Analyzing the SOFM Clustering Approach for Precipitation Detection and Estimation

Ali Behrangi

George J. Huffman
(mentor)

NASA-GSSP2008

1. Introduction

1.1 Importance of precipitation estimation

Advancement in measurement/estimation of the amount, temporal and spatial distribution of precipitation is one of the most challenging research problems in earth science due to its wide range of applications in different scales from global-climate down to local-weather. Precipitation is one of the most critical components of the Earth's hydrological cycle and it is essential to provide the fresh water that sustains life. On the other side, floods are among the most frequent and costly natural disasters in terms of human hardship and economic loss. Flooding results in damage to property, crops and negative impacts on human welfare. Since 1997, annually more than 5000 lives are lost, at least 15,000,000 people are displaced and more than $10 billion damage is reported worldwide due to occurrence of flood after a heavy rainfall (Brakenridge et al., , 1997-2006 data). According to the report by NWS, flooding is the 1st ranked source of damage in the US comparing to the other weather related events with an average of $4.6 billion a year over 20 years of 1984-2003. (http://www.weather.gov/floodsafety/floodsafe.shtml).

1.2 Methods of precipitation measurement/estimation

Different instruments are used to measure/estimate rainfall amount and intensity. Rain gauges have been a primary source of precipitation measurement for a long time. Although rain gauges directly measure the amount of rainfall, they have significant shortcomings for flash-flood forecasting purposes (Smith et al., 1996). Gauge networks are inadequate in many parts of the world and can not cover all area of interest to make a real time observation. Meanwhile, it is impractical and not cost-effective to have a dense network of gauges that can measure the highly variable spatial distribution of rainfall in a large domain of area.

Ground-based radars provide additional support for regional precipitation monitoring with high spatial and temporal resolution. Radars observe the back-scatter of electromagnetic radiation from liquid water drops. However, they cover only a limited area and because of this, numbers of radar are required to cover a wide area of interest. Such a system is not necessarily cost- effective. Furthermore, it is not feasible to site ground based radars to measure rainfall over oceans and also they are not efficient over mountainous regions due to wave blockage by mountains. Other issues include attenuation, beam-filling, and beam overshoot (i.e., Sauvageot, 1994).

Instead, the space-based technology for large-scale measurement of rainfall is a promising tool. Satellites can provide near-real time observation of precipitation

at fairly high resolution. The history of rainfall estimation from space dates back to the 1970s (Hossain, 2006). Since then a variety of sensors have been developed and launched to observe the atmosphere. These sensors fall into two main categories: a) visible(VIS) and Infrared (IR) sensors available from Geostationary Earth Orbit (GEO) and Low-Earth Orbiting (LEO) satellites and b) Passive microwave (PMW) and active radar sensors currently only available onboard LEO satellites. VIS/IR imagers from GEO satellites can provide high resolution images both in time (one hour or less) and space (0.05° or less), However they only give cloud-top information and can not be used for direct measurement of rainfall. As a result many GEO-based algorithms, for precipitation estimation, rely on the effect of scale averaging for accuracy improvement(i.e., Arkin and Meisner, 1987).

In contrast to the VIS/IR techniques, more direct estimation of precipitation is obtained via techniques utilizing passive microwave (PMW) data. PMW data have strong physical interaction with hydrometeors in the entire atmosphere and therefore, provide more accurate instantaneous estimate of precipitation. However, their poor temporal and spatial resolution imposes serious limitations. In addition, the heterogonous and in general unknown emissivity of land surfaces(Ferraro et al., 1994) limits their reliability over the land surface.

2. Research Motivation

Due to capability of PMW sensors to provide more accurate estimate of precipitation number of PMW sensors has been increased since last ten years. However, even with the future Global Precipitation Measurement (GPM) Mission the temporal and spatial resolution of precipitation estimation is still not enough for many applications. For extreme hydrologic events such as flash floods, precipitation observation at high sampling frequency and spatial resolution during the storm evolution stage is critical as the life of a convective storm from growing to dissipation can occur in a matter of 1 hour or less.

Given the ability of GEO satellite to provide frequent VIS/IR imagery, Attempt to improve accuracy, coverage and resolution of precipitation estimate via integration of information from GEO (VIS/IR) and LEO (PMW) has been an active area of research since almost last three decades (Adler et al., 1993: Huffman et al., 2001: Huffman et al., 2007: Joyce et al., 2004: Kidd et al., 2003: Kummerow and Giglio, 1995: Levizzani et al., 1996: Miller et al., 2001: Sorooshian et al., 2000: Todd et al., 2001: Turk et al., 2000: Turk et al., 2003: Xu et al., 1999). Joyce et al.(2004) proposed The Climate Prediction Center morphing (CMORPH) method that exclusively estimates precipitation from PMW observations. They used spatial lag correlations on successive GEO-IR images to propagate PMW-derived precipitation estimate in space and time via "morphing" for times when updated PMW data are not available. However, one

