Scientific Research and Essays Vol. 6(15), pp. 3172-3181, 11 August, 2011
Available online at http://www.academicjournals.org/SRE
DOI: 10.5897/SRE11.071
ISSN 1992-2248 ©2011 Academic Journals

Full Length Research Paper

Some application of principal component analysis on Malaysian wind data

F. Hussain, Y. Z. Zubairi and A. G. Hussin*

Centre for Foundation Studies in Science, University of Malaya, 50603 Kuala Lumpur, Malaysia.

Accepted 30 June, 2011

Understanding the meteorological characteristics helps in predicting the weather conditions; for example, the open burning by local farmers in the South East Asia caused adverse weather conditions in which hazardous haze affect the health conditions of the population. By looking at the multivariate weather variables such as wind speed, relative humidity, pressure, temperature at dew, temperature at dry, geo-potential meter, height above mean sea level and location, the dimensionality of the data is reduced to give a simpler understanding of the data. A Matlab program is written to perform the principal component analysis. Using diagrammatical outputs from scree plot, biplot, three dimensional scatter plot and loading plot, it is found that six components are needed to represent about 83% of the total variance of all components in the multivariate datasets obtained at the Kuala Lumpur International Airport and Bayan Lepas Airport at three different pressures. For the Bayan Lepas Airport Station, we found some modest negative correlation between the geo-potential variables. The components can be described as the variation of geo-potential at all levels, relative humidity at all levels and variables at 1000 level.

Key words: Principle component analysis, wind data, MATLAB, multivariate variables.

INTRODUCTION

Demand for energy from renewable energy resources has become the popular approach in response to the growing concern of the impact to the global climate change. Wind, for example, is a potential energy source as it does not exploit the earth's natural resources and is an environmental friendly approach on harnessing energy. Countries such as Denmark, Germany and United Kingdom have been successful in promoting innovation in wind energy with the use of wind turbine technology for electricity. In promoting the potential of wind energy, it is essential that in the research and development stage, statistical information of the existing wind characteristics is obtained. Such information is useful in evaluating the potential of source of energy.

In Malaysia, wind energy is one of the many renewable energy sources considered (Ong et al., 2011; Sopian et al., 2005). In fact, studies on green power technology from wind started as early as the 1980s when a Solar Energy Research Group from University Kebangsaan Malaysia (UKM) collected wind data from ten stations distributed all over Malaysia. The some success has been obtained with the use of a 150 kW wind turbine in an isolated coral reef island, northeast of Sabah's capital, Kota Kinabalu. However, the potential of wind energy generation largely depends on the availability of wind resource. It has been reported that wind resource varies from one location to another (Ong et al., 2011; Sopian et al., 2005). Therefore, as pointed out by Sopian et al. (1995), it is crucial to obtain a detailed knowledge nature of wind before embarking a wind energy project.

Statistical studies on the behavior of wind in Malaysia include the second Markov chain model to produce time series wind speed using data at two meteorological stations (Shamshad et al.,2005) and identification of wind characteristics (Lee,1993). Zaharim et al. (2008, 2009a) consider Rayleigh, Gamma and Weibull distributions to describe wind data obtained at two locations in Malaysia. Additionally, Hassan et al. (2009) looked at the wind direction of Malaysian. Other statistical works include Zaharim et al. (2009b) and Kamisan et al. (2010).

The data obtained from the Meteorological Services Department comprise of a large dataset where readings recorded over the years comprise of numerous variables related to the wind behavior. One of the common

***Corresponding author. E-mail: ghapor@um.edu.my.**

statistical approaches in analyzing the multivariate data is to obtain a simplified form of the data with few variables that captures the structure of the whole dataset. One classical approach in statistics is by using the principal component analysis method. These principal components, in turn, may be used in the subsequent statistical analyses.

Principal component analysis (PCA), also known as empirical orthogonal function (EOF) analysis, essentially is used to reduce the dimensionality of large dataset which consists of a large number of interrelated variables to smaller components (Jolliffe,1986). In other words, PCA is a one of the statistical technique used on multivariate linear data whereby data transformation is applied in search of the relationships in multivariate data sets. Thus, the aim of PCA is to determine a few linear combinations of the original variables in order to summarize the data. Examples of the application of PCA include the classification of vegetable oils (Rusak et al., 2003), identifying the sources of dimensional variation in the automotive body industry, modeling meteorological data (Mohanan, 2000), visualization of trace elemental pattern in vegetable after different cooking procedure (Pradova, 2001) and many more.

The meteorological data used in this study includes daily recordings of wind speed, relative humidity, temperature at dew and dry point and geo-potential readings collected at three different heights of the observatory stations located throughout the peninsular of Malaysia. In short, the data obtained are of multivariate in nature. The purpose of the study is to design a statistical analysis for the multivariate wind data using principal component analysis which focuses on the graphical plots that can be obtained from the analysis. Using simple diagrammatical outputs written using Matlab, the study aims to describe the variations of the multivariate data by reducing the dimensionality of the principal components.

In the study, the statistical model used in the PCA is described. Based on the eigenvalues obtained in the analysis, the significant components are identified. The numerical results obtained in the analysis are presented in graphical form in which meaningful interpretations can be made from the numerical outputs.

MATERIALS AND METHODS

Principle component analysis

Principal component analysis is one of the methods that can be used to analyse multivariate dataset. It can reduce the dimensionality of large dataset which consists of a number of interrelated variables to smaller components.

Essentially, the steps involved in the analysis of PCA include the method of getting the data, standardizing the data, calculating the covariance matrix, calculating the eigenvectors and eigenvalues of the covariance matrix and visualizing the results.

