Atmospheric Research 212 (2018) 240–258

Contents lists available at ScienceDirect

Atmospheric Research

journal homepage: www.elsevier.com/locate/atmosres

Statistical downscaling of precipitation using machine learning techniques

D.A. Sachindra[a,*], K. Ahmed[b], Md. Mamunur Rashid[c], S. Shahid[d], B.J.C. Perera[a]

[a] Institute for Sustainability and Innovation, College of Engineering and Science, Victoria University, P.O. Box 14428, Melbourne, Victoria 8001, Australia
[b] Faculty of Water Resources Management, Water and Marine Sciences, Lasbela University of Agriculture, Uthal, Balochistan, Pakistan
[c] Civil, Environmental, and Construction Engineering Department, University of Central Florida, Orlando, Florida 32816-2450, USA
[d] Faculty of Civil Engineering, Universiti Teknologi Malaysia, Johor Bahru 81310, Malaysia

ARTICLE INFO

Keywords:
Statistical downscaling
Machine learning
Precipitation
Australia
Floods
Droughts

ABSTRACT

Statistical models were developed for downscaling reanalysis data to monthly precipitation at 48 observation stations scattered across the Australian State of Victoria belonging to wet, intermediate and dry climate regimes. Downscaling models were calibrated over the period 1950–1991 and validated over the period 1992–2014 for each calendar month, for each station, using 4 machine learning techniques, (1) Genetic Programming (GP), (2) Artificial Neural Networks (ANNs), (3) Support Vector Machine (SVM), and (4) Relevance Vector Machine (RVM). It was found that, irrespective of the climate regime and the machine learning technique, downscaling models tend to better simulate the average (compared to other statistics) and under-estimate the standard deviation and the maximum of the observed precipitation. Also, irrespective of the climate regime and the machine learning technique, at the majority of stations downscaling models showed an over-estimating trend of low to mid percentiles (i.e. below the 50th percentile) of precipitation and under-estimating trend of high percentiles of precipitation (i.e. above the 90th percentile). The over-estimating trend of low to mid percentiles of precipitation was more pronounced at stations located in dryer climate, irrespective of the machine learning technique. Based on the results of this investigation the use of RVM or ANN over SVM or GP for developing downscaling models can be recommended for a study such as flood prediction which involves the consideration of high extremes of precipitation. Also, RVM can be recommended over GP, ANN or SVM in developing downscaling models for a study such as drought analysis which involves the consideration of low extremes of precipitation. Furthermore, it was found that irrespective of the climate regime, the SVM and RVM-based precipitation downscaling models showed the best performance with the Polynomial kernel.

1. Introduction

The assessment of water resources in catchments under changing climate is important as the spatial and temporal variability of water resources is highly influenced by the changes in the climate (Pascual et al., 2015). General Circulation Models (GCMs) are considered as the most advanced tools available for obtaining global scale climate change projections of hydroclimatic variables (Bates et al., 2010). GCMs are forced with likely future GHG emission scenarios in order to produce scenarios of global climate likely to occur in the future. Owing to the coarse spatial scale at which GCMs operate, they are unable to resolve sub-grid scale processes such as cloud physics and land surface processes, and also the topography of the Earth is coarsely represented within the structure of GCMs (Iorio et al., 2004). Therefore, projections of GCMs cannot be readily used in catchment scale applications such as hydrologic modelling or water resources allocation modelling.

In order to bridge the spatial scale gap between the coarse scale GCM outputs and catchment scale hydroclimatic variables, statistical and dynamic downscaling approaches have been developed (Wilby and Wigley, 1997). In statistical downscaling, empirical statistical relationships between GCM outputs and catchment scale hydroclimatic variables are developed to bridge the spatial scale gap between GCM outputs and catchment scale hydroclimatic variables (Benestad et al., 2008). In dynamic downscaling, physics based equations are used for the same purpose (Fowler and Wilby, 2010). Statistical downscaling has gained wide popularity due to its low computational cost and simplicity (Okkan and Inan, 2014; Rashid et al., 2015; Sachindra et al., 2016), compared to its counterpart dynamic downscaling.

According to Wilby et al. (2004), statistical downscaling approaches can be further sub-divided into three categories; regression-based approaches, weather classification-based approaches

* Corresponding author at: Institute for Sustainability and Innovation, College of Engineering and Science, Victoria University, Footscray Park Campus, P.O. Box 14428, Melbourne, Victoria 8001, Australia.
E-mail address: sachindra.dhanapalaarachchige@vu.edu.au (D.A. Sachindra).

https://doi.org/10.1016/j.atmosres.2018.05.022
Received 27 September 2017; Received in revised form 10 May 2018; Accepted 10 May 2018
Available online 27 May 2018

Table 1
Precipitation stations considered in the study.

Climate regime	Station ID	Station name	Lat	Long	Avg	Stdev	M
Relatively wet	83032	Whitlands (Burder's Lane)	−36.8	146.3	115.7	83.9	0.3
	86090	Warburton (O'Shannassy Reservoir) (Quarte))	−37.7	145.8	115.5	64.7	0.1
	88060	Kinglake West (Wallaby Creek)	−37.4	145.2	100.8	58.0	0.8
	85026	Erica	−38.0	146.4	91.5	49.8	0.8
	85096	Wilsons Promontory Lighthouse	−39.1	146.4	89.8	55.2	0.6
	85040	Fish Creek (Hoddle Range)	−38.7	146.1	87.6	47.6	0.6
	82008	Callaghan Creek Station	−36.5	147.4	86.4	59.2	1.2
	85023	Drouin Bowling Club	−38.1	145.9	83.2	41.3	1.0
	86127	Wonthaggi	−38.6	145.6	80.1	43.7	0.1
	90015	Cape Otway Lighthouse	−38.9	143.5	77.7	42.5	0.5
	88020	Daylesford	−37.3	144.2	74.8	48.5	0.9
	90060	Nullawarre	−38.5	142.8	72.6	41.4	0.9
	90021	Cobden (Post Office)	−38.3	143.1	71.1	40.9	0.3
	84005	Buchan	−37.5	148.2	68.5	45.7	1.0
	90013	Cape Bridgewater	−38.3	141.4	68.0	42.1	0.9
	86117	Toorourrong Reservoir (Toorourrong)	−37.5	145.2	68.0	38.7	0.0
Intermediate	87046	Scotsburn (Mount Boninyong)	−37.7	143.9	66.4	37.2	0.5
	88042	Malmsbury Reservoir	−37.2	144.4	63.0	41.8	0.0
	84044	Black Mountain	−37.0	148.3	62.2	39.7	0.5
	86018	Caulfield (Racecourse)	−37.9	145.0	62.1	35.8	0.6
	90063	Penshurst (Post Office)	−37.9	142.3	61.6	34.8	0.5
	84014	Dellicknora (Tellicura)	−37.1	148.7	60.3	38.6	1.0
	82010	Chiltern (Post Office)	−36.1	146.6	59.4	41.6	0.4
	84015	Ensay	−37.4	147.8	59.1	42.0	0.4
	89019	Mirranatwa (Bowacka)	−37.4	142.4	56.9	35.8	0.1
	89002	Ballarat Aerodrome	−37.5	143.8	56.7	34.3	0.1
	90033	Dergholm (Hillgrove)	−37.4	141.2	55.6	37.7	0.3
	85034	Glenmaggie Weir	−37.9	146.8	53.5	41.6	0.6
	88067	Yea	−37.2	145.4	52.2	31.7	0.9
	89009	Cavendish (Post Office)	−37.5	142.0	51.2	31.2	0.1
	85072	East Sale Airport	−38.1	147.1	49.7	34.0	0.0
	81033	Molka (Lowana)	−36.6	145.4	45.1	33.4	0.1
Relatively dry	81008	Colbinabbin	−36.5	144.8	45.1	33.8	0.8
	87009	Bannockburn	−38.0	144.2	44.0	27.5	1.7
	79008	Clear Lake	−36.9	141.9	41.9	30.2	0.6
	81085	Dunolly	−36.9	143.7	41.0	31.2	0.4
	81046	Stanhope	−36.4	145.0	39.0	30.8	0.4
	79016	Warranooke (Glenorchy)	−36.7	142.7	38.1	28.9	0.3
	81058	Bridgewater (Post Office)	−36.6	143.9	38.1	30.5	1.0
	78078	Kaniva	−36.4	141.2	37.1	27.5	0.3
	80027	Korong Vale (Burnbank)	−36.3	143.7	35.7	28.2	1.0
	78010	Dimboola	−36.5	142.0	33.4	26.4	1.0
	77030	Narraport	−36.0	143.0	30.4	25.1	0.5
	77051	Rainbow (Werrap (Oak-Lea))	−35.9	141.9	29.1	24.2	0.9
	76064	Walpeup Research	−35.1	142.0	28.3	24.3	0.5
	76043	Nulkwyne Kiamal	−34.9	142.2	26.1	23.8	1.0
	76037	Murray Lock Number 9	−34.2	141.6	24.1	23.4	0.5
	76015	Irymple (Arlington)	−34.2	142.2	23.9	24.1	0.5

*Avg = monthly average of precipitation over period 1950–2014 in mm, Stdev = monthly standard deviation of precipitation over period 1950–2014 in mm, Lat = latitude, Long = longitude. M = percentage of missing data in the record of observations of monthly precipitation over period 1950–2014. Station ID = station identification number assigned by Bureau of Meteorology of Australia.

and approaches based on weather generators. Regression-based statistical downscaling approaches have gained popularity out of the above three categories owing to their simplicity in application. The regression techniques widely used in statistical downscaling include Multi Linear Regression (MLR) (Sachindra et al., 2014a), Generalized Linear Models (GLMs) (Beecham et al., 2014), Artificial Neural Networks (ANNs) (Tripathi et al., 2006; Ahmed et al., 2015), Support Vector Machine (SVM) (Sachindra et al., 2013; Goly et al., 2014), Relevance Vector Machine (RVM) (Ghosh and Mujumdar, 2008; Okkan and Inan, 2014), Genetic Programming (GP) (Coulibaly, 2004; Sachindra et al., 2018) and Gene Expression Programming (GEP) (Hashmi et al., 2011; Sachindra et al., 2016). Owing to the learning abilities from data and their use in computer algorithms, techniques such as ANN, SVM, RVM, and GP are often called machine learning techniques.

In the past literature, studies have been documented on the comparison of performance of different downscaling approaches developed with machine learning techniques and traditional statistical techniques. Some examples for such studies are provided in this paragraph. In a downscaling exercise, Coulibaly (2004) found that GP-based downscaling models were able to better simulate both daily minimum and maximum temperature in comparison to that by MLR-based downscaling models. In a streamflow downscaling study, Sachindra et al. (2013) discovered that a Least Square Support Vector Machine (LSSVM)

based downscaling model was able to better capture the observed streamflow in comparison to that by a MLR-based model. Duhan and Pandey (2015) found that, SVM-based downscaling models are able to better perform in simulating the observed monthly maximum and minimum temperature in comparison to that by ANN and MLR-based models. Goly et al. (2014) employed MLR, positive coefficient regression (PCR), stepwise regression (SR), and SVM for downscaling large scale atmospheric variables to monthly precipitation, and concluded that SVM-based downscaling models outperform models developed with all other techniques in simulating statistics of monthly observed precipitation. According to above studies, downscaling models developed with machine learning techniques perform better in comparison to downscaling models developed with traditional statistical regression techniques.

Though downscaling literature contains the details of many comparison studies of models developed with various techniques, the literature lacks the details of a single study which assesses the performances of models developed using machine learning techniques; GP, ANN, SVM and RVM for downscaling large scale atmospheric information to catchment scale precipitation under diverse climate (relatively wet, intermediate and relatively dry). Also, the current literature does not contain a detailed investigation on the selection of a suitable kernel in the application of SVM and RVM techniques in developing downscaling models. This paper is dedicated to the assessment of the effectiveness of the use of GP, ANN, SVM and RVM in the development of models for downscaling large scale atmospheric information to catchment scale monthly precipitation under diverse climate. In addition to that, this paper presents an investigation on the assessment of suitability of number of different kernel functions in SVM and RVM-based downscaling models under diverse climate.

2. Study area and data

For the case study, 48 precipitation observation stations located across Victoria (237,000 km^2), Australia were selected. These precipitation observation stations were selected in such a way that they contain records of observations over the period 1950–2014 with the minimum missing data and they represent relatively wet, intermediate and relatively dry climate regimes. Names of the precipitation observations stations, their locations along with the long-term statistics of observed precipitation over the period 1950–2014 are shown in Table 1. The first 16 stations (first 1/3 of stations) in Table 1 were considered as relatively wet stations and the second and the third 16 stations were classified as intermediate and relatively dry stations respectively. In order to categorise the 48 precipitation stations into wet, intermediate and dry climate regimes, stations were sorted in the descending order of the monthly average of observed precipitation pertaining to the period 1950–2014. Once the 48 stations were sorted in the descending order of the monthly average of observed precipitation, the first 1/3 of the 48 stations was considered as the stations which represent the relatively wet climate and the last 1/3 of stations was considered as the stations which represent the relatively dry climate, rest of the stations represented the intermediate climate. The categorization of the 48 precipitation stations into relatively wet, intermediate and relatively dry climate regimes was based on the notion that each climate regime should contain equal number of stations, so that it allows a fair comparison of performance of downscaling models between the three climate regimes. This kind of categorization enables the separation of relatively wet and relatively dry stations clearly into two groups, as stations belonging to the intermediate regime depict a transition zone between wet and dry climate.

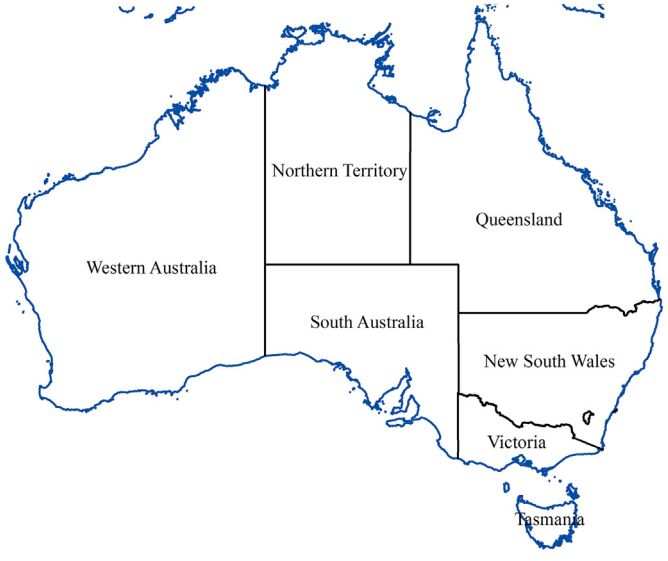

Fig. 1. Location of Victoria in Australia.

