MAUSAM, **54**, 1 (January 2003), 67-74

551.501.777:551.506

Multifeature classification based rainfall estimation using visible infrared TRMM data

P. K. PAL, C. M. KISHTAWAL and NEERAJ AGARWAL

Meteorology and Oceanography Group

Space Applications Centre, Ahmedabad-380015, India

e mails : pkpal25@hotmail.com & cmk307@rediffmail.com

सार – इस अध्ययन में बहुस्पैक्ट्रल वर्गीकरण पद्धति द्वारा दृश्य और अवरक्त स्पैक्ट्रल चैनलों में संकेतों का उपयोग करके मेघों के वर्षा विभव का आकलन करने का प्रयास किया गया है। इस अध्ययन के लिए उष्णकटिबंधीय वर्षा मापन मिशन (टी.आर.एम.एम.) उपग्रह में प्रयुक्त होने वाले दृश्य अवरक्त स्कैनर (वी.आई.आर. एस.) और टी.आर.एम.एम. सूक्ष्म तरंग उपकरण (टी.एम.आई.) से आँकड़े प्राप्त किए गए थे। वर्गीकरण के लिए 15 वी.आई.आर.एस. व्युत्पन्न प्राचलों का उपयोग किया गया है और के–माध्य वर्गीकरण एलगोरिथ्म का उपयोग करके मेघों को 24 वर्गों में विभजित किया गया। कुल वर्षा (~80 प्रतिशत) की तुलना में इन 24 वर्गों में से 6 में उच्च वर्षा प्रायिकता और उच्च संचयी योगदान पाया गया। इन 6 वर्गों के लिए वी.आई. आर.एस. प्रेक्षणों से प्राप्त किए गए लक्षणों के संबंध में टी. एम. आई. वर्षा की दर की व्याख्या करने के लिए समाश्रयण विश्लेषण किया गया। इसमें वर्गीकरण और जॉच के परिणामों पर चर्चा की गई है।

ABSTRACT. In this study an attempt has been made to estimate the rain potential of clouds using the signatures in visible and infrared spectral channels by multi spectral classification approach. The data for this study was obtained from Visible Infrared Scanner (VIRS) and TRMM Microwave Instrument (TMI) onboard Tropical Rainfall Measuring Mission (TRMM) satellite. Fifteen VIRS derived parameters have been used for classification and the clouds were separated into 24 classes using K-Mean classification algorithm. Six out of these 24 classes were found to have high raining probability as well as high cumulative contribution to the total rainfall (~80%). A regression analysis has been performed to explain the TMI rainfall rate in terms of features derived from VIRS observations for these six classes. The results of classification and verification have been discussed.

Key words – TRMM, Cloud classification, Rainfall, VIRS, Remote sensing.

1. Introduction

Measuring rainfall on a variety of temporal and spatial scales is one of the greatest challenges in atmospheric sciences. Accurate rain-rate measurements are required for wide range of applications, climate monitoring, drought detection, model verification, hydrological budget, modeling and severe weather and flash flood warnings. In remote sensing the retrieval of rainfall is most complex because of its high variability on both large and small scales. Much attention is now focused on the potential use of rain-rates to aid the initialization of forecast models by prescribing vertical heating rate profiles that corresponds to the observed rain-rates (Krishnamurti *et al*., 1990, Pal *et al*., 1999). Visible infrared data obtained from satellites can be used to derive reasonable estimates of rainfall (Arkin 1979). But it is noted that visible infrared algorithms for measuring rain rates are less physically direct than passive microwave algorithm thereby becoming less accurate than microwave techniques. Visible and infrared sensors detect more information about clouds than about rain. Microwave sensors are effectively giving better and direct estimates of rain-rates but due to their limited swath and longer repetivity time we cannot get the continuous information about rainfall for a particular place. Therefore geostationary satellites, which are continuously monitoring certain region of the globe, should be taken into use for rainfall estimation. Since geostationary satellites provide Visible/IR data only so there is a need to find a relation between different precipitating systems and rain-rates.

With the launch of Tropical Rainfall Measuring Mission (TRMM) satellite in 1997, it has become possible for the first time to make simultaneous observations of

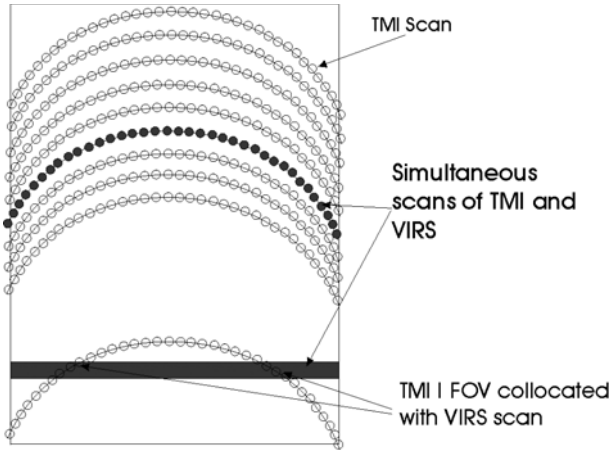

Fig. 1. Scan geometry of TMI and VIRS

raining systems using a variety of sensors like TRMM Microwave Imager (TMI), Visible Infrared Scanner (VIRS), and Precipitation Radar (PR). This combination ensures that one obtains valuable and mutually complimentary information about the rainfall from different sensors of TRMM, and that may facilitate a number of new combined algorithms, using both microwave and VIS/IR observations, for estimating rainfall from space.

