



A Novel Hybrid Approach for Predicting Western Australia's Seasonal Rainfall Variability

Farhana Islam¹ · Monzur Alam Imteaz¹

Received: 24 February 2022 / Accepted: 7 June 2022 / Published online: 22 June 2022
© The Author(s) 2022

Abstract

In this paper, 100 years of uninterrupted rainfall data for 12 rainfall stations (four rainfall stations from each region) in Western Australia were analyzed against respective dominant climate indices, and representative prediction models were developed using ARIMAX, GEP, and a hybrid technique (GEP-ARIMAX). Statistical performance evaluators such as Pearson correlation (r), root mean square error ($RMSE$), mean absolute error (MAE), and refined Willmot index of agreement (d_r) were used to evaluate the prediction performance of the developed models. These models demonstrated their capability to predict up to four months in advance with Pearson correlation (r) values ranging from 0.53 to 0.83, 0.75 to 0.85, and 0.87 to 0.95 for ARIMAX, GEP, and hybrid (GEP-ARIMAX) models respectively. While compared, the hybrid (GEP-ARIMAX) model showed superior prediction performance in both calibration and validation periods with Pearson correlation (r) and refined Willmot index of agreement (d_r) values were as high as 0.96 and 0.84 respectively. This paper demonstrated a novel hybrid GEP-ARIMAX model showing significantly good rainfall forecasting capability than conventional linear and non-linear models.

Keywords Rainfall forecasting · Climate indices · ARIMAX · GEP · Hybrid · Western Australia

1 Introduction

Australian rainfall has a distinct nature as coastal regions experience wetter winter while inland west encounters less rainfall at that time. Several studies around the world explored teleconnections between the climate indices with regional rainfalls (Ghamariadyan and Imteaz 2021; Calado et al. 2019). Around Australia, primarily IOD and SAM have been found as influential drivers for rainfall generation in south-eastern and western parts, Blocking Highs for southern parts, and ENSO Modoki and Madden Julian Oscillation

✉ Monzur Alam Imteaz
mimteaz@swin.edu.au

Farhana Islam
farhanaislam@swin.edu.au

¹ Department of Civil and Construction Engineering, Swinburne University of Technology, John Street, Hawthorn, VIC 3122, Australia

(MJO) for north-western and northern parts (Ashok et al. 2003; Marshall and Hendon 2014; Risbey et al. 2009). Among these drivers, ENSO is found as a significant contributor to rainfall generation all over Australia (Montazerolghaem et al. 2016). Pacific Ocean SST anomalies significantly affect rainfall generation in tropical and eastern regions, whereas Indian Ocean SST anomalies influence the rainfall generation in southern and western regions (Risbey et al. 2009). Current literature suggests that most of the attempts related to seasonal rainfall forecasting in Western Australia (WA) are region-based, with a majority of them being developed for South West Western Australia (SWWA) (Evans et al. 2020; Feng et al. 2015; Islam and Imteaz 2020; Ummenhofer et al. 2008). The majority of these studies attempted to demonstrate the concurrent relationship between climate indices and SWWA rainfall as the lagged relationships were not explored in detail.

Improved rainfall forecasting techniques aimed at optimum precision are an evolving process, where both linear and non-linear techniques are widely adopted (Bagirov and Mahmood 2018; Hossain et al. 2020; Islam and Imteaz 2019). As rainfall generation is a complicated phenomenon, any linear or non-linear model with overwhelming limitations may not be able to develop an accurate forecast model by itself. Hybrid modelling can overcome those limitations and is usually performed by trialling either linear or non-linear forecasting first and then analyzing the remains (i.e., residuals) using the other technique at a later stage. Previous studies have demonstrated that the hybrid model is more successful in developing an accurate forecast model compared to the individual forecast models (Wu et al. 2021; Xu et al. 2019). A combination of Gene Expression Programming (GEP) and other model techniques such as Auto-Regressive (AR), Auto-Regressive Moving Average (ARMA), Auto-Regressive Integrated Moving Average (ARIMA), Bayesian Network (BN), Linear Regression (LR), Multiple Linear Regression (MLR), Genetic Algorithm (GA) was attempted for estimating rainfall, flood, and streamflow trends (Al-Juboori 2022, Barbulescu and Bautu 2009, Kumar and Sahay 2018, Mehdizadeh and Sales 2018, Mehr 2018, Mozaffari et al. 2022, Sharghi et al. 2019).

Mehdizadeh and Sales (2018) reported that hybrid models (GEP-AR, GEP-ARMA, BN-AR, BN-ARMA, MLR-AR, and MLR-ARMA) demonstrated more accurate prediction compared to single models (GEP, BA, AR, ARMA, and MLR) for streamflow forecasting. A study performed by Mehr (2018) used a hybrid GEP-GA model for streamflow forecasting and demonstrated the superior performance of the hybrid GEP-GA model. Similarly, Karimi et al. (2016) also proposed a hybrid Wavelet-Genetic Programming (WT-GP) model that showed better performance in both short-term and long-term streamflow prediction compared to ARIMA, Adaptive Neuro-Fuzzy Interference System (ANFIS), and Artificial Neural Network (ANN) models. Barbulescu and Bautu (2009) applied a hybrid ARIMA-GEP model for precipitation analysis, where a combination of these two techniques improved the model's accuracy. Furthermore, a study conducted by Kumar and Sahay (2018) showed that the hybrid Wavelet-GEP (WT-GEP) model successfully predicted extreme flood cases which remained undetected in AR and GP models.

This study aimed at developing a hybrid model to improve the forecasting ability for WA rainfall prediction. The influence of lagged correlation between the climate indices and seasonal autumn rainfall variability for the south coast and north coast regions of South West Division (SWD) and summer rainfall variability for the Kimberley region of North West Western Australia (NWWA) was investigated. Later, the prediction outcome of the hybrid models was compared with a time series linear ARIMAX and non-linear GEP models.

2 Methodology

In this study, a correlation analysis was conducted between rainfall data and climate indices first. Afterwards, climate indices with high correlation were used as an input set for model development. Linear time series (ARIMAX) and non-linear (GEP) models were developed to see their capability to forecast seasonal rainfall in WA. The residuals derived from these two models were further analyzed using a different technique resulting in a hybrid model. The selection of methods was independent and aimed at describing the rainfall variability as per the inherent model structure, its capability to explain the relationship type (linear/ non-linear), and residual characteristics.

2.1 Autoregressive Integrated Moving Average with Exogenous Input

The ARIMA model is made up of the combination of 'AR', 'I', and 'MA' where 'AR' refers to Auto-Regressive, 'I' means for Integrated, which is a time series that must be differenced to make a non-stationary series stationary, and 'MA' is for Moving Average. The ARIMA model is generally expressed by the following expression, ARIMA (p, d, q) *(P, D, Q). The expression consists of two segments, where the first part is the non-seasonal part, and the second part is the seasonal part. Non-seasonal auto-regressive order is denoted by 'p,' non-seasonal differencing is denoted by 'd,' and the non-seasonal moving average is denoted by 'q,' whereas seasonal auto-regressive order is denoted by 'P,' seasonal differencing is denoted by 'D,' and the seasonal moving average is denoted by 'Q.' In this study, the time series did not demonstrate any seasonality, thus only the non-seasonal portion has been considered.

Two different kinds of input orders namely: ARIMA order (dependent variable; in this case, autumn/summer rainfall) and transfer function order (predictors; i.e., climatic indices) were required for ARIMAX model development. A detailed description of these input orders and model development can be found in the previous study of Islam and Imteaz (2020).

2.2 Gene Expression Programming

Gene Expression Programming (GEP) is a combination of the principles of genetic algorithms (GA) and genetic programming (GP). There is a fundamental disagreement among these three algorithms: GA utilizes a linear string of fixed length of chromosomes, GP uses non-linear entities of tree-based chromosomes with different sizes and shapes (parse tree), and GEP is encoded as a simple linear string of fixed length chromosomes and expressed as nonlinear entities of different sizes and shapes (Ferreira 2001).

