Artificial Intelligence-Based Regional Flood Frequency Analysis Methods: A Scoping Review

Amir Zalnezhad [1], Ataur Rahman [1,*], Nastaran Nasiri [2], Khaled Haddad [1], Muhammad Muhitur Rahman [3], Mehdi Vafakhah [4], Bijan Samali [1] and Farhad Ahamed [5]

[1] School of Engineering, Design and Built Environment, Western Sydney University, Office XB.3.43, Kingswood (Penrith Campus), Locked Bag 1797, Penrith South DC 1797, NSW, Australia

[2] Independent Researcher, Third Floor, No. 12, Miri Alley, Ghasemi Street, Ghadiri Street, Banihashem Avenue, Tehran 1663773415, Iran

[3] Department of Civil and Environmental Engineering, College of Engineering, King Faisal University, Al-Ahsa 31982, Saudi Arabia

[4] Department of Watershed Management, Faculty of Natural Resources, Tarbiat Modares University, Noor P.O. Box 46417-76489, Iran

[5] School of Computer, Data and Mathematical Sciences, Western Sydney University, Building ER, Victoria Rd., Rydalmere 2116, NSW, Australia

* Correspondence: a.rahman@westernsydney.edu.au

Abstract: Flood is one of the most destructive natural disasters, causing significant economic damage and loss of lives. Numerous methods have been introduced to estimate design floods, which include linear and non-linear techniques. Since flood generation is a non-linear process, the use of linear techniques has inherent weaknesses. To overcome these, artificial intelligence (AI)-based non-linear regional flood frequency analysis (RFFA) techniques have been introduced over the last two decades. There are limited articles available in the literature discussing the relative merits/demerits of these AI-based RFFA techniques. To fill this knowledge gap, a scoping review on the AI-based RFFA techniques is presented. Based on the Scopus database, more than 1000 articles were initially selected, which were then screened manually to select the most relevant articles. The accuracy and efficiency of the selected RFFA techniques based on a set of evaluation statistics were compared. Furthermore, the relationships among countries and researchers focusing on AI-based RFFA techniques are illustrated. In terms of performance, artificial neural networks (ANN) are found to be the best performing techniques among all the selected AI-based RFFA techniques. It is also found that Australia, Canada, and Iran have published the highest number of articles in this research field, followed by Turkey, the United Arab Emirates (UAE), India, and China. Future research should be directed towards identification of the impacts of data quantity and quality, model uncertainty and climate change on the AI-based RFFA techniques.

Keywords: regional flood frequency analysis; artificial neural networks; flood; artificial intelligence

1. Introduction

Flood is one of most devastating natural disasters, resulting in significant economic losses including human deaths [1,2]. This damages both rural and urban infrastructure like bridge and drainage systems [3,4]. Flood generally leaves undesirable sediments and debris in the affected lands [5,6], which can disrupt transportation networks [7], clog drainage infrastructure and sewers [8,9] and may make lands unproductive. The cleaning up of flood debris is usually costly, not to mention the disruption to the daily lives of the community involved [10,11]. Due to climate change, the frequency and magnitude of floods are increasing [12].

Flood forecasting requires significant efforts, and it is usually the responsibility of a large government organisation. Governments spend a significant amount on various

projects to identify flood-safe areas, which are used to build cities. Researchers have developed numerous methods to estimate design floods, which are used to build flood-safe infrastructure [13,14]. Design flood is defined as a flood level or discharge associated with a return period or annual exceedance probability such as a 100-year flood.

In addition to traditional techniques, like the rational method, physical and numerical [15,16] models have been proposed for design flood estimation. Most of the physical models require in-depth knowledge of flood processes [17,18], making them difficult to use in practice. Van den Honert and McAneney [19] pointed out the common limitations associated with these physical models [20,21], which include model inaccuracies resulting in systematic errors (over or underestimation of design floods) [22,23]. On the other hand, data-driven models have been quite popular for flood estimation in recent years [24]. Examples include a quantile regression technique and a probabilistic rational method [25]. This is because they usually consider climate factors and catchment characteristics in developing models, which are easier to apply [26,27]. A flood frequency analysis (FFA) is the most popular method to estimate design floods, which uses observed peak discharge data disregarding catchment characteristics [28,29]. A normal distribution [30,31], log-normal distribution [32,33], Gumbel distribution [34,35], generalised extreme value distribution and log-Pearson type III distribution [36,37] are some of the most commonly used flood frequency distributions in FFA. One of the major limitations of FFA is the lack of long and good quality recorded flood data at the location of interest. To overcome data limitations, hydrologists have proposed a regional flood frequency analysis (RFFA), which attempts to estimate design floods at an ungauged catchment based on the concept of a homogeneous region, which pools observed flood data from a group of similar catchments to estimate design floods at the ungauged catchment [38,39]. This method became more popular among researchers than physical models because it saves time and resources [40]. Probabilistic rational method (PRM) [41], multiple linear regression (MLR) [42,43], quantile regression techniques (QRT) [44,45], and index flood method (IFM) [46,47] are some of the most commonly used RFFA techniques. However, some of the early RFFA techniques (e.g., rational method) have lost their popularity due to their inconsistency and inappropriate model assumptions.

In the past two decades, scientists suggested hybrid or mixed methods to increase the relative accuracy of RFFA models [48,49]. Although some early linear models have been improved, they may not be accurate under some circumstances as flood generation is basically a non-linear process [50]. Hydrologists attempted to apply non-linear methods in RFFA such as a non-linear regression analysis (where log-transformation of the variables is considered). Artificial intelligence (AI)-based methods are also non-linear, but more powerful than simple non-linear models like log-log ones as they can consider many different combinations of variables and complex non-linear processes in model building. Given that the majority of flood estimation methods are data driven, they require a great deal of simplification and assumptions to be practical, accessible, and implementable [51,52]. They require relatively fewer input data and minimal knowledge of fundamental physical processes involved. Over the last two decades, non-linear AI-based RFFA methods have grown in popularity over physical models as these provide more accurate results and are easier to apply [53,54]. Artificial neural networks (ANNs) [55,56], support vector regression (SVR) [57–60], adaptive neuro-fuzzy inference system (ANFIS) [61,62], genetic algorithms (GA) [63,64] and hybrid, mixed and combined approaches [65,66] are some of the most popular AI-based flood estimation methods. As AI-based models are relatively new in flood estimation, it is not easy to decide which one is to be applied for a given problem [67,68].

There are several important aspects to consider when building models based on AI. Firstly, these models like all other data-driven models need enough data to develop and test the model [67]. If adequate data exist, it is often possible to build, test, and evaluate an AI-based model (similar to many other RFFA models) by dividing the data into training, test, and evaluation data sub-sets [69,70]. Cross-validation is also often used in building RFFA models when less data samples are available [71]. The more data used in the modeling,

the less generalization error occurs, meaning that the final model can be used on different sites with limited or no data available. Other benefits of having adequate data include the simplicity of using different distribution methods, the ability to account for lost data or missing variables, and, most crucially, the ability to train and validate the model multiple times to develop the best possible model [72,73]. However, it should be noted that data quality is of significant importance in developing and testing accurate models.

A scoping review aims to identify gaps in the current research on a selected field and highlight areas requiring further research. This is different to a systematic review, which aims to identify, appraise, and synthesize all available research studies that are relevant to a given question in review. In this study, a scoping review is assumed to be adequate as this is dedicated to find gaps in the current AI-based RFFA studies and point out areas that need further investigation. Although there have been few previous reviews on flood frequency analysis [74–77], there has been no review paper on AI-based RFFA. Hence, this paper will fill this knowledge gap and will serve as a basis for conducting future research on AI-based RFFA techniques.

2. Methodology of the Scoping Review

The standard methodology of a scoping review was followed to carry out this review, which consisted of (a) the formulation of research questions; (b) determining a list of relevant keywords; (c) the selection of relevant databases; (d) the selection of inclusion/exclusion criteria; (e) the selection of criteria for article selection; and (f) the selection of criteria to evaluate the finally extracted literature. The following research questions were formulated: (a) Why are AI-based methods needed in RFFA? (b) Do AI-based RFFA methods provide more accurate flood quantile estimates? (c) What levels of uncertainties are associated with the AI-based RFFA methods? (d) Who are the most active researchers in this field and how to map their collaborations? The following keywords were selected: flood, floods, flood frequency, regional flood frequency, artificial intelligence, ANN, ANFIS, SVM, GA, estimation, frequency, prediction, and analysis. Scopus and Google Scholar were selected as preferred database since it was believed that these would capture all the relevant articles on AI-based RFFA.

Journal rankings, citations, authorship, and comparative studies were the primary criteria for selecting relevant articles. Initially, 1033 articles were identified, however, after screening, only 30 articles were found to be on the AI-based RFFA. These articles were thoroughly examined, particularly those that use statistical criteria to compare various AI-based RFFA methods. Figure 1 illustrates the method of article selection in this study.

It should be noted that the relative accuracy of an RFFA study/method cannot be compared with another study/method directly using the reported error statistics, since in most cases they use different dataset to develop and test the methods. For example, an approach trained using quality data from real world floods reporting 80% accuracy is far better in practice in comparison to another approach trained over very few samples of lab simulated data reporting 99% accuracy. Hence, the comparison of AI-based RFFA methods made in the following section is taken to be as a guide only. In real-world applications, several methods should be applied and compared to select a preferred method for design flood estimation. One method found to be better in one geographic region does not guarantee its superiority at another location.

Identification

Screening

Eligibility

Included

Keywords Selected for Searching via Scopus and Google Scholar

Records Identified Through Database Searching (n=1033)

Duplicates Removed (n=765)

Unrelateds Removed (n=537)

Excluded Research Topics (n=35)

Focus on Flood Quantile Estimation (n=29)

Final Number of Articles (n=30)

