

International Journal of Hydrology Science and Technology

ISSN online: 2042-7816 - ISSN print: 2042-7808

<https://www.inderscience.com/ijhst>

Nonlinear multiple regression analysis for predicting seasonal streamflow using climate indices for New South Wales

Rijwana Esha, Monzur A. Imteaz

DOI: [10.1504/IJHST.2024.10060429](https://doi.org/10.1504/IJHST.2024.10060429)

Article History:

Received:	23 September 2021
Last revised:	23 February 2022
Accepted:	11 August 2022
Published online:	01 December 2023

Nonlinear multiple regression analysis for predicting seasonal streamflow using climate indices for New South Wales

Rijwana Esha and Monzur A. Imteaz*

Department of Civil and Construction Engineering,
Swinburne University of Technology,
Melbourne, VIC 3122, Australia
Email: rijwana.esh@gmail.com
Email: mimteaz@swin.edu.au
*Corresponding author

Abstract: This paper presents development of streamflow prediction models with long-lead timescale using the Multiple Non-Linear Regression (MNL) technique. Four major climate indices which were found to be influencing the streamflow of New South Wales (NSW) are used for this purpose. The developed models with all the possible combinations show good results in terms of Pearson correlation(r), Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Willmott index of agreement (d). The outcomes of MNL models are compared to the best models of Multiple Linear Regression (MLR) analysis. MNL models are evident to outperform the MLR models in terms of Pearson correlation (r) values, confirming the non-linear relationship between seasonal streamflow and large-scale climate drivers. Though the correlation values are not very high, they are statistically significant. The correlations obtained varied from 0.38 to 0.53 during calibration period, while it improved during the validation period, ranging from 0.52 to 0.63.

Keywords: multiple nonlinear regression; MNL; multiple linear regression; MLR; climate indices; streamflow; seasonal forecast; New South Wales; NSW.

Reference to this paper should be made as follows: Esha, R. and Imteaz, M.A. (2024) 'Nonlinear multiple regression analysis for predicting seasonal streamflow using climate indices for New South Wales', *Int. J. Hydrology Science and Technology*, Vol. 17, No. 1, pp.101–116.

Biographical notes: Rijwana Esha completed her Bachelor of Engineering in the field of Water Resources Engineering from Bangladesh University of Engineering and Technology (BUET) in 2014. After few years of consulting services, in 2016, she commenced her Master of Engineering study at Swinburne University of Technology, Melbourne, Australia. Through her significant research progresses, in 2017, she was converted to a PhD student. She was awarded the PhD degree in 2020. Her PhD research was to forecast streamflow for the state of New South Wales (Australia) using long-term climate indices applying linear and nonlinear techniques.

Monzur A. Imteaz is an Associate Professor in the Department of Civil and Construction Engineering at Swinburne University of Technology, Melbourne, Australia. He has completed his PhD in 1997 on Lake Water Quality Modelling from Saitama University (Japan). After his PhD, he was working with the Institute of Water Modelling (Bangladesh). He was working with several Australian Universities and Government Departments. At Swinburne, he is the

Head of 'Urban Environmental and Transport Systems' group. He is the author of the book *Urban Water Resources* published by CRC Press, USA. He has been actively involved with various researches on sustainability, water recycling and modelling, developing decision support tools and rainfall forecasting using artificial neural networks.

This paper is a revised and expanded version of a paper entitled 'Nonlinear multiple regression analysis for predicting seasonal streamflow using large scale climate modes', presented at 22nd International Congress on Modelling and Simulation (MODSIM), Hobart, Australia, 3–8 December 2017.

1 Introduction

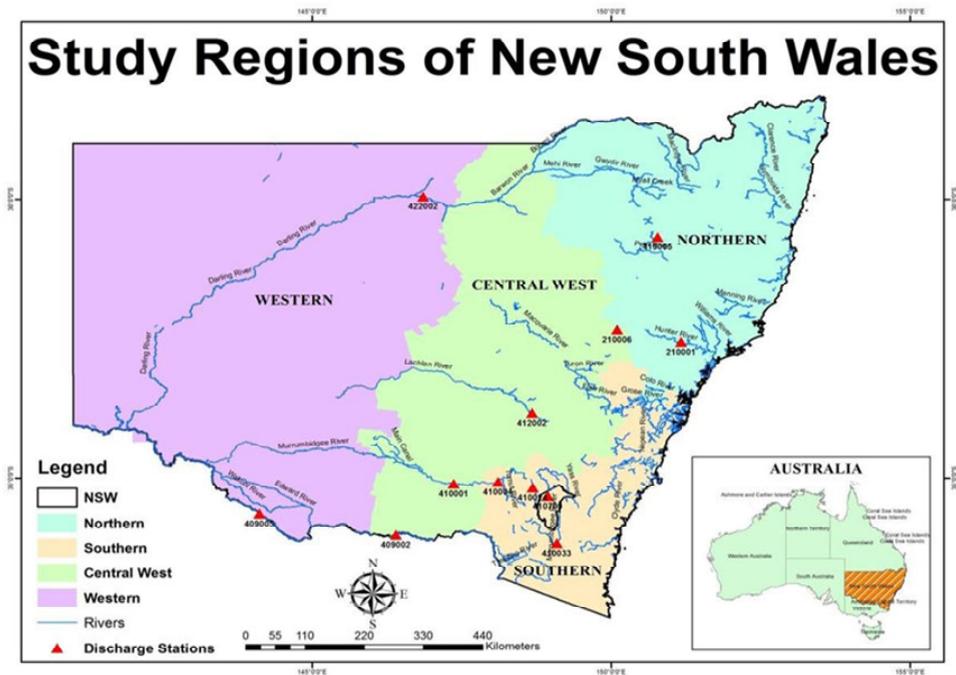
Australia's geographic location and extensive topographic variations present high climatic inconsistency, which results in significant inter-annual rainfall and subsequently streamflow variability across the country (Hossain et al., 2020a). Such inter-annual streamflow variability causes many difficulties to irrigators, agricultural producers, water managers and planners to allocate irrigation water and environmental flows, manage and operate reservoirs, supply municipal water, estimate future hydroelectricity supply, etc.