GEP genes are combined with two elements, one is the head, and the other is the tail. The head encoded the functions for expression. It represents both the function set (F) and the terminal set (T). On the other hand, the tail represents the only terminal set (T). This terminal set from the tail acts as a reservoir for an argument. The argument is required by the function used in the head while there is a shortage of terminals. Therefore, the head contains functions, variables, and constants, but the tail contains only

variables and constants. For any problem, head length (h) can be selected manually, and tail length (t) needs to be calculated using the following Eq. (1) (Ferreira 2001):

$$t = h(n - 1) + 1 \quad (1)$$

where n is the number of variables/arguments required by the functions, h is the head length, t is the tail length. For example, any gene consists of function $[Q, *, /, -, +, a, b]$, head length is selected as 10, and the number of arguments is 2, in that scenario, the tail length is $t = 10 * (2 - 1) + 1 = 11$. Therefore, the length of the genes is $10 + 11 = 21$.

2.3 GEP Model Setup for Rainfall Forecasting

In this study, the GEP methodology has been applied to develop a model to represent the relationship between climate indices and rainfall. The GEP form of the prediction model has been presented in Eq. (2) (Hashmi et al. 2011):

$$Y = f(X_1, X_2, \dots, X_n) \quad (2)$$

where Y is the dependent or response variable (seasonal summer/autumn rainfall), X_1, X_2, \dots, X_n are the predictors or independent variables (large-scale climate indices).

The significant steps to performing GEP are presented in Hashmi et al. (2011). In this process, the selection of the terminal and function set is also of great importance for better prediction. The terminal set contains independent variables that get selected from the correlation analysis. The climate indices that showed the highest significant correlation with seasonal rainfall were selected as a terminal set for this study. Selection of functional set is usually performed considering the nature of the problem, simplicity to use, and past evidence of the function as an efficient and effective tool. Table 1 illustrates the functional set and genetic operators used to create genetic variation in the chromosome population.

2.4 GEP-ARIMAX Hybrid Model Development

To develop the hybrid model, the GEP modelling technique was first applied to forecast the non-linear component of the seasonal rainfall. The second step involved calculating the residuals and using the residuals as input in the ARIMAX model to predict the linear part of the seasonal rainfall. The methodology to obtain the data stationarity, AR, and MA order in ARIMAX model development for residuals has been demonstrated in detail in Islam and Imteaz (2020). Finally, the non-linear outcome of the GEP model and the residual forecast outcomes were combined to obtain the final forecast. The general structure of the model development is presented in Eq. (3) (Zhang 2003):

$$Y_{ot} = L_{ot} + N_{ot} \quad (3)$$

where, L_{ot} is the linear component and N_{ot} is the non-linear components of the time series Y_{ot} (observed value) at time t .

While developing the hybrid (GEP-ARIMAX) model, the GEP Model was first developed to explain the non-linear components. Afterwards, the residual from the GEP model were calculated using Eq. (4) (Zhang 2003):

$$E_t = Y_{ot} - N_{\hat{t}} \quad (4)$$

Table 1 Initial setting of GEP model in the training period

Initial Setting	Symbol or Value	Arity
Function Set		
Addition	+	2
Subtraction	-	2
Multiplication	*	2
Division	/	2
Square Root	Sqrt	1
Exponential	Exp	1
Natural logarithm	Ln	1
The logarithm of base 10	Log	1
Inverse	Inv	1
x to the power of 2,3	x^2, x^3	1
Cubic root	3Rt	1
Minimum of 2 inputs	Min2	2
Maximum of 2 inputs	Max2	2
Average of 2 inputs	Avg2	2
Sine	Sin	1
Cosine	Cos	1
General Setting		
Chromosome	30	
Genes Number	3–6	
Head size	7–10	
Linking function	Addition	
Fitness function error type	RMSE with Parsimony Pressure	
Genetic Operator	Optimal Evaluation	
Mutation rate	0.00138	
Inversion rate	0.00546	
IS Transposition	0.00546	
RIS Transposition	0.00546	
One-point recombination rate	0.00277	
Two-point recombination rate	0.00277	
Gene Recombination rate	0.00277	
Gene Transposition rate	0.00277	
Numerical Constants	± 10	

where, Y_{ot} is the observed value and N_{ft} is the forecasted value from the GEP model at time t . Residuals unexplained in the GEP model were later used as input in the ARIMAX model to obtain the linear component of the time series. At this stage, the forecasted values from the ARIMAX model were combined with the predicted values of the GEP model. The combined forecast of the hybrid (GEP-ARIMAX) model is presented in Eq. (5) (Zhang 2003):

$$Y_{ft} = L_{ft} + N_{ft} \quad (5)$$

where, N_{ft} is the forecasted value from the GEP model and L_{ft} is the residual forecasted value from the ARIMAX model. The proposed methodology of the hybrid (GEP-ARIMAX) model is presented in Fig. 1.

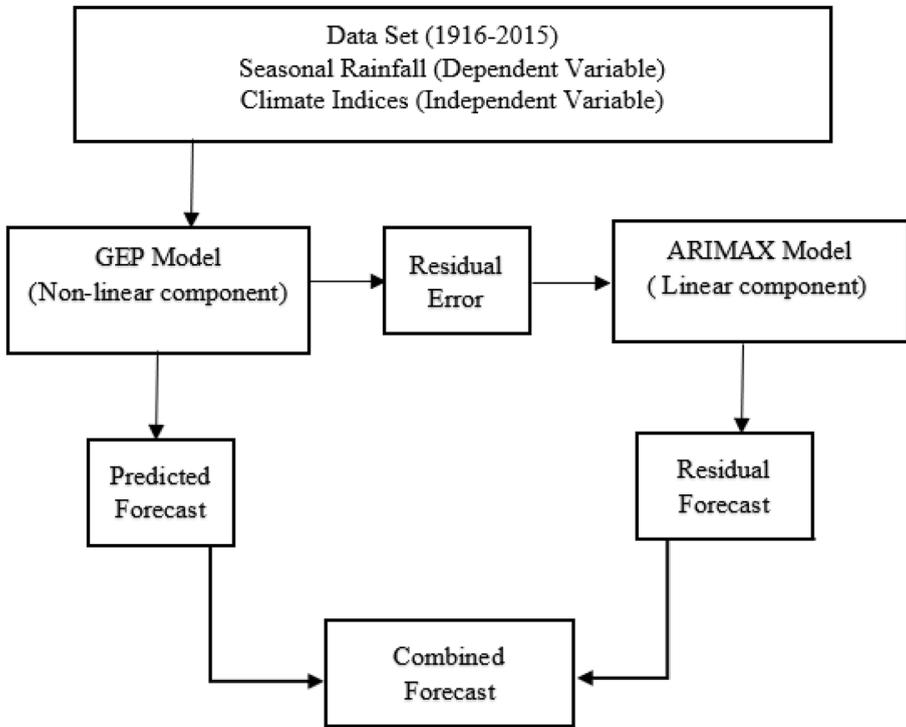


Fig. 1 Hybrid (GEP-ARIMAX) model development

2.5 Performance Metrics

The development of prediction models requires model performance measures and calculating statistical error parameters to evaluate the model performances. Among them, *RMSE* and *MAE* are the most prominent measure of evaluating errors in hydro-informatics, where a lower value of *RMSE* and *MAE* indicates a better predictability performance of the model (Saigal and Mehrotra 2012; Singh et al. 2005; Shabani et al. 2018). However, these tests have some limitations, which can be subjugated using an improved refined index of agreement (d_r) developed by Willmott et al. (2012). The d_r is the remodification of previously developed Willmott's index of agreement (d).