Algebraically, principal components are particular linear combinations of the p random variables $X_1, X_2, ..., X_p$. Geometrically, these linear combinations represent the selection of a new coordinate system obtained by rotating the original system with $X_1, X_2, ..., X_p$ as the coordinate axes. The new axes represent the directions with maximum variability and provide a simpler and more parsimonious description of the covariance structure.

The principal components depend solely on the covariance matrix Σ (or correlation matrix *ρ*) of $X_1, X_2, ..., X_p$. Their development does not require a multivariate normal assumption. On the other hand, principal components derived for multivariate normal populations have useful interpretations in terms of the constant density ellipsoids.

Further discussions of principal component technique can be found in Richard and Dean (2002). The following are the summary of the analysis performed using the Matlab:

Step 1: Get the data

Consider the linear combinations

$$\begin{aligned} Y_1 &= a_1^{'} X = a_{11} X_1 + a_{12} X_2 + ... + a_{1p} X_p \\ Y_2 &= a_2^{'} X = a_{21} X_1 + a_{22} X_2 + ... + a_{2p} X_p \\ &\vdots \qquad\qquad \vdots \\ Y_p &= a_p^{'} X = a_{p1} X_1 + a_{p2} X_2 + ... + a_{pp} X_p \end{aligned} \quad (1)$$

Step 2: Standardize the data

Sometimes it makes sense to compute principal components for raw data. This is appropriate when all the variables are in the same units. Standardizing the data is often preferable when the variables are in different units or when the variance of the different columns is substantial. This can be done by subtracting the mean of each column and dividing by its standard deviation namely,

$Z_i = \frac{(X_i - \mu_i)}{\sqrt{\sigma_{ii}}}, \ i = 1, 2, 3, \ldots p$. In matrix notation, it is given by:

$$Z = (V^{1/2})^{-1}(X - \mu) \quad (2)$$

Where $V^{1/2}$ is the diagonal standard deviation matrix. From this, we obtain mean of *Z* equals to zero, $E(Z) = 0$.

Step 3: Calculate the covariance matrix.

Further, the covariance matrix of Z is calculated using the formula below

$$Cov(Z) = (V^{1/2})^{-1} \Sigma (V^{1/2})^{-1} = \rho \quad (3)$$

Where ρ also known as correlation.

Step 4: Calculate the eigenvectors and eigenvalues of the covariance matrix. The principal components of Z maybe obtained from eigenvectors of the correlation matrix ρ of X. The ith principal component of the standardized variables $Z' = [Z_1, Z_2, ..., Z_p]$ with Cov(Z)= ρ, is given by:

$$Y_i = e_i' Z = e_i' (V^{1/2})^{-1} (X - \mu) \qquad (4)$$

$$i = 1, 2, ..., p$$

The eigenvectors of correlation matrix ρ are also known as principal components coefficient or principal component loadings. Moreover:

$$\sum_{i=1}^{p} Var(Y_i) = \sum_{i=1}^{p} Var(Z_i) = q \qquad (5)$$

And

$$\rho_{Yi,Zk} = e_{ik} \sqrt{\lambda_i} \qquad i, k = 1, 2, ..., p \qquad (6)$$

In this case, $(\lambda_1, e_1), (\lambda_2, e_2), ..., (\lambda_p, e_p)$ are the eigenvalue-eigenvector pairs for ρ, with $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p \geq 0$.

As seen from Equation 5, the total (standardized variables) population variance is simply q, the sum of the diagonal elements of the matrix ρ. Then, the proportion of the total variance explained by the *k*th principal component of *Z* is:

$$\frac{\lambda_k}{q}, \qquad k = 1, 2, ..., p \qquad (7)$$

Where the λ_k's are the eigenvalues of ρ.

In short, PCA consists of finding linear transformations $Y_1, Y_2, ..., Y_p$ of the original variables $X_1, X_2, ..., X_p$, that have the property of being uncorrelated. The y variables are chosen in such a way that y_1 has maximum variance, y_2 has maximum variance to being uncorrelated with y_1, and so on. The transformation is obtained by finding the latent roots and vector of the correlation matrix. The latent roots, arranged in descending order of magnitude, are equal to the variances of the corresponding y-variates, these being principal components. It is also quite often that the first few components account for a large proportion of the total variance of the *x*-variates and may therefore be used to summarize the original data.

Graphical plots of PCA

As mentioned earlier, the utilization of graphical plots in presenting the results of the analysis is described. In this paper, three plots are described in which meaningful interpretation and conclusion can be obtained from such plots. In the study, the principal components coefficient can be visualized using scree plot, biplot and loading plot.

Scree plot

Scree plot is one of the suitable tools for visualizing the percentage of variance represented by every component in analysis using the principal components method. From Equation 6 one could plot the proportion of the total variance explained by the k-th principal component of Z. From this plot, one can analyse how many components explain a certain percentage of variation. Normally, the main components are represented by components that give an approximation of two-third of the total variability in the standardized ratings.

Biplot

The principal components loading can be visualized by two dimensional biplot or three dimensional scatter plot. The two dimensional biplot shows the diagrammatical representation of the first two components and three dimensional scatter plot visualise the first three components.

Loading plot

Loading plot is useful in identifying what variables are represented in each component. In other words, loading plot provide a histogram for loading value for each of the component in the analysis.

The development of principle component in matlab

A program called "principal_components_analysis (data_pca,column_label)" was developed in MATLAB environment to analyse multivariate data using principal component methods and the details of this program is available at the following URL: http://asasi.um.edu.my/download/PCA_program.doc .

