ESTIMATION OF FAILURE TIME FOR EMBANKMENT DAMS

SAAD SHAUKET SAMMEN (1,2) THAMER AHMEED MOHAMED (1) ABD ALHALIM GHAZALI(1) LARIYAH MOHAMMED SIDEK(3) Ir. AZLAN Abdul Aziz(1)

Corresponding Author Email: SAAD123engineer@yahoo.com
(1) Department of Civil Engineering, Faculty of Engineering, University Putra Malaysia, Serdang, Selangor, Malaysia
(2) Department of Civil Engineering, College of Engineering, Diyala University, Diyala Governorate, Iraq.
(3) Department of Civil engineering, College of Engineering, University Tenaga Nasional, Kajang, Selangor, Malaysia

ABSTRACT

Several approaches have been used for estimation of failure time when dam failure occurs. Generalized Regression Neural Network (GRNN) model has been proposed in this present study as a novel method to estimate the dam failure time. It has also made a comparison of the results of the GRNN models with the results obtained from the existing approaches. Various reservoir and dam characteristics have been used in the development of the GRNN models in order to estimate the dam failure time. To obtain the optimum results from the GRNN models, all models was optimized to smooth factor values. The values were found to range from 0.001 to 0.03. Furthermore, training the network took up 85% of the total data and testing the network took up the 15% of the total data. Three statistical indices were used to assess the results of GRNN models. They were the Root Mean Square Error (RMSE), Mean Relative Error (MRE), and Coefficient of Correlation (R²). The results showed that the value of the MRE could be decreased more than 50% by using the GRNN models as compared to the values of the existing empirical techniques.

Key words: Dam safety, Dam failure, Failure Time, generalized regression neural network.

1 INTRODUCTION

Dams are built to improve human life. They are constructed as multipurpose structures and are used to produce hydroelectric power, provide water for irrigation and water supply, for improvement of economics, and for the control of flooding (Hooshyaripor et al., 2014). Thus, they have become a vital component of any country’s infrastructure (Wahl, 2010). However, the huge volume of water that is store in the reservoir behind the dam wall can lead to a damaging flood to the population and their properties located downstream from the dam if the stored water is suddenly released (Razad et al., 2013). Therefore, an analysis of the dam break is considered as tremendously important so that the peak outflow and its associated Emergency Action Plan (EAP) can be determined (Wahl, 2010). There are three main stages in a dam break analysis; they are the geometric breach parameter estimations (height, width, and side slope), hydrologic breach parameter estimations (failure time and peak outflow), and downstream routing of the flood. A physical model and numerical modelling methods can be employed to study the analysis of the dam break. As the physical model is quite expensive, numerical modelling is thought to be a better and less expensive approach (Wahl, 2010).

To carry out an analysis of the failure of the embankment dam, the parameters of the breach are considered to be the most important key parameters that must be estimated with accuracy because of their effect on the degree of the failure risk as well as the consequence of the peak outflow. As a result, many of the contributions to the literature in the past few decades have been focused on the development of precise and simple methods to handle these kinds of problems (Bentaher, 2013; Xu and Zhang, 2010; Wahl, 1998, 2004, and 2010; and Froehlich, 1995 and 2008).

Over the last few decades, research has been carried out on noncohesive and cohesive embankment dams by employing experimental research works which have been able to explain how breaches develop. In addition, these experimental research works have contributed to the provision of data which is necessary for both statistical and numerical studies (Gaucher et al., 2010; Coleman et al., 2002). Moreover, some numerical and analytical tools have been utilised to solve the Sainte Venant equations for the study and analysis of flood waves which result from the dam failure (Xia et al., 2010; Tsai, 2005; and Ponce et al., 2003). Besides that, numerical techniques which solve the shallow water equation have been employed to estimate the hydrograph of the peak outflow, and to determine the velocity along the distance, time, and inundation area (Gallegos, 2009; Liang et al., 2007). The eroding capacity of the flow can be determined by using a more reasonable methodology which takes the mean velocity of the flow and the mean shear stress as independent variables (Macchi, 2008). However, the field as well as the experimental research works have given evidence that this methodology could possibly be suitable for only a few stages in the development of the breach, but cannot be applied for all the stages of the development breach (Wahl, 1998).