important shortcoming of CMORPH, as mentioned by the authors, is that formation and dissipation of precipitation between PMW overpasses can not be detected. Another approach is using precipitation estimates from PMW to calibrate the GEO-derived algorithms. Kidd et al. (2003) argued that because of sample size requirements, combined/blended algorithms, in which PMW data is used to calibrate IR-based rainfall relationships, pose a tradeoff problem between the temporal and spatial details. As they indicated, when longer time scales are used for calibration, the spatial details can be retained at the expense of short-term variation in the IR-rainfall relationship. On the other hand, when calibration is performed using coincident PMW-IR images, the short term variability of IR Rainfall relationships are captured at the expense of information regarding their spatial variability. Another combination strategy, which focuses on obtaining the "best" local estimate at the times of PMW observations, as opposed to finer temporal resolution time series (Hufman, et. al., 2007), relies on using available PMW estimates from various satellites and then filling the gaps (spatially) with calibrated IR estimates. Due to the recognized importance of employing GEO-based VIS/IR information for precipitation estimation, whether in stand-alone mode or in combination, Behrangi et al. (2008a,b) incorporated other spectral bands (in addition to IR alone) into a rain detection and rainfall estimation algorithms using the self *organizing Feature Map (SOFM)* classification technique (Kohonen, 1982). They reported that multi-spectral data,

available from GEO satellite, can yield additional information relevant to precipitation. Therefore, using additional bands was recognized as an alternative to improve GEO-based estimation algorithms, which to some extent provide better foundation for combining PMW-VIS/IR rain-estimates. As a part of the GSSP program and in continuation of the author's previous work, we decided to analyze the clustering technique by visualizing the SOFM clusters in a selected precipitation event using data obtained from our previous multi-spectral experiments. This would help us to understand how effective SOFM clusters can represent precipitation field and observed input features. The clustering approach was also used to study diurnal cycle of precipitation over Florida. A brief review of the clustering technique for precipitation retrievals is provided in section 3. In section 4 clustering approach is analyzed using a precipitation event and result of the diurnal cycle experiment is reported in section 5. Finally, Section 6 reflects our ongoing efforts towards combining VIS/IR and PMW information for precipitation estimation.

3. Multi-spectral precipitation retrieval algorithm

Multi-spectral rain detection and estimation algorithms (Behrangi et al., 2008) use the SOFM technique for classification of input features. Since in the present study we aimed to analyze this clustering approach, a brief overview of SOFM is provided below.

3.1. *Self-organizing Feature Map (SOFM)*

Breaking up the input space into a number of classes is a typical statistical pattern recognition problem, which can be accomplished using a number of methods. A neural network based classification algorithm, the (SOFM, Kohonen, 1982) can be used to classify n-dimensional input (hereafter referred as feature) space into a number of groups (hereafter referred as clusters). SOFM permits the classification without dependency to precipitation field (see Hsu et al. 1999). SOFM can divide a multi-dimensional feature space into a predetermined number of clusters according to local clustering. These clusters are arranged into two dimensional discrete map which preserves topological order.

The process of training the SOFM consists of presenting input vectors one by one from the training dataset to the network. The input vector must be normalized in each feature dimension so that they are comparable. The distance (d) between each normalized input vector (x_i, $i=1,\ldots n_0$) and the SOFM cluster centers calculated as

$$d_j = \left[\sum_{i=1}^{n_o} (x_i - w_{ij})^2 \right]^{1/2}, \; j = 1,\ldots,n_1 \qquad (1)$$

where n_1 is predefined total number of clusters and w_{ij} is the connection parameter (weight) of SOFM from input feature *i* to the specified cluster *j*.

The best matching SOFM cluster, *c* (winning node), is defined from the minimum distance between the input feature vector and the SOFM connection weights w_{ij} as follow

$$d_c = \min(d_j) \qquad , j = 1,\ldots,n_1 \tag{2}$$

Through a recursive process of competitive cluster selection and weight adjustment, location of the cluster centers becomes stable. After training, the trained SOFM has the ability to assign any arbitrary feature vector x_i to a SOFM cluster (with fixed weights) according to their minimum distance.

3.2. Precipitation detection

The SOFM-based precipitation detection algorithm (Behrangi, 2008) is capable of applying multi-spectral data (or textural information) to distinguish rain and no rain areas by implementing a sequence of three main steps: (1) unsupervised classification of input features from satellite imagery to a number of predetermined clusters; (2) supervised computation of the probability of

precipitation (POP) for each classified cluster, and (3) separation of rain and no-rain clusters using the pattern matching technique.

POP for each cluster (*k*) is calculated using available rain-rate as:

$$POP_k = R_k / (R_k + D_k) * 100, \quad \text{where } k=1...n_1$$

Where, R_k and D_k are the total number of precipitation and no-precipitation counts of each cluster respectively.

