Improving streamflow forecast using optimal rain gauge network-based input to artificial neural network models

Sajal Kumar Adhikary, Nitin Muttil and Abdullah Gokhan Yilmaz

ABSTRACT

Accurate streamflow forecasting is of great importance for the effective management of water resources systems. In this study, an improved streamflow forecasting approach using the optimal rain gauge network-based input to artificial neural network (ANN) models is proposed and demonstrated through a case study (the Middle Yarra River catchment in Victoria, Australia). First, the optimal rain gauge network is established based on the current rain gauge network in the catchment. Rainfall data from the optimal and current rain gauge networks together with streamflow observations are used as the input to train the ANN. Then, the best subset of significant input variables relating to streamflow at the catchment outlet is identified by the trained ANN. Finally, one-day-ahead streamflow forecasting is carried out using ANN models formulated based on the selected input variables for each rain gauge network. The results indicate that the optimal rain gauge network-based input to ANN models gives the best streamflow forecasting results for the training, validation and testing phases in terms of various performance evaluation measures. Overall, the study concludes that the proposed approach is highly effective to achieve the enhanced streamflow forecasting and could be a viable option for streamflow forecasting in other catchments.

Key words | artificial neural network, input selection, optimal rain gauge network, streamflow forecasting, Yarra River catchment

Sajal Kumar Adhikary (corresponding author)
Nitin Muttil
College of Engineering and Science,
Victoria University,
Footscray Park Campus, P.O. Box 14428,
Melbourne, Victoria 8001,
Australia
and
Institute for Sustainability and Innovation,
Victoria University,
P.O. Box 14428, Melbourne, Victoria 8001,
Australia
E-mail: sajal.adhikary@live.vu.edu.au

Abdullah Gokhan Yilmaz
College of Engineering,
University of Sharjah,
P.O. Box 27272, Sharjah,
United Arab Emirates

INTRODUCTION

Streamflow is one of the key variables in hydrology. Accurate forecasting of streamflow is essential for many of the activities associated with the efficient planning and operation of the components of risk-based water resources systems. In particular, flood control and operational river management systems highly depend on accurate and reliable forecasting of streamflow. The analysis and design of dams and bridges, management of extreme events including floods and droughts, optimal operation of reservoirs encompassing irrigation, hydropower generation, domestic and industry water supply objectives are a few examples where information regarding short-term and long-term streamflow forecasting is vital (Londhe & Charhate 2010). Hence, there is a growing need to improve the short-term and long-term streamflow forecasting for

the efficient optimization of water resources systems (Akhtar et al. 2009).

The approaches used for streamflow forecasting cover a wide range of methods from completely black box (data-driven or machine learning) models to detailed conceptual or physically based models (Porporato & Ridolfi 2001). The conceptual or physically based models usually require extensive data and huge computational efforts, and are influenced by the effects of overparameterization and parameter redundancy (Linares-Rodriguez et al. 2015). Furthermore, such models could not be applied to a slightly different system. As a result of these limitations, data-driven methods have been increasingly preferred for hydrological modelling and forecasting (Khu et al. 2001; Yilmaz & Muttil 2014). In particular, a data-driven method that has gained significant

attention of researchers in recent years is the artificial neural network (ANN)-based streamflow forecasting technique (e.g., Zealand *et al.* 1999; Dibike & Solomatine 2001; Birikundavyi *et al.* 2002; Huang *et al.* 2004; Kumar *et al.* 2004; Wu *et al.* 2005; Kişi 2007; Srinivasulu & Jain 2009; Londhe & Charhate 2010; Abrahart *et al.* 2012; Sivapragasam *et al.* 2014; Linares-Rodriguez *et al.* 2015; Taormina *et al.* 2015).

The majority of the aforementioned studies have confirmed that ANN is able to outperform traditional statistical methods. ANN is perhaps the most popular machine learning method with flexible mathematical structure, which is capable of identifying a direct mapping between inputs and outputs without detailed consideration of the internal structure of the physical process (Maier & Dandy 2000; Dibike & Solomatine 2001). ANN models are computationally fast and reliable, and yield results comparable to conceptual models. These models can extract the complex nonlinear relationships between the inputs and outputs of a process without the physics being explicitly provided. Furthermore, ANN models for streamflow forecasting require only a limited number of input variables, such as rainfall and flow data (e.g., Londhe & Charhate 2010; Talei *et al.* 2010; Yilmaz & Muttil 2014), which makes them suitable for forecasting applications in practice. For a detailed description of ANNs with their modelling processes and applications in hydrology and water resources, readers are referred to Govindaraju & Rao (2000), ASCE Task Committee (2000a, 2000b), Dawson & Wilby (2001), Maier *et al.* (2010) and Tayfur (2012).

This study mainly focuses on the important hydrological aspects of rainfall input for streamflow simulation within the framework of ANN-based streamflow forecasting models. Rainfall is one of the most important inputs in the development of ANN models for streamflow forecasting. Since streamflow is a consequence of rainfall, using accurate rainfall input to ANN models is vital in order to achieve enhanced streamflow forecasting. However, many of the water resources systems are large in spatial extent and often consist of a rain gauge network that is very sparse due to economic, geological and logistic factors. This may cause inaccuracy in the collected rainfall information (Zealand *et al.* 1999). Therefore, it is necessary to establish an optimal rain gauge network, which can give high quality rainfall estimates for accurate streamflow forecasting. An optimal rain gauge network refers to a balanced network that never suffers from station shortages, or from over-saturations caused by redundant stations (Mishra & Coulibaly 2009; Adhikary *et al.* 2015). If rainfall information can be more accurately estimated through the optimal network and used in ANN-based streamflow forecasting models, it is likely that enhanced streamflow forecasting can be achieved, a conclusion supported by the works of Andréassian *et al.* (2001), who tested the sensitivity of watershed models to the imperfect knowledge of rainfall input.

Rainfall is often considered independent of streamflow forecasting in many hydrological studies such as average areal rainfall estimation over a catchment (e.g., Bras & Rodriguez-Iturbe 1976; Bastin *et al.* 1984; Seed & Austin 1990; Adhikary *et al.* 2016a, 2017) or the design of rain gauge networks (e.g., Papamichail & Metaxa 1996; Pardo-Igúzquiza 1998; Tsintikidis *et al.* 2002; Chen *et al.* 2008; Cheng *et al.* 2008; Adhikary *et al.* 2015; Feki *et al.* 2017). However, this does not allow one to focus on the strengths and weaknesses of an established network that really matter when rainfall data are fed into a streamflow forecasting model. Furthermore, Bras (1979) and Storm *et al.* (1989) emphasized that watersheds act as low-pass filters, attenuating the rainfall variability. It is thus necessary to take this filter into account to determine the quality and quantity of rainfall data required to achieve a certain degree of accuracy in streamflow forecasting. Hence, it is logical to design a rain gauge network for providing a satisfactory solution to the specific needs (enhanced streamflow forecasting in the current study) for which the network is being established. Based on the aforementioned considerations, it is thus hypothesized that use of the optimal rain gauge network-based input to streamflow forecasting models can contribute to the improved streamflow forecasting.

To date, many studies have been devoted to the impact of rainfall input, varying rain gauge network density and distribution on the performance of streamflow forecasting (e.g., Faurès *et al.* 1995; St-Hilaire *et al.* 2003; Dong *et al.* 2005; Anctil *et al.* 2006; Xu *et al.* 2006, 2013; Bárdossy & Das 2008; Ekström & Jones 2009; Moulin *et al.* 2009; Volkmann *et al.* 2010; Tsai *et al.* 2014; Linares-Rodriguez *et al.* 2015). However, none of these studies used rainfall input from the optimal rain gauge network for streamflow forecasting. Therefore, the objective of this study is to use rainfall information from an optimally designed rain gauge network in

combination with streamflow observations as the input to ANN-based streamflow forecasting models for enhanced streamflow forecasting. The specific focus is to evaluate the effectiveness of integrating an optimal rain gauge network within the framework of ANN models to achieve the improved streamflow forecasting. The experimental approach is planned in two phases and demonstrated through an application to the Middle Yarra River catchment in Victoria, Australia. First, the optimal rain gauge network is established from the current operational rain gauge network in the catchment by using the well-known kriging-based geostatistical technique (presented in Adhikary *et al.* 2015). Next, streamflow forecasting is undertaken one day in advance at the catchment outlet based on the selected significant input variables (rainfall and streamflow) for each of the current and optimal rain gauge networks. Such an approach could be scalable to other catchments contingent upon addressing the local contextual issues, which is expected to be a viable option to achieve the enhanced streamflow forecasting.

The remainder of the paper is structured as follows. First, the study area and dataset used are described in detail. This is followed by the detailed description of the methodology adopted in this study. The results are summarized next and, finally, the conclusions drawn from the study are presented.

