Development of a Weather-Type Classification Scheme for Karachi by using Multivariate Techniques

Naeem Sadiq

Abstract

Weather typing or categorization continues to be popular and numerous methods have been developed over the past century for this investigation. This paper attempts to classify the types of weather for Karachi on the basis of diurnal data according to seasonal classification. The idea of using weather classification in climate change research is inspired by both uncertainties accompanying climate simulation on a regional scale, and the conflicting results of examining long-term instrumental series by traditional statistical methods. Multivariate techniques of Principal component analysis and Cluster analysis are used to obtained different types of weather for each season separately. Calculations show that we have 14, 5, 8, 7 and 8 different weather types for Monsoon, winter, summer, spring and autumn seasons respectively. Noticeable greater variations came into view for monsoon season and winter appears as least varied season. The aim of this research is to investigate the properties of the adopted clustering procedure, the consequences of the modifications introduced, and the physical interpretation of the weather types in terms of meteorological variables.

Key Words: Weather Typing, Principal Component Analysis, Cluster Analysis

Introduction

In general, there was an approach, in which after circulation pattern classification, local weather elements were related to them (Barry and Perry, 1973; Yarnal and White, 1987) while Christensen and Bryson (1966) used only surface meteorological parameters for the same classification and avoid the stage of circulation typing. Old methods of statistics (Houghton et al., 1990) produce inconsistent results of examining long term instrumental sequence and create uncertainty of climate simulations on a regional scale.

Kalkstein et al. (1990) used very different approach in this regard. In his study of air masses he used principle component analysis for the reduction of variable parameters and then apply clustering procedure in order to create weather types i.e. with similar weather, for example Kalkstein and Corrigan (1986) and Stone (1989). The recent increased interest in the procedure is especially for the purpose of understanding possible implications of climate change, has stimulate greater in number and better weather typing schemes (Huth et al., 1993).

Data

To obtain weather types following variables are used in the study on diurnal basis: Precipitation amount (Ppt), Temperature {both maximum (Tx) and minimum (Tn)}, Relative humidity (Rh), cloud amount (Cl), and wind speed (Ws). Data used in the study comprises of 16 years that is for the period of 1990 to 2005 inclusive, which is sufficiently long enough period of time to produce a reliable and representative weather types for Karachi. Thus we have 5840 data points of each of six parameters (i.e. data set consists of 6 variables and 5840 days). All the meteorological data used in this weather type study is provided by Pakistan Meteorological Department, Karachi. 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).

In this paper weather types are obtained and analyzed according to the seasonal classification of Karachi (Sadiq, 2007), which is summarizes in Table 1.

Table 1: Seasonal Classification

#	**Period of Months**	**Seasons**
1	Mid of December to February	Winter
2	March to Mid April	Transitional (Spring)
3	Mid of April to June	Summer (first)
4	July to Initial September	Monsoon
5	September to october	Summer (second)
6	November to Mid of December	Transitional (Autumn)

Principal Component Analysis

Principal components analysis (PCA) is a procedure for finding hypothetical variables (components), which account for as much of the variance in multidimensional data as possible (Davis et al.1986, Harper 1999). These new variables are linear combinations of the original variables (Johnson and Wichern, 1998).

The PCA routine finds the eigenvalues and eigenvectors of the variance-covariance matrix or the correlation matrix. Either choose var-covar if all variables are measured in the same units (e.g. centimeters) or choose correlation (normalized var-covar) if variables are measured in different units. This implies normalizing all variables using division by their standard deviations (Hair et. al., 1998). The eigenvalues, giving a measure of the variance accounted for by the corresponding eigenvectors (components) are given for all components. The percentages of variance accounted for by these components are also given.