3 Study Area and Preliminary Analysis

3.1 Study Area and Data

In this study, four rainfall stations from each region were chosen based on the availability of continual monthly rainfall data and fewer missing values. The other regions of WA were disregarded as most of them are dry central locations. Monthly rainfall data were obtained from the Australian Bureau of Meteorology (BoM) database for the past 100 years (1916

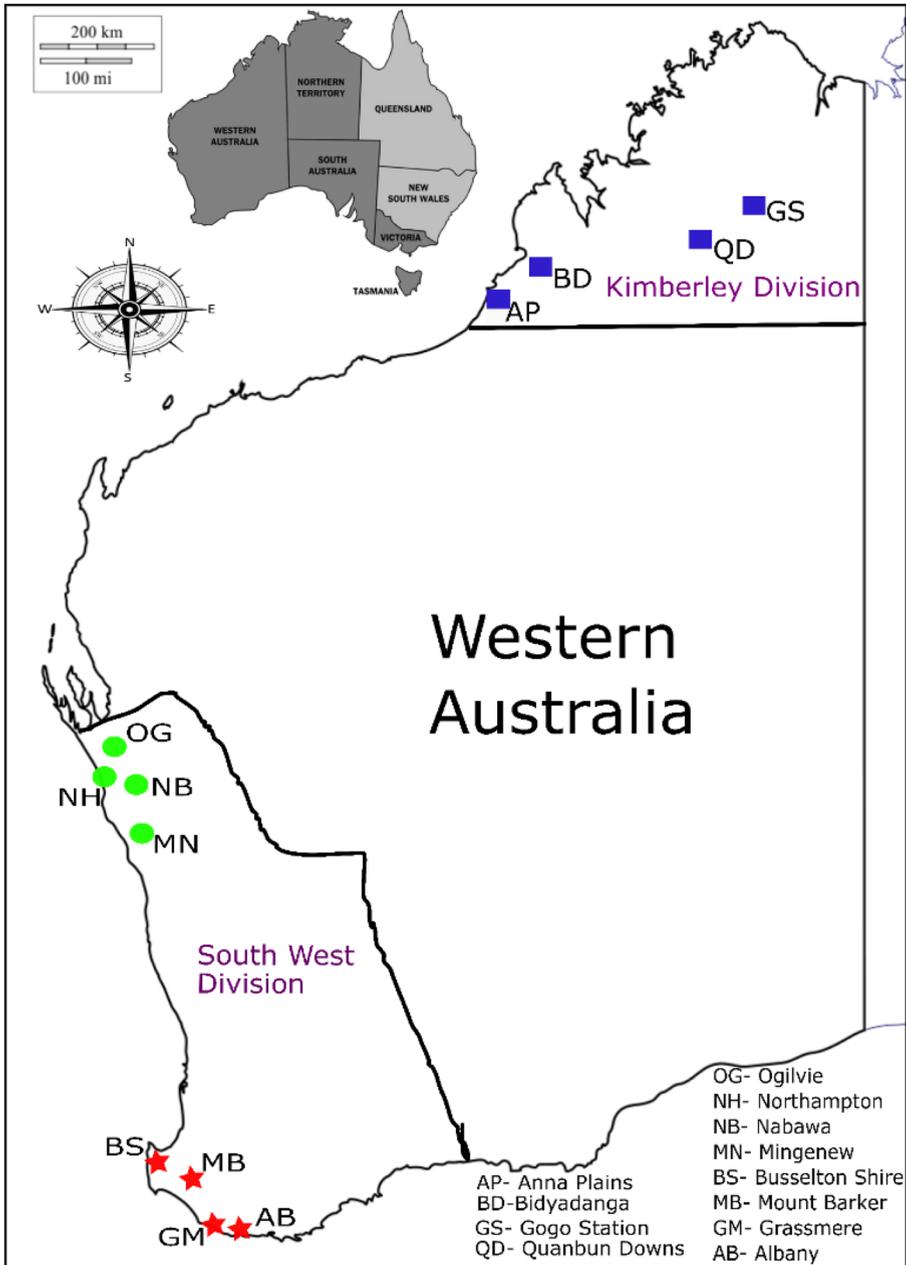


Fig. 2 The geographical location of the study area and selected rainfall stations

to 2015). Autumn (March to May) and summer (December to February) rainfall data for the selected stations were extracted and refined for analysis. The geographical location of the study area is presented in Fig. 2 and an overview of the selected rainfall stations is presented in Table 2.

Besides, 100 years (1916–2015) climate data for the climate drivers such as the Southern Oscillation Index (SOI) (SLP based), ENSO indices Nino3.4, Nino4, Nino3 (SST based), El Nino Modoki index (EMI), Dipole Mode Index (DMI), and Western Tropical Indian Ocean (WTIO) were obtained from the climate explorer website (<http://climexp.knmi.nl/>). Data partitioning for calibration and validation model sets were set at a 70:30 ratio (Ferranti 2012).

3.2 Preliminary Analysis

Previous research suggests that climate indices such as DMI, WTIO, Nino3, Nino3.4, Nino4, SOI, EMI, SAM, and SWAC have significant influences on Western Australian seasonal rainfall (Montazerolghaem et al. 2016; Risbey et al. 2009; Taschetto and England 2009; Ummenhofer et al. 2008). Due to 100 years of continuous data availability, this study investigated DMI, WTIO, Nino3, Nino3.4, Nino4, SOI, and EMI as viable climate indices for WA seasonal rainfall. However, SAM and SWAC were not included in this study as long-term continuous data for these indices were not available.

At first, single correlation analyses were performed between climate indices and seasonal rainfall to determine the set of potential predictors. It was observed that for SWD, maximum rainfall occurred during the winter and autumn seasons, whereas for Kimberley, summer and autumn are the dominant rainfall seasons. Climate indices with statistical significance (at 1% and 5% levels) were considered for analysis. However, for SWD, winter seasonal rainfall showed no correlation with the selected indices, thus, no prediction models were developed. The initial analysis showed that SWD's autumn rainfall and the Kimberley region's summer rainfall significantly correlated with the selected climate indices. Both SLP-based and SST-based ENSO indices showed a significant correlation with the South Coast and North Coast autumn rainfall with a five-month lagged period (October to February). Besides, DMI also showed significant correlations with autumn rainfall for both regions. However, ENSO Modoki Index (EMI) did not show any correlation for the south coast region, while, EMI showed significant correlations with the north coast rainfall stations with three months lagged period. These results are consistent with previously reported findings (Taschetto and England 2009; Fierro and Leslie 2013).

Single correlation analyses showed that the SLP-based ENSO index (i.e., SOI) has a significant influence on North West Western Australia's (NWWA) summer rainfall. Fierro and Leslie (2013), also reported a similar observation, stating SOI has the most robust relationship with November to April rainfall for the region. On the other hand, Nino3.4, Nino3, Nino4, and EMI exhibited very little influence on summer rainfall in NWWA. Moreover, DMI (the indicator of IOD) did not show any effect (except for the station- Quanbun Downs) confirming the SLP-based climate index (SOI) increases NWWA summer rainfall and the SST-based ENSO, IOD, and ENSO Modoki Index (EMI) has no significant impact on the rain. However, meteorological observation suggests that tropical Indian ocean indices may positively impact rainfall generation for the region (Lin and Li 2012; Shi et al. 2008). The analysis results also confirm that WTIO significantly correlates with summer rainfall for all the selected stations in NWWA.

Table 2 Overview of the selected rainfall stations

Region	Station Number	Station Name	Latitude	Longitude	Elevation (m)	Annual Mean Rainfall (mm)	Summer Rainfall (mm)	Autumn Rainfall (mm)	Winter Rainfall (mm)	Spring Rainfall (mm)
South Coast	9500	Albany	35.03° S	117.88° E	3.0	938.2	80.7	225.0	399.4	228.7
	9581	Mount Barker	34.63° S	117.64° E	300.0	733.3	80.7	157.6	283.5	189.0
	9551	Grassmere	35.02° S	117.76° E	10.0	987.8	85.8	236.8	421.7	238.5
North Coast	9515	Busselton Shire	33.66° S	115.35° E	4.0	811.6	32.7	177.2	446.5	149.2
	8104	Ogilvie	28.15° S	114.67° E	280.0	387.7	26.7	99.6	206.9	53.3
	8028	Nabawa	28.50° S	114.79° E	145.0	450.6	25.6	104.8	251.0	67.1
	8088	Mingenew	29.19° S	115.44° E	153.0	402.1	28.4	97.9	211.3	61.4
Kimberley	8100	Northampton	28.35° S	114.64° E	180.0	450.6	22.8	122.7	269.2	68.4
	3028	Anna Plains	19.25° S	121.49° E	10.0	417.92	266.03	122.38	24.98	7.79
	3030	Bidyadanga	18.68° S	121.78° E	11.0	521.56	329.29	149.29	25.85	8.10
	3014	Gogo Station	18.29° S	125.59° E	150.0	487.30	349.17	103.3	14.35	31.76
	3022	Quanban Downs	18.38° S	125.23° E	100.0	500.80	356.5	110.1	13.55	34.81

Data Source: www.bom.gov.au/climate/data/

Data Period: 1916–2015

4 Result and Discussion

4.1 ARIMAX Model Development

Several ARIMAX model sets were built using various combinations of climate drivers and respective correlation parameters are presented in Table 3. For all rainfall stations in the south coast region, the DMI-Nino3 model showed strong and consistent significant Pearson correlations (r) values. Similarly, the DMI-Nino3 model set has been identified as the best model illustrating the highest significant correlations (r) values for the north coast as well. Therefore, the DMI-Nino3 model has been considered the best-proposed model for both the north and south coast regions. For the Kimberley region, the WTIO-SOI model combination displayed the strongest correlation statistics and hence was selected as the best model set.

