Clim Dyn (2016) 47:2653–2666
DOI 10.1007/s00382-016-2990-1

A statistical downscaling model for summer rainfall over Pakistan

Dildar Hussain Kazmi[1,5] · Jianping Li[2,3] · Chengqing Ruan[4] · Sen Zhao[1,5] · Yanjie Li[1]

Received: 27 August 2015 / Accepted: 14 January 2016 / Published online: 1 February 2016

Abstract A statistical approach is utilized to construct an interannual model for summer (July–August) rainfall over the western parts of South Asian Monsoon. Observed monthly rainfall data for selected stations of Pakistan for the last 55 years (1960–2014) is taken as predictand. Recommended climate indices along with the oceanic and atmospheric data on global scales, for the period April–June are employed as predictors. First 40 years data has been taken as training period and the rest as validation period. Cross-validation stepwise regression approach adopted to select the robust predictors. Upper tropospheric zonal wind at 200 hPa over the northeastern Atlantic is finally selected as the best predictor for interannual model. Besides, the next possible candidate 'geopotential height at upper troposphere' is taken as the indirect predictor for being a source of energy transportation from core region (northeast Atlantic/western Europe) to the study area. The model performed well for both the training as well as validation period with correlation coefficient of 0.71 and tolerable root mean square errors. Cross-validation of the model has been processed by incorporating JRA-55 data for potential predictors in addition to NCEP and fragmentation of study period to five non-overlapping test samples. Subsequently, to verify the outcome of the model on physical grounds, observational analyses as well as the model simulations are incorporated. It is revealed that originating from the jet exit region through large vorticity gradients, zonally dominating waves may transport energy and momentum to the downstream areas of west-central Asia, that ultimately affect interannual variability of the specific rainfall. It has been detected that both the circum-global teleconnection and Rossby wave propagation play vital roles in modulating the proposed mechanism.

Keywords Interannual model · ENSO · Summer monsoon · CVSR · CGT

✉ Dildar Hussain Kazmi
dhkazmi@gmail.com

✉ Jianping Li
ljp@bnu.edu.cn

[1] State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

[2] College of Global Change and Earth System Sciences (GCESS), Beijing Normal University, Beijing 100875, China

[3] Joint Center for Global Change Studies, Beijing 100875, China

[4] North China Sea Marine Forecasting Center of State Oceanic Administration, Qingdao 266061, China

[5] University of Chinese Academy of Sciences, Beijing 100049, China

1 Introduction

Pakistan is located at the western edge of Summer Monsoon affected region in South Asia. During 2010, the country has faced an abnormal flooding condition due to heavy rainfalls in the northern parts (Webster et al. 2011). The summer monsoon over Pakistan is mainly concentrated during July to August (Ding and Ke 2013). Summer monsoon in Pakistan covers the period from July to September and contributes nearly 60 % of the annual total rainfalls. Pre-monsoon (generally May–June) is considered very hot and dry season with only localized convective rains occur occasionally (Rasul et al. 2004).

The Asian summer monsoon (ASM) has been a most significant aspect of the summer circulation in northern hemisphere (Li and Zhang 2009). Its relevant teleconnection

patterns have got pronounced importance in recent past (Zhang 1999). In most of the Asian countries the summer monsoon stands for major rainfall of the annual total. Large scale phenomena like El Nino Southern Oscillation (ENSO) have shown close interaction with ASM especially in modulating interannual variability. Moreover, as its year to year variation largely affects the annual gross product therefore impacts on the socioeconomic sector on global level as well (Wang et al. 2006).

Climate change and subsequently the threat of global warming is a bitter reality. Its aftermaths in the countries with agriculture based developing economy are more pronounced, not only in the present time but in the coming era as well (Rasul and Kazmi 2011). The areas which lie in active monsoon zone in the country are often blessed with over 200 mm rainfall but such events turns into disaster where this heavy rain makes the total of the year. Southern half of Pakistan holds Balochistan and Sindh provinces where annual total rainfall remained just a few hundreds of millimeters that occur during summer in 15–20 days (Rasul and Kazmi 2013). The consequences of climate change have more evidently been observed in case of summer monsoon over south Asia in the recent years, poor rainfalls during 2009 and historic heavy rains in 2010 may be the best examples.

Based on observational analysis and model simulations it is concluded that circumglobal teleconnection (CGT) and summer North Atlantic Oscillation in the extra-tropics are mainly responsible for interannual variability in South Asian Monsoon (SAM), covers the area northwest India and north Pakistan (Syed et al. 2011). Ding and Wang (2005) have observed a mid-latitude CGT pattern during summer over the northern hemisphere with significant correlations with SAM and that is independent of ENSO. They found that the positive phase of the CGT have close interaction with summer monsoon over northwestern India and Pakistan. According to Syed et al. (2011), the positive (negative) phases of the proposed CGT have shown corresponding responses over increased (decreased) Summer Monsoon over Pakistan and north western India.

Syed et al. (2010) found that an upper tropospheric warm anticyclone anomaly develops at the north west of Pakistan a few days before the start of the active phase of monsoon over western parts of SAM. Summer monsoon over Pakistan and north western India has been observed as closely associated with the upper level anomalous high over west-central Asia (Ding and Wang 2005).

Based on observational analysis it is expected that on interannual/interseasonal levels the strong variability in the upper-tropospheric circulation over the northeastern Atlantic in the Jetstream exit region may be associated with the barotropic instability of the summer mean flow (Ding and Wang 2007). According to Simmons et al. (1983), over the northeastern Atlantic the instabilities in the jet stream exit region may be excited by barotropic energy conversions that are closely connected to the zonal gradients of the basic flow, based on the particular location. Teleconnection patterns (either zonally or meridionally oriented) are considered like stationary waves originating from a source region. The zonally oriented ones are developed along the westerly jet, such as the Eurasian and the CGT patterns (Hoskins and Ambrizzi 1993).

In Pakistan besides climatic constraints, inadequate water supply throughout the cropping season has become a challenging task. Localized and reliable information in terms of atmospheric and hydrological prospectus is required for near future for better planning. Accordingly, we have to get local climate projections besides GCM outputs through best available sources like downscaling (Kazmi et al. 2014). Downscaling has been proved as a better option for obtaining regional climate conditions by employing atmospheric parameters (Christensen et al. 2007; Li et al. 2013). Statistical downscaling is based on the assumption that the past time associations may hold in future as well (Wilby 1997). The choice of predictors and study area is very important for any statistical downscaling model and based on some recent studies it is critical for future studies as well (Frias et al. 2006; Schmidli et al. 2007). For future studies regarding climate as well agro-meteorology, reliable atmospheric data is direly needed for the study region (Kazmi et al. 2015). The method is quite inexpensive and practicable for a developing country like Pakistan, which is facing the consequent impacts of climate change with lesser resources and poor infrastructure.

In the present study, we worked out an interannual model (IAM) for JA rainfall over Pakistan. The first section of this study holds the relevant introduction regarding the specified rainfall, techniques and scope. In the second chapter, we have given the data description and methods adopted. It includes a comprehensive discussion about the data sets incorporated for comparison along with important equations, as well. The third chapter holds an elaboration about the analysis applied in the study and the main findings. The fourth and last section covers a short discussion to summarize all the significant results. We have included the relevant figures and tables as per requirements.

1.1 Main rainfall season and study area

Generally, summer monsoon in Pakistan occurs during June–September with main concentration in July–August period (Suleman et al. 1995). Climatic data record also shows that two months (July and August) hold the prominent rainfalls in the country (Fig. 1a), accordingly same period has been adopted in the present study.

