ORIGINAL PAPER

Fuzzy algorithm for estimating average breach widths of embankment dams

Hasan G. Elmazoghi

Abstract Dam breach width significantly influences peak breach outflow, inundation levels, and flood arrival time, but uncertainties inherent in the prediction of its value for embankment dams make its accurate estimation a challenging task in dam risk assessments. The key focus of this paper is to provide a fuzzy logic (FL) model for estimating the average breach width of embankment dams as an alternative to regression equations (RE). The FL approach is capable of handling nonlinear behavior, imprecision in discrete measurements, and parameter uncertainty. Historical data from 69 embankment dam failures are used in the development and testing of the FL model. Application of the FL model is also presented for estimating average breach widths of two case studies that have adequately documented data. The accuracy of the FL rule-based model is investigated using uncertainty analysis: the mean prediction error between the FL estimates and the observed average breach widths is very small (=0.03) and comparable to that achieved using the best available RE. Moreover, the FL uncertainty band is found to be approximately ±0.51 order of magnitude smaller than the ±0.56 order of magnitude achieved with the best available RE. The simulation results indicate the potential of the FL model to be used as a predictive tool for estimating the average breach width of embankment dams.

List of symbols

B_{av}	Observed average breach width (m)
$\hat{B}_{av}$	Estimated average breach width (m)
C_b	A coefficient that depends on the reservoir volume
$\bar{e}$	Mean prediction error
e_i	Individual prediction errors, log cycles
h_b	Breach height (m)
h_d	Dam height (m)
h_w	Depth of water above breach invert at time of failure (m)

H. G. Elmazoghi (✉)
Civil Engineering Department, Faculty of Engineering, University of Benghazi, Benghazi, Libya
e-mail: hasan.elmazoghi@benghazi.edu.ly

K_o 1.4 if there is overtopping, else 1 (in Froehlich 1995 equation)
K_o 1.3 if there is overtopping, else 1 (in Froehlich 2008 equation)
S Reservoir storage (m^3)
S^* Nondimensional reservoir storage $= (S/h_b^3)$
S_e Standard deviation of the errors
S_{MAD} Estimator of scale derived from the median of the absolute deviations analogous to standard deviation
T Median of the errors, an estimator of location
V_w Volume of water stored above breach invert at time of failure (Mm^3)
Z_i Standardized error

1 Introduction

Embankment dams (earthfill or rockfill) are the most commonly built type of dams in many countries. Their construction is preferred under certain circumstances, especially when sufficient materials are available near the dam site, the foundation is pervious, and the ratio of dam length to height is high. The main problem facing this kind of dam is piping or overtopping that may cause erosion of materials and ultimately breaching of the embankment. Accurate estimations of breach characteristics are needed as a basis in dam risk assessments. In order to carry out an embankment failure analysis, the average breach width in the dam is one of the key parameters that should be accurately estimated because it influences the severity of failure and affects the magnitude of the peak discharge. Singh and Snorrason (1984) used DAMBRK and HEC-1 models on 8 hypothetical breached dams and assessed that changes in breach width were more significant for large dams because it produced larger changes (35–87 %) in peak outflow than for smaller reservoirs (6–50 %). The breach shape of an embankment dam is assumed to vary from triangular to trapezoidal as the breach progresses (Wahl 1998). Historic embankment failure data report either the breach width at the top and bottom of the breach section or simply the average breach width resulting from passage of the complete breach hydrograph. The average breach width (B_{av}) is one-half the sum of the trapezoid top and bottom widths. Methods of estimating B_{av} are based on either case study data from past dam failures or physically based numerical models. Case study methods include parametric models, regression equations (RE), and analysis by comparison. Most of these methods are based on small dams having heights of <15 m. Parametric models (e.g. NWS DAMBRK by Fread 1977) use empirical observations of previous dam failures to develop the outflow hydrograph. Physically based models such as BREACH (Fread 1988), BEED (Singh and Scarlatos 1985), and FLDWAV (Fread 1993) relay on sediment erosion and water flow formulas and generally suffer from insufficient understanding of breach development (Wahl 1998). However, recent researches considering hydrodynamics and soil mechanics for embankment erosion are offering opportunities to better understand and model breach mechanisms (e.g. Froehlich 2004; Hanson et al. 2005). In practice, the most widely applied methods to predict B_{av} are based on regression analysis of recoded data from embankment dam failures, for example, the Bureau of Reclamation (1988), Von Thun and Gillette (1990), and Froehlich (1995). RE provide simple and convenient algorithms, especially in cases of strong linear relationships between the input and output variables and when detailed simulations are not required. The results of the available RE vary widely depending on the

assumptions and subsets of data used in their formulation and internal uncertainties that are not taken explicitly into consideration. The linear regression approach assumes that the scatter of points around the best-fit line is approximately Gaussian and has the same standard deviation all along the line, the data points are independent of one another, and any imprecision in measuring the values of the independent variables is very small compared to the variability in the values of the dependent variable. If these assumptions are violated, then the linear regression approach leads to biased relationships. In practice, the breach extent depends on the embankment and the reservoir geometry in addition to other factors such as embankment material, type of protective cover, and mode of failure. Many of the available RE assumed the breach width as a linear function of only the dam (or the breach) height and/or the reservoir volume. This assumption may be valid for small embankments having similar geometry and soil characteristics. As the material changes, more uncertainties become included in the overall breaching process. In the regression line, fitting such a distinction is not considered and each point in the scatter diagram is treated equally in fixing the best straight line. Uncertainty is also included in determining the reservoir water volume and the breach height at the time of failure that will be used to predict the breach width. It is not possible to consider such variations in the coefficients through regression analysis. The fuzzy approach, on the other hand, can provide an alternative methodology for considering such uncertainties through vaguely defined membership functions. Fuzzy logic (FL) modeling is effectively utilized in applications ranging over perhaps all branches of engineering. However, there is currently no solution to predict B_{av} using this technique. A FL model is a logical-mathematical procedure based on linguistic variables (e.g., low, high, wide, etc.) and a system of IF–THEN rules that mimics the human way of thinking in computational form: an overall process called fuzzy inference. The idea behind fuzzy inference is to interpret (fuzzify) the crisp values of the system variables to express them in linguistic terms and, based on a set of IF–THEN rules, to assign values to the output vector. For each rule, inferencing looks up some membership values according to the condition of the rule and an implication method is employed to combine the IF and the THEN parts. The outputs of each rule are aggregated to produce a single fuzzy set that must be converted (defuzzified) to a crisp number representing the desired output. More details about the FL approach *will* be given in *subsequent sections*.

The main objective of the present paper is the development of a FL model to predict B_{av} as an alternative to the RE. This general objective includes: (1) developing a Mamdani's FL inference system for predicting B_{av}, (2) comparing the results of the FL model with those of the best RE for predicting B_{av}, (3) performing an uncertainty analysis for the results of the FL model and the RE, and (4) applying the FL model on some of the available case studies that have adequate data.

