

Nuevas técnicas para el análisis de datos de auscultación de presas y la definición de indicadores cuantitativos de su comportamiento

F. Salazar, E. Oñate

*Centro Internacional de Métodos Numéricos en Ingeniería. Campus Norte UPC. Ed. C-1.
Barcelona 08034*

L. Morera, R. Morán y M.Á. Toledo
Universidad Politécnica de Madrid

1. Introducción

El control de la seguridad de presas es un aspecto fundamental de su explotación. El elemento clave es el sistema de auscultación, considerado como el conjunto de a) los aparatos instalados en la presa que registran determinadas variables, b) el sistema de transmisión de datos y c) una metodología para evaluar el estado de seguridad en función de los datos registrados. En la práctica, el sistema debe determinar si los datos registrados se corresponden con una situación normal, o si existe un riesgo de avería.

Para ello, se seleccionan los aparatos más representativos del comportamiento general de la estructura, y se definen unos umbrales que se corresponden con el estado de seguridad de la presa (situación normal o algún grado de emergencia), que se incluyen en el Plan de Emergencia de la Presa como indicadores cuantitativos del estado de seguridad (Ministerio de Medio Ambiente 2001).

La propia Guía Técnica para Elaboración de Planes de Emergencia de Presas incluye una clasificación de los métodos que pueden emplearse para definir los indicadores, señalando algunos de los pros y los contras de cada uno de ellos:

- Estadísticos, que establecen la relación entre las variables externas e internas a partir de las series de observaciones reales. Solo pueden aplicarse a presas en servicio “durante un periodo largo de tiempo”.
- Deterministas, que se basan en la modelización del comportamiento de la presa (típicamente con elementos finitos). Pueden aplicarse a presas de nueva construcción, pero solo si es posible “caracterizar con fiabilidad la realidad estructural de la presa y su cimiento y su comportamiento y, por tanto, los parámetros que lo condicionan”.

- Mixtos, que combinan los dos anteriores.

La Guía recomienda también (y los autores lo suscriben) aplicar el juicio ingenieril antes de declarar cualquier escenario de emergencia. Sin embargo, del mismo modo establece que “cualquier cambio significativo que pueda tener un efecto negativo sobre la seguridad y no sea explicable directamente en función de otros parámetros (nivel de embalse, temperatura, etc.) debe considerarse causa suficiente para la declaración del Escenario 0 de control de la seguridad”.

En la práctica, es relativamente frecuente que ocurran errores de lectura que producen registros fuera de rango de situación normal. El responsable de explotación puede aplicar su juicio para establecer que dicha medición no constituye una anomalía y por tanto no requiere la declaración de un escenario de emergencia. Sin embargo, sería útil disponer de herramientas capaces de ayudar a tomar una decisión que en algunos casos puede ser comprometida.

Es difícil establecer umbrales a partir de modelos deterministas, porque la información disponible de la presa y el cimiento es muchas veces insuficiente para obtener resultados precisos. Por ello los indicadores cuantitativos suelen basarse en modelos estadísticos. Las herramientas de este tipo que se utilizan más habitualmente son relativamente simples, limitándose frecuentemente a la regresión lineal múltiple. El modelo más utilizado es el Hydrostatic-Season-Time (HST), basado en una combinación lineal del nivel de embalse, la estación del año y el tiempo de vida de la presa. En ocasiones se añaden variables derivadas de las registradas para aumentar la complejidad del modelo y hacerlo más flexible con el objetivo de aumentar la precisión de las predicciones.

En otros campos de la ciencia en que el volumen de datos disponible es mucho mayor, se han desarrollado herramientas para su tratamiento y para la generación de modelos de predicción. Algunas de ellas, como las redes neuronales, ya han sido aplicadas al caso de la auscultación de presas (Santillán *et al.* 2014, Mata, 2011, Simon *et al.* 2013, Salazar *et al.*, 2015a, Salazar *et al.*, 2015b) con resultados prometedores.

La aplicación de estas técnicas puede ayudar a mejorar la precisión de los modelos de predicción, y a entender mejor el comportamiento de la presa (Toledo *et al.*, 2013). Un modelo más preciso puede ayudar además a definir umbrales de emergencia más fiables.

