

Interpretación de los datos de auscultación de presas por métodos no convencionales

*Miguel Ángel Toledo**, *Fernando Salazar***, *León Morera**, *Alfonso Roa**, *David Santillán**
y *Rafael Morán **

** Universidad Politécnica de Madrid*

*** Centro Internacional de Métodos Numéricos en Ingeniería (CIMNE)*

1. Sobre la complejidad del comportamiento del conjunto presa-cimiento

Todo profesional familiarizado con las presas conoce la dificultad de analizar su comportamiento, a veces sorprendente, con el grado de aproximación deseado. Esta dificultad emana de la complejidad de todos los elementos que afectan al modo en que la presa y el terreno de apoyo responden a las solicitaciones, y también a la complejidad de las propias solicitaciones. Muchas de las características del conjunto presa-terreno escapan a las posibilidades de modelación mediante los modelos disponibles de base teórica.

Basta fijarse en la realidad de un cimiento rocoso, con sus sistemas de diaclasas de espaciamiento, grado de apertura y relleno variable a lo largo y ancho de la cerrada. A veces incluso con distintas formaciones presentes en la cerrada y siempre con características cambiantes en profundidad. Incluso el hormigón de las presas de fábrica, considerado generalmente como un material de características conocidas y estables, presenta características diferentes en las distintas partes de cada uno de los bloques, ya que su edad es distinta, en función de su fecha de puesta en obra, y los materiales utilizados para su fabricación son similares sólo en una primera aproximación. También difieren las condiciones de puesta en obra de cada tongada, lo que afecta a las características del hormigón resultante. Es además conocido el efecto de endurecimiento del hormigón que se produce con el paso del tiempo. Los materiales naturales utilizados en las presas de materiales sueltos tienen características todavía mucho más cambiantes y comportamientos más complejos que el hormigón.

Si nos fijamos en las solicitaciones, destaca la complejidad de la solicitación térmica en las presas de fábrica. Durante la puesta en obra e inicio de la explotación de la presa el efecto del calor de fraguado, junto con las condiciones de puesta en obra y del resto de factores siempre implicados en la transmisión del calor, hace difícil conocer con buena aproximación la solicitación térmica. Incluso años después, durante la explotación ordinaria de la presa, la solicitación térmica se ve afectada por factores como las oscilaciones del nivel en el embalse, la orientación de la presa, el número de días de sol, la radiación solar del entorno o la velocidad del viento. En la Figura 1 puede apreciarse la importancia del efecto de la orientación.

Por otra parte, la respuesta de la presa y el terreno al cambio de las solicitaciones es siempre diferida. Este retardo de la respuesta es evidente en el caso de la solicitación térmica, pues la inercia térmica de la presa es enorme, pero tampoco es instantánea la respuesta a cambios en el nivel del embalse. Además, el conjunto presa-

terreno es un sistema dinámico, en continuo cambio. Las diaclasas se van colmatando a lo largo del tiempo o su relleno puede ser arrastrado por el agua. Los sucesivos ciclos de llenado y vaciado del embalse modifican la deformabilidad del terreno con un cierto efecto de histéresis y esto afecta también a las tensiones y deformaciones de la presa. El conjunto presa-terreno evoluciona como un ser vivo a lo largo de todo el período de explotación. El cambio es generalmente lento y esto dificulta su percepción, pero los cambios de comportamiento pueden ser muy significativos en algunos casos.

Puede argumentarse que cualquier elemento o sistema, analizado con la profundidad necesaria, acaba por presentar un comportamiento complejo, porque así es la realidad. Los modelos introducen las simplificaciones necesarias en el elemento o sistema a estudiar para cumplir un determinado objetivo. La función de un modelo no es representar la realidad del objeto de estudio, sino simplificarlo de modo que esas simplificaciones permitan su análisis con una finalidad concreta. Si el modelo permite cumplir con su finalidad, el modelo resulta útil, con independencia de que la aproximación al objeto de estudio real sea mayor o menor. El campo de la ingeniería de presas tiene una larga tradición en el uso de modelos cuya finalidad es el diseño. Hay numerosos modelos numéricos, a veces de gran simplicidad, que han demostrado su utilidad para diseñar presas. Los modelos físicos a escala reducida, a pesar del inevitable efecto de escala, también han demostrado su utilidad para el diseño de aliviaderos y desagües. Para diseñar no es necesario conocer el comportamiento real, ni siquiera estar próximos a él, ya que el coeficiente de seguridad proporciona el margen necesario para que el diseño funcione a pesar de que el conocimiento sea sólo aproximado, a veces bastante alejado de la realidad.

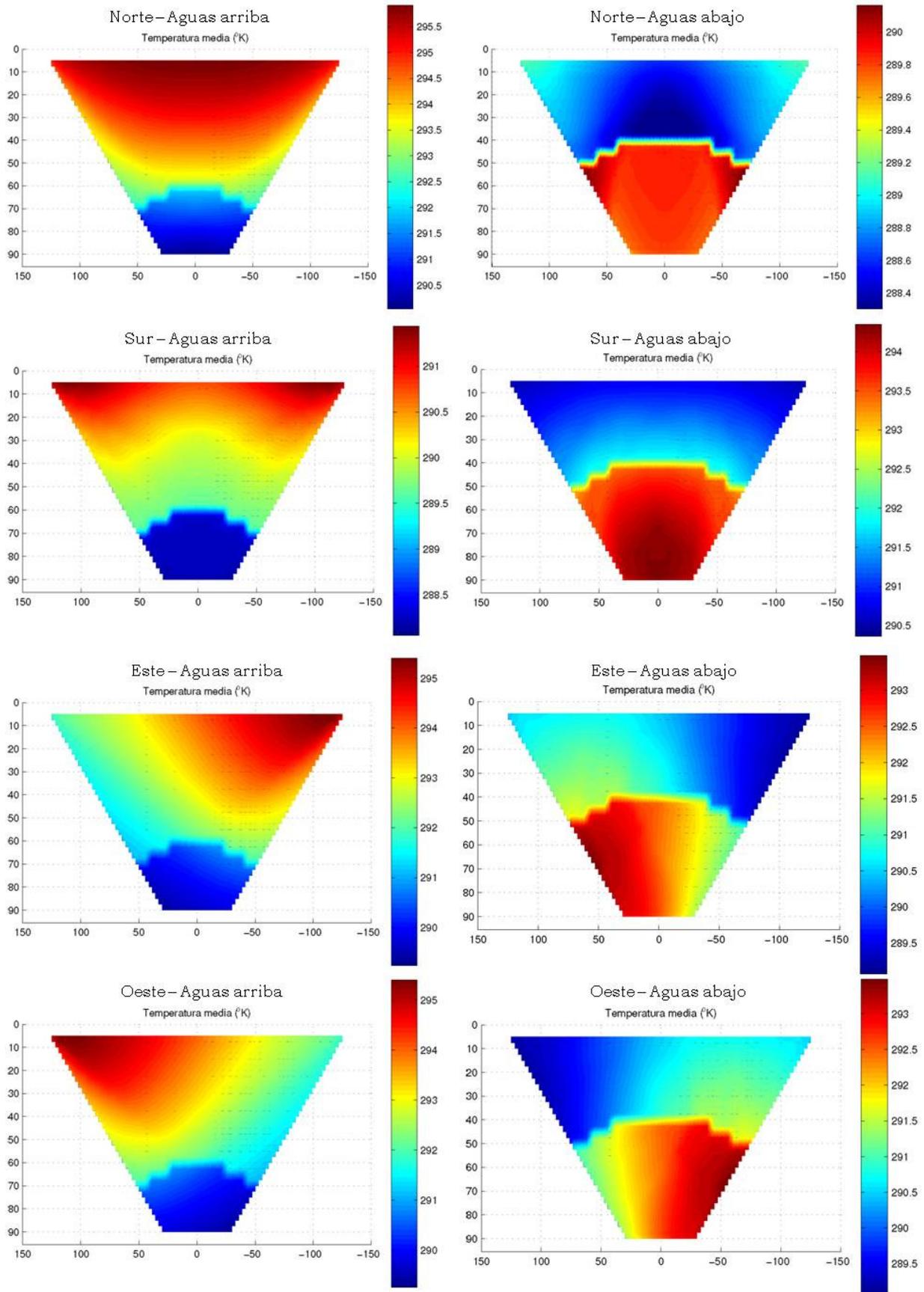


Figura 1. Distribución de las temperaturas medias en ambos paramentos de una presa bóveda para distintas orientaciones (direcciones hacia las que mira el paramento de aguas abajo). (Fuente: Santillán, 2013)

El análisis de la seguridad de una presa presenta características propias que lo diferencian del diseño. Cuando el explotador de una presa tiene que decidir, ante una incidencia o irregularidad de comportamiento, si comunica que hay riesgo de rotura de la presa, lo que implicará en muchas ocasiones la evacuación de poblaciones, necesita conocer con tanta aproximación como sea posible el comportamiento real de la presa. Lo mismo le sucede al propietario de un parque presístico, cuando tiene que decidir a qué presa o presas destina el presupuesto disponible para reparaciones, pues debe atender primero aquéllas que presentan un mayor riesgo real de avería o rotura. El análisis de la seguridad de una presa entra dentro del campo de los que han dado en llamarse *problemas del mundo real*, que escapan a la modelación teórica basada en las leyes de la física y deben analizarse a partir de los datos disponibles sobre su comportamiento. En las presas, la mayor parte de estos datos se obtiene mediante el sistema de auscultación.

2. Panorama actual

La finalidad principal del sistema de auscultación de una presa es evitar que llegue a producirse un accidente con consecuencias graves para la seguridad y bienes de la población asentada aguas abajo de la presa y también de su propietario. La idea esencial es que cualquier anomalía de comportamiento debe ser detectada en su inicio y con tiempo suficiente para evaluar su gravedad y posibles consecuencias y para, en caso de resultar necesario, tomar las medidas preventivas o correctoras necesarias. Si la anomalía no se detecta con tiempo suficiente, el papel de los datos de auscultación queda reducido a facilitar el análisis forense, que no es poco pero no es la finalidad principal.

Para el análisis del comportamiento de la presa y la detección de anomalías puede contarse con los datos de auscultación, además de la experiencia del profesional o profesionales encargados de su interpretación o de un modelo numérico de base física. Los sistemas de auscultación son muy dispares en su configuración, magnitud y fiabilidad. En las presas importantes los datos recogidos suelen ser abundantes en número, y siempre la bondad de los registros debe estar sometida a contraste, pues cualquier defecto de instalación o manipulación posterior puede causar que los registros sean incorrectos. La abundancia de los datos y la existencia de registros incorrectos o poco precisos, unida a la complejidad del comportamiento de la presa, hace difícil y costosa la interpretación por métodos convencionales. Ello ha llevado a un cierto descrédito de los datos de auscultación y su utilidad entre algunos profesionales del sector, algunos de ellos muy experimentados.

