Principal Component Analysis of Summer Rainfall and Outgoing Long-Wave Radiation over Pakistan

Muhammad Athar Haroon and Ghulam Rasul

Absract

The relationship between the inter-annual variability of precipitation and satellite derived Outgoing Long-wave Radiations (OLR) is studied over Pakistan during summer months for the period 1980-2007. Principal component analysis technique was applied in order to identify the major modes of oscillations present in the data series. The first four principal components explain 95% of the total variance for the OLR data while first eight principal components for precipitation data explain 80% of the total variance. The relationship between major principal mode for precipitation and OLR has remained around -0.78 for the summer season. The possible reason for this inverse relationship could be the overcast skies with clouds and frequent precipitation during the pre-monsoon, monsoon and post-monsoon season which might have contributed to lower OLR values.

Key Words: Principal Component Analysis, Outgoing Long Wave Radiation (OLR), Precipitation,

Introduction

Summer precipitation in South Asia is associated with monsoon circulation and convective activity which experiences a significant variability over the spatial and temporal scale. In order to investigate the inter-annual variability of summer rainfall and Outgoing Long-wave Radiation (OLR), the data was analyzed by using Principal Component Analysis (PCA) / Empirical Orthogonal Function Analysis (EOF). This method was developed by Pearson (1901) & Hotelling (1933). PCA is a way of identifying the anomalous patterns in the data. The other main advantage of PCA is that once these patterns are found in the data sets, they can be compressed by reducing the number of dimensions without much loss of information. The Eigen Vector with the highest Eigen value is the first Principal Component of the data set. Since the new variables are derived from the original variables, they are called components. Principal components are a sequence of uncorrelated linear combinations of the original data sets, each with a variance smaller than the previous one, that collectively preserve the total variation of the original data. PCA decides which, amongst all the possible projections, are the best for representing the structure of data. The projections are chosen so that the maximum amount of information, measured in terms of its variability is retained in the smallest number of dimensions.

A variety of statistical and analytical techniques have been made on rainfall data by several investigators under different objectives. Iyengar (1991) applied PCA to understand variability of rainfall over India. Liu & Yin (2000) used PCA in order to find teleconnections between the North Atlantic Oscillation (NAO) and summer precipitation over Eastern Tibetan Plateau. Singh (2004 & 2006) compared the dominant rainfall and OLR patterns in order to identify spatial and temporal variability. Based on PCA, Bansod (2004) incorporated composite cycle of the OLR to predict summer monsoon rainfall over Indian subcontinent. Diagnoses of OLR and precipitation data sets in relation to spatial and temporal variability were investigated by Haque & Lal, 1991 and Murakami, 1980.

Data and Analysis

Monthly Precipitation data for summer season (June to September) for the period 1980 to 2007 over Pakistan for fifty stations have been used in this study. This data have been taken from Pakistan Meteorological Department (PMD). Monthly values are averaged over the summer season in order to use in this study. Also outgoing long wave radiation (OLR) data set for the period (1980-2007) on monthly basis have been taken at a 2.5o longitude x 2.5o latitude from NCEP reanalysis archives. The OLR values have been extracted over the same fifty stations by using GRADS software. These monthly values are then averaged over the summer season (June-Sep) which mainly pertain Pakistan's summer monsoon.

For analysis of the data over the whole country, let P be a (r x s) anomaly matrix of OLR data over a series of r = 28 years and s = 50 stations. The elements of the matrix P represent departure of OLR values from their mean values. Principal component analysis is performed using P as the data matrix. It was also desirable to examine the coupling between OLR and Precipitation. Therefore another anomaly matrix of the precipitation data was taken into consideration.

Let P & Q represents the time and space variability of OLR & precipitation, then it can be defined as

$$P=UV \qquad \text{------- (i)}$$

&

$$Q=MN \qquad \text{------- (ii)}$$

Where the matrix U& M represent the time variation and V&N represent the space variation of both the data sets.

If P_{it} is the actual OLR data at stations i (i =1, 2…M) in the year t (t=1, 2…N), the mean value is

$$m_i = (1/N)\sum_{i=1}^{N} P_{it} \qquad \text{------- (iii)}$$

The centered data are

$$d_i = \{P_i - m_i\} \qquad \text{-------- (iv)}$$

The covariance matrix is

$$C_{ij} = (1/N)\sum_{t=1}^{N} d_{it} d_{jt} \qquad \text{------- (v)}$$

The eigen values {λ_j} and the eigen vectors {ϕ_{ij}} of this symmetric matrix are extracted. The principal components are

$$(PC)_{jt} = \sum_{t=1}^{M} d_{it}\phi_{ij};(j = 1, 2...M) \qquad \text{------ (vi)}$$

This transforms the original time series d_{jt} into the new time series $(PC)_{jt}$.First few $(PC)_{jt}$'s are generally sufficient to explain the variation in the original data.

The eigen vectors {ϕ_{ij}} represent spatial patterns while $(PC)_{jt}$ can be used to understand temporal variability in the data.