El principal inconveniente de estos modelos es que son más difíciles de interpretar que el HST. Mientras que en éste puede extraerse directamente la contribución de cada una de las acciones (nivel de embalse, temperatura, tiempo) en la respuesta de la presa, en otros modelos más complejos los efectos no son aditivos, suelen ser no-lineales, y basarse en un conjunto de variables mayor. Todo ello ha provocado que sean calificados con frecuencia como modelos de “caja negra” (p.e. Olden y Jackson, 2002). La mayor parte de los trabajos publicados sobre la aplicación de estas técnicas en seguridad de presas se limitan a evaluar el error de predicción. No se interpreta el modelo obtenido y por tanto no se

pueden extraer conclusiones sobre el estado de seguridad de la presa. Como excepción, cabe citar los trabajos de Santillán *et al* (2014) y Mata (2011).

Nos encontramos por tanto ante una disyuntiva: el modelo HST es ampliamente conocido y utilizado y fácilmente interpretable. Sin embargo, se basa en unas hipótesis que no son ciertas, como que la temperatura es independiente del nivel de embalse (Tatin *et al.*, 2013), y su precisión es limitada. Por otra parte, se dispone de unas técnicas más flexibles y precisas, pero más difíciles de implementar y analizar.

Ante esta situación, el presente trabajo tiene por objetivo proponer un método para interpretar un modelo de predicción del comportamiento de presas basado en bosques aleatorios (Breiman, 2001). El resultado de dicha interpretación se compara con el obtenido con el método HST.

2. Metodología

2.1. Caso de estudio

Se dispone de datos de auscultación de la presa de La Baells en el periodo 1.980-2.008. La figura 1 contiene el alzado esquemático y la sección de la presa, con la situación de algunos de los aparatos de auscultación. La descripción completa de los datos disponibles puede encontrarse en (Salazar *et al.*, 2015a).

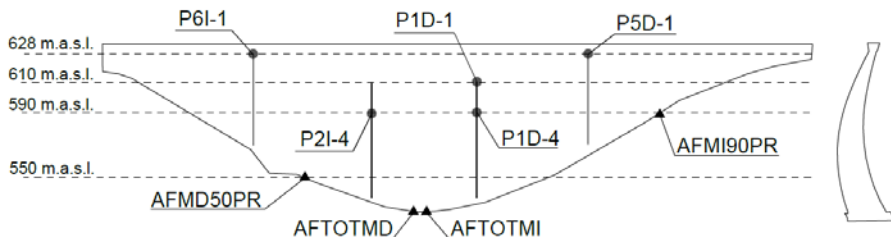


Figura 1. Alzado esquemático y sección de la presa de La Baells. Sobre el primero se señala la situación de algunos de los aparatos de auscultación.

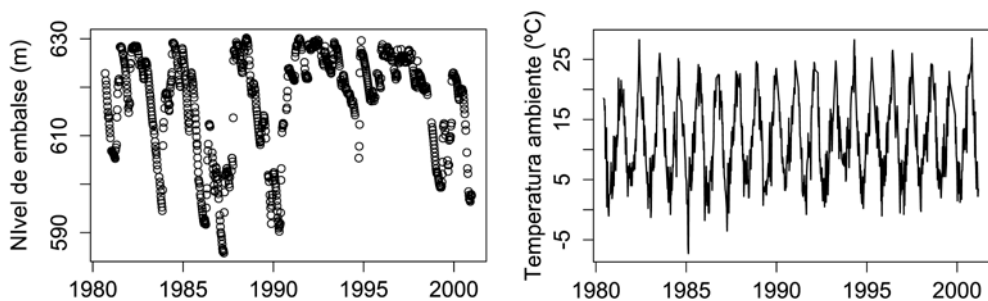


Figura 2. Nivel de embalse y temperatura media diaria en el periodo de análisis

Para el presente análisis se han seleccionado los datos de nivel de embalse y temperatura, para analizar los desplazamientos radiales registrados en el péndulo P1D1 (Figura 1). Se considera el periodo 1.980 – 2.000, durante el que se dispone de registros con frecuencia aproximadamente semanal (figura 2).