Analytical Approach

Assuming that there are p variables and we are interested in forming the following p linear combinations:

$$\begin{aligned} \xi_1 &= w_{11}x_1 + w_{12}x_2 + \dots\dots\dots\dots + w_{1p}x_p \\ \xi_2 &= w_{21}x_1 + w_{22}x_2 + \dots\dots\dots\dots + w_{2p}x_p \\ &: \\ &: \\ \xi_p &= w_{p1}x_1 + w_{p2}x_2 + \dots\dots\dots\dots + w_{pp}x_p \end{aligned} \tag{1}$$

Where ξ_1, $\xi_2,\ldots\xi_p$ are the p principal components and w_{ij} is the weight of the *jth* variable for the *ith* principal component. The weights, w_{ij}, are estimated such that:

i. The first principal component ξ_1, accounts for the maximum variance in the data, the second principal component, ξ_2, accounts for the maximum variance that has not been accounted for by the first principal component, and so on.

ii. $w_{i1}{}^2 + w_{i2}{}^2 + \ldots + w_{ip}{}^2 = 1$ $\quad i = 1,2,\ldots, p$ (2)

iii. $w_{i1}{}^2 + w_{i2}{}^2 + \ldots + w_{ip}{}^2 = 1$ $\quad$ *for all* $i \neq j$ (3)

The condition of Eq. 1 requires that the squares of the weights sum to one and is somewhat arbitrary. The condition of Eq. 3 ensures that the new axes are orthogonal to each other.

Cluster Analysis

Cluster analysis identifies and classifies variables on the basis of the similarity of the characteristics they possess. It seeks to minimize within-group variance and maximize between-group variance. The result of cluster analysis is a number of heterogeneous groups with homogeneous contents (Aldenderfer and Blashfield, 1984). There are substantial differences between the groups, but the individuals within a single group are similar.

Data may be thought of as points in a space where the axes correspond to the variables. Cluster analysis divides the space into regions characteristic of groups that it finds in the data. Euclidean distance (*d*) is appropriate for variables that are uncorrelated and have equal variances. This distance (between rows) is a robust and widely applicable measure. Values are divided by the square root of the number of variables. The formula for computing Euclidean distance for p variables is given by

$$D_{ij} = \sqrt{\sum_{k=1}^{p} \left(x_{ik} - x_{jk}\right)^2} \qquad (4)$$

where D_{ij} is the distance between subjects i and j , x_{ik} is the value of the k^{th} variable for the i^{th} subject, x_{jk} is the value of the k^{th} variable for the j^{th} subject, and p is number of variable (Sharma, 1996).

An algorithm for partitioning (or clustering) N data points into K disjoint subsets S_j containing N_j data points so as to minimize the sum-of-squares criterion

$$J = \sum_{j=1}^{k} \sum_{n \in S_j} \left| x_n - \mu j \right|^2 \qquad (5)$$

where x_n is a vector representing the n^{th} data point and μj is the geometric centroid of the data points in s_j. In general, the algorithm does not achieve a global minimum of J over the assignments. In fact, since the algorithm uses discrete assignment rather than a set of continuous parameters, the "minimum" it reaches cannot even be properly called a local minimum. Despite these limitations, the algorithm is used fairly frequently as a result of its ease of implementation.

The algorithm consists of a simple re-estimation procedure as follows. Initially, the data points are assigned at random to the *k* sets. For step 1, the centroid is computed for each set. In step 2, every point is assigned to the cluster whose centroid is closest to that point. These two steps are alternated until a stopping criterion is met, i.e., when there is no further change in the assignment of the data points.

Correlations

To check the suitability of the data, correlations performed. The correlation matrices indicate the interdependence of the variables (cf. Tables 2-6). It suggests that meteorological variables are sufficiently correlated to each other. Some prominent correlations regarding these computations are summarizing as follows:

(a) Monsoon Season

Correlations of cloud cover with relative humidity and minimum temperature with maximum temperature are same i.e. 0.51 for the both relations.

Table 2: Correlations (Pearson) for Monsoon

	Relative humidity	Max Temperature	Wind Speed	Cloud cover	Rainfall
Max temperature	-0.19				
Wind Speed	-0.13	-0.06			
Cloud cover	0.51	-0.17	0.02		
Rainfall	0.27	-0.03	-0.21	0.12	
Min temperature	-0.22	0.51	0.10	-0.09	-0.22

(b) Winter Season

Minimum temperature with maximum temperature has a correlation of 0.45 while cloud cover with relative humidity and minimum temperature are 0.33 and 0.36 respectively.

