Principal Component and Clustering Analyses for Seasonal Classification of Karachi

Naeem Sadiq[1]

Abstract

Queer and vigorous features regarding the weather of Karachi inspire to undertake the seasonal classification of it. There are several methods formulated for this purpose, usually consist of two stages. In the first stage, atmospheric circulation pattern are classified into distinct types and in the next, the local weather elements are related to them. Obviously, a different approach is that one avoids the stage of circulations type and attempts to classify weather itself according to measured surface variables. In this research clustering analysis is employed to the diurnal data of several climate parameters of the cosmopolitan city. Calculations show that Karachi has six different seasons including double summer in a year. Further the meteorological characteristics of these seasons are discussed in the paper. The aim of this research is to offer a preliminary discussion on this subject and has been projected in the climatic conditions obtainable under different heads, that is, the factors and elements of the climate of Karachi.

Key words: Seasonal classification. Cluster analysis, Double summer

Introduction

Human activities during the last century – particularly the burning of fossil fuels – have changed the composition of the atmosphere in ways that threaten to dramatically alter the global climate in the years to come (WMO, 2006). Global warming is caused by the increasing greenhouse effect, a natural phenomenon in which gases in the earth's atmosphere, including water vapor and carbon dioxide, trap heat emitted from the planet's surface (Money, 1992). Without a natural greenhouse effect, temperatures on earth would be too low for life to survive (Michael et. al., 1996).

Karachi is one of the major cities in the world and its temperature is increasing due to green house effect (Sadiq & Qureshi, 2010). It is also important in the sense that the region has a significantly long coast. Moreover, the climate here is rather unusual in as much as it has two summers every year (Shamshad, 1988). This asks for the study of its seasons and weather classification.

The seasonal classification of any region is the problem of interest for several decades (Huth, 2001; Iqbal, 2005,). In this regards several methods have been developed over many years to identify seasons. Manual procedures involve the subjective classification of circulation patterns for which the examples of Muller classification (Muller, 1977), the

Lamb catalogue (Lamb, 1972), and Grosswetterlagen (Hess and Brezowsky, 1977) are well known, but they are quite laborious and time consuming.

Christensen and Bryson, 1966 were among the first not to use this approach to circulation type but instead attempted to classify the weather itself according to measured surface variables (Huth et al., 1993). In the recent decades computer revolution has resulted in the development of many automated synoptic classification methods (Malmgren and Winter, 1999). Once some initial thresholds are set, then with the help of computer, classification groups and assign individual cases entirely based on statistical criteria (Sheridan, 2002).

As for seasons of Pakistan, the typical European or middle latitude seasons, namely winter, spring, summer and autumn can be found only in the extreme north of the country or in northeast Balochistan (Shamshad, 1988). However, Karachi lies in subtropical double season climate zones and climate of Karachi is unusual in that it has two summers every year (Shamshad, 1988). Karachi is chiefly sub-tropical, arid and yet maritime. Although the Arabian Sea constantly washes its shore, yet the city is

[1] naeemimg@yahoo.com, Pakistan Meteorological Department

deficient in its own water supply. The climate of Karachi is affected by several factors and since the city is now expanding on all the sides and new suburbs, satellite towns, etc., are coming into being, it is necessary to investigate the unusual pattern of season of cosmopolitan city.

Section 2 deals with data reduction problem. Classification of season of Karachi performed through clustering in Section 3. Section 4 provides the characteristics of different seasons found in Karachi. The last section comprises summary and conclusion.

Data Reduction

Diurnal data of precipitation (ppt), Temperature {both maximum (T_x) and minimum (T_n)}, Relative humidity (Rh), cloud amount (Cl), and wind speed (W_s) for the period of 1981-2010 was employed. All of the data used in this study have been provided by Pakistan Meteorological Department and all parameters are in their standard units (i.e. precipitation amount in mm, temperatures in degrees Celsius, relative humidity in percent, amount of cloud in Oktas and wind speed in Knots).

