Relative Significance of Various Factors in Lake Evaporation : A Case Study using Principal Component Analysis

S. D. Khobragade[1], C.S.P. Ojha[2], Omkar Singh[1] and R.D. Singh[1]

[1]National Institute of Hydrology, Roorkee (Uttarakhand)-247 667
[2]Indian Institute of Technology, Roorkee (Uttarakhand)-247 667
e-mail : suhas@nih.ernet.in

ABSTRACT

Evaporation in lakes is influenced by various factors such as temperature, radiation, atmospheric pressure, wind, sunshine hours, relative humidity, water quality etc. However, relative significance of the various factors in the process of lake evaporation is not known. There are hardly any reported studies on this aspects. The present study has been carried out for Pichhola Lake which is located in Udaipur in the state of Rajasthan in India. The major objective of the study is to identify some of the important variables that influence the process of lake evaporation and their relative significance for the process. A multivariate statistical technique namely principal component analysis has been used for the purpose. Five years of daily data (2002-2006) relating to eight variables viz. maximum temperature, minimum temperature, mean temperature, maximum relative humidity, minimum relative humidity, mean relative humidity, bright sunshine hours and wind were analyzed for the study.

The results of study indicate that three factors having dominance of temperature, relative humidity and wind respectively, control the evaporation from the lake. Out of the three the first two factors are very significant as they explain about 85 percent of the total variance in the original data. The first factor explains about 39 % and the second factor explains about 45 percent of the total variance. The three factors together explain about 97 percent of the total variance. Detailed analysis of the principal factors based on varimax factor rotation technique brings out the fact that lake evaporation is a combined effect of temperature, relative humidity (vapour pressure) and wind, with temperature and relative humidity (vapour pressure) being relatively highly significant than the wind. The study is useful in understanding the process of lake evaporation in a better way.

INTRODUCTION

Evaporation is one of the significant components of the hydrologic cycle. Therefore, proper understanding of the process of evaporation is necessary for proper understanding of the hydrologic cycle. However, since precise estimation of lake evaporation still remains one of the challenging tasks for the hydrologists and water resources engineers, most of the reported studies on evaporation lay emphasis on the estimation of evaporation and only a few refer to the understanding of the complexity of interrelationships of the various variables involved. Evaporation depends on complex interrelationships between various hydro-meteorological factors. The factors that influence the process include temperature, radiation, vapour pressure (humidity), bright sunshine hours, wind, water quality, atmospheric pressure etc. These interrelationships vary in different climatic settings. For example, although the controlling factors for evaporation are not quite different in tropical and temperate climatic settings, yet the evaporation rates are not the same for the two regions. This is because, the nature of controlling factors is different in the two climatic settings, implying varying significance of the controlling factors in different climatic settings. As rightly observed by Mohan and Arumugam (1996), "knowledge of the relative effects of these factors on evaporation, is very important for irrigation management. So analysis of the available variables is necessary, for understanding the relative importance of these factors". Multivariate statistical technique is a useful tool in this direction. Factor analysis and principal component analysis, which are the most popular multivariate statistical techniques available, have been used in the past to understand the mechanisms of various hydrological, hydro-meteorological and hydro-chemical phenomena (for example, Diaz et. Al., 1968; Morin et al., 1979; Deverel, 1989; Rao, 1990; Iyenger, 1991; Pandzic and Tminic, 1992; Melloul and Collin, 1992; Vajrappa and Srinivas 1994; Rao, 2001; Bhatia et. al., 2005 etc). The main advantage of these techniques is their suitability for simultaneous analysis of large number of variables and observations. However, application of such techniques has not been adequately explored for evaporation studies. The only reported study on the use of such a technique for evaporation is by Mohan and Arumugam (1996) who applied the factor analysis approach to understand the relative importance of various meteorological variables in evapo-transpiration for eight stations of Karnataka and Tamilnadu in India. Southern Rajasthan has a different climatic setting than these regions. So, the present study has been taken up for lake Pichhola, Udaipur, Rajasthan with a view to investigate the relative significance of the various meteorological factors in the process of lake evaporation. A multivariate statistical tool called Principal Component Analysis has been used for the purpose. Work of Mohan and Arumugam (1996) has been used as a guiding tool for the study.