Table 3: Correlations (Pearson) for winter

	Relative humidity	Maximum temperature	Wind Speed	Cloud cover	Rainfall
Max temperature	0.11				
Wind Speed	0.09	0.04			
Cloud cover	0.33	0.00	-0.03		
Rainfall	0.08	-0.17	0.08	0.07	
Min temperature	0.36	0.45	0.22	0.22	0.05

(c) Summer Season

Correlations of relative humidity with maximum temperature, wind speed, cloud cover and minimum temperature are comes out with the values of -0.44, 0.34, 0.43 and 0.53 respectively while minimum temperature with wind speed and cloud cover are 0.46 and 0.34 respectively.

Table 4: Correlation (Pearson) for summer

	Relative humidity	Maximum temperature	Wind Speed	Cloud cover	Rainfall
Max temperature	-0.44				
Wind Speed	0.34	-0.15			
Cloud cover	0.43	-0.28	0.24		
Rainfall	0.07	-0.04	-0.04	0.11	
Min temperature	0.53	-0.00	0.46	0.34	-0.04

(d) Spring Season

Minimum temperature with relative humidity, maximum temperature and wind speed has correlations of 0.35, 0.40 and 0.37 respectively.

Table 5: Correlation (Pearson) for spring

	Relative humidity	Maximum temperature	Wind Speed	Cloud cover	Rainfall
Max temperature	-0.21				
Wind Speed	0.26	0.03			
Cloud cover	0.26	-0.22	-0.00		
Rainfall	0.08	-0.10	-0.05	0.06	
Min temperature	0.35	0.40	0.37	-0.01	-0.01

(e) Autumn Season

Correlation of minimum temperature with maximum temperature is found 0.51.

Table 6: Correlation (Pearson) for Autum

	Relative humidity	Maximum temperature	Wind Speed	Cloud cover	Rainfall
Max temperature	0.02				
Wind Speed	0.21	0.024			
Cloud cover	0.06	-0.18	-0.01		
Rainfall	0.1	-0.18	0.06	0.11	
Min temperature	0.27	0.51	0.25	0.02	0.03

Identification of Weather Types

The aim of principal component analysis (PCA) before clustering procedure is to convert the input variables into uncorrelated components. It is applied to enhance the disconnecting power of the method. In this connection unrotated factor loadings and communalities for each season (cf. Tables 7-11) are calculated. The resulting percentage variance, unrotated factor loadings and communalities for each season suggested four and five statistically significant components for the monsoon and other seasons respectively, for the further process.

Table 7: Unrotated Factor Loadings and Communalities, Monsoon

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
Relative humidity	-0.74	-0.30	0.33	0.08	0.43	-0.24
Max Temperature	0.59	-0.65	0.10	-0.12	0.26	0.36
Wind Speed	0.21	0.55	0.57	-0.56	0.12	0.02
Cloud cover	-0.6	-0.19	0.63	0.12	-0.33	0.27
Precipitation	-0.49	-0.44	-0.35	-0.64	-0.17	-0.07
Min temperature	0.66	-0.43	0.41	0.02	-0.22	-0.40
Variance	1.98	1.23	1.13	0.76	0.46	0.43
% Variance	0.33	0.21	0.19	0.13	0.08	0.07

Table 8: Unrotated Factor Loadings and Communalities, winter

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
Relative humidity	-0.68	-0.33	-0.18	-0.14	-0.60	0.14
Max Temperature	-0.55	0.65	-0.04	0.34	0.11	0.38
Wind Speed	-0.32	-0.01	0.80	-0.48	0.12	0.13
Cloud cover	-0.49	-0.47	-0.48	-0.23	0.50	0.10
Precipitation	-0.08	-0.68	0.40	0.60	0.07	0.1
Min Temperature	-0.83	0.17	0.12	0.16	0.07	-0.48
Variance	1.81	1.23	1.07	0.81	0.64	0.43
% Variance	0.30	0.21	0.18	0.13	0.11	0.07