The available case studies of embankment dam failures presented by Froehlich (2008) constitute the basis for the development of this FL rule-based model. The FL system becomes more complex when many inputs and outputs are chosen for a single implementation. This adds to the difficulty of building the IF–THEN rules that control the system. For this reason, the present FL system is selected to have two inputs and one output. Since the height of water above the breach invert (breach base) at the dam at time of failure (h_w) and the volume of water stored above the breach invert at time of failure (V_w) have been documented and used in a number of prediction equations, they are taken as the input parameters in the development of the present FL model for estimating B_{av}. Even with these parameters, there is vagueness in the data (imprecision in discrete measurements and parameter uncertainty), FL set theory is especially well suited for handling such vagueness. The construction of the conditional statements 'IF–THEN rules' and the

membership functions, based on the observed measurements and the relation between the input and the output variables, are the things that make FL useful and capable of representing the actual system behavior.

2 Review of available approaches

Several physically based models are available in literature to simulate the breach of embankment dams. The majority of those models are based on different erosion and sediment transport formulas that in turn assume different flow conditions (quasi-steady state or unsteady-state that may lead to numerical instability). Some researchers used 1-D cross-section-averaged flow models (e.g., Cristofano 1965; Brown and Rogers 1977; Ponce and Tsivoglou 1981; Fread 1984; Visser 1998; Hanson et al. 2005), while others used 2-D depth average flow models (e.g., Froehlich 2004). Although physically based models can provide better understanding and extensive information about the breach, they are complex, require several assumptions and inputs, make use of empirical coefficients to describe material and flow resistance, and the results of some of them do not adequately simulate observed case studies. For more practical and easily applied models, many researchers gathered detailed case studies of breached embankment dams and developed expressions to predict the characteristics and consequences of the breach. From those studies, Johnson and Illes (1976) were the first to predict the breach shapes for earth fill dams assuming that the breach begins as a triangle and ends as a trapezoid. They recommended a range for the breach width as a linear function of the dam height (h_d). Similarly, Singh and Snorrason (1984) plotted the breach widths versus dam heights for 20 case studies and stated a range for the breach width as a linear function of h_d. MacDonald and Langridge-Monopolis (1984) used 42 case studies and suggested that the breach shape could be trapezoidal or triangular depending on whether the breach has reached the bottom of the dam or not. The Federal Energy Regulatory Commission (FERC 1987) also proposed a range for the breach width as a function of h_d. Froehlich (1987) used nondimensional analysis and developed an equation that estimates the average breach width as a function of the non-dimensional reservoir storage (S^*). Froehlich realized that overtopping causes the most breach extension, which erodes at a higher rate than by any other mode of failure. The Bureau of Reclamation (1988) developed an equation for breach width of earthen dams depending on the height measured from the initial reservoir water level to the breach bottom elevation. This equation assumes a linear relationship between B_{av} and h_w. Von Thun and Gillette (1990) used the data of MacDonald and Langridge-Monopolis (1984) and Froehlich (1987) and proposed a relation for estimating B_{av} knowing the depth of water at the dam at time of failure (h_w) and a coefficient (C_b) that depends on the reservoir storage. Later on, Froehlich (1995) published a revised equation that has better estimated coefficients to predict B_{av}. The independent variables in this equation are the volume of water stored above the breach invert at time of failure (V_w), the breach height (h_b), and a factor (K_o) that accounts for failure mode. Wahl (1998, 2004) provided a summary of the available RE for predicting the breach width, performed an uncertainty analysis, and compared state-of-the-art prediction equations. Wahl stated that Froehlich's (1995) equation had the best prediction performance for cases with observed breach widths <50 m. In 2008, Froehlich proposed another equation that will likely be accurate enough in application to estimate B_{av} as a function of $V_w^{1/3}$ and K_o. Most of the previous RE relate B_{av} to one or more characteristics of the dam and reservoir at failure, such as h_w (some investigators used h_d or h_b) and V_w or

combination of the two. Based on the above cited researches, it can be inferred that models based on conventional mathematical tools (e.g., regression) require several assumptions to deal with nonlinear and uncertain systems. Hence, application of FL modeling offers an alternative that allows the modeler to include imprecise data and parameters without the need for any assumption.

3 Data used in developing the FL model

The ability of the FL rule-based model to accurately predict B_{av} depends upon the amount of the available historical data and experience in 'IF–THEN' rules construction. Generally, B_{av} is one of the most documented breach parameters in dam failure case studies. Froehlich (2008) presented tabulated information about 74 breached embankment case studies compiled from numerous sources in the literature. The development of the present FL model is mainly based on this information. Since the breach width is missing from four case studies and the reservoir volume is missing from another, the data set that is used in the FL model and the RE consists of the measured values of V_w, h_w, h_b and B_{av} given in the remaining 69 case studies. This data set is subdivided into two sets without any special selection process. The bigger set, consisting of 51 case studies, is used in the training phase of the FL model. The smaller set, consisting of 18 case studies, is used in the testing phase of the FL model. In order to save the space in this paper, these sets will be presented later in the training and testing phases of the FL model. The 'estimated' average breach widths are presented under the symbol $\hat{B}_{av}$. The developed FL model is also applied to two embankment dam failures cited in recent publications (i.e., Jamestown and Big Bay dams).

4 Methodology (fuzzy sets, membership functions, and FL inference system)

As classical logic is based on classical set (crisp set) theory, FL is based on fuzzy set theory. An element x in a crisp set A from a given universe X may be defined by a membership function μ_A where $\forall_{x \in X} \mu_A(x) = 1 \Leftrightarrow x \in A$ and $\mu_A(x) = 0 \Leftrightarrow x \notin A$. This notation shows that crisp sets allow only full membership or no membership at all. Zadeh (1965) stated that in many systems, very precise numerical inputs are not always required, yet highly acceptable outputs are feasible. For this reason, Zadeh introduced the concept of the fuzzy set by defining partial membership, thus extending the degree of membership from {0, 1} to the continuous interval [0, 1]. A fuzzy set $\underline{A}$ in a universe of discourse U may be represented by a set of ordered pairs consisting of a generic element x and its degree of membership to the fuzzy set $\underline{A}$, that is, $\underline{A} = \{(x, \mu_{\underline{A}}(x)) : x \in U, \mu_{\underline{A}}(x) \in [0, 1]\}$. The notation $\underline{A} = \{(x_1, 0.8); (x_2, 0.25); (x_3, 0)\}$, for example, denotes the fact that elements x_1 and x_2 belong to a corresponding degree (0.8 and 0.25) to the fuzzy set $\underline{A}$, while element x_3 does not belong to $\underline{A}$. The function that defines to what degree an element x belongs to the fuzzy set $\underline{A}$ is called the membership function, which is essentially a curve that can be selected as triangular, trapezoidal, or Gaussian, etc. The shape is generally less important than the number of curves and their placement. From 3 to 7 curves are generally appropriate to cover the required range of an input or an output variable (Ross 1995). In order to process the input to get the output reasoning, there are three main steps involved in the development of a FL inference system: fuzzification, rules processing, and defuzzification. The fuzzification step allows the system inputs and outputs to be expressed in

