

COMITÉ NACIONAL ESPAÑOL DE GRANDES PRESAS

AVANCES EN EL TRATAMIENTO Y ANÁLISIS DE DATOS DE AUSCULTACIÓN DE PRESAS

Fernando Salazar¹

León Morera²

Miguel Ángel Toledo³

Rafael Morán⁴

Eugenio Oñate.⁵

RESUMEN: Los avances en los instrumentos de medida y en las técnicas de transmisión y almacenamiento de información han permitido aumentar el control de la seguridad de las presas, con medidas más fiables, precisas y frecuentes. Sin embargo, los métodos de tratamiento y análisis de los datos de auscultación no han evolucionado tanto, y con frecuencia se limitan a gráficos y modelos estadísticos sencillos.

Existen multitud de herramientas desarrolladas en diversos campos, generalmente alejados de la ingeniería civil, que facilitan el análisis y modelación de sistemas complejos: redes neuronales, redes complejas o bosques aleatorios son algunos ejemplos.

Estas técnicas han sido exploradas, implementadas y aplicadas a casos de ejemplo reales. Las estimaciones obtenidas son en general más precisas que las que resultan de aplicar los métodos convencionales, con lo que se puede definir mejor el rango de comportamiento normal de la presa. Además, permiten desvelar interacciones complejas entre variables de distinto tipo, más allá de la relación con el tiempo y el nivel de embalse. Por tanto, estos métodos, manejados e interpretados por expertos en ingeniería de presas, pueden ser de gran ayuda para conocer mejor el comportamiento de la presa y aumentar su seguridad.

¹ Centro Internacional de Métodos Numéricos en Ingeniería (CIMNE)

² Universidad Politécnica de Madrid (UPM)

³ Universidad Politécnica de Madrid (UPM)

⁴ Universidad Politécnica de Madrid (UPM)

⁵ Centro Internacional de Métodos Numéricos en Ingeniería (CIMNE)

1. INTRODUCCIÓN

El objetivo principal de los sistemas de auscultación de presas es la detección temprana de anomalías, de modo que sea posible tomar las medidas adecuadas para su corrección, y en última instancia, para evitar averías graves y la rotura.

La normativa requiere definir unos valores máximos (también mínimos en algunos casos) admisibles para determinados aparatos de auscultación, de modo que al verse superados deben activarse unos determinados protocolos de actuación.

Este rango de funcionamiento “normal” se basa en la estimación del comportamiento de la presa en unas determinadas condiciones de operación (principalmente nivel de embalse y temperatura). Para el cálculo de la respuesta en situación normal y la determinación del rango admisible se utilizan modelos de predicción de diversa naturaleza.

Los modelos determinísticos se basan en las leyes de la física, y generalmente consisten en un modelo estructural de elementos finitos, que considera con cierto grado de detalle las acciones sobre la presa. Son fundamentales en la fase de proyecto y el inicio de la explotación de la presa, si bien requieren adoptar simplificaciones importantes con respecto a las acciones [1] y a la respuesta estructural de la presa y el cimiento [2], que en la inmensa mayoría de los casos se consideran medios continuos elásticos lineales. Esto lleva a discrepancias entre sus resultados y la respuesta real de la presa, medida con los aparatos de auscultación.

Los modelos estadísticos utilizan los datos realmente medidos por el sistema de auscultación de la presa durante un periodo determinado para predecir su respuesta en un periodo posterior. Ello implica que no pueden aplicarse durante la fase inicial de la vida de la presa, hasta que no se han recopilado datos suficientes para ajustar los parámetros del modelo (el periodo necesario depende de cada caso, habiendo estudios que lo cifran en 5 [3], 10 [4] y 12 años [5]). Los métodos estadísticos convencionales presentan limitaciones importantes, como se ha puesto de manifiesto en trabajos recientes [9].

En los últimos años, la UPM y el CIMNE han puesto en marcha una investigación conjunta que pretende mejorar la seguridad de presas extrayendo la máxima información de los datos de auscultación. Se basa en dos ideas fundamentales:

- El conjunto presa-cimiento es un sistema complejo cuyos elementos son las series temporales de registro (de variables externas e internas). A partir del análisis de dichas series de registros pueden definirse conexiones o relaciones entre los elementos del sistema.
- Existen herramientas de inteligencia artificial, redes complejas y sistemas expertos que pueden ser útiles para analizar el sistema en conjunto, estudiar las relaciones entre sus elementos, y generar modelos de predicción de las variables de respuesta.