Table 9: Unrotated Factor Loadings and Communalities, summer

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
Relative humidity	0.84	0.12	0.09	-0.04	-0.40	-0.34
Max Temperature	-0.49	-0.56	-0.58	-0.23	0.01	-0.24
Wind Speed	0.64	-0.38	-0.07	0.54	0.37	-0.09
Cloud cover	0.68	0.22	-0.09	-0.52	0.46	-0.01
Precipitation	0.07	0.64	-0.72	0.23	-0.05	0.04
Min Temperature	0.72	-0.45	-0.27	-0.11	-0.28	0.32
Variance	2.34	1.14	0.95	0.69	0.59	0.29
% Variance	0.39	0.19	0.16	0.11	0.10	0.05

Table 10: Unrotated Factor Loadings and Communalities, spring

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
Relative humidity	0.65	0.54	-0.03	0.02	0.46	-0.28
Max Temperature	0.27	-0.78	0.2	-0.40	-0.10	-0.32
Wind Speed	0.69	-0.00	-0.17	0.49	-0.49	-0.09
Cloud cover	0.14	0.66	-0.15	-0.62	-0.38	0.01
Precipitation	-0.01	0.32	0.93	0.11	-0.13	-0.01
Min Temperature	0.84	-0.26	0.13	-0.18	0.14	0.39
Variance	1.70	1.51	0.97	0.84	0.64	0.33
% Variance	0.28	0.25	0.16	0.14	0.11	0.06

Table 11: Unrotated Factor Loadings and Communalities, Autumn

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
Relative humidity	-0.47	0.51	0.11	0.19	0.68	0.11
Max Temperature	-0.70	-0.49	-0.26	-0.19	-0.06	0.39
Wind Speed	-0.47	0.38	0.52	0.32	-0.50	0.11
Cloud cover	0.13	0.54	-0.73	0.33	-0.21	0.10
Precipitation	0.05	0.64	0.03	-0.75	-0.10	0.09
Min Temperature	-0.86	0.03	-0.24	-0.12	-0.09	-0.42
Variance	1.70	1.34	0.94	0.86	0.77	0.38

% Variance	0.28	0.22	0.16	0.14	0.13	0.06

PCs altogether accounting for more than 86% of the total variance in the diurnal data for monsoon while winter, summer, spring and autumn seasons acquired 93%, 95% 94%, and 93% respectively for the same. As regards monsoon season (cf. Table 7) relative humidity (Rh), maximum temperature (Tx), speed (Ws), cloud (Cl), precipitation (Ppt), and minimum temperature (Tn) shows the dominant values of –0.74, -0.65, 0.57, 0.63 –0.49, 0.66, for factor 1 (f1), factor 2 (f2), factor 3 (f3), factor 3 (f3), factor 1 (f1) and factor 1(f1). For winter season (cf. Table 8) the mentioned meteorological variables get the dominant values of –0.68, 0.65, 0.80, 0.50, -0.68, -0.83 for f1, f2, f3, f5, f2 and f1 respectively. Aimed at summer, spring and autumn seasons (cf. Table 9-12) domineering values of 0.84, -0.58, 0.64, 0.68, -0.72, 0.72 for f1, f3, f1, f1, f3, f1; 0.65, -0.78, 0.69, 0.66, 0.93, 0.84 for f1, f2, f1, f2, f3, f1 and 0.68, -0.70, 0.52, -0.73, -0.75, -0.86 for f5, f1, f3, f3, f4, f1 found respectively.

Analysis of Weather Types

After PCs operation *k-mean* clustering is applied to obtained different number of clusters for each season representing different types of weather (cf. Tables 12-16). Clustering analysis show that there are 14, 5, 8, 7 and 8 different weather types in monsoon, winter, summer, spring, and autumn seasons respectively. Some features of few important clusters/ weather types for each season are summarized in the following sub-sections:

