

Análisis del volumen útil diario del embalse de El Peñol de 2010
a 2023 a partir de datos funcionales.

Sebastián Giraldo Gómez

sgiraldog7@eafit.edu.co

Director

Santiago Ortiz

sortiza2@eafit.edu.co

Maestría en Ciencia de los Datos y Analítica

Escuela de Ciencias Aplicadas e Ingeniería

Universidad EAFIT, Medellín, Colombia

Palabras clave: embalse peñol,
hidroelectricidad, volumen útil diario, datos
funcionales, generación hidroeléctrica.

Análisis del volumen útil diario del embalse de El Peñol de 2010 a 2023 a partir de datos funcionales.

1 Resumen

Este estudio analiza el comportamiento hidroeléctrico del embalse El Peñol, con énfasis en su dinámica histórica. Se realizaron comparaciones con cuatro embalses colombianos: El Peñol, Playas, Punchiná y San Lorenzo. Para ello, se aplicaron técnicas de estadística funcional a datos históricos del período 2010-2023 proporcionados por XM, junto con información sobre los fenómenos de El Niño y La Niña obtenida del Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM). Las variables analizadas incluyen el volumen turbinado, el volumen útil diario, la generación total de energía y los precios de mercado, con el objetivo principal de identificar patrones temporales, tendencias estacionales y relaciones funcionales entre dichas variables.

El análisis abarcó el cálculo de medias funcionales, la estimación de varianzas funcionales y la aplicación de análisis de componentes principales funcional (PCA funcional). Estas técnicas permitieron reducir la dimensionalidad de los datos y comprender los principales factores que influyen en el comportamiento hidroeléctrico. Como parte de la metodología, se utilizó el suavizado de Fourier para representar las variables en curvas continuas, facilitando la eliminación de ruido y la captura de tendencias subyacentes. Este enfoque permitió realizar comparaciones funcionales entre los embalses, destacando tanto similitudes como diferencias en su operación.

Los resultados de este análisis funcional proporcionan una base sólida para interpretar los patrones hidrológicos en la región de Antioquia, con especial atención al embalse El Peñol, y su impacto en la eficiencia hidroeléctrica regional. Este embalse, uno de los más importantes del país, enfrenta desafíos significativos derivados de las fluctuaciones en la disponibilidad hídrica y los efectos del cambio climático, subrayando la necesidad de estrategias de gestión sostenibles. En este contexto, se desarrollaron indicadores funcionales para evaluar la sostenibilidad de la operación del embalse y proponer mejoras en su gestión. Este estudio contribuye al avance de herramientas analíticas específicas para la gestión hidroeléctrica en Colombia, estableciendo además precedentes para investigaciones futuras dirigidas a embalses con características similares tanto a nivel regional como internacional.

2 Descripción del Proyecto

2.1 Planteamiento del Problema

La escasez de agua representa un desafío global con implicaciones significativas para el bienestar humano y los sistemas económicos. Según estimaciones, para el año 2025, aproximadamente 1,8 mil millones de personas vivirán en regiones con escasez absoluta de agua, mientras que dos tercios de la población mundial podrían enfrentarse a condiciones de estrés hídrico (Rosegrant et al., 2002). Este problema es particularmente relevante en regiones vulnerables como América Latina, donde el cambio climático, el crecimiento poblacional y la urbanización acelerada incrementan la presión sobre los recursos hídricos disponibles.

Esta problemática encuentra respaldo en el trabajo de Pérez et al. (2019), quienes, mediante el análisis de la escasez hídrica subregional utilizando la metodología de huella hídrica, identificaron conflictos significativos en diversas regiones con severos déficits de agua. A nivel rural, Leroy (2019) resalta la percepción de los agricultores sobre la creciente escasez de agua en los páramos de Colombia y Venezuela, destacando cómo la reducción de este recurso afecta directamente a la sostenibilidad agrícola y las comunidades locales.

En el caso colombiano, diversos estudios han evidenciado un aumento gradual en la temperatura promedio durante las últimas décadas. Según el IDEAM, la temperatura media anual ha aumentado en aproximadamente 0.8°C en los últimos 30 años. Las proyecciones climáticas indican que esta tendencia continuará, con incrementos potenciales de entre 1.5°C y 2.5°C hacia el año 2050, dependiendo de los escenarios de emisiones de gases de efecto invernadero (Hincapie et al., 2019).

Este aumento en la temperatura impacta directamente la disponibilidad de agua. Una mayor evaporación y la disminución en la acumulación de nieve en zonas montañosas reducen la recarga de acuíferos y el caudal de los ríos, elementos fundamentales para el abastecimiento de agua potable y la generación de energía hidroeléctrica (Rodríguez et al., 2020).

La hidroeléctrica, como principal fuente de energía renovable en Colombia, depende en gran medida del recurso hídrico. Este proceso, que aprovecha la energía cinética y potencial del agua en movimiento para producir electricidad, requiere la construcción de represas que permitan almacenar agua en embalses. Cuando es

necesario generar electricidad, el agua almacenada se libera a través de turbinas que accionan los generadores eléctricos (Laghari et al., 2013).

Colombia ocupa una posición estratégica en términos de recursos energéticos, con una matriz energética diversificada donde predominan las fuentes hidroeléctricas. Actualmente, cerca del 70% de la electricidad del país proviene de plantas hidroeléctricas, gracias a la abundancia de recursos hídricos (Desbureaux & Rodella, 2019), Sin embargo, el crecimiento económico, la industrialización y el aumento poblacional han elevado la demanda de energía, lo que plantea desafíos importantes para garantizar la sostenibilidad y seguridad del suministro energético futuro.

La alta dependencia de la hidroelectricidad convierte al sistema energético de Colombia en un sistema particularmente vulnerable a las variaciones hidrológicas. Fluctuaciones en los patrones de precipitación y temperatura pueden ocasionar reducciones significativas en los caudales de los ríos y en los niveles de agua de los embalses, afectando directamente la capacidad de generación de energía (Tzanakakis et al., 2020).

Estas proyecciones también reflejan un aumento continuo en la demanda de electricidad. Según el *Plan Energético Nacional 2050*, la demanda de energía crecerá a una tasa promedio anual del 2.7%, impulsada por el desarrollo económico, la urbanización y el crecimiento poblacional (Rodríguez Hernández et al., 2015).

Por consiguiente, se destaca la importancia de gestionar de manera eficiente y responsable el recurso hídrico. Es crucial implementar planes y criterios que evalúen el

comportamiento del agua en distintas áreas geográficas y que identifiquen cómo las limitaciones en la disponibilidad de este recurso podrían convertirse en un obstáculo para el desarrollo sostenible de las comunidades (Gleik, 2018).

En el caso de Antioquia, el embalse El Peñol-Guatapé se rige como un componente estratégico en la generación hidroeléctrica en Colombia, aunque es una de las fuentes más importantes de energía en la región, enfrenta desafíos derivados del cambio climático, como el fenómeno de El Niño, que reduce los niveles de agua y afecta la generación energética, impactando la economía y la estabilidad social de la región (Jaramillo Gómez, 2020).

Por lo tanto, esta investigación busca fortalecer la resiliencia del sistema hidroeléctrico colombiano, en particular del embalse El Peñol, frente a los desafíos del cambio climático. La capacidad de prever cambios en los patrones hidrológicos y desarrollar estrategias de gestión adaptativas es esencial para garantizar un suministro eléctrico sostenible y eficiente (Angarita et al., 2018). Además, los hallazgos de este estudio tienen el potencial de aplicarse en otros contextos geográficos con desafíos similares, ampliando su impacto y relevancia.

2.2 Justificación

Colombia depende en gran medida de la hidroelectricidad para su suministro eléctrico, con aproximadamente el 70% de su electricidad proveniente de fuentes hidroeléctricas (Desbureaux & Rodella, 2019). Esta alta dependencia hace que el país sea vulnerable a las variaciones en los patrones de precipitación y a los efectos del cambio

climático, que pueden alterar significativamente la disponibilidad de agua para la generación de energía. Por lo tanto, la gestión eficiente y sostenible de los recursos hidroeléctricos es crucial para garantizar la estabilidad energética del país.

El cambio climático es el desafío más importante de la humanidad para el presente siglo, por lo que se espera que este altere los patrones de precipitación y aumente la frecuencia y severidad de eventos extremos, como sequías e inundaciones. Estos cambios impactan directamente la capacidad de generación hidroeléctrica (Tzanakakis et al., 2020). Por esta razón se realizará análisis de datos funcionales que puede proporcionar una comprensión más detallada de estos patrones y, por consiguiente, ayudar a desarrollar estrategias adaptativas más efectivas (Rodriguez et al., 2020).

Además, con el crecimiento poblacional y económico, la demanda de energía en Colombia está en aumento. Es crucial optimizar la generación hidroeléctrica para satisfacer esta demanda sin comprometer la sostenibilidad ambiental (Desbureaux & Rodella, 2019). Los modelos de series temporales multivariados permiten prever la demanda y ajustar la generación en consecuencia (Leroy, 2019).

Frente a lo mencionado, el uso de técnicas avanzadas de análisis de datos puede mejorar significativamente la gestión y operación de estos sistemas energéticos (Asprilla Echeverría, 2021). Es por esto por lo que estas técnicas son necesarias para lograr una visión que pueda encaminar otros proyectos para la gestión eficiente de los embalses hidroeléctricos es fundamental para maximizar su capacidad de generación y minimizar los impactos ambientales.

Esta investigación es crucial para abordar el problema de la gestión de embalses hidroeléctricos en Colombia ante la escasez de agua y el cambio climático. Los resultados permitirán analizar los impactos del cambio climático en la disponibilidad de agua y la generación hidroeléctrica a nivel regional y nacional. Lo anterior para comprender los cambiantes patrones hidrológicos para desarrollar estrategias de gestión adaptativas, esenciales para un suministro eléctrico sostenible y eficiente, y así mejorar la resiliencia del sistema hidroeléctrico.

Justamente se pretende conocer el comportamiento las variables que puedan afectar la generación hidroeléctrica y su relación entre ellas, con esta información obtenida se puede revisar el estado actual de las hidroeléctricas estudiadas, la disponibilidad de agua, por lo que se podría plantear nuevas tácticas para la gestión para estos recursos hídricos.

Por lo que se evaluarán datos que pueden ser de gran reflexión para considerar la escasez hídrica, el cambio climático y las necesidades energéticas del país. Con lo anterior, el objetivo es proporcionar datos que mejoren la comprensión de la dinámica hidrológica de los embalses estudiados, mediante la evaluación de la capacidad de almacenamiento y de la regulación de los embalses, además de Identificar los impactos de los vertimientos de estos lo que permitirá, en un futuro, una mejor planificación y gestión de los recursos.