STUDY AREA AND DATASET USED

The study area

In the current study, the middle segment of the Yarra River catchment (referred to as the Middle Yarra River catchment) located in Victoria, Australia, is selected as the case study area. Approximate location of the catchment is shown in Figure 1. The catchment is located northeast of Melbourne, and covers an area of 4,044 km^2. The catchment is home to more than one-third of Victoria's population (approximately 1.8 million). Although the Yarra River
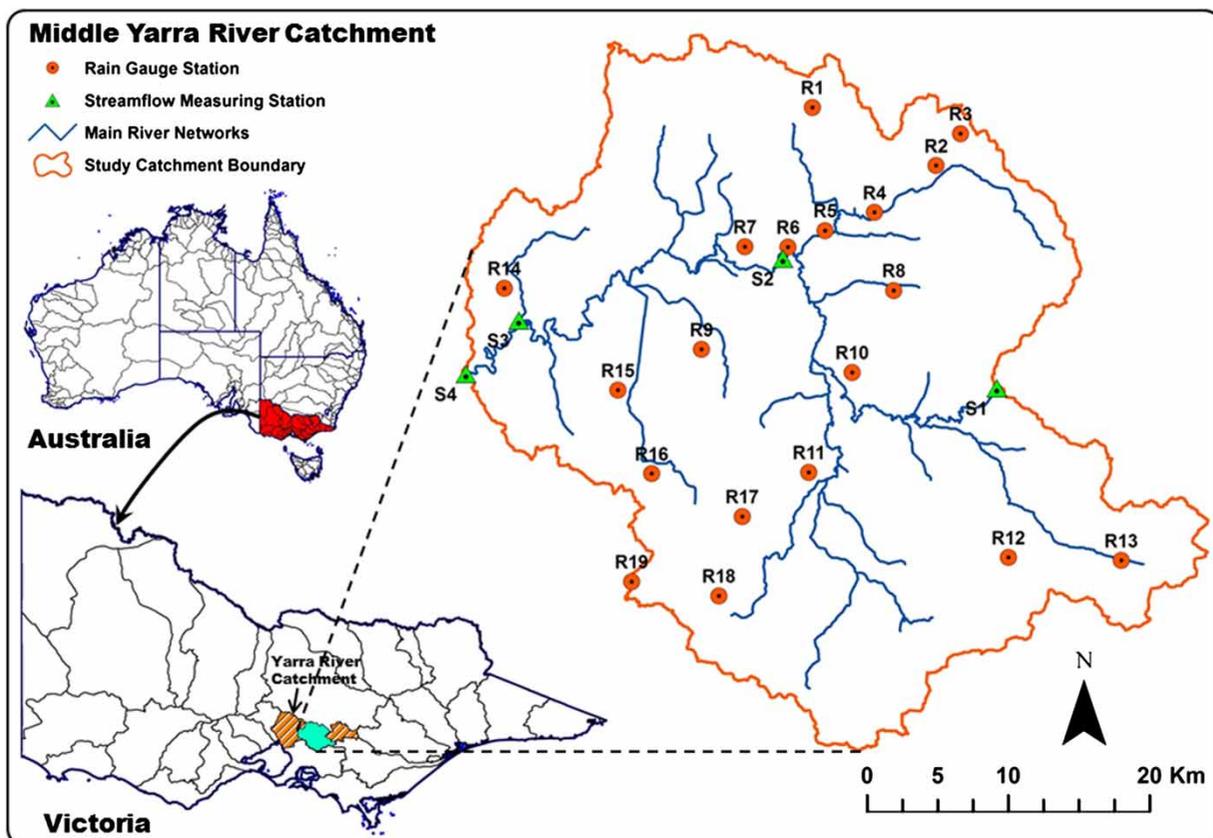

Figure 1 │ Location of the study area (Middle Yarra River catchment) with hydrometric stations.

catchment is not large with respect to other Australian catchments, it produces the fourth highest water yield per hectare of the catchment in Victoria, which makes it a very productive catchment. The Yarra River thus plays a key role in the way Melbourne has developed and grown (Adhikary *et al.* 2016b).

The Yarra River catchment is divided into three distinctive sub-catchments (as shown in Figure 1), namely Upper Yarra, Middle Yarra and Lower Yarra segments based on the different land use patterns. The Upper Yarra segment of the catchment consists of mainly forested and mountainous areas with minimum human settlement. Approximately 70% of Melbourne's drinking water supply comes from this pristine upper segment (Barua *et al.* 2012). The Middle Yarra segment is distinguished as the only part of the catchment with an extensive flood plain, which is mainly used for agricultural activities. The Lower Yarra segment is mainly characterized by the urbanized floodplain areas of Melbourne city. The average annual rainfall varies across the Yarra River catchment from about 1,100 mm in the Upper Yarra segment to 600 mm in the lower Yarra segment (Daly *et al.* 2013). Hence, water resources management in the catchment is of great importance considering the diverse water use activities and high variability in rainfall.

The Middle Yarra segment (the case study area as shown in Figure 1) covers an area of 1,511 km². There are three storage reservoirs, namely, Maroondah, Silvan and Sugarloaf, in the study area that support water supply for a range of activities including urban and agricultural activities. The main aim of the reservoir operation in Australia is to store as much water as possible to meet water demands during droughts while keeping provision for flood control during floods. Lower rainfall causes reduction in streamflows, which obviously results in the shortage of reservoir inflows and affects the overall water availability. In addition, reduction in streamflows may cause increased risk of bushfires. On the other hand, the occurrence of higher or extreme rainfall results in an excess amount of streamflows that may cause flash floods in the urbanized lower segment of the catchment and make it vulnerable and risk-prone. The urbanized lower segment also depends on the water supply from the storage reservoirs mainly located in the middle and upper segments of the catchment (Adhikary *et al.* 2015). Therefore, accurate streamflow forecasting is of great

significance for optimal operation of storage reservoirs, and planning for effective flood control and mitigation measures, particularly in the urbanized lower segment of Yarra River catchment.

Dataset used

Available literature suggests that many different variables are used as input to ANN models. Rainfall and antecedent streamflow are the most frequently used inputs for ANN-based streamflow forecasting models. The antecedent streamflow acts indirectly as a descriptor of the moisture state in the watershed (Anctil *et al.* 2004). The input also consists of air temperature or potential evapotranspiration in combination with rainfall information. However, some studies have shown that model results are nearly insensitive to the potential evapotranspiration or temperature and thus their usage as input are unnecessary (e.g., Oudin *et al.* 2005, 2006; Xu *et al.* 2006). Therefore, rainfall data together with streamflow observations are used as the necessary input to develop ANN-based streamflow forecasting models in the current study.

In the current study, the dataset is based on the historical rainfall records from the rain gauge network of the Australian Bureau of Meteorology (BoM) and streamflow observations from the streamflow measuring network of Melbourne Water Corporation (MWC). Spatial location of the hydrometric stations within the study area is shown in Figure 1. There are 19 rain gauge stations (indicated by R1 to R19) in the BoM's current network and four streamflow measuring stations (indicated by S1 to S4) along the main course of the Yarra River in the study area. Table 1 presents the particulars of the hydrometric (rain gauge and streamflow) stations. Thirty years of daily meteorological and hydrological data (from 1980 to 2009) including rainfall and streamflow are used in this study. The choice of this study period is based on the availability of high quality data with no missing records for an extended period. Daily rainfall data of all 19 rain gauge stations were collected from the Scientific Information for Land Owners (SILO, http://www.longpaddock.qld.gov.au/silo/) climate database. The SILO database has been selected for this study because SILO data are quality controlled and completely free from missing records. The missing records in this database are filled up during a quality control process

Table 1 | Summary of the hydrometric stations in the Middle Yarra River catchment

Station no.[a]	Station details		Easting (m)	Northing (m)
	Site ID	Name of station		
Rain gauge stations				
R1	86142	Toolangi (Mount St Leonard Department of Primary Industries)	367,665	5,840,620
R2	86366	Fernshaw	376,433	5,836,534
R3	86009	Black Spur	378,165	5,838,779
R4	86070	Maroondah Weir	372,048	5,833,250
R5	86385	Healesville (Mount Yule)	368,559	5,831,973
R6	86363	Tarrawarra	365,931	5,830,821
R7	86364	Tarrawarra Monastery	362,905	5,830,845
R8	86219	Coranderrk Badger Weir	373,425	5,827,770
R9	86383	Coldstream	359,825	5,823,625
R10	86229	Healesville (Valley View Farm)	370,480	5,822,015
R11	86367	Seville	367,398	5,815,000
R12	86358	Gladysdale (Little Feet Farm)	381,535	5,809,020
R13	86094	Powelltown Department of Natural Resources and Environment	389,545	5,808,810
R14	86059	Kangaroo Ground	345,855	5,827,920
R15	86066	Lilydale	353,900	5,820,765
R16	86076	Montrose	356,285	5,814,905
R17	86106	Silvan	362,717	5,811,901
R18	86072	Monbulk (Spring Road)	361,051	5,806,323
R19	86266	Ferny Creek	354,874	5,807,326
Streamflow measuring stations				
S1	229212	Yarra River at Millgrove	380,730	5,820,906
S2	229653	Yarra River at Yarra Grange	365,590	5,830,000
S3	229608	Watsons Creek at Kangaroo Ground South	346,900	5,825,660
S4	229200	Yarra River at Warrandyte	343,157	5,821,896

[a]Station nos. are the same as in Figure 1.

based on the ordinary kriging and thin plate spline interpolation techniques using available records in the nearby surrounding stations. The SILO (Scientific Information for Land Owners) database gives an additional benefit of data

drill opportunity using the aforementioned interpolation techniques by which one can obtain the necessary rainfall data at any ungauged location in the catchment (Jeffrey *et al.* 2001). Streamflow data of all four streamflow measuring stations were collected from the MWC database. The average annual rainfall in the study area during the 1980–2009 period varies from 710 mm to 1,422 mm with a mean rainfall of 1,063 mm. Approximately 60% of the mean rainfall occurs in the winter (June–August) and spring (September–November) seasons, which contributes mostly to streamflow.

METHODOLOGY

This study presents an approach of streamflow forecasting in an attempt to achieve the enhanced streamflow forecasting using the optimal rain gauge network-based input to ANN models. The methodological framework of the proposed approach is shown in Figure 2, which is demonstrated through an application to the Middle Yarra River catchment in Victoria, Australia. As can be seen from the figure, the framework has two parts and in the first part of the framework, an optimal and an augmented rain gauge network are established from the BoM's current operational rain gauge network. The second part consists of streamflow forecasting, which focuses on the impact of optimal rain gauge network-based input on the performance of streamflow forecasting. In general, the framework is implemented through the following four steps: (i) optimal rain gauge network design, (ii) augmentation of the optimal rain gauge network, (iii) ANN-based input variable selection and (iv) streamflow forecasting and assessment. These steps are described in the following subsections.

Optimal rain gauge network design

An optimal network should essentially consist of sufficient number of rain gauge stations with suitable locations in such a way that the network can provide optimum rainfall information with minimum uncertainty and cost. Adequate station density as well as location in the network equally plays a vital role in determining whether the rain gauge network is optimal and sufficient information is gained
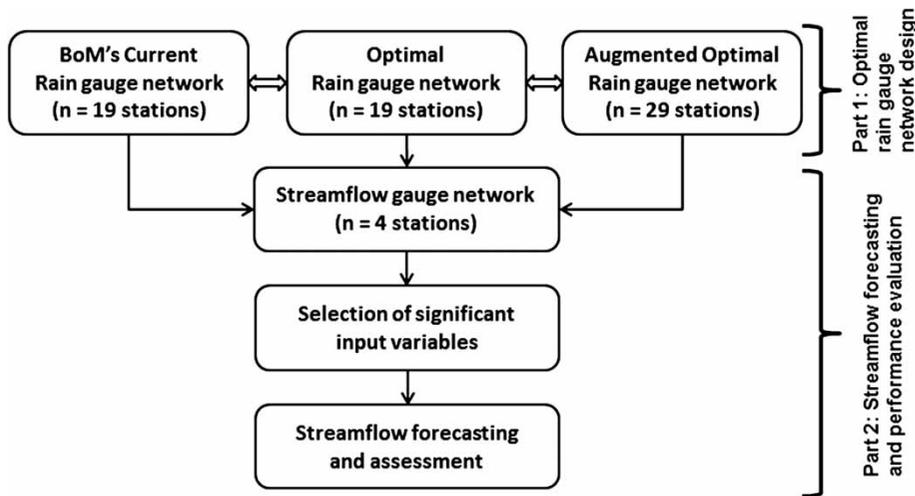
Figure 2 | Framework of methodology adopted in this study.

