GEOSPATIAL WORLD FORUM
Theme
Dimensions and Directions of Geospatial Industry
18 - 21 January 2011, Hyderabad, India

Title: Delineation of Climate Divisions for Peninsular Malaysia

Fariza Yunus, Meteorological Officer, Malaysian Meteorological Department, Malaysia

Abstract

The climate divisions in Peninsular Malaysia were delineated by integrating in-situ and Geographical Information System (GIS) raster data. Principal component (PC) analysis was applied to long-term mean monthly temperature elements data for monsoon seasons. The first three principal components were chosen to be statistically significant, accounted for 96.51% of the variability in the 27 variables. These three components are related to mean monthly variation in minimum temperature during monsoon seasons (first PC), mean monthly variation in maximum and mean temperature in southwest monsoon (second PC) and mean monthly variation in maximum temperature during northeast monsoon (third PC). Cluster analyses were applied to create cluster of meteorological stations, where six clusters were formed. GIS raster data of factor scores were utilized to place the cluster borders, where interpolation analysis was applied to generate these GIS raster data. The result of a maximum likelihood classification produced three clusters when summarized by 82 areas (districts). The resultant climate divisions showed rational climate regionalization that reveals controls on temperature. The use of factor score GIS raster data effectively assisted the delineation of meteorological station clusters grouped using only in-situ data.

1.0 Introduction

Generally, climate division is referring to areas which has similar characteristic of climatic elements (AMS, 2010). The most referred example of climate division is Thorthwaite (1931) where a variety of climate elements was used. However, most of standard approaches are directly based on temperature and precipitation (Rhee *et al.*, 2008, Paruelo et al., 1995, Fovel and Fovel, 1993, DeGaetano, 2001). These approaches frequently employ a combination of principal component analysis (PCA) and cluster analysis (DeGaetano, 1996; Gong and Richman, 1995). Many studies (DeGaetano, 1996, Fovell 1997, Gerstengarbe *et al.*, 1999, Rhee *et al.*, 2008) used a combination of hierarchical and nonhierarchical clustering analysis to overcome the unrealistic assumption of hierarchical clustering analysis (Hair *et al.*, 2009). Compared to other countries, limited studies were carried in delineating climate divisions for Peninsular Malaysia. The existing climate division which was delineated by Dale (1959) used only precipitation data to divide the area into five divisions (Tick and Abu Samah, 2004). Therefore classification of Peninsula Malaysia using temperature data has to be generated. This paper presents an approach to classify Peninsula Malaysia based on temperature data towards climate homogenous divisions.

Most of the previous studies (DeGaetano, 2001; Fovel, 1997) defined climate homogenous divisions based on meteorological stations data only. In Peninsular Malaysia, most of meteorological stations located at low lying areas and by using in-situ data only, it is impossible to cluster hilly areas with very limited meteorological station present. Furthermore it may introduces bias since the uneven distribution of meteorological stations creates a problem in determining borderlines among groups of stations and in identifying climate characteristic of divisions with low density of meteorological stations (DeGaetano, 1996; Rhee *et al.*, 2008). To place the cluster border, DeGaetano (1996) used a discriminant function analysis by linking a group of stations with locations variables. This technique may also introduce bias as the variables are only indirectly related to climate variables. Rhee *et al*. (2008) used remotely sensed data of monthly land surface temperature in determining climate cluster border. Although surface air temperature and land surface temperature are highly correlated, its measure different entities and land surface temperature is not a climate variable (Thorthwaite, 1931). Furthermore, different period for both types of data were utilized in the study. GIS raster data of factor scores may be useful to solve this problem. Although these factor scores are only arbitrary values, it is important information since these values measured the temperature elements. Furthermore, this study used the same set of PC data, for clustering meteorological stations and area of Peninsular Malaysia. The latent bias caused by the sparse and irregular distribution of stations can also be reduced.

2.0 Study Area and In-situ Temperature Elements Data

Peninsular Malaysia is located in south of Thailand, north of Singapore and east of Indonesia island of Sumatra. The main mountain range is Titiwangsa Mountain where the highest point is Mount Tahan (2,187m). Generally, the climate is hot and humid throughout the year with annual mean temperature of 26.5^{o}C and average annual rainfall exceeded 2000mm (Tick and Abu Samah, 2004). Figure 1 shows a network of active meteorological stations in the study areas is in irregular pattern.

