

COMITÉ NACIONAL ESPAÑOL DE GRANDES PRESAS

ANÁLISIS DE LOS DATOS DE AUSCULTACIÓN DE LA PRESA DE LA BAEELS CON TÉCNICAS DE APRENDIZAJE AUTOMÁTICO

Fernando Salazar¹
André Conde²
Carlos Barbero³
Nathalia Silva-Cancino⁴

RESUMEN: La presa de La Baells es de tipo bóveda de doble curvatura, con una altura sobre cimientos de 102 m. Entró en servicio en 1976 y desde entonces ha tenido un funcionamiento correcto, que ha sido controlado con un sistema de auscultación muy completo y técnicas de interpretación convencionales. En esta comunicación se presentan los resultados de un estudio piloto en el que se han aplicado técnicas de aprendizaje automático (machine learning) para el análisis de sus datos de auscultación. Estas técnicas ofrecen mayor flexibilidad que los procedimientos convencionales para la generación de modelos de predicción de la respuesta de la presa, así como para identificar relaciones entre aparatos de distinta naturaleza. Ello permite un mejor conocimiento del comportamiento de la estructura y un mayor control de la seguridad.

¹ International Center for Numerical Methods in Engineering (CIMNE), fsalazar@cimne.upc.edu.

² International Center for Numerical Methods in Engineering (CIMNE), aconde@cimne.upc.edu.

³ Agència Catalana de l'Aigua, cbarberol@gencat.cat

⁴ International Center for Numerical Methods in Engineering (CIMNE), nsilva@cimne.upc.edu.

1. INTRODUCCIÓN

La seguridad de las presas es un aspecto relevante, por los servicios esenciales que ofrecen estas estructuras, y por las graves consecuencias que pueden derivarse de su mal funcionamiento. El parque de presas está envejeciendo, lo que genera preocupación en la sociedad, más aún con la publicación de determinados estudios, como el reciente informe de Perera et al. [1], que ha tenido repercusión en los medios. Las cifras presentadas en dicho trabajo están claramente sesgadas por mostrar cifras de roturas que incluyen conjuntamente presas de embalse y de residuos mineros. Sin embargo, si analizamos únicamente las primeras, se observa también un incremento en el número de casos de avería grave en las últimas décadas. La figura 1 muestra el número de roturas de presas de embalse desde 1.950 (sin incluir las de residuos mineros), con indicación de la edad media de las mismas.

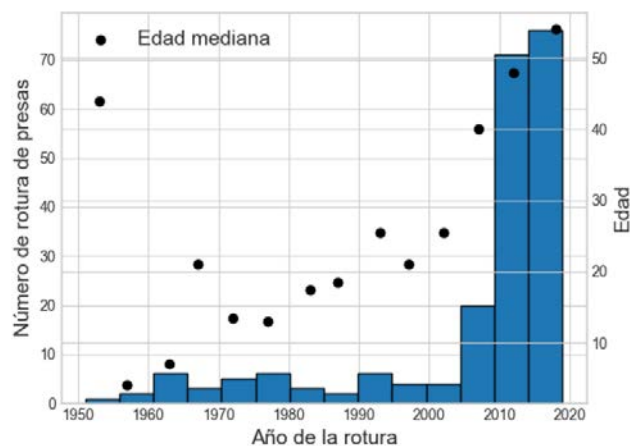


Figura 1: Roturas de presa y mediana de su edad desde 1.950. Fuente: elaboración propia a partir de la base de datos de ICOLD [2], ASDSO [3] y World Heritage Encyclopedia [4]

Los sistemas de auscultación modernos suelen incluir una gran cantidad de aparatos, con intervalos de lectura pequeños, lo que genera una gran cantidad de información. En este contexto de expansión de dispositivos y sistemas de tratamiento de datos, está creciendo el uso de tecnologías de aprendizaje automático para el tratamiento de datos de auscultación y la generación de modelos de predicción. Estos modelos pueden emplearse, como ocurre con los modelos numéricos, para comparar las predicciones con los datos registrados y tomar decisiones relacionadas con la seguridad.

Suele pensarse que estos modelos son difíciles de interpretar (se emplea el término “caja negra”), por la complejidad intrínseca de algunos de los algoritmos en que se basan. Aunque son indudablemente más complejos que la regresión lineal, también es evidente que existen herramientas para su análisis que permiten extraer conclusiones sobre el comportamiento de la presa. Algunos ejemplos se han publicado recientemente tanto en la literatura científica [5], como en la comunidad presística [6].

