HIGH RESOLUTION SATELLITE PRECIPITATION ESTIMATE USING CLUSTER ENSEMBLE CLOUD CLASSIFICATION

*Majid Mahroogh*y[1, 2], *Nicolas H. Younan*[1, 2], *Valentine G. Anantharaj*[3], *and James Aanstoos*[2]
[1]Department of Electrical Engineering, Mississippi State University, Mississippi
[2]Geosystems Research Institute, Mississippi State University, Mississippi
[3]National Center for Computational Sciences, Oak Ridge National Laboratory, Tennessee

ABSTRACT

The link-based cluster ensemble (LCE) method is applied to a high resolution satellite precipitation estimation (HSPE) algorithm, a modified form of the Precipitation Estimation from Remotely Sensed Imagery using an Artificial Neural Network Cloud Classification (PERSIANN-CCS) algorithm. The HSPE involves the following four steps: 1) segmentation of infrared cloud images into patches; 2) cloud patch feature extraction; 3) clustering and classification of cloud patches using cluster ensemble technique; and 4) dynamic application of brightness temperature (Tb) and rain rate relationships, derived using satellite observations. The LCE method combines multiple data partitions from different clustering in order to cluster the cloud patches. The results show that using the cluster ensemble increase the performance of rainfall estimates if compared to the HSPE algorithm using Self Organizing Map (SOM). The Heidke Skill Score (HSS) is improved 5% to 7% at medium and high level of rainfall thresholds.

Index Terms— Clustering method, neural networks, feature extraction, image texture analysis

1. INTRODUCTION

Rainfall estimation at high spatial and temporal resolutions is beneficial for research and applications in the areas of weather, climate, hydrology, water resources management, and agriculture. Ground-based estimates from weather radars and in-situ measurements from rain gages facilitate routine monitoring of rainfall across much of the continental areas of the world. But the coverage of the ground-based observation systems is not spatially and temporally uniform. For example, radar coverage is sparse in areas across mountain ranges and tropical rain forests that take up large areas of the globe; and most of the in-situ rainfall measurements are reported only as daily accumulated values. Besides, estimation of rainfall over the oceans is also important for climate studies which cannot be provided by ground-based estimates. On the other hand, satellite-based observing systems are used for the routine monitoring of the earth's environment. Hence, precipitation estimation based on satellite observations offers a viable solution for monitoring global precipitation patterns at sufficient spatial and temporal resolutions.

Many different satellite precipitation estimation (SPE) algorithms have been developed to integrate information from diverse sensors and platforms, including satellite measurements from active and passive radars, visible and infrared imagery (IR), in-situ measurements, and estimates from ground-based radars [1]. Despite the fact that active and passive microwave sensors on satellites can provide physical information about clouds, their temporal resolution is not appropriate for high temporal applications [2]. Passive microwave (PMW) sensors for precipitation measurements are generally deployed on low earth orbiting (LEO) satellites with coarse temporal sampling. Moreover, infrared sensors on-board geostationary (GEO) platforms can provide high temporal observation, but their cloud top information is not always physically related to precipitation microphysical properties [2]. Studies show that using infrared data with radar calibration can provide more accurate estimation [2-5].

Rainfall estimation algorithms using infrared data can also be classified into three groups depending on the level of information extracted from infrared cloud images: (a) pixel-based; (b) local-texture-based; and (c) patch–based algorithms. In pixel-based algorithms, a rain rate (fixed or variable) is assigned to every pixel of the cloud and just that pixel alone is considered. The cloud local-texture-based technique calculates pixel rain rates by considering a range of the neighborhood pixel coverage. Cloud-patch-based techniques use cloud coverage under a specified temperature threshold [1].

Cluster Ensemble techniques combine multiple data partitions from different clustering. There are different cluster ensemble methods [6] such as voting-based [7-8], evidence accumulation [9], and link-based [10].

In this work, the LCE method which is recently developed [10] is employed to a cloud-patch-based HSPE to cluster cloud patches. The scope of this paper is as follows. Section 2 describes the methodology and the cluster ensemble. Section 3 discusses the validation results, and section 4 presents a summary of the paper.