Flood

Artificial Intelligence Methods

Estimation

Frequency

Prediction

Analysis

Regional Flood Frequency

ANN, ANFIS, SVM, GA

Flood Forecasting

Daily, Hourly Flood Analysis

Linear Methods
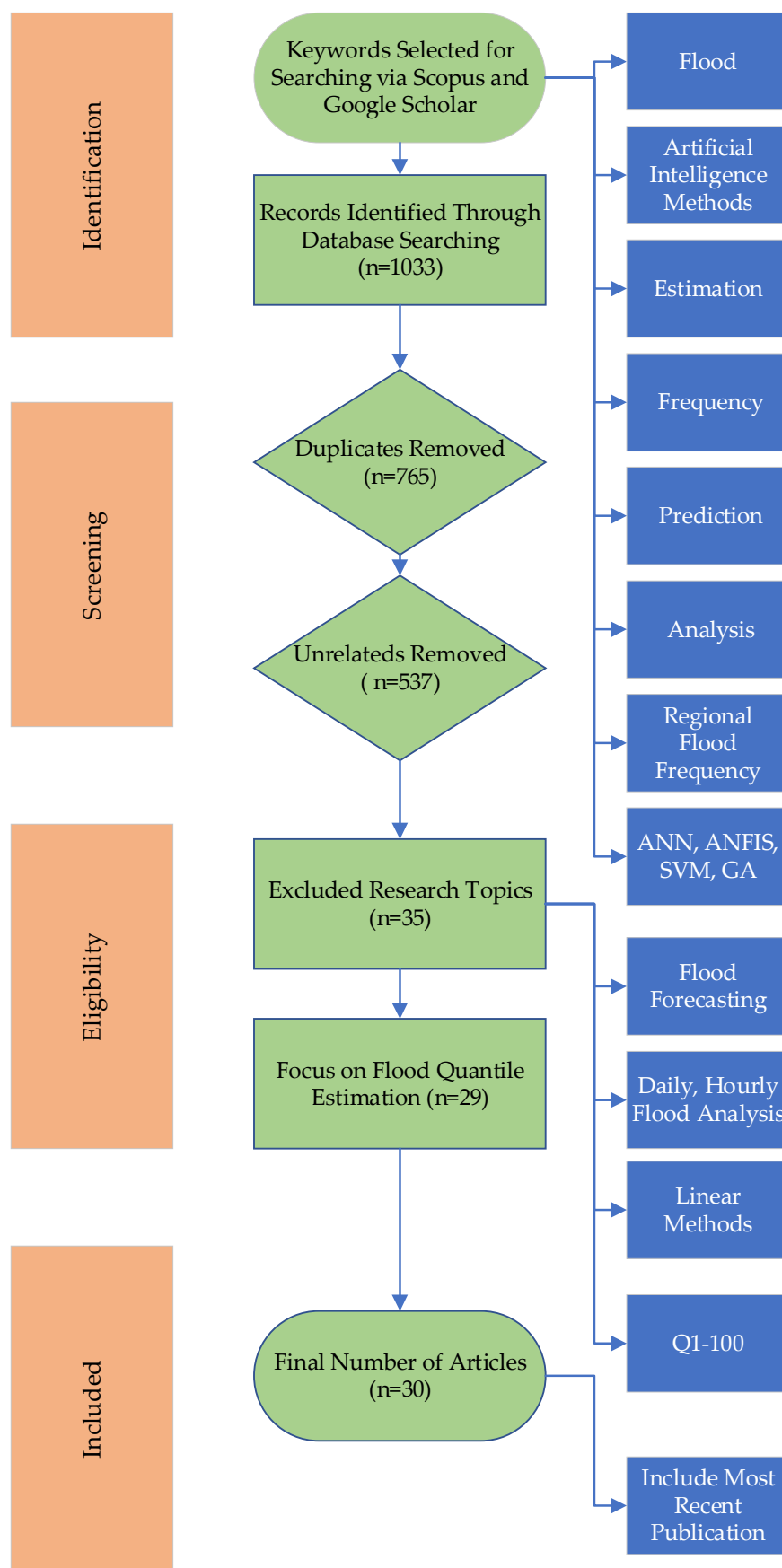
Q1-100

Include Most Recent Publication

Figure 1. Article selection method in this study.

3. AI-Based RFFA Methods

Figure 2 illustrates how to develop an AI-based RFFA model. It is important to identify input variables. Some of the most used input variables include catchment area (A), longitude (LON), latitude (LAT), elevation (EV), drainage density (DD), average annual maximum daily precipitation (AP), rainfall intensity (I), vegetation coverage (VC), slope (SL), and relative elevation (RE), fraction forested area (F), mean annual evapotranspiration (MAE), shape factor (SF), and stream density (SDEN). Output variables include maximum stream flow, flood quantiles, and time to peak. Collected data are then standardised to avoid a scaling problem. To build a reliable model, training, validation [69,70], and test data are required. Different statistical measures are used to compare alternative models such as RMSE, RMSNE, and R^2.
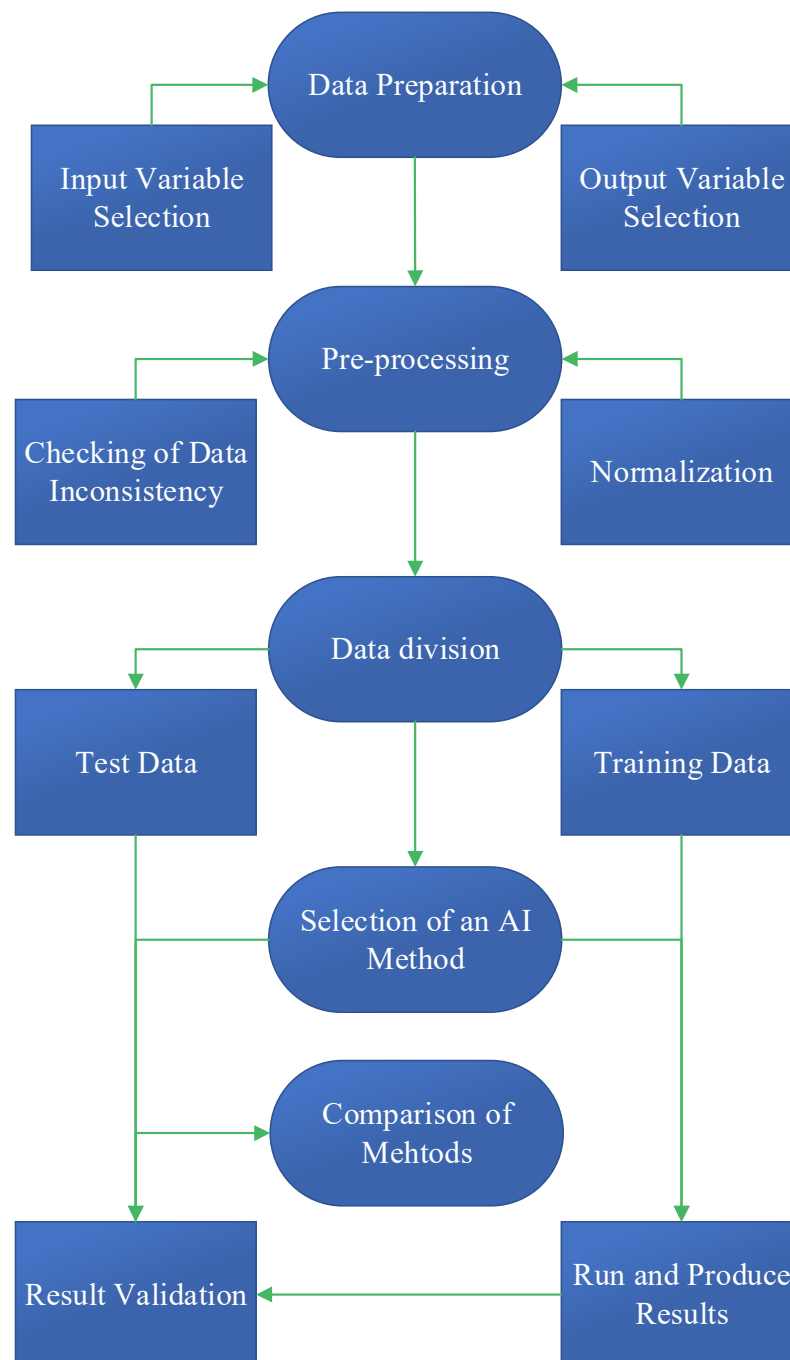

Figure 2. Steps in building an AI-based RFFA model.

3.1. ANN-Based RFFA Models

The ANN performs like a human nervous system in that it learns from previous trials and decides how to come up with a better model by exploiting the best possible links between dependent (flood quantiles such as Q_{10}) and independent variables (such as rainfall) in a series of steps. ANN, as a data-driven tool, does not require any physical knowledge of flood processes involved [78,79]. One of the limitations of this method is lack of physical interpretation of the developed models.

Shu and Burn [51] compared the ANN with a parametric regression analysis in one of the first articles on the AI-based RFFA. They found that a properly developed ANN model outperforms both linear (REG-OLS) and non-linear (REG-NONLINEAR) regression-based methods. They also compared the results of a single ANN to those of ANN ensembles, concluding that the latter provided more accurate flood estimates. Jingyi and Hall [80] compared four different models, including the residuals method, Ward's method, fuzzy c-mean, and a variation of the ANN, known as the Kohonen network. They found that, while other methods may be somewhat useful, the ANN method produced the lowest standard error of estimate and could be a useful method if adequate data from enough sites are available.

Dawson et al. [81] applied ANN using data from 850 stations. They compared the results of the ANN method to those of multiple regression models and found that ANN outperformed the other models. They noted that because there is little need to understand the physics of flood generation processes, scientists from all disciplines, not just hydrologists, could use the ANN method. Shu and Ouarda [56] developed RFFA models based on ANN and CCA using data from 151 catchments and found that the ANN–CCA combination provided better generalisation and accuracy. Srinivas et al. [49] used AI-based RFFA and regression methods involving various AI-based algorithms. To determine the best approach for data clustering, a regression analysis, CCA, and FCM algorithms were compared. They found that leave-one-out cross-validation based on the FCM algorithm produced better results when evaluating the accuracy of the estimated flood quantities.

Ouarda and Shu [82] estimated regional low flows at ungauged sites using a single ANN, an ANN ensemble, and a multiple regression analysis. When they compared single and ANN ensemble methods to the traditional regression analysis, they found that the AI-based methods are more accurate. Furthermore, the ANN ensemble outperformed the single ANN in terms of the generalisation of results.

Singh et al. [83] developed two types of AI-based RFFA methods (BNN and M5) using data from 93 catchments in India. They observed that the M5 method performed better than the BNN. They also mentioned that the M5 method has some advantages over the BNN, such as providing more information on the generated model and being highly efficient in training. Seckin et al. [84] compared the performance of various linear, non-linear, and AI-based RFFA methods, including MLP, RBNN, GRNN, MLR, and MNLR, using data from 13 basins. They observed that MLP outperformed other models by having the lowest RMSE for test data and the best accuracy, as well as requiring less computational effort and fewer input data to run the model.

Aziz et al. [85] also applied the ANN in RFFA in Australia and found that the ANN method worked well with only two predictor variables. This performed better when all the data from seven different regions were combined to form one region rather than studying them separately. This study shows better performance when using relatively large datasets in ANN. To estimate flood quantiles in ungauged catchments, Alobaidi et al. [86] proposed a new ensemble architecture based on the ANN by using data from 151 hydrometric stations in Canada. They observed that the proposed ANN-ensemble methods (G-EANN and EANN) outperformed the previous case study by Shu and Ouarda [56]. Based on evaluation statistics, the proposed G-EANN was determined to be the best method.

Durocher et al. [87] proposed PPR, a new method that combines ANN and the generalised additive model (GAM). They compared the results with eight different methods proposed by other researchers, including traditional [88,89], GAM [90], spatial [88,91], and

ANN [56]. Using at least 15 years of recorded data collected from 151 hydrometric stations in Canada, they reported that, despite having a simpler structure than ANN, the PPR method produced comparable results. While benefiting from the advantages of ANN and GAM, the simplicity of GAM and the ability of ANN to consider relationships between different characteristics help to overcome disadvantages associated with both models in terms of parsimony and interpretability.

Ouali et al. [92] used over 15 years of data collected from three different regions for a total of 424 catchments located in Canada and the United States to find the best possible RFFA method. They compared the performance of various ANN-based linear and non-linear method combinations with hybrid linear, non-linear, and semi-linear methods. In comparison to other studies, such as Ouarda et al. [93], Chebana et al. [90], and Ouali et al. [94], they found that non-linear methods were the most efficient, and while linear methods were the least accurate, semi-linear methods performed slightly better than linear methods. Kordrostami et al. [95] evaluated the performance of five different types of ANN methods in RFFA in Australia using data from 88 gauging stations spanning 25–82 years. They observed that ANN techniques with fewer predictor variables performed better than those with more, unless all the eight variables were utilized.