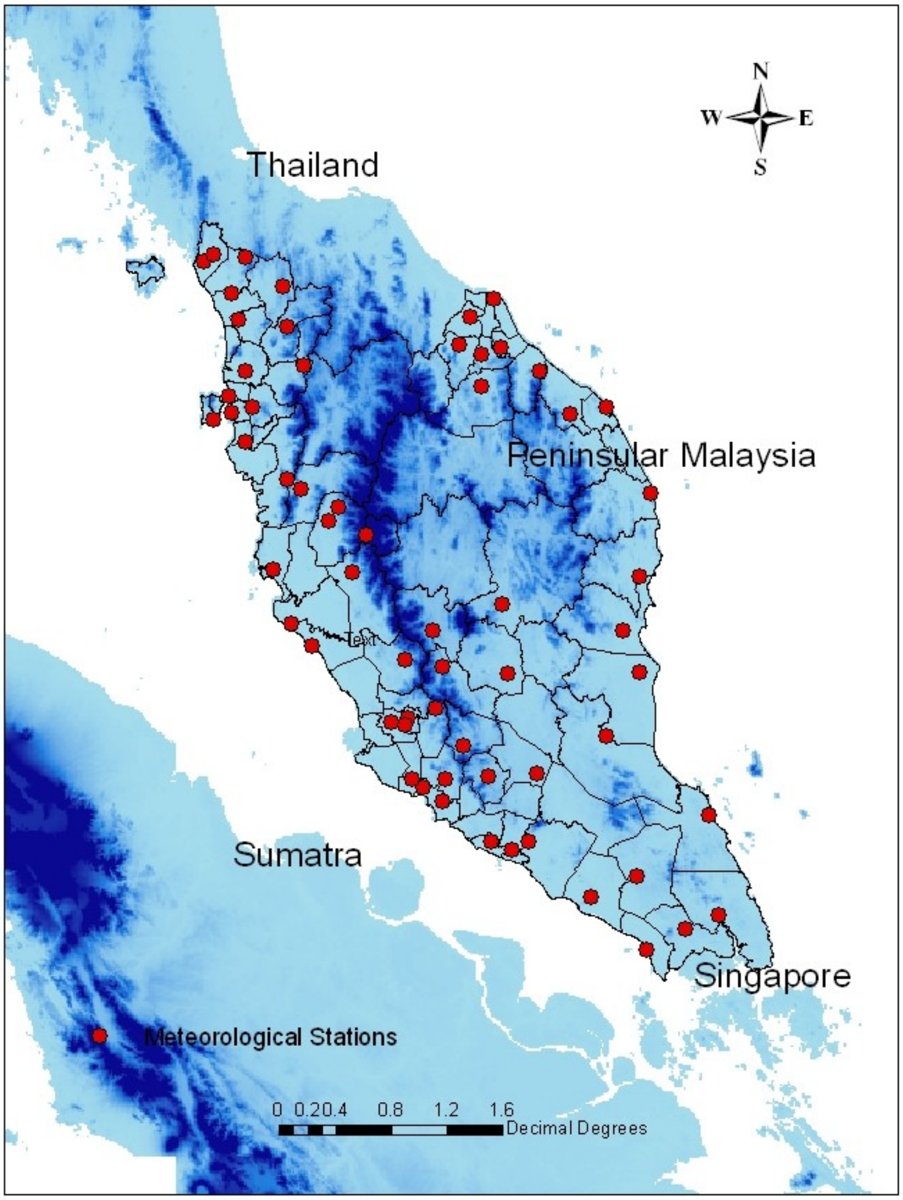

Figure 1: Distribution of meteorological stations in study area of Peninsula Malaysia

Referring to Figure 1, the meteorological stations distribution is relatively dense over the low lying areas where 60 out of 62 meteorological stations are located at below 250m from mean sea level (MSL). Temperature elements data of maximum, minimum and mean values were used in this study. Daily data were obtained for 10 years period (1st January 1999 to 31st December 2008). All stations had sufficient data where 55 stations (89%) had at least 80% of completed daily data (Malmgren, 1999). These data underwent quality control at Climate Division, Malaysian Meteorological Department (MMD). Since the seasonal and spatial temperature variations are relatively small and there is a definite variation during the monsoons seasons (MMD, 2010), this study used temperature data during the monsoon seasons. There are two monsoon seasons in Peninsular Malaysia (Tick and Abu Samah, 2004), namely southwest monsoon which occur in May to September (five months) and northeast monsoon occur during November to February (four months). The 10-year monthly means per station for these selected months were computed over the years. Therefore the analysis was based on a total of 27 variables which arises from three temperature elements of nine months.

3.0 Methodology

Based on long-term monthly average of temperature elements for monsoon seasons, climatologically homogenous divisions were generated using a multi-step approach (Rhee *et al*, 2008). Firstly, a principal component analysis (PCA) was performed (DeGaetano, 1996; Fovell and Fovell, 1993) among the 27 variables. Although the use of long-term means data and PCA will forfeit some information (Rhee *et al*., 2008), the effect is not significant for Peninsular Malaysia areas since the temperature is almost uniform with annual variation of less than 3^{o}C (MMD, 2010). The scree plot and latent root criterion was used to select the appropriate PC (Hair *et al*., 2009). To be statistically significant varimax rotation was employed to the selected PCs (Malmgern, 1999; Hair *et al*., 2009).