Input features can include multi-spectral data and textural features (i.e., STD and mean) extracted from neighborhood pixels for each spectrum. The pattern-matching technique introduced by Lovejoy and Austin (1979) can be used to define the optimum POP threshold. This method was also used by Cheng et al (1993). Through this technique the SOFM clusters, are sorted in order of decreasing POP. Starting from a cluster with the highest POP, total number of precipitation counts (obtained from all clusters) is distributed to each cluster equal to its total count (obtained from both precipitation and no-precipitation counts). The distribution is continued until the sum of the clusters count (precipitation counts) becomes closest to the total precipitation counts. The POP of the cluster at which this occurs, is defined as the optimum POP threshold. Therefore, Clusters having POP greater than optimum POP threshold are

defined as rain clusters and clusters having POP less than optimum POP threshold are defined as no-rain clusters.

3.3. Estimation of precipitation rate using multi-dimensional inputs

This algorithm is also based on the SOFM classification. The algorithm consists of six separate stages: (1) extraction of input features from multi-spectral data, (2) unsupervised classification of the input features to a number of predetermined clusters, (3) supervised computation of mean precipitation rate (MPR) for each classified cluster, (4) ranking and sorting the clusters based on their MRR, (5) redistribution of sorted precipitation rates to the ranked clusters such that the cluster with the highest MRR will contain the highest precipitation rates, and (6) averaging the sorted precipitation rates in each cluster (MRR2), resulting in a rain rate estimator.

4. Analyzing the clustering approach in a precipitation event

Figure 1 visually demonstrates SEVIRI images of brightness temperature (IR10.8 μm) and albedo (from VIS 0.6μm) from a precipitation event at 1157 UTC 9 January 2005. The Rain rate estimates from PMW, single IR, and bi-spectral scenarios are also shown in figures 2a-c respectively. Since detailed comparison of the selected scenarios may not be an easy task, ellipses and squares are shown in the images to represent areas with major false detection and major misses. In

other words, ellipses and squares show cases of non-raining cold-thin clouds and raining warm-cloud respectively, which are both problematic for single IR algorithms. Thus, using both brightness temperature and albedo images precipitation retrieval is improved.

For the present experiment, rain estimate from PMW is assumed to be "truth" and was used as reference. The bi-spectral scenario incorporates visible (0.6um) and IR (10.8 um) bands and Figure 3 demonstrates its corresponding SOFM clusters, arranged in a 2-D network with a size of 20x20. Comparison using maps of the albedo (Fig. 3a), brightness temperature (Fig 3b), MRR (Fig 3c), and MRR2 (Fig 3d), agrees with our expectation in such the brighter and the colder cloud is associated with higher rain rate (see zone B) and no precipitation is expected for hot and low-reflective regions (see zone D).

In order to track clusters in the precipitation field a 2D-colorbar was generated and a map of corresponding clusters, colored with the 2D-colorbar was plotted (the 2D-colorbar is shown in the bottom-right of the figures 1c and 1d). In Figures 1c, d the black line delineates raining areas, detected by PMW sensor and the SOFM-based rain estimator, using IR +VIS) respectively. Comparison between Figures 1, 2 and 3 reveals how clusters can classify satellite images to different groups, providing basis for rain detection and estimation. Similarity of neighborhood clusters is also inferred from Figures 1 and 3, which results in having a fairly smoothed field of estimated rainfall (Figure 2c and 2d). Indeed, by

implementing this approach to different events, it is possible to see how different-cloud-types are classified by the clusters, providing a valuable tool to evaluate the performance of the described estimation algorithm.

5. Diurnal cycle study using clustering approach

Using the described clustering approach, diurnal cycle of precipitation was also studied. As found previously, water vapor band in conjunction with IR band can improve IR-only methods and visible band adds significant information to a single IR band during. Figure 4 shows diurnal cycle of precipitation obtained using GOES data (spatial and temporal resolution: 0.04° lat-long grid-boxes and half hourly respectively) during June-Aug 2005. This study was conducted over Florida peninsula and hourly NEXRAD (0.04° lat-long grid-boxes) rain estimate were used as ground truth for reference.

As seen in figure 4, improvement in representing the diurnal cycle is achieved by addition of water vapor (WV, 6.7μm) and albedo channels to the single IR band during daytime and nighttime. As shown in Figure 4, by using additional channels, in general, improvement is more obvious for rain detection than rain estimation.