(Adhikary *et al.* 2015). Thus, the optimal network is achieved through optimal positioning of additional stations (i.e., network extension) together with redundant stations or simply removing redundant stations (i.e., network rationalization) (St-Hilaire *et al.* 2003; Mishra & Coulibaly 2009). In this study, the kriging-based geostatistical technique is used for optimal rain gauge network design. Kriging is a well-known stochastic interpolation technique that provides unbiased estimates of a variable at unsampled locations based on the sampled values at surrounding locations as well as kriging variance of estimation. The optimal rain gauge network is achieved through minimizing the kriging variance of the current network under the framework of variance reduction principle. The principle demonstrates that optimal positioning of additional as well as redundant stations in the high variance zones of the network reduces network variance and thus improves the network performance.

Details of the optimal rain gauge network design in the Middle Yarra River catchment can be found in an earlier study conducted by Adhikary *et al.* (2015). The optimal network in that study was established through a methodical search for the optimal number and locations of stations in the current network using the network extension and rationalization procedures. The optimal network established in this way for the study catchment is shown in Figure 3. As can be seen from the figure, the optimal network consists of 19 rain gauge stations including 16 original stations

(stations R1–R4, R6, R7–R17) in their current positions, two additional stations (stations R18a and R19a), and a redundant station (station R5b) in their corresponding new optimal positions. The rainfall data at the identified optimal locations of the additional and redundant stations (stations R18a, R19a and R5b) in the optimal network are also obtained from the SILO database through their data drill option based on the ordinary kriging technique (Jeffrey *et al.* 2001). A major finding in the study of Adhikary *et al.* (2015) was that the established optimal network provides more accurate areal average and point rainfall estimates in the Middle Yarra River catchment. Now, the objective of the current study is to answer the questions whether the optimal network-based rainfall information could produce enhanced streamflow forecasting.

Augmentation of optimal rain gauge network

In rain gauge network design, it is commonly believed that a denser network with more rain gauge stations causes reduction of network variance and thus results in the improved estimate of areal average or point rainfalls in a catchment (e.g., Papamichail & Metaxa 1996; Cheng *et al.* 2008). Furthermore, the network density often influences the quality of flow simulations (St-Hilaire *et al.* 2003). It is worth mentioning that unlike the past studies, no additional fictitious rain gauge stations to increase the network density were considered for optimal rain gauge network design
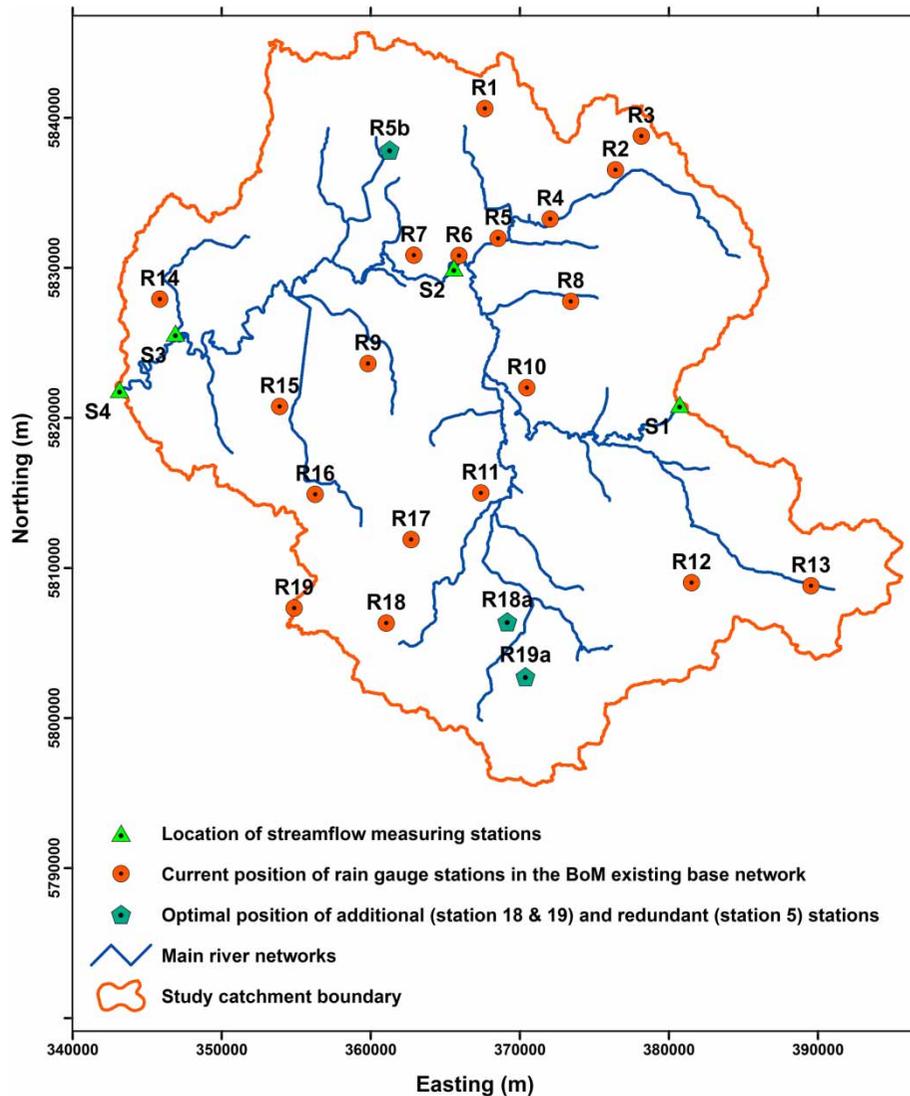
Figure 3 | Optimal rain gauge network as presented in Adhikary *et al.* (2015) for the Middle Yarra River catchment.

presented in Adhikary *et al.* (2015). Considering these factors, additional fictitious stations are incorporated to augment the optimal network of Adhikary *et al.* (2015) to increase the network density, which will be called the augmented optimal rain gauge network in the current study. The main intention is to investigate the potential of an augmented or dense network in enhancing the performance of streamflow forecasting. This strategy facilitates exploring the impact of a relatively denser network on the streamflow forecasting accuracy. This also helps to identify the locations of key fictitious stations in addition to rain gauge stations in

the optimal network, which have greater influence on the accurate streamflow forecasting.

In order to augment the optimal network presented in Adhikary *et al.* (2015), the study catchment is first delineated into a number of sub-catchments based on the digital elevation model using the ArcGIS software. Additional fictitious stations are then placed in such a way that each sub-catchment comprises at least one rain gauge station. Ten additional fictitious stations are considered for the network augmentation. Thus, the resulting augmented optimal network consists of 29 rain

gauge stations, and is shown in Figure 4. The rainfall data at the locations of fictitious stations (stations P1–P10) in the augmented optimal network are also collected from the SILO database. The data are generated through the data drill option based on the ordinary kriging technique (Jeffrey *et al.* 2001). For further details of the rainfall estimation at ungauged locations using the ordinary kriging technique, readers are referred to Adhikary *et al.* (2016b).

ANN-based input variable selection

ANN model

ANNs are biologically inspired general computational models that have been roughly based on the functioning of the human brain. ANN is highly beneficial over conventional hydrological models because it has flexible structures that are able to simulate not only the linear but also the complex
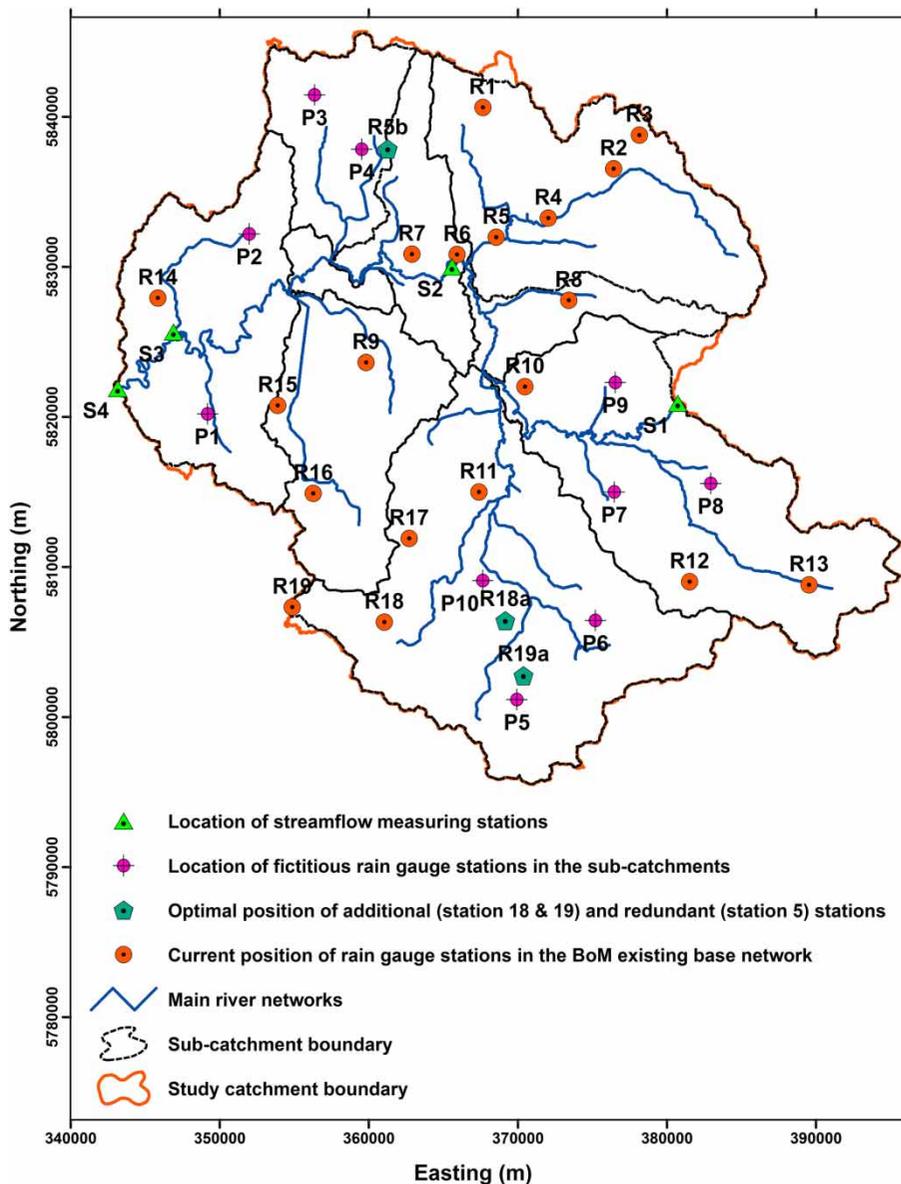

Figure 4 | The augmented optimal rain gauge network with additional fictitious stations in the study area.

nonlinear hydrologic relationship between a model's input and output variables. In addition, ANN is capable of adapting itself to changing conditions leading to enhanced model performance, shorter computation times and faster model development (Yilmaz & Muttil 2014). Once trained properly, the ANN model can be used to make forecasting of a future output for a set of given inputs. Detailed background of the ANN theory can be found in Govindaraju & Rao (2000) and Tayfur (2012).