Australia is greatly influenced by climatic anomalies that originate in the surrounding Pacific, Indian and Southern Oceans. It is accepted by numerous hydrologists that there exists strong correlation between streamflow and large-scale atmospheric circulation patterns. The climate of southeast Australia is influenced by four major climate drivers (Mekanik and Imteaz, 2018) originating in the Pacific, Indian and Southern Oceans. The dominating climate drivers in these regions are the El Niño Southern Oscillation (ENSO), the Interdecadal Pacific Oscillation (IPO) or Pacific Decadal Oscillation (PDO), Southern Annular Mode (SAM) and Indian Ocean Dipole (IOD).

The ENSO phenomenon, which results from the large-scale interactions between ocean and atmospheric circulation processes in the equatorial Pacific Ocean, has direct influences on the climate variability over many parts of the world (Islam and Imteaz, 2020; Ghamariadyan and Imteaz, 2020; Rasel et al., 2016). El Nino and La Nina events are responsible for the different climatic conditions around the Pacific, including eastern Australia (Ghamariadyan and Imteaz, 2021; Hossain et al., 2018). Several studies revealed the influences of ENSO on streamflow throughout Australia (Piechota et al., 1998; Chiew et al., 1998; Dutta et al., 2006). Chiew et al. (1998) and Piechota et al. (1998) found that ENSO based (SOI and SST) streamflow predictions in northeast Australia are better than the forecasts from climatology. Robertson and Wang (2009) determined that the anomalies related to ENSO are the best predictors of seasonal streamflow, while the greatest predictability achieved between September and December. These findings were similar to the previous findings (McBride and Nicholls, 1983) which reported the strongest correlations between seasonal rainfall of NSW and ENSO during spring. The study of Mekanik et al. (2003) demonstrated that spring rainfall and runoff had high correlation with winter SOI throughout eastern Australia.

Some recent studies show that Eastern Australia is also influenced by IOD as well as interdecadal modulation of ENSO because of the low frequency variability in the Pacific Ocean, which is referred to as PDO (Westra and Sharma, 2009). Some researchers (e.g., Power et al., 1999; Kiem et al., 2003) have demonstrated the influence of IPO to be significant on rainfall and streamflow variation on a decadal to multidecadal timescale. King et al. (2013) suggested that the IPO played a significant role in the frequency of major floods during the 1950s, 1970s and 2010–2011. Verdon et al. (2004) explained that the enhanced rainfall and streamflow in eastern Australia were the consequence of the combined impact of ENSO (La Nina) and IPO negative phase. Duc et al. (2017) showed in their study that IPO alone does not have any significant impact on rainfall of NSW but its combination with ENSO can make a significant impact on rainfall.

Figure 1 Selected locations of the streamflow stations within NSW (see online version for colours)



According to Qi and Chang (2011), the existing forecasting methods can be categorised into five categories: time series analysis, regression analysis, artificial intelligence method (e.g., ANN, fuzzy logic, etc.), the hybrid and Monte Carlo simulation methods (Haque et al., 2013). Various nonlinear methods, such as ANFIS, ANN are widely applied in the field of hydro-climatology and water resources (Hossain et al., 2020b; Mekanik et al., 2016; Dastorani et al., 2010). Recently, artificial intelligence has attained popularity to forecast streamflow (Kumar et al., 2005; Kişi, 2007; Mutlu et al., 2008). However, dealing with artificial intelligence method is difficult, which encourages researchers to apply comparatively simple and straightforward statistical

methods such as regression models (Rezaeianzadeh et al., 2014). Data driven statistical models have gained popularity for their simplicity, accuracy, lower information requirement and fast pace in model development (Adamowski, 2008). However, the limitations of these models are evident when the data become complex. While forecasting streamflow, Mekanik et al. (2003) suggested that nonlinear regression methods can be applied to obtain higher correlations between streamflow and climate indicators, by capturing the nonlinear relationships between them. MNLR methods are usually applied for the accurate and fast prediction of random periodic events (Adamowski et al., 2012). In the study of Miyagishi et al. (1999) while forecasting temperature, MNLR models outperformed the radial bias functions and numerical weather forecast methods.

Esha and Imteaz (2019) have used MLR for forecasting seasonal streamflow in agriculture dominating regions of NSW. With the aim of improving the accuracy of those forecasting, this study has applied MNLR analysis for the same regions of NSW, which is a major agricultural state of Australia. Nonlinear regression analyses were carried out using cubic, quadratic and exponential functions, where multiple climate indices were used as predictors of spring streamflow of the study regions. As independent predictor variable all the relevant climate indices were explored. Eventually, results of MNLR analysis were compared with the MLR analysis reported earlier.

