Flash flood modeling using the artificial neural network (Case study: Welang Watershed, Pasuruan District, Indonesia

To cite this article: Suhardi *et al* 2020 *IOP Conf. Ser.: Earth Environ. Sci.* **419** 012123

ICCEE 2019 IOP Publishing
IOP Conf. Series: Earth and Environmental Science **419** (2020) 012123 doi:10.1088/1755-1315/419/1/012123

Flash flood modeling using the artificial neural network (Case study: Welang Watershed, Pasuruan District, Indonesia

Suhardi[1], E Hidayah[1], and G Halik[1]

[1]Departement of Civil Engineering, Faculty of Engineering, University of Jember, Indonesia

E-mail: suhardi89@gmail.com

Abstract. The Artificial Neural Network (ANN) has been widely used in flood modeling and has proven to be good accuracy. This research aims to flash flood modeling using ANN. The flash flood modeling was conducted at Welang Watershed, Pasuruan District, East Java, Indonesia. The input of flash flood using ANN consists of rainfall and runoff coefficient. The runoff coefficient was derived by the Normalized Difference Vegetation Index (NDVI) value from the Landsat 8 Operational Land Imager (OLI). The output ANN model was flash flood discharge. The ANN architecture model uses a backpropagation neural network. The period of training and testing model ANN using data from January to February 2017 period and November to December 2017 period, respectively. The Result of flash flood modeling with ANN showed the good of fitness pattern between output model and observation data.

1. Introduction

The flash floods were a sudden local flood with an unpredictable high peak [1,2]. One of the causes of flooding was the change of vegetation function into residential and other functions [3]. This change in land function causes the volume of runoff flow to increase while the infiltrated water was very low [4]. Changes in land use also affect the increase in volumes of runoff and flood discharge [5]. The structure of the rain-flow model can be built on two approaches: a theory-based approach (physical and conceptual based), and a data-driven approach (empirical and black-box). Physically-based models are based on a relevant understanding of hydrological systems based on first-order principles, phenomena or simulated-based models [6,7]. The conceptual model describes the physical mechanism of the hydrological cycle, without considering the spatial variability and stochastic nature of the rainfall and runoff process [6]. Simple mathematical equations are often used in conceptual models to describe hydrological processes such as evapotranspiration, infiltration, percolation, basal flow, and runoff [8]. Meanwhile, the data-driven model is data-driven modeling (DDM) based on data analysis to search for input and output relationships based only on a number of assumptions about the physical behavior of the system [7]. The accuracy of the concept model depends heavily on the accuracy of the model parameters [9]. Therefore, this concept model has not been able to represent real field conditions. On the other hand, Forecasting and flood warnings often use a physical model. However, for a conventional flood forecasting process using a rain-flow physical model requires extensive information and data, and includes uncertainty that can accumulate errors during the modeling process [10,11]. Based on the description, an appropriate flood prediction model is needed so that it can produce a reliable outcome model. Several studies have shown that ANN was excellent in flood modeling and shows a good model time response to forecasting flash floods in its limited data watershed [12]. ANN can also precisely predict hydrographs, but it is still necessary to create a good modeling view to obtain information about the aquifer physical condition [13]. The ANN modeling is easy to develop and does not require detailed parameters such as the hydrological extent and geological catchment area required in the application of conceptual models [14]. JST modeling with rainfall input data and observation discharge gives good results in river discharge prediction [15]. Similarly for future flood forecasting, the modeling of ANN with precipitation and previous discharge input data was better than using one variable data input, such as rain or debit data only [16]. Excellent flood modeling and highly favored was a model that does not require many parameters in its application but has good accuracy. Therefore, the application of Artificial Neural Network (ANN) was a decision based on the above considerations. The selection of ANN model to model the flash floods with rainfall input parameters and runoff coefficient was the right choice. The runoff coefficient resulted from the decrease of the Normalized Difference Vegetation Index (NDVI). NDVI was generated from remote sensing to identify vegetation density by interpreting satellite image maps [17]. NDVI can also be used as an approach to predict the value of runoff coefficient (C) effectively and efficiently [18]. This paper will discuss the reliability of the flash floods model using Artificial Neural Network (ANN) with variable input of rainfall and runoff coefficient (C) derived from Normalized Difference Vegetation Index (NDVI) value in Welang Watershed.