6. Integration of information from VIS/IR and PMW sensors

As described in section 2, precipitation-related PMW data have strong physical relationship to the hydrometeors and result in more accurate precipitation estimation. However, even taken together, average coverage in the current 3-hourly PMW estimates is about 80% of the earth's surface in the latitude band 50ºN-S (Huffman et al. 2007). This will not fulfill many scientific communities' request (e.g., hydrology) due to the infrequent global precipitation estimate. Some of the previous attempts to ameliorate the problem using combined PMW-VIS/IR imagers were reviewed in section 2. As another alternative solution, which is currently under study, is using a high resolution cloud tracking technique (Bellerby, 2004) to capture the advection of grid points in two subsequent GEO-based images to define motion streamlines for each pixel. At each grid point along a streamline, an empirical cloud-motion and rainfall estimation model is assigned to incorporate convection, advection, and orographic enhancement into the precipitation estimation. The parameters of the above-described model are sequentially adjusted by Kalman filtering when PMW rainfall is available. <need a reference?>

Another approach, discussed during GSSP, lead to the idea of using the above mentioned tracking algorithm to interpolate PMW rain-estimates in time and employing that in conjunction with VIS/IR rain-estimates. As shown in figure 5, within a fixed grid-box, significant drop in correlation between consecutive half-hourly PMW rain-estimates exists. This implies that any attempt to use PMW

estimate after half-hour is associated with significant error, unless we track the precipitation field. The experiment was conducted within a box (longitude: 30°W-0°, latitude: 15°S-15°N) under SEVIRI's coverage. For this study, intercalibrated PMW-derived rain rate estimates at 1/8° lat-long resolution (provided by NOAA CPC (Joyce et al., 2004) were used during the first 9 days of June 2005.

7. Summary

Analyzing the clustering approach described above, both event-based cluster-visualization and diurnal cycle studies show that the clustering approach is a reasonable and effective way for multi-spectral precipitation retrievals. It was also found that implementing a proper tracking algorithm is important for interpolating PMW rain-estimate in time. Different approaches were discussed for combining PMW-VIS/IR, which initiated a follow up research.

8. Acknowledgements: I especially would like to thank my mentor Dr. George J. Huffman for his comments and encouragement during GSSP2008. I am also thankful to Ms. Valerie Casasanto for making GSSP a smooth and nice experience.

References

Adler, R. F., A. J. Negri, P. R. Keehn, and I. M. Hakkarinen, 1993: Estimation of Monthly Rainfall over Japan and Surrounding Waters from a Combination of Low-Orbit Microwave and Geosynchronous IR Data. *Journal of Applied Meteorology*, **32,** 335-356.

Arkin, P. A., and B. N. Meisner, 1987: The Relationship between Large-Scale Convective Rainfall and Cold Cloud over the Western-Hemisphere during 1982-84. *Monthly Weather Review*, **115,** 51-74.

Behrangi, A., K. L. Hsu, S. Sorooshian, and R. J. Kuligowski, 2008: Evaluating the Utility of Multi-Spectral Information in Delineating the Areal Extent of Precipitation, submitted to J. Hydrometeorology, July 2008.

Bellerby, T. J., 2004: A feature-based approach to satellite precipitation monitoring using geostationary IR imagery. *Journal of Hydrometeorology*, **5,** 910-921.

Brakenridge, G. R., E. Anderson, and S. Caquard: Flood Inundation Map DFO 2003-282, Dartmouth Flood Observatory, Hanover, USA, digital media, http://www.dartmouth.edu/~floods/ (1997_2006 tables).

Cheng, M., R. Brown, and C. G. Collier, 1993: Delineation of Precipitation Areas Using Meteosat Infrared and Visible Data in the Region of the United Kingdom. *Journal of Applied Meteorology*, **32,** 884-898.

Ferraro, R., N. Grody, and G. Marks, 1994: Effects of surface conditions on rain identification using the DMSP-SSM/I. *Rem. Sens. Rev.*, **11,** 195-209.

Huffman, G. J., R. F. Adler, M. M. Morrissey, D. T. Bolvin, S. Curtis, R. Joyce, B. McGavock, and J. Susskind, 2001: Global Precipitation at One-Degree Daily Resolution from Multisatellite Observations. *Journal of Hydrometeorology*, **2,** 36-50.

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

Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xie, 2004: CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution. *Journal of Hydrometeorology*, **5,** 487-503.

Kidd, C., D. R. Kniveton, M. C. Todd, and T. J. Bellerby, 2003: Satellite Rainfall Estimation Using Combined Passive Microwave and Infrared Algorithms. *Journal of Hydrometeorology*, **4,** 1088-1104.

Kohonen, T., 1982: Self-organized formation of topologically correct feature maps. Biol. Cybernetics, 43, 59–69.

Kummerow, C., and L. Giglio, 1995: A Method for Combining Passive Microwave and Infrared Rainfall Observations. *Journal of Atmospheric and Oceanic Technology*, **12,** 33-45.

Levizzani, V., F. Porcú, F. S. Marzano, A. Mugnai, E. A. Smith, and F. Prodi, 1996: Investigating a SSM/I microwave algorithm to calibrate Meteosat infrared i nstaneous rainrate estimates. *Meteorol. Appl.* , **3,** 5-17

Miller, S. W., P. A. Arkin, and R. Joyce, 2001: A combined microwave/infrared rain rate algorithm. *International Journal of Remote Sensing*, **22,** 3285-3307.