La investigación propuesta sobre la gestión de embalses hidroeléctricos en Colombia frente a la escasez de agua y el cambio climático es de gran relevancia para

el presente y futuro del país. Los resultados esperados contribuirán a la seguridad hídrica, energética y alimentaria, promoviendo el desarrollo sostenible en un contexto de cambio climático global.

2.3 Objetivos

2.3.1 Objetivo general:

Realizar un diagnóstico del comportamiento hidroeléctrico del embalse El Peñol estableciendo comparaciones con los embalses Punchina, San Lorenzo y Playas, mediante el estudio funcional de variables clave como el volumen turbinado, el volumen útil diario, la generación total de energía y los precios, utilizando datos históricos recopilados entre 2010 y 2023 con el fin de determinar los factores que afectan la sostenibilidad y eficiencia del sistema hidroeléctrico en Colombia.

2.3.2 Objetivos específicos:

- Describir la variabilidad temporal del volumen útil diario del embalse El Peñol examinando la variabilidad utilizando la media funcional y gráficos representativos para identificar tendencias clave.
- Comparar las dinámicas de comportamiento de años atípicos entre los embalses evaluando las dinámicas del embalse El Peñol frente a tres embalses colombianos, mediante el análisis de variables como volumen útil, volumen turbinado, generación total y precios.
- Identificar patrones en el volumen útil del embalse El Peñol mediante FPCA utilizando el análisis de componentes principales funcionales (FPCA) para

identificar en el volumen útil del embalse El Peñol, explorando su comportamiento integrado durante el período estudiado.

- Agrupar patrones de comportamiento mediante análisis de clúster funcional analizando el comportamiento y caracterizar las tendencias predominantes en los grupos formados.

2.4 Estado del arte y Marco teórico

El cambio climático está generando alteraciones significativas en los patrones de precipitación, aumentando la frecuencia de eventos extremos como sequías e inundaciones. Estas variaciones impactan directamente la disponibilidad y calidad del agua, afectando sectores clave como la generación de energía, la agricultura y el suministro de agua potable (Tzanakakis et al., 2020). En este contexto, resulta crucial analizar y gestionar los recursos hídricos para garantizar su sostenibilidad y eficiencia, particularmente en países como Colombia, donde la hidroelectricidad representa aproximadamente el 70% de la matriz energética (Desbureaux & Rodella, 2019).

A continuación, se presenta un estado del arte que analiza las principales metodologías y estudios relevantes sobre la gestión de recursos hidroeléctricos, con énfasis en el uso de modelos avanzados de análisis, incluyendo **series temporales multivariadas** y **análisis de datos funcionales (FDA)**, destacando sus aplicaciones en el contexto colombiano.

Datos Longitudinales

Los datos longitudinales se refieren a observaciones recolectadas repetidamente de las mismas unidades o sistemas a lo largo del tiempo, lo que permite capturar la evolución y dinámica de las variables de interés. Este enfoque es fundamental para el análisis hidrológico, ya que facilita la identificación de patrones temporales, tendencias de largo plazo y el impacto de eventos atípicos en sistemas como los embalses (Verbeke & Molenberghs, s. f.).

En el contexto de este trabajo, los datos longitudinales del volumen útil diario y otras variables clave de los embalses estudiados permiten evaluar tanto la variabilidad intraanual como las diferencias interanuales, proporcionando información valiosa para comprender las dinámicas de operación y los efectos de fenómenos climáticos como El Niño y La Niña. Además, su uso mejora la precisión de los modelos predictivos y optimiza la toma de decisiones para la gestión sostenible de los recursos hídricos (Fitzmaurice et al., 2011).

Modelos de Series Temporales Multivariados

Los modelos de series temporales multivariados son herramientas analíticas que permiten evaluar de manera conjunta múltiples variables que influyen en la generación hidroeléctrica, como la precipitación, el caudal de los ríos, la temperatura y la demanda de energía. Estos modelos son capaces de capturar la interdependencia entre las variables, mejorando la precisión de las predicciones tanto a corto como a largo plazo (Leroy, 2019).

En América Latina, Desbureaux y Rodella (2019), analizaron el impacto económico de la escasez de agua en áreas metropolitanas utilizando estos modelos. Su enfoque permitió prever la demanda de agua y ajustar la oferta según la disponibilidad hídrica y las condiciones climáticas. En Colombia, Pérez et al. (2019) aplicaron modelos multivariados junto con el concepto de huella hídrica para evaluar la escasez hídrica subregional, destacando la importancia de integrar múltiples fuentes de datos para una gestión más efectiva de los recursos.

Análisis de Datos Funcionales (FDA)

FDA se centra en la representación de datos continuos en el tiempo, como series temporales hidrológicas y climáticas, proporcionando herramientas para identificar patrones subyacentes y tendencias a largo plazo (Rodríguez et al., 2020). Este enfoque es especialmente valioso en contextos de recursos hídricos, donde las dinámicas temporales son complejas y requieren técnicas avanzadas para su modelado.

En el ámbito internacional, diversos estudios han aplicado el FDA en contextos hidrológicos. Por ejemplo, (Quintela-del-Río & Francisco-Fernández, 2024) utilizaron métodos no paramétricos de FDA para predecir el caudal mensual promedio y analizar valores extremos en ríos de Estados Unidos, contribuyendo a una mejor gestión del riesgo de inundaciones. Asimismo, (Massetot et al., 2016) implementaron modelos de regresión funcional para pronosticar el caudal de ríos, demostrando la eficacia de este enfoque en la previsión hidrológica.

Estas aplicaciones ilustran cómo el FDA puede extenderse a la gestión hidroeléctrica para abordar desafíos específicos, como la predicción de caudales, la evaluación de riesgos de inundación y la planificación sostenible de los recursos hídricos.

Impacto del Cambio Climático en la Hidroelectricidad

El cambio climático es un factor crítico que afecta la disponibilidad y gestión de los recursos hídricos, impactando directamente la generación hidroeléctrica. Las variaciones en los patrones de precipitación y temperatura reducen la capacidad de generación de las plantas hidroeléctricas, especialmente en regiones donde estas representan una proporción significativa del suministro energético (Bluma, 2020).

Diversos estudios han demostrado la necesidad de integrar modelos predictivos avanzados con datos climáticos y funcionales para desarrollar estrategias de adaptación y mitigación efectivas. Unfried et al. (2022). subrayan el papel de las tecnologías de análisis de datos y los sistemas de información geográfica (SIG) en la gestión de recursos hídricos, facilitando la recopilación, análisis y visualización de grandes volúmenes de datos, lo que mejora la capacidad de respuesta y la toma de decisiones.

2.5 Metodología

En la metodología de este trabajo se utiliza la guía que propone CRISP-DM (Cross Industry Standard Process for Data Mining) (Chapman, 1999), esto debido a que es una estrategia general y muy utilizada, además puede tener adaptaciones a las aplicaciones de minería de datos y proyectos de ciencia de datos de diferentes dominios (Martinez-

Plumed et al., 2021) ,de este estudio se estructura en varias etapas clave , que incluyen la recopilación y preprocesamiento de datos, el desarrollo y aplicación de modelos , la validación y comparación de resultados. A continuación, se detallan cada una de estas etapas.

La recopilación de datos se realiza mediante la API de XM, administrador del mercado eléctrico colombiano, y las bases históricas del IDEAM. Las observaciones abarcan variables clave como el volumen útil diario (x_{1t}) el volumen turbinado (x_{2t}) ,la generación total (x_{3t}) y los precios del mercado (x_{4t}) entre 2010 y 2023.

El conjunto de datos se representa como series temporales longitudinales:

$$\mathbf{X} = \{x_{it} \mid i = 1, \dots, N; t = 1, \dots, T\},$$

donde N es el número de variables y T representa los puntos temporales. Este enfoque captura dinámicas interanuales y estacionales.

El preprocesamiento garantiza que los datos estén en un formato adecuado para su análisis funcional, siguiendo los pasos propuestos por Han et al. (2011):

2.5.1 Limpieza de Datos

- Eliminación de duplicados: Identificación y eliminación de registros x_i repetidos.
- Imputación de valores faltantes: La imputación de valores faltantes utilizando la metodología **hacia adelante** (*forward-fill*) y **hacia atrás**

(*backward-fill*) consiste en reemplazar los valores faltantes en una serie temporal con el valor válido más cercano en una dirección específica.

$$x_{i,t} = \begin{cases} x_{i,t-1} & \text{si } x_{i,t} \text{ está faltante y } t > 1 \text{ (forward - fill)} \\ x_{i,t+1} & \text{si } x_{i,t} \text{ está faltante y } t < 1 \text{ (backward - fill)} \end{cases}$$

2.5.2 Construcción de Bases Funcionales

Para representar los datos longitudinales como funciones continuas $f_i(t)$, se utilizan bases funcionales. Entre las opciones, destacan las bases Fourier para datos periódicos y las bases B-spline para datos con cambios abruptos (Ramsay & Silverman, 2005).

Una función aproximada se define como:

$$f_{i(t)} = \sum_{\{k=1\}}^K c_k \phi_{k(t)},$$

Donde:

- $f_i(t)$ es la función suavizada o aproximada correspondiente a la i -ésima unidad de análisis (por ejemplo, el i -ésimo año o el i -ésimo embalse).
- $\phi_k(t)$ denota la k -ésima función base (por ejemplo, funciones de Fourier o B-splines).
- C_k es el coeficiente asociado a la k -ésima función base en la i -ésima curva.
- K es el número total de funciones base utilizadas para representar la curva.

El error de ajuste o MSE:

$$\text{MSE} = \frac{1}{T} \sum_{t=1}^T (x_{it} - f_i(t))^2.$$

Donde:

- x_{it} es el valor observado (por ejemplo, el volumen útil) para la i -ésima unidad de análisis en el tiempo t .
- T es el número total de instantes en los que se observan datos (p. ej., 365 días).
- $f_i(t)$ es la curva funcional que aproxima esos datos.

2.5.3 Suavizado Funcional

El suavizado elimina el ruido en los datos, destacando tendencias clave. Se emplean técnicas como penalización en B-splines y Fourier (Ramsay & Silverman, 2005).