Table 1: Variables Mean and St. Dev. (1981-2010)

Variable	Mean	Standard Deviation
Seasonal Rain (mm)	6.14	15.73
Relative Humidity (%)	47.70	14.08
Cloudiness (Oktas)	2.64	1.66
Wind Speed (Knots)	7.38	2.15
Max Temperature (°C)	32.32	3.06
Min Temperature (°C)	21.23	5.81

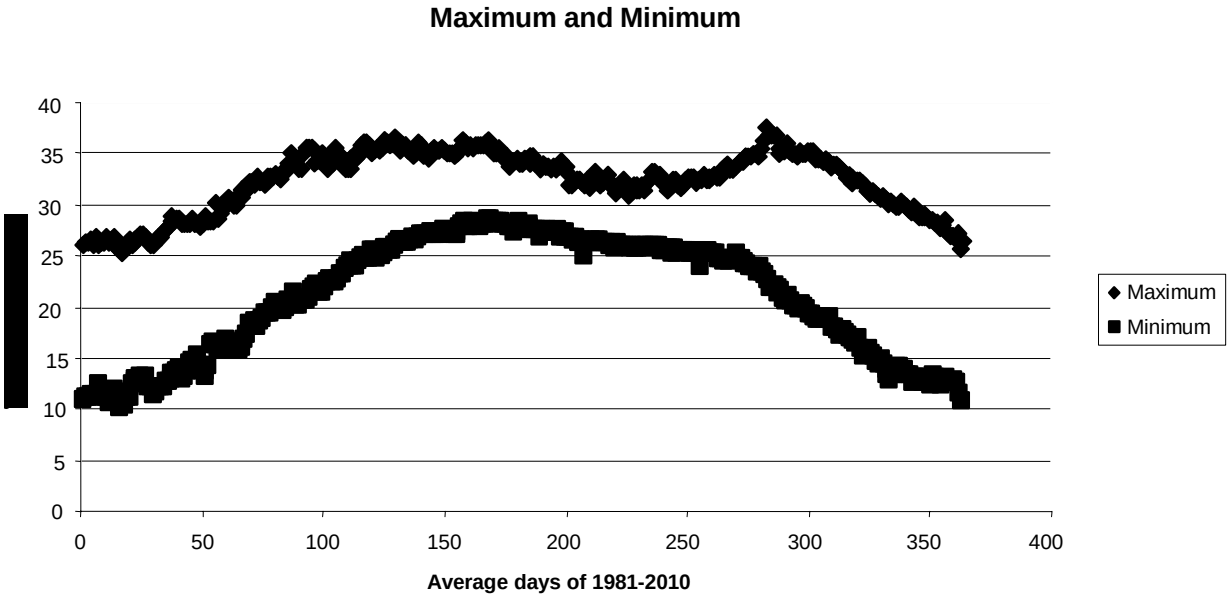

Figure 1: Diurnal distribution of Maximum and Minimum Temperature

To characterize the relation, the means and standard deviations of all the variables are calculated and are given in Table1. The period from 1981 to 2010 is characterized by somewhat higher Minimum temperatures deviation than maximum (Table 5 & Figure 2).

To make classification of season, firstly daily mean of each of six meteorological parameters, mentioned above, computed. Thus there were 365 data point of each of six parameters (i.e. data set consists of 6 variables and 365 days).

