Greenness Pattern Analysis with the Remote Sensing Index Clustering

Nittaya Kerdprasop, Kacha Chansilp, and Kittisak Kerdprasop

Abstract—**We present in this paper the analysis results of surface greenness patterns over some metropolitan areas in the East and Southeast Asia through remotely sensed measurement index obtained from the NOAA satellites. Remote sensing index used in our case studies is a vegetation health index that had been proven a proper proxy for Earth surface greenness assessment. This index had been computed and recorded as time series in a weekly timeframe. The greenness patterns of Bangkok, Beijing, and Ho Chi Minch City are used as demonstration cases of our analysis methodology, which is based on the time series clustering. The remotely observed vegetation health indices over Bangkok, Beijing, and Ho Chi Minch City during the years 1982 to 2016 had been clustered with three clustering algorithms: k-means, two-step, and Kohonen self-organizing network. The number of appropriate clusters is automatically determined from the Silhouette coefficient evaluation. Based on this coefficient, clustering results of greenness trends over Beijing and Ho Chi Minh City areas show the formation of three clusters, whereas the greenness pattern of Bangkok is more fluctuate with the formation of five clusters. The more number of clusters, the higher variation in greenness patterns.**

Index Terms—**Greenness patterns, NOAA remote sensing data, Time series clustering, Vegetation health index.**

I. Introduction

Earth observation satellites provide valuable products for monitoring environmental and ecosystems. In this work, we study the vegetation monitoring index derived from the sensors called Advanced Very High Resolution Radiometer (AVHRR) equipped onboard the polar-orbiting satellites of the National Oceanic and Atmospheric Administration (NOAA) of the United States of America.

The time series data and images are provided weekly with a 4 km spatial resolution [1], [2]. Vegetation Health Index (VHI) is a remote sensing product valuable for several purposes including drought monitoring [3]-[5], irrigation system [6], crop productivity [7], mosquito-borne disease [8], and ecological resources [9]-[16]. Most of these works adopt satellite images as major source of analysis [17]. We, on the contrary, employ the vegetation index as numeric data.

In this work, we use VHI for observing greenness of vegetation cover over the major metropolitan areas of the two Asian capitals (Beijing and Bangkok) and one populated city (Ho Chi Minh City). Greenness patterns over a long period of 35 years have been analyzed with cluster analysis method. The remotely sensed data are time series in ASCII numeric format. We thus propose a novel idea of using (slope, residual)-pairs obtained from the linear regression as time series representation for computing distance and forming similar data instances in the same cluster and placing dissimilar instances in different clusters. The details of research methodology are explained in Section II, and the cluster analysis results are illustrated in Section III. We conclude this paper and point out future research direction in Section IV.

Manuscript received September 9, 2017; revised December 2, 2017. This work was supported in part by grants from the Thailand Toray Science Foundation, the National Research Council of Thailand, and Suranaree University of Technology through the funding of Data Engineering Research Unit and the Knowledge Engineering Research Unit.

N. Kerdprasop is with the School of Computer Engineering and head of Data Engineering Research Unit, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand (e-mail: nittaya@sut.ac.th).

K. Chansilp is with the School of Computer Engineering, Suranaree University of Technology (e-mail: kacha@sut.ac.th).

K. Kerdprasop is with the School of Computer Engineering and Knowledge Engineering Research Unit, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand (e-mail: kerdpras@sut.ac.th).

II. Research Methodology

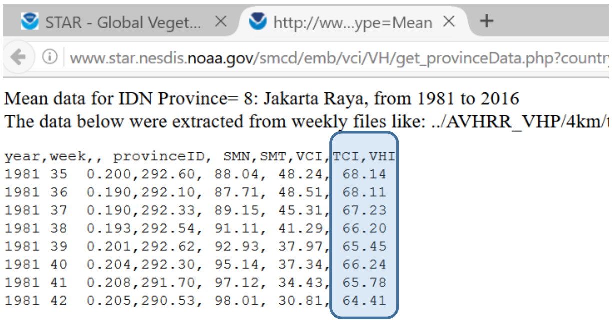
A. Study Area and Data Source

We focus our study of vegetation greenness patterns over the metropolitan areas of China (Beijing), Thailand (Bangkok), and Vietnam (Ho Chi Minh City). The locations of these three cities are shown in Fig. 1.

