Volume 4, No. 7, July 2013

Journal of Global Research in Computer Science

REVIEW ARTICLE

Available Online at www.jgrcs.info

ISSN-2229-371X

CLOUD COVERAGE IDENTIFICATION USING SATELLITE DATA AND K-MEAN CLUSTERING ALGORITHM

T. Gomasathit

Department of Physics, Faulty of Science,
Mahasarakham University, Khamriang, Kantarawichai,
Mahasarakham, 44150, Thailand
tosak.g@msu.ac.th, msuscienceteam@hotmail.com

Abstract: The primary objective of this study is to identify cloud coverage over the Thailand region by using satellite data and the K-mean clustering algorithm. The successful development software tools for K-mean unsupervised classification were based software ERDAS Imagine using the ERDAS Macro Language and Spatial Modelling Language. The clear rate for the Thailand region was identification from cloud coverage analysis of 8,760 digital data items from meteorological satellite data on October 2010 to September 2011. The analysis of the mean rate and the standard deviation of the clear rate data for each month were related to weather conditions in Thailand.

Keywords: Cloud coverage, identification, satellite data

INTRODUCTION

Clouds are one of the most important forces of Earth's heat balance and hydrological cycle, and at the same time one of the least understood. It is well known that low clouds produce negative feedback while high, thin clouds produce positive feedback on the amount of radiation. The net effect of clouds, however, is still unknown, which causes large uncertainties in climate models and climate predictions [1], [2]. Cloud detection from satellite imagery is important for remote sensing retrieval algorithms. Cloud free pixels must be identified before the retrieval of the atmospheric and surface variables (e.g. land surface temperature and vegetation index). When measuring high value spatial and temporal surface solar cloud detection is indispensable ,radiation [3], [4]. Satellite Remote Sensing data, including high spatial coverage and high temporal sampling, plays an important role in monitoring precipitation in flood-prone areas and other severe weather events [5]. Many researchers have developed numerous methods of cloud detection which have been successfully used to detect clouds for satellite data. These methods include International Satellite Cloud Climatology Project [6], Automated cloud screening of AVHRR[7], Clouds from AVHRR (CLAVR) [8], the AVHRR Processing method above the Clouds, Land and Ocean (APOLLO)[9], MSG/SEVIRI cloud mask [10], Cloud Detection with MODIS and improvements in the MODIS Cloud Mask for Collection [11], Estimation of instantaneous net surface long wave radiation from MODIS cloud-free data [12], synergistic use of POLDER and MODIS for multilayered cloud identification [13], and Generation of high-resolution East Antarctic land fast sea-ice maps from cloud-free MODIS satellite composite imagery [14]. The detection of cloud coverage using a full scene of MTSAT-2 data is computationally expensive because of the significant volume of data to be handled for each scene multiplied by the need throughout the whole year.

THE STUDY AREA

A. The study area

Thailand's boundaries are located from latitudes 05° 37' to 20° 27'N and longitudes 97° 22'E to 105° 37'E, with an area of 513,115.029 km^2, in the warm and tropical climate region (see in Figure.1). The average annual rainfalls range from 1000 to 1500 mm for the north eastern and central parts. But on the eastern tip and the southern peninsula, rainfall is higher, averaging from 2000 to 3000 mm [15].

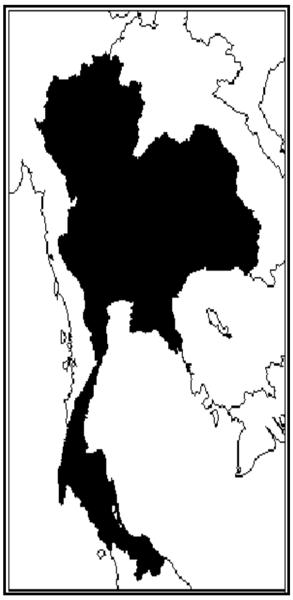

Figure 1. Thailand boundary

B. Data usage

The Multi-functional Transport Satellite-2 (MTSAT-2) is a Japanese geostationary satellite with an operation period from 2010 to 2015 [16]. It has five spectral channels, which are in the following wavelength ranges: visible channel VIS (0.55-0.90 µm), Thermal-Infrared channel IR1 (10.3-11.3 µm), IR2 (11.5-12.5 µm), Water Vapor channel IR3 (6.5-7.0 µm), and Shortwave Infrared channel IR4 (3.5-4.0 µm). The spatial resolution of the visible image is about 1x1 km^2 (nadir) and the TIR image is about 4x4 km^2. An example of the satellite's infrared image above Thailand is shown in Figure 2. The MTSAT-2 has provided information to most countries in the Asia-Pacific region including

(1) Data for monitoring the distribution and motion of clouds

(2) Sea surface temperatures, and

(3) Distribution of water vapor.