linguistic terms (e.g., *low*, *medium*, *high*, etc.). Using membership functions, crisp input data are converted (fuzzified) against the appropriate linguistic fuzzy terms thus generating membership degrees for a fuzzy variable. The degree(s) of membership of an input parameter is determined by plugging the value of that input into the horizontal axis (universe of discourse) of the desired variable, projecting vertically to the upper boundary of the membership function(s) and then reading the corresponding membership degree(s) from the vertical axis representing the continuous interval [0, 1]. The output also has a set of membership functions that define the possible responses and outputs of the system. The rules processing step calculates the response from the system inputs according to the constructed rule base under which the FL system will operate. The rule base consists of IF–THEN conditional statements that relate the input to the desired output. The IF-part of a rule is called the antecedent, while the THEN-part is called the consequent. The fuzzified inputs are applied to the antecedents of the fuzzy rules. In case the fuzzy rule has multiple antecedents, the fuzzy AND or OR logical operators are used in order to obtain 'a single number' that represents the result of the antecedent for that rule. For most applications, the fuzzy membership function for a given rule with the AND operator is obtained with the minimum implication as proposed by Mamdani ([1977](#)) and given as: $\mu_{A \cap B}(x) = \min[\mu_A(x), \mu_B(x)]$. This single number will then be used in shaping (clipping) the consequent membership function (implication from the antecedent to the consequent). Implication occurs for each rule. The outputs of each rule are then unified (aggregated) into a single fuzzy set. Thus, the input of the aggregation process is the list of clipped consequent membership functions and the output is a single fuzzy set. The defuzzification step then chooses the desired output crisp number from this aggregate fuzzy set. However, there are several defuzzification methods (centroids, bisectors, middle of maximum, etc.). The most popular method is the centroid calculation, which returns the center of area under the curve of the aggregate fuzzy set, thereby moving from a fuzzy set to a crisp number representing the desired output (Ross [1995](#)).

5 Development and implementation of the present FL inference system

In order to develop a FL inference system, it is essential to decide the inputs, outputs, each of their domains, input and output membership functions, overlap between these functions, fuzzy inference rules, implication and aggregation methods, and the defuzzification method. The inputs to the present FL model are h_w and V_w, while the output is the estimated B_{av}. The ranges "universe of discourse" of the fuzzy input and output subsets in the present study are identified based on the available data of 69 breached embankment dams taken from Froehlich ([2008](#)). Statistical analysis of this data shows that there are no obvious trends between the input variables and the output, and the correlation coefficient between each of h_w, V_w, h_b, and B_{av} is around 0.58. This enhances the applicability of the FL model for the considered database to give better results for estimating B_{av} in comparison with the RE. That is because FL modeling can control nonlinear systems that are difficult to model mathematically, does not require any assumption of linearity, and are capable of handling imprecision in discrete measurements and parameter uncertainty. The numbers and ranges of the fuzzy subsets of the input and output variables are proposed after classification of their data (i.e., arranging the data of each variable in an ascending order, considering only one data point from those values having similar magnitudes and drawing the ascending bar chart of the remaining values of that variable). Figure [1](#)a, for example, is drawn after classification of the available heights of water (h_w). Similarly,
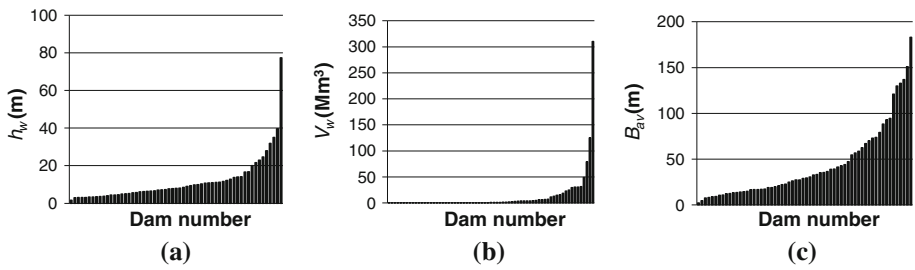
Fig. 1 Distribution of input and output variables: **a** for input variable h_w, **b** for input variable V_w, and **c** for output variable B_{av}

Fig. 1b, c is drawn using the volumes of water (V_w) and the average breach widths (B_{av}), respectively.

The horizontal axis in each of these figures represents the names of the embankments, which are omitted to save space. Such data distribution may help the modeler to get an overview of the data and identify obvious separations or clusters to be used in creating the possible number of fuzzy subsets with their ranges for the input and output variables. The proposed fuzzy sets and their spans for the input and output variables are given in Table 1.

The term set of the first input variable (h_w) is selected as {short (SH), medium (M), high (H)} in the universe of discourse [1.68–77.4] in meters. The term set of the second input (V_w) is {very low (VL), low (L), low medium (LM), medium (M), high medium (HM), high (H)} in the universe of discourse [0.0139–310], which represents the minimum and maximum storage volumes above breach invert at time of failure in Mm³. Similarly, the term set of the output variable (B_{av}) is selected as {very tight (VT), tight (T), medium (M), wide (W), very wide (VW)} in the universe of discourse [2.29–183] in meters. The present FL model for estimating the average breach width is implemented in MATLAB 6.5.1 (ITU Computer Center) using Mamdani's (1977) inference method, which expects the output to be fuzzy sets. The MATLAB fuzzy toolbox allows for the creation of input membership functions, fuzzy rules, and output membership functions through a graphical user interface. Based on the data of Table 1, the membership functions of the input and output variables are drawn using the Membership Function Editor. The membership functions should overlap to allow smooth mapping of the system. Straight line functions (e.g., triangular membership functions) are commonly used because they are the simplest and easiest to implement. In the present model, triangular membership functions are selected for the input variables (h_w and V_w) and the output variable (B_{av}), as shown in Fig. 2a–c, respectively. In order to easily repeat the process by other users, it is preferred to show these figures as originally obtained from the framework of MATLAB. It is to be noted that the fuzzy terms of some membership functions (e.g., VL, L and LM in Fig. 2b) are written over each other by MATLAB because of their very small ranges.

Each particular input or output can be interpreted from such fuzzy sets and a degree of membership is read. For example, using Fig. 2a, an input with $h_w = 10$ m will have about a 15 % membership in the SH function and about a 35 % membership in the M function. Once the input and output membership functions are defined, the fuzzy rules that control the system can now be prepared using the Rule Editor. The fuzzy rules are in the form of IF–THEN statements that look at both inputs (h_w and V_w) and determine the desired output ($\hat{B}_{av}$). For each output, several different rules will usually be used since the inputs will usually be in more than one membership function. For the present FL model, the rule base

Table 1 Fuzzy terms and corresponding ranges of input and output variables

Input no. 1: h_w (m)

Fuzzy terms	Short	Medium	High
Fuzzy symbols	SH	M	H
Fuzzy range	1.68–12.30	6.50–27.30	25.50–77.40

Input no. 2: V_w (Mm^3)

Fuzzy terms	Very low	Low	Low medium	Medium	High medium	High
Fuzzy symbols	VL	L	LM	M	HM	H
Fuzzy range	0.01–0.33	0.29–1.60	1.10–6.35	5.50–31.20	26.20–92.70	55.00–310.00

