Available online at www.sciencedirect.com

ScienceDirect

Journal of Hydro-environment Research 8 (2014) 292–303

Journal of Hydro-environment Research

www.elsevier.com/locate/jher

Research paper

Application of copula method and neural networks for predicting peak outflow from breached embankments

Farhad Hooshyaripor [a], Ahmad Tahershamsi [a,*], Saeed Golian [b]

[a] Faculty of Civil and Environment Engineering, Amirkabir University of Technology, Tehran, Iran
[b] School of Civil and Architectural Engineering, Shahrood University of Technology, Shahrood, Iran

Received 29 June 2012; revised 20 November 2013; accepted 21 November 2013
Available online 1 December 2013

Abstract

The limited number of available data is a common problem in most hydrologic and hydraulic studies, typically dam breach analysis. Construction of a probabilistic model is a key step in most decision making analyses to overcome such limitation. To analyze peak outflow from breached embankments, this paper has utilized two sets of data, original and synthetic datasets. Original datasets were collected from numerous historical dam failures and synthetic datasets were generated by copula method after incorporating the dependence structure among effective variables (height and volume of water behind the dam at failure and peak outflow discharge). The databases were separately employed to train two artificial neural networks (ANNs) as well as two statistical relations. Analyzing the results showed that the ANN model trained with synthetic datasets was the most competitive model for predicting peak outflows having R^2 of 0.96 and 0.95 for calibration and testing steps, respectively. The other ANN model was also better than statistical relations with R^2 of 0.94 and 0.87 respectively for calibration and testing steps.

Keywords: Dam breach; Peak outflow discharge; Dimensional analysis; Empirical formula; Copula; Neural network

1. Introduction

Dams are built to improve the quality of human life by providing drinking water, supporting economic growth and diverting water for power generation, navigation, flood control, and irrigation. Due to possible risk associated with the downstream life and economical losses, dam failure analysis is essential in the case of large dam design. Peak outflow rate due to dam failure is an important parameter in prediction of inundation water level for the emergency program planning and flood mitigation. Hence, dam failure studies have been of great interest among many researchers for several decades. In the last years, relevant experimental studies have been made with significant contributions such as laboratory tests on cohesive and non-cohesive embankments (Coleman et al., 2002; Gaucher et al., 2010). These are done to comprehend complexity of the phenomenon and also for achieving experimental data for application in analytical, numerical and statistical studies. Moreover, some investigations were focused on studying the flood wave propagation due to dam breaks using solution of Saint–Venant equations, analytically and/or numerically (Ponce et al., 2003; Tsai, 2005; Xia et al., 2010). Numerical solutions of the shallow water equations have also been used to predict flood inundation extent, flood hydrograph, and velocity distributions due to dam break flows (Liang et al., 2007; Gallegos et al., 2009). Despite a great deal of research on numerical hydrodynamic models, they are not entirely reliable and can only give guidance for better understanding of the phenomenon (Carling et al., 2009).

For predicting the hydrograph, a simple approach is to simulate generation of the outflows as the emptying of a

* Corresponding author. Tel.: +98 917 791 1364; fax: +98 2164543013.
E-mail addresses: hooshyarypor@gmail.com (F. Hooshyaripor), tshamsi@aut.ac.ir (A. Tahershamsi), s.golian@shahroodut.ac.ir (S. Golian).

1570-6443/$ - see front matter
http://dx.doi.org/10.1016/j.jher.2013.11.004

reservoir through a weir (Walder and O'Connor, 1997). A more realistic approach takes into account the eroding flow capacity as a function of the mean shear stress or average flow velocity on the breach (Macchione, 2008). This approach may be appropriate for some stages of the breach process, but is not consistent with the mechanics of much of the breaching process as observed in the field and in the laboratory (Wahl, 1998).

Statistical modeling is another approach which is based on analysis of the datasets of historical dam failures. In this approach, the estimation of flow characteristics and breach geometry are described as functions of a number of representative quantities of the dam and the reservoir, such as the water depth and the storage volume of the reservoir (USBR, 1982; Evans, 1986; MacDonald and Langridge-Monopolis, 1984; Xu and Zhang, 2009; Pierce et al., 2010). The equations which are derived based on case studies will be more preferable, if a database of numerous well-documented dam failure cases exists (Xu and Zhang, 2009). Since the 1970s, various empirical formulas have been developed using regression analysis (Table 1).