Weather types of Monsoon Season

Monsoon season is characterized by 14 weather types (cf. Table 12) and these are the largest number of types obtain among all seasons, pointing that the monsoon comprises much more variations (14 clusters for 1104 days) in its nature than any other season for Karachi. First weather type is the rainiest consists of 17 days and total amount of rainfall is 845.9 mm in this interval. Such a great amount of rainfall within the small period is also signifying the variable nature of this season. It is characterized by greatest amount of clouds and maximum relative humidity of 7 oktas and 80.59 % respectively accompanied by less temperature tendency (maximum temperature minus minimum temperature) of 7.88 °C. Oppositely there is least rainwater (total 1.8 mm rain in 79 days) in fourth type, which is slightly windy (above 6 knots), less cloudy (more than 3 oktas) and least humid (61.42 %) behavior. Fifth cluster is the relatively coldest (minimum temperature = 24.02 °C) and have second highest rainfall (total 138.4 mm within 40 days) in this season. Characteristic wise this cluster is humid (71.72 %) and slightly windy (above 6 knots) with slightly high temperature tendency (9.15 °C). The fourteenth weather type consists of 143 days (biggest cluster), which receives 25.7 mm total rain. This weather type is cloudy (6 oktas) with high wind speed (11 knots) and relative humid (66.64 %). The temperature tendency in this type is again relatively less (6.60 °C). Sixth type is relatively warmest (maximum temperature = 37.70 °C) with slightly greater temperature tendency (10.78 °C) with cloudy weather (greater than 6 oktas) and sufficient amount of rain (total 105.4 mm rain within 24 days)

Table 12: 12 K-mean result for monsoon categorization

#	Relative Humidity	Max Temp	Wind Speed	Cloud Cover	Rainfall	Total Rainfall	Min Temp	No of Clusters	No. of Days
1	80.59	32.85	4.94	7.00	49.76	845.9	24.97	1	17
2	70.18	30.76	6.39	7.13	0.63	66.6	25.48	2	106
3	63.00	32.21	11.75	3.45	0.07	7.3	26.11	3	106
4	61.42	34.66	6.57	3.18	0.02	1.8	28.1	4	79
5	71.72	33.18	6.5	5.57	3.46	138.4	24.02	5	40
6	67.38	37.70	8.33	6.54	4.39	105.4	26.92	6	24
7	66.56	33.12	15.83	5.88	0.12	9.6	27.50	7	78
8	67.14	30.75	6.81	4.44	1.66	94.8	27.05	8	57
9	69.26	35.57	5.13	7.37	0.05	2.1	28.38	9	38
10	69.47	30.99	11.93	6.98	0.78	97.4	26.59	10	124
11	62.34	32.99	7.82	0.76	0.15	9.5	26.41	11	62
12	68.19	32.71	7.23	6.65	0.57	67.1	27.43	12	118
13	65.29	34.15	10.52	5.56	0.15	17.1	28.54	13	112
14	66.64	32.96	10.83	6.71	0.18	25.7	26.36	14	143

Weather types of Winter Season

Winter season is characterized by 5 weather types (cf. Table 13) only and hence these are the smallest (only 5 clusters) number of clusters obtain among all seasons. It indicates the least varied nature of winter season than any other season for Karachi.

Table 13: K-mean result for winter categorization

#	Relative Humidity	Max Temp	Wind Speed	Cloud Cover	Rainfall	Total Rainfall	Min Temp	No of Clusters	No. of Days
1	42.38	29.46	5.43	0.51	0.13	39.1	13.46	1	293
2	24.92	28.37	4.52	4.79	0.64	117.3	13.21	2	183
3	24.70	25.89	3.51	0.52	0.17	59.7	11.78	3	359
4	57.04	26.70	5.69	4.96	0.66	129.0	14.35	4	194
5	26.64	26.458	10.50	0.70	0.45	76.4	12.98	5	171

Third weather type is the most important, as it is characterized by two significant properties i.e. biggest (359 days) and coldest (minimum temperature = 11.77 °C). Again this character shows that winter season is least varied in its nature. Other properties of this cluster are nominal clouds (less than 1 okta), less humidity (24.70 %), greater temperature tendency (14.12 °C) and nominal rain rainfall (59.7 mm in 359 days). First weather type is relatively hottest (maximum temperature = 29.45 °C) with high temperature tendency (16 °C) and nominal precipitation (total 39.1 mm rain in 293 days). It is average humid (42.38 %) with nominal clouds (less than 1 okta). Fourth weather type is relatively rainiest (total 129 mm rain in 194 days) in this season but cannot be considered rainy in general. Consequently this weather type is cloudy (almost 5 oktas), and highly humid (57.04 %). Temperature tendency is slightly high (12.34 °C) Fifth weather type is dissimilar to all others, owing highly wind speed (above 10 knots) although it is smallest in size (comprising 171 days) for this season. This cluster shows the days when season changes its state from one form to other.