Secondly, cluster analysis of hierarchical and nonhierarchical (DeGeatano, 1996; Rhee et al., 2008; Fovell, 1997; Hair *et al*, 2009) were applied to the results of PCA. The most popular hierarchical clustering method of average linkage method (Rhee *et al*., 2008; DeGaetano, 1996) was carried out. The appropriate numbers of clusters were determined by a number prior to the tremendous increase the coefficient values, where the large increase in heterogeneity exists (Hair *et al*, 2009). Initial seeding points for each of cluster were determined. A nonhierarchical analysis was implemented to the same truncated PCA data but using the results of previous hierarchical analysis as number of cluster and cluster seed points. Using these initial cluster seed points, nonhierarchical clustering was conducted by adopting algorithm used by DeGaetano (1995).

Thirdly, generation of GIS raster data for PC truncated results to place the cluster boundary. This GIS raster data was generated using widely used interpolation analysis of inverse distance technique (IDW) (Mitas and Mitasova, 1999; Tomczak, 1998). Fourthly, supervised classification of raster GIS data of factor scores was performed using the results of previous nonhierarchical clustering analysis as training data. A traditional technique of maximum likelihood classification (Walter, 2003; Walter, 1998; Huang and Jensen, 1997) was used as the supervised classification method. In this analysis the data assumed to be normally distributed using central limit theorem since the number of sample is large enough. Finally, climatologically homogenous divisions were delineated by using zonal analysis of district boundaries for Peninsular Malaysia meant for management purposes.

4.0 Results and Discussions

Correlations between the original 27 variables are relatively high, ranging from 0.59 in the case of January mean monthly minimum temperature ($mintt_1$) and February mean monthly maximum temperature ($maxtt_2$) to 0.998 for Jun mean monthly meant temperature ($meantt_6$) and July mean monthly meant temperature ($meantt_7$). In most cases, the correlation values exceeded 0.7 and 34% of the data had correlations of more than 0.9, thus it shows the presence of redundant information. The result of PCA was required to reduce information bias in the final clusters resulting from these redundancies. Figure 2 is scree plot result of PCA.

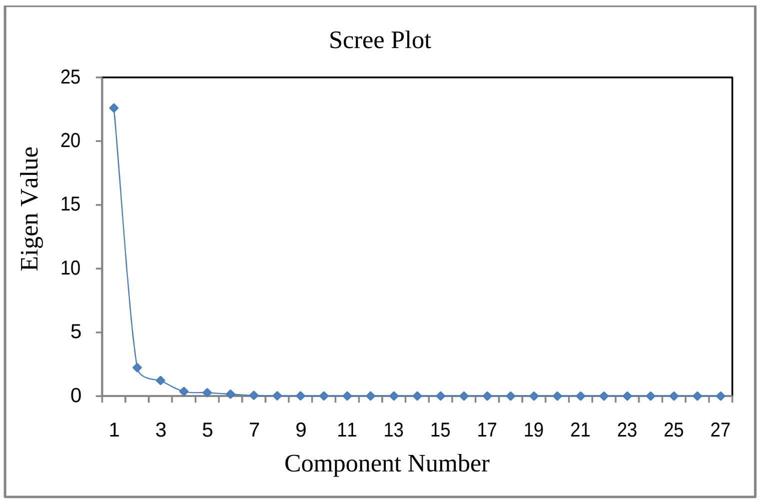

Figure 2: Scree Plot

Referring to Figure 2, scree plot shows component number four and above has very slight change in Eigen values. Furthermore, only the first three PCs had Eigen values of greater than one. Therefore, from the scree plot and latent root criterion the first three PCs were selected to be analyzed in this study.

To enhance the interpretation of selected principal components, varimax orthogonal rotation was applied (DeGaetano, 1996). For each of PC, only the selected variables are considered because according to guidelines by Hair *et al.* (2009), in sample size of 60 cases, factor loadings of 0.7 and above are significant. Table 1 shows factor loading of varimax rotated principal component for three selected PCs, where only variables corresponding to factor loadings of more than 0.7 were considered.