STUDY AREA

Lake Pichhola is a very significant lake of Udaipur, Rajasthan, because of the various socio-economic and cultural uses. It is a major source for drinking and domestic water supply for the people of the region. Of the total water supply to Udaipur city, about 85% is met from Lake Pichhola alone. Besides being a major tourist attraction, the lake is also ecologically very significant as it supports diverse fauna and flora. Lake Pichhola is a manmade lake constructed in the 14th century. It was on the bank of this lake that the city of Udaipur was developed. Due to improper water management, the water availability in the lake is reducing continuously in recent times. Heavy evaporation losses from the lake, particularly in summer, being one of the reasons for the shrinking water availability. So there is a need for a better understanding of the hydrological behaviour of the lake for its conservation and management.

The study area is a semi-arid climatic region. There are three distinct seasons viz. winter (October to mid February), summer (mid February to mid June) and monsoon (mid June to September). Maximum temperature is around 43^0 C in May-June while minimum can be as low as 1.5^0 C. Most of the rainfall occurs during the monsoon months of June-September. Distribution of annual rainfall is uneven and shows large variations. Air is generally dry except for the monsoon period when the humidity is around 70%. Summer months are the driest of the year when the humidity goes to about 20-25%. Winds are generally light with some strengthening in the latter half of summer and the monsoon. Dust-storms and thunderstorms occur sometimes in the hot months of summer (Khobragade, 1996). Figure 1 shows the location map of study area. Table 1 presents the salient features of the lake.

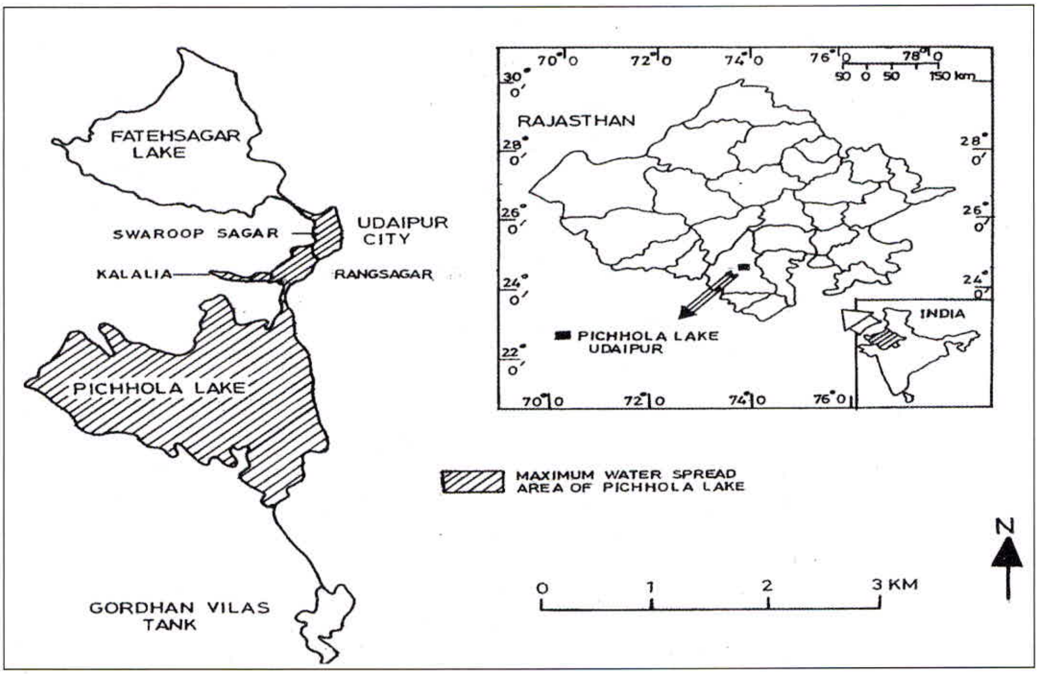


Fig. 1. Location map of Lake Pichhola, Udaipur (Rajasthan)