Linh et al. [96] evaluated the performance of WNN models as a hybrid of ANN in design flood estimation using 37 years of data from three hydrometric stations. They used RMSE, R^2, and NASH to determine the best method among the MLR, ANN, and WNN, and found that the latter (hybrid) performed significantly better than the single ANN. They also reported that WNN performed better in terms of the generalisation of results. Desai and Ouarda [97] developed various AI-based RRFA methods using data from 151 catchments in Canada, including different combinations of CCA with RFR, ANN, ANN ensemble, and MLR methods. The CCA technique, in general, increased the efficiency of selected methods by delineating homogeneous regions of stations. Based on statistical indices such as NASH, RMSE, and RRMSE, they found that CCA-PFR was the best method with less complexity, greater reliability, and accuracy when compared to other single or ensemble ANN-based RFFA methods.

Aziz et al. [67] used ANN, GEP, and OLS-based QRT methods to estimate flood quantiles using a recorded dataset spanning 25 to 75 years from 452 Australian catchments. They used statistical indicators such as median RE to evaluate the performance of the selected methods and found ANN and GEP to be more accurate than the QRT, particularly ANN for smaller ARIs.

It can be seen from the above discussion that ANN has been widely used in RFFA. The relative accuracy of the developed ANN-based RFFA methods has been found to vary from study to study as this largely depends on the data length and accuracy.

3.2. ANFIS-Based RFFA Models

Several hydrologists have used ANFIS to estimate design floods. This model clusters data in linguistic terms using fuzzy systems [98], which is useful in the absence of a comprehensive set of data [99,100]. Fuzzification, Rule, Normalization, Defuzzification, and Summation are the five layers of this method. To build a network between different nodes and a set of parameters, ANFIS employs the Takagi and Sugeno methods, which employ if-then fuzzy rules to modify certain parameters to reduce overall prediction error [101]. This method gradually trains data based on its learning capabilities. Some of the common types of ANFIS used in the literature [102] are grid partitioning (GP), subtractive clustering (SC), and fuzzy c-mean (FCM) clustering.

Shu and Ouarda [55] estimated design floods at ungauged sites using the ANFIS method. They compared this method to other non-linear ones such as ANN and found that ANFIS performed better in terms of accuracy and learning capability than other models. Aziz et al. [103] used a CANFIS-based RFFA method using data from 452 catchments in Australia. They noted that these methods work very well for regional flood quantile esti-

mates with reasonable accuracy. They did note, however, that the QRT method developed by Haddad and Rahman [104] performed better than the AI-based methods.

Bozchaloei and Vafakhah [105] assessed the performance of ANFIS, ANN, and NLR at their site of interest using 20 years of data from 33 hydrometric stations in Iran. They found that the ANFIS method performed better than the other two methods, yielding more robust and accurate results with the lowest RMSE and highest NASH coefficients.

Kumar et al. [61] used a dataset of 15–29 years collected on 17 catchments in India to compare the performance of two RFFA methods based on AI, FIS, and ANN. Using statistical indices such as RMSE and mean absolute error (MAE), they discovered that ANFIS outperformed ANN. In terms of accuracy and reliability, both AI-based RFFA methods outperformed QRT. Aziz et al. [106] used similar methods based on data from 452 Australian catchments and noted that ANN produced the most accurate and robust results based on statistical indicators such as RE, and hence they proposed that ANN could be a viable method in Australia's east.

Zalnezhad et al. [102] used data from 181 catchments in eastern parts of Australia to compare the performance of ANFIS based methods (FCM, SC and GP) with QRT method and noted that ANFIS (FCM) is the best performing method. They suggested that it could be a viable alternative to the conventional methods used in eastern parts of Australia.

From the above discussion, it is evident that ANFIS can be used to divide the available catchment data set into more homogeneous groups, which is likely to enhance overall prediction accuracy. The available number of catchments should be large enough so that meaningful sub-groups can be formed as a too small group with a handful of catchments will result in misleading error statistics.

3.3. SVM-Based RFFA Models

The SVM method is widely used for classification, which examines data at higher dimensions [107,108]. Several types of kernels assist SVM in classifying data by minimising data margins, eliminating outliers, and focusing on relationships between the test and training data. The most common kernel types used for developing SVM-based models include linear, polynomial, radial basis function (RBF), and sigmoid function. Among these, the SVM-based RBF kernel is the most used method that produces robust and consistent results.

Gizaw and Gan [109] developed RFFA-based ANN and SVR methods using data collected from 49 stations in Canada. When the results of these two methods were compared, they found that the SVR method outperformed the ANN in terms of consistency and generalisation ability. They also mentioned that better SVR performance could be attributed to smaller datasets, whereas ANN would most likely produce more accurate results for larger datasets. Sharifi Garmdareh et al. [110] estimated design floods using SVR, ANFIS, ANN, and NLR methods using more than 20 years of recorded data from 55 hydrometric stations in Iran. They tested various strategies for determining the best combination of input variables and found that gamma testing (GT) was the most effective, which can improve the result of ANFIS and SVR over a single method and that using GT reduced the number of input variables. They also noted that combining GT with the ANFIS produced the best results, followed by GT + SVR.

Ghaderi et al. [111] used ANFIS, SVM, and GEP to estimate flood quantiles with a 50-year return period. From 21 years of data collected from 47 catchments in Iran, they used GM and M-test to identify the most important predictor variables and the best ratio of test and training data. They compared the results of the three methods and noted that all three were "good" in terms of NASH, with the SVM method slightly outperforming the others in terms of R^2 and RMSE. Vafakhah and Bozchaloei [112] used SVR, ANN, and NLR to estimate design floods using data collected from 33 stations in Iran over 20 years. They noted that, according to RRMSE and NASH, SVR is the most efficient method of the three and can be used for regional flood duration curve analysis.

Haddad and Rahman [65] used 25 to 82 years of data from 202 catchments in Australia to evaluate 15 different combinations of multidimensional scaling (MDS), bayesian generalised least squares (BGLSR), and SVR methods to estimate design floods. They found that the MDS-based SVR method with RBF kernel outperforming others, including linear, polynomial, RBF, and sigmoid kernels, in terms of consistency and accuracy of the results. They also noted that using MDS improved the overall performance of all the methods.

Allahbakhshian-Farsani et al. [59] used 19 years of data from 54 hydrometric stations in Iran to compare the performance of several AI-based RFFA methods. This study employed methods such as SVR, multivariate adaptive regression spline (MARS), boosted regression trees (BRT), and projection pursuit regression (PPR). Using various statistical indices such as NASH, RMSE, RMSE, and R^2, they noted that the SVR model based on the RBF kernel outperformed all the others, including non-linear regression.

From the above discussion it can be stated that both SVM and SVR were used in RFFA. A large set of catchments are needed to group them into homogeneous sub-sets which can then be subjected to SVR to estimate flood quantiles.

3.4. GA and Hybrid Type of AI-Based RFFA Models

Hybrid models typically produce better results. As shown in Table 1, many scientists have conducted experiments based on combining various AI-based RFFA models. Some of the most common hybrid models include genetic algorithm (GA) combined with ANN or ANFIS. The GA is commonly used as a hybrid method in conjunction with other methods, particularly ANN [106]. Another popular hybridisation technique used in RFFA is the combination of canonical correlation analysis (CCA) with ANN and ANN ensembles, as well as ANFIS methods. CCA improves the performance and reduces the complexity of ANN-based RFFA models by exploiting regional flood data [92,97].

Seckin and Guven [113] used data from 543 catchments in Turkey to compare two genetic programming-based techniques (GEP and LGP) with the linear regression (LR). They found that GEP was the best operating method, closely followed by LGP and that both soft programming methods outperformed the LR method. Aziz et al. [114] evaluated the developed RFFA method, a combination of GA and ANN called GAANN, using data from 452 stations in Australia. They also compared the results of their proposed method to BPANN and noted that both methods produced similar results. When the results were compared to QRT, they concluded that the proposed AI-based RFFA could be a viable alternative to the traditional QRT method in Australia.

Table 1. Summary of AI-based RFFA studies (* indicates the best model) (ANN = Artificial neural network; GA = Genetic algorithm, BGLS-QRT-ROI: Bayesian generalized least squares QRT combined with region of influence approach, BNN = Backpropagation neural network, CANFIS = Co-active neuro fuzzy inference system, GEP = Gene-expression programming, GRNN = generalized regression neural networks, LGP = Linear genetic programming (LGP), LR = Linear regression, M5 = M5 model tree, MLP = Multi-layer perceptrons, MLR = Multiple linear regression, MNLR = Multiple non-linear regression, QRT = Quantile regression technique, RBNN = Radial basis function-based neural networks, G-EANN = generalized ANN-Ensembles, EANN = ANN-Ensembles, GAANN = GA-based ANN, BPANN = Back propagation for ANN, FIS = Fuzzy inference system, CCA = canonical correlation analysis, NLCCA = Non-linear canonical correlation analysis, BGLSR = Bayesian generalised least squares, MDS = multidimensional scaling, MARS = multivariate adaptive regression spline, BRT = boosted regression trees, PPR = projection pursuit regression, WNN = wavelet neural network and RFR = random forest regression).

Reference	Author, Year	Model	Predictor Variables (Inputs)	Model Output	Catchment, Year	Journal	Country (Catchment)	RMSE *	RRMSE/NASH *	R² *
[102]	Zalnezhad et al., 2022	ANFIS(FCM) * ANFIS(SC) ANFIS(GP) QRT	A, I, MAR, SF, MAE, SDEN, S1085, FOR	Q_{2-100}	181 Stations 40–89 Year	*Water*	Australia	50.88	RRMSE = 0.78	NA
[97]	Desai and Ouarda, 2021	CCA-RFR * PFR CCA-GAM EANN ANN CCA-MLR CCA-Kriging CCA-EANN CCA-ANN	A, MBS, FAL, AMP, AMD	Q_{10-100}	151 stations, ≥15 year	*Journal of Hydrology*	Canada (Quebec)	0.05	NASH = 0.57 RRMSE = 29.44	NA
[96]	Linh et al., 2021	WNN * ANN	SLP, SST	Max monthly discharge (MAD)	3 stations, 37 years	*Acta Geophysica*	Iran (Golestan Dam, Madarsoo)	0.68	NASH = 0.99	0.99
[59]	Allahbakhshian-Farsani et al., 2020	SVR * MARS BRT PPR NLR	A, AA, AMP, MXP, NDP, CC, CR, TC, P, SL, DD, SS, MBS, PF, SDT, RA, BL, FLA, FOR, RLA, DA, WA, EL, MXEL, MNEL	Q_{2-200}	54 stations, 19 years	*Water Resources Management*	Iran (Karun and Karkhe River)	50.70	NASH = 0.94 RRMSE = 63.93	0.96
[95]	Kordrostami et al., 2020	ANN	A, AEV, AMP, FOR, I, SS, SF and DD	Q_{5-100}	88 stations, 25–82 years	*Geosciences*	Australia (New South Wales)	NA	RRMSE = 0.48	0.74
[65]	Haddad and Rahman, 2020	MDS-SVR * MDS-BGLSR	A, AEV, SF, DD, SS, FOR, I and AMP	Q_{2-100}	202 stations, 25–82 years	*Natural Hazards*	Australia (New South Wales and Victoria)	NA	RRMSE = 56	0.78
[112]	Vafakhah and Khosrobeigi Bozchaloei, 2020	SVR * ANN NLR	A, AA, AEV, P, MBS, MXEL, MNEL, EL, SL, DD, SS, AMP, T, PF, RLA, BL, GA, RA	Q_{2-90}	33 Stations, 20 years	*Water Resources Management*	Iran (Namak Lake)	0.11	NASH = 0.91 RRMSE = 1.45	0.96