Penalización en B-splines: El ajuste de suavizado incorpora un término de penalización λ para controlar la complejidad:

$$J[f] = \sum_{t=1}^T (x_t - f(t))^2 + \lambda \int (f''(t))^2 dt,$$

Donde:

- $\sum_{t=1}^T (x_t - f(t))^2$ es la suma de residuos al cuadrado (mide el ajuste de la función a los datos).
- $\int (f''(t))^2 dt$ es el término de curvatura que penaliza la excesiva ondulación de la curva
- λ es el parámetro de suavizado

donde $f''(t)$ es la derivada segunda de $f(t)$ La selección de λ se realiza minimizando el criterio de validación cruzada (CV):

$$CV = \frac{1}{N} \sum_{i=1}^N (x_{it} - f_{-i}(t))^2.$$

Donde:

- x_{it} denota los valores reales observados para la i -ésima serie o individuo en un instante temporal t .
- $f_{-i}(t)$ hace referencia a la función ajustada **excluyendo** el i -ésimo individuo del conjunto de entrenamiento
- N es el número total de observaciones

Representación Fourier: Para datos periódicos, el suavizado Fourier representa $f(t)$ como:

$$f(t) = a_0 + \sum_{k=1}^K [a_k \cos(2\pi kt) + b_k \sin(2\pi kt)].$$

Donde:

- a_0 representa la componente promedio (offset), mientras que a_k y b_k son los coeficientes de amplitud para la k -ésima frecuencia.
- K indica cuántos términos de Fourier se emplean

2.5.4 Cálculo de Medias y Varianzas Funcionales

La media funcional describe el comportamiento promedio a lo largo del tiempo:

$$\mu(t) = \frac{1}{N} \sum_{i=1}^N f_i(t).$$

Donde:

- $\mu(t)$ es la **curva promedio** obtenida a partir de un conjunto de N funciones $f_i(t)$.
- $f_i(t)$ representa la evolución de la variable de estudio en el tiempo para el i -ésimo año, embalse.

La varianza funcional mide la dispersión en el tiempo:

$$\sigma^2(t) = \frac{1}{N} \sum_{i=1}^N [f_i(t) - \mu(t)]^2.$$

Estas métricas permiten identificar patrones generales y periodos de alta variabilidad.

2.5.5 Análisis de Componentes Principales Funcionales (FPCA)

El FPCA reduce la dimensionalidad al identificar las principales fuentes de variación en las funciones. Se basa en la descomposición espectral de la matriz de covarianza funcional $C(s,t)$:

$$C(s, t) = \frac{1}{N} \sum_{i=1}^N [f_i(s) - \mu(s)][f_i(t) - \mu(t)].$$

Las componentes principales $\phi_k(t)$ se obtienen resolviendo:

$$\int C(s, t) \phi_k(s) ds = \lambda_k \phi_k(t),$$

donde λ_k son los valores propios que indican la varianza explicada por cada componente. La proporción de varianza explicada se calcula como:

$$\text{Varianza explicada} = \frac{\sum_{k=1}^m \lambda_k}{\sum_{k=1}^N \lambda_k}.$$

2.5.6 Clustering Funcional

Se utiliza k-means funcional para agrupar curvas similares. La distancia entre curvas $f_i(t)$ y $f_j(t)$ se mide como:

$$d(f_i, f_j) = \sqrt{\int (f_i(t) - f_j(t))^2 dt}.$$

El algoritmo minimiza la suma de distancias intra-cluster (WCSS):

$$WCSS = \sum_{k=1}^K \sum_{f_i \in C_k} d^2(f_i, \bar{f}_k),$$

Donde (\bar{f}_k) , es el centroide del cluster C_k . El número óptimo de clusters (K) se determina mediante el método del codo encontrando el punto de inflexión en la curva.

Consideración de correlación entre diciembre y enero:

Si bien cada año se trató como una curva independiente (del 1 de enero al 31 de diciembre), es probable que exista correlación entre los valores de diciembre y los de enero del siguiente año, debido al carácter cíclico de la disponibilidad hídrica.

Para esta investigación, se mantuvo cada año como unidad de análisis, reconociendo que se podrían aplicar enfoques cíclicos para mejorar la modelación de la transición entre finales de un año e inicios del siguiente.

2.6 Plan de Gestión de Datos

Para el desarrollo del proyecto se usaran datos extraídos de las bases de datos de la empresa administradora del mercado eléctrico colombiano (XM) , el conjunto de registros será extraído a través de consultas específicas a la interfaz de aplicaciones (API) propia de XM, contendrá datos estructurados, además datos del IDEAM, es importante destacar que los datos son de acceso público y se pueden encontrar en el

portal web de las empresas, luego se creara una base de datos con estas dos bases de datos mencionadas.

2.7 Aspectos éticos

Las bases de datos seleccionadas para la elaboración de este trabajo son de carácter público, obtenidos de la fuente de datos de XM y del IDEAM. Asimismo, no presentan ninguna restricción legal para el procesamiento y análisis de estos datos, además el análisis es netamente para carácter académico a menos de considerar necesario se puede profundizar y compartir los análisis resultantes, de los cuales se obtendrán beneficios al ofrecer un panorama que impacte en estrategias que puedan mejorar la gestión sostenible de los recursos hídricos en Colombia.

3 Resultados

Para la recolección de datos, se utilizó la API de XM para extraer información relacionada con los embalses Peñol, Punchiná, San Lorenzo y Playas, correspondiente a los años 2010 a 2023. Las variables recopiladas incluyeron: volumen útil turbinado, volumen turbinado, generación total y precio, especificadas para cada embalse.

Las variables tienen unidades de la siguiente manera

XM	Generacion_Recurso	kWh
XM	Precio_Oferta_Despacho	COP/kWh
XM	Volumen útil diario por Embalse	m ³

XM	Volumen Turbinado por Embalse	Miles de m ³
----	-------------------------------	-------------------------

Tabla 1. Unidades de medida de las variables de los embalses.

La Tabla 1 presenta una descripción clara de las fuentes de datos y las variables analizadas en el estudio. En la primera columna se detalla el origen de los datos, garantizando su trazabilidad y confiabilidad. La segunda columna enumera las variables clave, que abarcan aspectos como energía, costos y volúmenes hídricos, esenciales para el análisis. En la tercera columna se especifican las unidades de medida, como kilovatios hora (kWh), pesos colombianos por kilovatio hora (COP/kWh), metros cúbicos (m³) y miles de metros cúbicos (miles de m³).

Este proceso se llevó a cabo mediante un script desarrollado en Python, que permitió la extracción automática de los datos. Posteriormente, se realizaron tareas de limpieza y aseguramiento de la integridad de los datos, lo cual incluyó la identificación y manejo de datos duplicados y faltantes. Los registros duplicados fueron eliminados, mientras que los datos faltantes, que representaron menos del 3% del total, se imputaron utilizando la técnica de llenado hacia adelante y hacia atrás ("forward-fill" y "backward-fill").

Se plantea un enfoque principal en el embalse El Peñol, debido a su gran tamaño e importancia dentro del sistema energético de Colombia. Para este análisis, se considerarán los datos históricos de las variables relevantes, desglosados por año.

Los datos serán transformados en datos funcionales utilizando diversas funciones y herramientas proporcionadas por las siguientes librerías de R especializadas que ofrecen herramientas avanzadas para cada etapa del proceso. La librería **fda** (Ramsay

et al., 2009): constituye la base fundamental del análisis funcional, proporcionando métodos para construir bases funcionales, como Fourier y B-splines, así como técnicas de suavizado y ajuste de funciones.

Para el ajuste de modelos de regresión funcional, se utilizó `refund` (Goldsmith et al., 2024), que permite trabajar con modelos lineales, aditivos y multivariados, incluyendo datos funcionales longitudinales. Asimismo, `fda.usc` (Febrero Bande & Oviedo de la Fuente, 2024): amplió estas capacidades al ofrecer métodos para clasificación, regresión, simulación, análisis exploratorio y clustering de datos funcionales, complementando la funcionalidad básica de `fda`.

Por otro lado, la librería `roahd` (Ieva et al., 2019) aportó técnicas robustas para el análisis de datos funcionales en alta dimensión, enfocándose en la detección de valores atípicos y el cálculo de profundidad funcional. Para representar datos no periódicos o con comportamientos complejos, se emplearon las herramientas de splines ofrecidas por la librería `splines` (R: The R Project for Statistical Computing, n.d.).

Finalmente, las librerías `dplyr` (Wickham et al., 2024) y `readr` (Wickham et al., 2014) complementaron el análisis funcional al facilitar la manipulación eficiente de datos y la lectura de archivos, optimizando el preprocesamiento de la información. En conjunto, estas librerías permitieron realizar un análisis integral y robusto de los datos funcionales.

Para la suavización de datos se crea una base Fourier con 365 bases. A continuación, se define un operador diferencial para la suavización, utilizando coeficientes específicos para crear un operador diferencial de tipo armónico acelerado, además se ajusta un parámetro de penalización, λ , con un valor de (1×10^6) .

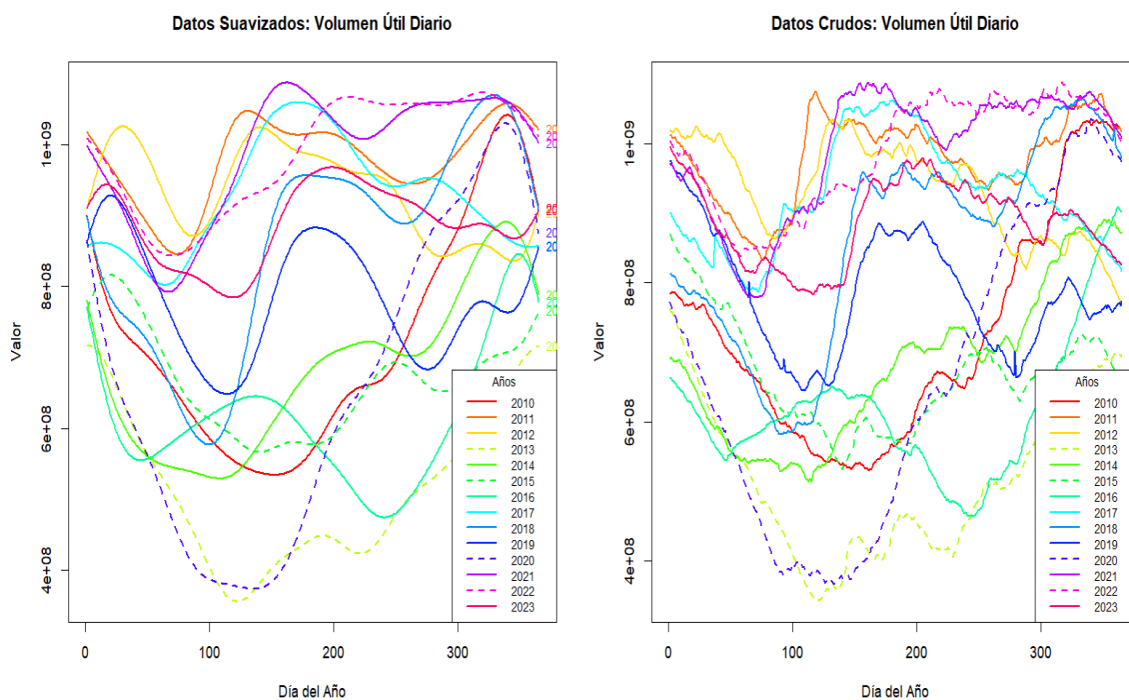


Figura 1. Comparativa del volumen útil diario del embalse El Peñol (2010-2023): datos suavizados y crudos.