DATA USED

Data of maximum temperature (Tmax), minimum temperature (Tmin), mean temperature (Tmean), maximum relative humidity (RHmax), minimum relative humidity (RHmin), mean relative humidity (RHmean), wind, bright sunshine hours (BSS), pan evaporation and rainfall have been initially used in the study. 5 years of monthly data (2002-2006) were obtained from the meteorological observatory at the College of Technology and Engineering, Maharana Pratap University of Agriculture and Technology, Udaipur; which is located at about 5 kms to the east of the lake. Data

Table 1 : Salient features of Lake Pichhola, Udaipur

Parameter	Value
Longitude	$73^0\ 42'$
Latitude	$24^0\ 35'$
Altitude (m)	587
Normal rainfall (mm)	635
Storage capacity (MCF)	485
Water Spread Area (Sq. Km)*	6.96
Maximum depth (m)	8
Mean depth (m)	4.5
Maximum length (km)	3.6
Maximum width (km)	2.61
Mean width (km)	1.93
Length of shoreline (km)	12.9

** The water spread area fluctuates annually and seasonally*

were available upto October, 2006. Lake evaporation was obtained from the pan evaporation data by applying suitable coefficients. Since the pan evaporation data is of Class a Pan covered with mesh, a correction factor of 1.144 was applied. Mean temperature and mean relative humidity were additionally considered besides their maximum and minimum values so as to analyze the significance of their average values for the day.

A simple linear correlation of each of these parameters with lake evaporation was first carried out. The results of the correlation analysis are given in Table 2. It can be observed that maximum temperature has a maximum correlation with lake evaporation (r=0.89) while rainfall has almost no correlation with lake evaporation (r=-0.1). Besides the maximum temperature; minimum temperature, mean temperature, maximum relative humidity and wind also show a significant correlation. While mean relative humidity shows a reasonable correlation, a little weak correlation is observed in the case of bright sunshine hours and minimum relative humidity. Based on the correlation coefficients, variables with a reasonable correlation ($R = \pm 0.3$) among the various variables, were considered as influencing the process of evaporation from the lake with more or less significance and as such were considered for principal component analysis to study their relative significance. Thus, eight of the nine variables (excluding rainfall) were considered. Since data for November and December, 2006 were not available, in all 58 number of observations (n=58) were used for the study.

Table 2 : Correlation coefficient of various variables with lake evaporation

S.No.	Variable	Correlation coefficient (r)
1	Tmax	0.89
2	Tmin	0.70
3	Tmean	0.82
4	RHmax	-0.77
5	RHmin	-.30
6	RHmean	-0.57
7	Wind	0.76
8	BSS	0.29
9	Rain	-0.10

METHODOLOGY

In principal component analysis (PCA) a set of relevant factors (called principle components) is extracted from correlation matrix of the various variables and the interpretation of these factors is then carried out to identify the relative significance of the various variables. All the extracted factors are uncorrelated (i.e. orthogonal). Each factor is a linear combination of the original variables. The total variation explained by the variables is equal to the total variation in the original variables. The first factor accounts for the largest amount of variation in the data, the second factor accounts for the greatest residual variance and so on. Thus, since no information is lost during the transformation, a small number of factors contain most information in the original data (Mohan and Arumugam, 1996). The detailed methodology of PCA is discussed by Davis (1986).

