Regional Flood Frequency Analysis Using An Artificial Neural Network Model

Sasan Kordrostami 1, Mohammad A Alim 1, Fazlul Karim 2,* and Ataur Rahman 1

1 School of Engineering, Western Sydney University, Sydney, NSW 2751, Australia; sasan.kordrostami@gmail.com (S.K.); M.Alim@westernsydney.edu.au (M.A.A.); A.Rahman@westernsydney.edu.au (A.R.)
2 CSIRO Land and Water, Commonwealth Scientific and Industrial Research Organization, Canberra, ACT 2601, Australia
* Correspondence: fazlul.karim@csiro.au; Tel.: +61-02-6246-4526
Received: 16 December 2019; Accepted: 31 March 2020; Published: 1 April 2020

Abstract: This paper presents the results from a study on the application of an artificial neural network (ANN) model for regional flood frequency analysis (RFFA). The study was conducted using stream flow data from 88 gauging stations across New South Wales (NSW) in Australia. Five different models consisting of three to eight predictor variables (i.e., annual rainfall, drainage area, fraction forested area, potential evapotranspiration, rainfall intensity, river slope, shape factor and stream density) were tested. The results show that an ANN model with a higher number of predictor variables does not always improve the performance of RFFA models. For example, the model with three predictor variables performs considerably better than the models using a higher number of predictor variables, except for the one which contains all the eight predictor variables. The model with three predictor variables exhibits smaller median relative error values for 2- and 20-year return periods compared to the model containing eight predictor variables. However, for 5-, 10-, 50- and 100-year return periods, the model with eight predictor variables shows smaller median relative error values. The proposed ANN modelling framework can be adapted to other regions in Australia and abroad.

Keywords: floods; artificial neural network; regional flood frequency; ungauged catchments

1. Introduction

Globally, floods are the most damaging natural disasters that cause enormous economic loss and social disruptions across the landscape. In the last decade, floods accounted for roughly 45% of all disasters (and people affected by them) and caused an average of 6000 casualties in each year [1]. Floods cause billions of dollars of damage annually worldwide, and even in the world’s driest inhabited continent, Australia, flooding is the costliest natural disaster [1–3]. In recent years, floods have become more frequent and highly disastrous due to global climate change [1].

One of the key steps in flood risk assessment process is the estimation of design floods [2]. A design flood is the peak discharge used to design hydraulic structures (e.g., bridge, culvert, retaining wall) and the magnitude of the flood is represented by the annual exceedance probability (AEP) [4]. At-site flood frequency analysis is the commonly accepted method to estimate design flood in a gauged catchment if observed streamflow data are available for a number of years. However, there are numerous catchments where observed streamflow data are not available. Regional flood frequency analysis (RFFA) is considered as the best option to estimate design flood for these catchments [5–7].

The efficiency of a RFFA technique primarily depends on two factors: (i) sufficiency and accuracy of historical streamflow data in terms of record length and spatial coverage over the study
area; and (ii) the adopted regionalization method that explores flood characteristics in gauged catchments and transfers the relevant flood attributes to ungauged catchments. Some of the most successful RFFA techniques include (i) an index flood estimation method [8], (ii) a quantile regression method [9] and (iii) a parameter regression method [10,11]. All of these techniques are essentially linear, such that floods are linearly related with catchment characteristics either in a log domain or domain with raw data [12,13].

While efforts have been made to develop nonlinear RFFA methods to estimate design floods, the application of such methods are quite limited. In the past few years, some non-linear methods, such as an artificial neural network (ANN), gene expression programming and fuzzy models, were developed and their efficiency was evaluated [12,13]. Aziz et al. [13] investigated the performances of ANN and GEP methods using streamflow records from 452 gauging stations across Australia. They compared the results with quantile regression technique, which is a linear RFFA method, and concluded that a non-linear method produced better results compared to a linear method.

The working principle of ANN is similar to that of the human neural system [14]. Unlike other data processing methods which learn through programming, an ANN model investigates the patterns in a data set and correlate them. The ANN consists of simple computing units called ‘artificial neurons’. Each unit is connected to the other units via weight connectors. These units calculate the sum of all weighted inputs and bias. It then produces output of previously weighted input and bias using an activation function.

In the last few decades, the ANN model (introduced by McCulloch and Pitts [15]) has been extensively used to solve various mathematical problems, especially in the field of medical science [14,16,17]. In recent years, the method has been used in engineering fields for forecasting and data compression. Lapedes and Farber [18] applied the ANN model to investigate non-linear data series and found better generalization capabilities of ANN models compared to a regression-based model. ANN model is more capable of identifying non-linear connections between observed and predicted data sets [19,20].

In the Australian context, studies on ANN-based RFFA modelling are limited. The majority of previous RFFA studies have been based on linear models [5,10]. Aziz et al. [21,22] applied ANN-based RFFA modelling; however, they applied a limited set of catchment characteristics in model building. The objectives of this study are: (i) to develop and test ANN-based RFFA models using a higher number of catchment attributes; and (ii) to recommend the best ANN-based RFFA model containing an optimum number of predictor variables for the catchments in New South Wales.

2. Principle of ANN

The concept of an ANN model is schematically represented in Figure 1. The first column represents various input variables (\(X_i\)) and the second column represents the specific weight of input variables (\(W_i\)). The output is determined by considering a given vector. There is a constant value (1) among the inputs which has been introduced to the neuron by its unique weight, known as bias (b). Bias allows the ANN to change the activation function. It is important to note that bias is not essential for a network but it helps in improving the performance of a network significantly [23].

![Figure 1. Conceptual representation of an artificial neural network (ANN) model.](image-url)
Mathematically, the input (I) and output (Y) of a neuron in an ANN model can be represented as [23]:

\[ I = X \times W + b \]  

\[ Y = \begin{cases} 1 \text{ if } I \geq \theta \\ 0 \text{ if } I \leq \theta \end{cases} \]  

where X is the input, W is the weight, b is the bias and \( \theta \) is the sum of all weighted inputs.

