CrossMark

ORIGINAL PAPER

Data-Based Models for the Prediction of Dam Behaviour: A Review and Some Methodological Considerations

Fernando Salazar[1] · Rafael Morán[1,2] · Miguel Á. Toledo[2] · Eugenio Oñate[1]

Abstract Predictive models are an important element in dam safety analysis. They provide an estimate of the dam response faced with a given load combination, which can be compared with the actual measurements to draw conclusions about dam safety. In addition to numerical finite element models, statistical models based on monitoring data have been used for decades for this purpose. In particular, the hydrostatic-season-time method is fully implemented in engineering practice, although some limitations have been pointed out. In other fields of science, powerful tools such as neural networks and support vector machines have been developed, which make use of observed data for interpreting complex systems. This paper contains a review of statistical and machine-learning data-based predictive models, which have been applied to dam safety analysis. Some aspects to take into account when developing analysis of this kind, such as the selection of the input variables, its division into training and validation sets, and the error analysis, are discussed. Most of the papers reviewed deal with one specific output variable of a given dam typology and the majority also lack enough validation data. As a consequence, although results are promising, there is a need for further validation and assessment of generalisation capability. Future research should also focus on the development of criteria for data pre-processing and model application.

1 Introduction

Behaviour models are a fundamental component of dam safety systems, both for the daily operation and for long-term behaviour evaluation. They are built to calculate the dam response under safe conditions for a given load combination, which is compared to actual measurements of dam performance [71]. The result is an essential ingredient for dam safety assessment, together with visual inspection and engineering judgement [27].

Numerical models based on the finite element method (FEM) are widely used to predict dam response, in terms of displacements, strains and stresses. They are based on the physical laws governing the involved phenomena, which gives them some interesting features: (a) they are useful for the design and, more importantly, for dam safety assessment during the first filling, and (b) they can be conveniently interpreted, provided that their parameters have physical meaning.

On the contrary, some relevant indicators of dam safety, such as uplift pressure and leakage flow in concrete dams, cannot be predicted accurately enough with numerical models [38, 39]. In addition, the knowledge of the stress-strain properties of the dam and foundation materials is always limited [75], and so is the prediction accuracy of FEM models [27].

✉ Fernando Salazar
fsalazar@cimne.upc.edu

Rafael Morán
rmoran@caminos.upm.es

Miguel Á. Toledo
matoledo@caminos.upm.es

Eugenio Oñate
onate@cimne.upc.edu

1 CIMNE Centre Internacional de Metodes Numerics en Enginyeria, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain

2 Department of Civil Engineering: Hydraulics, Energy and Environment, Technical University of Madrid (UPM), Profesor Aranguren s/n, 28040 Madrid, Spain

These limitations, together with the availability of monitoring data, have fostered the application of statistical models to predict dam response. They have been used in dam safety analysis for decades as a complement to visual inspection and numerical models, to support decision making.

In recent years, there is a tendency towards automatising dam monitoring devices [27], which allows for increasing the reading frequency and results in a greater amount of data available. Although it encourages extraction of as much information as possible in relation with dam safety conditions [57], it has revealed certain limitations of traditional statistical tools to manage dam monitoring data [58].

On another note, advanced tools have been developed in the machine learning (ML) community to build data-based predictive models. They have been applied in various fields of science and engineering, where similar problems have emerged more dramatically, provided that the amount of data is much larger or the underlying phenomena is much less understood. This is the case, for example, of medicine, e-commerce, smartphone applications, econometrics or business intelligence, among others. Most of these tools exclusively rely on data to build predictive models, i.e., no prior assumptions on the physics of the phenomenon have to be made beforehand [25].

The limitations of traditional statistical tools and the availability of these advanced learning algorithms have motivated dam engineers to search the possibilities of the latter for building dam behaviour models, as well as for analysing dam behaviour.

This paper reports a review on dam behaviour models based on monitoring data. The work focuses on prediction accuracy, although it also refers to model suitability for interpreting dam performance. The most popular techniques are dealt with in Sect. 2, whereas some common issues in building data-based models and evaluating their results are analysed in Sect. 3. The analysis is performed on the basis of the review of 41 papers on the field.

2 Statistical and Machine Learning Techniques Used in Dam Monitoring Analysis

The aim of these models is to predict the value of a given variable $Y \in \mathbb{R}$ (e.g. displacement, leakage flow, crack opening, etc.), in terms of a set of inputs[1] $X \in \mathbb{R}^d$:

$$Y = \hat{Y} + \varepsilon = F(X) + \varepsilon \tag{1}$$

ε is an error term, which encompasses the measurement error, the model error, and the deviation of the dam response from the expected behaviour [71]. This term is important, given that it is frequently used to define safety margins and warning thresholds [27].

The models are fitted on the basis of a set of observed input data x_i, and the correspondent registered outputs y_i, where $i = 1, \ldots, N$ and N is the number of observations. Note that each x_i is a vector of d components, being d the number of inputs.

The inputs may be of different nature, depending on the method:

- Raw data recorded by the monitoring system, which in turn can be:

  - External variables: reservoir level (h), air temperature (T), etc.
  - Internal variables: temperature in the dam body, stresses, displacements, etc.

- Variables derived from observed data. For example:

  - Polynomials
  - Moving averages
  - Derivatives

2.1 Hydrostatic-Seasonal-Time (HST) Model

The most popular data-based approach for dam monitoring analysis is the hydrostatic-seasonal-time (HST) model. It was first proposed by Willm and Beaujoint [76] to predict displacements in concrete dams, and has been widely applied ever since. It is based on the assumption that the dam response is a linear combination of three effects:

$$\hat{Y} = F_1(h) + F_2(s) + F_3(t) \tag{2}$$

- A reversible effect of the hydrostatic load which is commonly considered in the form of a fourth-order polynomial of the reservoir level (h) [4, 67, 71]:

$$F_1(h) = a_0 + a_1 h + a_2 h^2 + a_3 h^3 + a_4 h^4 \tag{3}$$

- A reversible influence of the air temperature, which is assumed to follow an annual cycle. Its effect is approximated by the first terms of the Fourier transform:

---

[1] Traditionally, the statistical models applied in dam monitoring analysis were based on causal variables, e.g. hydrostatic load and temperature, which are often termed "independent variables". On the contrary, other algorithms make use of transformed variables (such as gradients or moving averages), and non-causal observations (e.g. the previous value of the output). This has led to the use of various terms

---

Footnote 1 continued

to refer to the model inputs, such as "predictors", "covariates", and "features". In this paper they are used indistinctly.

$$F_2(s) = a_5 cos(s) + a_6 sen(s) + a_7 sen^2(s) \\ + a_8 sen(s)cos(s) \tag{4}$$

where $s = 2\pi d/365.25$ and d is the number of days since 1 January.

- An irreversible term due to the evolution of the dam response over time. A combination of monotonic time-dependant functions is frequently considered. The original form is [76]:

$$F_3(t) = a_9 log(t) + a_{10} e^t \tag{5}$$

The model parameters $a_1, \ldots, a_{10}$ are adjusted by the least squares method: the final model is based on the values which minimise the sum of the squared deviations between the model predictions and the observations.

Some authors used variations of the original HST model, by using some heuristics or after a trial-and-error process. Mata [40] considered the irreversible effect by means of $F_3(t) = a_9 t + a_{10} e^{-t}$. Chouinard and Roy [12] used a linear term in t and a third-order polynomial of h. Simon et al. [67] chose $F_3(t) = a_9 e^{-t} + a_{10}t + a_{11}t^2 + a_{12}t^3 + a_{13}t^4$, whereas Yu et al. [80] used $F_3(t) = a_9 t + a_{10}t^2 + a_{11}t^3$. Carrère applied a variation of HST in which the possibility of a sudden change in the dam response at a certain time is considered by adding a step function to the irreversible term [9].

The method makes use of strong assumptions on the response of the dam, which might not be fulfilled in general. In particular, the three effects are considered as independent, although it is well known that certain collinearity exists. The reservoir level affects the thermal response of the dam, provided that the air and water temperatures differ [73]. In some cases, the reservoir operation follows an annual cycle due to the evolution of the water demand, so there is a strong correlation between h and the air temperature [13, 33, 38, 66]. Collinearity may lead to poor prediction accuracy and, more importantly, to misinterpretation of the results [1].

Another limitation of the original form of HST model is that the actual air temperature is not considered. On one hand, this makes it more flexible, because it can be applied in dams where air temperature measurements are not available. On the other hand, it reduces its prediction accuracy for particularly warm or cold years [66, 73].

Several alternatives have been proposed to overcome this shortcoming. Penot et al. [50] introduced the HSTT method, in which the thermal periodic effect is corrected according to the actual air temperature. This procedure has been applied at Electricité de France (EDF) [20, 73] with higher accuracy than HST, especially during the 2003 European heat wave. Although the proposal of this method has been frequently attributed to Penot et al.,

Breitenstein et al. [8] applied a similar scheme 20 years earlier.

Tatin et al. [73, 74] proposed further corrections of HSTT. The HST-Grad model takes into account both the mean and the gradient of the temperature in the dam body, considered as a one-dimensional domain. They are estimated from the air temperature in the downstream face, and from a weighted average of the air and water temperatures in the upstream one. A similar and more detailed approach was applied by the same authors, called the SLICE model [73]. It considers different thermal conditions for the portion of the dam body located below the pool level to that situated above, which is not affected by the water temperature.

Other common choice is to replace the periodic function of the thermal component by the actual temperature in the dam body, resulting in the hydrostatic-thermal-time (HTT) method. One difficulty of this approach is how to select the appropriate thermometers among those available. In arch dams, some authors only consider the thermometers in the central cantilever, assuming that it represents the thermal equilibrium between cantilevers in the right and left margins [66]. Mata et al. [42] solved this issue by applying principal component analysis (PCA), while other authors [33] considered all the available instruments. Li et al. [34] proposed an error correction model (ECM), featuring a term which depends on the error in the estimation of previous output values.

Although HST was originally devised for the prediction of displacements in concrete dams, it has also been applied to predict other variables. Simon et al. [67] estimated uplifts and leakage with HST, although they obtained more accurate results with neural networks (NN). Guedes and Coelho [24] built a model for the prediction of leakage in Itaipú Dam with the form $a_1 h_{6,11}^2 + a_2 t + a_3 t^2 + a_4 log(1 + t)$, where $h_{6,11}$ is the average reservoir level between 6 and 11 days before the measurement. Breitenstein et al. [8] also studied leakage, although they discarded both the seasonal and the temporal terms. Yu et al. [80] combined HST with PCA to predict the opening of a longitudinal crack in Chencun Dam.

A common feature to HST and its variations is that the output is computed as a linear combination of the inputs. Hence, they are all multi-lineal regression models (MLR), so their coefficients can be fitted by least squares. Other approaches based on MLR have been applied in dam safety, considering a larger set of inputs (e.g. [19, 69]).

2.2 Models to Account for Delayed Effects

It is well known that dams respond to certain loads with some delay [39]. The most typical examples are:

- The change in pore pressure in an earth-fill dam due to reservoir level variation [6].
- The influence of the air temperature in the thermal field in a concrete dam body [67].

Other phenomena have been identified which are governed by similar processes. For example, Lombardi [38] noticed that the structural response of an arch dam to hydrostatic load comprised both elastic and viscous components. Hence, the displacements not only depended on the instantaneous reservoir level, but also on the past values. Simon et al. [67] reported that leakage flow at Bissorte Dam responded to rainfall and snow melt with certain delay.