Output: B_{av} (m)

Fuzzy terms	Very tight	Tight	Medium	Wide	Very wide
Fuzzy symbols	VT	T	M	W	VW
Fuzzy range	2.29–28.60	13.80–32.40	29.50–51.00	38.50–118.00	91.00–183.00

contains all linguistic rules that are extracted from the available numeric data processed in Excel worksheets and provided by experience as follows:

1. Type the values of h_w, V_w, and B_{av} in an Excel worksheet in vertical order in three separate columns (say columns B, D, and F).
2. Fix a particular function in the in between columns (i.e., columns C, E, and G) in front of the values of each variable in order to produce linguistic variables for each value depending on the boundaries of the previously selected membership functions. Suppose that the values of h_w are entered in the cells B2:B52, then fix the following function at the cell C2 and scroll down to the cell C52.

$$= IF(AND(B2 \geq 1.68, \ B2 \leq 12.3), \text{``SH''}, \ IF(AND(B2 > 6.5, \ B2 \leq 27.3),$$
$$\text{``M''}, \ IF(AND(B2 > 25.5, \ B2 \leq 77.4), \text{``H''})))$$

3. Repeat step 2 for V_w and B_{av} in order to produce linguistic variables for each of them. Recognize that the functions for V_w and B_{av} will have different forms depending on the selected boundaries and number of membership functions for each.

By doing so, and looking at the resulting linguistic variables line by line (i.e., for each embankment dam), one can build an idea concerning the possible IF–THEN rules that may control the available numeric data. This can be achieved by selecting those rules that are most repeated for each category of linguistic variables. However, one may find that some of the rules contradict each other, that is, although the linguistic variables in the 'IF' part of some rules are identical, the output linguistic variables in the 'THEN' part of the same rules are very different. For such reasons, the modeler has to modify some of the rules and exclude others from the whole list, taking into consideration the overlap between the membership functions, the possibility that some data may not trigger the rules, and the strength of some rules. This process depends on trial and error and experience and becomes tedious and difficult in the case of a very large number of rules. The full set of rules in the
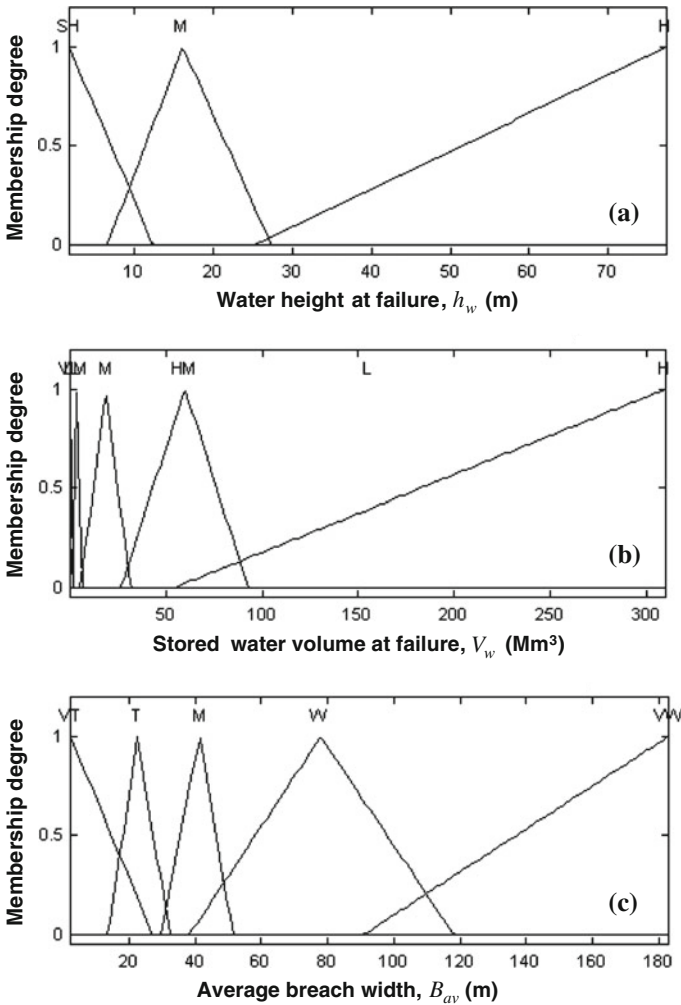
Fig. 2 Membership functions: **a** for input variable h_w, **b** for input variable V_w, and **c** for output variable B_{av}

present FL model is stated as shown in Table 2. The number of rules in this study is 18, where the 'AND' fuzzy operator is used in the formation of the antecedents of those rules. The fuzzy membership function for a given rule using the 'AND' fuzzy operator is obtained with the minimum implication as proposed by Mamdani (1977). It determines the degree to which the antecedent is satisfied for each rule. This is illustrated in Fig. 3, which assumes that rules 5 and 6 are activated when the two inputs $h_w = 5$ m and $V_w = 80$ Mm^3 are considered. If it is assumed that the AND operator (min at work) evaluating the antecedent of rule 5 (h_w is SH and V_w is HM) yielded the fuzzy membership degrees 0.6 and 0.4, respectively, then the fuzzy AND operator simply selects the minimum of the two values, 0.4, and the fuzzy operation for that rule is complete. This minimum value, 0.4, is then used in clipping the output membership function (B_{av} is W) as shown in Fig. 3.

Each activated rule will provide a particular clipped membership function for the output variable; see also the output of rule 6. The clipped membership functions of all the

Table 2 List of fuzzy rules

| Rule no.	Rule
1	If h_w (m) is SH and V_w (Mm^3) is VL then B_{av} (m) is VT
2	If h_w (m) is SH and V_w (Mm^3) is L then B_{av} (m) is T
3	If h_w (m) is SH and V_w (Mm^3) is LM then B_{av} (m) is M
4	If h_w (m) is SH and V_w (Mm^3) is M then B_{av} (m) is W
5	If h_w (m) is SH and V_w (Mm^3) is HM then B_{av} (m) is W
6	If h_w (m) is SH and V_w (Mm^3) is H then B_{av} (m) is VW
7	If h_w (m) is M and V_w (Mm^3) is VL then B_{av} (m) is VT
8	If h_w (m) is M and V_w(Mm^3) is L then B_{av} (m) is T
9	If h_w (m) is M and V_w (Mm^3) is LM then B_{av} (m) is M
10	If h_w (m) is M and V_w (Mm^3) is M then B_{av} (m) is W
11	If h_w (m) is M and V_w (Mm^3) is HM then B_{av} (m) is W
12	If h_w (m) is M and V_w (Mm^3) is H then B_{av} (m) is VW
13	If h_w (m) is H and V_w (Mm^3) is VL then B_{av} (m) is T
14	If h_w (m) is H and V_w (Mm^3) is L then B_{av} (m) is T
15	If h_w (m) is H and V_w (Mm^3) is LM then B_{av} (m) is VW
16	If h_w (m) is H and V_w (Mm^3) is M then B_{av} (m) is W
17	If h_w (m) is H and V_w (Mm^3) is HM then B_{av} (m) is VW
18	If h_w (m) is H and V_w (Mm^3) is H then B_{av} (m) is VW
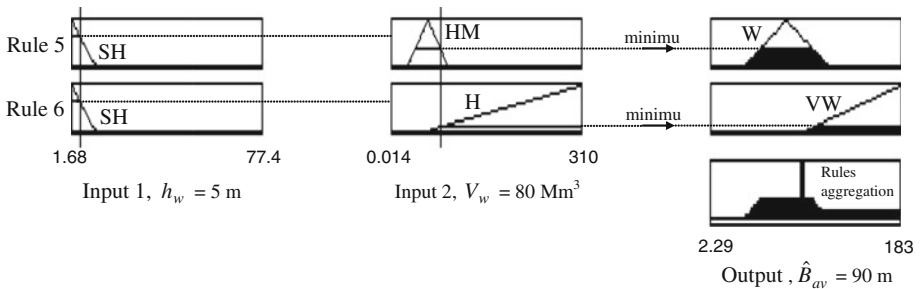
Fig. 3 Schematic representation showing the FL inference process