Weather types of Summer Season

We combine summer 1 and 2 for this categorization and get 8 weather types in total (cf. Table 14). Although number of weather types for this season seems greater than winter season, but this is not the case.

Table 14: K-mean result for summer categorization

#	Relative Humidity	Max Temp	Wind Speed	Cloud Cover	Rainfall	Total Rainfall	Min Temp	No of Clusters	No. of Days
1	56.77	34.77	12.55	1.07	0.06	18.4	26.36	1	318
2	38.38	36.20	5.57	0.50	0.03	6.1	19.96	2	215
3	60.96	33.49	6.26	5.59	0.08	18.9	25.82	3	244
4	55.21	35.62	6.77	0.90	0.01	4.0	27.10	4	378
5	43.51	39.71	6.87	1.74	0.01	3.0	25.87	5	202
6	54.37	32.76	6.66	1.02	0.03	12.4	24.49	6	358
7	64.09	33.98	12.75	5.49	0.06	17.2	27.56	7	293
8	69.37	34.00	5.5	5.5	24.02	192.2	23.81	8	8

Similar to winter, summer season also shows its less varied nature to Karachi, because it comprises much more days (8 clusters for 2016 days) than winter (5 clusters for 1200 days). Fifth weather type is hottest (maximum temperature = 39.70 °C) with slightly elevated temperature tendency (13.83 °C). Very less clouds (1.74 oktas), least nominal rain (total 3 mm rain within 202 days) for summer season, slightly greater windy (6.87 knots) and relatively less relative humid (43.50 %) are the other important characteristics of this weather type. Fourth weather type consists of 378 days (biggest cluster) having nominal clouds (less than 1 okta). 8.53 °C is the temperature tendency. Other characteristics involves slightly windy (above 6 knots) and average humid (55.21 %) almost no rain (total 4 mm 378 days) are the other characteristics of this weather type. Second weather type is relatively coldest (minimum temperature = 19.96 °C) in summer acquiring greater temperature tendency 16.23 °C. It shows high value because inspite of least minimum temperature, maximum temperature is 36.19 °C which is due to seasonal nature of summer. Other properties are nominal clouds (less than 1 okta), nominal rain (total 6.1 mm in 215 days), least humid (38.38 %) for summer season and slightly windy (more than 5 knots). Eighth weather type comprises 8 days only (smallest cluster) and rainiest (192.2 mm total rain in 8 days). Consequently it is humid (69.37 %) cloudy (greater than 5 oktas), slightly windy (above 5 knots) weather type. Temperature tendency is 10.19 °C. First and seventh weather type have distinct feature of highest wind speeds (greater than 12 knots) for summer season.

Weather types of Spring Season

Spring season is characterized by 7 weather types (cf. Table 15). Although it also seems of less varied nature because number of weather types are comparable to winter and summer, but truly speaking it has greater variations than both because it comprises of relatively less number of days (784 days).

Table 15: K-mean result for spring categorization

#	Relative Humidity	Max Temp	Wind Speed	Cloud Cover	Rainfall	Total Rainfall	Min Temp	No of Clusters	No. of Days
1	50.80	31.43	9.03	0.95	0.01	1.1	20.40	1	167
2	31.40	33.83	12.55	0.91	0.05	3.6	20.82	2	78
3	22.68	30.41	5.46	0.76	0.27	35.1	17.93	3	129
4	52.78	34.21	5.41	1.67	0	0	20.93	4	107
5	23.37	36.91	7.03	0.44	0	0	20.20	5	139
6	54.32	31.24	10.64	6.04	0.06	4.2	20.63	6	69
7	29.07	33.05	5.70	5.31	0	0	18.73	7	95