Table 1. *Cont.*

Reference	Author, Year	Model	Predictor Variables (Inputs)	Model Output	Catchment, Year	Journal	Country (Catchment)	RMSE *	RRMSE/NASH *	R² *
[111]	Ghaderi et al., 2019	SVM * ANFIS GEP	A, P, MBS, EL, L, SL, SS, DD, MXSO, FF, L, CR, CC, AMP, MXP, BL, FOR	Q_{50}	47 stations, 21 years	*Arabian Journal of Geosciences*	Iran (South-west)	239.94	NASH = 0.75	0.76
[110]	Sharifi Garmdareh et al., 2018	ANFIS * SVR ANN NLR	A, AEV, P, DD, MXEL, MNEL, MBS, EL, SL, SS, T, AMP,	Q_{2-100}	55 stations, 20 years	*Hydrological Sciences Journal*	Iran (Namak Lake)	8.40	NASH = 0.90	0.95
[67]	Aziz et al., 2017	ANN * GEP * QRT	A, AEV, AMP, SS, I	Q_{2-100}	452 stations, 25–75 years	*Stochastic Environmental Research and Risk Assessment*	Australia (New South Wales, Victoria, Queensland and Tasmania)	Na	NASH for ANN for smaller ARIs = 0.78 NASH for GEP for larger ARIs = 0.73	NA
[92]	Ouali et al., 2017	NLCCA-GAM * NLCCA-EANN CCA-ANN CCA-EANN NLCCA-ANN NLCCA-GAM/ STPW	A, MBS, FAL, AMP, AMD	Q_{10-100}	151, 204 and 69 stations, ≥15 years	*Journal of Advances in Modeling Earth Systems*	Canada and United states (Quebec, Arkansas, Texas)	NA	RRMSE = 0.28 NASH > 0.8	NA
[109]	Gizaw and Gan, 2016	SVR * ANN	A, SS, SL, TC, I, AMP	Q_{10-100}	26 and 23 stations, ≥15 years	*Journal of Hydrology*	Canada (British Columbia, Ontario)	46.2	NA	0.7
[106]	Aziz et al., 2016	ANN * GAANN CANFIS GEP	A, AEV, I, AMP, SS,	Q_{2-100}	452 Stations, 25–75 years	*Artificial Neural Network Modelling (Book)*	Australia (New South Wales, Victoria, Queensland and Tasmania)	NA	NASH = 0.69	NA
[61]	Kumar et al., 2015	FIS * ANN L-moments (PE3)	A, AMP, SDT, EL	Q_{2-1000}	17 stations, 15–29 years	*Water Resources Management*	India (Godavari river)	2.32	Na	NA
[114]	Aziz et al., 2015	GAANN BPANN	A, I	Q_{2-100}	452 stations, 25–75 years	*Natural Hazards*	Australia (New South Wales, Victoria, Queensland, and Tasmania)	NA	NA	NA
[105]	Bozchaloei and Vafakhah, 2015	ANFIS * ANN NLR	A, AA, AEV, P, MBS, MXEL, MNEL, EL, SL, DD, SS, AMP, T, PF, RLA, BL, GA, RA	Q_{2-92}	33 stations, 20 years	*Journal of Hydrologic Engineering*	Iran (Namak Lake)	0.008	NASH = 0.92	0.99

Table 1. *Cont.*

Reference	Author, Year	Model	Predictor Variables (Inputs)	Model Output	Catchment, Year	Journal	Country (Catchment)	RMSE *	RRMSE/NASH *	R² *
[87]	Durocher et al., 2015	PPR *	A, SL, SS, MBS, FOR, FAL, AMP, AMPS, AMPL, MLS, AMD	Q_{10-100}	151 stations, ≥15 years	*Journal of Hydrometeorology*	Canada (Quebec)	NA	RRMSE = 0.40	NA
[86]	Alobaidi et al., 2015	G-EANN * EANN	A, MBS, FAL, AMD, AMP	Q_{10-100}	151 stations, ≥15 years	*Advances in Water Resources*	Canada (Quebec)	NA	RRMSE = 0.34	NA
[85]	Aziz et al., 2014	ANN * QRT	A, AEV, AMP, SS, I	Q_{2-100}	452 stations, 25–75 years	*Stochastic Environmental Research and Risk Assessment*	Australia (New South Wales, Victoria, Queensland, Tasmania)	NA	NA	NA
[103]	Aziz et al., 2013	BGLS-QRT-ROI * CANFIS	A and I	Q_{2-100}	452 stations, 25–75 years	*Journal of Hydrological Environment Resources*	Australia (New South Wales, Victoria, Queensland, and Tasmania)	NA	NA	NA
[84]	Seckin et al., 2013	MLP * L-moment RBNN GRNN MLR MNLR	A, EL, LAT, LON, and RP	$Q_{1.111-1000}$	13 stations, 10-39 years	*Water Resources Management*	Turkey (East Mediterranean River)	0.173	NA	0.84
[113]	Seckin and Guven, 2012	GEP * LGP LR	A, EL, LAT, LON, and RP	$Q_{25.7-174.3}$	543 stations, ≥15 years	*Water Resource Management*	Turkey (Rivers across the country)	NA	NA	0.57
[83]	Singh et al., 2010	BNN * M5	A, MRD, AMP, RP, MBS and FOR	$Q_{2.33}$	93 stations, 10–83 years	*Water Resources Management*	India (Catchments across the country)	NA	NA	NA
[82]	Ouarda and Shu, 2009	ANN * Multiple regression model ANFIS *	A, FAL, FOR, AMD, AMPL, NT27, CN	Q_{2-10}	134 stations, ≥10 years	*Water Resources Research*	Canada (Quebec)	27.33	NASH = 0.96, RRMSE = 36.17	NA
[55]	Shu and Ouarda, 2008	ANN NLR NLR-R	A, MBS, FAL, AMP, AMD, HDB, TOPO	Q_{10-100}	151 stations- ≥15 years	*Journal of Hydrology*	Canada (Quebec)	316	NASH = 0.85 RRMSE = 57	NA
[49]	Srinivas et al., 2008	SOFM * CCA Regional regression	A, SS, SRC, SSC, AMP, SL, EL, FOR, R24h	Q_{2-100}	11 stations, 6–42 years	*Journal of Hydrology*	United states (Indiana)	NA	RRMSE = 0.276	NA
[56]	Shu and Ouarda, 2007	ANN * ANN-CCA	A, AMD, AMP, FAL, MBS	Q_{10-50}	151 stations, ≥15 year	*Water Resources Research*	Canada (Quebec)	0.053	NASH = 0.82 RRMSE = 38	NA

Table 1. *Cont.*

Reference	Author, Year	Model	Predictor Variables (Inputs)	Model Output	Catchment, Year	Journal	Country (Catchment)	RMSE *	RRMSE/NASH *	R² *
[81]	Dawson et al., 2006	ANN * MLR	A, AMP, L, DA, IF	$Q_{10, 20, 30}$	850 stations, 20 years	*Journal of Hydrology*	United kingdom (Catchment across the UK)	NA	NA	NA
[80]	Jingyi and Hall, 2004	ANN * Cluster analysis	A, AMP, MXP, SL, SS, EL, GFI and PLN	Q_{50}	86 stations 15–36 years	*Journal of Hydrology*	China (Jiangxi and Fujian, Gan and Ming rivers)	47	NA	NA
[51]	(Shu and Burn, 2004)	ANN * Ordinary least squares regression (REG_OLS) Non-linear regression (REG_NONLINEAR)	A, AMP, SDT, FARL	Q_{10}	404 stations 29 years	*Water Resources Management*	United Kingdom (England, Scotland, and Wales)	NA	NA	NA

4. Comparative Assessment

Table 1 summarises AI-based RFFA methods, while Table 2 provides a summary of the symbols and abbreviations used in Table 1 for the adopted predictor variables. Despite the popularity of AI-based methods in other fields, such as rainfall-runoff modelling, flood forecasting, and global circulation phenomena, very few publications mention AI-based RFFA.

Table 2. Symbols and abbreviation of predictor variables used in Table 1.

Symbol/Abbreviation	Unit	Name of Variable
A	km^2	Catchment area
AA	km^2	Agricultural area of catchment
AEV	mm	Annual evaporation (mean)
AMD	degree-day	Annual mean degree-day above 0 °C
AMP	mm	Annual mean total precipitation
AMPL	mm	Mean liquid precipitation during Jul–Dec, Summer mean liquid precipitation, Spring Precipitation—the total basin precipitation from the start of active snowmelt to the start of the spring crest
AMPS	mm	Mean solid annual precipitation, Winter precipitation—the total basin precipitation from November 1st of previous year to the start of active snowmelt during the flood year, measured in inches;
BL	%	Percentage of barren land
CC	NA	Watershed compactness coefficient
CN	NA	Curve number
CR	NA	Watershed circulatory ratio
DA	km^2	Developed area
DD	NA	Drainage density
EL	m	Elevation
EP	mm	Equivalent precipitation at the time of flood event
FAL	%	Fraction of catchment area occupied by lake
FARL	NA	Reservoir/lake effects
FF	NA	Form factor
FLA	km^2	Fallow land area
FOR	%	Percentage of catchment covered by forest
GA	km^2	Garden area
GFI	NA	Geological feature index
HDB	NA	Hydrological database
I	mm/s	Design rainfall intensity
IF	NA	Index flood
L	km	Catchment length
LAT	NA	Latitude
LON	NA	Longitude
MAR	mm	Mean Annual Rainfall
MAE	mm	Mean annual evapo-transpiration
MBS	NA	Mean basin slope
MLS	mm	Mean level of snow on the 30th of March
MNEL	m	Minimum watershed elevation
MRD	day	Average annual rainfall duration
MXEL	m	Maximum watershed elevation
MXP	mm	Maximum 24 h rainfall
MXSO	NA	Maximum stream order
NDP	Number (NA)	Number of days of precipitation
NT27	Number (NA)	The average number of days with a temperature above 27 °C
P	km	Watershed perimeter
PF	km^2	Permeable formation area
PLN	NA	Plantation cover index
R24h	mm	24-h rainfall having a recurrence interval of 2 years
RA	km^2	Rock area
RLA	km^2	Rangeland area

Table 2. *Cont.*

Symbol/Abbreviation	Unit	Name of Variable
RP	Year	Return period
S1085	m/km	Slope of central 75% of the mainstream
SDEN	km^{-1}	Stream density
SDT	NA	Soil drainage type
SF	NA	Shape factor
SL	km	Mainstream length
SLP	Mbar	Sea-level pressure
SRC	NA	Soil runoff coefficient
SS	NA	Slope of the main channel in the drainage basin
SSC	NA	Soil storage coefficient
SST	°C	Sea surface temperature
T	°C	Mean annual temperature
TC	Hour	Time of concentration
TOPO	NA	Topographic digital maps
WA	%	Water area