The data called "(data_pca,column_label)" have two types of data where "data_pca" is numerical type of data and "column_label" is string type of data. Using the steps mentioned earlier, the numerical data will be standardized as mentioned by Equations 2 and 3 using the following call functions:

```
"stdr = std(data_pca);
sr = data_pca./repmat(stdr,length(data_pca),1);"
```

The instructions produce a standardised data called "sr". The data can be analysed using the principal components method using the call function "[coefs,scores,variances,t2] = princomp(sr);" that produces four numerical outputs, namely "coefs", "scores", "variances" and "t2" respectively. The output "coefs" refers to the principal components coefficients and "scores" refers to the original data mapped into the new coordinate system defined by principal components. The components variance is a vector containing variance explained by the corresponding principal components and 't2' is Hotelling, a statistical measure of the multivariate distance of each observation from the center of the dataset. Three of these four numerical outputs namely, "coefs", "scores" and "variances" are used in the graphical visualizations which produce several plots namely scree plot, biplot, three dimensional scatter plots and loading plot respectively.

From variable "variances", the percentages of variance can be calculated using the instruction "percent_explained = 100*variances/sum(variances)". The "percent_explained" is sorted in descreasing order and be plotted as histogram using the call function "pareto (percent_explained)" to produce a plot called scree plot. To view the principal components coefficients and principal components scores, two dimensional and three dimensional biplot can be used using call functions as follows:

```
"biplot(coefs(:,1:2),'scores',scores(:,1:2),'varlabels',column_label);
```

```
biplot(coefs(:,1:3),'scores',scores(:,1:3),'varlabels',column_label);"
```

Furthermore, loading plot can provide a better visualisation for the user to determine which variables belong to which components. The program below can create eight main variables for the first five main components:

```
"for baris = 1:5
   subplot(5,1, baris);
   loadingplot (coefs(:,baris),column_label);
   baris = baris + 1;
end"
```

However, the number of component to be displayed can be altered in accordance to the needs of the user. In this illustration, the variables of the five principal components are displayed.

RESULTS AND DISCUSSION

As mentioned earlier, the PCA is applied to the wind data obtained from the Malaysian Meteorological Services Department. The data comprise of wind speed, relative humidity, temperature at dew, dry temperature and geo-potential meter which are recorded at three different levels namely at 500, 850 and 1000 hpa respectively.The wind data are obtained for 121 days in Bayan Lepas Airport (January, February, July and August 2005, Time = 1200) and in Kuala Lumpur International Airport (January, February, July and August 2005, Time = 0000). The results of the analyses are given in the form of principal components coefficient, scree plot, biplot, three dimensional scatter plot and loading plot.

Principal components coefficient

The principal components coefficient is the main numerical result of the principal component analysis. It is the eigenvectors of correlation matrix ρ and also known as principal component loadings. Table 1 shows the principal components coefficient value as shown by every column for every component. Each of the fifteen variables produces principal components coefficient values. The first column displays first component followed by the second column for the second component and so on. The principal components coefficient is the main result of the principal component analysis, which gives eigenvectors. It is possible to look at the eigenvectors directly but most researchers would rather analyse the output based on correlation between components, variance and percentage of variance explained.

Scree plot

As mentioned before from Table 1, each column represents the principal components coefficients for each of the fifteen variables used in the analysis. From each column, the variance and percentages of variance of each component can be calculated and is shown in Table 2. In the first column of Table 2, a list of the fifteen components for the fifteen variables is given. The second column gives the variance in the correlation matrix in which it is expressed into fifteen eigenvalues. Each eigenvalue represents the amount of variance that has been captured by one component. From the output, component one gives the highest variance explained followed by component two which gives the second highest variance explained and so on. The second component is formed from the variance remaining after those associated with the first component has been extracted, thus it accounts for the second largest amount of variance. It is worthwhile to note that the principal components coefficient which gives the variance explained for each component gives a value of less than 30% of the variance explained. Therefore, more than one component is needed to describe the variability of the data. In order to obtain a meaningful interpretation of the principal component analysis, we need to reduce to fewer than 15 components. Several approaches can be applied to determine the number of components to retain (Jolliffe, 1986; Jackson, 1993). In this study, we use the common approach in which we retain only the components with eigenvalues of one or more. Therefore, from Column 2, we observed that six components are retained and in Column 3, the percentage of variance explained by each component is given. The cumulative variance as given in column 4 shows that the first six components account for about 83.36% of the total variability in the data.

To provide a better understanding of the results, a scree plot given in Figure 1 shows a histogram of the variance explained for each component and the increasing curve in the diagram represents the cumulative percentage variance of the component. The number of component to be retained can be determined by looking at the point whereby the rate of change in the slope of the polygon drawn by joining the top of the histogram begins to decrease. Alternatively, one can observe the point where the slope of the increasing curve begins to be small. In Figure 1, we noted that the point is where the principal component is about six. It can also be seen that the first six components contributed towards an approximately 83.4% of the total variability in the standardised ratings.

Biplot

Biplot is also commonly used to visualize the principal component coefficients and the principal component scores for each observation in a single plot for the first two components or the first two columns of Table 1. As shown in Figure 2, the two axes represent the first two components namely for component one and component two, respectively. Each of the coordinates from the first two columns of Table 1 is represented by a star and a vector. The direction and length of the vector indicates

Table 1. Principal components coefficient for 15 variables of wind data.