An ANN is characterized by its architecture, training or learning algorithm and by its activation function. The ANN model constructed in this study is the feed-forward multilayer perceptron (MLP), which is the most commonly used network topology in hydrological forecasting (ASCE Task Committee 2000a, 2000b). The MLP is organized as layers of computing elements, known as neurons, connected between layers via weights. A single hidden layer is considered in this study because a single hidden layer with sufficient neurons is often sufficient in many cases to fit multi-dimensional mapping problems well (Wu *et al.* 2005). Thus, the resulting MLP network configuration, as shown in Figure 5, consists of an input layer that receives inputs from the environment, an intermediate hidden layer, and an output layer that produces the network's response (Muttil & Chau 2006, 2007). The number of neurons in the hidden layer depends on the problem complexity, number of input and output variables. Having a large number of hidden neurons usually gives the network flexibility to

solve complex systems but this may cause overfitting. Therefore, it is essential to identify the optimal number of nodes in the hidden layer, which greatly influences the performance of the trained network. In this study, the optimum number of neurons in the hidden layer is identified using a trial-and-error approach by varying the number of hidden layer neurons.

In the MLP network, processing in neurons is done from the input layer through hidden layers to the output layer. Nonlinearity of the system is captured with activation functions in the ANN model. Among many types of activation functions, the sigmoid and the hyperbolic-tangent activation functions are the most commonly used functions in hydrological modelling (Dawson & Wilby 2001). In this study, the sigmoid activation function is used in the hidden layer and a linear activation function is used in the output layer.

A backpropagation algorithm is used to train the ANN model, which is a supervised learning algorithm that adjusts the connection weights and biases in the backward direction. A number of training algorithms have been developed for error backpropagation learning. In this study, the Levenberg–Marquardt (LM) backpropagation algorithm is used. The LM algorithm is more reliable than any other backpropagation variants because it has the fastest convergence among all algorithms and is also able to obtain the lowest mean square error in many cases (Linares-Rodriguez *et al.* 2015). The ANN model is implemented through the MATLAB Neural Network Toolbox.

A common practice in ANN modelling is to split the input dataset into appropriate training, validation and testing subsets. This often helps to avoid overfitting problems and guarantee generalization capability of ANN (Linares-Rodriguez *et al.* 2015). Thus, the sampled dataset (i.e., 9,667) of this study is divided according to the proportions 70% (i.e., 6,767), 15% (i.e., 1,450) and 15% (i.e., 1,450) for training, validation and testing datasets, respectively. More data (two-thirds of total data) are considered in the training set because in an ideal situation a larger-input dataset is preferable for training an ANN model. This approach often helps to achieve a better calibrated ANN model by capturing all the maximum and minimum values in the data series. The training dataset is used to train the ANN model. The validation dataset is used during the training process to confirm that the model does not cause an overtraining problem.
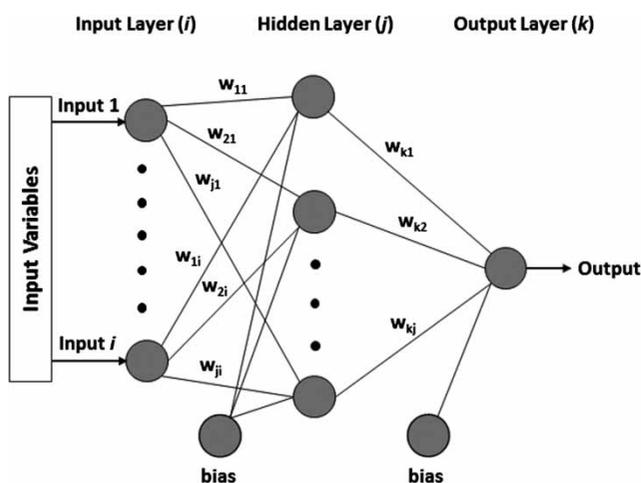

Figure 5 │ Configuration of a three-layer feed-forward multilayer perceptron (MLP) neural network architecture.

In other words, when validation error increases for a specified number of iterations, the training is stopped. Finally, the performance of the trained ANN model is tested using the testing datasets. ANN weights and biases are also initialized using a fixed random seed value so that the same ANN model structure can reproduce the same network response at all times. The backpropagation training of the ANN is terminated after 1,000 epochs, which is expected to be satisfactory in this study.

Identification of significant input variables based on ANN weights

One of the most important steps in ANN modelling is the identification of an appropriate set of input variables that essentially defines the output of a system (Muttil & Chau 2006, 2007). If relevant input variables cannot be accurately identified, it is likely that the desired input–output relationships cannot be accurately captured by the ANN model. On the contrary, when excessive numbers of variables are used as the input, the highly correlated variables dominate the model and hence it is not possible to use information from all the measurements available. In addition, too many inputs may cause overparameterization problems (Akhtar *et al.* 2009; Linares-Rodriguez *et al.* 2015). This is usually addressed by different pre-processing and/or input selection techniques that attempt to reduce the input space by selecting the most significant input variables. The commonly used input selection techniques include correlation-based analysis, mutual information analysis, data mining techniques (e.g., principal component analysis, cluster analysis) and forward selection and backward elimination techniques (Bowden *et al.* 2005; Muttil & Chau 2007).

In the recent past, an ANN-based input selection technique has been demonstrated by Muttil & Chau (2006, 2007) to identify the most significant input variables, which offers several advantages. Since ANN itself is used for significant input variable selection, no further analytical procedures are necessary for the same. A major advantage of ANN model is that it is able to learn problems involving very nonlinear and complex data. Therefore, the model can identify correlated patterns between input data and corresponding target values. The ANN-based input selection technique overcomes some of the limitations associated

with the aforementioned commonly used input selection techniques. For example, ANN can take into account the interaction among variables in the input space and thus identify variables that may not be significant by itself, but are significant in combination with other variables (Muttil & Chau 2007). Thus, the ANN-based technique is ideally suited for identifying significant input variables for streamflow forecasting.

In this study, a two-step procedure is adopted to identify the most significant input variables. First, a set of candidate inputs is prepared based on a priori knowledge of the system being modelled. Before training, it is often useful to scale the inputs and targets so that they always fall within a specified range. In this study, input and output data are standardized between 0 and 1. The MLP network is then trained with the standardized data and the ANN-based input selection technique is used to select the best set of input variables that predominantly describes the streamflow at the catchment outlet. According to this technique, an interpretation of the connection weights along the paths from the input layer to the hidden layer of the trained network is undertaken. The inputs with the largest weight values indicate the most significant input variables. An input significance measure, known as the contribution factor, is used to determine the relative predictive importance of the independent variables in predicting the network's output. The contribution factor of the nth variable, CF_n is defined by Equation (1) as:

$$CF_n = \frac{\sum_{j=1}^{nH} ABS(w_{jn})}{\sum_{i=1}^{nG} \sum_{j=1}^{nH} ABS(w_{ji})} \times 100 \qquad (1)$$

where nG is the number of input variables, nH indicates the number of hidden nodes, w_{ji} are the weights from input layer i to the hidden layer j (as shown in Figure 5) and ABS refers to the absolute function. The summation of absolute values of network weights is used because some weight values may be positive and others are negative (Muttil & Chau 2007).

Streamflow forecasting and assessment

In the current study, streamflow forecasting is achieved through developing the rainfall–runoff (R-R) relationship between the future streamflow at the catchment outlet,

and rainfall and streamflow records available up to the current time t. Mathematically, the R-R relationship can be expressed as:

$$Q_{t+V\Delta t} = f(R_t, R_{t-\Delta t}, \cdots, R_{t-U\Delta t}, Q_t, Q_{t-\Delta t}, \cdots, Q_{t-U\Delta t}) \qquad (2)$$

where Q is the streamflow (m^3/s), R is the rainfall (mm), V (with $V = 1, 2, 3, \ldots$) denotes how far into the future the streamflow forecasting is sought, U (with $U = 1, 2, 3, \ldots$) indicates how far back the recorded data in the time series likely affect the streamflow forecasts while Δt stands for time interval. The neural network structure for the ANN model as generalized in Figure 5 is used to forecast the 1-day-ahead streamflow at the catchment outlet. It is important to note that for a simple demonstration of the proposed methodology, only 1-day-ahead streamflow forecasting is undertaken in the current study and thus 7-days-ahead and/or seasonal forecast of streamflows are not the scope of this work.

In the current study, three different ANN-based streamflow forecasting models are formulated, which are described below:

- **ANN model-1:** This ANN model includes the current rain gauge network-based rainfall data together with streamflow observations as the input (see Figure 1). This model is designated as the base case for comparison in order to test the robustness and efficacy of the proposed approach.
- **ANN model-2:** This ANN model uses the optimal rain gauge network-based rainfall data along with streamflow observations as the input (see Figure 3). This model is indicated as test case-1 wherein no additional fictitious stations are incorporated in the optimal network design.
- **ANN model-3:** This ANN model includes the augmented optimal rain gauge network-based rainfall data in combination with streamflow observations as the input (see Figure 4). This model is designated as test case-2, in which additional fictitious stations are considered to augment the optimal rain gauge network.