The output of a neuron can be presented by different activation functions. A sigmoid function (Equation (2)) is a commonly used activation function for such outputs [24]. This sigmoid function maps the data between 1 and 0 (Figure 2). A sigmoid function is adopted in this study because it is a bounded differentiable function. It is defined for all real inputs and at each point it produces a non-negative derivative [21,22]. Other activation functions could have been adopted; however, the use of the sigmoid function is deemed adequate in this study. The dependent and predictor variables in this study are normalized to achieve zero mean and unit variance for each of the variables.

\[ S(t) = \frac{1}{1 + e^{-t}} \]  

Figure 2. Representation of a sigmoid function.

Figure 3 shows a configuration of an ANN model with interlinking between the input and output data layers. It consists of three input neurons, two hidden layers of three and two perceptron and an output layer with two output neurons.

In recent years, the application of ANN model has increased in rainfall-runoff modelling, groundwater modelling, streamflow forecasting and water quality modelling (e.g., [25–30]). Table 1 presents a list of relevant literature which used ANN successfully in solving hydrological and
related problems. It can be seen that the majority of studies used the feed forward (FF) algorithm in the ANN architecture, back propagation (BP) for optimization, and log normal as a transfer function (Table 1). In FF neural networks, the information only travels forward in the neural network through the input nodes, then through the hidden layers and finally through the output nodes [23]. Optimization is used for the fine-tuning of weights factors in ANN modelling based on the errors in the previous iteration [23].

| Author(s) and Year | Variable | Time Step | Architecture | Optimization | T-Function |
|--------------------|----------|-----------|--------------|--------------|------------|
| French et al., 1992 [31] | Rainfall | Hour | FF | BP | Log |
| Crespo and Mora, 1993 [32] | Flow | Day | FF | BP | HT |
| Allen and le Marshall, 1994 [33] | Rainfall | Day | FF | BP | Log |
| Karunanithi et al., 1994 [34] | Flow | Day | FF | QP | Log |
| Hsu et al., 1995 [19] | Flow | Day | FF | BP | Log |
| Raman and Sunilkumar, 1995 [35] | Flow | Month | FF | BP | Log |
| Smith and Eli, 1995 [7] | Flow | N/A | FF | BP | Log |
| Clair and Ehrman, 1996 [36] | Flow | Year | FF | BP | N/A |
| Minns and Hall, 1996 [37] | Flow | Hour | FF | BP | Log |
| Poff et al., 1996 [38] | Flow | Day | FF | BP | HT |
| Chow and Cho, 1997 [39] | Rainfall | Hour | RC | Mod BP | HT |
| Hsu et al., 1997 [40] | Rainfall | Hour | HYB | MCP | N/A |
| Loke et al., 1997 [41] | Rainfall | Min | FF | Mod BP | N/A |
| Shamseldin, A. Y., 1997 [42] | Rainfall-R unoff | Day | FF | BP | Log |
| Tawfik et al., 1997 [43] | Flow | Day | FF | BP | Lin |
| Venkatesan et al., 1997 [44] | Rainfall | Year | FF | BP | Log |
| Xiao and Chandrasekar, 1997 [45] | Rainfall | Min | FF | RSL | TL |
| Dawson and Wilby, 1998 [46] | Rainfall-R unoff | 15 min | FF | BP | Log |
| Fernando and Jayawardena, 1998 [47] | Flow | Hour | FF | BP | N/A |
| Golob et al., 1998 [48] | Flow | Hour | FF | BP | Log |
| Jayawardena and Fernando, 1998 [49] | Flow | Hour | FF | BP | N/A |
| Phien And Sureerattanan, 1999 [50] | Flow | Day | FF | BP | Log |
| Tokar and Johnson, 1999 [51] | Rainfall-R unoff | Day | FF | BP | N/A |
| Zealand et al., 1999 [52] | Flow | Week | FF | BP | Log |
| Luk et al., 2000 [53] | Rainfall | Day | FF | BP | Log |
| Coulibaly et al., 2000 [54] | Inflow | Day | FF | BP | Log |
| Luk et al., 2001 [55] | Rainfall | Day | FF-PRI-TD² | BP | Log |
| Zhang and Govindaraju, 2003 [56] | Runoff | Hour | FF | BP | Log |
| Rajurkar et al., 2004 [57] | Rainfall-R unoff | Day | FF | BP | Log |
| Dawson et al., 2006 [58] | RFFA | Year | FF | BP | Log |
| Jain and Kumar, 2006 [59] | Infiltration | N/A | FF | BP | N/A |
| Riad et al., 2004 [60] | Rainfall-R unoff | Day | FF | BP | Log |
3. Materials and Methods

Data from 88 stream gauging stations across New South Wales (NSW) were used to develop the ANN model (Figure 4). Selected gauges were situated on natural streams and free from any major regulation. Drainage area between the selected gauging stations varied from 8 to 1010 km². The first, second (median) and third quartile drainage area were 142.5, 260 and 537.25 km², respectively. The streamflow data of the selected stream gauging stations were downloaded from the WaterNSW website [85]. The periods of annual maximum (AM) flow records at these stream gauging stations varied between 25 to 82 years. This study was conducted based on eight predictor variables that included (i) mean annual rainfall (MAR); (ii) areal potential evapo-transpiration (MAE), (iii) drainage area (AREA) (iv) 6-hour duration rainfall for a 2-year return period (I62); (v) shape factor (SF); (vi) stream density (SDEN); (vii) river slope (S1085) and (viii) proportion of forest (FOREST).