The satellite derived vegetation indexes provided by the NOAA-AVHRR include the no noise normalized difference vegetation index (SMN), no noise brightness temperature (SMT), vegetation condition index (VCI), temperature condition index (TCI), and vegetation health index (VHI). In this study, we use only VHI, shown as the last column in Fig. 2.

Fig. 1. The metropolitan locations to be studied on their greenness patterns.

doi: 10.18178/ijmlc.2017.7.6.643



Mean data for IDN Province= 8: Jakarta Raya, from 1981 to 2016
The data below were extracted from weekly files like: ../AVHRR_VHP/4km/

```
year,week,, provinceID, SMN,SMT,VCI,TCI,VHI
1981 35  0.200,292.60, 88.04, 48.24, 68.14
1981 36  0.190,292.10, 87.71, 48.51, 68.11
1981 37  0.190,292.33, 89.15, 45.31, 67.23
1981 38  0.193,292.54, 91.11, 41.29, 66.20
1981 39  0.201,292.62, 92.93, 37.97, 65.45
1981 40  0.204,292.30, 95.14, 37.34, 66.24
1981 41  0.208,291.70, 97.12, 34.43, 65.78
1981 42  0.205,290.53, 98.01, 30.81, 64.41
```

Fig. 2. Example of VHI time series data obtained from the NOAA-AVHRR.

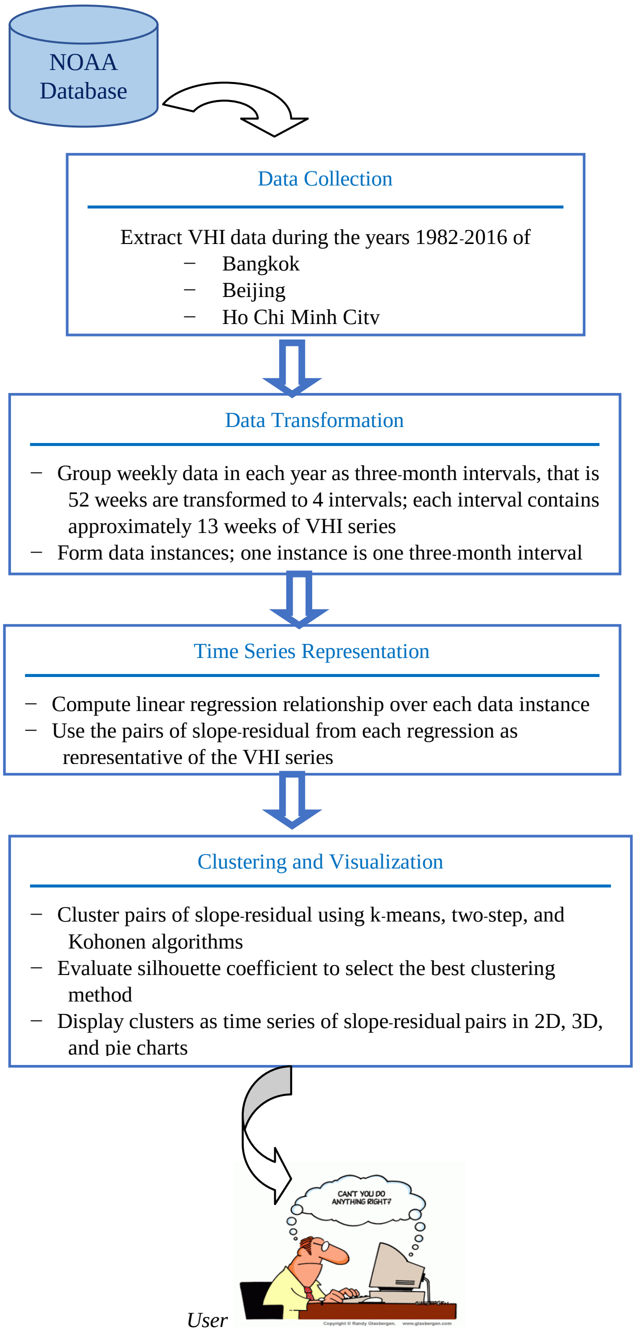


Fig. 3. The steps in our time series representing and clustering analysis.