2.2. Método HST

La fórmula más habitual del modelo HST, que es la que se ha utilizado en este trabajo, es:

$$\hat{f}(h, s, t) = a_1 \cdot h + a_{21} \cdot h^2 + a_3 \cdot h^3 + a_4 \cdot h^4 + a_5 \cdot e^{-t} + a_6 \cdot t + a_7 \cdot \cos(s) + a_8 \cdot \text{sen}(s) + a_9 \cdot \text{sen}(2s) + a_{10} \text{sen}(s) \cos(s) \quad [1]$$

donde \hat{f} es la función que define el comportamiento de la variable objetivo que se pretende evaluar, h es el nivel de embalse, t es el tiempo en días desde la puesta en carga de la presa, y s se calcula como:

$$s = \frac{d}{365,25} \cdot 2\pi \quad [2]$$

siendo d el número de días transcurridos desde el 1 de enero del año correspondiente.

Para identificar el efecto de cada una de las variables exteriores (temperatura, carga hidrostática y tiempo), basta con extraer los términos de [1] que dependen de cada una de ellas y dibujar su contribución parcial a la variable objetivo. Es decir, dibujar por separado $f(h)$, $f(s)$ y $f(t)$.

2.3. Bosques aleatorios

Los bosques aleatorios (BA) pertenecen a un tipo de modelos basados en datos denominados “no paramétricos”. El motivo es que no se hace ninguna suposición a priori sobre el tipo de relación entre las variables exteriores (en nuestro caso, nivel, temperatura y tiempo) y la respuesta del sistema (desplazamiento radial). Esta característica contrasta con el método HST, en el que debe fijarse por ejemplo el orden del polinomio de la variable nivel de embalse, así como el tipo de función dependiente del tiempo.

Los BA aproximan el valor de la variable objetivo a partir del promedio de la predicción de un gran número (generalmente varios centenares) de modelos sencillos del tipo árbol de decisión. Son por tanto un *modelo de conjunto*. La descripción de la base teórica puede encontrarse en diversas fuentes (p.e. Hastie *et al.*, 2009), así como en el artículo seminal de Breiman (2001). Se han publicado ejemplos de aplicación a seguridad de presas (Salazar *et al.*, 2013).

Aunque su interpretación no es fácil, existen herramientas que permiten extraer información útil sobre el funcionamiento del sistema. La herramienta principal de interpretación de los modelos basados en BA es el índice de importancia de las variables.

Una vez construido el modelo, se permuta por separado cada una de las variables predictoras y se calcula el incremento en el error de predicción que supone. Se basa en la idea de que si una variable no afecta a la respuesta, la precisión dependerá poco de si se utiliza la serie original o la permutada. Al revés, el error aumenta más cuanto mayor es la importancia de la variable permutada. Esta medida se puede utilizar para seleccionar variables, si bien se ha demostrado que tiene sesgo en determinadas circunstancias (Strobl *et al*, 2008).

También puede extraerse información del modelo mediante los gráficos de dependencia parcial (Friedman 2000). En estos gráficos, para cada variable predictora, se selecciona un conjunto de valores distribuidos uniformemente a lo largo de su rango. Para cada uno de esos valores, se calcula la media de la predicción del modelo considerando los valores reales del resto de predictoras:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n f(x, x_{iC}) \quad [3]$$

donde x es la variable cuyo efecto se quiere evaluar, y x_{iC} representan el resto de variables. Por ejemplo, si x es el nivel de embalse, se define un conjunto de valores equiespaciados a lo largo de la carrera de embalse x_1, \dots, x_p . Se sustituye la serie original del nivel de embalse por un valor constante $x=x_j$. Se calcula la predicción del modelo manteniendo el resto de variables con su valor original (x_{iC}), y se extrae la media de esas predicciones. Se repite el proceso para cada uno de los valores de $x=x_p$. Con ello se obtiene una serie de puntos que reflejan el efecto promedio de la variación del nivel de embalse en la predicción del modelo.

Los BA son la base de los denominados “quantile regression forests” (Meinshausen 2006), que permiten calcular la función de densidad de la variable objetivo, a partir de la cual pueden definirse intervalos de confianza de la predicción, que pueden ser útiles para definir umbrales de emergencia.

3. Resultados y discusión

3.1. Interpretación del comportamiento

La figura 3 contiene la contribución parcial al desplazamiento radial P1DR1 del tiempo, la temperatura y la carga hidrostática según el modelo HST. Se observa la relación no-lineal con el nivel de embalse, con desplazamiento hacia aguas abajo mayor con embalse lleno (sentido decreciente del eje y). Lo mismo ocurre en los meses fríos, como se comprueba en el gráfico parcial de la temperatura. Por último, el modelo identifica una deriva temporal hacia aguas abajo sensiblemente lineal de aproximadamente 0,2 mm/año.