Several approaches have been proposed to account for these effects. The most popular consists of including moving averages or gradients of some explanatory variables in the set of predictors. In the above mentioned study, Guedes and Coelho [24] predicted the leakage flow on the basis of the mean reservoir level over the course of a five-days period. Sánchez Caro [62] included the 30 and 60 days moving average of the reservoir level in the conventional HST formulation to predict the radial displacements of El Atazar Dam. Popovici et al. [53] used moving averages of 3, 10 and 30 days of the air temperature, together with the pool level in the previous 3 days to the measurement in order to predict displacements in a buttress dam with neural networks (NN). Crépon and Lino [15] reported significant improvement in the prediction of piezometric levels and leakage flows by considering the accumulated rainfall and the derivative of the hydrostatic load as predictors.

This approach requires a criterion to determine which moving averages and gradients should be considered for each particular case. Demirkaya and Balcilar [19] performed a sensitivity analysis to select the number of past values to include both in an MLR and in a NN model. They used the same period for the external and internal temperatures, as well as for the reservoir level, and found that the most accurate results were obtained with an MLR model considering data from 30 previous days. Although their results compared well to those proposed by the participants in the 6th ICOLD Benchmark Workshop[2] [81], they lacked physical meaning: they would imply that the dam responded with the same delay to the water level, the air temperature, and the internal temperature field.

Santillán et al. [64] proposed a methodology to select the optimal set of predictors among various gradients of air temperature and reservoir level. They used the gradients instead of the moving averages to ensure independence among predictors (moving averages are correlated with the original correspondent variables). They combined it with NN to predict leakage flow in an arch dam.

A more formal alternative to conventional HST to account for delayed effects was proposed by Bonelli and Royet [7]. It is based on the hypothesis that the delayed effect depends on the convolution integral of the impulse response function (IRF) and the loadings:

$$\hat{Y} = \alpha \frac{1}{t_0} \int_0^t e^{-\left(\frac{t-t'}{t_0}\right)} h(t') \partial t' \tag{6}$$

where α is a damping coefficient, t_0 is the characteristic time, which depends on the phenomenon, and $h(t')$ is the reservoir level at time t'. Although the analytical integration of this function is cumbersome, it can be solved by means of numerical approximation. The advantage of this approach is that the coefficients have physical meaning: the characteristic time provides insight into the lag with which the dam reacts to a variation in the input variable, whereas the damping reflects the relation between the amplitude of the reservoir level variation and that of the pore pressure in the location considered within the dam body.

A similar approach was followed by the same author in the frame of the above mentioned 6th ICOLD Benchmark Workshop [4]. In this case, it was intended to account for the delayed response of the dam in terms of the temperature field, with the final aim of predicting radial displacements.

Lombardi [38] suggested an equivalent formulation, also to compute the thermal response of the dam to changes in air temperature. Although the development was slightly different, the numerical approximation to the integral is equivalent. Lombardi arrived at the following expression [39]:

$$\hat{Y}(t) = \alpha \cdot Y(t - \Delta t) + \left(1 + \frac{\alpha}{\beta} - \frac{1}{\beta}\right) X(t) + \left(\frac{1}{\beta} - \frac{\alpha}{\beta} - \alpha\right) X(t - \Delta t) \tag{7}$$

where $\alpha = e^{\frac{-\Delta t}{t_0}}$, $\beta = \frac{\Delta t}{t_0}$, and Δt is the measurement interval. It should be noted that the numerical integration of (6) by means of (7) leads to a predictive model which is a linear combination of:

- the value of the predictors at t and $t - \Delta t$
- the value of the output variable at $t - \Delta t$

This is the conventional form of a first order auto-regressive exogenous (ARX) model. In general, these models require specific algorithms to determine the appropriate

---

[2] In the 6th ICOLD Benchmark Workshop, the participants were asked to provide a data-based model for predicting the radial displacement of Schlegeiss arch dam for the period 1999–2000. Time histories of water level, air temperature and concrete temperatures at various locations were provided for the period 1992–2000, as well as the observed values of the target variable for the period 1992–1998.

order of the model for a given case, i.e., the amount of past values to consider for the output and each of the input variables. The next section is devoted to this aspect and to auto-regressive models.

In practice, an input transformed by Eq. (6) is similar to a weighted moving average (WMA) [39]. Figure 1 shows the comparison between both transformations of 4 inputs: (a) a sinusoidal, (b) a random variable, (c) a cyclic variable with random noise and (d) an isolated pulse. It can be seen that the transformed sinusoidal can be accurately modelled with an appropriate moving average. The difference between IRF and WMA is greater for random inputs, and the discrepancy increases as the signal-to-noise ratio decreases.

IRF has the advantage of its physical meaning, and has offered accurate results for determined outputs. Nonetheless, given that it makes a strong assumption on the characteristics of the phenomenon, it is restricted to specific processes. Even when applied to a similar phenomenon, such as the effect of precipitation on the pore pressure on an earth-fill dam, the accuracy decreases [7]. Moreover, the coefficients lose their physical meaning in this case.

2.3 Auto-Regressive (AR) Models

The use of the previous (lagged) value of the output to calculate a prediction for current record may induce to question (a) whether the observed previous value or the precedent prediction should be used, and (b) whether the model parameters should be readjusted at every time step.

In general, using the actual previous value and refitting the model should provide better prediction accuracy, but such a model would not be able to detect gradual anomalies [79]: it would learn the abnormal behaviour and treat it as ordinary [38]. Riquelme et al. [59] improved the accuracy of a NN model by several orders of magnitude by applying this approach.

The opposite alternative is to fit the model to data gathered for a given time period, and make long-term predictions on a step-by-step basis [48], i.e., predict the output at $t + 1$, and use it (the prediction; not the observation) to estimate the value at $t + 2$. This procedure may fail in error propagation [10], but in principle should be appropriate to unveil gradual anomalies.
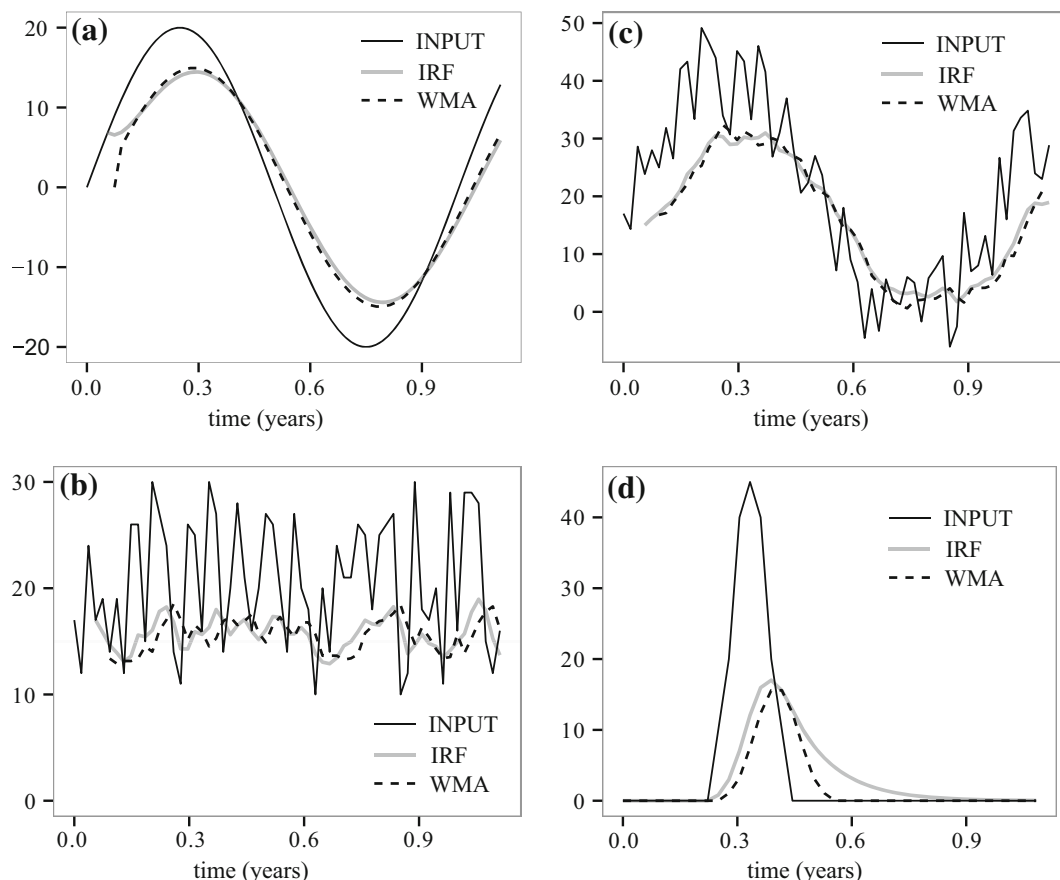

Fig. 1 Comparison between impulse response function (IRF) and weighted moving average (WMA) for various inputs: **a** sinusoidal, **b** random, **c** sinusoidal with random noise and **d** impulse

An intermediate choice is to use the actual measurement of the output variable, without readjusting the model parameters. In this case, the coefficients obtained on the basis of a period of normal behaviour are applied to future observations, hence the model could detect changes in the relation between current and next values of the output.

Although several authors built predictive models based on lagged output values, most of them did not mention which of the described approaches applied. Palumbo et al. [48], should have used the previous prediction, given that they presented a solution to the 6th ICOLD Benchmark Workshop, and the observed values of the output were not provided to the participants beforehand.

If the possibility of including past values of the variables is considered, a criterion to select some of the available shall be defined. Otherwise, the amount of predictors is quite high. For example, Piroddi and Spinelli [52] considered the most general form of a non-linear autoregressive exogenous model (NARX), which depended on current and previous values of the input variables, on precedent values of the output, as well as on linear and non-linear combinations of them. They applied a specific algorithm for selecting 11 predictors in the final model.

In general, these models prioritise prediction accuracy over explanatory capability. The greater the number of variables in the model, the harder it is to interpret and to isolate the effect of each component. Nonetheless, some procedures have been proposed to interpret models whose parameters do not have physical meaning, as described in Sect. 3.2.

2.4 Neural Networks (NN)

Linear models are not well suited to reproduce non-linear behaviour, even though some actions are considered in the form of high order polynomials [12]. On the contrary, NN models are flexible, and allow modelling complex and highly non-linear phenomena. Although there are various types of NN models [3], the vast majority of applications for dam monitoring data analysis are based on the multi-layer perceptron (MLP). Such models, as their name suggests, are comprised by a number of perceptrons (also called "units", or "neurons") organised in different layers: input, hidden, and output (Fig. 2). In principle, several hidden layers can be used (see Sect. 2.6), but one is mostly adopted in practice [3].

The input of each unit U_l is a linear combination of the predictors X^j:

$$c_l = \sum_{j=1}^{d} X^j \cdot w_l^j + b_l \tag{8}$$

which is later transformed by an activation function g to compute the neuron's output:

$$z_l = g(c_l) \tag{9}$$

Several forms of g can be chosen (non-linear in general), although sigmoid functions are often employed, such as the logistic (10) and the hyperbolic tangent (11) (Fig. 3). As an exception, Su et al. [70] selected Mexico-hat wavelet functions (12) to obtain a wavelet neural network (WNN) model, otherwise similar to conventional NN models described in this section.

$$g(c_l) = \frac{1}{1 + e^{-c_l}} \tag{10}$$

$$g(c_l) = \frac{e^{c_l} - e^{-c_l}}{e^{c_l} + e^{-c_l}} \tag{11}$$

$$g(c_l) = \left(1 - c_l^2\right) \cdot e^{\left(1 - \frac{c_l^2}{2}\right)} \tag{12}$$

The output layer may be composed of one of the described neurons, although a linear transform is frequently chosen, so that the overall model output is computed as:

$$\hat{Y} = \sum_{l=1}^{L} w_{out}^l \cdot g\left(\sum_{j=1}^{d} X^j w_l^j + b_l\right) + b_{out} \tag{13}$$

NN models can be thought of as an extension of MLR, which output c_l is expanded by the perceptron through a non-linear transformation g [25]. It should be noted (Fig. 3) that the sigmoid functions have a linear interval, thus an unit with small weights performs a linear transformation. On the contrary, they have horizontal asymptotes, which may cause numerical problems. While it is widely acknowledged that the variables shall be normalised before fitting an NN model, some authors restrict them to the range [0.1, 0.9] to avoid the above mentioned problems [23, 56, 75].