La Figura 1 muestra las observaciones temporales del volumen útil diario del embalse El Peñol. En el panel derecho se presentan las series de tiempo, mientras que el panel izquierdo ilustra la suavización de Fourier de estas observaciones. El comportamiento del volumen útil diario evidenció variaciones significativas estrechamente relacionadas con los fenómenos climáticos de El Niño y La Niña, los cuales afectan de manera decisiva el régimen de precipitaciones en Colombia. Durante este período, se identificaron patrones específicos en los niveles de agua del embalse que reflejan la interacción entre estas condiciones climáticas globales y los procesos de gestión locales.

Entre 2010 y 2012, se observó un aumento sostenido en el volumen útil del embalse, particularmente marcado en 2011 y 2012. Este comportamiento coincidió con un evento de La Niña reportado por el IDEAM entre 2010 y 2011, caracterizado por un incremento significativo en las precipitaciones en gran parte del país. Estas condiciones húmedas favorecieron la recuperación hídrica del embalse, manteniendo niveles altos durante ese período.

En contraste, durante 2015 y 2016 se registró una disminución en los niveles de agua del embalse, con volúmenes útiles más bajos en diversas fases del año. Esta situación puede atribuirse a la influencia de un evento de El Niño que afectó a Colombia en ese período. Este fenómeno, conocido por reducir las lluvias y aumentar las temperaturas, generó un contexto de sequía que impactó negativamente la disponibilidad hídrica del embalse.

Entre 2020 y 2023, las variaciones interanuales en el volumen útil se tornaron más dispersas y marcadas, sugiriendo una creciente variabilidad climática. Este comportamiento coincidió con un evento de La Niña entre 2020 y 2021, seguido de condiciones climáticas neutrales y la transición hacia un nuevo evento de El Niño en 2023. La alternancia entre estas condiciones extremas probablemente amplificó las fluctuaciones en los niveles del embalse, a lo que podrían sumarse posibles ajustes en las estrategias operativas de gestión.

Un caso particular ocurrió en 2017, cuando se observaron patrones atípicos en el volumen útil, con desviaciones significativas respecto a las tendencias generales de otros años. Este comportamiento no parece estar directamente relacionado con los fenómenos de El Niño o La Niña, ya que las condiciones climáticas predominantes fueron mayormente neutrales. Las anomalías observadas podrían explicarse por factores locales, como eventos climáticos extremos puntuales o decisiones específicas en la operación del embalse.

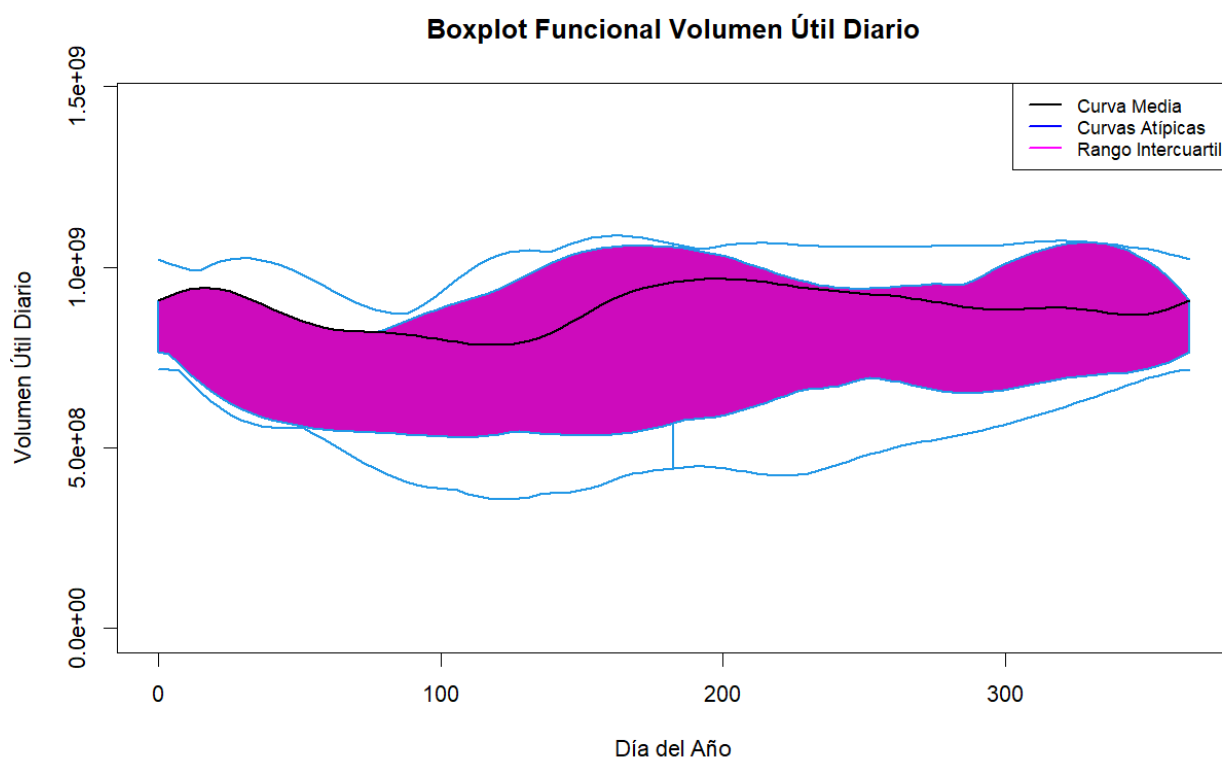


Figura 2. Boxplot funcional del volumen útil diario del embalse El Peñol (2010-2023).

La Figura 2 presenta el boxplot funcional del volumen útil diario, ofreciendo una síntesis visual integral que representa la dispersión, variabilidad y el comportamiento general de los datos observados entre 2010 y 2023. La mediana funcional, representada

por la línea negra central, resume el comportamiento central del volumen útil a lo largo del año. Esta curva indica un volumen relativamente estable con valores intermedios, que oscilan entre 5×10^8 y 10×10^9 durante la mayor parte del año.

El rango intercuartil funcional, representado por el área sombreada en magenta, captura la variabilidad central de los datos, mostrando que los valores típicos fluctúan en un intervalo considerablemente amplio. No obstante, en ciertos períodos del año, particularmente en los meses mediados y finales, se observa una dispersión más pronunciada que refleja una variabilidad significativa, posiblemente vinculada a fenómenos estacionales como lluvias intensas o sequías.

Los outliers funcionales, de existir, están representados como curvas que se sitúan fuera de los límites exteriores delineados en azul claro. Estos indican eventos atípicos, como sequías extremas o lluvias extraordinarias, que afectan el volumen útil del embalse.

En términos de estacionalidad y picos, el rango funcional es más amplio en períodos específicos del año, lo que sugiere una mayor variabilidad posiblemente atribuible a eventos climáticos extremos que alteraron significativamente el volumen útil del embalse. Al comparar esta representación con las gráficas de curvas anuales, el boxplot funcional condensa la información de las 14 curvas anuales, resaltando las áreas de mayor dispersión y la tendencia central, lo que facilita la identificación de períodos críticos y la evaluación de la consistencia temporal.

Aunque la mediana funcional muestra estabilidad relativa, los límites exteriores reflejan fluctuaciones notables en ciertos años, lo cual podría atribuirse a cambios climáticos o ajustes operativos en la gestión del embalse. Además, las áreas de mayor separación entre los límites exteriores se interpretan como períodos de mayor incertidumbre en el manejo del volumen útil, destacando patrones de riesgo potencial que requieren atención especial en la planificación y operación del sistema hídrico.

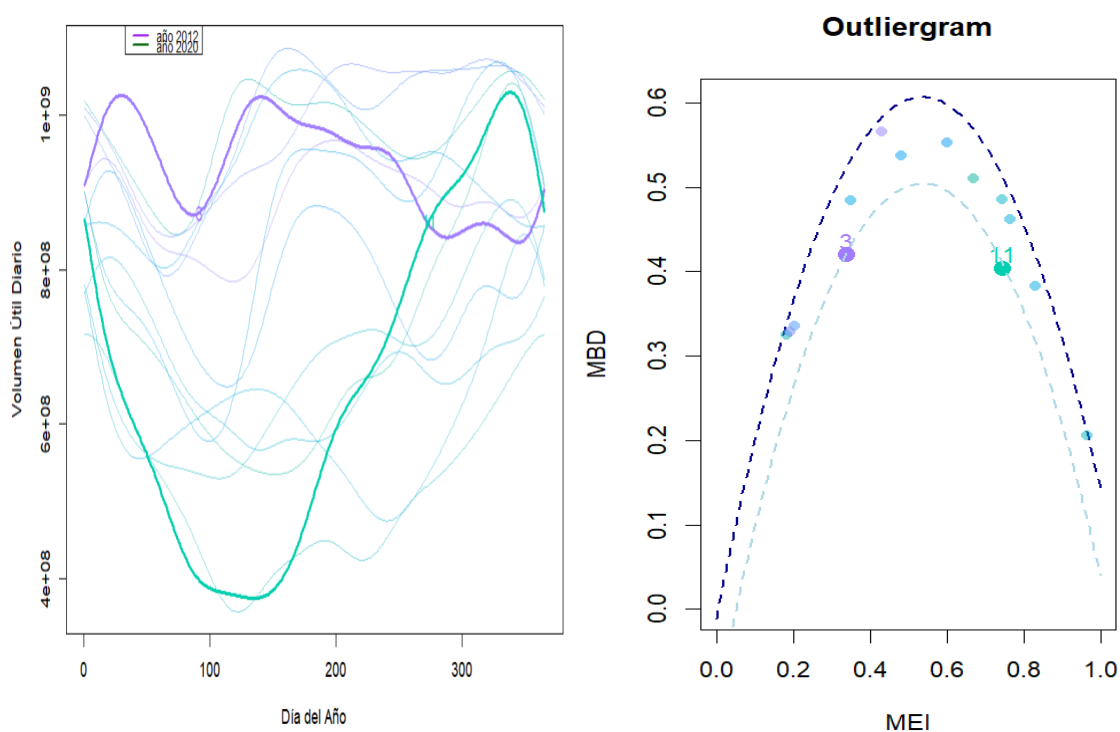


Figura 3. Outliergram del volumen útil diario del embalse El Peñol (2010-2023).