The statistical analysis method is thought of as a traditional technique which can be employed to predict the characteristics of the flow and the breach in the dam. In this technique, the characteristics of the reservoir, such as the water volume and the depth were used as the dependent variables for determining the
characteristics of the flow as well as the parameters of the breach (Xu and Zhang, 2009; Froehlich, 1995a, 2004, 2008; Von Yhun and Gillette, 1990; and USBR, 1988). When the database of dam failure cases is highly documented, preferred equations can result from a case study analysis (Xu and Zhang, 2009). After the 1970s, as shown in Table 1, various observational equations have been developed using the regression analysis. These relations, which are not very complex, are still needed, especially in situations where the intention is not to have detailed simulations or they are impossible to be applied easily or conveniently (Froehlich, 2008). However, these relations need to be represented with the relevant uncertainty such as confidence limit. For instance, On the other hand, relations such as these must be represented with the relevant uncertainty like when confidence is limited. For instance, Wahl (2004) has given a description of an uncertainty analysis technique and has compared the uncertainty estimates for the equations of the peak outflow prediction and breach parameters. Following the work by Wahl (2004), Pierce et al. (2010) also gave a description of the uncertainty analysis of the parameters of a breach. Limited data sets and their low adaptability to data varieties cause the uncertainty of these equations. Furthermore, the databases do not cover a very wide scope of included

<table>
<thead>
<tr>
<th>Investigator</th>
<th>Number of case studies</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Von Thun and Gillette (1990)</td>
<td>36</td>
<td>$T_f = 0.15H_w$ (for high erodible)</td>
</tr>
<tr>
<td>Von Thun and Gillette (1990)</td>
<td>36</td>
<td>$T_f = 0.15H_w + 0.25$ (for erosion resistance)</td>
</tr>
<tr>
<td>Froehlich (1995a)</td>
<td>34</td>
<td>$T_f = 0.00254(V_w)^{0.53}(H_w)^{0.9}$</td>
</tr>
<tr>
<td>Froehlich (2008)</td>
<td>unknown</td>
<td>$T_f = 0.0176([V_w/(gH_w^2)]^{0.5}$</td>
</tr>
<tr>
<td>Xu and Zhang (2009)-1</td>
<td>unknown</td>
<td>$T_f = C_1 T_r(H_w/H_r)^{0.654} (V_w^{1/3}/H_w)^{1.246}$</td>
</tr>
<tr>
<td>Xu and Zhang (2009)-1</td>
<td>unknown</td>
<td>$T_f = 0.304 e^{B} T_r(H_w/H_r)^{0.654} (V_w^{1/3}/H_w)^{1.246}$</td>
</tr>
</tbody>
</table>

Where:
- $H_w$: Height of dam (m), $H_b$: Height of breach (m), $V_w$: Volume of water in reservoir at failure time (m³), $H_r$: Height of of water in reservoir at failure time (m), $C_1$: Factor related to dam erodibility, $H_r$: Dam height reference = 15m., $T_r$: Time failure reference = 1hr.

variables as they do not incorporate very many cases of failures involving large dams (Nourani, 2012). Because of this, the mathematical statements made by employing such restricted databases are not very dependable for most cases.