B. The Induction of VHI Time Series Patterns

On analyzing the VHI time series patterns of metropolitan areas, we design analysis process (Fig. 3) composing of four major steps:

- data collection,
- data transformation,
- time series representation,
- clustering and visualization.

Data collection: The 35-year VHI values of each city are extracted and recorded as separate data files to be used in our analysis steps. The VHI data are collected weekly starting from January 1982 to September 2016.

Data transformation: VHI data in each year are then divided as four quarters; each quarter is a three-month period. This quarter formation is for detecting seasonal change of VHI patterns. The weekly time unit is then encoded as quarter unit and represented as year-month. For instance,

198201 means January-March of 1982,
198204 means April-June of 1982,
198207 means July-September of 1982, and
198210 means October-December of 1982.

At the end of data transformation step, our data file for each city contains 138 records of VHI from 1982-2016 in quarter unit. Each data record is VHI series for the three-month period. Note that the last quarter of 1994 is unavailable and we ignore this missing value.

Time series representation: We then calculate regression relationship within the quarter. The relationship is represented by a slope of linear line and the residual, which is the error from linear curve fitting. Example of (slope-residual)-pair values is illustrated as input table in Fig. 4. The purpose of representation step if for compactness of the series that help accelerating the series analysis.

Clustering and visualization: The subsequent step is clustering over the time series representation. Intuitive idea is that series showing the same pattern of slope, either positive or negative, with close amount of magnitude and residual should be placed in the same cluster. The clustering process has been performed with three methods: k-means [18], two-step [19], and Kohonen self-organizing network [20]. The best method is assessed through the silhouette coefficient (SC) [21]. The higher SC, the better cluster formation. Results are finally displayed with visualization tools.

III. Running Results and Discussions

The comparative results regarding clustering methods and their efficiency can be explained and discussed as follows.

Beijing greenness pattern:

The best method for clustering Beijing's VHI is two-step clustering with the optimal number of 3 clusters (Fig. 5). The cluster sizes are quite evenly distributed (31.4%, 33.6%, and 35.0%). This three-group partition means there are three greenness variations over the Beijing metropolitan area.

The Silhouette coefficient (SC) that has been used as an evaluation metric to assess cluster formation of Beijing area is 0.487. Based on this SC value, it can be interpreted that the grouping is quite fair and reasonable. The SC value ranges from -1 to 1. The desirable SC measure is the one that is close to 1. The VHI patterns (displayed as time series plot) of Beijing ground area show clearly three different patterns.

The 3D plot of VHI-regression residual in Fig. 5 shows the distinct three groups of greenness patterns. These patterns can be clearly seen from the 2D graph at the bottom part of the figure. The 2D graph is the plot of VHI slope during 1982 to 2016. For cluster 1, the VHI slope is ranging in the positive region from 0.2 up to 4.6. The positive slope refers to the increase in surface greenness. That means cluster 1 (graph on the left hand side) represents the spring season of Beijing. The greenness peak in the decade 2000-2010 is quite noticeable, but the greenness situation seems to decay from the years 2011.

For cluster 2 (the middle graph), the slope of VHI is in the negative region. This negative VHI-slope means the decrease in surface greenness. Therefore, cluster 2 represents the periods of fall and winter that trees turn brown.

Cluster 3 (a graph on the right hand side) includes both positive and negative trends of VHI regression lines. But the magnitudes of the positive trend is not high as in the cluster 1, likewise the negative trend is not strong as in the cluster 2. That means this group of VHI patterns contains a slight change of greenness. This kind of pattern may represent the early spring and the early fall or even the summer time over the Beijing area.