Sauvageot, H., 1994: Rainfall Measurement by Radar - A Review. *Atmospheric Research*, **35,** 27-54.

Smith, L. C., B. L. Isacks, A. L. Bloom, and A. B. Murray, 1996: Estimation of discharge from three braided rivers using synthetic aperture radar satellite imagery: Potential application to ungaged basins. *Water Resources Research*, **32,** 2021-2034.

Sorooshian, S., K. L. Hsu, X. Gao, H. V. Gupta, B. Imam, and D. Braithwaite, 2000: Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bulletin of the American Meteorological Society*, **81,** 2035-2046.

Todd, M. C., C. Kidd, D. Kniveton, and T. J. Bellerby, 2001: A Combined Satellite Infrared and Passive Microwave Technique for Estimation of Small-Scale Rainfall. *Journal of Atmospheric and Oceanic Technology*, **18,** 742-755.

Turk, F. J., J. Hawkins, E. A. Smith, F. S. Marzano, A. Mugnai, and v. Levizzani, 2000: Combining SSM/I, TRMM and infrared. geostationary satellite data in a near real-time fashion for rapid precipitation updates: Advantages and limitations. *Proc., EUMETSAT Meteorological Satellite Data Users' Conf., Vol. 2, Bologna, Italy, EUMETSAT, 705–707.*

Turk, F. J., E. E. Ebert, H. J. Oh, B. J. Sohn, v. Levizzani, E. A. Smith, and R. R. Ferraro, 2003: Validation of an operational global precipitation analysis at short time scales. *12th Conf. Sat. Meteor. and Ocean., AMS ,9-13 February, Long Beach, CA, CD-ROM.*

Xu, L., X. Gao, S. Sorooshian, P. A. Arkin, and B. Imam, 1999: A Microwave Infrared Threshold Technique to Improve the GOES Precipitation Index. *Journal of Applied Meteorology*, **38,** 569-579.