Estimating these parameters is considered to be a very complex task as the relationships of the parameters of the dam breach are highly nonlinear and because of their time variations. Thus, if an appropriate database is available, the black-box models can be employed as an alternative method. With the black-box method, the targets inside the model and the inputs are mapped directly with no need to consider the details of the structure inside the physical process (Hakimzadeh et al., 2014). An Artificial Neural Network (ANN) is known as a black-box model that has usefulness that exceeds the normally used statistical models like the free-pattern of forecasting model, a data-driven nature, and a toleration to data inaccuracy (Azmatullah et al., 2005). Besides its capability and simplicity, ANN has been employed in several common areas. These include hydraulics (Karunanithi et al., 1994), (Avramidis et al., 2001), (Liriano and Day, 2001), (Bateni and Jeng, 2006), (El-Shafie et al., 2007), (Zounemat-Kermani et al., 2009), and (Azamathulla and Gani, 2011); flood prediction (Chang et al., 2007); groundwater flow remediation (Ranjithan et al., 1993); rainfall forecasting (Ramirez et al., 2005); the design of reservoir operating policies (Raman and Chandramouli, 1996), (Chandramouli and Raman, 2001), (Cancelliere et al., 2002), (Liu et al., 2006); water resources management (Li and Huang, 2012); and the prediction of the peak outflow from a dam breach (Babaeyan et al., 2011; Nourani, 2012; Hooshyaripor et al., 2012; and Hoo shyari por et al., 2014) in. In an ANN model that was developed recently, Nourani (2012) gathered data from laboratories, historical cases, and a numerical model that was physically-based, and then a simulation of an outflow hydrograph from the earth dam breach was created using the ANN model. The sensitivity analysis of his work verified that, when dealing with the breach process, not only is the depth of the water in the reservoir when the failure occurs a very important physical parameter but also the volume of the reservoir as compared to others. According to the review of the literature, Artificial Neural Networks are not employed to predict the width of a dam breach although they have the capability of estimating the nonlinearity of the relationship for any kind of data and although the already available techniques, which are employed to estimate the width of a beach in a dam, have a degree of uncertainty. As a result, this present study has been proposed for the interduce the Generalized Regression Neural Network (GRNN) which is usually adopted for the function of approximating in order to improve the accuracy of the dam failure time estimation. This proposed GRNN has been trained and tested by applying real data so that its performance accuracy could be demonstrated.

3 Data Collection and Existing Approaches

The magnitude of the peak outflow is the main attribute of the hydrograph of the breach outflow which affects the losses of property and life in a dam break phenomena. This attribute is affected by the various parameters of the breach, such as the dam failure time (Costa, 1985). There are a few models available in the literature which are physically-based and used to simulate how breaches in dams develop (Hanson, 2005;
Froehlich, 2004; Visser, 1998; Fread, 1984, 1993; Ponce and Tsivoglou, 1981; and Cristofano, 1965). These models have a poor understanding of the development of the breach although they are dependent on water discharge and sediment erosion equations (Wahl, 2004). The database applied in this work included data from the failures of more than 140 real dams, which were gathered from a variety of sources. The results from various related research works (Froehlich, 1995; Singh, 1988; Wahl, 1998; Taher Shamsi et al., 2003; Xu and Zhang, 2009; and Pierce et al., 2010) were used in this present work.

A large number of the currently available methods employ the regression analysis to depict the value of the breach width as a function of the volume of water in the reservoir as well as the water depth behind the wall of the dam. The reported outcomes of the abundant regression analyses are illustrated for the incorporation of the prediction equation as well as the number of case study investigations employed as an element in the analysis, and can be seen in Table 1. Wahl (2004, 1998) introduced the case study data that were employed by previous researchers to determine the equations which can be utilised for the prediction of the peak outflow of the breach in a dam.

