Decision support system for predicting flood characteristics based on database modelling development (case study: Upper Citarum, West Java, Indonesia)

I. K. Hadihardaja[1], D. Indrawati[1], Y. Suryadi[1] & N. S. Griggs[2]
[1]*Faculty of Civil and Environmental Engineering,*
Institut Teknologi Bandung, Indonesia
[2]*Department of Civil and Environmental Engineering,*
Faculty of Engineering, Colorado State University (CSU), USA

Abstract

Floods, one of classical disaster occurrences, are difficult to identify accurately; most of the incidents have no recorded predictions and floods are generally frequent events with poor data acquisition. Development of a flood database system has become important for improving the management of information with regard to a flood early warning system for vulnerable communities along the flood area. In the case of the Upper Citarum river basin, flooding has occurred since the 1960s and was still happening up until 2010, even though a lot of study has been done and infrastructure has been built. Numerical modelling (using hydrology and hydraulics model) are usually employed to simulate the flood area (Ag), flood depth (hg), and travel time (tpr). However, numerical modelling for flood prediction is too time-consuming to be useful as an early warning system for the mitigation of flood-related damage and loss. Therefore, this model was used to develop a flood database system, based on hypothetical data, in order to develop a recognition pattern for a neural network learning process that will improve the speed and accuracy of flood prediction. To improve the accuracy of numerical modelling and observation, an adjustment was established for an advanced training process which used a generalized regression neural network. In other words, the training and testing data sets were obtained by correcting near-field hydrograph numerical models from hypothetical rainfall.

The case study was performed on the Upper Citarum river basin in West Java, Indonesia.

1 Introduction

In line with global climate change, rain with an intensity of more than 200 mm per day causes floods in many countries in the world, from China, Bangladesh, Taiwan, Singapore and Indonesia. During floods in China 701 people died and 347 were lost, as well as the destruction of infrastructure and moral hazards. The condition is the same with Indonesia and other countries, that the floods cause varying amount of loss such as moral, material and of peoples' lives. The essential fact of these disasters is not only can they not be predicted, but that they cause destruction and tragedy for people and the infrastructure all around the world. The rain prompts a flood and devastation for many communities along floodplains and floodlines.

The major reasons for these disasters are a lack of data and minor capability regarding a flood early warning system, especially with climate change phenomena. Thus, when rainfall becomes unpredictable, the water drainage system consisting of rivers and tributaries of natural and artificial drainage channels or existing flood reservoirs cannot accommodate the accumulation of rain water and make an overflow. The present study aims to develop a flood warning database system for improving decision making in the issuance of an early warning. The impact of climate change on flood events is infrequent, with most incidents passing with poor or no data acquisition. Therefore, it is necessary to develop a database system to analyse flood behaviour to obtain more accurate predictions.

This paper has developed technique of hard computing for learning flood behaviour from several rainfalls and time of rainfall data. The proposed technique was also examined based on historical flood events. In this case, numerical methods were established to develop a database management system after calibration and adjustment based on observations. The developed numerical modelling is the basis for generating potential flood events from numerous instances of rainfall and the time of rainfall, and to simulate the flood area, flood depth, and travel time of flood. Because numerical modelling is too time-consuming to act as a flood warning system, neural network modelling was applied for learning flood behaviour using a database developed from the numerical modelling data. This study is specifically related to flood events in the Upper Citarum watershed, West Java, Indonesia.

2 Case study

Upper Citarum is one of the commodious watersheds in West Java, Indonesia with an area of 172,100 ha. Flooding in the Upper Citarum River Basin is an area of frequent disaster since the 1960's and has become an important project of the

Indonesian government until the present time. This river basin resides in five residences in West Java, with nine rainfall gauge station whose data has influenced the system. The observed rainfall from nine rainfall gauge station was the point of rainfall. Thus, it was transformed to become a uniform rainfall using Thiessen Polygon method.

