Guadalmena

Residual MAE 0.177 / RMSE 0.472 /  $R^2$  0.992; Directo MAE 4.273 /  $R^2$  0.202; persist 0.997.

### ■ Modelo residual para embalse de Jándula

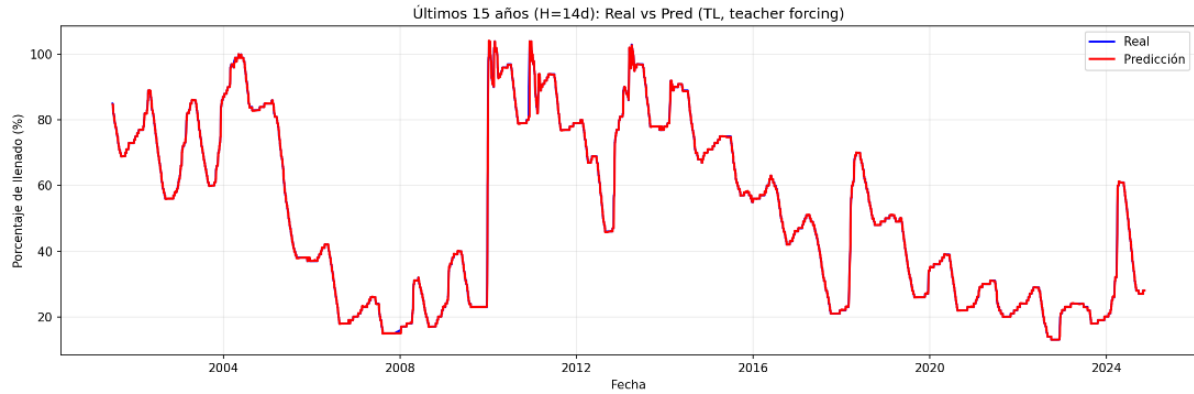


Figura 101: Evolución del porcentaje de llenado en los últimos 15 años en modo residual para el embalse de Jándula

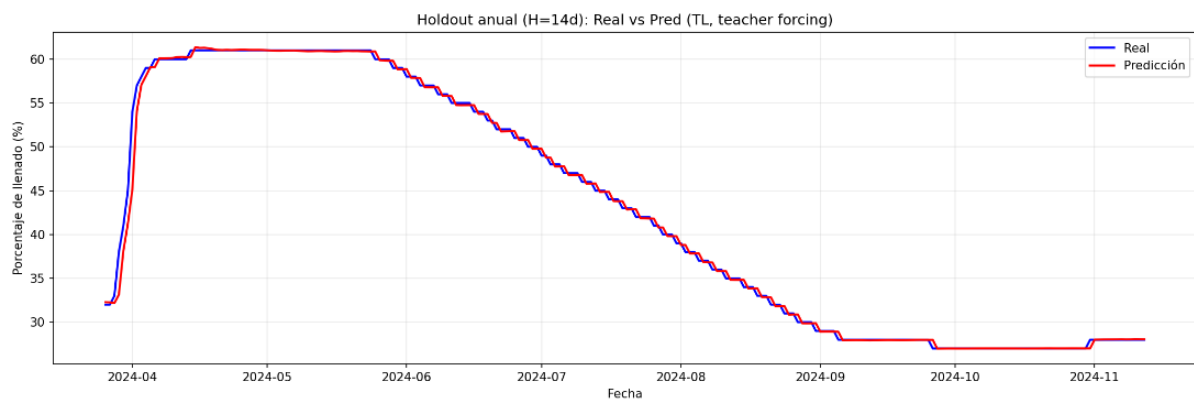


Figura 102: Evolución del porcentaje de llenado en un año en modo residual para el embalse de Jándula

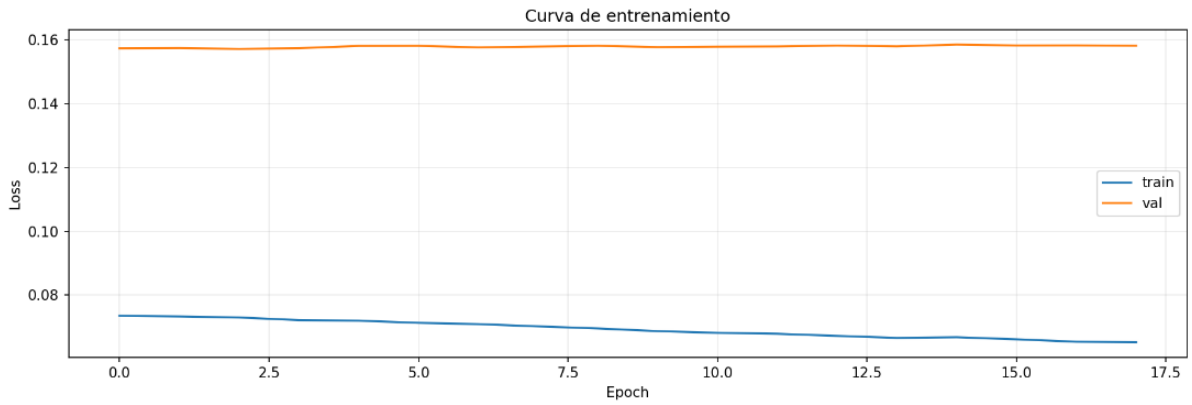


Figura 103: Curva de entrenamiento del modelo predictivo en modo residual para el embalse de Jándula

Residual MAE 0.330 / RMSE 0.843 /  $R^2$  0.996; Directo-transfer learning-82 MAE 1.725 / RMSE 1.946 /  $R^2$  0.970; persist 0.998.

■ **Modelo residual para embalse de Casasola**

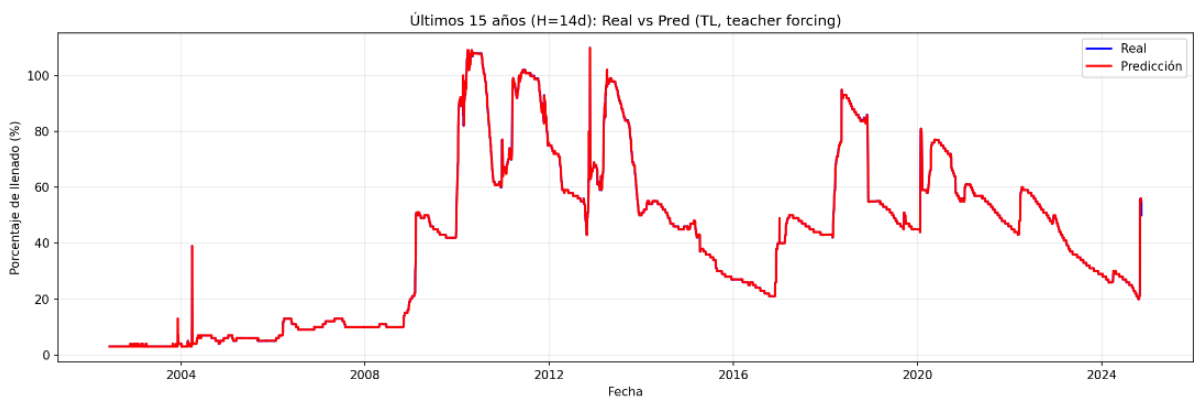


Figura 104: Evolución del porcentaje de llenado en los últimos 15 años en modo residual para el embalse de Casasola

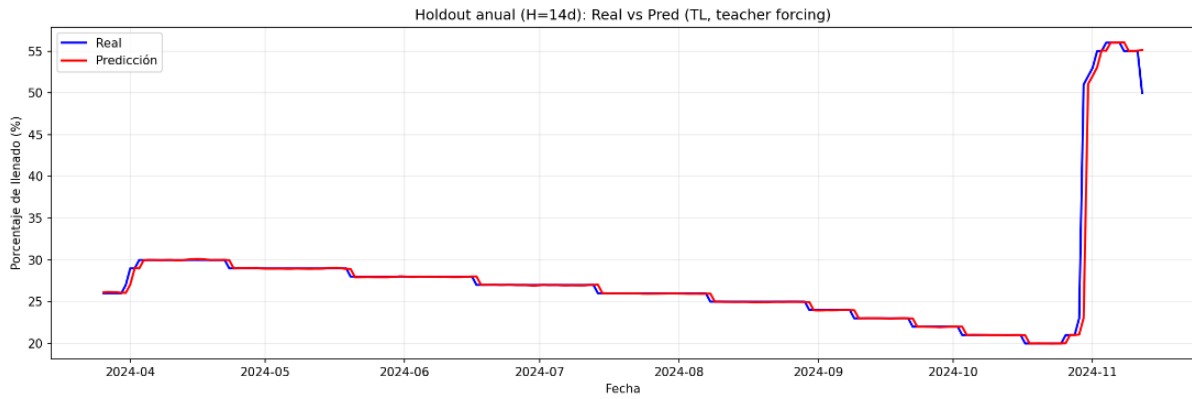


Figura 105: Evolución del porcentaje de llenado en un año en modo residual para el embalse de Casasola

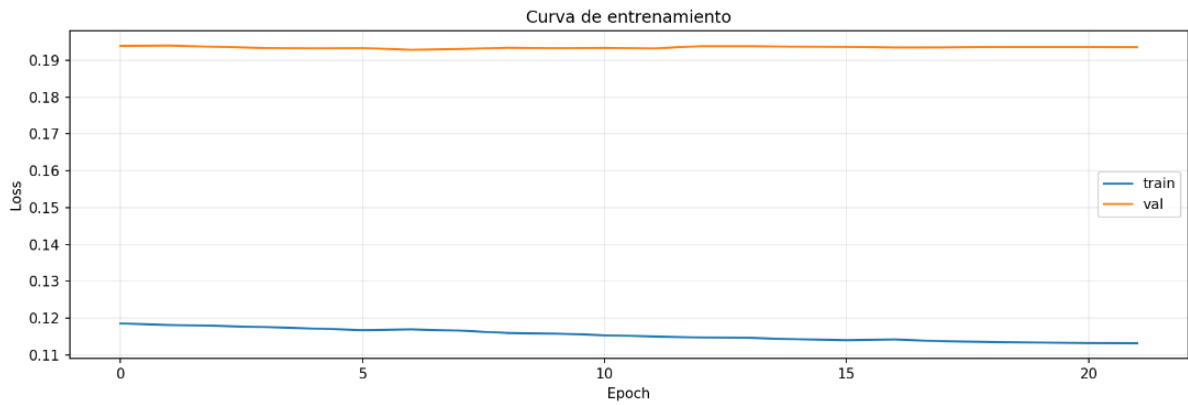


Figura 106: Curva de entrenamiento del modelo predictivo en modo residual para el embalse de Casasola

Residual  $R^2$  0.933 / MAPE 0.72 % ( $\approx$  persistencia  $R^2$  0.933); Directo-transfer learning-82  $R^2$  -0.297 / MAPE 9.44 %.

- **Modelo residual para embalse de Guadalhorce**

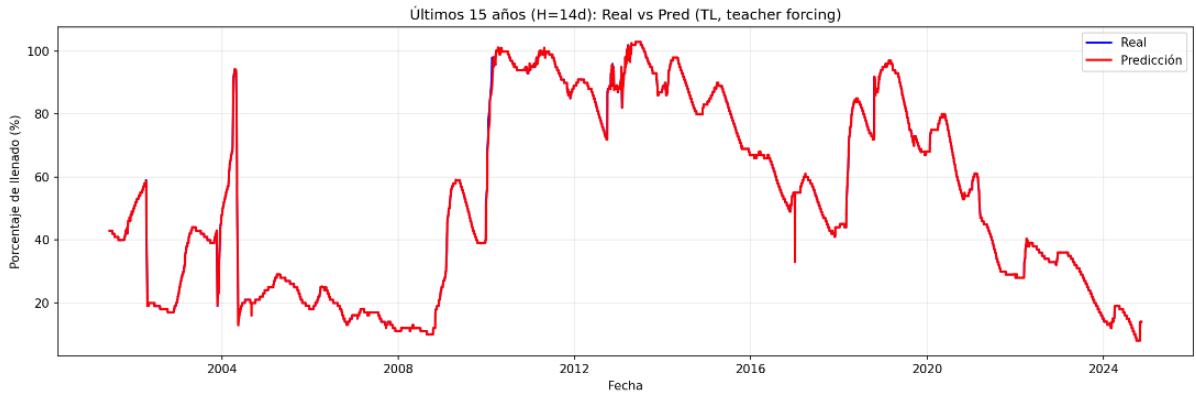


Figura 107: Evolución del porcentaje de llenado en los últimos 15 años en modo residual para el embalse de Guadalhorce

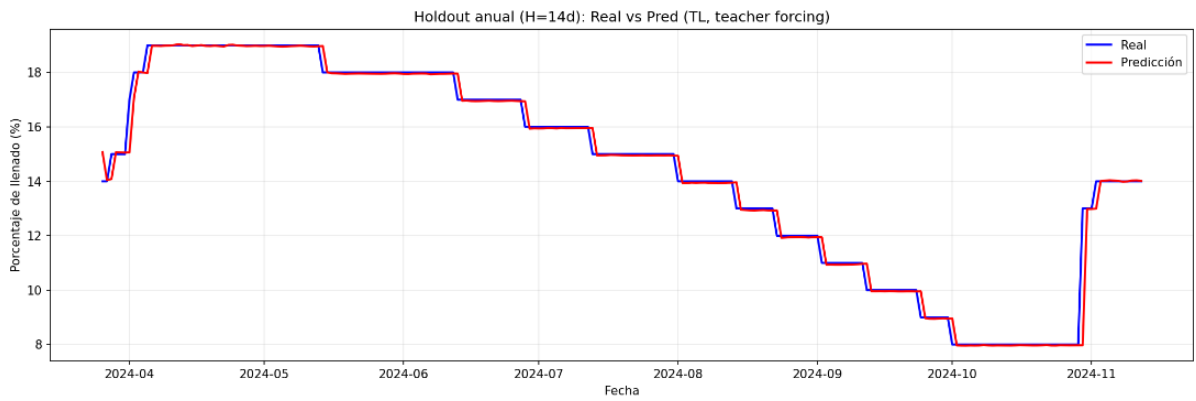


Figura 108: Evolución del porcentaje de llenado en un año en modo residual para el embalse de Guadalhorce

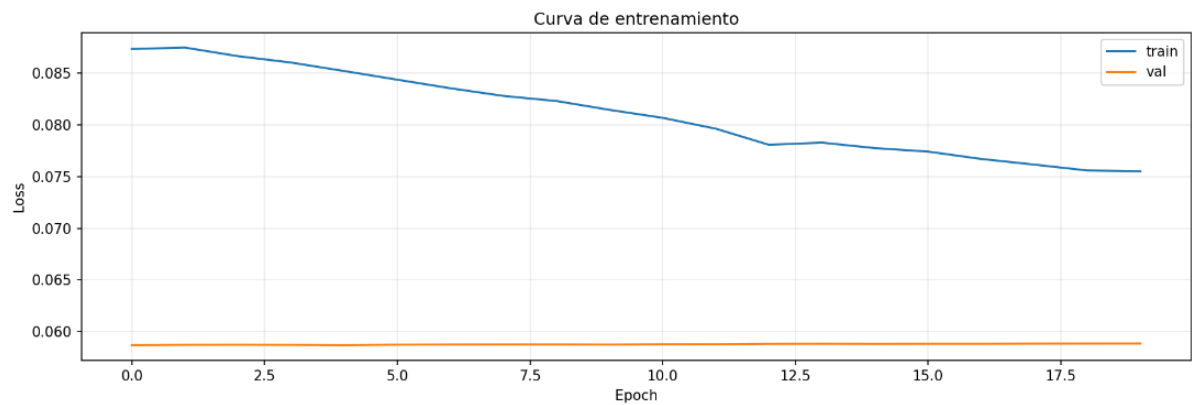


Figura 109: Curva de entrenamiento del modelo predictivo en modo residual para el embalse de Guadalhorce

Residual MAE 0.134 / RMSE 0.437 /  $R^2$  0.986; Directo-transfer learning-82  $R^2$  0.765 (errores un orden de magnitud mayores); persist 0.981.

■ **Modelo residual para embalse de Guadalcacín**

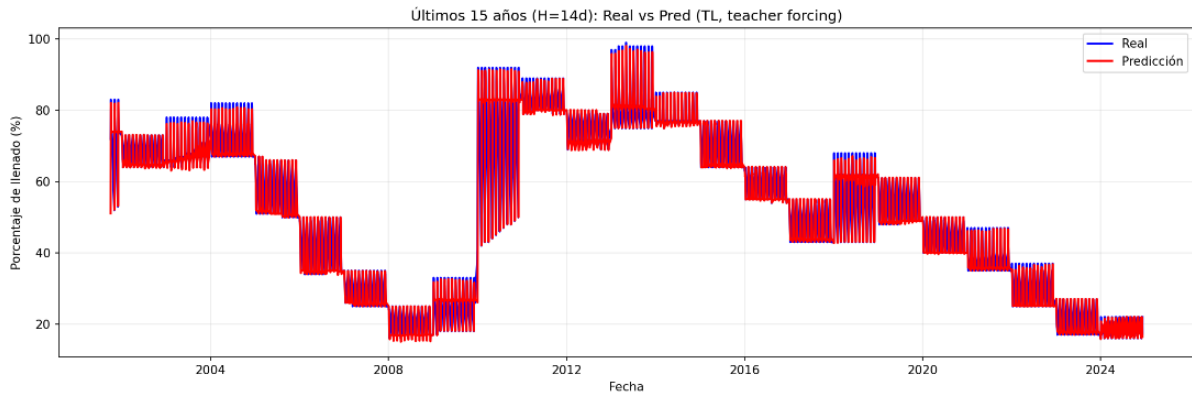


Figura 110: Evolución del porcentaje de llenado en los últimos 15 años en modo residual para el embalse de Guadalcacín

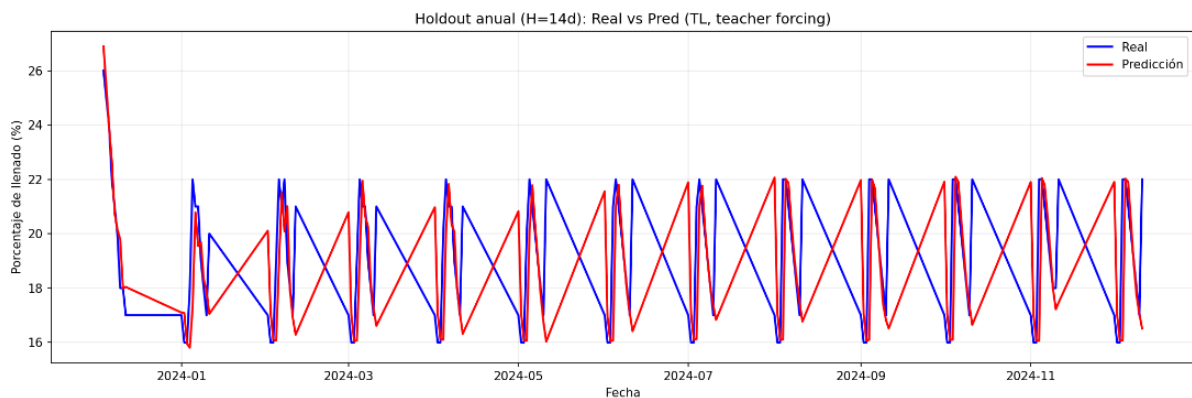


Figura 111: Evolución del porcentaje de llenado en un año en modo residual para el embalse de Guadalcacín

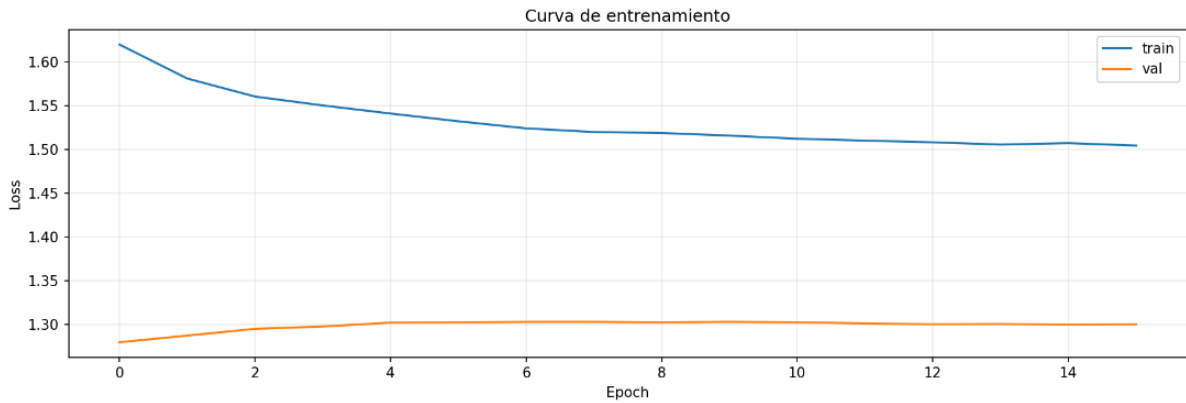


Figura 112: Curva de entrenamiento del modelo predictivo en modo residual para el embalse de Guadalcacín

Residual  $R^2$  -0.155 / MAPE 8.25 %; Directo- Transfer learning -82  $R^2$  0.793 / MAPE 13.06 %.