4 Artificial Neural Network

Neural Networks are normally thought of as black boxes which have been trained on a substantial number of data sets for a specific function. They are the worldview of data handling, which is motivated by the manner in which the natural biological nervous system processes information. Each network consists of numerous interconnected handling elements (neurons) working with each other as a single unit to deal with various problems. Neural systems have an impressive ability to determine the significance in loose or confounding information. In addition, they can be employed to focus on designs and identify patterns which are too complex to be noticed in any manner, either by people or even other computer strategies (Datt, 2012). The general structure of the neural network consists of three layers of neurons, which are the hidden layer and the output and input layers. An abundance of laboratory and hypothetical research works have explained that a complex nonlinearity function can be adequately approximated with an ANN having only one hidden layer. Similarly, it has been proposed that the furthest point for any number of the neurons in the layer that is hidden be less than $2n + 1$, where $n$ is the number of the input neurons (Hecht-Nielsen, 1987).

GRNN is a kind of new neural system which was introduced by Specht. GRNNs have a design that is similar to the Multilayer Perceptron Neural Networks (MLNNs), however, there is an important distinction between them. GRNNs carry out regression where there is a consistent objective variable whilst MLNNs carry out characterization where there is a categorical objective variable. GRNNs are set up to evaluate any approximate function of data which has been previously recorded. The GRNN is a modification of the Radial Basis Function Neural Network (RBFNN), which is dependent on kernel regression networks (Cigizoglu, 2005; Celikoglu, 2006). No iterative preparation methodology as back propagation networks is needed by GRNNs. Any discretionary function that has two or more variable vectors is approximated by the GRNN, drawing the function which has been evaluated in a straightforward manner from the data used for training. Moreover, it is consistent enough that, when a large data set has been used for training, the error of the estimation will approach zero and only slight limitations will be put on the function (Celikoglu, 2005). GRNNs are comprised of four layers. These layers are the input and output layers, the design layer, and the summation layer. The general structure of the GRNN which was employed in this present study is illustrated in Figure 1. GRNNs are thought of as standardized RBFNNs where there is a unit centered during each training stage.
The network is designed as a one-pass learning algorithm with a structure that is very parallel, Remarkably, the calculation provides smooth moves starting from one target value and moving to another even if using a poor data set in an estimation space which is multi-dimensional. An algorithmic structure can be employed with any problem of regression as a part of which there is no confirmation of a supposition of linearity. GRNNs are comprehensive approximations for a smooth function. As such they are able to solve any problem of smooth function estimation (Disorntetiwat, 2001). Initial weight values cannot be arbitrarily appointed as the feedforward backpropagation method’s execution is tremendously sensitive. However, this problem was not encountered in the GRNN simulation in any case (Celikoglu, 2005). It was explained by Specht that GRNNs do not require any iterative preparation process as is needed in the back-propagation method. There was no problem of a local minima being encountered when the GRNN model was used. As a result, the preference was given to the GRNN model instead of the feedforward backpropagation. The total number of data units present in the input layer is dependent upon the total number of parameters which have been employed. The pattern layer and the input layer are connected to each other; and in the pattern layer, each neuron represents a training pattern and the pattern’s output. The pattern layer is also joined to the summation layer. There are two distinct kinds of summation in the summation layer; they are the summation units and the division unit. The summation layer along with the output layer carries out the standardization of the output data. During the training of the network, the radial basis is employed in the hidden layer. At the same time, in the output layer, the function of linear activation is employed. In the summation layer, every two neurons (S and D) are joined to one of the units located in the pattern layer. All of the weighted output of the pattern layer is processed by the S summation neuron; whilst, the unweighted output is computed by the D summation neuron (Kim 2004). The following equation is used to determine the output value from the GRNN model.

\[
\text{Output} = \frac{\sum_{i=1}^{n} y_i \cdot \exp\left(\frac{-\sum_{i=1}^{m} (x_i - x_{ij})^2}{2\sigma^2}\right)}{\sum_{i=1}^{n} \exp\left(\frac{-\sum_{i=1}^{m} (x_i - x_{ij})^2}{2\sigma^2}\right)} \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (1)
\]

Where: 
- \(n\): No. of training cases,
- \(m\): No. of output data.

The term of \((x_i - x_{ij})\) is equal to the difference between the training data \(x_{ij}\) and the point of estimation \(x_i\) of the \(i\)th data.