A 1:100,000 topographic map was used to delineate catchment boundary and drainage area for individual gauge. Rainfall intensity data (I62) were obtained from the online portal of Australian Bureau of Meteorology (BOM) using intensity-frequency-duration (IFD) estimation tool [86]. The MAR and MAE data at the catchment centroid were extracted from Australian Bureau of Meteorology website [86]. The shape factor (SF) for a catchment was considered as the shortest distance between catchment outlet and centroid divided by the square root of the drainage area. The
SDEN was calculated as a ratio of the stream length (total length for all streams in a catchment) and drainage area. Stream length was measured on a 1:100,000 topographic map using a digital distance meter. The forested area for a catchment was measured on 1:100,000 topographic map using a planimeter. The minimum, maximum and median parameter values are presented in Table 2.

Table 2. Summary of catchment data used in this study.

| Input Factors                                      | Units | Min   | Max   | Median |
|----------------------------------------------------|-------|-------|-------|--------|
| Drainage area (AREA)                               | km²   | 8.00  | 1010.00 | 260.00 |
| Rainfall intensity (I62)                           | mm/h  | 31.30 | 87.30 | 43.10  |
| Mean annual rainfall (MAR)                         | mm    | 626.17 | 1953.23 | 909.92 |
| Shape factor (SF)                                  | -     | 0.25  | 1.62  | 0.76   |
| Mean annual areal potential evapo-transpiration (MAE) | mm    | 980.40 | 1543.30 | 1185.55 |
| Stream density (SDEN)                              | km/km²| 0.51  | 5.47  | 2.71   |
| River slope (S1085)                                | m/km  | 1.53  | 49.85 | 9.07   |
| Proportion of forest (FOREST)                      | -     | 0.010 | 0.99  | 0.51   |

Five different ANN-based RFFA models were selected to assess their performances (Table 3). Model 1 consisted of all the eight predictor variables. Other models consisted of more than one but less than eight variables. It is important to note that models were selected based on the best potential combination, but there could be other combinations as well. As recommendations in previous RFFA studies, rainfall intensity was included in all five models and drainage area was selected as the predictor variable [70,87].

Six flood events of discharge magnitudes 2 (Q2), 5 (Q5), 10 (Q10), 20 (Q20), 50 (Q50) and 100 (Q100) years return period were used as dependent variables. Design flood discharges were estimated by fitting the Log Pearson type III (LP3) probability distribution function to the observed annual maximum flood discharge [88,89]. The analyses were conducted using the FLIKE software, which is commonly used for flood frequency modelling in Australia [90]. The main advantage of FLIKE is that it fits five probability distribution functions for a given set of flood data and identifies the best frequency model for that data set [91]. In this study, LP3 was selected because it was found as the best fit distribution for Australian stream gauge data [10].
Table 3. Five different ANN based regional flood frequency analysis (RFFA) models with adopted catchment characteristics represented by green colour (predictor variables).

| Predictor Variable | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|--------------------|---------|---------|---------|---------|---------|
| Area               |         |         |         |         |         |
| I62                | ✓       |         |         |         |         |
| MAR                |         | ✓       |         |         |         |
| SF                 |         |         | ✓       |         |         |
| MAE                |         |         |         | ✓       |         |
| SDEN               |         |         |         |         | ✓       |
| S1085              |         |         |         |         |         |
| Forest             | ✓       |         |         |         |         |

A 2-layer feed forward neural network and with a backpropagation algorithm was selected in this study.

The selected 88 stream gauging stations were sub-divided into three sets. Group 1 consisted of 62 stations (70%) and they were used for model training, Group 2 consisted of 13 stations (15%) and they were used for validation. The third group (13 stations), were used to evaluate model performances in predictions. It is important to note that the gauges in each group were selected randomly. The analyses were conducted using a two-layer FF neural network with two hidden layers in each model setup. The Levenberg–Marquardt algorithm was used for model training. The activation function was represented in the model using a hyperbolic sigmoid function. The entire analyses were carried out using MATLAB software. A set of statistical tests were performed to evaluate the model performance. These include, root mean squared error (RMSE), root mean squared normalised error (RMSNE), relative root mean squared error (RRMSE), coefficient of determination ($R^2$), mean bias (BIAS), relative mean bias in percent (rBIAS), absolute relative error (abs-RE) and quantile ratio ($r$, ratio of predicted and observed discharge) as presented in Equations (3–10). These evaluation statistics were adopted from Bloschl et al. [92].

\[
\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{Q}_i - Q_i)^2} 
\]

\[
\text{RMSNE} = \sqrt{\frac{1}{n} \sum \left( \frac{\hat{Q}_i - Q_i}{Q_i} \right)^2} 
\]

\[
\text{RRMSE} = \sqrt{\frac{1}{n} \sum \left( \frac{\hat{Q}_i - Q_i}{\hat{Q}_i} \right)^2} 
\]

\[
R^2 = 1 - \frac{\sum (\hat{Q}_i - Q_i)^2}{\sum (\hat{Q}_i - \hat{Q}_i)^2} 
\]

\[
\text{BIAS} = \frac{1}{n} \sum (Q_i - \hat{Q}_i) 
\]

\[
\text{rBIAS} = \frac{1}{n} \sum \left( \frac{Q_i - \hat{Q}_i}{Q_i} \right) \times 100 
\]

\[
\text{abs-RE} = \left| \frac{\hat{Q}_i - Q_i}{Q_i} \right| \times 100 
\]
\[ r = \frac{\hat{Q}_i}{Q_i} \]  

where \( \hat{Q}_i \) and \( Q_i \) are the predicted and observed flood discharge, respectively, \( \hat{Q}_i \) is the mean observed discharge and \( n \) is the number of stations.

### 4. Results and Discussion

Models performances in predicting flood quantiles are presented in Table 4 in terms of abs-RE and \( R^2 \). The results show that all the models perform well for the majority of quantiles, except for \( Q_2 \) by Model 1 (61%), \( Q_5 \) by Model 4 (80%) and \( Q_2 \) (320%) and \( Q_5 \) (85%) by Model 5. Model 2 (two predictor variables) performed reasonably well with the lowest and highest RE of 32.9% and 47.7%, respectively. For all five quantiles, Model 3 performed very well (RE values in the range of 29.48%–52.24%). Model 4 performed well for most quantiles, except \( Q_5 \). Model 5 predictions are relatively poor for all quantiles, with the worst prediction for \( Q_2 \) (RE of 319.61%). Overall, Model 5 produced the poorest results among the five models.