In this paper, beside maximum daily rainfall, models also used series rainfall because the maximum daily rainfall did not produce the flood area the same as the observation flood and a broad flood area was produced by series rainfall. So, that series rainfall is most likely happen in the future.
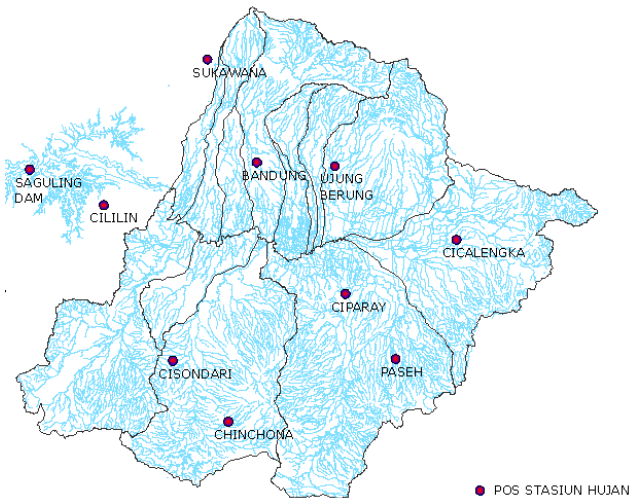

Figure 1:   Upper Citarum River Basin [1].

The rainfall and a discharge database system using data from September 2001 until August 2002 was twice calibrated. First was using discharge in Nanjung station and the second using flood event data on November 2001, and January and March 2002, and multivariate it into a wide system of flood database. The evaporation is using the same time data and the topography using 2005 data.

3   Hard computing technique within flood numerical modelling

Flood prediction is an emergency unit for the operational guideline of decision making. An early warning system for floods provide for immediate action through regional authorities to mitigate the damage from a predicted major inundation and flood area in a floodplain. More people lives and properties can be saved effectively when the warning system can send on information about flood characteristics immediately and accurately as soon as possible. The initial decisions of flood commonly are peak discharge, time peak and time base were

the generated characteristics of flood, like the flood area, flood depth, time of inundation, and travel time. However, this common method cannot be used for an early warning system because the generating flood characteristics need more time to build and the warning would arrive too late for evacuation measures and result in an ineffective local response to the flood. So, this paper is bidding a technology for predicting flood characteristics based on rainfall and the time of rainfall. This technology expects the user to input some of the rainfall and time of rainfall data, thus the flood characteristics can be the output and it can be very effective for an early warning system.

The hard computational method using numerical modelling until the present time has been of a high standard for flood studies. This modelling has a high precise simulation with robust performance in demonstrations based on field data. Carrivick [3] applied 2D hydrodynamics of SOBEK for predicting a high-magnitude outburst flood in Kverkfjoll, Iceland, and later Suryadi [1] applied numerical modelling of MIKEFLOOD for computing flood characteristics in the Upper Citarum river basin, West Java, Indonesia.

This model uses numerical modelling of MIKEFLOOD and divides into a two step method. First, from rainfall data, evaporation and watershed properties results in the peak discharge, time peak and time base, and step two with that data, the model is generated of flood characteristics in two dimensional modelling. In the first step, this paper using NAM model which is one of the modules from MIKE to calculate storm hydrograph. Thus, the storm hydrographs were generated into one dimensional modelling in the river using MIKE11 and two dimensional floods in the land with MIKE21. Furthermore, both of the results use MIKEFLOOD for generating floods into the river basin area.

4   Soft computing technique within neural network modelling

In the late 1980s, the artificial neural network (ANN) was used to solve problems in civil engineering [6]. Meanwhile application to simulating and forecasting problems in water resources is relatively recent [7, 9, 11]. ANN modelling has a specific characteristic that can solve problems without any prior assumption and will extract any irregularities or patterns that may exist and use them to form a relationship between input and output. Additional benefits include data error tolerance and the characteristic of being data-driven, thereby providing a capacity to learn and generalise a pattern in noisy and ambiguous input data. All this time NN was applied for modelling of rainfall runoff using GRNN and back propagation [13], prediction height of Tsunami [8], modelling suspended sediment estimations in Juniata catchment in USA [5], and various other applications.