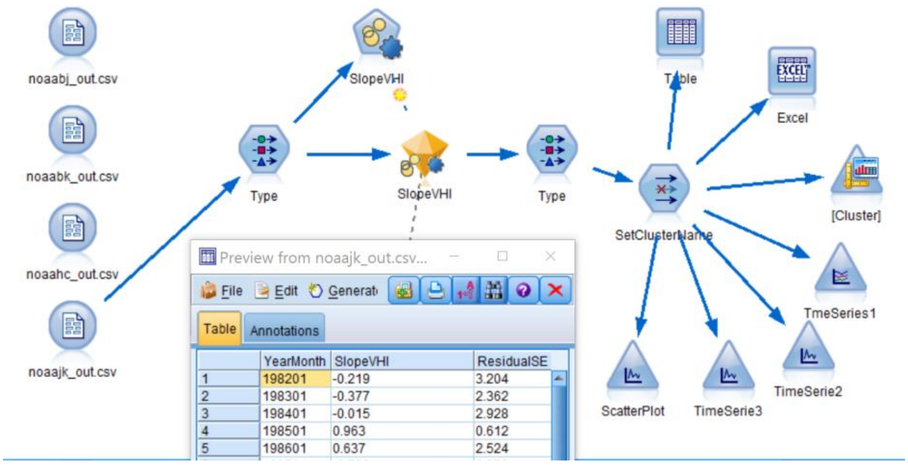


Fig. 4. Input data table and analysis steps implemented with the SPSS Modeler.

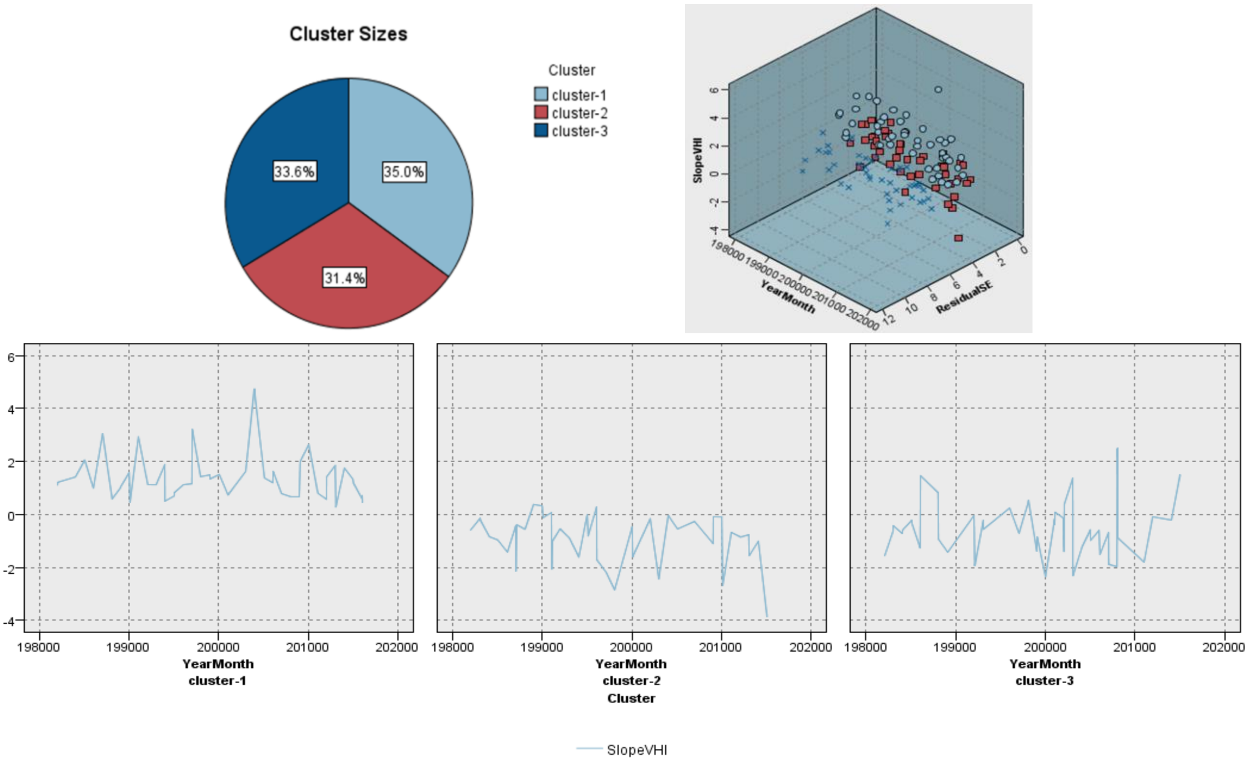


Fig. 5. Clustering result of Beijing greenness pattern.