Following the development of ARIMAX models, a diagnostic check (Ljung-Box test) was carried out for the selected rainfall stations to ensure the adequacy of the developed models. The p-values for all these developed models were implied as larger than 0.05, validating the null hypothesis of white noise is true (Ljung and Box 1978). Another alternative approach to determining the autocorrelation between residuals is to produce residual ACF and PACF plots. From the diagnostic check, it was found that the residuals do not have any autocorrelation and the developed ARIMAX models are adequate.

4.2 GEP Model Development

Similar to ARIMAX model development, several GEP model sets were developed for different climate driver combinations and their correlation parameters are presented in Table 3. All these analyses were performed using the 'GeneXpro-Tools 5' software.

From Table 3, it is observed that the DMI-Nino3 models have demonstrated consistent and significant correlations (r) for both the north and south coast regions. On the other hand, for the Kimberley region, the WTIO-SOI model shows the best correlation.

GEP models explicitly offer the functions utilized in the system and an easy-to-understand mathematical presentation in terms of Expression Trees (ETs). Table 4 illustrates the output equation for the developed GEP models for all the selected rainfall stations in three different regions. GeneXpro modelling tool involves thousands of iterations performed within the system and the structure of the equation depends on the functions and degree of the equation selected during the iteration process.

Nonetheless, the prediction performance of the GEP model to predict SWWA's north and south coast's autumn rainfall and NWWA's Kimberley region's summer rainfall is adequate, however, further studies were performed to explore the significant unexplained variability (i.e., residuals) by developing a hybrid model.

4.3 Hybrid Model Development

As single linear or nonlinear models are unable to explain all the underlying mechanisms involved in a complex rainfall generation system, therefore, GEP model residuals were fed into the ARIMAX models, resulting in novel hybrid model developments that offered enhanced rainfall forecasting for the region.

Once the linear component of the residual was obtained from the ARIMAX model, it was combined with the non-linear outcome of the GEP model to get the final

Table 3 Pearson correlation (r) results with the different model sets in ARIMAX and GEP

Model Set	Region	Station name	Pearson correlation (r) for different model sets										
			DMI-Nino3	DMI-Nino4	DMI-Nino3.4	DMI-SOI	DMI-EMI	WTIO-SOI	WTIO-Nino3	WTIO-Nino4			
ARIMAX	South Coast	Albany	0.60	0.55	0.57	0.62	---	---	---	---	---	---	---
		Mount Barker	0.67	0.59	0.57	0.55	---	---	---	---	---	---	---
		Grassmere	0.64	0.64	0.60	0.57	---	---	---	---	---	---	---
	North Coast	Busselton Shire	0.58	0.56	0.56	---	---	---	---	---	---	---	---
		Northampton	0.82	0.81	0.81	0.75	0.78	---	---	---	---	---	---
		Mingenew	0.56	0.55	0.54	0.56	0.54	---	---	---	---	---	---
		Nabawa	0.69	0.66	0.68	0.57	0.61	---	---	---	---	---	---
	Kimberley	Ogilvie	0.66	0.61	0.64	0.58	0.53	---	---	---	---	---	---
		Anaplains	---	---	---	---	---	0.84	---	---	---	---	0.84
		Bidyadanga	---	---	---	---	---	0.68	---	---	---	---	---
GEP	South Coast	Gogo Station	---	---	---	---	---	0.65	---	---	---	---	---
		Quantun Downs	0.47	---	---	0.50	---	0.53	0.44	---	---	---	---
		Albany	0.79	0.72	0.69	0.78	---	---	---	---	---	---	---
	North Coast	Mount Barker	0.78	0.69	0.72	0.67	---	---	---	---	---	---	---
		Grassmere	0.75	0.66	0.60	0.70	---	---	---	---	---	---	---
		Busselton Shire	0.79	0.74	0.74	---	---	---	---	---	---	---	---
	Kimberley	Northampton	0.82	0.68	0.67	0.69	0.63	---	---	---	---	---	---
		Mingenew	0.82	0.74	0.71	0.71	0.70	---	---	---	---	---	---
		Nabawa	0.80	0.67	0.71	0.70	0.79	---	---	---	---	---	---
		Ogilvie	0.79	0.72	0.79	0.75	0.46	---	---	---	---	---	0.74
Kimberley	Anaplains	---	---	---	---	---	0.82	---	---	---	---	---	
	Bidyadanga	---	---	---	---	---	0.85	---	---	---	---	---	
	Gogo Station	---	---	---	---	---	0.76	---	---	---	---	---	
	Quantun Downs	0.74	---	---	0.70	---	0.76	0.58	---	---	---	---	

Table 4 Output equation of best developed GEP model for the selected rainfall stations

Region	Rainfall Station	Model Set	Output Equation
South Coast	Albany	$DMI_{Occ} \cdot Nino3_{Nov}$	$\left[\cos(-8.29 * d_1) * \text{Max}\left\{5.61 - d_0, \frac{1}{d_0}\right\} + (8.06 - d_0)^2 \right] \left[\frac{d_0}{\text{Avg}\{\text{Log}1.48, \text{Max}(d_1, d_0)\} - \frac{\text{Log}1.48}{d_1 + 1.48}} \right]$ $+ \sin\left\{ [6.99 * d_0 - \text{Avg}(d_1, d_0)] * 6.45 \right\} + \left[7.53 - \frac{1}{d_1} - (d_0 - 0.35) + (\sin d_1 - d_0) \right]$ <p>Here, $d_0 = DMI_{Occ}$ and $d_1 = Nino3_{Nov}$</p>
	Mount banker	$DMI_{Occ} \cdot Nino3_{Nov}$	$\left[\cos\{(d_1 + 6.25) * (-12.94 + d_0)\} * \left\{ \left(\frac{1}{d_0} + 19.50 \right) * \text{Min}(d_1, d_0) \right\} \right] + \frac{1}{\cos\{\sqrt{d_0} * 2.49^2\} * \{3.85^3 + (-8.81 - d_1)\}}$ $+ + \frac{1}{\text{Avg}\{\cos(d_0 * -72.02), \frac{1}{d_0}\} * \text{Avg}\{(-4.32 * d_0), \sin 7.59\}} \left[\text{Min}(d_1, d_0) * (d_0 - 8.56) \right] - (d_0 - 9.69) + \text{Avg}\{-9.69^2, (d_1 - d_0)\}$ <p>Here, $d_0 = DMI_{Occ}$ and $d_1 = Nino3_{Nov}$</p>
	Grassmere	$DMI_{Occ} \cdot Nino3_{Nov}$	$\left[\frac{1}{\cos\{(-67.16 + \text{Exp}d_0) * \text{Avg}(-0.02, d_1^{-1})\}} \right] + \left[\{(1.17 - d_1)^2 * \text{Avg}(-0.97, d_0)\} * \text{Min}\{d_0^2, \text{Avg}(2.54, 3.11)\} \right]$ $+ [\cos d_0 * \{(\sin d_0 * -32.39) + (d_0^2 * 14.34)\}] + (-8.54)^2 + [d_0 * \{\text{Max}(3.75, d_0^{-1})\}^2]$ <p>Here, $d_0 = Nino3_{Nov}$ and $d_1 = DMI_{Occ}$</p>
	Busselton Shire	$DMI_{Occ} \cdot Nino3_{Nov}$	$d_1^4 - \left\{ \frac{1}{(d_1 - d_0)} + (-17.90 + d_1) * 3.83 \right\} + \left[\sqrt{6.03} * \sin\{(d_1 - 4.70) * (4.70 * 6.76)\}^2 \right]$ $+ \text{Max} \left[-10.67 - (d_0 + d_1)^3, \left\{ (-8.66 - 1.33) - \frac{d_0}{d_1} \right\}^3 \right] + \frac{\text{Exp}d_1}{\cos[\text{Avg}\{(d_0 + 6.01)^3, \frac{d_1^3}{6.76}\}]}$ <p>Here, $d_0 = DMI_{Occ}$ and $d_1 = Nino3_{Nov}$</p>

