A methodology for dam safety evaluation and anomaly detection based on boosted regression trees

Fernando SALAZAR[1], José M. GONZÁLEZ[1], Miguel Á. TOLEDO[2], Eugenio OÑATE[1]

[1] Centre Internacional de Mètodes Numèrics en Enginyeria (CIMNE), UPC. Campus Nord UPC.
Gran Capitán s/n, Barcelona 08034, SPAIN fsalazar@cimne.upc.edu,
josem.gonzalez@cimne.upc.edu, onate@cimne.upc.edu
[2] Civil Engineering Department: Hydraulics,
Energy and Environment, Technical University of Madrid (UPM)
Profesor Aranguren s/n, Madrid 28040, SPAIN matoledo@caminos.upm.es

Key words: Statistical analysis, Dam safety, Machine learning, Anomaly detection.

Abstract

Many countries are implementing new dam safety regulations that often include more restrictive standards. This, together with the increasing average age of dams, results in a greater need for dam control and maintenance works. The advances in information and communications technologies improved the performance of dam monitoring systems, so a large amount of information on the dam behaviour can be collected. This has led to the use of more powerful tools for its analysis, many of which were first developed in the field of machine learning (e. g. neural networks). They offer some advantages over the conventional statistical methods. However, their capacity for early detection of anomalies has seldom been studied. As a result, they are far from being fully accepted by practitioners, whose analyses are often restricted to the interpretation of simple plots of time series data, together with basic statistical models. The present work describes a methodology for anomaly detection in dam behaviour, with the following features: a) The prediction model is based on boosted regression trees (BRTs). b) Causal and auto-regressive models are combined to detect different types of anomalies. c) It is checked whether the values of the external variables fall within the range of the training data. The performance of the proposed methodology was assessed through its application to a test case corresponding to an actual 100-m height arch dam, in operation since 1980. Artificial data were generated by means of a finite element model. Different anomalies were later added in order to test the anomaly detection capability. The method can be applied to other response variables and dam typologies, due to the great flexibility of BRTs, which automatically select the most relevant inputs.

1 INTRODUCTION

The advances in information and communication technologies have improved the performance of dam monitoring systems in terms of accuracy, reliability and reading frequency. This results in more comprehensive information about the behaviour of the structure [1].

The increase in the amount of information available led to the application of advanced tools for data analysis, most of which provide from the machine learning community, e.g. neural networks [2], support vector machines [3], the adaptive neuro-fuzzy inference systems

(ANFIS) [4], among others [5], [6].

Nonetheless, these tools have not been introduced among practitioners, who typically employ graphical data exploration [7], together with simple statistical models [1].

The vast majority of the examples of application of advanced tools focus on the development of a behaviour model that predicts the value of a given response variable (e. g. radial displacement). The prediction is compared with the actually observed data and some error index is extracted. In many cases the results are compared with those obtained by conventional statistical methods (e.g. [2]).

These advanced tools offer some advantages in terms of greater accuracy, flexibility, or ability to interpret the dam behaviour [8]. However, an accurate predictive model is just one of the necessary ingredients of an anomaly detection system. Some criterion needs to be developed to determine whether a given discrepancy between prediction and observation shall be considered as anomalous. This aspect was seldom considered, with a few exceptions for particular cases [9], [10], [11], which nonetheless were evaluated over a short validation period. Jung et al [10] and Mata et al [9] also generated artificial data with numerical models representative of abnormal situations.

In these work, a similar procedure is employed: a predictive model is built, and prediction intervals are derived from the standard deviation of the prediction error. This is the main ingredient of a methodology for anomaly detection in dam behaviour, with the following innovative features:

- The predictive model is based on boosted regression tress (BRTs from now on). This tool offered higher accuracy than other conventional and advanced tools in previous works [6].
- Both causal and auto-regressive models are assessed and the correspondent efficiency is compared in terms of anomaly detection capability.
- Artificial data are taken as reference, obtained with a numerical finite element (FE) model. It represents an actual dam currently in operation. The numerical results were compared to monitoring data to verify that they represent the actual dam behaviour.
- The value of the external variables was compared to the range of training data: during model application, abnormal values correspondent to extraordinary loads (e.g. reservoir level) are considered as due to lack of training data. This contributes to reduce the amount of false positives.