The accuracy of the function resulting from the training stage of the GRNN was influenced by the factor for smoothing (spread), which was found to be equal to the standard deviation \(\sigma\). The nature of the data is the key factor influencing the selection of the smoothing factor’s value. For example, a small value of the smoothing factor has to be chosen for regular data; whereas, a large value is appropriate for irregular data so that a good performance can be achieved from a GRNN. The GRNN was selected for this study because it learns quickly and has the capability of converging to the optimum regression surface. Simply put, the GRNN is a method employed to assess any function when presented with only the data used for training. As there is a probability of the density function coming from the training stage without any biases in regards to its shape, the framework is completely general. There is no problem if the functions are created out of multiple disjointed non-Gaussian locals, in a variety of measurements, and in addition to those of easier dispersions (Wasserman, 1993).

**Methodology**

Figure 2 below shows the proposed Methodology that was utilized to build the GRNN models so that the dam failure time could be estimated and a comparison of the results with the results that were achieved by employing existing approach could be made. A database of 140 breached dams was employed to create the ANN models. The first stage in the building of the ANN model was to divide the data of the dam failures into a training and a testing set. To achieve this, the training set used 85% of the data and the testing set used the remaining 15%.

For better accuracy of the results and to create a more efficient neural network, a preprocessing step had to be performed on the variables of the network. It is often helpful, to scale the targets and inputs so that they will, generally, fall inside a limit that is predetermined before beginning to train a network. In this present research work, the targets and inputs had been scaled so that they would fall in the range of -1 and +1. The preprocessing step was performed by applying the equation that follows:

\[
X = \frac{2(X_i - X_{\text{min}})}{X_{\text{max}} - X_{\text{min}}} - 1 \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (2)
\]

Where: \(X\) is the standardized value of \(X_i\), \(X_{\text{min}}\) is the minimum value of data and \(X_{\text{max}}\) is the maximum value of data.

The main explanation behind needing to standardise the data is that measurement of the variables is usually made in different units. By taking this step, the variance in the levels amongst the data could be avoided (Romesburg, 1984; Sudheer et al., 2002).
After the simulation was completed, all of the output values were de-standardized by multiplying them by the respective factor of standardization to get the actual values of the peak outflow. This preprocessing step made for a more efficient stage of training by enhancing the process of learning (Demuth, 2001). A variety of standard statistical performance evaluation measures were employed to evaluate how the developed GRNN models performed. The three statistical performance indices employed in this study have been presented below:-

1- Mean Relative Error (MAE): measures the mean absolute error between the observed and the predicted values.

\[ \text{MRE} = \frac{1}{n} \sum_{i=1}^{n} \frac{(x_o - x_p)}{x_o} \] ...

\[ \text{MRE} = \frac{1}{n} \sum_{i=1}^{n} \frac{(x_o - x_p)}{x_o} \] ...

2- Root Mean Square Error (RMSE): measures the root square of the error.

\[ \text{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{(x_o - x_p)^2}{n}} \] ...

\[ \text{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{(x_o - x_p)^2}{n}} \] ...

3- Coefficient of Determination ($R^2$): this coefficient of efficiency is calculated as:

\[ R^2 = \frac{\sum_{i=1}^{n} (x_o - x_{m1})(x_p - x_{m2})^2}{\sum_{i=1}^{n} (x_o - x_{m1})^2(x_p - x_{m2})^2} \] ...

where: $x_o$ is the observed values, $x_p$ is the predicated values, $x_{m1}$ is the average of the observed values and $x_{m2}$ is the average of the predicated values.

The statistical parameters were calculated by employing the data of the dam failure time which were estimated with the GRNN model and other existing approaches. Use of the optimum value of the factor for smoothing (spread) is necessary if more accurate results are to be obtained with the GRNN model. Thus, to obtain a more efficient GRNN model, the researcher carried out a sensitivity analysis for various values of the factor for smoothing.