**Criterio final.** Aunque el residual domina en embalses muy inerciales, puede **inflar  $R^2$**  y volverse **frágil** en despliegues multi-embalse y transferencias; el **modo directo** aprende la serie futura y mostró **más estabilidad y realismo**, por lo que se adopta como **referencia operativa y para *transfer learning***, reservando **residual** para **casos puntuales** muy inerciales.

### Comparativa entrenamiento completo vs. *transfer learning* para el embalse de Jándula

En este punto, se ha entrenado el modelo predictivo correspondiente al **embalse de Jándula** al completo. De esta forma, podemos comparar los resultados obtenidos con el **modelo entrenado mediante *transfer learning***.

La configuración aplicada en este entrenamiento ha sido la siguiente: H=14, *steps*=180, modo directo, Huber, *early stopping*, *scheduler*, *z-score*.

- **Modelo independiente (82 *features*)**

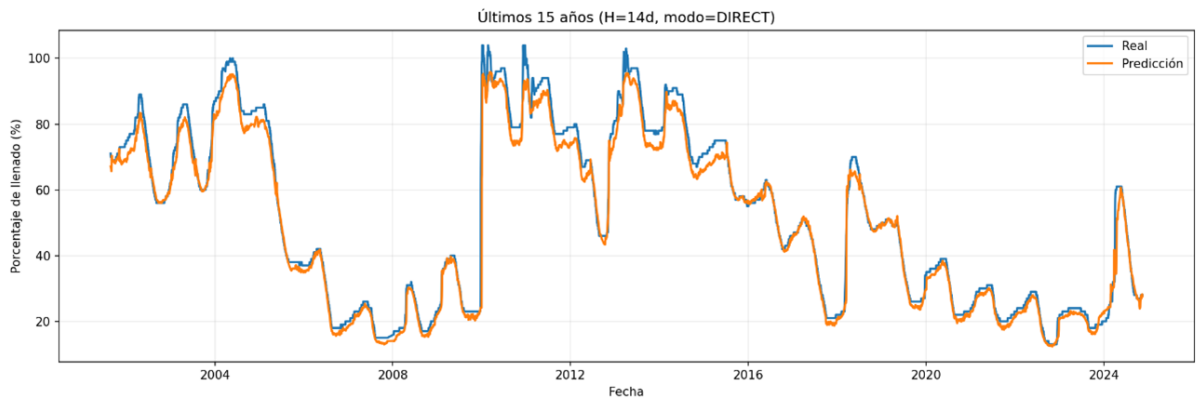


Figura 113: Evolución del porcentaje de llenado en los últimos 15 años para el embalse de Jándula (Entrenamiento completo del modelo predictivo)

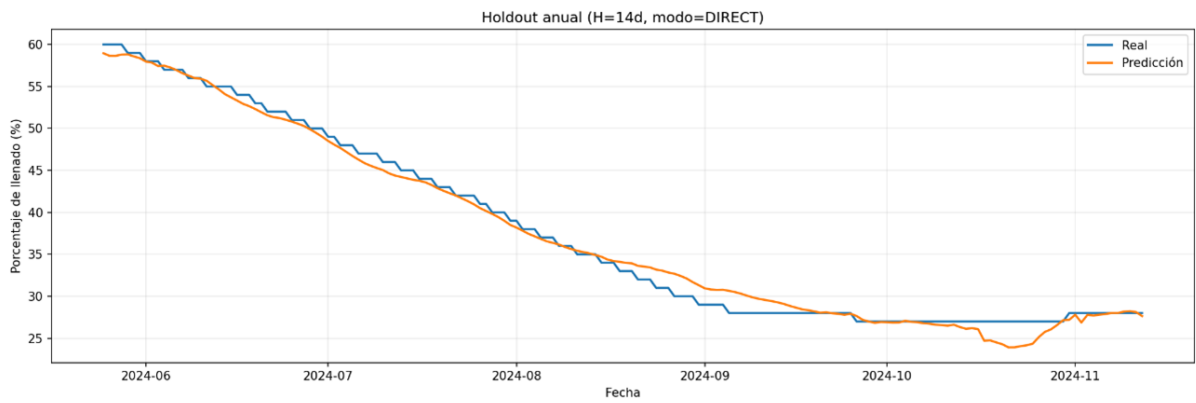


Figura 114: Evolución del porcentaje de llenado en un año para el embalse de Jándula (Entrenamiento completo del modelo predictivo)

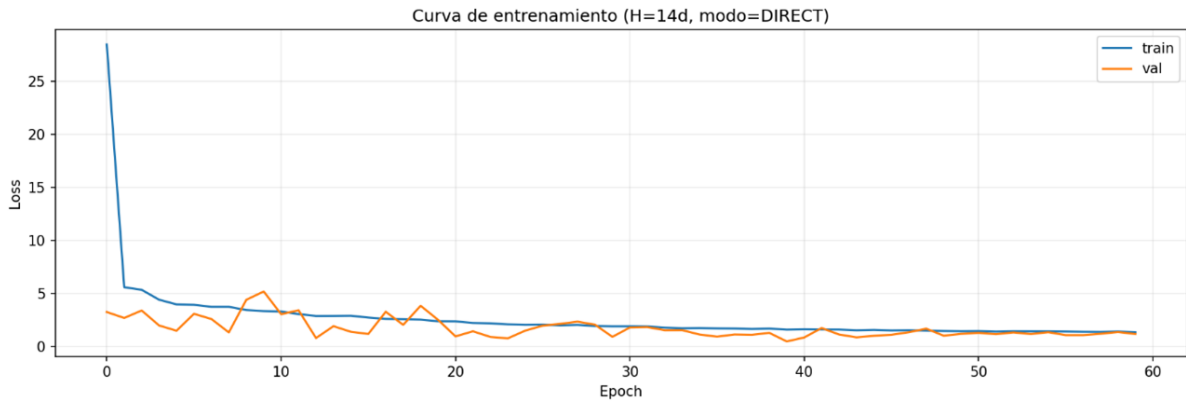


Figura 115: Curva de entrenamiento del modelo predictivo para el embalse de Jándula (Entrenamiento completo del modelo predictivo)

Configuración: horizonte **H=14**, ventana **steps=180**, esquema **directo**, pérdida **Huber**, normalización **z-score**, *early stopping* y *scheduler* activados; evaluación con **hold-out anual**. En prueba se obtuvieron **MAE = 0.852 p.p.**, **RMSE = 1.147 p.p.**, **MAPE = 2.57 %** y **R<sup>2</sup> = 0.989**. Para el mismo periodo, las referencias alcanzaron  $R_{persistente}^2 = 0.998$  y  $R_{climatología}^2 = -0.260$  ( $Skill_{climatología} = +1.250$ ). El **tiempo de entrenamiento ≈ 4390.1 s**. Los historiales de entrenamiento mostraron descenso sostenido de la *loss* en *train* y *val* hasta el punto de parada, sin oscilaciones marcadas.

- **Entrenamiento básico del Modelo Predictivo mediante *transfer learning* (32 features desde Guadalmena)**

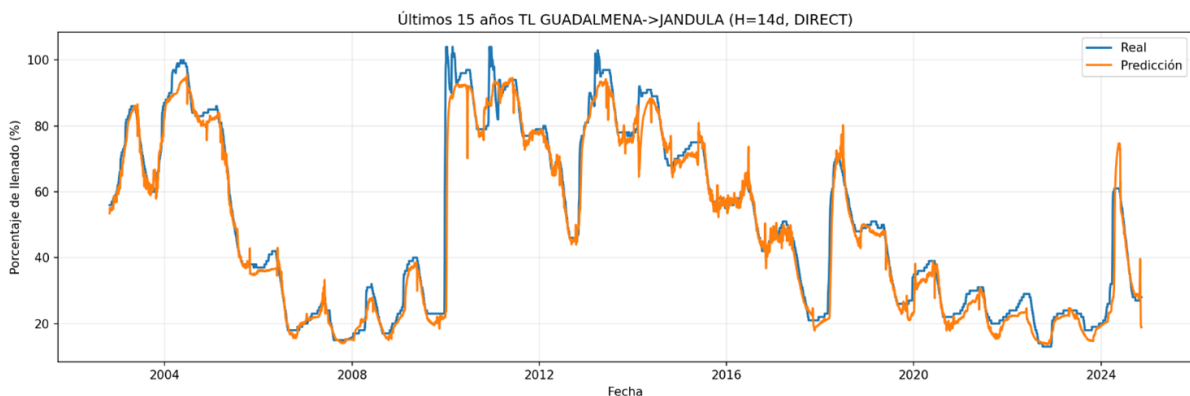


Figura 116: Evolución del porcentaje de llenado en los últimos 15 años para el embalse de Jándula (Entrenamiento por *transfer learning* – 32 features)

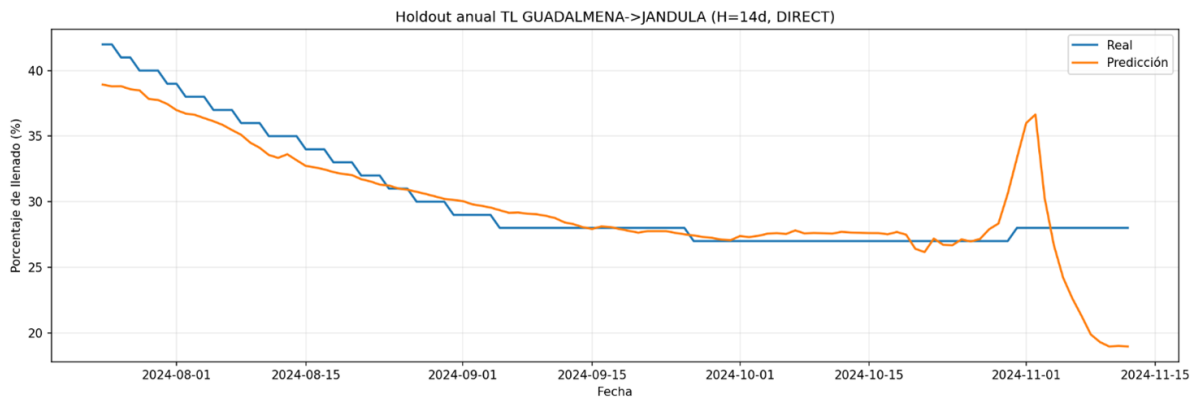


Figura 117: Evolución del porcentaje de llenado en un año para el embalse de Jándula (Entrenamiento por *transfer learning* – 32 *features*)

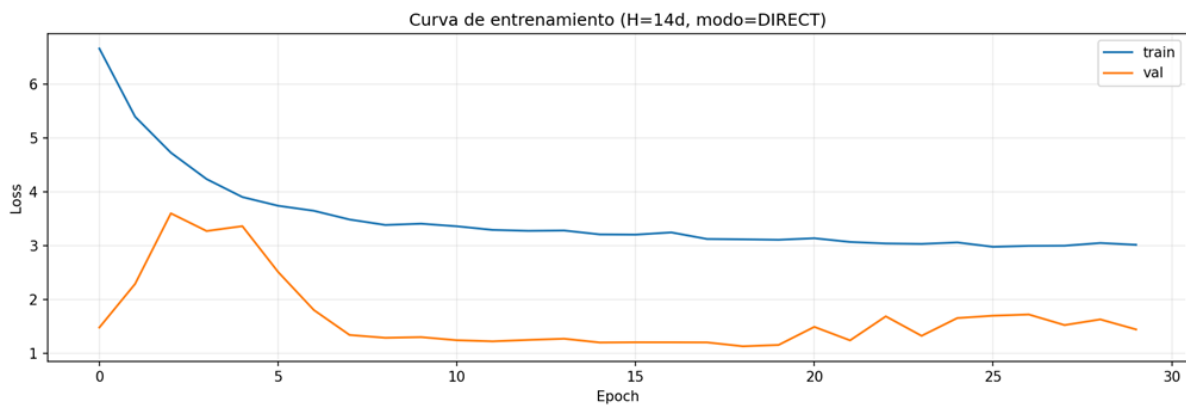


Figura 118: Curva de entrenamiento del modelo predictivo para el embalse de Jándula (Entrenamiento por *transfer learning* – 32 *features*)

Configuración: **misma ventana y horizonte** (steps=180, H=14), esquema **directo**, pérdida Huber, normalización *z-score*; inicialización con pesos del donante (**Guadalmena**), **warm-up** con bloque recurrente congelado y posterior *fine-tuning*. En prueba se registraron **MAE = 1.518 p.p.**, **RMSE = 2.624 p.p.**, **MAPE = 5.06 %** y **R<sup>2</sup> = 0.629**. El **tiempo de entrenamiento ≈ 4239.0 s**. Se conservaron los artefactos de entrenamiento (historial y curvas de pérdida) bajo el mismo esquema de validación temporal empleado en el resto de experimentos.

- **Entrenamiento extendido del Modelo Predictivo mediante *transfer learning* (82 *features* desde Guadalmena)**

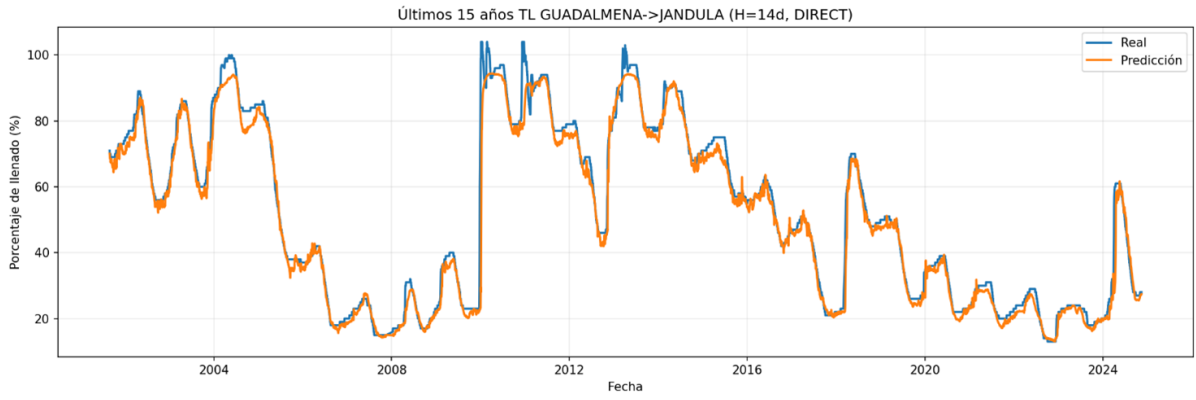


Figura 119: Evolución del porcentaje de llenado en los últimos 15 años para el embalse de Jándula (Entrenamiento por *transfer learning* - 82 features)

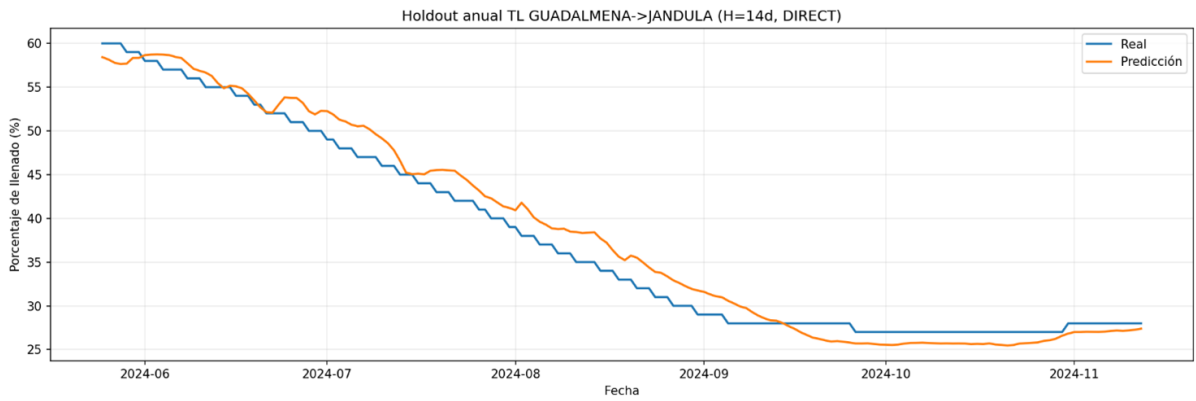


Figura 120: Evolución del porcentaje de llenado en un año para el embalse de Jándula (Entrenamiento por *transfer learning* - 82 features)

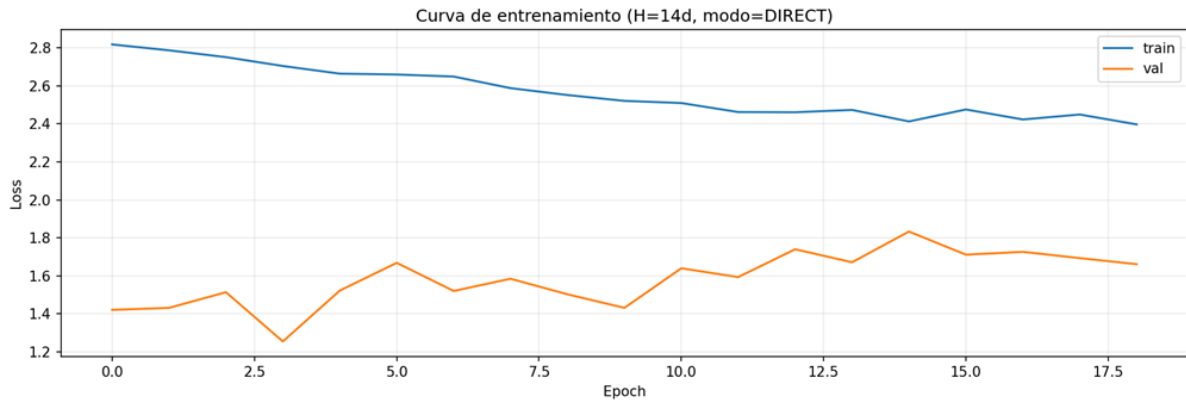


Figura 121: Curva de entrenamiento del modelo predictivo para el embalse de Jándula (Entrenamiento por *transfer learning* - 82 features)

Configuración: **steps=180**, **H=14**, esquema **directo**, pérdida **Huber**, normalización *z-score*; misma estrategia de inicialización (pesos del donante + *warm-up* con congelación) y *fine-tuning* con el **conjunto extendido**. Resultados en prueba: **MAE = 1.725 p.p.**, **RMSE = 1.946 p.p.**, **MAPE = 4.87 %** y  **$R^2 = 0.970$** . Para el mismo periodo, referencias  $R_{\text{persistente}}^2 = 0.998$  y  $R_{\text{climatología}}^2 = -0.260$  ( $Skill_{\text{climatología}} = +1.230$ ). El **tiempo de entrenamiento  $\approx 2357.3$  s**. Los historiales reflejaron convergencia estable y parada por mejora mínima en validación conforme a los criterios de *early stopping*.