2. Research sites

The research was conducted in Welang Watershed District of Pasuruan Regency, East Java Province precisely -7°36’00” to -7°54’00” South Latitude and 112°36’00” to 112°54,00” East Longitude. Area of river basin 414,9 km2 with land use consists of vetagasi, settlement, road, and empty land. The location of the research can be seen in figure 1.

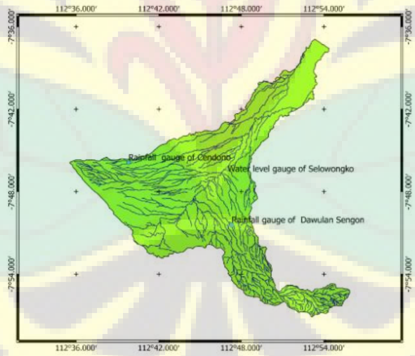

Figure 1. Map of Welang Watershed, Pasuruan Regency, Indonesia

3. Research methods

3.1 Research Data

This study uses secondary data in the form of rainfall data at 2 rain stations location (Cendono and Dawulan Sengon), observation discharge data (AWLR of Selowongko Station) and Landsat 8 satellite images. Training of ANN model used rainfall data for the period of January - February 2017 while model testing used data from November to December 2017. Previous research on estimation of runoff coefficient value using NDVI transformation of Landsat ETM + image used image data of recording resulted on December 22, 2001, and April 8, 2003. While hydrological data used data on the period of November 15, 2001 - January 28, 2002 [19]. Based on this, the satellite imagery period was adjusted to the needs of the research. In determining the period of satellite imagery always pay attention to the cloud conditions that cover it. Based on these conditions, the satellite image was not always the same month as the hydrological data period. Nevertheless, satellite imagery remains in the period of the rainy season as representative of the hydrological data period was May 7, 2017. The Landsat 8 satellite images were downloaded via EarthExplore USGS. The Landsat 8 satellite images were used to identify the vegetative density of the Welang Watershed. The runoff coefficient (C) value of the NDVI derivative with the widest area was used as the input data in the ANN application.

3.2 Research Stages

3.2.1 Conversion of runoff coefficient (C)

The NDVI calculation results from a comparison between Near Infrared and Red with the equation [20].

$$\text{NDVI} = \frac{(NIR-red)}{(\text{NIR}+red)} \tag{1}$$

NDVI calculations use bands 5 and 4, where band 5 as NIR and band 4 as Red. The value of vegetation canopy density based on NDVI analysis result was classified into three classes were rare, medium and lush density [21]. In general, the use of a combination of band Landsat 8 for color infrared calculation (vegetation) was 543 [22]. The best response of healthy plants to the electromagnetic spectrum was found in bands 5 and 4 [23]. Percentage of Watertight Surface Cover (PWS) and Percentage of Vegetation Density (PVD) was used to calculate the conversion of NDVI value to runoff coefficient (C) with the following conditions [24, 17]:

1. Percentage of Watertight Surface Cover, if NDVI < -0,0607

$$PWS = -63{,}16\text{x}^2 - 116{,}6\text{x} + 46{,}9 \tag{2}$$

x = NDVI value

$$\text{C} = 0{,}05 + 0{,}91\ \text{PWS} \tag{3}$$

C= runoff coefficient

2. Percentage of Vegetation Density (PVD) with runoff coefficient (C) calculated by the equation::

$$\text{C} = -\text{PVD} + 1 \tag{4}$$

C= runoff coefficient

The percentage of Vegetation Density was calculated based on the ratio between the width of the vegetation density class and the total area of vegetation density multiplied by 100 percent [21] Percentage of Vegetation Density in Welang Watershed can be calculated by equation

$$\text{PVD} = 37{,}705\text{x} - 1{,}596 \tag{5}$$

x = NDVI Value

3.2.2 Data processing

One method often used in time series data analysis to know the time lag response of discharge to the input of rainfall is cross-correlation method. A series relationship of input and output data in a system can be predicted by the cross-correlation method [25]. The highest cross-correlation value with positive time lag was a correlation to be considered in the grouping of time series data of rainfall and discharge. This high-correlation rain and discharge series data were used as input and target data in an ANN application. The calculation of data normalization using the min-max method, so that the input data and target scaled with a range of 0 to 1.