Entre los métodos convencionales de interpretación de los datos de auscultación, el más básico, que podemos denominar *método gráfico* y está ampliamente extendido, consiste en la representación de la evolución temporal comparada de distintas magnitudes, y de la relación entre variables, generalmente un parámetro de respuesta (como el movimiento medido en un péndulo) en función de otro causal (como el nivel de embalse). El *método gráfico* es interesante, pues permite detectar anomalías de un solo vistazo, pero la anomalía tiene que ser evidente para que la mera observación de los gráficos permita su detección. No obstante, los gráficos de evolución temporal y de relación entre variables permiten dar una idea intuitiva de cómo funciona la presa y ofrecen información importante que puede ser interpretada por un profesional experimentado. Naturalmente, los gráficos de relación entre variables quedan limitados a las posibilidades de una representación bidimensional, lo que deja fuera de este método numerosas relaciones.

El *método estadístico convencional*, que siempre va acompañado del método gráfico, permite cuantificar las relaciones entre parámetros, sin estar sujeto a las restricciones de una representación 2D, y constituye un modelo predictivo causal. El establecimiento de leyes de tipo estadístico entre *variables causales* (como el nivel de embalse o la temperatura) y *variables de respuesta* (como el caudal de filtración medido en un aforador o la deformación registrada por un extensómetro) permite obtener una fórmula para predecir el valor esperable de la variable respuesta en función de las variables causales. Además proporciona una medida sobre la bondad de la predicción (por ejemplo, el error cuadrático medio). No es un requisito del método que la relación deba establecerse entre variables causales y variables de respuesta, pero de hecho este es el esquema que se repite en cada modelo estadístico convencional. En la mente del ingeniero la causalidad está muy arraigada, y tiende a establecer en cualquier fenómeno relaciones entre las causas y sus consecuencias o efectos.

Para establecer las leyes estadísticas entre variables causales y variables de respuesta mediante los métodos estadísticos convencionales utilizados, es necesario hacer una hipótesis previa respecto de la naturaleza de la relación. Si se realiza una regresión lineal, se está asumiendo *a priori* que la relación entre las variables es lineal. Con frecuencia se hace la hipótesis, con mayor o menor fundamento según el caso, de que la respuesta viene dada por una combinación lineal de un determinado tipo de función, denominada *predictor*. Los coeficientes de la combinación lineal se determinan mediante un análisis de regresión lineal múltiple. En cierto modo podemos decir que este método *fuerza* a los datos a adaptarse a un tipo de relación preestablecido, aunque cabe esperar que en ocasiones la complejidad de la relación haga difícil o imposible conocer la verdadera naturaleza de la relación. Las posibilidades de los modelos estadísticos son inmensas pero en gran medida inexploradas todavía en el campo de la interpretación de los datos de auscultación de presas. Por otra parte, muchos de los modelos que se presentan bajo nombres específicos son en realidad modelos de tipo estadístico.

Las presas más importantes pueden disponer de un *modelo numérico de base física* del comportamiento tenso-deformacional, típicamente un modelo de elementos finitos o similar. Ya se comentó que la complejidad del comportamiento real de la presa y las propias limitaciones del sistema de auscultación dificultan que el ajuste entre el modelo y los datos registrados sea bueno en general. A pesar de ello, este tipo de modelos puede resultar de enorme utilidad para el explotador de la presa. Al ser modelos de base física, pueden proporcionar las *funciones predictoras* antes mencionadas para su uso en un *modelo estadístico convencional*. Un modelo numérico de base física puede ser útil para el análisis de las posibles causas de una anomalía detectada al interpretar los datos de auscultación, ya que permite simular el comportamiento de la presa en distintos escenarios. La calibración del modelo a partir de los datos registrados por el sistema de auscultación es esencial para que el ajuste sea el mejor posible. Cabe esperar que en el futuro próximo se produzcan avances en los procedimientos de calibración y en la capacidad y velocidad de computación que permitan introducir en los modelos de base física complejidades que habitualmente no se tienen en cuenta, como la variabilidad de la deformabilidad del material en las distintas zonas de la presa y su cimiento. Por otra parte, los modelos basados en datos funcionan bien en el rango en el que se dispone de datos suficientes, pero generalmente funcionan mal fuera de él, de modo que los modelos numéricos de base física resultan imprescindibles para el análisis de las situaciones extraordinarias. Finalmente, no hay que olvidar que durante la fase crítica de primer llenado del embalse, que supone la puesta en carga, apenas se dispone de datos registrados por el sistema de auscultación, por lo que los métodos basados en ellos resultan ineficaces en el momento más comprometido y el modelo numérico de base física puede ser casi el único apoyo.

Durante el primer llenado todavía no se tienen datos suficientes para conocer el comportamiento normal de la presa, y el modelo numérico de base física puede proporcionar una estimación de las variables de respuesta en función de las variables causales. No obstante, hay que tener en cuenta que este modelo aún no estará calibrado o lo estará de forma imperfecta con los datos de la fase inicial de llenado. Así se presenta el difícil problema de los *umbrales de alarma* en fase de primer llenado. El razonamiento subyacente parece ser el siguiente: para unos ciertos valores de las variables causales (básicamente las solicitaciones, aunque no exclusivamente), el valor de cada variable de respuesta debe estar en un *rango de respuesta normal*; si el valor se sale del rango esto se considera una *anomalía* de comportamiento que exige un análisis y en algunos casos una actuación. Frente a este planteamiento, que es razonable, cabe hacer alguna reflexión. El rango de respuesta normal, comoquiera que se defina, es desconocido en fase de primer llenado, pues la complejidad de comportamiento del conjunto presa-terreno impide su establecimiento *a priori* con un margen de error aceptable y útil. Para llenar este hueco se puede disponer del modelo numérico de base física y de la experiencia relativa a presas similares. Ambas fuentes deben complementarse, pues el modelo es específico de la presa pero aún no está calibrado y escapan a él diversas complejidades *del mundo real*, y los datos de presas similares se refieren a presas distintas de la presa en estudio, con semejanzas y diferencias. El estudio de un parque presístico suficientemente amplio podría permitir ahondar en el concepto de *semejanza de presas* y definir criterios cuantitativos fundamentados para el establecimiento de los umbrales de alarma con base en las medidas tomadas en presas reales, pero teniendo en cuenta las características particulares de la presa en estudio.

3. El conjunto presa-cimiento como sistema complejo

Ya se comentó la complejidad del comportamiento de las presas. Los sistemas complejos se caracterizan por la presencia de *propiedades emergentes*, que sólo pueden comprenderse a la luz de las relaciones entre los distintos elementos que conforman el sistema, y que es imposible comprender mediante división del sistema en elementos independientes que se analicen de forma separada (visión reduccionista). El enfoque sistémico o visión de red, frente al enfoque reduccionista, tiene en cuenta las relaciones entre todos los elementos del sistema como factor esencial para la comprensión del sistema en su conjunto. Pero si el conjunto presa-cimiento puede entenderse como un sistema, ¿cuáles son los elementos que lo componen? Una presa no es como un mecano, que pueda descomponerse en piezas sueltas, ni como una red de líneas aéreas, en la que las ciudades (elementos) están relacionados o conectados por los vuelos que van de una ciudad a otra (relaciones o conexiones entre elementos). La cuestión anterior se vuelve sencilla si consideramos que el comportamiento de la presa queda expresado a través del sistema de auscultación, de modo que estudiando el sistema de auscultación podremos comprender el comportamiento de la presa. La complejidad del comportamiento del conjunto presa-terreno queda patente a través de la relación existente entre las series de registros de los instrumentos del sistema de auscultación. Así pues, los *elementos* de nuestro sistema son las *series temporales de registros* y las *conexiones o relaciones* entre elementos vienen dadas por las *relaciones de dependencia* entre dichas series. Hay que resaltar que las relaciones no tienen por qué ser de causa a efecto. De hecho la mayor parte de las relaciones no lo son. La relación de dependencia entre las lecturas de dos estaciones de un mismo péndulo no es causal, y sin embargo es muy intensa.

Una vez establecida la configuración del enfoque sistémico para el análisis del comportamiento de una presa a través del tratamiento de sus datos de auscultación es posible utilizar técnicas desarrolladas en diversos campos como el de la inteligencia artificial, los sistemas expertos o en general la minería de datos, y por supuesto también los desarrollados en el campo de la estadística. Entre estas técnicas, que son muy numerosas, se encuentran las *redes bayesianas*, las *redes neuronales* y las *redes complejas*, que trataremos aquí junto con los *bosques aleatorios*. A pesar del nombre a veces intimidatorio, la idea que subyace en cada una de estas técnicas es sencilla. Ninguna de ellas constituye una solución milagrosa, pero la combinación de las técnicas convencionales con técnicas novedosas puede resultar muy potente. Hay que aclarar que las técnicas que aquí denominamos novedosas lo son en la interpretación de datos de auscultación de presas, pero todas ellas han sido ya aplicadas con éxito a diferentes campos del conocimiento. Contrariamente a lo que pudiera esperarse, una de las ventajas de algunas de estas técnicas es la simplicidad de planteamiento y uso. A continuación se destaca el fundamento y algunas aplicaciones de cada una de las técnicas antes mencionadas.

4. Redes complejas

La *teoría de redes complejas* se ha aplicado fundamentalmente a campos del conocimiento en los que resulta evidente la imposibilidad de modelar el fenómeno en estudio mediante modelos de base física, como la sociología o la biología. El fundamento es sencillo: el sistema queda definido por una serie de elementos (nudos) conectados por algún tipo de relación (arista, conexión o relación). Así se ha estudiado la propagación de las noticias en una red social o la de virus informáticos en internet. El conjunto de conexiones entre los nudos queda recogido en una matriz cuadrada, denominada *matriz de adyacencia*. Si el elemento a_{ij} de la matriz es nulo, eso indica que los elementos i y j del sistema no están relacionados o conectados. Si existe relación, dicho elemento será la unidad o bien un número o peso que expresa la intensidad de la conexión. En el segundo caso la red se denomina *ponderada*. La distribución de las conexiones permite caracterizar el tipo de red, y cada tipo de red tiene unas propiedades como tal, que son propiedades emergentes. Resulta fácil medir la importancia o *centralidad* de cada uno de los elementos (nudo) en el conjunto del sistema. Todo ello se basa en el álgebra matricial y la *teoría de grafos*, ya que el sistema puede representarse mediante un conjunto de círculos, que representan los elementos, unidos mediante líneas que representan las conexiones o relaciones. Esta representación se denomina *grafo*.