La Figura 3 presenta el outliergram. En el panel izquierdo se muestran las medidas de profundidad funcional (MBD, Modified Band Depth) y en el panel derecho, la

centralidad funcional (MEI, Modified Epigraph Index), herramientas utilizadas para identificar curvas que se desvían significativamente del comportamiento típico de los datos. Este análisis resulta especialmente valioso para detectar patrones anómalos en el volumen útil diario del embalse El Peñol.

En la gráfica de curvas de datos (izquierda), las observaciones identificadas en verde y morado se consideran posibles outliers. Estas curvas presentan patrones notablemente diferentes respecto al grupo central:

Curva verde (año 2020): Refleja un volumen útil inusualmente bajo hacia la mitad del año, lo que podría estar relacionado con un evento de sequía severa o una interrupción operativa en la gestión del embalse.

Curva morada (año 2012): Exhibe un comportamiento anómalo en el último tercio del año, caracterizado por un volumen útil inusualmente alto, posiblemente atribuible a lluvias extremas o cambios operativos significativos que permitieron un llenado extraordinario.

En el outliergram (derecha), el eje MEI representa la centralidad funcional, mientras que el eje MBD mide la profundidad funcional de las curvas en el rango general de los datos. Las observaciones que se sitúan fuera del rango típico, delimitado por líneas punteadas, se clasifican como posibles outliers.

En este caso, se identificaron dos observaciones atípicas que llaman la atención por su comportamiento divergente en relación con el volumen útil del embalse. La primera corresponde a un outlier inferior (curva verde), caracterizado por valores anormalmente bajos de volumen útil. Este comportamiento puede explicarse por un evento extraordinario, como una sequía extrema o una gestión operativa inusual que redujo considerablemente los niveles de agua disponibles.

Por otro lado, la segunda observación se clasifica como un outlier superior (curva morada), asociado con valores excepcionalmente altos de volumen útil. Este fenómeno puede atribuirse a condiciones excepcionales, como lluvias intensas que aumentaron el caudal del embalse, o a ajustes operativos específicos que permitieron un llenado extraordinario. Estas dos observaciones representan escenarios diametralmente opuestos en la gestión y disponibilidad hídrica, subrayando la importancia de analizar estos eventos para optimizar la operación del sistema.

En relación con el boxplot funcional, es probable que estas observaciones correspondan a las curvas que se acercaron o superaron los límites exteriores del rango intercuartil funcional. Esto resalta la utilidad complementaria de ambas herramientas, permitiendo identificar tanto tendencias generales como anomalías específicas en los datos.

El análisis de los dos atípicos presentes en el outliergram comienza con el año 2012, representado en color morado en la gráfica 3, y revela que

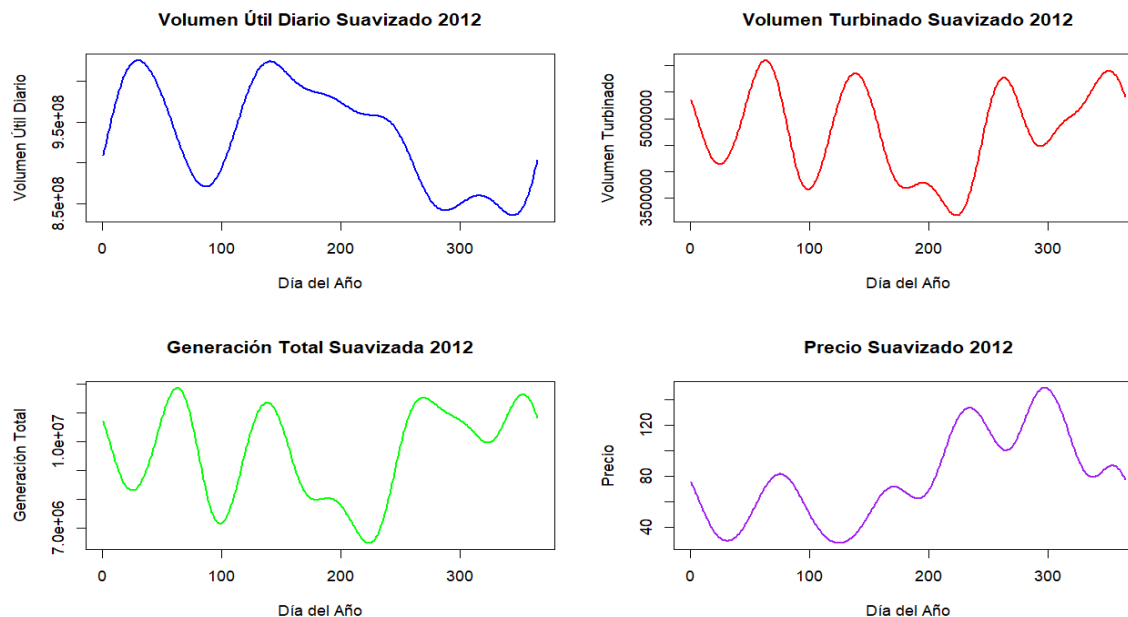


Figura 4. Datos suavizados del Volumen útil, turbinado turbando, Generación total y Precio del embalse El Peñol (2012).

En la Figura 4 se presentan las curvas suavizadas de las cuatro variables correspondientes al año 2012, destacando interacciones clave entre el volumen útil diario, el volumen turbinado, la generación total y los precios de la energía. Estas relaciones permiten identificar dinámicas esenciales en la gestión de recursos hídricos y la producción energética.

El volumen útil diario muestra una tendencia descendente significativa a lo largo del año, alcanzando un punto mínimo alrededor del día 300, seguido de un leve repunte. Esta reducción podría atribuirse a una disminución en las precipitaciones o a un aumento

en las demandas hídricas. Por otro lado, el volumen turbinado se mantiene relativamente constante, con oscilaciones periódicas y picos definidos, lo que indica una operación continua para satisfacer la demanda energética. La estabilidad del volumen turbinado, a pesar de la reducción en el volumen útil, sugiere una presión operativa que pudo haber acelerado el agotamiento de los recursos hídricos disponibles.

La generación total sigue un patrón similar al volumen turbinado, con oscilaciones periódicas sincronizadas. Sin embargo, se observa una disminución significativa hacia el día 200, posiblemente asociada con las limitaciones derivadas de la caída en el volumen útil. Esto refleja una relación directa entre la disponibilidad de agua en el embalse y la capacidad de generación energética, subrayando la dependencia del sistema en los recursos hídricos.

En contraste, los precios de la energía exhiben mayor variabilidad y picos moderados durante el año. Hacia el final del período analizado, los precios tienden a aumentar, coincidiendo con la disminución en el volumen útil. Esto sugiere que la escasez de recursos hídricos impactó en los costos. Este incremento podría estar relacionado con la necesidad de complementar la oferta energética con fuentes alternativas, posiblemente más costosas, aunque el aumento en los precios no es proporcional a la reducción en la generación total. Esto indica que se adoptaron medidas para mitigar el impacto de la disminución en la producción hidroeléctrica, como ajustes operativos o el uso de energía de respaldo.

A continuación, se analizarán los cuatro embalses: El Peñol, Playas, Punchiná y San Lorenzo, de manera conjunta, centrándose específicamente en el año 2012, que ha sido identificado como un outlier

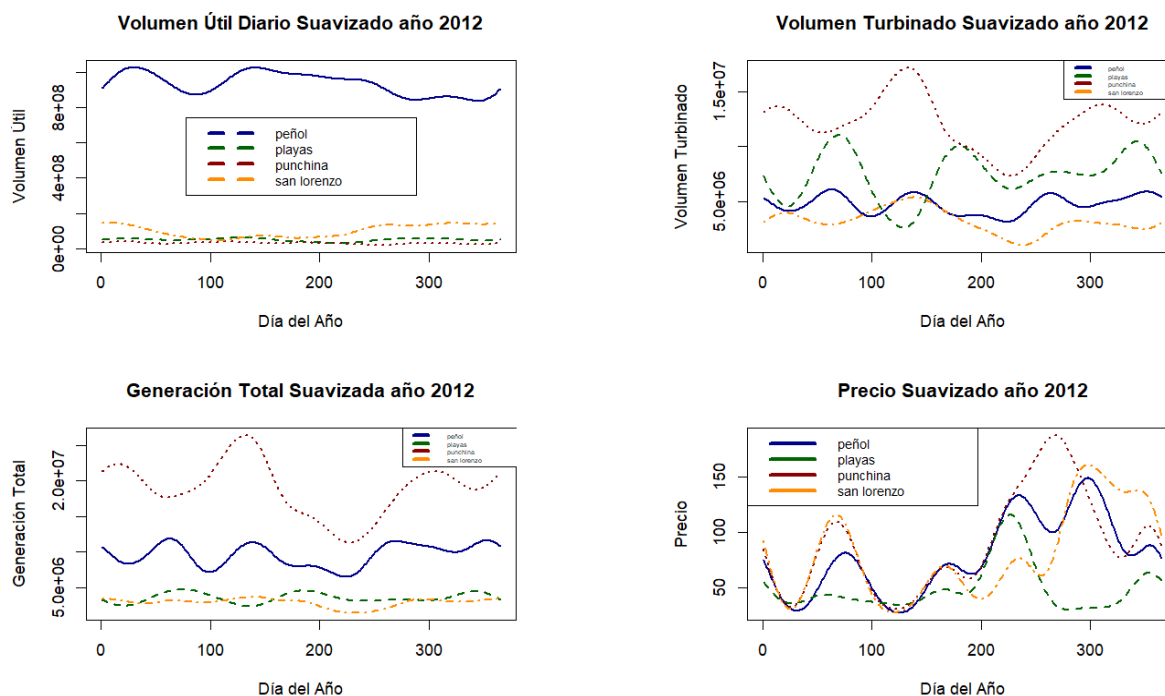


Figura 5. Comparativo año 2012, embalses Peñol, Punchina, Playas, San lorenzo.

La Figura 5 presenta una comparación de los cuatro embalses con datos suavizados, destacando el comportamiento de los embalses Peñol, Playas, Punchiná y San Lorenzo durante el año 2012. Se observan diferencias significativas en su operación y contribución al sistema energético, con una marcada dependencia en Punchiná como principal generador.

Aunque el embalse Peñol mantuvo el volumen útil más alto durante todo el año, su contribución a la generación de energía fue limitada, posiblemente debido a

restricciones operativas, a pesar de registrar los precios más altos del sistema, con picos hacia la mitad del año. En contraste, Punchiná, con un volumen útil significativamente más bajo, presentó los mayores picos de volumen turbinado y lideró la generación total. Esto evidencia una operación intensiva y eficiente para satisfacer la alta demanda energética.

El embalse Playas desempeñó un rol secundario, con volúmenes turbinados y generación relevantes, aunque más moderados en comparación con Punchiná. Por otro lado, San Lorenzo tuvo una contribución mínima tanto en generación como en volumen turbinado, reflejando su baja participación en la operación general del sistema.