3.2.3 Artificial neural network

The ANN architecture uses the backpropagation learning method with weight improvement using Levenberg-Marquardt method. This Levenberg-Marquardt method was very effective when training the network with a lot of weight [26]. Levenberg-Marquardt backpropagation algorithm is often used

IOP Conf. Series: Earth and Environmental Science **419** (2020) 012123 doi:10.1088/1755-1315/419/1/012123

for network training [27]. Training automatically stops when the generalization process stops, this was usually indicated by an increase in Mean Square Error (MSE). The Mean Squared Error (MSE) is the average value of the quadratic difference between output and target. The lower the MSE value gets better while zero means no error[26]. The input layer consists of hourly rainfall data and the runoff coefficient (C), while the hourly discharge was the target. Data in matrix format m x n, where m the number of rows of data and n number of data columns. The distribution of data composition for the ANN modeling consists of 70% training, 15% validation, and 15% testing. The input data was rainfall data from 2 automatic rain gauge and the runoff coefficient (C) resulting from the conversion of the NDVI value. The NDVI value was converted to the runoff coefficient was the NDVI value with the widest vegetation within the Welang watershed. The rainfall data was 31 rows and 2 columns, while the runoff coefficient (C) and the target data were 31 rows and 1 column. The network architecture with 1 hidden layer with the number of neurons 20 as shown in figure 2.

.

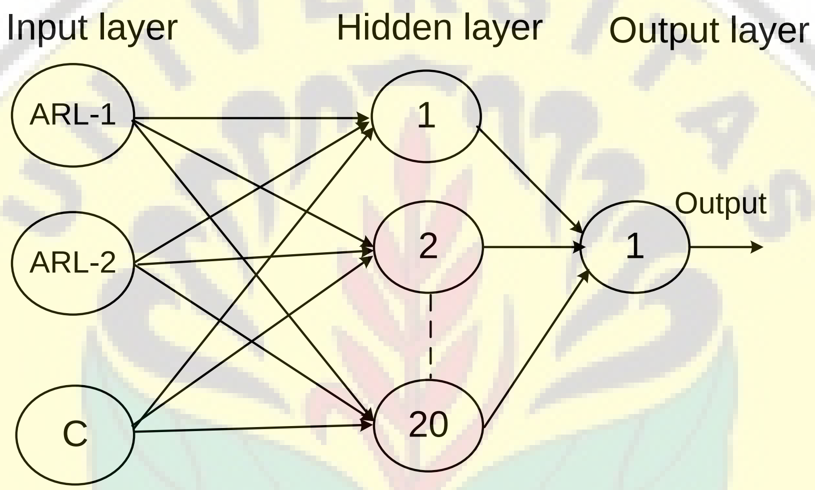

Figure 2. The network architecture of ANN

4. Results and discussion

4.1 NDVI conversion to a runoff coefficient (C)

NDVI value was generated from remote sensing using a satellite image of Landsat 8 OLI. Some stages in the determination of NDVI include: (a) cutting of satellite imagery, this cutting aims to limit the area of image analysis according to the location of the study; and (b) NDVI analysis by inputting the NIR data input with band 5 and Red with band 4. The result of NDVI analysis will give a different color images. This color difference identifies the value of NDVI. The higher the NDVI value indicates that the area of the vegetarian density was dense, but if the low NDVI values indicate rare vegetation density as shown in Fig 3. Figure 3 shows that the widest area was NDVI value of 0.47 ± 14,000 pixels. The lowest NDVI value was -0.01 and the highest was 0.60, so the NDVI value was converted to the runoff coefficient value by the Percentage of Vegetation Density (PVD) approach. The calculation result of PVD was converted to the runoff coefficient with equation 4. The Graph of NDVI relationship with a coefficient of runoff (C) in Welang Watershed as shown in figure 4.

The vegetation density class based on NDVI is grouped into 4 classes: (a) clouds -0.232 - 0.022; (b) rare vegetation 0.022 - 0.188; (c) moderate vegetations 0.188 - 0.398; and (d) dense vegetation 0.398 - 0.593 (Purwanto, 2015). Based on the description, it can be argued that welding watershed includes areas with dense vegetation with rarely runoff coefficient values of 0.81 to 0.98. This shows that the lower of the NDVI value so the runoff coefficient value greater than the run off in the middle part of the ocean.

