DOI:10.28974/idojaras.2023.3.4

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 127, No. 3, July – September, 2023, pp. 321–346

Impacts of large scale climate drivers on precipitation in Sindh, Pakistan using machine learning techniques

Sapna Tajbar*, **Ali Mohammad Khorshiddoust**, and **Saeed Jahanbakhsh Asl**

Department of Climatology
University of Tabriz
Tabriz 51664, East Azerbaijan, Iran

**Corresponding author E–mail: sapnatajbar@gmail.com*

(Manuscript received in final form August 8, 2022)

Abstract— Sindh province of Pakistan has a long history of severe droughts. Several large scale climate drivers (LSCD) are known for their effect on precipitation worldwide but studies in the Sindh region are missing; wide variety of LSCDs and lagged associative information. This study aimed to identify the significant LSCDs in Sindh province of Pakistan and improve the forecast skill of monthly precipitation by employing the principal component analysis (PCA), artificial neural network (ANN), Bayesian regularization neural network (BRNN), and multiple regression analysis (MRA), while considering the 12 months lagged LSCDs such as Nino-1+2, Nino-3, Nino-3.4, Nino-4, Quasi-Biennial Oscillation (QBO) at 30 and 50hPa (QBOI and QBOII), sea surface temperature (SST), 2m air temperature (T2M), 500 hPa and 850 hPa geopotential heights (H500 and H850), surface and 500 hPa zonal velocity (SU and U500), latent and sensible heat fluxes over land (LHFOL and SHFOL), and surface specific humidity (SSH). Global Land Data Assimilation System (GLDAS), Tropical Rainfall Measuring Mission (TRMM), Modern-Era Retrospective Analysis for Research and Application (MERRA-2), NOAA, Freie University Berlin, and Hadley Centre Sea Ice and Sea Surface Temperature (HadISST) datasets were used. Results manifested that significant LSCDs with 99% confidence level were SU, U500, T2M, SST, SHFOL, LHFOL, SSH, and H850. During test period, compared with MR models of 0.39 to 0.64 and principal components of 0.31 to 0.57, the ANN and BRNN models had better predictive skills with correlation coefficients of 0.57 to 0.83 and 0.52 to 0.76, respectively. It can be concluded that the ANN and BRNN models enable us to predict monthly precipitation in Sindh region with lagged LSCDs.

Key-words: artificial neural network, Bayesian regularization neural network, LSCD, multiple regression analysis, sea surface temperature, Sindh province of Pakistan

1. Introduction

Precipitation is a complex global atmospheric process, which is dependent on space and time, its increase or decrease can have several impacts on the society in the form of floods and droughts, and is not easy to predict (*Kumar et al.,* 2021). Due to the visible random characteristics of precipitation series, they are mostly described by a stochastic process (*Chinchorkar et al.,* 2012). It plays an important role in the economy of Pakistan being an agricultural country (*Aamir* and *Hassan,* 2018), where agriculture sector contributes 26% to the gross domestic product (*Rehman et al.,* 2015). In Pakistan, precipitation as well as thermal regimes have experienced variations particularly in the recent couple of decades when sharp jump of global atmospheric temperatures was noticed (*Rasul et al.,* 2012). The southern part of Pakistan which comprises of Sindh and Balochistan provinces has an arid climate and usually receive less amount of precipitation throughout the year as compared to other parts of the country. This study's main focus is Sindh province which has a long history of severe droughts and is the second important province in terms of agriculture. It has a considerable agricultural base along the Indus River (*Solomon,* 2019). Annual total rainfall is only 160 mm. Very less amount falls during the winter season (13 mm normal), while chief amount falls during the summer season and is highly variable. Its rainfall is a result of the monsoon depressions which forms in the Bay of Bengal and occasionally moves westward into the lower Sindh (*Muslehuddin* and *Faisal,* 2006).

The first and important stage for the management of water resources at any region is to identify the prospective climate variables which are affecting the future water resource situation (*Gholami Rostam et al.,* 2020). Worldwide, various large scale climate phenomena influence the occurrence of precipitation (*Hossain et al.,* 2015), and numerous studies have been carried out in this context, but such studies over the Sindh province of Pakistan are few (*Rashid,* 2004; *Mahmood et al.,* 2006; *Sarfaraz,* 2007; *Iqbal* and *Athar,* 2018). *Sarfaraz* (2007) discussed the monsoon over Pakistan by investigating its features and components that is El Nino Southern Oscillation (ENSO), heat low, Southern Oscillation Index, Tropical Easterly Jet, Low-level Jet, westerly Sub-tropical Jet, Tibetan anticyclone over Himalayas at 200 hPa, and high pressure region over the Indian Ocean. *Rashid* (2004) plotted normalized standard deviation (NSD) versus time to know the association between distinct intensities of El Nino and rainfall. *Mahmood et al.* (2006) dealt with the impacts of El Nino on summer monsoon rainfall over Pakistan and calculated percent rainfall departure and correlation coefficients. They found that the deficiency in rainfall is significant during August in Sindh province. *Iqbal* and *Athar* (2018) studied the variability and teleconnections of precipitation over Pakistan using percentile analysis, Mann-Kendall trend test, Sen's slope estimator, and correlation techniques. They found that ENSO did not show significant impact.

Various researches around the world identified different LSCDs and used Machine learning techniques such as artificial neural networks (ANN), Bayesian regularization neural networks (BRNN), and multiple regression models for the prediction of precipitation (*Awan* and *Maqbool,* 2010; *Shukla et al.,* 2011; *Mekanik et al.,* 2013; *Venkata Ramana et al.,* 2013; *Ahmadi et al.,* 2014; *Kashiwao et al.,* 2017; *Lee et al.,* 2018; *Doranalu Chandrashekar et al.,* 2019). In the current study, artificial neural networks and Bayesian regularization neural networks are selected as they are known for their capability of finding the complex non-linear associations between the input and output parameters without taking into consideration the nature of the physical processes. As the processes influencing the rainfall are non-linear and highly complex, therefore they can help in solving the complexity (*Adamowski* and *Sun,* 2010).

To the best of author's knowledge, there are no detailed studies in Sindh province in this regard. Past studies in the region are limited in their approaches; firstly because of the absence of wide variety of LSCDs. The El Nino Southern Oscillation with some other climate variables has been discussed, and only one study analyzed the effect of the Quasi-Biennial Oscillation(QBO) phenomena on precipitation which is known for its impacts in different regions around the world. Secondly, none of the studies investigated the time extent in which the LSCDs effect the prediction of monthly precipitation. To forecast precipitation, it is essential to have the lagged associative information. In addition to this, the usage of LSCDs as potential predictors for the future precipitation has not given much attention. Finally, the monthly prediction of precipitation is not carried out using powerful machine learning techniques while utilizing LSCDs; the studies have either used linear regression analysis or probabilistic/categorical analysis. Therefore, there is a need of further research.

The main objective of the present study is to identify the significant LSCDs in Sindh province of Pakistan and improve the forecast skill of monthly precipitation by application of the principal component analysis, artificial neural network, Bayesian regularization neural network and multiple regression analysis while considering the lagged association of LSCDs. This paper is arranged in four sections; Section 1 gives a brief introduction, Section 2 explains the study area and data used with methods of prediction. Section 3 presents the results of the application of models and the discussion of the results. Concluding remarks are mentioned in Section 4.

2. Materials and methods

2.1. Study area

Sindh province is located between 23–35° and 28–30°N and 66–42° and 71–1°E (*Muslehuddin* and *Faisal,* 2006). It is bordered by Punjab in the north, the Indian states of Gujarat and Rajasthan in the east, the Arabian Sea in the south, and

Balochistan in the west (*Fig. 1*). Its landscape comprises mostly of alluvial plains surrounding the Indus River, but along with the border with India and mountains of Kirthar in the west side, it encompasses the Thar desert (*Solomon,* 2019). The province is located in a region that is subtropical; it is cold in the winter and hot in the summer. During the months of December and January, the minimum temperature of 2 ºC occurs and during May and August the maximum temperature of 46 ºC occurs. About 178 mm annual average rainfall falls mainly during July and August. The region lies between two monsoons such as the southwest monsoon from the Indian Ocean, and the northeast or retreating monsoon from the Himalayan mountains (*SWP,* 2021).

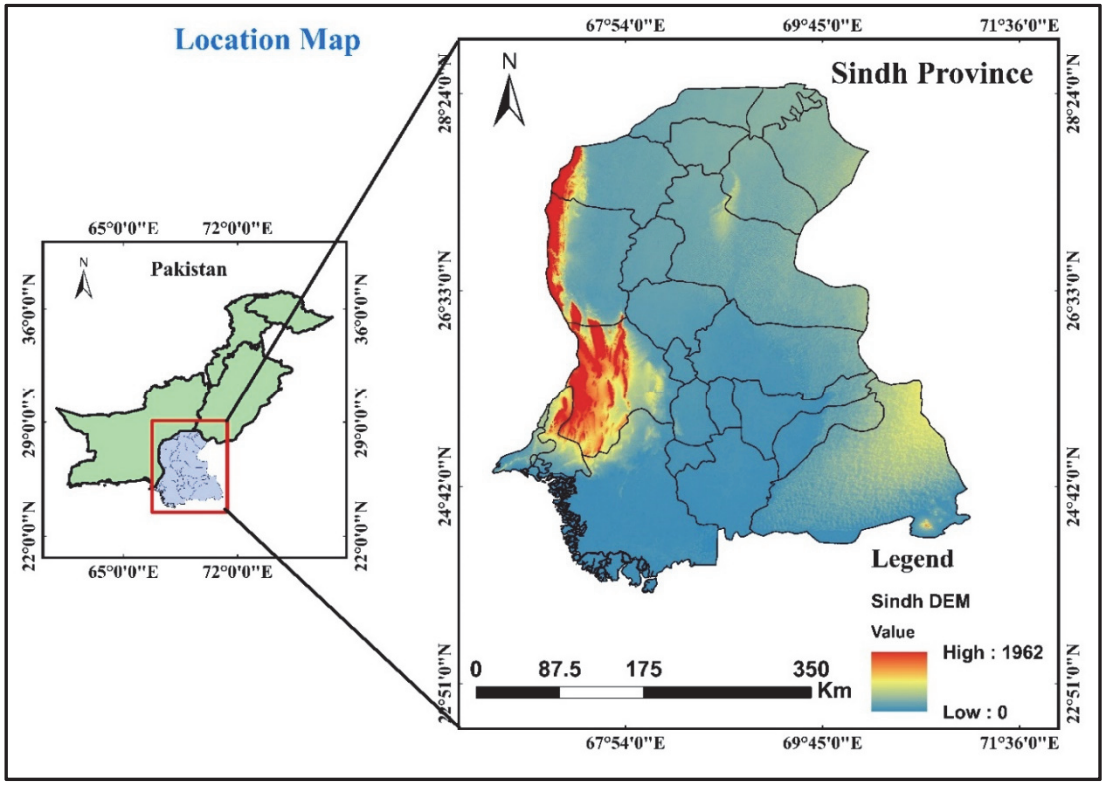

Fig. 1. Location of the study area.