En la comunicación se incluyen algunos de los resultados obtenidos hasta el momento.

2. EXPLORACIÓN DE DATOS

La primera operación a realizar sobre los datos de auscultación, una vez recibidos y almacenados, es el análisis gráfico. Es muy frecuente dibujar la evolución de los registros a lo largo del tiempo, así como en relación con las variables externas más influyentes, que generalmente son el nivel de embalse y la temperatura.

A primera vista, estos gráficos permiten tener una idea de los rangos de variación de las variables, el volumen de datos disponible, y las lagunas en las series de datos. Si además se lleva a cabo por un técnico experto y cualificado, pueden detectarse cambios claros de tendencia en las medidas, y tener una primera idea de si el comportamiento responde a lo esperado.

Las herramientas utilizadas para la generación de estos gráficos suelen ser las ofimáticas convencionales. En ocasiones se han desarrollado también herramientas específicas que incorporan ciertas funcionalidades, como la indicación de determinadas incidencias, ya sean del sensor o generales de la presa [6].

Los avances informáticos permiten de una manera fácil generar gráficos muy flexibles e interactivos. En el curso de la investigación, se han integrado diversas tecnologías para generar una herramienta de exploración de datos interactiva y adaptable, que permite:

1. Seleccionar las variables que se quieren visualizar en cada eje del gráfico.
2. Seleccionar qué variables utilizar para determinar el color y el tamaño de la visualización.
3. Navegar de forma dinámica sobre el gráfico, ampliando las regiones de interés en cada uno de los ejes.
4. Acceder al gráfico e interactuar con él desde cualquier dispositivo con acceso a internet, en cualquier momento.

La Figura 1 muestra una imagen de la aplicación. En el menú de la izquierda se pueden seleccionar las variables que se quieren representar en los ejes, así como las utilizadas para definir el tamaño y el color de los puntos.

En la Figura 2 se incluye un ejemplo de visualización dinámica de series temporales, con dos niveles de aumento diferentes.

Estas herramientas gráficas permiten además tomar decisiones para la generación de modelos de predicción. Como ejemplo, en la Figura 1 se observa la disminución del caudal de filtración (eje de ordenadas) con el tiempo (mapa de color) y su variación cuasi lineal con el nivel de embalse (eje de abscisas) para los registros más recientes. A la vista de este gráfico parece razonable restringir la selección a los datos más recientes para después ajustar una regresión lineal dependiente del nivel de embalse.

Data Exploration

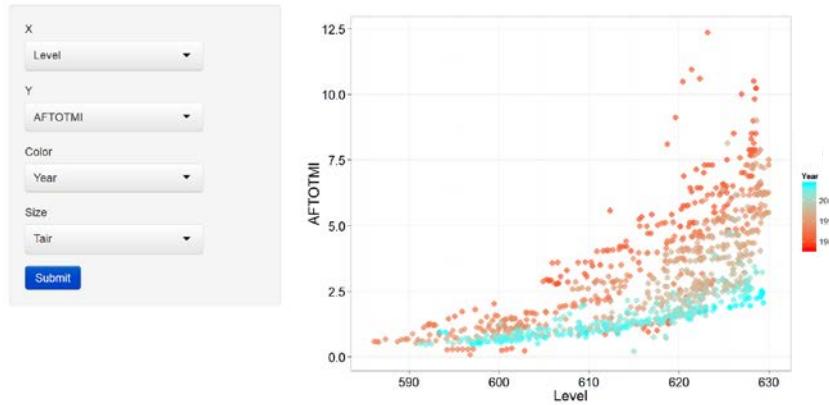


Figura 1. Aforo de filtración en función del nivel de embalse. Los colores se corresponden con el año de registro. Se observa que el caudal ha disminuido con el tiempo, y que su relación con el nivel de embalse es sensiblemente lineal en el periodo más reciente.