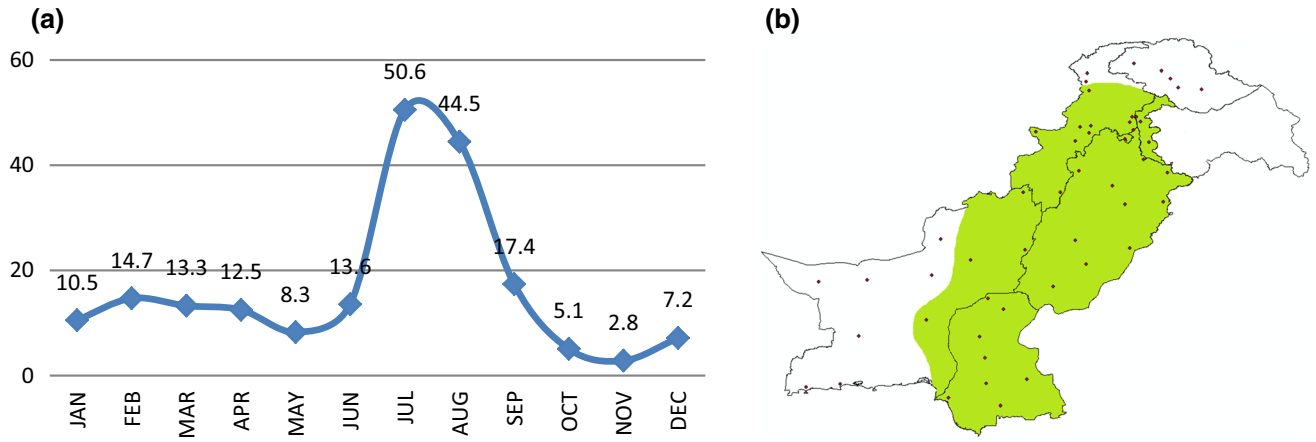

Fig. 1 Monthly area weighted rainfall (mm) for Pakistan (**a**), Location of meteorological stations in Pakistan (**b**) available for climate data, *green color* shows the monsoon zone

The summer rainfall in Pakistan is not homogenous but mainly confined in upper half and southeast parts (Ding and Ke 2013). Climatic data shows that summer monsoon mainly affects some specific parts of Pakistan, so the present study focus only on the relevant areas (green part) of the country (Fig. 1b). Presently data of 35 stations of Pakistan have been employed which includes; stations located in two provinces as a whole (Punjab and Sindh), Khyber Pakhtunkhwa "KP" (excluding north western parts), Azad Jammu Kashmir "AJK" and four stations from Baluchistan province. These additional 4 stations are located near Punjab and KP provinces and receive considerable rainfall during summer.

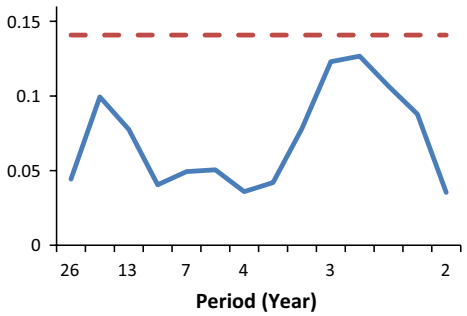

Fig. 2 Power spectrum for JA rainfall, *dashed line* shows the 95 % confidence level for *red/white* noise

2 Data and methods

Observed monthly rainfall data for July–August of all station of Pakistan (as provided by Pak Met Department) for the period 1960–2014 (55 years) has been incorporated. For comparison; the available JA rainfall data for the same period from GPCC and CRU (0.5° × 0.5°) has also been utilized.

Predictors data on monthly basis for the same period (1960–2014) is mainly taken from NCEP (2.5° × 2.5°) including the data for recommended climate indices; NAO, ONI, TNI, MEI, Nino 1 + 2, Nino 3, Nino 3.4, SST, SLP, HGT (850, 500, 400, 300 and 200 hPa), zonal and meridional winds (850 and 200 hPa), air temperatures at 850 hPa and relative humidity at 850 hPa. All these data sets are utilized for 3 months mainly (April–June) based on monthly, two monthly and three monthly formats except HGT at 200 hPa which is utilized for July and August additionally and HGT, zonal and meridional winds at 850, 400 and 300 hPa which are utilized for June only. Besides for model's validation, monthly data for HGT and zonal wind (for June) at 200 hPa is taken from JRA-55 (1.25° × 1.25°) for the whole study period. Data for three climate indices SASMI, SASMI1 and SASMI2 are utilized for June only for the whole study period, which is available online at; http://ljp.gcess.cn/dct/page/65576.

Primarily, 55 years data is incorporated for Power Spectrum. The result of power spectrum shows that there exists two pecks (equivalent values) in the curve indicating the interannual and interdecadal variability for rainfall (Fig. 2). Whereas, dashed line indicates the 95 % confidence level for red or white noise. Consequently, rainfall data for the whole study period 1960–2014 (N = 55) is distributed into the calibration/training period 1960–1999 (n = 40) and independent test period 2000–2014 (n = 15). And to calibrate the model, for the whole study period both the predictand and the predictors data is decomposed to get interannual component through Fourier decomposition filtering.

For investigation of spatial and temporal variation of JA rainfall, EOF analysis has been adopted; the relevant outcome will be discussed later. In the next step, the rainfall and the predictors data sets are decomposed into annual and decadal parts. To find out the potential predictors all the specified recommended climate indices are employed for the same period and dimension. Generally, it is considered that the summer monsoon in south Asia (Pakistan) is largely affected by ENSO, therefore initially all of them have been taken as candidate predictors ignoring their significance levels in correlation with the rainfall during training period (1960–1999). In the next step, the rainfall has been tested for any correlation with the listed meteorological parameters on global basis like SLP, SST and indices developed accordingly for all the grid points showing significant correlations.

Following Guo et al. (2012, 2014) from all the candidate predictors the robust ones are selected through correlation based cross-validation stepwise regression (CVSR) method for calibration of annual model. This approach of downscaling has mainly two stages; (1) through correlation analysis the identification of potential predictors from the set of indices for global field data and the climate indices, (2) selection of robust predictors from the set of potential predictors through CVSR method. It is based on leave-one-out cross validation approach to select the robust predictors from the potential predictors by reducing the false possibility. To verify the reliability of the predictor, root-mean-square error occurred between the observation and the cross-validated estimate has been set as criteria.

Dispersion equation for propagation of (barotropic) Rossby wave with horizontally non-uniform flow and holding the meridional component (e.g., Li and Li 2012; Li et al. 2015; Zhao et al. 2015) may be written as,

$$\omega = \bar{u}_M k + \bar{v}_M l + \frac{l\partial\bar{q}/\partial x - k\partial\bar{q}/\partial y}{K^2} \tag{1}$$

where $(\bar{u}_M, \bar{v}_M) = (\bar{u}, \bar{v})\cos\varphi$ is the ambient velocity for Mercator projection, $\partial\bar{q}/\partial x$ and $\partial\bar{q}/\partial y$ are the zonal and meridional gradients (respectively) of absolute vorticity.

Group velocity vector C_g, from Eq. (1) may be written as

$$u_g = \bar{u}_M + \left[\left(k^2 - l^2\right)\partial\bar{q}/\partial y - 2kl\partial\bar{q}/\partial x\right]\Big/ K^4 \tag{2a}$$

Similarly,

$$v_g = \bar{v}_M + \left[2kl\partial\bar{q}/\partial y + \left(k^2 - l^2\right)\partial\bar{q}/\partial x\right]\Big/ K^4 \tag{2b}$$

Lighthill (1978) states that, wave ray is a kind of trajectory which is tangential to the group velocity on smaller scales. Accordingly, wavenumbers k and l, which are variables for ray trajectory, can be found by following kinetic wave theory as stated by Whitham (1960), as

$$\frac{d_g k}{dt} = -k\frac{\partial\bar{u}_M}{\partial x} - l\frac{\partial\bar{v}_M}{\partial x} + \frac{k\partial^2\bar{q}/\partial y\partial x - l\partial^2\bar{q}/\partial x^2}{K^2} \tag{3a}$$

Similarly,

$$\frac{d_g l}{dt} = -k\frac{\partial\bar{u}_M}{\partial y} - l\frac{\partial\bar{v}_M}{\partial y} + \frac{k\partial^2\bar{q}/\partial y^2 - l\partial^2\bar{q}/\partial x\partial y}{K^2}, \tag{3b}$$

where, $d_g/dt = \partial/\partial t + \mathbf{c}_g \cdot \nabla$, represents the material derivative moving with the group velocity.

2.1 Employment of modeled data of precipitation for comparison

Long time past data for precipitation over Pakistan is not available as per WMO standards, on spatial and temporal scales. Therefore, for comparison and reliability of results from the core study, the data from two state of the art sources is utilized; CRU and GPCC (Fig. 3).

Based on more similarity with the observed rainfall (Fig. 3), GPCC data has been selected for further analysis like EOF over the whole country. To proceed, GPCC data is extracted (1960–2010, as available) for the same locations as utilized for observed data over the study area, accordingly a comparison of both these is shown in Fig. 4.

2.2 EOF for observed rainfall

EOF analysis has been applied for observed data (1971–2010, as since 1971 the consistent data is available) of monsoon period JA for Pakistan. The leading mode of EOF is contributed about 40 % of the total variance and the first seven modes contribute more than 83 % (not shown), so rests may be neglected. All over the country there are positive loadings with two maxima in north eastern parts (Fig. 5a). These areas are connected with north western India and receive heavy rains during the monsoon period. The temporal values are also shown in Fig. 5b, for leading mode of EOF (observed rainfall), showing almost the same pattern as by the observed rainfall.