2 METHODS

2.1 Boosted regression trees

BRT models are built by combining two algorithms: a set of models are fitted by means of simple decision trees [12], whose output is combined to compute the overall prediction using boosting [13]. This tool was employed in previous works [8], where its main features were described. A more comprehensive introduction can be found in [14].

The main properties of BRTs are:

- They are robust against outliers.
- They require little data pre-processing.
- They can handle numerical and categorical predictors.
- They are appropriate to model non-lineal relations, as well as interaction among predictors.

All the calculations were performed in the R environment [15].

2.2. Prediction intervals

As mentioned above, a method for anomaly detection requires determining which magnitude of the discrepancy between prediction and observation is considered abnormal. In this work, the density function of error was computed and the normal interval defined as

$$[\hat{y}+\bar{e}+2\cdot sd_e,\ \hat{y}+\bar{e}-2\cdot sd_e] \tag{1}$$

where $\hat{y}$ is the model prediction, $\bar{e}$ is the mean error and sd_e is the error standard deviation. If the error density function follows a normal distribution, this margin contains 95% of the normal values. This criterion is heuristic, and was determined after some preliminary tests, considering both the true anomalies detected (true positives) and the normal situations considered abnormal (false positives).

However, a more relevant issue is the proper computation of the model prediction error. Since BRTs are non-parametric and typically feature a large number of parameters, they are susceptible to over fit the training data. It is well known that the training prediction error for machine learning tools results in an optimistic estimate of the actual model generalisation capability.

Cross validation is a conventional method to overcome this drawback. However, it cannot be directly applied to dam monitoring data, since they are time series: the training period shall be precedent to the validation interval. Moreover, the dam and its foundation feature behaviour changes over time in the general case.

We propose a method based on the hold-out cross validation described by Arlot and Celisse [16] for non-stationary time-series data. It comprises an iterative procedure:

- Take a minimum training period of 5 years. Build a BRT model, and compute the prediction errors for the sixth year.
- Build a new BRT model with 5+1 years of training data. Compute prediction errors for 7th year, and aggregate them to those obtained in the previous step.
- Repeat step 2 until a model is built with all the available data but the most recent year, and aggregate the prediction errors.
- Compute the density function of the aggregated-error and its statistics ($\bar{e}$ and sd_e). They are employed to define the interval for normal behaviour with Eq. (1).
- Build a BRT model with all the available data. Generate predictions and normal intervals for new (validation) data and compare to observations for anomaly detection.

2.3. Training range verification

Machine learning models typically produce highly inaccurate results when extrapolating, i.e., when new data falls outside the range of the training data set. In the case of dam monitoring, this situation corresponds with external loads above (or below) the maximum (minimum) value registered during its service life. Since dam response depends on several actions, also a combination of them may result in an “out-of-range” situation, even if none of the values are out of range when considered separately. To account for this issue, we chose the two principal environmental loads, i.e., hydrostatic (reservoir level) and thermal (air temperature), and build a two-dimensional density function via kernel density estimation [17]. The training sample with the lower density was computed, and its correspondent iso-line plotted. New inputs falling outside this line are classified as out-of-range. In practice, they are not considered as anomalies even though the deviation between prediction and observation fell outside the normal interval. Figure 1 is an example of the density plots generated, with the training data, the iso-line and the validation data, part of which are out of range according to the described criterion.