Table 2: Correlations (Pearson) for (1981-2010)

	Rh	**T_x**	**W_s**	**Cl**	**ppt**
T_x	0.48				
W_s	0.81	0.54			
Cl	0.77	0.05	0.60		
ppt	0.33	-0.05	0.18	0.48	
T_n	0.88	0.81	0.84	0.54	0.18

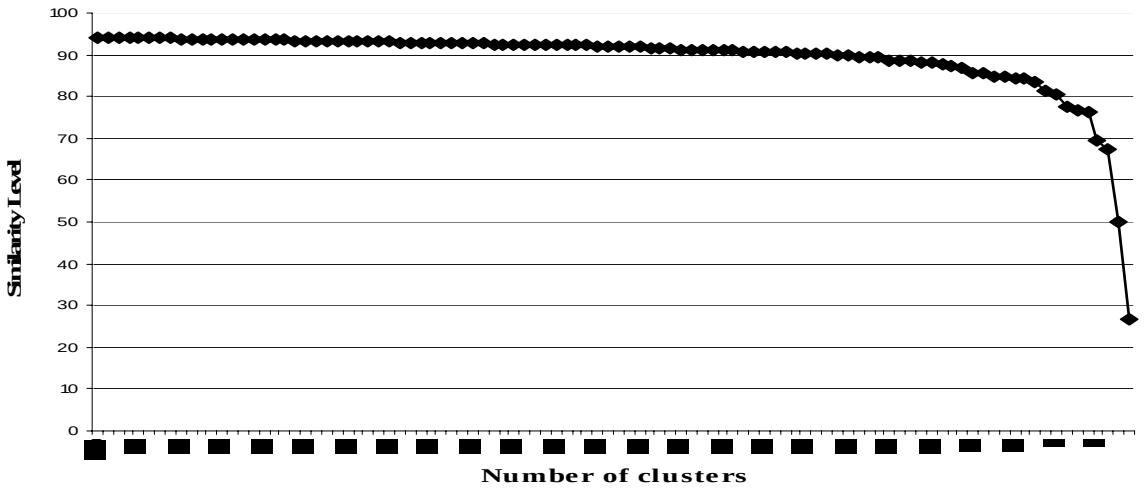

Figure 2: Plot of Similarity level against the number of clusters

To reduce the dimensionality of considered data of six parameters, the principles of PCA (principal components analysis) are invoken. PCA approach is a data reduction technique, which can be used to identify a small set of variables, which account for a large proportion of the total variance in the original variables (Sharma, 1996; Malmgren and Winter, 1999). To check the suitability Pearson's correlations were performed. As the correlations between six parameters are high (Table 2), they seem to be sufficiently correlated. This suggests an application of the technique of varimax rotation to compute the involved components. The procedure reduces the six variables to six independent principal components. The eigenvalues suggest that there are four statistically significant principal components because only the first four PCs have eigenvalues that are greater than one altogether accounting for about 94% of the total variance in the diurnal data (Table 3).

Table 3: Rotated PC Loadings and Communalities (varimax rotation)

Variable	PC1	PC2	PC3	PC4	PC5	PC6
Relative humidity	0.40	0.62	-0.42	-0.17	-0.50	-0.00
Max temperature	0.98	-0.01	-0.19	0.05	-0.03	-0.03
Min temperature	0.75	0.39	-0.44	-0.07	-0.27	0.12
Cloudiness	0.03	0.92	-0.26	-0.28	-0.06	0.01
Precipitation	-0.01	0.21	-0.05	-0.97	-0.04	0.00
Wind Speed	0.39	0.37	-0.83	-0.06	-0.11	0.00
Variance	1.84	1.57	1.17	1.07	0.34	0.02
% Variance	0.31	0.26	0.19	0.18	0.06	0.00

The relationship between each PC and six weather variables was then assessed from the component loadings (Table 3). Overall, such an influence on a PC seems to be significant. More specifically, from the correlation matrix between the given dataset and the rotated PCs and the coefficients of PCs, It is inferred that maximum temperature (T_x) dominates in the first rotated PC with the highest coefficient, 0.98. Similarly, PC2, PC3, PC4, PC5 and PC6 appear to represent Cl, W_s, Rh, ppt and T_n with correlation 0.92, -0.83, -0.97, -0.50 and 0.12 respectively (Tables 3).