Results and Discussion

The first principal component for OLR explains 66% of the total variance having positive loadings throughout the country. It is noted from Fig.1 that the area of large positive values is concentrated over Balochistan, Sindh and Punjab while western parts of NWFP, northern areas including Kashmir are showing low OLR values Balochistan is covered with intense seasonal heat-low due to which OLR values are higher. This intense heat-low with little or no clouds would contribute to higher values. Highest value is on Eastern Sindh where Thar Desert is located. Maxima over the eastern parts of the Sindh province share with arid plains of the southern Punjab which comprised sandy desert of Cholistan. The agricultural plains of upper Sindh, bound by Cholistan desert on the north and Thar Desert on the south although behaved slightly in a mild way but have shown similar characteristics.

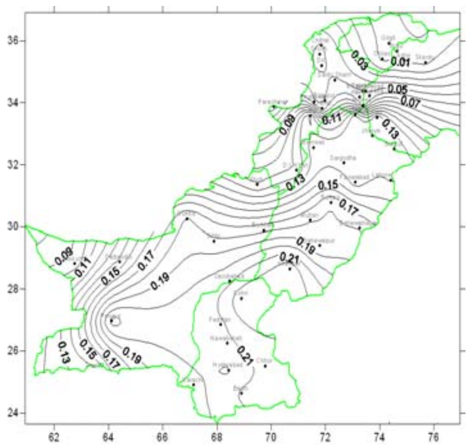

Figure 2: First Principal Component for OLR Pattern (Variance explained 66%)

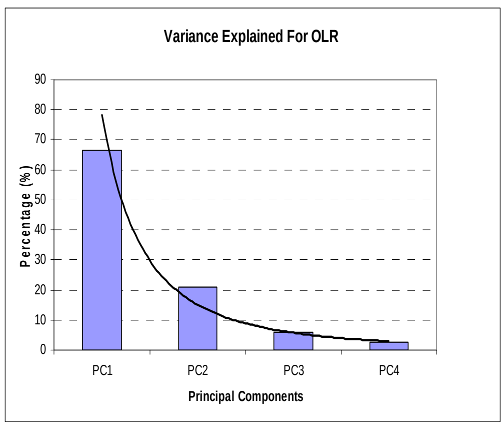

Figure 2: First four principal components explain 95% of the total variance

Northern parts of the country are mountainous, the precipitation frequently occurs as monsoon air mass prevails and presence of clouds contributes to lower values of OLR. Plain areas of Punjab also experience intense heating in summer therefore reflect higher values. First four principal components of OLR data explain 95% of the total variance as shown in the Fig 2.First PC reflects the largest variance and the fourth ends with the smallest value. Since remaining principal components explain only 5% of the total variance so we can neglect them.

Fig.3 shows first principal component for precipitation pattern explaining 32% of the total variance, again having positive loadings throughout the country. This PC is considered as the dominant pattern of summer precipitation. It has been observed that for precipitation, the maxima exist over the northeastern and extreme southeastern parts of Pakistan. These large values are mainly due to south westerly monsoon currents from Arabian Sea and northeasterly and easterly currents along Himalayas originating from Bay of Bengal.

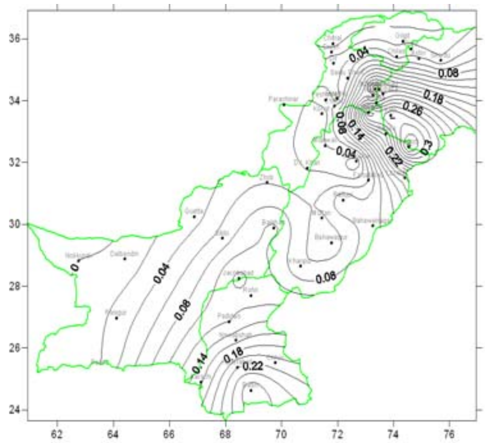

Figure 4: First Principal Component for Precipitation Pattern (Variance explained 32%)

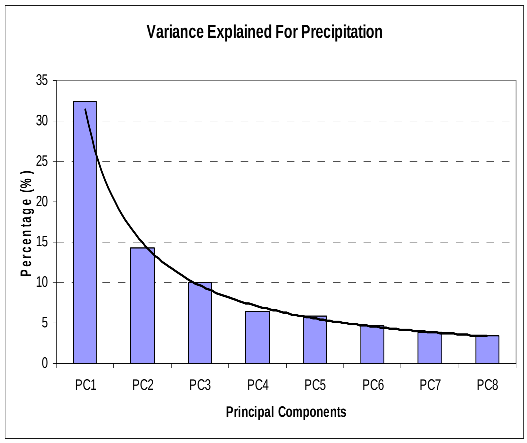

Figure 4: First eight principal components explain 80% of the total variance

Northern Punjab, north eastern NWFP and Kashmir having higher values due to the persistence of monsoon precipitation in these areas, where maximum amount of total rain occurs during summer. First eight components of the precipitation data contribute 80% of the total variance as shown in the Fig.4.Since remaining PC's are contributing only 20 % of the total variance therefore can be neglected. In this respect we have reduced the original data up to the eight principal components without losing much information. It is interesting to observe that not more than eight components are required to represent the rainfall over the whole country.