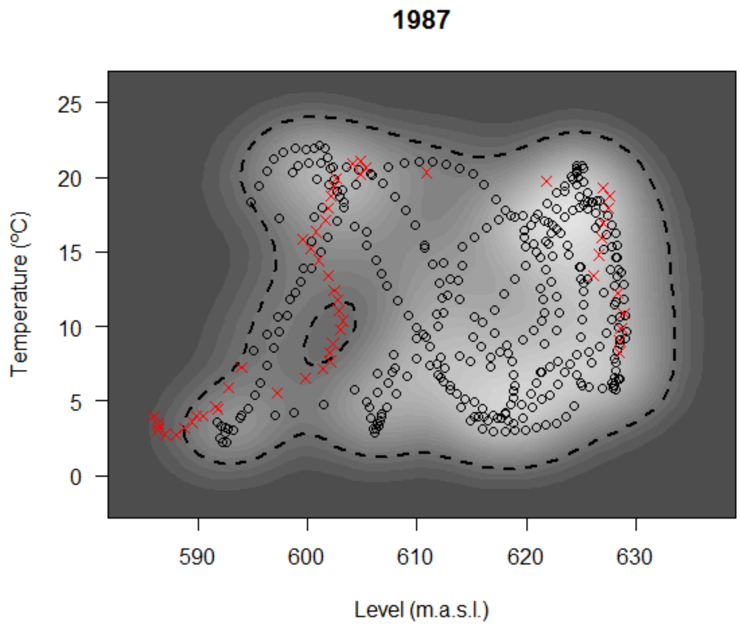

Figure 1: Example of density function for the main environmental loads. The circles represent the training set, and the red crosses the validation set. The dotted line is the iso-line with the lower density for the training set. It should be noted that some of the validation data are considered as out of range because they fall in low-density areas, even though they do not correspond to extraordinary high (or low) hydrostatic load (or air temperature).

2.4 Case study

La Baells dam is a double-curvature arch dam located in the Llobregat river, in the Barcelona region (Spain). The crest length is 403 m, whereas the maximum height above foundation is 102 m. Monitoring data were provided by the Catalan Water Agency for the period 1981-2008. These data correspond both to environmental and response variables. In this work, the air temperature and the reservoir level time series were considered as inputs to a finite element (FE) model. The results of this model in terms of radial displacements at the location of the pendulums were extracted and compared to the actual measurements. The objective was to check that the FE model could provide realistic data to generate reference time series of dam behaviour. This procedure allows obtaining data corresponding to the "ground truth", i.e., input-output time series without anomalies.

Time series data for hydrostatic load and air temperature were available for the period 1980-2008. Some derived variables were computed and considered as inputs, as described in previous works [5], [6] and summarised in Table 1. These are the inputs for the causal models.

The radial displacements measured at eight locations within the dam body were considered as outputs (Figure 2).

Moreover, a non-causal model was built for each output, taking as inputs both the environmental variables and all the outputs except that to be predicted in each case. Also the lagged output variables were included as inputs. This is, to predict the radial displacement $R(t_k,x_i)$ (at time t_k, location x_i), the following variables were added to those considered for the causal model:

- R($t_{k\text{-}1}$, x_i)
- R($t_{k\text{-}2}$, x_i)
- R(t_k, x_j; $j \neq i$)
- R($t_{k\text{-}1}$, x_j; $j \neq i$)
- R($t_{k\text{-}2}$, x_j; $j \neq i$)

Hence, both a causal and a non-causal model are built for each output variable.

Id	Type	Period (days)
Level	Hydrostatic load-original	-
Lev007	Hydrostatic load-moving average	7
Lev014	Hydrostatic load-moving average	14
Lev030	Hydrostatic load-moving average	30
Lev060	Hydrostatic load-moving average	60
Lev090	Hydrostatic load-moving average	90
Lev180	Hydrostatic load-moving average	180
Tair	Air Temperature-original	-
Tair007	Air Temperature-moving average	7
Tair014	Air Temperature-moving average	14
Tair030	Air Temperature-moving average	30
Tair060	Air Temperature-moving average	60
Tair090	Air Temperature-moving average	90
Tair180	Air Temperature-moving average	180
NDay	Time	-
Month	Season	-