Its data are currently available on the internet in several places, such as JMA's MTSAT data website (http://dmss. tksc.jaxa.jp/sentinel/ contents/ MTSAT.html).

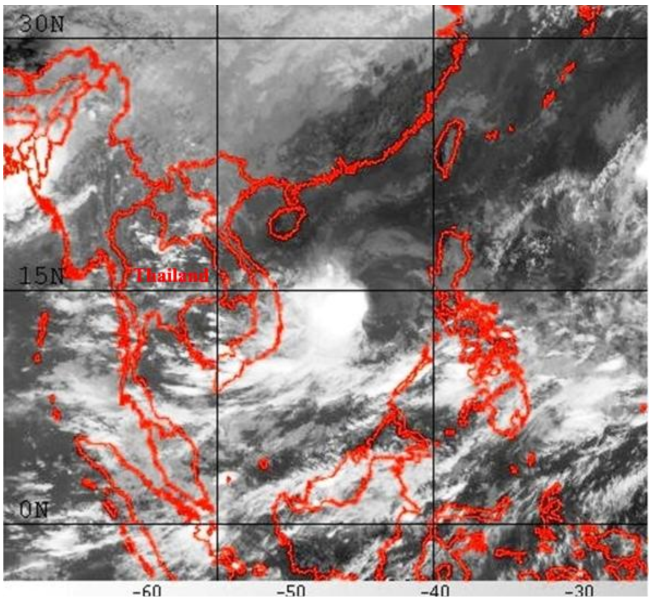

Figure 2. The Multi-functional Transport Satellite-2 data cover Thailand

METHODOLOGY

A. Cloud covers identification method

To determine cloud coverage, the pixels in the image were classified into various classes according to their digital numbers. Unsupervised classification using the K-means algorithm was used to distinguish between cloud and non-cloud pixels. The K-means algorithm is a non-hierarchical clustering unsupervised classification method [17], [18], [19], [20]. An arbitrary number of clusters should be temporally selected. Using the "least distance method" number belonging to each cluster, the author assigns the cluster to the group that corresponded to the cluster mean nearest to it. At the end of the first iteration, the mean for each group is calculated and the process is repeated again. The non-hierarchical clustering method has been used for cloud, land and water area detection because it is advantageous for a huge volume of data over large areas, as it is a computer-generated automated technique. Due the huge number of images to be processed, it would be very demanding on time and effort to perform the above-discussed processes on each and every image manually. Consequently, batch processing was considered as a better option to accomplish the various tasks. SML and EML, as a kind of macro language supported by ERDAS, are used to defining the client interface structure and content. It can provide some editing capabilities of the basic process [21], [22], [23], [24]. In this work, the author successfully developed a novel GUI by using SML and EML scripting language, for the K-mean unsupervised classification. In Figure 3 shows an example of the novel GUI.

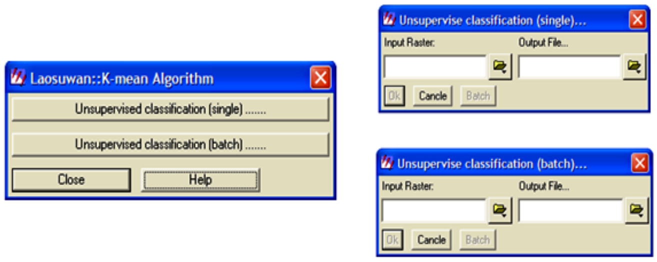

Figure 3. Development tool of K-mean unsupervised classification

The MTSAT-2 browse image was formed using three bands (visible, mid-infrared and thermal infrared) to classify the image data. The false color composite images were generated by assigning the red, green and blue colors to three channels of a MTSAT-2 image: visible is red, mid-infrared (or water vapor channel) is green, and thermal infrared is blue. The classified image was randomly interpreted by RGB color image (false color composite). Initially, the number of classes was increased to eight to study the effect of the number of classification classes on giving more accurate results. Finally eight classes were divided into two groups, namely cloud and non-cloud, using the cluster nearest mean value technique. The percentage of the cloud pixels over the total pixels present in the image was noted. The identification of the study area over Thailand is shown in Figure 4. It is clear that there are distinct differences in cloud coverage percentage along Thailand's boundary.