Ho Chi Minh City greenness pattern:

The best method for clustering Ho Chi Minh City's VHI is the same as the Beijing area; that is, two-step clustering with the optimal number of 3 clusters (Fig. 6). But the cluster sizes are not evenly distributed. The majority of greenness events (51.8%) are in cluster 2. The quality of cluster formation is quite good with the SC metric at 0.538. From the trend in cluster 2, it can be inferred that the greenness over Ho Chi Minh area are increasing almost all of the time, except in some season that the greenness is decreased (cluster 3).

Bangkok greenness pattern:

Greenness pattern of Bangkok (Fig. 7) is quite different from Beijing and Ho Chi Minh City. The method best clustering Bangkok is k-means with the SV value equal to 0.487. The optimal number of clusters is five. The 3D scatter plot and the 2D graph show that greenness pattern over

Bangkok metropolitan area is five distinctive styles.

Cluster 1 is the majority (34.8%) with positive trend of greenness. The second largest group is cluster 4 (23.2%) that greenness is decreasing. This may represent the cold and dry season in Bangkok. The noticeable group of greenness trends is those in cluster 5 (15.9%) that show high slope in VHI increasing. This should be the rainy season that trees are green all over Bangkok. To validate the clustering results over the three metropolitan areas, we also check the real VHI values during the past 35 years. Fig. 8 shows the actual VHI values from the NOAA-AVHRR. The possible range of VHI is 0 to 100. The VHI higher than 60 indicates good health of vegetation. The VHI below 40 shows vegetation stress due to drought.

It can be noticed that VHI, which is the proxy for greenness, over Beijing fluctuates from 24 to 76 with the mean value around 50. The greenness indicator of Ho Chi Minh City is also fluctuated between 24 and 75. But the majority of VHI values is below 50.

The VHI pattern of Bangkok is different from Beijing and Ho Chi Minh City in that the fluctuation does not center around some mean horizontal line. The VHI trend is instead declining from 70 down to around 28. This means the less greenness of Bangkok metropolitan over the past decades.

From the comparative analysis of clustering results and the actual trends of VHI values, we can confirm the reliable results obtained from our experiment. The confirmation is that the 3-group clustering over Beijing and Ho Chi Minh City and the 5-group forming of VHI values over Bangkok metropolitan areas are showing accordance, comparative to the real trends of VHI values.