Table 1: External variables considered

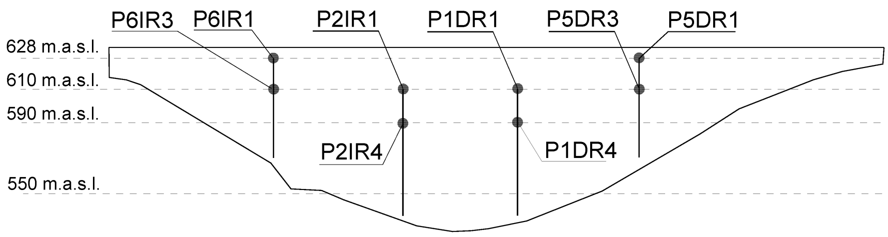

Figure 2: Location of the devices analysed within the dam body (view from downstream)

2.5 Anomalies

The results of the numerical model were artificially modified to simulate anomalies of two types and variable magnitude, which were introduced at a certain date:

- Offset, equivalent to adding a constant value (0.5, 1.0 or 1.5 mm) to the FEM model result.
- Incremental drift, where the added value is variable, growing linearly along time (0.5, 1.0 or 1.5 mm/year).

The abnormal period started at some random date between January 1st 1986 (5 years of minimum training period) and 2007 September 10th (one year before the end of the available period).

These criteria were applied to generate 1,000 abnormal time series, each one with random values of a) the target variable (among those depicted in Figure 2), b) the type and magnitude of anomaly, and c) the initial date of the abnormal period.

Each test case was presented both to the causal and the non-causal models. The time lapse since the initiation of the abnormal behaviour to the first observation identified as anomaly by each model was registered as “detection time”. The test period was limited to two years of abnormal behaviour. If the correspondent model did not identify any observation as abnormal during that period, the detection time was set to 730 days (2 years) as regards results analysis.

Also the amount of false positives (regular values considered anomalous) were computed for each model and output. It should be reminded that the out-of-range instances (according to the criterion described in section 2.3) were not considered as anomalies.

3. RESULTS AND DISCUSSION

Figure 3 shows an example of application of the methodology. The vertical line was plotted at the initial date of anomaly. The shaded area represents the normal range (Eq. 1), and the anomalous values are depicted with red asterisks. The blue hollow circles represent out-of-range input data.

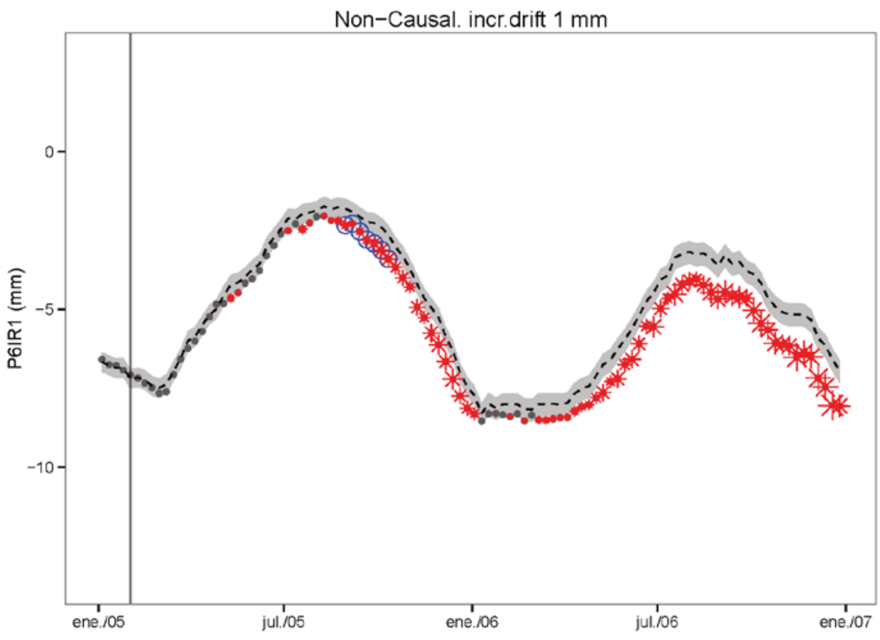

Figure 3: Anomaly detection. Example of application. Blue hollow circles depict out-of-range values. Red asterisks show observations out of the normal area (shaded). The vertical line is located at the starting of the abnormal interval.