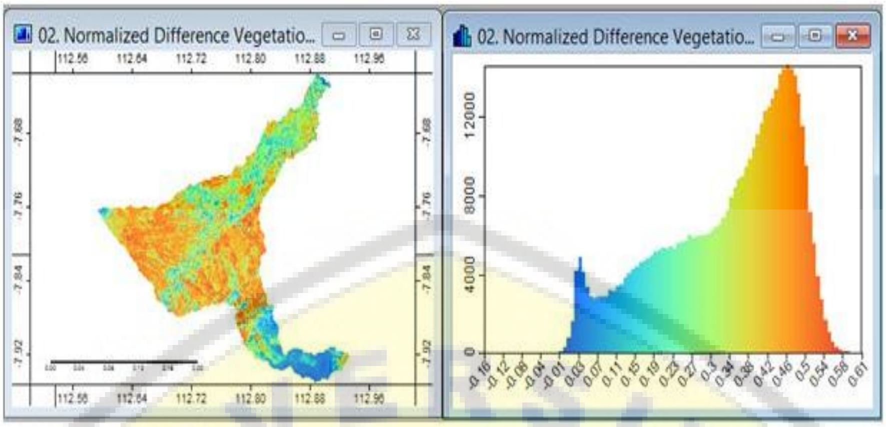

Figure 3. NDVI Welang watershed

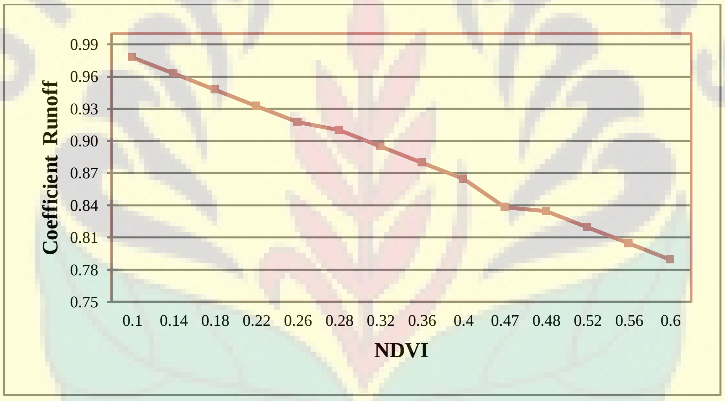

Figure 4. Relation of NDVI value with runoff coefficient in Welang watershed

4.2 Selection of rain and discharge data

The occurrence of rain in the upper watershed sometimes does not correlate with an increase in downstream discharge. Therefore it is necessary to select rain and discharge data that have a correlation in a time lag. Time lag can be known by cross-correlation analysis using rainfall series input data and output discharge. The result of cross-correlation with the input data of hourly rainfall and the discharge period of January 16, 2017 is shown in Figure 5. In figure 5, the positive cross-correlation value indicates that the rainfall affects the discharge, while the negative correlation value shows the inverse relationship where the discharge influences the rainfall. A correlation value of 0 shows no relationship between rainfall and discharge. Based on the results of cross-correlation was known that the rainfall has a positive correlation to the discharge that occurred at time lag 2 to 5. Based on the results of this cross-correlation series data of rain and discharge at time lag 2 up to time lag 5 was inserted in the input data and target applications ANN . The first data of rainfall and discharge as data at time lag 0. The second data of rainfall and discharge as data time lag 1, and so on. Based on the description, the third to sixth data of rainfall and discharge were selected as input and target data in ANN.