Table 1: Varimax rotated principal component factor loadings for three selected PCs. Only variables corresponding to factor loadings of more than 0.7 were considered. Cumulative variance is given in parenthesis

Principal Component (PC)					
PC1 (83.71%)		PC2 (91.98%)		PC3 (96.51%)	
Var	Load	Var	Load	Var	Load
mintt$_5$	0.89	maxtt$_7$	0.87	maxtt$_1$	0.88
mintt$_6$	0.88	maxtt$_8$	0.86	maxtt$_{12}$	0.83
mintt$_1$	0.88	maxtt$_9$	0.86	maxtt$_2$	0.83
mintt$_2$	0.88	maxtt$_6$	0.85	maxtt$_{11}$	0.72
mintt$_7$	0.88	maxtt$_5$	0.85		
mintt$_8$	0.87	meantt$_9$	0.75		
mintt$_{12}$	0.87	meantt$_7$	0.73		
mintt11	0.86	meantt$_8$	0.73		
mintt$_9$	0.86	meantt$_6$	0.72		
		meantt$_5$	0.70		

The first three selected PCs in Table 1 are accounting for 96.51% of the variation in the data. The first PC (PC1) which explain the most variation in the data (83.712%), contain significant contribution of mean monthly variation in minimum temperature during monsoon seasons (mintt$_{11}$, mintt$_{12}$, mintt$_1$ mintt$_2$ and mintt$_5$, mintt$_6$, mintt$_7$, mintt$_8$ mintt$_9$). The second PC (PC2), which explained 8.27% of variance, is represented by mean monthly variation of maximum (maxtt$_5$, maxtt$_6$, maxtt$_7$, maxtt$_{8,}$ maxtt$_9$) and mean temperature (meantt$_5$, meantt$_6$, meantt$_7$, meantt$_{8,}$ meantt$_9$) in southwest monsoon. The third PC (PC3), accounting for 4.53% variance is related to mean monthly distribution of maximum (maxtt$_{11}$, maxtt$_{12}$, maxtt$_{1,}$ maxtt$_2$) during northeast monsoon.

Average linkage hierarchical clustering method was applied to factor scores of the three PCs. Table 2 shows agglomeration schedule of PC1, PC2 and PC3.

Table 2: Agglomeration schedule for PC1, PC2 and PC3

Stage	Cluster		Coefficients	Number of Cluster After Combining	Increase in Coefficient to Next Stage	Proportionate Increase in Heterogeneity to Next Stage
	Cluster 1	Cluster 2				
50	16	19	1.01	12	0.10	9.90%
51	16	17	1.11	11	0.03	2.70%
52	21	25	1.14	10	0.35	30.70%
53	16	21	1.49	9	0.05	3.36%
54	1	5	1.54	8	0.25	16.23%
55	1	13	1.79	7	0.07	3.91%
56	3	11	1.86	**6**	**1.26**	**67.74%**
57	3	18	3.12	5	2.37	75.96%
58	1	3	5.49	4	0.07	1.28%
59	6	8	5.56	3	0.58	10.43%
60	1	16	6.14	2	25.56	416.29%
61	1	6	31.70	1		
					Average	58.05%

The average proportionate increase for stage 50 to 61 was 58.05%. The largest increase in coefficient occurred when moving to final stage (25.56) where the solution result is two clusters. However this is not an appropriate solution because it involved the largest change in heterogeneity (Hair *et al*., 2009). The second largest increase in coefficient was 2.37 or 75.96% of proportionate increase. However, the different in proportionate increase between the third largest and second largest is not distinct where increase in coefficient for the third largest was 1.26 (67.74% of proportionate increase). This third largest occurred when moved from stage 56 to 57. Furthermore, six-cluster solution was associated with proportionately less heterogeneous than the five-cluster solution. Based on these results, the stopping point for cluster analysis was at stage 56 and six-cluster solution was selected as the appropriate number of cluster.

The locations of meteorological station clusters using six-cluster solution were examined. Figure 3 shows solution of six-cluster for 62 meteorological stations.