2 Materials and methods

2.1 Study area

NSW, which is situated in the east coast of Australia covering a land area of 8,800,642 km² is the most populous state of Australia with a population of 7.5 million. The state is bordered on the north by Queensland, on the west by South Australia, on the south by Victoria and on the east by the Tasman Sea. The two most important features of NSW are the Great Dividing Range (GDB) and Murray Darling Basin (MDB) which accounts for nearly 40% of the value of agricultural production in Australia and 65% of irrigated land (Abbot and Marohasy, 2015). The coastal regions, which are in the east of the state adjacent to the Tasman Sea, have a rainfall variation of around 800 mm to 3,000 mm. Rainfall is moderate (600 mm–1,500 mm) and evenly distributed throughout the year in the highlands which is a part of the GDB. The main agricultural region of NSW is the western inland slopes, which have a less dense population than coastal areas. This area receives high rainfall (600 mm) throughout the year. The western arid or semi-arid plains, which cover almost two-thirds of the state, experience an average rainfall of 150 mm to 500 mm in almost all the time of the year across the whole region. Considering the geographical location, regional climatic variation and the agricultural importance, NSW is divided into four regions; Northern NSW (NNSW), Southern NSW (SNSW), Central West NSW (CWNSW) and Western NSW (WNSW). For the current study, six locations from all the four regions were selected. Locations of selected streamflow gauge locations along with the map of NSW is shown in Figure 1. Details of the selected locations are provided in Table 1. Most of the selected stations are located along the eastern part of NSW due to the predominance of coastal rivers as well as the agricultural importance of this region.

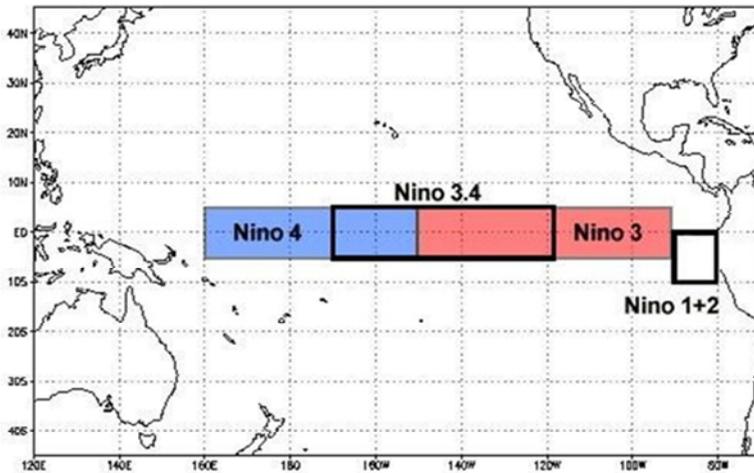
Table 1 Overview of the selected discharge stations

Station number	Latitude	Longitude	River name	Station name
210001	-32.56°S	151.17°E	HUNTER	SINGLETON
419005	-30.68°S	150.78°E	NAMOI	NORTH CUERINDI
410004	-35.07°S	148.11°E	MURRUMBIDGEE	GUNDAGAI
410700	-35.32°S	148.94°E	COTTER	KIOSK
410001	-35.10°S	147.37°E	MURRUMBIDGEE	WAGGA WAGGA
422002	-29.95°S	146.86°E	BARWON	BREWARRINA

2.2 Data sources

Streamflow data: The starting point for any statistical analysis should be the data collection; useful quality data are required for the development, calibration and validation of any model. Historical streamflow data was collected from the Australian Bureau of Meteorology website. Observed monthly streamflow in cumec (cubic metre per second) was collected for 102 years, (1914–2015) for all the stations, except for North Cuerindi, where data was available for 99 years. These stations have less than 0.5% missing values, which are filled by the series mean of the streamflow data. Using this data, seasonal mean discharge data is derived for the spring (September–October–November) season.

Figure 2 Map showing ENSO region (see online version for colours)



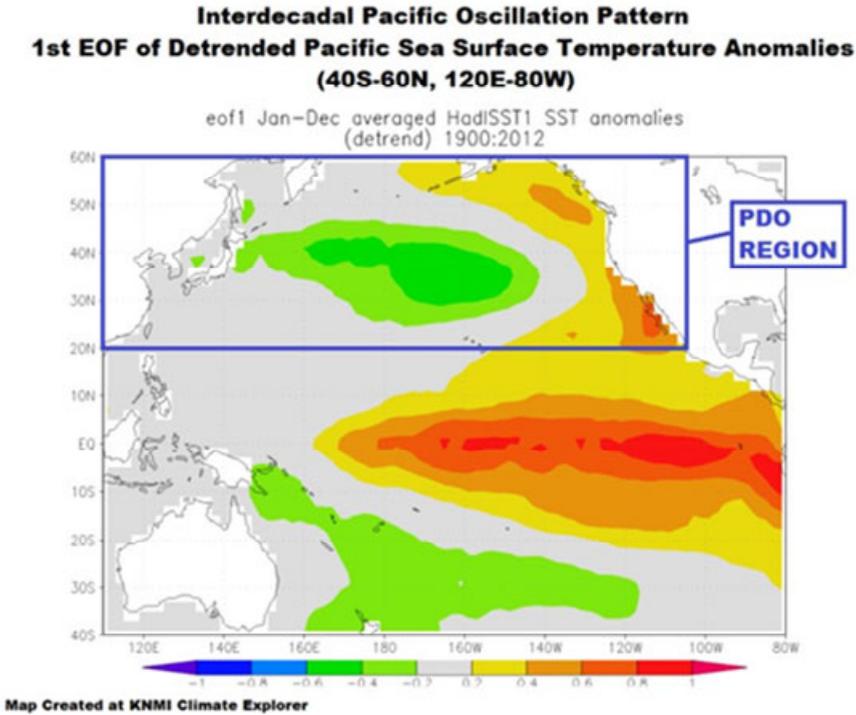
Source: NOAA

Climate indices data: five climate drivers, ENSO based NINO3.4, EMI, IPO, PDO, and DMI (IOD) were selected for MNLr analysis, considering the previous research works on rainfall and streamflow in this region as well as the concurrent and lagged correlation analysis in the preliminary stage of the current research.