The most common learning algorithm is called back-propagation: NN model parameters $\{w_l^j, b_l, w_{out}^l, b_{out}\}$ are randomly initialised, and iteratively updated to minimise a cost function (typically the sum of the squared errors), by means of the gradient descent method [25].

The issues to be considered for building an NN model are the following:

1. The best network architecture, i.e., number of layers and perceptrons in each layer, is not known beforehand. Some authors focus on the definition of an efficient algorithm for determining an appropriate network architecture [64], whereas others use conventional cross-validation [40] or a simple trial and error procedure [75].
2. The training process may reach a local minimum of the error function. The probability of occurrence of this event can be reduced by introducing a learning rate parameter [75].
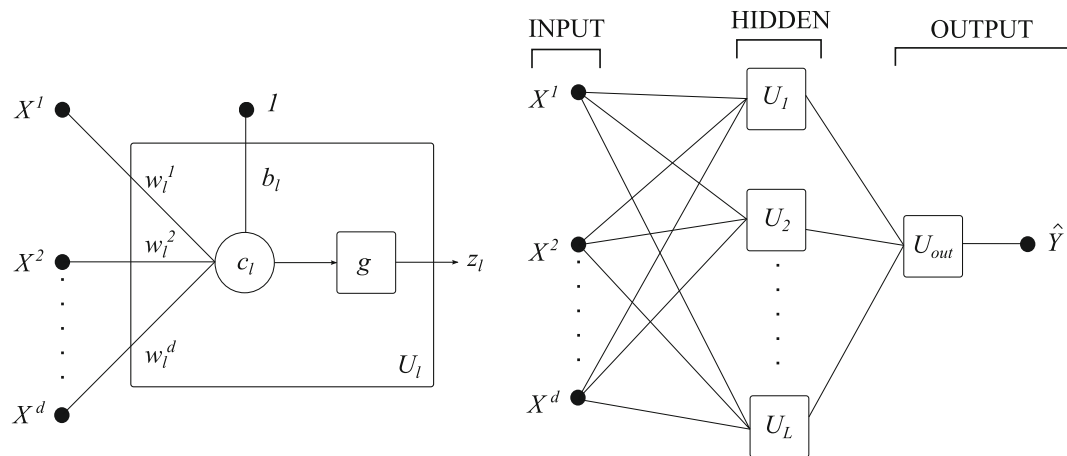
Fig. 2 *Left* schematic model of a perceptron U_l. *Right* Multilayer Perceptron formed by L units, $U_1, \ldots, U_L$
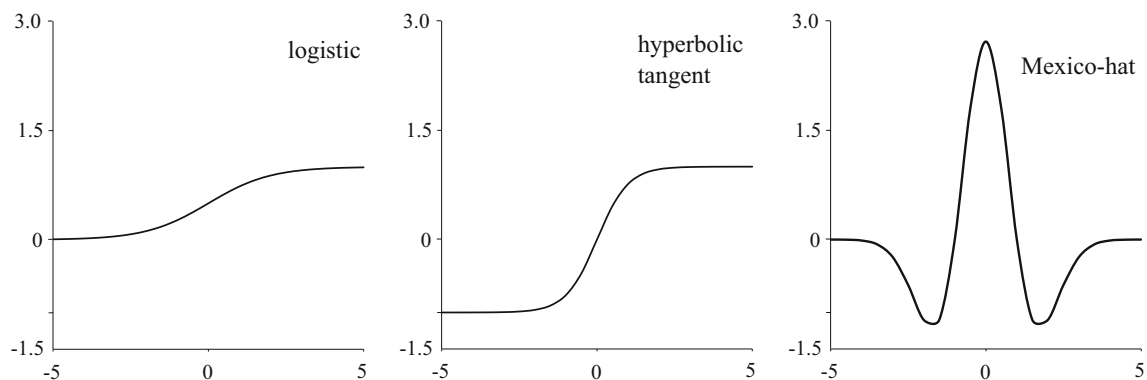

Fig. 3 Common activation functions in NN models

3. The NN models are prone to over-fitting. Various alternatives are suitable for solving this issue, such as *early stopping* and *regularisation* [25].

The fitting procedures greatly differ among authors. While Simon et al. [67] trained an MLP with three perceptrons in one hidden layer for 200,000 iterations, Tayfur et al. [75] used regularisation with 5 hidden neurons and 10,000 iterations. Neither of them followed any specific criterion to set the number of neurons. For his part, Mata [40] tested NN architectures with one hidden layer having 3–30 neurons on an independent test data set. He repeated the training of each NN model 5 times with different initialisation of the weights.

Kao and Loh [30] proposed a two-step procedure: first, the number of neurons was fixed whereas the optimal amount of iterations was computed. Second, NN models with different numbers of hidden nodes were trained with the selected amount of iterations, and the final architecture was chosen as the one which provided the lowest error in a validation set.

The results of the different studies are not comparable, due to the specific features of each case. Nonetheless, the lack of agreement on the training process suggests that similar results can be obtained with different criteria, provided enough care is taken to avoid over-fitting. This is in accordance with Hastie et al. [25], who stated that in general it is enough to set the architecture and compute the appropriate regularisation parameter, or vice versa.

NN models have been used regularly in dam monitoring in recent years. There is an increasing number of published studies, both in academic and professional journals. The most recent ICOLD bulletin on dam surveillance [27] mentions NN as an alternative to HST and deterministic models, although it terms the tool as a "possible future alternative" to be developed, which suggests that it is far from being implemented in the daily practice.

2.5 Adaptive Neuro-Fuzzy Systems (ANFIS)

Fuzzy logic allows inclusion of prior knowledge of the phenomenon, as opposed to the NN, who "learn" from the data. ANFIS models bring together the flexibility and ability to learn of the NN with the feasibility of

interpretation of fuzzy logic. In fact, ANFIS can be considered a class of NN [60]. They are meant for highly non-linear, complex phenomena which vary with time [28].

Among the different types of ANFIS schemes, most previous references in dam monitoring used Takagi–Sukeno (T–S) type, whose singularity is that its output is a combination of linear functions [72]. As an exception, Opyrchal [47] used fuzzy logic to qualitatively locate seepage paths in Tresna and Dobczyce dams.

Fuzzy logic is based on the concept of membership functions (MF). Each continuous variable X^j is decomposed into K^j classes (for example, the reservoir level, which is continuous, can be transformed into "low", "medium" and "high"; see Fig. 4). The particularity of fuzzy logic is that these classes have certain overlapping. Thus, a given reservoir level will generally have a different degree of membership (DOM), between zero and one, for more than one class. For Gaussian MF:

$$DOM_{jk}(X^j) = \frac{1}{1 + \left[\left(\frac{X^j - v_{jk}}{\lambda_{jk}}\right)^2\right]^{\mu_{jk}}}$$ (14)

$$j = 1, \ldots, d; k = 1, \ldots, K^j$$

The number of classes for each input (K^j, which can be different among inputs X^j) are prescribed by the modeller, whereas the shape and position of their MF are determined by the *premise* parameters v, λ and μ (Eq. 14), to be determined during training.

The other essential component in an ANFIS model is a set of rules, which take the form:

$$R_1 : if\ X^1 \in MF_{11} \wedge X^2 \in MF_{21} \wedge \cdots \wedge X^d \in MF_{d1} \Rightarrow$$
$$f_1 = p_{10} + p_{11}X^1 + p_{12}X^2 + \cdots + p_{1d}X^d$$
$$R_2 : if\ X^1 \in MF_{11} \wedge X^2 \in MF_{21} \wedge \cdots \wedge X^d \in MF_{d2} \Rightarrow$$
$$f_r = p_{20} + p_{21}X^1 + p_{22}X^2 + \cdots + p_{2d}X^d$$
$$\cdots$$
$$R_R : if\ X^1 \in MF_{1K^1} \wedge X^2 \in MF_{2K^2} \wedge \cdots$$
$$\cdots \wedge X^d \in MF_{dK^d} \Rightarrow$$
$$f_R = p_{R0} + p_{R1}X^1 + p_{R2}X^2 + \cdots + p_{Rd}X^d$$ (15)

where $p_{10}, \ldots, p_{Rd}$ are the *consequent* parameters, to be adjusted during model training. It should be noted that there can be up to $\prod_{j=1}^{d} K^j$ rules.

The model output is computed by means of 5 steps:

1. Compute the *DOM* of every input to each fuzzy category (14).
2. Compute the product of the correspondent DOM_{jk}, in accordance with the rules. In ANFIS terminology, these terms are referred to as the *firing strengths* ($w_r; r = 1, \ldots, R$) for each rule:
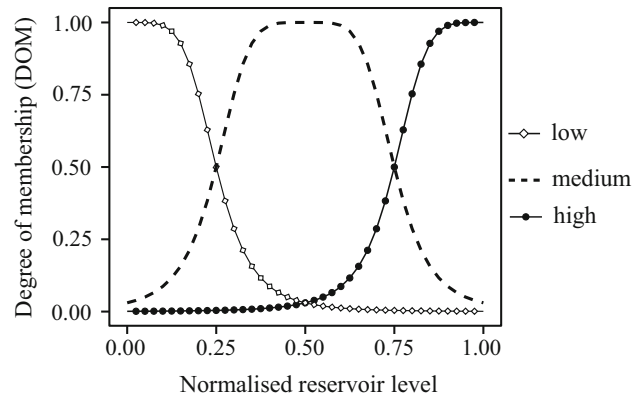

Fig. 4 Possible transformation of the normalised reservoir level into three fuzzy sets with Gaussian form: "low", "medium" and "high"

$$w_1 = DOM_{11} \cdot DOM_{21} \cdot \ldots \cdot DOM_{d1}$$
$$w_2 = DOM_{11} \cdot DOM_{21} \cdot \ldots \cdot DOM_{d2}$$ (16)
$$\cdots$$
$$w_R = DOM_{1K^1} \cdot DOM_{2K^2} \cdot \ldots \cdot DOM_{dK^d}$$

3. Normalise the firing strengths:

$$\overline{w_r} = \frac{w_r}{\sum w_r}$$ (17)

4. Compute the output of each rule, as a linear function of the consequent parameters:

$$O_r = \overline{w_r}f_r = \overline{w_r}(p_{r0} + p_{r1}X^1 + p_{r2}X^2 + \cdots + p_{rd}X^d)$$
$$r = 1, \ldots, R$$ (18)

5. Combine the outputs of each rule to compute the overall output of the ANFIS model:

$$\hat{Y} = \sum_{r=1}^{R} O_r$$ (19)

The final result is a combination of linear functions of the input variables. The non-linearity is modelled in the MF, which are typically Gaussian, as shown in the example of Fig. 4. Each MF is determined on the basis of 3 *premise* parameters, fitted with a hybrid method, in which the following steps are alternated:

1. The MF are fixed, and the consequent parameters are adjusted by least squares.
2. The premise parameters are modified by means of the gradient descent method.