2.2. Data used

2.2.1. Precipitation

The Global Land Data Assimilation System (GLDAS) is producing a series of land surface conditions such as moisture of the soil and temperature of the surface as well as fluxes such as sensible heat flux and evaporation products simulated by the Community Land Model (CLM), as well as the Mosaic, Noah, and Variable Infiltration Capacity (VIC) land surface models (*Fang et al.,* 2009). Monthly precipitation data of GLDAS from 1983 to 2020 and the Tropical Rainfall Measuring Mission (TRMM) from 1998 to 2019 with spatial resolution at

0.25° × 0.25° was used. GLDAS precipitation data was in $kgm^{-2}s^{-1}$ which was converted to mm. TRMM records data by utilizing three instruments, precipitation radar (PR), the TRMM microwave image (TMI), and the visible infrared scanner (VIRS) (*Begum et al.,* 2021). GLDAS and TRMM data services are provided at the NASA Giovanni portal.[1]

2.2.2. Climate indices

2m air temperature (T2M), 500 hPa and 850 hPa geopotential heights (H500 and H850), mean sea level pressure (SLP), surface zonal velocity (SU), 500 hPa zonal velocity (U500), latent heat flux over land (LHFOL), sensible heat flux over land (SHFOL), and surface specific humidity (SSH) were downloaded from Modern-Era Retrospective Analysis for Research and Application (MERRA-2) model at the NASA Giovanni portal.[1] It provides data at 0.5° × 0.625° spatial resolution from 1980 till present. *Fig. 2* shows the climate domains depicting the points where potential predictor data were extracted from MERRA-2. From MERRA-2 model, different variables were selected for different districts on the basis of their correlation with precipitation (*Table 1*). Climate indices from other sources were used same for all districts. The NOAA monthly Optimum Interpolation (OI) Sea Surface Temperature (SST) version 2.0 (1° × 1° resolution) data covering 89.5N-89.5S and 0.5E-359.5E region was downloaded from the NOAA Physical Sciences Laboratory and is available from December 1981onwards.[2] The QBO index of Freie University Berlin, which is the mean monthly zonal wind components computed for the levels of 10, 15, 20, 30, 40, 50, and 70 hPa at three radiosonde stations close to the equator, i.e., Canton Island, Gan Island in Maledives, and Singapore are accessible from 1953 onwards. QBO at 30 hPa (QBOI) and 50 hPa (QBOII) were utilized.[3] Nino-1+2, Nino-3, Nino-3.4, and Nino-4 were obtained from the NOAA Global Climate Observing System (GCOS) Working Group on Surface Pressure (WG-SP) which calculated all indices at 1° × 1° from the Hadley Centre Sea Ice and Sea Surface Temperature (HadISST).[4] Nino-1+2 is the Nino SST eastern-most region from 0N-10S and 90W-80W. Nino-3, Nino-3.4, and Nino-4 are the area-averaged SSTs from 5S-5N and 150W-90W, 5S-5N and 170-120W, and 5S-5N and 160E-150W, respectively and are available from 1870. In the current study, all datasets were downloaded for the period of 1982 to 2020.

[1] https://giovanni.gsfc.nasa.gov/giovanni/
[2] https://psl.noaa.gov/data/gridded/data.noaa.oisst.v2.html
[3] https://www.geo.fu-berlin.de/met/ag/strat/produkte/qbo/qbo.dat
[4] https://psl.noaa.gov/gcos_wgsp/Timeseries/

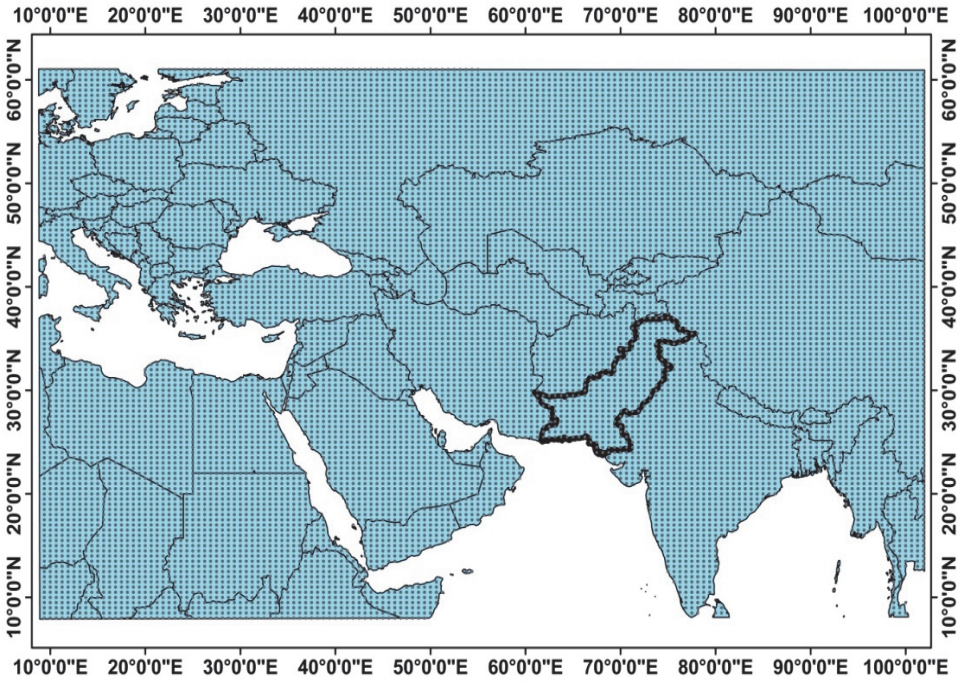

Fig. 2. Climate domains depicting the points where potential predictor data are extracted from MERRA-2.

Table 1. Selected variables for different districts from MERRA-2 model

Districts	**Variables**	**Districts**	**Variables**
Badin	H850, SSH, LHFOL, and SHFOL	Kashmore	T2M, LHFOL, and SHFOL
Dadu	T2M, LHFOL, and SHFOL	Larkana	H850, LHFOL, and SHFOL
Ghotki	H850, LHFOL, and SHFOL	Mithi	U500, SSH, LHFOL, and SHFOL
Hyderabad	U500, SSH, LHFOL, and SHFOL	Sanghar	SSH, LHFOL, and SHFOL
Jacobabad	SSH, LHFOL, and SHFOL	Shikarpur	T2M, LHFOL, and SHFOL
Jamshoro	U500, SSH, LHFOL, and SHFOL	Thatta	U500, SSH, LHFOL, and SHFOL
Karachi	SU, SSH, LHFOL, and SHFOL	Umerkot	SSH, LHFOL, and SHFOL

2.3. Methods

The techniques applied in this study are robust and are utilized in many studies around the world. Detailed explanations are given below. *Fig. 3* shows the flow chart of adopted methodology.

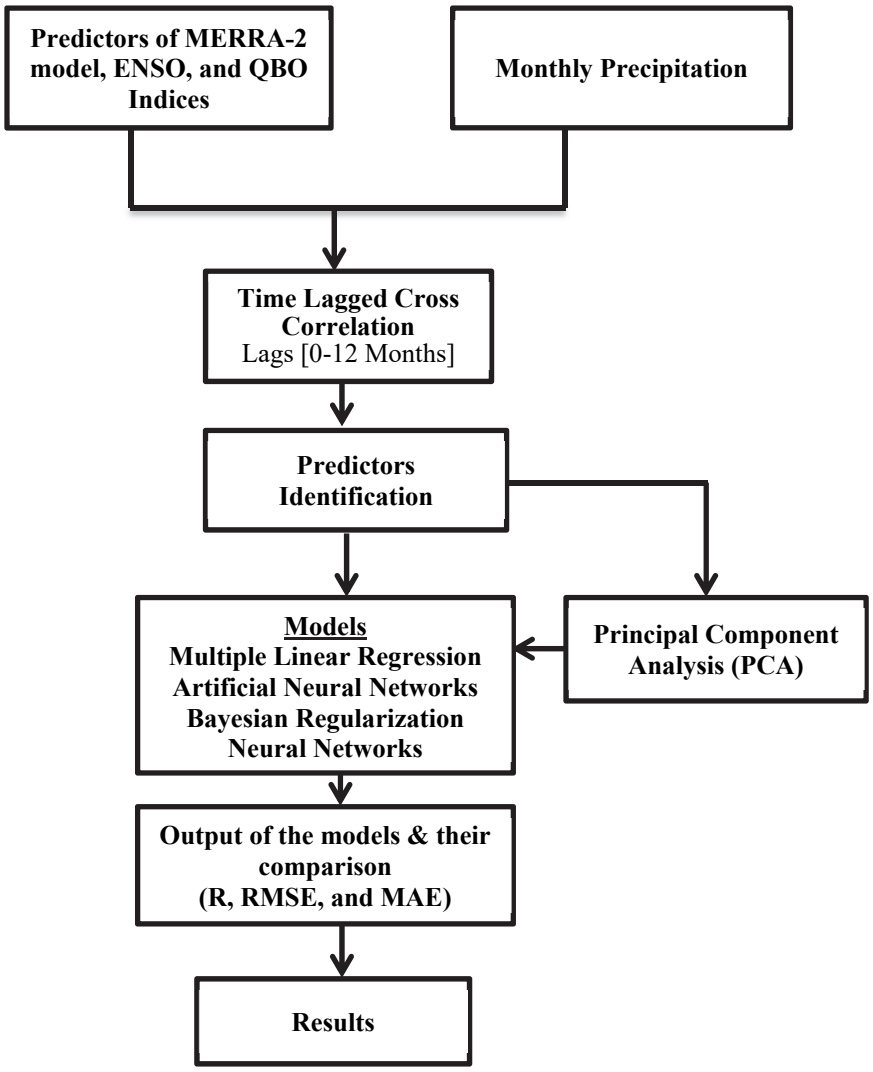

Fig. 3. Flow chart of the adopted methodology for monthly precipitation estimation teleconnected with large scale climate drivers.