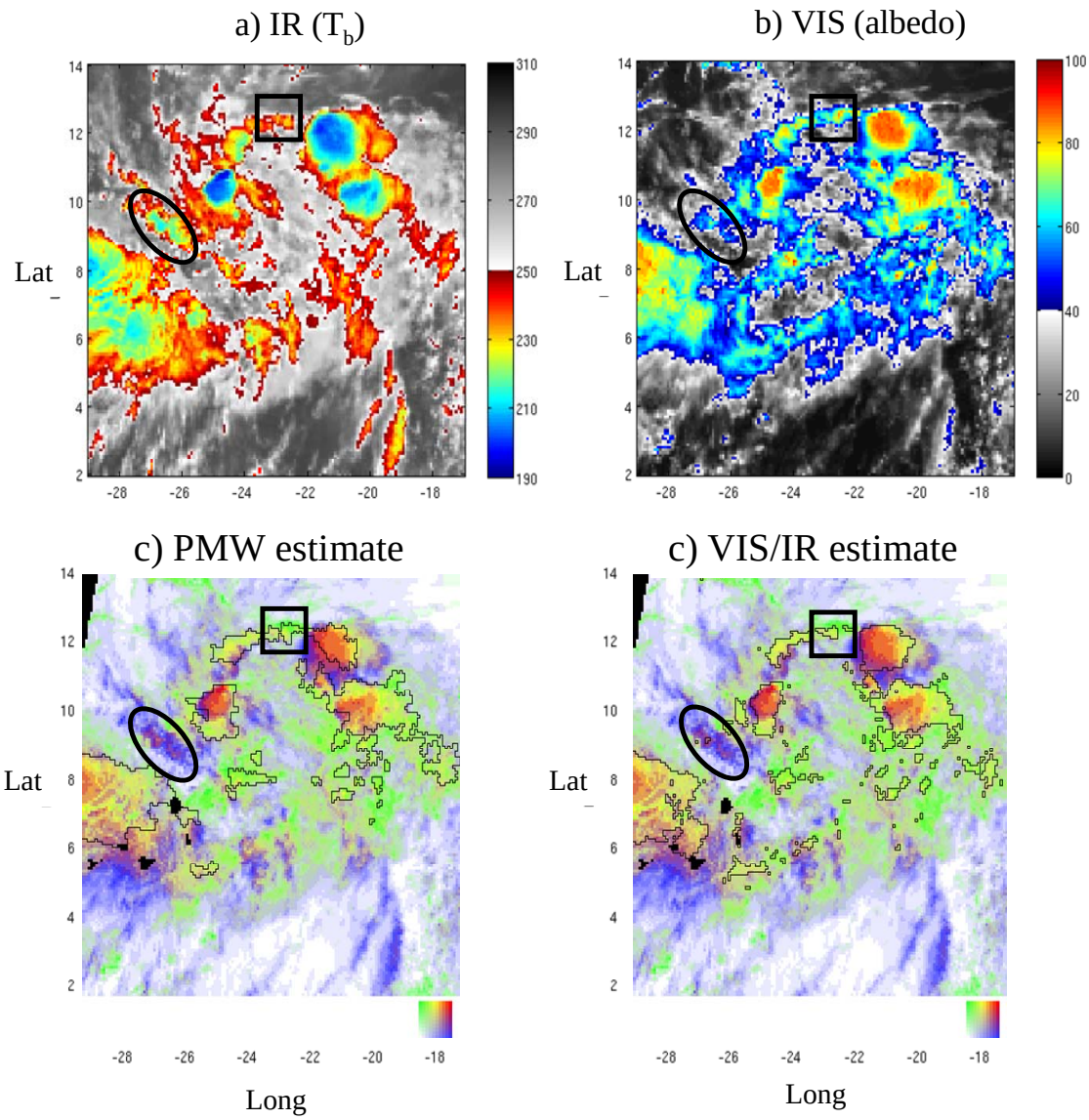

Figure 1. Analyzing the clustering approach during a precipitation event on January 9, 2005 at 1157 UTC: a) map of brightness temperature (10.8 μm), b) map of normalized albedo. In the background of figures c and d, each grid-box of the satellite image is replaced by the corresponding cluster with colors shown in the right-bottom corner of the figures. Raining areas are delineated in c and d using PMW and VIS/IR estimates.

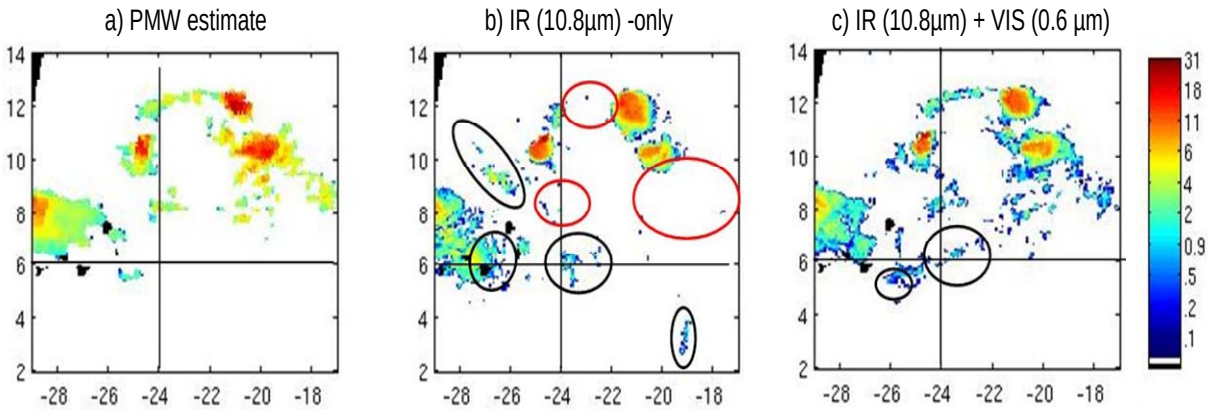

Figure 2. Comparison of rain-rate estimates. a) PMW-derived rain rate. Figures b and c represent rain rate derived from b) single IR and c) VIS/IR imagery through the clustering approach.