First weather type comprises 167 days (biggest cluster). It is windy (greater than 9 knots). The important point to be noted is that inspite of nominal clouds (less than 1 okta) and nominal rain (total rain 1.1 mm) it sufficiently humid (i.e. 50.80 %). It describes that although moisture content is sufficient in atmosphere but the factors for the formation and stimulating the convective clouds are absent and hence rain is also awfully nominal. Fifth weather type is hottest (maximum temperature = 36.91 °C). This factor causes less relative humidity (23.37 %), nominal clouds (less than 1 okta) and no rain, while wind is slightly high (about 7 knots). Third Weather type is coldest (minimum temperature = 17.93 °C) with slightly high temperature tendency is 12.48 °C. Note that maximum temperature is also reaches its least value (30.41

°C) in this season. It is because of the nature of this spring season. Other individuality involves nominal clouds (less than 1 okta), slightly windy (greater than 5 knots), least relative humidity (22.68 %) for this season and greatest amount (infect nominal) of rain (35.1 mm in 129 days) for this season. Second weather type has a separate feature of greater wind speed (greater than 12 knots) with almost no rain (3.6 mm total rain in 78 days) and nominal clouds (less than 1 okta). Sixth and seventh clusters are very cloudy weather (above 6 and 7 oktas) respectively.

Weather types of Autumn Season

Autumn season is characterized by 8 weather types (cf. Table 16) and as in case of spring, it is also relatively greater varied nature because relatively large number of clusters (8) obtained in less days (736 days).

Table 16: K-mean result for autumn categorization

#	Relative Humidity	Max temp	Wind Speed	Cloud Cover	Rainfall	Total Rainfall	Min Temp	No of Clusters	No. of Days
1	58.19	31.19	4.60	1.02	0.01	1.2	16.83	1	87
2	38.10	34.65	3.92	0.23	0	0	16.57	2	89
3	22.18	34.33	4.78	0.62	0	0	17.24	3	121
4	19.47	30.17	8.31	1.49	0.01	0.5	15.23	4	57
5	32.70	31.01	3.68	5.48	0.09	8.3	15.69	5	93
6	25.05	31.41	2.31	0.35	0.03	4.4	15.23	6	135
7	47.40	31.72	7.23	1	0	0	17.59	7	87
8	25.82	27.66	3.63	0.40	0.00	0.3	12.85	8	67

Sixth weather type comprises of 135 days (biggest cluster) with marked temperature tendency (16.17 °C). Calm wind (2.3 knots), nominal rain (total 4.4 mm), nominal clouds (less than 1 okta) and less humid (25.05 %) are the other personality features of this weather type. fourth type is unlike because of least relative humidity (19.47 %) for autumn. Almost no rain (total 0.5 mm in 57 days), nominal clouds (less than 2 oktas) and slightly high windy (above 8 knots) are other features. The values of relative humidity and wind speed in this type shows the extreme dryness and windiness for this season. Eighth weather type is coldest (minimum temperature = 12.85 °C) in autumn with greater temperature tendency (14.8 °C). Other factors are nominal clouds (less than 1okta), almost no rain (total just 0.3mm in 67 days) and less humid (25.82 %). Second weather type is warmest (maximum temperature = 34.64 °C) with very greater temperature tendency (18.07 °C). As almost no clouds (less than 1 okta) are there, accordingly there is no rain, calm wind (3 knots) and average humidity (38.10 %). In this season Fifth weather type is the only type with cloudiness (more than 5 oktas) and hence appears rainiest (total rain 8.3 mm in 93 days) for this season (in fact rain is nominal), otherwise for whole autumn it is almost equal or less than 1 okta.

Conclusions

The geographical location of Karachi plays very important role in respect of weather types. The main scope of this weather types study was to categorize the weather feature in the area of Karachi formed by each prevailing weather type. The main characteristics of all categories are summarized in Tables 12-16.

In this paper an attempt was made to apply a modification of Kalkstein's procedure (originally designed for the North American Arctic), to the diurnal data for Karachi. The procedure applies principal component analysis to the data of 16 years, and clustering is achieved by the average linkage technique.