5. Bibliometric Analysis

Figure 3 depicts the most commonly used AI-based RFFA methods from 2004 to 2021. Since 2004, the number of publications based on AI has gradually increased, with the highest number of publications in 2015. As shown in Figure 3, ANN has been the most popular AI-based RFFA method; however, the popularity of other methods such as SVM and ANFIS has increased in recent years, with SVM being the most popular model in 2020.
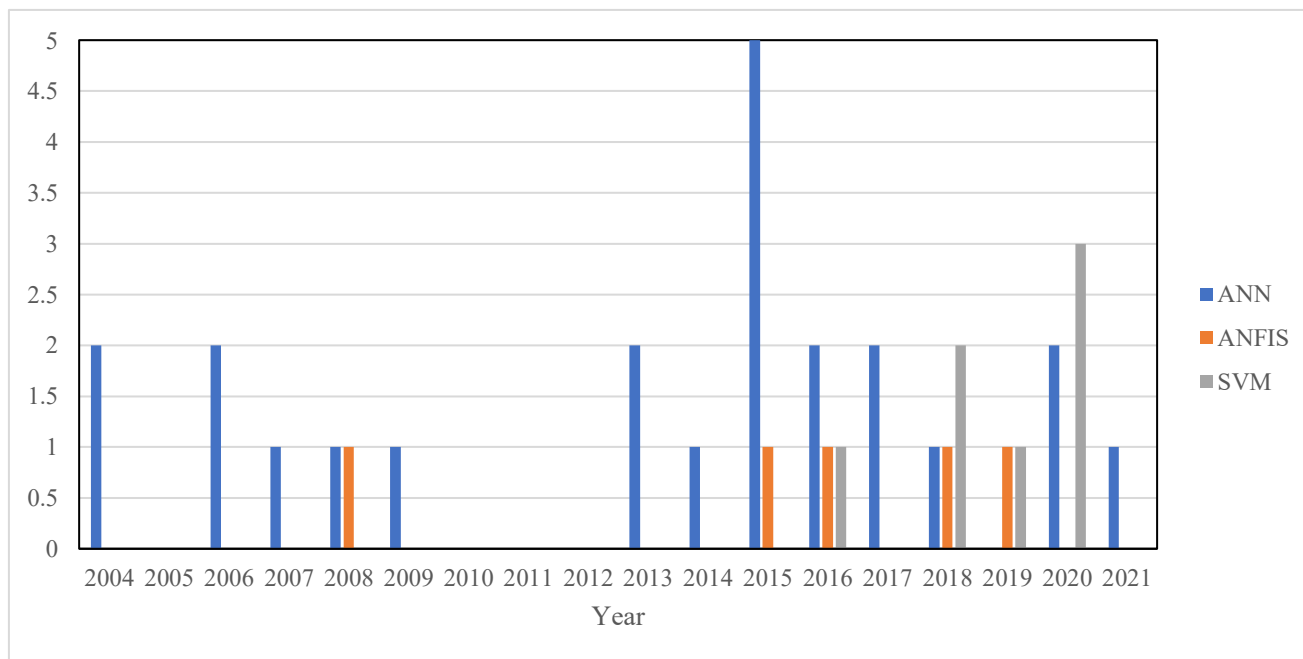
Figure 3. Usage frequency of AI-based RFFA methods during 2004–2021.

Figure 4 presents a bar graph of the most used predictor variables in estimating design floods. The A and AMP were the most used predictor variables in these studies, with 29 appearances for A and 24 for AMP, followed by other variables such as MBS, EL and SS.
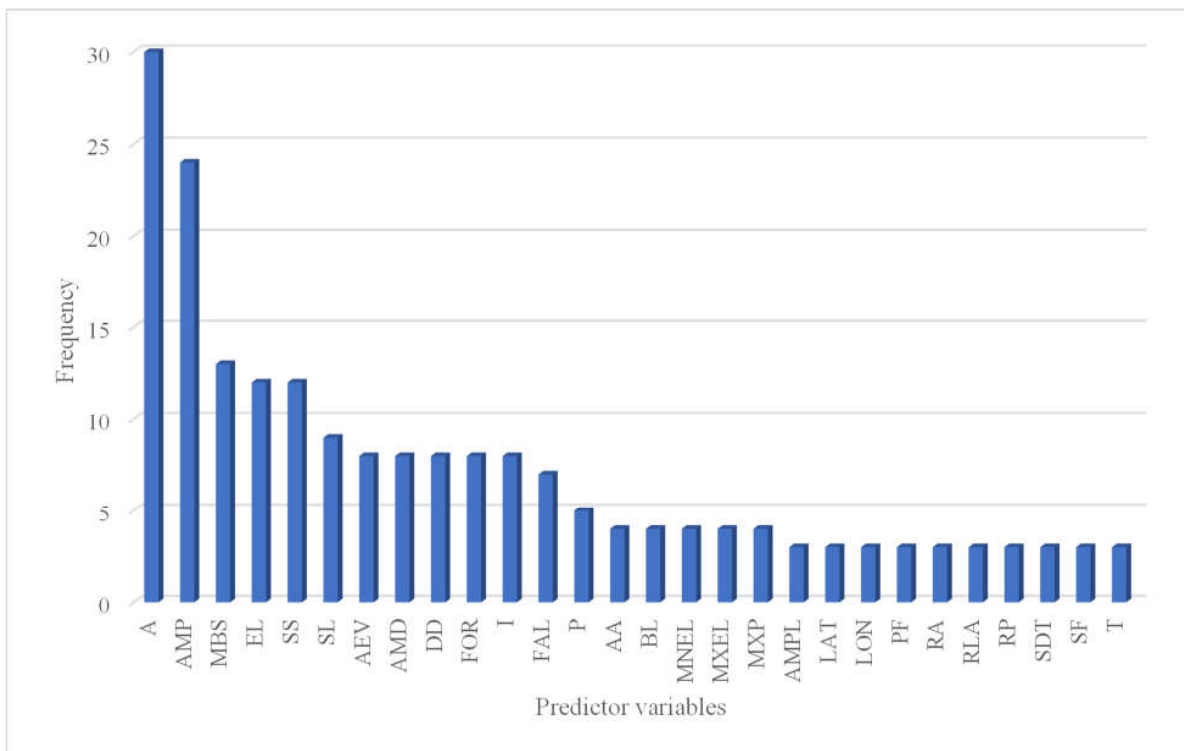
Figure 4. Predictor variables used in flood estimations.

Figure 5 shows map of authorship of the AI-based RFFA articles and the network of co-authorships. Here, the bigger the circle, the greater the number of papers published by the country. It can be seen that Australia, Canada, and Iran have published the highest number of articles in this research field, which is followed by Turkey, United Arab Emirates (UAE), India, and China. The third most important group consists of USA, the Netherlands, Germany, Malaysia, and Ethiopia. It can also be seen that Iran has the strongest network of co-authorship (linked with Germany, Malaysia, India, Ethiopia, China and Ireland), followed by India (linked with USA, Ethiopia, Iran, and China).
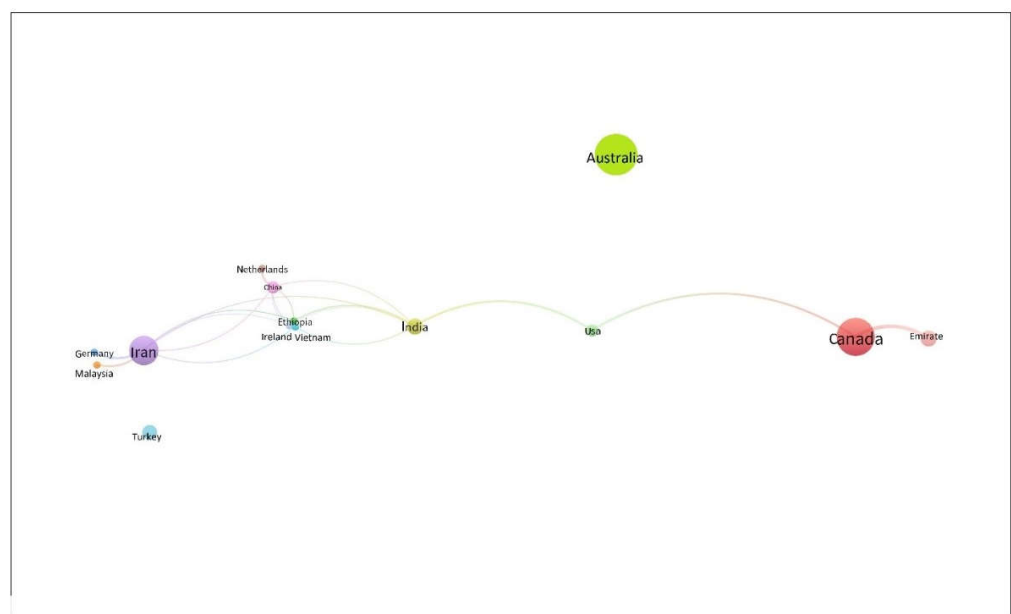
Figure 5. Network visualization map of country co-authorships on AI-based RFFA research field.

Figure 6 depicts some of the world's most active researchers working on flood research using AI-based models. Most of the authors are working in groups, and as it can be seen in this figure, Prof Vafakhah (Iran), Prof Taha Ouarda (Canada) and Prof Rahman (Australia) are the most active authors in this area. Figure 7 shows the authors' preferred destinations for publishing their studies on AI-based flood research. The most popular journals are *Water Resources Management and Journal of Hydrology*, followed by other journals such as *Water Resources Research* and *Natural Hazards*. These are prestigious journals in hydrology which indicates that AI-based RFFA methods have scientific merits to be included in these journals.
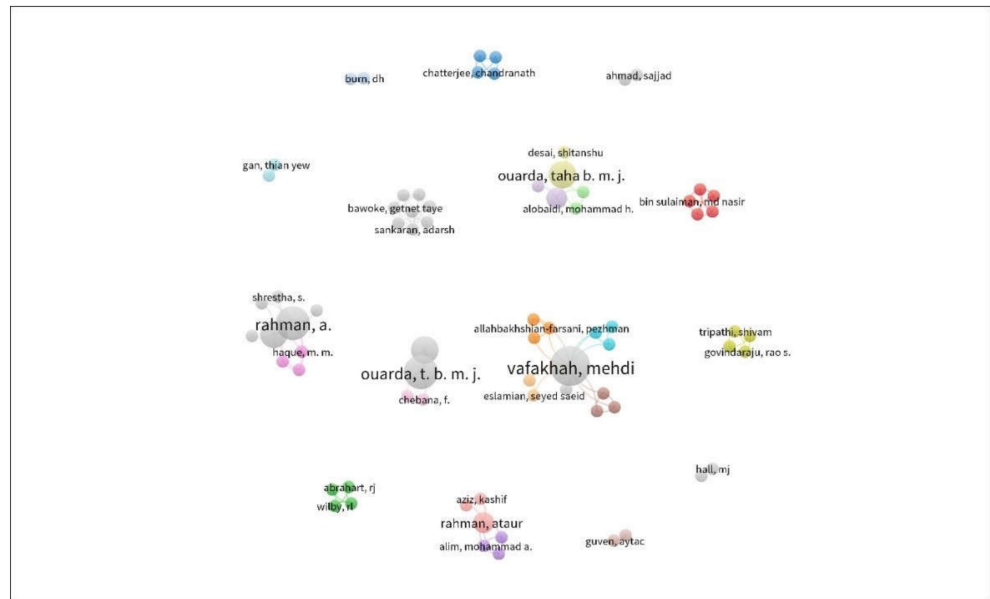