activated rules are combined together (aggregated) into a single fuzzy set. In Fig. 3, the process proceeds from the inputs in the upper left, then across each rule line by line and then down the rule outputs to finish with a single fuzzy set in the lower right that should be defuzzified, or resolved to a single number representing the desired output. This can be done in many ways, but the most common method used is the center of gravity of that single set. If one value (singleton position, B_{av_i}) is associated with each level ($i = 1, \ldots, n$) in the single fuzzy set of the output instead of a range of values, the centroid ($\hat{B}_{av}$) can be approximated by $\hat{B}_{av} = \sum_{i=1}^{n} B_{av_i} \mu(B_{av_i}) / \sum_{i=1}^{n} \mu(B_{av_i})$.

6 Results of the training phase of the FL model

In the training phase of the FL model, more work needs to be done before it becomes available in its final form for estimating the average breach width. The model should be

trained with the measured data through several simulations. In each simulation, one should evaluate the results and tune the model until satisfactory outputs are obtained. Tuning the model can be done by changing some of the rules consequents or strengths, changing the centers of the input and/or output membership functions, or slightly adjusting the ranges of the input and/or output membership functions. Table 3 shows the estimated and calculated average breach widths resulting from the final accepted training phase of the present FL model and some of the available RE. From Table 3, it is seen that in general the observed 'measured' average breach widths are successfully estimated by implementing the present FL approach. Moreover, the FL estimates are either close to the observed values or to those calculated by Froehlich's (1995) regression equation. Since breach formation depends on numerous factors (material properties, construction conditions, protective cover, etc.), the available data show contradiction in the outputs at some dams, although they have approximately the same height and the reservoir volume as inputs.

7 Results of the testing phase of the FL Model

For the testing phase of the developed FL model, a subset of eighteen dams is randomly selected from the database (Table 4) and used as a means of providing a comparison for assessing the performance of the FL model. Data from the same subset are also tested in some of the available RE for predicting average breach width. Table 4 summarizes the simulation results of this phase. From the values presented in Table 4, it is seen that the FL model gives comparable estimates to the observed average breach widths. The other notable feature in Table 4 is that the fuzzy estimates in most of the cases are approximately equal to those calculated by Froehlich's (1995) regression equation.

Figure 4a, b presents visual inspection of the results in order to compare the relations of the estimates of the FL model and Froehlich (1995) equation with case study data as obtained from the training and testing phases, respectively. While there is a reasonable match between the estimates of the FL model and the Froehlich (1995) equation at some dam sites (e.g., Hatchtown-Utah, Lake Avalon-N.M., Prospect-Colo., Wheatland No. 1-Wyo. in the training phase and Butler-Ariz., Castlewood-Colo., Coedty-UK, Fogelman-Tenn. in the testing phase), these estimates are different from the observed values. For the FL model, this can be attributed to the rules that are built to suit the larger number of data within a group of dams.

8 Uncertainty analysis of FL and RE estimates (testing phase)

Using the available equations, Wahl (2004) prepared log–log plots between the observed and predicted breach widths and noticed that the data are scattered approximately evenly above and below the lines of perfect prediction, suggesting that uncertainties would best be expressed as a number of log cycles on either side of the predicted value. He stated that the available RE for predicting breach width had absolute mean prediction errors <1/10th of an order of magnitude, indicating that on average their predictions are on target with the Froehlich (1995) equation having the smallest uncertainty.

An uncertainty analysis on the FL estimates ($\hat{B}_{av}$) resulted from the testing phase is performed with a similar procedure to that conducted by Wahl (2004). Application of the

Table 3 Estimated average breach widths (training phase)

Dam	h_w (m)	V_w (Mm³)	h_b (m)	Observed breach width B_{av} (m)	Estimated average breach widths $\hat{B}_{av}$ (m)			
					FL model	Froehlich (1995)	Bureau of Reclamation (1988)	Von Thun and Gillette (1990)
FP&L Martin Plant, Fla.	5.09	125.000	9.14	183.0	146.0	107.1	15.3	77.8
Hart, Mich.	10.70	6.350	10.80	73.9	78.2	42.6	32.1	69.7
Hatchtown, Utah	16.80	14.800	18.30	151.0	78.1	61.7	50.4	100.7
Hell Hole, Calif.	35.10	30.600	56.40	121.0	126.8	96.4	105.3	195.9
Horse Creek, Colo.	7.01	12.800	12.80	73.1	78.1	55	21.0	86.9
Hutchinson Lake, Ga.	4.42	1.170	3.75	33.4	25.0	28.4	13.3	15.5
Iowa Beef Processors, Wash.	4.42	0.333	4.57	16.8	23.1	14.1	13.3	17.5
Ireland No. 5, Colo.	3.81	0.160	5.18	13.5	11.5	11.4	11.4	19.1
Jacobs Creek, Penn.	20.10	0.423	21.30	17.5	23.0	20.5	60.3	59.4
Johnston City, Ill.	3.05	0.575	5.18	8.2	22.9	17.2	9.2	19.1
Kelly Barnes, Ga.	11.30	0.777	12.80	27.3	23.0	22.5	33.9	38.1
Kraftsmens Lake, Ga.	3.66	0.177	3.20	14.5	12.0	15	11.0	14.1
La Fruta, Tex.	7.90	78.900	14.00	58.8	92.2	100.2	23.7	89.9
Lake Avalon, NM	13.70	31.500	14.60	130.0	78.2	75.3	41.1	91.4
Lake Francis, Calif.	14.00	0.789	17.10	18.9	22.9	23.8	42.0	48.9
Lake Genevieve, Ky.	6.71	0.680	7.92	16.8	22.1	20.0	20.1	25.9
Lake Latonka, Penn.	6.25	4.090	8.69	39.2	40.0	35.5	18.8	40.0
Lake Philema, Ga.	9.00	4.780	8.53	47.2	41.0	52.1	27.0	39.6
Lambert Lake, Tenn.	12.80	0.296	14.30	7.6	17.4	16.8	38.4	41.9
Laurel Run, Penn.	14.10	0.555	13.70	35.1	23.0	28.6	42.3	40.4
Lawn Lake, Colo.	6.71	0.798	7.62	22.2	22.9	20.5	20.1	25.2
Lily Lake, Colo.	3.35	0.093	3.66	10.8	10.6	9	10.1	15.3
Little Deer Creek, Utah	22.90	1.360	27.10	29.6	29.1	31	68.7	86.1