2.3 EOF for GPCC rainfall

Figure 6a shows the spatial distribution of leading mode of EOF (GPCC data for 1971–2010) with about 49 % of the total variance. Same as in case of observed data, all over the country there are positive loadings, with maxima in north eastern parts depicting higher concentration from rest of the country (Fig. 6a). Its is also observed that first

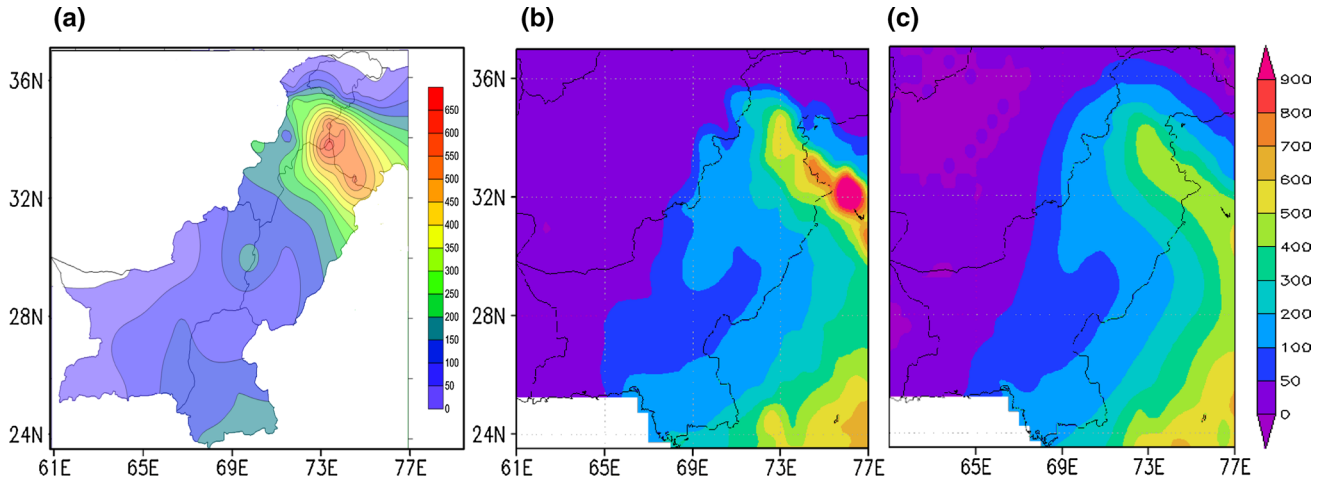

Fig. 3 Average observed rainfall for JA (**a**) for the period 1960–2010, GPCC (**b**) and CRU (**c**)

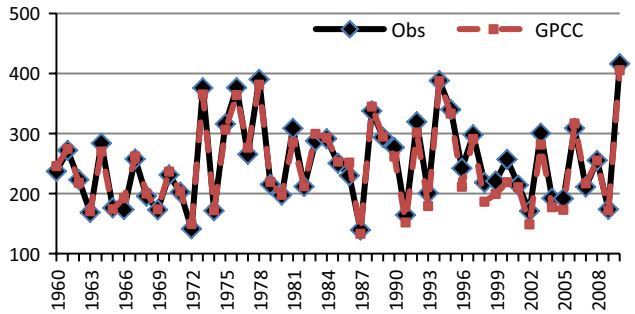

Fig. 4 Observed and GPCC rainfall for JA for the period 1960–2010

5 PCs contribute about 83 % of the total variance (not shown) so rest may be neglected. The temporal values are also shown (Fig. 6b) for leading mode of EOF (for GPCC data), displaying almost the same pattern as by the observed rainfall.

2.4 Cross-validation of model's output

Before physical verification of the robustness of the predictors we have applied two techniques, to reconfirm the model's output. Primarily, following Gutiérrez et al. 2013, the whole data period of 55 years is fragmented into five non-overlapping K-folds (each fold contains 11 years) test samples. The first sample contains the years 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005 and 2010. Similarly, the next sub group will start from 1961 with 1966 and 1971 as the second and third years. The purpose behind is to run the final model for different segments of study period in the ratio of 80 % (training) to 20 % (test period). In this way each test sample (11 years) will be set as independent test period by taking the rest (44 years) as training period. Secondly, in the light of recommendations from some recent studies like Manzanas et al. (2014, 2015) (to avoid any contradiction in the results) JRA-55 data is incorporated for potential predictors besides NCEP.

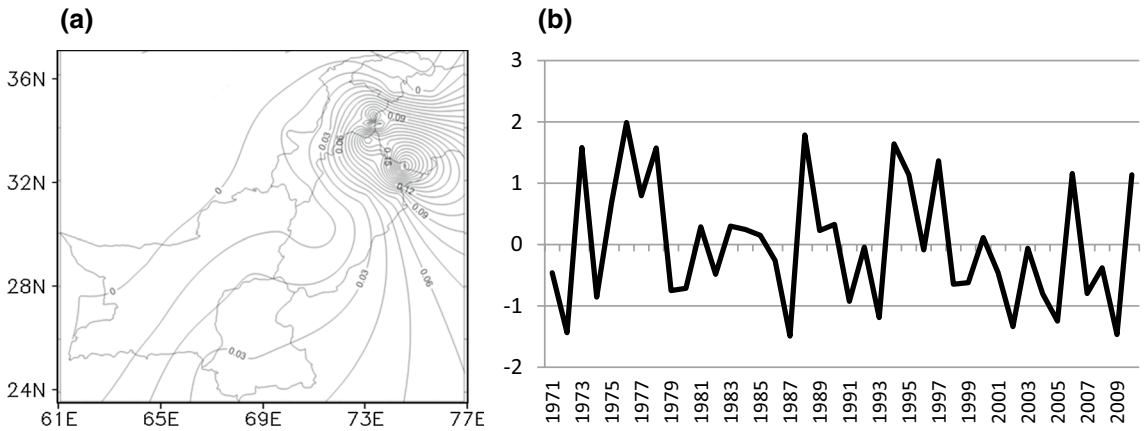

Fig. 5 Leading EOF mode for JA observed rainfall for the period 1971–2010; spatial (**a**) and temporal (**b**), for Pakistan

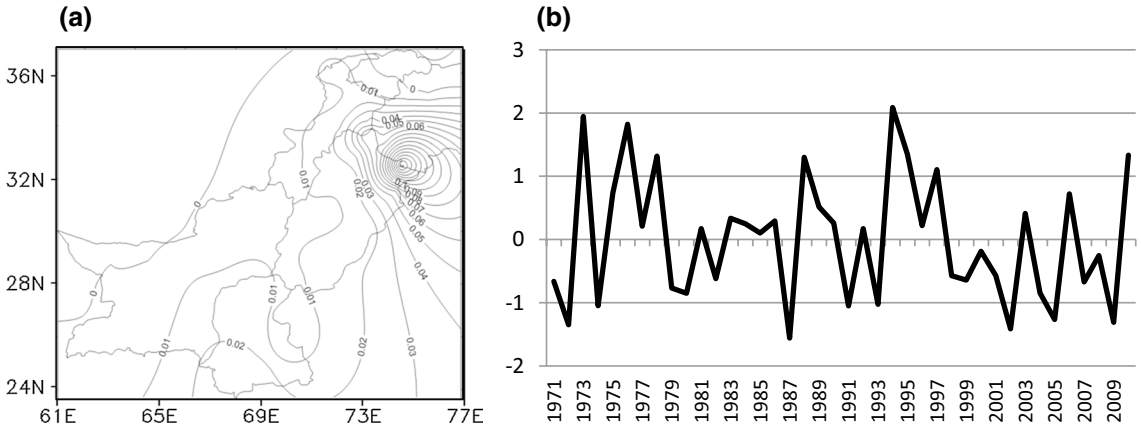

Fig. 6 Leading EOF mode for JA GPCC-rainfall for the period 1971–2010; spatial (**a**) and temporal (**b**), for Pakistan

Table 1 Correlation between climate indices and JA rainfall, over the study area

Period	ONI	Nino1+2	Nino3	Nino 3.4	NAO	SASMI	SASMI1	SASMI2
Training (1960–1999)	.49**	.30	−.46**	−.47**	.43*	.31	.27	.26
Validation (2000–2014)	.68**	.52**	−.74**	−.76**	−.49	−.09	.01	−.71**

** Correlation is significant at the .01 level, * Correlation is significant at the .05 level

3 Results and discussion

All the specified climate indices with their relevant association with JA rainfall are listed in Table 1. It is noticed that only three of them (ONI, Nino 3 and Nino 3.4) have shown significant correlation in both the training as well as independent test periods. In comparison to validation period the correlations remained weaker during the training period. However, as the ENSO and NAO are considered very much associated with the summer rainfall over the south Asian region therefore, all of those three and NAO have been preselected for final stage of CVSR. Syed et al. (2011) stated that CGT and summer NAO in the extra-tropics are two important factors for interannual variability of summer monsoon over northwestern India and Pakistan. To seek more potential predictors, interannual correlation of time series between all the specified fields and JA rainfall during the training period (1960–1999) is calculated.