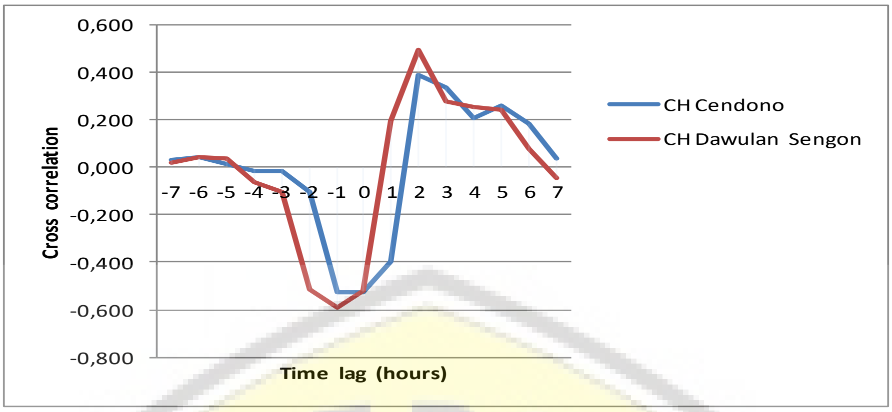

Figure 5. Correlogram cross-correlation of rain and discharge data

4.3 Result of artificial neural network (ANN)
Modeling of the flash flood in Welang watershed using ANN with variable input rain series data and runoff coefficient (C) showed pattern corresponding to observation with a very strong correlation coefficient (R) and lower mean squared error (MSE). The training process aims to optimize the correlation performance of a variable such as Mean Squared Error and correlation coefficient (R) as shown in Figure 6a. Meanwhile, The flash flood modeling with variable input rain series data shows a pattern corresponding to observation with a very strong correlation coefficient (R) and lower mean squared error (MSE). The value of correlation (R) and MSE at the training stage (R = 0,92657 and MSE= 0,00826), validation (R = 0,91477 and MSE= 0,06712), and testing (R = 0,90498 and MSE= 0,043162) as Figure 6b. Figure 7a showed The flash floods modeling with variable input hourly rain series data and runoff coefficient (C) was greater the observation discharge at a certain period. The flash flood discharge modeling results on January 30, 2017 at 9.00 PM was 48,494 m^3/sec and on November 25, 2017 at 8.00 PM was 48,494 m^3/sec lower than observation discharge. While the result of flash flood modeling with variable input hourly rain series data was shown in Figure 7b. The flash flood discharge from modeling on January 30, 2017 at 8.00 PM was 26,58953 m^3/sec still lower when compared with the observation discharge. While the flash flood discharge modeling results on November 2, 2017 showed greater than observation discharge 52,44968 m^3/sec.