2.3.1. Cross correlations

To show lagged association between the LSCDs and precipitation, the cross correlation method was employed. This method is useful when we want to align two time series in a way that one is delayed with respect to other and their peaks occurs at a lag where both series match up best (correlate well) (*Menke* and *Menke,* 2009). In the present study, monthly cross-correlations were computed between the LSCDs and the precipitation up to 12 months range for different districts, and the best correlations were identified at different lags. The formula is:

$$r_m = \frac{\sum(x_i - \bar{x})(y_i - m - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - m - \bar{y})^2}}, \quad (1)$$

where r_m shows the m lag time and $\{x_i\}$ and $\{y_i\}$ are random variables (*Taweesin* and *Seeboonruang,* 2019). Heatmaps were prepared in Origin 2021b software to show significant lagged correlations in different districts.

2.3.2. Principal component analysis (PCA)

World widely multivariate data analysis is carried out using principal component analysis (PCA) (*Shukla et al.,* 2011). It transforms several correlated parameters into a smaller number of uncorrelated parameters which are called principal components (PCs). It is suitable for the application where data records are few compared to the complexity of the model. The number of free parameters is large in ANN models, and if the training set is not large enough to accommodate for the suitable optimization of those parameters, then overfitting occurs (*Shukla et al.,* 2011; *Doranalu Chandrashekar et al.,* 2019). In the present study, the number of predictors is 10 in some districts and 11 in other, and hence the number of neurons is 10 and 11, respectively, in the input layer. The output layer possesses only one neuron. Due to the inadequate number of present data points, the free parameters optimization without the danger of overfitting is not possible, and PCA can help by reducing the number of predictors. Dataset of 456 points (1983–2020) was divided into two parts, 75% (1983–2011) for training and 25% (2012–2020) for testing. PCA was applied using a Python script on the training data only. The eigen vector which corresponds to the predictors covariance matrix's highest eigen value was determined. Selected PCs were then used to represent the data in two and three dimensional spaces (different locations selected PCs were different in number). The first two principal components explained maximum variances for all districts, and the third one was considered in some districts. 114 points (test points for ANN and BRNN) were still in 10 and 11 dimensional spaces. We projected them one at a time on the ascertained eigen vector which gave us test points in two- and three-dimensional spaces. *Table 2* and *Fig. 4* show the variances and eigen values for each of the components calculated by PCA.

Table 2. Eigen values of principal component analysis

Districts		PC1	PC2	PC3	PC4	PC5	PC6
	Eigenvalue	1.7203	1.2491	0.8954	0.1572	-	-
Badin	Variability (%)	42.7716	31.0547	22.2631	3.9104	-	-
	Cumulative (%)	42.7716	73.8264	96.0895	100	-	-
	Eigenvalue	1.7562	0.9804	0.2658	-	-	-
Dadu	Variability (%)	58.4917	32.6549	8.8532	-	-	-
	Cumulative (%)	58.4917	91.1467	100	-	-	-
	Eigenvalue	2.9136	0.8340	0.1485	0.0608	-	-
Ghotki	Variability (%)	73.6290	21.0779	3.7546	1.5383	-	-
	Cumulative (%)	73.6290	94.7070	98.4616	100	-	-
	Eigenvalue	2.0521	0.7845	0.1682	-	-	-
Hyderabad	Variability (%)	68.2920	26.1091	5.5988	-	-	-
	Cumulative (%)	68.2920	94.4011	100	-	-	-
	Eigenvalue	1.8909	1.0050	0.0747	-	-	-
Jacobabad	Variability (%)	63.6520	33.8302	2.5176	-	-	-
	Cumulative (%)	63.6520	97.4823	100	-	-	-
	Eigenvalue	3.8062	1.1217	0.7665	0.2232	0.0235	0.0125
Jamshoro	Variability (%)	63.9286	18.8402	12.8751	3.7498	0.3957	0.2102
	Cumulative (%)	63.9286	82.7689	95.6441	99.3940	99.7897	100
	Eigenvalue	2.6010	0.9405	0.4006	0.1090	-	-
Karachi	Variability (%)	64.2021	23.2158	9.8894	2.6925	-	-
	Cumulative (%)	64.2021	87.4179	97.3074	100	-	-
	Eigenvalue	1.8909	1.0050	0.0747	-	-	-
Kashmore	Variability (%)	63.6520	33.8302	2.5176	-	-	-
	Cumulative (%)	63.6520	97.4823	100	-	-	-
	Eigenvalue	2.6994	1.0496	0.5963	0.4172	0.1364	-
Larkana	Variability (%)	55.0995	21.4259	12.1729	8.5161	2.7854	-
	Cumulative (%)	55.0995	76.5254	88.6984	97.2145	100	-
	Eigenvalue	1.7168	1.2600	0.8968	0.2341	-	-
Mithi	Variability (%)	41.7938	30.6744	21.8311	5.7004	-	-
	Cumulative (%)	41.7938	72.4683	94.2995	100	-	-
	Eigenvalue	2.0521	0.7845	0.1682	-	-	-
Thatta	Variability (%)	68.2920	26.1091	5.5988	-	-	-
	Cumulative (%)	68.2920	94.4011	100	-	-	-
	Eigenvalue	2.1588	1.2723	0.5007	0.0747	-	-
Shikarpur	Variability (%)	53.8809	31.7562	12.4977	1.8650	-	-
	Cumulative (%)	53.8809	85.6371	98.1349	100	-	-
	Eigenvalue	1.7168	1.2600	0.8968	0.2341	-	-
Sanghar	Variability (%)	41.7983	30.6744	21.8311	5.7004	-	-
	Cumulative (%)	41.7938	72.4683	94.2995	100	-	-
	Eigenvalue	1.9619	1.2582	0.7409	0.0739	-	-
Umerkot	Variability (%)	48.6216	31.1820	18.3628	1.8334	-	-
	Cumulative (%)	48.6216	79.8037	98.1665	100	-	-

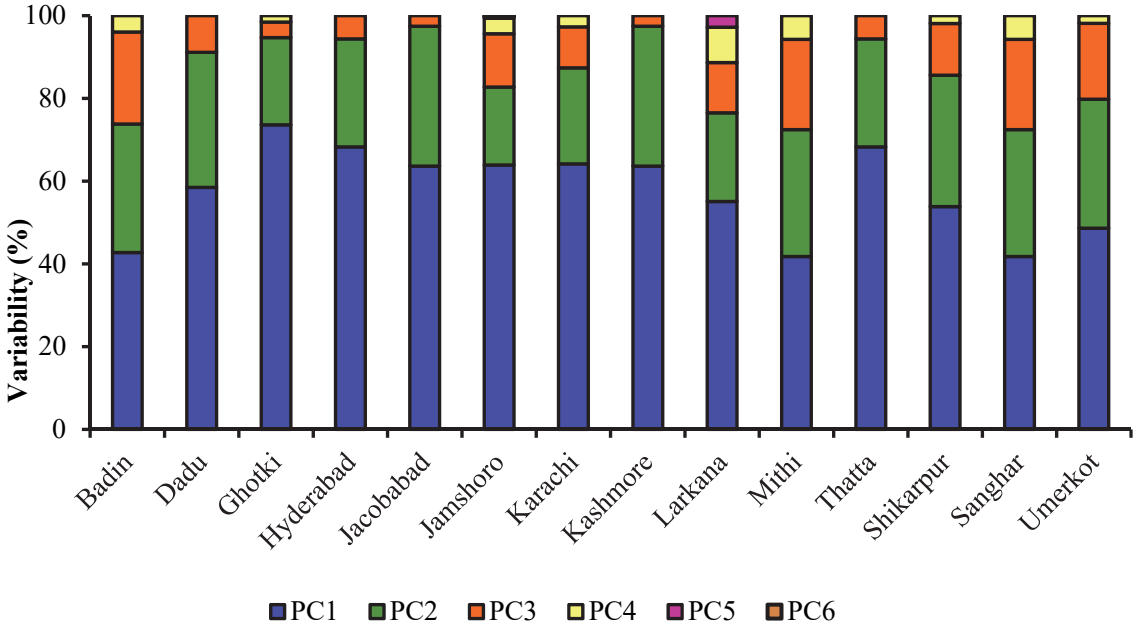

Fig. 4. Explained variability (%) of the principal components (PC1, PC2, PC3, PC4, PC5, and PC6).

2.3.3. Multiple regression analysis (MRA)

The data was normalized in the range of 0 and 1 using the following equation, which is a min-max formula.

$$\bar{x}_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{2}$$

where $\bar{x}_i$ represents normalized values, x_i are given values, x_{min} represents minimum value and x_{max} represents maximum value. 75% of the data was used for training and 25% for testing. Distinct regression models (*Wilks,* 1995) were developed using stepwise MRA. *Table 3* shows these regression models for each district.