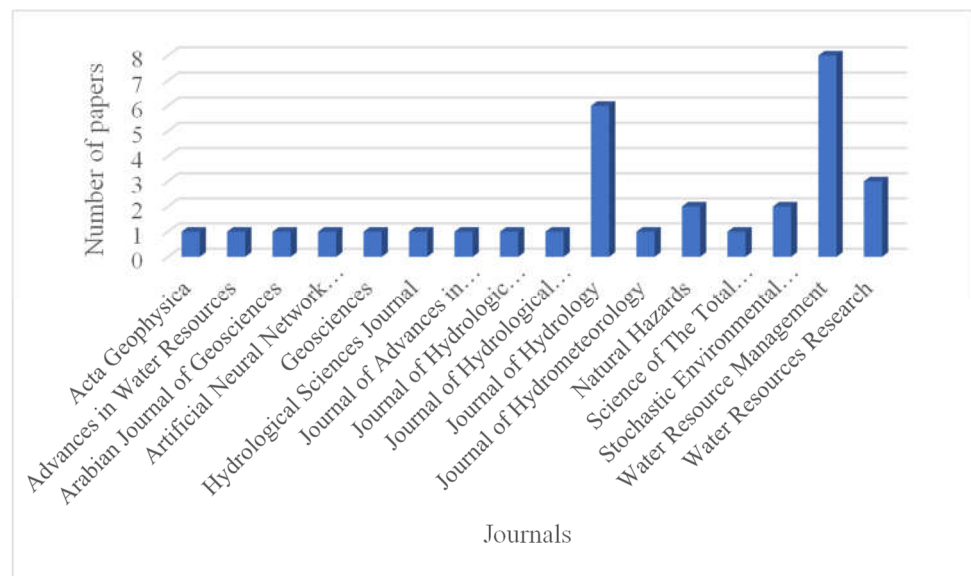**Figure 6.** Map showing clusters of authors.

Figure 7. Distribution of articles in different journals.

6. Challenges and Future Research Directions

In RFFA research, the main challenge is the lack of available data to train and test a model adequately. For example, in the Australian continent, about 800 stations are available having data with sufficient length and quality, which is too little to develop a reliable RFFA model covering an area of 7.69 km^2 (22 times bigger than Germany).

For most of the AI-based RFFA methods, quality data is very important, similar to other RFFA methods. The reported streamflow data often suffers from rating curve error, gaps and inconsistency, which must be considered in developing an AI-based RFFA model for practical applications. It is often better to use an integrated AI-based approach where several methods are combined and tested instead of a single approach.

For the interpretation of results of an AI-based RFFA technique, an expertise is needed as model assumptions must be well understood in relation to derived results before accepting results for design application. To tackle with the bias in the final results is often a challenging task, which depends on both the data quality and quantity and the optimization methods adopted to train and test the model. The sampling error in the dependent variable (e.g., the estimated flood quantile) is often ignored; however, it could be a significant source of uncertainty in RFFA in particular when streamflow data lengths of the selected catchments are too small (e.g., smaller than 20 years). The impact of climate change on streamflow data is becoming more prominent, which will make previously recorded streamflow data unsuitable. To incorporate non-stationarity in the AI-based RFFA methods is a challenge and needs further research. Another challenging aspect is the identification of the main sources of uncertainty, which stems from data reliability and model parameter estimation methods. A confidence interval estimation for the estimated flood quantiles is desirable with all the AI-based RFFA methods, which needs further research.

7. Conclusions

This study demonstrates that AI-based RFFA methods are becoming more appealing for estimating design floods than the linear models such as regression analysis. The majority of the articles chosen compare the accuracy and performance of various linear, non-linear, and AI-based RFFA methods. Because AI is a promising technique, most of these articles are published in prestigious journals. The most common AI-based methods for estimating design floods are ANN, ANFIS, and SVM, which have grown in popularity over the last decade. Hydrologists have experimented with various methods to improve the performance of the AI-based RFFA methods and the accuracy of estimation. Some of the common approaches used to improve the performance of AI-based methods include hybrid methods, data clustering, and the use of optimisation algorithms. According to evaluation statistics such as RMSE, NASH, RRMSE, and R^2, most AI-based RFFA methods outperform many traditional methods in terms of accuracy and reliability. They are typically data-driven models and require little to no understanding of the physics of flood generation processes, making them user-friendly. It has also been noticed that some AI-based models perform better in specific situations and locations, implying that it is preferable to test various AI-based RFFA methods to find the best possible model for the area of interest. Australia, Canada, and Iran have published the highest number of articles on AI based RFFA techniques, followed by Turkey, the United Arab Emirates (UAE), India, and China. Future research should concentrate on the effects of data quantity and quality, model uncertainty, and climate change, which includes rising temperatures, rising sea levels, and changes in rainfall intensity, on the overall results of AI-based RFFA techniques.

Author Contributions: A.Z. conducted data analysis and drafted the manuscript, A.R. assisted in interpreting results and revised the manuscript, N.N. assisted in data collection, referencing and revising the manuscript, K.H. assessed in revising the manuscript, M.M.R. assisted in revising the manuscript, M.V. assisted in interpreting the results, B.S. edited and improved the manuscript and F.A. edited and enhanced presentation. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare they have no conflict of interest.

References

1. Fernandez, A.; Black, J.; Jones, M.; Wilson, L.; Salvador-Carulla, L.; Astell-Burt, T.; Black, D. Flooding and mental health: A systematic mapping review. *PLoS ONE* **2015**, *10*, e0119929. [CrossRef] [PubMed]

2. Longman, J.; Bennett-Levy, J.; Matthews, V.; Berry, H.; Passey, M.; Rolfe, M.; Morgan, G.; Braddon, M.; Bailie, R. Rationale and methods for a cross-sectional study of mental health and wellbeing following river flooding in rural Australia, using a community-academic partnership approach. *BMC Public Health* **2019**, *19*, 1255. [CrossRef] [PubMed]

3. Mejia Manrique, S.A.; Harmsen, E.W.; Khanbilvardi, R.M.; González, J.E. Flood Impacts on Critical Infrastructure in a Coastal Floodplain in Western Puerto Rico during Hurricane María. *Hydrology* **2021**, *8*, 104. [CrossRef]

4. Ogie, R.I.; Holderness, T.; Dunn, S.; Turpin, E. Assessing the vulnerability of hydrological infrastructure to flood damage in coastal cities of developing nations. *Comput. Environ. Urban. Syst.* **2018**, *68*, 97–109. [CrossRef]

5. Ilinca, V. Using morphometrics to distinguish between debris flow, debris flood and flood (Southern Carpathians, Romania). *Catena* **2021**, *197*, 104982. [CrossRef]

6. Marcato, G.; Bossi, G.; Rivelli, F.; Borgatti, L. Debris flood hazard documentation and mitigation on the Tilcara alluvial fan (Quebrada de Humahuaca, Jujuy province, North-West Argentina). *Nat. Hazards Earth Syst. Sci.* **2012**, *12*, 1873–1882. [CrossRef]

7. Sung, C.-H.; Liaw, S.-C. A GIS-based approach for assessing social vulnerability to flood and debris flow hazards. *Int. J. Disaster Risk Reduct.* **2020**, *46*, 101531. [CrossRef]

8. Kellermann, P.; Schöbel, A.; Kundela, G.; Thieken, A.H. Estimating flood damage to railway infrastructure–The case study of the March River flood in 2006 at the Austrian Northern Railway. *Nat. Hazards Earth Syst. Sci.* **2015**, *15*, 2485–2496. [CrossRef]

9. Kellermann, P.; Schönberger, C.; Thieken, A.H. Large-scale application of the flood damage model RAilway Infrastructure Loss (RAIL). *Nat. Hazards Earth Syst. Sci.* **2016**, *16*, 2357–2371. [CrossRef]

10. Jato-Espino, D.; Sillanpää, N.; Pathak, S. Flood modelling in sewer networks using dependence measures and learning classifier systems. *J. Hydrol.* **2019**, *578*, 124013. [CrossRef]

11. Beg, M.N.A.; Rubinato, M.; Carvalho, R.F.; Shucksmith, J.D. CFD modelling of the transport of soluble pollutants from sewer networks to surface flows during urban flood events. *Water* **2020**, *12*, 2514. [CrossRef]

12. Rajkhowa, S.; Sarma, J. 14—Climate change and flood risk, global climate change. In *Global Climate Change*; Singh, S., Singh, P., Rangabhashiyam, S., Srivastava, K.K., Eds.; Elsevier: Amsterdam, The Netherlands, 2021; pp. 321–339.

13. Lyu, H.-M.; Shen, S.-L.; Zhou, A.; Yang, J. Perspectives for flood risk assessment and management for mega-city metro system. *Tunn. Undergr. Space Technol.* **2019**, *84*, 31–44. [CrossRef]

14. Kuhlicke, C.; Seebauer, S.; Hudson, P.; Begg, C.; Bubeck, P.; Dittmer, C.; Grothmann, T.; Heidenreich, A.; Kreibich, H.; Lorenz, D.F. The behavioral turn in flood risk management, its assumptions and potential implications. *Wiley Interdiscip. Rev. Water* **2020**, *7*, e1418. [CrossRef]

15. Didier, D.; Baudry, J.; Bernatchez, P.; Dumont, D.; Sadegh, M.; Bismuth, E.; Bandet, M.; Dugas, S.; Sévigny, C. Multihazard simulation for coastal flood mapping: Bathtub versus numerical modelling in an open estuary, Eastern Canada. *J. Flood Risk Manag.* **2019**, *12*, e12505. [CrossRef]

16. Vacondio, R.; Aureli, F.; Ferrari, A.; Mignosa, P.; Dal Palu, A. Simulation of the January 2014 flood on the Secchia River using a fast and high-resolution 2D parallel shallow-water numerical scheme. *Nat. Hazards* **2016**, *80*, 103–125. [CrossRef]

17. Costabile, P.; Costanzo, C.; De Lorenzo, G.; Macchione, F. Is local flood hazard assessment in urban areas significantly influenced by the physical complexity of the hydrodynamic inundation model? *J. Hydrol.* **2020**, *580*, 124231. [CrossRef]

18. Paprotny, D.; Morales-Nápoles, O.; Jonkman, S.N. Efficient pan-European river flood hazard modelling through a combination of statistical and physical models. *Nat. Hazards Earth Syst. Sci.* **2017**, *17*, 1267–1283. [CrossRef]

19. Van den Honert, R.C.; McAneney, J. The 2011 Brisbane floods: Causes, impacts and implications. *Water* **2011**, *3*, 1149–1173. [CrossRef]