The crucial difference between Kalkstein's procedure and the procedure in this study consists in the method of determining the threshold dissimilarity measure (or the aggregation level) at which the clustering procedure should be terminated. Kalkstein's procedure made use of the R2 criterion, implying a fixed threshold aggregation level throughout the whole data set. This led to the snowballing effect i.e. one huge cluster was formed, accompanied by a large number of much smaller clusters and unclustered days. To avoid this undesirable effect, in this study variation in the threshold aggregation level is allowed. The clustering procedure is then terminated at different dissimilarity threshold aggregation level to vary. The clustering procedure is then terminated at different dissimilarity measures in various parts of the data set

Six variables describing the weather were used as the input data. The principal component analysis reduced them to four and five uncorrelated components for monsoon and other seasons correspondingly. After the average linkage clustering procedure had been applied 14, 5, 8, 7 and 8 different clusters with sizes of eight or more days were found for the period 1990-2005 for monsoon, winter, summer, spring, autumn seasons respectively. Some features of few important clusters for each season were summarized in the paper. The results confirmed the expectation that the weather in the Karachi is much more varied than in the Arctic.

The results concerning the distinguishability of the clusters from the parent data justify the application of the procedures used (both PCA and average linkage clustering technique) and prove the proposed method to be suitable for determining clusters. The results imply that an optimum number of synoptic types may be greater that suggested by Yarnal White (1987), who pointed out that some measures would have to be taken to reduce the numbers of weather types. They also claimed "one of the thorniest problems facing the computer assisted synoptic climatologist is to minimize the number of synoptic categories while maximizing the percentage of day classified". We succeeded in maximizing the number of days classified with out being obliged to reduce the number of categories. In author's opinion, the variety of weather conditions in Karachi is so sufficient that the weather cannot be squeezed greater, say, to only more than 1 type. To do so would result in categories not likely to be distinguished sufficiently from the whole data set, and therefore they would not represent meaningful weather types.

The clusters obtained by the average linkage technique for these available, 16 years, period may serve as a basis for a non-hierarchical method of clustering. Such methods, as compared with the average linkage, are simpler and may enable us to elaborate much longer time series. Then, the temporal changes in frequencies of individual categories as well also the changes in their physical character can be investigated, both indication the presence or absence of significant long term temperature trends. The future outlook work includes an expansion of the system spatially wherever data availability is sufficient.

References

Aldenderfer M. S. and R. K. Blashfield (1984): Cluster Analysis, Sage Publications, Newbury Park. Cal.

Barry, R.G. and A. H. Perry, 1973.Synoptic Climatology: Methods and Applications, Methuen, London.

Christensen, W.1. and R.A. Bryson, 1966: 'An investigation of the potential of component analysis for weather classification', Mon. Wea. Rev. 94, 697—709.

Davies T.D., Kelly P.M., P. Brimblecombe, G. Farmer and R.J. Barthelmie, 1986: Acidity of Scottish rainfall influenced by climatic change. Nature 322: 359-361.

Hair, Anderson, Tatham and Black (1998): Multivariate Data Analysis. 5th edition, Prentice Hall, US.

Harper (1999) Numerical Palaeobiology. John Wiley & Sons.

Houghton, J.T., G.J. Jenkins and J.J. Ephramus, 1990: Climate Change: the IPCC scientific Assessment, Cambridge University Press, Cambridge.

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

Johnson, R., and D. Wichern. 1998. Applied Multivariate Statistical Analysis. Prentice Hall, Upper Saddle City, NJ.

Kalkstein, L. S., P.C. Dunne and R. S. Vose, R. S. 1990: ' Detection of climatic change in the western North American Arctic using a synoptic climatiological approach', J. Climate, 3, 1163-1167.

Kalkstein, L.S. and P. Corrigan, 1986: 'A synoptic climatological approach for geographical analysis: assessment of sulfur dioxide concentrations', Ann. Assoc. Am. Geogr., 76, 381-395.

Sadiq N, (2007): "Classification of weather at Karachi region", M.Phil. Thesis, University of Karachi.

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

Stone, R.C.1989: 'Weather types at Brisbane, Queensland: an example of the use of principal components and cluster analysis', Int .J. Climatol., 9, 3-32.

Yarnal, B. and D. A. White 1987: 'Subjectivity in a computer-assisted synoptic climatiology I: classification results', J. Climatolol., 7, 119-128.