Fig. 1 shows the location of Victoria which is the study area of the present investigation.

The locations of the 48 precipitation observation stations in Victoria are shown in Fig. 2, along with topography and average annual long-term precipitation.

Monthly observations of precipitation for the above 48 stations were obtained from the Australian Bureau of Meteorology database (www.bom.gov.au/) for the period 1950–2014. Precipitation observations at these stations contained < 1.7% missing data per station over the period 1950–2014 (see Table 1), and the missing data were filled by obtaining the average of the precipitation values of preceding and succeeding months.

The north western region of Victoria is located close to the driest part of Australia which is in the central region of the country. Hence, north western region of Victoria experiences relatively dry climate. On the other hand, the south eastern region of Victoria is bound by the sea and is separated from the rest of the state by a mountain range called the “Great Dividing Range” (see elevation and precipitation maps in Fig. 2). This causes the south eastern region of Victoria to be relatively wet in comparison to the north western region. According to Table 1, it was seen that the stations those showed higher/lower monthly average precipitation also displayed higher/lower monthly standard deviation denoting higher/lower degree of fluctuations in the precipitation regime. In the relatively wet, intermediate and relatively dry climate regimes considered in this study, there are 4 distinct seasons with respect to temperature. These seasons are summer (December–February), autumn (March–May), winter (June–August) and spring (September–November) (Min et al., 2013). Winter and summer are the wettest and the driest seasons in all three climate regimes, respectively.

NCEP/NCAR reanalysis data set is regarded as the most widely used reanalysis data set in climate research (Brands et al., 2012). Hence, for

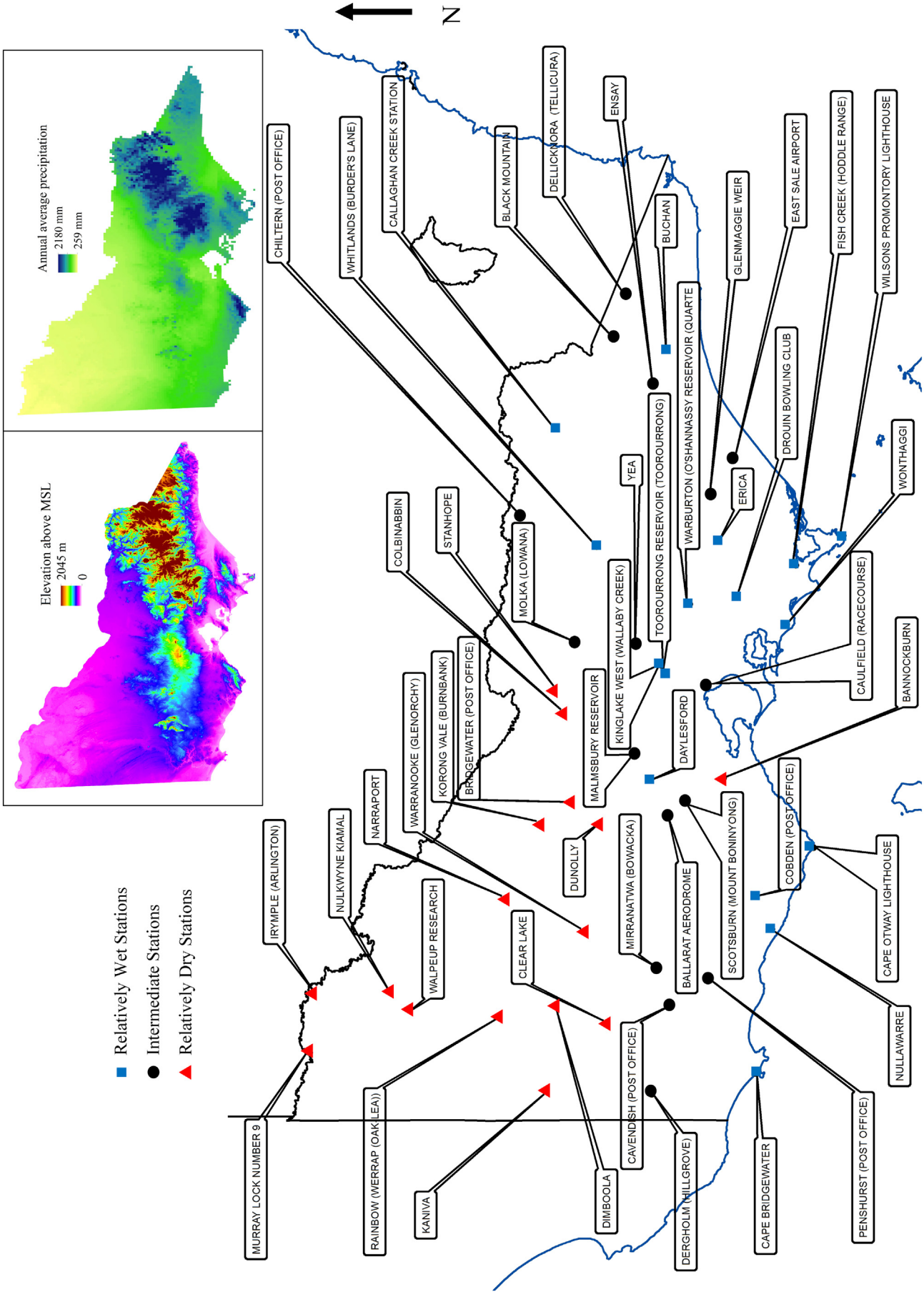

Fig. 2. Locations of observation stations in Victoria along with topography and average annual precipitation.

the development of downscaling models for each calendar month, for each station, monthly NCEP/NCAR reanalysis data pertaining to a set of probable predictors for the period 1950–2014 were downloaded from the Physical Sciences Division of the Earth System Research Laboratory of National Oceanic and Atmospheric Administration (NOAA) website at http://www.esrl.noaa.gov/psd/.

3. Methodology

Sub-section 3.1 provides the details of the theory of machine learning techniques used in this study, and Sub-section 3.2 details the application of the machine learning techniques in the development of downscaling models.

3.1. Theory of machine learning techniques

In this study, for each station, for each calendar month downscaling models were developed using 4 different machine learning techniques, namely; (1) GP, (2) ANN, (3) SVM and (4) RVM. Sub-sections 3.1.1, 3.1.2, 3.1.3 and 3.1.4 provide a brief theoretical background of GP, ANN, SVM and RVM respectively.

3.1.1. Genetic programming (GP)

GP is an evolutionary algorithm which is based on the Darwin's theory of evolution (Koza, 1992). In the GP algorithm, initially, a set of equations (e.g. downscaling models) is randomly generated for relating the predictors with the predictand. Then these equations (models) are evolved by performing various genetic operations on them until their fitness reaches a maximum (Whigham and Crapper, 2001) or a predefined fitness threshold or in certain instances up to a predefined number of generations (Selle and Muttil, 2011).

For the implementation of GP, training (calibration) and testing (validation) data sets, population size, tree size (depth of a tree), a set of terminals, a set of mathematical functions, a fitness measure, criterion for selecting models for the mating pool, values of genetic operators (e.g. probabilities of crossover, mutation and replication) and termination/stopping criterion should be predefined. Once above attributes of GP are defined, both model structure and optimum predictors are simultaneously determined by the GP algorithm through model evolution. Table 2 shows details of the main attributes of GP.

The implementation of the GP algorithm involves the following steps: (1) random generation of an initial population of mathematical equations (e.g. downscaling models), (2) evaluation of fitness of each equation in the initial population, (3) selection of equations for the mating pool from the initial population based on their fitness, (4) generation of a new population of equations by performing genetic operations on the above selected equations, and (5) repetition of steps 2 to 4 until a predefined stopping criterion is satisfied (Koza, 1992). The main steps of GP algorithm are shown in Fig. 3. More details on the GP algorithm can be found in Koza (1992).

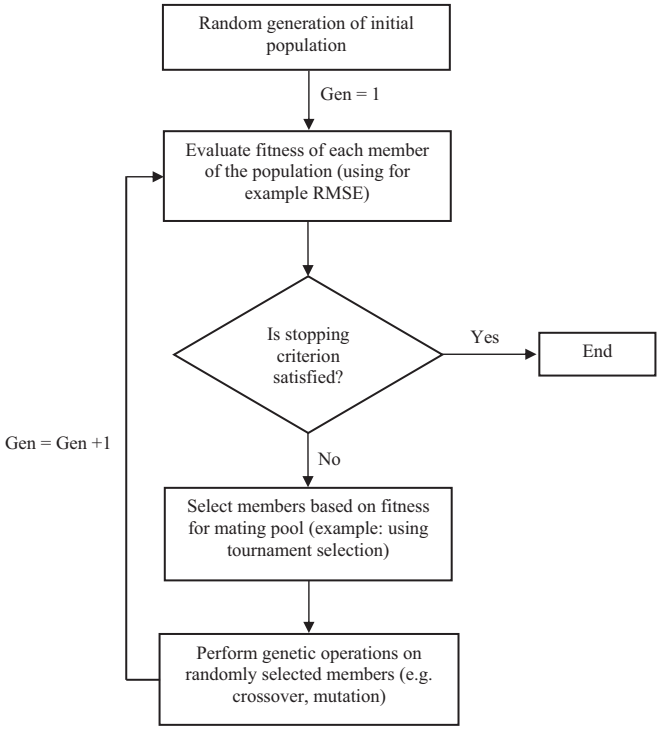

Fig. 3. Genetic programming algorithm.

3.1.2. Artificial neural networks (ANNs)

Artificial Neural Networks (ANNs) have been inspired by the function of the biological nervous system in the human brain (Agatonovic-Kustrin and Beresford, 2000). The ability of ANN to learn and generalize relationships between data of predictors and predictands makes it possible to solve large-scale complex problems which are either linear or nonlinear in nature (e.g. downscaling of precipitation). This provides

Table 2
Main attributes of Genetic Programming Algorithm.

GP attribute	Description
Training and testing data %	Percentage of data used for training (calibration) and testing (validation) of models
Population size	Number of models in a generation
Program size/tree size	The maximum limit on the size of GP trees
Terminals	Inputs, number of constants (values of constants determined through evolution of models)
Mathematical Function set	Mathematical functions such as +, −, x, ÷, √, power, sin, cos and Boolean functions such as AND, OR and NOT
Fitness measure	Measurement of fitness of models in a generation (e.g. root mean square error)
Selection criterion for mating pool	How members of a generation are selected for the mating pool (e.g. roulette selection, tournament selection)
Mutation probability	Probability of replacing a sub-tree of a member with a new sub-tree
Crossover probability	Probability of swapping sub-trees between two members
Replication probability	Probability of copying a member in one generation to next generation
Stopping/termination criterion	This decides when to stop the evolution (e.g. maximum fitness, predefined fitness threshold, predefined number of generations)

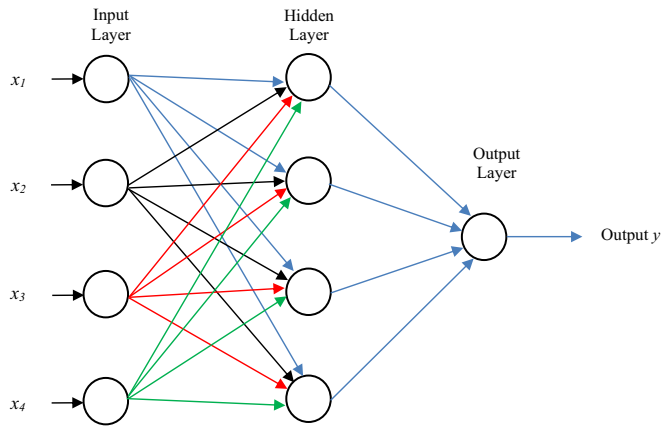

Fig. 4. A typical artificial neural network.

ANN an advantage over the traditional statistical regression techniques such MLR and GLM. ANNs have been successfully applied in the fields of hydrology and climatology widely (Kişi, 2008; Nourani et al., 2009; Govindaraju and Rao, 2013; Belayneh et al., 2014; Okkan and Fistikoglu, 2014; Sahay and Srivastava, 2014; Ahmed et al., 2015; Vu et al., 2016).

Fig. 4 shows a typical ANN consists of an input layer with 4 neurons (4 input variables; x_1, x_2, x_3 and x_4), one hidden layer with 4 neurons and an output layer with 1 neuron (1 output variable, y) in a graphical manner. The number of hidden layers and nodes are usually determined using a trial and error procedure (Ham and Kostanic, 2000).

An ANN can be expressed in the form of Eq. (1) in general, for an input vector $x = (x_1, x_2, x_3 \ldots, x_n)$ with weight vector for a node j, $w_j = (w_{1j}, w_{2j}, w_{3j} \ldots, w_{nj})$, to compute the output y_j at node j.

$$y_j = f(x.\, w_j - b_j) \tag{1}$$

In Eq. (1), $f(.)$ is the activation function and b_j is the bias associated with node j. The Hyperbolic tangent sigmoid function shown in Eq. (2) is one of the most commonly used activation functions in ANN architecture (Mekanik et al., 2013).

$$f(.) = \frac{2}{1 + e^{-2(.)}} - 1 \tag{2}$$

In hydrology and climatology research, Multilayer Perceptions (MLP) is one of the most widely used ANN architectures (Ahmed et al., 2015). Typically, a MLP is a feed forward neural network which consists of an input layer, hidden layer and an output layer. Mathematically a MLP neural network can be represented as Eq. (3) (Kim and Valdés, 2003), where y_k and x_j refer to the outputs and the inputs to the network, respectively.

$$y_k = f_2\left[\sum_{i=1}^{m} w_{ki} f_1\left(\sum_{j=1}^{n} w_{ij} x_j + b_i\right) + b_k\right] \tag{3}$$

In Eq. (3), subscripts i, j and k refer to hidden, input and output layers respectively, and n and m refer to the number of neurons on the input and hidden layers respectively. w_{ij} are the weights between neurons on input and hidden layers and w_{ki} are the weights between neurons on hidden and output layers. f_1 and f_2 are the activation functions for the hidden and output layers, respectively, and b_i and b_k are the bias associated with the neurons on hidden and output layers, respectively.