Modelo	MAE (p.p.)	RMSE (p.p.)	MAPE (%)	$R^2$	Tiempo entrenamiento (s)
Modelo independiente	0,852	1,147	2,57 %	0,989	4390,1
<i>Transfer learning</i> Básico (32 features)	1,518	2,624	5,06 %	0,629	4239,0
<i>Transfer learning</i> Extendido (82 features)	1,725	1,946	4,87 %	0,970	2357,3

Tabla 5: Métricas de evaluación de entrenamiento para los distintos modelos predictivos analizados

### Experimentos fallidos

Con el objetivo de documentar con rigor los límites del enfoque en horizonte diario, se realizaron tres ensayos en el embalse de **Cala (H=1)** cuyos resultados se incorporan aquí de forma objetiva. El primero empleó **formulación directa** con una **ventana de 180 días** y un **conjunto amplio de exógenas**; los dos siguientes utilizaron **formulación residual** con **steps = 120**, primero con una entrada mínima (solo *lags* del nivel y estacionalidad) y después

añadiendo **precipitación, radiación y temperatura con sus retardos**. En todos los casos se comparó con **climatología y persistencia**, evaluando **MAE, RMSE, MAPE y  $R^2$** , y se revisaron las **curvas de entrenamiento/validación** y las gráficas **real vs. predicción**.

### ■ Experimento 1



Figura 122: Evolución del porcentaje de llenado en los últimos 15 años para el embalse de Cala (Entrenamiento directo y solo variables exógenas)



Figura 123: Evolución del porcentaje de llenado en un año para el embalse de Cala (Entrenamiento directo y solo variables exógenas)

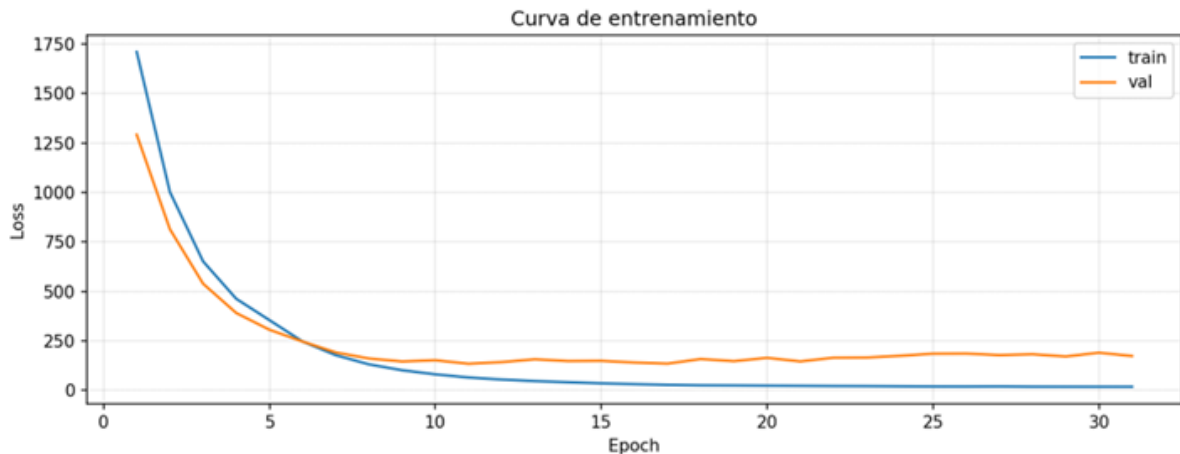


Figura 124: Curva de entrenamiento del modelo predictivo para el embalse de Cala (Entrenamiento directo y solo variables exógenas)

En el **Experimento 1 (directo, steps = 180, con exógenas)**, el modelo arrojó **MAE = 9.28 p.p.**, **RMSE = 11.57 p.p.**, **MAPE = 21.45 %** y  **$R^2 = 0.488$** , superando solo de forma marginal a la **climatología ( $R^2 = 0.534$ )** y quedando muy por debajo de la **persistencia ( $R^2 = 0.994$ )** con un tiempo de entrenamiento aproximado de 1651 s. Las curvas mostraron **mínimo de validación temprano y empeoramiento** tras 10–15 épocas (indicativo de **sobreajuste**), mientras que las gráficas evidenciaron **suavizado de picos** y una respuesta incapaz de reproducir cambios bruscos del nivel. En el informe se atribuye este comportamiento, de forma no prescriptiva, a la **dificultad intrínseca** de batir a la persistencia en diario, al **uso de MSE** (que tiende a **aplanar picos**) y a un posible **exceso de dimensionalidad** en la entrada (*lags* y *rolling* prolongados).

- **Experimento 2**



Figura 125: Evolución del porcentaje de llenado en los últimos 15 años para el embalse de Cala (Entrenamiento residual y solo lags)

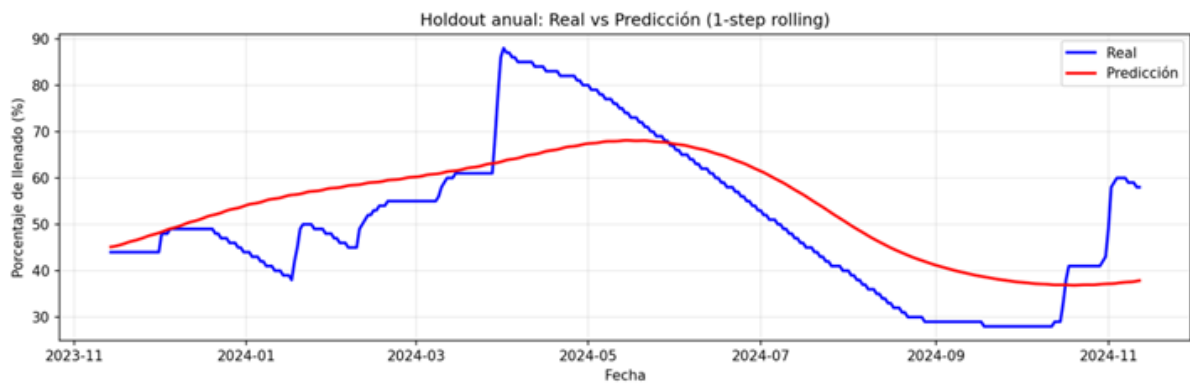


Figura 126: Evolución del porcentaje de llenado en un año para el embalse de Cala (Entrenamiento residual y solo lags)

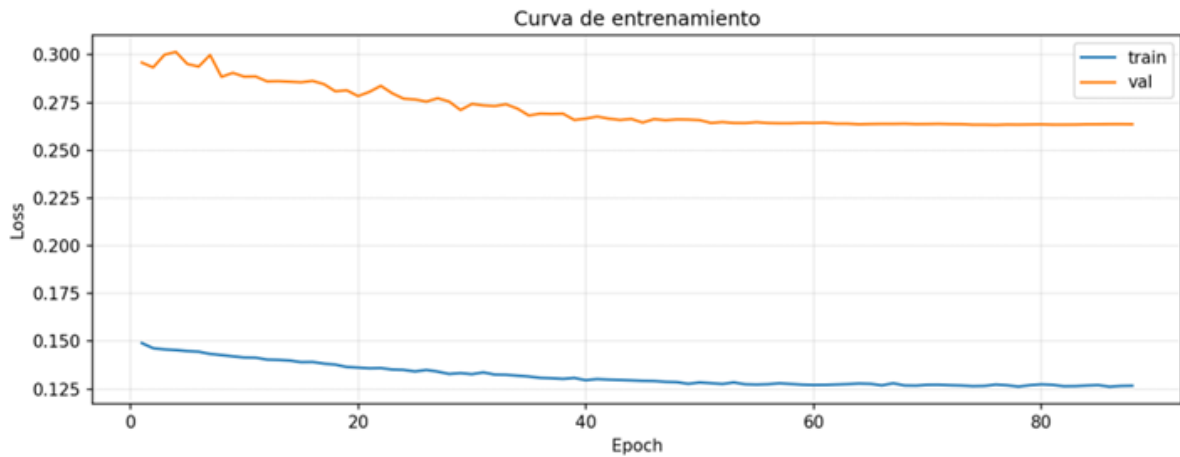


Figura 127: Curva de entrenamiento del modelo predictivo para el embalse de Cala (Entrenamiento residual y solo lags)

El Experimento 2 (residual, steps = 120, solo lags + estacionalidad) mejoró respecto a la climatología pero permaneció lejos de la persistencia:  $MAE = 8.70$  p.p.,  $RMSE = 10.30$  p.p.,  $MAPE = 19.03\%$ ,  $R^2 = 0.594$ , con  $Skill_{climatología} = +0.06$  y  $Skill_{persistente} \approx -0.40$  (la persistencia seguía en  $R^2 \approx 0.994$ ); el **entrenamiento** duró en torno a **2949 s**. La serie predicha adoptó un **patrón cuasi sinusoidal**, señal de que el modelo captó sobre todo la **estacionalidad** del año pero **sin capacidad** para anticipar eventos **abruptos por falta de exógenas**.

### ■ Experimento 3

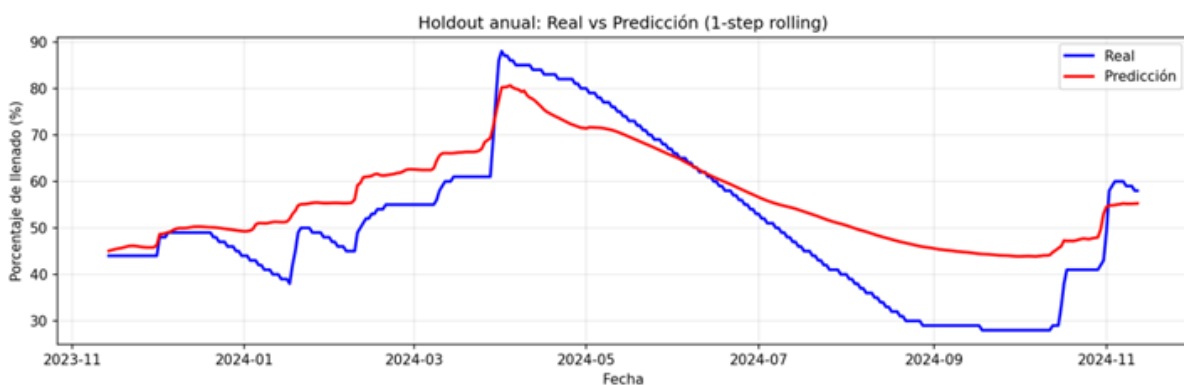


Figura 128: Evolución del porcentaje de llenado en los últimos 15 años para el embalse de Cala (Entrenamiento residual con variables exógenas y lags)



Figura 129: Evolución del porcentaje de llenado en un año para el embalse de Cala (Entrenamiento residual con variables exógenas y lags)

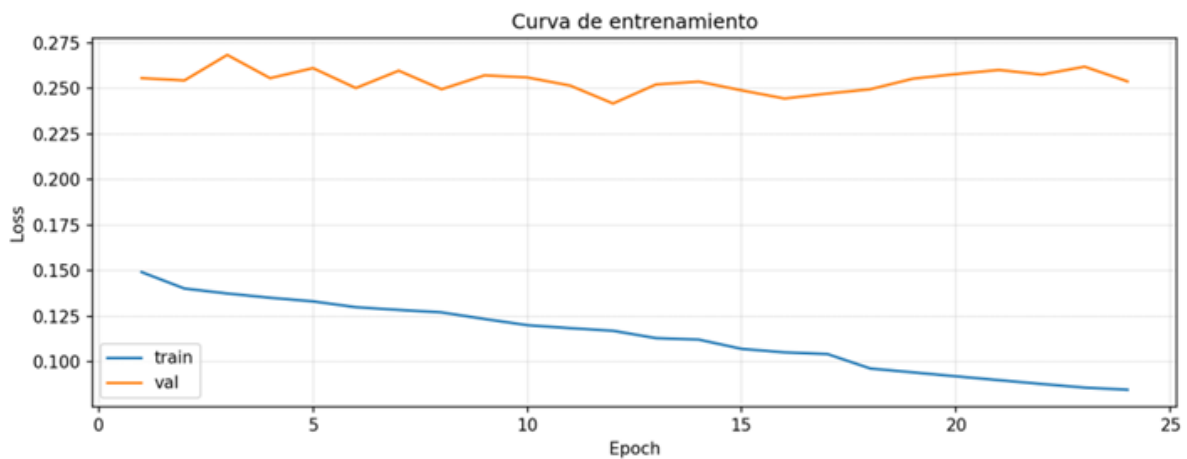


Figura 130: Curva de entrenamiento del modelo predictivo para el embalse de Cala (Entrenamiento residual con variables exógenas y lags)

En el **Experimento 3 (residual, steps = 120, con exógenas)**, la incorporación de **precipitación, radiación y temperatura con sus retardos** elevó el desempeño hasta  $R^2 = 0.688$  y un  $Skill_{climatología} = +0.15$ , si bien el **MAPE se mantuvo  $\approx 19\%$**  y la **persistencia** continuó **inalcanzable ( $R^2 \approx 0.994$ )**. Se observaron **desfases temporales** en picos y descargas, y además se detectó en la representación de **15 años una deriva negativa** debida a un **error de acumulación** en el modo residual cuando se encadenaban predicciones; este problema quedó **corregido** al forzar **teacher forcing** en la generación larga (alimentando con el **último valor real** en lugar de la predicción previa). El **tiempo de entrenamiento fue  $\approx 981$  s** por parada temprana.

## 3.4. Discusión general

### 3.4.1. Selección de características

La lectura conjunta de los análisis confirma que la relación instantánea entre meteorología y volumen embalsado es tenue, en línea con la fuerte inercia propia de un sistema de almacenamiento. Cuando se consideran desfases temporales, emergen patrones retardados coherentes (especialmente vinculados a temperatura, humedad y precipitación) que refuerzan la conveniencia de representar la señal con retardos explícitos y agregados móviles. En paralelo, se observa una colinealidad marcada dentro de cada familia de variables (p. ej., distintos indicadores de temperatura o de humedad), lo que aconseja una selección parca que escoja representantes claros por familia y reserve variables adicionales sólo cuando aporten información diferencial.

Los ensayos de importancia convergen en priorizar la memoria de corto plazo del propio sistema y los forzamientos hidrometeorológicos integrados, mientras que otros predictores muestran contribuciones menos estables. A partir de esta evidencia se perfila un diseño de *features* sobrio y causal: incluir retardos breves en las meteorológicas y en el objetivo, incorporar agregados multiescala donde la dinámica lo requiera y apoyarse en ventanas secuenciales suficientemente largas para captar respuestas más lentas sin inflar la dimensionalidad. Todo ello, acompañado de un filtrado previo que elimine variables de baja calidad o redundantes, ofrece una representación más limpia y alineada con el comportamiento físico del sistema.

### 3.4.2. Clustering

La comparación entre configuraciones muestra un patrón claro: **K=4** ofrece el mejor compromiso entre calidad métrica e interpretación hidrológica. Frente a **K=3**, los incrementos en **Silhouette** y **Calinski–Harabasz** y el descenso del **Davies–Bouldin** indican que cuatro grupos capturan la heterogeneidad sin caer en agregaciones demasiado amplias; frente a **K=5**, evita la **sobre-segmentación** en micro-clústeres poco representativos y con mayor solapamiento. La **distribución del Silhouette** (media  $\approx 0.24$ ) es coherente con un problema de alta dimensionalidad y ruido hidrometeorológico: separación moderada, fronteras próximas en algunos casos y un subconjunto de embalses muy bien definidos.

La lectura hidro-climática de los clústeres valida la partición:

- **C0** agrupa **embalses grandes y estables**, con variaciones suaves y alta persistencia.
- **C1** reúne **sistemas más pequeños y variables**, sensibles a cambios meteorológicos de corto plazo.
- **C2** concentra embalses en contextos **áridos/semiáridos**, con **baja precipitación, alta radiación/temperatura** y niveles medios más bajos.
- **C3** identifica embalses **con alto porcentaje medio de llenado y dinámica asociada a acumulados de precipitación**.

La **selección automática de másters** aporta trazabilidad y robustez al flujo de *transfer learning*: la propuesta inicial (Guadalmena–C0, Andévalo–C1, Guadalteba–C2, Breña–C3) se ajustó tras el control de estabilidad reciente, sustituyéndose **Breña por Barbate en C3**. Esta corrección mantiene el criterio de que los másters sean **representativos, estables y con buen skill** frente a referencias simples, condición necesaria para transferencias fiables.

El tratamiento de **outliers** (San Rafael de Navallana, Cala, Canales, Iznájar) fuera del agrupamiento principal evita que dinámicas singulares distorsionen las fronteras y, en términos operativos, señala casos que **requieren modelos específicos** o transferencias puntuales cuidadosamente validadas.

En conjunto, adoptar **K=4** proporciona un **mapa de clústeres operativo** para transfer learning: prioriza transferencias **intra-clúster**, reduce costes de entrenamiento al apoyarse en másters representativos y establece un **marco reproducible y escalable**. Dado que la separación entre grupos es moderada, se recomienda **revisiones periódicas** (p. ej., anuales) ante posibles cambios de régimen (sequías prolongadas, alteraciones de operación) y la incorporación gradual de **nuevas covariables estáticas/dinámicas** que puedan afinar la segmentación sin perder coherencia física.

### 3.4.3. Redes neuronales

Se evaluaron varias familias de arquitecturas secuenciales. La decisión final se inclinó por *Gated Recurrent Units (GRU)* por su equilibrio entre capacidad de representación, estabilidad de entrenamiento y eficiencia computacional. Este compromiso resulta clave con **ventanas largas (120–180 días)**, **horizontes prácticos (7–14–30 días)** y la necesidad de **reentrenar/transferir** modelos a múltiples embalses.

1. **Memoria suficiente con menor complejidad.** Las GRU incorporan compuertas de **actualización y reseteo** que controlan eficazmente el flujo temporal y mitigan la desaparición/explosión del gradiente. Frente a LSTM, su arquitectura más compacta (menos parámetros) ofrece prestaciones comparables en dependencias de medio plazo. Con 15–25 años de datos diarios por embalse y *feature sets* extensos, esta parquedad se traduce en convergencias más rápidas, menor uso de memoria y menor propensión al sobreajuste al añadir exógenas.
2. **Robustez y reproducibilidad del entrenamiento.** Con **Adam**, **pérdida Huber** y **callbacks** (parada temprana y reducción por meseta), las GRU mostraron curvas estables sin *tuning* fino por embalse. Esto reduce la prueba-error y facilita la estandarización del flujo: misma arquitectura, mismas ventanas y mismo preprocesado.
3. **Idoneidad para el aprendizaje por transferencia.** El menor número de parámetros favorece el *fine-tuning* entre embalses del mismo clúster: menor riesgo de *overfitting* al dominio origen, adaptación más rápida al destino y menor coste de cómputo. En la práctica, permite mantener un catálogo de másteres por clúster y actualizar los destinos con ciclos de *fine-tuning* acotados.
4. **Comparativa con alternativas.**
  - **LSTM.** Gran capacidad de memoria, pero con más compuertas y parámetros, y tiempos de entrenamiento/inferencia superiores. En nuestro caso, la penalización no aportó ganancias sustantivas frente a GRU con ventanas 120–180 días.
  - **TCN / 1D-CNN con dilataciones.** Competitivas y paralelizables, pero exigen diseñar con cuidado el **campo receptivo** (dilataciones y profundidad) para cubrir

retardos hidrológicos de 30–60 días. Además, su comportamiento fuera de dominio (transferencia entre embalses) resultó más sensible a cambios en la distribución temporal.