ENSO phenomenon has two components: sea surface temperature and atmospheric pressure which are intensely correlated and can be represented by two types of indicators, the SLP indicator and the SST indicator (Duc et al., 2017). The most common of those is

the Troup Southern Oscillation Index (SOI) which measures the difference between sea-level atmospheric pressures at Tahiti and Darwin. There are various SST anomalies, which were derived using different areas of the equatorial Pacific Ocean (Kiem and Franks, 2001). Generally, the SST anomalies are monitored in 3 geographic regions (Figure 2) of the equatorial Pacific and defined as NINO3 (5°S – 5°N, 150° – 90°W), NINO3.4 (5°S – 5°N, 170° – 120°W) and NINO4 (5°S – 5°N, 160° – 150°W) (Risbey et al., 2009).

Figure 3 Map showing IPO and PDO region (see online version for colours)



Source: Timmermann and Trenberth (2014)

The El Nino Modoki is an ocean-atmosphere coupled process, which results in unique tripolar sea level pressure pattern during the evolution, similar to the Southern Oscillation phenomenon of El Nino (Ashok et al., 2007). Therefore, this phenomenon is named as El Nino–Southern Oscillation Modoki (EMI) and expressed by the following equation (Ashok et al., 2007).

$$EMI = SSTX - (0.5 * SSTY) - (0.5 * SSTZ) \tag{1}$$

where $X = 165^{\circ}E-140^{\circ}W, 10^{\circ}S-10^{\circ}N, Y = 110^{\circ}W-70^{\circ}W, 15^{\circ}S-5^{\circ}N, Z = 125^{\circ}E-145^{\circ}E, 10^{\circ}S-20^{\circ}N$.

The IOD represents the oceanic-atmospheric variability in the tropical Indian Ocean, which is classified by SST anomalies of reverse sign in the east and west (Saji et al., 1999; Webster et al., 1999). The dipole mode index (DMI) is a measure of the IOD and is defined by the difference in SST anomaly between the tropical western Indian Ocean

(10°S–10°N, 50°–70°E) and the tropical south-eastern Indian Ocean (10°S–equator, 90°–110°E).

The IPO is described as the Pacific ENSO-like pattern of SST, which is found in the analysis of near-global inter-decadal SST (Folland et al., 1999). IPO has a cycle of 15–30 years and characterised with two phases, namely the positive and negative phases (Henley et al., 2015; Salinger et al., 2001). While IPO is defined for the whole Pacific Basin, PDO is defined for the North Pacific, poleward of 20°N (Figure 3).

The oceanic and atmospheric climate indices data were obtained from the climate explorer website (<http://climexp.knmi.nl>), while the EMI data was collected from the website of JAMSTEC (<http://www.jamstec.go.jp/frcgc/research/dl/iod/modoki>) for the duration of 102 years (1914–2015).

2.3 Methodology of MLR and MNLR modelling

Regression analysis modelling is one of the popular statistical approaches and widely used for this kind of analysis (Mekanik et al., 2013). Regression methods can explain the relationships between a response (dependent) variable and several regressor (independent) variables (Tabari et al., 2010). In MLR, the function is linear which can be explained by the following equation:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (2)$$

where Y is the dependent variable (e.g. streamflow for the current study) and X_1, X_2, \dots, X_n are the independent variables (climate indices e.g. ENSO, PDO, IPO, etc.). $\beta_1, \beta_2, \dots, \beta_n$ are the coefficients of the independent variables while α is the intercept or error and n is the number of observations.

Unlike traditional MLR methods, the MNLR function is the nonlinear combination of model parameters and depends on one or more independent variables (Bilgili, 2010). The general form of a MNLR function can be represented by the following equation:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_j + \beta_3 X_i^2 + \beta_4 X_j^2 \dots + \beta_n X_i X_j \quad (3)$$

where, α is the intercept and $\beta_1, \beta_2, \dots, \beta_n$ are the coefficients of the independent variables while n is the number of observations.

In order to find out the suitable relationship of each independent variable (climate indices), a series of simple regression analysis between the streamflow and climate variables were performed. Based on the correlation values (Pearson correlation, r) of this analysis, the appropriate nonlinear relationship for each variable was selected to develop the multiple nonlinear equations for predicting streamflow. Different functions including the exponential, cubic, quadratic and linear functions are used to identify the best relation (Table 2).

The MNLR analysis in the present study was performed using Minitab software. Different combinations of input variables were trialled to explore the best MNLR model for each location. Also, as the relationships are to be used for forecasting, lagged values of the climate indices were used with the goal of forecasting at least 3–4 months in advance. At first, every MNLR model was calibrated using 96 years of data (1914–2009) which was followed by the validation of the models with the remaining six years (2010–2015) of data.

Table 2 List of nonlinear equations used in this study

Function	General equation
Cubic	$y = ax^3 + bx^2 + cx + d$
Quadratic	$y = ax^2 + bx + c$
Exponential	$y = ae^{bx}$

The performances of the developed MNL models were assessed with several statistical performance measures such as Pearson correlation value (r), mean absolute error (MAE), root mean squared error (RMSE) and Willmott index of agreement (d). A similar approach for validating the results was applied by Mekanik et al. (2013) while predicting seasonal rainfall using climate indices.