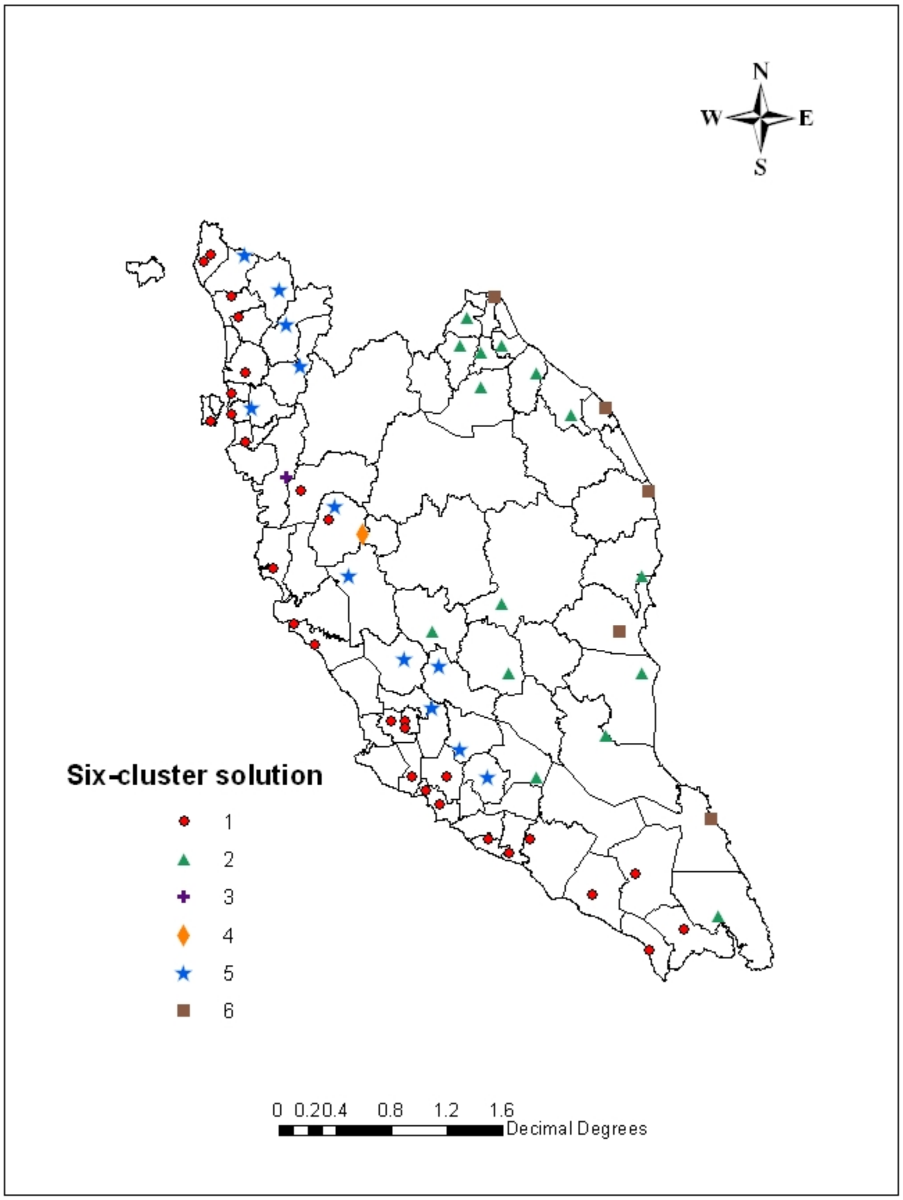

Figure 3: Six-cluster solution of meteorological station

From Figure 3, cluster one consists of 28 stations and located at west coast. Cluster two consists of 15 stations and majority of these stations are located at northeast and central part. Cluster three and cluster four has only a single station and both station are located at hilly areas. Cluster five has 12 stations located at inland areas of west Peninsula. Cluster six has five stations which located at coastal areas of east Peninsula.

Centroids for each of the six clusters were obtained and used as seed points for k-means nonhierarchical clustering. The reassignment of meteorological stations was examined. Most of stations are group in the same cluster as in hierarchical cluster analysis, except Hospital Parit Buntar's station and Hospital Kota Tinggi's station. Both stations were regroup in cluster five in k-means nonhierarchical cluster analysis, where originally Hospital Parit Buntar's station was in cluster one and Hospital Kota Tinggi's station was in cluster two.

Simple linear regression models of discrete data for factor scores and elevation at every station were developed for every PC to estimate lapse rate values for every PC. Gradient values for regression models of -0.003,-0.003 and -0.002 were used to estimate the lapse rate value for PC1, PC2 and PC3 respectively. Factor score at zero elevation at all stations were interpolated using IDW interpolation method to develop surface of factor score at zero elevation for every PC. To generate raster factor scores, the respective lapse rate was timed with digital elevation model (DEM), and this surface was then subtracted to raster data of factor scores at zero elevation.

Results from nonhierarchical cluster analysis were utilized in supervised classification where all 62 meteorological stations were used as training data for factor scores. Supervised classification of a maximum likelihood classification was applied to the corresponding raster data of factor scores. Figure 4 shows result of supervised classification.