Cluster Analysis

In order to sort out the seasons of Karachi, principal component scores by applying average-linkage method of hierarchical clustering analysis were clustered. In average-linkage method, it is, simply, the maximum Euclidean distance (MAXD) between two clusters. Figure 2 gives the plot of series of clusters against similarity level. Obviously, MAXD should be small to merge two clusters.

A large value for MAXD would indicate that two dissimilar groups are being merged. From the figure 2, it is clear that there is a big change in value of the distance when going from a five-cluster to a four-cluster solution. Consequently, it appears that seasons of Karachi can be divided into five (summer appears twice) different categories.

Table 4: Clusters for 6 Seasons

#	Period of Month	Season
1	Mid of Dec to Feb	Winter
2	March to Mid April	Transitional (Spring)
3	Mid of April to June	Summer (first)
4	July to Initial Sep	Monsoon
5	September to October	Summer (second)
6	Nov to Mid of Dec	Transitional (Autumn)

The next problem is a grouping of observations into a pre-defined number of clusters. For this, K-means clustering on the four PC scores are applied. Usually, as the result of a K-means clustering analysis, one would examine the means for each cluster on each dimension to assess how distinct the K clusters happen to be (Sharma, 1996). This helps to differentiate members from 365 days into six seasons. Thus first cluster is classified from December to February as winter season. Similarly the second (March to Mid

April), third (April to End of June), fourth (End of June to Initial September), fifth (September to Mid of October) and sixth's season (Mid of October to November) are classified as spring, first summer, monsoon, second summer and autumn respectively (Table 4). Third and fifth cluster shows that there are two summers in a year, first summer begins from April to June (Pre monsoon summer) and second summer occurs from September to Mid of October (post monsoon summer).

Characteristics of Seasons

Finally, mean values accompanied by standard deviation, maximum and minimum of each meteorological variable are computed separately and tabulated in Table 5. Meteorological characteristics of six seasons found above are discussed in the following subsequent sections.

Table 5: Some Statistical Characteristics of Six Seasons (1981-2010)

Parameter	Mean	St. Dev.	Min	Max	Mean	St. Dev.	Min	Max
	Winter Season				**Spring (Transitional 1) Season**			
W	5.48	1.16	3.25	9.12	7.68	0.94	5.56	9.62
Cl	1.94	0.46	0.94	3.06	1.90	0.57	1.00	3.31
ppt	5.62	13.13	0.00	100.20	0.90	2.99	0.00	18.00
T_n	12.95	1.42	10.36	16.67	19.89	2.21	16.07	23.51
T_x	27.35	1.10	25.29	30.09	33.03	1.73	29.59	35.64
Rh	34.56	4.36	25.06	49.31	37.33	5.06	28.25	50.12
	First Summer Season				**Monsoon Season**			
W	9.56	1.12	7.44	11.50	9.36	0.79	7.56	11.00
Cl	2.44	1.05	0.87	5.00	5.56	0.69	3.94	7.12
ppt	1.41	4.87	0.00	30.00	21.57	27.13	0.00	122.60
T_n	26.92	1.44	23.36	28.80	26.81	0.79	25.32	28.52
T_x	35.36	0.69	33.45	36.57	32.77	1.00	31.06	34.96
Rh	57.64	4.72	48.69	65.37	66.76	2.42	60.44	72.56
	Second Summer Season				**Autumn (Transitional 2) Season**			
W	6.73	1.39	4.62	9.37	4.52	0.70	3.06	6.19
Cl	1.97	1.29	0.31	4.75	1.28	0.47	0.37	2.06
ppt	3.15	8.21	0.00	33.80	0.32	1.25	0.00	7.30
T_n	23.65	2.02	19.62	25.81	16.03	2.19	12.80	19.42
T_x	34.40	1.46	32.06	37.80	31.80	1.84	28.88	34.75
Rh	50.04	11.31	30.37	66.62	33.32	3.41	27.31	40.06