Estas dinámicas operativas resaltan un desequilibrio en el uso de los embalses, donde Punchiná asumió la mayor carga operativa, mientras que los otros embalses, especialmente Peñol y San Lorenzo, jugaron roles mucho más limitados. Las asimetrías en los precios también reflejan estas diferencias. Peñol registró los costos más altos, posiblemente debido a su limitada capacidad para aumentar la generación en momentos críticos. En contraste, Punchiná, a pesar de ser el principal generador, mantuvo precios relativamente bajos, destacando su eficiencia operativa.

En conjunto, este análisis subraya cómo una fuerte dependencia de un único embalse, como Punchiná, puede generar un estrés operativo significativo y desequilibrios en el sistema. La subutilización de otros embalses limita la capacidad del sistema para aliviar dicha presión. Este comportamiento resalta la necesidad de comprender las

interacciones entre los embalses para evitar concentraciones operativas y mejorar la resiliencia del sistema ante eventos atípicos, como los ocurridos en 2012. Además, se señala que otra curva representativa en el outliergram corresponde al año 2020.

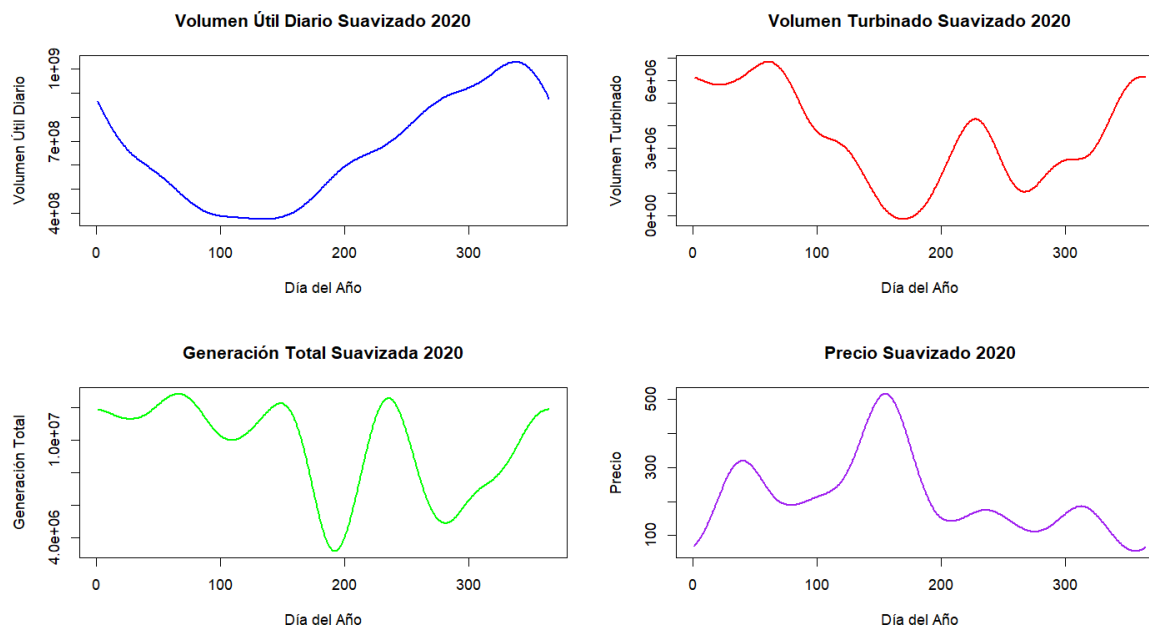


Figura 6. Datos suavizados del Volumen útil, turbinado turbando, Generación total, Precio del embalse El Peñol (2020).

La Figura 6 presenta las cuatro variables suavizadas correspondientes al año 2020, destacando interdependencias clave que explican el comportamiento del sistema energético durante ese período. El volumen útil diario muestra una caída pronunciada en los primeros 150 días del año, seguida de una recuperación gradual hacia el final del mismo. Esta disminución inicial parece haber tenido un impacto directo en los precios de la energía, que alcanzaron su valor máximo alrededor del día 150, probablemente debido a una menor disponibilidad hídrica que incrementó los costos de generación. Conforme

el volumen útil se recuperó, los precios comenzaron a estabilizarse, lo que indica una relación inversa entre estas dos variables.

Por otro lado, el volumen turbinado presenta fluctuaciones notables a lo largo del año, con picos destacados al inicio y al final del período, así como descensos significativos alrededor del día 200. Estas oscilaciones se reflejan directamente en la generación total, cuyo comportamiento sigue un patrón similar, evidenciando una dependencia operativa. La reducción en ambos indicadores hacia la mitad del año podría estar vinculada a ajustes estratégicos para preservar los recursos hídricos en respuesta a la disminución del volumen útil.

En conjunto, estos resultados evidencian cómo la recuperación progresiva del volumen útil hacia el cierre del año permitió estabilizar tanto la generación como los precios. Este análisis subraya la relación crucial entre la disponibilidad hídrica y las dinámicas del mercado energético, destacando la importancia de una gestión eficiente de los recursos hídricos para garantizar la sostenibilidad del sistema y mitigar las fluctuaciones en los precios.

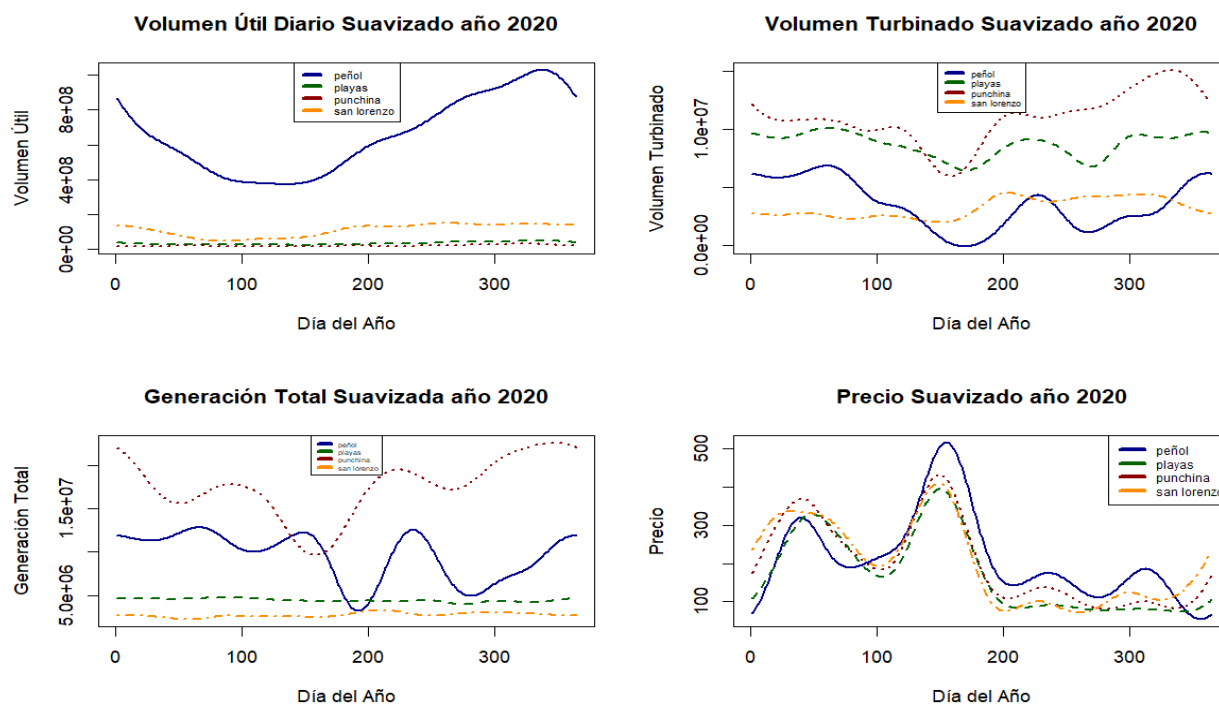


Figura 7. comparativo año 2020, embalses Peñol, Punchina, Playas, San lorenzo.

La Figura 7 ilustra un comparativo de las cuatro variables frente al año atípico 2020 y el comportamiento de los embalses. Asimismo, refleja una operación caracterizada por una fuerte dependencia de Punchina y Playas como principales generadores de energía, mientras que Peñol y San Lorenzo jugaron roles más secundarios. Peñol, gracias a su gran capacidad de almacenamiento, presentó inicialmente una tendencia descendente seguida de un repunte significativo después del día 200, manteniendo el volumen útil más alto y sostenido durante el año. Por el contrario, Punchina, Playas y San Lorenzo mostraron volúmenes útiles bajos y relativamente constantes, lo que limitó su capacidad para actuar como soporte operativo en periodos de alta demanda.

En términos de volumen turbinado, Punchina lideró con picos importantes a mediados y hacia el final del año, consolidándose como el principal generador. Playas también mostró un buen desempeño, aunque con volúmenes turbinados relativamente consistentes a lo largo del año. En contraste, Peñol y San Lorenzo registraron bajos volúmenes turbinados, lo que reflejó un menor grado de utilización operativa. Esto impactó directamente en la generación total, donde Punchina y Playas asumieron la mayor parte de la carga, mientras que la contribución de Peñol y San Lorenzo fue significativamente más limitada.

Las mismas dinámicas se reflejan en los precios de la energía. Peñol registró los precios más altos, con picos evidentes alrededor del día 100, probablemente debido a su menor contribución durante momentos críticos de generación, lo que incrementó el estrés operativo del sistema. En contraste, los precios en Punchina, Playas y San Lorenzo fueron más bajos y estables, reflejando una operación más eficiente y uniforme en estos embalses.

En conjunto, el año 2020 se caracterizó por una alta concentración de la generación en Punchina y Playas, lo que incrementó el estrés hídrico y redujo la capacidad de respuesta hacia el final del año. Paralelamente, la baja participación de Peñol y, especialmente, de San Lorenzo, limitó el equilibrio operativo, contribuyendo a picos de precios y a un aprovechamiento desigual de los recursos hídricos. Estas dinámicas subrayan la necesidad de alcanzar un balance más uniforme en la operación del sistema hidráulico para mitigar los efectos de años atípicos como este.