The performance of each ANN model for streamflow forecasting is assessed and compared using four different evaluation metrics given in Equations (3)–(6): normalized root mean squared error (NRMSE), mean absolute error (MAE), Nash–Sutcliffe coefficient of efficiency (NSCE),

correlation coefficient (CC). Further details on these metrics can be found in Dawson *et al.* (2007) and Moriasi *et al.* (2007).

$$\text{NRMSE} = \sqrt{\frac{\sum_{t=1}^{N} [Q_{obs}(t) - Q_{est}(t)]^2}{\sum_{t=1}^{N} [Q_{obs}(t) - \bar{Q}_{obs}]^2}} \qquad (3)$$

$$\text{MAE} = \frac{\sum_{t=1}^{N} \|Q_{obs}(t) - Q_{est}(t)\|}{N} \qquad (4)$$

$$\text{NSCE} = 1 - \frac{\sum_{t=1}^{N} [Q_{obs}(t) - Q_{est}(t)]^2}{\sum_{t=1}^{N} [Q_{obs}(t) - \bar{Q}_{obs}]^2} \qquad (5)$$

$$\text{CC} = \frac{\sum_{t=1}^{N} [Q_{obs}(t) - \bar{Q}_{obs}] \times [Q_{est}(t) - \bar{Q}_{est}]}{\sqrt{\sum_{t=1}^{N} [Q_{obs}(t) - \bar{Q}_{obs}]^2} \times \sqrt{\sum_{t=1}^{N} [Q_{est}(t) - \bar{Q}_{est}]^2}} \qquad (6)$$

where $Q_{obs}(t)$ is the observed streamflow at time t, $Q_{est}(t)$ is the forecasted streamflow at time t, $\bar{Q}_{obs}$ and $\bar{Q}_{est}$ are the mean value of observed and forecasted streamflow, respectively, and N is the number of observations in the time series data.

RESULTS AND DISCUSSION

Current and optimal rain gauge network-based input for ANN model

The three-layer feed-forward MLP neural network, as generalized in Figure 5, is first trained to formulate the ANN model-1 using inputs that comprise data from the current rain gauge network-based rainfall and available streamflow records in the Middle Yarra River catchment (Figure 1). The neural network is then trained to formulate the ANN model-2 using inputs that include data from the optimal rain gauge network-based rainfall and available streamflow records in the catchment (Figure 3). As mentioned earlier, there are 19 rain gauges and 4 streamflow measuring stations in the current rain gauge network (Figure 1 and Table 1). The optimal rain gauge network as described in Adhikary *et al.* (2015) also consists of the same number of rain gauge and streamflow measuring stations (Figure 3) because no additional fictitious stations were considered for the optimal network design in that study. A major advantage of the optimal network is that stations are optimally located

in the optimal network and hence it provides improved rainfall estimates (see Adhikary *et al.* (2015) for details).

According to the Bransby Williams formula (Wanielista *et al.* 1997), it is estimated that the catchment has a time of concentration of approximately 3 days. Rainfall occurring within a duration equal to time of concentration would exhibit the greatest influence on streamflows. In addition, streamflow values from the preceding duration provide the antecedent flow information prior to the onset of a rainfall event (Wu *et al.* 2005). Therefore, a time lag of 3 days (t, $t - 1$ and $t - 2$) is adopted in this study to obtain the time-lagged input (rainfall and streamflow) values for forecasting ($t + 1$) streamflows at the catchment outlet. Hence, for each of the current and optimal networks, rainfall and streamflow data from $19 + 4$ (=23) stations gives a total of 23×3 (=69) inputs from which the significant input variables are to be selected to formulate ANN model-1 and ANN model-2.

The input layer of the neural network for both the current and optimal networks consists of 69 nodes based on the 69 inputs. The output layer consists of a single node, which is streamflow at the catchment outlet that is to be forecasted. The neural networks are then trained with the training details and data division described earlier. The backpropagation training of the neural networks is terminated after 1,000 epochs, which is found to be sufficient in this study. In order to find the optimum number of hidden nodes, a trial-and-error procedure is adopted in the training of neural networks by gradually varying the number of nodes in the hidden layer from two to ten. The optimal number of hidden nodes is found to be six for both the current and optimal rain gauge networks. Hence, the resulting neural network based on the current and optimal rain gauge network-based input has a 69-6-1 structure.

ANN weights for each of the trained neural networks with 69-6-1 structure are obtained from the simulation. The ANN weights are inserted in Equation (1) to calculate the contribution factor of each of the 69 inputs for both the current and optimal rain gauge networks, which are presented in Table 2. The sum of the contribution factors of all the 69 input variables should be 100%, which can be seen in Table 2. As explained earlier, the definition of the contribution factor demonstrates that the higher its value for an input variable, the more that input contributes to the forecasting. In other words, if all input variables are considered to have

equal significance, then each input exhibits a significance of 1/69 (equivalent to contribution factor of 1.45%) of the total contribution factor (=100%) of all input variables. Thus, the input variables with a contribution factor greater than 1.45% are considered as the relatively more significant variables, which are indicated with bold font in Table 2. It is evident from Table 2 that the influence of the significant input variables decreases in most cases with an increase of time lag for both the current and optimal rain gauge networks. It is also seen from the table that the optimally located stations in the optimal rain gauge network have a higher contribution factor than that given by those in the current rain gauge network. In other words, the significant input variables based on the optimal network describes the outlet streamflow relatively better than does the current network. This indicates the significance of incorporating the optimal rain gauge network-based input for accurate streamflow forecasting in a catchment, which is the main focus of the current study.

Augmented optimal rain gauge network-based input for ANN model

The MLP neural network, as generalized in Figure 5, is also trained using inputs that include data from the augmented optimal rain gauge network-based rainfall and available streamflow values in the study catchment to formulate the ANN model-3. The augmented optimal network consists of 29 rain gauges and 4 streamflow measuring stations, as shown in Figure 4. Therefore, rainfall and streamflow data from $19 + 10 + 4$ (=33) stations in the augmented optimal network gives a total of 33×3 (=99) inputs considering the adopted 3-day time lag, from which the significant input variables are to be selected for the ANN model-3. Thus, the input layer of the neural network for the augmented optimal network comprises 99 nodes and the output layer consists of a single node based on the outlet streamflow that is to be forecasted. The optimal number of nodes in the hidden layer is found to be four based on the trial-and-error process by gradually varying the number of hidden nodes from two to ten. Thus, the resulting neural network has a 99-4-1 structure for the augmented optimal network, which is trained using the same training specification and data division explained earlier.

ANN weights of the trained neural network with a 99-6-1 structure are then obtained from the simulation. The

Table 2 | Contribution factor based on the trained ANN weights for the current and optimal rain gauge networks for 1-day-ahead streamflow forecasting

		Current rain gauge network (see Figure 1) (BoM's existing base network)				Optimal rain gauge network (see Figure 3) (additional fictitious stations are not considered in the network design)				
		Contribution factor[a] (*CFn*) of the input variables (%)					Contribution factor[a] (*CFn*) of the input variables (%)			
Sl. no.	Input variables	(t)	(t − 1)	(t − 2)	Sum	Input variables	(t)	(t − 1)	(t − 2)	Sum
1	R1	**2.30**	1.18	0.86	4.35	R1	**2.82**	1.41	**1.64**	5.88
2	R2	1.32	**1.84**	0.91	4.06	R2	**2.60**	1.38	0.82	4.80
3	R3	**2.09**	**1.57**	1.32	4.99	R3	**1.93**	**1.68**	1.42	5.03
4	R4	1.18	1.05	**1.50**	3.74	R4	1.04	1.24	0.83	3.12
5	R5	0.81	**2.28**	**1.58**	4.68	R5b[c]	**1.64**	1.01	0.97	3.61
6	R6	0.89	0.86	0.53	2.28	R6	1.28	0.97	0.86	3.10
7	R7	**2.15**	0.71	1.42	4.29	R7	**2.08**	1.00	1.22	4.31
8	R8	**1.79**	**2.52**	**1.55**	5.86	R8	**2.39**	**2.19**	**1.84**	6.43
9	R9	**1.61**	1.42	**1.51**	4.54	R9	1.09	1.02	**1.76**	3.87
10	R10	**2.49**	1.15	**1.79**	5.42	R10	1.16	1.39	0.89	3.44
11	R11	0.96	0.86	1.22	3.04	R11	**1.83**	**1.58**	**1.52**	4.93
12	R12	1.34	0.69	0.61	2.63	R12	1.32	0.92	0.78	3.02
13	R13	**1.90**	1.44	0.73	4.06	R13	0.79	1.40	**1.55**	3.74
14	R14	1.28	**2.22**	1.27	4.76	R14	0.84	0.79	1.16	2.79
15	R15	**2.05**	1.44	0.90	4.39	R15	1.22	0.68	**1.46**	3.36
16	R16	**1.56**	0.65	1.09	3.30	R16	**2.06**	0.64	1.24	3.94
17	R17	**1.86**	0.70	1.21	3.76	R17	**1.78**	1.08	0.76	3.63
18	R18	1.34	**1.98**	0.98	4.30	R18a[b]	**2.10**	**1.62**	0.63	4.35
19	R19	**2.21**	0.56	0.92	3.70	R19a[b]	**1.91**	0.83	1.34	4.08
20	S1	**1.76**	**2.53**	**1.58**	5.88	S1	**1.45**	**1.72**	**1.51**	4.68
21	S2	**2.68**	**2.83**	**1.76**	7.27	S2	**2.79**	**2.44**	**2.35**	7.58
22	S3	**2.91**	1.06	1.23	5.19	S3	**2.62**	1.30	1.03	4.96
23	S4	1.37	1.11	1.03	3.51	S4	**2.49**	**2.29**	0.57	5.35
	Sum of contribution of all variables =				100	Sum of contribution of all variables =				100

[a]Bold font shows variables having a contribution factor greater than 1/69 = 1.45%.
[b]Optimal position of additional rain gauge stations (stations 18 and 19, see Figure 3) as identified by Adhikary *et al.* (2015).
[c]Optimal re-located position of redundant rain gauge station (station 5, see Figure 3) as identified by Adhikary *et al.* (2015).

contribution factor of each of the 99 input variables is calculated using Equation (1) based on the ANN weights for the augmented optimal network, presented in Table 3. The sum of the contribution factors of all the 99 input variables should be 100%, which can be seen in Table 3. In general, if all input variables are considered to have equal significance, then each input has a significance of 1/99 (equivalent to contribution factor of 1.01%) of the total contribution factor (=100%) of all input variables. Thus, the input variables with a contribution factor greater than

1.01% are considered as the relatively more significant variables in this case, which are indicated with bold font in Table 3. As can be seen from the table, apart from the selected other significant input variables, some additional fictitious stations in the augmented optimal network are seen to have influence on the outlet streamflows. This indicates that the optimal locations of rain gauge stations should be decided in the final operational network after satisfying the objectives of accurate rainfall estimations as well as enhanced streamflow forecasting simultaneously.