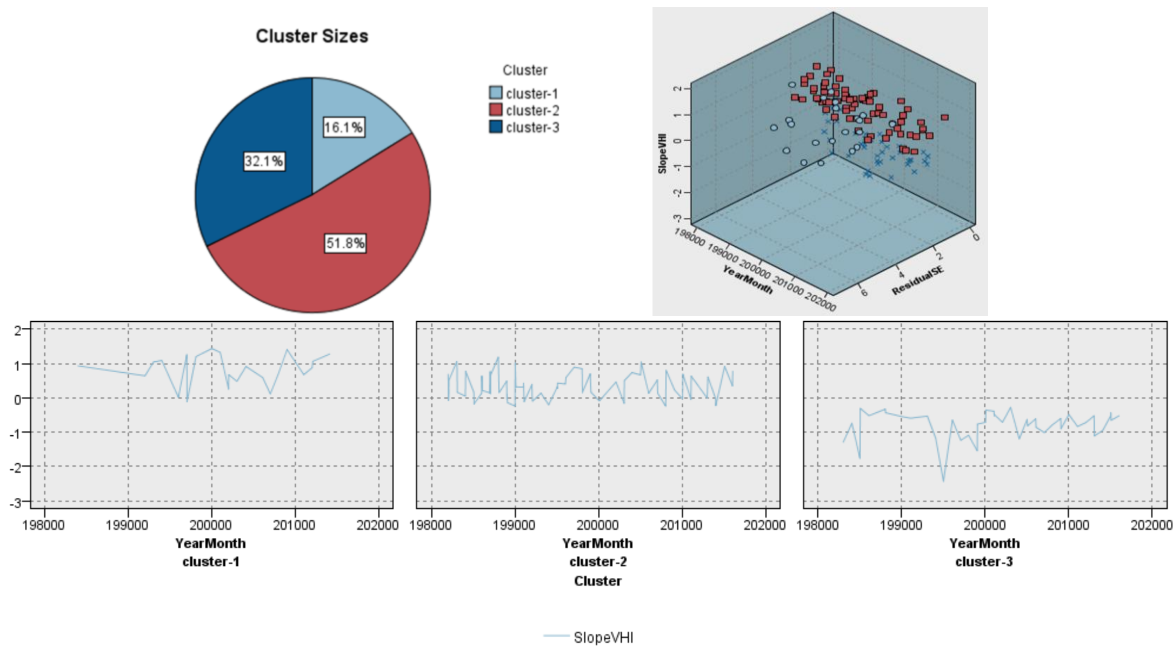


Fig. 6. Clustering result of Ho Chi Minh City greenness pattern.

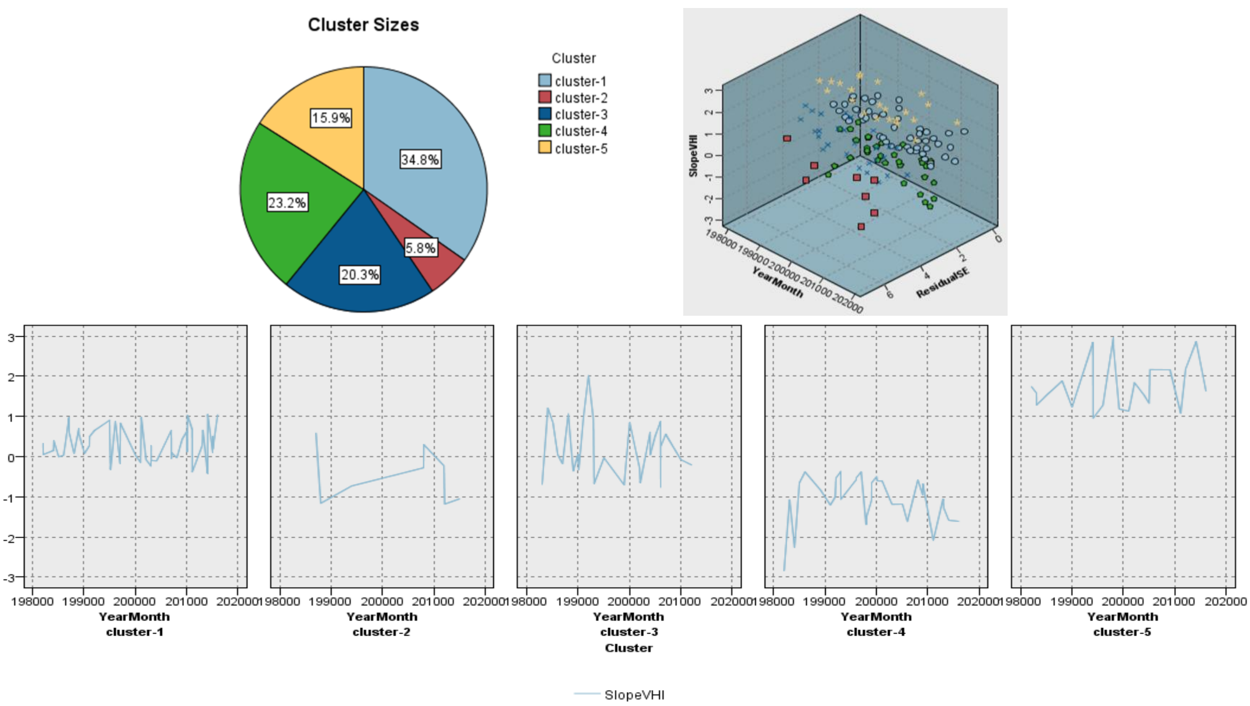


Fig. 7. Clustering results of Bangkok greenness pattern.