An increasing deviation between observations and predictions can be clearly identified, correspondent to the artificial anomaly introduced. The model is re-trained at the beginning of each year, and the observations in the precedent year are added to the training data set. In this case, the model considers them as normal response, and partially adapts to the abnormal behaviour. This is the reason why the observations in January 2006 fall within the normal interval. However, the anomalous behaviour is again detected in February 2006. In practice, the re-training at January 2006 should be modified accounting for the abnormal behaviour previously identified.

Table 2 shows some statistics of the performance of both models across the 2,000 cases analysed. The non-causal model outperforms the causal one for all the indicators considered.

Model	Undetected anomalies	Detection time. Mean/median (days)	False positives (average per year)
Causal	20%	276/186	0.85
Non causal	10%	163/72	0.56

Table 2: Results of the performance of causal and non-causal models

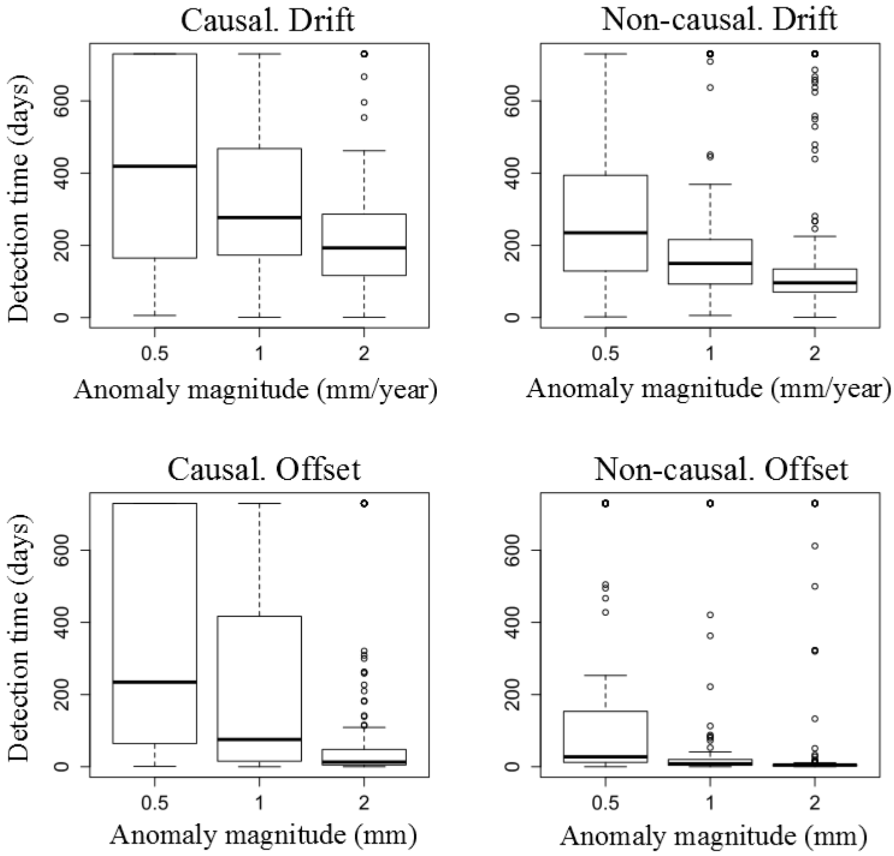

Figure 4: Detection time per type of anomaly, predictive model and magnitude. The non-causal model performs better, especially for the “offset” type.

The detection time per type of model and anomaly is depicted in Figure 3. It is lower for the non-causal model in all cases. As expected, the anomalies with lower magnitude are harder to detect, which results in a higher detection time (it should be reminded that the detection time is limited to 730 days). Also, the “incremental drift” anomalies require a longer time to be detected than the “offset”.

Figure 4 shows the results per type of model and anomaly, as well as per year of initiation of the abnormal behaviour. The latter factor determines the size of the training set, which in turn affects the model accuracy and its ability to detect anomalies.

For the “offset” anomaly, the performance of both models neatly improves for later date of anomaly initiation. This effect is remarkable for the non-causal model, whose median detection time is close to zero for anomalies starting after 1995. The tendency is less clear for the “incremental drift” anomaly.

