Journal of
Climatology & Weather Forecasting

Adnan et al., J Climatol Weather Forecasting 2016, 4:3
DOI: 10.4172/2332-2594.1000185

Research Article Open Access

Predicting the Frequency and Intensity of Climate Extremes by Regression Models

Adnan M[1], Rehman N[1] and Shahbir J[2]

[1]*Global Change Impact Studies Centre (GCISC), Islamabad, Pakistan*

[2]*Department of Statistics, Quaid-I-Azam University, Islamabad, Pakistan*

[*]**Corresponding author:** Muhammad Adnan, Global Change Impact Studies Centre (GCISC), 6th Floor, Emigration Tower, 10-Mauve Area, G-8/1, Islamabad, Pakistan, Tel: +92-321-5553378; E-mail: scarpion982@gmail.com

Received date: Sep 01, 2016; **Accepted date:** Nov 29, 2016; **Published date:** Dec 9, 2016

Abstract

This study examines the relationship of climate extremes indices with the large-scale factors like Sea Level Pressure (SLP) and Sea Surface Temperature (SST). The prediction of extreme indices is carried out and is based on statistical downscaling using the extreme indices data, National Centers for Environmental Prediction (NCEP) monthly SLP and SST reanalysis data. For this purpose, five extreme indices (PRCPTOT, R95p, RX5day, TN90p and TX90p) are developed by using homogenized and high quality daily data of temperature and precipitation for the period 1961-2010 of 10 meteorological stations of monsoon-dominated region of Pakistan. These indices are then average to develop an average time series of each extreme index. To check the assumption of regression model, extreme indices data are tested for heteroscedasticity, auto-correlation and normality. All extreme indices are independent, normal and homogeneous. These indices data are then used as predictand and SLP & SST datasets are used as predictors in regression model. Data for period 1961-2000 and 2001-2010 is used for training and validation purpose respectively. Stepwise regression procedure is adopted to compute regression coefficients based on algorithm of Jennrich. Predictors having strong correlation with extreme indices are identified and a regression model is developed using these predictors and also apply cross-validation technique. Performance of regression model and cross validation models is tested by using statistical measures (Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and bias). The performance is seen reasonably high both in training and validation period. The actual and estimated values show a close agreement. It is seen that ensemble mean prediction obtain from cross-validation models well estimated the extreme indices than the regression model. This study is useful because extremes have a large impact on human society & economy and causing huge losses of the country. The timely prediction of extremes is a major factor and will help the policy makers to take necessary measures for reducing the huge losses.

Keywords: Climate extreme indices; Predictors; Regression model; Root mean square error; Sea surface temperature; Cross validation; Sea level pressure

Introduction

Some natural climate variations can significantly alter the behavior of extreme events (Intergovernmental Panel on Climate Change (IPCC)), Third Assessment Report [1]. Extreme weather events can have major impacts on society, the economy and the environment [2]. Any change in the frequency or severity of extreme climate events could have profound impacts on nature and society. It is thus very important to analyze extreme events [3].

The climate extreme Index (CEI) was first introduced in early 1996 [4] with the goal to summarize and present a complex set of multivariate and multidimensional climate changes in United States so that the results could be easily understood and used in policy decisions made even by non-specialists in the field. This is the first tool developed as a framework for quantifying observed changes in climate extremes. Over the last few decades, the globe has witnessed numerous extreme weather events, including hurricanes, severe cyclonic storms, floods, droughts, heat waves and cold spans [5].

The South Asia region especially Pakistan was a no exception to that and witnessed too a good number of weather extremes in the recent past. Pakistan experienced the last century's worst flood in Jhelum River in 1992. According to IPCC TAR [1], that over the period 1990 to 2100, the average global temperature would increase by 1.4-5.8°C and would be subject to increase in frequency and intensity of extreme climate events (floods, droughts, extreme temperatures etc.). Pakistan also faced the country's worst drought during the period 1998-2001 [6].

Conversely, a record 620 mm of rain fell in Islamabad, Pakistan during 10 h in July 2001 bringing urban storm flooding and causing catastrophic losses to life and property [7]. Such events have led to many studies of observed changes in temperature and precipitation extremes.

Possible changes in extreme event frequency receive considerable attention along with the global warming, because extremes directly impact human society and the economy. For most societally sensitive extremes and related changes in their variability, an analysis based on daily data becomes necessary.

In this study R Clim Dex and RH Test were used in which daily digitized data of temperature (maximum and minimum) and total precipitation for the period 1961-2010 is used for calculating Expert Team on Climate Change Detection and Monitoring Indices

Citation: Adnan M, Rehman N, Shahbir J (2016) Predicting the Frequency and Intensity of Climate Extremes by Regression Models. J Climatol Weather Forecasting 4: 185. doi:10.4172/2332-2594.1000185

(ETCCDMI) core climate indices [8]. The quality control and homogeneity test has performed by using above mentioned tests to ensure the data quality, variability and sufficiency.

These indices are based on daily temperature and precipitation data. These indices were then used to develop the average indices of monsoon-dominated region and further used these average indices to develop the regression model for the prediction purpose. Nicholls et al. [9] found a strong relationship between El-Nino Southern Oscillation and climate extreme events in the East Asia-west Pacific region.

Strong correlations were observed between extremes indices and El-Nino Southern Oscillation index in months prior to the occurrence of the extremes, indicating that predictions of extreme temperatures should be feasible. It was suggested that the relationship between the El-Nino Southern Oscillation and extremes indices would be a useful. Interest in climate variations has experienced a significant increase in recent years due to the important economic and social consequences connected with extreme weather events [10].

Extreme events have a large impact on the society and ecosystems. Therefore, the scientific community and different end-users are interested in future changes of extreme events [11]. According to IPCC TAR [1] some natural climate variations such as ENSO (El-Nino Southern Oscillation), PDO (Pacific Decadal Oscillation), IOD (Indian Ocean Dipole) and NAO (Northern Atlantic Oscillation)/NAM (Northern Hemisphere Annular Mode), can significantly alter the behavior of extreme events, including floods, droughts, hurricanes and cold waves. Studies from throughout much of the world have shown a general increase in extreme precipitation events over the past few decades [12].

