Empirical formulas in the prediction of breach parameters

Lin Zhang i), W. M. Yan ii) and Ka-Veng Yuen iii)

i) Ph.D Student, Department of Civil Engineering, The University of Hong Kong, Pokfulam, Hong Kong, China
ii) Assistant Professor, Department of Civil Engineering, The University of Hong Kong, Pokfulam, Hong Kong, China.
iii) Professor, Department of Civil and Environmental Engineering, University of Macau, Macao, China

ABSTRACT

Failure of dams could lead to not only severe economic loss but also human casualties. Breaching analysis is therefore an important component in the risk assessment of dam safety. Traditionally the breaching process is often characterized by the breach parameters including the breach geometry and hydrograph. Over the decades, a large variety of empirical formulas have been proposed on the basis of different dam failure databases to predict the parameters. In most cases, the goodness-of-fit criteria (e.g., minimizing the sum of squares of the fitting errors) remains the only measure of the formula suitability. It may therefore lead to the use of over-complicated formula due to its higher capability of fitting. In this paper, the Bayesian probabilistic approach is used to revisit the problem. The study concludes a set of empirical formulas for the dam breach parameters which are well-balanced between formula fitting capacity and complexity.

Keywords: Bayesian analysis, probability, dam failure, breach parameters, model class selection

1 INTRODUCTION

Dams are vital to the development of cities for its contribution to flood control and generation of hydropower. Flooding resulting from the failure of a dam often leads to devastating disasters which cause not only economic loss but also human casualties. The failure of Banqiao Dam (China) in 1975 killed approximately 171,000 people while 11 million people became homeless (Osnos 2011). The failure of Teton Dam (USA) in 1976 resulted in the deaths of 13,000 heads of cattle and 11 people. The total damage was estimated up to 2 billion US dollars (Reisner 1993). Analysis of the dam breach process and the outflow from the breach is essential for risk assessment of the resulting flood. Prediction of the breach parameters, which are used to describe the breach geometry and outflow hydrograph, therefore plays an important role. The breach parameters include the breach geometric parameters and hydrograph (Wahl 1998; Xu and Zhang 2009). Singh and Snorrasen (1984) studied the sensitivity of the key breach parameters with DAMBRK and HEC-1 models, which employ the St. Venant equations and hydraulics for routing. Petrascheck and Sylder (1984) presented a mathematical model to study the process of flood flow and demonstrated the sensitivity of discharge, inundation levels, and flood arrival time to the changes in the breach width and breach formation time. These studies showed that accurate prediction of breach parameters is necessary for reliable estimation of the resulting downstream inundation in close proximity to the dam.

Different approaches have been employed to estimate the breach parameters. Wahl (1998) summarized them into three groups: (1) comparative analysis, (2) physically-based model, and (3) prediction equation. The comparative analysis method is expected to work well only for the dams with similar geological and hydrological conditions and the physically-based models always over-simplify the failure process due to the lack of understanding of the mechanism. Therefore, the prediction equation approach remains the most widely used method nowadays. The prediction equations are often obtained by using minimum fitting error as the sole criterion in the regression analysis of limited records from a database. It may result in over-complicated formulas due to over-fitting and it will affect the predictability of the equations.

This paper presents a Bayesian probabilistic method to derive a set of the prediction formulas of the breach parameters. This method strives for a balance between predictability and model complexity through a robust Bayesian consideration of the data. A database containing 141 failures cases of earthfill dams, ranging from small to large dams with a reference height of 15 m as given by ICOLD (1988), has been compiled from the literature (Wahl 1998; Xu and Zhang 2009). Both additive and multiplicative forms of prediction equations are examined and the plausibility and robustness for each model are investigated.
2 BREACH PARAMETERS

Breach parameters are used to describe the breach geometry and the outflow hydrograph. A dam breach is usually idealized as a trapezoid such that the geometric parameters include the breach depth $H_b$, breach top width $B_t$, breach bottom width $B_b$, breach average width $B_{ave}$ and side slope $Z$ (Fig. 1). These parameters are geometrically related and therefore any three parameters are adequate to fully describe the geometry. In this paper, $H_b$, $B_t$ and $B_{ave}$ are used.

![Fig. 1. Geometric parameters of idealized breach](image)

The hydrographic parameters include the peak outflow rate $Q_p$, and failure time $T_f$. The reported peak outflow was determined either from the stage records of reservoir levels or by the slope area measurements (Froehlich 1995). The failure time is defined as the duration from inception to the completion of the breach (Singh and Snorrason 1984).

By reviewing the failure mechanism and the data availability, only four factors are considered in this paper for the breach parameters prediction. They are (1) height of dam $H_d$, (2) height of water above the invert of the breach $H_a$, (3) capacity of the reservoir behind the dam $V_d$, and (4) volume of water above the invert of breach $V_a$.