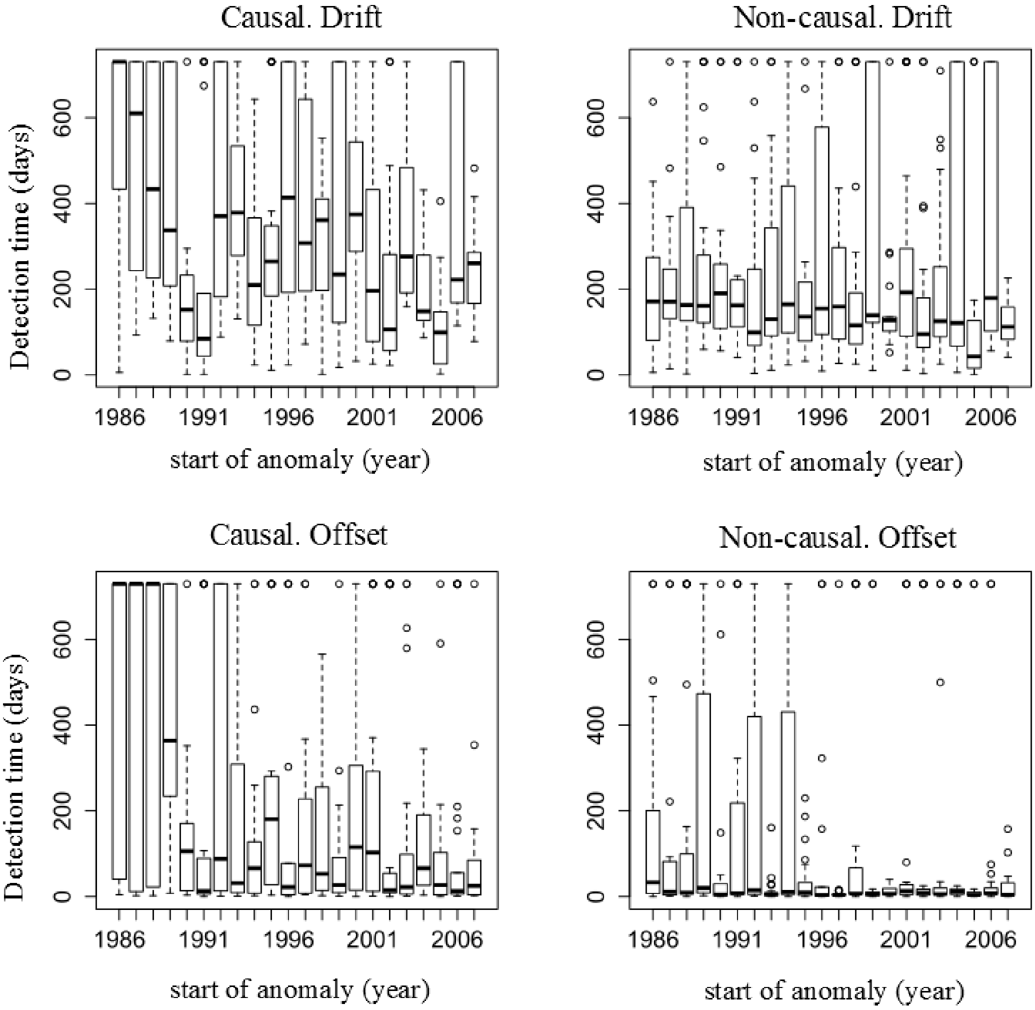

Figure 5: Detection time per year of anomaly initiation, type of model and anomaly. The performance is better along time for both models and “offset” anomalies, whereas the tendency is less clear for “incremental drift”.

9 SUMMARY, CONCLUSIONS AND FUTURE WORK

A methodology for anomaly detection in dam behaviour based on BRTs was presented. It is based on a criterion for defining a range of normal behaviour, based in turn on the model prediction and the statistics of the training error.

The occurrence of extraordinary loads is accounted for by computing a density function of the most relevant input variables (hydrostatic load and air temperature) via kernel density estimation. This reduces the amount of false positives due to lack of model accuracy for extrapolation.