The extreme events play an important role in nature and in our daily life because they are often associated to destructive events, e.g. hurricanes, strong earthquakes, etc. In this respect, the predictability of extreme events is urgently desired but also intensely debated [13]. The timely prediction of extremes is a major factor and will help the policy makers to take necessary measures for reducing the huge losses.

In this study we are analyzing the relationship of the climate extremes indices with the large scale factors like mean sea level pressure and sea surface temperature for the prediction of climate extreme indices.

For this purpose first, we developed the climate extreme indices using the daily station data and then these climate extreme indices data and NCEP monthly mean sea level pressure (MSLP) and monthly sea surface temperature (SST) reanalysis data sets prepared for different months/combination of months are used to develop correlation graph to identify the initial predictors for regression models and to predict extreme indices. Different predictors are used to develop a scheme for the prediction of extreme indices.

Predictors having strong correlation with extreme indices are identified and a regression model is developed using these predictors.

Climate Extreme Indices

Climate extreme indices used in the study with their definition are shown in Table 1. The definitions of climate extreme indices used for the analysis are given below.

ID	Indicator Name	Definitions	Units
TN90p	Warm nights	Percentage of days when TN>90th percentile	Days
TX90p	Warm days	Percentage of days when TX>90th percentile	Days
PRCPTOT	Annual total wet-day precipitation	Annual total precipitation in wet days (daily rainfall ≥1 mm)	mm
R95p	Very wet days	Annual total PRCP when RR>95th percentile	mm
RX5day	Maximum 5-day precipitation amount	Monthly maximum consecutive 5-day precipitation	mm

Table 1: Extreme indices/indicators used in the study.

Tn90p

Let $Tnij$ be the daily minimum temperature on day in period and let be the calendar day 90th percentile centered on a 5-day window. The percentage of time is determined where: $Tnij > Tnin90$.

Tx90p

Let $Txij$ be the daily maximum temperature on day in period and let be the calendar day 90th percentile centred on a 5-day window. The percentage of time is determined where: $Txij > Txin\,90$.

Rx5day

Let $RRkj$ be the precipitation amount for the 5-day interval ending, period. Then maximum 5-day values for period j are: $Rx5dayj = \max(RRkj)$.

R95p

Let $RRwj$ be the daily precipitation amount on a wet day $w(RR \geq 1.0\ mm)$ in period and let $RRwn95$ be the 95th percentile of precipitation on wet days in the 1961-1990 period. If W represents the number of wet days in the period, then: $R95pj = \sum_{w=1}^{W} RRwj$ where $RRwj > RRwn\,95$.

PRCPTOT

Let $RRij$ be the daily precipitation amount on day i in period j. If I represents the number of days in j, then $PRCPTOTj = \sum_{i=1}^{I} RRij$.

Station Data

The region selected for the study is Zone-1 (b) sub-mountain region (monsoon-dominated region) of Pakistan. This region is located within the latitudes 31.5°N-35°N and longitudes 72°E-74.5°E. The variables used are Maximum temperature (T_{max}), Minimum temperature (T_{min})

Citation: Adnan M, Rehman N, Shahbir J (2016) Predicting the Frequency and Intensity of Climate Extremes by Regression Models. J Climatol Weather Forecasting 4: 185. doi:10.4172/2332-2594.1000185

and Precipitation (Prec). The daily data of these three variables of 10 stations that are located in the monsoon dominated region of Pakistan for the period (1961-2010) are used. Table 2 shows the details of station data used for the analysis including the variables used, time period and coordinates of the stations. Figure 1 shows the location and the name of the stations in the study region.

Station	Variables used	Time Period	Latitude	Longitude
Balakot	T_{max}, T_{min}, Prec	1961-2010	34.38°N	73.35°E
Garhi Dupatta	T_{max}, T_{min}, Prec	1961-2010	34.22°N	73.62°E
Islamabad	T_{max}, T_{min}, Prec	1961-2010	33.62°N	73.10°E
Jhelum	T_{max}, T_{min}, Prec	1961-2010	32.93°N	73.72°E
Kakul	T_{max}, T_{min}, Prec	1961-2010	34.18°N	73.25°E
Kotli	T_{max}, T_{min}, Prec	1961-2010	33.52°N	73.90°E
Lahore	T_{max}, T_{min}, Prec	1961-2010	31.50°N	74.33°E
Murree	T_{max}, T_{min}, Prec	1961-2010	33.92°N	73.38°E
Muzaffarabad	T_{max}, T_{min}, Prec	1961-2010	34.37°N	73.48°E
Sialkot	T_{max}, T_{min}, Prec	1961-2010	32.50°N	74.53°E

Table 2: Details of the station data used.

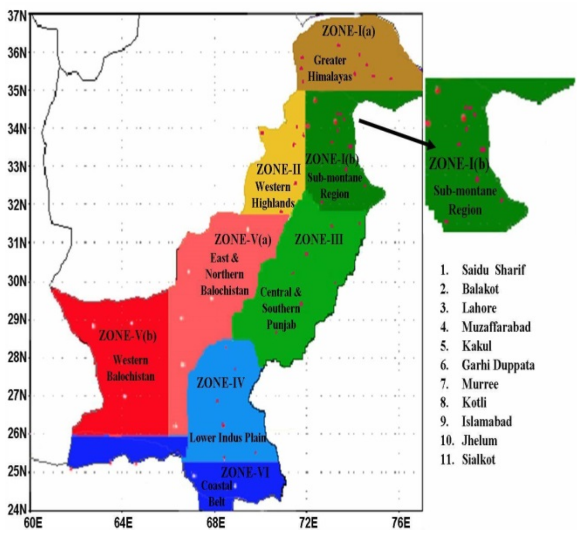

Figure 1: Locations and names of the stations used for the analysis.

Gridded Data

National Centre for Environment Prediction (NCEP) reanalysis Global Gridded data sets of Monthly Mean Sea Surface Temperatures (SST) and Monthly Mean Sea Level Pressure (MSLP) of the resolution 2° × 2° and 2.5° × 2.5° respectively for the period of 1960-2010 are used for analysis in this study. The NCEP data provide a consistent long-term representation of the large-scale climate [14].