The criterion of the user is more important for building ANFIS than for other kinds of models. Both the prediction accuracy and the possibility of interpreting the results may vary greatly according to the number of inputs (d), MF (K^j) and rules (R). It should be noted that the number of parameters in a first order T–S ANFIS model can be up to:

$$3 \cdot \sum_{j=1}^{d} K^j + (d+1) \cdot \prod_{j=1}^{d} K^j \qquad (20)$$

Ranković et al. [54] prioritised prediction accuracy over model interpretation, by considering lagged values of both the input and output variables as predictors, resulting in an ANFIS model with $d = 5, K^j = 2, \forall j$ and $R = 32$. They used a zero-order T–S model, in which $f_r = p_{r0}, \forall r \in [1, R]$, and two-sided Gaussian MF, defined by 4 parameters each. No attempt was made to interpret the 32 rules.

On the contrary, Xu and Li [78] considered only 9 rules and could identify the worst environmental conditions for crack opening in the Chencun Dam.

For his part, Demirkaya [18] chose $d = 5$ and $K = 4$. Although he limited the number of rules to 4, the final model had 84 parameters.

ANFIS models can be as flexible and accurate as NN, while allowing for introducing engineering knowledge to some extent. If the amount of rules and MF is low, the resultant model can be interpreted. Furthermore, an ANFIS model can be used for qualitatively describing dam behaviour, especially if the output is "fuzzyfied" into linguistic variables [78].

On the contrary, they may comprise a high number of parameters, even with a few rules, which results in a high risk of over-fitting and low interpretability.

2.6 Principal Component Analysis (PCA) and Dimensionality Reduction

PCA is a well known technique in statistics. It was devised to transform a set of partially dependent variables into independent features called principal components (PCs), which are linear combination of the original variables. It is acknowledged that the first PCs contain the relevant information, whereas the less influential correspond to the signal noise. It has been used in dam monitoring for various purposes.

Mata et al. [42] used PCA to select the most useful thermometers to predict radial displacements in an arch dam. They pointed out the potentiality of this tool to select a group of sensors to be automatised in a given dam.

Yu et al. [80] applied PCA to a group of sensors to measure the opening of a longitudinal crack in an arch dam. They reported that PCA was useful for reducing the dimensionality of the problem, as well as to separate the signal from the noise. They also defined alarm thresholds as a function of the first PCs. Cheng and Zheng [11] followed a similar procedure: they analysed the covariance matrix of the outputs to separate the effect of the environmental variables from the signal noise.

Similar applications were due to Chouinard et al. [13], and Chouinard and Roy [12], who extracted PCs from a set of outputs (radial displacements at pendulums) to better understand the behaviour of the structure. They focused on the model interpretation, rather than on the prediction accuracy. In this line, Nedushan [44] extracted PCs from a group of sensors to analyse them jointly, as well as to identify the correlations by means of stepwise linear regression. He defined a set of predictors (reservoir level, temperature and time), and built linear regression models by adding the most relevant one by one.

A limitation of PCA is that only linear relations between variables are considered. If the dependency is non-linear, it may lead to misinterpretation of the results. Non-linear principal component analysis (NPCA) can be an alternative, as showed by Loh et al. [37] and Kao and Loh [30], who applied it by means of auto-associative neural networks (AANN) to predict radial displacements in an arch dam.

AANN are a special kind of NN models, formed by 5 layers (Fig. 5), which can be viewed as two NN models put in series. The intermediate (bottleneck) layer has fewer neurons than the number of model inputs, and the target outputs equal the inputs. Thus, the first part of the model reduces its dimensionality, computing some sort of non-linear PCA. The right-hand-side of the AANN is a conventional NN whose inputs are the non-linear PCs.

Jung et al. [29] developed a methodology to identify anomalies in piezometric readings in an earth-fill dam by means of moving PCA (MPCA), which is conventional PCA applied to different time periods. The goal was to detect significant variations in the PCs over time, which would reveal a change in dam behaviour.

PCA is mostly applied to input or output variable selection. The first option may increase the prediction accuracy, whereas the second can be useful for managing very large dams with a large amount of devices. For example, more than 8,000 instruments were installed to control the behaviour of the Three Gorges Dam [80].

2.7 Other ML Techniques

There is a wide variety of ML algorithms which can be useful for dam monitoring data analysis. Their accuracy depends on the specific features of every prediction task. Given that research on ML is a highly active field, the algorithms are constantly improved and new practical applications are reported each year. Some of them have been applied to dam monitoring analysis. They are considered in this section more briefly than others, in accordance with their lower popularity in dam engineering so far. This does not mean that they can not offer advantages over the methods described previously.

Fig. 5 Architecture of an auto-associative neural network. There are 3 hidden layers between the inputs and the output. The central one is called "bottleneck" layer, and shall have fewer nodes than model inputs, so that each one can be considered a non-linear principal component of the inputs
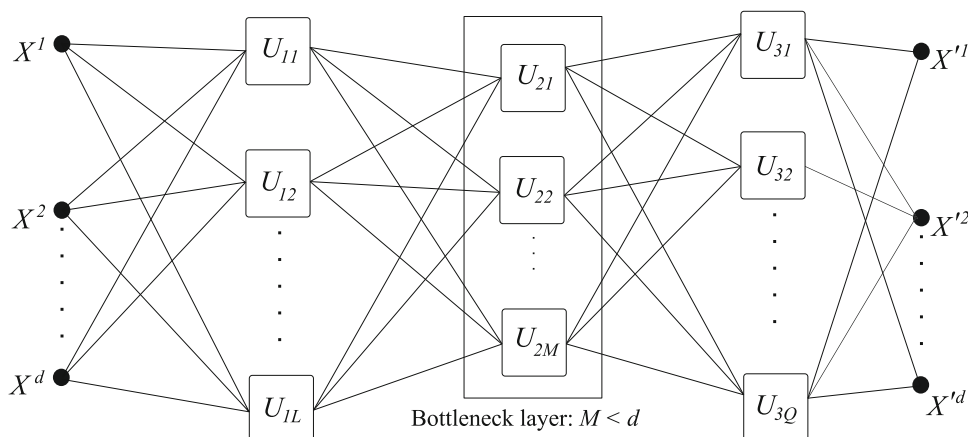
Support vector machines (SVM) stand among the most popular ML algorithms nowadays. They combine a non-linear transformation of the predictor variables to a higher dimensional space, a linear regression on the transformed variables, and an ε-insensitive error function that neglects errors below a given threshold [68]. Cheng and Zheng [11] used SVM in combination with PCA for short-term prediction of the response of the Minhuatan gravity dam. Although the results were highly accurate, the computational time was high. Rankovic et al. [55] built a behaviour model based on SVM for predicting tangential displacements.

K-nearest neighbours (KNN) is a non-parametric method which requires no assumptions to be made about the physics of the problem; it is solely based on the observed data. The KNN method basically consists on estimating the value of the target variable as the weighted average of observed outputs in similar conditions within the training set. The similarity between observed values is measured as the Euclidean distance in the d-dimensional space defined by the input variables.

A clear disadvantage of this type of model is that if the Euclidean distance is used as a measure of similarity, all the predictors are given the same relevance. Hence, including a low relevant variable may result in a model with poor generalisation capability. As a consequence, variable selection is a critical aspect for fitting a KNN model.

Saouma et al. [65] presented a solution to the 6th ICOLD Benchmark Workshop based on KNN. To determine the similarity of observations, they used only two significant predictors (the reservoir level and a thermometer in the dam body) among the eight available. This selection of variables was performed by trial and error, although other criteria exist, as described in the next section.

Stojanovic et al. [69] combined greedy MLR with variable selection by means of genetic algorithms (GA). Unlike HST, they considered all the observed variables in various forms (e.g. $h, h^2, h^3, \sqrt{h}$, etc.). They defined a

methodology to select the best set of predictors which could be useful to update the predictive model in case of missing variables. A similar approach was followed by Xu et al. [77], though with a smaller set of potential inputs.

Salazar et al. [61] performed a comparative study among various statistical and ML methods, including HST, NN, and others which had never been used before in dam monitoring, such as random forests (RF) or boosted regression trees (BRT). It was reported that innovative ML algorithms offered the most accurate results, although no one performed better for all 14 outputs analysed, which corresponded to radial and tangential displacements and leakage flow in an arch dam.

3 Methodological Considerations for Building Behaviour Models

While each model has specific issues to take into account, there are also common aspects to consider when developing a prediction model, regardless of the technique. They are discussed in this section, in relation with a selection of 59 studies corresponding to 41 papers presented at conferences and in scientific journals. It is not an exhaustive review: the studies were selected on the basis of their relevance and interest, following the authors' criterion.

The Tables 1 and 2 summarise the main characteristics of the studies reviewed. It was found that most of them (38/59) considered radial displacements, especially in arch dams (31/59). This reflects the greater concern of dam engineers for this variable and dam typology, although other indicators such as leakage or uplift are acknowledged as equally relevant for dam safety [39]. The lower frequency with which the latter are chosen as target variables may be partly due to their more complex behaviour, which makes them harder to reproduce and interpret [39]. The HST and MLR methods, which have been the only ones

Table 1 Review summary: case studies

Id	Author	Year	Dam(s)	Country	Typology	Output	# Outputs
1	Breitenstein [8]	1985	Limberg, Mooser, Drossen	Switzerland	ARC, GRA, ARC	RAD	7
2	Breitenstein [8]	1985	Limberg, Mooser, Drossen	Switzerland	ARC, GRA, ARC	LEAK	6
3	Guedes [24]	1985	São Simão	Brazil	EF+GRA	RAD	1
4	Guedes [24]	1985	Água Vermelha	Brazil	EF+GRA	RAD	1
5	Guedes [24]	1985	Funil	Brazil	ARC	PIEZ	1
6	Guedes [24]	1985	Sobradinho	Brazil	EF+GRA	JOINT	1
7	Guedes [24]	1985	Itaipú	Brazil	GRA	LEAK	1
8	Bonelli [7]	2001	Alzitone, Chamboux, La Verne	France	EF	PIEZ	9,6,4
9	Bonelli [4]	2001	Schelegeis	Austria	ARC	RAD	1
10	Carrere [9]	2001	Schelegeis	Austria	ARC	RAD	1
11	Saouma [65]	2001	Schelegeis	Austria	ARC	RAD	1
12	Palumbo [48]	2001	Shclegeis	Austria	ARC	RAD	1
13	Nedushan [44]	2002	Chute-à-Caron	Canada	GRA	RAD, TAN, VERT	1,1,1
14	Piroddi [52]	2003	Schelegeis	Austria	ARC	RAD	1
15	Tayfur [75]	2005	Jeziorsko	Poland	CFRD	PIEZ	4
16	De Sortis [17]	2006	Ancipa	Italy	BUT	RAD	5
17	De Sortis [17]	2006	Sabbione	Italy	BUT	RAD	3
18	De Sortis [17]	2006	Malga Bissina	Italy	BUT	RAD	5
19	S. Caro [62]	2007	El Atazar	Spain	ARC	RAD	46
20	Léger [33]	2007	Schelegeis	Austria	ARC	RAD	1
21	Su [70]	2007	?	China	AG	VERT	1
22	Panizzo [49]	2007	Pieve di Cadore	Italy	AG	RAD	1
23	Lombardi [39]	2008	?	?	ARC	RAD	1
24	Lombardi [39]	2008	?	?	ARC	LEAK	1
25	Bonelli [5]	2007	?	?	EF	PIEZ	14
26	Bonelli [6]	2008	?	?	EF	PIEZ	16
27	Yu [80]	2010	Chencun	China	AG	CRACK	5
28	Perner [51]	2010	Zillergruendl	Austria	ARC	RAD	2
29	Demirkaya [18]	2010	Schelegeis	Austria	ARC	RAD	1
30	Riquelme [59]	2011	La Baells	Spain	ARC	RAD	1
31	Mata [40]	2011	Alto Rabagão	Portugal	ARC	RAD	1
32	Rankoviĉ [54]	2012	Bocac	Bosnia Herzegovina	ARC	RAD	2
33	Xu [77]	2012	Chencun	China	AG	CRACK	1
34	Demirkaya [19]	2012	Schelegeis	Austria	ARC	RAD	1
35	Demirkaya [19]	2012	Schelegeis	Austria	ARC	RAD	1
36	Cheng [11]	2013	Mianhuatan	China	GRA	RAD	12
37	Cheng [11]	2013	Mianhuatan	China	GRA	UP	16
38	Popovici [53]	2013	Gura Râului	Romania	BUT	RAD, TAN, ROCK	2, 2, 3
39	Tatin [73]	2013	Castelnau	France	GRA	RAD	1
40	Tatin [73]	2013	Castelnau	France	GRA	RAD	1
41	Li [34]	2013	Wanfu	China	ARC	RAD	4
42	Li [34]	2013	Wanfu	China	ARC	RAD	4
43	Simon [67]	2013	Pareloup	France	ARC	PIEZ	1
44	Simon [67]	2013	Bissorte	France	GRA	LEAK	4
45	Simon [67]	2013	Monteynard	France	ARC	RAD	1
46	Simon [67]	2013	Monteynard	France	ARC	RAD	1
47	Nourani [45]	2013	Sahand	Iran	EF	PIEZ	4
48	Kao [30]	2013	Fei-Tsui	Taiwan	ARC	RAD	13