- **Transformers.** Muy potentes para secuencias largas con *multi-head attention*, pero su complejidad cuadrática y la necesidad de grandes volúmenes de datos los hacen menos eficientes aquí (horizontes corto-medios y actualización recurrente por embalse). El beneficio marginal no compensó el mayor coste ni la ingeniería adicional (enmascarados, *positional encodings*, regularización).

5. **Integración operativa.** GRU se integra de forma natural en el *pipeline*:

- funciona en **modo directo** (predicción del nivel futuro) y **modo residual** (corrección sobre persistencia),
- admite *teacher forcing* en validación para reconstrucciones *one-step* realistas
- y se despliega con facilidad en **TensorFlow/Keras**, lo que favorece reproducibilidad, automatización y mantenimiento.

En conclusión, **GRU** ofrece el mejor compromiso entre **calidad predictiva, coste de entrenamiento e interoperabilidad con el transfer learning** para nuestro caso de uso. Mantiene la memoria necesaria para capturar la dinámica a 7–14–30 días, reduce la complejidad frente a LSTM, evita la sobre-ingeniería que exigen *TCN/Transformers* y simplifica el ciclo completo: **entrenamiento, transferencia, validación y despliegue.**

#### 3.4.4. Entrenamiento de modelos predictivos

La evidencia acumulada en este trabajo muestra que, para la predicción quincenal del porcentaje de llenado de los embalses andaluces, el *transfer learning* es una palanca eficaz siempre que exista **afinidad físico-climática** entre donante y destino y se disponga de un **conjunto de features suficientemente rico**. Frente a entrenamientos desde cero, el *transfer learning* reduce típicamente entre **un quinto y un tercio** el tiempo/épocas, **estabiliza** la convergencia al partir de representaciones ya aprendidas y **mejora** especialmente en episodios no triviales (llenados o vaciados súbitos) cuando se incorporan variables meteorológicas y de memoria hidrológica. Operativamente, la pauta es **mantener un catálogo de másters**

**por clúster**, ejecutar un piloto corto para comprobar compatibilidad y, si la señal responde, **completar el *fine-tuning***; si no, **cambiar de donante** y/o **enriquecer las *features***. Este flujo es **reproducible y escalable**.

La **elección de horizonte** es clave. Aunque el *skill* frente a climatología tiende a empeorar según aumenta el plazo, **dos semanas (H=14)** resultan el mejor compromiso entre **utilidad y error**: permiten planificar maniobras de explotación, bombeos y comunicación con antelación suficiente; agregan información meteorológica y memoria hidrológica que el corto plazo apenas aprovecha; y evitan el salto de error típico del mes. Con el **conjunto enriquecido de *features*** (que amplía clima, *lags*, acumulados y estacionalidad) se refuerza la capacidad del modelo para **sostener predicciones quincenales**. Por todo ello, **H=14** se adopta como **estándar** tanto para entrenar másters y *outliers* como para transferir a embalses del mismo clúster, garantizando **comparabilidad y eficiencia**. En conjunto, estos tres ensayos muestran de manera consistente que, para **H=1**, la **persistencia** establece un **techo extremadamente alto ( $R^2 \approx 0.994$ )** que los modelos probados no superan, incluso tras enriquecer la entrada con exógenas. Además, el diseño de la pérdida (MSE en el primer caso) y la alta dimensionalidad contribuyeron a **suavizado de picos y sobreajuste/infraajuste** según la configuración. Por todo ello, y en coherencia con la evidencia empírica, se **descarta el horizonte** diario y se centra el análisis experimental en **H=7, 14 y 30 días**, intervalos en los que la persistencia pierde competitividad y los modelos (especialmente en **formulación residual** con *features* extendidas) ofrecen un **valor añadido tangible** para la predicción operativa.

En los **modelos entrenados desde cero**, los *outliers* ofrecen una lectura clara en contextos muy inerciales: **San Rafael de Navallana** y **Cala** logran **precisión operativa** con errores contenidos aunque, como es habitual a 14 días, la **persistencia** fija un techo alto difícil de superar; **Canales** destaca por **bajo error relativo** pese a quedar ligeramente por debajo de la climatología en varianza explicada; **Iznájar**, por su gran capacidad y variación lenta, presenta **errores absolutos pequeños** y un  $R^2$  razonable para ese régimen. Entre los másters, **Barbate** es **aceptable** como referencia de su clúster.

La **transferencia embalse-embalse** confirma el valor del enfoque cuando hay **compa-**

**tibilidad. Jándula** es paradigmático: con *features* extendidas el modelo **roza el límite impuesto por la persistencia**, con **errores muy bajos** y **coste de entrenamiento menor** que un *scratch* equivalente. **Guadalhorce** evidencia **incrementos claros en varianza explicada y descensos de error** respecto a su referencia, además de **ahorros de tiempo** relevantes. **Guadalcacín** ilustra otro patrón útil: incluso partiendo de un **donante débil**, la transferencia puede resultar en un **modelo operativo** si se **enriquece la entrada** en el destino. La excepción es **Casasola**: ni con 32 ni con 82 *features* se alcanzó la calidad mínima, señalando **incompatibilidad de dominio** y/o **regularización insuficiente**; en estos casos conviene **cambiar de donante, aumentar información o endurecer el fine-tuning**.

En el plano metodológico, la comparación entre **modo residual** (aprender la corrección sobre la persistencia) y **modo directo** (predecir el nivel futuro) aporta matices importantes. El **residual**, en la mayoría de embalses muy inerciales, ofrece **métricas excelentes** y a veces **igual a la persistencia**; sin embargo, puede **inflar  $R^2$**  al apoyarse demasiado en ella y volverse **frágil** cuando se amplían ventanas de validación, se despliega de forma multi-embalse o se transfiere entre dominios con residuales de distinta naturaleza. El **modo directo**, en cambio, **aprende la serie futura completa**, mostró **mayor estabilidad y realismo temporal**, y se adopta como **configuración de referencia para explotación y transfer learning**. El residual queda como **opción táctica** en casos muy inerciales y con datos homogéneos.

El caso **Jándula** sirve como lectura comparada para elegir estrategia: **entrenado desde cero** con *features* extendidas, alcanza **precisión casi perfecta** y **generaliza con solidez**, aunque con **coste de entrenamiento alto**; mediante **transfer learning extendido desde Guadalmena** se obtiene una **calidad cercana** con **menos coste** y, además, **mejor control de errores grandes**. Si se **reduce** el conjunto de *features*, la transferencia **pierde** capacidad para captar la dinámica interna. En consecuencia, si se persigue **precisión máxima**, el *scratch* rico en *features* es idóneo; si se busca **equilibrio precisión–coste**, el **transfer learning extendido desde un donante** competente es la **mejor alternativa** [14][15].

En conjunto, estas conclusiones se traducen en **pautas operativas** claras: **estandarizar H=14 y el modo directo**; aplicar **transfer learning** siempre que exista **afinidad físico-**

**climática**, priorizando **donantes sólidos** y **features extendidas**; desplegar con la cadena **catálogo de másters - fine-tuning - verificación** frente a persistencia/climatología, documentando siempre el **ahorro temporal**; gestionar excepciones (como Casasola) cambiando donante y reforzando señal/regularización; y comunicar límites cuando persistencia/climatología sean muy fuertes, centrando el valor del modelo en episodios no triviales y en la consistencia regional del sistema. Con este marco, el proyecto ofrece una vía robusta y escalable para implantar modelos predictivos de gestión sostenible del agua en los embalses de Andalucía.

# 4

## Desarrollo de la aplicación

### 4.1. Ingeniería del software

#### 4.1.1. Plataforma web

En la fase final del proyecto se implementó una aplicación web cuyo propósito es centralizar y facilitar el acceso a la información histórica y en tiempo real de los embalses y de las variables meteorológicas de Andalucía. La plataforma combina visualización en formato tabular con filtros avanzados y representaciones gráficas interactivas, lo que permite explorar la evolución de los datos en distintos intervalos temporales. La aplicación se compone de un *frontend* y un *backend* conectados a una base de datos de series temporales InfluxDB, que contiene aproximadamente 25 años de registros. Tanto la parte cliente como la parte servidor se desarrollaron en **JavaScript**, utilizando herramientas específicas para el manejo de datos temporales y para el despliegue de los servicios.

El objetivo es disponer de un entorno ágil, moderno y accesible que pueda ser utilizado por:

- Estudiantes e investigadores interesados en el análisis de recursos hídricos.
- Profesionales vinculados a la gestión de embalses y sistemas hidráulicos.
- Ciudadanos que quieran seguir de forma sencilla la evolución de los embalses y la meteorología regional.

En los apartados siguientes se describe con más detalle la arquitectura y las funcionalidades de la plataforma.

### 4.1.2. Estructura del proyecto

El sistema se organiza en cuatro capas bien diferenciadas (datos, modelado, servicio (API) y presentación) empaquetadas con Docker para facilitar el despliegue reproducible (véase Figura 131).

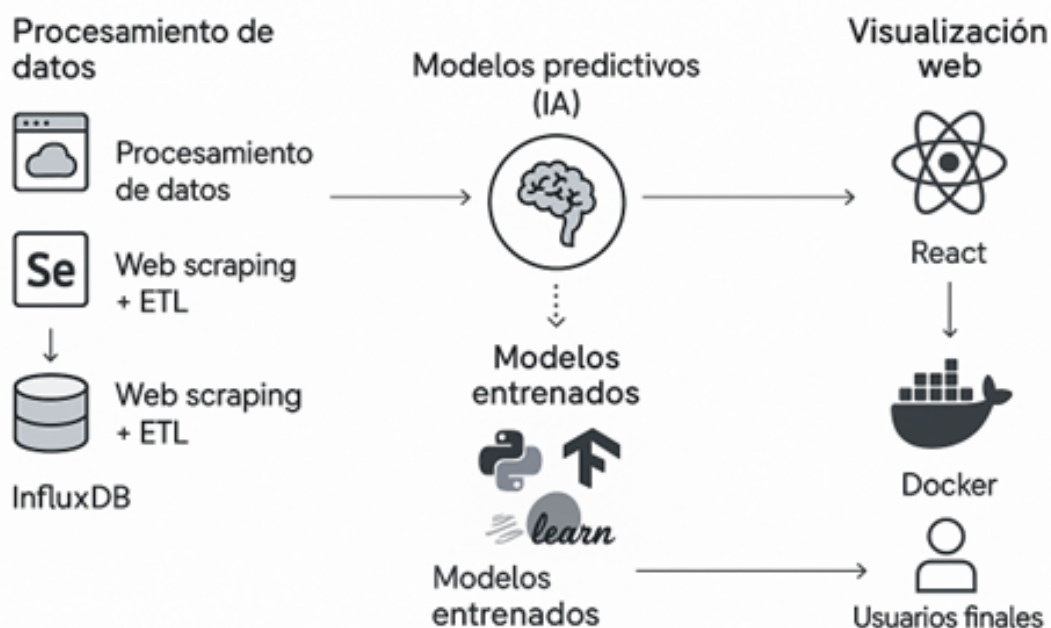


Figura 131: Arquitectura general del sistema: de la ingestión de datos a la visualización web

#### ▪ Capa de datos (ETL + Almacenamiento)

Las series hidrológicas y meteorológicas se obtienen mediante *web scraping* y un proceso *ETL* automatizado. Tras la depuración y normalización, se almacenan en **InfluxDB** como series temporales, separando *measurements* de meteorología y embalses y añadiendo tags normalizados (códigos, provincias y metadatos de estación/embalse). Esta capa expone un contrato estable de campos y tipos que consumen tanto la API como los *jobs* de modelado.

#### ▪ Capa de modelado

Los modelos predictivos se entrenan *offline* en Python (scikit-learn / TensorFlow) y se versionan como artefactos serializados. Unos *jobs* programados generan predicciones y agregados

a partir de InfluxDB y escriben los resultados en ficheros o *buckets* específicos para su consulta desde la web. Esta separación mantiene desacopladas la inferencia y la interacción del usuario.

- **Capa de servicio *backend***

El *backend* es una **API REST** desarrollada con **Node.js + Express**. Orquesta las consultas a InfluxDB mediante la librería oficial y Flux, y ofrece *endpoints* para series, agregados y predicciones filtrables por embalse, estación, provincia y rango temporal. Incorpora validación de parámetros, tratamiento de errores y CORS. La configuración sensible (*tokens* de InfluxDB, URL, *buckets*) se gestiona mediante variables de entorno (.env). El servicio se ejecuta en contenedor y se publica mediante *docker-compose.yml*.

- **Capa de presentación (*frontend*)**

La interfaz de usuario se implementa en **React**, con componentes reutilizables para tablas y gráficos interactivos. El usuario puede filtrar por provincia, embalse, estación y ventana temporal, así como comparar variables de forma simultánea mediante selectores y multiselección. La lógica de presentación está desacoplada del acceso a datos y consume exclusivamente la *API REST*, lo que facilita el mantenimiento y la evolución.

- **Despliegue y operación**

Aunque inicialmente los servicios se ejecutaban en local, con el tiempo se optó por *Docker* para garantizar la portabilidad y aislar componentes. El archivo *docker-compose.yml* coordina el arranque de los contenedores necesarios: uno para el *frontend*, otro para el *backend* y otro para la base de datos InfluxDB.

#### 4.1.3. Despliegue de servicios

Aunque inicialmente los servicios se ejecutaban en local, con el tiempo se optó por **Docker** para garantizar la portabilidad y aislar componentes. El archivo *docker-compose.yml* coordina el arranque de los contenedores necesarios: uno para el *frontend*, otro para el *backend* y otro para la base de datos InfluxDB.

#### 4.1.4. Funcionamiento de la aplicación web

La plataforma ofrece principalmente dos funcionalidades:

##### 1. Visualización de datos en tablas

- o Tabla de embalses: con filtros de selección por embalse y fechas (únicas o rangos)
- o Tabla meteorológica: con filtros más avanzados, incluyendo provincia, estación y variables meteorológicas específicas.

##### 2. Visualización de datos en gráficas

- Gráfica de embalses: permite comparar reservas en hectómetros cúbicos y porcentajes de llenado en diferentes embalses.
- Gráfica meteorológica: ofrece múltiples filtros (fechas, provincias, estaciones y variables como precipitaciones, temperaturas, humedad, viento y radiación).

Además, se añadió un **mapa interactivo de Andalucía** que muestra la ubicación de embalses y estaciones meteorológicas, facilitando la exploración geográfica de los datos.

#### 4.1.5. Problemas encontrados y soluciones

Durante el desarrollo se identificaron varias incidencias que requirieron ajustes técnicos:

- **Uso incorrecto de range() en InfluxDB** La función generaba errores al aplicarse en consultas sobre etiquetas que no incluían la columna *\_time*. La solución fue emplear funciones específicas como *schema.tagValues()* para este tipo de consultas.
- **Fallo de carga de datos tras la migración a Docker** Al ejecutar la aplicación en contenedores, los datos dejaron de mostrarse correctamente. Se resolvió modificando la lógica: ahora los datos se cargan de una sola vez desde InfluxDB y los filtros se aplican en el frontend, sin depender de endpoints adicionales.
- **Selección por defecto en los checkboxes** Inicialmente todas las estaciones aparecían seleccionadas por defecto, lo que sobrecargaba la aplicación. Se corrigió de modo que, al entrar en la plataforma, ningún elemento aparezca preseleccionado y sea el usuario quien elija manualmente.

## 4.2. Procesos *ETL*

En este proyecto se diseñó un *pipeline ETL* (extracción, transformación y carga) que conduce los datos desde las fuentes originales hasta su almacenamiento final en InfluxDB. A continuación, se describen cada una de las fases y las decisiones técnicas adoptadas.

### 4.2.1. Fase de extracción

La materia prima del sistema son las **series históricas de embalses y estaciones meteorológicas** publicadas por la Junta de Andalucía en distintos portales. Para recopilar esos datos de forma reproducible se implementó un proceso de *web scraping* que automatiza la navegación, la lectura del HTML y la descarga de ficheros para su posterior tratamiento. Qué es el *web scraping*.

Consiste en programar un cliente que solicita páginas al servidor (vía HTTP), analiza su estructura (HTML/XML/JSON) para localizar los elementos de interés y persiste la información en formatos manejables (CSV/JSON o directamente en BD). La principal ventaja es la **automatización masiva**: grandes volúmenes de datos pueden capturarse en poco tiempo y de manera programática.

**Herramientas empleadas** Se optó por **Python** por su ecosistema de librerías. En concreto:

- **Selenium** para automatizar el navegador y acceder a contenidos dinámicos (JavaScript, formularios, paginación).
- **pandas** para lectura y manipulación inmediata de los ficheros descargados.

(Se valoraron alternativas como Scrapy o servicios sin código tipo ParseHub, pero no se emplearon en la versión final.)

### Extracción de datos meteorológicos (RIA)

1. Configuración del navegador. Chrome en modo *headless* con opciones de estabilidad (sin *GPU/sandbox*, descargas sin diálogo, directorio por defecto y *safe-browsing* habilitado).

2. **Arranque del WebDriver.** Gestión automática del *chromedriver* y esperas explícitas de hasta 30 s.
3. **Acceso y aceptación de cookies.** Navegación al listado de estaciones RIA y cierre del *banner* si aparece.
4. **Raspado de enlaces.** Recorremos la tabla y paginación hasta obtener todos los pares (nombre de estación, URL de ficha).
5. **Descarga por estación.** Para cada estación se accede a “Descarga de ficheros”, se marcan las variables (temperaturas, humedades, viento, radiación, precipitación, ETo) y se lanza la descarga.
6. **Gestión de ficheros.** Se identifica el último archivo CSV/TXT descargado, se normaliza el nombre de la estación y se renombra con el patrón *<Estacion>.csv*.
7. **Tolerancia a fallos.** Errores puntuales se registran y el proceso continúa; al finalizar se cierra el navegador.

#### **Extracción de datos de embalses (REDIAM)**

1. **Inicialización.** Chrome con ventana detach para depuración y driver autogestionado.
2. **Acceso al portal.** Carga del buscador principal.
3. **Listado de embalses.** Se extraen todos los nombres de la lista de sugerencias.
4. **Utilidades**
  - Detección de fin de descarga observando el directorio.
  - Normalización de textos (acentos, mayúsculas, signos) para comparar nombre buscado vs. marcador del mapa.
5. **Bucle de descarga con reintentos ( $\leq 5$ ).** Búsqueda, selección del marcador correcto, verificación del *popup* informativo y descarga del CSV. Si falla, se reintenta.
6. **Registro.** Conteo de descargas exitosas y listado de embalses sin descargar.

7. **Cierre.** Terminación controlada y resumen por consola.

### **Beneficios del enfoque**

Automatización (ahorro de tiempo), eficiencia en ingestas masivas, actualización periódica (datos recientes), flexibilidad frente a heterogeneidad de portales y conversión de contenido no estructurado en **CSV** listos para análisis.

#### **4.2.2. Fase de transformación**

En esta fase se convirtieron las extracciones brutas (meteorología y embalses) en conjuntos de datos limpios y coherentes, listos para su ingesta directa en InfluxDB. El objetivo fue doble:

- normalizar nombres, fechas y tipos para eliminar ambigüedades y
- estructurar cada fila con el esquema *measurement-tags-fields-timestamp* exigido por el *Line Protocol*.