Table 3. Summary of the best multiple regression models

Districts	Equation	r	VIF	DW
Badin	$Y_t = 0.003 + 0.151SST_{t-12} - 0.113H850_{t-12} + 0.183SST_{t-1} + 0.044SST_{t-6}$	0.59	1.53	2.05
Dadu	$Y_t = -0.053 + 0.190SST_{t-11} + 0.127SST_{t-2} + 0.046LHFOL_{t-7}$	0.47	1.27	2.15
Ghotki	$Y_t = 0.024 + 0.397SST_{t-11} - 0.215SST_{t-10} - 0.346SHFOL_{t-2} + 0.381LHFOL_{t-11}$	0.58	1.52	2.07
Hyderabad	$Y_t = -0.105 + 0.246SST_{t-12} + 0.094SST_{t-2} + 0.086SSH_{t-7}$	0.54	1.40	2.04
Jacobabad	$Y_t = -0.059 + 0.519SST_{t-11} - 0.291SST_{t-10} + 0.127LHFOL_{t-8}$	0.46	1.26	2.16
Jamshoro	$Y_t = 0.393 - 0.492SHFOL_{t-2} + 0.375SST_{t-1} + 0.208SST_{t-11} - 0.355SST_{t-8} - 0.218LHFOL_{t-7} - 0.130SSH_{t-6}$	0.65	1.72	2.05
Karachi	$Y_t = 0.102 + 0.150SST_{t-12} + 0.404SST_{t-1} - 0.279SST_{t-8} - 0.243SSH_{t-2}$	0.60	1.56	1.99
Kashmore	$Y_t = -0.068 + 0.572SST_{t-11} - 0.300SST_{t-10} + 0.119LHFOL_{t-8}$	0.51	1.36	2.13
Larkana	$Y_t = 0.080 + 0.188SST_{t-11} - 0.201H850_{t-1} + 0.131H850_{t-9} - 0.137SHFOL_{t-1} + 0.193SST_{t-1}$	0.55	1.43	2.07
Mithi	$Y_t = -0.031 + 0.272SST_{t-11} - 0.219SST_{t-8} + 0.183SST_{t-6} + 0.113SST_{t-1}$	0.62	1.62	2.02
Sanghar	$Y_t = -0.035 + 0.124SST_{t-1} + 0.190SST_{t-6} + 0.252SST_{t-11} - 0.191SST_{t-8}$	0.59	1.53	2.07
Shikarpur	$Y_t = -0.086 + 0.520SST_{t-11} - 0.292SST_{t-10} + 0.138LHFOL_{t-8} + 0.164SST_{t-1}$	0.49	1.32	2.08
Thatta	$Y_t = -0.105 + 0.089SSH_{t-7} + 0.108SST_{t-2} + 0.232SST_{t-12}$	0.54	1.40	2.04
Umerkot	$Y_t = 0.065 - 0.470SST_{t-8} + 0.503SST_{t-1} + 0.338SSH_{t-11} - 0.232SHFOL_{t-2}$	0.59	1.54	2.05

Significant lagged climate indices were included as predictors. Another set of models were developed using the selected PCs for different districts in MRA. These models are presented in *Table 4*.

Table 4. Summary of the best PC based multiple regression models

Districts	Equation	r	VIF	DW
Badin	$Y_t = 0.031 – 0.036PC1 + 0.000PC2$	0.48	1.29	2.02
Dadu	$Y_t = 0.040 – 0.030PC1 – 0.004PC2$	0.40	1.19	2.11
Ghotki	$Y_t = 0.061 + 0.035PC1 + 0.017PC2$	0.45	1.25	2.21
Hyderabad	$Y_t = 0.029 - 0.029PC1 – 0.009PC2$	0.44	1.24	2.09
Jacobabad	$Y_t = 0.069 + 0.033PC1 – 0.001PC2$	0.32	1.11	2.15
Jamshoro	$Y_t = 0.027 + 0.019PC1 – 0.006PC2$	0.43	1.23	2.05
Karachi	$Y_t = 0.025 + 0.019PC1 + 0.020PC2$	0.43	1.23	2.05
Kashmore	$Y_t = 0.072 + 0.044PC1 – 0.005PC2$	0.40	1.19	2.22
Larkana	$Y_t = 0.061 – 0.036PC1 + 0.002PC2$	0.46	1.26	2.09
Mithi	$Y_t = 0.047 – 0.043PC1 + 0.037PC2 + 0.002PC3$	0.57	1.47	2.02
Sanghar	$Y_t = 0.043 – 0.038PC1 – 0.038PC2 + 0.003PC3$	0.54	1.40	2.08
Shikarpur	$Y_t = 0.066 + 0.038PC1 + 0.005PC2 + 0.007PC3$	0.38	1.17	2.05
Thatta	$Y_t = 0.027 + 0.027PC1 + 0.007PC2$	0.45	1.25	2.05
Umerkot	$Y_t = 0.048 + 0.045PC1 + 0.011PC2$	0.48	1.29	2.04

Verification of the multicollinearity is significant in the MR modeling, which is observed in case of highly correlated predictors and can result in substantial variation in the estimates of the parameter in response to small variations in the data or the model. The utilized indicators are tolerance (*T*) and variance inflation factor (*VIF*):

$$Tolerance = 1–R^2 \quad , \quad VIF = \frac{1}{Tolerance}, \tag{3}$$

where R^2 represents the coefficient of multiple determination:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}, \tag{4}$$

where *SST* represents the total sum of squares, *SSR* represents the regression sum of squares, and *SSE* shows the error sum of squares. Tolerance value less than 0.20–0.10 or a *VIF* value greater than 5–10 shows the problem of multicollinearity (*Lin,* 2008). To assess the errors independence of the models, the Durbin-Watson test was used which looks for the serial correlations between errors and ranges from 0–4. Values >3 or <1 create problems (*Field,* 2009).

2.3.4. Artificial neural network (ANN)

Multi-layer feed-forward neural network with back propagation algorithm was used for the prediction of monthly precipitation with the first two PCs in most cases and three PCs in some cases as the predictors. The model comprised three layers, the input layer having two neurons and three neurons in some cases, the output layer having one neuron, and one hidden layer were selected, and the number of neurons was determined by a trial and error method. It is known that a single hidden layer is ample to estimate any non-linear function to arbitrary accuracy (*Cybenko,* 1989). The number of neurons in hidden layer significantly affects the performance of the model, for example, less number causes underfitting, while greater number causes overfitting (*Hussain,* 2020). The ANN model development process is 1) to find suitable data set for input, 2) to ascertain the hidden layers number as well as neurons, and 3) to train, validate, and test the network. It can be expressed mathematically as:

$$y_j = f_2[\sum_{j=1}^{J} w_i \, f_1\left(\sum_{i=1}^{I} w_i \, x_i\right)] \, , \tag{5}$$

where the output of the network is represented by y_j, the input by x_i, w_i and w_j represent the weights between the input neurons and the hidden layer and between the hidden layer and output neurons, respectively, the activation functions for the hidden layer and output layer are f_1 and f_2, respectively. Sigmoidal and linear kinds of transfer functions are appropriate for the hidden and output layers (*Maier* and *Dandy,* 2000). In the current study, f_1 is a sigmoid function, which is basically a nonlinear function, and f_2 is a linear purelin function, given as:

$$f_1 = \frac{1}{1+e^{-x}} \, , \tag{6}$$

$$f_2(x) = x. \tag{7}$$

The models of ANN were trained employing the Levenberg-Marquardt technique, and the early stop approach was employed to avoid the chance of overfitting while training and validating. In this method, the training process stops when the validation set error start to increase, while the training set error is still reducing (*Mekanik et al.,* 2013; *Luk et al.,* 2021). *Fig. 5* shows schematic diagram of the ANN.

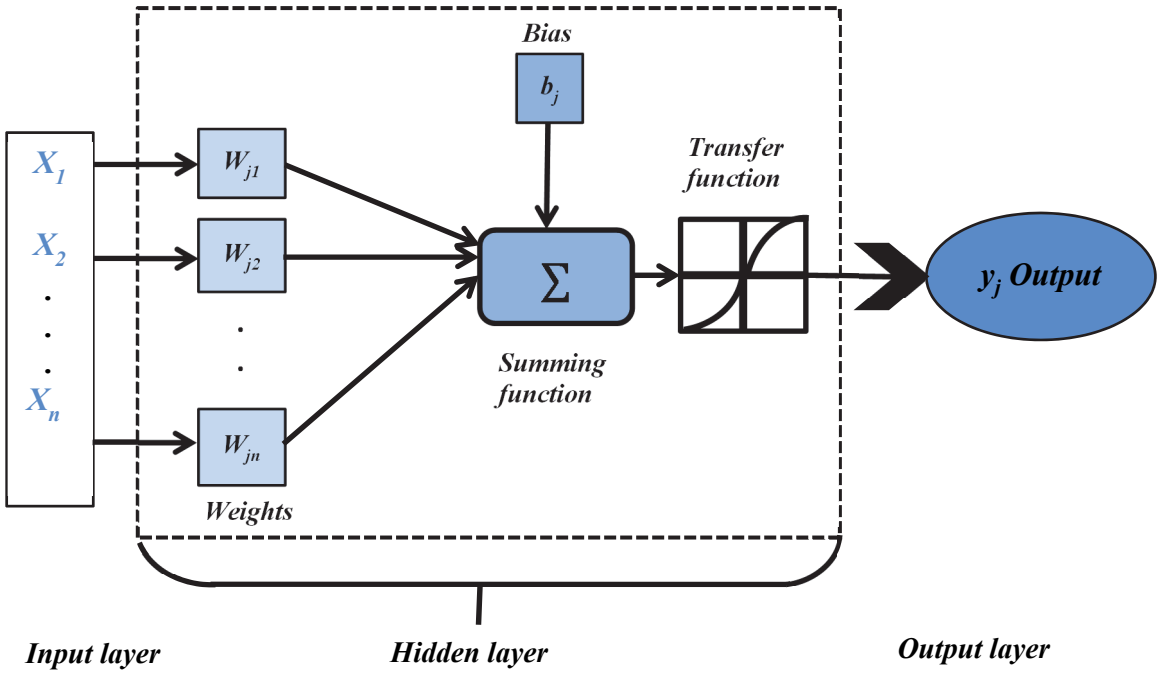

Fig. 5. Schematic diagram of the artificial neural network (ANN).

2.3.5. Bayesian regularization neural network (BRNN)

The Bayesian regularization neural network (BRNN) is the version of artificial neural networks (ANN), which is a more powerful method as compared to the conventional ANN. A common error function (E_d) of ANN utilizing early stop technique can be represented as:

$$E_D(D|w,M) = \sum_{i=1}^{n}(t_i - \hat{t}_i)^2 \,, \tag{8}$$

where w represents the weight, M represents the structure of ANN, n represents the size of training data, t_i represents the i_{th} target, whereas $\hat{t}_i$ represents the output. ANN's immature convergence results in overfitting. Its regularization in using Bayesian technique helps optimization of the ANN parameters by utilizing their prior values. For this reason, an additional term (E_w) is incorporated as:

$$E_D(D|w,M) = \sum_{i=1}^{n}(t_i - \hat{t}_i)^2 + E_w \,, \tag{9}$$

where E_w deals with the unrealistic weights for better generalization and cautious conversion. This kind of optimization method is utilized to minimize the function:

$$F = \beta E_D(D|w,M) + \alpha E_w(w|M) \,, \tag{10}$$

where $E_w(w|M)$ represents the sum of ANN architecture square, and α and β show the hyperparameters for the optimization process (*Ye et al.,* 2021). This technique decreases a combination of squared errors and weights and identifies the accurate combination. It does not require validation set and is potentially suitable algorithm for limited data (*Beale et al.,* 2011). In the current study, the same PCs used in ANN were utilized in BRNN. Scripts were developed for both ANN and BRNN in the MATLAB R2015a environment.