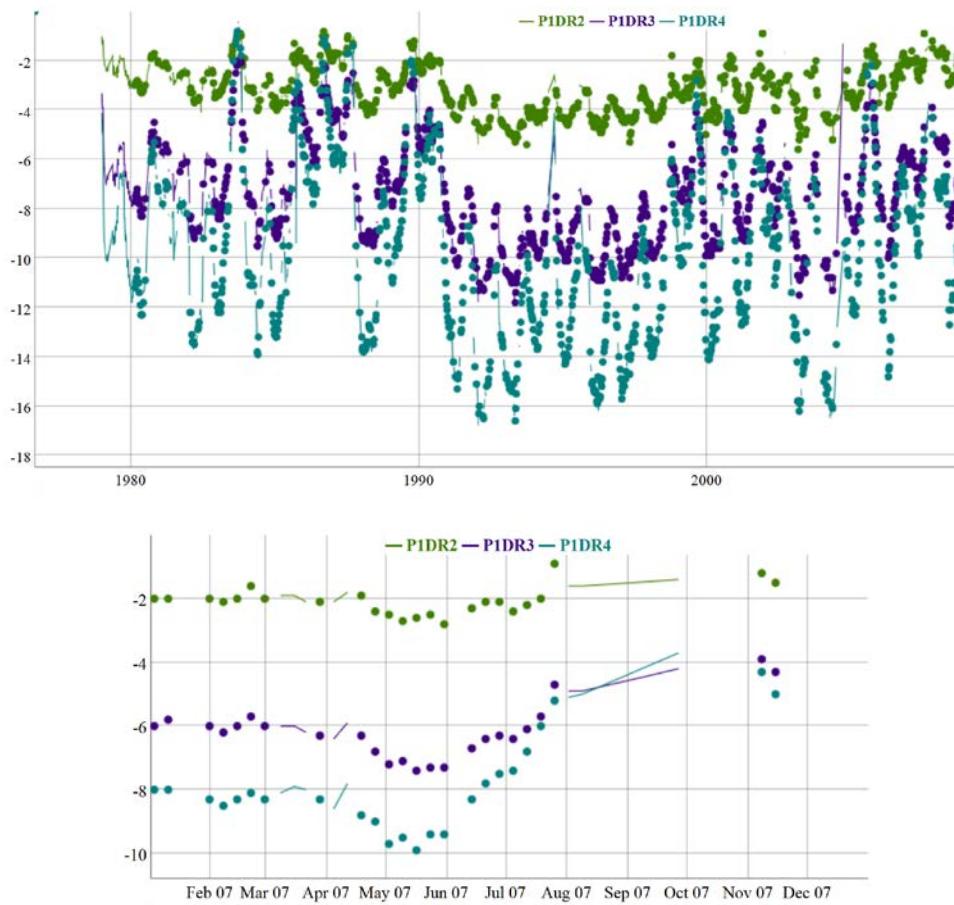


Figura 2. Interfaz de la aplicación para exploración dinámica de datos. Se muestra la evolución temporal de tres movimientos en péndulos. Arriba: series completas. Abajo: ampliación de los datos del año 2007.

3. INTERPRETACIÓN DEL COMPORTAMIENTO CONJUNTO. LAS REDES COMPLEJAS

La aplicación de la Teoría de Redes Complejas al análisis de los registros de auscultación se plantea como una posibilidad adicional dentro de la exploración de datos permitiendo el alcance de un punto de vista global o sistémico, a diferencia de otros procedimientos exploratorios más específicos (como los anteriormente descritos). Es posible definir la red compleja a partir de las variables o las series registradas en cada aparato (que serán los elementos o nudos de la red), mientras que las posibles conexiones entre las parejas de nudos se obtendrán a través de la comparación de los datos contenidos en cada una de las series, mediante la fijación de un determinado criterio. En las primeras aplicaciones [7], el criterio escogido para valorar la intensidad de la relación entre cada pareja de series fue el valor del mejor coeficiente de determinación resultante del ajuste de polinomios de segundo grado a los datos coincidentes en fecha entre ambos aparatos (se recuerda que el citado coeficiente establece una medida normalizada de la bondad del ajuste obtenido).