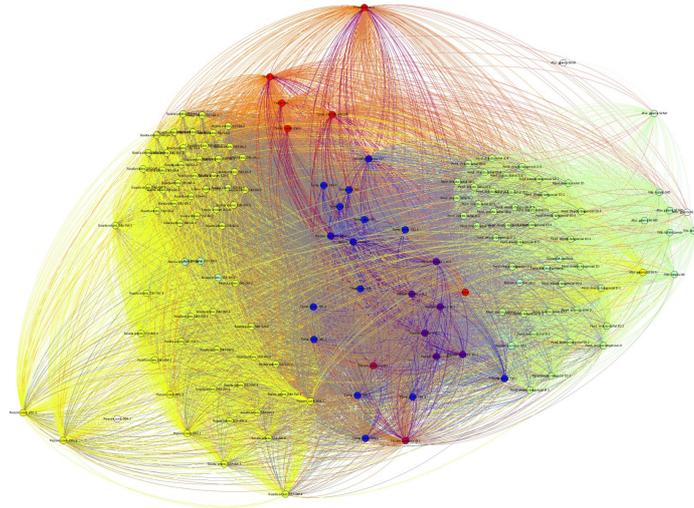


Figura 2. Ejemplo de grafo que representa las relaciones entre series registradas en una presa bóveda

Comenzamos presentando esta técnica porque la representación del sistema complejo presa-cimiento a través del sistema de auscultación mediante el grafo correspondiente permite tener una idea de conjunto que puede facilitar el análisis. Como hemos visto antes, los elementos del sistema son las series de registros realizadas a lo largo del tiempo. Hay que definir el criterio para cuantificar la relación entre dos series (peso en una red ponderada). Aunque las posibilidades son diversas, como una de las opciones más evidentes se ha considerado el *coeficiente de determinación* (R^2) que mide la bondad del ajuste de una función que relaciona los valores de ambas series (X e Y):

$$R^2 = \frac{\sum_i (f_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2}$$

donde f_i es el valor predicho por la función ajustada para el valor x_i , e \bar{y} es la media de los registros de la variable y . El valor de este coeficiente está comprendido entre 0 y 1, de manera que cuanto más próximo a la unidad se encuentre, mejor será el ajuste conseguido por la función. Así, la relación entre las series se considera máxima si $R^2=1$ y mínima si $R^2=0$. Naturalmente, la relación entre dos series cualesquiera, si existe, no tiene por qué ser lineal. Por eso se ha definido un listado de funciones de ajuste, y se considera como peso de la conexión el R^2 de la función que proporcione un valor mayor, y por tanto un mejor ajuste. Esta operación se repite para todas las parejas de nudos de la red y así se obtiene la matriz de adyacencia que define el sistema. Existen diversos algoritmos que permiten obtener el grafo correspondiente a partir de la matriz de adyacencia.

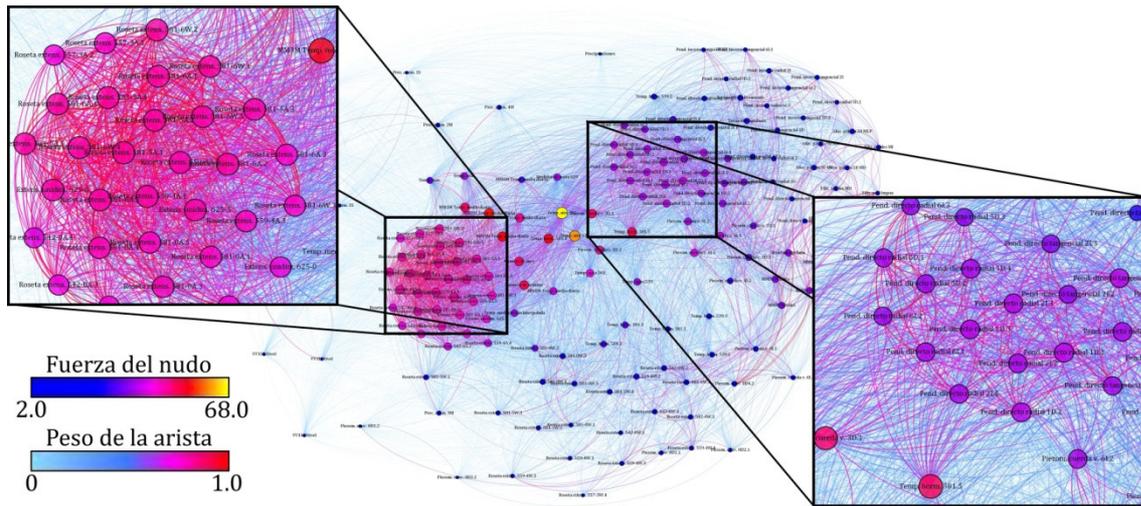


Figura 3. Grafo de la red compleja resultante del análisis de 163 series de registros de una presa bóveda. Los recuadros muestran el detalle de algunas agrupaciones de aparatos y series intensamente interconectadas

Son especialmente interesantes las representaciones en que el tamaño de cada nudo es mayor si el nudo tiene una mayor importancia en la red de acuerdo con algún tipo de medida de *centralidad*, y el grosor de las aristas de conexión resulta indicativo de la intensidad de la relación. Además, los elementos más relacionados se sitúan más próximos, quedando agrupados en familias o *clusters*. Los grafos de este tipo proporcionan una imagen global del sistema de auscultación muy descriptiva que puede facilitar su interpretación en combinación con otros métodos de análisis. La topología definida puede entenderse como un mapa que destaca la interrelación entre los distintos elementos del sistema, poniendo de manifiesto relaciones existentes entre sus distintas partes, a veces insospechadas.

Se comentó antes que la respuesta de la presa a los cambios de las solicitaciones no es inmediata. Esta complejidad puede introducirse en la red añadiendo variables que podemos denominar *variables de inercia*: medias móviles o velocidades de variación de las variables causales, típicamente el nivel del embalse y la temperatura ambiente. Se puede observar cómo en una presa bóveda la intensidad de la conexión del movimiento en un péndulo con una media móvil de la temperatura correspondiente a un período de de dos o tres meses es mucho mayor que la intensidad de la conexión con el valor diario de la temperatura, como cabía esperar. El sistema facilita el análisis de la inercia de la respuesta mediante la introducción de diversas variables de inercia y el análisis de la intensidad de las conexiones resultantes.

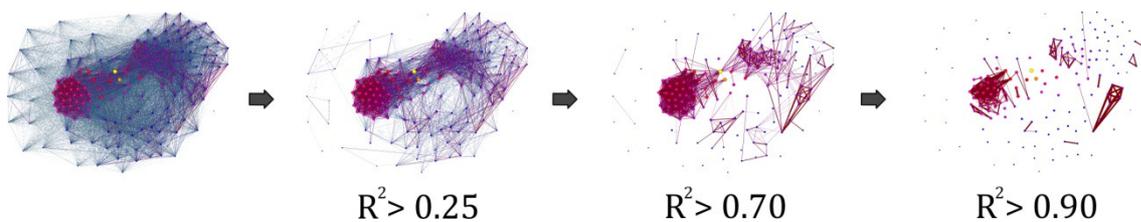


Figura 4. Apariencia de la red completa (izquierda) que representa el sistema de auscultación de una presa bóveda, y de la red filtrada para diferentes valores del coeficiente de determinación. Las redes filtradas mantienen las conexiones de mayor intensidad

La evolución temporal del comportamiento del conjunto presa-cimiento constituye otra complejidad que debe tenerse en cuenta. Puede abordarse su consideración inicial del modo más sencillo introduciendo *variables evolutivas*, como el año, el número de día contado secuencialmente desde el inicio de los registros o el número de ciclo de llenado o vaciado del embalse. Se ha observado que la intensidad de la conexión (recordemos que viene dada simplemente por el coeficiente de determinación) del número de día con una serie o nudo del sistema resulta indicativa de la importancia de la *deriva* temporal. Esta deriva, una vez identificada, puede estudiarse por otros procedimientos.

En primera instancia los modelos de redes complejas no son más que mapas de relaciones que sirven de base para el análisis posterior mediante técnicas diversas bajo la mirada atenta de un profesional experimentado. No obstante, los modelos de redes complejas pueden servir de base topológica para la definición de modelos predictivos, basados en métodos estadísticos o de redes neuronales, por ejemplo.

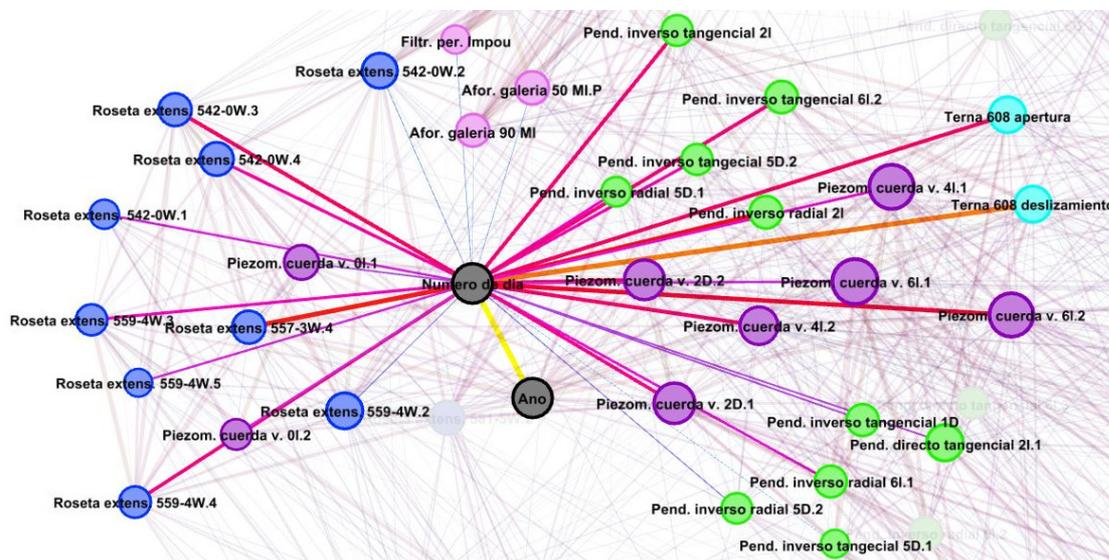


Figura 5: Detalle de las conexiones del elemento/serie "número de día" con diversos elementos/series. El "número de día" indica simplemente la posición absoluta de cada día en la serie total de registros. La intensidad de la conexión con esta serie es indicativa de la magnitud de la deriva temporal de la serie correspondiente. La intensidad de la conexión está relacionada con el grosor y el color de la arista de conexión. La representación gráfica facilita el análisis conjunto, poniendo de manifiesto semejanzas y diferencias entre instrumentos del mismo tipo y de unos tipos con otros

5. Redes neuronales

Los modelos de redes neuronales se desarrollaron dentro del campo de la inteligencia artificial, si bien simultáneamente se desarrolló también en el campo de la estadística otra técnica cuya idea central es básicamente la misma: extraer relaciones lineales de las variables de entrada al modelo y modelar los resultados como una función no lineal de ellas. Por tanto, una característica importante de esta técnica es que permite detectar relaciones no lineales entre las variables de entrada y las variables de salida del modelo. Así, la técnica de RRNN queda emparentada con la técnica de regresión no lineal múltiple. Su potencia y simplicidad de uso ha llevado a su utilización en multitud de disciplinas, incluida la ingeniería civil, con algunas aplicaciones a la ingeniería de presas.