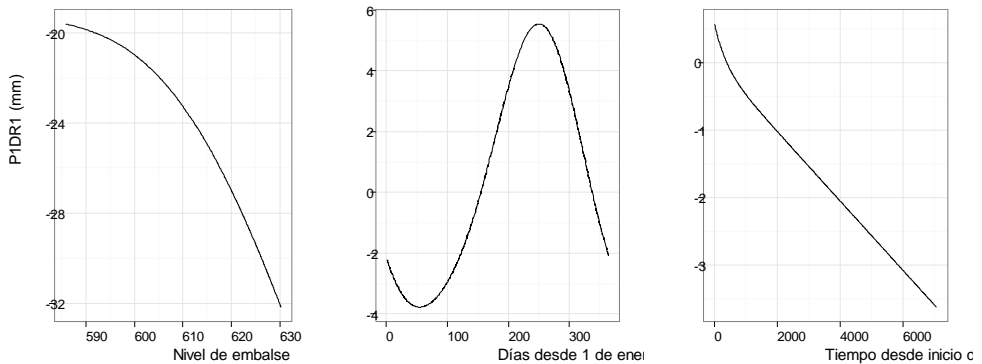


Figura 3. Contribución parcial de las variables exteriores al desplazamiento radial según el modelo HST.

Los gráficos de dependencia parcial que se obtienen del modelo BA se incluyen en la figura 4. Se observan tendencias muy similares en cuanto al efecto de la temperatura y el nivel de embalse. Sin embargo, la influencia del tiempo es cualitativamente diferente. En este caso, se registra una variación brusca sobre el año 1992, y un comportamiento sensiblemente constante entre ese momento y el año 2000.

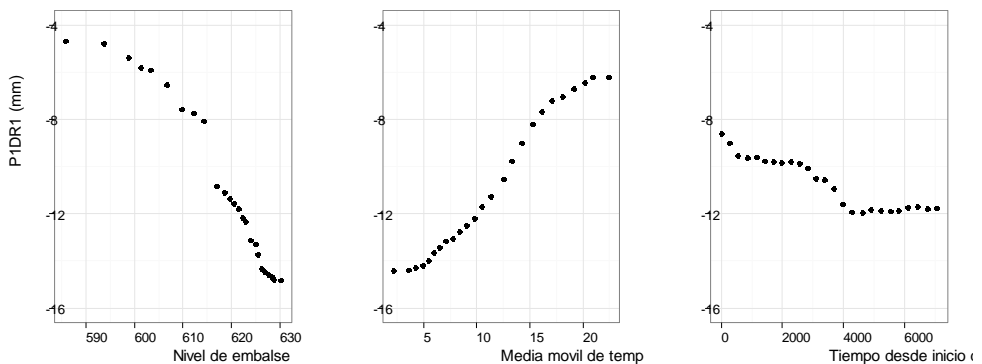


Figura 4. Gráficos de dependencia parcial del desplazamiento radial en P1DR1 (datos registrados) calculados con el modelo BA

La interpretación de estos resultados y su implicación con respecto del comportamiento de la presa son muy importantes y deben hacerse a partir de la máxima información sobre las incidencias en la estructura y en su sistema de auscultación.

Ya se ha mencionado que en el modelo HST debe definirse a priori la forma de la función dependiente del tiempo (en este caso, se ha optado por una combinación de exponencial negativa y lineal), mientras que en el modelo BA puede tener en principio cualquier forma.

Por lo tanto, podría pensarse que el comportamiento real de la presa es el que muestra el modelo BA, y que el resultado del HST se debe a la restricción impuesta a priori. En ese caso, la curva de contribución parcial del tiempo de la figura 3 sería el mejor ajuste del modelo HST al comportamiento real (escalón) que muestra la figura 4.

No obstante, cabe cuestionarse esta conclusión, ya que los gráficos de dependencia parcial no son equivalentes a los de la figura 3. Además, se da la circunstancia de que el nivel de embalse promedio en el periodo 1992-2000 fue sensiblemente mayor que en el periodo 1980-1992 (figura 2). Esta podría ser la causa del escalón registrado en la figura 4, que denota un desplazamiento mayor hacia aguas abajo en el periodo más reciente (coherente con el mayor nivel de embalse).