The Pearson correlation coefficient (r) is calculated using the following equation:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} * \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \tag{4}$$

where, ‘ r ’ is the correlation coefficient, x and y are two variables, x_i and y_i are the value of the i^{th} observation and \bar{x} and \bar{y} are the average values of x and y respectively.

The MAE is calculated using the following equation:

$$MAE = \frac{1}{n} \sum_{i=1}^{i=n} |P_i - O_i| \tag{5}$$

The RMSE is calculated using the following equation:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2} \tag{6}$$

where O_i is the observed value, P_i is the predicted value, and n is the number of observations.

Willmott index of agreement (d) is calculated through the following equation:

$$d = 1 - \frac{\left[\sum |\hat{y}_i - x_i|^2 \right]}{\left[\sum (|\hat{y}_i - x_i| + |x_i - \bar{x}_i|^2) \right]} \tag{7}$$

where \hat{y}_i refers to the predicted value corresponding to i^{th} observation and x_i refers to i^{th} value of observation. The ideal value for the Pearson correlation is one, which will refer to the best association between two variables, whereas a value of zero will indicate there is no association. Lower values of the MAE and RMSE indicate better performance of the model. The closer the ‘ d ’ value to one, the better the model accuracy.

Table 3 Statistical performance measures of MNL models

Station	Model	Calibration period				Validation period			
		r	MAE	RMSE	d	r	MAE	RMSE	d
Singleton	PDO _{June} -NINO3.4 _{June}	0.47	12.67	17.96	0.58	0.56	12.92	13.71	0.83
North Cuerindi	PDO _{April} -NINO3.4 _{June}	0.53	5.18	7.83	0.64	0.62	7.63	8.98	0.57
Gundagai	PDO _{Mar} -NINO3.4 _{Mar}	0.47	53.91	68.98	0.57	0.54	27.97	31.23	0.66
Kiosk	IOD _{Jun} -NINO3.4 _{Jun}	0.38	3.19	4.11	0.48	0.56	3.84	4.13	0.57
Wagga Wagga	PDO _{Mar} -NINO3.4 _{Mar}	0.43	65.74	84.42	0.19	0.52	0.35	36.94	5.04
Brewarrina	EMI _{Jun} -NINO3.4 _{Jun}	0.51	30.73	41.00	0.63	0.63	0.61	35.02	40.99

Table 4 Comparison of MLR and MNL models

Station name	MLR models				MNL models			
	Best models	Pearson correlation (r)		Best models	Pearson correlation (r)			
		Calibration	Validation		Calibration	Validation		
Singleton	PDO _{Mar} -NINO3.4 _{Jun}	0.43	0.51	PDO _{June} -NINO3.4 _{Jun}	0.47	0.56		
North Cuerindi	PDO _{Jul} -NINO3.4 _{Jul}	0.51	0.56	PDO _{Apr} -NINO3.4 _{June}	0.53	0.62		
Gundagai	IPO _{Jul} -NINO3.4 _{Jul}	0.40	0.43	PDO _{Mar} -NINO3.4 _{Mar}	0.47	0.54		
Kiosk	PDO _{Aug} -NINO3.4 _{Jul}	0.44	0.53	IOD _{Jun} -NINO3.4 _{Jun}	0.38	0.56		
Wagga Wagga	IPO _{Jul} -NINO3.4 _{Jul}	0.41	0.20	PDO _{Mar} -NINO3.4 _{Mar}	0.43	0.35		
Brewarrina	IOD _{Jul} -NINO3.4 _{Jul}	0.41	0.59	EM1 _{Jun} -NINO3.4 _{Jun}	0.51	0.61		

3 Results and discussion

3.1 MLNR model results

For northern NSW (Singleton and North Cuerindi), through single nonlinear regression analysis, it was found that the maximum correlation value was obtained with a cubic function for all the indices. NINO3.4_{July} provided the highest correlation (0.52), whereas through single linear correlation analysis, the maximum correlation value obtained was 0.48. For MNLN analysis, the maximum value of correlation coefficient in the calibration stage was found to be 0.53 with PDO_{April}–NINO3.4_{June} combination, whereas in the validation stage, the same combination yielded a maximum correlation of 0.62. Table 3 shows the detailed statistical performance measures for all the stations. For southern NSW (Gundagai and Kiosk), through single nonlinear regression analysis, all indices yielded highest correlations through cubic function. The highest correlation value obtained was 0.451 with PDO_{Mar}, whereas through the linear regression, the maximum correlation obtained was 0.375. For Gundagai, through MNLN, the highest correlation value was obtained with PDO_{Mar}–NINO3.4_{Mar} combination, where correlation values for calibration and validation periods were 0.47 and 0.54 respectively (Table 3). This model enables the prediction of streamflow five months in advance. However, for Kiosk, the highest correlation values were obtained with the IOD_{Jun}–NINO3.4_{Jun} combination. Correlation values for calibration and validation periods were 0.38 and 0.56 respectively (Table 3). For central-west NSW (Wagga Wagga), through single nonlinear regression analysis, all indices showed the highest correlations for cubic function. The highest correlation value was 0.414 with IOD_{Aug}, whereas through single linear correlation analysis, the maximum correlation obtained was 0.39. With MLNR model, PDO_{Mar}–NINO3.4_{Mar} combination provided highest correlations (0.434 in calibration period and 0.52 in validation period). Also, this model enables predictions five months in advance. For western NSW (Brewarrina), through single nonlinear regression analysis the maximum correlations were obtained using cubic function for all the indices, while the highest correlation was 0.44 with NINO3.4_{June}. Through MNLN analysis, the highest correlations were obtained for EMI_{July}–NINO3.4_{July} combination and the correlation values for calibration and validation periods were 0.51 and 0.61 respectively.