Table 3 continued

Dam	h_w (m)	V_w (Mm^3)	h_b (m)	Observed breach width B_{av} (m)	Estimated average breach widths $\hat{B}_{av}$ (m)			
					FL model	Froehlich (1995)	Bureau of Reclamation (1988)	Von Thun and Gillette (1990)
Long Branch Canyon, Calif.	3.17	0.284	3.66	9.1	14.0	13.0	9.5	15.3
Lower Latham, Colo.	5.79	7.080	7.01	79.2	78.2	40.6	17.4	60.2
Lower Otay, Calif.	39.60	49.300	39.60	133.0	143.4	147	118.8	153.9
Lower Two Medicine, Mont.	11.30	29.600	11.30	67.0	78.2	70.3	33.9	83.2
Lynde Brook, Mass.	11.60	2.880	12.50	30.5	40.7	34	34.8	49.6
Melville, Utah	7.92	24.700	9.75	32.8	78.1	64.5	23.8	79.3
Merimac Upper Lake, Ga.	3.44	0.069	3.05	14.2	11.0	11.1	10.3	13.7
Mossy Lake, Ga.	4.41	4.130	3.44	41.5	41.0	41.8	13.2	26.9
Otter Lake, Tenn.	5.00	0.109	6.10	9.3	10.5	10.4	15.0	21.4
Pierce Reservoir, Wyo.	8.08	4.070	8.69	30.5	40.7	49.6	24.2	40.0
Potato Hill, NC	7.77	0.105	7.77	16.5	12.0	15.1	23.3	25.5
Prospect, Colo.	1.68	3.540	4.42	88.4	41.0	29.8	5.0	29.4
Quail Creek, Utah	16.70	30.800	21.30	70.0	78.2	80.3	50.1	108.2
Rainbow Lake, Mich.	10.00	6.780	9.54	38.9	78.2	59.5	30.0	66.6
Renegade Resort Lake, Tenn	3.66	0.014	3.66	2.29	10.3	6.8	11.0	15.3
Rito Manzanares, NM	4.57	0.025	7.32	13.3	11.0	6.7	13.7	24.4
Schaeffer, Colo.	31.90	4.440	30.50	137.0	140.2	64.8	95.7	94.6
Scott Farm Dam No. 2, Canada	10.40	0.086	11.90	15.0	12.0	10.9	31.2	35.9
South Fork, Penn.	24.60	18.900	24.40	94.5	78.2	98.7	73.8	115.9
Statham Lake, Americus, Ga.	5.55	0.564	5.12	21.0	22.9	23.8	16.7	18.9
Teton, Idaho	77.40	310.000	86.90	151.0	152.9	219.6	232.2	272.2
Timberlake, Va.	7.33	1.800	7.32	56.7	40.6	37	22.0	36.6
Trial Lake, Utah	5.18	1.480	5.18	21.0	33.3	23.5	15.5	31.3

Table 3 continued

Dam	h_w (m)	V_w (Mm³)	h_b (m)	Observed breach width B_{av} (m)	Estimated average breach widths $\hat{B}_{av}$ (m)			
					FL model	Froehlich (1995)	Bureau of Reclamation (1988)	Von Thun and Gillette (1990)
Trout Lake, NC	8.53	0.493	8.53	26.2	23.0	25.2	25.6	27.4
Upper Pond, Conn.	5.18	0.222	5.18	16.5	12.7	17.7	15.5	19.1
Wheatland No. 1, Wyo.	12.20	11.600	13.70	35.4	58.0	55	36.6	77.0
Wilkinson Lake, Ga.	3.57	0.533	3.72	29.0	23.0	15.7	10.7	15.4
Winston, NC	6.40	0.662	6.10	19.8	22.9	25.9	19.2	21.4

Table 4 Estimated average breach widths (testing phase)

Dam	h_w (m)	V_w (Mm^3)	h_b (m)	Observed breach width B_{av} (m)	Estimated average breach widths $\hat{B}_{av}$ (m)			
					FL model	Froehlich (1995)	Bureau of Reclamation (1988)	Von Thun and Gillette (1990)
Apishapa, Colo.	28.00	22.800	31.10	93.0	78.6	78.4	84.0	112.7
Baldwin Hills, Calif.	12.20	0.910	21.30	25.0	23.5	26.0	36.6	36.6
Bearwallow Lake, NC	5.79	0.049	6.40	12.2	11.5	11.4	17.4	20.6
Buckhaven No. 2, Tenn.	6.10	0.025	6.10	4.7	11.1	9.1	18.3	21.4
Bullock Draw Dike, Utah	3.05	0.740	5.79	12.5	22.9	26.6	9.2	13.7
Butler, Ariz.	7.16	2.380	7.16	62.5	40.2	40.3	21.5	36.2
Castlewood, Colo.	21.60	6.170	21.30	44.2	69.5	67.2	64.8	72.3
Caulk Lake, Ky.	11.10	0.698	12.20	35.1	23.5	21.5	33.3	33.9
Clearwater Lake, Ga.	4.05	0.466	3.78	22.8	23.0	21.2	12.2	16.2
Coedty, UK	11.00	0.311	11.00	42.7	23.7	22.8	33.0	33.6
East Fork Pond River, Ky.	9.80	1.870	11.40	17.2	39.0	29.1	29.4	42.8
Elk City, Okla.	9.44	1.180	9.14	36.6	25.3	33.7	28.3	41.9
Emery, Calif.	6.55	0.425	8.23	10.8	22.0	17.0	19.7	22.5
Fogelman, Tenn.	11.10	0.493	12.60	7.6	23.0	20.0	33.3	33.9
French Landing, Mich.	8.53	3.870	14.20	27.4	39.0	38.3	25.6	39.6
Frenchman Creek, Mont.	10.80	16.000	12.50	54.6	76.0	82.4	32.4	81.9
Grand Rapids, Mich.	6.40	0.026	6.40	19.0	11.2	9.2	19.2	22.1
Haas Pond, Conn.	2.99	0.023	3.96	10.7	10.3	5.9	9.0	13.6