Variables	Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8
geo_500	-0.4174	-0.0761	-0.0317	0.1725	-0.3459	-0.0998	0.1041	0.0286
tempdry_500	-0.2159	-0.0254	0.4658	-0.1290	-0.1068	0.1994	0.0777	-0.0628
tempddew_500	0.3412	-0.0434	-0.4303	-0.0522	-0.1533	-0.1117	0.0297	-0.1596
hum_500	0.3205	-0.1440	-0.4825	0.0213	-0.1393	-0.0940	0.0302	-0.0514
spd_500	-0.0978	0.0836	0.0415	-0.0547	0.4019	-0.6472	0.6220	-0.0054
geo_850	-0.4080	-0.1267	-0.2760	0.2492	-0.1789	0.0301	0.1248	0.1332
tempdry_850	0.0073	0.3994	0.0482	0.3587	-0.2513	-0.2601	-0.1912	0.0232
tempddew_850	-0.0868	-0.4419	0.0793	-0.2641	-0.2136	-0.2859	-0.1628	-0.1641
hum_850	-0.0676	-0.4865	0.0673	-0.3344	-0.0834	-0.1380	-0.0551	-0.1523
spd_850	-0.2692	0.0194	-0.0915	0.2917	0.3440	-0.0835	-0.3301	-0.7579
geo_1000	-0.3507	-0.2454	-0.3092	0.2237	-0.0202	0.1670	0.2055	0.1285
tempdry_1000	-0.1443	0.3200	0.0036	-0.2266	-0.4418	-0.3905	-0.1740	-0.0172
tempddew_1000	0.2152	-0.2522	0.2870	0.4311	-0.1355	-0.3157	-0.1269	0.1403
hum_1000	0.2375	-0.3575	0.2245	0.4255	0.0885	-0.0704	-0.0326	0.1146
spd_1000	-0.2298	-0.0461	-0.1902	-0.1472	0.4203	-0.2209	-0.5605	0.5242

Variables	Column 9	Column 10	Column 11	Column 12	Column 13	Column 14	Column 15
geo_500	0.0424	-0.3801	-0.6952	0.1353	0.0054	-0.0019	0.0036
tempdry_500	0.7207	-0.2115	0.2906	0.0944	0.0092	0.0037	-0.0189
tempddew_500	0.4274	-0.1893	-0.0969	-0.6344	0.0161	-0.0299	0.0084
hum_500	0.2284	-0.0845	0.1367	0.7264	-0.0291	0.0606	-0.0362
spd_500	0.0656	-0.0569	0.0585	0.0172	0.0047	0.0098	0.0025
geo_850	0.0026	0.1778	0.3086	-0.1681	-0.1286	0.6340	-0.1983
tempdry_850	-0.1993	-0.4958	0.4233	-0.0418	-0.2120	-0.1674	0.0458
tempddew_850	-0.2485	-0.2032	0.2393	-0.0439	0.6115	0.0609	-0.0224
hum_850	-0.1373	-0.0356	0.0518	-0.0435	-0.7413	-0.1253	0.0177
spd_850	0.1108	0.0929	-0.0379	0.0511	-0.0027	0.0059	-0.0011
geo_1000	0.0338	0.1639	0.2351	-0.0528	0.1160	-0.6624	0.2141
tempdry_1000	0.1579	0.5366	-0.0288	0.0446	-0.0033	-0.0192	0.3620
tempddew_1000	0.1563	0.3188	-0.0799	-0.0248	0.0062	-0.2093	-0.5395
hum_1000	0.0662	-0.0118	-0.0335	-0.0243	-0.0149	0.2509	0.6986
spd_1000	0.2285	-0.1392	-0.0346	0.0005	-0.0097	-0.0021	-0.0074

Key: geo_500 (850) (1000) = geo-potential at 500 hp (850 hp) (1000 hp).vtempdry_500 (850) (1000) = Dry Temperature at dry at 500hp (850 hp) (1000 hp). hum_500 (850) (1000) = humidity at 500 hp (850 hp) (1000 hp). spd_500 (850) (1000) = wind speed at 500 hp (850 hp) (1000 hp). tempddew_500 (850) (1000) = temperature at dew at a 500 hp (850 hp) (1000 hp).

Table 2. Variance explained for principal components coefficients for 15 variables of wind data.

Components	Variances	Percentages of variance explained	Cumulative percentage of variance
1	3.7502	25.0012	25.0012
2	3.1817	21.2115	46.2127
3	1.8972	12.6480	58.8607
4	1.4205	9.4699	68.3306
5	1.1626	7.7509	76.0815
6	1.0919	7.2792	83.3607
7	0.7370	4.9130	88.2737
8	0.6172	4.1147	92.3884
9	0.5695	3.7968	96.1852
10	0.3195	2.1302	98.3154
11	0.1596	1.0640	99.3794

Table 2. Contd.

12	0.0698	0.4657	99.8451
13	0.0093	0.0619	99.907
14	0.0089	0.0591	99.9661
15	0.0051	0.0338	99.9999