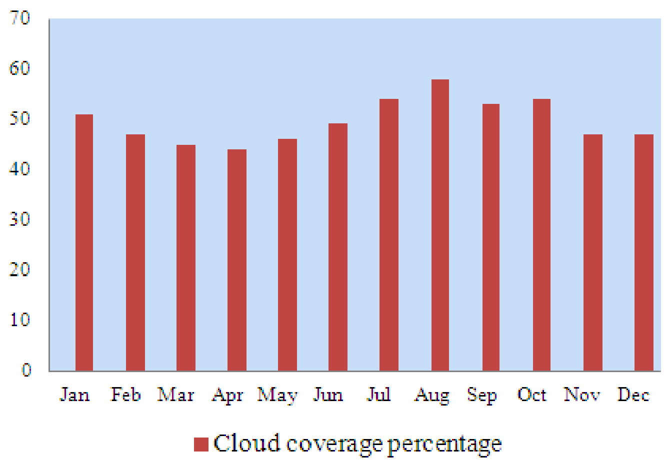

Figure 4. Illustrate cloud coverage over Thailand

B. Clear rate equation

In this study the author uses the clear rate technique (see in equation 1) as follows:

$$CL=10-(C/10) \qquad (1)$$

Where CL is the clear rate and C is cloud coverage (expresses as a percentage).

When C = 0%, then CL = 10, and when C = 100%, then CL = 0. A clear rate of 10 is considered as clear weather, and conversely, a clear rate of 0 is considered as bad weather (overcast). Clear rate was estimated from the non-cloudiness of the data for each pixel of each image. Clear rate was used instead of a cloud percentage for the simplicity analysis of weather conditions. The estimated clear rates for the three regions are shown in Figure 5.

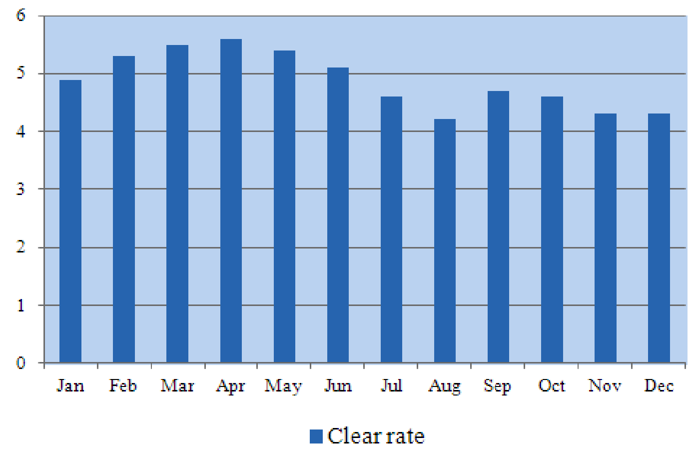

Figure 5. Clear rate in twelve months

In Figure 6. Illustrates the relationship between the standard deviation and the mean value of the clear rate for each month of the year. The relative frequency of occurrence of clear rate for Thailand regions was estimated and different probability distribution function were tried for a best-fit curve, the results of which are shown in Figure 7.

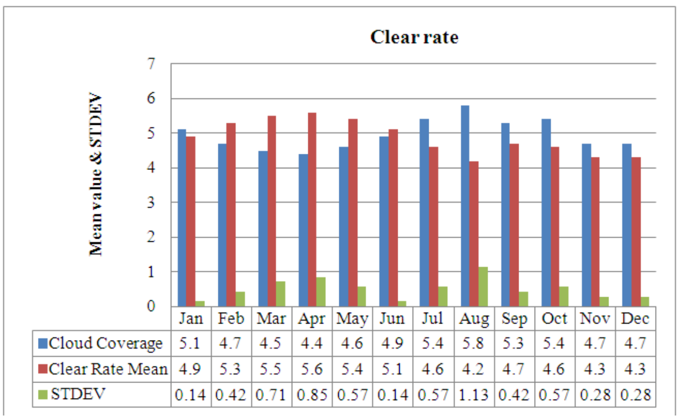

Figure 6. Illustrates the relation between the mean value and the standard deviation