In the present study, as mentioned earlier, eight variables viz. maximum temperature, minimum temperature, mean temperature, maximum relative humidity, minimum relative humidity, mean relative humidity, wind and bright sunshine hours have been considered. Since the variables have different units, so the data have been first standardized. This has been done by subtracting the mean of the observations for a variable from the value of that variable for an observation and then by dividing it by the standard deviation of that variable. This is a prerequisite for principal component analysis. As there are 8 variables under consideration , so a 8 X 8 correlation matrix [R] is first constructed in the PCA. This matrix represents the basis for an orthogonal transformation of the observed variables into factors, called principal components.

From this correlation matrix [R], eigenvectors and eigenvalues are extracted. Trace of the matrix (sum of the 8 eigenvalues) and percentage of trace for each eigenvalue are computed. This explains the variance in the data set. Number of principal components (factors) (generally small in number than the total variables) are decided on the basis of the total variance explained. A matrix of loading of factors is derived from the eigenvalues for the chosen number of factors. To remove unnecessary orthogonal axes a process called factor rotation is employed. In the present study varimax factor rotation has been utilized. Rotated factor loadings are then examined for qualitative interpretation. How well the variance is described by a particular set of factors is defined by communality.

RESULTS AND DISCUSSION

General statistical characteristics of the data set are presented in Table 3. During the period of study the maximum temperature reached was 39.49^0 C while minimum temperature reported was as low as 4.06^0 C., the average temperature being 24.33^0 C. The highest relative humidity recorded was 90.97 % while the lowest value of relative humidity was 13.29. The mean relative humidity was 54.39%. Similarly variations can be observed for other parameters also. Standard deviation and variance for the data indicate that the variations are highest for the relative humidity and are lowest for wind. These variations result in consequent variations in the rates of evaporation from the lake.

The results of the principal component analysis are presented in Table 4. It can be observed from the table that the eigenvalue of the first component is 4.016 and it has a trace value of 50.19% . The next two components have eigenvalues of 3.292 and 0.425

Table 3 : General statistical characteristics of the data

Statistic	Tmax	Tmin	Tmean	RHmax	RHmin	RHmean	Wind	BSS
n	58	58	58	58	58	58	58	58
Min	22.88	4.06	13.49	37.73	13.29	25.68	0.90	1.97
Max	39.49	26.36	32.85	90.97	78.23	84.10	9.53	10.95
Mean	31.85	16.80	24.33	71.44	37.34	54.39	4.05	8.28
median	31.60	17.09	25.80	78.10	31.23	55.60	3.81	8.94
STDDEV	4.38	7.60	5.60	17.01	18.85	16.44	2.13	2.18
Variance	19.17	57.75	31.40	289.27	355.22	270.39	4.55	4.75

Table 4 : Eigenvalues and trace of the various components

Components	Eigenvalues	% of trace	Cumulative % of trace
1	4.016	50.19	50.19
2	3.292	41.14	91.34
3	0.425	5.31	96.65
4	0.201	2.51	99.16
5	0.055	0.69	99.85
6	0.012	0.15	100.00
7	0.000	0.00	100.00
8	0.000	0.00	100.00

respectively. These two variables have trace values of 41.14% and 5.31% respectively. Thus, the first two components only have eigen values of more than one and as such are significant. For further finer interpretation, factor analysis employing varimax factor rotation technique was to be employed. The factor analysis needs the knowledge of the number of factors before employing the factor rotation. These loading numbers of factors are to be decided based on the PCA analysis. From the PCA analysis it is obvious that only two components have eigenvalues of more than 1.00 and as such are sufficient to be considered. The third component has an eigenvalue of less than one (0.425). However, it was decided to include the third component also for better interpretation as well as to have a higher degree of precision in the analysis.

The results of rotated principal component analysis are presented in Table 5. Factors loading of various variables for the first three factors and corresponding communality values are presented. The geographical representation of the factors is presented in Fig. 2. Fig. 3 shows the graphical representation of the factor loadings.