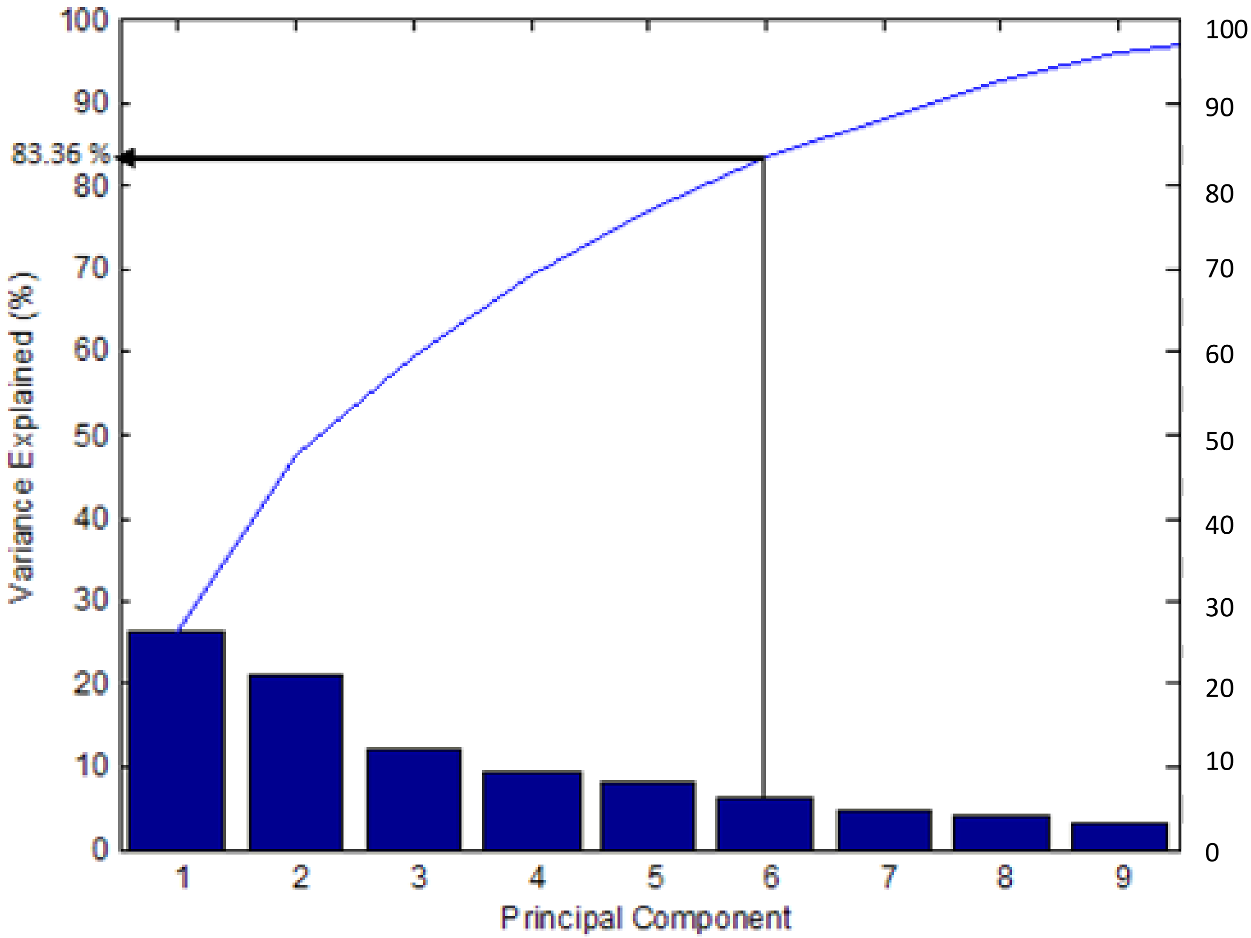

Figure 1. Scree plot for 15 variables of wind data.

how each variable contributes towards the two principal components in the plot. The dot shows the coordinates of the first two components score for all fifteen variables. According to the percentages of total variance explained in Table 2, the first two components only represent 46.21% of the cumulative variance of this analysis.

Three-dimensional scatter plot

Alternatively, a three-dimensional scatter plot or three-dimensional biplot as shown in Figure 3 can be used to visualize the first three principal component coefficients for each variable and the principal component scores for each observation in a three dimensional single plot. This plot can be useful if the first two components do not explain enough of the variance in the data.

This plot has three axes which represent the first three components. The star and the vectors represent the coordinates of first three columns in Table 1 and the dot shows the coordinates of the first three components score for all the fifteen variables. Similar to the previous two dimensional biplot, the three dimensional biplot represents the first three components and only accounts for about 58.86% of the cumulative variance of this analysis and this is considered insufficient.