978-1-4577-1005-6/11/$26.00 ©2011 IEEE

2. METHODOLOGY

A block diagram of the HSPE algorithm, a modified PERSIANN-CCS [1] is depicted in Figure 1. The infrared images from GOES-12 are calibrated into brightness temperatures. The next step is to segment the clouds into patches by using a region growing segmentation method [1]. Figure 2.a depicts the clouds at 0615 on 04 February 2008 and the corresponding cloud patches are shown in Figure 2.b. Then, the shape, statistic, and texture features of each patch are extracted. The statistic features are minimum mean and standard deviation of the brightness temperature of each patch. The texture features, including wavelet, Grey-Level Co-occurrence Matrix (GLCM), local mean and local standard deviation, are calculated. In the next step, the features are classified using a cluster ensemble. Finally, a Temperature–Rain Rate (T-R) curve generated by a polynomial curve fitting technique and Probability Matching (PMM) is assigned to each cluster. Note that the big differences between the HSPE algorithm and the PERSIANN-CCS is that the HSPE is enriched with more texture features such as wavelet and GLCM and also it uses a polynomial curve fitting instead of an exponential curve fitting. In addition, the PERSIANN-CCS employs a SOM [11] for cloud patch clustering.

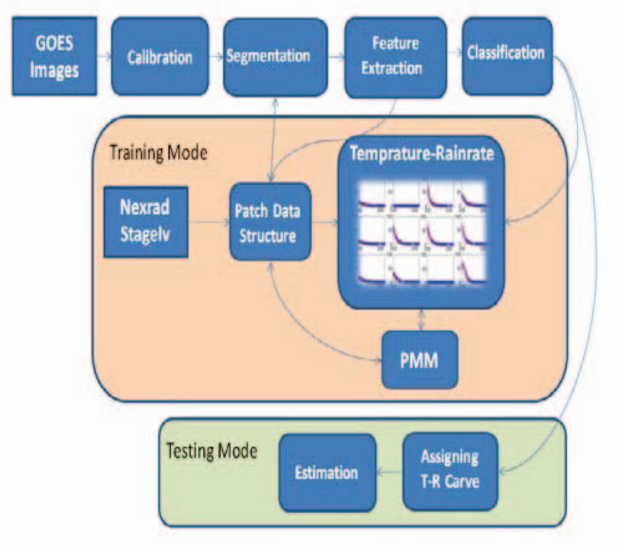
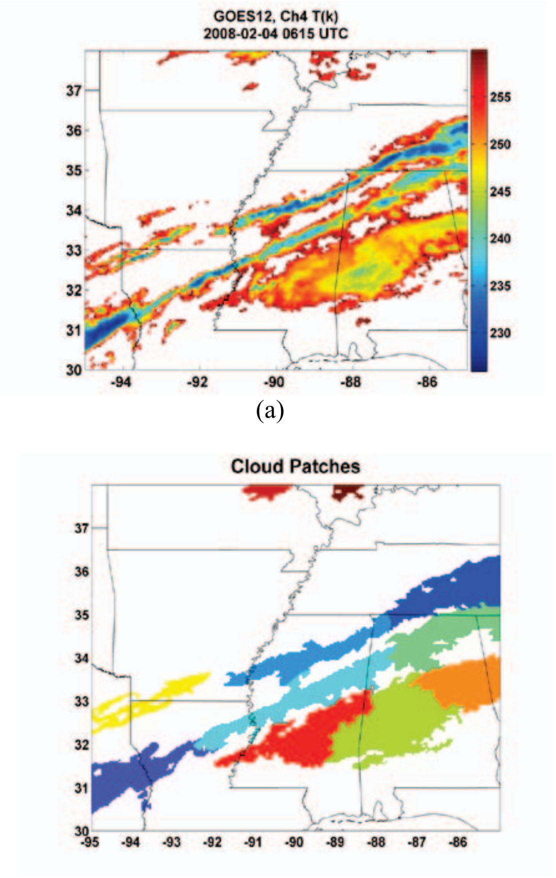