Coefficient of determination (\( R^2 \)) values for \( Q_5 \), \( Q_{10} \), \( Q_{20} \) and \( Q_{50} \) show a moderate model accuracy for the Model 1 (\( R^2 \) values ranging from 0.71 to 0.76). Model 2 performed well except for the \( Q_{100} \) (\( R^2 \) value of 0.16). Model 3 performed well for \( Q_5 \), \( Q_{10} \) and \( Q_{20} \). Overall, Model 1 and 3 produced better results compared to other models. Model 4 performed the best for \( Q_2 \) (\( R^2 = 0.74 \)) and worst for \( Q_{20} \). As expected, \( R^2 \) values for smaller ARIs (average recurrence interval) are higher than those of the large ARIs. Model 5 performed the worst for \( R^2 \) value which is consistent with RE value.

**Table 4.** Model performance in terms of absolute relative error (abs-RE) and the coefficient of determination (\( R^2 \)) in predicting flood quantile.

| T      | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|--------|---------|---------|---------|---------|---------|
|        | Median absRE(%) | \( R^2 \) | Median absRE(%) | \( R^2 \) | Median absRE(%) | \( R^2 \) | Median absRE(%) | \( R^2 \) | Median absRE(%) | \( R^2 \) |
| \( Q_2 \) | 61.31   | 0.63    | 36.13   | 0.71    | 29.48   | 0.77    | 38.28   | 0.74    | 319.61   | 0.42    |
| \( Q_5 \) | 23.39   | 0.73    | 32.93   | 0.74    | 52.24   | 0.48    | 79.96   | 0.51    | 84.89    | 0.47    |
| \( Q_{10} \) | 10.25   | 0.76    | 34.19   | 0.69    | 33.23   | 0.73    | 29.84   | 0.62    | 44.94    | 0.49    |
| \( Q_{20} \) | 34.06   | 0.71    | 39.50   | 0.57    | 33.43   | 0.69    | 46.99   | 0.39    | 38.44    | 0.21    |
| \( Q_{50} \) | 33.99   | 0.74    | 38.09   | 0.59    | 38.90   | 0.37    | 30.69   | 0.55    | 46.69    | 0.52    |
| \( Q_{100} \) | 33.09   | 0.57    | 47.72   | 0.16    | 38.32   | 0.42    | 35.42   | 0.51    | 67.21    | 0.05    |

Model performances in terms of RMSNE and RRMSE are presented Table 5 for all five models. In regard to RMSNE, models performed well for most quantities, except \( Q_2 \) in the case of Model 1, \( Q_{20} \) for Model 4 and \( Q_2 \), \( Q_5 \) and \( Q_{100} \) for Model 5. The RMSNE value for Model 1 is highest for \( Q_5 \), which indicates low accuracy of the prediction. Model 2 predictions are relatively better with RMSNE values of between 0.89 (\( Q_2 \)) and 2.47 (\( Q_{100} \)). In addition, the best value of RRMSE is found for \( Q_5 \) (0.48) and the poorest value is found for \( Q_{100} \) (0.85). Model 3 performed well for \( Q_{10} \) with the value of 0.75; however, for \( Q_{100} \), the RMSNE is poor with the value of 3.5. The RRMSE values for \( Q_2 \), \( Q_{10} \) and \( Q_{20} \) are approximately the same (0.48). Furthermore, the values of RRMSE are the same (0.69) for \( Q_{30} \) and \( Q_{50} \). The highest error is associated with \( Q_5 \) with a value of 0.82. Model 4 performed well for \( Q_2 \) and \( Q_{10} \) with the values of 0.53 and 0.56, respectively, while the best performance was for \( Q_{30} \). Like other indicators, Model 5 performed the worst for RMSNE and RRMSE.
Table 5. Model performance in terms of root mean squared normalised error (RMSNE) and relative root mean squared error (RRMSE) in predicting flood quantiles.

|   | Model 1 |   | Model 2 |   | Model 3 |   | Model 4 |   | Model 5 |
|---|---------|---|---------|---|---------|---|---------|---|---------|
|   | RMSNE   |   | RMSNE   |   | RMSNE   |   | RMSNE   |   | RMSNE   |
| Q2| 5.28    |   | 0.76    |   | 0.89    |   | 0.54    |   | 1.18    |
| Q5| 1.51    |   | 0.50    |   | 1.23    |   | 0.48    |   | 0.98    |
| Q10| 1.17   |   | 0.46    |   | 1.79    |   | 0.51    |   | 0.75    |
| Q20| 2.25   |   | 0.52    |   | 1.30    |   | 0.58    |   | 1.30    |
| Q50| 1.65   |   | 0.45    |   | 1.87    |   | 0.55    |   | 2.25    |
| Q100| 2.05  |   | 0.61   |   | 2.47    |   | 0.85    |   | 3.50    |

Table 6 presents the comparison between five models in terms of model predictions for different floods. There is no clear trend of underprediction or overprediction, nor any trend for small or large floods. All models overpredicted for some floods and underpredicted for some other floods. Overall, Model 3 produced the best result (relative bias of 33%) and Model 4 produced the worst result (relative bias of 1258%).