Development of Climate Extremes Indices

Five (05) climate extreme indices are used for the study to check prediction out of 27 core climate indices recommended by ETCCDMI [8]. Two (02) of these indices (TN90p, TX90p) relate to temperature and three (03) (PRCPTOT, R95p, RX5 day) relate to precipitation. Climate extreme indices were developed on annual basis by using R-ClimDex for a time series data of daily temperature (maximum and minimum) and precipitation of total of 10 meteorological stations for the period of 1961-2010. After developing the annual climate extreme indices of each station, the indices (TN90p, TX90p, PRCPTOT, R95p and RX5day) are averaged over all the 10 stations that are located in the monsoon dominated region of Pakistan.

Preparation of Gridded data

For identification of predictors and to prepare spatial correlation maps the monthly NCEP reanalysis global gridded data of Mean Sea level pressure (MSLP) and Sea Surface Temperature (SST) for the period 1959 to 2010 is prepared for each month, combinations of two months and three months [15]. The individual months is from January (Jan) to December (Dec) and the combinations of two months (averages of two consecutive months) used are Dec-Jan (DJ), Jan-Feb (JF), Feb-Mar (FM), Mar-Apr (MA), Apr-May (AM), May-Jun (MJ), Jun-July (JJ), July-Aug (JA), Aug-Sep (AS), Sep-Oct (SO), Oct-Nov (ON), Nov-Dec (ND). Similarly, the combination of three months (average of three consecutive months) used are Nov-Dec-Jan (NDJ), Dec-Jan-Feb (DJF), Jan-Feb-Mar (JFM), Feb-Mar-Apr (FMA), Mar-Apr-May (MAM), Apr-May-Jun (AMJ), May-Jun-Jul (MJJ), Jun-Jul-Aug (JJA), Jul-Aug-Sep (JAS), Aug-Sep-Oct (ASO), Sep-Oct-Nov (SON), Oct-Nov-Dec (OND). Each time series consists of the values only for the relevant month or the combination of the months from each year. For example, the Dec data are the mean value of month December for each of the 51 years at each grid point. We have a total 36 combination of global gridded data for each of SST and SLP, including 12 monthly files, 12 two-monthly files and 12 three-monthly files and then the data for these months, combination of two months and combination of three months are used to developed the correlation graph and identify the initial predictors.

Identification and Selection of Predictors for Regression Models

The relationship of each averaged climate extreme indices data (PRCPTOT, R95p, RX5day, TN90p and TX90p) and the climate variables such as NCEP global gridded monthly mean sea level pressure (MSLP) and monthly sea surface temperature (SST) reanalysis data sets is observed. For the extreme events prediction purpose, NCEP global gridded data prepared for different months, combination of two months and three months are used to develop the correlation graph of each climate extreme indices with these months and combinations of two & three months. Predictors having significantly high correlation (at 95% confidence level) with each climate extreme index (PRCPTOT, R95p, RX5day, TN90p and TX90p) are identified and used these predictors in a regression model. After identifying the predictors, multiple linear regression models are developed for each averaged climate extreme index, using the index data as dependant variable and predictors as independent variables. The training period for the regression model is 1961-2000. In building a regression model to check the prediction of each averaged climate extreme indices, as such we need to check the assumptions underlying the Classical Linear

Regression Model, these are: i) The number of observations must be greater than the number of regressors in the model, ii): Regressors have no significant linear correlation among them. Addition of all the predictors (independent variables) in the regression model could increase R2 but decrease the reliability of regression. Step-wise regression is adopted to overcome this problem and help select the final predictors for the regression model. Step-wise regression method works by adding the predictors in the model one by one and allows only those predictors that are uncorrelated with each other and give the optimum value of R2, keeping in view the highest possible accuracy level. Step-wise multiple regression procedure is adopted to compute regression coefficients based on the algorithm of Jennrich [16]. Predictors significant at 5% level are entered in the model and an equation is developed for each averaged climate extreme indices. The final regression models thus obtained are shown below. The predictors in the regression models are having the month/bi-month/three-monthly combination name and the data used either SST or SLP (e.g.: Dec(sst) means predictor of the month December for SST & NDJ (slp) means predictor of the three-monthly combination NDJ for SLP).

$$PRCPTOT = -153105 - 118.497Dec(sst) + 23.553May(slp) + 23.632Sep(slp) + 22.978NDJ(slp) + 47.86SON(slp) + 36.157June(slp) - 47.791MJJ(sst) \quad (1)$$

$R^2 = 72.7\%$, $Adj.R^2 = 67.6\%$,
Multiple Correlation Coefficient = 0.853

The regression model for PRCPTOT explains 72.7% variation of the data with multiple correlation co-efficient is 0.853. The regression coefficients along with the respective predictors and corresponding months/combination of months of this regression model are shown in Table 3 and the final predictors are shown in Figure 2. The Figure 2 shows that five predictors are from the variable MSLP in the month of May, September, June and combination of months SON, NDJ while two predictors are from the variable SST in the month of December and combination of months MJJ [17-22]. The analysis shows that MSLP and SST have an important role in the prediction of extreme events in Pakistan. The predictors Dec (sst) and NDJ (slp) are from North Pacific Ocean, predictors May (slp) and SON (slp) are from South Pacific Ocean whereas the predictors MJJ (sst), June (slp) and Sep (slp) are from North Atlantic Ocean, South Atlantic Ocean and Indian Ocean respectively.

Predictors	Longitudes	Latitudes	Month/combination of months	Coefficients
Constant	-	-	-	−153105
Dec(sst)	179W:167W	27N:37N	December (Dec)	−118.497
May(slp)	156E:166E	36S:26S	May	23.553
Sep(slp)	41E:54E	41S:34S	September (Sep)	23.632
NDJ(slp)	152W:132W	35N:45N	NDJ (Nov, Dec, Jan)	22.978
SON(slp)	120W:105W	30S:20S	SON (Sep, Oct, Nov)	47.86
June(slp)	5E:14E	29S:21S	June (June)	36.157
MJJ(sst)	51W:41W	45N:51N	MJJ (May, June, July)	−47.791

Table 3: Final predictors of the regression model for PRCPTOT.