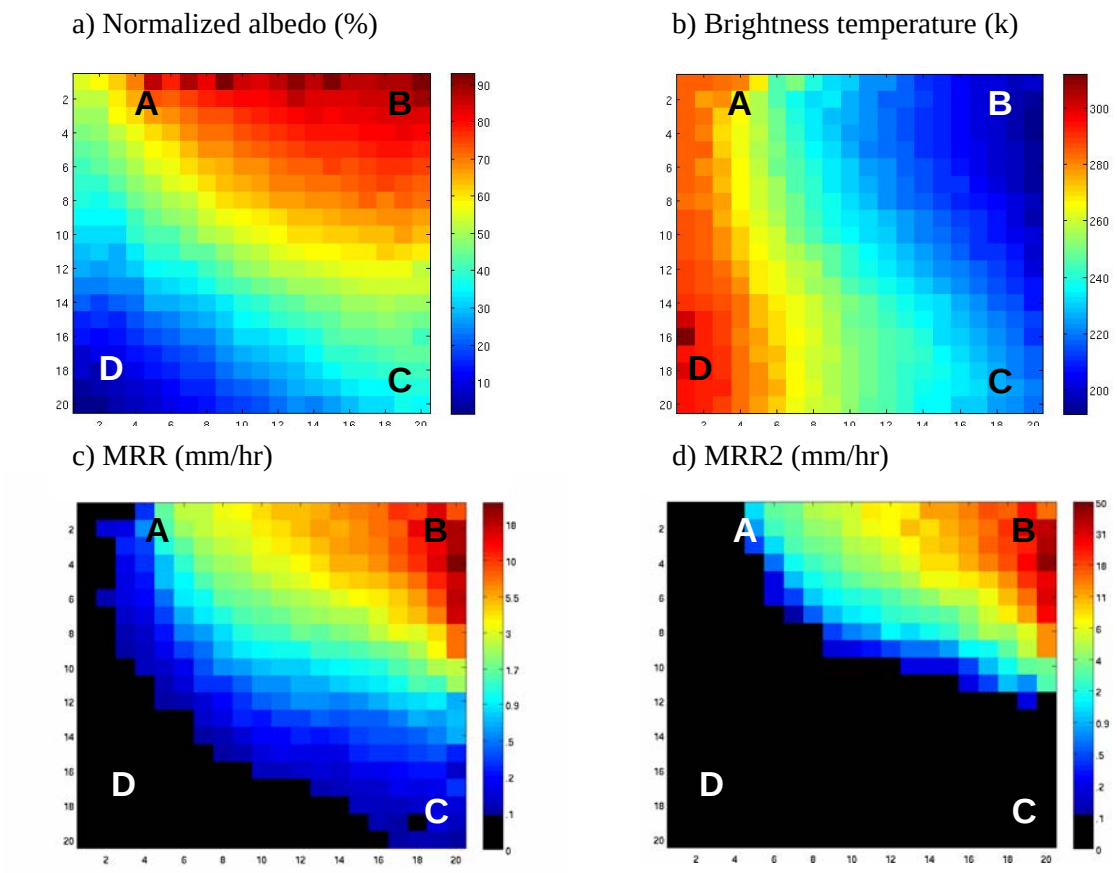

Figure 3: Representation of corresponding SOFM clusters, arranged in a 2-D network with a size of 20x20: a) Albedo(%), b) BT(k), c) MRR (mm/hr), and d) MRR2 (mm/hr). Note that, the brighter and the colder cloud is associated with higher rain rate (see zone B) and no precipitation is expected for hot and low-reflective regions (see zone D). Also, a relatively warm and thick cloud (high albedo) can produce rainfall (see Zone A) while in almost no precipitation is obtained for cold and low-reflective cloud (i.e., Cirrus cloud. see zone C)