In ANN, initial values of the weights are randomly assigned to each node and the bias is estimated as the difference between the values simulated by the model and the corresponding observations. Then the optimum values of weights are determined by iteratively changing them to minimize the bias and this process is known as training. A detailed description on ANNs can be found in Haykin (2009) and Shanmuganathan and Samarasinghe (2016).

3.1.3. Support vector machine (SVM)

SVM (Vapnik, 1998) can be represented as a two-layer neural network (Tripathi et al., 2006) that can be used for linear and non-linear regression. The principle of SVM is based on the statistical learning theory and the method of structural risk minimization (Haykin, 2009). A brief description of the SVM regression theory is provided in this section, and for more details readers are referred to Vapnik (2000). A comprehensive review on applications of SVM in hydrological research can be found in Raghavendra and Deka (2014).

Consider a set of data $\{x_i, y_i\}_{i=1}^{n}$, where x_i are the input vectors and y_i is the target vector (i.e. observations). $x_i \subset R^n$ and $y_i \subset R$ for $i = 1,\ldots, n$ where n refers to the size of the data set. The generic SVM regression function can be expressed as Eq. (4) where w is the weight vector and b is the bias. A non-linear transformation function Φ (.) is defined to map the input space into a higher dimensional feature space.

$$f(x_i) = w.\, \Phi(x_i) + b \tag{4}$$

The aim of SVM regression is to find the values of w and b such that values of $f(x)$ can be determined by minimizing the empirical risk $R_{empirical}$ as given in Eq. (5).

$$R_{empirical} = \frac{1}{n}\sum_{i=1}^{n} \left|(y_i - f(x_i))\right| \tag{5}$$

In Eq. (5), $|(y_i - f(x_i))|$ is defined as Vapnik's ε - insensitive loss function as shown in Eq. (6).

$$|(y_i - f(x_i))| = \begin{cases} 0 & \text{if } |(y_i - f(x_i))| \le \varepsilon \\ |(y_i - f(x_i))| - \varepsilon & \text{otherwise} \end{cases} \tag{6}$$

The parameters w and b are estimated by minimizing the cost function ψ_ε shown in Eq. (7), where ξ_i and ξ_i^* are slack variables and C is a positive real constant.

$$\psi_\varepsilon = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{7}$$

Eq. (7) is subject to constraints given below for all $i = 1,\ldots,n$

$$y_i - f(x_i) \le \varepsilon + \xi_i$$

$$-y_i + f(x_i) \le \varepsilon + \xi_i^*$$

$$\xi_i \ge 0$$

$$\xi_i^* \ge 0$$

w in Eq. (7) can be expressed as Eq. (8) where α_i and α_i^* are Lagrange multipliers, which are positive real constants.

$$w = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\Phi(x_i) \tag{8}$$

By substituting Eq. (8) into Eq. (4), the generic equation can be rewritten as Eq. (9).

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)(\Phi(x_i).\, \Phi(x_i)) + b \tag{9}$$

The dot product in Eq. (9) can be replaced with function $k(x, x_i)$, known as the kernel function and can be.

expressed as Eq. (10).

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) k(x, x_i) + b \tag{10}$$

Different kernels such as Linear, Polynomial, Sigmoid, Spline and Radial Basis Function (RBF) can be used in Eq. (10). However, RBF shown in Eq. (11) where σ is the width of the kernel is identified as the most widely used kernel in the past downscaling studies (Ghosh and Mujumdar, 2008; Sachindra et al., 2013). A detailed description on different kernels can be found in Hofmann et al. (2008).

$$k(x, x_i) = exp\left[-\frac{1}{2\sigma^2} \|x - x_i\|^2\right] \tag{11}$$

3.1.4. Relevance vector machine (RVM)

RVM can be regarded as a variant of SVM as it is based on the same statistical learning framework used in SVM. However, RVM has an alternative functional algorithm to allow for probabilistic regression unlike its predecessor SVM (Tipping, 2001). A brief description of the RVM theory is provided in this section, and for more details readers are referred to Tipping (2001). Applications of the RVM technique in hydroclimatology can be found in Okkan and Inan (2015a), Joshi et al. (2015) and Deo et al. (2016).

Consider a set of data $\{x_i, y_i\}_{i=1}^{n}$, where x_i are the input vectors and y_i is the target vector. $y(x)$ which is the model output can be express as Eq. (12), where w_i represents the weight vector, $k(x, x_i)$ represents a kernel function and ε_n is the bias which is assumed to be Gaussian with a zero mean and a variance of σ^2.

$$y(x) = \sum_{i=1}^{n} w_i k(x, x_i) + \varepsilon_n \tag{12}$$

The likelihood of the dataset can be expressed in the form of Eq. (13), where $\Phi = [1, k(x_i,x_1), k(x_i,x_2), \ldots, k(x_i,x_n)]^T$ is a matrix $n\ x\ (n+1)$ in size, with $\Phi_{nm} = k(x_n, x_{m-1})$ and $\Phi_{n1} = 1$.

$$p(y \mid w, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} exp\left[-\frac{1}{2\sigma^2} \|y - \Phi\|^2\right] \tag{13}$$

The maximum likelihood estimation of w and σ^2 in Eq. (13), may result in model over-fitting in training. As a solution to this issue, imposition of some prior constraints on the parameter w by adding complexity to the likelihood or error function is recommended by Tipping (2001). This prior information governs the generalization ability of the learning process. As expressed in Eq. (14), an explicit prior probability distribution over the weights to improve the generalization ability of RVM model is used. In Eq. (14), α is a vector of $(n+1)$ hyperparameters, and it controls how far from zero each weight is allowed to stray.

$$p(w \mid \alpha) = \prod_{i=0}^{n} N((w_i \mid 0, a_i^{-1})) \tag{14}$$

By using Bayes' rule, the posterior over all unknowns can be obtained given the defined non-informative prior distribution as in Eq. (15) (Tipping, 2001; Ghosh and Mujumdar, 2008).

$$p(w, \alpha, \sigma^2 \mid y) = \frac{p(y \mid w, \alpha, \sigma^2).\ p(w, \alpha, \sigma^2)}{\int p(y \mid w, \alpha, \sigma^2) p(w, \alpha, \sigma^2) dw d\alpha d\sigma^2} \tag{15}$$

Direct resolution of Eq. (15) is impossible as the integral in the denominator of right hand side of Eq. (15) cannot be solved. As a solution to the above issue, posterior is decomposed as given in Eq. (16).

$$p(w, \alpha, \sigma^2 \mid y) = p(w \mid y, \alpha, \sigma^2).\ p(\alpha, \sigma^2 \mid y) \tag{16}$$

The posterior distribution of the weights can be expressed as Eq. (17), where Σ and μ denote posterior covariance and mean, respectively.

$$p(w \mid y, \alpha, \sigma^2) = 2\pi^{-(n+1)/2}\ |\Sigma|^{-1/2}.\ exp\left\{-\frac{1}{2}(w - \mu)^T.\ \Sigma^{-1}(w - \mu)\right\} \tag{17}$$

The posterior covariance and mean are expressed in Eqs. (18) and (19) respectively, where $A = diag(\alpha_0, \alpha_1 \ldots \alpha_n)$.

$$\Sigma = (\sigma^{-2}\Phi^T\Phi + A)^{-1} \tag{18}$$

$$\mu = \sigma^{-2}\Sigma\Phi^T y \tag{19}$$

For uniform hyperpriors, the maximization of $p(y|\alpha, \sigma^2)$ is required, and it is achieved through Eq. (20).

$$p(y \mid \alpha, \sigma^2) = \int p(y \mid w, \sigma^2) p(w \mid \alpha) dw \tag{20}$$

3.2. Application

This section presents the application of the machine learning techniques detailed in Section 3.1 to develop downscaling models for the 48 precipitation stations.

3.2.1. Defining an atmospheric domain and selection of predictors

In any statistical downscaling study, definition of the atmospheric domain (region for which large scale atmospheric information is obtained as inputs to downscaling models) and the selection of predictors (inputs to downscaling models) are regarded as key tasks. In this study, an atmospheric domain with 9 × 8 grid points in the longitudinal and latitudinal directions respectively, spaced at an interval of 2.5° was selected. The atmospheric domain was selected in such a way that it adequately captures the movement of the sub-tropical ridge, north-west cloud bands, cutoff-low, and the east coast lows, which are the main large scale atmospheric phenomena influential on the precipitation over the study area (Bureau of Meteorology, 2017). The atmospheric domain used in this study is shown in Fig. 5.

A set of probable predictors which is common for all stations and all calendar months was selected based on the past studies by Timbal et al. (2009) and Sachindra et al. (2013, 2014b, 2016) on downscaling of large scale atmospheric information to catchment scale hydroclimatic information (i.e. precipitation, evaporation, temperature and streamflow) over Victoria. This set of probable predictors contained; air temperature at surface, air temperature at 1000, 925, 850, 700, 600, 500, 400, 300 hPa pressure levels, geopotential heights at 1000, 925, 850, 700, 600, 500, 400, 300 hPa pressure levels, relative and specific humidity at surface, relative and specific humidity at 1000, 925, 850, 700, 600, 500, 400, 300 hPa pressure levels, zonal and meridional wind speeds at 1000, 925, 850, 700, 600, 500, 400, 300 hPa pressure levels, sea level pressure, pressure at surface and precipitable water content.

Potential predictors are the most influential predictors on the predictand of consideration, and they are a subset of probable predictors (Anandhi et al., 2008). A set of potential predictors was separately identified for each station for each calendar month, from the common pool of probable predictors by employing the following procedure. Initially, the Pearson correlations between each probable predictor data set obtained from the NCEP/NCAR reanalysis database for each grid point in the atmospheric domain (see Fig. 5) and the observations of precipitation at each station were determined for the period 1950–2014. Then, for each station, for each calendar month, 6 probable predictors which displayed statistically significant high correlations at the 95% confidence level were selected as potential predictors.

Sachindra et al. (2013, 2014a, 2014b, 2016) detail some downscaling studies conducted over Victoria, Australia. In those studies, it was found that 3–12 potential predictors per calendar month are adequate in developing a downscaling model. The selection of too few variables can cause the failure of the model to adequately explain the predictor–predictand relationships and the selection of too many variables can introduce redundant information to the model (Sachindra

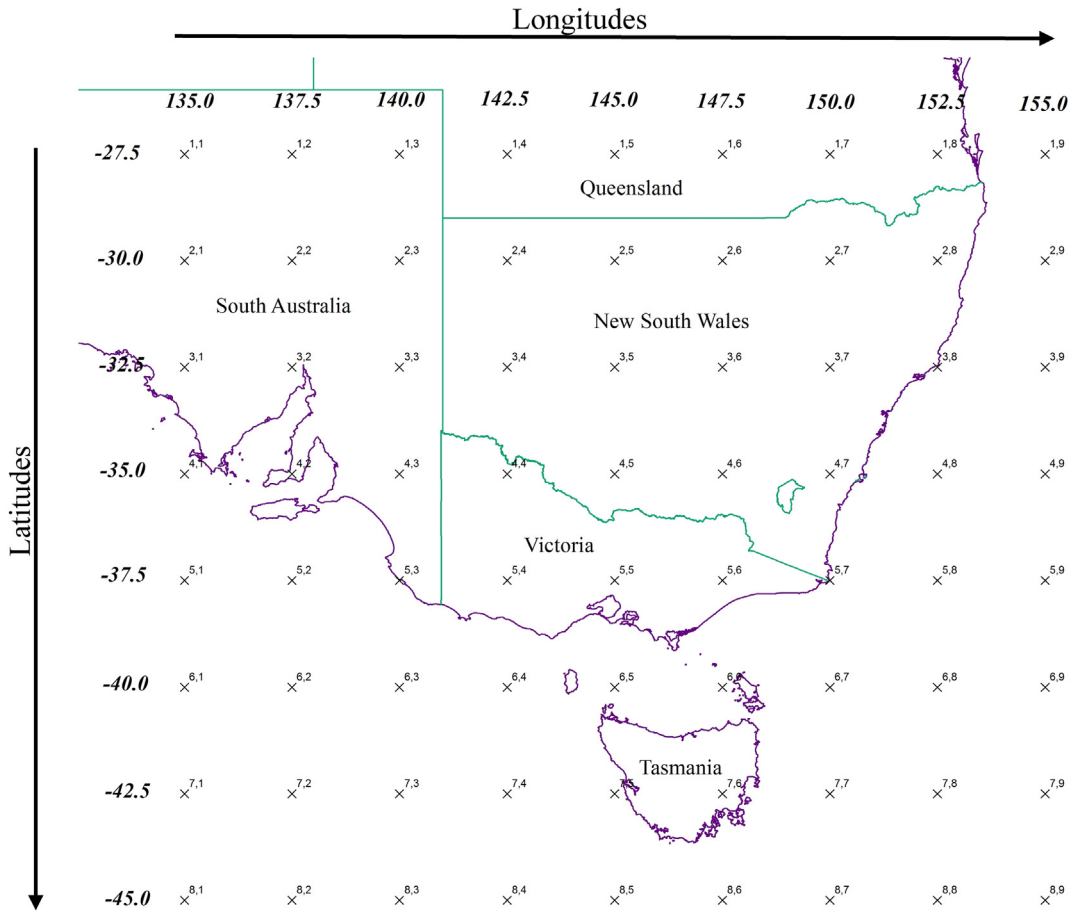

Fig. 5. Atmospheric domain used in this study.

Table 3
Details of main attributes of Genetic Programming Algorithm used this study.

GP attribute	Description
Training and testing data %	Training 65% and testing 35%
Population size	500 members per generation
Program size/tree size	The maximum size of a model (maximum tree depth) = 6
Terminals	Maximum number of inputs = 12, maximum number of constants in a model = 6
Mathematical Function set	+, −, x, ÷, √, x^2, x^3, sin, cos, e^x, and ln
Initial population generation	Ramped half-half initialization
Fitness measure	Root mean square error (RMSE)
Selection criterion for mating pool	Fitness proportionate selection/roulette wheel selection
Mutation probability	0.05
Crossover probability	0.9
Replication probability	0.2
Stopping/termination criterion	Minimize RMSE

et al., 2016). Therefore, in this study 6 potential predictors were used as inputs to the downscaling models for each calendar month for each station.