Una dificultad cierta para el uso de estos modelos, que no se pone tan frecuentemente de manifiesto, es la necesidad de disponer series de datos de auscultación largas y de calidad. En presas con varias décadas en funcionamiento, es frecuente que coexistan en las bases de datos registros antiguos, en muchos casos manuales, con otros obtenidos con medios modernos, mayor frecuencia de lectura y mejor control. En estos casos, para aprovechar toda la información disponible es necesari-

rio realizar una labor de tratamiento, corrección, integración de datos de distintas fuentes que requiere muchos recursos y que debe realizarse por personal experto.

Los problemas que pueden aparecer son diversos y frecuentes. Además de los posibles errores de lectura o registro, en presas antiguas es relativamente común que falten los datos en meses de verano, cuando la persona responsable de la toma manual de datos estaba ausente. Es también muy frecuente que diferentes conjuntos de aparatos fueran leídos en distintos días de la semana, lo que impide que puedan analizarse conjuntamente. La integración de bases de datos antiguas con las modernas, con distinta frecuencia de lectura, es también un reto que debe acometerse con sumo cuidado.

En este contexto, desde CIMNE se mantiene una línea de investigación con el objetivo de desarrollar soluciones útiles para resolver estos problemas de forma eficaz [7]. El objetivo final es poder elaborar modelos de predicción flexibles, precisos e interpretables, para tomar decisiones relacionadas con la seguridad de presas. Sin embargo, siendo conscientes de la problemática asociada a la preparación de los datos, y los efectos relevantes que pueden tener en los resultados obtenidos, se han desarrollado también herramientas para optimizar esa fase preliminar.

Este trabajo se ha materializado en dos aplicaciones informáticas con funcionalidades específicas. La primera se centra en el tratamiento e integración de datos, además de la generación de gráficos exploratorios (PREDATOR), y la segunda (SOL-DIER) permite elaborar e interpretar modelos de predicción para obtener información útil acerca del comportamiento de la presa como tendencias, efecto de las cargas y su interacción.

En esta comunicación se presenta el resultado de la aplicación de estas herramientas para el análisis de datos de auscultación de la presa de La Baells, que ha sido seleccionada como caso piloto.

2. DESCRIPCIÓN DE LA PRESA Y DEL SISTEMA DE AUSCULTACIÓN

La presa de La Baells es una bóveda de doble curvatura que entró en servicio en 1.976. Tiene 102 m de altura sobre cimientos y un aliviadero regulado por compuertas en la margen izquierda.

Su comportamiento ha sido satisfactorio durante toda su vida útil, según se refleja en los estudios de seguridad.

El sistema de auscultación incluye las siguientes familias de aparatos: péndulos directos e inversos, que miden desplazamientos radiales y tangenciales; piezómetros abiertos y de cuerda vibrante para control de la subpresión; aforadores de filtraciones; control de apertura de juntas internas y externas; extensómetros; termómetros en cuerpo de presa; variables exteriores: nivel, temperatura del aire, precipitación.

En el año 2.012 se rehabilitó la pantalla de drenaje y se completó el sistema de auscultación en relación a nuevos piezómetros y aforos.

Aunque podría plantearse la opción de analizar todos los aparatos disponibles, para esta prueba piloto se han considerado aquellos más relevantes, de acuerdo con los estudios de seguridad. Son los siguientes: péndulos radial y tangencial en las cotas superiores; piezómetros de medida de subpresión en la base de los bloques 2 y 4; aforos de filtraciones en ambas márgenes. A su vez, se ha seleccionado un aparato de cada tipo como ejemplo, cuyos resultados se muestran en la sección 5. Su localización se muestra en la Figura 2.