Se trata de un modelo matemático inspirado en el funcionamiento del cerebro de los animales. Un modelo de RRNN está formado por un conjunto de unidades de procesamiento denominadas neuronas que trabajan en paralelo y realizan operaciones matemáticas sencillas. Las neuronas se organizan en capas, están interconectadas entre sí y son capaces de aprender, adaptarse y generar soluciones a partir de un conjunto de datos de entrenamiento. Existen una gran variedad de modelos de RRNN en función de la estructura de la red. El más extendido como modelo de predicción de comportamiento es el perceptrón multicapa. Se trata de un modelo neuronal unidireccional compuesto por un mínimo de tres capas de neuronas: una capa de entrada, una capa de salida y una o varias capas ocultas. En la figura 6 se representa un perceptrón con una capa oculta. Las variables de entrada en este ejemplo quedan integradas en el vector $\mathbf{X} = [x_1, x_2]$ de dos componentes (por ejemplo, x_1 podría ser la temperatura ambiente y x_2 el nivel de agua en el embalse) y la variable de salida, única en este caso, es $\mathbf{Y} = [y_1]$; podría tratarse, por ejemplo, del caudal de filtración registrado en un aforador. Los círculos representan las neuronas y las líneas la interconexión entre ellas. Cada neurona está conectada a todas las situadas en la capa siguiente. La fuerza de la conexión está cuantificada por un valor, denominado peso, ponderándose así la información que se transmite de una neurona a otra.

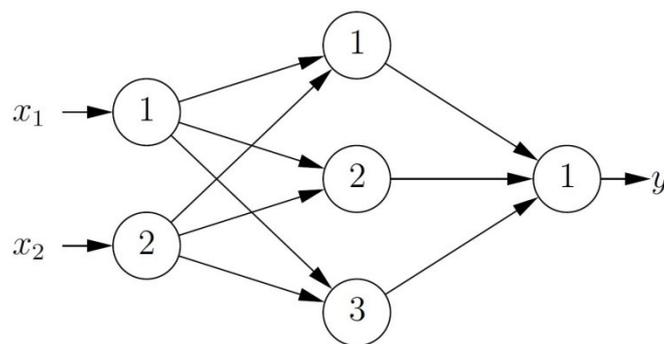


Figura 6. Perceptrón multicapa

En consecuencia, la entrada total a una neurona es la suma de sus entradas multiplicada cada una de ellas por el peso de la conexión entre la neurona origen de la información y la neurona destino, menos una cantidad denominada *sesgo*. La entrada total a la neurona es el argumento de una función sencilla (como una función lineal o sigmoideal) cuyo resultado es la salida de la neurona. El entrenamiento de la red neuronal, equivalente a la calibración de un modelo, consiste en determinar los pesos y sesgos que proporcionan el mejor ajuste. Para ello es necesario disponer de un conjunto de datos (entradas y salidas del modelo) suficientemente amplio.

Al entrenar un modelo de redes neuronales se procede de forma semejante a como se ajusta cualquier otro modelo similar: el modelo se entrena con un conjunto de datos extraído del total de datos disponibles (datos de entrenamiento), mientras que el resto se reserva para comprobar la capacidad predictiva del modelo una vez entrenado, y por tanto determinados sus pesos y sesgos (datos de validación).

En el marco del proyecto de investigación SEPRISIS, desarrollado por CIMNE, Ofiteco y la UPM (SERPA, subgrupo del Grupo de Hidroinformática y Gestión del Agua), se aplicaron técnicas de redes neuronales para modelar las filtraciones y los movimientos de los péndulos de una presa bóveda, a partir de los datos de un caso piloto. La respuesta del sistema en filtraciones a los cambios de las variables (fundamentalmente el nivel de

embalse) resulta de una gran complejidad. En la figura 7 se muestra el resultado proporcionado por el modelo de redes neuronales al considerar el descenso y el aumento del nivel del embalse a lo largo de un período de cinco meses con distintas velocidades de variación del nivel del embalse. Puede verse cómo la red ha captado la influencia de dicha velocidad en el caudal de filtración, con un curioso efecto de inercia y rebote para niveles ascendentes, más acusado cuando la velocidad es mayor. En la figura 8 puede verse un efecto parecido con datos registrados en un aforador y su modelación con la red neuronal.

Para la modelación de los movimientos registrados en los péndulos se plantearon diversos tipos de modelos, desde *modelos simples*, cuyas entradas quedaban reducidas al nivel de embalse y la temperatura, representada mediante una media móvil para tener en cuenta el fuerte efecto de la inercia térmica, hasta *modelos completos* que incorporaban gran número de variables. Los mejores resultados para el péndulo analizado se obtuvieron utilizando la media móvil de tres meses, lo que sugiere que considerar medias mensuales de temperatura en fase de diseño puede resultar inexacto. También se plantearon *modelos de corto plazo*, que incorporaban como entradas al modelo los registros de movimientos en los días anteriores al de la predicción. Lógicamente estos modelos incrementan enormemente la precisión de la predicción respecto de los modelos tradicionales, como puede verse en la figura 9, a costa de la necesidad de realizar un entrenamiento diario. Pueden resultar útiles para detectar el inicio de averías, pues pequeñas desviaciones respecto del valor pronosticado por el modelo serán indicativas de una anomalía, dada la gran precisión del modelo. En cambio, este tipo de modelos resultará inútil para la detección de variaciones progresivas del sistema, como las derivas temporales, ya que el propio modelo irá adaptándose día a día.

Hay que destacar que todos los modelos desarrollados en el marco del proyecto SEPRISIS fueron causales, es decir, las entradas al modelo eran variables causales, como el nivel, la temperatura ambiente o medias móviles o velocidades de variación de ambas, y las salidas del modelo eran variables de respuesta del sistema, como filtraciones o movimientos.

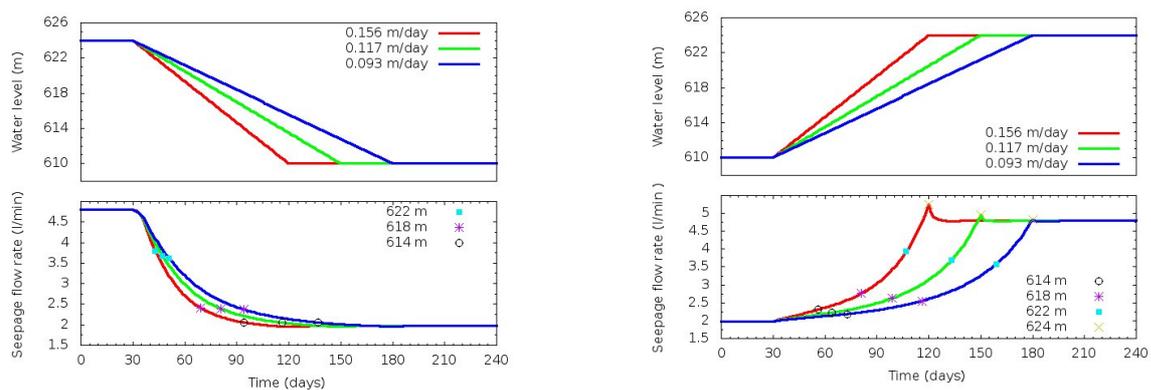


Figura 7. Resultado proporcionado por un modelo de redes neuronales al considerar el descenso y el aumento del nivel del embalse a lo largo de un período de cinco meses con distintas velocidades de variación del nivel del embalse

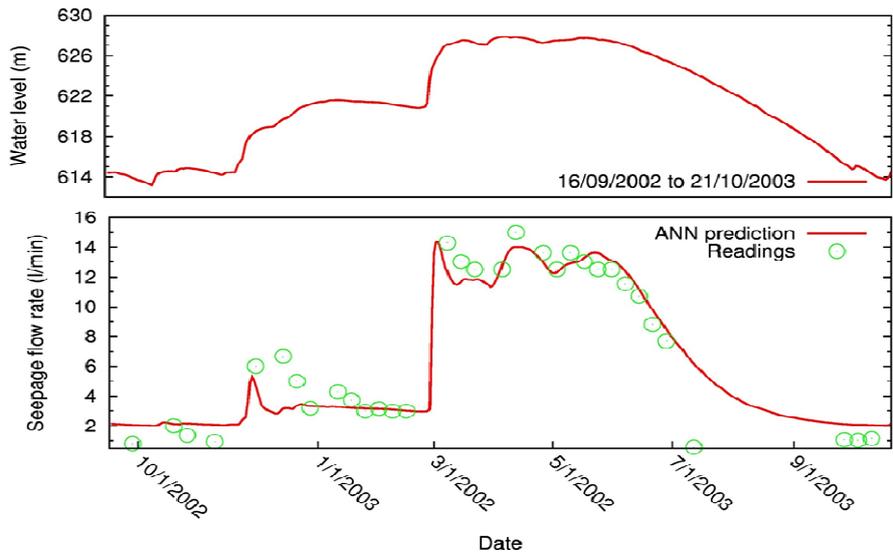


Figura 8. Datos registrados en un aforador y su modelación mediante una red neuronal, que parece haber capturado el comportamiento complejo con efecto de "rebote" después de un ascenso de nivel prolongado