Table 4 (continued)

Region	Rainfall Station	Model Set	Output Equation
North Coast	Northampton	$DMI_{Jan_Nino3_{Nov}}$	$\left[\frac{Min[-0.40, \{(0.39 + d_1) + Max(-1.06, d_1)\}]}{Exp(-0.34) + \sqrt[3]{d_0}} \right] + [-9.05 - \{\sqrt[3]{d_0} + (d_0 + d_1)\} + \cos d_1^2]$ $+ [Exp\{Max(\frac{1}{4.42}, d_0) + d_0^3\} - \{(-4.05 + 3.29) + \sqrt{8.26}\}] + [\frac{\{(d_1 - d_0) - (d_0 * 2.71)\} - \{(-0.29 + d_0)\}}{(d_1 * d_1) + \frac{5.26}{(-8.17)}}]$
			Here, $d_0 = DMI_{Jan}$, and $d_1 = Nino3_{Nov}$
Mingenew		$DMI_{Nov_Nino3_{Nov}}$	$\left[\left\{ \frac{d_1}{Avg(d_1 * 6.41)} - (d_0 * -5.54) \right\} * Min\{(1.95 - d_0), (d_0 - d_1)\} + \sin\left(\frac{-8.99}{d_1}\right) * \{(d_0 * -2.85) + 9.91\} - \sin 5.09] \right]$ $+ [\{-6.30^2 + (d_0 - d_1)\} - d_1] * \cos\left(\cos\frac{-8.89}{d_1}\right) + \frac{1}{\{(d_1 * d_0) - \frac{1}{0.99}\} - \{\sin 5.29 + Max(d_0, d_1)\}}$
			Here, $d_0 = Nino3_{Nov}$, and $d_1 = DMI_{Nov}$
Nabawa		$DMI_{Feb_Nino3_{Nov}}$	$\left[-4.46 + \sqrt[3]{\cos\frac{(d_0 + 2.36)}{d_1} - (d_0 + d_0)} \right]^2 + \sqrt[3]{\frac{[\{(d_1 * (-9.33) + d_0) + \sqrt[3]{(-1.72)}]}{Avg\{(d_0 + 0.21), (d_0 * 0.21)\}} + [Exp\sqrt[3]{Min\{(-7.10 * d_1), 0.96\}}]^2 - \left\{ \frac{Min(1.79, d_0)}{d_1} \right\}]}$ $+ [\cos(d_0 + 5.13)^3 - Avg\{(\sqrt[3]{d_0} - 4.67), Avg(d_0, 3.53)\}]^3]$
			Here, $d_0 = Nino3_{Nov}$, and $d_1 = DMI_{Feb}$
Ogilvie		$DMI_{Feb_Nino3_{Nov}}$	$Max\left[Exp\left\{ Min\left(-3.64, \frac{-1.24}{d_0}\right) \right\} * \frac{(14.44 + 14.44)}{(-0.16 - d_0)} \right] + [d_0^2 + Exp\{d_1 - Avg(\sqrt[3]{d_0}, -4.36)\}] - d_1]$ $+ [[d_0^3 + Exp\{d_1 - Avg(\sqrt[3]{d_0}, -4.36)\}] - d_1] + [\frac{(d_1^3 - d_0^3)}{7.46 - Max\{(d_0 * -7.71), 7\}}] + [\cos\{Max(\frac{d_0 * 1.70}{\sqrt[3]{d_1}}, \frac{-6.36}{d_0})\} - 1.70] * (-12.70)$
			Here, $d_0 = Nino3_{Nov}$, and $d_1 = DMI_{Feb}$

Table 4 (continued)

Region	Rainfall Station	Model Set	Output Equation
Kimberley	Anna Plains	WTIO _{Aug} -SOL _{Mar}	$25 * \left[\sin(6.93d_0 + \sqrt[3]{d_1}) - \sqrt[3]{d_1} \right] + \left[\frac{3.65}{\sin(d_1) - \sin(5.25)} * \{\cos(d_1) - \exp(d_0)\} \right] + \exp(4.78)$ $+ \left[\frac{d_1}{\sqrt[3]{(d_1 + d_0) * (d_0 - 2.23)}} - \sqrt[3]{2.38} \right] + \left[\text{Max} \left\{ \exp(3.49), \frac{1}{d_0} \right\} + \{(-3.14)^3 * d_0\} \right] * \sqrt[3]{-1.81 - d_1}$ <p>Here, $d_0 = WTIO_{Aug}$ and $d_1 = SOL_{Mar}$</p>
Bidyadanga		WTIO _{Aug} -SOL _{May}	$\frac{1}{\cos \left[\frac{2.46}{\text{Min}(d_1, \cos d_1)} \right]^2} + \frac{1}{\cos \left[\frac{5.96}{\text{Min}(d_0 + 5.96d_1, d_1)} \right]^2} + \frac{1}{\cos[\text{Avg}(d_1 * (8.1 - d_0), d_0 * (78.03 + d_0))]} + \{\text{Min} \left(3.94, \frac{1}{-7 * d_0} \right) + 2.92\} * d_1$ $+ \sqrt{\exp\{\text{Min}(d_0, \cos d_1^3)\}} * 9.74^2$ <p>Here, $d_0 = WTIO_{Aug}$ and $d_1 = SOL_{May}$</p>
Gogo Station		WTIO _{Aug} -SOL _{Mar}	$\left[\text{Avg} \left\{ -14.79 * (d_1 - 4.35), \text{Max} \left(-\frac{2.96}{d_1}, -4.5 \right) \right\} * \sqrt[3]{d_0} \right] + \left[\text{Avg} \left\{ -11.22, \sin(-6.25 * d_1^2) * \exp(1.35)^2 \right\} \right]^2 + \frac{1}{\text{Log} \left(\frac{\text{Min}(d_0}{25}, \sqrt[3]{d_1} \right) + d_0}$ $+ \left[\frac{d_1 + 3.65}{\text{Avg}(d_1, 2.19)} - \exp(d_0) \right] * d_1 + \left[\{-7.52 + \text{Min}(0.49, d_1)\}^2 - \{ \sqrt[3]{\sin(d_1)} * (-18.19 * d_1) \} \right]$ <p>Here, $d_0 = WTIO_{Aug}$ and $d_1 = SOL_{Mar}$</p>
Quanbun Downs		WTIO _{Oct} -SOL _{Nov}	$\text{Avg} \left[\left\{ \sin(d_1), d_1^2 \right\} - 3.91 \right] * d_1^4 + \frac{1}{\text{Avg} \left\{ d_1^2, (d_0 - 3.02) \right\} * \text{Avg} \left\{ d_1^3, \sqrt[3]{(-0.94)} \right\}} + \{ \text{Avg}(7.64, d_0) \}^3 * 2.26$ $+ \text{Max} \left[\text{Avg} \left\{ (-25.83 * d_1), \frac{1}{d_0} \right\} * \frac{1}{d_1}, \frac{1}{\text{Avg}(d_1, 0.58)} \right] + \text{Min} \left[-21.93, [-21.93 * 4.22 * \{ \text{Avg}(d_1, d_0) + d_0 \}] \right]$ <p>Here, $d_0 = WTIO_{Oct}$ and $d_1 = SOL_{Nov}$</p>

forecast. Performance evaluators such as correlation statistics, error, and fitness of model statistics such as NSE and d_r , were considered to evaluate the hybrid model's performance. Once the hybrid (GEP-ARIMAX) model was developed, a validation test was performed for the selected model set. Table 5 presents the model description for the developed hybrid models in both calibration and validation periods. It is to be noted that an alternative approach (ARIMAX-GEP) to the above methodology has also been evaluated, however, only the best model outcomes (i.e., GEP-ARIMAX) are presented in this paper.

From Table 5, it is evident that for the south coast region, the Pearson correlation (r) increased in the validation period for all the selected stations except for Busselton shire. For all the stations, Pearson correlation (r) values ranged from 0.87 to 0.91 and 0.88 to 0.92 in the calibration and validation period, respectively. However, a decrease in the refined Willmot index of agreement (d_r) was reported in the validation period for all the selected stations except for Albany. For all these stations, a refined Willmot index of agreement (d_r) values ranged from 0.74 to 0.78 and 0.69 to 0.77 in the calibration and validation period. Simultaneously, a relatively high Pearson correlation (r) and refined Willmot index of agreement (d_r) value along with relatively low $RMSE$, MAE , $RRSE$, and RAE values was found in the validation period compared to the calibration period. The observation indicates that the developed models are good prediction models apart from their relatively high r and positive d_r values. Also, for the calibration and validation periods, the NSE values ranged from 0.68 to 0.75 and 0.59 to 0.77, respectively, suggesting the model is a good fit. All these outcomes indicate that the developed hybrid model is a good prediction model. Moreover, these models showed a rainfall prediction capability of up to four months in advance for the region.