A continuación, se analizará el volumen útil diario del embalse El Peñol, desglosado en componentes principales funcionales (FPCA)

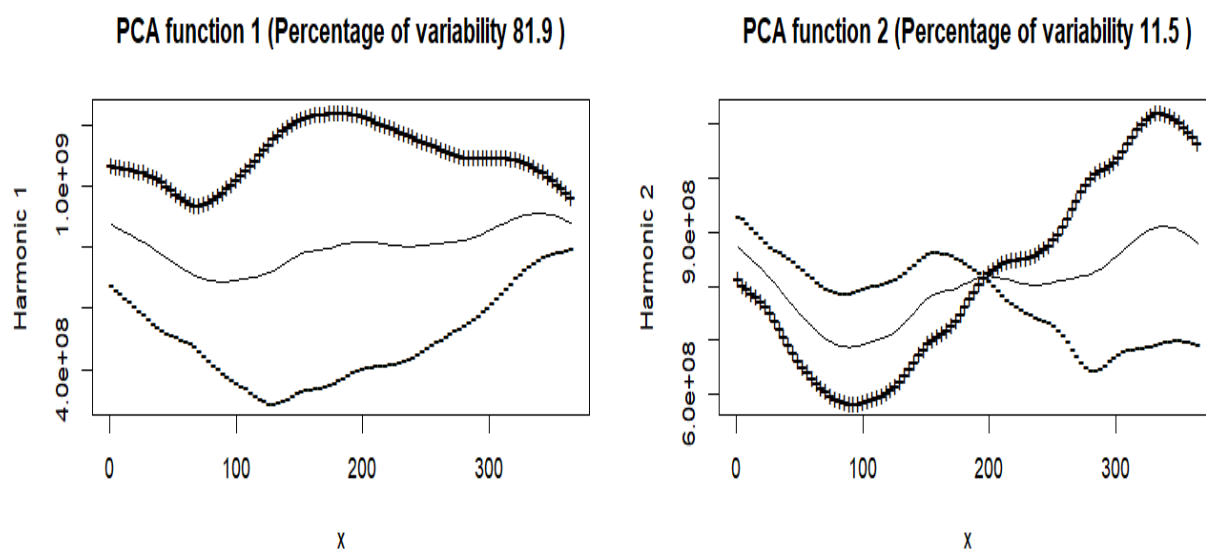


Figura 8. Análisis de componentes principales funcionales (FPCA) para volumen útil diario del embalse El Peñol (2010-2023).

En la Figura 8 se muestran las dos componentes principales. En la parte izquierda se encuentra la primera componente principal (FPCA 1), que captura el 81.9% de la variabilidad y refleja las tendencias generales del volumen útil a lo largo del año. La curva media, ubicada en el centro del gráfico, evidencia un descenso gradual en la primera mitad del año, seguido de una recuperación hacia el final. Este patrón representa el comportamiento típico del sistema hídrico, mientras que las desviaciones superior e inferior reflejan escenarios extremos: años con alta disponibilidad hídrica y años críticos

con estrés hídrico. Esta componente es crucial para interpretar las fluctuaciones más generales y recurrentes en los datos.

Por otro lado, en la misma figura, se encuentra la segunda componente principal (FPCA 2), que explica el 11.5% de la variabilidad y resalta fluctuaciones específicas, principalmente en la segunda mitad del año. A diferencia de la FPCA 1, su curva media es más plana, pero las oscilaciones se intensifican después del día 200. Esto sugiere que captura eventos interanuales diferenciados, como lluvias tardías, ajustes operativos o fenómenos climáticos no recurrentes. Las desviaciones superior e inferior de esta componente destacan, respectivamente, años con recuperaciones excepcionales hacia el final del año y años con condiciones adversas persistentes.

En conjunto, ambas componentes ofrecen un análisis complementario. Mientras la FPCA 1 identifica las tendencias dominantes y recurrentes, la FPCA 2 revela variaciones más sutiles y específicas, particularmente relacionadas con la dinámica hacia el final del año.

El análisis de las dos primeras componentes principales del FPCA confirma la naturaleza atípica de los años 2012 y 2020, destacando diferencias clave en sus comportamientos. La primera componente principal (FPCA 1) refleja las particularidades del año 2012, mientras que la segunda componente principal (FPCA 2) captura las del año 2020. Operativamente, esto implica que la FPCA 1 es útil para la planificación a largo

plazo basada en patrones generales, mientras que la FPCA 2 permite ajustar estrategias ante eventos inusuales o condiciones particulares de ciertos años.

El siguiente análisis es un clúster funcional, en el cual se utiliza una metodología de k-means funcional y una metodología para determinar el clúster óptimo, también conocido como método del codo. A continuación, se describe cómo se llevará a cabo este análisis.

Método del Codo

El método del codo es una técnica para determinar el número óptimo de clusters (K) al realizar análisis de clustering. Este enfoque se basa en la suma de distancias intra-cluster (WCSS), que mide cuán cerca están los puntos dentro de un cluster de su centroide.

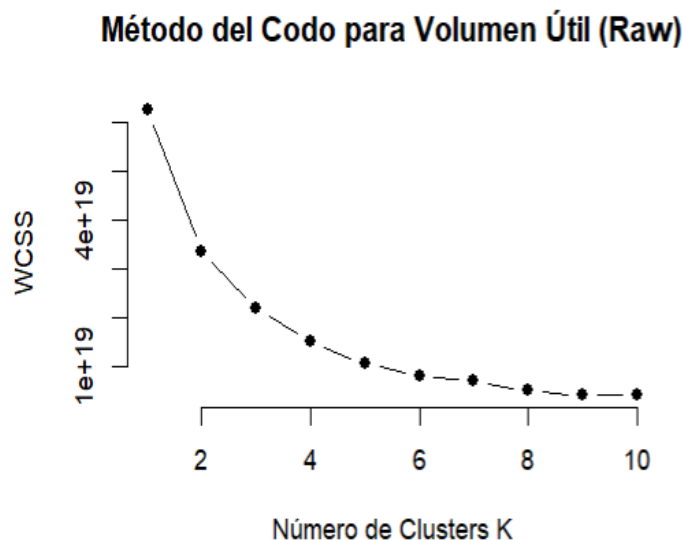


Figura 9. Método del codo para el volumen útil diario del embalse El Peñol (2010-2023)

En la Figura 9 se encuentra el eje X que representa el número de clusters (k) y el eje Y la suma de distancias intra-cluster (WCSS), la curva muestra una disminución rápida de WCSS a medida que aumenta el número de clusters (k), estabilizándose gradualmente el método del codo en este caso, el punto de inflexión ("codo") se encuentra alrededor de $k=3$ o $k=4$. Esto sugiere que dividir las curvas en 3 o 4 grupos es suficiente para capturar la mayor parte de la variabilidad sin sobreajustar.

Una vez encontrado el cluster óptimo $K=3$ procedemos a encontrar la clusterización con Kmeans funcional.

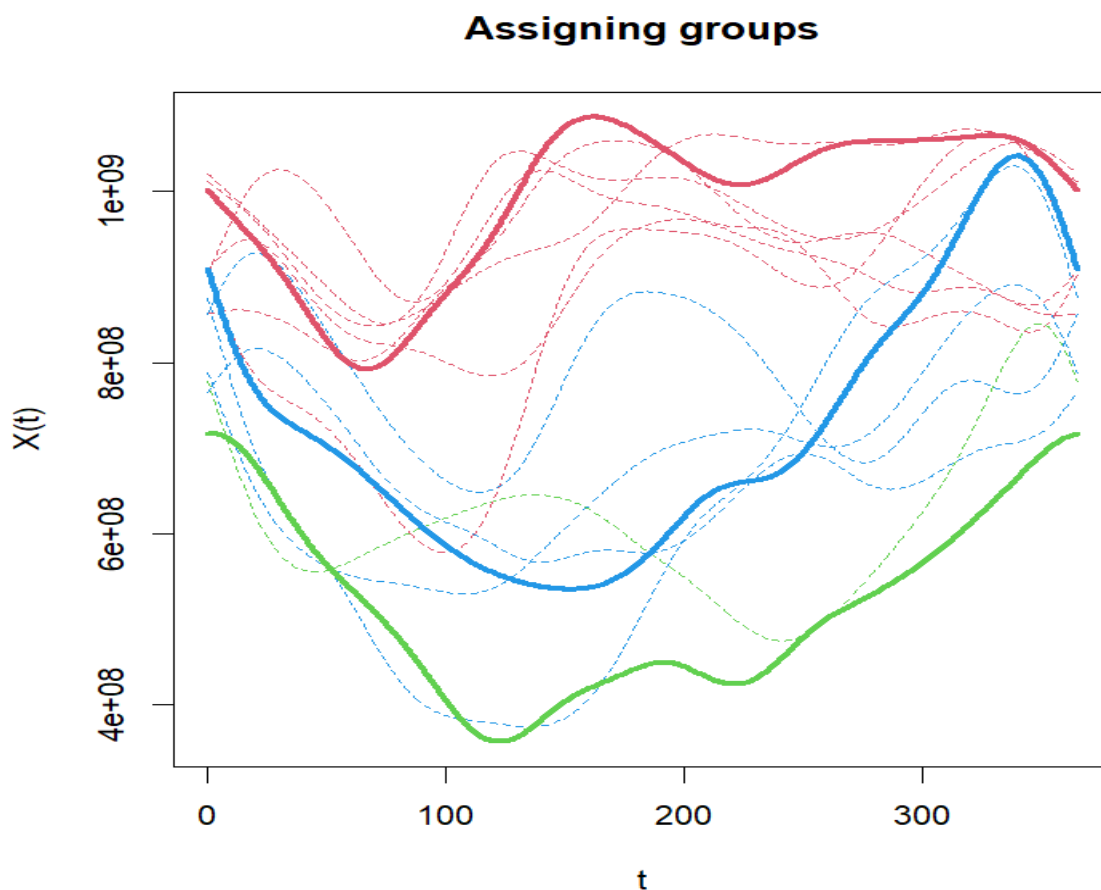


Figura 10. Clustering funcional del volumen útil diario del embalse El Peñol (2010-2023).

En la Figura 10 se observan las trayectorias de los grupos asignados, las cuales se dividen en tres patrones principales: años con volúmenes consistentemente altos (rojo), años con volúmenes moderados y recuperación hacia finales del año (azul), y años con volúmenes persistentemente bajos (verde).

El grupo rojo refleja años con condiciones climáticas favorables y alta disponibilidad hídrica, ideales para maximizar la generación y reservar agua como medida de contingencia. El grupo azul representa periodos de balance hídrico moderado, en los que es necesario optimizar la operación para garantizar la estabilidad, especialmente durante la segunda mitad del año. Finalmente, el grupo verde identifica años críticos de estrés hídrico, asociados principalmente con sequías o alta demanda, donde la prioridad es conservar el recurso y reducir el volumen turbinado.

Es importante resaltar que, debido a la naturaleza estacional y cíclica de los embalses, podría existir una fuerte correlación entre los niveles de diciembre y enero de años consecutivos. Para este estudio, se tomó cada año calendario como unidad de análisis, reconociendo que los eventos de final de año pueden estar fuertemente ligados a la disponibilidad de agua a inicios del siguiente. En consecuencia, para el análisis de agrupamiento (k-means funcional), se asume cada año de forma independiente, pero se reconoce que futuras investigaciones podrían incorporar un modelo cíclico que capture la continuidad temporal entre años.