To validate the performance of all developed models, the TRMM dataset was used. It was similarly normalized, and the PCs were calculated by projecting the data points on the previous ascertained eigen vectors. In the last, time series graphs and radar charts were prepared for comparison.

2.3.6. Model evaluation

The constructed models were evaluated using three performance metrics such as root mean square error (*RMSE*), mean absolute error (*MAE*), and correlation coefficient (*R*) between the GLDAS precipitation values and models predicted values as well as between TRMM precipitation values and models predicted values. They can be expressed as follows.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Yi-Xi)^2}{n}}, \tag{11}$$

$$MAE = \frac{\sum_{i=1}^{n}|Y_i-X_i|}{n}, \tag{12}$$

$$\mathrm{R} = \frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2\sum_{i=1}^{n}(Y_i-\overline{Y})^2}}, \tag{13}$$

where *Yi* are the predicted values, *Xi*are the GLDAS or TRMM values, $\overline{X}$ is the mean GLDAS or TRMM value, and $\overline{Y}$ is the mean predicted value.

3. Results and discussion

3.1. Identification of significant large scale climate drivers

The cross-correlations between monthly precipitation and LSCD values are presented in *Fig. 6*. We observe that in Badin district, five LSCDs (H850, SST, SSH, LHFOL, and SHFOL) have significant positive (negative) cross-correlations at various lags with 99% confidence interval, revealing that there exists a direct (inverse) physical association. Dadu, Kashmore, and Shikarpur

districts have four (T2M, LHFOL, SHFOL, and SST), Ghotki and Larkana have four (H850, LHFOL, SHFOL, and SST), Hyderabad, Jamshoro, Thatta and Mithi have five (U500, SSH, LHFOL, SHFOL, and SST), Jacobabad, Sanghar, and Umerkot have four (SSH, LHFOL, SHFOL, and SST), and Karachi has five (SSH, LHFOL, SHFOL, SST, and SU) significant LSCDs.

We notice that for most of the LSCDs, the relationship is persistent and statistically significant at lags 1, 2, and 4 to 12, with the strongest correlations in the SST, LHFOL, SSH, and U500. For H850, we observe that at lags 1, 2, and 12, the correlation was negative, while from 8 to 11 it was positive. In case of SSH, T2M, LHFOL, SHFOL, SST, and SU, positive correlation was observed from lags 1 to 3 and 10 to 12, while negative from 4 to 9. For U500, it was positive at lags 4 to 9, while negative at 1 to 3 and 10 to 12. Moreover, this analysis showed that the cross-correlations of Nino-1+2, Nino-3, Nino-3.4, Nino-4, QBOI, and QBOII were not significant in the Sindh region, and the most significant LSCDs were SU, U500, T2M, SST, SHFOL, LHFOL, SSH, and H850.

Till now, some studies were carried out in Sindh province (*Mahmood et al.,* 2006; *Iqbal* and *Athar,* 2018; *Bhutto* and *Wei,* 2009) and the last mentioned study (*Iqbal* and *Athar,* 2018) obtained the same results for the Sindh region, but they found concurrent correlations, and did not find antecedent correlations (lagged correlations) where the QBO as well as ENSO did not show correlation with the monthly precipitation of Sindh province stations. The lagged association was only discussed by one study for the period of 1960–1990, and some of the used climate indices (Nino-1+2 and Nino-4) were not significant over the region which is in accordance with this study (*Muslehuddin et al.,* 2005).

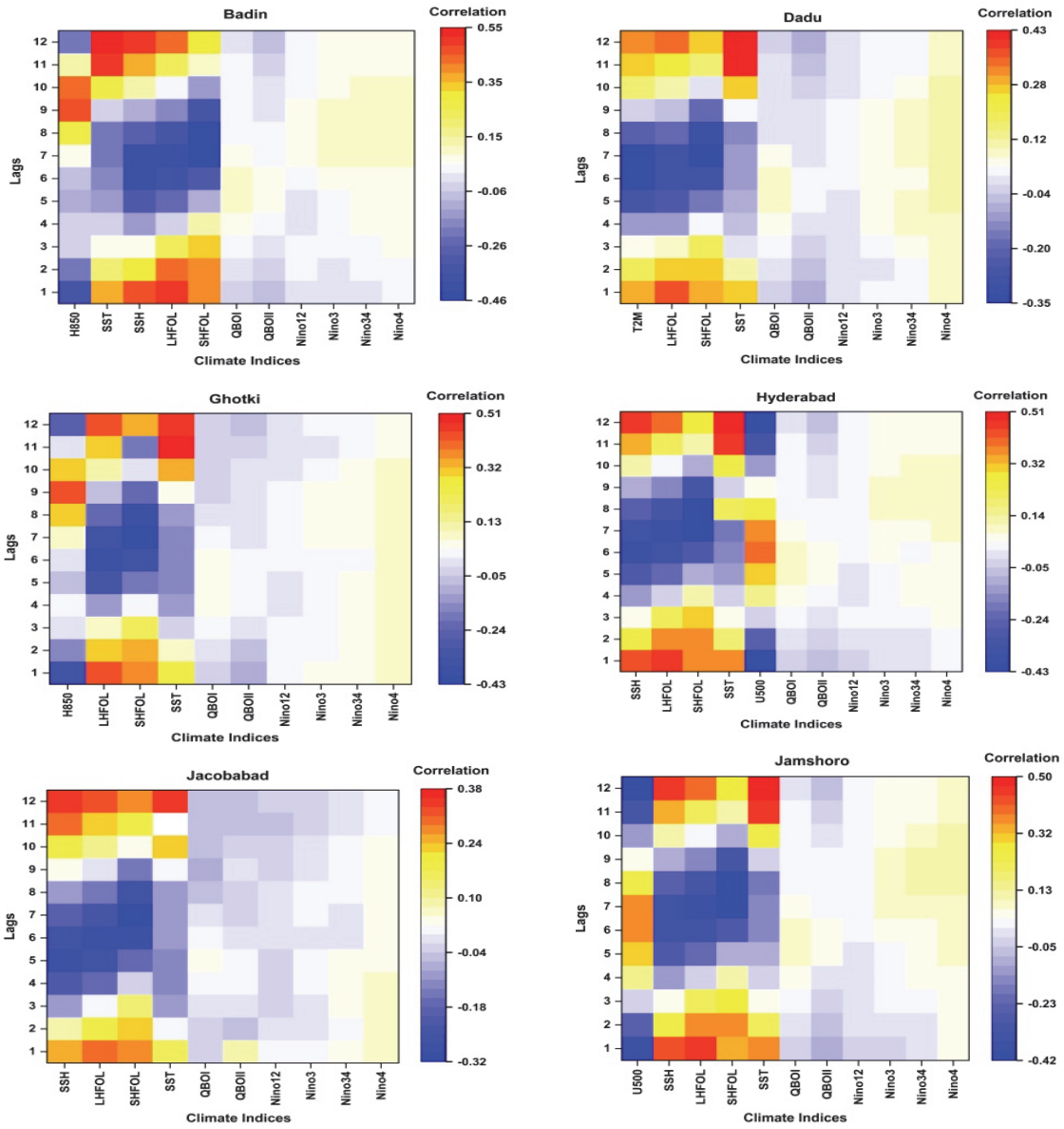

Fig. 6. Correlations of lagged climate indices and monthly precipitation for the first 6 districts.

3.2. Results of multiple regression analysis

In the present study, MLR equations were developed using the identified significant LSCDs at different lag times to predict monthly precipitation. The models were selected as the best models, when they did not violate the limits of statistical significance and recorded lower errors. *Table 3* shows the best developed multiple regression (MR) models using lagged LSCDs for each district, with Pearson correlation (*r*), variance inflation factor (*VIF*), and Durbin-Watson

statistics (*DW*), and similarly, *Table 4* shows the same statistics for MR models developed with PC components (MR-PC). It can be seen that the *VIF* values are just above one which shows that there is no multicollinearity among the predictors, and the *DW* statistics confirmed the goodness-of-fit of the models as the residuals have no autocorrelation. *Table 5* manifests the performance of MR and MR-PC models. The highest Pearson correlation was observed for Mithi district in both sets of the models. In training case, it was 0.62 and 0.57, and in test case it was 0.64 and 0.57, respectively. Overall, MR models performed better than MR-PC models, and all districts' *RMSE* and *MAE* values are relatively low. In the research of *Muslehuddin et al.* (2005), MR models were developed for Sindh, Pakistan with correlation values ranging from 0.53–0.71 which is in accordance with the results of the current study. Studies around the world evaluated oceanic and atmospheric climate indices were lagged at period of several months as predictors of precipitation. In the study of *Taweesin* and *Seeboonruang* (2019), MR models were prepared for Thailand with and without lagged LSCDs, and it was noticed that the rainfall's response to climatic factors was delayed and the models were capable of forecasting monthly rainfall accurately. In the study of *Kim et al.* (2020) in South Korea, it was observed that the models predicted some summer seasons well, but satisfactorily performed during other seasons and long periods which were ascribed to irregular characteristics of rainfall such as heavy rains resulted by monsoonal front and typhoon. The reasons could be a failure to identify the associated climate index signal or in the historical data, or the absence of a significant teleconnection between the climate indices and precipitation. It was also stated that, because of the models statistical nature, the predictability is most likely to be decreased if distinct statistical characteristics climate phenomena appear in the predictors or predictands instead like the past. In the study of *Choubin et al.* (2014) MR model was used also, and it was found that the fluctuations or standard deviation of the observed station data cannot be predicted by the models, therefore, it is not able to forecast droughts and wet years. Some other studies were also carried out (*Hossain et al.,* 2015; *Shukla et al.,* 2011; *Mekanik et al.,* 2013).