Una vez establecida la red compleja, la Teoría comprende diferentes técnicas que permiten efectuar una descriptiva del sistema:

- Medidas de centralidad para valorar la importancia de cada nudo o serie dentro del conjunto conforme a diferentes criterios.
- Algoritmos de partición o de clasificación para la detección de conjuntos homogéneos de aparatos.
- Caracterización de la red, de forma que puedan establecerse paralelismos entre fenómenos de diversos campos científico-técnicos.

Otro de los aspectos esenciales de las redes complejas es su representación en forma de grafos (conjuntos de nudos y lazos), donde la posición de los elementos puede definirse mediante diferentes algoritmos de distribución, y en los que es posible establecer mapas de colores y distribuciones de tamaños asociados a los nudos o a los lazos, aumentando considerablemente la información que se visualiza en una sola imagen. También es posible representar la red compleja tomando como base la ubicación real de los aparatos sobre la presa. Todas estas posibilidades de representación pueden combinarse con un enfoque dinámico o evolutivo de la red, para detectar así pautas de comportamiento cronológico o derivas temporales, así como anomalías.

De este modo, las redes complejas y sus representaciones en forma de grafo pueden servir como un soporte de visualización global del sistema presa de gran elocuencia, que además puede ser descompuesto o sometido a diferentes criterios de filtrado para la realización de análisis pormenorizados por parte del ingeniero experto. (En relación a la cantidad de información representada, se recuerda aquí que, si el número de nudos del sistema es de orden n , el número de lazos o conexiones representadas será de orden $n^2/2$).

Actualmente se está explorando la posibilidad de utilizar las redes complejas así obtenidas como soporte para la selección de variables de entrada para modelos de predicción, gracias a que los diferentes algoritmos de partición pueden ayudar a evitar la multi-colinealidad así como a enriquecer la información de entrada empleada para la predicción.

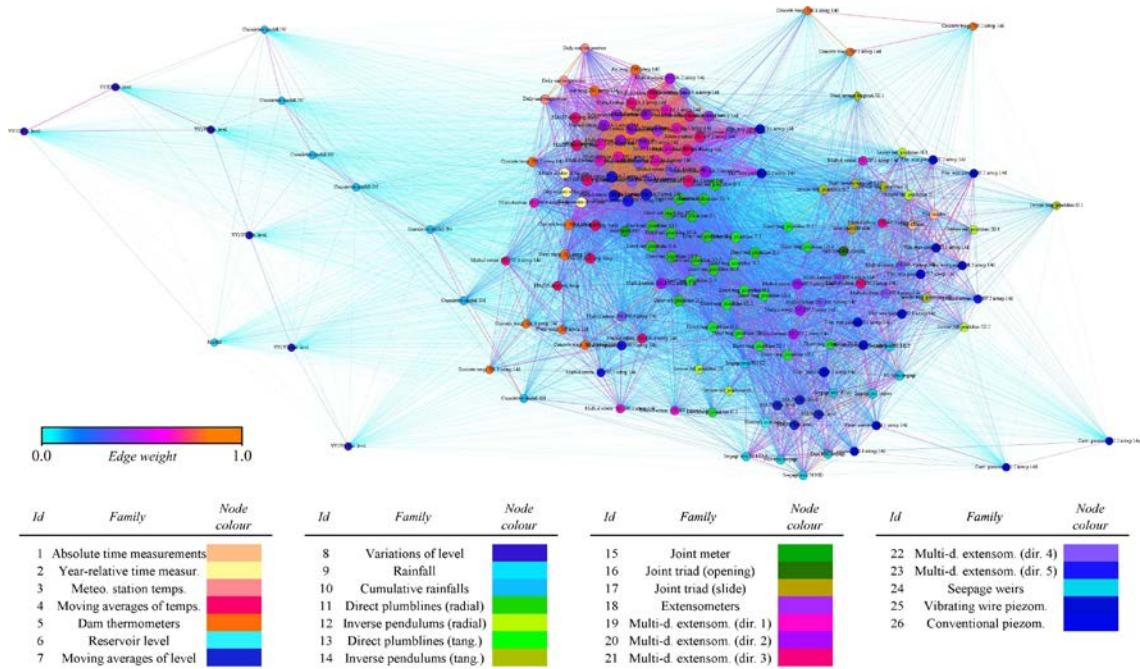


Figura 3. Ejemplo de representación de una red compleja obtenida a partir del análisis de los datos procedentes de una presa bóveda. Los colores de los nudos representan en este caso el tipo de aparato o variable. Los colores de las conexiones dan una idea de la intensidad de la relación entre los nudos que une. El algoritmo de distribución sobre el plano tiene en cuenta los grupos de aparatos densamente interconectados.