4 Conclusiones

Los resultados destacaron que los años 2012 y 2020 fueron atípicos, mostrando como las condiciones climáticas extremas afectaron la operación del embalse, ocasionando variaciones considerables en los precios de la energía. Adicionalmente, el clustering funcional permitió identificar tres patrones operativos distintivos, cumpliendo con el objetivo de caracterizar dinámicas recurrentes y diferenciadas, proporcionando una base sólida para optimizar la planificación y el manejo eficiente de los recursos hídricos.

Desde el punto de vista estadístico, las herramientas funcionales empleadas demostraron su efectividad para reducir la dimensionalidad de los datos y destacar las principales fuentes de variabilidad en el sistema hídrico. Los patrones identificados en el volumen útil podrían ser usados para diseñar estrategias de conservación hídrica y planificar respuestas operativas ante eventos extremos, también se estable un marco metodológico que puede replicarse en otros embalses, tanto en Colombia como a nivel internacional, particularmente en contextos con retos similares relacionados con el cambio climático y las demandas energéticas.

En cuanto a aplicaciones futuras, este trabajo abre múltiples líneas de investigación. La inclusión de variables adicionales, como índices de calidad del agua o tasas de sedimentación, permitiría evaluar los impactos a largo plazo de cambios climáticos en la sostenibilidad del sistema hidroeléctrico. Asimismo, se recomienda implementar modelos predictivos basados en redes neuronales podría mejorar la

capacidad de respuesta ante condiciones críticas, mejorando la toma de decisiones estratégicas. La metodología empleada puede extenderse a otros embalses en regiones vulnerables al cambio climático, promoviendo estrategias de gestión conjunta para mejorar la resiliencia hídrica. Por último, la integración de tecnologías emergentes, como sistemas inteligentes y plataformas IoT, podría automatizar la gestión de los recursos hídricos, optimizando su uso y asegurando un suministro energético sostenible.

5 Referencias bibliográficas

- Angarita, H., Wickel, A. J., Sieber, J., Chavarro, J., Maldonado-Ocampo, J. A., Herrera-R, G. A., Delgado, J., & Purkey, D. (2018). Basin-scale impacts of hydropower development on the Mompós Depression wetlands, Colombia. *Hydrology and Earth System Sciences*, 22(5), 2839-2865. <https://doi.org/10.5194/HESS-22-2839-2018>
- Asprilla Echeverría, J. M. (2021). Plan B water assessment: Efficiency and circularity for agricultural and municipal adaptation to water scarcity. *Groundwater for Sustainable Development*, 14, 100602. <https://doi.org/10.1016/J.GSD.2021.100602>
- Bluma, M. (2020). Struggling to Survive: Water Scarcity and Poor Sanitation in Colombia and Kenya. *Global Majority E-Journal*, 11(1), 20-33.
- Chapman, P. ; C. J. ; K. R. ; K. T. ; R. T. ; S. C. ; & W. R. (1999). *CRISP-DM 1.0 Step-by-step data mining guide*.
- Desbureaux, S., & Rodella, A. S. (2019). Drought in the city: The economic impact of water scarcity in Latin American metropolitan areas. *World Development*, 114, 13-27. <https://doi.org/10.1016/J.WORLDDEV.2018.09.026>
- Febrero Bande, M., & Oviedo de la Fuente, M. (2024). Functional Data Analysis and Utilities for Statistical Computing [R package fda.usc version 2.2.0]. *CRAN: Contributed Packages*. <https://doi.org/10.32614/CRAN.PACKAGE.FDA.USC>
- Fitzmaurice, G., Laird, N., & Ware, J. (2011). *Applied longitudinal Analysis* (2nd Edition). Wiley, <http://biosun1.harvard.edu/~fitzmaur/ala/>.
- Gleik, P. H. (2018). Water and Conflict Update - Events, Trends, and Analysis. *The World's Water Volume 9: The Report on Freshwater Resources*, 141-145.

https://www.researchgate.net/publication/336882048_The_World's_Water_Volume_9_The_Report_on_Freshwater_Resources

Goldsmith, J., Scheipl, F., Huang, L., Wrobel, J., Di, C., Gellar, J., Harezlak, J., McLean, M. W., Swihart, B., Xiao, L., Crainiceanu, C., Reiss, P. T., & Cui, E. (2024). Regression with Functional Data [R package refund version 0.1-37]. *CRAN: Contributed Packages*.

<https://doi.org/10.32614/CRAN.PACKAGE.REFUND>

Han, J., Kamber, M., & Pei, J. (2011). *Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)*.

Hincapie, J. A., Mejia, C. Z., & Prieto, L. E. (2019). Climate change and water resources in Colombia. *Revista U.D.C.A Actualidad and Divulgacion Cientifica*, 22(2).

<https://doi.org/10.31910/RUDCA.V22.N2.2019.1368>

Ieva, F., Paganoni, A. M., Romo, J., & Tarabelloni, N. (2019). Roahd package: Robust analysis of high dimensional data. *R Journal*, 11(2), 291-307. <https://doi.org/10.32614/RJ-2019-032>

Jaramillo Gómez, E. A. (2020). *Conflictos por usos no extractivos en embalses de almacenamiento energético en Colombia. Caso de estudio: Embalse Peñol-Guatapé*. 200.

Laghari, J. A., Mokhlis, H., Bakar, A. H. A., & Mohammad, H. (2013). A comprehensive overview of new designs in the hydraulic, electrical equipments and controllers of mini hydro power plants making it cost effective technology. *Renewable and Sustainable Energy Reviews*, 20, 279-293. <https://doi.org/10.1016/j.rser.2012.12.002>

Leroy, D. (2019). Farmers' Perceptions of and Adaptations to Water Scarcity in Colombian and Venezuelan Páramos in the Context of Climate Change. <https://doi.org/10.1659/MRD-JOURNAL-D-18-00062.1>, 39(2), R21-R34. <https://doi.org/10.1659/MRD-JOURNAL-D-18-00062.1>

- Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N., Ramirez-Quintana, M. J., & Flach, P. (2021). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 33(8), 3048-3061.
<https://doi.org/10.1109/TKDE.2019.2962680>
- Masselot, P., Dabo-Niang, S., Chebana, F., & Ouarda, T. B. M. J. (2016). Streamflow forecasting using functional regression. *Journal of Hydrology*, 538, 754-766.
<https://doi.org/10.1016/j.jhydrol.2016.04.048>
- Pérez, A. J., Hurtado-Patiño, J., Herrera, H. M., Carvajal, A. F., Pérez, M. L., Gonzalez-Rojas, E., & Pérez-García, J. (2019). Assessing sub-regional water scarcity using the groundwater footprint. *Ecological Indicators*, 96, 32-39.
<https://doi.org/10.1016/J.ECOLIND.2018.08.056>
- Quintela-del-Río, A., & Francisco-Fernández, M. (2024). River flow modelling using nonparametric functional data analysis. *Journal of Flood Risk Management*, 11, S902-S915.
<https://doi.org/10.1111/jfr3.12282>
- R: *The R Project for Statistical Computing*. (s. f.). Recuperado 4 de diciembre de 2024, de <https://www.r-project.org/>
- Ramsay, J., Hooker, G., & Graves, S. (2009). Functional Data Analysis with R and MATLAB. *Functional Data Analysis with R and MATLAB*. <https://doi.org/10.1007/978-0-387-98185-7>
- Ramsay, J. O., & Silverman, B. W. (2005). *Functional Data Analysis*.
<https://doi.org/10.1007/B98888>
- Rodriguez, E. A., Sanchez-Rodriguez, I., Duque-Gardeazabal, N., Felipe, P., & Obando, A. (2020). *Combined Use of Local and Global Hydro Meteorological Data with Hydrological*

Models for Water Resources Management in the Magdalena-Cauca Macro Basin-Colombia. <https://doi.org/10.1007/s11269-019>

Rodríguez Hernández, A., Herrera, B., Subdirectora De Hidrocarburos, J., García, C., Subdirector De Demanda, B., Cárdenas, S., Subdirectora De Minería, C., Corredor, R. B., Camilo, J., Bejarano, B., Fernando, J., Mahecha, A., Arellano, R., Omar, P., Daza, A. B., Bonilla, M., German, M., Camacho, L., Marco, A., ... Alfaro, V. (2015). *PLAN ENERGÉTICO NACIONAL COLOMBIA: IDEARIO ENERGÉTICO 2050 REPÚBLICA DE COLOMBIA* Ministerio de Minas y Energía Tomás González Estrada, Ministro UNIDAD DE PLANEACIÓN MINERO ENERGÉTICA Ángela Inés Cadena Monroy Directora General Secretario General COLABORADORES.

Rosegrant, M. W., Cai, X., & Cline, S. A. (2002). *GLOBAL WATER OUTLOOK TO 2025 Averting an Impending Crisis* . Intl Food Policy Res Inst.

Tzanakakis, V. A., Paranychianakis, N. V., & Angelakis, A. N. (2020). Water Supply and Water Scarcity. *Water 2020, Vol. 12, Page 2347, 12(9), 2347.* <https://doi.org/10.3390/W12092347>

Unfried, K., Kis-Katos, K., & Poser, T. (2022). Water scarcity and social conflict. *Journal of Environmental Economics and Management, 113.*
<https://doi.org/10.1016/J.JEEM.2022.102633>

Verbeke, G., & Molenberghs, G. (s. f.). *Linear Mixed Models for Longitudinal Data*. Recuperado 9 de diciembre de 2024, de https://books.google.com.co/books/about/Linear_Mixed_Models_for_Longitudinal_Dat.html?id=ha8_AAAAQBAJ&redir_esc=y

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2014). *dplyr: A Grammar of Data Manipulation*. *CRAN: Contributed Packages*.

<https://doi.org/10.32614/CRAN.PACKAGE.DPLYR>

Wickham, H., Hester, J., & Bryan, J. (2024). Read Rectangular Text Data [R package readr version 2.1.5]. *CRAN: Contributed Packages*.

<https://doi.org/10.32614/CRAN.PACKAGE.READR>

Quintela-del-Río, A., & Francisco-Fernández, M. (2018). *River flow modelling using nonparametric functional data analysis*. *Journal of Flood Risk Management*, 11(S2), S902–S915. <https://doi.org/10.1111/jfr3.12282>

Masselot, P., Dabo-Niang, S., Chebana, F., & Ouarda, T. B. M. J. (2016). *Streamflow forecasting using functional regression*. *Journal of Hydrology*, 538, 754–766.

<https://doi.org/10.1016/j.jhydrol.2016.04.048>