Table 5. Performance of the regression models

Districts	Training						Test					
	Correlation		RMSE		MAE		Correlation		RMSE		MAE	
	MR	MR-PC	MR	MR-PC	MR	MR-PC	MR	MR-PC	MR	MR-PC	MR	MR-PC
Badin	0.59	0.50	0.08	0.09	0.03	0.04	0.55	0.49	0.08	0.08	0.04	0.04
Dadu	0.47	0.39	0.09	0.09	0.04	0.05	0.39	0.36	0.11	0.11	0.06	0.06
Ghotki	0.59	0.45	0.12	0.13	0.06	0.06	0.50	0.43	0.13	0.14	0.07	0.08
Hyderabad	0.54	0.45	0.08	0.09	0.03	0.04	0.45	0.44	0.10	0.09	0.05	0.04
Jacobabad	0.46	0.34	0.13	0.14	0.08	0.08	0.41	0.31	0.17	0.19	0.10	0.10
Jamshoro	0.65	0.43	0.07	0.08	0.03	0.03	0.52	0.45	0.08	0.08	0.04	0.05
Karachi	0.60	0.42	0.07	0.08	0.03	0.03	0.44	0.31	0.08	0.09	0.04	0.04
Kashmore	0.51	0.41	0.13	0.14	0.07	0.08	0.45	0.37	0.16	0.17	0.09	0.10
Larkana	0.55	0.46	0.12	0.12	0.07	0.07	0.41	0.38	0.15	0.15	0.08	0.08
Mithi	0.62	0.57	0.10	0.10	0.04	0.05	0.64	0.57	0.09	0.10	0.05	0.06
Sanghar	0.59	0.54	0.10	0.10	0.04	0.05	0.57	0.49	0.10	0.11	0.06	0.06
Shikarpur	0.49	0.39	0.13	0.14	0.07	0.08	0.39	0.31	0.16	0.17	0.09	0.09
Thatta	0.54	0.44	0.08	0.08	0.03	0.03	0.46	0.46	0.08	0.07	0.04	0.04
Umerkot	0.59	0.49	0.11	0.12	0.05	0.05	0.56	0.51	0.11	0.12	0.06	0.06

3.3. Results of artificial neural networks and Bayesian regularization neural networks analysis

Table 6 shows the performance of ANN and BRNN models based on correlation, *RMSE* and *MAE* errors for the training and test cases. The highest correlation coefficient of ANN and BRNN for training case was 0.76 and 0.74 for Mithi district. In the test case, the highest for ANN was 0.83 for Badin and 0.76 for Mithi. All districts' *RMSE* and *MAE* values were relatively low for both types of models. The training case correlation coefficient varied between 0.46–0.76 for ANN and 0.40–0.74 for BRNN. *RMSE* ranged between 0.05–0.14 for ANN and 0.06–0.13 for BRNN. MAE ranged between 0.02–0.09 for ANN and 0.03-0.08 for BRNN. In the test case, correlation coefficient ranged between 0.57–0.83 for ANN and 0.61–0.76 for BRNN. *RMSE* varied between 0.04–0.12 for ANN and 0.07–0.14 for BRNN. *MAE* varied between 0.03–0.07 for ANN models and 0.03-0.08 for BRNN models. The higher values of correlation coefficient of ANN and BRNN models manifest that both models are capable of finding the pattern and trend of the GLDAS precipitation compared to MR and MR-PC models. There is no study in the Sindh region for comparison, but some research carried out in Pakistan and across the world is mentioned here. In the study of *Choubin et al.* (2014), multilayer perceptron (MLP) neural network was employed with MR and ANFIS models for Iran, and the lagged LSCDs were used as inputs. The MLP model showed better performance than the other two models, which is in

agreement with the results of the current study, because the MLP utilizes the Levenberg-Marquardt technique which is faster and powerful than the ANFIS with gradient decent technique for computing the membership function parameters. In the study of *Bello* and *Mamman* (2018), a linear and an ANN model were developed for Kano, Nigeria, and it was found that the ANN is preferable and can be used confidently with the ENSO indices. Similar results were obtained by *Shukla et al.* (2011) and *Doranalu Chandrashekar et al.* (2019) over India. In the study of *Ahmed et al.,* (2015), an MLP was utilized to downscale the rainfall in the Balochistan region of Pakistan. The observed and downscaled rainfall showed good agreement, while it was found that the model underpredicted the variance of rainfall. The study of *Awan* and *Maqbool* (2010) over Islamabad Pakistan revealed better performance of neural network approaches in terms of accuracy, greater lead time, and fewer requirements of resources. In the study of *Ye at al.* (2021), it was observed that BRNN performed well compared with other models but was not capable to reproduce extreme rainfall showing that it cannot be utilized for extreme rainfall and flash flood, prediction, and it overpredicted low rainfall.

Table 6. Performance of ANN and BRNN models

Districts	**Training**						**Test**					
	Correlation		**RMSE**		**MAE**		**Correlation**		**RMSE**		**MAE**	
	ANN	**BRNN**	**ANN**	**BRNN**	**ANN**	**BRNN**	**ANN**	**BRNN**	**ANN**	**BRNN**	**ANN**	**BRNN**
Badin	0.62	0.60	0.06	0.07	0.03	0.04	0.83	0.63	0.08	0.08	0.04	0.03
Dadu	0.57	0.49	0.09	0.08	0.05	0.04	0.61	0.64	0.07	0.10	0.04	0.05
Ghotki	0.64	0.60	0.11	0.12	0.05	0.06	0.66	0.62	0.12	0.10	0.05	0.06
Hyderabad	0.69	0.64	0.08	0.07	0.04	0.03	0.71	0.63	0.04	0.09	0.03	0.04
Jacobabad	0.46	0.40	0.14	0.13	0.09	0.08	0.57	0.61	0.09	0.17	0.07	0.08
Jamshoro	0.65	0.62	0.06	0.07	0.02	0.03	0.79	0.75	0.09	0.07	0.04	0.03
Karachi	0.68	0.46	0.07	0.06	0.03	0.03	0.66	0.68	0.09	0.09	0.04	0.04
Kashmore	0.60	0.46	0.12	0.13	0.07	0.08	0.61	0.61	0.12	0.14	0.07	0.08
Larkana	0.55	0.52	0.11	0.11	0.06	0.07	0.62	0.70	0.12	0.11	0.06	0.07
Mithi	0.76	0.74	0.08	0.08	0.04	0.04	0.79	0.76	0.12	0.09	0.04	0.04
Sanghar	0.65	0.72	0.08	0.09	0.05	0.04	0.74	0.70	0.09	0.08	0.05	0.04
Shikarpur	0.50	0.50	0.14	0.12	0.08	0.07	0.60	0.52	0.16	0.16	0.08	0.08
Thatta	0.65	0.62	0.05	0.06	0.02	0.03	0.74	0.70	0.07	0.09	0.03	0.03
Umerkot	0.62	0.58	0.10	0.11	0.05	0.05	0.65	0.71	0.07	0.10	0.04	0.05

3.4. *Evaluation of model generalization ability on the TRMM set*

After calibration and validation of the models, to assess the generalization capability of the constructed MR, MR-PC, ANN, and BRNN models, the TRMM dataset was used for the period of 1998–2019 (*Fig. 7*). It can be seen that MR and MR-PC models are showing relatively low performance as compared to ANN and BRNN. The correlation coefficient values for the MR and MR-PC based models ranged between 0.22–0.57 and 0.17–0.56, respectively. *RMSE* ranged between 0.11–0.20 and 0.11–0.21. *MAE* varied between 0.06–0.15 and 0.05–0.13. ANN and BRNN models' correlation coefficient varied between 0.27–0.65 and 0.28–0.60, respectively. *RMSE* ranged between 0.10–0.14 and 0.11–0.14 while *MAE* ranged between 0.05–0.08 and 0.04–0.08.

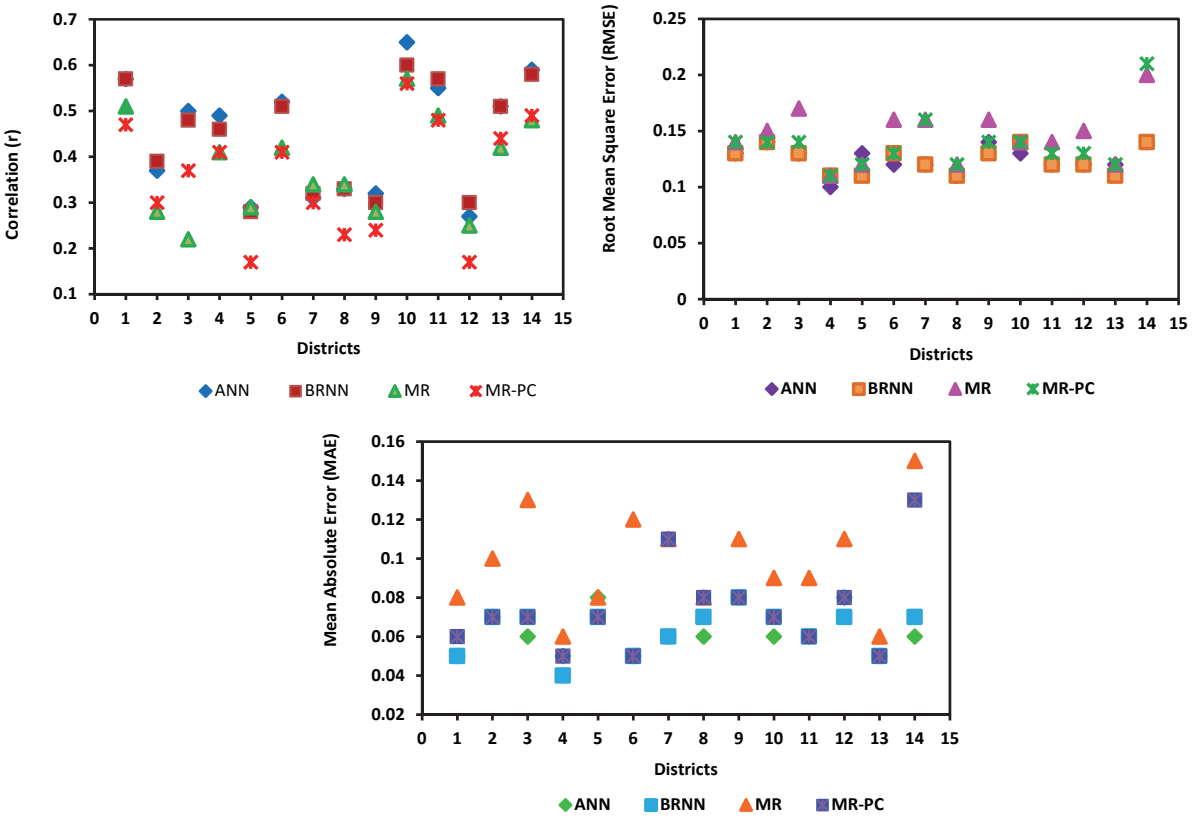

Fig. 7. Performance of ANN, BRNN, and multiple regression models for the test set (TRMM dataset).