Based on the higher and significant correlation and lower RMSE only four predictors have been selected from the global field data. Figure 7a–d show the interannual correlation between the specified detrended predictors with the JA rainfall over the study area during the training period 1960–1999. On the same format GPCC data is also taken to find correlation with the same four predictors and almost similar correlation with same area and coefficient are resulted (not shown). Accordingly, indices for these four grid points are selected for the final run of IAM in addition to the mentioned climate indices.

3.1 Outcome of IAM

When we fed the predictors data into IAM, at first stage of CVSR, $U2J_{NEA}$ (Zonal Wind at 200 hPa over Northeast Atlantic during June) has been selected as the robust one through CVSR screening out procedure. Whereas, the relevant correlation coefficient is 0.71 with RMSE within the considerable limits. But the model couldn't processed the second step (no second predictor has been selected) due to failure of F and T tests even with highest correlation coefficient and lowest RMSC. The next possible candidate '$H2J_{NEA}$' (Geopotential Height at 200 hpa over Northeast Atlantic during June) couldn't selected on statistical grounds but later for physical explanation of IAM that appears as the most important indirect predictor. Accordingly, the equation for IAM may be written as

$$Y_A(t) = 36.45 \times U2J_{NEA}(t), \tag{4}$$

where, Y_A (t) is the interannual component of rainfall for 't^{th}' year (t = 1, 2… 40) over the period 1960–1999, $U2J_{NEA}$ is the 't^{th}' value of zonal wind at 200 hPa over northeast Atlantic during June.

Downscaled results obtained from IAM for training (1960–1999) and independent test periods (2000–2014) are shown in the Table 2. Corr is correlation between observation and prediction; Var (Coefficient of Variance) is the ratio of RMSE (mm) to the climatology July–August rainfall during 1960–2014. The projected time series for JA rainfall is plotted against the observed in Fig. 8.

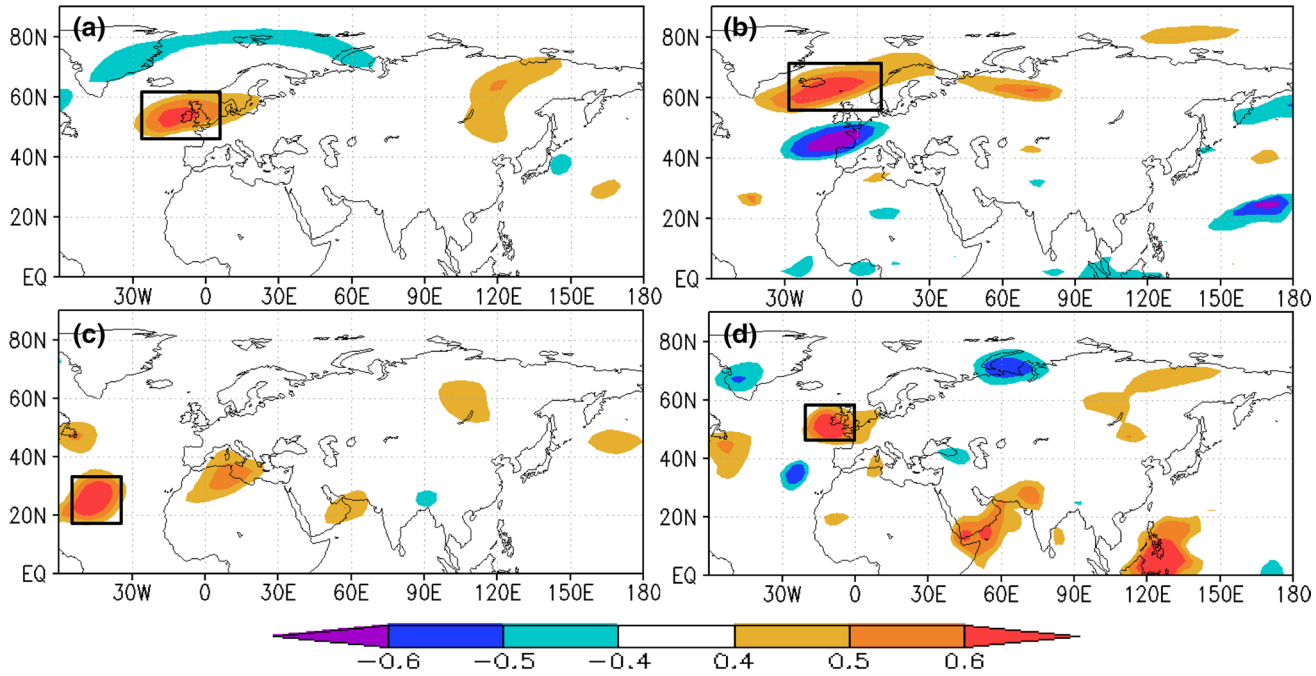

Fig. 7 Interannual correlation between detrended time series of JA rainfall for study area with HGT-200 hPa for June (**a**), Zonal Wind-200 hPa for June (**b**), HGT 500 hPa for May (**c**) and Air Temperature-850 hPa for May–June (**d**), for the training period of 40 years (1960–1999). *Rectangles* show the areas with comparatively higher correlation and lower RMSE, during both the training and independent test period

Table 2 Important statistics regarding IAM output

Model	Training period			Test period		
	Corr	RMSE	Var (%)	Corr	RMSE	Var (%)
IAM	0.71	36.25	14.1	0.71	36.91	14.4

Corr is correlation between observation and prediction, Var (Coefficient of Variance) is the ratio of RMSE (mm) to the climatology July–August rainfall during 1960–2014

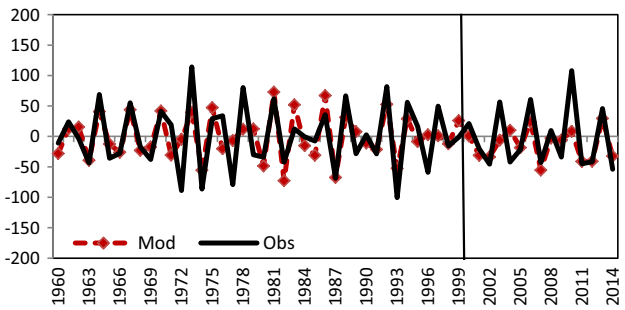

Fig. 8 Time series for both observed and modeled data of IAM, *bar line* separates the training and independent test periods

3.1.1 Output of IAM for simultaneous circulations

In the later section we are to provide physical significance of the robust predictor 'U2J$_{NEA}$' for modifying the targeted rainfall by enhancing the positive circulation over the west-central Asia at 200 hPa during JA (H2JA$_{WCA}$). Therefore, it is better to provide some relevant information regarding the association of 'H2JA$_{WCA}$' for the rainfall. For that we run the model by including 'H2JA$_{WCA}$' as well as a candidate predictor. Accordingly it has been selected as a robust predictor with more improved results, which are shown in Table 3. Figure 9 depicts the relevant correlation region 'west-central Asia' for more elaboration.

3.2 Reconfirmation of the statistical results

For cross checking the outcome of the statistical model, K-folds approach is adopted (as briefed in Sect. 2). We have rerun IAM for each of the independent test samples and entered the corresponding statistics in Table 4. This table shows that IAM has successfully predicted the rainfall for all the sub periods. Although RMSE is quite higher for test period in case of K3 and K4 but the relevant variance is still considerable.