Table 1 continued

Id	Author	Year	Dam(s)	Country	Typology	Output	# Outputs
49	Kao [30]	2013	Fei-Tsui	Taiwan	ARC	RAD	13
50	Kao [30]	2013	Fei-Tsui	Taiwan	ARC	RAD	13
51	Mata [42]	2013	Alto Lindoso	Portugal	ARC	RAD	5
52	Jung [29]	2013	?	USA	EF	PIEZ	1
53	Stojanovic [69]	2013	Bocac	Bosnia Herzegovina	ARC	RAD	1
54	Ranković [56]	2014	Iron Gate 2	Serbia/Romania	EF+GRA	PIEZ	2
55	Ranković [56]	2014	Iron Gate 2	Serbia/Romania	EF+GRA	PIEZ	2
56	Santillán [64]	2014	La Baells	Spain	ARC	LEAK	1
57	Salazar [61]	2014	La Baells	Spain	ARC	RAD, TAN, LEAK	5, 5, 4
58	Ranković [55]	2014	Iron Gate 2	Serbia/Romania	EF+GRA	TAN	2
59	Tatin [74]	2015	Eguzon, Izourt, Roselend, Tignes, Vouglans, Bissorte, Gittaz, Sarrans	France	GRA,ARC, BUT	RAD	1

Typology: ARC = arch; GRA = gravity; EF = earth-fill; AG = arch-gravity; BUT = buttress; CFRD = concrete-faced rockfill dam; ? = Not specified. Outputs: RAD= radial displacements; LEAK = leakage flow; PIEZ = pore pressure; JOINT = joint opening; TAN = tangential displacements; VERT = vertical displacements; UP = uplift pressure; ROCK = rockmeter displacements; CRACK = crack opening

available for a long time, are not suitable to model them [67], although some references exist [8, 24].

3.1 Input Selection

In previous sections, it was pointed out that the model performance depended on the predictor variables considered. The range of options for variable selection is wide. In most of the papers reviewed, no specific method was applied for variable selection, apart from user criterion (e.g. [49]) or "a priori knowledge" (e.g. [54]).

This issue has arisen in combination with the use of NN [19, 30, 37, 49, 56], NARX [37, 52], MLR [69] and ANFIS models [54].

First, the selection is limited by the available data. While the reservoir level and the air temperature are usually measured at the dam site, other potentially influential variables, such as precipitation, are frequently not available. One of the advantages of the HST method is that only the reservoir level is required.

Second, it must be decided whether or not to use the lagged values of the target variable for prediction. The consequences of making predictions from the output itself have already been mentioned in sect. 2.3, regardless whether the observed or the estimated previous value is used. It can be concluded that the AR models prioritise prediction accuracy over model interpretation.

Third, the possibility of adding derived variables (and which ones), such as moving averages and gradients, can be considered. They can be set beforehand, on the basis of engineering judgement, or selected by means of some performance criterion from a wide set of variables.

Finally, consideration should be given to include non-causal variables in the model. For example, is it appropriate to base the prediction of radial displacements at a given location on the displacement recorded at another point of the dam? Will it improve the model accuracy? What consequences would it have in the interpretation of the results?

Some models like the HST are often used with a set of specific predictors, and therefore variable selection is restricted to the order of the polynomial of the reservoir level, and the shape of the time dependent functions. The opposite case is the NARX method, which can be used with a high amount of predictor variables.

Hence, the criterion to be used depends on the type of data available, the main objective of the study (prediction or interpretation), and the characteristics of the phenomenon to be modelled. Again, engineering judgement is essential to make these decisions.

The selection of predictors may be useful to reduce the dimensionality of the problem (essential for NARX models), as well as to facilitate the interpretation of the results. PCA can be used for this purpose [42], as well as AANN [37]. Some specific methods for variable selection in dam monitoring analysis have been proposed, by means of backward elimination [64] genetic algorithms (GA) [69], and singular spectrum analysis (SSA) [37], although the vast majority of authors applied trial and error or engineering judgement.

3.2 Model Interpretation

The main interest of this work focuses on model accuracy: a more accurate predictive model allows defining narrower thresholds, and therefore reducing the number of false anomalies. Nonetheless, once a value above (or below, if

Table 2 Review summary. Methods

Id	Model	Inputs	Training set (years/# samples)	Validation set (years/# samples)	Validation data (%)	Error metric (see Table 3)
1	MLR	$h, S, t, T_{air}, \partial(T_{air}), \partial(h)$	10/14600	0.0/	0	R^2
2	MLR	h	10/14600	0.0/	0	R^2
3	MLR	$h, mav(T_{air})$	1/103	0.0/	0	r
4	MLR	$h, mav(T_{air})$	0.5/63	0.0/	0	r
5	MLR	$t, mav(h)$	2/230	0.0/	0	r
6	MLR	$t, mav(T_c)$	2.5/66	0.0/	0	r
7	MLR	$t, mav(h)$	0.5/86	0.0/	0	r
8	IRF	$h, lag(h), P, lag(P), t$	Var/Var	0.0/	0	–
9	IRF	h, T_{air}	7/2557	2.0/730	22	–
10	HST	h, T_{air}	7/2557	2.0/730	22	r, R^2, σ_ϵ
11	KNN	h, T_c	7/2557	2.0/730	22	r, R^2, σ_ϵ
12	NARX	$h, T_{air}, T_c, lag(h), lag(T_c), lag(T_{air})$	7/2555	2.0/730	22	$RMSE$
13	NN	T_c, t	1.5/548	1.5/548	50	R^2
14	NARX	$h, T_{air}, T_c, lag(h), lag(T_c), lag(T_{air})$	7/2555	2.0/730	22	MSE
15	NN	h	1/26	2.0/52	67	$RMSE, MAE, R^2$
16	HST	h, S, t	2 to 15/730 to 5475	0.0/	0	$r, \sigma_\epsilon, \frac{\sigma_\epsilon}{D/2}$
17	HST	h, S, t	5/1825	0.0/	0	$r, \sigma_\epsilon, \frac{\sigma_\epsilon}{D/2}$
18	HST	h, S, t	9/3285	0.0/	0	$r, \sigma_\epsilon, \frac{\sigma_\epsilon}{D/2}$
19	MLR	$h, mav(h), S, T_{air}, mav(T_{air})$	24.5/8943	0.0/	0	σ_ϵ, MSE
20	HTT	h, T_c, t	5/1825	0.0/	0	r
21	WNN	h, S, t	11/44	2.0/8	15	MAE
22	NN	$h, lag(rad), T_{air}, T_c$	7/2555	0.0/	0	$R^2, pdf(\epsilon), MSE$
23	IRF	$h, lag(h), lag(rad), T_{air}$	4/?	0.0/	0	σ_ϵ
24	IRF	$h, lag(h), lag(out)$	5/?	0.0/	0	–
25	IRF	$h, lag(P)$	3/167	0.0/	0	R^2
26	IRF	$h, hd, lag(P)$	var/var	0.0/	0	R^2
27	HST	h, S, t	10/1200	0.0/	0	r
28	HYB	h, T_c, t	22/8030	0.0/	0	–
29	ANFIS	h, T_{air}, T_c	6/2044	1.0/365	15	$r, RMSE, MAE$
30	NN	$h, T, mav(T), lag(out)$	18/706	12.0/470	40	$MAPE$
31	NN	h, S	23/914	1.8/69	7	$MAE, MaxAE, r$
32	ANFIS	$lag(h), lag(S), lag(out)$	9/657	2.0/140	18	$r, MAE, RMSE$
33	ANFIS	T_{air}, h	15/400	?/?	0	$RMSE$
34	MLR	$h, T_{air}, T_c, lag(h), lag(T_{air}), lag(T_c)$	7/2555	2.0/730	22	ME, σ_ϵ, R^2
35	NN	$h, T_{air}, T_c, lag(h), lag(T_{air}), lag(T_c)$	7/2555	2.0/730	22	ME, σ_ϵ, R^2
37	PCA, SVM	h, T_{air}, P	3/900	0.2/56	6	–
38	NN	$t, h, T_{air}, lag(h), mav(T_{air})$	14/?	2.0/?	13	r, R^2, σ_ϵ
39	GRAD	$h, S, t, IRF, T_{air}, T_w$	12/?	0.0/	0	σ_ϵ
40	SLICE	$h, S, t, IRF, T_{air}, T_w$	12/?	0.0/	0	σ_ϵ
41	HTT	h, S, T_c	3.2/169	0.4/20	11	$R^2_{adj}, \sigma_\epsilon, pdf(\epsilon)$
42	ECM	$h, S, T_c, \epsilon(t-1)$	3.2/169	0.4/20	11	$R^2_{adj}, \sigma_\epsilon, pdf(\epsilon)$
43	NN	h, S, t	?/429	0.0/	0	σ_ϵ, MSE
44	IRF+NN	$h, S, t, T_{air}, IRF(P), IRF(M)$	?/?	0.0/	0	R^2
45	NN	$h, S, t, IRF(T_{air})$	?/?	0.0/	0	σ_ϵ
46	HSTT	$h, S, t, IRF(T_{air})$	?/?	0.0/	0	σ_ϵ
47	NN	$h, h_d, lag(P)$	1.1/58	0.4/18	24	R^2

Table 2 continued

Id	Model	Inputs	Training set (years/# samples)	Validation set (years/# samples)	Validation data (%)	Error metric (see Table 3)
48	NN	h, T_c	22/8120	0.3/62	1	$R^2, pdf(\varepsilon), MSE$
49	NARXNN	$h, lag(h), lag(out)$	22/8120	0.3/62	1	$R^2, pdf(\varepsilon), MSE$
50	AANN	$lag(rad)$	22/8120	0.3/62	1	$R^2, pdf(\varepsilon), MSE$
51	HTT	h, T_c	5/95	0.0/	0	$R^2_{adj}, \sigma_\varepsilon, \varepsilon_{max}, \varepsilon_{min}, SSE$
52	MPCA	h	6/4380	0.0/	0	–
53	MLR	h, T_c, T_{air}, P, t	6/2550	1.0/365	13	$R^2_{adj}, RMSE$
54	NN	$h_d, lag(h_d)$	8/163	1.0/20	11	r, r^2, MSE, MAE
55	MLR	$h_d, lag(h_d)$	8/163	1.0/20	11	r, MSE, MAE
56	NN	$h, T_{air}, \partial(h), \partial(T_{air})$	25.5/918	3.0/103	10	$RMSE$
57	NN, MARS, RF, BRT, SVM	$h, T_{air}, S, t, mav(h), mav(T_{air}), P, \partial(h)$	18/600	10.0/400	40	MAE, ARV
58	SVM	$h, h_d, lag(h), lag(h_d), lag(out)$	11/573	3.0/156	21	r, MAE, MSE
59	GRAD	$h, S, t, IRF, T_{air}, T_w$	8/?	2/?	20	σ_ε