Figure 1. Block diagram of the HPSE algorithm

The cluster ensemble using link-based applied in this study includes three steps [10]: 1) Creating M-base clustering; this can be performed either using a single clustering, for instance the Kmeans with different initialization or multiple clustering algorithms such as SOM, Kmeans, and Fuzzy Cmean. In this work, we have examined the kmeans with different initialization and 2) based on the results of step 1, a cluster-association matrix is generated. Each entry in this matrix represents an association degree between each sample and each cluster of the base clustering. If a sample belongs to a cluster, the corresponding entry of the cluster-association matrix for the sample and the cluster is one; otherwise, similarities between the clusters are considered. The following formula shows how the cluster-association matrix (RM) is calculated [4]:



(a)

(b)

Figure 2. (a) Cloud top brightness temperature at0615 UTC on 04 Feb 2008, (b) the corresponding cloud patches

.

$$RM(x_i, cl) = \begin{cases} 1 & if\ cl = C^* \\ sim(cl, C^*) & otherwise \end{cases} \qquad (1)$$

where x_i and cl are a sample and a cluster, respectively. If the x_i belongs to cl , $RM(x_i, cl) = 1$. If not, $RM(x_i, cl)$ is the similarity between the cluster cl and the cluster C^*, which x_i belongs to it. The similarity between any pair of clusters is defined based on the Connected-Triple method [10], where a subgraph of three clusters with two non-zero edges are considered for each pair of clusters. The following formula shows how the similarity between two clusters is calculated.

$$sim\left(C_i\, , C_j\right) = \frac{WCT_{ij}}{WCT_{max}} \times DC \qquad (2)$$

where

$$WCT_{ij} = \sum_{k=1}^{q} WCT_{ij}^{k}\, , \qquad (3)$$

$$WCT_{ij}^{k} = min\left(w_{ik}, w_{jk}\right), \qquad (4)$$

and $w_{ij} = \frac{|L_i \cap L_j|}{|L_i \cup L_j|}$. L_i denotes the samples belonging to cluster C_i, and q represent all triples between the C_i and C_j. DC is also a constant delay factor [10].

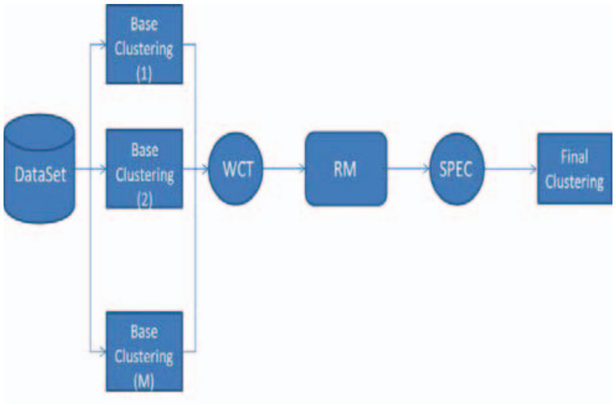


Figure 3. block diagram of the link-based cluster ensemble

In step 3, a consensus function is applied to obtain final clustering. The consensus function is a graph-based clustering so the cluster-association matrix is transformed to the weighted bipartite graph, and then spectral graph partitioning is performed. Figure 3 shows a block diagram of the link-based cluster ensemble. The data are partitioned by different base clusters, and the WCT are calculated. Then, the RM matrix is obtained and final consensus clustering is applied using spectral clustering (SPEC).

3. RESULTS AND VALIDATION

The study region covers an area of the United States extending between 30N to 38N and -95E to - 85E during January and February 2008. The training data is obtained one month before the respective testing month. The IR brightness temperature observations are obtained from the GOES-12 satellite. The National Weather Service Next Generation Weather Radar (NEXRAD) Stage IV precipitation products are used for training and validation. Also, we use the PERSIANN-CCS precipitation estimates (obtained from the PERSIANN group) for comparing the results. The IR data from GOES-12 (Channel 4) has 30-minute interval images that cover the entire area of study.