It can be observed from Table 5 and Fig. 3 that RHmean has the highest loading for factor 1 with a loading of 0.978. The other variables which have significant loadings on factor 1 are RHmin, RHmax and BSS. While the relative humidity variables are positively loaded for factor 1, BSS is negatively loaded. This means that as the impact of relative humidity increases, the effect of bright sunshine hours decreases. Tmean has the highest loading for factor 2, while wind has the highest loading for factor 3. Tmean (mean temperature) has the highest loading of 0.973 on factor 2, while it has the lowest loading of -0.004 for factor 1. It also has a low factor loading for factor 3. This means that while it the most significant variable for factor 2, it not so significant for the remaining factors. Similarly while minimum temperature has a high loading of 0.920 for the second factor it has low loadings for the factors 1 and 3. In case of maximum temperature, the loading is significantly high for factor 2 while it is low for factor 1 and 3. It can also be observed that whereas Tmin has a positive factor loading on all the three rotated factors, the behaviour is different for Tmin and Tmax. Both Tmean and Tmax show a positive loading for factor 2 and 3 but show a negative for factor 1. This means that where temperature

Table 5 : Factor loadings of variables on rotated factors

Variable				
	1	2	3	Communality
TMEAN	-0.004	0.973	0.228	0.999
TMIN	0.243	0.920	0.299	0.995
TMAX	-0.433	0.894	0.066	0.991
WIND	0.015	0.592	0.772	0.947
RHMEAN	0.978	-0.122	-0.130	0.988
RHMIN	0.973	0.145	0.080	0.974
BSS	-0.905	-0.077	-0.283	0.905
RHMAX	0.812	-0.397	-0.339	0.932

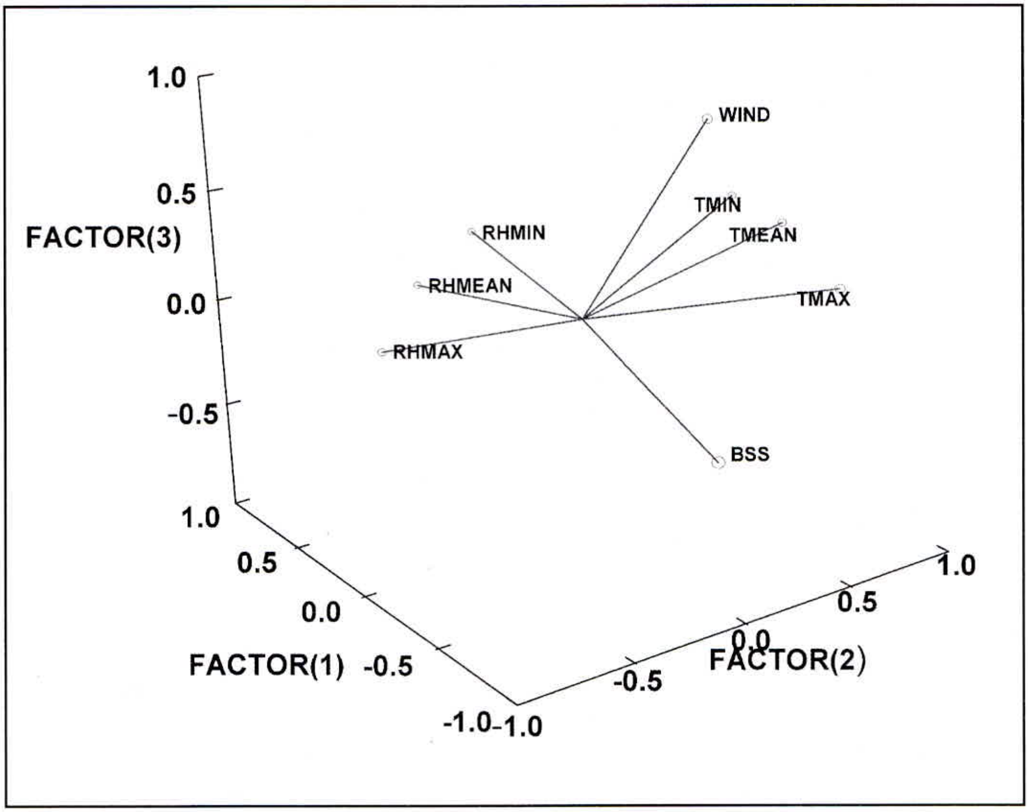
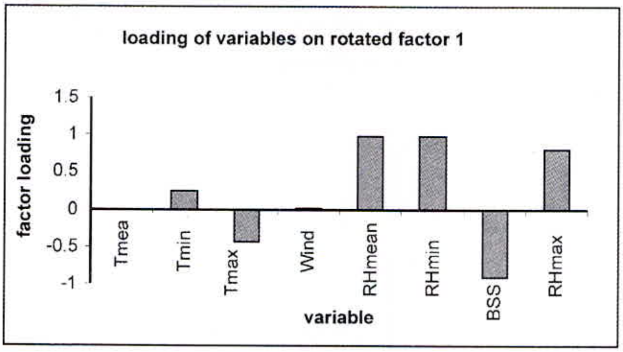