5   GRNN for learning database model

After the flood characteristics resulting from hard computing processing, the next step is learning neural network which uses generalised regression as

the learning algorithm. In this study, that system developed by Delphi 7.0 as a compiler and the algorithms are continued from GRNN-ITB that is predicting height of Tsunami by Hadihardaja and Mulya [8]. Specht [12] was the first build GRNN that is neural network architecture capable of solving any problem of estimating a probability distribution function. The learning process of GRNN is similar with forecasting for finding optimal regression with nearest approximation.

The basic theory of GRNN explain by Amrouche and Rouvaen [2] where a function of regression was performed by X as independent variable for finding dependent variable Y based on observations data X with associated values of Specht [10, 12] provided same equation as follow which based on sample values X^i and Y^i of the random variables x and y:

$$\hat{f}(X,Y) = \left[\frac{1}{(2\pi)^{(p+1)/2} \sigma^{p+1}} \right] \frac{1}{n} \sum_{i=1}^{n} \exp\left[-\frac{(X-X^i)^T (X-X^i)}{2\sigma^2} \right] \exp\left[-\frac{(Y-Y^i)^2}{2\sigma^2} \right], \qquad (1)$$

where n is the number of observations data, p is the dimension of the vector variable, σ is the smoothing factor, and Y^i is the desired scalar output given the observed input X^i. The denumerator D_i^2 is defined by:

$$D_i^2 = (X-X^i)^T (X-X^i). \qquad (2)$$

The desired conditional mean is produced by combining eqs. (1) and (2) and interchanging the order of integration and summation, with N for number of input pattern, D_i^2 is an Euclidean square from the new and training input pattern, and σ is a smoothing factor as follows:

$$\hat{Y}(X) = \frac{\sum_{i=1}^{N} Y^i \exp\left(-\frac{D_i^2}{2\sigma^2} \right)}{\sum_{i=1}^{N} \exp\left(-\frac{D_i^2}{2\sigma^2} \right)}, \qquad (3)$$

This equation known as the Nadaraya-Watson kernel regression estimator and for problems involving numerical data that equation is directly applicable. Specht [10] was visualisation weighted average for all observed data by $\hat{Y}(X)$ with weighted exponentially based on Eucledian distance from X.

6 Performance of NN

For appraise the stability of NN performance, this study using Root Mean Square Error (RMSE) which has the formula as follows:

$$RMSE = \sqrt{\frac{\sum_{p=1}^{P}(t_p - H_p)^2}{P}} \tag{4}$$

where P is the total number of input-output data patterns, t_p is the target output and H_p is the predicted output.

7 Neural network architect

Implementation of generalised regression in neural network performance proposed by Specht [10, 12] and Amrouche and Rouvaen [2] using the equation:

$$y_j = \frac{\sum_{i=1}^{n} w_{ij} h_i}{\sum_{i=1}^{n} h_i} \tag{5}$$

$$h_i = \exp\left[-\frac{D_i^2}{2\sigma^2}\right] \tag{6}$$

where w_{ij} is the target output, x_i is input training vector and J^{th} as the output.

GRNN has specialist architect (figure 2) which consists of:

1). Input layer which fully connected with pattern layer.
2). Pattern layer which has one neuron for each pattern that is calculating h_i (σ,C_i) function and expressed in equation (2) and using C_i center.
3). Summation layer which has two units N and D. First unit has input weight equal with X^i, then calculate the numerator (N) with summation exponential form multiplied by Yi which has relationship with Xi. Both of them have input weight equal to one, with denominator D is summary of all exponential form.
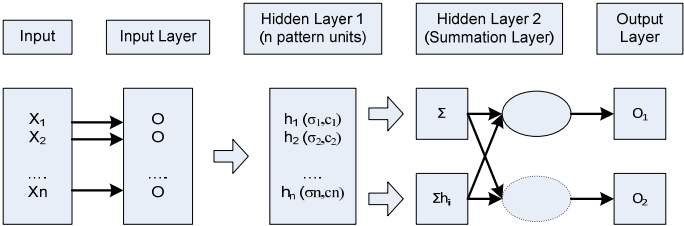4). The prediction outcome provided by output unit divided by N and D.