Models: MLR = multilineal regression; IRF = impulse response function; HST = hydrostatic seasonal time; KNN = k-nearest neighbours; NN = neural networks; WNN = wavelet neural networks; NARX = non-linear autoregressive exogenous; HTT = hydrostatic thermal time; HYB = hybrid; ANFIS = adaptive neuro-fuzzy system; PCA = principal component analysis; MPCA = moving PCA; SVM = support vector machine; ECM = error correction model; HSTT = hydrostatic seasonal thermal time; NARXNN = non-linear autoregressive exogenous neural network; AANN = auto-associative neural network; RR = robust regression; MARS = multivariate adaptive regression splines; RF = random forest; BRT = boosted regression trees; WNN = wavelet neuran networks; ECM = error correction method. Inputs: h = upstream pool level; S = season; t = time; $\partial(\cdot)$ = time derivative; T_c = concrete temperature; T_{air} = air temperature; T_w = water temperature; $IRF(\cdot)$ = impulse response function; $lag(\cdot)$ = lagged variable; P = precipitation; out = output; $mav(\cdot)$ = moving average; h_d = downstream pool level; M = snow melt; $pdf(\varepsilon)$ = probability density function of error

appropriate) the warning threshold is registered, an engineering analysis of the situation is needed to assess its seriousness. The ability of the model to interpret dam behaviour may be useful for this purpose.

The HST method has been traditionally used to identify the effect on the response of the dam of each considered action: hydrostatic load, temperature and time (e.g. [40]). However, it is clear that this analysis is only valid if the predictor variables are independent, which is not generally true [38, 66].

On the contrary, the ability of NN and similar models for interpreting dam behaviour is often neglected. They are frequently termed "black box" models, in reference to its lack of interpretability.

It turns out that NN models are well suited to capture complex interactions among inputs, as well as non-linear input-output relations. If an NN model offers a much better accuracy than the HST for a given phenomenon, it is probable that it does not fulfil the hypothesis of HST (input independence, linearity). Hence, it would be more appropriate to extract information on the dam behaviour from the interpretation of the NN model.

The effect of each predictor can be analysed by means of *ceteris paribus* analysis [40]: the output is computed for the range of variation of the variable under consideration, while keeping the rest at constant values. They can be set either to the correspondent mean or to several other values, in order to gain more detailed information on the dam response. Analyses of this kind can be found in the pertinent literature: Mata [40] calculated the effect of the reservoir level on the radial displacements of an arch dam for each season of the year, and the effect of temperature when setting the pool level at several constant values. Similar studies are due to Santillán et al. [63], Simon et al. [67] and Popovici et al. [53].

More complex algorithms have been proposed in related fields to unveil the relevance of each input in NN models (see for example [14, 22] and [46]), which may be helpful in dam monitoring.

Therefore, even though NN and similar models must be interpreted with great care, their ability to extract information on the dam behaviour should not be underestimated.

3.3 Training and Validation Sets

It is common and convenient to divide the available data into two subsets: the training set is used to adjust the model parameters, whereas the validation set is solely used to measure the prediction accuracy[3]. In statistics, this need is

---

[3] the terminology is not universal; the data which are not used to fit the model is sometimes called test or prediction set.

well known, since it has been proven that the prediction accuracy of a predictive model, measured on the training data, is an overestimation of its overall performance [2]. Any subsetting of the available data into training and validation sets is acceptable, provided the data are independent and identically distributed (i.i.d.). This is not the case in dam monitoring series, which are time-dependant in general.

The amount of available data is limited, what in turn limits the size of the training and validation sets. Ideally, both should cover all the range of variation of the most influential variables. This is particularly relevant for the training set of the more complex models, as they are typically unable to produce accurate results beyond the range of the training data [21].

It is not infrequent that reservoir level follow a relatively constant yearly cycle by which situations from the lowest to the highest pool level are presented each year. Temperature, which is the second most influential variable on average, responds to a more defined annual cycle. As a consequence, many authors measure the size of the training and validation sets in years.

Moreover, dam behaviour models are used in practice to calculate the future response, on the basis of the observed, normal functioning, and draw conclusions about the safety state. Therefore, it seems reasonable to estimate the model accuracy with a similar scheme, i.e., to take the most recent data as the validation set. This is the procedure used in the vast majority of the reviewed papers (40/41), with the unique exception of Santillán et al. [64], who made a random division of the data.

Models based on the underlying physics of the phenomenon and those with fewer parameters (HST, IRF and MLR), are less prone to over-fitting. As a result, a higher value can be given to the training error. This is probably the reason why most studies do not consider a validation set, but rather use all the data for the model fit e.g. [7, 42] (Fig. 6a).

When a validation set is used, 10 % of the available data is reserved for that purpose on average. The higher frequency observed around 20 % corresponds to the papers dealing with the data from the 6th ICOLD Benchmark Workshop, where the splitting criterion was fixed by the organisers.

Tayfur et al. [75] reserved only one year for training, but explicitly mentioned that it contained all the range of variation of the reservoir level. Some authors proposed to set a minimum of 5 to 10 observations per model parameter to estimate [71].

A fundamental premise for the successful implementation of any prediction model is that the training data correspond to a period in which the dam has not undergone significant changes in its behaviour. In practice, it is not easy to ensure that this condition is fulfilled. While the history of major repairs and events is usually available, it is well known that the behaviour in the first years of operation usually corresponds to a transient state, which may not be representative of its response in normal operation afterwards [38]. Therefore, the use of data corresponding to the first period to adjust the model parameters may lead to an increase in prediction error. Lombardi [38] estimated that 12 years from dam construction are required for a data-based model to be effective.

This issue can be checked by analysing the training error: ideally, errors shall be independent, with zero mean and constant variance [71]. Some authors compute some of these values for evaluating the goodness of fit (e.g. [30, 34, 67]).

On another note, a minimum amount of data is necessary to build a predictive model with appropriate generalisation ability. De Sortis and Paoliani [17] run a sensitivity analysis of the prediction error as a function of the training set size. They concluded that 10 years were necessary for obtaining stable results. For their part, Chouinard and Roy [12] performed a similar work on a dam set. Provided that most of them were run-of-the-river small dams, which remained full most of the time, the thermal effect was the preponderant variable. As this is almost constant every year, 5 years of data were enough for most cases to achieve high accuracy.

According to the Swiss Comittee on Dams [71], a minimum of "5 yearly cycles" should be available, which suggests that they refer to filling-emptying cycles throughout a year (to account for the thermal variation). On the contrary, ICOLD [27] recommended to set thresholds as a function of the prediction error along "2 or 3 years of normal operation".

Salazar et al. performed a similar analysis for 14 instruments in an arch dam [61], and reported that the prediction accuracy was higher in some cases for models trained over the most recent 5 years of data (the maximum training set length was 18 years).

The size of the validation set ranges from 1 to 25 years (Fig. 6b), and depends on the amount of data available, rather than on the type of model.

Such verifications regarding the training and testing data sets are not performed in general in dam monitoring analysis, probably due to (a) the number of data available at a given time cannot be arbitrarily increased, and (b) the validation data shall be the most recent. In practice, there is not agreement on the appropriate criterion to define training and validation sets. Consequently, the comparison between models which predict different variables has limited reliability, although it was sometimes considered [56, 69].

Again, engineering judgement is essential to assess the appropriateness of the train and validation sets, as well as the model performance.
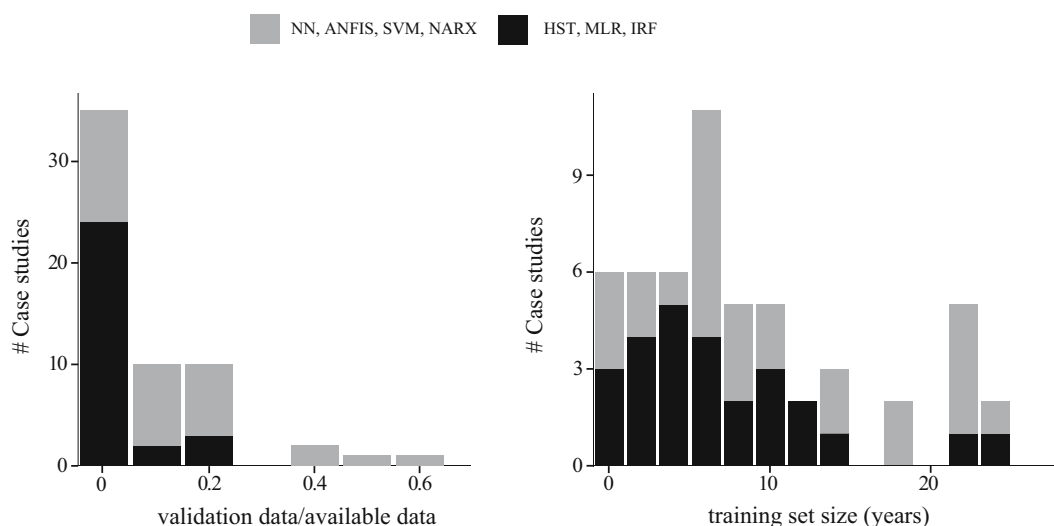
Fig. 6 Training and validation sets in the papers reviewed. *Left* ratio of validation data with respect to available data. *Right* training set size (years)

3.4 Missing Values

There are several potential sources of data incompleteness, such as insufficient measurement frequency [16, 42] or fault in the data acquisition system [41, 69]. Although there is a tendency towards increasing the quality of measurements and the frequency of reading, there are many dams in operation with long and low-quality monitoring data series to be analysed. According to Lombardi [38], only a small minority of the world population of dams feature adequate, properly-interpreted monitoring records. Curt and Gervais [16] showed the importance of controlling the quality of the data on which the dam safety studies are based, although they focused on proposing future corrective measures rather than on how to improve imperfect time series.

However, the vast majority of published articles overlooked this issue. They limited to the selection of some specific time period for which complete data series were available. For example, Mata et al. [42] only considered the period 1998–2002 for their analysis of the Alto Lindoso dam, due to the absence of simultaneous readings of displacements and temperatures in subsequent periods. In general, the need for simultaneous data of both the external variables and the dam response reduces the amount of data available for model fitting and limits the prediction accuracy.

If the missing values correspond to one of the predictors, these models are inapplicable, which limits their use in practice. If lagged variables are considered, there is also a need for equally time spaced readings. The above mentioned adaptive system proposed by Stojanovic et al. [69] can be applied in the event of failure of one or several devices.

Faults in the data acquisition process can also result in erroneous readings [36] which should be identified and eventually discarded or corrected. During model fitting, this would improve the model accuracy and increase its ability to interpret the dam response. Once a behaviour model is built, it can be used for that purpose [11].

Numerous statistical techniques have been developed to impute missing values. Their review is beyond the scope of this work, as they were not employed in the papers analysed. Moreover, their application should be tailored to the specific features of the problem, as well as to the nature of the variable in question. For example, missing values of air temperature can be reasonably filled from the average historical temperature for the period, or interpolated from available data [64]. By contrast, daily rainfall may change largely between consecutive readings, so that one missing value cannot be imputed with similar confidence.