Table 3 | Contribution factor based on the trained ANN weights for the augmented optimal rain gauge network for 1-day-ahead streamflow forecasting

Augmented optimal rain gauge network (see Figure 4) (additional fictitious stations are considered in the optimal network design)

| Sl. no. | Input variables | Contribution factor[a] (CFn) of the input variables (%) | | | |
		(t)	(t − 1)	(t − 2)	Sum
1	R1	0.82	0.84	0.87	2.53
2	R2	**2.36**	**1.11**	0.54	4.01
3	R3	**1.83**	**1.21**	0.87	3.91
4	R4	0.78	**1.46**	**1.43**	3.68
5	R5b[c]	**1.22**	0.88	0.63	2.74
6	R6	0.60	0.64	**1.02**	2.26
7	R7	**1.79**	0.73	0.98	3.51
8	R8	**1.44**	**1.61**	**2.13**	5.18
9	R9	0.87	**1.28**	0.33	2.47
10	R10	0.57	**1.10**	0.59	2.26
11	R11	0.95	**1.51**	0.85	3.31
12	R12	**2.25**	0.64	0.55	3.44
13	R13	**1.35**	0.36	0.55	2.26
14	R14	0.25	0.24	0.98	1.46
15	R15	0.85	0.60	0.87	2.33
16	R16	**1.35**	0.26	0.58	2.20
17	R17	0.75	0.76	0.30	1.81
18	R18a[b]	0.74	**1.03**	**1.14**	2.91
19	R19a[b]	**1.61**	0.72	**1.06**	3.39
20	S1	**2.22**	**2.02**	**1.62**	5.86
21	S2	**2.55**	**2.55**	**2.00**	7.10
22	S3	**1.10**	**1.18**	0.89	3.17
23	S4	**1.41**	0.62	0.94	2.98
24	P1	**1.01**	0.46	**1.30**	2.78
25	P2	**1.40**	0.41	**1.22**	3.02
26	P3	0.73	0.55	**1.09**	2.37
27	P4	0.49	0.72	**1.02**	2.24
28	P5	0.32	**1.09**	0.38	1.80
29	P6	0.97	0.99	0.91	2.87
30	P7	0.39	0.80	**1.03**	2.22
31	P8	0.80	0.61	**1.22**	2.64
32	P9	0.49	0.84	0.82	2.14
33	P10	**1.30**	**1.20**	0.68	3.18
	Sum of contribution of all variables =				100

[a]Bold font shows variables having a contribution factor greater than $1/99 = 1.01\%$.
[b]Optimal position of additional rain gauge stations (stations 18 and 19, see Figure 3) as identified by Adhikary et al. (2015).
[c]Optimal re-located position of redundant rain gauge station (station 5, see Figure 3) as identified by Adhikary et al. (2015).

## Streamflow forecasting with current and optimal rain gauge network-based input

In order to forecast streamflow, the ANN-based streamflow forecasting models (i.e., ANN model-1 and -2) as explained by Equation (2) are formulated using the identified significant inputs for both the current and optimal rain gauge networks. It can be seen from Table 2 that 29 significant inputs are identified for the current rain gauge network whereas 30 significant inputs are identified for the optimal rain gauge network. The neural networks are trained once again with the training details and data division described earlier using the selected significant inputs. The optimum number of hidden neurons for the neural networks is also obtained through the trial-and-error process described earlier. The optimal number of hidden nodes is found to be two for both ANN model-1 and -2. Hence, the ANN model-1 consists of 29-2-1 structure whereas the ANN model-2 has 30-2-1 structure. The neural network simulations of the ANN model-1 and -2 are then carried out to generate 1-day-ahead streamflow forecasting at the catchment outlet.

For both ANN model-1 and ANN model-2, different performance evaluation measures are computed for the observed and predicted streamflows, which are presented in Table 4. It is evident from the results presented in the table that the ANN model-2 produces more accurate streamflow forecasts than that produced by the ANN model-1. The improvement is significant in terms of the four performance evaluation measures (NRMSE, MAE, NSCE and CC) in which NRMSE and MAE are decreased by 7.1% and 2.4%, respectively, whereas NSCE is increased from 0.919 to 0.930 and CC is improved from 0.961 to 0.969 for the testing dataset. Scatter plots of the ANN model-1 and ANN model-2 forecasting results in the testing phase are presented in Figure 6(a) and 6(b). As can be seen from the figures, scatter points for the ANN model-2 forecasted values are located more closely with the 45° calibration line and thus show a relatively better agreement between the observed and predicted streamflows than given by the ANN model-1 forecast. Time series plots of the observed and simulated streamflows in the testing phase for the ANN model-1 and ANN model-2 are shown in Figure 7(a) and 7(b). It is evident from the figures that ANN model-2 exhibits better agreement between the observed and streamflow time series than ANN

Table 4 │ Comparison of different performance measures of ANN modelling for the 1-day-ahead streamflow forecasting

Rain gauge network and ANN models used	Training phase				Validation phase				Testing phase			
	NRMSE	MAE	NSCE	CC	NRMSE	MAE	NSCE	CC	NRMSE	MAE	NSCE	CC
Current rain gauge network (BoM's base network): ANN model-1	0.251	2.236	0.937	0.968	0.296	1.511	0.913	0.955	0.284	0.946	0.919	0.961
Optimal rain gauge network considering no additional fictitious stations: ANN model-2	0.190	1.620	0.964	0.982	0.248	1.327	0.939	0.969	0.264	0.923	0.930	0.969
Augmented optimal rain gauge network considering additional fictitious stations: ANN model-3	0.183	1.425	0.967	0.983	0.250	1.115	0.938	0.968	0.232	0.658	0.946	0.974

NRMSE, normalized root mean squared error; MAE, mean absolute error; NSCE, Nash–Sutcliffe coefficient of efficiency; CC, correlation coefficient.
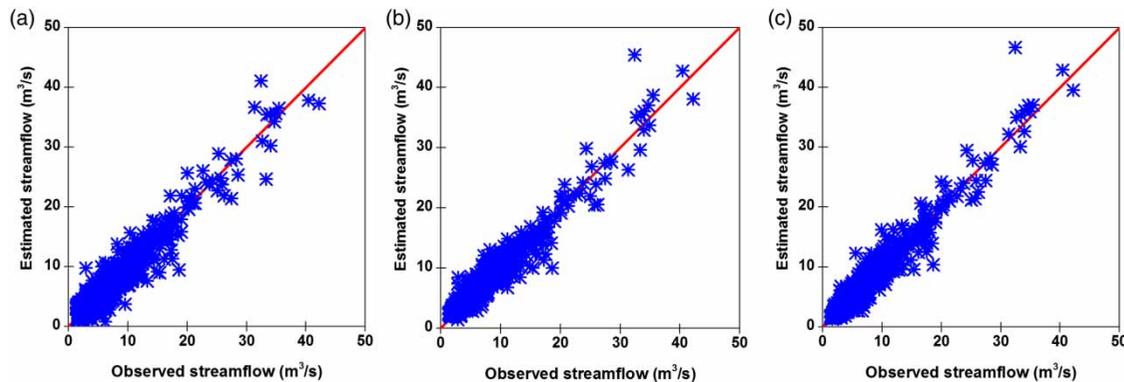
Figure 6 │ Scatter plots for the testing phase for (a) ANN model-1 based on the BoM's current rain gauge network, (b) ANN model-2 based on the optimal rain gauge network considering no additional fictitious stations and (c) ANN model-3 based on the augmented optimal rain gauge network considering additional fictitious stations.

model-1. Figure 7(a) and 7(b) (with the largest peak values zoomed) also indicate that the ability of ANN model-2 to capture the peak values are better than the ANN model-1. This conclusively proves that the optimal rain gauge network-based input (ANN model-2) produces better streamflow forecasts than the current rain gauge network-based inputs (ANN model-1) produce. All these findings reveal the effectiveness of using the optimal rain gauge network-based input in improving the streamflow forecasting of a catchment.

Streamflow forecasting with augmented optimal rain gauge network-based input

ANN model-3 as explained by Equation (2) is formulated using the identified significant inputs for the augmented optimal rain gauge network. Table 3 shows that 42 inputs are selected as the significant inputs for the augmented optimal network. The neural networks are trained once again with the training details

and data division described earlier using the 42 selected significant inputs. The optimum number of hidden neurons for the neural network is obtained through the trial-and-error process described earlier, which is found to be two for the ANN model-3. Thus, the ANN model-3 has 42-2-1 structure, which is then used to generate 1-day-ahead streamflow forecasting at the catchment outlet. Different performance evaluation measures are also computed for the observed and the ANN model-3 forecasting results, and is shown in Table 4. As can be seen from the table, the ANN model-3 outperforms all the models to generate accurate streamflow forecasting. The improvement is significant when compared to ANN model-1, in terms of the four performance evaluation measures (NRMSE, MAE, NSCE and CC) in which NRMSE and MAE are reduced by 18.3% and 30.4%, respectively, whereas NSCE is improved from 0.919 to 0.946 and CC is improved from 0.961 to 0.974 for the testing dataset. Although the improvement by the ANN model-3 is not significant compared
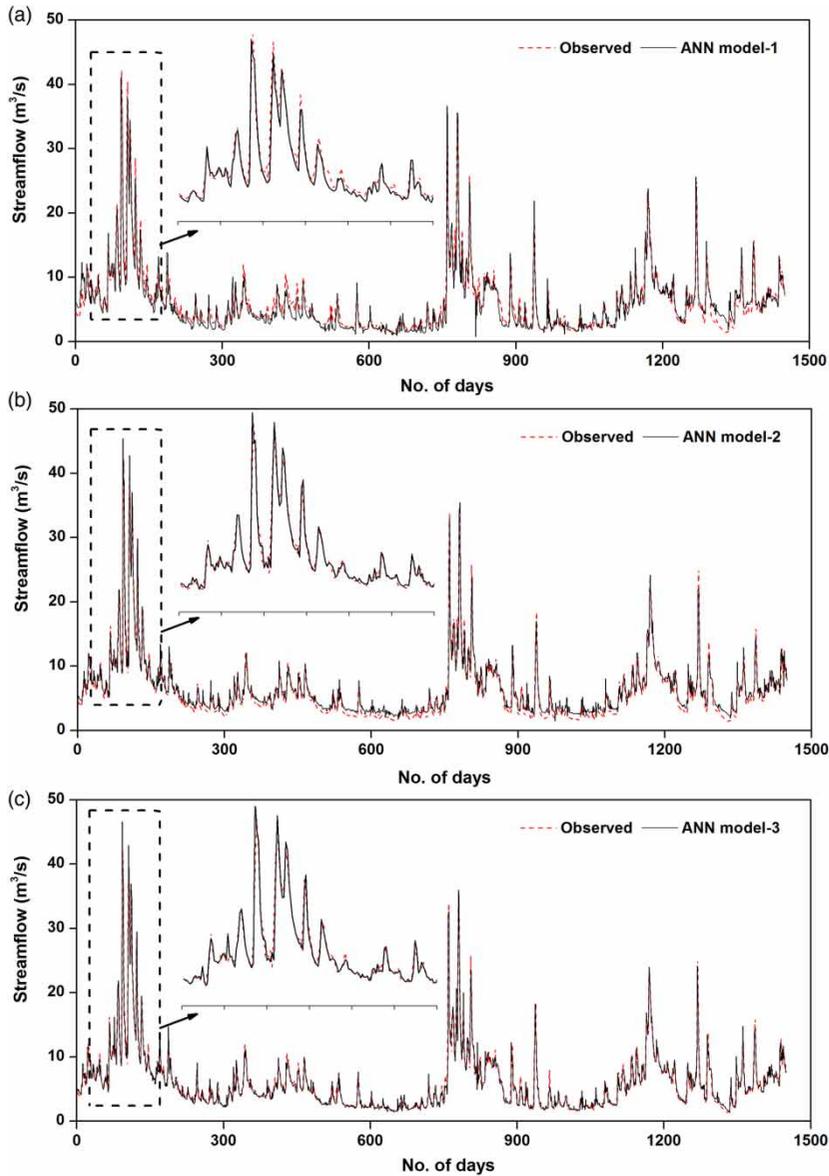
Figure 7 │ Time series plots of the observed streamflow vs simulated streamflow in the testing phase for (a) ANN model-1 based on the BoM's current rain gauge network, (b) ANN model-2 based on the optimal rain gauge network considering no additional fictitious stations and (c) ANN model-3 based on the augmented optimal rain gauge network considering additional fictitious stations.