Figure 2:   Architect of GRNN [2].

One of differentiation GRNN from another NN is smoothing factor σ. If σ is small that meaning only a few of samples play in role and when σ become large that meaning estimation become very smooth for samples. Thus, when σ has infinity value, the meaning $\hat{Y}(X)$ is the average of Y^i and has independent input.

8 Model input adjustment

For the numerical modelling, there is a two step algorithm for generating the flood characteristics. First rainfall and time of rainfall were generated for predicting time base, time peak and peak discharge of storm hydrograph. Second, the results were generated to obtain flood area, flood depth, time of inundation, and travel time, known as flood characteristics. For the first generated, this paper using NAM Model with rainfall and evaporation from September 2001 until August 2002 data from twelve rain gauge stations and being calibrated with discharge data from Nanjung station downstream of the Upper Citarum river basin at the same time. Meanwhile, for land use of soil, this paper using 2001 data where it is provided by BAPEDDA in West Java Province. For the second generated, this paper using MIKEFLOOD with topography data of Upper Citarum river basin in 2005. The topography was converted from arcview format (*.shp) to bathymetry file (dfs2). The space between grids in this hydrodynamics modelling using 50 m x 50 m and totally has 1100 x 600 grids. The grid seems too large for predicting flood characteristics accurately but by adjustment of the grid into a smaller size, the result of modelling needs a very large memory and high resolution of computer considering the area of this river basin is very broad.

And for GRNN modelling, the input is a state space denoted by X (rainfall and time of rainfall). The estimated, or target, value $\hat{Y}(X)$ is the flood characteristics number of observation sub river basin (12 points). Total data from the neural training is combination of input and target data, thus there are 105.238 rows of data, with 90% and 10% of the combinations used for neural network training and testing, respectively.

Table 1:   Input for GRNN.

Input	Number	Minimum	Maximum
Sub river basin	12	-	-
Rainfall	14	171	2997
Time of rainfall	14	1	303

9 Result and discussion

This study generated the flood database by numerical modelling which gave a satisfactory result. For the first step when rainfall and time of rainfall being generated by NAM model become storm hydrograph, root square value of

database got up to 0.80 for eight month data. And for the next step, the location of floods in the model resembled the observation flood data along November 2001, January 2002 and March 2002. But there is an inadequacy of observation data for the flood area, so the verification of the flood area just per sub-sub district without exact area of flood. And for this verification, two dimensional modelling has precision result.

The next step is to enlarge input rainfall and time of rainfall with various patterns for developing a flood characteristics database. The results of modelling are shown in Figures 3 and 4.

As seen in figure 3, the flood can be derived by three spatial location also called area 1, area 2 and area 3 (see fig. 4). Area 1 is the first flood and continued by other in line with rainfall increasing. And the spatial flood propagation not only in area side but also discharge and flood depth.
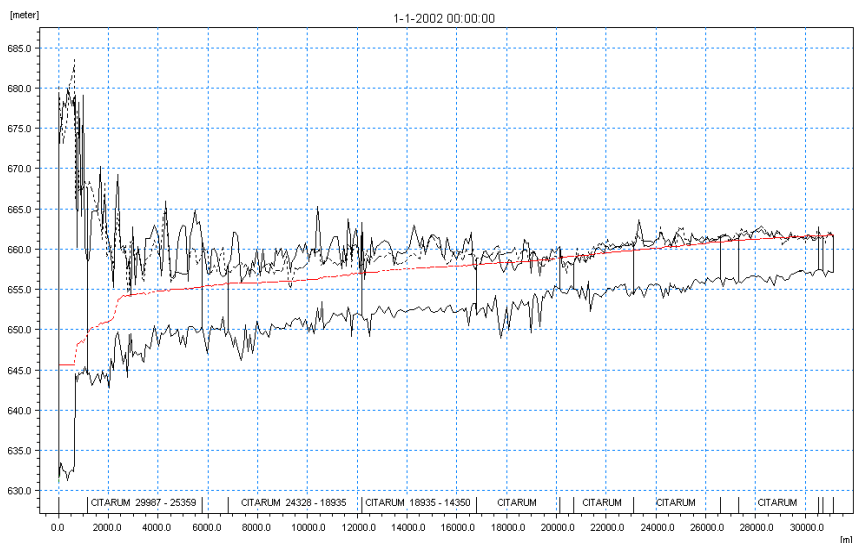

Figure 3:   Two dimensional modelling of Upper Citarum River Basin.