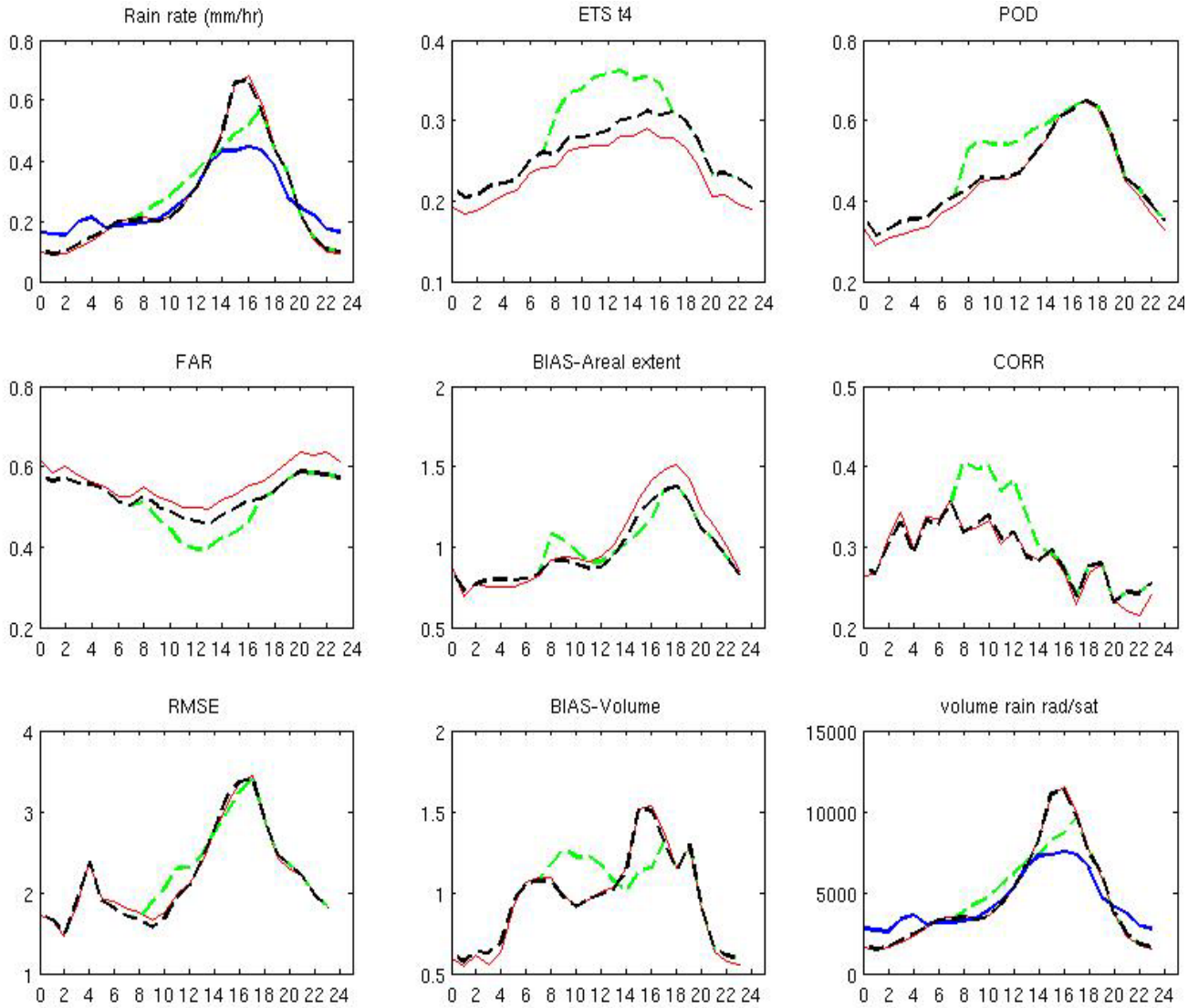

Figure 4. Statistical representation of Diurnal cycle of precipitation over Florida peninsula during June-August, 2006. Red, black and green lines respectively represent IR-only, IR+WV, and IR+VIS scenarios, used to estimate precipitation through using the clustering approach. Blue lines in the first and last images represents NEXRAD rain-rate estimate.

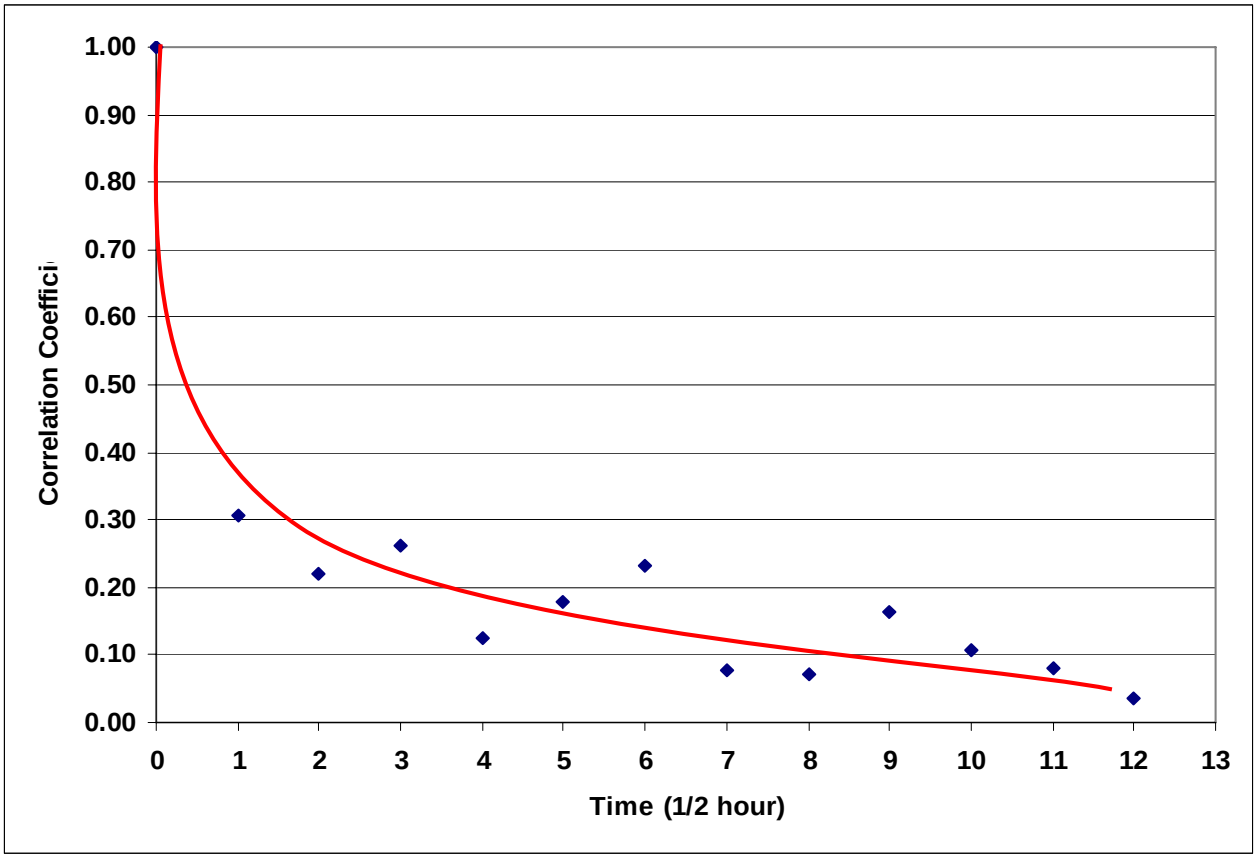

Figure 5. Decay in correlation between consecutive half-hourly PMW rain-estimates, once no tracking is involved.