3.2 Comparison with MLR results

Dataset for the same period was used for developing MLR forecast models for all the selected stations. The comparison of the performances of these models based on their correlation values has been presented in Table 4. It is to be noted that in all the cases, the same combination of indices did not yield the best performance in both the methods, i.e., best combination through MLR models was different than the best combination through MNLN models. The result shows that in the calibration stage, the MNLN models outperformed MLR models for all the stations, except one (Kiosk station), where correlation of MLR (0.44) model was slightly higher than MNLN model (0.38). In the validation stage, MNLN models always outperformed the performances of MLR models. However, in most cases, differences in the correlation values are insignificant.

For both the methods, the maximum number of best models were found with PDO and NINO3.4 combined indices, which indicates the strongest influence of these indices

on spring streamflow of NSW. However, influence of a particular index does vary spatially, which has been proved by many studies. For this particular scenario, the strong influences of IPO and IOD along with EMI are also observed in some other locations, where the best models were achieved with either of these indices along with NINO3.4. Hence, in general, the strong influence of all these five indices on spring streamflow of NSW is clear through this analysis.

3.3 Developed nonlinear equations

Table 5 shows the developed nonlinear equations representing best model for each location.

Table 5 Equations of the best-developed MNLR models

Station	Developed model
Singleton	$Q = 19.9093 + 0.405381 * PDO_{JUNE}^3 - 0.781302 * PDO_{JUNE}^2 - 2.70294 * PDO_{JUNE} + 12.3403 * NINO3.4_{JUNE}^3 - 1.24083 * NINO3.4_{JUNE}^2 - 24.9111 * NINO3.4_{JUNE}$
North Cuerindi	$Q = 8.11847 - 0.298752 * PDO_{APRIL}^3 + 0.813401 * PDO_{APRIL}^2 - 1.09289 * PDO_{AUG} + 4.37942 * NINO3.4_{JUNE}^3 - 0.118078 * NINO3.4_{JUNE}^2 - 11.1756 * NINO3.4_{JUNE}$
Gundagai	$Q = -8.403 * PDO_{Mar}^3 + 9.54 * PDO_{Mar}^2 + 13.462 * PDO_{Mar} + 18.246 * NINO3.4_{Mar}^3 - 1.842 * NINO3.4_{Mar}^2 - 29.314 * NINO3.4_{Mar} + 136.08$
Kiosk	$Q = -4.838 * IDO_{Jun}^3 - 0.413 * IDO_{Jun}^2 - 0.521 * IDO_{Jun} + 2.05461 * NINO3.4_{Jun}^3 + 0.53274 * NINO3.4_{Jun}^2 - 3.70504 * NINO3.4_{Jun} + 5.65868$
Wagga Wagga	$Q = -9.02781 * PDO_{Mar}^3 + 10.7172 * PDO_{Mar}^2 + 14.8737 * PDO_{Mar} + 20.3952 * NINO3.4_{Mar}^3 + 1.35147 * NINO3.4_{Mar}^2 - 34.972 * NINO3.4_{Mar} + 150.128$
Brewarrina	$Q = 44.6711 - 50.112 * EMI_{JUNE}^3 - 6.60371 * EMI_{JUNE}^2 + 17.5917 * EMI_{JUNE} + 14.0838 * NINO3.4_{JUNE}^3 + 5.33871 * NINO3.4_{JUNE}^2 - 51.1802 * NINO3.4_{JUNE}$

4 Conclusions

The multiple nonlinear regression (MNLR) analysis was carried out to explore the nonlinear relationship between streamflow and climate indices. Three different nonlinear functions along with the linear function were used to perform a single correlation analysis between spring streamflow and single lagged climate indices which was followed by MNLR analysis. The functions which provided the highest correlations were chosen to

develop the MNLR equations. Finally, the best model for each station was selected based on highest correlation values and better statistical performance.

From the single linear regression analysis, it is observed that all the indices for almost all stations showed highest correlations for the cubic function which implies that cubic function has comparatively more potential to explain the relationship between spring streamflow and lagged climate indices. For every station, the nonlinear function had higher correlation values than the linear function, which showed that the underlying relationship between spring streamflow and lagged climate indices is nonlinear.

Based on the outcomes of single linear regression analysis, combined MNLR models were developed with the combination of two different climate indices. The nonlinear function that showed the highest correlation in single nonlinear regression analysis was used for the combined model development. Out of the six stations, four stations showed the best models with combined PDO and NINO3.4 indices, implying that these two indices have a stronger influence on the spring streamflow of NSW. Among the other two stations, Kiosk had the best model consisting of IOD and NINO3.4 indices, while Brewarrina had the best model with EMI and NINO3.4 indices. Hence, it was evident that PDO and ENSO indices have the strongest impact on spring streamflow of NSW, and ENSO has an influence on the streamflow of all the stations.

Statistical performances of the developed models were analysed to ensure the reliability of the models. Different statistical measures, including r , MAE, RMSE and d were used to check the reliability of the models. The models with higher correlation values and lower errors were selected as the best models. The Pearson correlation values in calibration and validation stages were quite similar, which implies the good performance of the models. The best-developed models were able to predict streamflow from three to six months in advance depending on the geographical location.