10 Flood characteristics database

After hard computing simulation, the next step is generated soft computing using GRNN formula. With the various simulations the result of Numerical modeling compared with GRNN modeling and the result are illustrated in figure 5: a database system was developed using a numerical model generated in this study based on rainfall and time of rainfall which generated the flood, which were necessary to have large size data required for the database. This limitation of the GRNN is because the database system development generally needs a large amount of observation data for accuracy and learning process which cannot yet be provided in this paper.
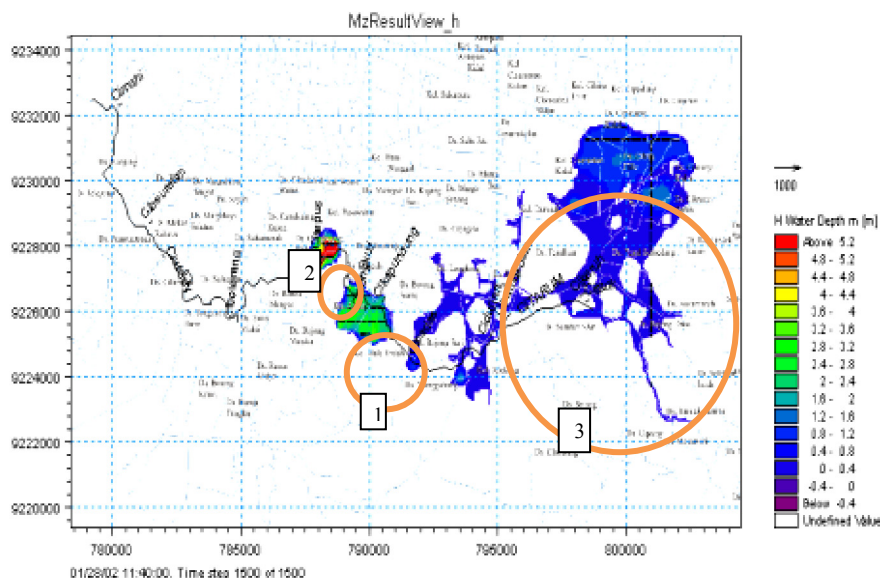
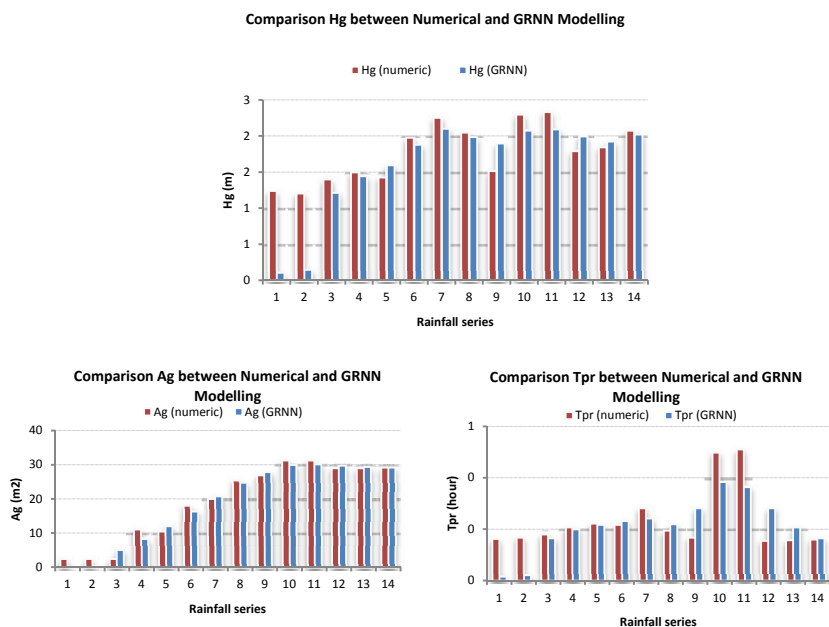Figure 4:   Flood simulation for 2D modelling.