### **Meteorología**

Para cada estación agroclimática se llevó a cabo una “fusión segura” de orígenes y una posterior adaptación al formato de carga. Primero se **normalizaron los identificadores** (unificación de tildes, guiones y mayúsculas) y se **detectaron duplicados**. Cuando existían varios ficheros de la misma estación, se **homogeneizaron cabeceras**, se **concatenaron los contenidos**, se **eliminaron filas repetidas** y se **ordenaron cronológicamente**; con un único origen, bastó con el **parseo de la fecha** (dd/mm/aaaa) y la escritura del fichero unificado.

A continuación, se aplicó la **estandarización de cabeceras en MAYÚSCULAS**, la **derivación del id de las estaciones** a partir del prefijo alfanumérico de las columnas y el mapeo de provincia a códigos normalizados (AL, CA, CO, GR, HU, JA, MA, SE). Se **depuraron nulos** frecuentes (“n/d” y vacíos), se **simplificó la nomenclatura** de variables (p. ej., TMAX, TMED, PRECIP, ETO) y se añadieron los **metadatos mínimos** como tags (*station\_id*, *provincia*, *station\_name*) junto con el **tiempo en ISO-8601 UTC**. La serialización final dejó cada registro preparado para convertirse sin ambigüedad al esquema lógico:

```
weather_data,station_id,provincia,station_name TMAX=...,TMED=...,PRECIP=...,ETO=... time=...
```

## Embalses

El tratamiento de las series de embalses se centró en preservar fidelidad nominal y consistencia temporal. La lectura se realizó con **codificación ISO-8859-1** para mantener caracteres locales. Se **repararon filas con comas internas** en los nombres (que podían fragmentar columnas), se **eliminaron comillas residuales** y las fechas se convirtieron a **ISO-8601 UTC**, descartando observaciones inválidas.

Cada fila se estructuró en un **measurement** único (embalses), con **tags** código y nombre, y **fields** reserva y porcentaje; el **timestamp** se derivó de la fecha ya normalizada. Por coherencia con el proceso de carga, los CSV resultantes incluyen línea **#datatype** y **tipados explícitos** en cabecera, de modo que el lote pueda enviarse a InfluxDB sin transformaciones adicionales.

### 4.2.3. Fase de carga

La inserción en InfluxDB adoptó dos estrategias complementarias:

- Para **meteorología**, se generó el *Line Protocol* de cada fichero completo y se envió en una única operación (modo *bulk* por archivo), lo que reduce llamadas y garantiza atomicidad a nivel de lote; la rutina de carga lee en UTF-8, omite las dos primeras líneas de metadatos, escapa adecuadamente los *tags* con espacios o comas, convierte los timestamps ISO-8601 a nanosegundos y escribe únicamente los *fields* numéricos válidos.
- Para **embalses**, se utilizó un esquema *batch* asíncrono, acumulando hasta 500 puntos o un intervalo de 10 segundos antes del *flush*. Cada registro se compone con los *tags* (código, nombre), los *fields* (reserva, porcentaje) y la marca temporal en nanosegundos, manteniendo el *throughput* alto y una buena tolerancia a la latencia de red. En ambos casos, el cierre de cliente se realiza con vaciado explícito de *buffers* para asegurar la persistencia de todas las observaciones.

### 4.2.4. Incidencias y soluciones

Durante el proyecto se produjo un cambio integral en la web oficial de embalses de la Junta, que invalidó los *endpoints* del *scraper* original. Se reprogramó el módulo de extracción para la nueva interfaz y sus rutas, restableciendo el flujo de datos. El repositorio mantiene

ahora dos *scrapers* operativos, meteorología y embalses (nueva web), junto al código histórico previo al cambio de portal. La modularidad del *ETL* y su documentación asociada permiten sustituir componentes sin afectar al resto de la cadena y conservan la trazabilidad de extremo a extremo.

### 4.3. Metodología empleada

El proyecto se desarrolló con una metodología ágil de tipo **Scrum**, organizada en *sprints* de aproximadamente dos semanas. Al inicio de cada ciclo se priorizaban las tareas en un *product backlog* único (extracción y gobierno del dato, análisis exploratorio, modelado, validación y despliegue), definiendo criterios de aceptación y riesgos asociados. Las revisiones quincenales con tutoría sirvieron para contrastar avances, incorporar ajustes y replanificar cuando aparecían incidencias (p. ej., cambios en las páginas origen de datos o necesidades adicionales de validación).

**Sprints de arranque (fundaciones y ETL)** Los primeros ciclos se centraron en preparar el entorno y asegurar la trazabilidad del dato: configuración de repositorios y *branching* en Git, definición de entornos reproducibles, y construcción del flujo **web scraping - InfluxDB - CSV por embalse**. En esta fase se normalizaron nomenclaturas (estandarización de nombres de embalses y estaciones), se estableció el **mapa de asociación embalse-estación** por proximidad geográfica y vigencia, y se implantaron **controles de calidad** (parseo coherente de fechas, tipificación numérica, detección de duplicados y rangos imposibles). Se documentó el proceso *ETL* y se añadieron pruebas unitarias ligeras para garantizar idempotencia de las cargas.

**Sprints intermedios (análisis y preparación de features)** Con el *pipeline* estable, el foco pasó al análisis exploratorio y a la **selección de características**. Se generaron *heatmaps* y matrices de **correlación de Pearson** a diferentes escalas (sin desfase y con *lags* de 0–60 días), así como mapas de **lag óptimo** por provincia. En paralelo, se evaluó la **importancia de variables** con Random Forest (impureza y permutation importance), lo que guio la ingenie-

ría de atributos: *lags* explícitos, agregados móviles multi-escala y codificaciones estacionales, manteniendo siempre causalidad para evitar *leakage*. Los dataframes enriquecidos se validaron con **particiones temporales** (*TimeSeriesSplit*), asegurando consistencia entre entrenamiento y validación.

**Sprints de clustering y diseño experimental** A continuación se abordó la **segmentación de embalses** para facilitar el *transfer learning*: detección de atípicos con HDBSCAN, agrupamiento jerárquico (*Ward*) y evaluación con **Silhouette, Calinski–Harabasz y Davies–Bouldin**. Con los clústeres fijados, se automatizó la **selección de másters por grupo** (representatividad, estabilidad y *skill* de *baselines*) y la elección **jerárquica de targets**. Se definió el diseño de experimentos: ventanas de secuencia (120/180 días), **H=14** como horizonte operativo, normalización **z-score**, **pérdida Huber**, *early stopping* y *learning-rate schedulers*.

**Sprints de modelado y validación** Se entrenaron modelos **GRU/LSTM** en dos líneas: **from scratch** (*outliers* y *másters*) y **transfer learning** (básico y extendido) intra-clúster. Cada iteración incluía: entrenamiento, registro de **métricas objetivas** (MAE, RMSE, MAPE,  $R^2$ , *skill* frente a climatología y persistencia), **hold-out anual** y validación *rolling*, más generación automática de curvas de pérdida y comparativas “real vs. predicho”. Las incidencias detectadas (p. ej., desajustes por acumulación en modo residual) se resolvieron y se repitieron los experimentos bajo el mismo protocolo para asegurar comparabilidad.

**Sprints finales (integración y entrega)** Con los modelos consolidados, se procedió a la **serialización de artefactos** (modelos y *scalers*), a la exposición de predicciones y a la capa de **visualización web en React**, integrando gráficos (*Chart.js*) y paneles filtrables por embalse, provincia y ventana temporal. Se prepararon **contenedores Docker** para *backend* y *frontend*, facilitando el despliegue reproducible. Cada *sprint* concluyó con una **revisión** (demostración de entregables) y una **retrospectiva** centrada en mejorar el flujo de trabajo, manteniendo la documentación técnica y científica actualizada.

En conjunto, esta estrategia **iterativa e incremental** permitió avanzar de forma controlada desde la captura y gobierno del dato hasta el entrenamiento, evaluación y despliegue de los modelos predictivos, incorporando de manera continua el *feedback* de los tutores y las restricciones operativas del sistema.

# 5

## Conclusiones y trabajos futuros

### 5.1. Conclusiones

Este trabajo implementa una cadena completa de valor del dato (desde la captura automática y la normalización hasta el modelado y la visualización) y, sobre esa base, escala modelos predictivos de embalses mediante aprendizaje por transferencia (*transfer learning*) cuando existe afinidad entre hidrología y climatología.

La preparación de datos ha sido determinante para la correcta finalización de este proyecto. La estandarización de las fuentes, junto con la integración meteorología–hidrología, permitió construir series diarias homogéneas por embalse. El análisis confirmó un rasgo estructural: la relación instantánea entre tiempo atmosférico y almacenamiento es débil, pero aparecen dependencias retardadas cuando se incorporan *lags* y acumulados. Esto guió una ingeniería de atributos (retardos explícitos y ventanas móviles multiescala) que estabiliza el entrenamiento y evita fugas de información.

Para favorecer la transferencia entre sistemas comparables, los embalses se clasificaron en clústeres hidroclimáticos, con un modelo máster por grupo y exclusión de casos singulares que alteran la transferencia directa. En la práctica, este esquema redujo costes de entrenamiento y aceleró la convergencia sin sacrificar precisión: en varios pares donante–destino la calidad se aproxima a la de modelos entrenados desde cero, lo que valida el enfoque de “másteres por clúster” como mecanismo operativo de escalado.

En el diseño de los predictores, las GRU ofrecieron el mejor equilibrio entre capacidad de memoria, número de parámetros y estabilidad de entrenamiento, algo clave cuando se reentrenan múltiples destinos. Se adoptó un horizonte de **14 días** por su valor operativo: evita el techo de la persistencia en el muy corto plazo y limita la degradación típica del horizonte mensual. Además, la **formulación directa** (que aprende la trayectoria futura sin apoyarse en la persistencia) fue más robusta entre embalses y bajo *transfer learning* que la alternativa **residual**, que se reserva para sistemas extremadamente inerciales.

La lectura crítica de resultados muestra que el *transfer learning* es eficaz, pero no universal. Funciona cuando la similitud entre donante y destino es real; de lo contrario, surgen transferencias negativas que aconsejan cambiar de máster o entrenar desde cero. Aquí cobra relevancia un factor ausente en los datos brutos: la **gestión humana** del embalse. En Andalucía, la operación se condiciona por decisiones institucionales (dotaciones, prioridades de uso, restricciones de sequía, caudales ecológicos, bombeos o derivaciones). Estos criterios pueden introducir cambios de régimen que un modelo puramente hidrometeorológico no capta. El caso de **Guadalquivir** es ilustrativo: comparte rasgos con su clúster y no es un atípico estadístico, pero su gestión en periodos de sequía explica desajustes persistentes incluso tras aplicar *transfer learning*. La conclusión es clara: sin covariables de operación, el techo predictivo queda limitado en sistemas fuertemente regulados.

De la experiencia se derivan varias lecciones metodológicas:

1. comparar siempre frente a **baselines** fuertes (persistencia y climatología) para valorar utilidad real a futuro;
2. mantener **paridad estricta** de variables y escaladores entre máster y destino para minimizar la desviación de covariables (*covariate shift*) y facilitar un **warm-up** estable antes del *fine-tuning*;
3. usar **clustering operativo** como marco reproducible para elegir donantes y detectar casos que requieren tratamiento específico;
4. versionar datos y artefactos para hacer el sistema **auditable** y facilitar actualizaciones ante cambios de régimen.

Las principales limitaciones provienen de la **información disponible**: ausencia de variables de operación (reglas y curvas de explotación, dotaciones por campaña, turbinados, trasvases), huecos y heterogeneidades en algunas series y alta colinealidad entre familias climáticas que obliga a una selección prudente de atributos. Además, la prolongación de la sequía y los cambios normativos introducen **rupturas temporales** que exigen recalibraciones periódicas del *clustering* y de los modelos.

Con todo, el proyecto deja una aportación tangible: un **ETL automatizado**, un repositorio temporal coherente, **criterios de segmentación** claros y un **estándar de modelización** (GRU, H = 14, modo directo) listo para explotación y despliegue web. Desde una perspectiva aplicada, el valor no reside solo en las métricas, sino en un **método trazable** para decidir (con evidencias) cuándo transferir, cuándo cambiar de donante y cuándo entrenar desde cero. En síntesis, el *transfer learning* es beneficioso en la mayoría de embalses andaluces porque ahorra recursos y acorta ciclos manteniendo la precisión; allí donde la política de desembalse y la gestión real dominan la señal, la mejora pasa por incorporar **covariables de operación y protocolos de validación** específicos.

## 5.2. Trabajos futuros

A partir de lo ya construido, un *ETL* reproducible, la segmentación por clústeres, modelos GRU y transferencia intra-clúster, se propone una hoja de ruta que refuerza la precisión, la robustez y la utilidad operativa. **El primer eje es el dato**. Se ampliará el conjunto de variables con **información operativa** (series de desembalses y captaciones, turbinados, bombeos, trasvases, reglas de explotación o curvas guía, siempre que estos datos estén disponibles para su consulta) para explicar desajustes en sistemas con gestión intensa. Se integrarán **fuentes de teledetección y reanálisis** (precipitación en rejilla por radar o reanálisis, radiación y evapotranspiración de referencia  $ET_0$  Penman–Monteith) [46] y se añadirán **variables hidrológicas** adicionales (caudales de entrada por aportes y escorrentías), así como **meteorológicas** como la humedad del suelo. El contexto físico se enriquecerá con **información del entorno** (tipos de suelos, zonas potenciales de filtración, condicionantes geomorfológicos). En paralelo, se incorporarán **factores humanos** que condicionan la demanda (consumo urbano, tipología de cultivos abastecidos, necesidades hídricas por cultivo y calendario de campañas, estacio-

alidad turística). Este eje incluye mejoras de **calidad de series** (imputación espaciotemporal probabilística y *gap-filling* por variable) y auditorías de **cambios de sensor/estación** mediante detección de *breakpoints*. Para garantizar frescura del dato se automatizará el **web scraping diario**, con trazabilidad y alertas ante fallos de ingesta.

El segundo eje es el **modelado**. Se evolucionará desde el esquema directo a H=14 hacia **predicción multi-horizonte** (7/14/30 días) con arquitecturas *seq2seq* tipo *encoder-decoder* (GRU/LSTM), *N-BEATS* o **Temporal Fusion Transformer**, contrastadas frente a enfoques *direct multi-output* para asegurar coherencia temporal. Se explorarán **mecanismos de atención** que prioricen ventanas informativas y **GNN** que codifiquen conectividades físicas y operativas (red fluvial e infraestructuras). Como refuerzo, los **modelos híbridos físico-ML** incorporarán restricciones suaves de balance de masas y regularización hidrológica (no negatividad y cotas físicas) directamente en la función de pérdida.

El tercer eje profundiza en el **aprendizaje por transferencia**. Antes del *fine-tuning*, se aplicará un *screening* automático basado en **índices de transferibilidad** (similitud climática, patrones residuales y proximidad en espacios de *embeddings*). Durante la adaptación, se empleará **layer freezing parcial** y *parameter-efficient tuning* (p. ej., LoRA), con búsqueda sistemática de tasas de aprendizaje y épocas para evitar *overfitting* en destinos con poca muestra. Cuando existan diferencias claras entre donante y receptor, la **adaptación de dominio** mediante reponderación o alineamientos corregirá el *covariate shift*. Este ciclo se apoyará en **clustering dinámico** (recalculado en ventanas móviles, por ejemplo anuales) para capturar cambios de régimen, como sequías prolongadas, manteniendo métricas internas (Silhouette, Calinski-Harabasz, Davies-Bouldin) y una **validación externa** basada en el rendimiento medio del *transfer learning* intra-clúster como criterio práctico de calidad de la partición.

El cuarto eje aborda **incertidumbre, explicabilidad y operación**. La incertidumbre se cuantificará con *quantile loss*, *ensembles* y **conformal prediction**, con informes de calibración y cobertura orientados a uso operativo. La **interpretabilidad** se reforzará con SHAP o permutaciones por familias de variables, *ablation* sistemática y mapas de sensibilidad ( $ET_0$ ), precipitación acumulada, lags del objetivo). En **MLOps**, se priorizarán la **trazabilidad y el**

**monitoring** (*model registry*, versionado de datos, alarmas por *data/concept drift* y *retrain* automático bajo umbrales de degradación) y la **integración en explotación**: una API de inferencia *batch/streaming*, **incorporación automática de noticias y avisos** (meteorología y estado de embalses) para soporte a la decisión, y una interfaz que permita **descarga de datos en JSON/CSV** y **filtrado avanzado** por embalse, variable y ventana temporal.

Finalmente, la **extensión y evaluación** del protocolo contemplará su **generalización geográfica** a otras demarcaciones (p. ej., Cuencas Mediterráneas y Guadiana) y una **validación robusta** mediante *backtesting* en periodos hidrológicos contrastados (años húmedos/secos), ventanas *rolling* y evaluación por subregímenes (alto/bajo almacenamiento) para medir estabilidad. En **gobernanza y datos abiertos**, se propondrá un catálogo de **metadatos y licencias** con linaje completo, políticas de actualización y publicación de artefactos no sensibles (esquemas, *codebooks*, *notebooks*), junto con una **línea ética y de comunicación** que documente los límites del modelo (en particular, la dependencia de decisiones humanas) y facilite materiales de apoyo para usuarios no expertos.



# Bibliografía

- [1] Junta de Andalucía – IFAPA. *Red de Información Agroclimática de Andalucía (RIA)*. URL: [https://www.juntadeandalucia.es/agriculturaypesca/ifapa/riaweb/web/inicio\\_estaciones](https://www.juntadeandalucia.es/agriculturaypesca/ifapa/riaweb/web/inicio_estaciones) (visitado 07-09-2025).
- [2] Junta de Andalucía – REDIAM. *Geoportal REDIAM & Visor de embalses*. URL: <https://www.juntadeandalucia.es/medioambiente/portal/acceso-rediam> (visitado 07-09-2025). Visor de embalses: <https://www.juntadeandalucia.es/medioambiente/portal/acceso-rediam/geoportal/visualizadores/visor-embalses-andalucia>.
- [3] Keith Beven. *Rainfall-Runoff Modelling: The Primer*. 2.<sup>a</sup> ed. Wiley-Blackwell, 2012.
- [4] Mike Sit et al. *A Comprehensive Review of Deep Learning Applications in Hydrology and Water Resources*. 2020. arXiv: [2007.12269](https://arxiv.org/abs/2007.12269). URL: <https://arxiv.org/abs/2007.12269> (visitado 07-09-2025).
- [5] Ian Goodfellow, Yoshua Bengio y Aaron Courville. *Deep Learning*. MIT Press, 2016. URL: <https://www.deeplearningbook.org/> (visitado 07-09-2025).
- [6] Sepp Hochreiter y Jürgen Schmidhuber. “Long short-term memory”. En: *Neural Computation* 9.8 (1997), págs. 1735-1780. DOI: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- [7] Kyunghyun Cho et al. “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)*. 2014, págs. 1724-1734. URL: <https://arxiv.org/abs/1406.1078> (visitado 07-09-2025).
- [8] Friedrich Kratzert et al. “Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks”. En: *Hydrology and Earth System Sciences* 22.11 (2018), págs. 6005-6022. URL: <https://hess.copernicus.org/articles/22/6005/2018/> (visitado 07-09-2025).
- [9] Da Feng, Kuai Fang y Chaopeng Shen. “Enhancing Streamflow Forecast and Extracting Insights Using LSTM with Data Integration at Continental Scales”. En: *Water Resources Research* 57.2 (2021). DOI: [10.1029/2020WR028091](https://doi.org/10.1029/2020WR028091).