Table 3 Statistics for IAM output based on simultaneous circulations, during core rainy season JA

Model	Training period			Test period		
	Corr	RMSE	Var (%)	Corr	RMSE	Var (%)
IAM	0.81	30.45	11.9	0.72	34.20	13.3

Corr is correlation between observation and prediction, Var (Coefficient of Variance) is the ratio of RMSE (mm) to the climatology July–August rainfall during 1960–2014

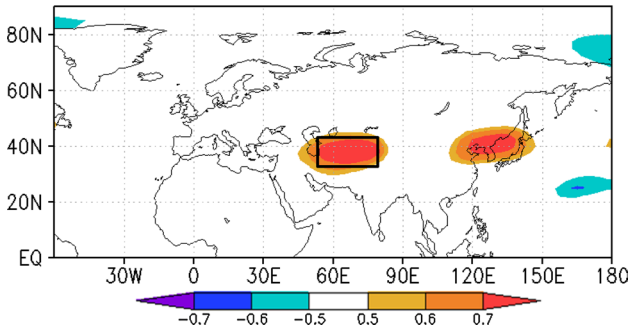

Fig. 9 Interannual correlation between detrended time series of JA-rainfall for study area with HGT-200 hPa, for Jul–Aug, for training period 1960–1999. *Rectangle* shows the area with best correlation and lowest RMSE values

Besides to avoid any contradiction in the results obtained, for comparison in addition to NCEP's data for predictors, we have utilized JRA-55 data set as well in the model. Figure 10 depicts that in case of zonal wind at 200 hPa (for June) both the data sets hold the similar association with the targeted rainfall. Most importantly, the positive circulation over the northeastern Atlantic region is very similar to that produced by NCEP data. Accordingly, IAM is rerun based on data from JRA-55 and the results are listed in Table 5. These results reassure the robustness of the model's predictor to downscale JA-rainfall in the study region. For assessment purpose, the observed as well as the generated time series of rainfalls are plotted in Fig. 11. It is well evident that in our case NCEP data set has shown

Table 4 Output of IAM for K-folds, to predict a particular 'K' test sample, the rest are taken as training period

Prediction set for	Training period (44 years)			Test period (11 years)		
	Corr	RMSE	Var (%)	Corr	RMSE	Var (%)
K1	0.73	38.96	15.2	0.76	31.80	12.4
K2	0.74	38.25	14.9	0.68	35.86	14.0
K3	0.73	35.17	13.7	0.65	45.84	17.9
K4	0.69	35.09	13.7	0.82	48.18	18.8
K5	0.72	38.14	14.9	0.84	36.56	14.3

Like to set 'K1' as test period, all the rest (K2, K3, K4 and K5) will be taken as training period and so on. Corr is correlation between observation and prediction, Var (Coefficient of Variance) is the ratio of RMSE (mm) to the climatology July–August rainfall during 1960–2014

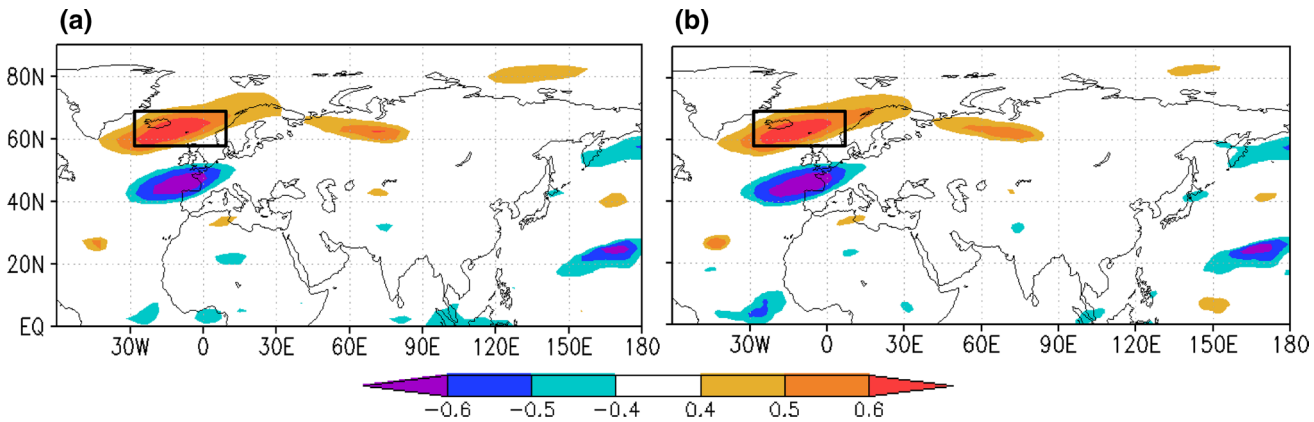

Fig. 10 Interannual correlation between detrended time series of JA-rainfall (observed) for study area with zonal wind at 200 hPa (for June) for training period 1960–1999; by utilizing NCEP data set (**a**) and by JRA-55 (**b**). *Rectangles* show the areas with best correlation and lowest RMSE values

Table 5 Output of IAM, through two independent data sets taken for the robust predictor '$U2J_{NEA}$'

Data source for $U2J_{NEA}$	Training period (1960–1999)			Test period (2000–2014)		
	Corr	RMSE	Var (%)	Corr	RMSE	Var (%)
NCEP	0.71	36.25	14.1	0.71	36.91	14.4
JRA-55	0.69	37.17	14.5	0.71	37.66	14.7

Corr is correlation between observation and prediction, Var (Coefficient of Variance) is the ratio of RMSE (mm) to the climatology July–August rainfall during 1960–2014

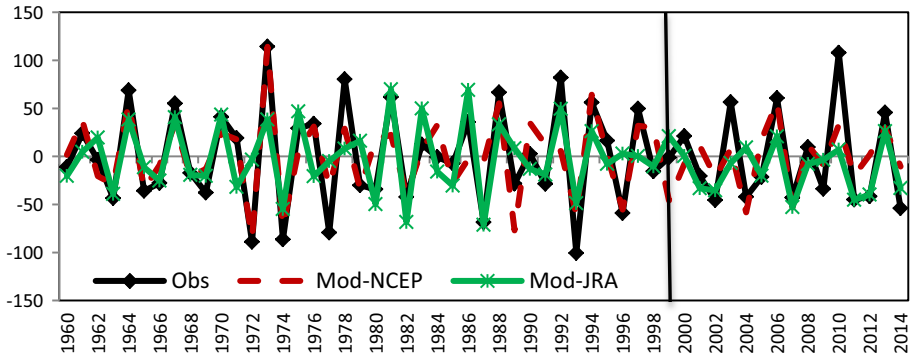

Fig. 11 Comparison between the observed and modeled rainfall for the whole study period, by taking two different data sets, bar line separates the training and independent test periods

better association with the observed rainfall against JRA-55 throughout the study period.

3.3 Physical verification of the model

In order to give physical reasoning behind some possible association and robustness of the selected predictor $U2J_{NEA}$, primarily all the relevant data sets are distributed into three categories based on ±standard deviation of summer monsoon (weak, strong and a difference between strong and weak period) over the study area. Ding and Wang (2005) proposed that near the exit region of westerly jet, strong barotropic instability is developed over the North Atlantic, that consequently help create an anomalous high over the northeast Atlantic region by upstream disturbances. Moreover, they proposed a CGT pattern favored by the basic state, through which the wave energy is transported from the core region to west-central Asia, which is a major factor behind interannual variability of the summer monsoon over the study area.

We suggest a similar mechanism for the reliability of our model's results. Primarily, we start with the observational analysis for zonal wind and geopotential height at 200 hPa for June. It is observed that both of these circulations have positive trends over the north Atlantic region during this month. This is especially true during the years yield strong monsoons in the study regions (Fig. 12b, e) and difference between strong to weak (Fig. 12c, f).

Table 6 shows the association of both $U2J_{NEA}$ and $H2J_{NEA}$ with each other and with the rainfall for different segments of the study period. It is evident that both of these are highly correlated with each other and may be originated from a single source. Accordingly, at the exit jet region, zonal flow may trigger anomalous high during June. And if that high may help enhance a positive circulation over the west-central Asia then it may considered as an important factor for summer rainfall over the study area. Consequently, in the core monsoon season JA '$H2JA_{WCA}$' is highly associated with the rainfall in the study area, for all the test samples from the study area except for K1 (Table 6).

The progressive movement of the HGT at 200 hPa during June, JJ and JA (Fig. 13a–c) may be more helpful to see the chain of associated circulation during strong monsoon period over the study area. It can be observed that during the particular period, an anomalous high develops over the northeast Atlantic region in June '$H2J_{NEA}$' and with the passage of time that help enhance another high over the west-central Asian region '$H2JA_{WCA}$'. However, the situation remains merely different in case of weak monsoon over the study area (Figures not shown). For elaboration of the mutual association between $H2J_{NEA}$ and $H2JA_{WCA}$, a map of point correlation is constructed (Fig. 13d) which projects that besides few other regions (as listed in proposed CGT by Ding and Wang 2005), pressure variations over the northeast Atlantic during summer is nearly in phase with $H2JA_{WCA}$. These findings strongly agree the outcome of some recent findings, which stated that positive circulation over the northeast Atlantic during June is mainly responsible for creating/enhancing $H2JA_{WCA}$. Whereas, the association of '$H2JA_{WCA}$' with JA rainfall is evident from the statistics listed in Tables 3 and 6.