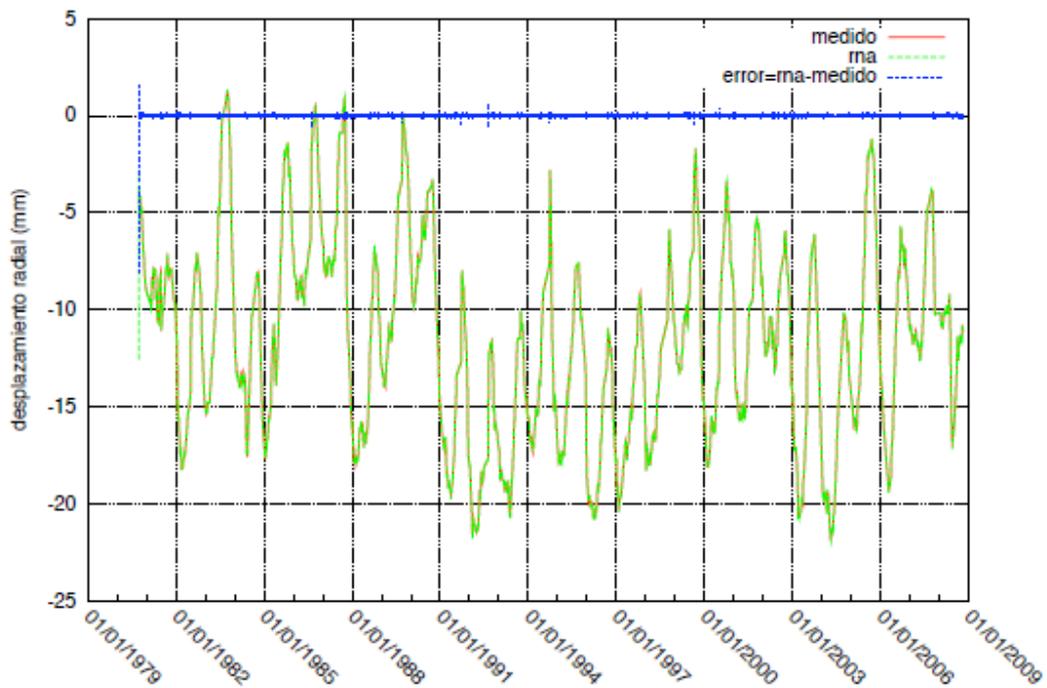


Figura 9. Serie de datos registrados de movimientos en una estación de un péndulo, y de los datos pronosticados por una red neuronal de corto plazo. Puede observarse que el error, mostrado en color azul, es mínimo, lo que facilitaría la detección del inicio de una anomalía

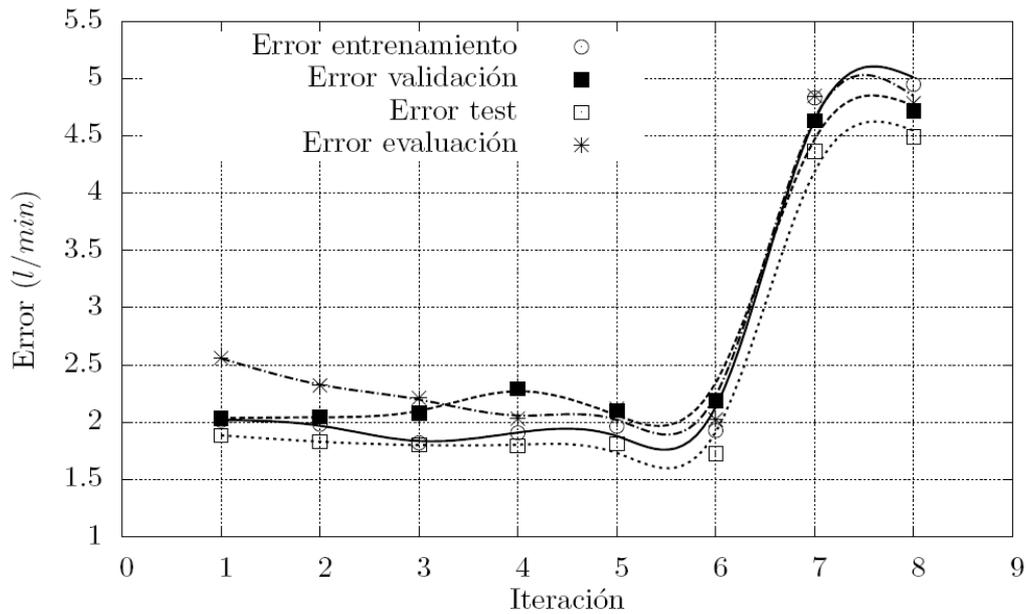


Figura 10. Las redes neuronales pueden también utilizarse para analizar la influencia de diversas magnitudes sobre una variable objetivo. La figura representa la evolución del error en la predicción del caudal registrado en un aforador durante el proceso de descarte de variables realizado por un algoritmo desarrollado para mantener únicamente las variables que afectan significativamente al ajuste de la predicción. En cada iteración se descarta una variable de las candidatas seleccionadas a priori, con variables de inercia de diversos períodos incluidas. Puede verse cómo en la iteración 7 el error aumenta de forma muy significativa, lo que indica que la última variable descartada debe mantenerse. El análisis de las variables mantenidas puede esclarecer aspectos del comportamiento de la presa y ayudar a su mejor comprensión global

6. Redes bayesianas

Las redes neuronales son modelos deterministas. Si bien el resultado del entrenamiento depende de los pesos de partida utilizados al inicio del proceso de optimización de la red, una vez determinados los pesos y sesgos óptimos, que proporcionan el mejor ajuste, las salidas de la red quedan expresadas mediante una fórmula sencilla en función de las variables de entrada. Los modelos de redes bayesianas, en cambio, adoptan un enfoque probabilístico a partir del conocido teorema de Bayes sobre probabilidad condicionada:

$$P(A|B) = P(B|A) \cdot P(A) / P(B)$$

Dados dos sucesos A y B que en general no sean independientes, la probabilidad de ocurrencia del suceso A condicionado a B es igual al producto de la probabilidad de ocurrencia de B condicionado a A por la probabilidad de ocurrencia del suceso A dividido por la probabilidad de ocurrencia del suceso B. La idea esencial de los modelos de redes bayesianas es que en un sistema cuyos elementos están relacionados la probabilidad de que un elemento se encuentre en un estado determinado queda condicionada por el estado de los demás elementos del sistema que estén relacionados con él. En el caso de nuestro sistema presa-cimiento, los elementos son series de registros (niveles, temperaturas, filtraciones, movimientos...) y el estado en cada momento viene dado por el registro correspondiente. Para cada registro de una serie, la probabilidad de que el registro adopte un valor determinado depende de los valores de las demás series, registrados por otro instrumento o incluso por el mismo, puesto que algunos instrumentos registran varias series de datos. Hay que destacar que la

dependencia no tiene por qué ser causal con carácter general, y en la mayor parte de los casos no lo es. Está claro que hay una dependencia entre el valor de la serie *nivel de embalse* y el de la serie *movimiento radial* de una estación de lectura de un péndulo, y en este caso la dependencia es causal. Pero no es menos cierto que hay una dependencia intensa entre las series *movimiento radial* correspondientes a dos estaciones de lectura de un mismo péndulo. En este segundo caso la relación de dependencia no es causal, pero está claro que una vez conocido el valor registrado en una estación, el valor esperable en la otra está ya fuertemente condicionado. Si el objetivo del modelo es detectar *anomalías* interesa tener una predicción lo más ajustada posible del valor o rango de valores esperables en una situación dada. Parece lógico pensar que la consideración de los condicionamientos impuestos *de facto* por cualquier elemento del sistema sobre el valor pronosticado, con independencia de su causalidad, ha de proporcionar un ajuste más fino que la consideración exclusiva de las variables causales.

Los modelos de redes bayesianas permiten considerar el condicionamiento probabilístico. Se dispone de diversos algoritmos para determinar la estructura de la red, que es arborescente. Cada nudo (variable, serie...) del sistema tiene en general unos *padres*, que son los nudos que lo condicionan. El valor del nudo se determina a partir de los valores de sus *padres*. A su vez, cada nudo puede tener una serie de *hijos*, cuyo valor condiciona. El valor del nudo *padre* interviene en la predicción del nudo *hijo*, junto con el resto de los *padres*. Así, el condicionamiento se propaga en cascada desde la parte superior del árbol (nudos sin padres) hacia la parte inferior (nudos sin hijos). Los nudos situados más arriba en el árbol que define la estructura del sistema resultan más condicionantes para el conjunto del sistema que los nudos situados en la parte inferior.

Hemos hablado de predicción de valores, pero en realidad el resultado del modelo no es un valor, sino una función de distribución de probabilidad. El valor más probable puede tomarse como el valor pronosticado y por ejemplo la desviación estándar puede tomarse como un índice de la calidad del pronóstico. Por otra parte, el valor realmente medido puede ubicarse en la función de distribución de probabilidad para juzgar si se trata de un valor *normal* o de una *anomalía*, de acuerdo con un criterio predefinido.

En la figura 11 se muestra a modo de ejemplo una representación de un modelo de redes bayesianas de una presa bóveda tomada como caso piloto. En este modelo se incluyen únicamente las series de niveles y temperaturas, y medias móviles de ambas magnitudes, junto con los movimientos medidos en los péndulos. Se ha podido comprobar que el ajuste mejora en casi todos los casos los resultados obtenidos mediante técnicas estadísticas convencionales, a veces de forma muy notable. Hay que tener en cuenta que un mejor ajuste de la predicción implica una mayor capacidad para detectar una anomalía, y por tanto una mejor posición para detectar el inicio de una avería en una fase temprana.

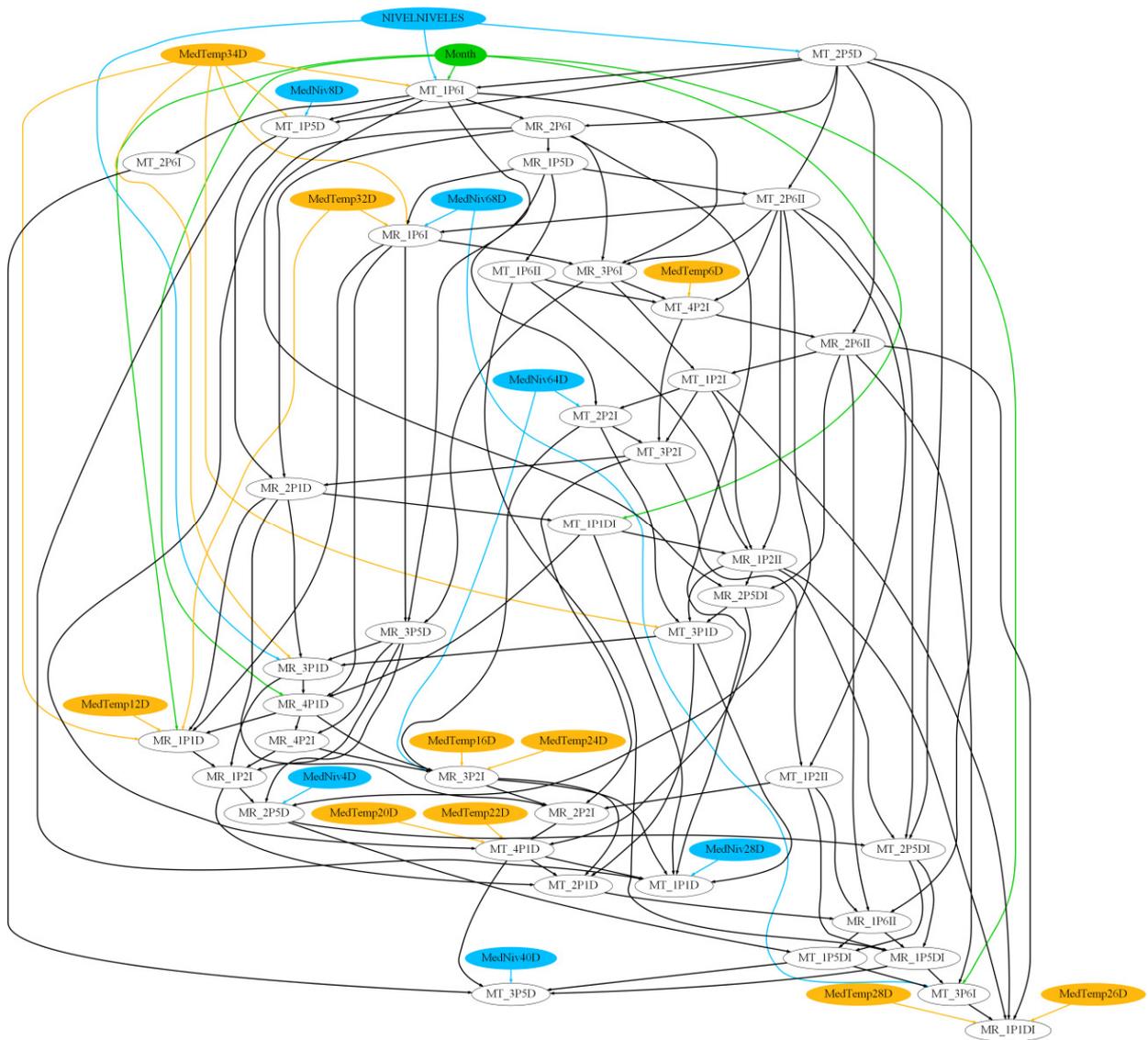


Figura 11. Representación de una red bayesiana que modela el sistema de auscultación de una presa bóveda.