3.5 Prediction Accuracy Measurement

It is important to appropriately estimate the prediction error of a model, since (a) it provides insight into its accuracy, (b) it allows comparison of different models, and (c) it is used to define warning thresholds.

There are various error measures to assess how well a model matches the observed data, among which the most commonly used are included in Table 3.

The result of using any of these indexes is frequently equivalent when referred to a given prediction task: the more accurate model will have a smaller *RMSE* value, but also the lowest *MSE*, and higher r and R^2. However, they also present differences which can be relevant, and are often not considered.

Provided that $MSE = (RMSE)^2$, they can be used indistinctly for model comparison. The only difference is that $RMSE$ can be compared to the target variable, given that both are measured in the same units. It should be noted that they are computed on the basis of the squared residuals, therefore they are sensitive to the presence of outliers, i.e., a few large prediction errors. In this sense, MAE could be considered a better choice, provided that it shares the advantage of $RMSE$ (it is measured in the same units as the output), and not its drawback. Mindful of this fact, both can be used interchangeably, if the analysis is complemented with a graphical exploration of the model fit, or other error measures.

The drawback to both MSE and $RMSE$ is that they are not suitable for comparing models fitting different variables, provided that they do not consider neither the mean nor the deviation of the output.

This limitation can be overcome by using the correlation coefficient r, since $r \in [-1, 1]$. On the contrary, it is not exactly an error rate, but rather an index of the strength of the linear relationship between observations and predictions. In other words, it indicates to what extent one variable increases as the other does, and vice versa. It can be checked that the value of r for a prediction calculated as $\hat{Y} = AY + B$ is equal to 1 for $A \neq 0$, while the error can be very large and will generally be non-zero (unless $A = 1$ and $B = 0$) [32]. As an example, Rankovic et al. [56] considered r and r^2, as well as MAE and MSE. While the results were similar for the training and validation sets in terms of r and r^2, both MAE and MSE were much greater in the validation set (as much as 7 times greater). These results may reflect some degree of over-fitting.

If r is used as a measure of goodness of fit, its value always increases with increasing number of model parameters (except in the highly unlikely event that the functions are completely independent of output). The R_{adj} coefficient can be used (e.g. [34, 69]) to account for the number of parameters of each model.

As an alternative, R^2, or its equivalent ARV can be chosen. They have the advantage over the correlation coefficient of being sensitive to differences in the means and variances of observations and predictions, while maintaining the ability to compare models fitted to different data [61].

Finally, it should be noted that the reading error of the devices (ε_r) may be relevant when predictions of variables of different nature are compared, although it is often ignored. It cannot be expected to obtain a model with an error below the measurement resolution [80]. Popovici et al. [53] reported that the overall accuracy of NN models was lower for tangential than for radial displacements, and attributed it to the lower range of variation of the former. It is possible that the reading error (which in principle should

be the same for tangential and radial displacements) were relevant in the first case and negligible in the second.

Salazar et al. found that models with relatively high ARV corresponded with very low MAE, close to ε_r [61].

Reading error should always be considered for evaluating model accuracy. One possibility would be to neglect the errors below that value before computing the prediction accuracy, by means of substituting $(y_i - F(x_i))$ by $|y_i - F(x_i)| - \varepsilon_r$, in the calculation of MSE, $RMSE$, r and R^2. Similarly, MAE could be computed as:

$$MAE^* = \frac{1}{N} \sum_{i=1}^{N} (|y_i - F(x_i)| - \varepsilon_r) \tag{21}$$

It is convenient to compute more than one error rate, especially if the aim is to compare models predicting variables of different kind. In addition, a graphical analysis of the error is highly advisable.

3.6 Practical Application

Despite the increasing amount of literature on the use of advanced data-based tools, very few examples described their practical integration in dam safety analysis. The vast majority were limited to the model accuracy assessment, by quantifying the model error with respect to the actual measured data. Only a few cases dealt with the interpretation of dam behaviour, by identifying the effect of each of the external variables on the dam response (e.g. [17, 35, 40]).

A detailed analysis of the results is always convenient [26], especially when complex models are employed. However, improvements in instrumentation and data acquisition systems allow the implementation of automatic warning generation schemes. The information provided by reliable automated systems, based on highly accurate models, can be a great support for decision making regarding dam safety [27, 31].

To achieve that goal, the outcome of the predictive model must be transformed into a set of rules that determine whether the system should issue a warning. In turn, these rules should be based on an overall analysis of the most representative instruments: a single value out of the normal-operation range will probably correspond to a reading error, if other instruments show no anomalies. However, the coincidence of out-of-range values in several devices may correspond to some abnormal behaviour. This is the idea behind the method proposed by Cheng and Zheng [11], which features a procedure for calculating normal operating thresholds ("control limits"), and a qualitative classification of potential anomalies: a) extreme environmental variable values, b) global structure damage, c) instrument malfunctions and d) local structure damage.

Table 3 Measures of accuracy

Error metric	Formula
Mean squared error	$MSE = \frac{\sum_{i=1}^{N}(y_i - F(x_i))^2}{N}$
Root mean squared error	$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - F(x_i))^2}{N}} = \sqrt{MSE}$
Mean absolute error	$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - F(x_i)|$
Correlation coefficient	$r = \frac{\sum_{i=1}^{N}(y_i - \bar{y})\left(F(x_i) - \overline{F(x_i)}\right)}{\left(\sum_{i=1}^{N}(y_i - \bar{y})^2\right)^{0.5}\left(\sum_{i=1}^{N}\left(F(x_i) - \overline{F(x_i)}\right)\right)^{0.5}}$
Coefficient of determination	$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - F(x_i))^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$
Standard error of estimate	$\sigma_\varepsilon = \sqrt{\frac{\sum(y_i - f(x_i))^2}{N}}$
Mean absolute percentage error	$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{y_i - F(x)}{y_i}\right|$
Maximum absolute error	$MaxAE = |\varepsilon|_{max}$
Adjusted R^2	$R^2_{adj} = R^2 - \left(1 - R^2\right)\frac{p}{N-p-1}$
Sum of squared error	$SSE = \sum_{i=1}^{N}(y_i - F(x_i))^2$
Average relative variance	$ARV = \frac{\sum_{i=1}^{N}(y_i - F(x_i))^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} = 1 - R^2$
Mean error	$ME = \frac{1}{N}\sum_{i=1}^{N}(y_i - F(x_i))$

p = number of parameters of the model. $\bar{\cdot}$ = mean

A more accurate analysis could be based on the consideration of the major potential modes of failure to obtain the corresponding behaviour patterns and an estimate of how they would be reflected on the monitoring data. Mata et al. [43] employed this idea to develop a methodology that includes the following steps:

- Identification of the most probable failure mode.
- Simulation of the structural response of the dam in normal and accidental situations (failure) by means of finite element models.
- Selection of the set of instruments that better identify the dam response during failure.
- Construction of a classification rule based on linear discriminant analysis (LDA) that labels a set of monitoring data as normal behaviour or incipient failure.

This scheme can be easily implemented in an automatic system. By contrast, it requires a detailed analysis of the possible failure modes, and their numerical simulation to provide data with which to train the classifier. Moreover, the finite element model must be able to accurately represent the actual behaviour of the dam, which is frequently hard to achieve.

4 Conclusions

There is a growing interest in the application of innovative tools in dam monitoring data analysis. Although only HST is fully implemented in engineering practice, the number of publications on the application of other methods has increased considerably in recent years, specially NN.

It seems clear that the models based on ML algorithms can offer more accurate estimates of the dam behaviour than the HST method in many cases. In general, they are more suitable to reproduce non-linear effects and complex interactions between input variables and dam response.

However, most of the papers analysed referred to specific case studies, certain dam typologies or determined outputs. More than a half of them focused on radial displacements in arch dams, although this typology represents roughly 5 % of dams in operation worldwide.

Moreover, the vast majority of articles overlooked the data pre-process. It is implicitly assumed that the monitoring data are free of reading errors and missing values, whereas that is not the case in practice. The development of criteria to fix imperfect data would allow to take advantage of a large amount of stored dam monitoring data.

An useful data-based algorithm should be versatile to face the variety of situations presented in dam safety: different typologies, outputs, quality and volume of data available, etc. Data-based techniques should be capable of dealing with missing values and robust to reading errors.

These tools must be employed rigorously, given their relatively high number of parameters and flexibility, what makes them susceptible to over-fit the training data. It is thus essential to check their generalisation capability on an adequate validation data set, not used for fitting the model parameters.

In this sense, most of the studies reviewed did not include an evaluation of the predictive model on an independent data set, and there are very few examples that used more than 20 % of the data for validation. This raises doubts about the generalisation capability of these models, in particular of those more strictly data-based, such as NN or SVM. It should be reminded that the main limitation of these methods is their inability to extrapolate, i.e., to generate accurate predictions outside the range of variation of the training data.

Before applying these models for predicting the dam response in a given situation, it should be checked whether the load combination under consideration lies within the values of the input variables in the training data set. Verifications of this kind were not reported in the reviewed papers, although they would provide insight into the reliability of the predictions.

From a practical viewpoint, data-based models should also be user-friendly and easily understood by civil engineering practitioners, typically unfamiliar with computer science, who have the responsibility for decision making.

Finally, two overall conclusions can be drawn from the review:

- ML techniques can be highly valuable for dam safety analysis, though some issues remain unsolved.
- Regardless of the technique used, engineering judgement based on experience is critical for building the model, for interpreting the results, and for decision making with regard to dam safety.

References

1. Amberg F (2009) Interpretative models for concrete dam displacements. In: 23th ICOLD congress, q91-R43
2. Arlot S, Celisse A et al (2010) A survey of cross-validation procedures for model selection. Stat Surv 4:40–79
3. Bishop CM (1995) Neural networks for pattern recognition. Oxford University Press, Oxford
4. Bonelli S, Félix H (2001) Delayed response analysis of temperature effect. In: Proceedings of the sixth ICOLD benchmark workshop on numerical analysis of dams, Salzburg, Austria
5. Bonelli S, Radzicki K (2007) The impulse response function analysis of pore pressures monitoring data. In: 5th international conference on dam engineering, Lisbon, Portugal
6. Bonelli S, Radzicki K (2008) Impulse response function analysis of pore pressure in earthdams. Eur J Environ Civ Eng 12(3):243–262
7. Bonelli S, Royet P (2001) Delayed response analysis of dam monitoring data. In: Proceedings of the fifth ICOLD European symposium on dams in a European context, Geiranger, Norway
8. Breitenstein F, Klher W, Widman R (1985) Safety control of the dams of the Glockner-Kaprun hydro-electric development. In: 15th ICOLD congress, pp 1121–1134, q56-R59
9. Carrère A, Noret-Duchêne C (2001) Interpretation of an arch dam behaviour using enhanced statistical models. In: Proceedings of the sixth ICOLD benchmark workshop on numerical analysis of dams, Salzburg, Austria
10. Chen BJ, Chang MW et al (2004) Load forecasting using support vector machines: a study on EUNITE competition 2001. IEEE Trans Power Syst 19(4):1821–1830
11. Cheng L, Zheng D (2013) Two online dam safety monitoring models based on the process of extracting environmental effect. Adv Eng Softw 57:4856
12. Chouinard L, Roy V (2006) Performance of statistical models for dam monitoring data. In: Joint international conference on computing and decision making in civil and building engineering, Montreal, pp 14–16
13. Chouinard L, Bennett D, Feknous N (1995) Statistical analysis of monitoring data for concrete arch dams. J Perform Constr Facil 9(4):286–301
14. Cortez P, Embrechts MJ (2011) Opening black box data mining models using sensitivity analysis. In: 2011 IEEE Symposium on computational intelligence and data mining (CIDM), IEEE, pp 341–348
15. Crépon O, Lino M (1999) An analytical approach to monitoring. Water Power Dam Constr 51(6):52–54
16. Curt C, Gervais R (2014) Approach to improving the quality of data used to analyse dams-illustrations by two methods. Eur J Environ Civ Eng 18(1):87–105
17. De Sortis A, Paoliani P (2007) Statistical analysis and structural identification in concrete dam monitoring. Eng Struct 29(1):110–120
18. Demirkaya S (2010) Deformation analysis of an arch dam using ANFIS. In: Proceedings of the second international workshop on application of artificial intelligence and innovations in engineering geodesy. Braunschweig, Germany, p 2131
19. Demirkaya S, Balcilar M (2012) The contribution of soft computing techniques for the interpretation of dam deformation. In: Proceedings of the FIG working week, Rome, Italy
20. Fabre J, Geffraye G (2013) Study and control of thermal displacements of Gage II dam (France) through the contribution of special heating and cooling devices. In: Proceedings of the seventh argentinian conference on dams, San Juan, Argentina, [in Spanish]
21. Flood I, Kartam N (1994) Neural networks in civil engineering. I: principles and understanding. J Comput Civ Eng 8(2):131–148
22. Gevrey M, Dimopoulos I, Lek S (2003) Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol Model 160(3):249–264
23. Govindaraju RS (2000) Artificial neural networks in hydrology II: hydrologic applications. J Hydrol Eng 5(2):124–137
24. Guedes Q, Coelho P (1985) Statistical behaviour model of dams. In: 15th ICOLD congress, pp 319–334, q56-R16
25. Hastie T, Tibshirani R, Firedman J (2009) The elements of statistical learning—data mining, inference, and prediction, 2nd edn. Springer, Berlin
26. Hill C, Sundaram M (2013) Instrumentation data collection, management and analysis. Tech. rep., United States Society on Dams (USSD) Committee on Monitoring of Dams and Their Foundations
27. International Commission on Large Dams (2012) Dam surveillance guide. Tech. Rep. B-158, ICOLD
28. Jang JS (1993) ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern 23(3):665–685
29. Jung IS, Berges M, Garrett JH, Kelly CJ (2013) Interpreting the dynamics of embankment dams through a time-series analysis of