Table 6. Model performance in terms of mean bias (BIAS) and relative mean bias (rBIAS) in predicting flood quantile.

|   | Model 1 |   | Model 2 |   | Model 3 |   | Model 4 |   | Model 5 |
|---|---------|---|---------|---|---------|---|---------|---|---------|
|   | BIAS    | rBIAS | BIAS    | rBIAS | BIAS    | rBIAS | BIAS    | rBIAS | BIAS    |
| Q2| -20.08  | -125.13| -22.09  | -9.91  | -40.08  | -97.66| 329.24  | -174.27| 53.26   |
| Q5| -5.89   | -2.67  | 10.63   | -40.19| 110.80  | 12.49 | -12.04  | 590.09 | -8.30   |
| Q10| -102.98| -26.97| -16.63  | -30.76| 21.09   | -24.73| -122.62| 2777.88| 109.56 |
| Q20| 13.66  | 18.83  | 3.44    | -59.29| 26.56   | -7.02 | -59.87  | 1057.57| 48.93   |
| Q50| 5.10   | 29.04  | -2.96   | -56.57| 44.02   | -68.93| -66.57  | 429.41 | 213.44 |
| Q100| 3.32  | 29.22  | 311.68  | -54.28| 187.66  | 46.73 | -62.77  | 2366.70| 164.90 |

Figure 5 shows a graphical comparison of model performance for predicting flood discharge for the flood events of different magnitudes. In general, Model 2 and 4 overpredicted, while other models underpredicted flood magnitude. Overall, Model 2 predictions are close to 1, which indicates a better performance. Model 5 performed the worst with predicted discharge for Q2, which is more than double the observed value.

Figure 6 shows a comparison between five models in predicting flood quantiles. The results show similar performances by Model 1 and Model 3, with little differences in the RMSE values for all six flood quantiles. Model 5 performs relatively poorly for Q2, Q50, Q20 and Q100 and Model 4
performs poor for the Q5. Overall, Model 3’s performances are better and it is considered as the best model for predicting design flood for the catchments in NSW.

The parameter setup for the Model 3 is similar to the current Australian Rainfall and Runoff (ARR) model for NSW which consists of AREA, I62 and SF as predictor variables. While Model 1 (which includes all eight predictor variables) performs well with respect to RE for all quantiles except Q2, Model 3’s performances are the best. It is important to note that none of the models are the best for all six quantiles. For example, with respect to RE, Model 3 is the best for predicting Q2 and Q20 but Model 1 is better for predicting Q5, Q10, Q50 and Q100. Model 4 performs poorly with respect to BIAS. Model 1 is the best based on $R^2$ except for Q2 and Q100.

Table 7 compares the performance of our two best models (Model 1 and Model 3) with similar RFFA studies based on ANN. It shows that both our models provide smaller median RE than Aziz et al. [22] for higher ARIs (10–100 years). Also, in terms of $R^2$, our models perform much better than those of Dawson et al. [58].

| Quantile | Present Work | Aziz et al. [22] | Present Work | Dawson et al. [58] |
|----------|--------------|------------------|--------------|--------------------|
|          | RE (%)       | $R^2$            | RE (%)       | $R^2$             |
|          | Model 1 | Model 3 | Aziz et al. | Model 1 | Model 3 | Dawson et al. |
| Q2       | 61.31  | 29.48  | 37.56        | *N/A    | N/A    | N/A            |
| Q5       | 23.39  | 52.24  | 40.39        | N/A     | N/A    | N/A            |
| Q10      | 10.25  | 33.23  | 44.63        | 76      | 73     | 66             |
| Q20      | 34.06  | 33.43  | 35.62        | 71      | 69     | 65             |
| Q50      | 33.99  | 38.90  | 39.09        | N/A     | N/A    | N/A            |
| Q100     | 33.09  | 38.32  | 44.53        | N/A     | N/A    | N/A            |

*N/A represents the unavailability of the value in the literature.

The best model (Model 3) contains three predictor variables: AREA (the main scaling factor in flood generation process), I62 (the main input that triggers runoff) and SF (the factor that affects travel time of generated runoff). It has been found that the other five predictor variables (MAR, MAE, SDEN, S1085 and FOREST) have minor roles in RFFA modelling in the study area.

5. Conclusions

In this study, ANN-based regional flood frequency models are developed to estimate the design floods. The models were tested using observed discharge data from 88 catchments in NSW. Five models were tested considering two to eight predictor variables. The best combinations of
predictor variables were identified based on model performance in predicting design floods of 2, 5, 10, 20, 50 and 100 years return periods. Model performances were evaluated based on nine statistical error metrics. The study found that model predictions are better when all eight predictor variables are included. However, the results based on three predictor variables were found to be close to the results for eight predictor variables. The study concludes that a high number of predictor variables does not always improve the model predictions. The use of MAR, MAE, SDEN, S1085 and FOREST as predictor variables provide little contribution in predicting design flood for an ungauged catchment. The key predictor variables are catchment area, rainfall intensity and slope factor. The findings are very useful because these predictor variables are readily available for the majority of the catchments. The results demonstrate the potential use of the ANN-based regional flood frequency model. However, further testing with larger data sets is necessary before it can be applied elsewhere.

Author Contributions: Conceptualization, S.K. and A.R.; methodology, S.K. and A.R.; software, S.K., F.K and A.R.; validation, S.K., F.K and A.R.; formal analysis, S.K., and M.A.; investigation, S.K., M.A, F.K and A.R.; resources, S.K. and A.R.; writing—original draft preparation, S.K. and A.R.; writing—review and editing, M.A and F.K.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare that there is no conflict of interest.