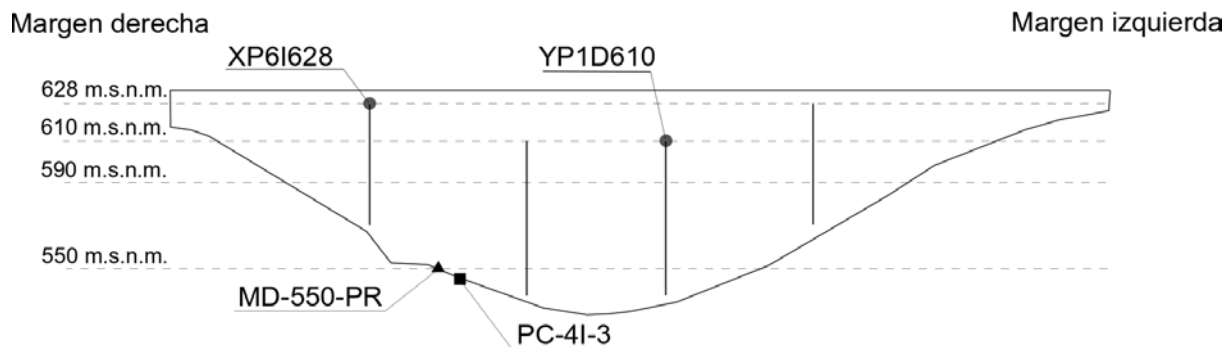


Figura 2: Situación de los aparatos analizados sobre el alzado de la presa de La Baells. Vista desde aguas abajo.

3. PREPARACIÓN DE DATOS

Para este trabajo, los datos tomados como base de partida se encontraban almacenados en hojas de cálculo separadas por familias de aparatos, si bien en todas ellas aparecen los registros de las variables exteriores: precipitación, nivel de embalse y temperatura ambiente. El primer objetivo es por tanto unificarlas en un único archivo, de modo que pueda trabajarse con ellas conjuntamente.

Esta unificación de datos, que en principio es una operación simple, implica sin embargo el problema de la no coincidencia de las fechas de medida entre familias de aparatos. Aunque la frecuencia de medida es aproximadamente semanal en todos los casos, las medidas no coinciden los mismos días. Así, por ejemplo, durante 2.002 se registraron los péndulos los jueves, los aforadores los viernes y los piezómetros los lunes.

Se ha implementado una funcionalidad que permite rellenar los datos entre medidas por interpolación lineal. Con ello se obtiene un archivo único, con una línea por cada día, y todas las series completas, combinando datos reales e interpolados. Se dejan sin interpolar los periodos de más de una semana sin datos. Sobre este archivo unificado se pueden identificar y corregir los datos erróneos, como se muestra en la figura 3, donde se muestra un ejemplo de corrección de la serie de nivel de embalse. La herramienta desarrollada permite corregir las series de forma interactiva, seleccionando los puntos y un método de modificación.

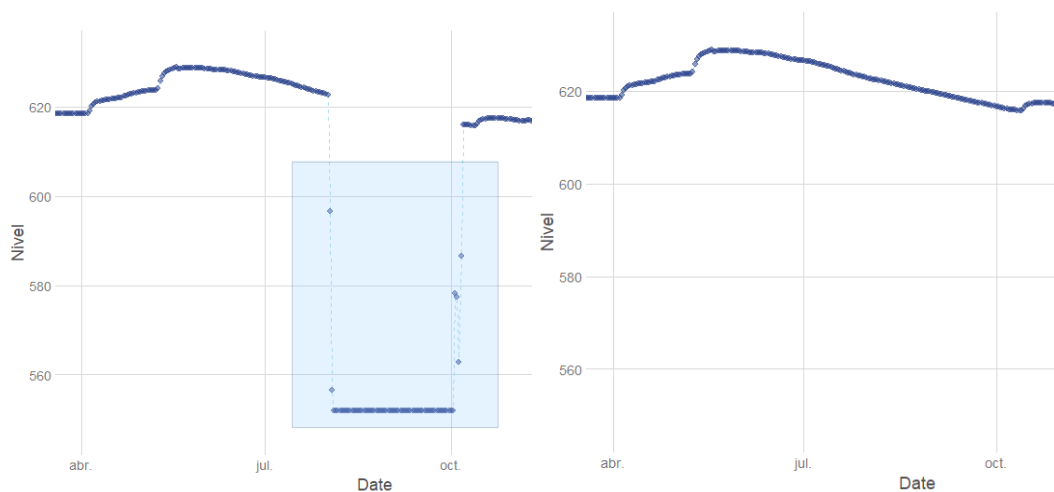


Figura 3: Corrección interactiva de datos erróneos por interpolación lineal. A la izquierda, la serie original, con la selección de los datos a corregir. A la derecha, resultado de la corrección.

En ocasiones, la interpolación lineal no es el mejor procedimiento para completar o corregir datos. Es el caso de la variable temperatura ambiente, que también tuvo que ser corregida por presentar periodos sin datos y otros erróneos. Dado que algunos de esos periodos tenían duración superior a la semana, y que es importante disponer de la serie completa de temperatura, en este caso se optó por utilizar el valor medio de la temperatura de ese mismo día del año, a partir de los datos registrados en otros años del periodo. Con ello se obtuvieron valores razonables.