20. Pappenberger, F.; Matgen, P.; Beven, K.J.; Henry, J.-B.; Pfister, L. Influence of uncertain boundary conditions and model structure on flood inundation predictions. *Adv. Water Resour.* **2006**, *29*, 1430–1449. [CrossRef]

21. Wu, Y.; Ding, Y.; Zhu, Y.; Feng, J.; Wang, S. Complexity to forecast flood: Problem definition and spatiotemporal attention LSTM solution. *Complexity* **2020**, *2020*, 7670382. [CrossRef]

22. Lavoie, B.; Mahdi, T.-F. Comparison of two-dimensional flood propagation models: SRH-2D and Hydro_AS-2D. *Nat. Hazards* **2017**, *86*, 1207–1222. [CrossRef]

23. Tsakiris, G.; Bellos, V. A numerical model for two-dimensional flood routing in complex terrains. *Water Resour. Manag.* **2014**, *28*, 1277–1291. [CrossRef]

24. Sahu, R.; Verma, M.; Ahmad, I. *Regional Frequency Analysis Using L-Moment Methodology—A Review*; Springer: Singapore, 2021; pp. 811–832. [CrossRef]

25. Rahman, A.; Haddad, K.; Zaman, M.; Kuczera, G.; Weinmann, P.E. Design Flood Estimation in Ungauged Catchments: A Comparison Between the Probabilistic Rational Method and Quantile Regression Technique for NSW. *Australas. J. Water Resour.* **2011**, *14*, 127–139. [CrossRef]

26. Darabi, H.; Haghighi, A.T.; Mohamadi, M.A.; Rashidpour, M.; Ziegler, A.D.; Hekmatzadeh, A.A.; Kløve, B. Urban flood risk mapping using data-driven geospatial techniques for a flood-prone case area in Iran. *Hydrol. Res.* **2020**, *51*, 127–142. [CrossRef]

27. Ahmadalipour, A.; Moradkhani, H. A data-driven analysis of flash flood hazard, fatalities, and damages over the CONUS during 1996–2017. *J. Hydrol.* **2019**, *578*, 124106. [CrossRef]

28. Villarini, G.; Smith, J.A.; Serinaldi, F.; Bales, J.; Bates, P.D.; Krajewski, W.F. Flood frequency analysis for nonstationary annual peak records in an urban drainage basin. *Adv. Water Resour.* **2009**, *32*, 1255–1266. [CrossRef]

29. Saghafian, B.; Golian, S.; Ghasemi, A. Flood frequency analysis based on simulated peak discharges. *Nat. Hazards* **2014**, *71*, 403–417. [CrossRef]

30. Yue, S. Applying Bivariate Normal Distribution to Flood Frequency Analysis. *Water Int.* **1999**, *24*, 248–254. [CrossRef]

31. Zhang, L.; Singh, V. Bivariate flood frequency analysis using the copula method. *J. Hydrol. Eng.* **2006**, *11*, 150–164. [CrossRef]

32. Haddad, K.; Rahman, A. Selection of the best fit flood frequency distribution and parameter estimation procedure: A case study for Tasmania in Australia. *Stoch. Environ. Res. Risk Assess.* **2011**, *25*, 415–428. [CrossRef]

33. Kousar, S.; Khan, A.R.; Ul Hassan, M.; Noreen, Z.; Bhatti, S.H. Some best-fit probability distributions for at-site flood frequency analysis of the Ume River. *J. Flood Risk Manag.* **2020**, *13*, e12640. [CrossRef]

34. Ozga-Zielinski, B.; Ciupak, M.; Adamowski, J.; Khalil, B.; Malard, J. Snow-melt flood frequency analysis by means of copula based 2D probability distributions for the Narew River in Poland. *J. Hydrol. Reg. Stud.* **2016**, *6*, 26–51. [CrossRef]

35. Ganamala, K.; Kumar, P.S. A case study on flood frequency analysis. *Int. J. Civil. Eng. Technol.* **2017**, *8*, 1762–1767.

36. Bhat, M.S.; Alam, A.; Ahmad, B.; Kotlia, B.S.; Farooq, H.; Taloor, A.K.; Ahmad, S. Flood frequency analysis of river Jhelum in Kashmir basin. *Quat. Int.* **2019**, *507*, 288–294. [CrossRef]

37. Drissia, T.K.; Jothiprakash, V.; Anitha, A.B. Flood Frequency Analysis Using L Moments: A Comparison between At-Site and Regional Approach. *Water Resour. Manag.* **2019**, *33*, 1013–1037. [CrossRef]

38. Młyński, D.; Wałęga, A.; Stachura, T.; Kaczor, G. A new empirical approach to calculating flood frequency in ungauged catchments: A case study of the upper Vistula basin, Poland. *Water* **2019**, *11*, 601. [CrossRef]

39. De Souza, G.R.; Merwade, V.; de Oliveira, L.F.C.; Viola, M.R.; de Sá Farias, M. Regional flood frequency analysis and uncertainties: Maximum streamflow estimates in ungauged basins in the region of Lavras, MG, Brazil. *Catena* **2021**, *197*, 104970. [CrossRef]

40. Lee, D.-H.; Kim, N.W. Regional flood frequency analysis for a poorly gauged basin using the simulated flood data and L-moment method. *Water* **2019**, *11*, 1717. [CrossRef]

41. Rahman, A.; Haddad, K.; Kuczera, G.; Weinmann, E. Regional flood methods. In *The Australian Rainfall and Runoff: A Guide Flood Estimation. Book 3 Peak Flow Estim*; Geoscience Australia: Canberra, Australia, 2019; pp. 105–146.

42. Esha, R.I.; Imteaz, M.A. Assessing the predictability of MLR models for long-term streamflow using lagged climate indices as predictors: A case study of NSW (Australia). *Hydrol. Res.* **2018**, *50*, 262–281. [CrossRef]

43. Tsakiri, K.; Marsellos, A.; Kapetanakis, S. Artificial neural network and multiple linear regression for flood prediction in Mohawk River, New York. *Water* **2018**, *10*, 1158. [CrossRef]

44. Rahman, A.S.; Rahman, A. Application of principal component analysis and cluster analysis in regional flood frequency analysis: A case study in New South Wales, Australia. *Water* **2020**, *12*, 781. [CrossRef]

45. Ahn, K.-H.; Palmer, R. Regional flood frequency analysis using spatial proximity and basin characteristics: Quantile regression vs. parameter regression technique. *J. Hydrol.* **2016**, *540*, 515–526. [CrossRef]

46. Formetta, G.; Over, T.; Stewart, E. Assessment of peak flow scaling and its effect on flood quantile estimation in the United Kingdom. *Water Resour. Res.* **2021**, *57*, e2020WR028076. [CrossRef]

47. Saf, B. Regional flood frequency analysis using L-moments for the West Mediterranean region of Turkey. *Water Resour. Manag.* **2009**, *23*, 531–551. [CrossRef]

48. Kiran, K.G.; Srinivas, V.V. Fuzzy Ensemble Clustering Approach to Address Regionalization Uncertainties in Flood Frequency Analysis. *Water Resour. Res.* **2021**, *57*, e2020WR028412. [CrossRef]

49. Srinivas, V.; Tripathi, S.; Rao, A.R.; Govindaraju, R.S. Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering. *J. Hydrol.* **2008**, *348*, 148–166. [CrossRef]

50. Rahman, A.; Charron, C.; Ouarda, T.B.M.J.; Chebana, F. Development of regional flood frequency analysis techniques using generalized additive models for Australia. *Stoch. Environ. Res. Risk Assess.* **2018**, *32*, 123–139. [CrossRef]

51. Shu, C.; Burn, D.H. Artificial neural network ensembles and their application in pooled flood frequency analysis. *Water Resour. Res.* **2004**, *40*, 1–10. [CrossRef]

52. Cassalho, F.; Beskow, S.; de Mello, C.R.; de Moura, M.M.; de Oliveira, L.F.; de Aguiar, M.S. Artificial intelligence for identifying hydrologically homogeneous regions: A state-of-the-art regional flood frequency analysis. *Hydrol. Process.* **2019**, *33*, 1101–1116. [CrossRef]

53. Kim, T.; Yang, T.; Gao, S.; Zhang, L.; Ding, Z.; Wen, X.; Gourley, J.J.; Hong, Y. Can artificial intelligence and data-driven machine learning models match or even replace process-driven hydrologic models for streamflow simulation?: A case study of four watersheds with different hydro-climatic regions across the CONUS. *J. Hydrol.* **2021**, *598*, 126423. [CrossRef]

54. Chen, W.; Li, Y.; Xue, W.; Shahabi, H.; Li, S.; Hong, H.; Wang, X.; Bian, H.; Zhang, S.; Pradhan, B.; et al. Modeling flood susceptibility using data-driven approaches of naïve Bayes tree, alternating decision tree, and random forest methods. *Sci. Total Environ.* **2020**, *701*, 134979. [CrossRef]

55. Shu, C.; Ouarda, T.B. Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system. *J. Hydrol.* **2008**, *349*, 31–43. [CrossRef]

56. Shu, C.; Ouarda, T.B. Flood frequency analysis at ungauged sites using artificial neural networks in canonical correlation analysis physiographic space. *Water Resour. Res.* **2007**, *43*, 1–12. [CrossRef]

57. Ul Hassan, M.; Noreen, Z.; Ahmed, R. Regional frequency analysis of annual daily rainfall maxima in Skåne, Sweden. *Int. J. Climatol.* **2021**, *41*, 4307–4320. [CrossRef]

58. Dhara, S.; Dang, T.; Parial, K.; Lu, X.X. Accounting for Uncertainty and Reconstruction of Flooding Patterns Based on Multi-Satellite Imagery and Support Vector Machine Technique: A Case Study of Can Tho City, Vietnam. *Water* **2020**, *12*, 1543. [CrossRef]

59. Allahbakhshian-Farsani, P.; Vafakhah, M.; Khosravi-Farsani, H.; Hertig, E. Regional Flood Frequency Analysis Through Some Machine Learning Models in Semi-arid Regions. *Water Resour. Manag.* **2020**, *34*, 2887–2909. [CrossRef]

60. Panahi, M.; Dodangeh, E.; Rezaie, F.; Khosravi, K.; Van Le, H.; Lee, M.-J.; Lee, S.; Thai Pham, B. Flood spatial prediction modeling using a hybrid of meta-optimization and support vector regression modeling. *Catena* **2021**, *199*, 105114. [CrossRef]

61. Kumar, R.; Goel, N.K.; Chatterjee, C.; Nayak, P.C. Regional Flood Frequency Analysis using Soft Computing Techniques. *Water Resour. Manag.* **2015**, *29*, 1965–1978. [CrossRef]

62. Zhou, Y.; Guo, S.; Chang, F.-J. Explore an evolutionary recurrent ANFIS for modelling multi-step-ahead flood forecasts. *J. Hydrol.* **2019**, *570*, 343–355. [CrossRef]

63. Kasiviswanathan, K.S.; He, J.; Tay, J.-H. Flood frequency analysis using multi-objective optimization based interval estimation approach. *J. Hydrol.* **2017**, *545*, 251–262. [CrossRef]

64. Tayfur, G.; Singh, V.P.; Moramarco, T.; Barbetta, S. Flood Hydrograph Prediction Using Machine Learning Methods. *Water* **2018**, *10*, 968. [CrossRef]