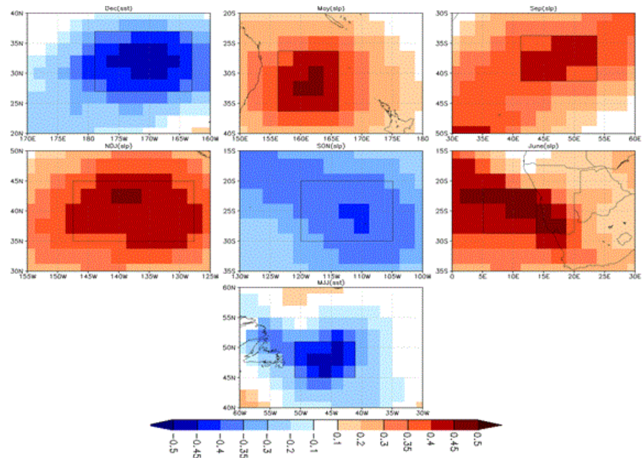

Figure 2: Location of final predictor selected in regression model for PRCPTOT.

$$R95p = -64961.3 - 55.692SON(slp) + 56.754June(slp) - 68.718Dec(sst) + 5.335May(slp) + 41.850ND(slp) + 21.455June(slp) - 7418Apr(sst) - 63.905Sep(sst) \quad (2)$$

$R^2 = 74.2\%$, $Adj.R^2 = 68.5\%$,
Multiple Correlation Coefficient = 0.861

The regression model for R95p explains 74.2% variation of the data with multiple correlation co-efficient is 0.861. The regression coefficients along with the respective predictors and corresponding months/combination of months of this regression model are shown in Table 4 and the final predictors are shown in Figure 3. The Figure 3 shows that five predictors are from the variable MSLP in the month of May, June and combination of months SON, OND while three predictors are from the variable SST in the month of April, September and December. The predictors SON (slp) and OND (slp) are from North Pacific Ocean, predictors June (slp) and Sep (sst) are from South Pacific Ocean, predictors May (slp) and Dec (sst) are from North Atlantic Ocean whereas the predictors June (slp) and Apr (sst) are from South Atlantic Ocean and Indian Ocean respectively.

Predictors	Longitudes	Latitudes	Month/combination of months	Coefficients
Constant	-	-	-	−64961.3
SON(slp)	166W:151W	26N:39N	SON (Sep, Oct, Nov)	−55.692
June(slp)	150E:170E	16S:4S	June (June)	56.754
Dec(sst)	29W:19W	33N:45N	December (Dec)	−68.718
May(slp)	11W:1E	59N:64N	May	5.335
OND(slp)	154W:144W	21N:29N	OND (Oct, Nov, Dec)	41.85
June(slp)	0:15E	29S:21S	June (June)	21.455
Apr(sst)	45E:56E	3S:3N	April (Apr)	−74.18
Sep(sst)	175W:159W	7S:0	September (Sep)	−63.905

Table 4: Final predictors of the regression model for R95p.

Citation: Adnan M, Rehman N, Shahbir J (2016) Predicting the Frequency and Intensity of Climate Extremes by Regression Models. J Climatol Weather Forecasting 4: 185. doi:10.4172/2332-2594.1000185

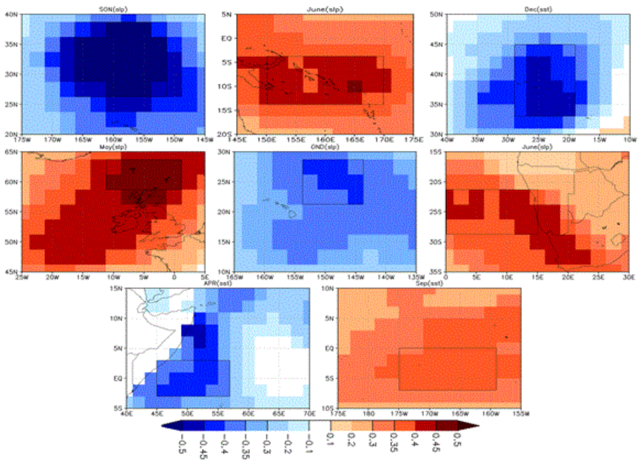

Figure 3: Location of final predictor selected in regression model for R95p.

$$RX5day = -41613.6 - 12.898SON(slp) + 21.378AS(slp) + 20.058MJ(slp) + 20.996July(sst) - 44.514Apr(sst) + 13.684DJ(slp) \quad (3)$$

$R^2 = 71.4\%$, $Adj.R^2 = 66.9\%$,
Multiple Correlation Coefficient = 0.845

The regression model for RX5day explains 71.4% variation of the data with multiple correlation co-efficient is 0.845. The regression coefficients along with the respective predictors and corresponding months/combination of months of this regression model are shown in Table 5 and the final predictors are shown in Figure 4. The Figure 4 shows that four predictors are from the variable MSLP in the combination of months SON, AS, DJ and MJ while two predictors are from the variable SST in the month of April and July. The predictors SON (slp), DJ (slp) and AS (slp) are from North Pacific Ocean, whereas the predictors MJ (slp), July (sst) and Apr (sst) are from South Pacific Ocean, South Atlantic Ocean and Indian Ocean respectively.

Predictors	**Longitudes**	**Latitudes**	**Month/combination of months**	**Coefficients**
Constant	-	-	-	−41613.6
SON(slp)	176W:161W	29N:39N	SON (Sep, Oct, Nov)	−12.898
AS(slp)	131W:121W	21N:34N	AS (Aug, Sep)	21.378
MJ(slp)	159E:176E	9S:4N	MJ (May, June)	20.058
July(sst)	17W:3W	3S:5N	July (July)	20.996
Apr(sst)	113E:123E	21S:11S	April (Apr)	−44.514
DJ(slp)	124E:136E	21N:29N	DJ (Dec, Jan)	13.684

Table 5: Final predictors of the regression model for RX5 day.