El algoritmo utilizado para la determinación de la estructura de la red permite introducir un cierto grado de aleatoriedad, de modo que partiendo del mismo conjunto de datos de partida, cada ejecución del algoritmo determinará una red distinta. Esto puede hacerse simplemente eligiendo los conjuntos de entrenamiento y validación de forma aleatoria y distinta en cada entrenamiento. Disponer de gran número de redes similares, pero no iguales, es posible hoy gracias a la capacidad de computación de los ordenadores actuales, y resulta interesante al menos por dos motivos. En primer lugar, porque el *modelo de redes bayesianas* puede convertirse en un *modelo de conjunto*, considerando como predicción el valor medio de las predicciones de todas las redes generadas. Al considerar el valor medio de un número suficientemente elevado de predicciones de modelos distintos, los errores, que no serán siempre en la misma dirección, tenderán a compensarse en cierta medida, resultando una predicción más precisa que la de un modelo concreto. Cuando los modelos individuales utilizados tienen errores importantes, la consideración del conjunto es imprescindible. Este no es el caso de los modelos de redes bayesianas. No obstante, se ha comprobado una reducción apreciable del error de la predicción del movimiento de los péndulos en una presa bóveda, considerando valores medios a partir de un conjunto de 1.000 modelos. Por tanto, parece factible e interesante manejar modelos de conjunto formados por redes bayesianas.

Por otra parte, la generación de un número elevado de modelos a partir del mismo conjunto de datos puede servir de base para la generación de una estructura de *red compleja*. El número de modelos en los que aparece una determinada conexión entre dos nodos o series puede tomarse como medida de la intensidad o fuerza de la conexión. De esta manera puede generarse una red compleja alternativa a la basada en el coeficiente de determinación que se comentó con anterioridad. A esta red puede aplicarse la teoría de redes complejas para su análisis como sistema. Hay que destacar que esta red ya no tiene en general forma de árbol, como sucede en las redes bayesianas. En la figura 12 se muestra una red compleja obtenida del modo descrito a partir de 1.000 modelos de redes bayesianas. La red se muestra filtrada para resaltar únicamente las conexiones de mayor intensidad. La red incluye numerosas variables tipo *media móvil*. Puede observarse que sólo algunas de ellas aparecen insistentemente en las redes bayesianas generadas, dando origen a conexiones fuertes.

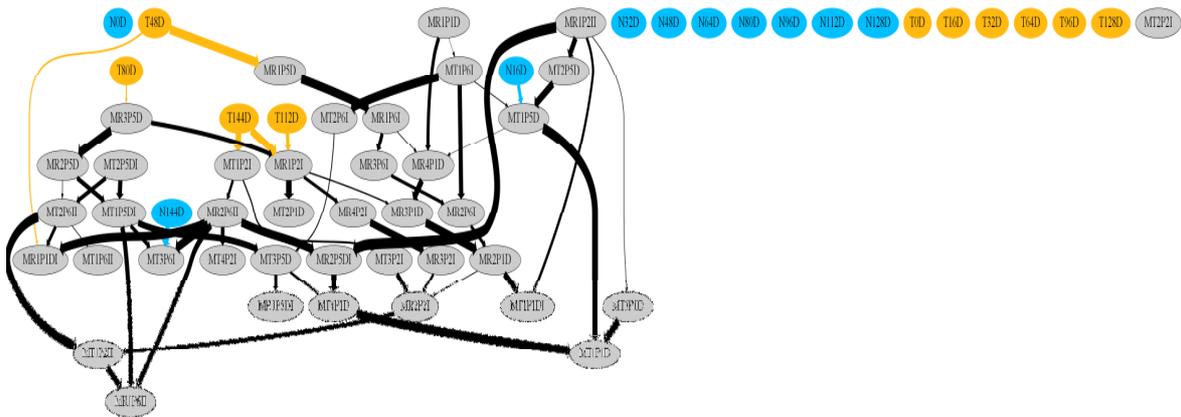


Figura 11. Red compleja obtenida a partir de 10.000 redes bayesianas y filtrada para representar únicamente las conexiones de mayor intensidad. Los elementos de la red incluyen la temperatura ambiente(T) y el nivel del embalse(N) y medias móviles de diversos períodos de ambos, y los movimientos en los péndulos (M). Puede verse cómo numerosas medias móviles de temperatura y de nivel ofrecen conexiones de baja intensidad y han quedado fuera del grafo, junto con un movimiento.

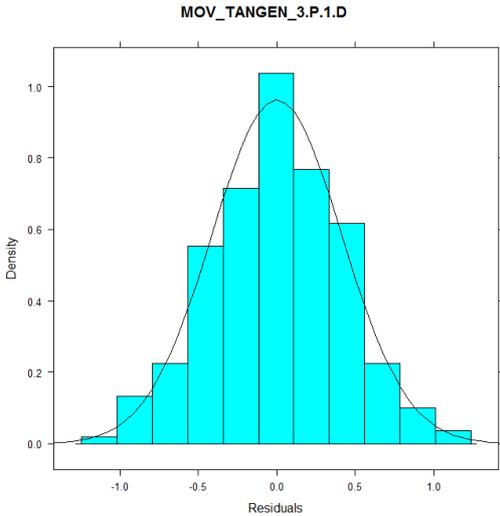


Figura 12. Distribución del residuo (diferencia entre el valor pronosticado y el valor medido) en la predicción del movimiento tangencial en uno de los péndulos de una presa bóveda mediante una red bayesiana.

7. Bosques aleatorios

Los bosques aleatorios son modelos que permiten predecir el valor de una determinada variable (*variable objetivo*) a partir de una serie de *variables predictoras*, cuyo valor es conocido. Como los demás métodos anteriormente comentados, requiere de unos datos de entrenamiento, a partir de los cuales se ajusta el modelo al caso de estudio. Un bosque aleatorio está formado por un conjunto de árboles de decisión. La predicción del bosque es el promedio de las predicciones de los árboles de decisión que lo forman. Por tanto, se trata de un *modelo de conjunto*.

Los árboles de decisión se basan en la división sucesiva del conjunto de datos observados en grupos de casos “similares”. Suelen denominarse *árboles de regresión* cuando la variable objetivo es continua, y *árboles de clasificación* si es discreta o categórica. En adelante se utilizará por tanto el término árbol de regresión, ya que en auscultación de presas se trata de predecir variables continuas. La predicción de un árbol de regresión es en general un valor constante para cada grupo, igual a la media de los valores observados. Para explicar el proceso de generación de un árbol de decisión, se utiliza un ejemplo sencillo relacionado con la auscultación de presas: supongamos que se desea ajustar un modelo para predecir el caudal de filtración en un determinado aforador a partir únicamente del nivel de embalse. La figura 13 muestra la relación entre las variables de entrada (nivel de embalse) y objetivo (caudal de filtración).

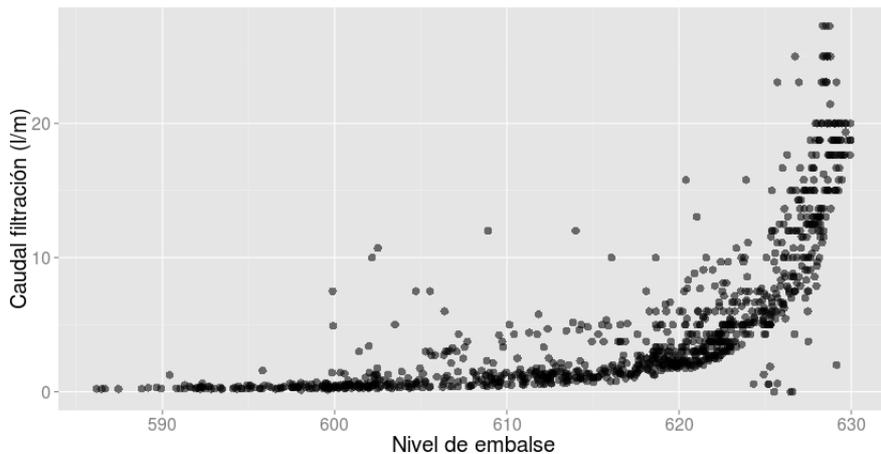


Figura 13. Caudal de filtración en función del nivel de embalse.

En primer lugar, se dividen los datos en dos regiones según el nivel de embalse (en el ejemplo, según sea mayor o menor de 625,6). A continuación, una de las regiones creadas se subdivide a su vez en dos, y el proceso continúa hasta que se alcanza algún criterio de parada. En el ejemplo, el resultado final es la división de los casos en 7 grupos. La predicción del modelo es la media de los valores observados en cada grupo, y por tanto el resultado es una sucesión de escalones (figura 15).

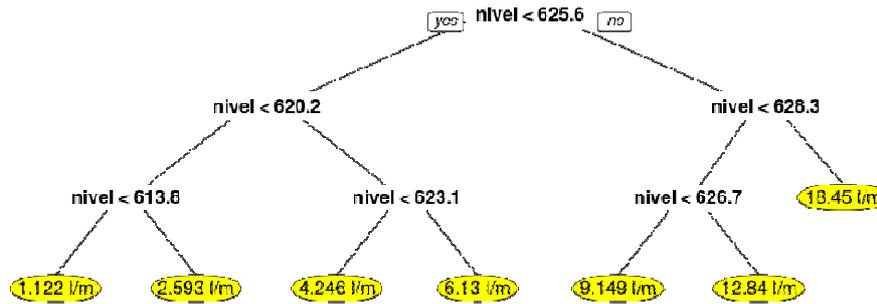


Figura 14. Árbol de regresión para la predicción del caudal de filtración en función del nivel de embalse.