For the north coast region, the Pearson correlation (r) was found to be increased in the validation period for all the selected stations except for Mingenew and Ogilvie. A similar observation was made for the refined Willmot index of agreement (d_r) values as a decrease for Mingenew and Ogilvie were reported in the validation period. Furthermore, relatively low $RMSE$, MAE , $RRSE$, and RAE values indicate the goodness of the developed models for both calibration and validation periods. The reported NSE values ranged from 0.78 to 0.86 and 0.72 to 0.83 in the calibration and validation period, respectively, indicating a good fit for all the models. Similarly, the reported refined Willmott index of agreement (d_r) value greater than 0.70 in both calibration and validation period also indicates the skillfulness of the developed model. The developed hybrid model showed seasonal autumn rainfall predictability up to four months in advance for Mingenew, two months in advance for Northampton, and only one month in advance for Nabawa and Ogilvie.

For the Kimberley region, the Pearson correlation (r) and refined Willmot index of agreement (d_r) were found to increase in the validation period for all the selected rainfall stations except for Anna Plains. For all these models, relatively low $RMSE$, MAE , $RRSE$, and RAE values confirm them as good prediction models. For the stations located in the region, the NSE values ranged from 0.69 to 0.84 and 0.72 to 0.84 in the calibration and validation period, respectively. Furthermore, in both calibration and validation periods, the refined Willmott index of agreement (d_r) value close to or more than 0.70 has been reported. All the developed hybrid models showed their rainfall prediction capability up to four months in advance except for the station- Quanbun Downs, in which the prediction was deemed possible only up to one month in advance.

Table 5 Model description of hybrid (GEP-ARIMAX) model in the calibration and validation period

Region	Station Name	Predictors	Pearson's Correlation (<i>r</i>)		Refined Willmot Index of Agreement (<i>d_r</i>)		RMSE		MAE		NSE	
			Cal	Val	Cal	Val	Cal	Val	Cal	Val	Cal	Val
South Coast	Albany	DMI _{Oct} -Nino3 _{Nov}	0.87	0.92	0.74	0.77	11.50	8.04	7.89	6.85	0.68	0.77
	Mount Barker	DMI _{Oct} -Nino3 _{Nov}	0.90	0.91	0.75	0.74	10.08	11.03	8.14	8.41	0.74	0.75
	Grassmere	DMI _{Oct} -Nino3 _{Nov}	0.89	0.91	0.75	0.70	11.44	11.30	8.54	8.80	0.71	0.66
North Coast	Busselton Shire	DMI _{Oct} -Nino3 _{Nov}	0.91	0.88	0.78	0.69	12.09	13.68	9.10	10.84	0.75	0.59
	Northampton	DMI _{Jan} -Nino3 _{Nov}	0.92	0.96	0.77	0.77	10.03	7.42	7.28	5.86	0.78	0.82
	Mingenew	DMI _{Nov} -Nino3 _{Nov}	0.92	0.90	0.77	0.70	8.23	10.27	6.02	7.73	0.78	0.72
Kimberley	Nabawa	DMI _{Feb} -Nino3 _{Nov}	0.93	0.93	0.77	0.73	8.02	9.40	6.43	7.45	0.80	0.75
	Ogilvie	DMI _{Feb} -Nino3 _{Nov}	0.95	0.94	0.81	0.81	7.40	7.91	5.64	5.77	0.86	0.83
	Anna Plains	WIO _{Aug} -SOI _{Mar}	0.94	0.89	0.79	0.78	26.37	34.92	19.68	25.69	0.83	0.74
	Bidyadanga	WIO _{Aug} -SOI _{May}	0.94	0.94	0.82	0.84	27.82	25.74	19.27	17.82	0.84	0.84
	Gogo Station	WIO _{Aug} -SOI _{Mar}	0.90	0.91	0.76	0.75	24.47	30.49	18.98	23.89	0.74	0.72
	Quambun Downs	WIO _{Oct} -SOI _{Nov}	0.88	0.92	0.72	0.80	29.62	25.92	23.79	21.07	0.69	0.81

5 Comparison of Model Performance

A comparative analysis of the statistical performance of different models in both calibration and validation periods has been presented in Table 6. Table 6 also presents the statistical performance parameters *RMSE*, *MAE*, and refined Willmot index of agreement (d_r) for the selected model sets.

For the south coast region, the $DMI_{Oct}-Nino3_{Nov}$ model developed using the ARIMAX technique showed its capability to forecast autumn rainfall up to four months in advance with Pearson correlation (r) values ranging from 0.58 to 0.60 and 0.62 to 0.80 in the calibration and validation period, respectively. For the GEP model, the same model set showed its capability of forecasting up to four months in advance with a correlation value ranging from 0.75 to 0.79 and 0.75 to 0.82 in the calibration and validation period, respectively. In contrast, the hybrid (GEP-ARIMAX) model showed much better forecasting capability with correlation values ranging from 0.87 to 0.91 and 0.88 to 0.92 during the calibration and validation period.

The $DMI-Nino3$ model set showed its ability to forecast autumn rainfall up to four months in advance with different lagged months for different rainfall stations in the north coast region. For Northampton, the $DMI_{Jan}-Nino3_{Nov}$ model set developed using ARIMAX showed its capability to forecast autumn rainfall up to two months in advance with a Pearson correlation (r) value of 0.82 and 0.70 in the calibration and validation period, respectively. The same model set developed using the GEP technique showed its capability of forecasting autumn rainfall with a correlation value of 0.82 and 0.84 in the calibration and validation period. However, the $DMI_{Jan}-Nino3_{Nov}$ model developed using the hybrid (GEP-ARIMAX) technique outperformed both ARIMAX and GEP models and showed better forecasting capability with a significant correlation of 0.92 and 0.96 in the calibration and validation period respectively.

For the Kimberley region, the WTIO-SOI model set showed its ability to forecast summer rainfall with different lagged months for different rainfall stations. For Anna Plains, the $WTIO_{Aug}-SOI_{Mar}$ model set developed using the ARIMAX technique showed its capability of forecasting up to four months in advance with a correlation of 0.83 and 0.65 in the calibration and validation period, respectively. The $WTIO_{Aug}-SOI_{Mar}$ model developed using the GEP technique showed similar forecasting capability to ARIMAX (i.e., four months in advance), with a significant correlation value of 0.82 during the calibration and validation period. The same model set (i.e., the $WTIO_{Aug}-SOI_{Mar}$) developed using the hybrid (GEP-ARIMAX) technique showed much better forecasting capability than ARIMAX and GEP. A significantly high correlation value ranging from 0.94 and 0.89 in both the calibration and validation period was obtained. For Bidyadanga, the $WTIO_{Aug}-SOI_{May}$ model set developed using the ARIMAX, and GEP techniques showed promising performance, as the ARIMAX model returned a correlation value of 0.68 and 0.77, wherein for the GEP model, the correlation values obtained were as high as 0.85 and 0.87 for the calibration and validation period, respectively. Both the model sets demonstrated their prediction capability up to four months in advance. On the other hand, the developed hybrid (GEP-ARIMAX) model resulted in a significantly high correlation of 0.94 in both the calibration and validation period. While Pearson correlation (r) values are compared, a Pearson correlation (r) value of more than 0.5 is considered a significant effect (Field 2013).