uncertainty analysis, for the same database considered in this phase, is also illustrated for the breach widths calculated by Froehlich (1995), Bureau of Reclamation (1988), and Von
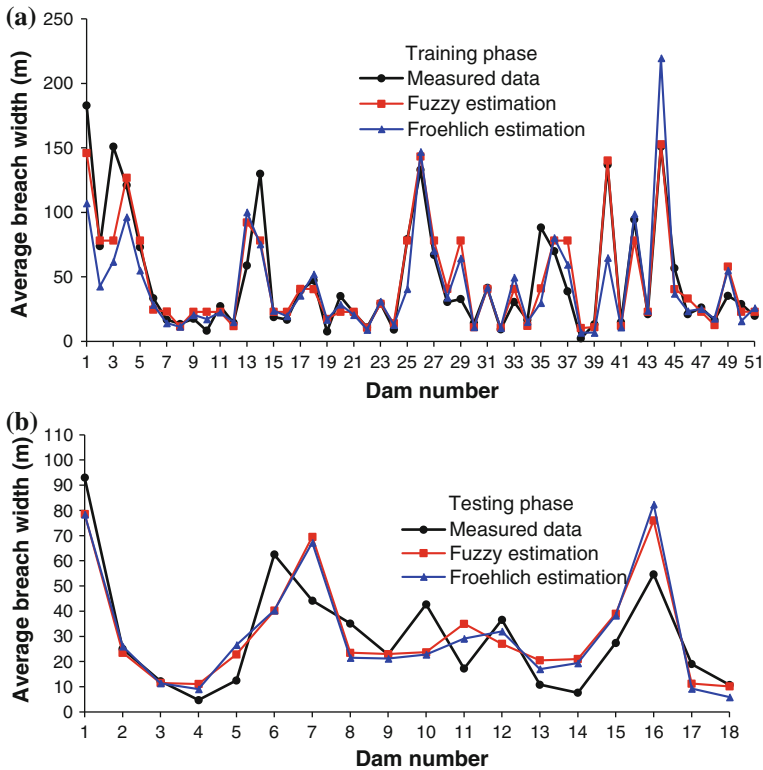
Fig. 4 **a** Comparison of FL and Froehlich (1995) estimates with measured data in the training phase. **b** Comparison of FL and Froehlich (1995) estimates with measured data in the testing phase

Thun and Gillette (1990) equations. The uncertainty analysis method consists of the following steps:

1. Compute $e_i = \log_{10}(\hat{B}_{av}/B_{av})$, where e_i's are individual prediction errors in terms of the number of log cycles separating the estimated and observed value, $\hat{B}_{av}$ and B_{av} are the estimated and the observed average breach widths, respectively.

2. Apply the outlier-exclusion algorithm to the series of prediction errors computed in step 2. The algorithm is described by Rousseeuw (1998):

   - Determine the estimator of location, $T = \text{median}(e_i)$.
   - Compute the deviations from the median and determine the median of these absolute deviations, $\text{MAD} = \text{median}|T - e_i|$.
   - Compute an estimator of scale, $S_{\text{MAD}} = 1.483 \times (\text{MAD})$. The 1.483 factor makes S_{MAD} comparable to the standard deviation, which is the usual scale parameter of a normal distribution. Compute a Z score for each observation, $Z_i = (e_i - T)/S_{\text{MAD}}$, then reject any observations for which $|Z| > 2.5$. If the samples are from a perfect normal distribution, this method rejects at the 98.7 % probability level.

3. Compute the mean, $\bar{e}$, and the standard deviation, S_e, of the remaining prediction errors. If the mean value is negative, it indicates that the prediction equation

Table 5 Uncertainty estimates of average breach widths from FL model and some RE (testing phase)

Parameter	FL model	Froehlich (1995)	Bureau of Reclamation (1988)	Von Thun and Gillette (1990)
Estimator of location, T	−0.01	−0.01	−0.03	0.13
Median of absolute deviations, MAD	0.17	0.20	0.19	0.09
Estimator of scale, S_{MAD}	0.26	0.29	0.28	0.14
Mean prediction error, $\bar{e}$	0.03	0.03	0.04	0.10
Two standard deviations, $\pm 2S_e$	0.51	0.56	0.56	0.28

underestimated the observed values, and if positive the equation overestimated the observed values. A confidence band around the predicted value is expressed using the values of $\bar{e}$ and S_e, as $\left\{ \hat{B}_{av} \cdot 10^{-\bar{e}-2S_e}, \hat{B}_{av} \cdot 10^{-\bar{e}+2S_e} \right\}$. The use of $\pm 2S_e$ approximately yields a 95 % confidence band.

Table 5 summarizes the results of the performed uncertainty analysis on the FL and RE estimates from the testing phase. From Table 5, it can be seen that the FL model and the Froehlich (1995) equation predicted the average breach widths with equal mean prediction error of $\bar{e} = +0.03$; however, the fuzzy uncertainty band of $\pm 2S_e = \pm 0.51$ order of magnitude is slightly smaller than the $\pm 2S_e = \pm 0.56$ order of magnitude obtained by the Froehlich (1995) equation. The results suggest that the FL model could be a good tool for estimating average breach width, although not replacing the best available regression equation of Froehlich (1995), but complementing it with extra information that may help the modeler. The FL approach provides the modeler with the capability to express the uncertainty inherent in such real systems, especially when confidence intervals of uncertainty concerning the data and the parameters are unknown. It also allows the modeler to use his intelligence, understanding, and descriptive capabilities for making decisions to solve nonlinear systems without complex mathematical formulation. Moreover, the FL model provides a comparatively effective tool for estimating the breach width.

Moreover, the uncertainty analysis showed that the FL approach and Froehlich (1995) equation yielded average breach widths with comparatively less mean prediction error than other RE. In light of the results presented in Tables 3, 4, and 5, one can conclude that the FL model can give reliable estimations for the average breach widths of embankment dams in comparison with the available RE. Linear regression methods performed well in case of strong linear relationship between the inputs and output. However, as the data and their functional relationships possess nonlinear behavior, the modifications become necessary for regression analysis. In contrast, the fuzzy approach can offer an alternative for considering such nonlinearity through vaguely defined membership functions.

9 Application

For validation of the FL model, two embankment failure case studies cited in recent publications are used. These are the risk assessment study for the Jamestown embankment dam, conducted in January 2001 by the Bureau of Reclamation, and the Big Bay dam failure that occurred on March 12, 2004. *The Jamestown dam* is a zoned-earth fill with a height of 24.7 m above the original streambed. The crest length is 432 m at an elevation of

Table 6 Estimated breach widths for Jamestown and Big Bay dams using the FL model and some RE

Reference	Estimated average breach widths $\hat{B}_{av}$ (m)	
	Jamestown dam top-of-flood-space, ElE. 443.18 m	Big Bay dam
FL model	155.0	78.1
Froehlich (1995) $\hat{B}_{av} = 0.1803 K_0 (V_w)^{0.32} (h_b)^{0.19}$	158.8	61.45
Bureau of Reclamation (1988) $\hat{B}_{av} = 3h_w$	58.5	40.50
Von Thun and Gillette (1990) $\hat{B}_{av} = 2.5h_w + C_b$	104.0	88.65
Recommended or observed values	165.0	83.20

Table 7 Error criteria using the estimates of the FL model and some RE (application phase)

Error criteria	FL model	Froehlich (1995)	Bureau of Reclamation (1988)	Von Thun and Gillette (1990)
MAE	7.6	14.0	74.6	33.2
MRE	−6.1	−14.9	−57.9	−15.2
MSE	63.0	255.8	6,582.8	1,875.4

448.36 m, the crest width is 9.14 m, and assumed initial elevation of piping failure is 423.7 m, Wahl (2004). One potential reservoir water surface elevation at failure is considered in this study, that is, the top-of-flood-space with elevation 443.18 m and reservoir capacity of about 273.3×10^6 m^3. *The Big Bay embankment* is approximately 576 m long and 15.6 m high. A pool elevation of 84.89 m, which corresponds to storage of 17.5×10^6 m^3, was used in the analysis. Site survey indicated an average breach width of 83.2 m, original ground elevation of 71.3 m, and assumed initial piping elevation of 71.3 m (Yochum et al. 2008).