Una vez revisadas, corregidas y completadas las series, se generaron las variables derivadas. En este caso, se crearon las medias móviles de distintos periodos del nivel de embalse y de la temperatura ambiente, que permiten posteriormente identificar la inercia térmica de la presa. Del mismo modo, se calculó la precipitación acumulada en diversos periodos de tiempo, para investigar su posible efecto en la respuesta de la presa. Finalmente, se tomó un dato por semana, de modo que la base de datos final tiene la misma frecuencia de lectura que la original. Este resultado puede exportarse en forma de hoja de cálculo para su uso posterior con otras herramientas, o bien, como en este caso, usarse para elaborar modelos de predicción basados en machine learning.

4. MODELO DE PREDICCIÓN

La aplicación se basa en el algoritmo denominado “Boosted Regression Trees” (BRT), que fue seleccionado a partir de un estudio comparativo con otros métodos de aprendizaje automático [8]. Para la selección del algoritmo, en dicho estudio se tuvo en cuenta tanto la precisión como la facilidad de implementación y la posibilidad de interpretación del modelo. El algoritmo es bien conocido en la comunidad de aprendizaje automático y su descripción puede encontrarse en [8]. Se basa en combinar el resultado de un gran número de modelos sencillos.

Una de las mayores ventajas de este método es su flexibilidad: puede considerarse un conjunto amplio de variables potencialmente influyentes en cada una de las respuestas de la presa sin que el resultado varía sustancialmente: el modelo automáticamente selecciona aquellas que tienen influencia real en la respuesta, y descarta el resto. Ello permite generar modelos de predicción para variables de distinta naturaleza (p. e. piezómetros, desplazamientos, aforos), de una forma ágil, simplemente cambiando la variable objetivo.

Igualmente, su flexibilidad permite analizar variables como los piezómetros, cuyo comportamiento es difícil de predecir con métodos convencionales como el Hydrostatic Season Time (HST) [10].

Así, se elaboraron modelos de predicción para cada una de las variables de respuesta consideradas, que fueron analizados posteriormente para interpretar el comportamiento de la presa.

En anteriores trabajos [8] se comprobó que estos modelos ofrecen mayor precisión en la predicción del comportamiento de la presa, lo cual permite su empleo para la definición de umbrales de aviso más ajustados, dependientes del estado de cargas en cada situación. Sin embargo, el objetivo principal de este trabajo es el análisis del comportamiento de la presa, identificando el efecto de cada una de las cargas sobre las variables de respuesta consideradas. Por ello, el análisis de resultados se centra en la interpretación de los modelos.

5. RESULTADOS

El análisis de los modelos permite identificar el efecto de cada una de las variables exteriores sobre cada aparato considerado, tanto en magnitud relativa como en la forma de ese efecto. Se resumen a continuación algunos de los resultados más relevantes para las variables seleccionadas.

5.1. DESPLAZAMIENTOS RADIALES

Los desplazamientos radiales responden a las variaciones de la carga hidrostática (considerada mediante la variable nivel de embalse), así como a los cambios de temperatura. Es bien conocido que las presas bóveda tienen una inercia térmica, según la cual los cambios de temperatura ambiente no se reflejan en los desplazamientos hasta pasado un determinado periodo. Este efecto se captura directamente con los modelos BRT cuando se incluyen diferentes medias móviles de la temperatura ambiente como variables predictoras. La ventaja en este caso es que puede capturarse igualmente la variación temporal de los desplazamientos, con independencia de que sea o no lineal. Como ejemplo, la figura 4 muestra el resultado de la interpretación del modelo para la lectura a la cota 610 del péndulo del bloque 1.