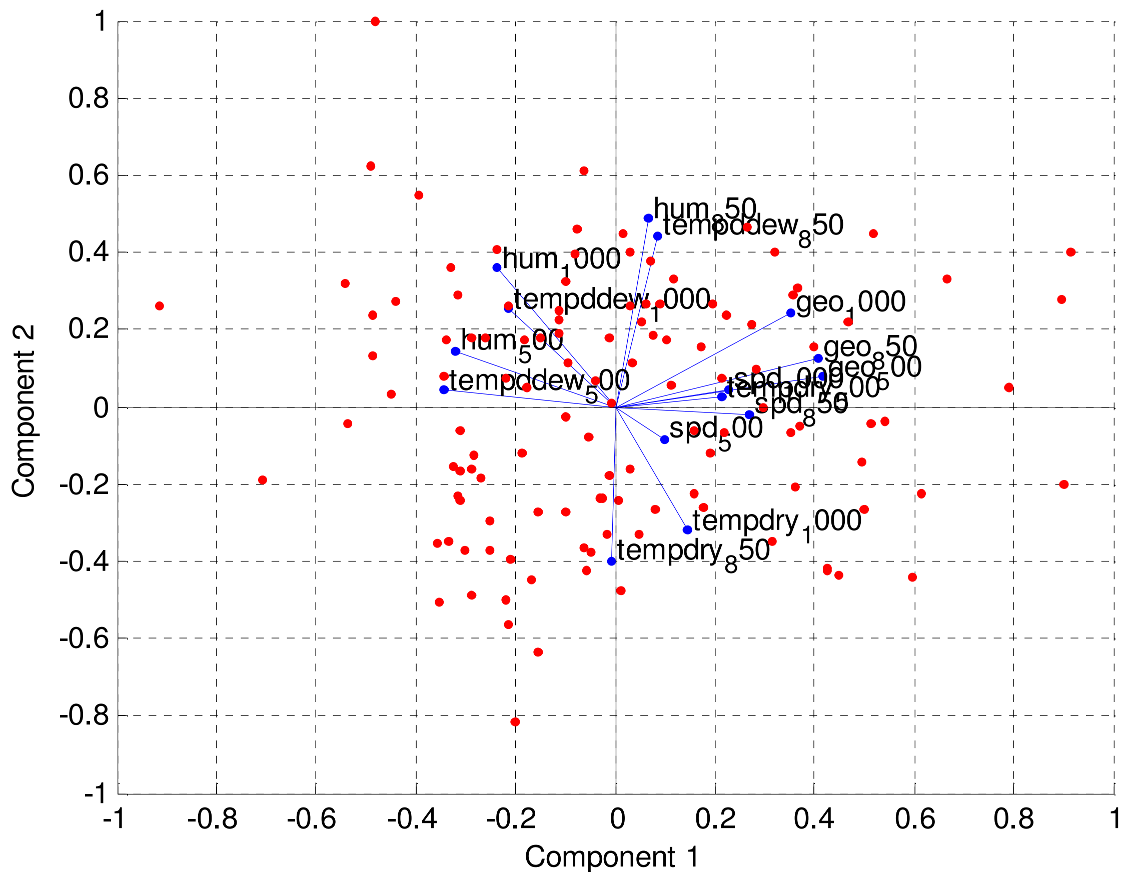

Figure 2. Biplot plot for 15 variables of wind data.

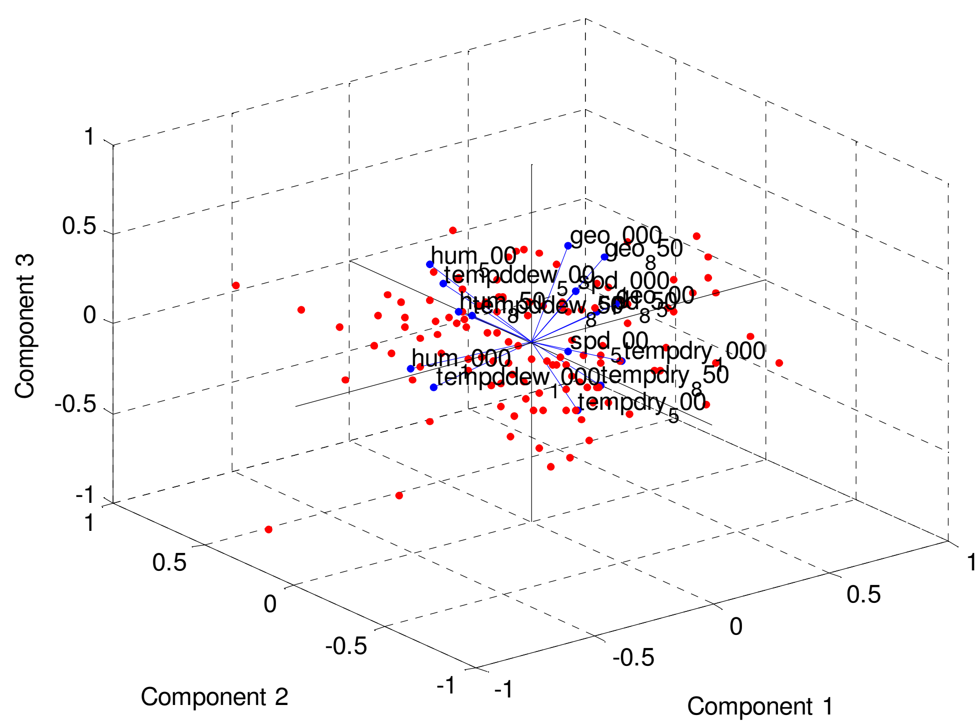

Figure 3. Three-dimensional scatter plot for 15 variables of wind data.