Para comprobar esta circunstancia, se ha realizado un experimento adicional. Se ha generado una serie de datos modificada del desplazamiento en P1DR1. Se ha obtenido introduciendo los valores reales de tiempo y nivel en [1], con los coeficientes calculados durante el ajuste del modelo HST, eliminando los términos dependientes del tiempo. Se obtiene por tanto una serie artificial, que representa cualitativamente la contribución del nivel de embalse y la temperatura, pero que es independiente del tiempo. Sí mantiene la particularidad de que la carga hidrostática en el periodo más reciente es mayor que en el periodo inicial. Se ha añadido al resultado un término aleatorio de valor medio nulo y desviación típica igual a 0,5 mm.

Se ha ajustado un modelo basado en BA para ajustar la serie artificial, y se incluyen los gráficos de dependencia parcial en las figura 5.

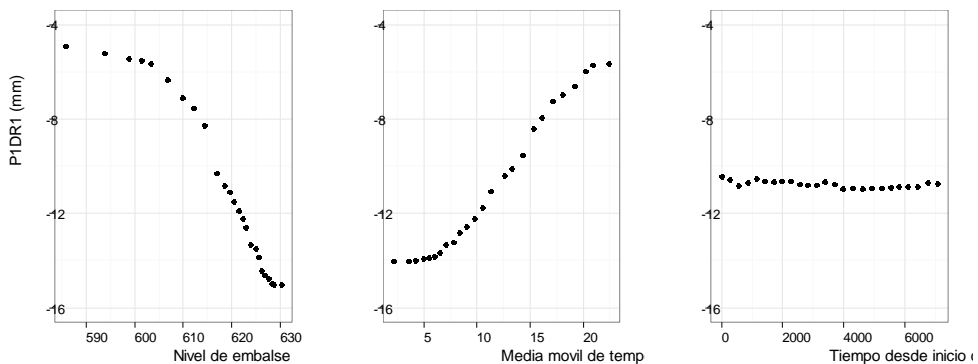


Figura 5. Gráficos de dependencia parcial del desplazamiento radial $P1DR1_{HS}$ (serie generada artificialmente eliminando la influencia del tiempo), calculados con el modelo BA

Se comprueba que la respuesta es prácticamente independiente del tiempo. Estos resultados confirman que el escalón registrado en el gráfico de dependencia parcial del desplazamiento con respecto del tiempo (figura 4) no se debe a que el nivel de embalse

promedio es mayor en el periodo 1992-2000. Ello demuestra la utilidad de los gráficos de dependencia parcial de modelos BA para interpretar el comportamiento de presas.

3.2. Definición de umbrales

La figura 6 muestra la predicción del modelo BA para el año 2000 con los datos reales registrados de nivel y temperatura (línea), junto con el rango del 95% de la función de densidad en cada punto (zona sombreada). Los datos registrados se representan por puntos. Se observa que los valores caen dentro del umbral así definido.

Si bien la declaración de un escenario de emergencia debe corresponder al responsable de seguridad de la presa, este tipo de gráficos pueden ser útiles para tal efecto. En particular, podría definirse una alerta que se activase al registrarse un valor fuera del rango de comportamiento normal del 95%.

Se trata de una prueba preliminar, pero muestra que el rango de confianza de la predicción del modelo varía dependiendo del valor de las variables exteriores.

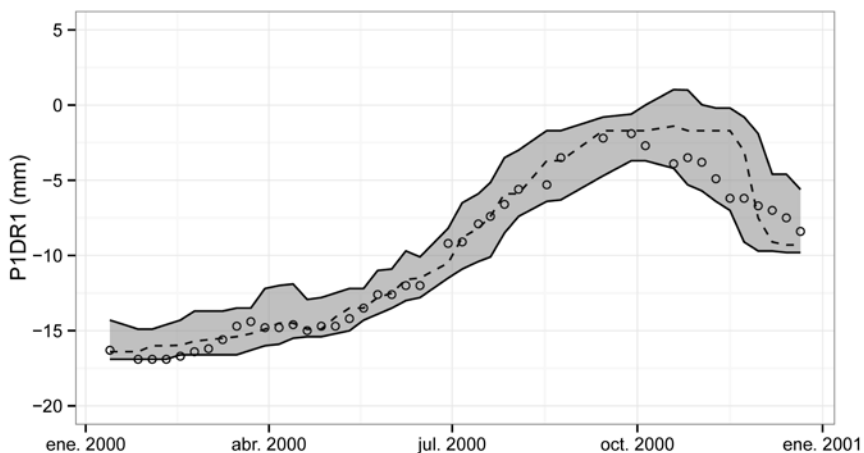


Figura 6. Predicción del modelo BA (línea discontinua) y su rango de confianza del 95% (zona sombreada). Los datos registrados se muestran con círculos.

