Engineering Structures 119 (2016) 230–251

Contents lists available at ScienceDirect

Engineering Structures

journal homepage: www.elsevier.com/locate/engstruct

Interpretation of dam deformation and leakage with boosted regression trees

Fernando Salazar [a,*], Miguel Á. Toledo [b], Eugenio Oñate [a], Benjamín Suárez [a]

[a] International Center for Numerical Methods in Engineering (CIMNE), Campus Norte UPC, Gran Capitán s/n, 08034 Barcelona, Spain
[b] Technical University of Madrid (UPM), Civil Engineering Department: Hydraulics, Energy and Environment, Profesor Aranguren s/n, 28040 Madrid, Spain

ARTICLE INFO

Article history:
Received 28 October 2015
Revised 15 February 2016
Accepted 6 April 2016
Available online 26 April 2016

Keywords:
Machine learning
Dam safety
Dam monitoring
Boosted regression trees

ABSTRACT

Predictive models are essential in dam safety assessment. They have been traditionally based on simple statistical tools such as the hydrostatic-season-time (HST) model. These tools are well known to have limitations in terms of accuracy and reliability. In the recent years, the examples of application of machine learning and related techniques are becoming more frequent as an alternative to HST. While they proved to feature higher flexibility and prediction accuracy, they are also more difficult to interpret. As a consequence, the vast majority of the research is limited to prediction accuracy estimation. In this work, one of the most popular machine learning techniques (boosted regression trees), was applied to model 8 radial displacements and 4 leakage flows at La Baells Dam. The possibilities of model interpretation were explored: the relative influence of each predictor was computed, and the partial dependence plots were obtained. Both results were analysed to draw conclusions on dam response to environmental variables, and its evolution over time. The results show that this technique can efficiently identify dam performance changes with higher flexibility and reliability than simple regression models.

1. Introduction

Dam monitoring is essential to ensure its proper operation and its long-term safety [1]. One of the main tasks to be carried out is the comparison between the expected response and that registered by the monitoring system, to understand the dam behaviour and to detect potential anomalies. In this context, predictive models are necessary to estimate the dam response in a given situation.

Data-based tools allow building predictive models based on monitoring data, i.e., without explicitly considering the physical properties of the dam and the foundation. The hydrostatic-season-time (HST) model [2] is the most widely applied, and the only generally accepted by practitioners.

HST is based on multiple linear regression considering the three most influential external variables: hydrostatic load, air temperature and time. The main advantages of HST are:

1. It frequently provides useful estimations of displacements in concrete dams [3].
2. It is simple and thus easily interpretable: the effect of each external variable can be isolated in a straightforward manner, since they are cumulative.
3. Since the thermal effect is considered as a periodic function, the time series of air temperature are not required. This widens the possibilities of application, as only the reservoir level variation is needs to be available to build an HST model.
4. It is well known by practitioners and frequently applied in several countries [3].

Nonetheless, HST also features conceptual limitations that damage the prediction accuracy [3] and may lead to misinterpretation of the results [4]. For example, it is based on the assumption that the hydrostatic load and the temperature are independent, whereas it is obviously not the case: the thermal field in the dam body, especially in the vicinity of the water surface, is strongly dependant on the water temperature in the upstream face [5]. In turn, the thermal load influences the stress and displacement fields.

Several modifications to the original HST model have been proposed to overcome these drawbacks. They focus on improving the consideration of the thermal load, by taking into account the actual air temperature instead of the historical mean [6], or the effect of the water temperature on the upstream face [3,7].

* Corresponding author.
E-mail addresses: fsalazar@cimne.upc.edu (F. Salazar), matoledo@caminos.upm.es (M.Á. Toledo), onate@cimne.upc.edu (E. Oñate), benjamin.suarez@upc.edu (B. Suárez).

http://dx.doi.org/10.1016/j.engstruct.2016.04.012
0141-0296/

In the recent years, non-parametric techniques have emerged as an alternative to HST for building data-based behaviour models [8], e.g. support vector machines (SVN) [9], neural networks (NN) [10], adaptive neuro-fuzzy systems (ANFIS) [11], among others [8]. In general, these tools are more suitable to model non-linear cause-effect relations, as well as interaction among external variables, as that previously mentioned between hydrostatic load and temperature. On the contrary, they are typically more difficult to interpret, what led them to be termed as "black box" models (e.g. [12]).