4. MODELOS DE PREDICCIÓN NO CONVENCIONALES

4.1. INTRODUCCIÓN

Los sistemas de aviso son claves en el sistema de auscultación. Con frecuencia, se implementan herramientas que lanzan un aviso cuando se registra un valor que queda fuera de un rango determinado previamente. Dicho rango de comportamiento “normal” se determina a partir de un modelo de predicción, que estima el valor más probable de la variable en cuestión en función de las variables exteriores de la presa: fundamentalmente, temperatura, nivel de embalse, y edad de la estructura.

Estos modelos suelen basarse en relaciones estadísticas sencillas de las variables mencionadas, siendo el método “HST” el más común. Se trata de una regresión lineal múltiple, de modo que la variable objetivo se calcula como una combinación lineal de:

- Varias potencias del nivel de embalse
- La temperatura ambiente, que suele considerarse como una función senoidal.
- Una función del tiempo, que considera los efectos no reversibles.

Este modelo se desarrolló originalmente para calcular el movimiento en presas bóveda, si bien se ha aplicado con algunas variantes para otro tipo de variables, como, por ejemplo, el aforo de filtraciones [8].

El método presenta limitaciones, según han puesto de manifiesto investigaciones recientes [5], [9]. Las principales son:

1. Asume que las variables son independientes, lo cual no siempre se cumple.
2. Asume que las relaciones entre las variables exteriores y la respuesta de la presa son lineales, lo cual tampoco es cierto con carácter general.

Para solventar estas limitaciones, se han comenzado a utilizar otras técnicas de generación de modelos, en general más flexibles, y que por tanto se permiten modelar relaciones no lineales y variables dependientes. Las más comunes son las redes neuronales, como muestra la cantidad de estudios presentados recientemente (por ejemplo, [10]).

Estos modelos no hacen suposiciones a priori sobre las relaciones entre variables. Por el contrario, pueden adaptarse y “descubrir” las interacciones existentes entre las series de datos. Como contrapartida, deben utilizarse con precaución, para evitar el efecto denominado “sobreajuste”, por el cual el modelo aproxima muy bien los datos de entrenamiento (los utilizados para ajustar los parámetros del modelo), pero no tienen buena capacidad de generalización. Es decir, el error de predicción aumenta considerablemente al aplicarlo a un caso diferente de los de entrenamiento. Un procedimiento aplicable siempre, de gran utilidad para evitar este problema, consiste en reservar una parte de los datos disponibles (datos de test), que no se utilizan para ajustar los parámetros del modelo. Al contrario, se compara la predicción del modelo en el periodo de test con los valores observados. Un aumento del error en este periodo con respecto al registrado en los datos de entrenamiento denota sobreajuste.

En el curso de la investigación, se ha realizado un estudio comparativo entre algunas de estas herramientas, enmarcadas en el campo de la inteligencia artificial [11]. Los resultados muestran que las redes neuronales ofrecen en general mayor precisión que el modelo HST, y también que otras técnicas permiten obtener una precisión incluso mayor en muchos casos.

La idea es que modelos más precisos permitirán definir con mayor precisión los umbrales de comportamiento normal, y por tanto unos niveles de aviso más realistas.

4.2. IMPLEMENTACIÓN

Se ha desarrollado una aplicación que permite generar modelos de predicción basados en distintas técnicas (incluso las convencionales y las redes neuronales), con diferentes valores de los parámetros que las definen.

Como ejemplo, para construir una red neuronal deben definirse el número de neuronas y los parámetros que controlan el proceso entrenamiento. La aplicación permite ajustar un modelo de este tipo, bien con unos parámetros por defecto (que suelen funcionar aceptablemente bien en el caso general), para que el usuario no iniciado pueda hacer pruebas, bien particularizándolos, para comprobar cómo afectan los cambios a la precisión del modelo.

Se incluye también documentación explicativa, en forma resumida, sobre a) las bases de funcionamiento del modelo, b) criterios para la definición de los

parámetros, y c) enlaces a textos explicativos de la base matemática de cada herramienta.