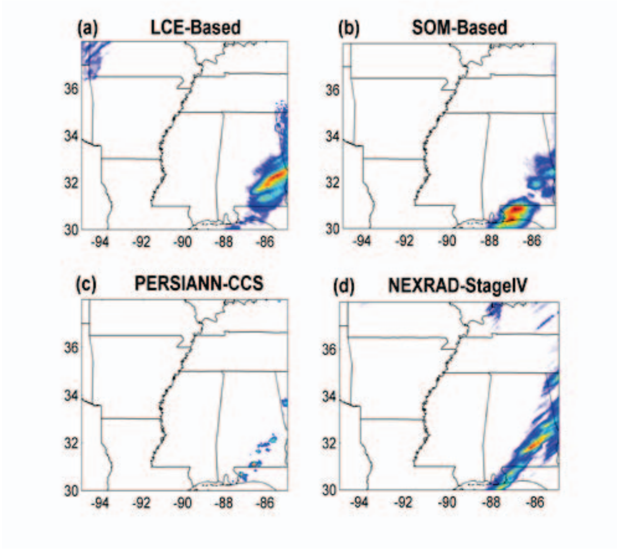


Figure 4. Estimated hourly rainy area ending at 1500 UTC on February 6, 2008: (a) LCE-based; (b) SOM-based; (c) PERSIANN-CCS; and (d) NEXRAD-Stage IV

Figure 4 shows an example of hourly rainfall estimates of the LCE-based, SOM-based, PERSIANN-CCS, and NEXRAD-Stage IV over the area of study ending at 1500 UTC on February 6, 2008. This figure shows that the hourly estimate based on the cluster ensemble is more similar to NEXRAD-Step IV than that of SOM-based in this case.

A set of 3 verification metrics, commonly used in the precipitation verification community [12], are used to compare the performance of the algorithms. These metrics include the Probability of Detection (POD), the False-Alarm Ratio (FAR), and the Heidke Skill Score (HSS).

Figure 5 shows the daily estimate verification for the HSPE algorithm using SOM and LCE as well as the PERSIANN-CCS algorithm (also called “CCS”) against the NEXRAD Stage IV at rainfall threshold levels of 0.01, 0.1, 1, 2, 5, 15, and 25 for winter 2008. Figure 5.a depicts the FAR for the three algorithms. As it is observed, the FAR ratio of the LCE-based algorithm is less than that of the SOM-based almost at all rainfall thresholds. The PERSIANN-CCS also has less FAR at high rainfall, but at low rainfall level, it is almost similar to other algorithms. Figure 5.b shows the POD for the three algorithms. As it can be seen, the POD of the LCE-based is larger than those of the two other algorithms almost at all rainfall thresholds. About 12 % increase in POD is obtained at medium and high rainfall thresholds when the cluster ensemble is used. The LCE and SOM-based algorithms have better POD compared to the PERSIANN-CCS at all rainfall thresholds. Figure 5.c shows the HSS of the algorithms. The HSS of the

LCE-based is larger than the two other algorithms at all rainfall threshold levels.

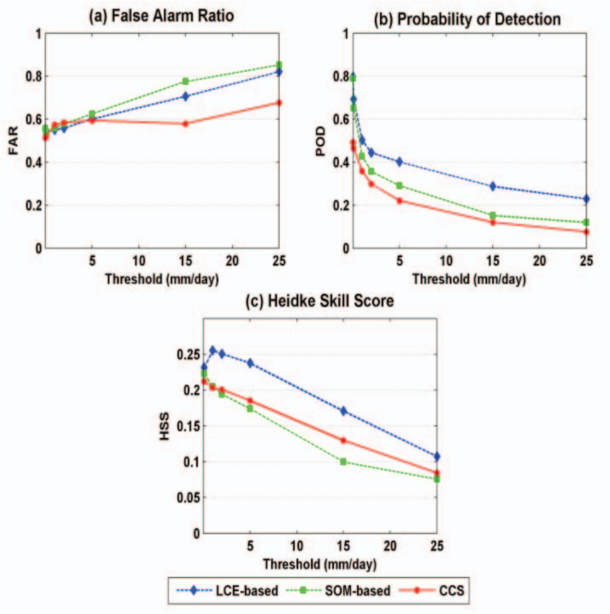


Figure 5– Verification result for January through March 2008: (a) False Alarm ratio; (b) Probability of Detection; and (c) Heidke Skill Score