to the ANN model-2, this gives an important insight about the usage of an augmented rain gauge network for the enhanced streamflow forecasting.

For the ANN model-3 forecasting results, the scatter plot in the testing phase, as shown in Figure 6(c), also shows that the best results are achieved through this model. It is also evident from the time series plot for the ANN model-3 in the testing phase, as shown in Figure 7(c), that the observed and simulated streamflows have the best agreement. As can

also be seen from Figure 7(a)–7(c), the ability of ANN model-3 to capture the peak values are better than the ANN model-1 and ANN model-2. These results indicate that improved streamflow forecasting can be achieved through an augmented rain gauge network. In other words, these results demonstrate that one should obtain the final operational rain gauge network from this augmented network, which is able to provide accurate rainfall estimates (rain gauge network design objective) as well as

give the enhanced streamflow forecasting (flow forecasting objective) simultaneously. However, it is emphasized that if cost is a concern, the optimal network, which provides significant improvement in streamflow forecasting over the current network, should be used in practice for flow forecasting since the optimal network consists of the optimal number of stations. The reason is that unlike the previous studies, the optimal network used in this study was designed without incorporating the additional fictitious rain gauge stations. Again, if forecasting accuracy is taken as the primary objective, the augmented network is recommended for flow forecasting in practice.

CONCLUSIONS

Four conclusions can be drawn based on the findings of the current study:

- The proposed approach of using the rainfall input to ANN-based streamflow forecasting models from the optimal rain gauge network appears to be effective for the enhanced streamflow forecasting, particularly when the current operational rain gauge network is not optimal. The study conclusively proves the significance of the optimal location of rain gauge station in a catchment for enhanced streamflow forecasting.
- The optimal locations of rain gauge stations in the final operational optimal network should be established after satisfying the accurate rainfall estimations and improved streamflow forecasting objectives simultaneously. The network design based on only accurate rainfall estimations objective may not always guarantee accurate streamflow forecasting.
- Further improvement of forecasting performance can be achieved through expansion or augmentation of the rain gauge network considering additional fictitious rain gauge stations. In fact, the best forecasting performances are achieved in this study when the augmented rain gauge network-based input is used in the ANN-based streamflow forecasting models.
- ANN-based input variable selection offers an indirect way of identifying the optimal locations of rain gauge stations in the final operational rain gauge network. The optimal locations of rain gauge stations can be identified from an augmented or expanded network by checking the selected significant input variables.

REFERENCES

Abrahart, R. J., Anctil, F., Coulibaly, P., Dawson, C. W., Mount, N. J., See, L. M., Shamseldin, A. Y., Solomatine, D. P., Toth, E. & Wilby, R. L. 2012 Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting. *Prog. Phy. Geo.* **36** (4), 480–513. DOI: 10.1177/0309133312444943.

Adhikary, S. K., Yilmaz, A. G. & Muttil, N. 2015 Optimal design of rain gauge network in the Middle Yarra River catchment, Australia. *Hydrol. Process.* **29** (11), 2582–2599. DOI: 10.1002/hyp.10389.

Adhikary, S. K., Muttil, N. & Yilmaz, A. G. 2016a Ordinary kriging and genetic programming for spatial estimation of rainfall in the Middle Yarra River catchment, Australia. *Hydrol. Res.* **47** (6), 1182–1197. DOI: 10.2166/nh.2016.196.

Adhikary, S. K., Muttil, N. & Yilmaz, A. G. 2016b Genetic programming-based ordinary kriging for spatial interpolation of rainfall. *J. Hydrol. Eng.* **21** (2), 04015062. DOI: 10.1061/(ASCE)HE.1943-5584.0001300.

Adhikary, S. K., Muttil, N. & Yilmaz, A. G. 2017 Cokriging for enhanced spatial interpolation of rainfall in two Australian catchments. *Hydrol. Process.* **31** (12), 2143–2161. DOI: 10.1002/hyp.11163.

Akhtar, M. K., Corzo, G. A., van Andel, S. J. & Jonoski, A. 2009 River flow forecasting with artificial neural networks using satellite observed precipitation pre-processed with flow lengths and travel time information: case study of the Ganges River basin. *Hydrol. Earth Syst. Sci.* **13**, 1607–1618. DOI: 10.5194/hess-13-1607-2009.

Anctil, F., Michel, C., Perrin, C. & Andréassian, V. 2004 A soil moisture index as an auxiliary ANN input for streamflow forecasting. *J. Hydrol.* **286** (1–4), 155–167. DOI: 10.1016/j.jhydrol.2003.09.006.

Anctil, F., Lauzon, N., Andréassian, V., Oudin, L. & Perrin, C. 2006 Improvement of rainfall-runoff forecasts through mean

areal rainfall optimization. *J. Hydrol.* **328** (3–4), 717–725. DOI: 10.1016/j.jhydrol.2006.01.016.

Andréassian, V., Perrin, C., Michel, C., Usart-Sanchez, I. & Lavabre, J. 2001 Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. *J. Hydrol.* **250** (1–4), 206–223. DOI: 10.1016/S0022-1694(01)00437-1.

ASCE Task Committee 2000a Artificial neural networks in hydrology. I: preliminary concepts. *J. Hydrol. Eng.* **5** (2), 115–137. DOI: 10.1061/(ASCE)10840699(2000)5:2(115).

ASCE Task Committee 2000b Artificial neural networks in hydrology. II: hydrologic applications. *J. Hydrol. Eng.* **5** (2), 115–123. DOI: 10.1061/(ASCE)10840699(2000)5:2(124).

Bárdossy, A. & Das, T. 2008 Influence of rainfall observation network on model calibration and application. *Hydrol. Earth Syst. Sci.* **12**, 77–89. DOI: 10.5194/hess-12-77-2008.

Barua, S., Muttil, N., Ng, A. W. M. & Perera, B. J. C. 2012 Rainfall trend and its implications for water resource management within the Yarra River catchment, Australia. *Hydrol. Process.* **27** (12), 1727–1738. DOI: 10.1002/hyp.9311.

Bastin, G., Lorent, B., Duqué, C. & Gevers, M. 1984 Optimal estimation of the average areal rainfall and optimal selection of rain gauge locations. *Water Resour. Res.* **20** (4), 463–470. DOI: 10.1029/WR020i004p00463.

Birikundavyi, S., Labib, R., Trung, H. T. & Rousselle, J. 2002 Performance of neural networks in daily streamflow forecasting. *J. Hydrol. Eng.* **7** (5), 392–398. DOI: 10.1061/(ASCE) 1084-0699(2002)7:5(392).

Bowden, G. J., Dandy, G. C. & Maier, H. R. 2005 Input determination for neural network models in water resources applications. Part 1 – background and methodology. *J. Hydrol.* **301** (1–4), 75–92. DOI: 10.1016/j.jhydrol.2004.06.021.

Bras, R. L. 1979 Sampling of interrelated random fields: the rainfall-runoff case. *Water Resour. Res.* **15** (6), 1767–1780. DOI: 10.1029/WR015i006p01767.

Bras, R. L. & Rodriguez-Iturbe, I. 1976 Network design for the estimation of areal mean of rainfall events. *Water Resour. Res.* **12** (6), 1185–1195. DOI: 10.1029/WR012i006p01185.

Chen, Y. C., Wei, C. & Yeh, H. C. 2008 Rainfall network design using kriging and entropy. *Hydrol. Process.* **22**, 340–346. DOI: 10.1002/hyp.6292.

Cheng, K. S., Lin, Y. C. & Liou, J. J. 2008 Rain-gauge network evaluation and augmentation using geostatistics. *Hydrol. Process.* **22**, 2554–2564. DOI: 10.1002/hyp.6851.

Daly, E., Kolotelo, P., Schang, C., Osborne, C. A., Coleman, R., Deletic, A. & McCarthy, D. T. 2013 Escherichia coli concentrations and loads in an urbanised catchment: the Yarra River, Australia. *J. Hydrol.* **497**, 51–61. DOI: 10.1016/j.jhydrol.2013.05.024.

Dawson, C. W. & Wilby, R. L. 2001 Hydrological modelling using artificial neural networks. *Prog. Phy. Geo.* **25** (1), 80–108. DOI: 10.1177/030913330102500104.

Dawson, C. W., Abrahart, R. J. & See, L. M. 2007 Hydrotest: a web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Env. Model. Soft.* **22** (7), 1034–1052. DOI: 10.1016/j.envsoft.2006.06.008.

Dibike, Y. B. & Solomatine, D. P. 2001 River flow forecasting using artificial neural networks. *Phy. Chem. Earth, Part B: Hydrol. Ocean. Atmosph.* **26** (1), 1–7. DOI: 10.1016/S1464-1909(01) 85005-X.