3.5. *Comparison of the models*

Fig. 8 shows comparisons between all four types of models (MR, MR-PC, ANN, and BRNN). To summarize results, only two out of the fourteen districts are

graphically presented. ANN and BRNN outperformed the other models, and in general, the regression models showed underestimation of the actual observations also discussed by *Mekanik et al.* (2013). However, it is noted that these models were not able to predict the extreme precipitation, which means that they are not applicable for extreme precipitation prediction. This is in accordance with the findings of *Ye et al.* (2021). *Fig. 9(a-f)* represents the radar charts of overall performance, where the first two charts are showing the correlation over two sets of the 14 districts (*Fig. 9a* and *b*). Correlation of the ANN with GLDAS precipitation is higher in most of the districts followed by BRNN and MLR, while MLR-PC one is low. To check the errors of the models, two radar charts of root mean square error (*RMSE*) (*Fig. 9c* and *d*) and two of the mean absolute error (*MAE*) (*Fig. 9e* and *f*) are also presented, which are manifesting low errors for ANN, BRNN, and MR and highest for MR-PC.

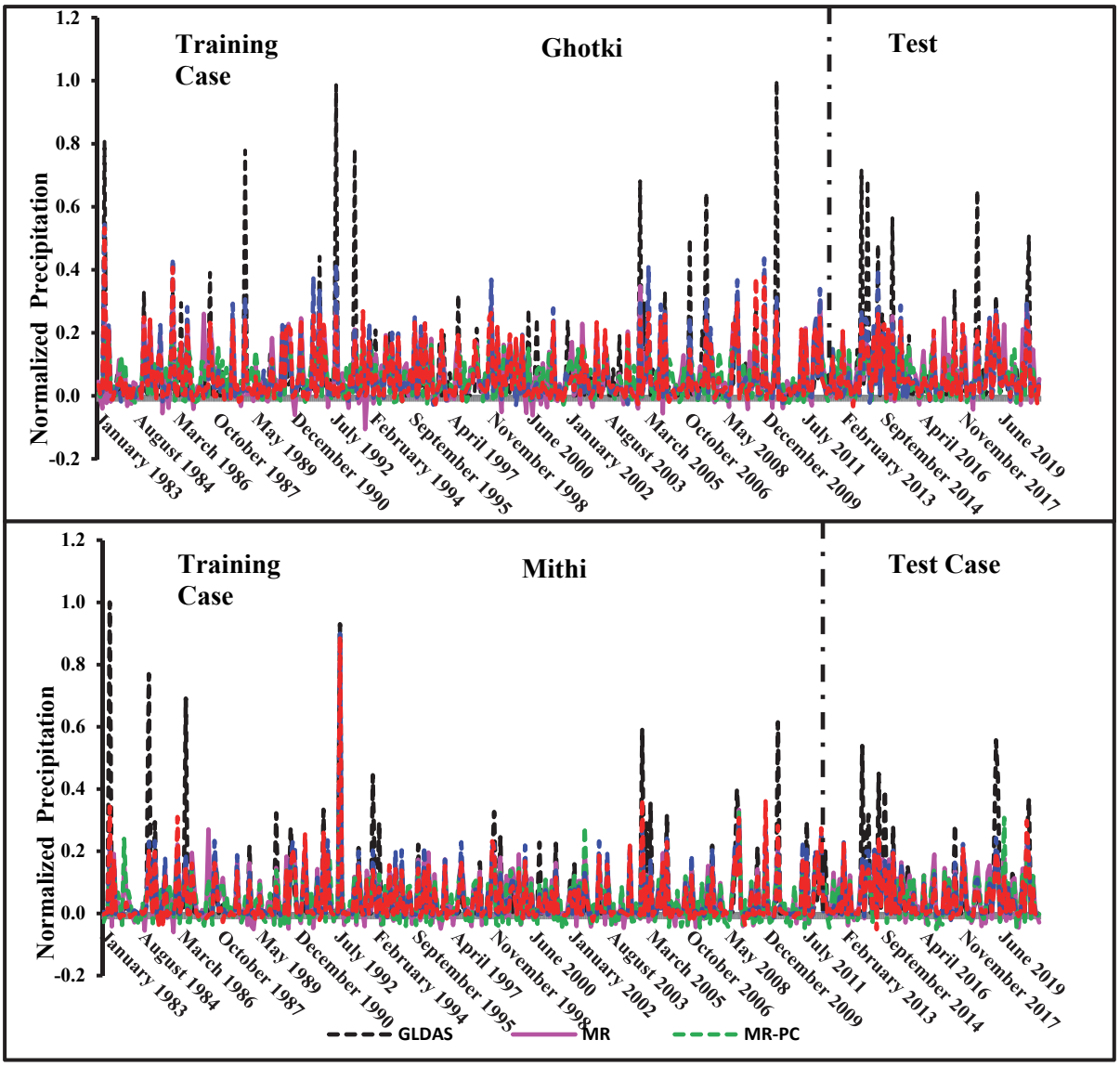

Fig. 8. Comparing ANN and BRNN models with MR and MR-PC models over two districts of Sindh province.

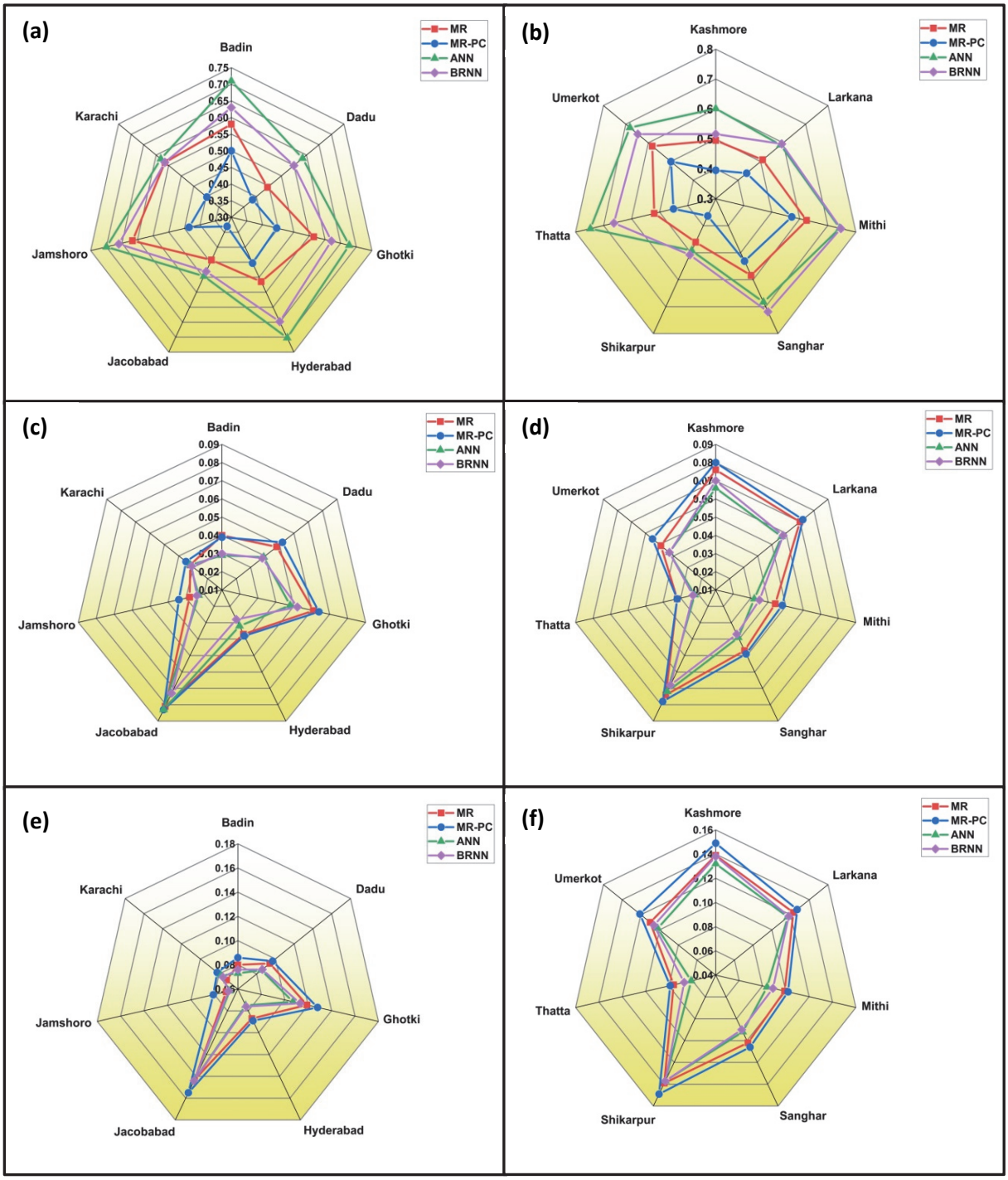

Fig. 9. Radar charts of GLDAS and predicted precipitation values by MR, MR-PC, ANN, and BRNN during 1983-2020; (a and b) correlation (c and d) mean absolute error, and (e and f) root mean square error.

4. Conclusion

This study focused on the identification of significant LSCDs in Sindh province of Pakistan, and improved the forecast skill of monthly precipitation by application of principal component analysis, artificial neural network, Bayesian

regularization neural network, and multiple regression analysis while taking into account the lagged association of LSCDs. Nino-1+2, Nino-3, Nino-3.4, Nino-4, SST, QBOI, QBOII, T2M, H500, H850, SU, U500, LHFOL, SHFOL, and SSH were selected as LSCDs. The significant LSCDs identified by a cross correlation analysis with 99% confidence level were SU, U500, T2M, SST, SHFOL, LHFOL, SSH, and H850. It was noticed that for most of the LSCDs, the relationship was persistent and statistically significant at lags 1, 2, and 4 to 12 with the strongest correlations in the SST, LHFOL, SSH, and U500. To predict monthly precipitation using lagged LSCDs and principal components, MR models were developed. The models that did not violate the limits of statistical significance and multicollinearity and had lower errors were selected. MR models performed better than MR-PC. Highest correlation was observed for Mithi district in both model sets. In the training case it was 0.62 and 0.57 and in the test case it was 0.64 and 0.57, respectively. ANN and BRNN models were developed using selected PC components and gave higher correlations as compared to regression models indicating their capability of finding the pattern and trend of the observations. They generally manifested lower errors and were more reliable for the purpose of prediction in the region. However, they were not able to predict very high precipitation events, which means that they are not applicable for extreme precipitation prediction. The highest correlation coefficient of ANN and BRNN for the training case was 0.76 and 0.74 for Mithi district. In the test case, the highest for ANN was 0.83 for Badin and 0.76 for Mithi. Their generalization ability was tested on TRMM dataset. In conclusion, this study divulged the possibility of monthly precipitation forecasting using ANN and BRNN and lagged LSCDs for the study region. It is explicit that the response of precipitation to climatic factors is delayed. Future studies are needed to be carried out for the improvement of prediction of the extreme and lower precipitation events with the addition of new climatic indices.