3 PREDICTION EQUATIONS AND MODEL CLASS SELECTION FRAMEWORK

Two forms (additive and multiplicative) of prediction equation are used in this study. The additive form of prediction equation is given in Eq. (1). By considering the modeling error and measurement noise, the output measurements can be written as the summation of the predicted value and error term:

$$y = \alpha^T x = \sum_{i=0}^{N_p} \alpha_i x_i$$

$$\hat{y} = y + \varepsilon = \sum_{i=0}^{N_p} \alpha_i x_i + \varepsilon$$

where $y$ and $\hat{y}$ are the predicted and measured output value of the system, respectively; $\alpha_i$ are the fitting coefficients of the system; and $x_i$ are the independent measurements. $\varepsilon$ represents the modeling error and measurement error, which can be modeled by Gaussian distribution with zero mean.

Eq. (3) shows the multiplicative form.

$$y = \prod_{i=0}^{N_p} x_i^\beta_i = x_0^\beta_0 x_1^\beta_1 \cdots x_{N_p}^\beta_{N_p}$$

To facilitate the regression analysis, Eq. (3) is transformed to an additive format with respect to $\ln y$ and $\ln x_i$ by taking log on both sides of the equation (Eq. (4)).

$$\ln y = \beta_0 \ln x_0 + \beta_1 \ln x_1 + \cdots + \beta_{N_p} \ln x_{N_p}$$

Bayesian model class selection framework (Yan et al. 2009; Yuen 2010; Chiu et al. 2012) was introduced in this paper to select the most suitable class of models to capture the relationship between system input and output among the given model candidates. Instead of minimizing the fitting error, the Bayesian framework aims at selecting the most suitable model by balancing the goodness-of-fit and robustness of the fitting. To achieve this goal, the plausibility of the each candidate model is evaluated using Eq. (5).

$$p(M_j|D,U) = \frac{p(D|M_j,U)p(M_j|U)}{p(D|U)}$$

where $p(M_j|U)$ represents the user’s judgment on the initial plausibility of the model class $M_j$. $p(D|U) = \sum_{j=1}^{N_M} p(D|M_j,U)p(M_j|U)$ is a denominator factor to normalized the plausibility of each model class. $p(D|M_j,U)$ is the evidence for the model class $M_j$ given data $D$. The most plausible model class is the one with the largest $p(D|M_j,U)p(M_j|U)$. With the help of the total probability theorem, this term can be calculated as Eq. (6).

$$p(D|M_j,U) = \int_{\Theta_j} p(D|\theta_j,M_j,U)p(\theta_j|M_j,U)d\theta_j$$
\[ p(D | \theta_j, M_j) = 
(2\pi)^{-N/2} \sigma^N \exp \left[ -\frac{N_j}{2\sigma^2} J_s(\theta_j | D, M_j) \right] 
\] (7)

where \( J_s(\theta_j | D, M_j) = \frac{1}{N_j} \sum_{n=1}^{N_j} \left[ y(\theta_j, x_{(n)}, M_j) - \bar{y}_{(n)} \right]^2 \)

is the goodness-of-fit function under fitting coefficient \( \theta_j \). \( N_j \) is the number of the uncertain parameters for the model class \( M_j \).

In practice, the calculation of \( p(D | M_j) \) is usually approximated by the Laplace’s method for asymptotic expansion (Papadimitriou et al. 1997), and it is given by (Beck and Yuen 2004):

\[ p(D | M_j) \approx p(D | \hat{\theta}_j, M_j) p(\hat{\theta}_j | M_j) (2\pi)^{-N/2} |H_j(\hat{\theta})|^{1/2} \] (8)

where \( \hat{\theta}_j \) is the optimal parameter vector, which is the most probable value of the parameter vector \( \theta_j \). \( \hat{\theta}_j \) can be obtained by maximizing the integrand \( p(D | \hat{\theta}_j, M_j) p(\hat{\theta}_j | M_j) \). In a linear system with uniform priors, it equals to minimizing the \( J_s(\theta_j | D, M_j) \) and the optimal fitting parameters \( \hat{\theta}_j \) can be obtained by solving the formula in Eq. (9)

\[ \frac{\partial J_s(\theta_j | D, M_j)}{\partial \theta_j} = 0 \] (9)

Eq. (10) gives the result of the fitting

\[
\hat{\theta}_j = \left[ \begin{array}{cccc}
    b_{11} & b_{12} & \cdots & b_{1N_j} \\
    b_{21} & b_{22} & \cdots & b_{2N_j} \\
    \vdots & \vdots & \ddots & \vdots \\
    b_{N_j1} & b_{N_j2} & \cdots & b_{NN_j}
\end{array} \right] \left[ \begin{array}{c}
    \sum_{n=1}^{N_j} x_{(n)} \bar{y}_{(n)} \\
    \sum_{n=1}^{N_j} x_{2(n)} \bar{y}_{(n)} \\
    \vdots \\
    \sum_{n=1}^{N_j} x_{N_n(n)} \bar{y}_{(n)}
\end{array} \right] 
\] (10)

where the coefficient \( b_{ij} = \sum_{n=1}^{N_j} x_{(n)} x_{n(i)} \). Moreover, the most probable value for the fitting error \( \hat{\sigma} \) is given by Eq. (11).