References

1. FitzGerald, G.; Du, W.; Jamal, A.; Clark, M.; Hou, X.Y. Flood fatalities in contemporary Australia (1997–2008). Emerg. Med. Australas. 2010, 22, 180–186.
2. Loveridge, M.; Rahman, A. Monte Carlo simulation for design flood estimation: A review of Australian practice. Australas. J. Water Resour. 2018, 22, 52–70.
3. Wang, Z.; Chengguang, L.; Chen, X.; Yang, B.; Zhao, S.; Bai, X. Flood hazard risk assessment model based on random forest. J. Hydrol. 2015, 527, 1130–1141.
4. Myronidis, D.; Ioannou, K. Forecasting the Urban Expansion Effects on the Design Storm Hydrograph and Sediment Yield using Artificial Neural Networks. Water 2018, 11, 31.
5. Ouarda, T.B.; Girard, C.; Cavadias, G.S.; Bobée, B. Regional flood frequency estimation with canonical correlation analysis. J. Hydrol. 2001, 254, 157–173.
6. Ribatet, M.; Sauquet, E.; Gressillon, J.-M.; Ouarda, T.B.M.J. A regional Bayesian POT model for flood frequency analysis. Stoch. Environ. Res. Risk Assess. 2006, 21, 327–339.
7. Smith, J.; Eli, R.N. Neural-Network Models of Rainfall-Runoff Process. J. Water Resour. Plan. Manag. 1995, 121, 499–508.
8. Thorvat, A.R.; Mujumdar, M.M. Design flood estimation for Upper Krishna Basin through RFFA. Int. J. Eng. Sci. Technol. 2011, 3, 5252.
9. Ouali, D.; Chebana, F.; Ouarda, T.B.M.J. Quantile Regression in Regional Frequency Analysis: A Better Exploitation of the Available Information. J. Hydrometeorol. 2016, 17, 1869–1883.
10. Haddad, K.; Rahman, A. Regional flood frequency analysis in eastern Australia: Bayesian GLS regression-based methods within fixed region and ROI framework – Quantile Regression vs. Parameter Regression Technique. J. Hydrol. 2012, 430, 142–161.
11. Ahn, K.-H.; Palmer, R.N. Regional flood frequency analysis using spatial proximity and basin characteristics: Quantile regression vs. parameter regression technique. J. Hydrol. 2016, 540, 515–526.
12. Srinivas, V.V.; Tripathi, S.; Rao, A.; Govindaraju, R. Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering. J. Hydrol. 2008, 348, 148–166.
13. Aziz, K.; Rahman, A.; Shamseldin, A.; Shoaiib, M. Regional flood estimation in Australia: Application of gene expression programming and artificial neural network techniques. In Proceedings of the 20th International Congress on Modelling and Simulation, Adelaide, Australia, 1–6 December 2013.
14. Agatonovic-Kustrin, S.; Beresford, R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. J. Pharm. Biomed. Anal. 2000, 22, 717–727.
15. McCulloch, W.S.; Pitts, W. A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biol. 1943, 5, 115–133.
16. Baxt, W.G. Use of an Artificial Neural Network for Data Analysis in Clinical Decision-Making: The Diagnosis of Acute Coronary Occlusion. Neural Comput. 1990, 2, 480–489.
17. Amato, F.; López, A.; Peña-Méndez, E.M.; Vanhara, P.; Hampl, A.; Havel, J. Artificial neural networks in medical diagnosis. *J. Appl. Biomed.* 2013, 11, 47–58.

18. Lapedes, A.; Farber, R. Nonlinear signal processing using neural networks: Prediction and system modelling. In Proceedings of the IEEE international conference on neural networks, San Diego, CA, USA, 21 Jun 1987; p. 52.

19. Hsu, K.-L.; Gupta, H.; Sorooshian, S. Artificial Neural Network Modeling of the Rainfall-Runoff Process. *Water Resour. Res.* 1995, 31, 2517–2530.

20. El-Shafie, A.; Mukhilsin, M.; Najah, A.A.; Taha, M.R. Performance of artificial neural network and regression techniques for rainfall-runoff prediction. *Int. J. Phys. Sci.* 2011, 6, 1997–2003.

21. Aziz, K.; Rahman, A.; Fang, G.; Shrestha, S. Comparison of artificial neural networks and adaptive neuro-fuzzy inference system for regional flood estimation in Australia. In *Hydrology and Water Resources Symposium 2012*; 2012 Engineers Australia, 19-22 Nov, Sydney, Australia.

22. Aziz, K.; Rahman, A.; Fang, G.; Shrestha, S. Application of artificial neural networks in regional flood frequency analysis: A case study for Australia. *Stoch. Environ. Res. Risk Assess.* 2013, 28, 541–554.

23. Araghinejad, S. *Data-Driven Modeling: Using MATLAB® in Water Resources and Environmental Engineering*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2014; Volume 67.

24. Ito, Y. Representation of functions by superpositions of a step or sigmoid function and their applications to neural network theory. *Neural Networks* 1991, 4, 385–394.

25. Maier, H.R.; Dandy, G. The Use of Artificial Neural Networks for the Prediction of Water Quality Parameters. *Water Resour. Res.* 1996, 32, 1013–1022.

26. Wu, C.; Chau, K.-W. Rainfall–runoff modeling using artificial neural network coupled with singular spectrum analysis. *J. Hydrol.* 2011, 399, 394–409.

27. Taormina, R.; Chau, K.-W.; Sethi, R. Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice lagoon. *Eng. Appl. Artif. Intell.* 2012, 25, 1670–1676.

28. Zemzami, M.; Benaabidate, L. Improvement of artificial neural networks to predict daily streamflow in a semi arid area. *Hydrol. Sci. J.* 2016, 61, 1–12.

29. Ioannou, K.; Myronidis, D.; Lefakis, P.; Stathis, D. The use of artificial neural networks (ANNs) for the forecast of precipitation levels of lake Doirani (N. Greece). *Fresenius Environ. Bull.* 2010, 19, 1921–1927.

30. Myronidis, D.; Ioannou, K.; Fotakis, D.; Dörflinger, G. Streamflow and Hydrological Drought Trend Analysis and Forecasting in Cyprus. *Water Resour. Manag.* 2018, 32, 1759–1776.

31. French, M.N.; Krajewski, W.F.; Cuyckendall, R.R. Rainfall forecasting in space and time using a neural network. *J. Hydrol.* 1992, 137, 1–31.