3.2.2. Development of downscaling models

For the development of downscaling models, initially the observed precipitation and NCEP/NCAR reanalysis data pertaining to the 6 potential predictors for each calendar month for each station were split into two sets in such way that the first 2/3 of data (period 1950–1991) for the model calibration and the rest of the data sets (period 1992–2014) for model validation. For each station, reanalysis data of the 6 potential predictors were standardised using their calendar monthly means and standard deviations (by subtracting the mean and normalizing by the standard deviation) pertaining to the model calibration period. In the current study, total of 576 monthly downscaling models (48 stations × 12 months) were developed using each machine learning technique (i.e. GP, ANN, SVM and RVM). The selection of potential predictors and development of downscaling models for each

Table 4
Percentage of selection of a kernel as the best for a given climate regime.

Machine learning technique	Climate regime	Kernel								
		Hyperbolic tangent	Polynomial	RBF	Laplacian	Bessel	ANOVA	Spline	String	Linear
SVM	Wet	0.0	34.4	13.0	25.0	18.8	8.9	0.0	0.0	0.0
	Intermediate	0.0	42.2	13.0	19.3	16.7	8.9	0.0	0.0	0.0
	Dry	0.0	39.1	10.4	25.0	14.6	10.9	0.0	0.0	0.0
RVM	Wet	1.0	33.9	13.0	25.0	18.2	8.9	0.0	0.0	0.0
	Intermediate	1.6	39.6	13.0	19.8	16.1	9.9	0.0	0.0	0.0
	Dry	0.5	38.0	10.4	25.5	15.1	9.9	0.5	0.0	0.0

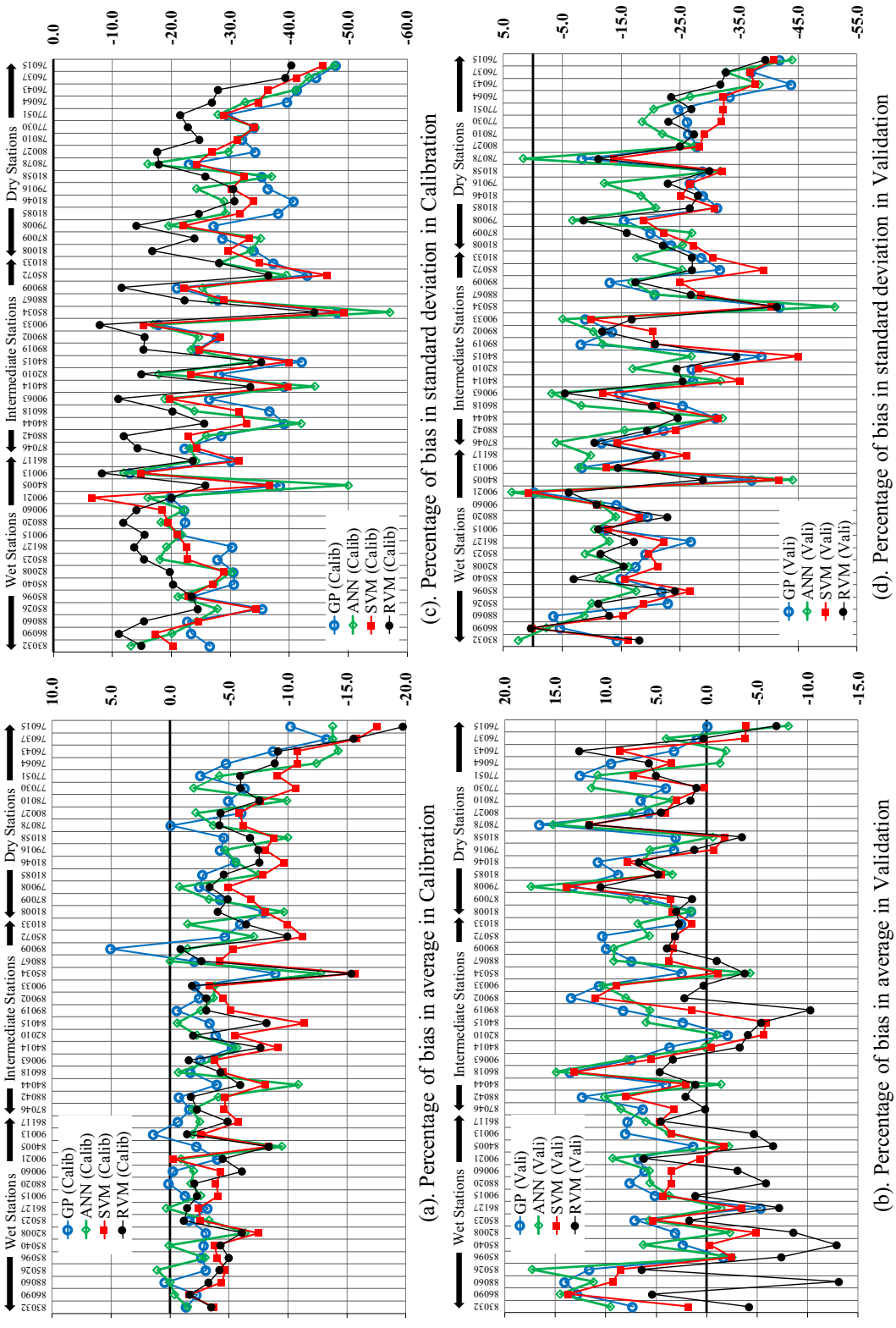

Fig. 6. Percentage of bias in the average and standard deviation of precipitation simulated by downscaling models.

calendar month is regarded as a better approach in comparison to selection of potential predictors and development of downscaling models for each season (i.e. summer, autumn, winter and spring) (Sachindra et al., 2018). This is because unlike the latter approach, the former approach enables capturing intra-seasonal variations in the predictor-predictand relationships.

3.2.2.1. GP based downscaling models. Once the 6 potential predictors for each station for each calendar month were identified and their NCEP/NCAR reanalysis data were standardised, the attributes of the GP algorithm were defined as shown in Table 3. The population size was limited to 500 (Havlíček et al., 2013) and the maximum size of a member/tree was limited to 6 in order to reduce any chance of model over-fitting in calibration and model under-fitting in validation due to

evolution of overly large models. Also, the limitation of member/tree depth to a too small value may cause the duplication of models in the initial population which reduces the genetic diversity in the initial population. The mathematical function set was selected in such way that it can be used to evolve a linear or a non-linear model (Havlíček et al., 2013) which is not overly complex (e.g. x^4, x^5 functions were avoided as they can lead to the evolution of overly complex models). Overly complex models can also cause over-fitting in calibration and under-fitting in validation. For the generation of the initial population the ramped half-half initialization proposed by Koza (1992) was selected, as it is known to produce a wide variety of models. As the fitness measure in the GP algorithm the root mean square error (RMSE) was used (Coulibaly, 2004). Following the usual practice the mutation probability was kept at a low value (Machado et al., 2015). Mutations introduce some new genetic information to the pool of models and in general it is kept at a minimum, as large mutation probabilities may delay the convergence of the GP algorithm. The cross-over probability was kept at a relatively high value (Selle and Muttil, 2011), and cross-over aids in recombining the models in the pool. The replication probability was kept at a low value as copying models too often from one generation to another may hinder the evolution of the GP algorithm. In this study, for each calendar month for each station, a GP-based downscaling model was calibrated (evolved) over the period 1950–1991 and validated for the period 1992–2014, using the observed precipitation and NCEP/NCAR reanalysis data pertaining to the 6 potential predictors with the GP attributes shown in Table 3.

3.2.2.2. ANN based downscaling models. In this study, the multilayer feed forward neural network (FFNN) algorithm (Coulibaly et al., 2000) was used to develop downscaling models for each calendar month for each precipitation station. In the current investigation, one hidden layer was used in the ANN architecture, as in the most of the previous studies in hydroclimatology one or two hidden layers have been successfully used (Govindaraju and Rao, 2013). The Levenberg-Marquardt Algorithm (LMA) was employed in training the ANN-based downscaling models as it enables faster converging of ANN (Hagan et al., 1996). In this study, the hyperbolic tangent sigmoid function (refer to Eq. (2) was used as the activation function at the neurons on the hidden and the output layers. The hyperbolic tangent sigmoid activation function has been widely used in the past studies for building ANN-based regression models (Mekanik et al., 2013; Deo and Şahin, 2015).

The optimum number of neurons in the hidden layer of the ANN-based downscaling models for each calendar month, for each station, was identified using a trial and error procedure. In this trial and error procedure, initially, an ANN-based downscaling model with only one hidden neuron was trained over the period 1950–1991, and its performance was quantified in terms of RMSE. Then, the number of hidden neurons was increased by one in each model run, up to twenty, and in each such model run the performance of the ANN-based downscaling model was determined in terms of RMSE. The number of hidden neurons in the ANN-based downscaling model which yielded the minimum RMSE was identified as the optimum number of hidden neurons for a given calendar month at a particular station. The above procedure enabled the determination of the optimum architecture for the ANN-based downscaling models for each calendar month for each station. Once the ANN-based downscaling models were calibrated over the period 1950–1991, using the observed monthly precipitation and NCEP/NCAR reanalysis data of the 6 potential predictors, they were validated over the period 1992–2014.

3.2.2.3. SVM and RVM based downscaling models. In general, the performance of SVM and RVM-based downscaling models are improved by optimizing the values of the tuning/hyper parameters in their algorithms (Tripathi et al., 2006; Ghosh, 2010). For this purpose, several optimization techniques such as genetic algorithm (Huang and Wang, 2006), probabilistic global search algorithm (Ghosh, 2010), simplex algorithm with leave-one-out cross-validation (Sachindra et al., 2013), Bayesian framework (Campozano et al., 2016) and grid search method (Tripathi et al., 2006; Halik et al., 2015) have been used in the past studies.

The grid search method is regarded as one of the most widely used techniques for the optimization of the tuning parameters in both SVM and RVM algorithms (Raghavendra and Deka, 2014). Hence, in the present study, the grid search method was employed to optimize the values of the tuning parameters in the SVM and RVM algorithms. The tuning parameters of the SVM and RVM algorithms varied depending on the kernel functions selected. In the SVM and RVM algorithms, the kernel function maps the non-linear problem into a linear problem in a multidimensional space. The selection of a suitable kernel function is quite important in the application of SVM and RVM algorithms (Okkan et al., 2014). The performance of a model is highly dependent on the kernel function (Erdal and Karakurt, 2013). In this investigation; Radial Basis Function (RBF), Polynomial, Linear, Hyperbolic tangent, Laplacian, Bessel, ANOVA, Spline and String kernel functions were tested with the SVM and RVM algorithms, in order to identify the best kernel function for each calendar month for each station. The objective of the use of several kernel functions was to identify the most suitable kernel for a particular climatic regime (e.g. relatively wet and relatively dry). In the past literature, the investigation of the suitability of few different kernel functions was seen in several downscaling exercises (Ghosh and Mujumdar, 2008; Kouhestani et al., 2016). However, in the published literature details of the suitability of RBF, Polynomial, Linear, Hyperbolic tangent, Laplacian, Bessel, ANOVA, Spline and String kernel functions with SVM and RVM algorithms in downscaling large scale atmospheric variables to precipitation under diverse climate have not been documented. In this study, for each calendar month, for each station, SVM and RVM-based downscaling models were calibrated with each of the above kernels by optimizing the tuning parameters employing the grid search method, over the period 1950–1991. The SVM and RVM-based downscaling models which displayed the smallest RMSE was selected for each calendar month for each station as the best model, and the kernel employed in the best model was identified as the best kernel. Once, the SVM and RVM-based downscaling models were calibrated, they were validated with the observed precipitation and NCEP/NCAR reanalysis data pertaining to the 6 potential predictors for each calendar month for each station over the period 1992–2014.

3.2.3. Post-processing of outputs of downscaling models and their performance assessment

Owing to outliers in the predictor data, statistical models on certain instances simulate outliers in their outputs (abnormally high/low values of predictands) (Sachindra et al., 2018). Another possible cause of outliers in the outputs of statistical models is the hyper-sensitivity of some of the predictor-predictand relationships to certain values of data of predictors. The hyper-sensitivity refers to simulation of very high/low values of the predictand by the predictor-predictand relationships for a certain range/s of values of predictor/s. As an example, the presence of an asymptotic region in a predictor-predictand relationship leads to high/low values of the predictand when it is introduced with values of data of predictor/s corresponding to that region. Since machine learning techniques tend to simulate highly complex non-linear predictor-predictand relationships for precipitation, there is a significant chance of having hyper-sensitive regions in the relationships, which could produce outliers.