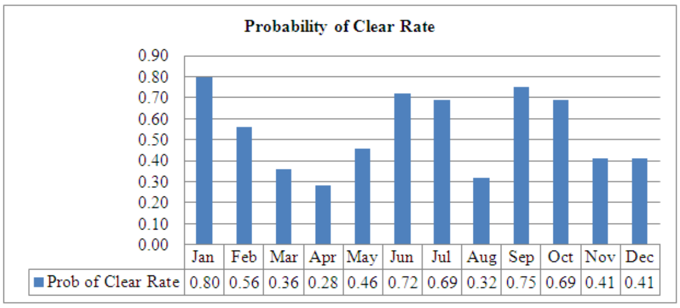

Figure 7. Results of probability of clear rate

RESULTE

A. *Clear rate analysis*

Thailand experiences the effects of monsoons from the South China Sea. It experiences northeast monsoons from November to March with a large amount of rainfall due to water vapor and clouds brought by the northeast wind from the South China Sea. It experiences southwest monsoons from April to September. These monsoons are relatively drier as most of the air comes from Sumatra instead of from a large sea body. It experiences inter-monsoons of moderate rainfall and light winds during the months of March and October.

B. *The clear rate of mean value and standard deviation*

Figure 6 shows the relationship between the mean value and standard deviation of clear rate for Thailand regions. A high mean value for the clear rate indicates good weather and a low mean value indicates cloudy weather. A high standard deviation indicates frequently changing weather and a low standard deviation indicates more stable weather conditions. Figure 6 also shows that the Thailand region experiences a more consistent weather pattern, with a narrow dispersion range of mean clear rate values.

C. *Probability Model*

Different probability distribution functions (normal, lognormal, extreme, and value) were tested to find the best fitted function of the clear-rate data for each region. The interval ranges used for clear rates and relative frequency of occurrence are shown in Figure. 7. The Figure 7 illustrates only the considered probability distribution function for the Southeast Asian region. The RMS error was used to decide the best-fitted curve of probability distribution function (see in Table 1). The RMS error rates obtained from the different probability distribution functions that were tested for each of the Thailand region. One can estimate the probability of occurrence of good or bad weather at one position by using the probability distribution function.

Table I. The RMS error was used to decide the best-fitted curve of probability distribution function

Normal	Log-normal	Extreme-max	Extreme-min
0.01500	0.01872	0.03607	0.03314

CONCLUSION

The aim of this study is to identify cloud coverage over the Thailand region by using satellite data and the K-mean clustering algorithm. The successful development software tools for K-mean unsupervised classification were based on the commercial software ERDAS Imagine using the ERDAS Macro Language and Spatial Modelling Language. The clear rate for the Thailand region was identificatied from cloud coverage analysis of 8,760 digital data items from October 2010 to September 2011. The results would be more conclusive if more data covering a longer span of between 3 to 5 years were available. However, the result described in this study provides some preliminary information about the possibility of weather analysis from cloud cover estimates using MTSAT-2 data, and would be useful for further studies when more data will be available.

ACKNOWLEDGMENT

This research was financially supported by budget fiscal year 2012 of the Faculty of Science, Mahasarakham University

REFERENCES

[1] S. Solomon, D. Qin, M. Manning, R.B. Alley, T. Berntsen, N.L. Bindoff, Z. Chen, A. Chidthaisong, J.M. Gregory, G.C. Hegerl, M. Heimann, B. Hewitson, B.J. Hoskins, F. Joos, J. Jouzel, V. Kattsov, U. Lohmann, T. Matsuno, M. Molina, N. Nicholls, J.Overpeck, G. Raga, V. Ramaswamy, J. Ren, M. Rusticucci, R. Somerville, T.F. Stocker, P. Whetton, R.A. Wood and D. Wratt, Technical Summary. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (Eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.

[2] E. Jansen, J. Overpeck, K.R. Briffa, J.-C. Duplessy, F. Joos, V. Masson-Delmotte, D. Olago, B. Otto-Bliesner, W.R. Peltier, S. Rahmstorf, R. Ramesh, D. Raynaud, D. Rind, O. Solomina, R. Villalba, and D. Zhang, Palaeoclimate. In Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor, and H.L. Miller, Eds. Cambridge University Press, 2007, pp 433-497.

[3] V. Laine, A. Ven¨al¨ainen, M. Heikinheimo, O. Hyv¨arinen, Estimation of surface solar global radiation from NOA AVHRR data in high latitudes, Journal of Applied and Meteorology, 38 (12), 1999, pp 1706–1719.

[4] L. Teerawong, N.K. Tripathi, Potentiality of data fusion techniques for flood identification. Proceeding of USPAM. Conference, 8-6March, 2007.