Fig. 2. Geometric representation of the rotated factors

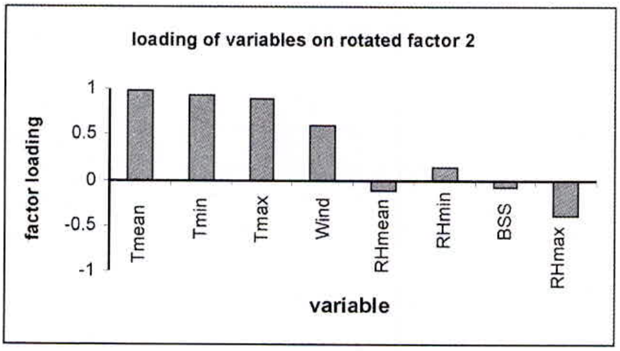
is the dominant factor, it affects the lake evaporation process positively in the presence of wind. However, in relation with relative humidity the role temperature becomes negative. This is obvious because the vapor pressure difference decreases causing the decrease in the rate of lake evaporation.

Communality data in Table 5 indicates how well the variance is explained by the factors. It can be observed that communality is more than 0.90 in all the cases and it is 0.99 in the case of mean temperature, maximum temperature, minimum temperature. The communalities for all the factors together is 1.0. But since only three factors have been considered, the communalities are less than 1. A communality of 0.99 indicates that of the total contribution of any variable in all the factors together, 99.00 has been accounted for in the first three factors itself.

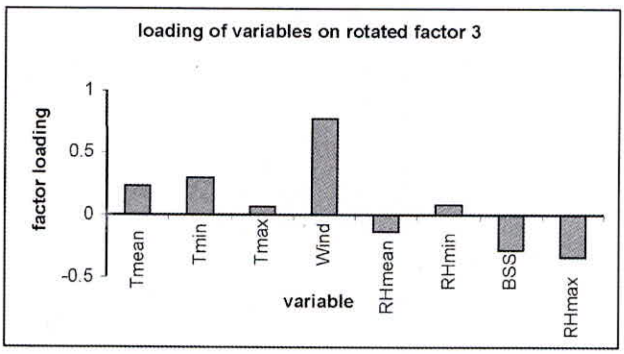
The percentage of total variance explained by each factor is presented in Table 6. As has been pointed out in the principal component analysis the first two factors are very significant while the third factor is marginally significant. The first component explains about 45.5 % of the total variance while the second factor explains 39.3 % of the total variance. Thus a little change in variance explained is observed by rotated factors than



(Fig. 3.1)



(Fig. 3.2)



(Fig. 3.3)

Fig. 3 : Graphical representation of factor loadings on the rotated factors

the principal components. In PCA the first factor explained 50.19 % of the variance while the second factor accounted for 41.14 % variance. In the rotated factor analysis, the first two rotated factors explain about 85 % of the total variance in the data set and the first three factors explain almost 96 % of the variance. Obviously the remaining factors are not substantial to significantly explain the variation and hence their exclusion is justified.