It is the common practice to detect abnormally high/low values in a time series (outliers) by determining an interval spanning about the average plus/minus three times the standard deviation (Leys et al., 2013). In the current study, the outliers in the precipitation simulated by the downscaling models were corrected employing linear interpolation prior to model performance assessment. In a statistical modelling exercise there is a need to conduct a comprehensive assessment of

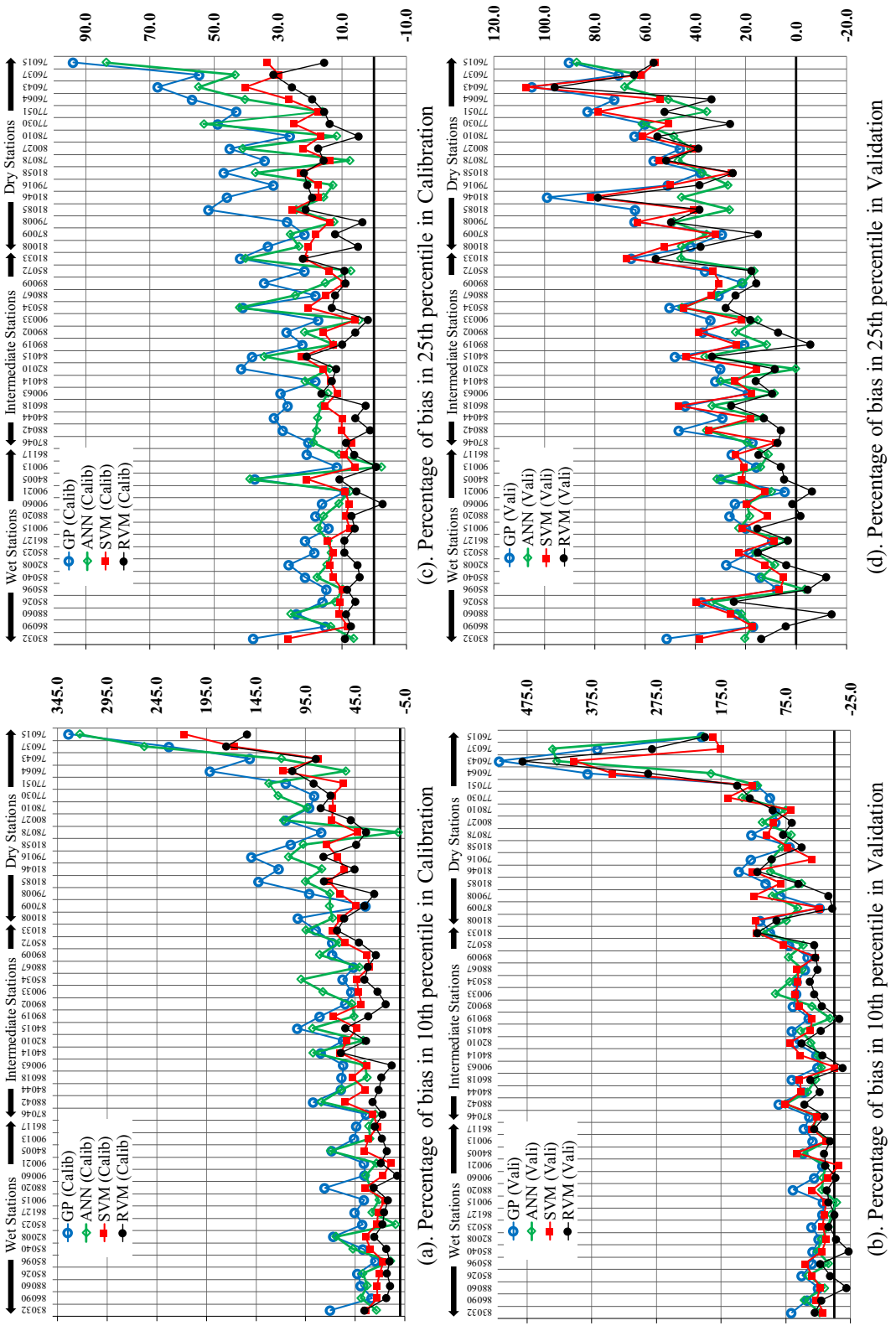

Fig. 7. Percentage of bias in the 10th and 25th percentiles of precipitation simulated by downscaling models.

model strengths and limitations rather than simply evaluating the overall model performance (Shortridge et al., 2016). In this study, the ability of downscaling models to simulate individual statistics was measured in terms of percentage of bias in the percentiles, average and standard deviation of simulated precipitation, for calibration and validation periods, separately. Also, the overall performance of the downscaling models was assessed in terms of normalised mean square error (NMSE) for calibration and validation periods separately. The use of percentages of bias in statistics and the NMSE enabled a fair comparison of downscaling models that are representative of different climate regimes.

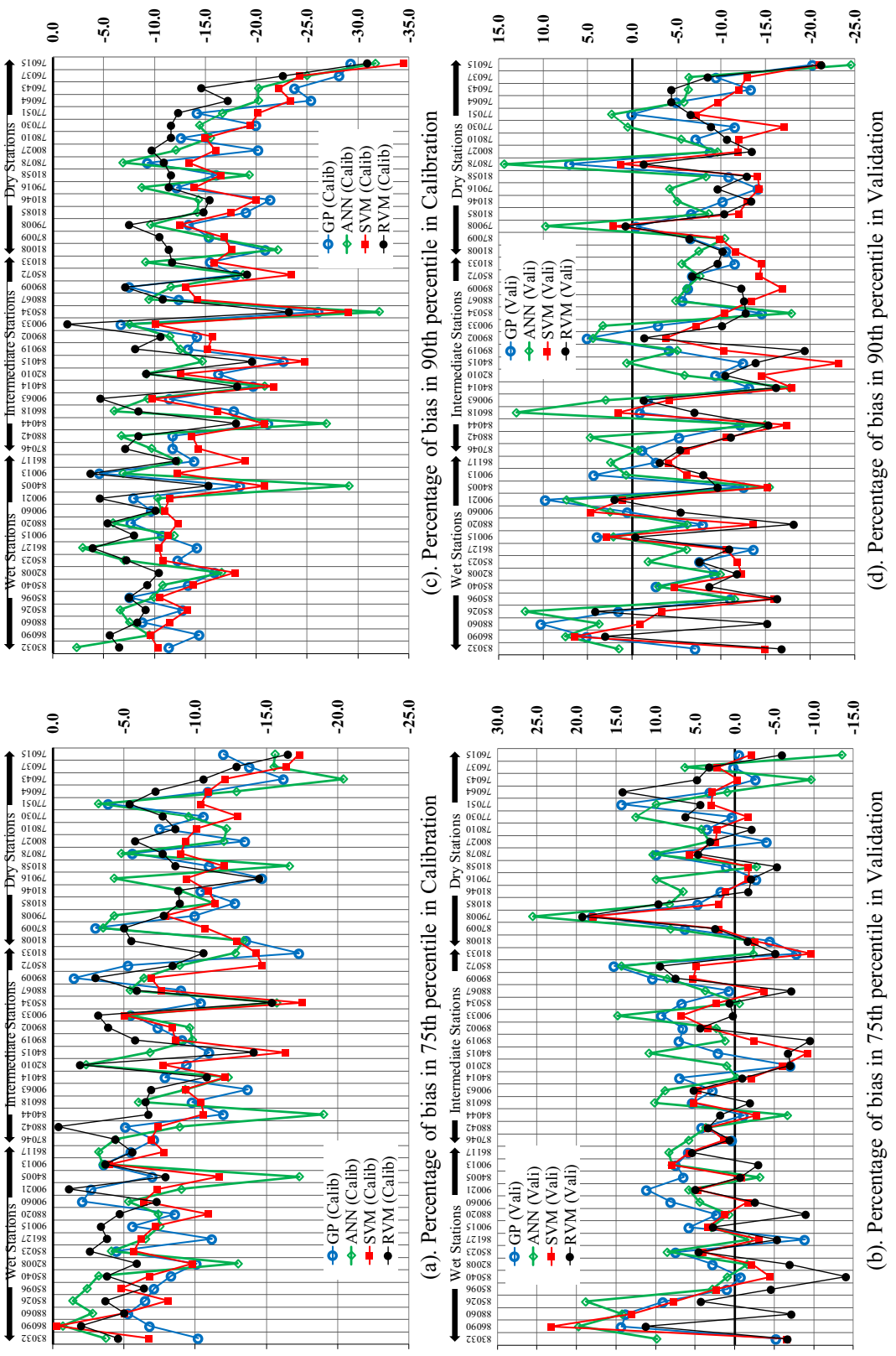

Fig. 8. Percentage of bias in the 75th and 90th percentiles of precipitation simulated by downscaling models.

4. Results and discussion

4.1. Results of selection of kernels for SVM and RVM-based downscaling models

It was observed that the kernels used in the SVM and RVM-based downscaling models which produced the best performance in calibration in terms of RMSE (the best kernel), varied from one calendar month to another, even at the same station. Table 4 shows the percentage of selection of a given kernel as the best in relatively wet, intermediate and relatively dry climate regimes for SVM and RVM. The percentage of selection of a kernel as the best for a given climate regime was calculated using Eq. (21).

$$\text{Percentage of selection of a kernel as the best for a given climate regime} = \frac{\text{(Total number of months a particular kernel was selected as the best for a climate regime)} \times 100}{\text{Number of stations in the climate regime (16)} \times \text{Number of calendar months (12)}} \quad (21)$$

According to Table 4, it was seen that irrespective of the climate regime both SVM and RVM showed the best performance with the Polynomial kernel followed by Laplacian, Bessel, RBF and ANOVA. Also, it was seen that Spline, String, Linear and Hyperbolic tangent kernels very rarely lead to the formation of the best downscaling model. The use of RBF kernel is widely seen in the applications of SVM and RVM (Roodposhti et al., 2017), however no comprehensive assessment of impacts of different kernels on precipitation downscaling has been conducted in the past, under diverse climate. Tripathi et al. (2006), Anandhi et al. (2008) and Halik et al. (2015) used RBF kernel with SVM in downscaling monthly precipitation and commented that it is computationally simpler in comparison to the Polynomial kernel. Also, Devak et al. (2015) used RBF kernel with LSSVM in a daily precipitation downscaling exercise, and Joshi et al. (2013) used RBF kernel with RVM in direct downscaling of low flow indices. In a downscaling study conducted over Iran, Kouhestani et al. (2016) commented that SVM with Polynomial kernel yields more accurate simulations of precipitation compared to that by SVM with RBF kernel, in terms of RMSE. Sarhadi et al. (2016) found that both SVM and RVM with RBF kernel yield models with low RMSEs in the model validation period. Similarly, in a precipitation downscaling exercise, Okkan and Inan (2015a) used RBF and Spline kernels with RVM and found that RVM with RBF kernel performs better in the model validation period in terms of RMSE. Tehrany et al. (2014) used linear, polynomial, RBF and sigmoid kernels with SVM in a flood susceptibility mapping study and commented that SVM with RBF kernel performs better compared to SVM with other kernels. In the current study, it was proven that, the Polynomial kernel can assist in developing a better performing downscaling model with both SVM and RVM in comparison to other kernels. According to the findings presented in Table 4, in a downscaling study it is recommended to investigate the suitability of Polynomial, Laplacian, Bessel and RBF kernels rather than the ANOVA, Spline, String, Linear and Hyperbolic tangent kernels. It is not possible to identify one single kernel as the best, as depending on the nature of the investigation the performance of kernels may vary (Han et al., 2007).

4.2. Results of statistical evaluation of performance of downscaling models

Once the downscaling models were calibrated and validated over the periods of 1950–1991 and 1992–2014 respectively, the percentage of bias in the statistics of precipitation simulated by the downscaling models was computed for the assessment of model performance. Climate change adaptation strategies such as the design of hydraulic structures (e.g. dams) for the management of water resources (e.g. flood mitigation), are partly based on the assessment of extreme hydroclimatological phenomena such as floods and droughts (Kao and Ganguly, 2011) and the statistics of hydroclimatic variables such as average and standard deviation. The performance measures such as RMSE, NMSE and Nash-Sutcliffe Efficiency (NSE) can provide an indication of the overall performance of a downscaling model. However, such measures cannot provide any idea of the ability of the model to simulate individual statistics such as average, standard deviation, low and high percentiles or the maximum of the predictand. Therefore, there is a need for the assessment of performance of downscaling models at the level of individual statistics, in addition to the assessment of overall performance. Hence, in this study the performance of downscaling models in simulating individual statistics of precipitation was quantified in terms of percentage of bias. The use of percentage of bias in statistics of precipitation enabled a fair comparison of ability of downscaling models in simulating statistics of precipitation, corresponding to observation stations located in significantly different regimes of climate.

The percentage of bias in the statistics of precipitation simulated by a downscaling model (for a station considering all 12 calendar monthly models) was calculated using Eq. (22) for the calibration and validation periods separately. According to Eq. (22), a positive bias percentage indicates an over-estimation of the statistic of interest and an under-estimation of the statistic of interest was indicated by a negative bias percentage.

$$\text{Percentage of bias in the statistic of interest} = \frac{\text{(model simulated value of statistic} - \text{corresponding observed value of statistic)} \times 100}{\text{corresponding observed value of statistic}} \quad (22)$$

Fig. 6 shows the percentages of bias in the average and standard deviation of precipitation simulated by downscaling models. In Fig. 6a and b, the percentages of bias in the average of precipitation downscaled by the models developed using different machine learning techniques are shown, pertaining to the calibration and validation periods respectively. As seen in Fig. 6a, it was understood that the percentage of bias in the average of model simulated precipitation was relatively larger at dry stations and relatively smaller at wet stations in calibration. Also, in calibration, GP-based models showed the smallest percentage of bias in the average of precipitation and SVM-based models showed the largest percentage of bias in the average of precipitation at the majority of the stations.

As seen in Fig. 6b, at several wet stations, in the validation period, RVM-based models showed large percentage of bias in the average of precipitation in comparison to that of models based on other machine learning techniques. However, at the majority of intermediate and dry stations RVM-based models showed smaller percentage of bias in the average of precipitation, during validation. The SVM-based models in general showed relatively small percentage of bias in the average of precipitation at the majority of stations during the validation period, particularly compared to that of GP-based downscaling models.