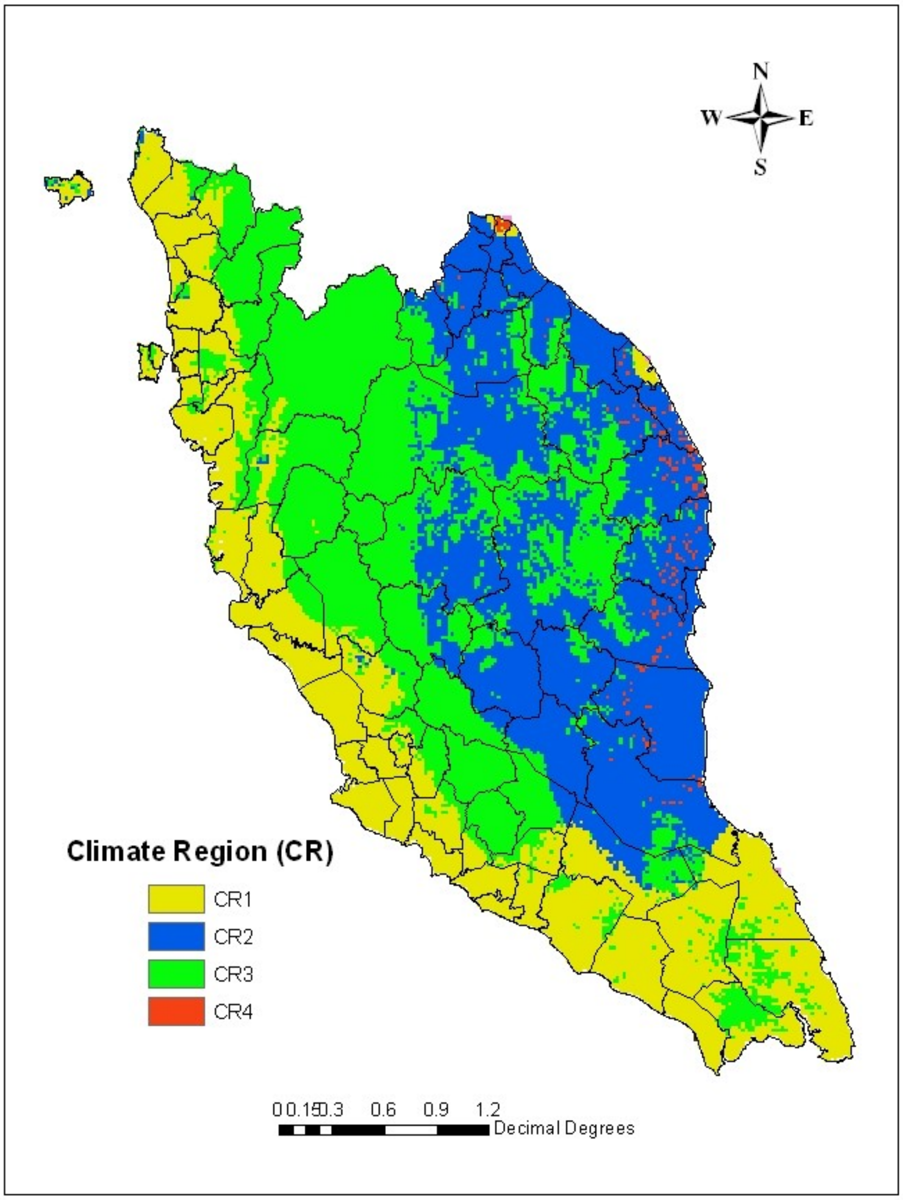

Figure 4: PCs classed based on maximum likelihood classification

Although there were six classes in the training data, classification results only four classes in analysis area (Figure 4). The single station cluster three and cluster four were not shown in classification analysis. These clusters were merged with cluster five, and become climate division three. Cluster six still existed but only in small fraction areas over east part of the Peninsula. The used of GIS raster factor scores data help in identifying climate characteristic of locations without meteorological stations and the delineation between cluster of stations where distribution were not dense. For example, there are areas in inland part of east Peninsular Malaysia without meteorological station, were classified using GIS raster data into climate division three.

82 areas of district boundaries were suggested as a delineation unit, since the country boundaries are too coarse. This delineation is useful for planning and management of climate-related projects as well as in critical decision making. Zonal analyses were performed with a majority filter (Rhee *et al.*, 2008). Figure 5 presents the three classes of Peninsular Malaysia's climate divisions which are corresponding to west coast (CD1), east coast (CD2) and the main range area (CD3).