Table 6 : Variance explained by the rotated factors

Factor	% of total variance explained	Cumulative % total of variance explained
Factor 1	45.367	45.367
Factor 2	39.294	84.66
Factor 3	11.990	96.65

It can be observed from table 4 above that in each factor, factor loadings are high for few variables while it very low for some variables. This is the result of the factor rotation. Factor rotation maximizes the variances of each factor and produces a set of orthogonal factors which are independent or uncorrelated. Maximizing the variance implies maximizing the range of loadings, which tends to produce either extreme (positive or negative) or near zero loadings, satisfying the purpose of rotation. The method operates by adjusting the component loadings so that they are either near ±1 or near zero. For each component, there will be a few significantly high loadings and many insignificant loadings. The variables with higher values of factor loadings are the ones which are dominating and are relatively significant on each factor. The dominanace of the variables is decided by the magnitude of the loading. Higher the magnitude greater is the dominance i.e. greater would be the association with that factor. In interpretation it helps in naming the factor. In this study 0.5 was taken as a line of demarcation. Variables with factor loading of 0.5 and more (in either direction) were treated as dominant variables. These variables were used in identifying and naming the factors. The significant variables for each factor in descending order of their significance are listed in Table 6. They are clearly visible in the graphical representation shown in Fig. 3.

It can be observed from Fig. 3 and Table 7 that variables RHmean, RHmin, BSS and RHmax are the main variables in the first factor. All the loadings are significantly high. RHmean had the highest factor loading for factor 1. BSS has a negative loading meaning that in combination with the relative humidity, it affects the lake evaporation negatively. Since three of the four variables are related to humidity, relative humidity is the key variable for factor 1. Thus this factor can be considered as relative humidity factor. Since relative humidity is the measure of vapour pressure, it tells that vapor pressure plays the most significant role in the process of lake evaporation for the study area.

Table 7 : Significant variables for rotated factors

Factor 1		Factor 2		Factor 3	
Variables	Factor loading	Variables	Factor loading	Variables	Factor loading
RHmean	0.978	Tmean	0.973	Wind	0.772
RHmin	0.973	Tmin	0.920		
BSS	-0.905	Tmax	0.894		
RHmax	0.812	Wind	0.592		

Variables Tmean, Tmin, Tmax and Wind were are significant for the second factor. The loading of mean temperature was the highest for this factor. It can further be observed that out of the four variables that were dominating for the factor 2 (Tmean, Tmin, Tmax and wind) the first three variables that had very high factor loadings of almost 0.9 and above, are related to temperature. The fourth variable wind has a relatively low factor loading (0.592). Thus this factor can be identified and named easily as temperature factor. For the third factor, which is less significant than the first two factors in respect of the percent of variance explained, wind was the only significant variable with a factor loading of 0.772. This factor can be considered as wind factor.

From the above analysis, it can be interpreted that lake evaporation is a combined effect of relative humidity (vapour pressure), temperature and wind. Factor 1, which is the most significant factor for the lake evaporation in the order of significance, explains about 45 % of the variance in the original data. It is related to vapour pressure. The second factor, which is a temperature factor, explains about 39 % of the variance and as such can be regarded as equally significant factor. Thus, it can be inferred that relative humidity (vapour pressure) and temperature are the most significant variable for the process of evaporation from lake Pichhola. Wind is the third significant variable. However, its significance in relation to temperature and humidity is far less.