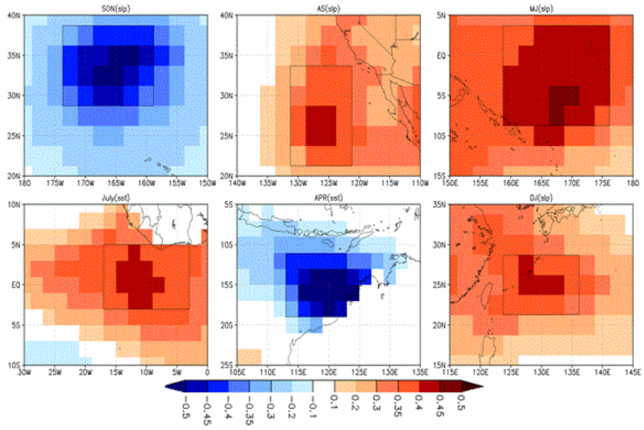

Figure 4: Location of final predictor selected in regression model for RX5 day.

$$TN90p = 1171.847 + 1.83JJA(slp) - 2.565JAS(slp) + 0.111Feb(slp) + 4.967ASO(slp) - 2.599Sep(slp) + 2.323Feb(sst) - 1.973AS(slp) + 0.942MJJ(slp) \quad (4)$$

$R^2 = 80.6\%$, $Adj.R^2 = 76.3\%$,
Multiple Correlation Coefficient = 0.898

The regression model for TN90p explains 80.6% variation of the data with multiple correlation co-efficient is 0.898. The regression coefficients along with the respective predictors and corresponding months/combination of months of this regression model are shown in Table 6 and the final predictors are shown in Figure 5. The Figure 5 shows that seven predictors are from the variable MSLP in the month of September, February and combination of months AS, MJJ, ASO, JJA, JAS while one predictor is from the variable SST in the month of February. The predictors MJJ (slp), Feb (sst), Sep (slp) and JJA (slp) are from North Pacific Ocean, predictor Feb (slp) is from South Pacific Ocean whereas the predictors AS (slp), JAS (slp) and ASO (slp) are from North Atlantic Ocean.

Predictors	**Longitudes**	**Latitudes**	**Month/combination of months**	**Coefficients**
Constant	-	-	-	1171.85
JJA(slp)	149E:166E	24N:34N	JJA (June, July, Aug)	1.83
JAS(slp)	39W:24W	26N:34N	JAS (July, Aug, Sep)	−2.565
Feb(slp)	89W:74W	31S:21S	February (Feb)	0.111
ASO(slp)	31W:16W	24N:36N	ASO (Aug, Sep, OCT)	4.967
Sep(slp)	111W:94W	4N:19N	September (Sep)	−2.599
Feb(sst)	145E:163E	21N:29N	February (Feb)	2.323
AS(slp)	66W:51W	19N:26N	AS (Aug, Sep)	−1.973
MJJ(slp)	149E:164E	24N:34N	MJJ (May, June, July)	0.942

Table 6: Final predictors of the regression model for TN90p.

Citation: Adnan M, Rehman N, Shahbir J (2016) Predicting the Frequency and Intensity of Climate Extremes by Regression Models. J Climatol Weather Forecasting 4: 185. doi:10.4172/2332-2594.1000185

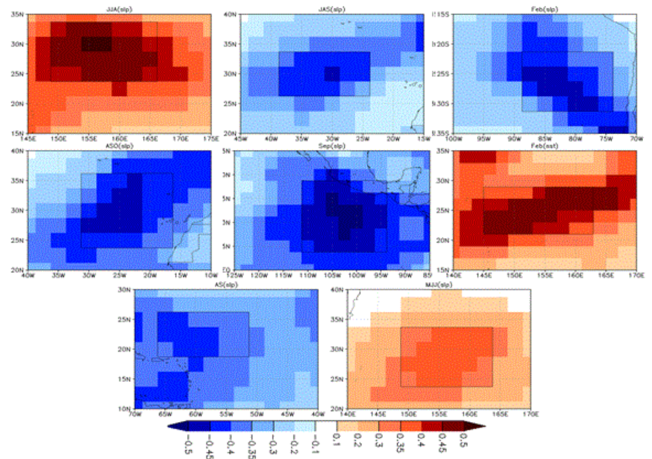

Figure 5: Location of final predictor selected in regression model for TN90p.

$$TX90p = 6245.59 - 2.856Dec(slp) - 1.957JAS(slp) - 1.736MA(slp) + 0.407July(slp) \quad (5)$$

$R^2 = 56.7\%$, $Adj.R^2 = 52.3\%$,
Multiple Correlation Coefficient = 0.753

The regression model for TX90p explains 56.7% variation of the data with multiple correlation co-efficient is 0.753. The regression coefficients along with the respective predictors and corresponding months/combination of months of this regression model are shown in Table 7 and the final predictors are shown in Figure 6.

The Figure 6 shows that four predictors are from the variable MSLP in the month of December, July and combination of months MA, JAS. The predictors MA (slp) and JAS (slp) are from North Atlantic Ocean whereas the predictors Dec (slp) and July (slp) are from South Pacific Ocean and Indian Ocean respectively.

Predictors	Longitudes	Latitudes	Month/combination of months	Coefficients
Constant	-	-	-	6245.59
Dec(slp)	171E:179W	21S:9S	December (Dec)	−2.856
JAS(slp)	41W:29W	31N:41N	JAS (July, Aug, Sep)	-1.957
MA(slp)	66W:54W	19N:31N	MA (Mar, Apr)	−1.736
July(slp)	104E:121E	41S:31S	July (July)	0.407

Table 7: Final predictors of the regression model for TX90p.

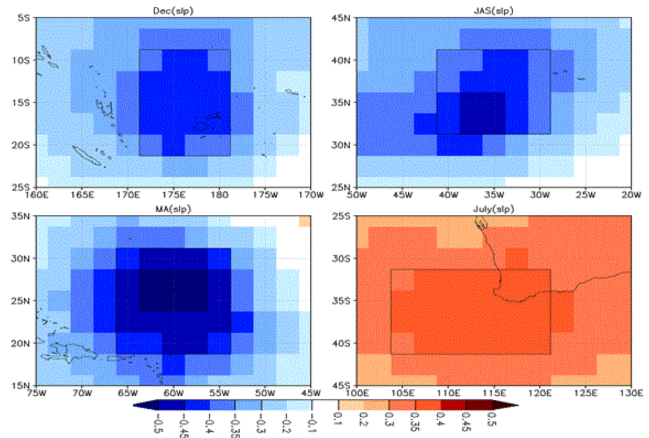

Figure 6: Location of final predictor selected in regression model for TX90p.

Validation of Regression Models

Performance of the regression models for each climate extreme index for the validation period 2001-2010 is tested by using four statistical techniques namely Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and bias to evaluate the skill of the prediction. The estimated values of each climate extreme index are validated with the values of their respective climate extreme index calculated from the observed data for the validation period 2001-2010. Table 8 shows the results of the validation for the regression models.