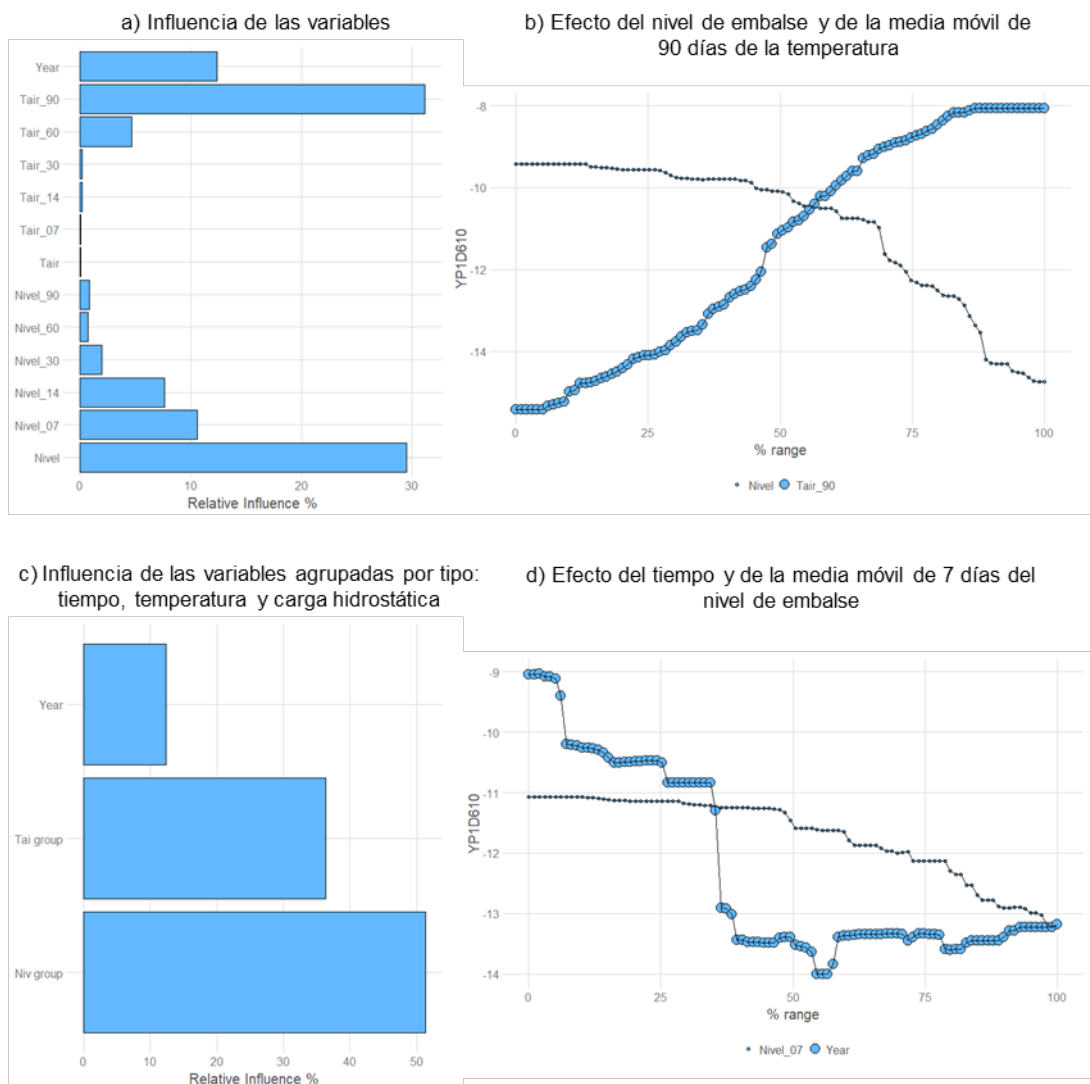


Figura 4: Análisis del modelo de predicción del desplazamiento radial en el péndulo YP1D610.

A la vista de los resultados de la figura 4, pueden hacerse algunas observaciones:

- a) La variable más influyente es la media móvil de 90 días de la temperatura ambiente, lo cual refleja la inercia térmica mencionada (Figura 4a).
- b) Si se agrupan las variables predictoras según su naturaleza (Figura 4c), es decir, el tiempo por una parte y por otra los grupos de medias móviles de temperatura y nivel de embalse, puede concluirse que, en el periodo considerado, la contribución de la carga hidrostática al desplazamiento radial es de algo más del 50%, la de la temperatura de alrededor del 35%, y la debida a cambios de comportamiento en el tiempo, de algo más del 10%.
- c) El efecto del tiempo se corresponde con un cambio no lineal hacia 1.991. Desde esa fecha, la respuesta ha sido estable y por tanto no se identifica ninguna tendencia (Figura 4d).
- d) El modelo identifica correctamente los efectos de la temperatura y de la carga hidrostática (Figura 4b): mayor nivel y menor temperatura se asocian a desplazamientos hacia aguas abajo y viceversa.
- e) Aunque se detecta cierta inercia frente a la carga hidrostática (influencia de medias móviles de 7 días del nivel), su efecto es mucho menor que el de la carga instantánea (Figura 4d).