Winter Season

The season occurs during December to February is known as winter season. It is the cold season because minimum temperature reaches its minimum value (10.36 °C) in this season. Due to Karachi's coastal location, the sea has a marked effect on it and so the winters are not so cool as at other upcountry places (northern areas reaching to negative temperatures) in Pakistan. However, four to five cold waves may reach the area during the winter season and out of these one or two may prove to be quite cold and make Karachi temporarily uncomfortable. Temperatures, however, under the influence of these cold waves, never approach the freezing point. This season is less humid (34.56%)

and its least average value found is 25.06%, fewer cloudy (less than 2 oktas) but slightly rainy (total 26.34 mm) due to the western disturbance. Slightly windy behavior (above 5 knots) causes the land breeze in the morning sometimes dry and chilly. Average maximum temperature does not reach upto such higher values (27.35 °C) as in other seasons.

Spring Season

After the winter, the city have moderate climate during March to mid April. This transitional season is pleasant as the temperature becomes moderate by reduced diurnal range and by increasing mean maximum (33.03 °C) & minimum (19.89 °C) temperatures, for the city. The mean increases rapidly for this period. Due to fast increase of greenery and some other factors (although not too many), this transitional season is recognized as spring. This season is windy (above 7 knots) and less humid (37.33%). Nominal rainfall (total 2.75 mm for the whole season) occur in this season, consequently, nominal clouds (less than 2 oktas) are there.

First Summer Season

After mid April, weather becomes hot. May, hotter than June is, generally the hottest month (as maximum temperature reaches upto 35.36 °C). The maximum value for maximum temperature is found 36.57 °C in this season. Although there is no fast increase in mean and maximum temperatures but minimum temperature increases faster (26.92 °C) comparable to other seasons and hence morning is also hot some times. The temperature amplitude decreases with appreciable amount. Some time, in May and June, when depressions take time in forming over the Arabian sea, a trough of low pressure appears over the North Arabian Sea, and causes sultry weather over the coast by reversing the wind flow. Comparatively relative humidity is greater (57.64%) than the winter and spring season. It is windy (almost 10 knots) and partial cloudy (greater than 2 oktas) season. Nominal rainfall (total 6.36 mm) is also found in this season.

Monsoon Season

By the end of June, monsoon winds produce cloudy weather, which persists till the beginning of September. But these “Stratus or Strato-Cumulus” clouds, as they are called, give no rain though they look quite dark. They, however, keep the Karachi weather pleasant with occasional drizzle at night or during cooler parts of the day. Less temperature tendency and less maximum (32.77 °C) and minimum (26.81 °C) temperatures also maintain this pleasant change.

Humidity is, generally, high (66.76%) during the summer season, but the presence of constant “sea breeze” (above 8 knots) makes one feel quite comfortable in the open air or in well-ventilated rooms. Towards the end of July, monsoon showers expected which give appreciable amount (total 93.04 mm) of rainfall. In fact, July and August are the rainiest and cloudiest (above 5 oktas) months of the year.

Second Summer Season

Generally after the monsoon season, Karachi experiences once again summer. The sky is again cleared of clouds (less than 2 oktas) and the breeze become relatively gentle (above 6 knots) now. The air is humid (50.04%) and some times calm. The weather frequently becomes oppressive and some times sudden sultry as maximum temperature reaches its extreme (maximum of the maximum temperature reaches up to 37.80 °C), while the average maximum (34.40 °C) and minimum (23.65 °C) temperatures are slightly lower than the pre-monsoonal summer. This post-monsoon period, from September to Mid of October, also experiences heat waves, which raise the morning and day temperatures. Beside the marked decreasing values of clouds the amount of rainfall seems greater (total 10.65 mm), comparable to first summer but still it may not be considered as rainy by any means. Actually it is aftermath of the monsoon.