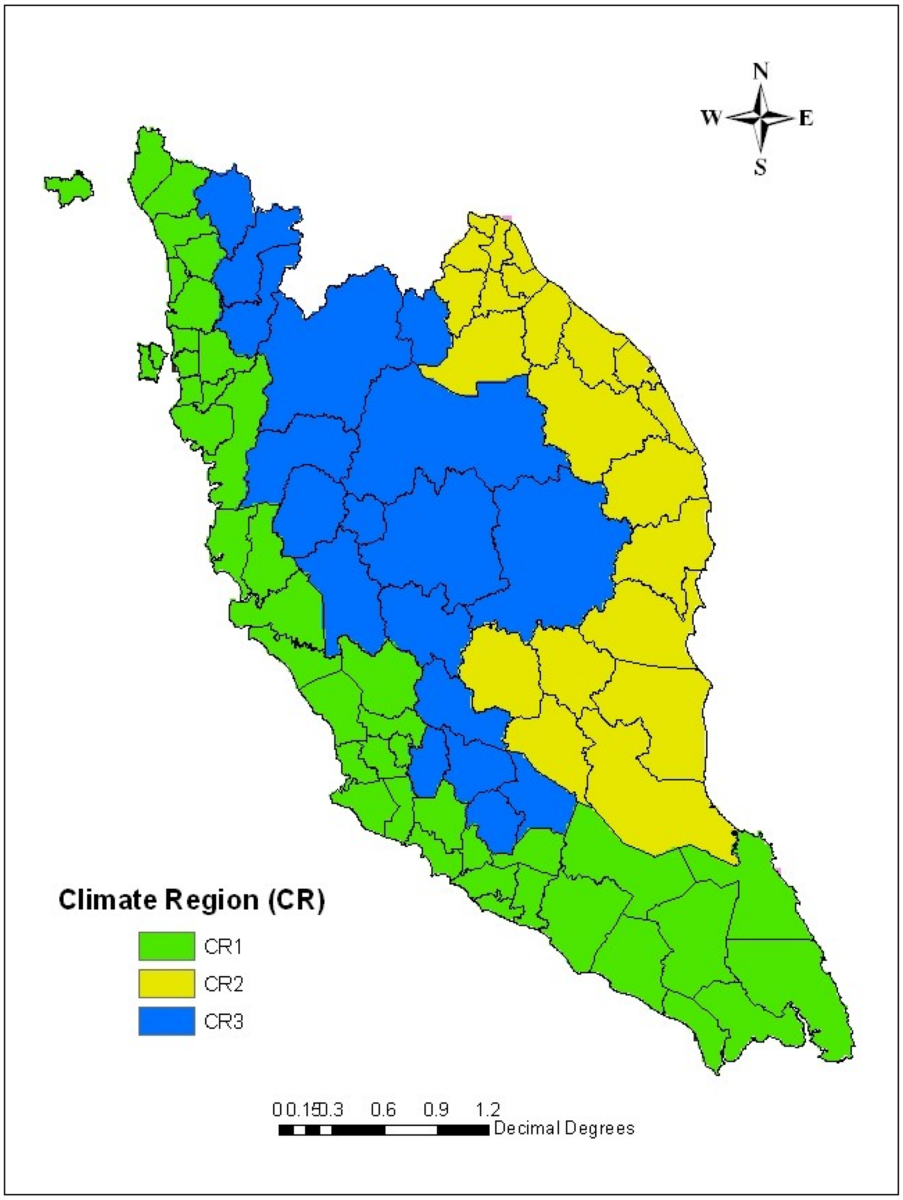

Figure 5: Climate divisions (CD) as a result of the three CDs classes using GIS raster data of factor scores and summarized by district

Since areas of climate division four was very small, this area was eliminated in this output and merged with climate division two (Figure 5). These three areas play an important role in determining temperature distribution pattern in Peninsular Malaysia. Annual variation of temperature are relatively higher in east coast areas compared to other areas since these areas are often affected by cold surges originating from Siberia during the northeast monsoon. Generally, the average daily temperature to the east of the Main Range is lower than that of the west of the Main Range. Obviously the generated climate divisions were conformed to climate control of Peninsular Malaysia, including topography, latitudinal locations and distance from coastal areas.

5.0 Conclusions and Recommendations

Climatologically homogenous divisions in Peninsular Malaysia were delineated by using in-situ temperature data and GIS raster data of factor scores. Resultant climate divisions are consistent with patterns suggested by climate record. The used of GIS raster data assisted the delineation of climate divisions particularly in areas where very limited or no meteorological stations are available. In generating GIS raster data of factor scores, only elevation was used as independent variable. The incorporation of other independent variables that affects temperature distribution, could improve the result. This study used long-term mean monthly temperature elements data during monsoon seasons only. By using a complete long-term monthly data for the whole year may improve the results. Furthermore, only temperature elements data were used in this study. By considering other climatic variables may also improve the final results of climate division. Since two clusters (cluster three and cluster four) in the training data of supervised classification included only a single station, consequently these two clusters could not be presented in climate divisions. The use of more meteorological stations specifically over hilly areas, even with relatively short histories needs to be considered. These climate divisions can be overlaid with existing climate divisions by using consensus clustering suggested by Fovell (1997). Accurate climate divisions can be used in many other applied climate studies in the future.

References

American Meteorological Society, 2010. Glossary of meteorology, Retrieved April 1, 2010, from http://amsglossary.allenpress.com/glossary/search?p=1&query=climatic+classifycation&submit=Search

Dale, W. L., 1959. The rainfall of Malaya part I, Journal of Tropical Geography, 13, 23 – 37.

DeGaetano, A. T., 1996. Delineation of mesoscale climate zones in the Northeastern United States using a novel approach to cluster analysis, Journal of Climate, 9, 1765 – 1782.

DeGaetano, A. T., 2001. Spatial grouping of United States climate stations using a hybrid clustering approach, International Journal of Climate, 21, 791 – 807.

Fovell, R. G., 1997. Consensus clustering of U.S temperature and precipitation data, Journal of Climate, 10, 1405 – 1427.

Fovell, R. G., Fovell, M-Y., 1993. Climate zones of the conterminous United States defined using cluster analysis, Journal of Climate, 6, 2103 – 2135.