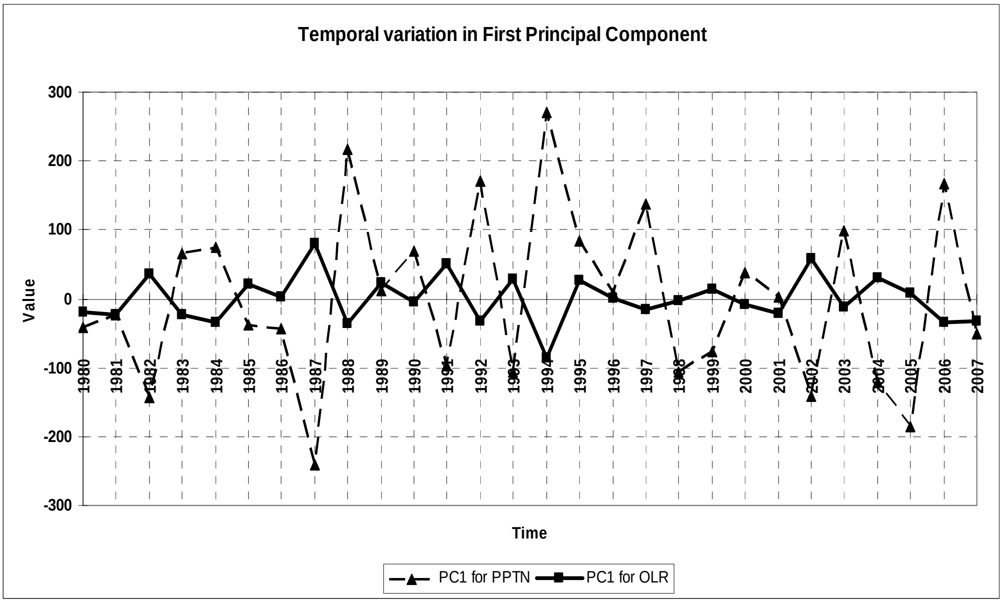

Figure 5: Temporal Variation in the first principal component for OLR (Solid Line) and precipitation (dashed line) from 1980-2007

A strong negative correlation having value -0.78 between first principal component of OLR and Precipitation has been observed. This result indicates a strong inverse relationship mainly due to the fact that OLR values remain low when there is a cloud cover. Temporal variability in the first principal component of both the data sets also revealed this result as shown in the Fig.5.Inverse relationship for 1994 is the best example of the wettest year when country's summer precipitation exceeded about 40% above normal. Another extreme is the prolonged drought in Pakistan which was triggered by the history's strongest El Nino (1997-98).It persisted for four consecutive years setting the new record as the worst drought ever occurred in last 100 years in Pakistan. Figure.5 reflects the dominant inverse correlation between OLR and precipitation.

Conclusions

Using principal component analysis, we identified the major modes of oscillation present in OLR data during the summer season and also interannual variability of summer precipitation over Pakistan. This suggests that OLR has higher spatial coherence than precipitation. When the temporal variation of the leading precipitation anomaly pattern is correlated with major OLR pattern, a strong negative correlation is found. This inverse relationship is an indication that presence of clouds and precipitation are responsible for lower OLR values.

References

Bansod.S.D, 2004. “Outgoing long-wave radiation over the Tropical Pacific and Atlantic Ocean and Indian summer monsoon rainfall”. Theor.Appl.Climatol. 77, 185-193.

Haque.M.A, M.Lal, 1991. “Space and time variability analyses of the Indian monsoon rainfall as inferred from satellite-derived OLR data”. Clim.Res. 1, 187-197.

Hotelling, H, 1933. “The most predictable criterion”. J.Educ.Psychol.26, 139-142.

Iyengar.R.N, 1991. “Application of principal component analysis to understand variability of rainfall”. Earth Planet.Sci.100, 105-126.

Liu.X and Zhi-Yong.Yin, 2000. “Spatial and temporal variation of summer precipitation over the eastern Tibetan Plateau and the North Atlantic Oscillation”. J.Climate.14, 2896-2909.

Murakami.T, 1980. “Empirical Orthogonal Function Analysis of Satellite-Observed Outgoing Long wave Radiation during summer”. Mon.Wea.Rev.108, 205-222.

Pearson.K, 1901. “On lines and planes of closest fit to systems of points in space”. Phil.Mag.B,2.559-572.

Singh.C.V, 2004. “Empirical Orthogonal Function (EOF) analysis of monsoon rainfall and satellite-observed outgoing long-wave radiation for Indian monsoon: a comparative study”. Meteorol Atmos Phys 85,227-234.

Singh.C.V, 2006. “Principal component analysis of satellite-observed outgoing long-wave radiation during the monsoon period (June-September) over India”. Theor.Appl.Climatol.84, 207-211.