5.2. AFORADORES

Se incluye como ejemplo la interpretación para el aforador MD-550-PR, que registra los caudales circulantes por la galería perimetral en el bloque 4 (margen derecha) a la cota 550 (Figura 5). De su observación pueden hacerse los siguientes comentarios:

- a) Aunque se han incluido como variables predictoras las que dependen de la temperatura y de la precipitación, el modelo las descarta automáticamente, como queda reflejado en que la influencia relativa es despreciable (Figura 5a).
- b) El nivel de embalse es la carga que más afecta al caudal aforado, si bien su efecto es claramente no lineal: con niveles inferiores al 5% de la carrera de embalse no se observan filtraciones; hasta el 75% de carga se observa un incremento moderado, y es con cargas superiores cuando se incrementa de forma significativa la filtración (Figura 5b). En los otros aforadores analizados se ha encontrado un efecto similar.
- c) El análisis de la evolución temporal muestra un comportamiento estable hasta el año 2.012, cuando se produce un incremento significativo, que muestra la efectividad de los trabajos de rehabilitación de la red de drenaje. Desde esa fecha, se observa un descenso aproximadamente lineal (Figura 5b). El modelo es capaz de capturar estos cambios de comportamiento de forma automática, sin necesidad de un análisis preliminar.

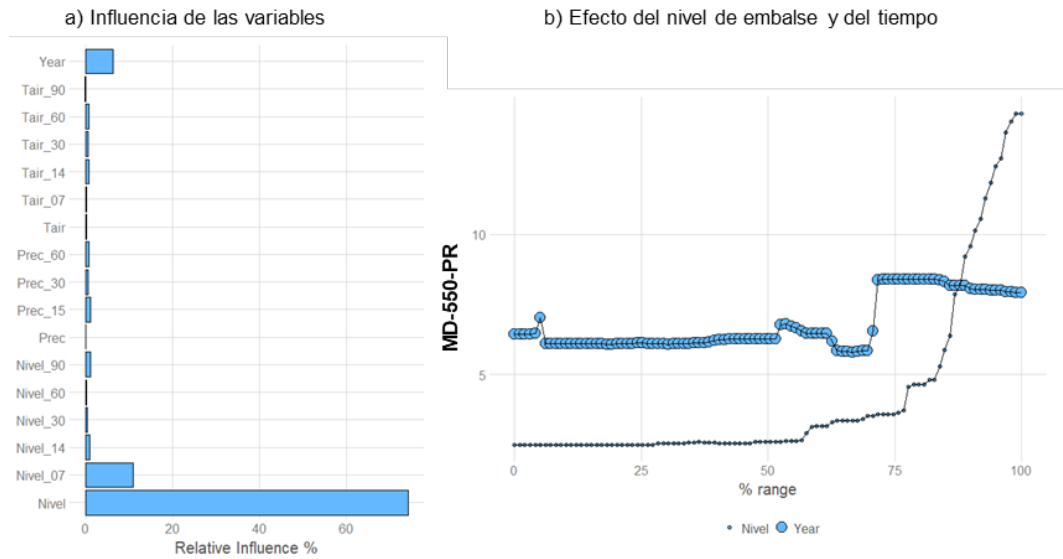


Figura 5: Análisis del modelo de predicción del aforador MD-550-PR.

5.3. DESPLAZAMIENTOS TANGENCIALES

Como es conocido, la magnitud de los desplazamientos tangenciales es en general inferior a la de los radiales. Esto ocurre también en la presa de La Baells, y hace que la interpretación del modelo sea más delicada. Sin embargo, como en el caso de los desplazamientos radiales, es posible separar la influencia de la temperatura y del nivel de embalse, como se observa en la Figura 6. Se comprueba que los desplazamientos hacia margen izquierda se producen con niveles bajos y temperaturas altas, y viceversa.