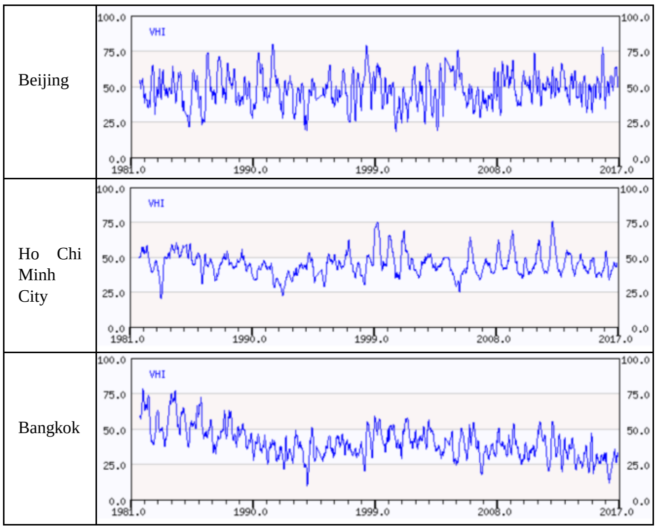


Fig. 8. The plot of actual VHI values during years 1982-2016 of the three metropolitan cities.
(Source: www.star.nesdis.noaa.gov/smcd/emb/vci/VH/vh_browseByCountry_province.php)

IV. Conclusion

This research work present the vegetation greenness analysis based on the remote sensing information obtained from the polar orbiting environmental satellites of the National Oceanic and Atmospheric Administration (NOAA), U.S.A. We consider vegetation greenness as a proxy indicator for fertile condition of the ecosystem and the good condition for human and other livings. The remotely sensed vegetation greenness is evaluated from the vegetation health index (VHI) computed from the Advanced Very High Resolution Radiometer (AVHRR).

From the literature, vegetation index of NOAA-AVHRR has been successfully applied to study drought and other climate conditions. We therefore have an initiative idea of applying VHI to study greenness patterns over 35-year periods of major populated cities in East and Southeast Asia. We study VHI patterns over Beijing, Bangkok, and Ho Chi Minh City. Our study is based on the cluster analysis method.

We also propose a new idea of using linear regression of VHI sub-intervals as time series representatives. The use of time series representation is for information compactness and for the temporal data transformation to the time-independent data that are free for data swap during cluster formation.

From the clustering results, we have found that the greenness patterns of Beijing and Ho Chi Minh City over the past three decades can be naturally formed the patterns as three groups based on the VHI trends. The greenness pattern of Bangkok is basically five groups according to the VHI trends. From the promising results of cluster analysis, we plan to further our study with additional data mining techniques to make an in-depth analysis of environmental situations over these major metropolitan cities.

References

[1] F. Kogan, "30-year land surface trend from AVHRR-based global vegetation health data," in *Use of Satellite and In-situ Data to Improve Sustainability*, F. Kogan *et al*., Eds., 2011, pp. 119-123.

[2] NOAA STAR. (August 2017). Global Vegetation Health Products, Center for Satellite Applications and Research. [Online]. Available: http://www.star.nesdis.noaa.gov/smcd/emb/vci/VH/

[3] Y. Bayarjargal, A. Karnieli, M. Bayasgalan, S. Khudulmur, C. Gandush, and C. Tucker, "A comparative study of NOAA-AVHRR derived drought indices using change vector analysis," *Remote Sensing of Environment*, vol. 105, pp. 9-22, 2006.

[4] K. Kerdprasop and N. Kerdprasop, "Remote sensing based model induction for drought monitoring and rainfall estimation," in *Proc. the 16th International Conference on Computational Science and Its Applications*, pp. 356-368, 2016.

[5] K. Kerdprasop and N. Kerdprasop, "Geo-information knowledge base system for drought pattern search," in *Proc. the 16th International Conference on Computational Science and Its Applications*, pp. 332-344, 2016.

[6] V. Boken, G. Hoogenboom, F. Kogan, J. Hook, D. Thomas, and K. Harrison, "Potential of using NOAA-AVHRR data for estimating irrigated area to help solve an inter-state water dispute," *International Journal of Remote Sensing*, vol. 25, no. 12, pp. 2277-2286, 2004.