In Fig. 6c and d the percentages of bias in the standard deviation of precipitation simulated by downscaling models developed using different machine learning techniques are shown pertaining to the calibration and validation periods, respectively. According to Fig. 6c and d, it was realised that the models based on all machine learning techniques under-estimated the standard deviation at the majority of stations in both calibration and validation periods. The under-estimation of the standard deviation of precipitation was more pronounced at relatively dry stations in both calibration and validation. The under-estimation of the standard deviation of predictands was reported in many past downscaling studies (Tripathi et al., 2006; Anandhi et al., 2008; Sachindra et al., 2013), and it is due to the fact that the large-scale climate information does not contain the variance required for the proper explanation of observed variance of catchment scale hydroclimatic variables. During the calibration period, RVM-based models showed the lowest bias percentage for the standard deviation of

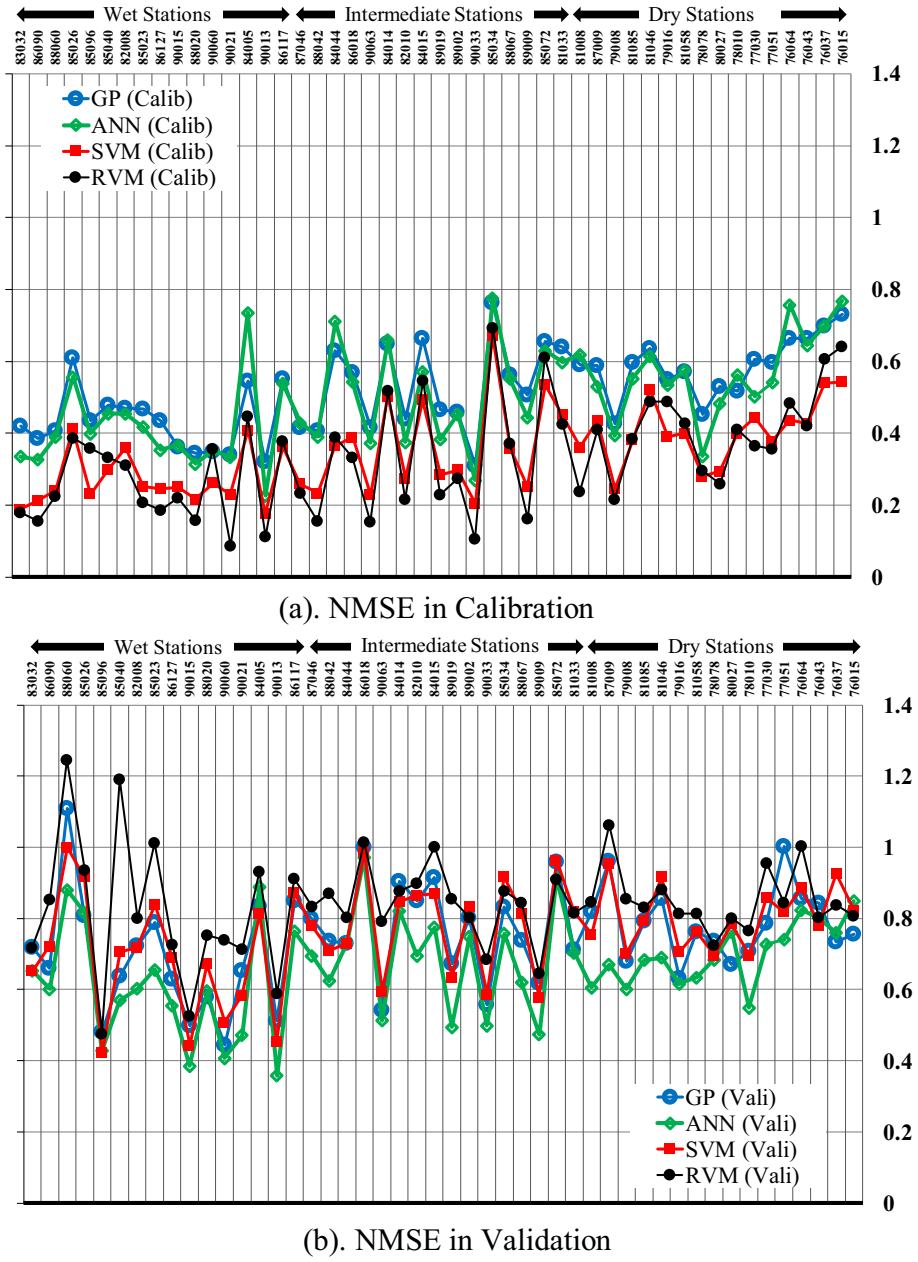

Fig. 9. Normalised mean square error of precipitation simulated by downscaling models.

precipitation at the majority of the stations. In validation, ANN-based models displayed the lowest bias percentage for the standard deviation of precipitation at the majority of the stations along with RVM-based models, while SVM-based models exhibited lager bias percentage for the standard deviation of precipitation along with GP-based models.

Fig. 7 shows the percentage of bias in the 10th and 25th percentiles of precipitation simulated by the downscaling models developed with different machine learning techniques.

According to Fig. 7, it was seen that in both calibration and validation at the majority of stations, the downscaling models developed using all machine learning techniques showed an over-estimating trend of the 10th and 25th percentiles of precipitation in both calibration and validation phases. Also, it was evident that the percentage of bias in the 10th and 25th percentiles of precipitation was significantly higher for dryer stations in both calibration and validation phases, irrespective of the machine learning technique used. Deo et al. (2017) used Multivariate Adaptive Regression Splines (MARS), LSSVM, and M5Tree models for forecasting droughts based on Standardised Precipitation Index (SPI) over eastern Australia. They also concluded that the performance of drought forecasting models developed with above machine learning techniques show significant dependence on the climate regime and the set of predictors. Belayneh et al. (2014) explored the potential of wavelet-ANN and wavelet-SVM based models in forecasting droughts based on SPI over the Awash River Basin in Ethiopia. They found that wavelet-ANN based drought forecasting models show better performance compared to that of wavelet-SVM based models in terms of both RMSE and mean absolute error (MAE). They commented that, the performances of models do not vary significantly depending on the geographic location of the weather stations. However, it should be noted that RMSE and MAE are sensitive to the order of magnitude of data at a station, unlike NMSE used in the current study. Hence, comparison of performances of models corresponding to stations located in different climate regimes using RMSE and MAE is unfair. Belayneh et al. (2016) found that the use of boosting ensemble technique with wavelet-ANN and wavelet-SVM can further improve the performance of drought forecasting models. Meyer et al. (2016) commented that in retrieving

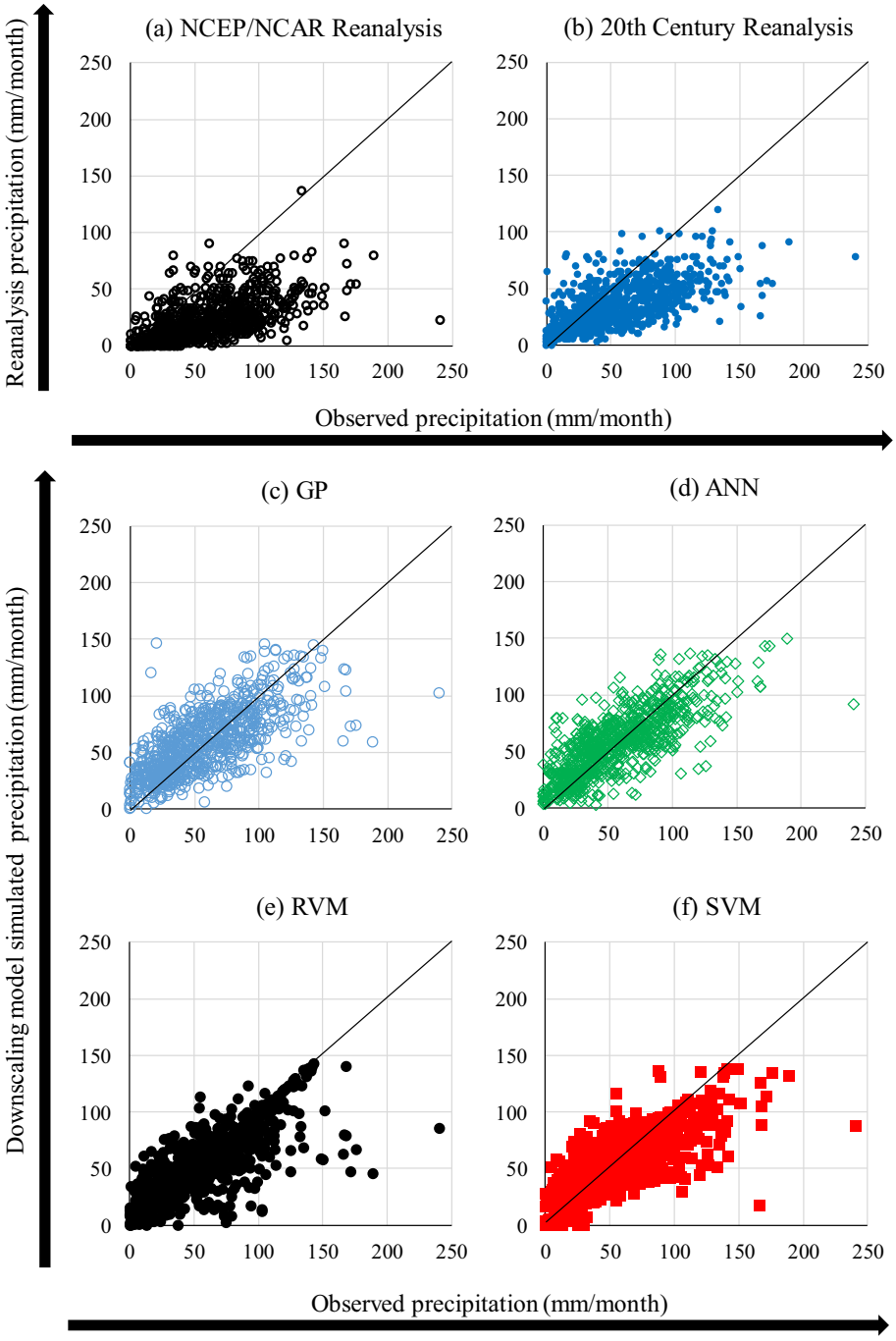

Fig. 10. Agreement between observed precipitation, reanalysis precipitation, and downscaled precipitation.

precipitation rate and determining precipitation area from satellite data ANN-based models performed better compared to SVM-based models, though the difference in performance was quite small. Furthermore, also in the current study, ANN-based downscaling models displayed smaller percentage of bias for the 10th and 25th percentiles compared to that of SVM-based models, in the validation period at the majority of the stations.

Although not shown, the 50th percentiles of precipitation simulated by the downscaling models developed using all machine learning techniques also displayed an over-estimating trend which was more pronounced at relatively dry stations. It was observed that, in the majority of the stations, the percentage of bias in the 10th percentile of precipitation > percentage of bias in the 25th percentile of precipitation > percentage of bias in the 50th percentile of precipitation, in both calibration and validation phases irrespective of the machine learning technique used in the development of the models.

Furthermore, a close examination revealed that at the majority of stations, in both calibration and validation, RVM-based downscaling models showed relatively small bias percentage for low percentiles of precipitation (i.e. 10th and 25th) in comparison to that of downscaling models developed with other machine learning techniques. GP-based models in particular displayed relatively higher bias percentage for low percentiles of precipitation in both calibration and validation. In studies such as drought analysis, accurate knowledge of the low percentiles of precipitation is a key requirement. According to the results of this investigation, it can be recommended to use RVM over GP, ANN or SVM in developing downscaling models for studies such as drought analysis.

Fig. 8 shows the percentage of bias in the 75th and the 90th

percentiles of precipitation simulated by the downscaling models developed using different machine learning techniques.

According to Fig. 8, it was observed that during the calibration period at all stations, models developed with all machine learning techniques showed an under-estimating trend of the 75th and the 90th percentiles of precipitation. The percentage of bias in the 75th and the 90th percentiles of precipitation was higher at relatively dryer stations during calibration (also the 95th percentile which is not shown). In validation, at the majority of the stations, the models developed with all machine learning techniques over-estimated the 75th percentile and under-estimated the 90th and the 95th percentiles. It was realised that the downscaling models developed with all machine learning techniques tend to under-estimate the high percentiles of precipitation such as the 90th and the 95th irrespective of the climate regime in which the station is located. Furthermore, it was observed that RVM-based downscaling models show better generalization skills in simulating higher percentiles of precipitation compared to that of downscaling models based on GP, ANN and SVM. However, GP and ANN-based downscaling models displayed relatively small percentage of bias in higher percentiles of precipitation in validation compared to that of SVM-based downscaling models.

Similar to standard deviation of model simulated precipitation, at the majority of the stations, models developed with all machine learning techniques showed an under-estimating trend for the maximum of observed precipitation in both calibration and validation periods (not shown). In a monthly streamflow forecasting study by Erdal and Karakurt (2013) it was seen that all models based on Classification And Regression Trees (CART), SVM, Bagged Regression Trees (BRT) and Gradient Boosted Regression Trees (GBRT) tended to underestimate the peaks of streamflow. In that study, lag 1, 2 and 3 streamflows were used as inputs to the forecasting models. In a daily streamflow forecasting study Bayesian Neural Network (BNN), SVM, Gaussian Process and MLR were employed by Rasouli et al. (2012). They commented that models based on all techniques irrespective of the lead time display difficulty in simulating extremes of streamflows defined as flows above the 95th percentile. In the current study, it was observed that RVM and ANN-based downscaling models consistently displayed relatively low bias percentage for the maximum of precipitation in both calibration and validation periods. Therefore, in a study such as flood prediction which involves the consideration of high extremes of precipitation, the use of RVM or ANN over SVM or GP for developing downscaling models can be recommended.

Fig. 9 shows the normalised mean square error (NMSE) for models developed with all machine learning techniques. The NMSE is used for the assessment of the overall performance of the downscaling models in the calibration and validation periods. In the calculation of the NMSE, the mean square error is normalised with the variance of the observed precipitation (Zhang and Govindaraju, 2000). This enables the intercomparison of performance of downscaling models developed for stations pertaining to different climate regimes (Anandhi et al., 2009).

According to Fig. 9, it was seen that in calibration, RVM-based downscaling models show the least NMSE, and GP and ANN-based models show relatively higher NMSE irrespective of the climate regime. In validation, the RVM-based downscaling models show relatively higher NMSE values particularly at several precipitation stations located in the wet regime. In a precipitation downscaling study over Turkey by Okkan and Inan (2015b) it was commented that RVM-based downscaling models show better performance compared to SVM-based downscaling models in terms of coefficient of determination (R^2), adjusted R^2 and RMSE, in the validation period. Kouhestani et al. (2016) investigated the potential of kernel Principal Component Analysis (PCA) as a pre-processing tool in developing SVM and RVM-based precipitation downscaling models over Iran. They also found that RVM-based downscaling models show better performance in terms of RMSE in validation compared to that of SVM-based models. In a study on modelling evaporative losses in reservoirs, Samui and Dixon (2012) commented that both SVM and RVM-based models show good generalization skills though RVM-based models are slightly better than the SVM-based models in terms of correlation coefficient. However, in the current investigation such tendency was seen only at a few stations in terms of NMSE. In an evaporative loss simulation study by Deo et al. (2016) it was concluded that RVM-based models show a higher correlation coefficient with observations and lower RMSE and MAE compared to those of models based on Extreme Learning Machine (ELM) and MARS. However, like in most of the other studies (e.g. Zhang et al., 2009; Ali Ghorbani et al., 2010; Yadav et al., 2016) it was seen that the differences in performance between models based on different machine learning techniques were marginal. Also, in that study models based on all machine learning techniques produced outliers, like in the case of the current investigation. In a streamflow forecasting study by Yaseen et al. (2016) models based on ELM, SVM and Generalized Regression Neural Network (GRNN) simulated outliers. In a daily precipitation downscaling study by Raje and Mujumdar (2011) which employed Conditional Random Field (CRF), K-Nearest Neighbour (KNN) and SVM techniques it was seen that in the validation phase SVM-based models did not simulate any outliers at any of the six observation stations. Furthermore, the number of outliers produced by the models varied depending on the observation station and the technique used in developing the downscaling models. In the current study, at several stations RVM and GP-based models simulated unphysically large outliers, nevertheless models based on other techniques also simulated outliers. It was understood that outliers in the outputs of downscaling models may have a profound impact on overall model performance and hence the conclusions of a study. Unlike in the above studies, in the current study the outliers in the outputs of downscaling models were corrected prior to the performance assessment.