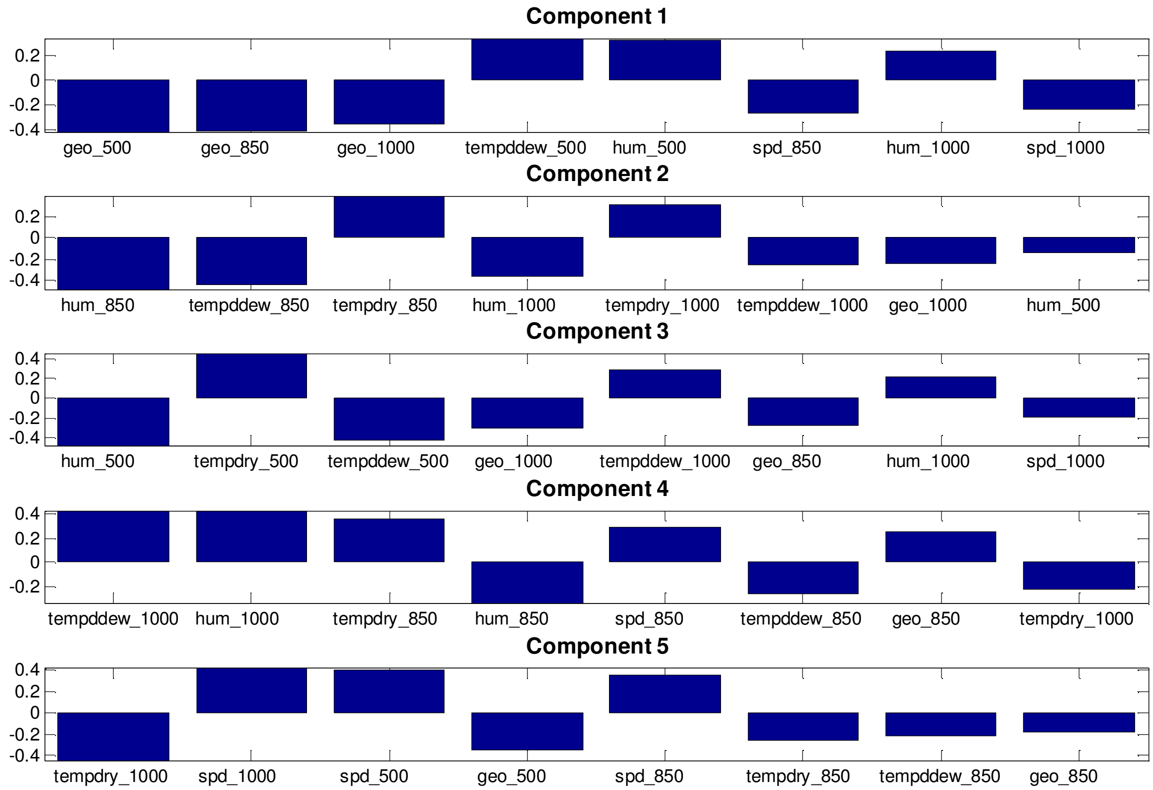

Figure 4. Loading plot for 15 variables wind data.

Table 3. Five main components 15 variables of wind data.

Component 1	Component 2	Component 3	Components 4	Components 5
Geo_ (500)	Hum_ (850)	Hum_(500)	Tempddew_ (1000)	Tempdry_ (1000)
Geo_ (850)	Tempddew_ (850)	Tempdry_(500)	Hum_ (1000)	Spd_ (1000)
Geo_ (1000)	Tempdry_ (850)	Tempddew_ (500)	Tempdry_ (850)	Spd_ (500)
Tempddew_ (500)	Hum_ (1000)	Geo_ (1000)	Hum_ (850)	Geo_ (500)
Hum_ (500)	Tempdry_ (1000)	Tempddew_ (1000)	Spd_ (850)	Spd_ (850)
Spd_ (850)	Tempddew_ (1000)	Geo_ (850)	Tempddew_ (850)	Tempdry_ (850)
Hum_ (1000)	Geo_ (1000)	Hum_ (1000)	Geo_ (850)	Tempddew_ (850)
Spd_ (1000)	Hum_ (500)	Spd_ (1000)	Tempdry_ (1000)	Geo_ (850)

Loading plot

Loading plot as shown in Figure 4 displays how each variable contributes towards the loading of each component. For the first five principal components coefficient in Table 1, the first eight largest magnitudes of the coefficient have been rearranged in decreasing order in the form of histograms. These five histograms show five main components of which the first row shows Component 1, the second row shows Component 2 and so on.

In other words, the variances with largest magnitude listed in Table 3 are plotted as histograms that give coefficient values of the variable as shown in Figure 4. It can be seen from loading values of Component 1, three variables with largest coefficient are geo_500(-0.4174), geo_850(-0.4080) and geo_1000(0.3507) respectively. From the plot, one can identify the specific variable of the main components which cannot be easily identified in the biplot and three dimensional biplot. Perhaps, in further analysis such as predicting the weather, the independent variable which includes the contribution of the wind factor into the model building could take into account the significant variables that have been identified in the PCA, such as the geo-potential reading at different levels. As a comparison, this method has been applied to the wind