4. Resumen y conclusiones

Los BA tienen gran flexibilidad para aproximar relaciones no lineales entre variables predictoras y respuesta. Aunque su complejidad dificulta su interpretación, los gráficos de dependencia parcial son una herramienta útil para identificar cambios en el comportamiento del sistema.

Su flexibilidad permite capturar variaciones de comportamiento de cualquier tipo. Esto supone una gran ventaja con respecto del modelo HST, en el que debe definirse a priori el

tipo de influencia entre cada variable predictora y la respuesta. El ejemplo presentado ilustra la ventaja que supone en la práctica. El análisis del modelo HST podría llevar a concluir que existe una deriva de la presa que se traduce en un desplazamiento hacia aguas abajo lineal con el tiempo y que dicha deriva no muestra síntomas de atenuación. Por el contrario, el modelo BA muestra que la variación temporal se produjo bruscamente, y que se ha estabilizado en los últimos años.

Otra conclusión importante es la posibilidad de asociar los umbrales de emergencia con funciones de densidad dependientes de la incertidumbre de las variables predictoras. Sobre esta línea se mantiene una investigación en marcha que pretende establecer criterios para la definición de dichos umbrales mediante funciones de probabilidad asociadas al modelo de comportamiento de la presa.

Agradecimientos

El trabajo ha sido financiado por el Ministerio de Economía y Competitividad de España (MINECO), a través de los proyectos iComplex (IPT-2012-0813-390000) y AIDA (BIA2013-49018-C2-1-R y BIA2013-49018-C2-2-R).

Los autores desean expresar su agradecimiento a Carlos Barbero, de la Agencia Catalana del Agua, por facilitar los datos de auscultación de la presa de La Baells.

Referencias

Breiman, L. (2001). Random forests, *Machine learning* 45(1) (2001) 05-32.

Friedman, J.H. (2000). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 29, 1189-1232.

Hastie, T., Tibshirani, R. y Friedman, J. 2009. The elements of statistical learning - Data mining, Inference and Prediction. Springer, 2ª edición.

Olden, J. D., Jackson, D. A. (2002). Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. *Ecol. Model.* 154(1), 135-150.

Mata, J. (2011). Interpretation of concrete dam behaviour with artificial neural network and multiple linear regression models. *Engineering Structures*, 33(3), 903-910.

Meinshausen, N. (2006). Quantile regression forests. *The Journal of Machine Learning Research*, 7, 983-999.

Ministerio de Medio Ambiente. 2001. Guía Técnica para Elaboración de Planes de Emergencia de Presas.

Salazar, F., Oñate, E. y Toledo, M.Á. (2013). Posibilidades de la inteligencia artificial en el análisis de auscultación de presas. III Jornadas de Ingeniería del Agua.

Salazar, F., Toledo, M. A., Oñate, E., & Morán, R. (2015a). An empirical comparison of machine learning techniques for dam behaviour modelling. *Structural Safety*, 56, 9-17.

Salazar, F., Morán, R., Toledo, M.Á., Oñate, E. (2015b). Data-based models for the prediction of dam behaviour. A review and some methodological considerations. *Archives of Computational Methods in Engineering*. Doi: 10.1007/s11831-015-9157-9.

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 de 2014, pp. 81-96.

Strobl, C., Boulesteix, A.L., Kneib, T., Augustion, T, Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics* 2008, 9:307

Tatin, M., Briffaut, M., Dufour, F., Simon, A., Fabre, J. (2013). Thermal displacements of concrete dams: Finite element and statistical modelling, 9th ICOLD European Club Symposium, Venice, Italy, 2013.

Toledo, M.Á., Salazar, F., Morera, L., Roa, A., Santillán, D. y Morán, R. 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. Madrid, junio de 2013. [http://oa.upm.es/30175/1/INVE MEM 2013 150013.pdf](http://oa.upm.es/30175/1/INVE_MEM_2013_150013.pdf)