The refined Willmot index of agreement (d_r) value is an indicator of the model fitness, where a relatively high positive value indicates a good fit (Willmott et al. 2012). For the developed hybrid model, d_r value in the calibration and validation period were ranging

Table 6 Comparison of Pearson's Correlation, refined Willmot index of agreement, RMSE and MAE between ARIMAX, GEP, and hybrid model

Region	Station Name	Model Type	Lag Month	Pearson's Correlation (<i>r</i>)		Refined Willmot Index of Agreement (<i>d_r</i>)		RMSE		MAE	
				Cal	Val	Cal	Val	Cal	Val	Cal	Val
				South Coast	Albany (DMI _{Oct} -Nino3 _{Nov})	ARIMAX	4	0.60	0.80	0.61	0.71
		GEP		0.79	0.82	0.69	0.70	14.44	8.71	11.39	8.71
		Hybrid		0.87	0.92	0.74	0.77	11.50	8.04	7.89	6.85
	Mount Barker (DMI _{Oct} -Nino3 _{Nov})	ARIMAX	4	0.67	0.66	0.61	0.60	16.92	16.61	12.13	12.89
		GEP		0.78	0.75	0.69	0.66	12.89	11.20	10.43	11.20
		Hybrid		0.90	0.91	0.75	0.74	10.08	11.03	8.14	8.41
	Grassmere (DMI _{Oct} -Nino3 _{Nov})	ARIMAX	4	0.64	0.72	0.61	0.64	18.85	13.28	14.48	10.62
		GEP		0.75	0.77	0.67	0.66	15.42	10.24	12.14	10.24
		Hybrid		0.89	0.91	0.75	0.70	11.44	11.30	8.54	8.80
	Busseton Shire (DMI _{Oct} -Nino3 _{Nov})	ARIMAX	4	0.58	0.62	0.61	0.61	21.58	16.87	15.56	13.69
		GEP		0.79	0.78	0.70	0.64	14.94	12.50	12.24	12.50
		Hybrid		0.91	0.88	0.78	0.69	12.09	13.68	9.10	10.84

Table 6 (continued)

Region	Station Name	Model Type	Lag Month	Pearson's Correlation (r)		Refined Willmot Index of Agreement (d_r)		RMSE		MAE	
				Cal	Val	Cal	Val	Cal	Val	Cal	Val
North Coast	Northampton (DMI _{Jan} -Nino3 _{Nov})	ARIMAX	2	0.82	0.70	0.70	0.61	13.59	12.37	9.79	9.90
		GEP		0.82	0.84	0.69	0.68	12.27	8.11	9.84	8.11
	Mingenew (DMI _{Nov} -Nino3 _{Nov})	Hybrid		0.92	0.96	0.77	0.77	10.03	7.42	7.28	5.86
		ARIMAX	4	0.56	0.62	0.56	0.50	15.55	15.65	11.53	12.63
	Nabawa (DMI _{Feb} -Nino3 _{Nov})	GEP		0.80	0.83	0.71	0.62	10.01	6.61	7.49	6.61
		Hybrid		0.92	0.90	0.77	0.70	8.23	10.27	6.02	7.73
	Ogilvie (DMI _{Feb} -Nino3 _{Nov})	ARIMAX	1	0.69	0.66	0.63	0.56	14.19	14.16	10.26	11.80
		GEP		0.80	0.82	0.68	0.65	10.69	9.39	9.06	9.39
		Hybrid		0.93	0.93	0.77	0.73	8.02	9.40	6.43	7.45
		ARIMAX	1	0.66	0.68	0.61	0.64	16.17	14.30	11.66	10.89
		GEP		0.79	0.78	0.71	0.65	12.28	12.68	8.76	9.90
		Hybrid		0.95	0.94	0.81	0.81	7.40	7.91	5.64	5.77

Table 6 (continued)

Region	Station Name	Model Type	Lag Month	Pearson's Correlation (<i>r</i>)		Refined Willmot Index of Agreement (<i>d_r</i>)		RMSE		MAE	
				Cal	Val	Cal	Val	Cal	Val	Cal	Val
Kimberley	Anna Plains (WTIO _{Aug} -SOI _{Mar})	ARIMAX	4	0.83	0.65	0.68	0.64	39.94	53.54	29.51	40.97
		GEP		0.82	0.82	0.70	0.75	36.74	39.63	27.52	28.96
	Bidiyadanga (WTIO _{Aug} -SOI _{May})	Hybrid		0.94	0.89	0.79	0.78	26.37	34.92	19.68	25.69
		ARIMAX	4	0.68	0.77	0.63	0.70	57.90	38.77	41.97	32.96
	Gogo Station (WTIO _{Aug} -SOI _{Mar})	GEP		0.85	0.87	0.75	0.74	36.92	34.99	27.20	28.20
		Hybrid		0.94	0.94	0.82	0.84	27.82	25.74	19.27	17.82
	Quanbun Downs (WTIO _{Oct} -SOI _{Nov})	ARIMAX	4	0.65	0.74	0.66	0.70	40.01	40.33	27.49	28.86
		GEP		0.76	0.78	0.69	0.68	30.58	37.01	23.95	31.06
		Hybrid		0.90	0.91	0.76	0.75	24.47	30.49	18.98	23.89
			ARIMAX	1	0.53	0.59	0.57	0.63	43.56	49.07	36.33
		GEP		0.76	0.74	0.69	0.65	35.82	47.06	27.56	35.76
		Hybrid		0.88	0.92	0.72	0.80	29.62	25.92	23.79	21.07

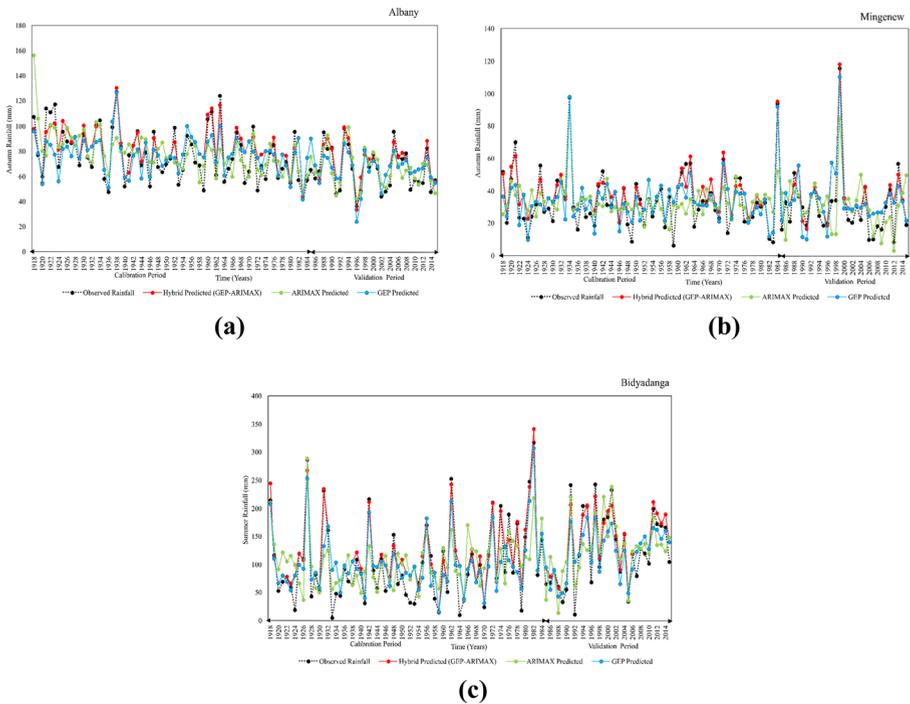


Fig. 3 Comparison of prediction performance of the developed best ARIMAX, GEP, and hybrid (GEP-ARIMAX) model for **a** Albany **b** Mingenew; and **c** Bidiyadanga

from 0.72 to 0.82 and 0.69 to 0.84 respectively. Also, the developed hybrid model returned relatively low *RMSE* and *MAE* values in both the calibration and validation periods than the ARIMAX and GEP models. An observed vs. predicted plots for ARIMAX, GEP, and hybrid models are presented in Fig. 3 for Albany, Mingenew, and Bidiyadanga. From observed vs. predicted plots, it is apparent that the hybrid (GEP-ARIMAX) model’s prediction performance demonstrated a similar pattern to the naturally occurring rainfall events. It is also notable that the developed hybrid model can predict all the severe rainfall cases; however, it could not predict some drought cases.

6 Conclusion

The findings of this study demonstrated that for WA’s south and north coast regions, a combined effect of DMI and Nino3 has a noticeable impact on seasonal autumn rainfall, whereas a different climate index set, namely WTIO-SOI, showed a significant contribution to the Kimberley region’s seasonal summer rainfall. To develop the rainfall forecast models for these regions, both linear (ARIMAX) and non-linear (GEP) relationships were evaluated. Even though the GEP model’s capability to predict seasonal autumn and summer rainfall for the selected regions is comparatively higher than ARIMAX models, residuals from the GEP models were fed into ARIMAX models, and the developed hybrid model showed improved forecasting for all the regions.