Dong, X., Dohmen-Janssen, M. & Booij, M. J. 2005 Appropriate spatial sampling of rainfall for flow simulation. *Hydrol. Sci. J.* **50** (2), 279–298. DOI: 10.1623/hysj.50.2.279.61801.

Ekström, M. & Jones, P. D. 2009 Impact of rainfall estimation uncertainty on streamflow estimations for catchments Wye and Tyne in the United Kingdom. *Int. J. Clim.* **29**, 79–86. DOI: 10.1002/joc.1665.

Faurès, J. M., Goodrich, D., Woolhiser, D. A. & Sorooshian, S. 1995 Impact of small-scale spatial rainfall variability on runoff modelling. *J. Hydrol.* **173**, 309–326. DOI: 10.1016/0022-1694 (95)02704-S.

Feki, H., Slimani, M. & Cudennec, C. 2017 Geostatistically based optimization of a rainfall monitoring network extension: case of the climatically heterogeneous Tunisia. *Hydrol. Res.* **48** (2), 514–541. DOI: 10.2166/nh.2016.256.

Govindaraju, R. S. & Rao, A. R. (eds). 2000 *Artificial Neural Networks in Hydrology.* Kluwer Academic Publishers, Boston, MA.

Huang, W., Xu, B. & Chan-Hilton, A. 2004 Forecasting flows in Apalachicola River using neural networks. *Hydrol. Process.* **18**, 2545–2564. DOI: 10.1002/hyp.1492.

Jeffrey, S. J., Carter, J. O., Moodie, K. B. & Beswick, A. R. 2001 Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Env. Modell. Softw.* **16**, 309–330. DOI: 10.1016/S1364-8152(01)00008-1.

Khu, S., Liong, S. Y., Babovic, V., Madsen, H. & Muttil, N. 2001 Genetic programming and its application in real-time runoff forecasting. *J. Am. Water Resour. Assoc.* **37** (2), 439–451. DOI: 10.1111/j.1752-1688.2001.tb00980.x.

Kişi, Ö. 2007 Streamflow forecasting using different artificial neural network algorithms. *J. Hydrol. Eng.* **12** (5), 532–539. DOI: 10.1061/(ASCE)1084-0699(2007)12:5(532).

Kumar, D. N., Raju, K. S. & Sathish, T. 2004 River flow forecasting using recurrent neural networks. *Water Resour. Manage.* **18** (2), 143–161. DOI: 10.1023/B:WARM.0000024727. 94701.12.

Linares-Rodriguez, A., Lara-Fanego, V., Pozo-Vazquez, D. & Tovar-Pescador, J. 2015 One-day-ahead streamflow forecasting using artificial neural networks and a meteorological mesoscale model. *J. Hydrol. Eng.* **20** (9), 05015001. DOI: 10.1061/(ASCE)HE.19435584.0001163.

Londhe, S. & Charhate, S. 2010 Comparison of data-driven modelling techniques for river flow forecasting. *Hydrol. Sci. J.* **55** (7), 1163–1174. DOI: 10.1080/02626667.2010.512867.

Maier, H. R. & Dandy, G. C. 2000 Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Env. Modell. Softw.* **15** (1), 101–124. DOI: 10.1016/S1364-8152(99) 00007-9.

Maier, H. R., Jain, A., Dandy, G. C. & Sudheer, K. P. 2010 Methods used for the development of neural networks for the prediction of water resource variables in river systems:

current status and future directions. *Env. Modell. Softw.* **25** (8), 891–909. DOI: 10.1016/j.envsoft. 2010.02.003.

Mishra, A. K. & Coulibaly, P. 2009 Developments in hydrometric network design: a review. *Rev. Geophy.* **47**, RG2001. DOI: 10.1029/2007RG000243.

Moriasi, D. N., Arnold, J. G., van Liew, M. W., Bingner, R. L., Harmel, R. D. & Veith, T. L. 2007 Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Tran. ASABE* **50** (3), 885–900. DOI: 10.13031/2013.23153.

Moulin, L., Gaume, E. & Obled, C. 2009 Uncertainties on mean areal precipitation: assessment and impact on streamflow simulations. *Hydrol. Earth Syst. Sci.* **13**, 99–114. DOI: 10. 5194/hess-13-99-2009.

Muttil, N. & Chau, K. W. 2006 Neural network and genetic programming for modelling coastal algal blooms. *Int. J. Env. Pollut.* **28** (3–4), 223–238. DOI: 10.1504/IJEP.2006.011208.

Muttil, N. & Chau, K. W. 2007 Machine learning paradigms for selecting ecologically significant input variables. *Eng. App.Artifi. Intell.* **20**, 735–744. DOI: 10.1016/j.engappai. 2006.11.016.

Oudin, L., Michel, C. & Anctil, F. 2005 Which potential evapotranspiration input for a lumped rainfall-runoff model?: part 1 – can rainfall-runoff models effectively handle detailed potential evapotranspiration inputs? *J. Hydrol.* **303** (1–4), 275–289. DOI: 10.1016/j.jhydrol. 2004.08.025.

Oudin, L., Perrin, C., Mathevet, T., Andréassian, V. & Michel, C. 2006 Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models. *J. Hydrol.* **320** (1–2), 62–83. DOI: 10.1016/j.jhydrol.2005.07.016.

Papamichail, D. M. & Metaxa, I. G. 1996 Geostatistical analysis of spatial variability of rainfall and optimal design of rain gauge network. *Water Resour. Manage.* **10** (2), 107–127. DOI: 10. 1007/BF00429682.

Pardo-Igúzquiza, E. 1998 Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing. *J. Hydrol.* **210** (1–4), 206–220. DOI: 10.1016/S0022-1694(98)00188-7.

Porporato, A. & Ridolfi, L. 2001 Multivariate nonlinear prediction of river flows. *J. Hydrol.* **248** (1–4), 109–122. DOI: 10.1016/ S0022-1694(01)00395-X.

Seed, A. W. & Austin, G. L. 1990 Sampling errors for rain gauge-derived mean areal daily and monthly rainfall. *J. Hydrol.* **118** (1–4), 163–173. DOI: 10.1016/0022-1694(90)90256-W.

Sivapragasam, C., Vanitha, S., Muttil, N., Suganya, K., Suji, S., Selvi, M. T. & Sudha, S. J. 2014 Monthly flow forecast for Mississippi River basin using artificial neural networks. *Neu. Comp. App.* **24** (7), 1785–1793. DOI: 10.1007/s00521-013-1419-6.

Srinivasulu, S. & Jain, A. 2009 River flow prediction using an integrated approach. *J. Hydrol. Eng.* **14** (1), 75–83. DOI: 10. 1061/(ASCE)1084-0699(2009)14:1(75).

St-Hilaire, A., Ouarda, T. B. M. J., Lachance, M., Bobee, B., Gaudet, J. & Gignac, C. 2003 Assessment of the impact of

meteorological network density on the estimation of basin precipitation and runoff: a case study. *Hydrol. Process.* **17** (18), 3561–3580. DOI: 10.1002/hyp.1350.

Storm, B., Høgh, J. K. & Refsgaard, J. C. 1989 Estimation of catchment rainfall uncertainty and its influence on runoff prediction. *Hydrol. Res.* **19** (2), 77–88.

Talei, A., Chua, L. H. C. & Wong, T. S. W. 2010 Evaluation of rainfall and discharge inputs used by adaptive network-based fuzzy inference systems (ANFIS) in rainfall-runoff modeling. *J. Hydrol.* **391** (3–4), 248–262. DOI: 10.1016/j.jhydrol.2010.07.023.

Taormina, R., Chau, K. W. & Sivakumar, B. 2015 Neural network river forecasting through baseflow separation and binary-coded swarm optimization. *J. Hydrol.* **529**, 1788–1797. DOI: 10.1016/j.jhydrol.2015.08.008.

Tayfur, G. 2012 *Soft Computing in Water Resources Engineering: Artificial Neural Networks, Fuzzy Logic and Genetic Algorithms.* WIT Press, Southampton, UK.

Tsai, M. J., Abrahart, R. J., Mount, N. J. & Chang, F. J. 2014 Including spatial distribution in a data-driven rainfall-runoff model to improve reservoir inflow forecasting in Taiwan. *Hydrol. Process.* **28** (3), 1055–1070. DOI: 10.1002/hyp.9559.

Tsintikidis, D., Georgakakos, K. P., Sperfslage, J. A., Smith, D. E. & Carpenter, T. M. 2002 Precipitation uncertainty, rain gauge network design within Folsom Lake watershed. *J. Hydrol. Eng.* **7** (2), 175–184. DOI: 10.1061/(ASCE)1084-0699(2002)7:2(175).

Volkmann, T. H. M., Lyon, S. W., Gupta, H. V. & Troch, P. A. 2010 Multicriteria design of rain gauge networks for flash flood prediction in semiarid catchments with complex terrain. *Water Resour. Res.* **46** (11), W11554. DOI: 10.1029/ 2010WR009145.

Wanielista, M., Kersten, R. & Eaglin, R. 1997 *Hydrology: Water Quantity and Quality Control,* 2nd edn. John Wiley & Sons, New York.

Wu, J. S., Han, J., Annambhotla, S. & Bryant, S. 2005 Artificial neural networks for forecasting watershed runoff and streamflows. *J. Hydrol. Eng.* **10** (3), 216–222. DOI: 10.1061/ (ASCE)1084-0699(2005)10:3(216).

Xu, C. Y., Tunemar, L., Chen, Y. D. & Singh, V. P. 2006 Evaluation of seasonal and spatial variations of conceptual hydrological model sensitivity to precipitation data errors. *J. Hydrol.* **324** (1–4), 80–93. DOI: 10.1016/j.jhydrol.2005.09.019.

Xu, H., Xu, C. Y., Chen, H., Zhang, Z. & Li, L. 2013 Assessing the influence of rain gauge network density and distribution on hydrological model performance in a humid region of China. *J. Hydrol.* **505**, 1–12. DOI: 10.1016/j.jhydrol.2013.09.004.

Yilmaz, A. G. & Muttil, N. 2014 Runoff estimation by machine learning methods and application to the Euphrates Basin in Turkey. *J. Hydrol. Eng.* **19** (5), 1015–1025. DOI: 10.1061/ (ASCE)HE.1943-5584.0000869.

Zealand, C. M., Burn, D. H. & Simonovic, S. P. 1999 Short-term streamflow forecasting using artificial neural networks. *J. Hydrol.* **214** (1–4), 32–48. DOI: 10.1016/S00221694(98) 00242-X.