[5] W.B. Rossow, and L.C. Garder, Validation of ISCCP cloud detections, Journal of Climate, 6, 1993, pp 2370-2393.

[6] Gallaudet, T .C., and J.J. Simpson, Automated cloud screening of AVHRR imagery using split-andmerge clustering. Remote Sensing of Environment, 38, 1991, pp 77 – 121,

[7] L. L. Stowe, P. A. Davis, and E. P. McClain, Scientific basis and initial evaluation of the CLAVR-1 global clear/cloud classification algorithm for the Advanced Very High Resolution Radiometer, Journal of Atmospheric and Oceanic Technology, 16**,** 1999**,** pp 656–681.

[8] K.T. Kriebel, G. Gessell, M. Kästner, H. Mannstein, The cloud analysis tool APOLLO: improvements and validations. International Journal of Remote Sensing, 24, 2003, pp 2389-2408.

[9] M. Derrien, and H. Le Gléau, MSG/SEVIRI cloud mask and type from SAFNWC, International Journal of Remote Sensing, 26, 2005, pp 4707 – 4732.

[10] I. Genkova, G. Zhao, G. Seiz, E. Snodgrass, M. Colon, L. Di Girolamo, and R. Rauber, Validation of trade wind cumulus cloud properties produced by meteorological satellites. Proceedings of SPIE, Remote Sensing of Clouds and the Atmosphere X, Vol. 5979, 0-8194-5999-2 Pub. Oct 2005; Softcover; In Print, 2005.

[11] B. Tang and ZL. Li, Estimation of instantaneous net surface longwave radiation from MODIS cloud-free data, Remote Sensing of Environment, 112(9), 2008, pp 3482-3492.

[12] Z. Yao, Z. Han, Z. Zhao, L. Lin, and X., Fan, Synergetic use of POLDER and MODIS for multilayered cloud identification, Remote Sensing of Environment, 114, 2010, pp 1910-1923.

[13] A.D. Fraser, R.A. Massom, K.J. Michael, Generation of high-resolution East Antarctic land fast sea-ice maps from cloud free MODIS satellite composite imagery, Remote Sensing of Environment, 114 (12),2010, pp 2888-2896.

[14] A.D. Fraser, R. A. Massom, K. J. Michael, B. K. Galton-Fenzi, J. L. Lieser, East Antarctic Landfast Sea Ice Distribution and Variability, 2000–08, Journal of Climate, 25,2012, pp 1137–1156.

[15] Thai Meteorological Department, The average annual rainfalls in Thailand, Available from: http://www.tmd.go.th [Last accessed on 2012 Jan 28].

[16] Meteorological Satellite Center (MSC), MTSAT-2, Available: http://mscweb.kishou.go.jp/index.htm [Last accessed on 2012 Jan 28].

[17] J.T. Tou, and R.C.Gonzalez, Pattern Recognition Principles (Reading, MA: Addison-Wesley), 1974.

[18] Teerawong Laosuwan, Pormate Chunpang, Seckson Sukkhasena and Yuttana Uraichun, The Study Change Detection of Land use/ Land cover by using Multi-Temporal Satellite Data, Journal of Science and Technology Mahasarakham University, Spatial issue 7, 2011, pp 1-6,

[19] Teerawong Laosuwan, Pormate Chunpang, Seckson Sukkhasena and Yuttana Uraichun, Improving Satellite Remote Sensing in Multi Spectral through Image Fusion Algorithm, Journal of Science and Technology Mahasarakham University, Spatial issue 8, 2012, pp 82-87.

[20] T.Laosuwan, S. Pattanasethanon and W. Sa-Ngiamvibool, Using GIS, RS for soil erosion mapping, Geospatial World Weekly 21 January 2013, Available from: http://www.geospatialworld.net/Paper/Application/Articl

eView.aspx?aid=30407 [Last accessed on 2013 March 02].

[21] Leica Geosystems, ERDAS Macro Language Reference Manual, ERDAS, Inc. Atlanta, Georgia, 2000.

[22] Leica Geosystems, ERDAS IMAGINE Configuration Guide for Windows, GIS & Mapping, LLC Atlanta, Georgia, 2003

[23] Leica Geosystems, Geospatial Imagine, Programming Reference Guide- Erdas Macro Language, 2010.

[24] Teerawong Laosuwan, Singthong Pattanasethanon, Worawat Sa-ngiamvibool, Automated Cloud Detection of Satellite Imagery Using Spatial Modeler Language and ERDAS Macro Language, IETE Technical Review,30, 2013, pp183-90.