Froehlich (1995) evaluated and compared several empirical equations and proposed a regression-base equation. Based on his analysis, Wahl (1998) suggests that the Froehlich equation is one of the best methods for direct prediction of peak outflow. However, such equations have significant uncertainty (Wahl, 2004) due to limited datasets and their low flexibility to data variations. Moreover, the existing databases include few cases of large dam failures, so they don't encompass a broad range for involved variables (Nourani et al., 2012). Therefore, the equations which are developed based on such limited databases cannot be reliable in most cases. To overcome the limitation of datasets, it may be helpful to generate a proper set of synthetic datasets using advanced statistical methods. Gupta and Singh (2012) assumed that H_w and V_w are independent and randomly generated a set of data having statistical characteristic similar to real dam breach case studies. Recently, using copula method has been increased because of its ability to model the dependence structure between variables. Favre et al. (2004) mentioned the advantages of using copulas in modeling the dependence between random variables in comparison with other traditional methods. De Michele and Salvodari (2003) utilized copulas to model the dependence between rainfall duration and intensity and declared that, in this way, both properties of marginal distributions (assumed to be of Generalized Pareto type) and dependence between storm duration and intensity were preserved. Salvadori and De Michele (2010) introduced a new model for multivariate return period analysis using copula method. Golian et al. (2011) considered spatial dependence of rainfall fields in the derivation of rainfall threshold curves. The structure of the spatial dependence among sub-watersheds' rainfalls was taken into account under different scenarios, namely bivariate copula and multivariate Gaussian copula. Golian et al. (2012) investigated the joint response of key hydrologic variables, including total precipitation depths and the corresponding simulated peak discharges, for different antecedent soil moisture conditions using copula method.

The present study uses copula method to generate a comprehensive set of synthetic datasets for water height, reservoir capacity, and relevant dam breach discharge. The resulting database should be analyzed with a proper method. Artificial neural network (ANN) is an alternative approach due to its simplicity and generalizing ability. ANN has a number of advantages over statistical models (e.g. its data-driven nature, model–free form of predictions, and tolerance to data errors) (Azmathullah et al., 2005). ANN has been widely used in the field of hydraulic engineering. Babaeyan Amini et al. (2011) compared the performance of ANN and some empirical formulas for estimation of peak outflow from breached embankment dams. Obtained results demonstrated the higher performance of ANN. They declared that the ANN predictions outperformed the estimates obtained by empirical formulas conventionally used in the literature and in the current engineering practices. They also mentioned that the datasets were too limited to model the phenomenon and it would better to use more experimental and historical datasets. Extending the datasets with experimental data and mathematical models-based outputs along with real breaches data, Nourani et al. (2012) used ANN to predict peak outflow from laboratory models and real breached embankments. Results showed that the ANN predictions are in good agreement with the observed values and the presented method produces better results than existing classical methods. However, they finally stated that due to the complexity of the process on one hand and the lack of sufficient data on the other hand, it cannot be claimed that the presented ANN model is a perfect approach to simulate the breach phenomenon. In another investigation, Hooshyaripor and Tahershamsi (2012) collected 93 case study data and evaluated the performance of different training algorithms, back propagation (BP) and imperialist competitive algorithm (ICA), in ANN modeling. Among different BP algorithms, Levenberg–Marquardt (LM) had the best performance and ICA was better than BP algorithms. Furthermore, they stated that lack of extreme datasets affects the performance of ANNs, so enlarging the database by adding values of new

Table 1
Empirical equations for peak outflow prediction.

Investigator	R^2	No. of case studies	Equation
SCS (1981) for dams>31.4 m	Not available	13	$Q_p = 16.6(H_w)^{1.85}$
USBR (1982)	0.724	21	$Q_p = 19.1(H_w)^{1.85}$
Evans (1986)	0.836	29	$Q_p = 0.72(V_w)^{0.53}$
MacDonald and Langridge-Monopolis (1984)	0.788	23	$Q_p = 1.154(H_w V_w)^{0.412}$
Froehlich (1995)	0.934	22	$Q_p = 0.607(H_w^{1.24} V_w^{0.295})$
Pierce et al. (2010)	0.850	87	$Q_p = 0.038(H_w^{1.09} V_w^{0.475})$

Note: Q_p = peak outflow discharge (m^3/s); H_w = height of water behind the dam (m); and V_w = volume of water behind the dam at the failure time (m^3).