- [10] Junyoung Chung et al. *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*. 2014. arXiv: [1412.3555](https://arxiv.org/abs/1412.3555). URL: <https://arxiv.org/abs/1412.3555> (visitado 07-09-2025).
- [11] iAgua. *¿Cómo se aplica la inteligencia artificial en la gestión del agua?* 2024. URL: <https://www.iagua.es/respuestas/como-se-aplica-inteligencia-artificial-gestion-agua> (visitado 07-09-2025).
- [12] Cetaqua. *Proyecto TERRA*. 2023. URL: <https://www.cetaqua.com/proyectos/terra/> (visitado 07-09-2025).
- [13] Junta de Andalucía. *Andalucía avanza en la mejora de la gestión del agua en la agricultura con inteligencia artificial (Proyecto GestIAGua)*. 20 de feb. de 2025. URL: <https://www.juntadeandalucia.es/presidencia/portavoz/201867/> (visitado 07-09-2025).
- [14] Sinno Jialin Pan y Qiang Yang. “A Survey on Transfer Learning”. En: *IEEE Transactions on Knowledge and Data Engineering* 22.10 (2010), págs. 1345-1359. URL: [https://www.cse.ust.hk/~qyang/Docs/2009/tkde\\_transfer\\_learning.pdf](https://www.cse.ust.hk/~qyang/Docs/2009/tkde_transfer_learning.pdf).
- [15] Haoran Jiang et al. “Transfer Learning for Flood Forecasting Across Watersheds Using Deep Learning”. En: *Journal of Hydrology* 585 (2020), pág. 124776. DOI: [10.1016/j.jhydrol.2020.124776](https://doi.org/10.1016/j.jhydrol.2020.124776).
- [16] Young-Gyu Ham, Jeong-Hwan Kim y Jing-Jia Luo. “Deep Learning for Multi-year ENSO Forecasts”. En: *Nature* 573.7775 (2019), págs. 568-572. DOI: [10.1038/s41586-019-1559-7](https://doi.org/10.1038/s41586-019-1559-7).
- [17] Joaquín Torres, Juan M. Cordero y Fernando Pérez. “Transfer Learning in Energy Systems: Applications and Challenges”. En: *Energy AI* 1 (2020), pág. 100005. DOI: [10.1016/j.egyai.2020.100005](https://doi.org/10.1016/j.egyai.2020.100005).
- [18] Confederación Hidrográfica del Guadalquivir. *Informes del SAIH del Guadalquivir (situación diaria de embalses, pluviometría)*. URL: <https://www.chguadalquivir.es/saih/Informes.aspx> (visitado 07-09-2025).
- [19] MITECO. *Observatorio Nacional de la Sequía – Informes y mapas mensuales 2025*. URL: <https://www.miteco.gob.es/es/agua/temas/observatorio-nacional-de-la-sequia/informes-mapas-seguimiento/infomapas2025.html> (visitado 07-09-2025).

- [20] Confederación Hidrográfica del Guadalquivir. *Histórico de Informes de Sequías (índice y documentación relacionada)*. URL: <https://www.chguadalquivir.es/historico-de-informes-de-sequias/> (visitado 07-09-2025).
- [21] Confederación Hidrográfica del Guadalquivir. *Acuerdos de la Comisión de Desembalse (campana 2025: desembalse de 1.200 hm<sup>3</sup> y dotaciones)*. URL: [https://www.chguadalquivir.es/inicio/-/asset\\_publisher/YCnOkIuwzWDc/content/id/2880410](https://www.chguadalquivir.es/inicio/-/asset_publisher/YCnOkIuwzWDc/content/id/2880410) (visitado 07-09-2025).
- [22] Junta de Andalucía. *Planes especiales de actuación en situaciones de alerta y eventual sequía (incluye Acuerdo de 05/09/2022)*. URL: [https://www.juntadeandalucia.es/medioambiente/portal/landing-page-%C3%ADndice/-/asset\\_publisher/zX2ouZa4r1Rf/content/planes-especiales-de-actuaci-c3-b3n-en-situaciones-de-alerta-y-eventual-sequ-c3-ada/20151](https://www.juntadeandalucia.es/medioambiente/portal/landing-page-%C3%ADndice/-/asset_publisher/zX2ouZa4r1Rf/content/planes-especiales-de-actuaci-c3-b3n-en-situaciones-de-alerta-y-eventual-sequ-c3-ada/20151) (visitado 07-09-2025).
- [23] ASAJA Sevilla. *La CHG aprueba un desembalse de 1.200 hm<sup>3</sup> y dotaciones máximas del 91,6 % por hectárea para la nueva campaña de riego*. 30 de abr. de 2025. URL: <https://www.asajasevilla.es/noticias/agua/item/9741-la-chg-aprueba-un-desembalse-de-1-200-hm3-y-dotaciones-maximas-del-91-6-por-hectarea-para-la-nueva-campana-de-riego.html> (visitado 07-09-2025).
- [24] MSMK University. *Mean Absolute Error (MAE)*. [Artículo]. URL: <https://msmk.university/mean-absolute-error/> (visitado 11-09-2025).
- [25] SAP SE. *SAP Analytics Cloud — documentación*. [Documentación en línea]. URL: [https://help.sap.com/docs/SAP\\_ANALYTICS\\_CLOUD/00f68c2e08b941f081002fd3691d86a7/12cda9ce48b049a8adda7a6f3c240fa4.html](https://help.sap.com/docs/SAP_ANALYTICS_CLOUD/00f68c2e08b941f081002fd3691d86a7/12cda9ce48b049a8adda7a6f3c240fa4.html) (visitado 11-09-2025).
- [26] scikit-learn. *sklearn.metrics.mean\_absolute\_percentage\_error*. [Documentación]. URL: [https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean\\_absolute\\_percentage\\_error.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean_absolute_percentage_error.html) (visitado 11-09-2025).
- [27] Universitat de València. *Coeficiente de determinación*. [Apuntes]. URL: [https://www.uv.es/webgid/Descriptiva/6\\_coeficiente\\_de\\_determinacin.html](https://www.uv.es/webgid/Descriptiva/6_coeficiente_de_determinacin.html) (visitado 11-09-2025).

- [28] scikit-learn. *Silhouette analysis for KMeans clustering*. [Ejemplo — Documentación]. URL: [https://scikit-learn.org/stable/auto\\_examples/cluster/plot\\_kmeans\\_silhouette\\_analysis.html](https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html) (visitado 11-09-2025).
- [29] Towards Data Science. *Calinski–Harabasz index for K-Means clustering evaluation (using Python)*. [Artículo]. URL: <https://towardsdatascience.com/calinski-harabasz-index-for-k-means-clustering-evaluation-using-python-4fefe2988e/> (visitado 11-09-2025).
- [30] Analytics Lane. *El índice de Davies–Bouldin en k-means e implementación en Python*. [Artículo]. URL: <https://www.analyticslane.com/2023/06/30/el-indice-de-davies-bouldin-en-k-means-e-implementacion-en-python/> (visitado 11-09-2025).
- [31] Universidad de Sevilla. *Correlación*. [Apuntes]. URL: <https://personal.us.es/vararey/adatos2/correlacion.pdf> (visitado 11-09-2025).
- [32] IBM. *Análisis de potencia: prueba de correlación de Spearman de una muestra*. [Documentación de SPSS Statistics]. URL: <https://www.ibm.com/docs/es/spss-statistics/31.0.0?topic=correlations-power-analysis-one-sample-spearman-correlation-test> (visitado 11-09-2025).
- [33] IBM. *Random Forest*. [Tema]. URL: <https://www.ibm.com/mx-es/think/topics/random-forest> (visitado 11-09-2025).
- [34] Python Software Foundation. *Python (sitio web en español)*. [Sitio web]. URL: <https://es.python.org/> (visitado 11-09-2025).
- [35] InfluxData. *InfluxData*. [Sitio web]. URL: <https://www.influxdata.com/> (visitado 11-09-2025).
- [36] Amazon Web Services. *¿Qué es JavaScript?* [Página informativa]. URL: <https://aws.amazon.com/es/what-is/javascript/> (visitado 11-09-2025).
- [37] Microsoft. *Visual Studio Code*. [Sitio web]. URL: <https://code.visualstudio.com/> (visitado 11-09-2025).
- [38] Google Research. *Google Colaboratory*. [Sitio web]. URL: <https://colab.google/> (visitado 11-09-2025).

- [39] *About GitHub and Git*. GitHub. URL: <https://docs.github.com/es/get-started/start-your-journey/about-github-and-git> (visitado 01-09-2025).
- [40] *Docker en AWS (introducción)*. Amazon Web Services. URL: <https://aws.amazon.com/es/docker> (visitado 01-09-2025).
- [41] Node.js. *Node.js (sitio web en español)*. [Sitio web]. URL: <https://nodejs.org/es> (visitado 11-09-2025).
- [42] Bootstrap. *Bootstrap*. [Sitio web]. URL: <https://getbootstrap.com/> (visitado 11-09-2025).
- [43] Express.js. *Express (sitio web en español)*. [Sitio web]. URL: <https://expressjs.com/es/> (visitado 11-09-2025).
- [44] W3Schools. *HTML Tutorial*. [Sitio web]. URL: <https://www.w3schools.com/Html/> (visitado 11-09-2025).
- [45] World Wide Web Consortium (W3C). *Cascading Style Sheets (CSS) – Overview*. [Sitio web]. URL: [https://www.w3.org/Style/CSS/Overview.en.html#:~:text=Cascading%20Style%20Sheets%20\(CSS\)%20is,from%20the%20CSS%20working%20group](https://www.w3.org/Style/CSS/Overview.en.html#:~:text=Cascading%20Style%20Sheets%20(CSS)%20is,from%20the%20CSS%20working%20group). (visitado 11-09-2025).
- [46] Sistema de Información Agroclimática para el Regadío (SIAR). *Cálculo de  $ET_0$ : Método de Penman–Monteith*. [Documento técnico]. Traducido por Tragsatec. URL: [https://www.mapa.gob.es/dam/mapa/contenido/desarrollo-rural/temas/gestion-sostenible-de-regadios/anteriorregadio\\_sep\\_22/siar-2011/documentos/calculo-et0dic12.pdf](https://www.mapa.gob.es/dam/mapa/contenido/desarrollo-rural/temas/gestion-sostenible-de-regadios/anteriorregadio_sep_22/siar-2011/documentos/calculo-et0dic12.pdf) (visitado 11-09-2025).



# Apéndice A

# Manual de Usuario

Este manual describe cómo navegar por la aplicación web, localizar la información disponible y aprovechar sus funciones principales (tablas de datos, visualización de gráficas y mapa interactivo).

## 1. Página de Inicio

En la figura 132 se muestra la visualización de miniaturas de noticias relacionadas con la meteorología y la gestión del agua en Andalucía. Al pulsar el botón «Ver más», el usuario es redirigido a la noticia completa.

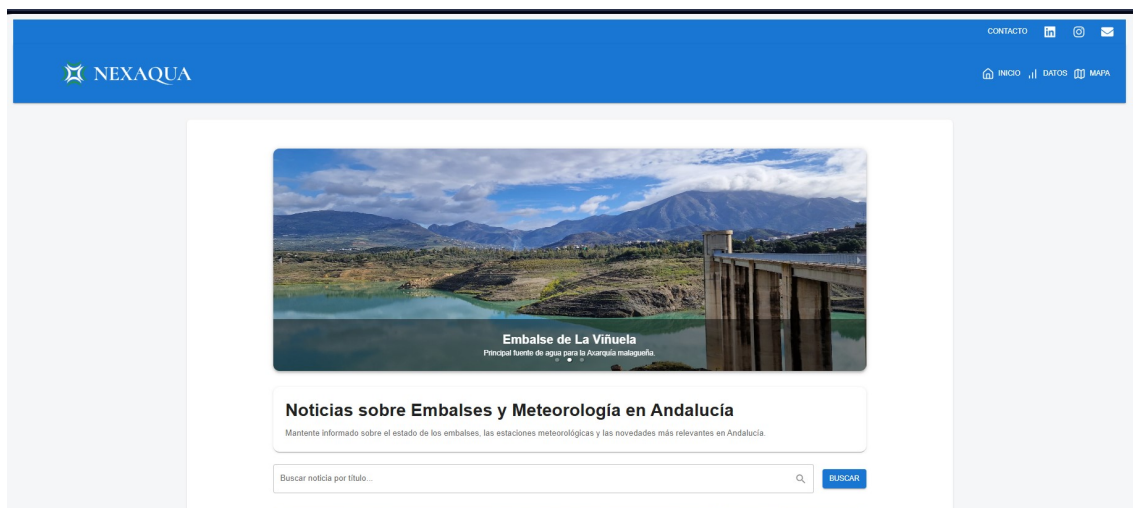


Figura 132: Ventana de inicio

En figura 133 se muestra la visualización de una miniatura de noticias referentes a la meteorología y a la gestión de agua en Andalucía. Si el usuario pulsa sobre el botón de “Ver más”, se redirige a la noticia completa.

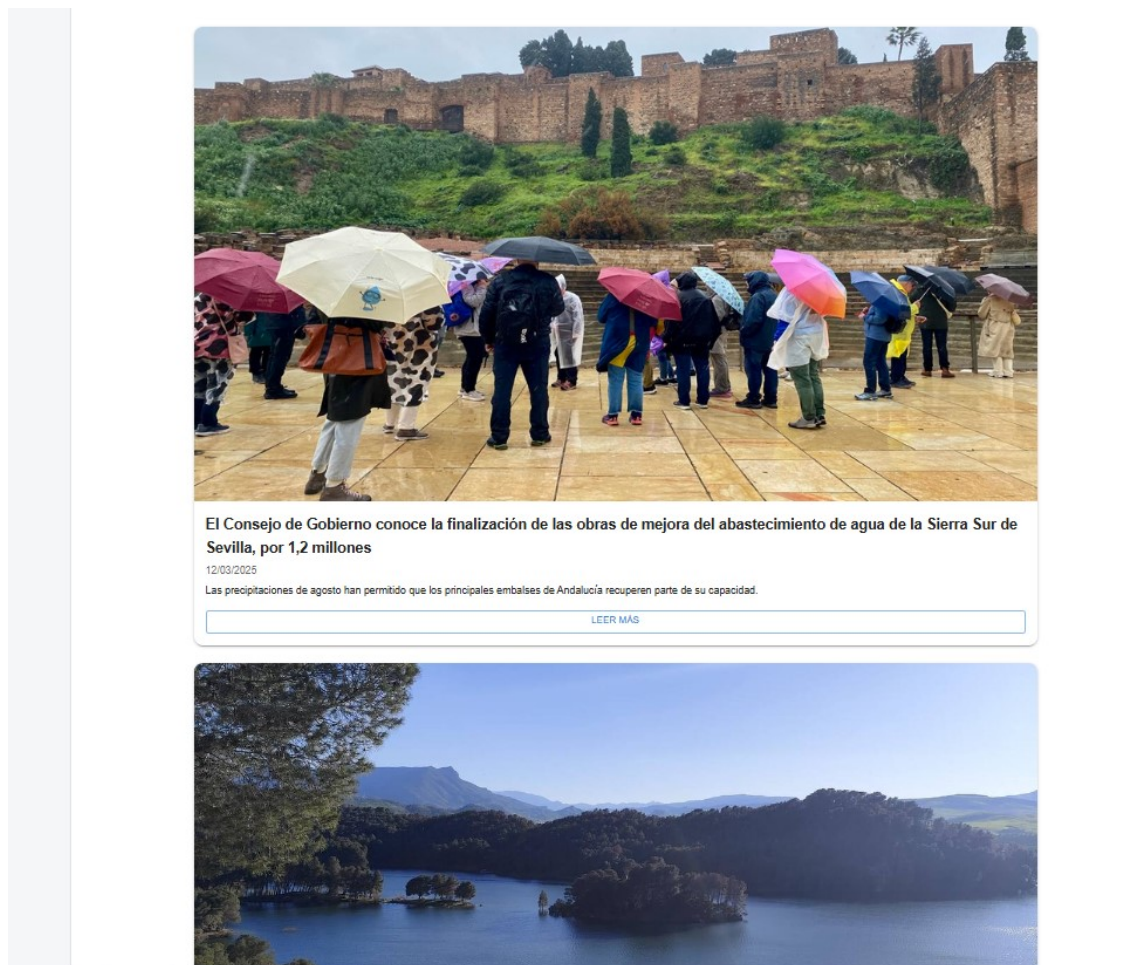


Figura 133: Ventana de inicio noticias

## 2. Páginas de avisos legales

En el pie de página se encuentran los enlaces a **Aviso legal**, **Política de cookies**, **Accesibilidad** y **Protección de datos**. Cada sección ofrece la información normativa aplicable y las condiciones de uso del sitio.

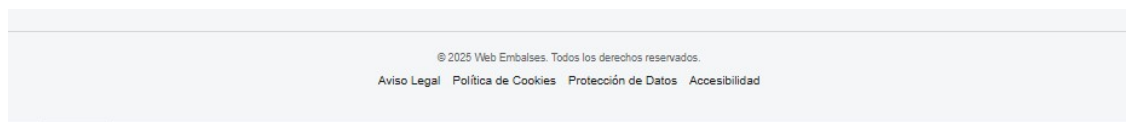


Figura 134: Pestañas de avisos legales

Las pestañas personalizadas de las políticas legales se pueden observar en las siguientes figuras [135](#) [139](#) [137](#) [138](#).

## Aviso Legal

El presente sitio web, **NEXAQUA**, tiene como finalidad ofrecer información y herramientas relacionadas con la gestión de embalses y la meteorología en Andalucía.

### Titularidad

Este sitio web es propiedad de **Patricia Martín Castillo** y **Rodrigo Hernández Barba**, con correos de contacto en: [0619993622@uma.es](mailto:0619993622@uma.es) y [0619894427@uma.es](mailto:0619894427@uma.es).

### Condiciones de uso

La utilización de este sitio web implica la aceptación plena de las condiciones aquí expuestas. El usuario se compromete a utilizar el portal de conformidad con la legislación vigente y con el presente Aviso Legal, absteniéndose de realizar cualquier actividad que pueda resultar ilícita o lesiva para terceros.

### Propiedad intelectual e industrial

Los contenidos de esta web, incluidos textos, imágenes, logotipos, diseño gráfico y código fuente, están protegidos por los derechos de propiedad intelectual e industrial, y no podrán ser reproducidos, distribuidos ni transformados sin autorización expresa del titular.

### Responsabilidad

**NEXAQUA** no se hace responsable del uso indebido de los contenidos de la web, ni de posibles daños o perjuicios derivados de la falta de disponibilidad o continuidad del sitio.

Figura 135: Ventana de aviso legal

## Protección de Datos y Privacidad

En cumplimiento con el Reglamento (UE) 2016/679 de Protección de Datos (RGPD) y la normativa española aplicable, se informa a los usuarios de lo siguiente:

### Responsable del tratamiento

**Patricia Martín Castillo** y **Rodrigo Hernández Barba**.

Correo electrónico de contacto: [0619993622@uma.es](mailto:0619993622@uma.es) y [0619894427@uma.es](mailto:0619894427@uma.es).

### Finalidad del tratamiento

Los datos personales recogidos a través de los formularios de contacto o comunicación serán utilizados únicamente para atender las consultas planteadas por los usuarios y mejorar la experiencia de navegación en el sitio web.

### Legitimación

La base legal para el tratamiento de los datos es el **consentimiento del usuario**, otorgado al remitir su consulta o al utilizar el sitio.

### Destinatarios

No se cederán datos a terceros, salvo en los casos en que exista una **obligación legal**.

### Derechos de los usuarios

Los usuarios pueden ejercer sus derechos de acceso, rectificación, supresión, limitación y oposición escribiendo a las direcciones de correo electrónico indicadas: [0619993622@uma.es](mailto:0619993622@uma.es) o [0619894427@uma.es](mailto:0619894427@uma.es).