32. Crespo, J.; Mora, E. Drought estimation with neural networks. *Adv. Eng. Softw.* 1993, 18, 167–170.

33. Allen, G.; le Marshall, J. An evaluation of neural networks and discriminant analysis methods for application in operational rain forecasting. *Aust. Meteorol. Mag.* 1994, 43, 17–28.

34. Karunanithi, N.; Grenney, W.J.; Whitley, D.; Bovee, K. Neural Networks for River Flow Prediction. *J. Comput. Civ. Eng.* 1994, 8, 201–220.

35. Raman, H.; Sunilkumar, N. Multivariate modelling of water resources time series using artificial neural networks. *Hydrol. Sci. J.* 1995, 40, 145–163.

36. Clair, T.A.; Ehrman, J.M. Variations in discharge and dissolved organic carbon and nitrogen export from terrestrial basins with changes in climate: A neural network approach. *Limnol. Oceanogr.* 1996, 41, 921–927.

37. Minns, A.W.; Hall, M.J. Artificial neural networks as rainfall-runoff models. *Hydrol. Sci. J.* 1996, 41, 399–417.

38. Poff, N.L.; Tokar, S.; Johnson, P. Stream hydrological and ecological responses to climate change assessed with an artificial neural network. *Limnol. Oceanogr.* 1996, 41, 857–863.

39. Chow, T.W.S.; Cho, S.Y. Development of a recurrent Sigma-Pi neural network rainfall forecasting system in Hong Kong. *Neural Comput. Appl.* 1997, 5, 66–75.

40. Hsu, K.-L.; Gupta, H.V.; Gao, X.; Sorooshian, S.; Imam, B. Self-organizing linear output map (SOLO): An artificial neural network suitable for hydrologic modeling and analysis. *Water Resour. Res.* 2002, 38, 38–1.

41. Loke, E.; Warnaars, E.; Jacobsen, P.; Nelen, F.; do Ceu Almeida, M. Artificial neural networks as a tool in urban storm drainage. *Water Sci. Technol.* 1997, 36, 101–109.

42. Shamseldin, A.Y. Application of a neural network technique to rainfall-runoff modelling. *J. Hydrol.* 1997, 199, 272–294.
43. Tawfik, M.; Ibrahim, A.; Fahmy, H. Hysteresis Sensitive Neural Network for Modeling Rating Curves. *J. Comput. Civ. Eng.* 1997, 11, 206–211.
44. Venkatesan, C.; Raskar, S.D.; Tambe, S.S.; Kulkarni, B.D.; Keshavamurty, R.N. Prediction of all India summer monsoon rainfall using error-back-propagation neural networks. *Theor. Appl. Clim.* 1997, 62, 225–240.
45. Xiao, R.; Chandrasekar, V. Development of a neural network based algorithm for rainfall estimation from radar observations. *IEEE Trans. Geosci. Remote. Sens.* 1997, 35, 160–171.
46. Dawson, C.; Wilby, R. An artificial neural network approach to rainfall-runoff modelling. *Hydrol. Sci. J.* 1998, 43, 47–66.
47. Fernando, D.A.K.; Jayawardena, A.W. Runoff Forecasting Using RBF Networks with OLS Algorithm. *J. Hydrol. Eng.* 1998, 3, 203–209.
48. Golob, R.; Štokelj, T.; Grgič, D. Neural-network-based water inflow forecasting. *Control. Eng. Pr.* 1998, 6, 593–600.
49. Jayawardena, A.W.; Fernando, D.A.K. Use of Radial Basis Function Type Artificial Neural Networks for Runoff Simulation. *Comput. Civ. Infrastruct. Eng.* 1998, 13, 91–99.
50. Phien, H.N.; Sureerattanan, S. Neural networks for filtering and forecasting of daily and monthly streamflows. In *Hydrologic Modeling. Proceedings of the International Conference On Water, Engineering, Ecology, Socio-Economics and Health Engineering*; Seol National University, Korea, 18–21 October, Singh, V.P., Won Seo, I., Sonu, J.H., Eds.; Water Resources Publications: Highlands Ranch, CO, USA, 1999.
51. Tokar, A.S.; Johnson, P.A. Rainfall-Runoff Model Using Artificial Neural Networks. *J. Hydrol. Eng.* 1999, 4, 232–239.
52. Zealand, C.M.; Burn, D.H.; Simonovic, S.P. Short term streamflow forecasting using artificial neural networks. *J. Hydrol.* 1999, 214, 32–48.
53. Luč, K.; Ball, J.; Sharma, A. A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting. *J. Hydrol.* 2000, 227, 56–65.
54. Coulibaly, P.; Ancil, F.; Bobee, B. Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. *J. Hydrol.* 2000, 230, 244–257.
55. Luč, K.C.; Ball, J.; Sharma, A. An application of artificial neural networks for rainfall forecasting. *Math. Comput. Model.* 2001, 33, 683–693.
56. Zhang, B.; Govindaraju, R. Geomorphology-based artificial neural networks (GANNs) for estimation of direct runoff over watersheds. *J. Hydrol.* 2003, 273, 18–34.
57. Rajurkar, M.; Kothyari, U.; Chaube, U. Modeling of the daily rainfall-runoff relationship with artificial neural network. *J. Hydrol.* 2004, 285, 96–113.
58. Dawson, C.; Abrahart, R.; Shamseldin, A.Y.; Wilby, R. Flood estimation at ungaged sites using artificial neural networks. *J. Hydrol.* 2006, 319, 391–409.
59. Jain, A.; Kumar, A. An evaluation of artificial neural network technique for the determination of infiltration model parameters. *Appl. Soft Comput.* 2006, 6, 272–282.
60. Riad, S.; Mania, J.; Bouchaou, L.; Najjar, Y. Rainfall-runoff model using an artificial neural network approach. *Math. Comput. Model.* 2004, 40, 839–846.
61. Kisi, O.; Cigizoglu, H.K. Comparison of different ANN techniques in river flow prediction. *Civ. Eng. Environ. Syst.* 2007, 24, 211–231.
62. Chua, L.H.; Wong, T.S.; Sriramula, L. Comparison between kinematic wave and artificial neural network models in event-based runoff simulation for an overland plane. *J. Hydrol.* 2008, 357, 337–348.
63. Shu, C.; Ouarda, T.B.M.J. Regional flood frequency analysis at ungaged sites using the adaptive neuro-fuzzy inference system. *J. Hydrol.* 2008, 349, 31–43.
64. Sciuto, G.; Bonaccorso, B.; Cancelleri, A.; Rossi, G. Quality control of daily rainfall data with neural networks. *J. Hydrol.* 2009, 364, 13–22.
65. Toth, E. Classification of hydro-meteorological conditions and multiple artificial neural networks for streamflow forecasting. *Hydrol. Earth Syst. Sci.* 2009, 13, 1555–1566.
66. Wang, W.-C.; Chau, K.-W.; Cheng, C.-T.; Qiu, L. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *J. Hydrol.* 2009, 374, 294–306.
67. Besaw, L.E.; Rizzo, D.M.; Bierman, P.; Hackett, W.R. Advances in ungaged streamflow prediction using artificial neural networks. *J. Hydrol.* 2010, 386, 27–37.
68. Shamseldin, A.Y. Artificial neural network model for river flow forecasting in a developing country. *J. Hydroinformatics* **2010**, *12*, 22–35.