**Results and Discussion**

Two scenarios were performed to predict the dam failure time; each scenario employed different variables as the input. In addition, a comparison was made between the already available techniques for the linear regression analysis of width estimation of dam breaches and these proposed models. The details and results of each model are presented below.

**First Scenario**

The water depth upstream (Hw) of the dam was used in this model to predict the dam failure time when dam break ia happen. Von Thun and Gillette employed a linear regression analysis in order to create a relationship between $T_f$ and $H_w$. For the smoothing factor (spread) value, the range from 0.0001 to 1 was utilised to obtain the optimum value to achieve results with better accuracy. In Figure 3, it can be seen that, the optimum value for the factor of smoothing (spread) with this model was observed to be equal to 0.001.
The performance of the GRNN model for each of the stages (training and testing) is presented in Figure 4. For further analysis, the statistical analysis results for the GRNN model as well as the other empirical equations which were used for the estimation of the dam failure time depending on the depth of the water ($H_w$) are presented in Table 2. As can be seen in the results in Table 2, when compared with the regression equation, the GRNN model had a high value of NSE and low value of RMSE. Furthermore, the MRE value was decreased by more than 90% when employing the GRNN model as compared to employing the Von Thun and Gillette equation.

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSE</th>
<th>$R^2$</th>
<th>MRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRNN</td>
<td>0.20</td>
<td>0.995</td>
<td>0.09</td>
</tr>
<tr>
<td>Von Thun and Gillette</td>
<td>2.093</td>
<td>0.614</td>
<td>0.57</td>
</tr>
</tbody>
</table>

### Second Scenario

In this scenario, an estimation was made of the dam failure time depending on the volume, depth of the water in the reservoir and the dam height. According to the literature, Xu and Zhang (2009) found the most well-known equations for the relationship between ($V_w$, $H_b$) and ($T_f$). Thus, a comparison was made between the results of the GRNN model and the results of that technique. The optimum spread value was found to be equal to 0.028 as can be seen in Figure 5.
Figure 5 MRE variation with different smoothing factors values for second scenario

Figure 6 shows the performance of the GRNN model for both the training stage and the testing stage. It is obvious in the figure that the GRNN model was able to predict the value of $T_f$ close to the value that was actually observed.

Table 3 tabulated the statistical evaluation of the GRNN model and Xu and Zhang (2009) for dam failure time estimation. As can be seen from the results in Table 3, the GRNN model can estimate the dam failure time with good accuracy comparison with the regression equation. It is very clear from the high value of the NSE and the low values of the RMSE and the MRE. In addition, when the GRNN model was used, the value of the MRE was reduced by about 50%.

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSE</th>
<th>NSE</th>
<th>MRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRNN</td>
<td>0.569</td>
<td>0.955</td>
<td>0.16</td>
</tr>
<tr>
<td>Xu and Zhang</td>
<td>1.038</td>
<td>0.836</td>
<td>0.34</td>
</tr>
</tbody>
</table>

5. Conclusion

In recent decades, the development of precise and simple models for the prediction of dam breach parameters have been quite a challenging task because of the importance of the preparation of emergency action plans as well as the assessment of risk when the failure of the dam occurs. A generalized regression artificial neural network model was used in this present study for the estimation of the dam failure time. The
results were later compared with the results of the existing empirical approaches. A study of the problem of the prediction of the dam failure time was carried out and the analysis was performed by employing data from more than 140 recorded dam failures worldwide, which were gathered from the related literature. Network training made use of 85% of the total data, and network testing made use of the remaining 15%.

The values of the RMSE, R2, and MRE, which resulted from the analysis, demonstrate the potential of employing the GRNN as a predictive tool for estimation of the breach width. Additionally, the results showed that the MRE values of the breach width prediction estimated using the common regression analysis techniques can be decreased about 50%.

References