For the south coast region, the maximum correlation obtained in the ARIMAX models was reported as 0.64 for Grassmere, which is relatively lower than the minimum correlation found in the GEP models (0.75 for Grassmere) and hybrid models (0.87 for Albany). For the north coast region, the maximum correlation obtained in both the ARIMAX and GEP models was 0.82 for Northampton, which is relatively lower than the maximum correlation obtained for hybrid models (0.95 for Ogilvie). Similarly, the hybrid models developed for the Kimberley region stations showed correlation values as high as 0.94 for Anna Plains and Bidyadanga, whereas, relatively low correlations were achieved in ARIMAX and GEP, in particular for Gogo Station and Quanbun Downs. A comparatively high positive Willmott index of agreement (d_r) values for all selected stations were also evident.

An overview of the error measurement for the developed models indicates that comparatively low RMSE and MAE values were obtained for hybrid models for all selected stations. From the observed vs predicted plots, many peaks and troughs were also well captured by the hybrid model if compared to ARIMAX and GEP. All these developed models have also shown robust prediction capability, as forecasting the rainfall as early as 1 month in advance is possible for the south coast (Nabawa and Ogilvie) and Kimberley (Quanbun Downs) regions. This is expected to offer greater flexibility in economic decision-making principles and better management of the agricultural and water resources. Furthermore, the developed models are expected to assist in disaster risk management and abnegating associated costly remediation, hence, creating robust disaster recovery and economic preparedness plans for Western Australia.

Authors Contributions F. Islam: Data collection, data analysis, model development and draft paper writing; M.A. Imteaz: Conceptualisation, supervision, project management.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. The study was not funded by any organisation/funder.

Availability of Data and Materials Data used in this study will be made available upon request.

Declarations

Ethical Approval Not required as the study did not involve human or animal.

Consent to Participate Authors have consent to participate any offer by the journal.

Consent to Publish Authors are giving consent to publish the article in the submitted journal.

Competing Interests There is no competing or conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Al-Juboori AM (2022) Solving complex rainfall-runoff processes in semi-arid regions using hybrid heuristic model. *Water Resour Manag* 1–12
- Ashok K, Guan Z, Yamagata T (2003) Influence of the Indian Ocean Dipole on the Australian winter rainfall. *Geophys Res Lett* 30:1821
- Bagirov AM, Mahmood A (2018) A comparative assessment of models to predict monthly rainfall in Australia. *Water Resour Manag* 32:1777–1794
- Barbulescu A, Bautu E (2009) Alternative models in precipitation analysis. *An St Univ Ovidius Math* 17:45–68
- Calado GG, Valverde MC, Baigorria GA (2019) Use of teleconnection indices for water management in the cantareira system - São Paulo – Brazil. *Environ Process* 6:413–431
- Evans FH, Guthrie MM, Foster I (2020) Accuracy of six years of operational statistical seasonal forecasts of rainfall in Western Australia (2013 to 2018). *Atmos Res* 233:104697
- Feng J, Li J, Li Y, Zhu J, Xie F (2015) Relationships among the monsoon-like southwest Australian circulation, the southern annular mode, and winter rainfall over southwest western Australia. *Adv Atmos Sci* 32:1063–1076
- Ferranti L (2012) Calibration and validation of seasonal forecasts. ECMWF Seminar on Seasonal Prediction, Shinfield Park, Reading
- Ferreira C (2001) Gene expression programming: a new adaptive algorithm for solving problems. *Complex Syst* 13:87–129
- Field A (2013) *Discovering statistics using IBM SPSS statistics*. Sage, Washington DC
- Fierro AO, Leslie LM (2013) Links between central west Western Australian rainfall variability and large-scale climate drivers. *J Clim* 26:2222–2246
- Ghamariadyan M, Imteaz MA (2021) Prediction of seasonal rainfall with one-year lead time using climate indices: a wavelet neural network scheme. *Water Resour Manag* 35:5347–5365
- Hashmi MZ, Shamseldin AY, Melville BW (2011) Statistical downscaling of watershed precipitation using Gene Expression Programming (GEP). *Environ Model Softw* 26:1639–1646
- Hossain I, Rasel H, Imteaz MA, Mekanik F (2020) Long-term seasonal rainfall forecasting using linear and non-linear modelling approaches: A case study for Western Australia. *Meteorol Atmos Phys* 132:131–141
- Islam F, Imteaz MA (2019) Development of prediction model for forecasting rainfall in Western Australia using lagged climate indices. *Int J Water* 13:248–268
- Islam F, Imteaz MA (2020) Use of teleconnections to predict western australian seasonal rainfall using ARIMAX model. *Hydrology* 7:52
- Karimi S, Shiri J, Kisi O, Shiri AA (2016) Short-term and long-term streamflow prediction by using “wavelet–gene expression” programming approach. *J Hydraul Eng* 22:148–162
- Kumar M, Sahay RR (2018) Wavelet-genetic programming conjunction model for flood forecasting in rivers. *Hydrol Res* 49:1880–1889
- Lin Z, Li Y (2012) Remote influence of the tropical Atlantic on the variability and trend in North West Australia summer rainfall. *J Clim* 25:2408–2420
- Ljung GM, Box GE (1978) On a measure of lack of fit in time series models. *Biometrika* 65:297–303
- Marshall A, Hendon H (2014) Impacts of the MJO in the Indian Ocean and on the Western Australian coast. *Clim Dyn* 42:579–595
- Mehdizadeh S, Sales AK (2018) A comparative study of autoregressive, autoregressive moving average, gene expression programming and Bayesian networks for estimating monthly streamflow. *Water Resour Manag* 32:3001–3022
- Mehr AD (2018) An improved gene expression programming model for streamflow forecasting in intermittent streams. *J Hydrol* 563:669–678
- Montazerolghaem M, Vervoort W, Minasny B, McBratney A (2016) Long-term variability of the leading seasonal modes of rainfall in south-eastern Australia. *Weather Clim Extremes* 13:1–14
- Mozaffari S, Javadi S, Moghaddam HK, Randhir TO (2022) Forecasting groundwater levels using a hybrid of support vector regression and particle swarm optimization. *Water Resour Manag* 1–18
- Risbey JS, Pook MJ, McIntosh PC, Wheeler MC, Hendon HH (2009) On the remote drivers of rainfall variability in Australia. *Mon Weather Rev* 137:3233–3253
- Saigal S, Mehrotra D (2012) Performance comparison of time series data using predictive data mining techniques. *Adv Inf Min* 4:57–66
- Shabani S, Candelieri A, Archetti F, Naser G (2018) Gene expression programming coupled with unsupervised learning: a two-stage learning process in multi-scale, short-term water demand forecasts. *Water* 10:142

- Sharghi E, Nourani V, Najafi H, Soleimani S (2019) Wavelet-exponential smoothing: a new hybrid method for suspended sediment load modeling. *Environ Process* 6:191–218
- Shi G, Cai W, Cowan T, Ribbe J, Rotstayn L, Dix M (2008) Variability and trend of North West Australia rainfall: observations and coupled climate modeling. *J Clim* 21:2938–2959
- Singh J, Knapp HV, Arnold J, Demissie M (2005) Hydrological modeling of the Iroquois river watershed using HSPF and SWAT 1. *J Am Water Resour Assoc* 41:343–360
- Taschetto AS, England MH (2009) El Niño Modoki impacts on Australian rainfall. *J Clim* 22:3167–3174
- Ummenhofer CC, Sen Gupta A, Pook MJ, England MH (2008) Anomalous rainfall over southwest Western Australia forced by Indian Ocean sea surface temperatures. *J Clim* 21:5113–5134
- Willmott CJ, Robeson SM, Matsuura K (2012) A refined index of model performance. *Int J Climatol* 32:2088–2094
- Wu X, Zhou J, Yu H, Liu D, Xie K, Chen Y, Hu J, Sun H, Xing F (2021) The development of a hybrid wavelet-ARIMA-LSTM model for precipitation amounts and drought analysis. *Atmosphere* 12:74
- Xu W, Peng H, Zeng X, Zhou F, Tian X, Peng X (2019) A hybrid modelling method for time series forecasting based on a linear regression model and deep learning. *Appl Intell* 49:3002–3015
- Zhang GP (2003) Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 50:159–175

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.