It is to be noted that only with two climate indices it is unlikely that such model will be able to capture the unusual phenomenon like severe droughts (e.g., millennium drought in Australia from 1994–2010) and floods. For this study, the millennium drought period fell during the calibration period, which impacted the models' performances during calibration period.

References

- Abbot, J. and Marohasy, J. (2015) 'Using lagged and forecast climate indices with artificial intelligence to predict monthly rainfall in the Brisbane Catchment, Queensland, Australia', *International Journal of Sustainable Development and Planning*, Vol. 10, No. 1, pp.29–41.
- Adamowski, J., Fung, C.H., Prasher, S.O., Ozga-Zielinski, B. and Sliusarieva, A. (2012) 'Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada', *Water Resources Research*, Vol. 48, No. 1.
- Adamowski, J.F. (2008) 'Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis', *Journal of Hydrology*, Vol. 353, No. 3, pp.247–266.
- Ashok, K., Behera, S.K., Rao, S.A., Weng, H. and Yamagata, T. (2007) 'El Niño Modoki and its possible teleconnection', *Journal of Geophysical Research: Oceans*, Vol. 112, No. C11007, pp.1–27.
- Bilgili, M. (2010) 'Prediction of soil temperature using regression and artificial neural network models', *Meteorology and Atmospheric Physics*, Vol. 110, Nos. 1–2, pp.59–70.

- Chiew, F.H., Piechota, T.C., Dracup, J.A. and McMahon, T.A. (1998) 'El Nino/Southern Oscillation and Australian rainfall, streamflow and drought: links and potential for forecasting', *Journal of Hydrology*, Vol. 204, Nos. 1–4, pp.138–149.
- Dastorani, M.T., Moghadamnia, A., Piri, J. and Rico-Ramirez, M. (2010) 'Application of ANN and ANFIS models for reconstructing missing flow data', *Environmental Monitoring and Assessment*, Vol. 166, No. 1, pp.421–434.
- Duc, H.N., Rivett, K., MacSween, K. and Le-Anh, L. (2017) 'Association of climate drivers with rainfall in New South Wales, Australia, using Bayesian model averaging', *Theoretical and Applied Climatology*, Vol. 127, Nos. 1–2, pp.169–185.
- Dutta, S.C., Ritchie, J.W., Freebairn, D.M. and Abawi, G.Y. (2006) 'Rainfall and streamflow response to El Niño Southern Oscillation: a case study in a semiarid catchment, Australia', *Hydrological Sciences Journal*, Vol. 51, No. 6, pp.1006–1020.
- Esha, R.I. and Imteaz, M.A. (2019) 'Assessing the predictability of MLR models for long-term streamflow using lagged climate indices as predictors: a case study of NSW (Australia)', *Hydrology Research*, Vol. 50, No. 1, pp.262–281.
- Folland, C.K., Parker, D.E., Colman, A.W. and Washington, R. (1999) 'Large scale modes of ocean surface temperature since the late nineteenth century', in *Beyond El Niño*, Springer Berlin Heidelberg.
- Ghamariadyan, M. and Imteaz, M.A. (2020) 'A wavelet artificial neural network method for medium-term rainfall prediction in Queensland (Australia) and the comparisons with conventional methods', *International Journal of Climatology*, Vol. 41, No. S1, pp.E1396–E1416.
- Ghamariadyan, M. and Imteaz, M.A. (2021) 'Monthly rainfall forecasting using temperature and climate indices through a hybrid method in Queensland, Australia', *Journal of Hydrometeorology*, Vol. 22, No. 5, pp.1259–1273.
- Haque, M.M., Rahman, A., Hagare, D. and Kibria, G. (2013) 'A comparison of linear and nonlinear regression modelling for forecasting long term urban water demand: a case study for Blue Mountains water supply system in Australia', *Proceedings of the 6th International Conference on Water Resources and Environment Research*, 3–7 June, Koblenz, Germany.
- Henley, B.J., Gergis, J., Karoly, D.J., Power, S., Kennedy, J. and Folland, C.K. (2015) 'A tripole index for the Interdecadal Pacific Oscillation', *Climate Dynamics*, Vol. 45, Nos. 11–12, pp.3077–3090.
- Hossain, I., Esha, R.I. and Imteaz, M.A. (2018) 'An attempt to use non-linear regression modelling technique in long-term seasonal rainfall forecasting for Australian Capital Territory', *Geosciences*, Vol. 8, No. 8, pp.1–12.
- Hossain, I., Rasel, H.M., Imteaz, M.A. and Mekanik, F. (2020a) 'Long-term seasonal rainfall forecasting using linear and non-linear modelling approaches: a case study for Western Australia', *Meteorology and Atmospheric Physics*, Vol. 132, No. 4, pp.131–141.
- Hossain, I., Rasel, H.M., Mekanik, F. and Imteaz, M.A. (2020b) 'Artificial neural network modelling technique in predicting Western Australian seasonal rainfall', *International Journal of Water*, Vol. 14, No.1, pp.14–28.
- Islam, F. and Imteaz, M.A. (2020) 'Use of teleconnections to predict Western Australian seasonal rainfall using ARIMAX model', *Hydrology*, Vol. 7, No. 52, pp.1–25.
- Kiem, A.S. and Franks, S.W. (2001) 'On the identification of ENSO-induced rainfall and runoff variability: a comparison of methods and indices', *Hydrological Sciences Journal*, Vol. 46, No. 5, pp.715–727.
- Kiem, A.S., Franks, S.W. and Kuczera, G. (2003) 'Multi-decadal variability of flood risk', *Geophysical Research Letters*, Vol. 30, No. 2, pp.1–4.
- King, A.D., Alexander, L.V. and Donat, M.G. (2013) 'Asymmetry in the response of eastern Australia extreme rainfall to low-frequency pacific variability', *Geophysical Research Letters*, Vol. 40, No. 10, pp.2271–2277.