piezometer data using a non-parametric spectral estimation method. In: Computing in civil engineering (2013), ASCE, p 2532, doi:10.1061/9780784413029.004

30. Kao C, Loh C (2013) Monitoring of long-term static deformation data of Fei-Tsui arch dam using artificial neural network-based approaches. Struct Control Health Monit 20(3):282–303

31. Legates DR, McCabe GJ (1999) Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resour Res 35(1):233–241

32. Léger P, Leclerc M (2007) Hydrostatic, temperature, time-displacement model for concrete dams. J Eng Mech 133(3):267277

33. Li F, Wang Z, Liu G (2013) Towards an error correction model for dam monitoring data analysis based on cointegration theory. Struct Saf 43:1220

34. Li F, Wang Z, Liu G, Fu C, Wang J (2014) Hydrostatic seasonal state model for monitoring data analysis of concrete dams. Struct Infrastruct Eng (ahead-of-print):1–16

35. Ljunggren M, Logan T, Campbell P (2013) Is your dam as safe as your data suggest? In: Proceedings of the NZSOLD/ANCOLD conference, vol I. Rotorua, New Zealand

36. Loh CH, Chen CH, Hsu TY (2011) Application of advanced statistical methods for extracting long-term trends in static monitoring data from an arch dam. Struct Health Monit 10(6): 587–601

37. Lombardi G (2004) Advanced data interpretation for diagnosis of concrete dams. Tech. rep., CISM

38. Lombardi G, Amberg F, Darbre G (2008) Algorithm for he prediction of functional delays in the behaviour of concrete dams. Hydropower Dams 3:111–116

39. Mata J (2011) Interpretation of concrete dam behaviour with artificial neural network and multiple linear regression models. Eng Struct 3(3):03–910. doi:10.1016/j.engstruct.2010.12.011

40. Mata J, de Castro AT, da Costa JS (2013) Time-frequency analysis for concrete dam safety control: correlation between the daily variation of structural response and air temperature. Eng Struct 48:658–665

41. Mata J, Tavares de Castro A, Sá da Costa J (2014a) Constructing statistical models for arch dam deformation. Struct Control Health Monit 21(3):423–437. doi:10.1002/stc.1575

42. Mata J, Leitão NS, de Castro AT, da Costa JS (2014b) Construction of decision rules for early detection of a developing concrete arch dam failure scenario. A discriminant approach. Comput Struct 142:45–53

43. Nedushan B (2002) Multivariable statistical analysis of monitoring data for concrete dams. PhD thesis, McGill University

44. Nourani V, Babakhani A (2012) Integration of artificial neural networks with radial basis function interpolation in earthfill dam seepage modeling. J Comput Civ Eng 27(2):183–195

45. Olden JD, Jackson DA (2002) Illuminating the black box: a randomization approach for understanding variable contributions in artificial neural networks. Ecol Model 154(1):135–150

46. On Large Dams IC (2000) Automated dam monitoring systems. guidelines and case histories. Tech. Rep. B-118, ICOLD

47. Opyrchal L (2003) Application of fuzzy sets method to identify seepage path through dams. J Hydraul Eng 129(7):546–548

48. Palumbo P, Piroddi L, Lancini S, Lozza F (2001) NARX modeling of radial crest displacements of the Schlegeis arch dam. In: Proceedings of the sixth ICOLD benchmark workshop on numerical analysis of dams, Salzburg, Austria

49. Panizzo A, Petaccia A (2009) Analysis of monitoring data for the safety control of dams using neural networks. In: New trends in fluid mechanics research, Springer, p 344347

50. Penot I, Daumas B, Fabre J (2005) Monitoring behaviour. Water Power Dam Constr 57(12):24–27

51. Perner F, Obernhuber P (2010) Analysis of arch dam deformations. Front Archit Civ Eng China 4(1):102–108

52. Piroddi L, Spinelli W (2003) Long-range nonlinear prediction: a case study. In: Proceedings of the 42nd IEEE conference on decision and control, vol 4. IEEE, pp 3984–3989

53. Popovici A, Ilinca C, Ayvaz T (2013) The performance of the neural networks to model some response parameters of a buttress dam to environment actions. In: Proceedings of the 9th ICOLD European club symposium, Venice, Italy

54. Ranković V, Grujović N, Divac D, Milivojević N, Novaković A (2012) Modelling of dam behaviour based on neuro-fuzzy identification. Eng Struct 35:107113. doi:10.1016/j.engstruct.2011.11.011

55. Ranković V, Grujović N, Divac D, Milivojević N (2014a) Development of support vector regression identification model for prediction of dam structural behaviour. Struct Saf 48:33–39

56. Ranković V, Novaković A, Grujović N, Divac D, Milivojević N (2014b) Predicting piezometric water level in dams via artificial neural networks. Neural Comput Appl 24(5):1115–1121

57. Restelli F (2010) Systemic evaluation of dam monitoring using PCA. In: Proceedings of the sixth argentinian conference on dams, Neuquén, Argentina, [in Spanish]

58. Restelli F (2013) Systemic evaluation of the response of large dams instrumentation. Application at El Chocón dam. In: Proceedings of the 9th ICOLD European Club Symposium, Venice, Italy

59. Riquelme F, Fraile J, Santillán D, Morán R, Toledo M (2011) Application of artificial neural network models to determine movements in an arch dam. In: Proceedings of the 2nd international congress on dam maintenance and rehabilitation. Zaragoza, Spain, pp 117–123

60. Ruiz H (2013) Fisher networks: a principled approach to retrieval-based classification. PhD thesis, Liverpool John Moores University

61. Salazar F, Toledo M, Oñate E, Morán R (2015) An empirical comparison of machine learning techniques for dam behaviour modelling. Struct Saf 56:9–17. doi:10.1016/j.strusafe.2015.05.001

62. Sánchez Caro FJ (2007) Dam safety: contributions to the deformation analysis and monitoring as an element of prevention of pathologies of geotechnical origin. PhD thesis, UPM, [In Spanish]

63. Santillán D, Fraile-Ardanuy J, Toledo M (2013) Dam seepage analysis based on artificial neural networks: The hysteresis phenomenon. In: The 2013 international joint conference on neural networks (IJCNN), IEEE, pp 1–8

64. Santillán D, Fraile-Ardanuy J, Toledo M (2014) Seepage prediction in arch dams by means of artificial neural networks. Water Technol Sci V(3), 81–96. [in Spanish]

65. Saouma V, Hansen E, Rajagopalan B (2001) Statistical and 3d nonlinear finite element analysis of Schlegeis dam. In: Proceedings of the sixth ICOLD benchmark workshop on numerical analysis of dams, pp 17–19

66. Silva Gomes AF, Silva Matos D (1985) Quantitative analysis of dam monitoring results. State of the art, applications and prospects. In: 15th ICOLD congres, pp 319–334, q56-R39

67. Simon A, Royer M, Mauris F, Fabre J (2013) Analysis and interpretation of dam measurements using artificial neural networks. In: Proceedings of the 9th ICOLD European club symposium, Venice, Italy

68. Smola AJ, Schlkopf B (2004) A tutorial on support vector regression. Stat Comput 14(3):199222

69. Stojanovic B, Milivojevic M, Ivanovic M, Milivojevic N, Divac D (2013) Adaptive system for dam behavior modeling based on linear regression and genetic algorithms. Adv Eng Softw 65:182190

70. Su H, Wu Z, Wen Z (2007) Identification model for dam behavior based on wavelet network. Comput Aided Civ Infrastruct Eng 22(6):438–448

71. Swiss Committee on Dams (2003) Methods of analysis for the prediction and the verification of dam behaviour. Tech. rep., ICOLD
72. Takagi T, Sugeno M (1985) Fuzzy identification of systems and its applications to modeling and control. IEEE Trans Syst Man Cybern 1:116–132
73. Tatin M, Briffaut M, Dufour F, Simon A, Fabre J (2013) Thermal displacements of concrete dams: finite element and statistical modelling. In: Proceedings of the 9th ICOLD European club symposium, Venice, Italy
74. Tatin M, Briffaut M, Dufour F, Simon A, Fabre JP (2015) Thermal displacements of concrete dams: accounting for water temperature in statistical models. Eng Struct 91:26–39
75. Tayfur G, Swiatek D, Wita A, Singh VP (2005) Case study: finite element method and artificial neural network models for flow through Jeziorsko earthfill dam in Poland. J Hydraul Eng 131(6):431440
76. Willm G, Beaujoint N (1967) Les méthodes de surveillance des barrages au service de la production hydraulique d'Electricité de France-Problèmes anciens et solutions nouvelles. In: 9th ICOLD Congres, pp 529–550, q34-R30. [in French]
77. Xu C, Yue D, Deng C (2012) Hybrid GA/SIMPLS as alternative regression model in dam deformation analysis. Eng Appl Artif Intell 25(3):468475
78. Xu H, Li X (2012) Inferring rules for adverse load combinations to crack in concrete dam from monitoring data using adaptive neuro-fuzzy inference system. Sci China Technol Sci 55(1):136141
79. Yao Y, Sharma A, Golubchik L, Govindan R (2010) Online anomaly detection for sensor systems: a simple and efficient approach. Perform Eval 67(11):1059–1075
80. Yu H, Wu Z, Bao T, Zhang L (2010) Multivariate analysis in dam monitoring data with PCA. Sci China Technol Sci 53(4):1088–1097. doi:10.1007/s11431-010-0060-1
81. Zenz G, Obernhuber P (2001) ICOLD benchmark workshops on dam safety. Hydropower Dams 8:75–78