Application of the FL model for estimating the breach width is illustrated for the mentioned data in these case studies. Two simulation runs for the FL model were made using the interactive View Rule option. The FL model requires the input data related to the water height and the reservoir storage in these case studies. Table 6 presents the results of these case studies using the present FL model and some of the available RE.

Using some error criteria such as the mean absolute error $\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} \left| \hat{B}_{av_i} - B_{av_i} \right|$, mean relative error $\text{MRE} = \frac{1}{N} \sum_{i=1}^{N} 100 \frac{(\hat{B}_{av_i} - B_{av_i})}{B_{av_i}}$, and mean square error $\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (\hat{B}_{av_i} - B_{av_i})^2$, where N is the total number of predicted outputs, $\hat{B}_{av_i}$ is the ith average breach width estimated by FL model or RE, and B_{av_i} is the corresponding ith observation, can help in comparisons between measured versus estimated average breach widths obtained using the FL model and the RE, (Table 7). In Table 7, the error values from RE are higher than those obtained from the FL model. The results show that the FL approach provides quite reasonable estimates for the observed average breach widths of these case studies compared with the applied RE. Moreover, the FL model and the Froehlich (1995) equation gave more accurate estimates for the average breach widths of these case studies than the other available RE.

10 Conclusions

This paper presented a FL rule-based model to estimate average breach widths of embankment dams. It provides an alternative solution to treat uncertainty and imprecision of breach data through the vaguely defined membership functions. The model is based on Mamdani's inference system, which expects the output to be a fuzzy set and is very good for the representation of human reasoning to describe systems in linguistic terms rather than in terms of complex relationships. Two data sets of 51 dams and 18 dams were used in the training and testing phases of the FL model, respectively. After performing an uncertainty analysis for its estimates, the developed FL model is applied through simulation of two case studies. Simulation results of the testing and application phases indicated that the proposed FL model exhibits reasonable accuracy and its estimates are in good agreement with the observed measurements and the results of the best available regression equation. The FL model offers a good tool that can be effectively used to predict the average breach width. Although the application of FL for estimating average breach width is promising, the proposed FL model can further be adjusted by searching for optimum key parameters and membership functions and modifying the rules accordingly.

References

Brown R, Rogers D (1977) A simulation of the hydraulic events during and following the Teton Dam failure. In: Proceedings of the dam-break flood routing model workshop. US Water Resources Council, Hydrology Committee, pp 131–163

Bureau of Reclamation (1988) Downstream hazard classification guidelines. ACER tech. memorandum no. 11. US Department of the Interior, Bureau of Reclamation, Denver, p 57

Cristofano E (1965) Method of computing erosion rate for failure of earthfill dams. US Department of the Interior, Bureau of Reclamation, Denver

Federal Energy Regulatory Commission (FERC) (1987) Engineering guidelines for the evaluation of hydropower projects. FERC 0119-1, Office of Hydropower Licensing

Fread D (1977) The development and testing of a dam-break flood forecasting model. In: Proceedings of the dam-break flood routing workshop. Water Resources Council, pp 164–197

Fread D (1984) DAMBRK—the NWS dam-break flood forecasting model. National Weather Service, hydrological technical note no. 4. Hydrol Res Lab, Silver Spring, MD

Fread D (1988) BREACH—An erosion model for earthen dam failures. National Weather Service, Office of Hydrology, Silver Spring, MD

Fread D (1993) NWS FLDWAV model—the replacement of DAMBRK for dam-break flood prediction. In: Proceedings of the 10th annual conference of the Association of State Dam Safety Officials, Lexington, KY, pp 177–184

Froehlich D (1987) Embankment-dam breach parameters. In: Proceedings of the ASCE national conference on hydraulic engineering, New York, pp 570–575

Froehlich D (1995) Embankment dam breach parameters revisited. In: Proceedings of the ASCE conference on water resources engineering, New York, pp 887–891

Froehlich D (2004) Two-dimensional model for embankment dam breach formation and flood wave generation. In: Proceedings of the 21st annual conf of the Association of State Dam Safety Officials, Lexington, KY

Froehlich D (2008) Embankment dam breach parameters and their uncertainties. J Hydraul Eng 134(12):1708–1721

Hanson G, Cook K, Hunt S (2005) Physical modeling of overtopping erosion and breach formation of cohesive embankments. Trans ASAE 48(5):1783–1794

Johnson F, Illes P (1976) A classification of dam failures. Int Water Power Dam Constr 28(12):43–45

MacDonald T, Langridge-Monopolis J (1984) Breaching characteristics of dam failures. J Hydraul Eng 110(5):567–586

Mamdani E (1977) Applications of fuzzy logic to approximate reasoning using linguistic synthesis. IEEE Trans Comput 26(12):1182–1191

Ponce V, Tsivoglou A (1981) Modeling gradual dam breaches. J Hydraul Div Proc ASCE 107(7):829–838

Ross J (1995) Fuzzy logic with engineering applications. McGraw-Hill, New York, p 593

Rousseeuw P (1998) Chapter 17: robust estimation and identifying outliers. In: Wadsworth HM Jr (ed) Handbook of statistical methods for engineers and scientists, 2nd edn. McGraw-Hill, New York, pp 17.1–17.15

Singh V, Scarlatos P (1985) Breach erosion of earthfill dams and flood routing: BEED model. Research Report, Army Research Office, Research Triangle Park, NC, p 131

Singh K, Snorrason A (1984) Sensitivity of outflow peaks and flood stages to the selection of dam breach parameters and simulation models. J Hydrol 68:295–310

Visser P (1998) Breach growth in sand-dikes. Communications on hydraulic and geotechnical engineering, report no. 98-1. Delft University of Technology, Faculty of Civil Engineering and Geosciences, Hydraulic and Geotechnical Engineering Division, Delft, The Netherlands

Von Thun J, Gillette D (1990) Guidance on breach parameters. Internal Memorandum, US Department of the Interior, Bureau of Reclamation, Denver, p 17

Wahl T (1998) Prediction of embankment dam breach parameters: a literature review and needs assessment. DSO-98-004, Dam Safety Research Report, US Department of the Interior Bureau of Reclamation

Wahl T (2004) Uncertainty of predictions of embankment dam breach parameters. J Hydraul Eng 130(5):389–397

Yochum S, Goertz L, Jones P (2008) Case study of the Big Bay dam failure: accuracy and comparison of breach predictions. J Hydraul Eng 134(9):1285–1293

Zadeh L (1965) Fuzzy sets. Inf Control 8:338–353