- Kişi, Ö. (2007) 'Streamflow forecasting using different artificial neural network algorithms', *Journal of Hydrologic Engineering*, Vol. 12, No. 5, pp.532–539.
- Kumar, A.R., Sudheer, K.P., Jain, S.K. and Agarwal, P.K. (2005) 'Rainfall-runoff modelling using artificial neural networks: comparison of network types', *Hydrological Processes*, Vol. 19, No. 6, pp.1277–1291.
- McBride, J.L. and Nicholls, N. (1983) 'Seasonal relationships between Australian rainfall and the Southern Oscillation', *Monthly Weather Review*, Vol. 111, No. 10, pp.1998–2004.
- Mekanik, F. and Imteaz, M.A. (2018) 'Variability of cool seasonal rainfall associated with Indo-Pacific climate modes: case study of Victoria, Australia', *Journal of Water and Climate Change*, Vol. 9, No. 3, pp.584–597.
- Mekanik, F., Imteaz, M.A. and Talei, A. (2016) 'Seasonal rainfall forecasting by adaptive network-based fuzzy inference system (ANFIS) using large scale climate signals', *Climate Dynamics*, Vol. 46, No. 9, pp.3097–3111.
- Mekanik, F., Imteaz, M.A., Gato-Trinidad, S. and Elmahdi, A. (2013) 'Multiple regression and artificial neural network for long-term rainfall forecasting using large scale climate modes', *Journal of Hydrology*, Vol. 503, pp.11–21.
- Miyagishi, K., Ohsako, M. and Ichihashi, H. (1999) 'Temperature prediction from regional spectral model by neurofuzzy GMDH', *Proc. of The Second Asia-Pacific Conference on Industrial Engineering and Management Systems*, Kanazawa, Japan, pp.705–708.
- Mutlu, E., Chaubey, I., Hexmoor, H. and Bajwa, S.G. (2008) 'Comparison of artificial neural network models for hydrologic predictions at multiple gauging stations in an agricultural watershed', *Hydrological Processes*, Vol. 22, No. 26, pp.5097–5106.
- Piechota, T.C., Chiew, F.H., Dracup, J.A. and McMahon, T.A. (1998) 'Seasonal streamflow forecasting in eastern Australia and the El Niño–Southern Oscillation', *Water Resources Research*, Vol. 34, No. 11, pp.3035–3044.
- Power, S., Casey, T., Folland, C., Colman, A. and Mehta, V. (1999) 'Inter-decadal modulation of the impact of ENSO on Australia', *Climate Dynamics*, Vol. 15, No. 5, pp.319–324.
- Qi, C. and Chang, N.B. (2011) 'System dynamics modeling for municipal water demand estimation in an urban region under uncertain economic impacts', *Journal of Environmental Management*, Vol. 92, No. 6, pp.1628–1641.
- Rasel, H.M., Imteaz, M.A. and Mekanik, F. (2016) 'Investigating the influence of remote climate drivers as the predictors in forecasting South Australian spring rainfall', *International Journal of Environmental Research*, Vol. 10, No. 1, pp.1–12.
- Rezaeianzadeh, M., Tabari, H., Yazdi, A.A., Isik, S. and Kalin, L. (2014) 'Flood flow forecasting using ANN, ANFIS and regression models', *Neural Computing and Applications*, Vol. 25, No. 1, pp.25–37.
- Risbey, J.S., Pook, M.J., McIntosh, P.C., Wheeler, M.C. and Hendon, H.H. (2009) 'On the remote drivers of rainfall variability in Australia', *Monthly Weather Review*, Vol. 137, No. 10, pp.3233–3253.
- Robertson, D.E. and Wang, Q.J. (2009) 'A Bayesian joint probability approach to seasonal prediction of streamflows: predictor selection and skill assessment', *32nd Hydrology and Water Resources Symposium*, Newcastle, Australia, Engineers Australia.
- Saji, N.H., Goswami, B.N., Vinayachandran, P.N. and Yamagata, T. (1999) 'A dipole mode in the tropical Indian Ocean', *Nature*, Vol. 401, No. 6751, pp.360–363.
- Salinger, M.J., Renwick, J.A. and Mullan, A.B. (2001) 'Interdecadal Pacific oscillation and south Pacific climate', *International Journal of Climatology*, Vol. 21, No. 14, pp.1705–1721.
- Tabari, H., Marofi, S. and Sabziparvar, A.A. (2010) 'Estimation of daily pan evaporation using artificial neural network and multivariate non-linear regression', *Irrigation Science*, Vol. 28, No. 5, pp.399–406.

- Verdon, D.C., Wyatt, A.M., Kiem, A.S. and Franks, S.W. (2004) 'Multidecadal variability of rainfall and streamflow: Eastern Australia', *Water Resources Research*, Vol. 40, No. W10201, pp.1-8.
- Webster, P.J., Moore, A.M., Loschnigg, J.P. and Leben, R.R. (1999) 'Coupled ocean-atmosphere dynamics in the Indian Ocean during 1997-98', *Nature*, Vol. 401, No. 6751, pp.356-360.
- Westra, S. and Sharma, A. (2009) 'Probabilistic estimation of multivariate streamflow using independent component analysis and climate information', *Journal of Hydrometeorology*, Vol. 10, No. 6, pp.1479-1492.