Year	PRCPTOT	R95p	RX5day	TN90p	TX90p
S.D	185.84	112.57	45.72	3.35	5.5
Bias	39.56	3.69	2.49	−0.89	−3.45
MAE	96.2	66.52	30.92	1.97	3.8
RMSE	123.58	85.77	35.65	2.39	4.69

Table 8: Validation results of the regression models.

$$\text{Bias} = \frac{1}{n}\sum_{t=1}^{n}(\widehat{Yt} - Yt)$$

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|\widehat{Yt} - Y_t|$$

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n}(Y_t - \widehat{Y}_t)^2}{n}}$$

Where Yt are the observed values and Yt are the estimated values.

The projected time series shows less variability than observed, which is reflected in the low RMSE and MAE. As both statistical measures (RMSE & MAE) identify the error between observed and estimated value therefore these errors tends to zero and or otherwise to be less than the variability of the observed data for a good performance of the models results. The model performs well if the MAE and RMSE

Citation: Adnan M, Rehman N, Shahbir J (2016) Predicting the Frequency and Intensity of Climate Extremes by Regression Models. J Climatol Weather Forecasting 4: 185. doi:10.4172/2332-2594.1000185

are found less than or equal to the standard deviation of the observed data. As it is seen from the above results, RMSE and MAE both are less than standard deviation of the observed data for each climate extreme index which in a broader sense shows the regression models perform better and there is a close agreement between estimated and observed values. It is seen that regression equation well estimated the climate extreme indices.

Figures 7a-7e show the Observed and estimated values for the training and validation periods for each climate extreme index.

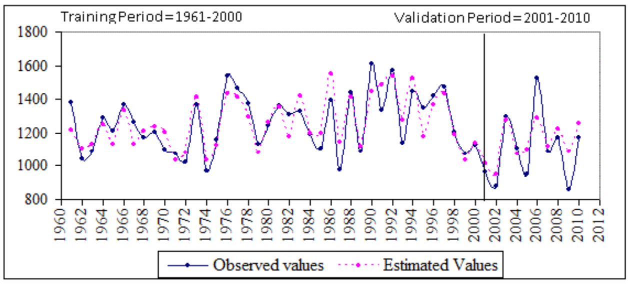

Figure 7a: shows the observed and estimated values for PRCPTOT.

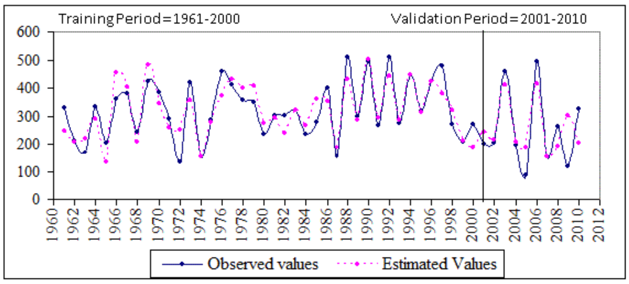

Figure 7b: Shows the observed and estimated values for R95p.

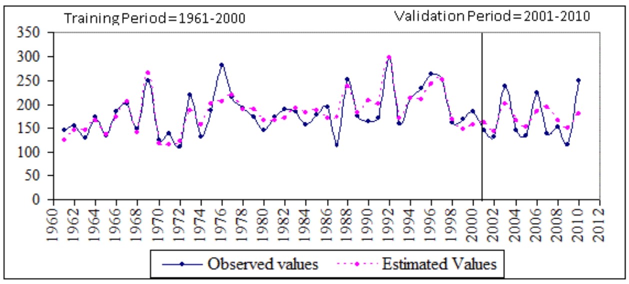

Figure 7c: Shows the observed and estimated values for RX5 day.

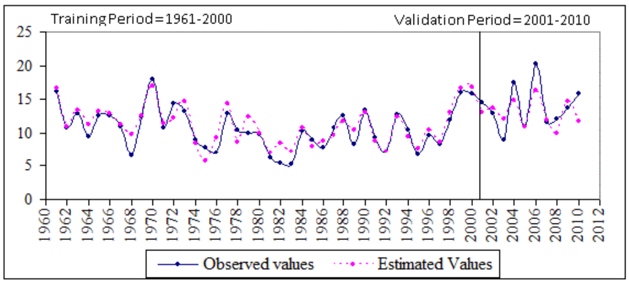

Figure 7d: Shows the observed and estimated values for TN90p.

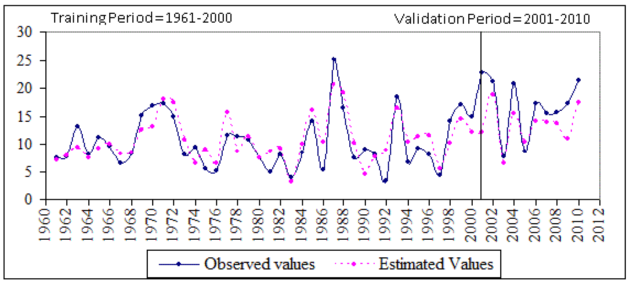

Figure 7e: Shows the observed and estimated values for TX90p.

Cross-Validation

The cross-validation method used in this study is defined as the correlation between the N-5 years for the predictors and the predictand over the monsoon dominated region of Pakistan. These correlations are calculated and the predictors are identified by using t-test at 95% confidence level. This method is used for whole training period (1961-2000). For example, if we have a data for the training period 1961-2000 like

61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, ………..…… ……..95, 96, 97, 98, 99, 00

61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, ………..…… ……..95, 96, 97, 98, 99, 00

and so on ……….