Build a Random Forest model

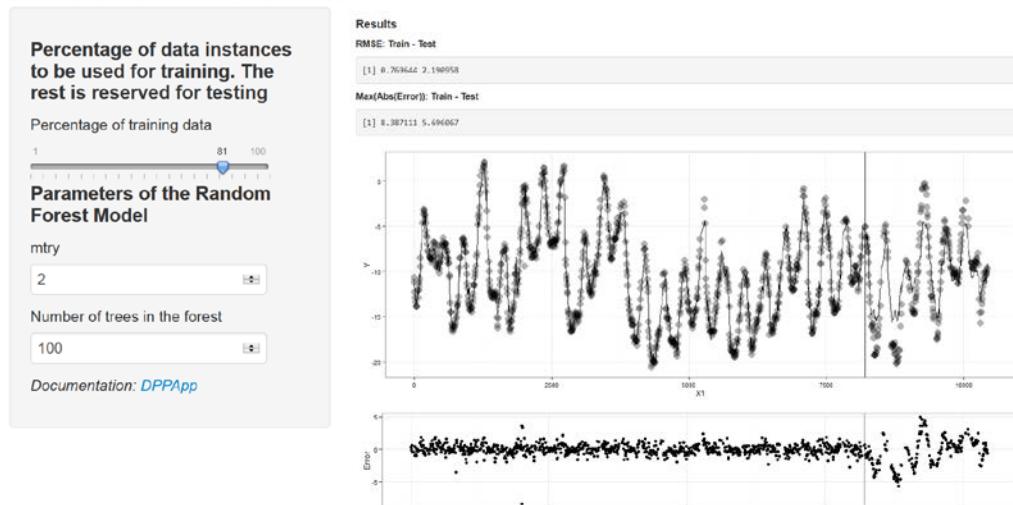


Figura 4. Aplicación para el cálculo de un modelo de predicción basado en bosques aleatorios. Los controles de la izquierda permiten particularizar la cantidad de datos utilizados para el entrenamiento, y los parámetros que definen el modelo (en este caso, el número de árboles y el parámetro “ntry”). Se incluye también un enlace a una documentación resumida, con criterios a seguir y enlaces a textos clave para entender la base matemática y la aplicación de cada modelo. En el ejemplo mostrado se observa sobreajuste del modelo, como indica el aumento claro del error entre los datos de entrenamiento (a la izquierda de la línea vertical) y los de test (no empleados para el ajuste del modelo).

Como en el ejemplo anterior, estas aplicaciones se han diseñado para ser accesibles a través de internet, desde cualquier dispositivo. La figura 4 muestra una imagen de la herramienta para la generación de modelos basados en bosques aleatorios [12].

5. CONCLUSIONES Y LÍNEAS DE INVESTIGACIÓN FUTURAS

Existen multitud de técnicas desarrolladas en diversos campos, generalmente alejados de la ingeniería civil, que han demostrado su utilidad para la interpretación y predicción del comportamiento de sistemas complejos. Su aplicación a sistemas de ingeniería civil ha sido relativamente escasa, muy escasa en el caso de las presas.

Es necesario considerar la complejidad del comportamiento del conjunto presa-terreno, y admitir las posibilidades que ofrecen los métodos no convencionales para comprender mejor la respuesta de la estructura. Se trata de combinar las técnicas habituales con otras nuevas en el campo de la ingeniería de presas, pero ya maduras gracias a su aplicación a diversos ámbitos del conocimiento.

Las posibilidades que ofrecen las técnicas de inteligencia artificial son enormes. Algunas de las que se han identificado, y que se prevé explorar en el futuro, son las siguientes:

- Selección de variables flexible: en algunos casos, se añaden al modelo variables derivadas de las exteriores, como por ejemplo la media móvil del nivel de embalse, o su velocidad de variación [13], en base a la intuición del modelador. Las herramientas de inteligencia artificial permiten seleccionar de forma automática las variables más útiles, y descartar las poco importantes [10].
- Generación de modelos no causales: pueden utilizarse algunas de las variables de respuesta de la presa para predecir otras (por ejemplo, calcular el movimiento en un péndulo a partir de otro). Ello puede reducir la utilidad del modelo para explicar el comportamiento, pero aumentar su aptitud para detectar determinadas anomalías.
- Modelos autorregresivos: son los que basan la predicción en el valor registrado en instantes de tiempo previos. También se llaman modelos de corto plazo, porque suelen utilizarse para predicción “paso a paso”. Su precisión disminuye en general para predecir a largo plazo, porque el error se propaga hacia adelante en el tiempo. Pueden ser útiles sin embargo en combinación con otro tipo de modelos.
- Interpretación del comportamiento de la presa. Muchos de estos modelos permiten calcular un índice de importancia de las variables, que normalmente se basa en criterios empíricos (miden cuánto aumenta el error al eliminar cada una de las variables por separado). La interpretación de estos resultados no es tan directa como la de un modelo lineal, donde los coeficientes correspondientes a cada variable pueden compararse directamente y representan la variación de la variable respuesta al modificarse cada una de las variables predictivas. Sin embargo, pueden ser de gran utilidad como apoyo para el encargado de seguridad de la presa, en combinación con su conocimiento previo de la estructura y su comportamiento.

Finalmente hay que destacar dos ideas fundamentales:

1. la inspección visual es irreemplazable y debe siempre formar parte de un sistema de auscultación
2. la aplicación de estas herramientas y la interpretación de sus resultados debe llevarse a cabo por profesionales con formación y experiencia suficientes.

4. REFERENCIAS

- [1] Santillán, D. (2014). Mejora de los modelos térmicos de las presas bóveda en explotación: aplicación al análisis del efecto del cambio climático. Tesis Doctoral, E.T.S.I. Caminos, Canales y Puertos (UPM).
- [2] International Commission on Large Dams (2012). Dam surveillance guide. Boletín 158, ICOLD.
- [3] Chouinard, L. y Roy, V. (2006). Performance of statistical models for dam monitoring data. In Joint International Conference on Computing and Decision Making in Civil and Building Engineering, Montreal. 2006 14–16.

- [4] De Sortis, A. y Paoliani, P. (2007). Statistical analysis and structural identification in concrete dam monitoring. *Engineering structures*. 29(1):110–120.
- [5] Lombardi, G. (2004). Advanced data interpretation for diagnosis of concrete dams. Technical report, CISM, 2004.
- [6] Mora, J., López, E. y de Cea, J.C. (2008). Estandarización en la gestión de datos de auscultación e informe anual en presas de titularidad estatal. VIII Jornadas Españolas de Presas
- [7] Morera, L. (2014). A systemic approach to dam data analysis using Complex Networks Theory. Tesina Fin de Máster Universitario en Sistemas de Ingeniería Civil. Universidad Politécnica de Madrid, España.
- [8] Simon, A., Royer, M., Mauris, F. y Fabre, J. (2013). Analysis and interpretation of dam measurements using artificial neural networks. 9th ICOLD European Club Symposium. Venecia, Italia.
- [9] Tatin, M., Briaut, M., Dufour, F., Simon, A. y Fabre, J. (2013). Thermal displacements of concrete dams: Finite element and statistical modelling, in: 9th ICOLD European Club Symposium, Venecia, Italia.
- [10] Santillán D., Fraile-Ardanuy J., Toledo M. Á. (2014). Predicción de lecturas de aforos de filtraciones de presas bóveda mediante redes neuronales artificiales. *Tecnología y Ciencias del Agua*, vol. V, núm. 3, mayo-junio, 2014, pp. 81-96
- [11] Toledo, M. Á., Salazar, F., Morán, R., Morera, L., Roa, A. y Santillán, D. (2013). Interpretación de los datos de auscultación de presas por métodos no convencionales. Jornada Técnica sobre Avances en investigación aplicada en seguridad hidráulica de presas. CEDEX, junio de 2013.
- [12] Salazar, F., Toledo M. Á., Oñate, E. y Morán, R. (2015). An empirical comparison of machine learning techniques for dam behaviour modelling. *Structural Safety*. Enviado.
- [13] Sánchez Caro, F. J. (2007). Seguridad de presas: aportación al análisis y control de deformaciones como elemento de prevención de patologías de origen geotécnico. Tesis Doctoral, E.T.S.I. Caminos, Canales y Puertos (UPM).