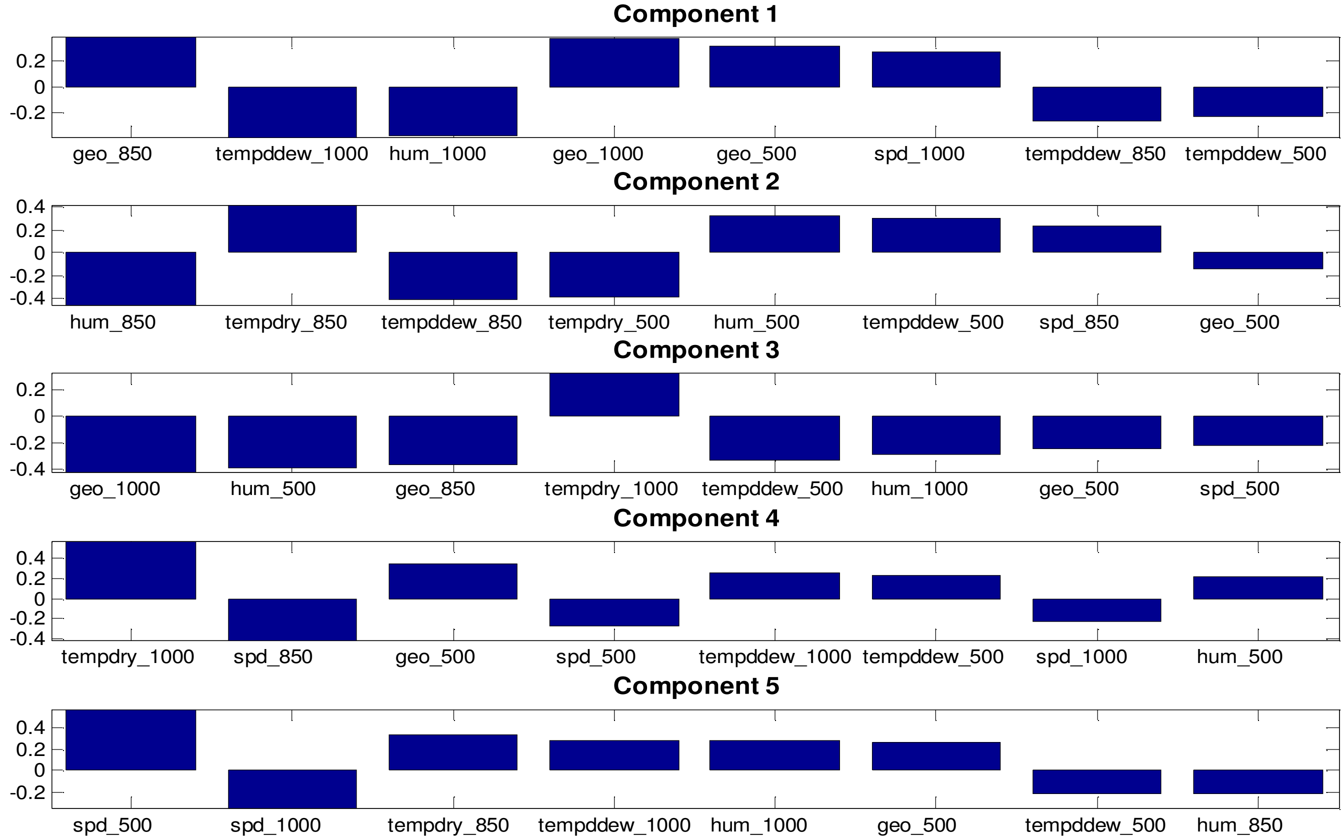

Figure 5. Loading plot for 15 variables of wind data for KLIA.

data for 121 days in Kuala Lumpur International Airport (KLIA) (January, February, July and August 2005, Time = 0000). The data also contained 15 variables measured at three different pressures of 500, 850 and 1000 hpa respectively. After calculating the cumulative percentages of variances, five components are needed to represent more than 77.0% of the total variability in the standardized ratings. The loading plot is shown in Figure 5. It can be seen from both Figures 4 and 5, that some of the variables appear as the main components, which in turn suggest some similarity between the two datasets.

From the developed scree plots, biplots and loading plots, the results show that six components are needed to represent about 83% of the total variance of all components in the analysis on datasets at Kuala Lumpur International Airport and for the dataset of Bayan Lepas Airport, five components represents about 77.0% of the total variation. For the Bayan Lepas Airport Station, we found some modest negative correlation between the variables namely the geo-potential variables. The components can be described as the variation of geo-potential at all levels, relative humidity at all levels and variables at 1000 level. As mentioned earlier, these results may have some important implications for a better understanding on the weather of some places in Malaysia.

Conclusion

This article describes the development of graphical plots for the principal component analysis using the MATLAB environment. It is shown that using several plots, a better explanation or interpretation can be obtained in comparison to the usual numerical output of the principal components. The coefficients of the principal components are displayed using scree plot, biplot, three-dimensional scatter plot and loading plot respectively. It is found that the first two principal components can be described as the variation of geo-potential values at all levels and relative humidity with temperature values at 850 and 1000 level respectively.

REFERENCES

Hassan SF, Hussin AG, Zubairi YZ (2009). Analysis of Malaysian Wind Direction Data, Modern Appl. Sci., 3(3): 115-119.

Jackson DA (1993). Stopping Rules in Principal Component Analysis: A Comparison of Heuristical and Statistical Approaches. Ecology, 74(8): 2204-2214.

Jolliffe IT (1986). Principal Component Analysis. Springer-Verlag, pp. 1-6.

Kamisan NAB, Hussin AG, Zubairi YZ (2010). Finding the Best Circular Distribution for Southwesterly Monsoon Wind Direction in Malaysia, Sains Malaysiana. 39(3): 387-393.

Lee ET (1993). Wind Power Potential in Malaysia, Energy Sources, Part A:Recovery, Utilization, and Enviromental Effects, 15(1): 23-26.
Mohanan AH (2000). Nonlinear Principal Component Analysis by Neural Networks: Theory and Application to the Lorenz System. J. Clim., 13(4): 821-835.
Ong HC, Mahlia TMI, Masjuki HH (2011). A review on energy scenario and sustainable energy in Malaysia, Renewable Sustain. Ener. Rev., 15: 639-647.
Pradova V (2001). Three-way Principal Component Analysis for the Visualization of Trace Element Patterns in Vegetables after Different Cooking Procedures. J. Food Compost. Anal., 14(2): 207-225.
Richard AJ, Dean WW (2002). Applied Multivariate Statistical Analysis. Pearson Education International, pp. 426-475.
Rusak DA, Brown L, Martin SD (2003). Classification of Vegetable Oils by Principal Component Analysis of FTIR Spectra. J. Chem. Educ., 80(5): 541-548.
Shamshad A, Bawadi MA, Wan Hussin WMA, Majid TA, Sanusi SAM (2005). First and second order Markov chain models for synthetic generation of wind speed time series, Energy, 30: 693-708.
Sopian K, Othman MY, Wirsat A (1995). The Wind Energy Potential of Malaysia. Renewable Energy, 6(8): 1005-1016.
Sopian K, Othman MY, Yatim B, Daud WRW (2005). Future Directions in Malaysian Environment Friendly Renewable Energy Technologies Research and Development, Science and technology Vision, 1(5): 30-36.
Zaharim A, Najid SK, Razali AM, Sopian K (2008). The Suitability of Statistical Distribution in Fitting Wind Speed, WSEAS Transaction on Mathematics, 7(12): 718-727.
Zaharim A, Najid SK, Razali AM, Sopian K (2009a). Wind Speed Analysis in the East Coast of Malaysia. Eur. J. Sci. Res., 3(2): 208-215.
Zaharim A, Najid SK, Razali AM, Sopian K (2009b). Analyzing Malaysian Wind Speed Data Using Statistical Distribution, Proceedings of the 4^{th} IASME/WSEAS International Conference on Energy and Environment, pp. 360-370.