\[
\hat{\sigma}^2 = \min \left[ J_s(\theta_j | D, M_j) \right] = J_s(\hat{\theta}_j | D, M_j) 
\] (11)

\( H_j(\hat{\theta}_j) \) is Hessian Matrix of the objective function \( J(\theta) = -\ln[p(D | \theta, M_j) p(\theta | M_j)] \) evaluated at \( \hat{\theta}_j \).

The factor \( p(\hat{\theta}_j | M_j)(2\pi)^{-N/2} |H_j(\hat{\theta}_j)|^{1/2} \) in Eq. (8) is called the Ockham factor (Gull 1988) and it is an evaluation of the robustness of the model class \( M_j \). It penalizes the model classes with parameters that are highly sensitive to the measurement, which means a very little variation in the fitting coefficients will cause a relatively large change in the prediction. A model class with higher Ockham factor is more robust. In other words, it is less sensitive to the measurement and system noise. More details of the Bayesian model class selection can be found in Yan et al. (2009) and Yuen (2010).

Due to different combinations of the control variables, there are 15 model classes (MCs) for each breach parameter in each form. Eq. (12) and Eq. (13) show the generalized formulas for the prediction equations.

\[
y_i = \sum_{j \in P} \alpha_j x_j + \alpha_0
\] (12)

\[
y_i = \beta_0 \prod_{j \in P} x_j^{\beta_j}
\] (13)

where \( y_i \) represents the breach parameter to be predicted in vector \( y = [H_s, B_s, B_{sd}, Q_p, T_j]^T \) . \( P \) denotes the set of selected variable index of model \( M_j \) and therefore, \( x_j \) represents the selected control variable from the vector \( x = [H_s, H_u, V_d, V_s]^T \). Table 1 shows the control variables in each model candidate.

| Table 1. Model class list |
|---------------------------|
| Model Class ID | Selected Variables | Model Class ID | Selected Variables |
|-----------------|--------------------|----------------|--------------------|
| 1               | \( H_s \)         | 9              | \( H_u, V_u \)    |
| 2               | \( H_s \)         | 10             | \( V_d \)         |
| 3               | \( V_d \)         | 11             | \( H_s, H_u, V_d \) |
| 4               | \( V_s \)         | 12             | \( H_s, H_u, V_u \) |
| 5               | \( H_s, H_u \)    | 13             | \( H_s, V_d, V_u \) |
| 6               | \( H_s, V_d \)    | 14             | \( H_u, V_d, V_u \) |
| 7               | \( H_u, V_u \)    | 15             | \( H_s, H_u, V_d \) |

Table 2 shows the uniform priors range used in this paper for both additive and multiplicative forms and it is sufficiently wide to cover the optimal values.

| Table 2. Priors for additive and multiplicative forms |
|---------------------------------|
| \( \alpha_0 \) (or \( \beta_0 \) ) | \( H_s \) | \( H_u \) | \( V_d \) | \( V_u \) |
| Add. | (-100,100) | (-10,10) | (-10,10) | (-10,10) |
| Mult. | (-10,10) | (-5.5) | (-5.5) | (-5.5) |

4 ANALYSIS AND DISCUSSION

4.1 Most plausible model

Following the procedures of the Bayesian model class selection framework, for each breach parameter, the values of plausibility and Ockham factor can be obtained for every model class candidate. Fig. 2 shows
the result of the prediction of $H_b$ in the additive form. The coefficient of determination (denoted as $R^2$), which represents the goodness-of-fit of the candidate model class, was also included. Log-form of Ockham factor was used to facilitate the comparison.

It can be seen that Model Class 15 shows the highest $R^2$ value yet the lowest Log-Ockham factor. It means that although it has the best goodness-of-fit (minimum fitting error), it exhibits the least robustness. The result is understandable since Model Class 15 has the largest number of control parameters (i.e., more fitting coefficients) which offer the model more flexibility to "fit" the measurements, including the modeling error and measurement noise. In other words, the model is more sensitive to the measurements, which leads to deterioration of the robustness in the prediction. As illustrated in Fig. 2, instead of Model 15, which has the highest goodness-of-fit, and Model 1, which has the highest robustness, Model Class 5 is the most plausible candidate for the prediction of $H_b$ in the additive form.