[7] A. Vina, A. A. Gitelson, A. L. Nguy-Robertson, and Y. Peng, "Comparison of different vegetation indices for the remote assessment of green leaf are index of crops," *Remote Sensing Environment*, vol. 115, pp. 3468-3478, 2011.

[8] N. Kerdprasop and K. Kerdprasop, "Remote sensing based modeling of dengue outbreak with regression and binning classification," in *Proc. the 2nd IEEE International Conference on Computer and Communications*, pp. 46-49, 2016.

[9] L. C. Alatorre, et al., "Temporal changes of NDVI for qualitative environmental assessment of mangroves: Shrimp farming impact on the health decline of the arid mangroves in the Gulf of California (1990-2010)," *Journal of Arid Environments*, vol. 125, pp. 98-109, 2016.

[10] P. M.Atkinson, J. Dash, and C. Jeganathan, "Amazon vegetation greenness as measured by satellite sensors over the last decade," *Geophysical Research Letters*, vol. 38, pp. 1-6, 2011.

[11] R. E. Burgan and R. A. Hartford, "Monitoring vegetation greenness with satellite data," General Technical Report INT-297, Intermountain Research Station, Forest Service, United States Department of Agriculture, 16 pp., 1993.

[12] F. Detsch, I. Otte, T. Appelhans, A. Hemp, and T. Nauss, "Seasonal and long-term vegetation dynamics from 1-km GIMMS-based NDVI time series at Mt. Kilimanjaro, Tanzania," *Remote Sensing of Environment*, vol. 178, pp. 70-83, 2016.

[13] S. Erasmi, A. Schucknecht, M. P. Barbosa, and J. Matschullat, "Vegetation greenness in northern Brazil and its relation to ENSO warm events," *Remote Sensing*, vol. 6, pp. 3041-3058, 2014.

[14] L. Ji and A. J. Peters, "Forecasting vegetation greenness with satellite and climate data," *IEEE Geoscience and Remote Sensing Letters*, vol. 1, no. 1, pp. 3-6, 2004.

[15] A. Smith *et al*., "Remote sensing the vulnerability of vegetation in natural terrestrial ecosystems," *Remote Sensing of Environment*, vol. 154, pp. 322-337, 2014.

[16] L. Zhou *et al*., "Widespread decline of Congo rainforest greenness in the past decade," *Nature*, vol. 509, pp. 86-90, 2014.

[17] Y. Xie, Z. Sha, and M. Yu, "Remote sensing imagery in vegetation mapping: A review," *Journal of Plant Ecology*, vol. 1, no. 1, pp. 9-23, 2008.

[18] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. the 5th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281-297, 1967.

[19] T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: An efficient data clustering method for very large databases," in *Proc. the ACM SIGMOD Conference on Management of Data*, pp. 103-114, 1996.

[20] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, pp. 59-69, 1982.

[21] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, New York: Wiley, 1990.

Nittaya Kerdprasop is an associate professor and the head of Data Engineering Research Unit, School of Computer Engineering, Suranaree University of Technology (SUT), Thailand. She received her B.S. in radiation techniques from Mahidol University, Thailand, in 1985, M.S. in computer science from the Prince of Songkla University, Thailand, in 1991 and Ph.D. in computer science from Nova Southeastern University, U.S.A., in 1999. Her research of interest includes data mining, artificial intelligence, logic and constraint programming.

Kacha Chansilp is an assistant professor at the School of Computer Engineering, SUT, Thailand. He received his bachelor degree in computer science from Queens College, U.S.A., in 1993, master degree in graphic design from New York Institute of Technology, U.S.A., in 1995 and doctoral degree in interactive multimedia technologies from Edith Cowan University, Australia, in 2003. His current research includes Web Databases and Educational Systems.

Kittisak Kerdprasop is an associate professor at the School of Computer Engineering, Chair of the School, and the head of Knowledge Engineering Research Unit, SUT. He received his bachelor degree in Mathematics from Srinakarinwirot University, Thailand, in 1986, MS in computer science from the Prince of Songkla University, Thailand, in 1991 and PhD in computer science from Nova Southeastern University, U.S.A., in 1999. His current research includes Machine Learning and Artificial Intelligence