Autumn Season

The period from mid October to November, however, is pleasant with its cool nights (minimum of minimum temperature reaches to 12.80 °C) and moderate mean maximum (31.80 °C) and minimum (16.03 °C) temperatures. Minimum temperature decreases rapidly by large numerical values with negligible amount of clouds (slightly above 1 okta). This season may be considered as the precursor for the incoming winter. Observations of Ecology also indicate the signs of autumn. This season is classified as transitional (autumn); of course, different transitional form earlier described (i.e. transitional/spring). Slightly windy (above 4 knots) less humid (33.32%) and almost no rain (total less than 1 mm) are the other characteristics of this season.

Summary and Conclusion

As demonstrated in Section 1, the problem of seasonal classification for today's populous Karachi is a rather unattended but serious issue because it has unusual weather behavior. As discussed in Section 2, it is important to first sort out the data problem and the methodology to tackle the seasonal problem afresh on a sound basis. In Section 3, the groundwork has been laid down to arrive at a solution of the problem by computing PCA scores of the diurnal data. The four PCs extracted from the matrix of correlation among the six meteorological variables i.e. W_s, Cl, ppt, T_n, T_x and Rh. As carried out in Section 4, a hierarchical clustering analysis, coupled with K-mean analysis, indicates that the seasons of Karachi should be divided into six seasons Our calculation shows that Karachi acquires annual occurrences of two summers: One is the usual summer and other in October and hence we also get the peculiar feature of double summer in Karachi. In the last section, the important meteorological characteristics of the seasons of Karachi have been discussed which may be useful for crops plantation and watering in the suitable seasons. This study is also supportive for aviation because there is a need to be more careful in first summer (gusty winds) and monsoon (gusty winds, cloudy and rainfall). On the basis of this study, it is imperative that the sighting of new moon in Karachi is not suitable during summers and monsoon seasons because of sudden and frequent cloud appearances.

References

Hess, P. and H. Brezowsky, 1977: Katalog der Grobwetterlagen Europas (1881- 1976), Berichte des Deutschen Wetterdienstes 15(113); (Selbstverlag des Deutschen Wetterdienstes, Offenbach am Main, Germany).

Huth, R., I. Nemesova and N. Klimeperova, 1993: Weather categorization based on the average linkage clustering technique: An application to European mid-latitudes, Journal of Climatology, 13, 817-835.

Huth, R., 2001: Disaggregating climatic trends by classification of circulation patterns, Journal of Climatology, 21, 135-153.

Iqbal, M. J., 2005: Detection and modeling of hydro-meteorological variability over Pakistan region, Ph.D. Thesis, University of Karachi.

Lamb, H. H., 1972: British Isles weather types and a register of the daily sequence of circulations patterns, Geophysical Memoris, 116; London.

Malmgren and Winter, 1999: Climatic zonation in Puerto Rico based on principal components analysis and an artificial neural network, Journal of Climate, 12, 977-985.

Micheal, A. J., A. Haines, **R. Sloff and S. Kovats, 1996:** Climatic change and human health, WHO, WMO, UNEP, Geneva.

Money, D. C., 1992: Climate and environment system, Collins Educational, London.

Muller, R. A., 1977: A synoptic climatology for environmental baseline analysis: New Orleans, Journal of Applied Meteorology, 16, 20-33.

Sadiq N. and M. S. Qureshi, 2010: 'Climatic variability and linear trend models for the five major cites of Pakistan', Journal of Geography and Geology., 2 (1), 83-92.

Shamshad, K. M., 1988: The Meteorology of Pakistan, Royal Book Company, Karachi.

Sharma, S., 1996: Applied Multivariate Techniques, John Wiley, USA.

Sheridan, S. C., 2002: The redevelopment of a weather-type classification scheme for North America, Int. J. Climatol., 22, 51-68.

WMO, 2006: WMO statement on status of the global climate in 2005, WMO No 998.