69. Wu, C.; Chau, K.-W.; Fan, C. Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques. *J. Hydro.** **2010**, *389*, 146–167.

70. Aziz, K.; Rahman, A.; Fang, G.; Shrestha, S. Application of Artificial Neural Networks for Regional Flood Estimation in Australia: Formation of Regions Based on Catchment Attributes. In Proceedings of The Second International Conference on Soft Computing Technology in Civil, Structural and Environmental Engineering, Chania, Crete, Greece, 6-9 September, 6-9 September 2011; Volume 97.

71. Yilmaz, A.G.; Imteaz, M.; Jenkins, G. Catchment flow estimation using Artificial Neural Networks in the mountainous Euphrates Basin. *J. Hydro.** **2011**, *410*, 134–140.

72. Kia, M.B.; Pirasteh, S.; Pradhan, B.; Mahmud, A.R.; Sulaiman, W.N.A.; Moradi, A. An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. *Environ. Earth Sci.* **2011**, *67*, 251–264.

73. Isik, S.; Kalin, L.; Schoonover, J.E.; Srivastava, P.; Lockaby, B.G. Modeling effects of changing land use/cover on daily streamflow: An Artificial Neural Network and curve number based hybrid approach. *J. Hydro.** **2013**, *485*, 103–112.

74. Kalteh, A.M. Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. *Comput. Geosci.** **2013**, *54*, 1-8.

75. Nourani, V.; Baghanam, A.H.; Adamowski, J.; Gebremichael, M. Using self-organizing maps and wavelet transforms for space–time pre-processing of satellite precipitation and runoff data in neural network based rainfall–runoff modeling. *J. Hydro.** **2013**, *476*, 228–243.

76. Aziz, K.; Rai, S.; Rahman, A. Application of artificial neural networks and genetic algorithm for regional flood estimation in Eastern Australia. In *Hydrology and Water Resources Symposium 2014*; Engineers Australia, 24-27 Feb, Perth, Australia, 2014.

77. Makwana, J.J.; Tiwari, M.K. Intermittent Streamflow Forecasting and Extreme Event Modelling using Wavelet based Artificial Neural Networks. *Water Resour. Manag.** **2014**, *28*, 4857–4873.

78. Aziz, K.; Rai, S.; Rahman, A. Design flood estimation in ungauged catchments using genetic algorithm-based artificial neural network (GAANN) technique for Australia. *Nat. Hazards** **2015**, *77*, 805–821.

79. Lee, J.; Kim, C.; Lee, J.E.; Kim, N.W.; Kim, H. Application of Artificial Neural Networks to Rainfall Forecasting in the Geum River Basin, Korea. *Water** **2018**, *10*, 1448.

80. Sadeghi, M.; Asanjian, A.A.; Faridzad, M.; Nguyen, P.; Hsu, K.; Sorooshian, S.; Breithwaite, D. PERSIANN-CNN: Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks–Convolutional Neural Networks. *J. Hydrometeorol.** **2019**, *20*, 2273–2289.

81. Australian Bureau of Meteorology, Available online: http://www.bom.gov.au/. (accessed on 21 May 2018)

82. Zaman, M.A.; Rahman, A.; Haddad, K. Regional flood frequency analysis in arid regions: A case study for Australia. *J. Hydro.** **2012**, *475*, 74–83.

83. Rahman, A.; Haddad, K.; Kuczera, G.; Weinmann, P.E. Regional flood methods, Chapter 3 in Book 3. In *Australian Rainfall and Runoff—A Guide to Flood Estimation*; Commonwealth of Australia: 2019, Canberra, Australia.
89. Ball, J.; Babister, M.; Nathan, R.; Weeks, W.; Weinmann, E.; Retallick, M.; Testoni, I. (Eds.) *Australian Rainfall and Runoff—A Guide to Flood Estimation*; Commonwealth of Australia: 2019, Canberra, Australia.

90. Kuczera, G.; Franks, S. At-site flood frequency analysis, Chapter 2 in Book 3. In *Australian Rainfall and Runoff—A Guide to Flood Estimation*; Commonwealth of Australia: 2019, Canberra, Australia.

91. Kuczera, G. Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference. *Water Resour. Res.* **1999**, *35*, 1551–1557.

92. Blöschl, G.; Sivapalan, M.; Savenije, H.; Wagener, T.; Viglione, A. *Runoff prediction in Ungauged Basins: Synthesis Across Processes, Places and Scales*; Cambridge University Press: Cambridge, UK, 2013.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).