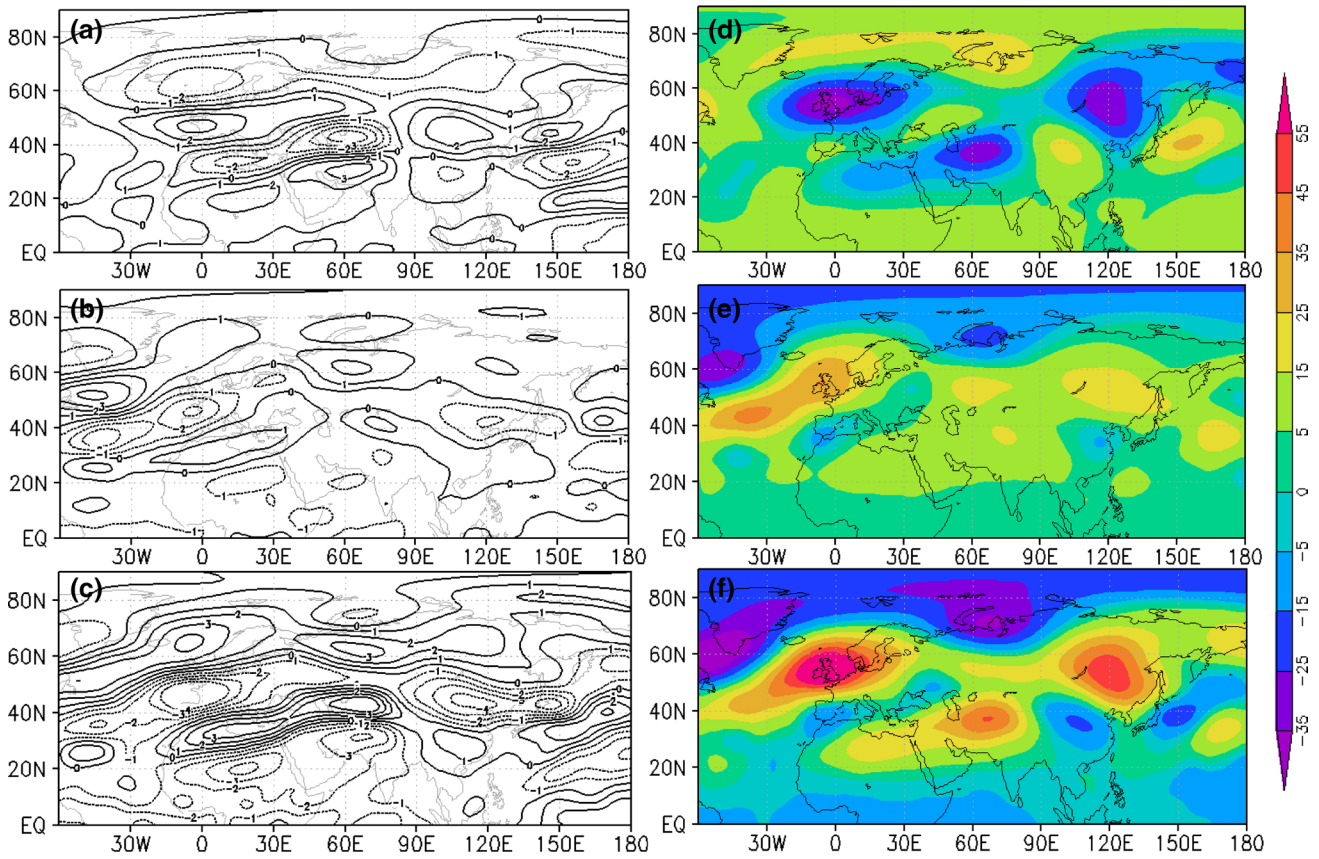

Fig. 12 Detrended zonal wind at 200 hPa for June during weak (**a**), strong (**b**) and Strong–Weak (**c**) monsoon over the study area, for 1960–2014. Similarly, detrended geopotential height at 200 hPa for June, during weak (**d**), strong (**e**) and Strong–Weak (**f**) monsoon

Table 6 Interannual correlation between detrended $U2J_{NEA}$, $H2J_{NEA}$ and $H2JAwc_A$ with targeted rainfall and between the first two predictors as well

Period	$U2J_{NEA}$ and Pre	$H2J_{NEA}$ and Pre	$H2JAwc_A$ and Pre	$H2J_{NEA}$ and $U2J_{NEA}$
Training (1960–1999)	0.71**	0.65**	0.81**	0.83**
Validation (2000–2014)	0.71**	0.76**	0.72**	0.80**
K1	0.76**	0.55	0.50	0.59
K2	0.68*	0.72*	0.77*	0.79*
K3	0.65*	0.58	0.71*	0.87**
K4	0.82**	0.67*	0.90**	0.81**
K5	0.83**	0.66*	0.74*	0.81**

Correlation covers different segments from the study area including training and test period and the specified K-folds

** Correlation is significant at the .01 level, * Correlation is significant at the .05 level

We applied EOF on geopotential height at 200 hPa during the period of interest JJA, leading mode (with 28 % variance) is shown in Fig. 14. The analysis domain covers the relevant regions only in order to be focused on the origin of important circulations and the rainfall area. It can be noticed that as per previous discussion, there are positive loadings over the north-east Atlantic to western Europe and west-central Asia. Leading mode is depicting a wave train originated from the core region and approaching the west-central Asia via western Siberian plains. This shows that atmospheric circulations over both the regions are in phase. The 2nd, 3rd and 4th modes have contributed 19, 11 and 10 % of the total variance, respectively.

3.4 Proposed CGT and the associated atmospheric circulations

The CGT may be responsible for climate variability like rainfall anomaly in the mid-latitude regions like

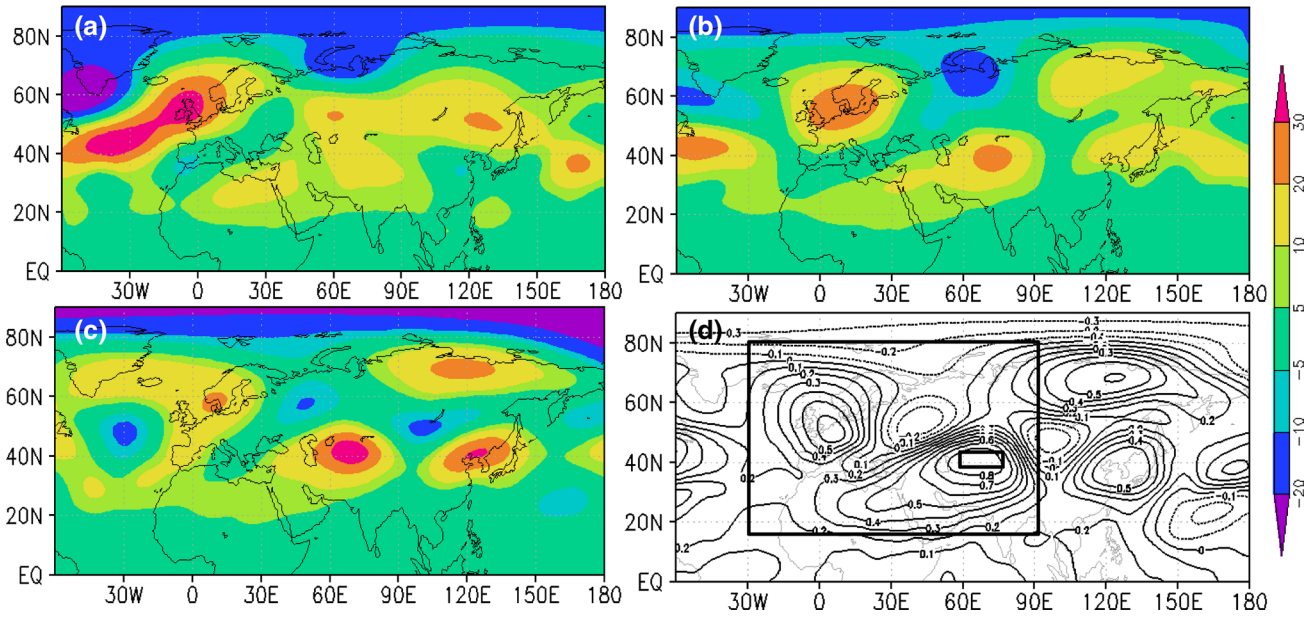

Fig. 13 Detrended geopotential height at 200 hPa for June (**a**), JJ (**b**) and JA (**c**), during strong monsoon years over the study area, for 1960–2014. Point correlation map (**d**) for hgt 200 hPa during JA for a specified region west central Asia (indicated with *small rectangle*) and global hgt at 200 hPa during JJA, area of interest is highlighted with large *rectangle*