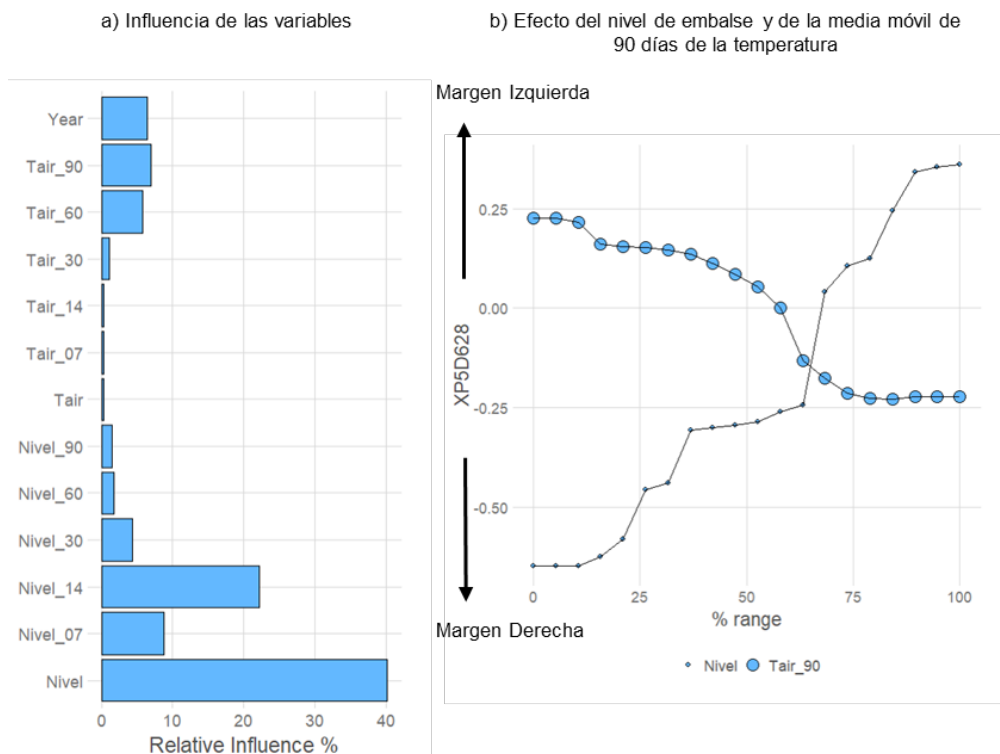


Figura 6: Análisis del modelo de predicción del desplazamiento tangencial en el péndulo XP5D628.

5.4. SUBPRESIONES

Las series de medida de subpresiones analizadas presentan una singularidad en el año 2.012, cuando se llevaron a cabo unas obras de rehabilitación de la red de drenaje que dieron como resultado una reducción apreciable de la subpresión en la margen derecha, donde previamente se habían registrado valores relativamente altos.

Como ejemplo se muestran los gráficos correspondientes al piezómetro PC-4I-3. En este caso, además del gráfico de influencia global del nivel de embalse y del tiempo (Figura 7b), se incluye el correspondiente gráfico tridimensional (Figura 7c), que ofrece información adicional: aunque en promedio hay una relación aproximadamente lineal entre el nivel de embalse y la subpresión, en el gráfico 3D se observa que éste solo se produce en el periodo anterior a 2.012. El efecto posterior a esa fecha es despreciable, como se puede observar en la serie de registros, que se mantiene prácticamente constante en unos 5 m.c.a.

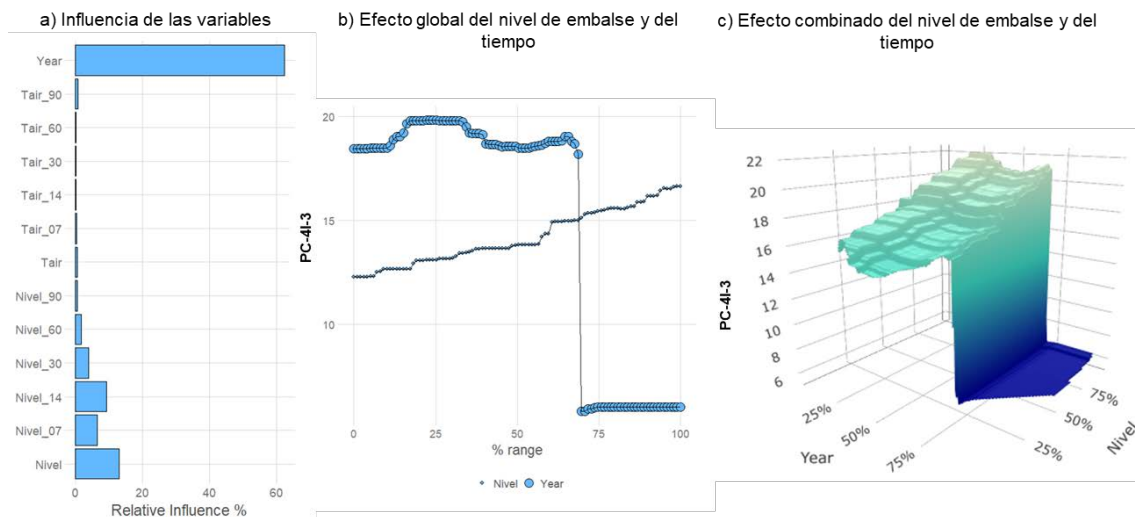


Figura 7: Análisis del modelo de predicción del piezómetro PC-4I-3.