Gerstengarbe, F. W., Werner, P. C., Fraedrich, K., 1999. Applying non-hierarchical cluster analysis algorithms to climate classification: some problems and their solution, Theor. Appl. Climatol., 64, 143 – 150.

Gong, X., Richman, M. B., 1995. On the application of cluster analysis to growing season precipitation data in North America East of the Rockies, Journal of Climate, 8, 897 – 931.

Hair, J. F., Black, W. C., Babin, J. B., Anderson, R. E., 2009. Multivariate data analysis, Pearson Education Inc., New Jersey, 91 – 150, 505 – 533.

Huang, X., Jensen, J. R., 1997. A Machine-Learning Approach to Automated Knowledge Base Building for Remote Sensing Image Analysis with GIs Data, Photogrammetric Engineering & Remote Sensing, 63, 1185-1194.

Malmgren, B. A., 1999. Climate zonation in Puerto Rico based on principal component analysis and an artificial neural network, Journal of Climate, 12, 977 – 985.

Mitas, L., Mitasova, H., 1999. Spatial Interpolation, In Longley P A, Goodchild M F, Maguire DJ and Rhind DW, Geographical Information Systems: Principles, Techniques, Management and Applications. New York, Wiley, 481 – 573.

Malaysian Meteorological Department, 2010. General climate of Malaysia, Retrieved April 1, 2010, from http://www.met.gov.my/index.php?option=com_content&task=view&id=75&Itemid=1089

Paruelo, J. M., Lauenroth, W. K., Epstein, H. E., Burke, I. C., Aguiar, M. R., Sala, O. E., 1995. Regional climatic similarities in the temperate zones of North and South America, Journal of Biogeography, 22, 915 – 925.

Rhee, J., Im, J, Carbone, G. J., Jensen, J. R., 2008. Delineation of climate regions using in-situ and remotely-sensed data for the Carolinas, Remote Sensing of Environment, 112, 3099 – 3111.

Thornthwaite, C. W., 1931. The climates of North America: according to a new classification, Geographical Review, 21, 633 – 655.

Lim, J. T., Abu Samah, A., 2004, Weather and climate of Malaysia, Univ. Malaya Press, Malaysia.

Tomczak. M., 1998. Spatial Interpolation and its Uncertainty Using Automated Anisotropic Inverse Distance Weighting (IDW) - Cross-Validation/Jackknife Approach, Journal of Geographic Information and Decision Analysis, 2, 2, 18 – 30.

Walter, V., 1998. Automatic classification of remote sensing data for GIS database revision, IAPRS, 32, 641 – 648.

Walter, V., 2004. Object-based classification of remote sensing data for change detection, ISPRS Journal of Photogrammetry & Remote Sensing, 58, 225 – 238.

Paper Reference No.	:	PN – 53
Title of the Paper	:	Delineation of Climate Divisions for Peninsular Malaysia
Name of the Presenter	:	Fariza Yunus

Authors Affiliation	:	Assoc. Prof. Sr. Dr. Jasmee Jaafar Assoc. Prof. Dr. Zamalia Mahmud Dr. Aziz Shafie
Mailing Address	:	Department of Surveying Science and Geomatics, Faculty of Architecture, Planning and Surveying, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia
Email Address	:	fariza@met.gov.my
Telephone Numbers	:	603 0123993579 603 55444347
Fax Number	:	603 55444353

Authors Photograph :

Fariza Yunus
Department of Surveying Science and Geomatics,
Faculty of Architecture, Planning and Surveying,
Universiti Teknologi MARA
40450 Shah Alam, Selangor, Malaysia

Assoc. Prof. Sr. Dr. Jasmee Jaafar
Academic Advisor,
Department of Surveying Science and Geomatics,
Faculty of Architecture, Planning and Surveying,
Universiti Teknologi MARA
40450 Shah Alam, Selangor, Malaysia.

Assoc. Prof. Dr. Zamalia Mahmud
Senior Lecture,
Faculty of Computer & Mathematical Sciences,
Universiti Teknologi MARA,
40450 Shah Alam, Selangor, Malaysia.

Dr. Aziz Shafie
Senior Lecturer,
Geography Department,
Faculty of Arts and Social Sciences,
University of Malaya,
50603 Kuala Lumpur, Malaysia

Presenter Bibliography

Fariza Yunus is attached to Malaysian Meteorological Department since December 1996. Her first duty was in Agrometerorological Division, she supervised Geographical Information System (GIS), produced GIS product and involved in research. She was graduated on August 1996 from University of Malaya with Bachelor Degree of Science in Mathematics. In 2005, she received her Master Degree of Science in Remote Sensing and GIS from University Putra Malaysia. Currently she is pursuing her Ph.D in Build Environment at University Technology MARA (UiTM). She has produced several technical papers related to geospatial topics.