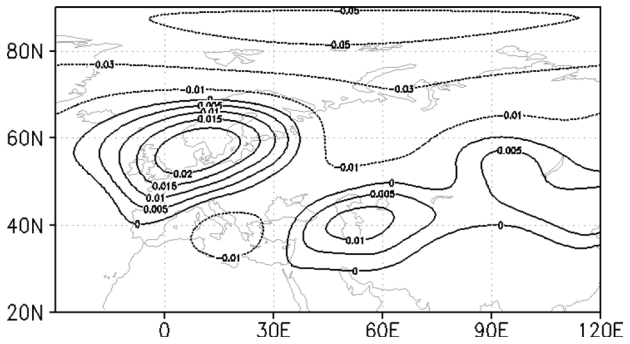

Fig. 14 Leading EOF mode for HGT at 200 hPa for JJA (Variance contributed 28 %)

western Europe, European Russia, India, east Asia, and North America (Ding and Wang 2005). To see the association of proposed CGT with the summer rainfall over the study area, HGT at 200 hpa for JJA have been employed. In Fig. 15a, b the atmospheric field is averaged and plotted for Weak/Strong monsoon years. It is evident that during the Strong monsoon years the field is very much following the same pattern as by proposed CGT (Fig. 15c) while the case is merely different for Weak monsoons. It depicts that the circulation is in phase with the proposed CGT. In the same way, 200 hpa zonal wind has been plotted along with HGT200 for JJA, for the whole study period (Fig. 15d). This figure favors the mutual correlation between the jet stream and HGT at 200 hpa, especially over the area between Atlantic and Pacific.

3.5 Rossby wave's based possible mechanism

To find some alternate reason for our predictors reliablility on physical grounds, for the same period JJA behavior of Stationary Rossby waves is investigated. Müller and Ambrizzi 2007; Naoe et al. 1997 stated that, at middle and high latitudes the zonally oritented winds have larger magniture than the meridional ones, here the westerly jets dominates the waves propogation. But the inclussion of meridionally oriented winds in the wave propogation along the jets may be helpful in demonstrating teleconnections (Li et al. 2015; Zhao et al. 2015). Figure 16 shows that apparently there are two distinct paths of waves originated through the northeast Atlantic/western Europe; the one that is flowing equator wards after origination may be considered responsible for transporting energy up to the positive circulation over west-central Asia. This wave track is practically supported by the wave guide effect of the subtropical jet stream (Hoskins and Ambrizzi 1993), in which a zonally confined wave with a limited meridional dimension (that may be trapped in the jet core) can be transported a long way in the downstream. The other wave branch is moving poleward (approaching ~70°N) and then refracting a little distance towards the equator. These wave ray tracings are quite consistent with the observational statistical results (Fig. 12f), indicating that the predictor in the northeast Atlantic/western Europe is physical based.

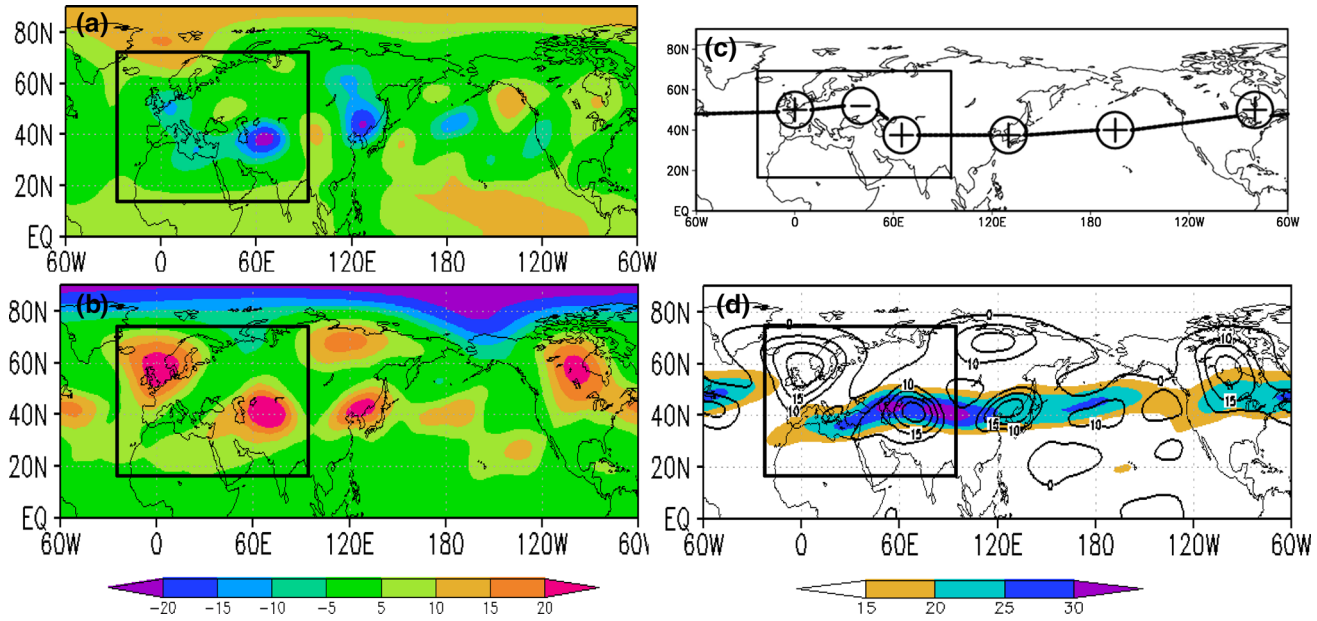

Fig. 15 HGT200 hpa (JJA) for study period during dry conditions of Summer Monsoon (**a**), wet period (**b**), (**c**) proposed CGT (Ding and Wang 2005) and (**d**) 200 hpa jet stream (zonal wind, shaded) and H200 (contours) for JJA during wet conditions, *large squares* show the areas of interest from proposed CGT

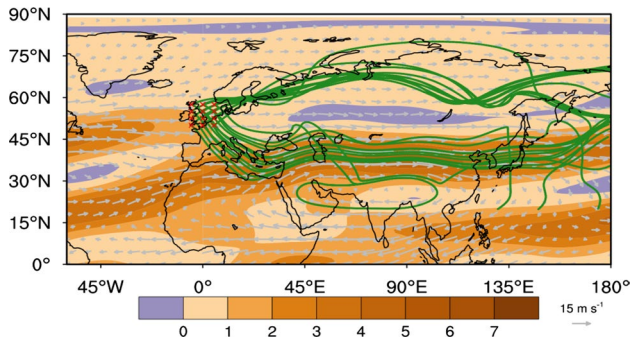

Fig. 16 The stationary Rossby wave rays (green curves) with initialed zonal wavenumber k = 3 in the JJA 200-hPa climatological flows (*gray vectors*), for the study period 1960–2014. *Red forks* denote wave source. The *color shadings* denote the meridional gradient of absolute vorticity in the basic state

Rossby waves originated from the Western Europe are supported by the basic state and are responsible for transporting wave energy from high latitude to west-central Asia (Ding and Wang 2005). Although, this proposed mechanism seems a little bit different from the proposed CGT (Fig. 15c) for the area between northeast Atlantic and the west-central Asia but the associated circulations over both the regions are well in phase. Moreover, our proposed mechanism may be more appealing based on a practical favor by the Rossby wave's trajectory (as the Model's predictors) in the specified season.

4 Summary

We started our work from searching reliable predictors for summer monsoon over Pakistan, primarily from the ENSO or NAO based climate indices. Secondly, more predictors have also been incorporated from the field data for different atmospheric and oceanic elements. Besides, CRU and GPCC data for JA rainfall is also taken to verify the quality as well as reliability of observed data. In the final stages only four grid points from the atmospheric circulations of May to June period came out with better results from all the candidate predictors. Out of which only $U2J_{NEA}$ is selected as the robust predictor with considerable results for IAM, whereas the next possible candidate $H2J_{NEA}$ is taken as the indirect predictor for being a source of energy transportation from point of origin (northeast Atlantic/western Europe) to the study area. Besides training period, the model performed well for the independent test period as well with correlation coefficient of 0.71, with considerable value of RMSE. Prior to physical verification, cross-checking of the results is applied in terms of data reliability and the robustness of the predictor. Besides NCEP, predictor's data from another recommended source JRA-55 is utilized to rerun the model. To test the robustness of the predictor non-overlapping K-folds approach has been adopted, as practiced by Gutiérrez et al. (2013). Accordingly, the model 'IAM' has been successfully cross-validated and put forward for physical significance.

Main features of our findings are given in the following lines.