CONCLUSIONS

In the present study, a multivariate statistical technique principal component analysis with varimax factor rotation has been employed to find the relative significance of different variables in the process of lake evaporation from Pichhola lake, Udaipur, Rajasthan which is located in semi arid/subtropical climatic setting. The results of study indicate that three factors control the evaporation from the lake. Out of the three, the first two factors are very significant as they explain about 85 percent of the total variance in the original data. The first factor which is related to vapour pressure explains about 45 % and the second factor which is related to temperature, explains about 39 percent of the total variance. The three factors together explain about 97 percent of the total variance. Thus, in a nutshell, lake evaporation is a combined effect of relative humidity (vapour pressure), temperature and wind, with relative humidity (vapour pressure) and temperature being highly significant than the wind. The study is useful in understanding the process of lake evaporation in the region and to understand the relative significance of the various factors that affect the process of lake evaporation in a particular climatic setting.

ACKNOWLEDGEMENT

The Authors are very much grateful to the Dr. R. C. Purohit, Head, Department of Soil and Water Engineering, College of Technology and Engineering, Maharana Pratap University of Agriculture and Technology, Udaipur, for supplying the required meteorological data of Udaipur.

REFERENCES

1. Bhatia, K.K.S., Singh, Omkar and S. D. Khobragade. 2005. Prioritization of water quality parameters for management of a typical lake in South India. Urban Lakes in India: Conservation, Management and Rejuvenation. Eds. K.K.S. Bhatia and S.D.Khobragade. NIH, Roorkee publication, pp 376-382.
2. Davis, J.C. 1986. Statistics and data analysis in geology (2nd Ed.). John Wiley & Sons, New York.
3. Deverel, S. J. 1989. Geo-statistical and principal components analysis of ground water chemistry and soil salinity data, San Joaquin Valley, California. Proc. of Baltimore Symposium, May 1989. Regional Characterisation of Water Quality. IAHS Publ. 182, pp. 11-18.
4. Diaz, G., Sewell, J. I. and C. H. Shelton. 1968. An application of principal component and factor analyses in the study of water yield. Water Re. Research. 4(2), pp. 29-306.
5. Iyenger, R.N. 1991. Application o principal component analysis to understand variability of rainfall. Proc. Ind. Acad. Sci. 100(2), pp. 105-126.
6. Khobragade, S. D. 1996. Major and important lakes of Rajasthan: status of hydrological research, SR-45. 1995-96. NIH, Roorkee.
7. Melloul, A. and M. Collin. 1992. The Principal Components statistical method as a complementary approach to geochemical methods in water quality factor identification, application to the Coastal Plain aquifer of Israel. Elsevier Science Publishers, Amsterdam. pp. 49-73.
8. Mohan, S. and N. Arumugam. 1996. Relative importance of meteorological variables in evapotranspiration: factor analysis approach. Water Res. Mgt.10, pp. 1-20.
9. Morin, G, Fortin, J.P., Sochanska, W., Lardeau, J.P. and R. Charbonneau. 1979. Use of principal component analysis to identify homogeneous precipitation stations for optimal interpolation. Water Res. Res. 15(6), pp. 1841-1850.
10. Rao, A.R. 1990. Empirical orthogonal function analysis of rainfall and runoff series. Wat. Res. Mgt. 4, pp. 235-250.

11. Pandzic, K. and D. Tminic. 1992. Principal component analysis of river basin discharge and precipitation anomaly fields associated with the global circulation. Jou. of Hydrology. 132, pp. 343-360.
12. Rao, S.V.L. 2001 Principal components of ground water quality in Venkatagiri taluq, Nellore district, A.P., India. Hydrology Journal. 24 (3), pp. 49-54.
13. Vajrappa, H. C. and G. Srinivas. 1994. Hydro-geochemistry of Kabini river basin in Karnataka.: Regional Workshop on Environ. Aspects of Ground Water Dev., Oct. 17-19, 1994. Department of Geology, Kurukshetra University, Ku.