Similarly, the analysis result of prediction of $H_b$ in the multiplicative form is given in Fig. 3. Model 15 with the most control variables was selected as the most plausible model for the prediction in the multiplicative form.

By the same procedure, the most plausible models for other four breach parameters in both additive and multiplicative forms can also be obtained. The selected control variables and $R^2$ values for the plausible models are summarized in Table 3. It is noted that the $R^2$ values for multiplicative form are calculated by the original values of the measurements instead of the natural logarithmic values.

4.2 Comparison between additive and multiplicative forms

As shown in Table 2, it is not necessary that the same set of control variables is selected for the same breach parameters in the two forms. Taking $H_b$ as an example, only $H_d$ and $H_v$ are selected for the prediction in the additive form. However, in the multiplicative form, all the four control variables are required. On the other hand, the same set of control variables was selected in both forms for the predictions of $Q_p$ and $T_f$. By comparing the $R^2$ values for each breach parameter in the two forms, it is suggested that the predictors in additive form are suitable for $H_b$ while the predictions made from the multiplicative form are suitable for $B_{ave}$ and $B_t$. The $R^2$ values for the predictions of $Q_p$ from both forms are very similar, so it is difficult to decide which form is superior. Furthermore, $R^2$ values for predicting $T_f$ in both forms are noticeably lower than that of the other breach parameters, which suggests that $T_f$ is weakly correlated to the studied control variables.

5 CONCLUSIONS

A Bayesian probabilistic approach is adopted to revisit the derivation of empirical formulas predicting the breach parameters. A database compiling 141 cases of earthenfill dam failure is analyzed. Model classes in both additive and multiplicative form are considered. This novel framework not only considers the
goodness-of-fit but also takes into account the robustness of each candidate model when selecting the most suitable formula. The most plausible prediction formula for $H_s$, $B_{ave}$ and $B_r$ is proposed among the model classes. It is found that neither the same control variables will be selected in the prediction equation having different forms nor the most complicated form will be selected. Furthermore, $T_f$ is found to be weakly correlated to the suggested control variables.

REFERENCES

1) Beck, J. L., and Yuen, K. V. (2004). Model selection using response measurements: Bayesian probabilistic approach. Journal of Engineering Mechanics (ASCE), 130(2), 192-203.

2) Chiu, C. F., Yan, W. M., and Yuen, K. V. (2012). Estimation of water retention curve of granular soils from particle-size distribution – a Bayesian probabilistic approach. Canadian Geotechnical Journal, 49(9), 1024-1035.

3) Costa, J. E. (1985). Floods from Dam Failures. U. S. Geological Survey Open-File Report 85-560, Denver, Colorado, pp 54.

4) Froehlich, D. C. (1995). Peak outflow from breached embankment dam. Journal of Water Resources Planning and Management, 121(1), 90-97.

5) Froehlich, D. C. (2008). Embankment dam breach parameters and their uncertainties. Journal of Hydraulic Engineering, 134(12), 1708-1721.

6) Gull, S. F. (1988). Bayesian inductive inference and maximum entropy. Maximum-entropy and Bayesian methods in science and engineering, Kluwer Academic, Boston, 53-74.

7) ICOLD (1988). World register of dams: 4th updating. International Commission on Large Dams, Paris.

8) Osnos, E. (2011). Faust, China, and Nuclear Power. The New Yorker, October 12, 2011.

9) Papadimitriou, C., Beck, J. L., and Katasfygiotis, L. S. (1997). Asymptotic expansions for reliability and moments of uncertain systems. Journal of Engineering Mechanics, 123(12), 1219-1229.

10) Petrascheck, A. W., and Sydler, P. A. (1984). Routing of dam break floods. International Water Power and Dam Construction, 36, 3.

11) Reisner, M. (1993). Cadillac Desert: The American West and its Disappearing Water. Penguin Books.

12) Singh, K. P., and Snorrason, A. (1984). Sensitivity of outflow peaks and flood stages to the selection of dam breach parameters and simulation-models. Journal of Hydrology, 68(1-4), 295-310.

13) Wahl, T. L. (1998). Prediction of Embankment Dam Breach Parameters: a Literature Review and Needs Assessment. Dam Safety Rep. No. DSO-98-004, Bureau of Reclamation, U.S. Dept. of the Interior, Denver.

14) Xu, Y., and Zhang, L. M. (2009). Breaching parameters for earth and rockfill dams. Journal of Geotechnical and Geoenvironmental Engineering, 135(12), 1957-1970.

15) Yan, W. M., Yuen, K. V., and Yoon, G. L. (2009). Bayesian probabilistic approach for the correlations of compression index for marine clays. Journal of Geotechnical and Geoenvironmental Engineering, 135(12), 1932-1940.

16) Yuen, K.-V. (2010). Bayesian Methods for Structural Dynamics and Civil Engineering, John Wiley & Sons, Singapore.