65. Haddad, K.; Rahman, A. Regional flood frequency analysis: Evaluation of regions in cluster space using support vector regression. *Nat. Hazards* **2020**, *102*, 489–517. [CrossRef]

66. Anaraki, M.V.; Farzin, S.; Mousavi, S.-F.; Karami, H. Uncertainty Analysis of Climate Change Impacts on Flood Frequency by Using Hybrid Machine Learning Methods. *Water Resour. Manag.* **2021**, *35*, 199–223. [CrossRef]

67. Aziz, K.; Haque, M.; Rahman, A.; Shamseldin, A.Y.; Shoaib, M. Flood estimation in ungauged catchments: Application of artificial intelligence based methods for Eastern Australia. *Stoch. Environ. Res. Risk Assess.* **2017**, *31*, 1499–1514. [CrossRef]

68. Khazaee Poul, A.; Shourian, M.; Ebrahimi, H. A Comparative Study of MLR, KNN, ANN and ANFIS Models with Wavelet Transform in Monthly Stream Flow Prediction. *Water Resour. Manag.* **2019**, *33*, 2907–2923. [CrossRef]

69. Fleming, S.W.; Bourdin, D.R.; Campbell, D.; Stull, R.B.; Gardner, T. Development and Operational Testing of a Super-Ensemble Artificial Intelligence Flood-Forecast Model for a Pacific Northwest River. *JAWRA J. Am. Water Resour. Assoc.* **2015**, *51*, 502–512. [CrossRef]

70. Bui, D.T.; Panahi, M.; Shahabi, H.; Singh, V.P.; Shirzadi, A.; Chapi, K.; Khosravi, K.; Chen, W.; Panahi, S.; Li, S.; et al. Novel Hybrid Evolutionary Algorithms for Spatial Prediction of Floods. *Sci. Rep.* **2018**, *8*, 15364. [CrossRef]

71. Jung, K.; Ouarda, T.B.M.J.; Marpu, P.R. On the Value of River Network Information in Regional Frequency Analysis. *J. Hydrometeorol.* **2021**, *22*, 201–216. [CrossRef]

72. Haddad, K.; Rahman, A.; Zaman, M.; Shrestha, S. Applicability of Monte Carlo cross validation technique for model development and validation using generalised least squares regression. *J. Hydrol.* **2013**, *482*, 119–128. [CrossRef]

73. Haddad, K.; Egodawatta, P.; Rahman, A.; Goonetilleke, A. Assessing uncertainty in pollutant wash-off modelling via model validation. *Sci. Total Environ.* **2014**, *497–498*, 578–584. [CrossRef]

74. Dawdy, D.; Webster, V.; Gupta, V. Regional Flood-Frequency Analysis: How We Got Here and Where We Are Going. *J. Hydrol. Eng.* **2012**, *17*, 953–959. [CrossRef]

75. Madsen, H.; Lawrence, D.; Lang, M.; Martinkova, M.; Kjeldsen, T. *A Review of Applied Methods in Europe for Flood-Frequency Analysis in a Changing Environment*; NERC: Atlanta, GA, USA, 2013.

76. Khairudin, N.M.; Mustapha, N.; Aris, T.N.M.; Zolkepli, M. In-Depth review on machine learning models for long-term flood forecasting. *J. Theor. Appl. Inf. Technol.* **2022**, *100*, 1–19.

77. Diaconu, D.C.; Costache, R.; Popa, M.C. An Overview of Flood Risk Analysis Methods. *Water* **2021**, *13*, 474. [CrossRef]

78. Lohani, A.K.; Kumar, R.; Singh, R.D. Hydrological time series modeling: A comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques. *J. Hydrol.* **2012**, *442–443*, 23–35. [CrossRef]

79. Tanty, R.M.; Desmukh, T.S. Application of Artificial Neural Network in Hydrology—A Review. *Int. J. Eng. Res. Technol.* **2015**, *4*, 184–188.

80. Jingyi, Z.; Hall, M. Regional flood frequency analysis for the Gan-Ming River basin in China. *J. Hydrol.* **2004**, *296*, 98–117. [CrossRef]

81. Dawson, C.W.; Abrahart, R.J.; Shamseldin, A.Y.; Wilby, R.L. Flood estimation at ungauged sites using artificial neural networks. *J. Hydrol.* **2006**, *319*, 391–409. [CrossRef]

82. Ouarda, T.B.; Shu, C. Regional low-flow frequency analysis using single and ensemble artificial neural networks. *Water Resour. Res.* **2009**, *45*, 1–16. [CrossRef]

83. Singh, K.K.; Pal, M.; Singh, V. Estimation of mean annual flood in Indian catchments using backpropagation neural network and M5 model tree. *Water Resour. Manag.* **2010**, *24*, 2007–2019. [CrossRef]

84. Seckin, N.; Cobaner, M.; Yurtal, R.; Haktanir, T. Comparison of artificial neural network methods with L-moments for estimating flood flow at ungauged sites: The case of East Mediterranean River Basin, Turkey. *Water Resour. Manag.* **2013**, *27*, 2103–2124. [CrossRef]

85. 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.* **2014**, *28*, 541–554. [CrossRef]

86. Alobaidi, M.H.; Marpu, P.R.; Ouarda, T.B.; Chebana, F. Regional frequency analysis at ungauged sites using a two-stage resampling generalized ensemble framework. *Adv. Water Resour.* **2015**, *84*, 103–111. [CrossRef]

87. Durocher, M.; Chebana, F.; Ouarda, T.B. A nonlinear approach to regional flood frequency analysis using projection pursuit regression. *J. Hydrometeorol.* **2015**, *16*, 1561–1574. [CrossRef]

88. Chokmani, K.; Ouarda, T.B. Physiographical space-based kriging for regional flood frequency estimation at ungauged sites. *Water Resour. Res.* **2004**, *40*, 1–13. [CrossRef]

89. Wazneh, H.; Chebana, F.; Ouarda, T.B.M.J. Optimal depth-based regional frequency analysis. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 2281–2296. [CrossRef]

90. Chebana, F.; Charron, C.; Ouarda, T.B.; Martel, B. Regional frequency analysis at ungauged sites with the generalized additive model. *J. Hydrometeorol.* **2014**, *15*, 2418–2428. [CrossRef]

91. Nezhad, M.K.; Chokmani, K.; Ouarda, T.B.; Barbet, M.; Bruneau, P. Regional flood frequency analysis using residual kriging in physiographical space. *Hydrol. Process.* **2010**, *24*, 2045–2055. [CrossRef]

92. Ouali, D.; Chebana, F.; Ouarda, T.B. Fully nonlinear statistical and machine-learning approaches for hydrological frequency estimation at ungauged sites. *J. Adv. Model. Earth Syst.* **2017**, *9*, 1292–1306. [CrossRef]

93. Ouarda, T.B.; Girard, C.; Cavadias, G.S.; Bobée, B. Regional flood frequency estimation with canonical correlation analysis. *J. Hydrol.* **2001**, *254*, 157–173. [CrossRef]

94. Ouali, D.; Chebana, F.; Ouarda, T.B. Non-linear canonical correlation analysis in regional frequency analysis. *Stoch. Environ. Res. Risk Assess.* **2016**, *30*, 449–462. [CrossRef]

95. Kordrostami, S.; Alim, M.A.; Karim, F.; Rahman, A. Regional flood frequency analysis using an artificial neural network model. *Geosciences* **2020**, *10*, 127. [CrossRef]

96. Linh, N.T.T.; Ruigar, H.; Golian, S.; Bawoke, G.T.; Gupta, V.; Rahman, K.U.; Sankaran, A.; Pham, Q.B. Flood prediction based on climatic signals using wavelet neural network. *Acta Geophys.* **2021**, *69*, 1413–1426. [CrossRef]

97. Desai, S.; Ouarda, T.B. Regional hydrological frequency analysis at ungauged sites with random forest regression. *J. Hydrol.* **2021**, *594*, 125861. [CrossRef]

98. Zadeh, L.A. Fuzzy sets. In *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi A Zadeh*; World Scientific: Singapore, 1996; pp. 394–432.

99. Takagi, T.; Sugeno, M. Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. Syst. Man. Cybern.* **1985**, *SMC-15*, 116–132. [CrossRef]

100. Mamdani, E.H.; Assilian, S. An experiment in linguistic synthesis with a fuzzy logic controller. *Int. J. Man-Mach. Stud.* **1975**, *7*, 1–13. [CrossRef]

101. Jang, J.S.R. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Trans. Syst. Man Cybern.* **1993**, *23*, 665–685. [CrossRef]

102. Zalnezhad, A.; Rahman, A.; Vafakhah, M.; Samali, B.; Ahamed, F. Regional Flood Frequency Analysis Using the FCM-ANFIS Algorithm: A Case Study in South-Eastern Australia. *Water* **2022**, *14*, 1608. [CrossRef]

103. Aziz, K.; Rahman, A.; Shamseldin, A.; Shoaib, M. Co-active neuro fuzzy inference system for regional flood estimation in Australia. *J. Hydrol. Environ. Res.* **2013**, *1*, 11–20.

104. 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. [CrossRef]

105. Bozchaloei, S.K.; Vafakhah, M. Regional analysis of flow duration curves using adaptive neuro-fuzzy inference system. *J. Hydrol. Eng.* **2015**, *20*, 06015008. [CrossRef]

106. Aziz, K.; Rahman, A.; Shamseldin, A. Development of artificial intelligence based regional flood estimation techniques for Eastern Australia. In *Artificial Neural Network Modelling*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 307–323.

107. Wu, J.; Liu, H.; Wei, G.; Song, T.; Zhang, C.; Zhou, H. Flash flood forecasting using support vector regression model in a small mountainous catchment. *Water* **2019**, *11*, 1327. [CrossRef]

108. Dodangeh, E.; Panahi, M.; Rezaie, F.; Lee, S.; Tien Bui, D.; Lee, C.-W.; Pradhan, B. Novel hybrid intelligence models for flood-susceptibility prediction: Meta optimization of the GMDH and SVR models with the genetic algorithm and harmony search. *J. Hydrol.* **2020**, *590*, 125423. [CrossRef]

109. Gizaw, M.S.; Gan, T.Y. Regional flood frequency analysis using support vector regression under historical and future climate. *J. Hydrol.* **2016**, *538*, 387–398. [CrossRef]

110. Sharifi Garmdareh, E.; Vafakhah, M.; Eslamian, S.S. Regional flood frequency analysis using support vector regression in arid and semi-arid regions of Iran. *Hydrol. Sci. J.* **2018**, *63*, 426–440. [CrossRef]

111. Ghaderi, K.; Motamedvaziri, B.; Vafakhah, M.; Dehghani, A.A. Regional flood frequency modeling: A comparative study among several data-driven models. *Arab. J. Geosci.* **2019**, *12*, 588. [CrossRef]

112. Vafakhah, M.; Khosrobeigi Bozchaloei, S. Regional analysis of flow duration curves through support vector regression. *Water Resour. Manag.* **2020**, *34*, 283–294. [CrossRef]

113. Seckin, N.; Guven, A. Estimation of peak flood discharges at ungauged sites across Turkey. *Water Resour. Manag.* **2012**, *26*, 2569–2581. [CrossRef]

114. 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. [CrossRef]