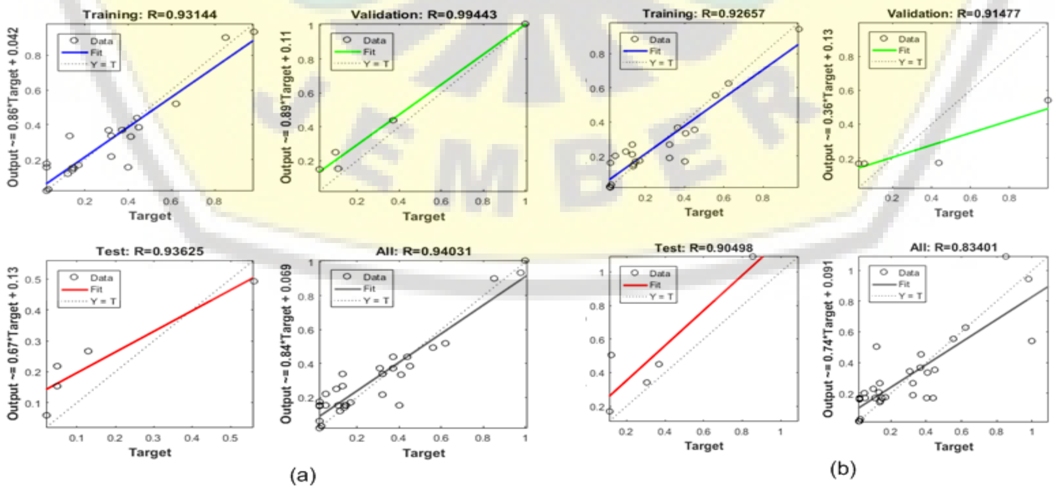

Figure 6. The correlation coefficient (R) of flash flood modeling using ANN

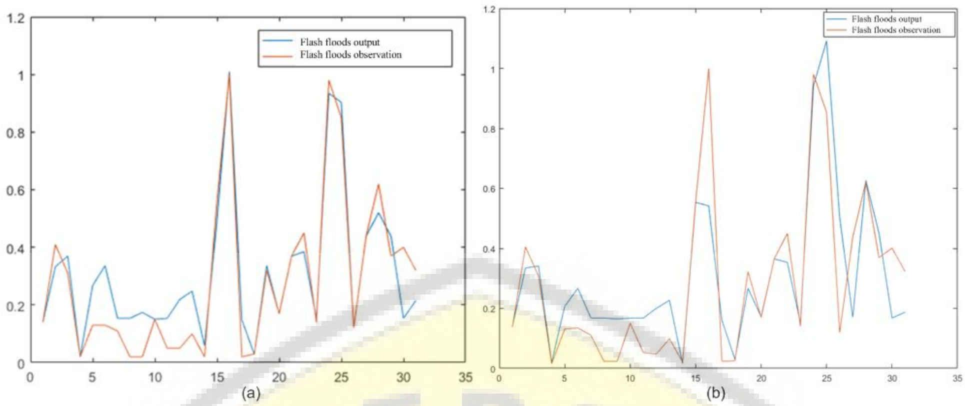

Figure 7. The observation and fash flood discharge result ANN modeling

5. Conclusion

The flash flood modeling using ANN with input series data of rainfall and runoff coefficient (C) from conversion NDVI showed significant results. This was indicated by the value of the correlation coefficient (R) and mean squared error (MSE) of training stage (R = 0,93144 and MSE= 0,00881), validation (R = 0,99443 and MSE= 0,00906), and testing (R= 0,9362 and MSE= 0.01284624). Thus, the model of flash flood discharge with input series data of rainfall and runoff coefficients (C) using ANN was reliable.

Acknowledgment

This research was funded by Post Grant Research program of Directorate General of Higher Education (Ditjen DIKTI). The author also expressed his gratitude to the leaders of the Master Program of Civil Engineering and the Dean of the Faculty of Engineering, the University of Jember which has provided support in writing this article.

References

[1] Elkhrachy, I 2015 Flash flood hazard mapping using satellite images and gis tools: a case study of najran city, kingdom of saudi arabia (ksa). *Egyptian Journal of Remote Sensing and Space Science*. **18**(2) 261–278.

[2] Hapuarachchi, H. A. P. dan Wang 2008 A review of methods and systems available for flashflood forecasting. **6**(1) 137 -142.

[3] Halik, G., S. Wahyuni, dan A. Maududie 2010 Dampak Perubahan Tata Guna Lahan Terhadap Respon Hidrograf Banjir Di Daerah Aliran Sungai Sampean Baru. **4** 259–265.

[4] Latuamury, B., T. Gunawan, dan S. Suprayogi 2012 Perubahan tutupan lahan akan berdampak pada berubahnya sifat-sifat hidrologi seperti koefisien aliran , debit dan karakteristik hidrograf aliran . indikator kerusakan hutan dapat dilihat dari karakteristik hidrograf . evaluasi respon das berupa hidrograf. *Majalah Geografi Indonesia Fakultas Geografi UGM*. **26**(2):98–118.

[5] Ohana-Levi, N., A. Karnieli, R. Egozi, A. Givati, dan A. Peeters 2015 *Modeling the Effects of Land-Cover Change on Rainfall-Runoff Relationships in a Semiarid, Eastern Mediterranean Watershed*. Advances in Meteorology

[6] Solomatine, D. P. 2003 Model trees as an alternative to neural networks in rainfall–runoff modelling. **48**(Hydrological Sciences–Journal–des Sciences Hydrologiques,) :299–411.

[7] Solomatine, D. P. dan A. Ostfeld 2008 Data-driven modelling: some past experiences and new approaches. *Journal of Hydroinformatics*. **10**(1):3.

[8] Singh, V. P. dan D. A. Woolhiser 2002 Mathematical modeling of watershed hydrology. *Journal of Hydrologic Engineering*. **7**(4):270–292.

[9] Jajarmizadeh, M., S. Harun, dan M. Salarpour. (2012). A review on theoretical consideration

IOP Conf. Series: Earth and Environmental Science **419** (2020) 012123 doi:10.1088/1755-1315/419/1/012123

and types of models in hydrology. *Journal of Environmental Sciences and Technology*. **5**(May):249–261.