Causal and non-causal models were compared, as regards their capability for detecting anomalies in radial displacements of an arch dam. Artificial anomalies were generated by adding certain values to the dam response, as computed by means of a FE model. The non-causal model showed better performance: fewer false positives, more anomalies detected and lower detection time. This is due to its higher accuracy, which results in narrower intervals for normal behaviour.

However, non-causal models rely on response variables to predict each output, i.e., the predictions for the radial displacement at a given location and time is based on other displacements, as well as on the previous values of the variable to predict. This may lead to poor performance when the dam undergoes abnormal behaviour affecting several devices. We are currently working on this issue, by means of applying the same methodology to more realistic abnormal data: the boundary conditions of the numerical model are modified to reproduce hypothetical, though feasible, anomalies, which are reflected in several output variables.

In any case, these techniques should be used as a tool to provide detailed and accurate information to the dam safety managers, rather than as a totally automatic detection system. All relevant decisions influencing dam safety shall be made by an expert and capable engineer, based on the analysis of all the relevant information available. In this context, plots as that depicted in Figure 3 can be highly valuable: they allow visually identifying the occurrence of a deviation from normal behaviour (increasing along time in the example presented). Thus, the automatic system can be used as an indicator to generate a warning which leads to intensify the dam safety monitoring.

ACKNOWLEDGEMENTS

The authors thank Carlos Barbero, dam safety manager at the Catalan Water Agency (ACA), for providing the information regarding La Baells dam.

The research was supported by the Spanish Ministry of Economy and Competitiveness (Ministerio de Economía y Competitividad, MINECO) through the projects iComplex (IPT-2012-0813-390000) and AIDA (BIA2013-49018-C2-1-R and BIA2013- 49018-C2-2-R).

REFERENCES

[1] International Commission on Large Dams (2012). Dam surveillance guide. Bulletin 158, ICOLD

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

[3] Ranković, V., Grujović, N., Divac, D., & Milivojević, N. (2014). Development of support vector regression identification model for prediction of dam structural behaviour. *Structural Safety*, *48*, 33-39.

[4] Ranković, V., Grujović, N., Divac, D., Milivojević, N., & Novaković, A. (2012). Modelling of dam behaviour based on neuro-fuzzy identification. *Engineering Structures*, *35*, 107-113.

[5] Salazar, F., Morán, R., Toledo, M. Á., & Oñate, E. (2015a). Data-Based Models for the Prediction of Dam Behaviour: A Review and Some Methodological Considerations. *Archives of Computational Methods in Engineering*, 1-21.

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

[7] S. D. Myers, B.K., Providing improved dam safety monitoring using existing staff resources: Fern Ridge Dam case study. Proceedings of 28th Annual USSD Conference, 2008.

[8] Salazar, F., Toledo, M. Á., Oñate, E. & Suárez, B. (2016). Interpretation of dam deformation and leakage with boosted regression trees. *Engineering Structures* doi: 10.1016/j.engstruct.2016.04.012

[9] Mata, J., Leitão, N. S., de Castro, A. T., & da Costa, J. S. (2014). Construction of decision rules for early detection of a developing concrete arch dam failure scenario. A discriminant approach. *Computers & Structures*, *142*, 45-53.

[10] Jung, I. S., Berges, M., Garrett, J. H., & Poczos, B. (2015). Exploration and evaluation of AR, MPCA and KL anomaly detection techniques to embankment dam piezometer data. *Advanced Engineering Informatics*, *29*(4), 902-917.

[11] Cheng, L., & Zheng, D. (2013). Two online dam safety monitoring models based on the process of extracting environmental effect. *Advances in Engineering Software*, *57*, 48-56.

[12] Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. CRC press.

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

[14] Leathwick, J. R., Elith, J., Francis, M. P., Hastie, T., & Taylor, P. (2006). Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology Progress Series*, *321*, 267-281.

[15] Team, R. C. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013.

[16] Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics surveys*, *4*, 40-79.

[17] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, *27*(3), 832-837.