Most of the published works focused on building predictive models whose accuracy was generally higher than that offered by HST (e.g. [10,13,14]). Since the resulting model was seldom analysed, little information was provided for dam safety assessment. Some exceptions worth mentioning, though simple, were due to Santillán et al. [15], Mata [10] and Cheng and Zheng [16].

Therefore, dam engineers face a dilemma: the HST model is widely known and used and easily interpretable. However, it is based on some incorrect assumptions, and its accuracy can be increased. On the other hand, more flexible and accurate models are available, but they are more difficult to implement and analyse. The same problem arose in the field of statistics [17].

The objective of this work is to investigate the possibilities of interpretation of one of these black box models to:

1. Identify the effect of each external variable on the dam behaviour.
2. Detect the temporal evolution of the dam response.
3. Provide meaningful information to draw conclusions about dam safety.

Among the plethora of machine learning techniques available [18], a previous comparative study [13] showed boosted regression trees (BRT) as one of the more appropriate tools for the prediction of dam response. In this paper, the technique was further explored, with focus on the interpretation of the results for dam behaviour identification. In particular, the partial dependence plots were examined to isolate the effect of each action, and the relative influence (RI) was computed to identify the strength of each input–output relation. Furthermore, the results were interpreted from an overall viewpoint to draw conclusions on the dam behaviour.

The method was applied to the analysis of La Baells Dam, as compared to the conventional HST model.

Table 1
Predictor variables considered for the initial BRT model (M1).

Code	Group	Type	Period (days)
Level	Hydrostatic load	Original	–
Lev007	Hydrostatic load	Moving average	7
Lev014			14
Lev030			30
Lev060			60
Lev090			90
Lev180			180
Tair	Air temperature	Moving average	1
Tair007			7
Tair014			14
Tair030			30
Tair060			60
Tair090			90
Tair180			180
Rain	Rainfall	Accumulated	1
Rain030			30
Rain060			60
Rain090			90
Rain180			180
NDay	Time	Original	–
Year			–
Month	Season	Original	–
n010	Hydrostatic load	Rate of variation	10
n020			20
n030			30

The rest of the paper is organised as follows. A brief introduction to BRT is presented, including the methods for interpretation. Then, the case study and the HST version taken as reference are described. The results are included and interpreted in terms of the dam behaviour, and the differences between both methods are discussed.

2. Methods

2.1. Boosted regression trees

The objective of a predictive model is to estimate the value of an output variable $Y \in \mathbb{R}$ (i.e. radial displacement or leakage), based on a set of predictors (reservoir level, air temperature, etc.)

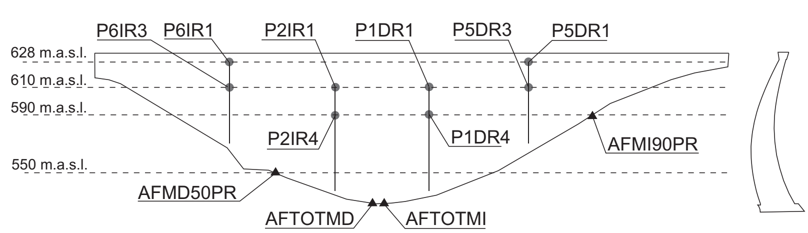

Fig. 1. Geometry and location of the monitoring devices in La Baells Dam. Left: view from downstream. Right: highest cross-section.

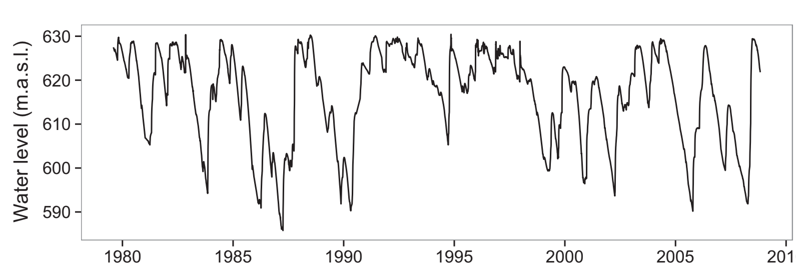

Fig. 2. Time series of the reservoir level at La Baells Dam.