At very low rainfall thresholds, the resulting HSS from the LCE-based algorithm is slightly larger than that of the SOM-based by approximately 1%. At thresholds of 1 to 10, around 5% improvement is obtained. Moreover, the percentage of improvement at a threshold of 15mm is about 8%. As it is observed, the HSS of the SOM-based and the PERSIANN-CCS are almost the same except at medium rainfall thresholds.

4. SUMMARY

In this study, the link-based cluster ensemble method is used and examined in a HSPE algorithm, which is similar to the PERSIANN-CCS algorithm. The LCE clustering includes three steps: 1) creating an M-base clustering, 2) generating a cluster-association matrix, and 3) applying a consensus function to obtain the final clustering. In comparison with the SOM-based and the PERSIANN-CCS algorithms, the cluster ensemble method improves the POD and HSS at all rainfall thresholds. This improvement is about 5% to 7% for HSS at medium and high level rainfall thresholds for winter 2008.

5. REFERENCES

[1] Y. Hong, K. L. Hsu, S. Sorooshian, and X. G. Gao, “Precipitation Estimation from Remotely Sensed Imagery using an Artificial Neural Network Cloud Classification System,” Journal of Applied Meteorology, vol. 43, pp. 1834-1852, 2004

[2] R. J. Joyce, J. E. Janowiak, P. A. Arkin, and P. Xie, “ CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution,” Journal of Hydrometeorology, vol. 5, pp. 487-503, 2004.

[3] G. J. Huffman, R. F. Adler, D. T. Bolvin, G. J. Gu, E. J. Nelkin, K. P. Bowman, Y. Hong, E. F.Stocker, and D. B. Wolff, “ The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales,” Journal of Hydrometeorology, vol. 8, pp. 38-55, 2007.

[4] F. J. Turk, and S. D. Miller, “Toward improved characterization of remotely sensed precipitation regimes with MODIS/AMSR-E blended data techniques,” IEEE Trans. Geosci. Remote Sens., vol. 43, pp. 1059–1069, 2005.

[5] Sorooshian, S., K. L. Hsu , X. Gao , H. V. Gupta , B.Imam and D. Braithwaite, “Evaluation of PERSIANN system satellite based estimates of tropical rainfall,” Bull. Amer. Meteorol. Soc., 81, 2035, 2000

[6] A. Topchy, A. K. Jain, W. Punch, "Clustering ensembles: models of consensus and weak partitions," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.27, no.12, pp.1866-1881, Dec. 2005.

[7] H. G. Ayad and M. S. Kamel, “On voting-based consensus of cluster ensemble,” Patter recognition J., vol 43, pp. 1943-1953, 2010.

[8] H. G. Ayad, M. S. Kamel, "Cumulative Voting Consensus Method for Partitions with Variable Number of Clusters," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.30, no.1, pp.160-173, Jan. 2008.

[9] A.L.N. Fred, A. K., Jain, "Combining multiple clustering using evidence accumulation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.27, no.6, pp. 835- 850, Jun 2005.

[10] N. Iam-on, T. Boongoen, and S. Garrett “LCE: a link-based cluster ensemble method for improved gene expression data analysis,” Bioinformatics, vol.26, pp. 1513-1519, 2010.

[11] T. Kohonen, “Self-organized formation of topologically correct features maps,” Biol. Cybernetics, vol. 43, pp. 59–69, 1982.

[12] E.E. Ebert, J.E. Janowiak, and C. Kidd. “Comparison of near-real-time precipitation estimates from satellite observations and numerical models,” Bull. Amer. Meteor. Soc., pp. 47-64, 2007.