61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, …………….... ……..95, 96, 97, 98, 99, 00

In the first step, remove 1st five (05) years (61-65) and select the predictors with remaining N-5 years for the regression. The climate extreme indices are then predicted for the whole data (1961-2010) with this regression and check the skill of the model. In next step, remove next five (05) years (66-70) and again select the predictors with remaining N-5 years for the regression. The climate extreme indices are then predicted for the whole data (1961-2010) with this regression and check the skill of the model. This procedure will continue till the last step where, remove last five (05) years (96-00) and again select the predictors with remaining N-5 years for the regression. The climate extreme indices are then predicted for the whole data (1961-2010) with this regression and check the skill of the model. In this study the forty (40) years i.e. 1961-2000 is used for training period therefore after applying the cross-validation method, there are eight (08) regression model based on N-5 years data. We make the ensemble mean of all the eight (08) predictions for the whole data (1961-2010) obtain from the cross-validated regression models. The performance of the cross-validation is also tested for the verification period (2001-2010). To measure the skill of the ensemble mean prediction of the cross-validation regression models on the basis of the results obtained from training period (1961-2000) following statistics are computed: Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Bias. The results obtain from the cross-validation regression models are shown in the Table 9. As it is seen from the results, RMSE and ABSE both are less than standard deviation of the observed data of each climate extreme index, which shows that there is a close agreement between estimated (ensemble mean prediction) and observed extreme indices. It is seen that ensemble mean prediction of cross-validation regression models well estimated the extreme events. (Figures 8a-8e) shows the Observed and estimated (ensemble mean) values for the training and validation periods for each climate extreme index.

Year	PRCPTOT	R95p	RX5day	TN90p	TX90p
S.D	185.84	112.57	45.72	3.35	5.5
Bias	29.72	3.04	2.19	−0.33	−3.09
MAE	71.08	45.95	25.31	1.52	3.44
RMSE	89.31	56.24	29.81	1.8	4.33

Table 9: Results of the cross-validation regression models.

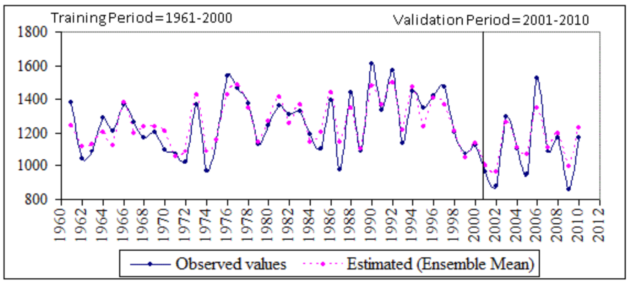

Figure 8a: Shows the observed and estimated (Ensemble Mean) values for PRCPTOT.

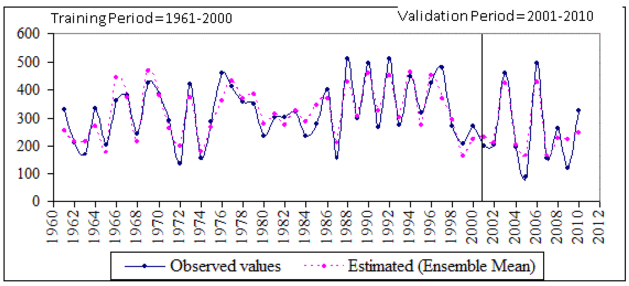

Figure 8b: Shows the observed and estimated (Ensemble Mean) values for R95p.

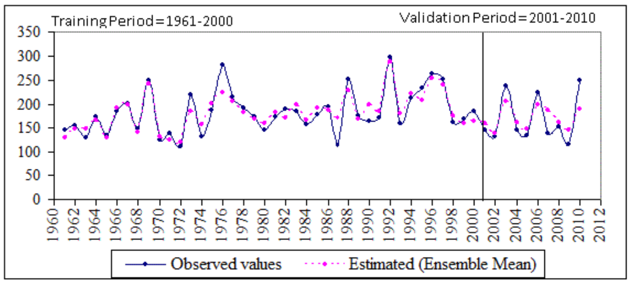

Figure 8c: Shows the observed and estimated (Ensemble Mean) values for RX5day.

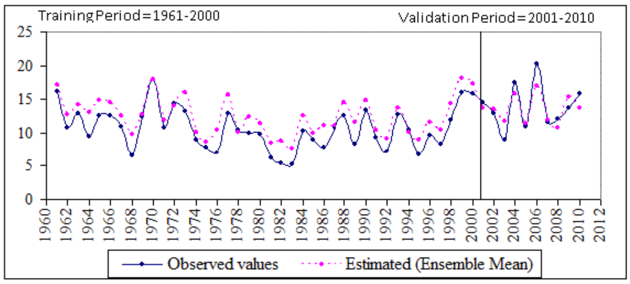

Figure 8d: Shows the observed and estimated (Ensemble Mean) values for TN90p.

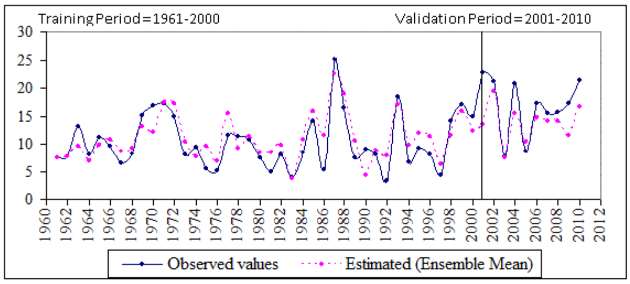

Figure 8e: Shows the observed and estimated (Ensemble Mean) values for TX90p.

Summary and Conclusion

Climate Extreme events have major impact on the economy and society and causes huge losses of the country, therefore the prediction of these extreme events are very important. Climate extreme indices are developed for a total of 10 stations lying in the monsoon dominated region of Pakistan by using the software R-ClimDex and then these indices are averaged over all the 10 stations. After developing the climate extreme indices, the predictors are identified and regression model is developed for each extreme index by using the stepwise regression procedure. The regression model for PRCPTOT, R95p, RX5day, TN90p and TX90p respectively explains 72.7%, 74.2%, 71.4%, 80.6% and 56.7% variation of the data, and with their respective multiple correlation coefficient are 0.853, 0.861, 0.845, 0.898 and 0.753.

The Figures 2-6 shows that MSLP and SST of the pacific and Indian ocean have an important role in the prediction of extreme events in monsoon dominated region of Pakistan which in a broad sense, indicate that strong easterly winds from pacific to Indian ocean that affect the atmospheric conditions of the region. However, no attempt is made to attribute any physical linkage to the predictors-extreme relationship. The regression models have estimated and captured the pattern of each extreme index during training period and verification period quite good. The observed and estimated values of climate extreme indices have shown the values of Bias, MAE, RMSE in validation period respectively as 39.56, 96.2 and 123.58 for PRCPTOT, 3.69, 66.52 and 85.77 for R95p, 2.49, 30.92 and 35.65 for RX5day, -0.89, 1.97 and 2.39 for TN90p and −3.45, 3.80 and 4.69 for TX90p. The standard deviation of the PRCPTOT, R95p, RX5day, TN90p and TX90p are respectively as 185.84, 112.57, 45.72, 3.35 and 5.50. As it is seen from the results, that the values of MAE and RMSE both are less than the standard deviation of the data for each climate extreme index which in a broader sense shows that the regression model perform better and there is a close agreement between the observe and estimated values of each climate extreme index.