Figure 5:   Comparison of flood depth (hg), flood area (Ag) and travel time (tpr).

But the GRNN give satisfactory results where average error of data is 6.12%. For each root mean square error are flood depth (hg) is 6.73%, flood area (Ag) is 2.20% and travel time (tpr) is 10.16%.

11 Conclusion and future work

The study indicates the possibility of predicting flood characteristics of flood area, flood depth, and travel time in river basin and has an accurate and fast time of fashion by developing a flood database model and associated neural network. A GRNN was successful in the generalisation of the target function based on few data set.

For future work, because in this paper the grid of DEM relatively broad and for detail design of flood database system need small grid, furthermore the flood database system can be developed with minimize grid size without divide into sub river basin because them are influent each other. It will not have escaped this model not to include climate change phenomena, so for future work the GRNN must build into an extrapolation learning database. Besides that, another flood characteristic that must be learned was the time of inundation.

Acknowledgements

References

[1] Suryadi, Yadi., *Pengaruh hujan terhadap debit puncak hidrograf inflow, luas genangan, kedalaman genangan dan waktu genang dalam menentukan indeks banjir*, Dissertation, 2008
[2] Amrouche, A., and Rouvaen, J. M., *Efficient System for Speech Recognition using General Regression Neural Network*, International Journal of Intelligent Systems and Technologies 1;2 © www.waset.org, 2006
[3] Carrivick, *Application of 2D hydrodynamic modelling to high-magnitude Outburst floods: An example from Kverkfjoll, Iceland*, Journal of Hydrology 321(2006) 187–199, 2005
[4] Cigizoglu, H.K., *Application of Generalized Regression Neural Networks to Intermittent Flow Forecasting and Estimation*, Journal of Hydrologic Engineering, ASCE, Vol. 10, No. 4, pp. 336–341, 2005
[5] Cigizoglu, H.K., and Alp, M., *Generalized Regression Neural Network in Modeling River Sediment yield*, Advances in Engineering Software,

Elsevier Science Ltd. Oxford, UK, UK, ISSN:0965-9978, Volume 37, Issue 2, Pages: 63–68, 2006

[6]   Flood, I., and Kartam, N., *Neural networks in civil engineering –I Principles and understanding*, J. Computing in Civil Engineering., ASCE, 8(2), 131–148, 1994

[7]   French, M.N., Krajewski, W.F., and Cuykendall, R.R., *Rainfall forecasting in space and time using a neural network*, J. Hydro.137, 1–31, 1992

[8]   Hadihardaja, I.K dan Mulya I.E., *Decision Support System for Predicting Tsunami Characteristics along Coastline Areas Based on Database Modelling Development*, Journal of hydroinformatics, 13.1.2011, 209

[9]   Hsu, K., Gupta, H.V., and Sorooshian, S., *Artificial neural networks modeling of the rainfall-runoff process*, Water Resour. Res., 31(10), 2517–2530, 1995

[10] Specht, D. F., *A General Regression Neural Networks*, IEEE Trans. on Neural Networks, Vol. 2, no. 6, pp. 568–576, Nov, 1991

[11] Thirumalaiah, K. and Deo, M., *River stage forecasting using artificial neural networks*, Journal of Hydrologic Engineering 3(1), 26-32, 1998

[12] Specht, D.F., *Probabilistic Neural Networks and General Regression Neural Networks*, Fuzzy Logic and Neural Network Handbook, Chap 3. McGraw Hill Inc., 1995

[13] Hadihardaja, I.K., and Sutikno, S., *Rainfall And Stream flow Discharge Relation Using Artificial Neural Network*, International Conference on Soft Computing, Intelligent System and Information Technology (ICSIIT), Denpasar, Indonesia, 2007