De nuevo, el modelo se construye con las mismas premisas que los correspondientes a las otras variables, cambiando simplemente la variable objetivo. El propio algoritmo selecciona las variables con efecto relevante en cada caso, y descarta el resto. Además, captura automáticamente efectos no lineales como es la evolución temporal de los piezómetros.

6. CONCLUSIONES

Aunque los modelos basados en aprendizaje automático (BRT en este caso) son más complejos que la regresión lineal convencional, la prueba piloto realizada demuestra que su interpretación es factible, y que pueden extraerse conclusiones útiles acerca del comportamiento de la presa. Además, ofrecen algunas ventajas sobre los modelos simples, de las que pueden destacarse las siguientes:

- Permiten considerar distintos tipos de respuesta de la presa con un mismo modelo, simplemente seleccionando la variable objetivo de interés. En este caso, se han analizado desplazamientos horizontales, caudales de filtración y subpresiones.

- Son capaces de seleccionar de forma automática las cargas que afectan a cada repuesta de la presa, sin necesidad de hacer una selección previa de variables.
- Son flexibles para capturar cualquier efecto no lineal de las cargas actuantes. Como ejemplo, en el medidor de filtraciones analizado, se ha identificado la evolución en el tiempo, que incluye cambios bruscos y periodos de variación lineal. No es necesario ninguna adaptación del modelo para estos casos.
- Además, los modelos basados en BRTs dependen muy poco de los parámetros de ajuste. Aunque cierto trabajo de adaptación podría dar como resultado un leve incremento de la precisión, los valores por defecto funcionan bien en la mayoría de los casos. Todos los resultados presentados se han obtenido con los valores por defecto.

El empleo de este tipo de modelos no evita la necesidad de que el análisis se realice por usuarios expertos en ingeniería de presas, y en particular familiarizados con el comportamiento de la estructura. Sin embargo, las herramientas desarrolladas pueden incrementar la eficacia del ciclo de trabajo completo, desde la preparación y unificación de los datos, hasta la construcción del modelo y su análisis.

7. REFERENCIAS

1. Perera, D., Smakhtin, V., Williams, S., North, T., Curry, A., 2021. Ageing water storage infrastructure: an emerging global risk. United Nations University. http://inweh.unu.edu/wp-content/uploads/2021/01/Ageing-Water-Storage-Infrastructure-An-Emerging-Global-Risk_web-version.pdf
2. ICOLD. Dam failure database.
3. ASDSO, 2020. Dam Incident Database. <https://damsafety.org/Incidents>
4. World Heritage Encyclopedia, 2017. List of dam failures. http://self.gutenberg.org/articles/list_of_dam_failures
5. Salazar, Fernando, et al. "Data-based models for the prediction of dam behaviour: a review and some methodological considerations." *Archives of computational methods in engineering* 24.1 (2017): 1-21.
6. Landstorfer, F., Wagner, E. (2021). Durlassboden – how old measurement data and new data processing methods improve the understanding of an 50 year old embankment dam with underseepage. 27th ICOLD Congress, Marsella, 2021.
7. Salazar, F. A machine learning based methodology for anomaly detection in dam behavior. Tesis Doctoral, UPC, 2017
8. Salazar, Fernando, et al. "An empirical comparison of machine learning techniques for dam behaviour modelling." *Structural Safety* 56 (2015): 9-17.
9. Elith, Jane, John R. Leathwick, and Trevor Hastie. "A working guide to boosted regression trees." *Journal of Animal Ecology* 77.4 (2008): 802-813.
10. Willm G, Beaujoint N (1967). Les méthodes de surveillance des barrages au service de la production hydraulique d'Electricité de France-Problèmes anciens et solutions nouvelles. In: 9th ICOLD Congress, pp 529{550, q34-R30. [inFrench]