After applying the cross-validation technique, the estimated (Ensemble Mean) values captured the pattern of each extreme index during training and verification period quite good. The Estimated (Ensemble Mean) prediction obtain from the cross validation regression for PRCPTOT, R95p, RX5day, TN90p and TX90p respectively explains 81.8%, 77.5%, 81.6%, 88.0% and 70.5% variation of the data, and with their respective multiple correlation coefficient are 0.905, 0.880, 0.903, 0.938 and 0.840. The estimated (ensemble mean) extreme indices of the training and verification period have shown a close agreement with the observed extreme indices data. The observed and estimated (ensemble mean) values of climate extreme indices have shown the values of Bias, MAE, RMSE in validation period respectively as 29.72, 71.08 and 89.31 for PRCPTOT, 3.04, 45.95 and 56.24 for R95p, 2.19, 25.31 and 29.81 for RX5day, −0.33, 1.52 and 1.80 for TN90p and −3.09, 3.44 and 4.33 for TX90p. The standard deviation of the PRCPTOT, R95p, RX5day, TN90p and TX90p are respectively as 185.84, 112.57, 45.72, 3.35 and 5.50. As it is seen from the results, that the values of MAE and RMSE both are less than the standard deviation of the data for each climate extreme index which in a broader sense shows that the regression model perform better and there is a close agreement between the observe and estimated values of each climate extreme index. From the above mentioned results, it is seen that ensemble mean prediction obtain from cross-validation regression models well estimated the climate extreme indices than the multiple linear regression model.

References

1. IPCC TAR (2001) Climate Change 2001. Impacts, adaptation and vulnerability. Cambridge University Press, London.
2. Manton MJ, Della-marta, PM Haylock, MR Hennessy, KJ Nicholls, et al. (2001) Trend in extreme daily rainfall and temperature in Southeast Asia and South Pacific 1961-1998. Int J Climatol 21: 269-284.
3. Zhang X, Hegerl G, Zwiers FW, Kenyon J (2005) Avoiding inhomogeneity in percentile-based indices of temperature extremes. JCLI 18: 1641-1651.
4. Karl TR, Richard WK, David RE, Robert GQ (1996) Indices of climate change for the United States. Bull Am Meteorol Soc 77: 279-292.
5. Easterling DR, Evans JL, Groisman PY, Karl TR, Kunkel KE, et al. (2000) Observed variability and trends in extreme climate events. A brief review. Bull Am Meteorol Soc 81: 417-425.
6. Sheikh MM (2001) Drought management and prevention in Pakistan. Science Vision Quarterly. COMSATS 1st Meeting on Water Resources in the South. Present Scenario and Future Prospects.
7. Rasul GR, Chaudhry QZ, Zhao S, Zeng Q (2004) A diagnostic study of record heavy rain in twin cities Islamabad-Rawalpindi. Adv Atmosp Sci 21: 976-988.
8. Alexander LV, Zhang X, Peterson TC, Caesar J, Gleason B, et al. (2006) Global observed changes in daily climate extremes of temperature and precipitation. J Geophys Res 111: D05109.
9. Nicholls N, Baek HJ, Gosai A, Chambers LE, Choi Y, et al. (2005) The El Niño–Southern Oscillation and daily temperature extremes in east Asia and the west Pacific. Geophys Res Lett 32: L16714.
10. Abaurrea J, Cebrian AC (2001) Trend and variability analysis of rainfall series and their extreme events. Chapter in detecting and modelling regional climate change. Springer-Verlag 191-202.
11. Lustenberger A, Knutti R (2010) Projections of extreme indices over Europe from a pattern scaling approach. 10th EMS Annual Meeting. 10th European Conference on Applications of Meteorology (ECAM) Abstracts held Sept. 13-17, 2010 in Zurich, Switzerland.
12. Peralta-Hernandez AR, Balling RC, Barba-Martinez LR (2009) Comparative analysis of indices of extreme rainfall events. Variation and trends from southern Mexico. Atmosfera 22: 219-228.
13. Caruso F, Kantz H (2010) Prediction of extreme events in the OFC model on a small world network. Eur Phys J B 79: 7-11.
14. Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, et al. (1996) The NCEP/NCAR 40-year Reanalysis Project. Bull Am Meteorol Soc 77: 437-471.
15. Munot AA, Krishna KK (2007) Long range prediction of Indian summer monsoon rainfall. J Earth Sys Sci 116: 73-79.
16. Jennrich R (1977) Stepwise Regression. Statistical methods for digital computers (eds.) John Wiley. New York pp: 58-75.
17. Amici M, Tomozeiu R, Cacciamani C, Morgillo A (2004) Statistical downscaling method for extreme temperature events in Emilia Romagna. Geophys Res Abs 6: 06273.
18. Easterling DR, Meehl GA, Parmesan C, Changnon SA, Karl TR, et al. (2000) Climate extremes: observations, modeling, and impacts. Science 289: 2068-2074.
19. IPCC AR (2007) Climate Change 2007. Impacts, adaptation and vulnerability. Cambridge University Press, London.
20. Senior CA, Jones RG, Lowe JA, Durman CF, Hudson D (2002) Predictions of extreme precipitation and sea-level rise under climate change. Philosophical Transactions of the Royal Society 360: 1301-1311.
21. Sheikh MM, Manzoor N, Adnan M, Ashraf J, Arshad MK (2009) Climate profile and past climate changes in Pakistan. GCISC-RR- 01. Global Change Impact Studies Centre. Islamabad. Pakistan.
22. Zhu Y, Zoltan T (2005) Extreme weather events and their probabilistic prediction by the NCEP ensemble forecast system. Environmental Modeling Center, NCEP. NWS/NOAA, Washington DC, p: 20233.