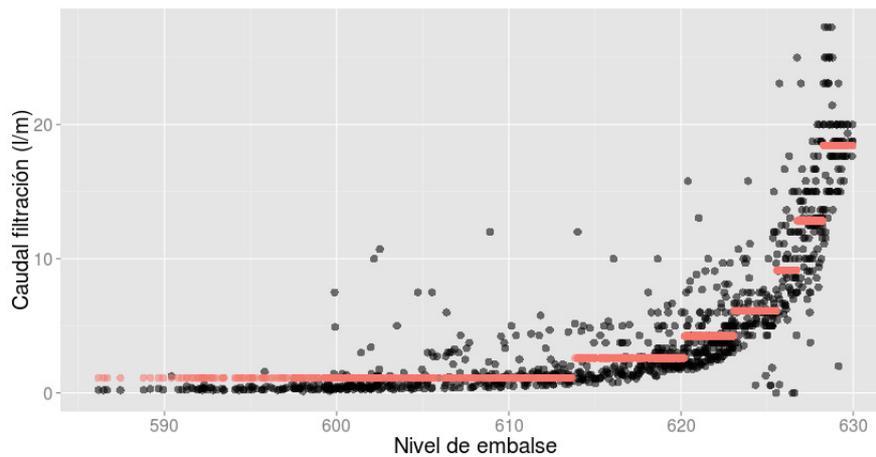


Figura 15. Predicción del árbol de regresión (en rojo), en comparación con los datos registrados (en negro).

El algoritmo de generación del árbol determina la mejor división posible en cada paso como la que minimiza el error de predicción. Si hubiera más de una variable predictora, se calcularía para cada una de ellas la división óptima y a continuación se selecciona la variable que produce una división que proporciona una predicción más precisa. Algunas de las propiedades más representativas de los árboles de decisión son las siguientes: a) su coste computacional es moderado; b) se adaptan bien a problemas no lineales; c) permiten manejar sin problemas mezclas de variables continuas y discretas; además, las continuas pueden tener rangos muy diferentes, lo que evita tener que transformarlas previamente, como ocurre con otros métodos; d) permiten considerar valores no medidos; e) los valores atípicos no modifican sustancialmente el resultado; f) son inestables, en el sentido de que una pequeña variación de los datos de entrenamiento puede producir una variación importante del resultado; este problema puede convertirse en una ventaja si se utilizan métodos avanzados basados en árboles, como es el caso de los bosques aleatorios.

Los *bosques aleatorios* constituyen uno de los denominados *modelos de conjuntos*, basados en la generación de un número (generalmente elevado) de modelos sobre una misma base de datos (o sobre subconjuntos de ella). La predicción se calcula como la media de las predicciones de los modelos individuales. Un bosque aleatorio está formado por un conjunto de árboles de decisión generados a partir de perturbaciones de los datos observados. El aspecto clave que caracteriza el método es que busca que los árboles sean independientes entre sí. Para ello, la diferencia principal con un árbol de decisión convencional es que en cada división, en lugar de considerar todas las variables predictoras disponibles para seleccionar la que minimiza el error, se analiza únicamente un subconjunto aleatorio de las mismas. De este modo, se aprovecha la propiedad de los árboles de decisión por la

cual pequeñas perturbaciones en las primeras divisiones del espacio provocan resultados apreciablemente diferentes. Al introducir aleatoriedad en la construcción de cada árbol, se consigue un conjunto de ellos sensiblemente independientes, de modo que se captura una proporción mayor de los patrones presentes en los datos de entrenamiento, y el resultado final mejora.

Los modelos de bosques aleatorios tienen algunas características singulares. El criterio de parada es muy laxo, con lo que se generan árboles que tienden a sobreaprender los datos de entrenamiento. Esto permite capturar patrones poco frecuentes en los datos de entrenamiento. Cada árbol se genera a partir de un conjunto de datos diferente, obtenido a partir de los originales tomando de éstos una muestra aleatoria con repetición. Es decir, cada nuevo árbol se construye a partir de una muestra en la que aproximadamente un tercio de los datos originales aparecen una vez, otro tercio aparece repetido, y el tercio restante no aparece.

Los bosques aleatorios han cobrado gran popularidad como método para generar modelos predictivos por su sencillez de programación y sus buenos resultados en diversas aplicaciones. Si bien suelen catalogarse como un modelo *de caja negra*, al igual que las redes neuronales, se han desarrollado herramientas que permiten cuantificar cómo afecta cada variable al resultado final. En concreto, se define el *índice de importancia* de una variable como la variación del error de predicción que se produce al modificar aleatoriamente su valor, manteniendo el resto de variables. Las variables más importantes provocarán un mayor aumento del error al ser permutadas.

Se ha elaborado un modelo de bosques aleatorios para la predicción de la filtración en los aforadores de una presa tomada como caso piloto. La variable que se pretende predecir es por tanto el caudal de filtración en cada uno de los aforadores. En la tabla 1 se muestra el número de datos disponible en cada uno de los seis aforadores considerados. Además de los datos de aforo, del nivel de embalse y de numerosas magnitudes que no son objeto de este trabajo, se miden en la presa variables ambientales como la precipitación y la temperatura ambiente.

Aforador	Margen	Nº datos disponible
md50pr	Derecha	1023
md90pr	Derecha	748
totmd	Derecha	1071
mi50p	Izquierda	1066
mi90pr	Izquierda	1001
totmi	Izquierda	1021

Tabla 1: Número de datos disponibles en cada aforador

Las variables utilizadas como entrada al modelo son: a) el número de día del registro, contado a partir de la puesta en carga de la presa, b) el año; c) el nivel de embalse el día de la lectura; d) la temperatura media ambiental; e) la precipitación acumulada durante los 30 días anteriores a la lectura; f) la velocidad media de variación del nivel de embalse en los 10 días anteriores a la lectura; g) la media móvil de 60 días del nivel de embalse. Los datos disponibles se han dividido en dos grupos: el primero se utiliza para ajustar los parámetros del modelo (datos de entrenamiento), y el segundo para comprobar la bondad del ajuste (datos de validación). La división se ha realizado de dos formas: a) El 60% de los datos más antiguos para entrenamiento, y el 40% más reciente para validación. Este criterio pretende simular la aplicación del modelo en condiciones de explotación de la presa, y es el criterio utilizado en el análisis de seguridad de la presa, así como en el trabajo de Santillán *et al.*, que utiliza redes neuronales; b) División aleatoria en todo el periodo registrado, con un 70% para

entrenamiento y un 30% para validación. Este criterio pretende garantizar que ambos grupos pertenecen a la misma muestra estadística. En la tabla 2 se muestra el error de predicción resultante en cada caso. Se observa que, si bien el modelo es capaz de adaptarse de forma similar a los datos de entrenamiento con ambos criterios de división, el error de predicción cuando el modelo se aplica al grupo de datos de validación es menor si la división es aleatoria, como cabía esperar. Esto indica que el sistema presa-cimiento evoluciona a lo largo del tiempo, y es por tanto un sistema dinámico.

División de los datos	RMSE (l/min)				Valor medio (l/min)
	60%-40% temporal		70%-30% aleatoria		
	Entrenamiento	Validación	Entrenamiento	Validación	
Aforador					
md50pr	2,15	2,64	2,12	1,77	5,08
md90pr	0,40	1,58	0,55	0,47	1,57
totmd	2,56	4,42	2,41	2,64	7,32
mi50p	0,48	0,41	0,45	0,41	2,21
mi90pr	0,16	0,24	0,17	0,10	0,64
totmi	0,67	1,05	0,62	0,60	2,91

Tabla 2: Errores de predicción del modelo (raíz del error cuadrático medio). Se muestra también el caudal medio de los registrados.

En la figura 16 se muestra el resultado para el aforador “md50pr”, que es al que corresponde también la figura 15. Puede apreciarse la mejora de la predicción del bosque aleatorio respecto del árbol de regresión individual.

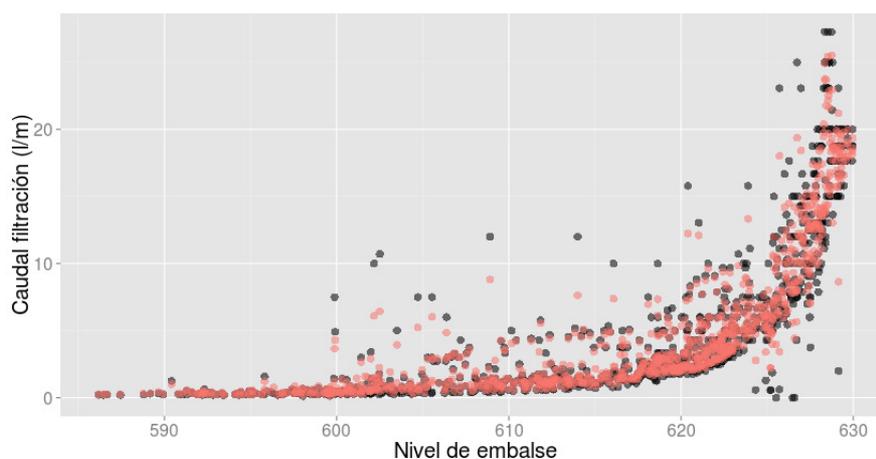


Figura 16: Predicción del modelo basado en bosques aleatorios (en rojo), en comparación con los datos registrados (en negro). Aforador “md50pr”.

Se ha calculado también el *índice de importancia* de las variables predictoras consideradas, y llama la atención el hecho de que las relativas al período de medición (número de día desde la puesta en carga y año de lectura) son siempre más relevantes que otras como la temperatura o la precipitación. En algunos casos, llegan a serlo más incluso que el nivel de embalse. Para verificar estos resultados, se ha dibujando la relación entre el nivel de embalse y el caudal de filtración separando los registros por intervalos temporales. La figura 17 muestra un gráfico de este tipo (correspondiente al aforador “totmi”), donde se observa claramente cómo el caudal de filtración depende en gran medida del período de la vida de la presa. En este caso concreto, la filtración ha disminuido con el tiempo de forma significativa.