In the current investigation, SVM, ANN and GP-based downscaling models displayed better generalisations skills at stations located in the wet regime in terms of NMSE, compared to that of RVM-based downscaling models. In general, NMSE was higher at dry stations compared to that at wet stations, in both calibration and validation.He et al. (2014) used ANN, Adaptive Neuro Fuzzy Inference System (ANFIS) and SVM for forecasting daily streamflows in a semi-arid mountainous region in China. In their study it was seen that, in validation, depending on the inputs the technique which led to the best forecasting model in terms of RMSE and NSE varied. However, in terms of Mean Absolute Relative Error (MARE), SVM-based forecasting models outperformed the models based on other techniques irrespective of the inputs used. Similarly, in a monthly streamflow forecasting study by Erdal and Karakurt (2013) it was observed that in the validation phase the performance of models based on CARTs and SVM were dependant on inputs as well as the model performance measure. This raises the argument that not only the inputs to a model but also how the model performance assessment is performed may influence the identification of a technique as the best. Witten et al. (2011) commented that there is no universally superior machine learning technique, which was also proven in the current investigation. It can be concluded that, a machine learning technique should be selected considering its ability to simulate statistics of high importance associated with the study of interest (e.g. flood/drought analysis), rather than just considering the overall performance of a model.

5. Caveats of the study

5.1. Differences in bias in the statistics of modelled precipitation in calibration and validation

It was seen that there are noticeable differences in the percentages of bias in the statistics of precipitation simulated by the downscaling models in the calibration and validation periods, at certain stations. It is natural for a downscaling model to display a relatively larger bias percentage in validation in comparison to that in calibration. This is

because during calibration, the model parameters are optimised (values of parameters are allowed to change freely) in order to achieve the best agreement between the model simulations and observations. On the other hand, in validation, the model is run with the optimum values of parameters derived in calibration.

Furthermore, the validation period of the downscaling models (1992–2014), was dominated by a significantly dry period called the "Millennium Drought" which spanned over 1997–2010. The "Millennium Drought" in Victoria is regarded as the most severe drought observed in the historic record (Verdon-Kidd and Kiem, 2009). It can be stated that, the inclusion of this significantly dry period which dominated the validation period of the downscaling models also partly contributed to the relatively higher degree of percentage of bias in the statistics of precipitation simulated by the downscaling models during validation. Also, Sachindra et al. (2013) developed models using MLR and LS-SVM techniques for a station located in the north western region of Victoria for downscaling NCEP/NCAR reanalysis outputs to monthly streamflows and commented that, in model validation over the period 1990–2010, the downscaling models tended to over-estimate the average of streamflow. In that study, it was further commented that even when the downscaling models were calibrated over the period 1990–2010, they tended to over-estimate the average of streamflow. Finally, they concluded that the NCEP/NCAR reanalysis data pertaining to the period 1990–2010 does not contain some vital information required for the proper simulation of streamflow in the north-western region of Victoria during which the State of Victoria suffered the "Millennium Drought".

5.2. Agreement between observed precipitation, reanalysis precipitation, and downscaled precipitation

In order to investigate the agreement between observed precipitation and reanalysis precipitation, and agreement between observed precipitation and downscaled precipitation, a comparison was performed. Fig. 10 shows the scatter plots between (1) observed precipitation and raw reanalysis precipitation obtained from the NCEP/NCAR reanalysis database and the 20th Century Reanalysis (20CR) database, and (2) observed precipitation and precipitation simulated by downscaling models developed with GP, ANN, SVM and RVM, over the period 1950–2014. For the above comparison observation station at Mirranatwa (−37.4°, 142.4°) was selected as it is the closest observation station to the point at latitude −37.1° and longitude 142.5° for which the raw reanalysis precipitation data were extracted. It should be noted that the spatial resolution of the raw reanalysis precipitation data was 1.875° × 1.875° in longitudinal and latitudinal directions, respectively, and these data were not interpolated into the grid shown in Fig. 5, for the above comparison.

The scatter of raw precipitation data obtained from both NCEP/NCAR reanalysis and 20CR database shown in Fig. 10a and b indicated a significant underestimation of the majority of precipitation values, in comparison to the scatter of precipitation simulated by downscaling models developed with GP, ANN, RVM and SVM shown in Fig. 10c, d, e and f. This indicated that precipitation simulated by the downscaling models is far more accurate than the raw precipitation obtained from both NCEP/NCAR reanalysis and 20CR database.

5.3. Possible impacts of the use of different sets of reanalysis data in downscaling model development

According to Fig. 10a and b, it was realised that the nature of the scatter pertaining to NCEP/NCAR reanalysis raw precipitation was significantly similar to that of 20CR raw precipitation. This hinted that there is a good agreement between the raw precipitation data obtained from NCEP/NCAR and 20CR reanalysis database over the atmospheric domain considered in this study. Hence, it can be argued that there is a good agreement between the data of other variables (e.g. mean sea level pressure) of the NCEP/NCAR reanalysis and 20CR database over this atmospheric domain. Furthermore, Compo et al. (2011), showed that there is a very close agreement between the 300 hPa geopotential height data obtained from the NCEP/NCAR reanalysis and 20CR database over large regions spanning across extratropics which essentially includes the atmospheric domain considered in this study. Also, in that study it was shown that in the tropics the agreement between 300 hPa geopotential height data obtained from the NCEP/NCAR reanalysis and 20CR database is relatively low. Considering the above facts, it can be stated that the conclusions derived in this study are also valid for downscaling models developed with 20CR data obtained for a region where there is a good agreement between NCEP/NCAR reanalysis and 20CR data exists.

NCEP/NCAR and European Centre for Medium Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) are the most widely used reanalysis database in modern climate research (Brands et al., 2012). Brands et al. (2012) conducted a worldwide study on the sensitivity of downscaling exercises to NCEP/NCAR reanalysis data and ERA-40 data considering temperature, geopotential and humidity at 500 and 800 hPa pressure levels. In that study, it was found that, in general, there is a close agreement between the daily NCEP/NCAR reanalysis and ERA-40 data of temperature, geopotential and specific humidity over extratropics in terms of Pearson correlation coefficients. However, it was also revealed that, in general, in the tropics the agreement between NCEP/NCAR reanalysis and ERA-40 data is relatively low. A similar conclusion was derived in a study by Sterl (2004) where monthly NCEP/NCAR reanalysis and ERA-40 data of sea level pressure and 500 hPa geopotential height were compared. Also, Compo et al. (2011) provided a comparison between NCEP/NCAR and ERA-40 reanalysis data of 300 hPa geopotential height and commented that very high correlations between the two data sets exist in the extratropical regions. Considering the above facts, it can be stated that the conclusions derived in this study are also valid for downscaling models developed with ERA-40 data obtained for a region where there is a good agreement between data of NCEP/NCAR reanalysis and ERA-40. However, areas in the tropics such as northern region of South America, Central Africa and regions in extratropics such as Tibet where there is a poor correlation between the data of NCEP/NCAR and ERA-40 (Brands et al., 2012), the validity of the conclusions derived in this study are not guaranteed. As a guideline, for checking validity of the conclusions derived in this study for a different study area for NCEP/NCAR and ERA-40 reanalysis data, readers are encouraged to refer to Brands et al. (2012).

6. Conclusions

Following conclusions were drawn from this investigation;

Irrespective of the climate regime and the machine learning technique, in both calibration and validation at the majority of the stations downscaling models showed an over-estimating trend of low to mid percentiles (i.e. below the 50th percentile) of precipitation and under-estimating trend of high percentiles of precipitation (i.e. above the 90th percentile). The over-estimating trend of low to mid percentiles and under-estimating trend of high percentiles of precipitation was significantly more pronounced at stations located in dryer climate, irrespective of the machine learning technique employed. Also, irrespective of the climate regime and the machine learning technique, in both calibration and validation at the majority of the stations the standard deviation of precipitation was under-estimated by the downscaling models and this trend was more pronounced at station located in dry climate.

In this study, it was observed that RVM and ANN-based downscaling models display relatively low bias percentage for the maximum of precipitation in both calibration and validation periods. Therefore, in a study such as flood prediction which involves the consideration of high extremes of precipitation, the use of RVM or ANN over SVM or GP for

developing downscaling models is recommended. Also, irrespective of the climate regime, in both calibration and validation, RVM-based downscaling models showed relatively small bias percentage for low percentiles of precipitation (i.e. 10th and 25th) in comparison to that of downscaling models developed with other machine learning techniques. Therefore, it can be recommended to use RVM over GP, ANN or SVM in developing downscaling models for studies such as drought analysis.

Irrespective of the climate regime, the SVM and RVM-based precipitation downscaling models showed the best performance with the Polynomial kernel followed by Laplacian, Bessel, Radial basis and ANOVA kernels. Furthermore, it was seen that Spline, String, Linear and Hyperbolic tangent kernels very rarely lead to the formation of an effective precipitation downscaling model.

References

Agatonovic-Kustrin, S., Beresford, R., 2000. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. J. Pharmaceut. Biomed. 22, 717–727.

Ahmed, K., Shahid, S., Haroon, S.B., Xiao-Jun, W., 2015. Multilayer perceptron neural network for downscaling rainfall in arid region: a case study of Baluchistan, Pakistan. J. Earth Syst. Sci. 124, 1325–1341.

Ali Ghorbani, M., Khatibi, R., Aytek, A., Makarynskyy, O., Shiri, J., 2010. Sea water level forecasting using genetic programming and comparing the performance with Artificial Neural Networks. Comput. Geosci. 36, 620–627. http://dx.doi.org/10.1016/j.cageo.2009.09.014.

Anandhi, A., Srinivas, V.V., Nanjundiah, R.S., Kumar, D.N., 2008. Downscaling precipitation to river basin in India for IPCC SRES scenarios using support vector machine. Int. J. Climatol. 28, 401–420.

Anandhi, A., Srinivas, V.V., Nanjundiah, R.S., Kumar, D.N., 2009. Role of predictors in downscaling surface temperature to river basin in India for IPCC SRES scenarios using support vector machine. Int. J. Climatol. 29, 583–603.

Bates, B.C., Walker, K., Beare, S., Page, S., 2010. Incorporating Climate Change in Water Allocation Planning, Waterlines Report. National Water Commission, Canberra Available at. http://www.nwc.gov.au, Accessed date: 7 August 2017.

Beecham, S., Rashid, M., Chowdhury, R.K., 2014. Statistical downscaling of multi-site daily rainfall in a south Australian catchment using a generalized linear model. Int. J. Climatol. 34, 3654–3670.

Belayneh, A., Adamowski, J., Khalil, B., Ozga-Zielinski, B., 2014. Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural networks and wavelet support vector regression models. J. Hydrol. 508, 418–429. http://dx.doi.org/10.1016/j.jhydrol.2013.10.052.

Belayneh, A., Adamowski, J., Khalil, B., Quilty, J., 2016. Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction. Atmos. Res. 172–173, 37–47. http://dx.doi.org/10.1016/j.atmosres.2015.12.017.

Benestad, R., Hanssen-Bauer, I., Chen, D., 2008. Empirical-Statistical Downscaling. World Scientific Publishing Company, Singapore, pp. 228.

Brands, S.S., Gutiérrez, J.M., Herrera, S.S., Cofiño, A.S., 2012. On the use of reanalysis data for downscaling. J. Clim. 25, 2517–2526. http://dx.doi.org/10.1175/JCLI-D-11-00251.1.

Bureau of Meteorology, 2017. Available at. http://www.bom.gov.au/watl/about-weather-and-climate/australian-climate-influences.shtml?bookmark=introduction, Accessed date: 24 September 2017.

Campozano, L., Tenelanda, D., Sanchez, E., Samaniego, E., Feyen, J., 2016. Comparison of statistical downscaling methods for monthly total precipitation: case study for the Paute River basin in southern Ecuador. Adv. Meteorol. http://dx.doi.org/10.1155/2016/6526341.

Compo, G.P., Whitaker, J.S., Sardeshmukh, P.D., Matsui, N., Allan, R.J., Yin, X., Gleason, B.E., Vose, R.S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R.I., Grant, A.N., Groisman, P.Y., Jones, P.D., Kruk, M.C., Kruger, A.C., Marshall, G.J., Maugeri, M., Mok, H.Y., Nordli, Ø., Ross, T.F., Trigo, R.M., Wang, X.L., Woodruff, S.D., Worley, S.J., 2011. The twentieth century reanalysis project. Q. J. Roy. Meteorol. Soc. 137, 1–28. http://dx.doi.org/10.1002/qj.776.

Coulibaly, P., 2004. Downscaling daily extreme temperatures with genetic programming. Geophys. Res. Lett. 31, L16203.

Coulibaly, P., Anctil, F., Bobee, B., 2000. Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. J. Hydrol. 230, 244–257.

Deo, R.C., Şahin, M., 2015. Application of the artificial neural network model for prediction of monthly standardized precipitation and evapotranspiration index using hydrometeorological parameters and climate indices in eastern Australia. Atmos. Res. 161, 65–81.

Deo, R.C., Samui, P., Kim, D., 2016. Estimation of monthly evaporative loss using relevance vector machine, extreme learning machine and multivariate adaptive regression spline models. Stoch. Env. Res. Risk A. 30, 1769–1784. http://dx.doi.org/10.1007/s00477-015-1153-y.

Deo, R.C., Kisi, O., Singh, V.P., 2017. Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model. Atmos. Res. 184, 149–175. http://dx.doi.org/10.1016/j.atmosres.2016.10.004.

Devak, M., Dhanya, C.T., Gosain, A.K., 2015. Dynamic coupling of support vector machine and K-nearest neighbour for downscaling daily rainfall. J. Hydrol. 525, 286–301. http://dx.doi.org/10.1016/j.jhydrol.2015.03.051.