Acknowledgements: This study is a part of the first author's doctoral research work conducted at the University of Tabriz, Iran. The authors would like to acknowledge the NASA Goddard Earth Sciences Information Services Center (GES DISC) for giving access to the GLDAS model satellite precipitation data and the MERRA-2 model, the Freie University Berlin, and the NOAA Global Climate Observing System (GCOS) Working Group on Surface Pressure (WG-SP) for the HadISST data for the research purpose on free basis.

References

Ahmed, K., Shahid, S., Haroon, S.B., and *Wang, X.J.,* 2015: Multilayer perceptron neural network for downscaling rainfall in arid region: A case study of Baluchistan, Pakistan. *J. Earth Sys. Sci.* 124, 325–1341. https://doi.org/10.1007/s12040-015-0602-9

Aamir, E., and *Hassan, I.,* 2018: Trend analysis in precipitation at individual and regional levels in Baluchistan, Pakistan. In IOP Conference Series: Material Science and Engineering 414, 1st International Conference on Advances in Engineering and Technology (ICAET-2018). https://doi.org/10.1088/1757-899X/414/1/012042

Adamowski, J., and *Sun, K.,* 2010: Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *J. Hydrol. 390*, 85–91. https://doi.org/10.1016/j.jhydrol.2010.06.033

Ahmadi, A., Moridi, A., Lafdani, E.K., and *Kianpisheh, G.,* 2014: Assessment of climate change impacts on rainfall using large scale climate variables and downscaling models – A case study. *J. Earth Sys. Sci. 123*, 1603–1618. https://doi.org/10.1007/s12040-014-0497-x

Awan, J.A., and *Maqbool, O.,* 2010: Application of Artificial Neural Networks for monsoon rainfall prediction. In 2010 6th International Conference on Emerging Technologies (ICET). 27–32.

Bello, A.A., and *Mamman, M.B.,* 2018: Monthly rainfall prediction using artificial neural network: A case study of Kano, Nigeria. *Environ. Earth Sci. Res. J. 5*(2), 37–41. https://doi.org/10.18280/eesrj.050201

Begum, B., Tajbar, S., Khan, B., and *Rafiq, L.,* 2021: Identification of relationships between climate indices and precipitation fluctuation in Peshawar City-Pakistan. *J. Res. Environ. Earth Sci. 10*, 264–278.

Beale, M.H., Hagan, M.T., and *Demuth, H.B.,* 2011: Neural network toolbox TM7: User's guide.

Bhutto, A., and *Wei, M.,* 2009: Impact of ENSO on summer monsoon in southern parts of Pakistan. In 2009 1st International Conference on Information Science and Engineering (ICISE). 4903–4906. https://doi.org/10.1109/ICISE.2009.654

Choubin, B., Khalighi-Sigaroodi, S., Malekian, A., and *Kisi, O.,* 2014: Multiple linear regression, multi-layer perceptron network and adaptive neuro-fuzzy inference system for forecasting precipitation based on large-scale climate signals. *Hydrol. Sci. J. 61*, 1001–1009. https://doi.org/10.1080/02626667.2014.966721

Chinchorkar, S.S., Patel, G.R., and *Sayyad, F.G.,* 2012: Development of monsoon model for long range forecast rainfall explored for Anand (Gujarat-India). *Int. J. Wat. Res. Environ. Engineer. 4*, 322–326.

Cybenko, G., 1989: Approximation of superpositions of a sigmoidal function. *Mathemat. Control. Sign. Sys. 2*, 303–314. https://doi.org/10.1007/BF02551274

Doranalu Chandrashekar, V., Shetty, A., and *Patel, G.C.M.,* 2019: Estimation of monsoon seasonal precipitation teleconnection with El Niño-Southern Oscillation sea surface temperature indices over the Western Ghats of Karnataka. *Asia-Pacific J. Atm. Sci.(2019)* https://doi.org/10.1007/s13143-019-00133-w

Field, A., 2009: *Discovering statistics using SPSS,* 3rd edn. Sage Publications Ltd, London.

Fang, H., Beaudoing, H.K., Rodell, M., Teng, W.L., and *Vollmer, B.E.,* 2009: Global Land Data Assimilation System (GLDAS) products, services and application from NASA Hydrology Data and Information Services Center (HDISC). In ASPRS 2009 Annual Conference. 151–159.

Gholami Rostam, M., Sadatinejad, S.J., and *Malekian, A.,* 2020: Precipitation forecasting by large-scale climate indices and machine learning techniques. *J. Arid Land* 12, 854–864. https://doi.org/10.1007/s40333-020-0097-3

Hossain, I., Rasel, H.M., Imteaz, M.A., and *Pourakbar, S.,* 2015: Effects of climate indices on extreme rainfall in Queensland, Australia. In 21st International Congress on Modelling and Simulation MODSIM. 1984–1990.

Hussain, M., 2020: Spatiotemporal construction safety planning by integrating Building Information Modeling (BIM) with Artificial Intelligence (AI). Dissertation, Sungkyunkwan University, South Korea.

Iqbal, M.F., and Athar, H., 2018: Variability, trends, and teleconnections of observed precipitation over Pakistan. *Theor. Appl. Climatol. 134*, 613–632. https://doi.org/10.1007/s00704-017-2296-1

Kumar, R., Singh, M.P., Roy, B., and *Shahid, A.H.,* 2021: A comparative assessment of metaheuristic optimized extreme learning machine and deep neural network in multi-step-ahead long-term rainfall prediction for all-Indian regions. *Wat. Res. Manage. 35*, 1927–1960. https://doi.org/10.1007/s11269-021-02822-6

Kashiwao, T., Nakayama, K., Ando, S., Ikeda, K., Lee, M., and *Bahadori, A.,* 2017: A neural network-based local rainfall prediction system using meteorological data on the Internet: A case study using data from the Japan Meteorological Agency. *Appl. Soft Comput. J. 56*, 317–330. https://doi.org/10.1016/j.asoc.2017.03.015

Kim, C-G., Lee, J., Lee, J.E., Kim, N.W., and *Kim, H.,* 2020: Monthly precipitation forecasting in the Han River Basin, South Korea, using large-scale teleconnections and multiple regression models. *Water 12*(6), 1590. https://doi.org/10.3390/w12061590

Lee, J., Kim, C.G., Lee, J.E., Kim, N.M., and *Kim, H.,* 2018: Application of artificial neural networks to rainfall forecasting in the Geum River Basin, Korea. *Water 10*(10), 1448. https://doi.org/10.3390/w10101448

Luk, K.C., Ball, J.E., and *Sharma, A.,* 2000: A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting. *J. Hydrol. 227*, 56–65. https://doi.org/10.1016/S0022-1694(99)00165-1

Lin, F.J., 2008: Solving multicollinearity in the process of fitting regression model using the nested estimate procedure. *Qual. Quant. 42*, 417–426. https://doi.org/10.1007/s11135-006-9055-1

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. *J. Hydrol. 503*, 11–21. https://doi.org/10.1016/j.jhydrol.2013.08.035

Muslehuddin, M., and *Faisal, N.,* 2006: Long range forecast of Sindh monsoon. *Pak. J. Meteorol. 3*, 35–44.

Mahmood, A., Khan, T.M.A., and *Faisal, N.,* 2006: Relationship between El Nino and summer monsoon rainfall over Pakistan. *Pak. J. Marine Sci. 15*, 161–178.

Menke, W., and *Menke, J.,* 2009: Environmental data analysis with MATLAB. Elsevier, London.

Maier, H.R., and *Dandy, G.C.,* 2000: Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications. *Environ. Model. Softw. 15*, 101–124. https://doi.org/10.1016/S1364-8152(99)00007-9

Muslehuddin, M., Mir, H., and *Faisal, N.,* 2005: Sindh Summer (June-september) Monsoon Rainfall Prediction. *Pak. J. Meteorol. 2*, 91–108.

Rehman, A., Jingdong, L., Shahzad, B., Chandio, A.A., Hussain, I., Nabi, G., Iqbal, M.S., 2015: Economic perspectives of major field crops of Pakistan: An empirical study. Pacific Sci. Rev. B: Humanit. Soci. Sci. 1, 145–158. https://doi.org/10.1016/j.psrb.2016.09.002

Rasul, G., Afzal, M., Zahid, M., and *Bukhari, S.A.A.,* 2012: Climate Change in Pakistan Focused on Sindh Province, Technical Report PMD 25/2012 55. Pakistan Meteorological Department, Islamabad, Pakistan, 2012, https://doi.org/10.13140/2.1.2170.6560

Rashid, A., 2004: Mpact of El-Nino on summer monsoon rainfall of Pakistan. *Pak. J. Meteorol. 1*, 35–43.

Solomon, S., 2019: Understanding the impacts of climate change on water access and the lives of women in Tharparkar District, Sindh Province, Pakistan: A Literature Review 1990-2018.

Sarfaraz., 2007: Monsoon dynamics: Its behavioral impact in Pakistan's perspective. *Pak. J. Meteorol. 4*, 55–73.

Shukla, R.P., Tripathi, K.C., Pandey, A.C., and *Das, I.M.L.,* 2011: Prediction of Indian summer monsoon rainfall using Niño indices: A neural network approach. *Atmos. Res. 102*, 99–109. https://doi.org/10.1016/j.atmosres.2011.06.013

SWP, 2021: Climate of Sindh. https://sindhweatherportal.wordpress.com/climate-of-sindh/. 1 January 2021.