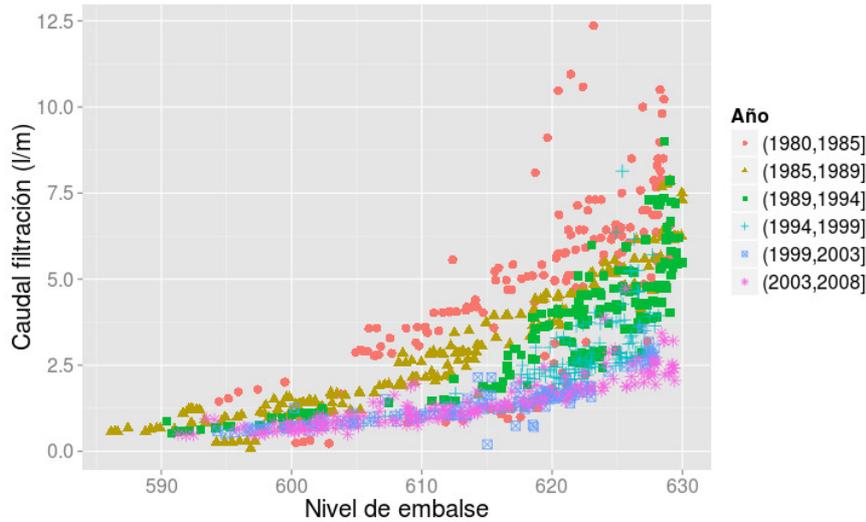


Figura 17: Caudal de filtración en el aforador “totmi” en función del nivel de embalse, con división de datos por períodos.

Los resultados obtenidos hasta el momento con los bosques aleatorios sugieren que pueden ser una herramienta útil no solo como modelo de predicción, sino también para obtener información acerca del comportamiento de la presa, y del efecto de las variables de entorno. En la actualidad se está trabajando para desarrollar criterios de selección de variables, así como para determinar cómo evoluciona la precisión del modelo en función del número de datos disponible.

8. Resumen y conclusiones

Hemos visto cuatro técnicas que facilitan el análisis y modelación de sistemas complejos. Son sólo ejemplos de una multitud de técnicas desarrolladas en diversos campos, generalmente alejados de la ingeniería civil, por lo que su aplicación a sistemas de ingeniería civil ha sido relativamente escasa, muy escasa en el caso de la ingeniería de presas. Todas ellas se han demostrado su utilidad y se han aplicado con éxito al análisis de multitud de sistemas de diversa índole. En el caso concreto de la ingeniería de presas es necesario un cambio de mentalidad, adoptando una visión sistémica y aceptando la complejidad del comportamiento del conjunto presa-terreno. No se trata de sustituir unos métodos por otros, sino 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. La capacidad de computación de los ordenadores permite actualmente aplicar diversas técnicas, con modelación incluso *en tiempo real*, para extraer mucha de la abundante información contenida en los datos de auscultación de las presas. Sólo analizando en profundidad y con técnicas adecuadas para el tratamiento de grandes masas de datos podremos conocer y comprender el *comportamiento real* de las presas.

Por otra parte, cabe esperar el desarrollo de técnicas de detección de instrumentos defectuosos, basadas en una visión de sistema, o de detección de patrones de avería. Considerando tanto el primer llenado como el período de explotación, cabe reflexionar sobre la interpretación de las *anomalías de comportamiento*, una vez detectadas. La anomalía de un registro puede deberse, como sucede en muchas ocasiones, a un funcionamiento defectuoso del instrumento que registra la medida o a un error del sistema de transmisión de datos. Es preciso desarrollar

procedimientos que permitan detectar este tipo de anomalías para evitar falsas alarmas, que restan eficacia y credibilidad al sistema. Estos procedimientos previsiblemente deberán basarse en el análisis de las relaciones cruzadas entre series de registros. Dada la interrelación entre las distintas series temporales de registros, en muchas ocasiones será poco probable que la anomalía de un solo registro pueda deberse a una anomalía de comportamiento de la presa. Generalmente una anomalía de comportamiento tendrá reflejo en distintas variables registradas, dando origen a un *cuadro de anomalía*. De la misma forma que el médico busca distintos síntomas para identificar la enfermedad de su paciente, al interpretar los datos de auscultación de una presa es necesario encontrar diversas anomalías en los registros para identificar la avería que pueda estar produciéndose. Los profesionales experimentados en el análisis de datos de auscultación saben que deben proceder de esta manera, pero lamentablemente el estado de nuestras técnicas de diagnosis está muy lejos del desarrollo alcanzado en la medicina. Esto es lógico por distintos motivos. Uno de ellos es que el número de personas enfermas es inmensamente mayor que el de *presas enfermas*. Todavía está por hacer un catálogo de síntomas cruzados que, una vez identificados, permita establecer el tipo de avería que está desarrollándose en una presa. Todo esto será posible utilizando técnicas de detección de patrones ya desarrolladas y que es necesario poner a disposición de los datos de auscultación de las presas.

De forma inmediata es posible mejorar sensiblemente los *modelos de predicción* para la detección de *anomalías*. Hay que tener en cuenta que cualquier mejora en el ajuste del modelo a los datos permitirá estrechar el *rango de valores normales*, y por tanto mejorará la capacidad de detección de una anomalía de comportamiento. El resultado es la detección más temprana de la anomalía, lo que resulta esencial para disponer del tiempo necesario para el análisis y la acción, si fuera necesaria. La creciente automatización de los sistemas de auscultación facilita la implantación de sistemas de modelación en tiempo real y la detección temprana de cualquier anomalía. Automatización, capacidad de computación de los ordenadores y técnicas de tratamiento masivo de datos se complementan para ofrecer posibilidades de análisis todavía insospechadas en el campo de la interpretación de los datos de auscultación de presas. El proyecto iCOMPLEX, actualmente en marcha, y en el que participan Dacartec, CIMNE y el grupo SERPA de la UPM, permitirá ahondar en la aplicación de las técnicas mencionadas al estudio de la *seguridad real* de las presas.

Finalmente hay que destacar dos ideas fundamentales: 1) la inspección visual es irremplazable y debe siempre formar parte de un sistema de auscultación; 2) todas las técnicas disponibles sólo pueden estar al servicio de profesionales experimentados en el campo de la ingeniería de presas, que deben guiar el análisis e interpretar los resultados obtenidos.

9. Agradecimientos

Los autores quieren agradecer a la Agencia Catalana de l'Aigua la autorización para el uso de los datos de una de sus presas con fines de investigación, y a Ofiteco las gestiones realizadas para ello.

Los autores quieren también agradecer al Ministerio de Medio Ambiente y Medio Rural y Marino, y al Ministerio de Economía y Competitividad su apoyo mediante la financiación de los siguientes proyectos del plan Nacional de Investigación, respectivamente:

ESTUDIO DE LA SEGURIDAD DE PRESAS E IDENTIFICACIÓN DE ESCENARIOS DE RIESGO MEDIANTE SISTEMAS INTELIGENTES (SEPRISIS), del Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica 2008-11; Número de expediente: 048/RN08/04.5.

DESARROLLO DEL SOFTWARE iCOMPLEX PARA EL CONTROL Y EVALUACIÓN DE LA SEGURIDAD DE INFRAESTRUCTURAS CRÍTICAS (iCOMPLEX), del Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica 2008-11; Número de expediente: : IPT-2012-0813-390000.

10. Referencias

Adeli, H. 2001, "Neural networks in civil engineering: 1989–2000", *Computer-Aided Civil and Infrastructure Engineering*, vol. 16, no. 2, pp. 126-142.

Ahmadi Nedushan, B. & Chouinard, L.E. 2003, "Use of artificial neural networks for real time analysis of dam monitoring data", *Proceedings, Annual Conference - Canadian Society for Civil Engineering*, pp. 1987.

ASCE Task Committee on Guidelines for Instrumentation & Measurements for Monitoring Dam Performance 2000, *Guidelines for instrumentation and measurements for monitoring dam performance*, Amer Society of Civil Engineers.

Breiman, L. Random forests. *Machine learning*, 45(1), 5-32, 2001.

Delgado-Hernández, D., Morales-Nápoles, O., & De-León-Escobedo, D., A continuous bayesian network for earth dams' risk assessment: an application. *Structure and Infrastructure Engineering*, (ahead of print), (1-14, 2012).

Flood, I. & Kartam, N. 1994, "Neural networks in civil engineering. I: Principles and understanding", *Journal of Computing in Civil Engineering*, vol. 8, no. 2, pp. 131-148.

Ganuer, R., Poggi, J.M., Tuleau, C. (2008). Random Forests: some methodological insights. arXiv: 0811.3619.

Junrui, C., Yanqing, W. & Shouyi, L. 2004, "Analysis of coupled seepage and stress fields in rock mass around the Xiaowan arch dam", *Communications in Numerical Methods in Engineering*, vol. 20, no. 8, pp. 607-617.

Kim, Y. & Kim, B. 2008, "Prediction of relative crest settlement of concrete-faced rockfill dams analyzed using an artificial neural network model", *Computers and Geotechnics*, vol. 35, no. 3, pp. 313-322.

Li, P., Lu, W., Long, Y., Yang, Z. & Li, J. 2008, "Seepage analysis in a fractured rock mass: The upper reservoir of Pushihe pumped-storage power station in China", *Engineering Geology*, vol. 97, no. 1, pp. 53-62.

Mata, J. 2011, "Interpretation of concrete dam behaviour with artificial neural network and multiple linear regression models", *Engineering Structures*, vol. 33, no. 3, pp. 903-910.

Miao, X.Y., Chu, J.K., Qiao, J. & Zhang, L.H. 2012, "Predicting Seepage of Earth Dams Using Neural Network and Genetic Algorithm", *Advanced Materials Research*, vol. 403, pp. 3081-3085.

Rafiq, M., Bugmann, G. & Easterbrook, D. 2001, "Neural network design for engineering applications", *Computers & Structures*, vol. 79, no. 17, pp. 1541-1552.

Santillán, D., Morán, R., Fraile, J.J., Toledo, M.Á. Forecasting of dam flow-meter measurements using artificial neural networks. En Romeo García et al. (eds) *Dam Maintenance and Rehabilitation II*, CRC Press, Londres, 2010, pp 183-189.

Santillán, D., Caracterización de la sollicitación térmica en presas bóveda en fase de operación, teniendo en cuenta su tridimensionalidad y el efecto de la radiación solar, (tesis doctoral en ejecución)

Tayfur, G., Swiatek, D., Wita, A. & Singh, V. 2005, "Case Study: Finite Element Method and Artificial Neural Network Models for Flow through Jeziorsko Earthfill Dam in Poland", *Journal of Hydraulic Engineering*, vol. 131, no. 6, pp. 431-440.

Waszczyszyn, Z. 2011, "Artificial neural networks in civil engineering: another five years of research in Poland", *Computer Assisted Mechanics and Engineering Sciences*, vol. 18, pp. 131-146.

Waszczyszyn, Z. 2006, "Artificial neural networks in civil and structural engineering: Ten years of research in Poland", *Computer Assisted Mechanics and Engineering Sciences*, vol. 13, no. 4, pp. 489-512.

Zurada, J.M., Malinowski, A. & Usui, S. 1997, "Perturbation method for deleting redundant inputs of perceptron networks", *Neurocomputing*, vol. 14, no. 2, pp. 177-193.