Duhan, D., Pandey, A., 2015. Statistical downscaling of temperature using three techniques in the Tons River basin in Central India. Theor. Appl. Climatol. 121, 605–622.

Erdal, H.I., Karakurt, O., 2013. Advancing monthly streamflow prediction accuracy of CART models using ensemble learning paradigms. J. Hydrol. 477, 119–128. http://dx.doi.org/10.1016/j.jhydrol.2012.11.015.

Fowler, H.J., Wilby, R.L., 2010. Detecting changes in seasonal precipitation extremes using regional climate model projections: implications for managing fluvial flood risk. Water Resour. Res. 46, W03525.

Ghosh, S., 2010. SVM-PGSL coupled approach for statistical downscaling to predict rainfall from GCM output. J. Geophys. Res. 115, D22102.

Ghosh, S., Mujumdar, P., 2008. Statistical downscaling of GCM simulations to streamflow using relevance vector machine. Adv. Water Resour. 31, 132–146.

Goly, A., Teegavarapu, R.S.V., Mondal, A., 2014. Development and evaluation of statistical downscaling models for monthly precipitation. Earth Interact. 18, 1–28.

Govindaraju, R.S., Rao, A.R., 2013. Artificial Neural Networks in Hydrology. Springer Science and Business Media.

Hagan, M.T., Demuth, H.B., Beale, M.H., De Jesús, O., 1996. Neural Network Design. PWS Publishing Company, Boston.

Halik, G., Anwar, N., Santosa, B., 2015. Reservoir inflow prediction under GCM scenario downscaled by wavelet transform and support vector machine hybrid models. Adv. Civil Eng. http://dx.doi.org/10.1155/2015/515376.

Ham, F.M., Kostanic, I., 2000. Principles of Neurocomputing for Science and Engineering. McGraw-Hill Higher Education.

Han, D., Chan, L., Zhu, N., 2007. Flood forecasting using support vector machines. J. Hydroinf. 9, 267. http://dx.doi.org/10.2166/hydro.2007.027.

Hashmi, M.Z., Shamseldin, A.Y., Melville, B.W., 2011. Statistical downscaling of watershed precipitation using Gene Expression Programming (GEP). Environ. Model Softw. 26, 1639–1646.

Havlíček, V., Hanel, M., Máca, P., Kuráž, M., Pech, P., 2013. Incorporating basic hydrological concepts into genetic programming for rainfall-runoff forecasting. Computing 95, 363–380.

Haykin, S.S., 2009. Neural Networks and Learning Machines. Prentice Hall/Pearson, New York.

He, Z., Wen, X., Liu, H., Du, J., 2014. A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. J. Hydrol. 509, 379–386. http://dx.doi.org/10.1016/j.jhydrol.2013.11.054.

Hofmann, T., Schölkopf, B., Smola, A.J., 2008. Kernel methods in machine learning. Ann. Stat. 36, 1171–1220.

Huang, C.L., Wang, C.J., 2006. A GA-based feature selection and parameters optimization for support vector machines. Expert Syst. Appl. 31, 231–240.

Iorio, J.P., Duffy, P.B., Govindasamy, B., Thompson, S.L., Khairoutdinov, M., Randall, D., 2004. Effects of model resolution and subgrid-scale physics on the simulation of precipitation in the continental United States. Clim. Dyn. 23, 243–258.

Joshi, D., St-Hilaire, A., Daigle, A., Ouarda, T.B.M.J., 2013. Databased comparison of sparse Bayesian learning and multiple linear regression for statistical downscaling of low flow indices. J. Hydrol. 488, 136–149. http://dx.doi.org/10.1016/j.jhydrol.2013.02.040.

Joshi, D., St-Hilaire, A., Ouarda, T.B.M.J., Daigle, A., 2015. Statistical downscaling of precipitation and temperature using sparse Bayesian learning, multiple linear regression and genetic programming frameworks. Can. Water Resour. J. 40, 392–408. http://dx.doi.org/10.1080/07011784.2015.1089191.

Kao, S.C., Ganguly, A.R., 2011. Intensity, duration, and frequency of precipitation extremes under 21st-century warming scenarios. J. Geophys. Res. 116, D16119.

Kim, T.W., Valdés, J.B., 2003. Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. J. Hydrol. Eng. 8, 319–328.

Kişi, Ö., 2008. Streamflow forecasting using neuro-wavelet technique. Hydrol. Process. 22, 4142–4152.

Kouhestani, S., Eslamian, S.S., Abedi-Koupai, J., Besalatpour, A.A., 2016. Projection of climate change impacts on precipitation using soft-computing techniques: a case study in Zayandeh-rud Basin, Iran. Glob. Planet. Chang. 144, 158–170. http://dx.doi.org/10.1016/j.gloplacha.2016.07.013.

Koza, J., 1992. Genetic Programming: On the Programming of Computers by Natural Selection. MIT Press, Cambridge.

Leys, C., Ley, C., Klein, O., Bernard, P., Licata, L., 2013. Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. J. Exp. Soc. Psychol. 49, 764–766.

Machado, P., Correia, J., Assunção, F., 2015. Graph-based evolutionary art. In: Handbook of Genetic Programming Applications. Springer International Publishing, pp. 3–36. http://dx.doi.org/10.1007/978-3-319-20883-1_1.

Mekanik, F., Imteaz, M.A., Gato-Trinidad, S., Elmahdi, A., 2013. Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes. J. Hydrol. 503, 11–21.

Meyer, H., Kühnlein, M., Appelhans, T., Nauss, T., 2016. Comparison of four machine learning algorithms for their applicability in satellite-based optical rainfall retrievals. Atmos. Res. 169, 424–433. http://dx.doi.org/10.1016/j.atmosres.2015.09.021.

Min, S.K., Cai, W., Whetton, P., 2013. Influence of climate variability on seasonal extremes over Australia. J. Geophys. Res. Atmos. 118, 643–654. http://dx.doi.org/10.1002/jgrd.50164.

Nourani, V., Alami, M.T., Aminfar, M.H., 2009. A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. Eng. Appl. Artif. Intell. 22, 466–472.

Okkan, U., Fistikoglu, O., 2014. Evaluating climate change effects on runoff by statistical downscaling and hydrological model GR2M. Theor. Appl. Climatol. 117, 343–361.
Okkan, U., Inan, G., 2014. Bayesian learning and relevance vector machines approach for downscaling of monthly precipitation. J. Hydrol. Eng. 20, 04014051.
Okkan, U., Inan, G., 2015a. Statistical downscaling of monthly reservoir inflows for Kemer watershed in Turkey: use of machine learning methods, multiple GCMs and emission scenarios. Int. J. Climatol. 35, 3274–3295. http://dx.doi.org/10.1002/joc.4206.
Okkan, U., Inan, G., 2015b. Bayesian learning and relevance vector machines approach for downscaling of monthly precipitation. J. Hydrol. Eng. 20, 04014051-1–04014051-13. http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0001024.
Okkan, U., Serbes, Z.A., Samui, P., 2014. Relevance vector machines approach for long-term flow prediction. Neural Comput. Appl. 25, 1393–1405. http://dx.doi.org/10.1007/s00521-014-1626-9.
Pascual, D., Pla, E., Lopez-Bustins, J.A., Retana, J., Terradas, J., 2015. Impacts of climate change on water resources in the Mediterranean Basin: a case study in Catalonia, Spain. Hydrol. Sci. J. 60, 2132–2147.
Raghavendra, N.S., Deka, P.C., 2014. Support vector machine applications in the field of hydrology: a review. Appl. Soft Comput. 19, 372–386.
Raje, D., Mujumdar, P.P., 2011. A comparison of three methods for downscaling daily precipitation in the Punjab region. Hydrol. Process. 25, 3575–3589. http://dx.doi.org/10.1002/hyp.8083.
Rashid, M.M., Beecham, S., Chowdhury, R.K., 2015. Statistical downscaling of rainfall: anon-stationary and multi-resolution approach. Theor. Appl. Climatol. 124, 919–933.
Rasouli, K., Hsieh, W.W., Cannon, A.J., 2012. Daily streamflow forecasting by machine learning methods with weather and climate inputs. J. Hydrol. 414, 284–293. http://dx.doi.org/10.1016/j.jhydrol.2011.10.039.
Roodposhti, M.S., Safarrad, T., Shahabi, H., 2017. Drought sensitivity mapping using two one-class support vector machine algorithms. Atmos. Res. 193, 73–82. http://dx.doi.org/10.1016/j.atmosres.2017.04.017.
Sachindra, D.A., Huang, F., Barton, A.F., Perera, B.J.C., 2013. Least square support vector and multi-linear regression for statistically downscaling general circulation model outputs to catchment streamflows. Int. J. Climatol. 33, 1087–1106.
Sachindra, D.A., Huang, F., Barton, A.F., Perera, B.J.C., 2014a. Multi-model ensemble approach for statistically downscaling general circulation model outputs to precipitation. Q. J. Roy. Meteorol. Soc. 140, 1161–1178.
Sachindra, D.A., Huang, F., Barton, A.F., Perera, B.J.C., 2014b. Statistical downscaling of general circulation model outputs to precipitation part 1: calibration and validation. Int. J. Climatol. 34, 3264–3281.
Sachindra, D.A., Ng, A.W.M., Muthukumaran, S., Perera, B.J.C., 2016. Impact of climate change on urban heat island effect and extreme temperatures: a case-study. Q. J. Roy. Meteorol. Soc. 142, 172–186.
Sachindra, D.A., Ahmed, K., Shahid, S., BJC, Perera, 2018. Cautionary note on the use of genetic programming in statistical downscaling. Int. J. Climatol (Accepted on 22/02/2018), Article in press. https://doi.org/10.1002/joc.5508.
Sahay, R.R., Srivastava, A., 2014. Predicting monsoon floods in rivers embedding wavelet transform, genetic algorithm and neural network. Water Resour. Manag. 28, 301–317.
Samui, P., Dixon, B., 2012. Application of support vector machine and relevance vector machine to determine evaporative losses in reservoirs. Hydrol. Process. 26, 1361–1369. http://dx.doi.org/10.1002/hyp.8278.
Sarhadi, A., Burn, D.H., Johnson, F., Mehrotra, R., Sharma, A., 2016. Water resources climate change projections using supervised nonlinear and multivariate soft computing techniques. J. Hydrol. 536, 119–132. http://dx.doi.org/10.1016/j.jhydrol.2016.02.040.
Selle, B., Muttil, N., 2011. Testing the structure of a hydrological model using genetic programming. J. Hydrol. 397, 1–9.
Shanmuganathan, S., Samarasinghe, S. (Eds.), 2016. Artificial Neural Network Modelling Series: Studies in Computational Intelligence. Springer International Publishing, Switzerland.
Shortridge, J.E., Guikema, S.D., Zaitchik, B.F., 2016. Machine learning methods for empirical streamflow simulation: a comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. Hydrol. Earth Syst. Sci. 20, 2611–2628. http://dx.doi.org/10.5194/hess-20-2611-2016.
Sterl, A., 2004. On the (in)homogeneity of reanalysis products. J. Clim. 17, 3866–3873.
Tehrany, M.S., Pradhan, B., Jebur, M.N., 2014. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. J. Hydrol. 512, 332–343. http://dx.doi.org/10.1016/j.jhydrol.2014.03.008.
Timbal, B., Fernandez, E., Li, Z., 2009. Generalization of a statistical downscaling model to provide local climate change projections for Australia. Environ. Model. Softw. 24, 341–358.
Tipping, M.E., 2001. Sparse Bayesian learning and the relevance vector machine. J. Mach. Learn. Res. 1, 211–244.
Tripathi, S., Srinivas, V.V., Nanjundiah, R.S., 2006. Downscaling of precipitation for climate change scenarios: a support vector machine approach. J. Hydrol. 330, 621–640.
Vapnik, V.N., 1998. Statistical Learning Theory. John Wiley and Sons, New York.
Vapnik, V.N., 2000. The Nature of Statistical Learning Theory. Springer-Verlag, New York.
Verdon-Kidd, D.C., Kiem, A.S., 2009. Nature and causes of protracted droughts in Southeast Australia: comparison between the federation, WWII, and big dry droughts. Geophys. Res. Lett. 36, L22707. http://dx.doi.org/10.1029/2009GL041067.
Vu, M.T., Aribarg, T., Supratid, S., Raghavan, S.V., Liong, S.Y., 2016. Statistical downscaling rainfall using artificial neural network: significantly wetter Bangkok? Theor. Appl. Climatol. 126, 453–467.
Whigham, P.A., Crapper, P.F., 2001. Modelling rainfall-runoff using genetic programming. Math. Comput. Model. 33, 707–721.
Wilby, R.L., Wigley, T.M.L., 1997. Downscaling general circulation model output: a review of methods and limitations. Prog. Phys. Geogr. 21, 530–548.
Wilby, R.L., Charles, S.P., Zorita, E., Timbal, B., Whetton, P., Mearns, L.O., 2004. Guidelines for Use of Climate Scenarios Developed from Statistical Downscaling Methods, Supporting Material to the IPCC. pp. 3–21. Available at. https://www.narccap.ucar.edu/, Accessed date: 1 September 2017.
Witten, I.H., Frank, E., Hall, M.A., 2011. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufman, Burlington, MA.
Yadav, B., Ch, S., Mathur, S., Adamowski, J., 2016. Discharge forecasting using an online sequential extreme learning machine (OS-ELM) model: a case study in Neckar River, Germany. Measurement 92, 433–445. http://dx.doi.org/10.1016/j.measurement.2016.06.042.
Yaseen, Z.M., Jaafar, O., Deo, R.C., Kisi, O., Adamowski, J., Quilty, J., El-Shafie, A., 2016. Stream-flow forecasting using extreme learning machines: a case study in a semi-arid region in Iraq. J. Hydrol. 542, 603–614. http://dx.doi.org/10.1016/j.jhydrol.2016.09.035.
Zhang, B., Govindaraju, R.S., 2000. Prediction of watershed runoff using bayesian concepts and modular neural network. Water Resour. Res. 36, 753–762.
Zhang, X., Srinivasan, R., Van Liew, M., 2009. Approximating SWAT model using artificial neural network and support vector machine. J. Am. Water Resour. 45, 460–474. http://dx.doi.org/10.1111/j.1752-1688.2009.00302.x.