Recent research in the study area (e.g., Ding and Wang 2005; Syed et al. 2011) concludes that the Eurasian wave train triggers through the anomalous high (200 hpa) over the northeastern Atlantic, which moves down (equator wards) in a curved path through the west Siberian plains to enhance the west-central Asian high. Consequently, this high help creates a strong convection/cloudiness over Pakistan and northeastern India. All that may primarily be ignited by the instability over the upper tropospheric northeastern Atlantic region, where (at the jet stream exit) the zonal flow has dominating role. Observed analysis also confirms this possible mechanism that modulates the summer rainfall in the study region. Moreover, EOF analysis depicts that the upper-tropospheric circulations are in phase over the jet exit region and the west-central Asia during summer season, starting from the time of prediction (June) up to the core rainfall season (JA) over the study area. Observational analysis also shows that how jet stream drives the upper-tropospheric circulation during JJA (Fig. 15d). Both have established very significant association between each other (mainly during June) and with the JA rainfall over Pakistan.

On the other hand, through stationary Rossby waves analysis it has been revealed that jet flow during JJA is mainly responsible for establishing the possible mechanism behind variability of specified rainfall. It is explored that after originating from the jet stream exit region through large vorticity gradients, zonally dominating waves may be responsible for transporting energy to the downstream areas. Li et al. (2015) have stated that, Rossby wave's propagation is very important for transportation of energy and momentum that may cause variations in the atmospheric circulations. Accordingly, $U2J_{NEA}$ and $H2J_{NEA}$ may help enhance $H2JA_{WCA}$ and ultimately affects the rainfall under study. Moreover, based on significance of ENSO over the rainfall in the study area, by applying few modifications and the inclusion of relevant circulations/indices may help improve the performance of IAM in future.

Acknowledgments This work was jointly sponsored by the National Natural Science Foundation of China (Grants 41530424, 41205034, and 41030961). The authors are grateful to the anonymous reviewers for their helpful comments and suggestions in improving the quality and scope of this study.

References

Christensen JH, Hewitson B, Busuioc A et al (2007) Regional climate projections. Climate Change (2007) The physical science basis, contribution of working group I to the fourth assessment report of the intergovernmental panel on climate Change. Cambridge University Press: Cambridge

Ding T, Ke Z (2013) A comparison of statistical approaches for seasonal precipitation prediction in Pakistan. Weather Forecast. doi:10.1175/WAF-D-12-00112.1

Ding Q, Wang B (2005) Circumglobal teleconnection in the northern hemisphere summer. J Clim 18:3483–3505

Ding Q, Wang B (2007) Intraseasonal teleconnection between the summer Eurasian wave train and Indian monsoon. J Clim 20:3551–3767

Frias MD, Zorita E, Fernandez J, Rodriguez-Puebla C (2006) Testing statistical downscaling methods in simulated climates. Geophys Res Lett. doi:10.1029/2006GL027453

Guo Y, Li JP, Li Y (2012) A time-scale decomposition approach to statistically downscale summer rainfall over North China. J Clim 25:572–591

Guo Y, Li JP, Li Y (2014) Seasonal forecasting of north China summer rainfall using a statistical downscaling model. J Appl Meteorol Clim 53:1739–1749. doi:10.1175/JAMC-D-13-0207.1

Gutiérrez JM, San-Martín D, Brands S, Manzanas R, Herrera S (2013) Reassessing statistical downscaling techniques for their robust application under climate change conditions. J Clim 26(1):171–188. doi:10.1175/JCLI-D-11-00687.1

Hoskins BJ, Ambrizzi T (1993) Rossby wave propagation on a realistic longitudinally varying flow. J Atmos Sci 50:1661–1671

Kazmi DH, Rasul G, Li JP, Cheema SB (2014) Comparative study for ECHAM5 and SDSM in downscaling temperature for a geo-climatically diversified region, Pakistan. Appl Math. doi:10.4236/am.2014.51016

Kazmi DH, Li JP, Rasul G, Tong J, Ali G, Cheema SB, Liu LL, Gemmer M, Fischer T (2015) Statistical downscaling and future scenario generation of temperatures for Pakistan region. Theor Appl Climatol 120:341–350. doi:10.1007/s00704-014-1176-1

Li Y, Li JP (2012) Propagation of planetary waves in the horizonal non-uniform basic flow (in Chinese). Chinese Journal of Geophysics 55:361–371

Li JP, Zhang L (2009) Wind onset and withdrawal of Asian summer monsoon and their simulated performance in AMIP models. Clim Dyn 32(7–8):935–968. doi:10.1007/s00382-008-0465-8

Li JP, Ren R, Qi Y, Wang F, Lu R, Zhang P, Jiang Z, Duan W, Yu F, Yang Y (2013) Progress in air-land-sea interactions in Asia and their role in global and Asian climate change. Chin J Atmos Sci 37(2):518–538. doi:10.3878/j.issn.1006-9895.2012.12322

Li Y, Li JP, Jin FF, Zhao S (2015) Interhemispheric propagation of stationary rossby waves in a horizontally nonuniform background flow. J Atmos Sci 72:3233–3256. doi:10.1175/JAS-D-14-0239.1

Lighthill J (1978) Waves in fluids. Cambridge University Press, London

Manzanas R, Amekudzi LK, Preko K, Herrera S, Gutiérrez JM (2014) Precipitation variability and trends in Ghana: an intercomparison of observational and reanalysis products. Clim Change 124(4):805–819. doi:10.1007/s10584-014-1100-9

Manzanas R, Brands S, San-Martín D, Lucero A, Limbo C, Gutiérrez JM (2015) Statistical downscaling in the tropics can be sensitive to reanalysis choice: a case study for precipitation in the Philippines. J Clim 28(10):4171–4184. doi:10.1175/JCLI-D-14-00331.1

Müller G, Ambrizzi T (2007) Teleconnection patterns and Rossby wave propagation associated to generalized frosts over southern South America. Clim Dyn 29:633–645. doi:10.1007/s00382-007-0253-x

Naoe H, Matsuda Y, Nakamura H (1997) Rossby wave propagation in idealized and realistic zonally varying flows. J Meteorol Soc Jpn 75:687–700

Rasul G, Kazmi DH (2011) Climate change and challenges to crop production. In: Proceedings of international workshop on plant conservation and reversing desertification. A way Forward, PMAS Arid Agriculture University, Rawalpindi, pp 81–99

Rasul G, Kazmi DH (2013) Wheat yield and climate change in potohar region of Pakistan. LAP LAMBERT Academic Publishing, Germany

Rasul G, Chaudhry QZ, Zhao SX, Zeng QC (2004) A diagnostic study of record heavy rainfall in twin cities Islamabad-Rawalpindi. Adv Atmos Sci 21(6):976–988

Schmidli J, Goodess CM, Frei C, Haylock MR, Hundecha Y, Ribalaygua J, Schmith T (2007) Statistical and dynamical downscaling of precipitation: an evaluation and comparison of scenarios for the European Alps. J Geophys Res 112:D04105. doi:10.1029/2005JD007026

Simmons AJ, Wallace JM, Branstator GW (1983) Barotropic wave propagation and instability, and atmospheric teleconnection patterns. J Atmos Sci 40:1363–1392

Suleman S, Wood MK, Shah BH, Murray L (1995) Development of a rainwater harvesting system for increasing soil moisture in arid rangelands of Pakistan. J Arid Environ 31:471–481

Syed FS, Yoo JH, Körnich H, Kucharski F (2010) Are intraseasonal summer rainfall events micro monsoon-onsets over the western edge of the South-Asian monsoon? Atmos Res 98:341–346

Syed FS, Yoo JH, Körnich H, Kucharski F (2011) Extratropical influences on the inter-annual variability of South-Asian monsoon. Clim Dyn. doi:10.1007/s00382-011-1059-4

Wang B et al (2006) The Asian monsoon. Paxis Publishing Ltd, Chichester. ISBN 3-540-40610-7

Webster PJ, Toma VE, Kim HM (2011) Were the 2010 Pakistan floods predictable? Geophys Res Lett 38:L04806. doi:10.1029/2010GL046346

Whitham G (1960) A note on group velocity. J Fluid Mech 9:347–352. doi:10.1017/S0022112060001158

Wilby RL (1997) Non-stationarity in daily precipitation series: implications for GCM downscaling using atmospheric circulation indices. Int J Climatol 17:439–454

Zhang RH (1999) The role of Indian summer monsoon water vapor transportation on the summer rainfall anomalies in the northern part of China during the El Niño mature phase (in Chinese). Plateau Meteorol 18:567–574

Zhao S, Li JP, Li Y (2015) Dynamics of an interhemispheric teleconnection across the critical latitude through a southerly duct during boreal winter. J Clim 28:7437–7456. doi:10.1175/JCLI-D-14-00425.1