Para procesar la solicitud será necesario adjuntar una copia de un documento acreditativo de identidad.

### Conservación de los datos

Los datos personales se conservarán únicamente durante el tiempo necesario para atender la consulta o petición realizada, y mientras el usuario no solicite su supresión.

Figura 136: Ventana de protección de datos

## Política de Cookies

Este sitio web utiliza cookies propias y de terceros con el objetivo de mejorar la experiencia de usuario, analizar el tráfico y ofrecer un servicio más personalizado.

### ¿Qué son las cookies?

Las cookies son pequeños archivos de texto que se almacenan en el navegador del usuario al visitar un sitio web. Sirven para reconocer al usuario en futuras visitas, recordar sus preferencias o recopilar información estadística sobre la navegación.

### Tipos de cookies utilizadas

- **Cookies técnicas:** necesarias para el correcto funcionamiento de la web y la prestación de los servicios solicitados por el usuario.
- **Cookies de análisis:** nos ayudan a comprender cómo interactúan los usuarios con la página, midiendo y analizando la audiencia con fines estadísticos.
- **Cookies de personalización:** permiten recordar preferencias de navegación como idioma, región o ajustes elegidos por el usuario.

### Gestión de cookies

El usuario puede configurar, bloquear o eliminar las cookies instaladas en su equipo a través de las opciones de configuración de su navegador:

- [Google Chrome](#)
- [Mozilla Firefox](#)
- [Microsoft Edge](#)
- [Safari](#)

### Actualizaciones de la política

La presente política de cookies puede ser actualizada para adaptarse a nuevas exigencias normativas o a cambios en la configuración del sitio web. Recomendamos a los usuarios revisarla de forma periódica.

Figura 137: Ventana de política de cookies

## Accesibilidad

En NEXAQUA estamos comprometidos con la accesibilidad digital y el diseño universal, promoviendo la igualdad de oportunidades y la no discriminación. Nuestro objetivo es que cualquier persona pueda utilizar este sitio web con independencia de sus capacidades, su dispositivo o su contexto de uso.

### Marco y nivel de conformidad

Trabajamos para cumplir con los criterios de conformidad de las **Pautas de Accesibilidad para el Contenido Web (WCAG) 2.1**, nivel **AA**, y la normativa vigente en materia de accesibilidad digital.

### Medidas adoptadas

- **Estructura semántica** del contenido (encabezados, listas y etiquetas correctamente anidadas).
- **Texto alternativo** en imágenes y controles no textuales relevantes.
- **Contraste de color** suficiente para garantizar la legibilidad.
- **Navegación por teclado** y **focus** visible en elementos interactivos.
- **Uso de etiquetas y descripciones** en formularios e inputs.
- Evitamos contenido que **parpadee** o pueda provocar molestias.
- Compatibilidad progresiva con **lectores de pantalla** y ayudas técnicas habituales.

### Contenidos y tecnologías

Este sitio emplea HTML5, CSS, JavaScript y componentes React/Material UI. Intentamos mantener la compatibilidad con los navegadores modernos más utilizados y con diferentes tamaños de pantalla.

### Observaciones y contacto

La accesibilidad es un proceso de mejora continua. Si encuentras alguna barrera o necesitas la información en un formato alternativo, escríbenos a:

- [0619993622@uma.es](mailto:0619993622@uma.es)
- [0619894427@uma.es](mailto:0619894427@uma.es)

### Limitaciones conocidas

Aunque realizamos revisiones periódicas, podrían existir páginas o componentes con:

- Etiquetas ARIA o textos alternativos pendientes de revisión puntual.
- Tablas con gran volumen de datos que requieren mejoras de navegación para lectores de pantalla.
- Elementos de terceros (por ejemplo, mapas incrustados) con controles no totalmente personalizables.

Si detectas algún problema, por favor avísanos para priorizar su corrección.

### Fecha de la declaración

Esta declaración de accesibilidad se actualizó el **1 de septiembre de 2025**.

---

Figura 138: Ventana de accesibilidad

## Protección de Datos y Privacidad

En cumplimiento con el Reglamento (UE) 2016/679 de Protección de Datos (RGPD) y la normativa española aplicable, se informa a los usuarios de lo siguiente:

### Responsable del tratamiento

Patricia Martín Castillo y Rodrigo Hernández Barba.

Correo electrónico de contacto: [0619993822@uma.es](mailto:0619993822@uma.es) y [0619894427@uma.es](mailto:0619894427@uma.es).

### Finalidad del tratamiento

Los datos personales recogidos a través de los formularios de contacto o comunicación serán utilizados únicamente para atender las consultas planteadas por los usuarios y mejorar la experiencia de navegación en el sitio web.

### Legitimación

La base legal para el tratamiento de los datos es el **consentimiento del usuario**, otorgado al remitir su consulta o al utilizar el sitio.

### Destinatarios

No se cederán datos a terceros, salvo en los casos en que exista una **obligación legal**.

### Derechos de los usuarios

Los usuarios pueden ejercer sus derechos de acceso, rectificación, supresión, limitación y oposición escribiendo a las direcciones de correo electrónico indicadas: [0619993822@uma.es](mailto:0619993822@uma.es) o [0619894427@uma.es](mailto:0619894427@uma.es).

Para procesar la solicitud será necesario adjuntar una copia de un documento acreditativo de identidad.

### Conservación de los datos

Los datos personales se conservarán únicamente durante el tiempo necesario para atender la consulta o petición realizada, y mientras el usuario no solicite su supresión.

Figura 139: Ventana de protección de datos

### 3. Cabecera de la web

La cabecera reúne los accesos a las principales secciones: **Tablas**, **Mapas** y **Gráficas**. En la parte superior también se dispone de un botón de Contacto **140** y enlaces a las redes sociales, desde los que se puede remitir consultas y seguir las novedades del proyecto.

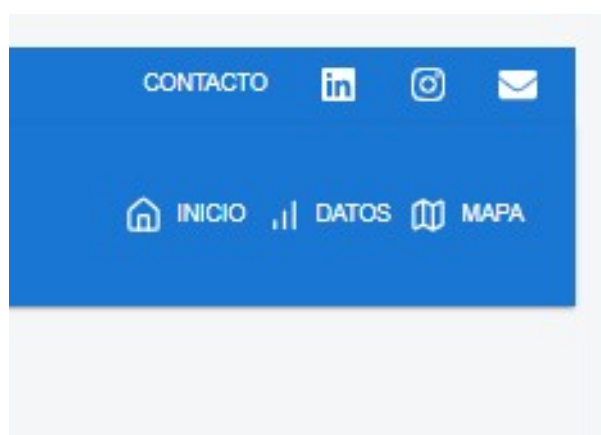


Figura 140: Header



Figura 141: Ventana de contacto

Uno de los pilares fundamentales de la web son las pestañas de datos que, a través de un desplegable, se muestran las páginas de las tablas. Estas tablas son:

- **Tabla de embalses**
- **Tabla de datos meteorológicos**
- **Tabla de gráficas**

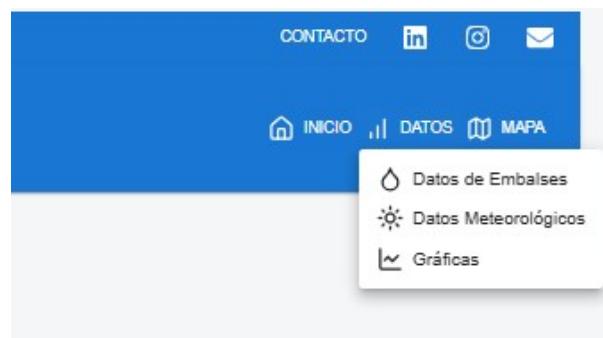


Figura 142: Desplegable

La aplicación incorpora un nombre y un logotipo personalizados; al pulsarlos, el usuario vuelve a la página de inicio. Nexaqua surge de la unión de “Nexo” y “Aqua”, subrayando el papel de la plataforma como puente entre datos, modelos y gestión del agua. El logotipo, en verde y blanco, colores de la bandera de Andalucía, refuerza su identidad territorial y el compromiso con la sostenibilidad. 143.



Figura 143: Logo y nombre de la página web

#### 4. Páginas de datos

Las páginas de datos permiten consultar, filtrar y explorar la información tabular y sus representaciones gráficas. Cada conjunto, embalses y meteorología, se muestra en una vista independiente con sus propios filtros (véanse las figuras 144 y 145). Para agilizar la navegación en tablas extensas, se incorporan flechas de desplazamiento rápido (esquina inferior derecha) que llevan al inicio o al final, así como paginación y un selector para ajustar el número de filas por página. Todas estas funcionalidades están disponibles en ambas tablas, garantizando una experiencia de consulta homogénea.

**Datos de Embalses**

RANGO DE FECHAS | FECHA ÚNICA

Fecha Inicio: 12/04/2024 | Fecha Fin: 13/04/2024 | BUSCAR

Filtrar por Embalse

Fecha	Embalse	Reserva (hm <sup>2</sup> )	Porcentaje (%)
2024-04-12	AGUASCEBAS	0.39 hm <sup>2</sup>	100%
2024-04-13	AGUASCEBAS	0.39 hm <sup>2</sup>	100%
2024-04-12	ALMODÓVAR	4.65 hm <sup>2</sup>	82%
2024-04-13	ALMODÓVAR	4.65 hm <sup>2</sup>	82%
2024-04-12	ANDÁVALO	220.47 hm <sup>2</sup>	35%
2024-04-13	ANDÁVALO	220.42 hm <sup>2</sup>	35%
2024-04-12	ARACENA	75.26 hm <sup>2</sup>	50%
2024-04-13	ARACENA	75.33 hm <sup>2</sup>	50%
2024-04-12	ARCOS	12.85 hm <sup>2</sup>	88%
2024-04-13	ARCOS	12.85 hm <sup>2</sup>	88%

Filas por página: 10 | 1-10 of 148

© 2025 Web Embalses. Todos los derechos reservados.  
 Aviso Legal | Política de Cookies | Protección de Datos | Accesibilidad

Figura 144: Ventana tabla de embalses

**Datos Meteorológicos**

RANGO DE FECHAS     FECHA ÚNICA  
 Fecha Inicio:     Fecha Fin:       
   

Fecha	Provincia	Estación	Temp. Media (°C)	Temp. Más (°C)	Temp. Mín (°C)	Humedad Media (%)	Humedad Más (%)	Humedad Mín (%)
2024-04-12	Almería	Adra	16.70°C	21.10°C	12.20°C	66.2%	85.3%	46.9%
2024-04-13	Almería	Adra	18.10°C	22.40°C	13.70°C	57.0%	72.4%	36.5%
2024-04-12	Jaén	Alcaudete	18.90°C	27.50°C	10.70°C	41.6%	61.7%	23.1%
2024-04-13	Jaén	Alcaudete	18.90°C	26.00°C	11.80°C	38.3%	55.5%	20.9%
2024-04-12	Almería	Almería	17.00°C	21.40°C	11.00°C	61.0%	88.1%	36.1%
2024-04-13	Almería	Almería	16.00°C	25.40°C	12.40°C	53.2%	81.1%	30.3%
2024-04-12	Huelva	Aimonte	16.00°C	26.30°C	10.20°C	63.8%	91.6%	26.8%
2024-04-13	Huelva	Aimonte	16.40°C	27.90°C	10.90°C	60.6%	90.6%	28.3%
2024-04-12	Huelva	Aimonte bajo plástico	16.00°C	36.60°C	6.20°C	66.3%	100.0%	20.9%
2024-04-13	Huelva	Aimonte bajo plástico	16.90°C	37.20°C	6.70°C	63.7%	100.0%	15.6%

Filas por página: 10    1-10 of 180

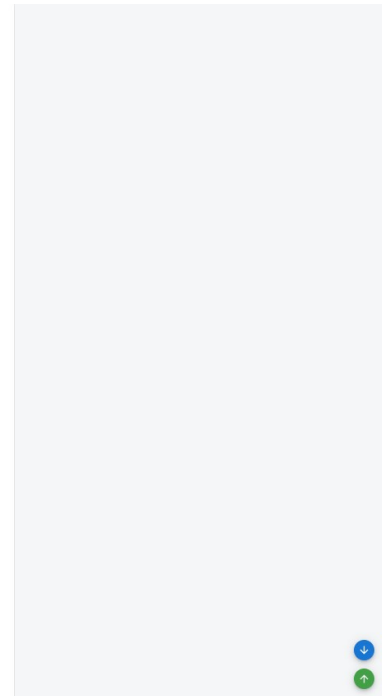


Figura 145: Ventana tabla de datos meteorológicos

Dentro del apartado Datos del menú se incluye la página de gráficas. Su funcionamiento replica el de las tablas: el usuario configura los filtros (p. ej., embalse/estación, variable y rango temporal) y la vista se actualiza con las representaciones correspondientes. Las Figuras 146 y 147 ilustran un ejemplo de resultado tras aplicar los filtros.

## Evolución del nivel de los embalses

Selecciona un embalse para ver la gráfica.

Fecha Inicio

02/04/2024

Fecha Fin

26/04/2024

Medida

Reserva de Agua (hm<sup>3</sup>)

### Filtrar por Embalse

Todos

TRANCO DE BEAS  AGUASCEBAS  SAN CLEMENTE  EL PORTILLO  LA BOLERA  NEGRATIN  FRANCISCO ABELLAN

DAÑADADOR  GUADALEN  LA FERNANDINA  GUADALMENA  GIRIBAILE  QUIEBRAJANO  RUMBLAR  JÁNDULA

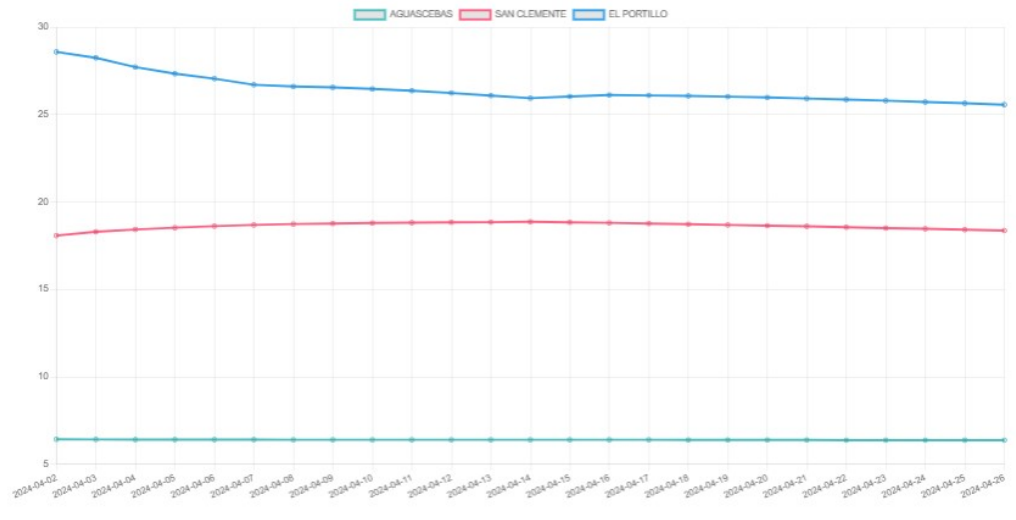


Figura 146: Ventana de grafica de embalses

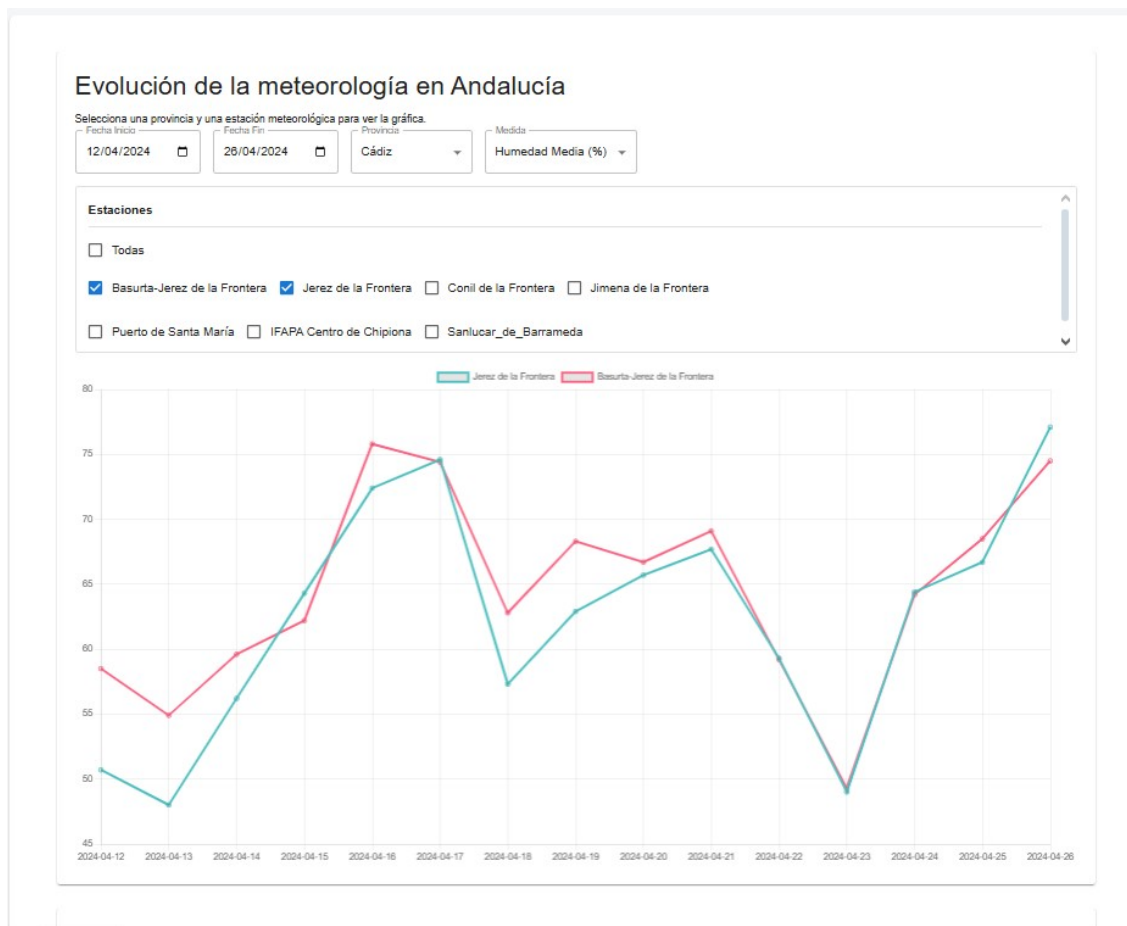


Figura 147: Ventana de grafica de datos meteorologicos

## 5. Mapa de Andalucía con embalses y estaciones meteorológicas

El mapa, accesible desde la cabecera, muestra la localización precisa de todos los embalses y estaciones meteorológicas de Andalucía. Los iconos azules identifican los embalses y los rojos las estaciones. Cada marcador es interactivo: al hacer clic se despliega una ficha con la información asociada.