[10] Ji, J., C. Choi, M. Yu, dan J. Yi. (2012). Comparison of a data-driven model and a physical model for flood forecasting. *WIT Transactions on Ecology and the Environment*. 159(WIT Transactions on Ecology and The Environment,) **7**:133–142.

[11] Jajarmizadeh, M., S. Harun, dan M. Salarpour 2012 A review on theoretical consideration and types of models in hydrology. *Journal of Environmental Sciences and Technology*. **5**(May):249–261.

[12] Artigue, G., A. Johannet, V. Borrell, dan S. Pistre 2012 Flash flood forecasting in poorly gauged basins using neural networks: case study of the gardon de mialet basin (southern france). *Natural Hazards and Earth System Science*. **12**(11):3307–3324.

[13] Siou, L. K. A., A. Johannet, S. Pistre, dan V. Borrell 2010 Advances in research in karst media. **32** 215–216.

[14] Ghumman, A. R., Y. M. Ghazaw, A. R. Sohail, dan K. Watanabe. 2011 Runoff forecasting by artificial neural network and conventional model. *Alexandria Engineering Journal*. **50**(4):345–350.

[15] Areerachakul, S. dan P. Junsawang 2014 Rainfall-Runoff Relationship for Streamflow Discharge Forecasting by ANN Modelling. *World Congress on Sustainable Technologies, WCST 2014*. 2014. 27–30.

[16] Mohammed, K. dan M. J. Uddin Khan 2017 Flash Flood Forecasting in the Northeast Region of Bangladesh Using Artificial Neural Network. *6th International Conference on Water & Flood Management*. (March). 2017

[17] Wibowo, H., M. P. Hadi, dan Suharyadi 2010 Transformasi n d v i untuk estimasi nilai koefisien aliran kasus di das citarum hulu. *LIMNOTEK*. 17:138–146.

[18] Thanapura, P., D. L. Helder, dan S. Burckhard 2007 Mapping urban land cover using quickbird ndvi and gis spatial modeling for runoff coefficient determination. *Photogrammetric Engineering & Remote Sensing*. **73**(1):57–65.

[19] Wibowo, H. dan D. Daruati 2008 Estimasi Nilai Koefisien Aliran DAS Citarum Hulu Menggunakan Transformasi NDVI Citra Landsat. *Prosiding Seminar Nasional Limnologi IV*. 2008. 244–259.

[20] Green, E. P., J. M. Peter, A. J. Edwards, dan C. D. Clrak 2000 Mapping Mangroves. United Nations Educational, Scientific and Cultural Organization. Remote Sensing Handbook for Tropical

[21] Purwanto, A 2015 Pemanfaatan citra landsat 8 untuk identifikasi normalized difference vegetation index (ndvi) di kecamatan silat hilir kabupaten kapuas hulu. Edukasi. **13**(1):27–36.

[22] University of Iceland 2014 Landsat B. 2014.

[23] Alba, E., E. P. Mello, J. Marchesan, dan E. A. Silva 2017 Spectral characterization of forest plantations with landsat 8 / oli images for forest planning and management. Pesq. Agropec. Bras., Brasília,. **52**(1):1072–1079.

[24] Dauwani, K. N. 2012 Analisis Nilai Koefisien Runoff Untuk Pengendalian Direct Runoff (Studi Kasus DAS Citarum Hulu). Bandung.

[25] Nugroho Adji, T., M. Prawira Yudha, dan B. Pandu Dewantara 2016 Distribusi Spasial Respon Debit Mataair Dan Sungai Bawah Tanah Terhadap Hujan Distribusi Spasial Respon Debit Mataair Dan Sungai Bawah Tanah Terhadap Hujan Untuk Prediksi Kapasitas Penyimpanan Air Oleh Akuifer Karst Di Sebagian Wilayah Karst Di Pulau Jawa. Yogyakarta: Fakultas Goegrafi Universitas Gadjah Mada

[26] Sapna, S., A. Tamilarasi, dan M. K. Pravin 2012 Backpropagation learning algorithm based on levenberg marquardt algorithm. *Computer Science & Information Technology (CS & IT)*. 393–398.