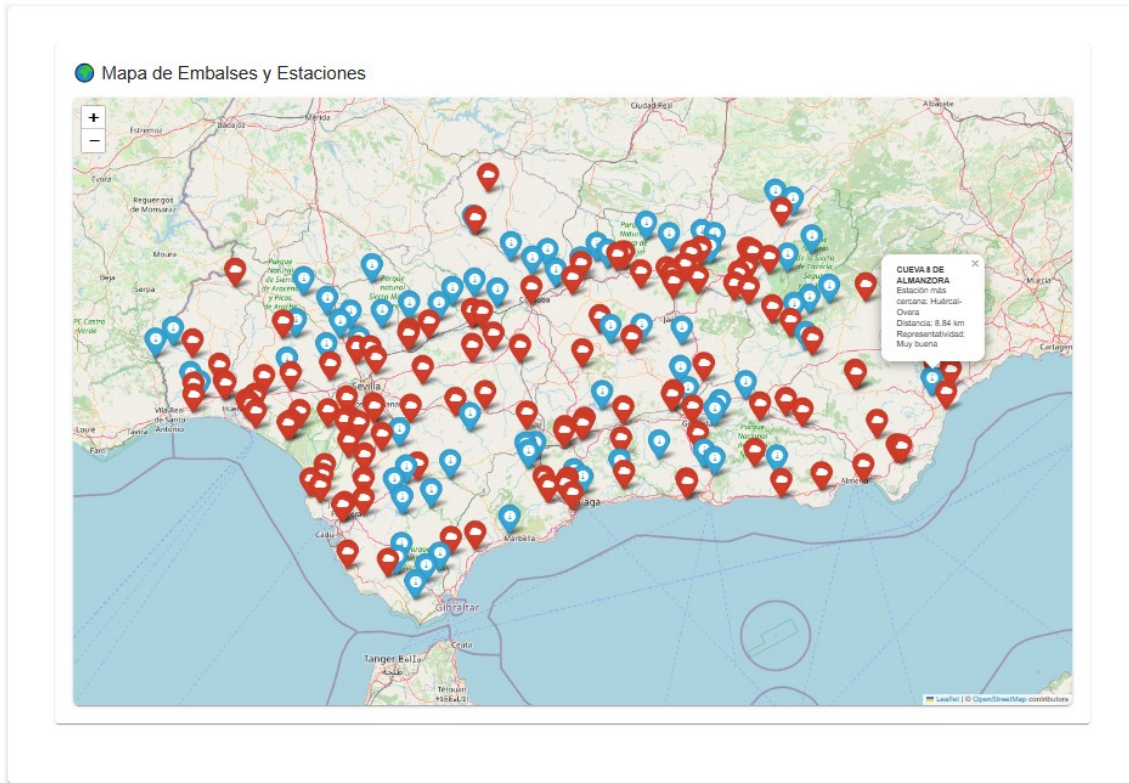


Figura 148: Ventana de mapa

# Apéndice B

## Contexto del estudio

### B.1. Red de embalses

A efectos de esta memoria, se han clasificado los 73 **embalses andaluces** y se presentan organizados por su demarcación u organismo gestor. Para cada uno se recogen sus **usos principales, río asociado, provincia y su sistema de explotación de referencia**.

#### ▪ Demarcación del Guadalquivir

Embalse	Uso	Río	Provincia	Sistema
Aguascebas	Abastecimiento	Arroyo De Las Aguascebas Chico	Jaen	Sistema De Regulacion General
Aracena	Abastecimiento, Hidroelectrico	Rivera De Huelva	Huelva	Abastecimiento De Sevilla
Arenoso	Abastecimiento, Hidroelectrico, Riego	Arenoso	Cordoba	Sistema De Regulacion General
Bembezar	Abastecimiento, Defensa Frente Avenidas, Riego, Ganadero	Bembezar	Cordoba	Bembezar- Retortillo
Cala	Abastecimiento, Hidroelectrico, Ganadero	Rivera De Cala	Sevilla	Abastecimiento De Sevilla

Continúa en la siguiente página

<b>Embalse</b>	<b>Uso</b>	<b>Río</b>	<b>Provincia</b>	<b>Sistema</b>
Canales	Abastecimiento, Defensa Frente Avenidas, Ganadero, Hidroelectico, Riego	Genil	Granada	Alto Genil
Colomera	Abastecimiento, Defensa Frente Avenidas, Riego, Ganadero	Colomera	Granada	Alto Genil
Cubillas	Riego, Hidroelectico, Abastecimiento	Cubillas	Granada	Alto Genil
Dañador	Abastecimiento	Dañador	Jaen	Sistema De Regulacion General
El Agrio	Riego, Industrial	Cripinejo O Agrio	Sevilla	Almonte- Marismas
El Pintado	Abastecimiento, Hidroelectico, Riego	Viar	Sevilla	Viar
El Portillo	Abastecimiento, Defensa Frente Avenidas, Riego	Castril	Granada	Alto Guadiana Menor
Francisco Abellan	Abastecimiento, Riego	Fardes	Granada	Hoya De Guadix
Gergal	Abastecimiento	Rivera De Huelva	Sevilla	Abastecimiento De Sevilla
Giribaile	Abastecimiento, Defensa Frente Avenidas, Riego, Hidroelectrico	Guadalimar	Jaen	Sistema De Regulacion General

Continúa en la siguiente página

<b>Embalse</b>	<b>Uso</b>	<b>Río</b>	<b>Provincia</b>	<b>Sistema</b>
Guadalen	Abastecimiento, Defensa Frente Avenidas, Riego, Hidroelectrico, Ganadero	Guadalen	Jaen	Sistema De Regulacion General
Guadalmellato	Abastecimiento, Hidroelectrico, Riego	Guadalmellato	Cordoba	Sistema De Regulacion General
Guadalmena	Abastecimiento, Hidroelectrico, Riego	Guadalmena	Jaen	Sistema De Regulacion General
Guadalupe	Abastecimiento, Defensa Frente Avenidas	Guadalupe	Cordoba	Sistema De Regulacion General
Huesna	Abastecimiento, Defensa Frente Avenidas, Ganadero	Rivera De Huesna	Sevilla	Rivera Del Huesna
Iznajar	Abastecimiento, Defensa Frente Avenidas, Hidroelectrico, Riego, Ganadero	Genil	Cordoba	Sistema De Regulacion General
Jandula	Abastecimiento, Riego, Hidroelectrico, Ganadero	Jandula	Jaen	Sistema De Regulacion General
Jose Toran	Abastecimiento, Defensa Frente Avenidas, Riego	Arroyo De Guadalbacar	Sevilla	Sistema De Regulacion General

Continúa en la siguiente página

<b>Embalse</b>	<b>Uso</b>	<b>Río</b>	<b>Provincia</b>	<b>Sistema</b>
La Bolera	Abastecimiento, Defensa Frente Avenidas, Riego	Guadalentin	Jaen	Alto Guadiana Menor
La Breña	Abastecimiento, Defensa Frente Avenidas, Hidroelectrico, Riego	Guadiato	Cordoba	Sistema De Regulacion General
La Fernandina	Abastecimiento, Hidroelectrico, Riego	Guarriza	Jaen	Sistema De Regulacion General
La Minilla	Abastecimiento, Defensa Frente Avenidas	Rivera De Huelva	Sevilla	Abastecimiento De Sevilla
Los Bermejales	Abastecimiento, Defensa Frente Avenidas, Riego, Hidroelectrico, Ganadero	Cacin	Granada	Alto Genil
Martin Gonzalo	Abastecimiento, Regulacion	Arroyo De Martin Gonzalo	Cordoba	Sistema De Regulacion General
Melonares	Abastecimiento, Riego	Viar	Sevilla	Abastecimiento De Sevilla
Negratin	Abastecimiento, Defensa Frente Avenidas, Riego, Hidroelectrico	Guadiana Menor	Granada	Sistema De Regulacion General

Continúa en la siguiente página

<b>Embalse</b>	<b>Uso</b>	<b>Río</b>	<b>Provincia</b>	<b>Sistema</b>
Puebla De Cazalla	Riego	Corbones	Sevilla	Sistema De Regulacion General
Puente Nuevo	Abastecimiento, Hidroelectrico, Riego	Guadiato	Cordoba	Sistema De Regulacion General
Quentar	Riego, Hidroelectico, Abastecimiento	Aguas Blancas	Granada	Alto Genil
Quiebrajano	Abastecimiento	Quiebrajano	Jaen	Abastecimiento De Jaen
Retortillo	Abastecimiento, Riego	Retortillo	Cordoba	Bembezar- Retortillo
Rumblar	Abastecimiento, Hidroelectrico, Riego	Rumblar	Jaen	Rumblar
San Clemente	Riego	Guardal	Granada	Alto Guadiana Menor
San Rafael De Navallana	Abastecimiento, Defensa Frente Avenidas, Riego	Guadalmellato	Cordoba	Sistema De Regulacion General
Sierra Boyera	Riego, Hidroelectrico	Guadiato	Cordoba	Sistema De Regulacion General
Torre Del Aguila	Riego	Arroyo Salado De Moron	Sevilla	Sistema De Regulacion General
Tranco De Beas	Abastecimiento, Hidroelectrico, Riego	Guadalquivir	Jaen	Sistema De Regulacion General

Continúa en la siguiente página

<b>Embalse</b>	<b>Uso</b>	<b>Río</b>	<b>Provincia</b>	<b>Sistema</b>
Vadomojon	Riego, Hidroelectrico	Guadajoz	Cordoba	Sistema De Regulacion General
Viboras	Abastecimiento	Viboras	Jaen	Abastecimiento De Jaen
Yeguas	Riego, Hidroelectrico	De Las Yeguas	Cordoba	Sistema De Regulacion General
Zufre	Abastecimiento, Hidroelectrico	Rivera De Huelva	Huelva	Abastecimiento De Sevilla

Tabla 6: Embalses Confederación Guadalquivir

■ **Demarcación de Guadalete-Barbate**

<b>Embalse</b>	<b>Uso</b>	<b>Río</b>	<b>Provincia</b>	<b>Sistema</b>
Almodovar	Abastecimiento, Defensa Frente Avenidas	Almodovar	Cadiz	Barbate
Arcos	Abastecimiento, Defensa Frente Avenidas	Guadalete	Cadiz	Guadalete
Barbate	Abastecimiento	Barbate	Cadiz	Barbate
Bornos	Abastecimiento, Defensa Frente Avenidas	Guadalete	Cadiz	Guadalete
Celemin	Abastecimiento	Celemin	Cadiz	Barbate

Continúa en la siguiente página

<b>Embalse</b>	<b>Uso</b>	<b>Río</b>	<b>Provincia</b>	<b>Sistema</b>
Guadalcacín	Abastecimiento, Defensa Frente Avenidas	Majaceite	Cádiz	Guadalete
Los Hurones	Abastecimiento, Defensa Frente Avenidas, Ganadero	Majaceite	Cádiz	Guadalete
Zahara-El Gastor	Abastecimiento	Guadalete	Cádiz	Guadalete

Tabla 7: Embalses Demarcación Guadalete -Barbate

■ **Demarcación de Tinto-Odiel-Piedras**

<b>Embalse</b>	<b>Uso</b>	<b>Río</b>	<b>Provincia</b>	<b>Sistema</b>
Andevalo	Abastecimiento, Defensa Frente Avenidas	Rivera De Malagon	Huelva	Tinto-Odiel- Piedras (Cuenca Del Guadiana)
Chanza	Abastecimiento, Defensa Frente Avenidas, Ganadero	Rivera De Chanza	Huelva	Tinto-Odiel- Piedras (Cuenca Del Guadiana)
Corumbel Bajo	Abastecimiento, Defensa Frente Avenidas	Corumbel	Huelva	Tinto-Odiel- Piedras (Tinto)
Jarrama	Abastecimiento, Defensa Frente Avenidas	Rivera Del Jarrama	Huelva	Tinto-Odiel- Piedras (Tinto)
Machos	Abastecimiento, Defensa Frente Avenidas	Piedras	Huelva	Tinto-Odiel- Piedras (Piedras)

Continúa en la siguiente página

<b>Embalse</b>	<b>Uso</b>	<b>Río</b>	<b>Provincia</b>	<b>Sistema</b>
Piedras	Abastecimiento, Defensa Frente Avenidas	Piedras	Huelva	Tinto-Odiel- Piedras (Piedras)

Tabla 8: Embalses Demarcación Tinto – Odiel – Piedras

■ **Demarcación del Mediterraneo**

<b>Embalse</b>	<b>Uso</b>	<b>Río</b>	<b>Provincia</b>	<b>Sistema</b>
Beninar	Abastecimiento, Defensa Frente Avenidas	Adra	Almeria	Sierra Nevada
Beznar	Abastecimiento, Defensa Frente Avenidas, Ganadero	Izbor	Granada	Sierra Nevada
Casasola	Abastecimiento, Defensa Frente Avenidas	Campanillas	Malaga	Serrania De Ronda
Charco Redondo	Abastecimiento, Defensa Frente Avenidas, Ganadero	Las Cañas O Palmones	Cadiz	Serrania De Ronda
Concepcion	Abastecimiento	Verde	Malaga	Serrania De Ronda
Conde De Guadalhorce	Abastecimiento	Turon	Malaga	Serrania De Ronda
Cuevas De Almanzora	Abastecimiento, Defensa Frente Avenidas, Ganadero	Almanzora	Almeria	Sierra De Los Filabres- Estancias

Continúa en la siguiente página

<b>Embalse</b>	<b>Uso</b>	<b>Río</b>	<b>Provincia</b>	<b>Sistema</b>
Guadalhorce	Abastecimiento, Defensa Frente Avenidas, Ganadero	Guadalhorce	Malaga	Serrania De Ronda
Guadalteba	Abastecimiento, Defensa Frente Avenidas, Ganadero	Guadalteba	Malaga	Serrania De Ronda
Guadarranque	Abastecimiento, Defensa Frente Avenidas, Ganadero	Guadarranque	Cadiz	Serrania De Ronda
Limonero	Abastecimiento, Defensa Frente Avenidas	Guadalmedina	Malaga	Serrania De Ronda
Rules	Abastecimiento, Defensa Frente Avenidas, Ganadero	Guadalfeo	Granada	Sierra Nevada
Viñuela	Abastecimiento, Defensa Frente Avenidas	Guaro	Malaga	Sierra De Tejada Y Almirajara

Tabla 9: Embalses Demarcación Cuencas Mediterráneas Andaluzas

## B.2. Red de estaciones agroclimáticas de Andalucía

A continuación, se listan las **estaciones utilizadas** en Andalucía **clasificadas por provincia**:

- **Temperatura máxima** (°C): temperatura máxima registrada en un día en particular. Se mide en grados celsius.
- **Temperatura mínima** (°C): temperatura mínima registrada en un día en particular. Se mide en grados celsius.

- **Temperatura media** (°C): temperatura media de un día en concreto. Se mide en grados celsius.
- **Humedad máxima** (%): Es la cantidad máxima de vapor de agua que se encuentra en un determinado volumen de aire en un día en concreto. En este caso se mide en tantos por cien.
- **Humedad mínima** (%): Es la cantidad mínima de vapor de agua que se encuentra en un determinado volumen de aire en un día en concreto. En este caso se mide en tantos por cien.
- **Humedad media** (%): Es la cantidad media de vapor de agua que se encuentra en un determinado volumen de aire en un día en concreto. En este caso se mide en tantos por cien.
- **Velocidad del viento** (m/s): Es el desplazamiento del aire en un espacio e instante determinado de un día en concreto. Se mide en metros por segundo.
- **Velocidad del viento máxima** (m/s): Es el desplazamiento máximo del aire en un espacio e instante determinado en un día en concreto. Se mide en metros por segundo.
- **Dirección del viento** (°): Nos indica de donde viene el viento. Se mide en grados, desde a 1° a 360°. Valores cercanos a 1° y 360° indican vientos procedentes del norte, a 90° vientos procedentes del este, a 180° vientos procedentes del sur y 270° del oeste [meteo-navarra-viento].
- **Radiación** (MJ/m<sup>2</sup>): Es la energía emitida por el sol que llega a la tierra en forma de ondas electromagnéticas. Se mide en milijúlios por metro cuadrado.
- **Precipitación** (mm): Es la cantidad de lluvia que se registra en un día en concreto. Se mide en milímetros.
- **Evapotranspiración de referencia (ETO)**: La evapotranspiración es la combinación de dos sucesos, la evaporación del agua del suelo y la transpiración de los cultivos. Esto nos ayuda a medir la cantidad de agua que se pierde por estos dos sucesos [sencrop-eto].

ETO viene del término evapotranspiración de referencia y es la cantidad de agua demandada de la atmósfera usando, como su nombre indica, un cultivo de referencia. Esto se usa para calcular la ETc, que es la evapotranspiración real de ese cultivo, junto con un coeficiente Kc. Cada tipo de planta tiene su Kc. Además, para calcular el ETO se usan la radiación, la velocidad del viento, la temperatura y la humedad.

Las estaciones se han clasificado por provincias. Andalucía cuenta con 110 estaciones , número que varía a lo largo del tiempo por bajas, altas y periodos de inactividad.

### **Almería:**

- Adra
- Almería
- Cuevas de Almanzora
- Fiñana
- Huércal-Overa
- La Mojonera
- Níjar
- Tabernas
- Virgen de Fátima–Cuevas de Almanzora
- Tíjola

### **Cádiz**

- Basurta–Jerez de la Frontera
- Conil de la Frontera
- IFAPA Centro de Chipiona
- Jerez de la Frontera
- Jimena de la Frontera
- Puerto de Santa María
- Sanlúcar de Barrameda
- Villamartín

- Vejer de la Frontera

### **Córdoba**

- Adamuz
- Baena
- Belmez
- Córdoba
- El Carpio
- Hornachuelos
- IFAPA Centro de Cabra
- IFAPA Centro de Hinojosa del Duque
- Palma del Río
- Santaella

### **Granada**

- Almuñécar
- Baza
- Cádiar
- Huéneja
- IFAPA Centro Camino del Purchil
- Iznalloz
- Jerez del Marquesado
- Loja
- Puebla de Don Fabrique
- Pinos Puente
- Zafarraya
- Padul
- Pinos Puente Casanueva

### **Huelva**

- Almonte

- Almonte bajo plástico
- Almonte bajo plástico frb
- Aroche
- El Campillo
- El Tojalillo–Gibraleón
- Gibraleón
- Gibraleón – Manzorales
- IFAPA Centro de Huelva. El Cebollar bajo plástico
- IFAPA Centro Huelva. Finca el Cebollar
- La Palma del Condado
- La Puebla de Guzmán
- Lepe
- Lepe bajo plástico
- Moguer
- Moguer El Fresno
- Niebla
- Palos de la Frontera bajo plástico

### **Jaén**

- Alcaudete
- Chiclana de Segura
- Huesa
- IFAPA Centro Mengíbar
- Jaén
- Jódar
- La Higuera de Arjona
- Linares
- Mancha Real
- Marmolejo

- Mengíbar
- Pozo Alcón
- San José de los Propios
- Sabiote
- Torreblascopedro
- Ubeda
- Santo Tomé
- Torreperogil
- Villacarrillo

### **Málaga**

- Antequera
- Archidona
- Cártama
- Estepona
- IFAPA Centro de Campanillas
- IFAPA Churriana
- Málaga
- Vélez-Málaga
- Sierra Yeguas
- Pizarra

### **Sevilla**

- Aznalcázar
- Ecija
- Guillena
- IFAPA Centro Las Torres-Tomejil
- IFAPA Centro Las Torres-Tomejil. Finca Tomejil
- IFAPA Centro de Los Palacios
- Isla Mayor

- La Luisiana
- La Puebla del Río
- La Puebla del Río II
- La Rinconada
- Las Cabezas de San Juan
- Lebrija I
- Lebrija II
- Lora del Río
- Los Molares
- Los Palacios y Villafranca
- Villanueva del Río y Minas
- Villafranco del Guadalquivir
- Osuna
- Sanlúcar la Mayor
- Puebla Cazallar



UNIVERSIDAD  
DE MÁLAGA

| [uma.es](http://uma.es)

E.T.S de Ingeniería Informática  
Bulevar Louis Pasteur, 35  
Campus de Teatinos  
29071 Málaga

E.T.S. DE INGENIERÍA INFORMÁTICA