In the "multispectral" approach, coincident observations in microwave, IR, and/or VIS spectrum are used for classification of clouds that characterize different types of raining systems. During the learning phase of classification, microwave, IR, and/or VIS measurements are from the coincident set of these observations, while, only VIS/IR measurements are used during the application phase, to locate the rain class of actual measurements within the higher dimension space of the learning phase. This arrangement gives the advantage of high temporal sampling of geostationary orbits as well as the higher physical content of microwave observations.

In the present paper, we have tried to understand how the rain rates derived from microwave observations are related to IR/VIS observations of different cloud types. For the present study, we have used only VIRS data from TRMM for the classification of clouds, based on several physical and textural parameters. This is followed by a statistical analysis of the relationship between different cloud types and corresponding microwave signatures and rainfall rates derived using only microwave observations. In section 2, a brief description is given about the TRMM data used in the present study. This section also describes a simple geometric arrangement used to collocate the observations from different TRMM sensors. Section 3 describes the physical and textural parameters, which were used as input to the clustering algorithm, and the statistical analysis used for making the appropriate selection. Second part of the same section also gives a brief description about the clustering. Results of the cloud classification and the relationship of cloud types with TMI observations of brightness temperatures and rain-rates are discussed in section 4. Finally the conclusions of the study and some suggestions for the improvement are presented in section 5.

2. Data and methodology

The VIRS is a five-channel imaging spectro-radiometer with wavelengths of 0.63, 1.61, 3.75, 10.80, and 12.00 μm. The VIRS is, in many ways, similar to the Advanced Very High Resolution Radiometer (AVHRR) that has flown since 1978 on the National Oceanic and Atmospheric Administration (NOAA) series of spacecraft in that both have the same center wavelengths and bandwidths. The major differences between the two systems is the 2.11 km nadir IFOV of VIRS in contrast to 1.1 km for the AVHRR and the fact that the VIRS has an onboard solar diffuser for post launch calibration of the two reflected solar bands. The noise equivalent differences in temperature (NEΔTs) listed are for a focal plane temperature of 107 K. Predictions for end-of-life hot (worst case) performance for channel 5 indicate a focal plane temperature of 122 K and a NEΔT of 0.13 K.

For the present study, we have used TRMM products 2A-12 (TMI observed passive microwave brightness temperatures, and instantaneous rain rates), and 1B-01 (VIRS radiances at five channels). TMI has a conical scan whereas the VIRS scan is across track. In order to collocate the TMI and VIRS observations, we collected different TMI pixels from different scan lines, which coincided with a VIRS scene, made up of 32 × 32 VIRS pixels. For this, we first located a "central TMI pixel" and a "central TMI scan line" as per the geometric configuration shown in Fig. 1. Central TMI pixel and central TMI scan lines were approximated as a sinusoidal function of VIRS scan line and pixels.

The above approximation function locates the central TMI pixels and scans lines within an average distance of about 0.1 degree from the center of the VIRS scene. The approximate time difference between the observations at this point by TMI and VIRS is less than 1 minute, which has been ignored in this study. Once this point is located, the TMI derived rain rates were averaged over 24 TMI pixels (3 points along and 8 points across the TMI scan

TABLE 1

Different features determined by analyzing VIRS observations
(VIRS Channels : 1 = 6.3μm, 2 = 1.61μm, 3 = 3.75μm, 4 = 10.8μm, 5 = 12.0μm)

Feature	Notation	Parameter	Comments
1	TCF	Cloud Fraction	Ratio of cloudy pixels to total pixels (32 × 32) in a scene
2	α_1	Channel-1 Albedo	Solar zenith angle corrections were made to VIRS raw counts to get albedo
3	CON	Contrast	Contrast of grey counts in channel-1
4	χ	$\chi = 1 + (\alpha_1 + \mathrm{NDVI} - \alpha_3)$	α_3 is the channel-3 albedo, NDVI is defined below
5	NDVI	NDVI	$(\alpha_1 - \alpha_2)/(\alpha_1 + \alpha_2)$. NDVI differentiates water and cloud surfaces better
6	$\delta^2_{4\text{-}5}$	$\delta^2_{4\text{-}5}$ Channel-4=10.8μm Channel-5=12.0 μm	Difference of squares of channel-4 and channel-5 brightness temperatures
7	CC	Cloud Connectivity	A textural feature indicating if a scene is made of a few large clouds, or many broken ones
8	BC	Background connectivity	Reverse, in nature, to cloud connectivity
9	Φ_1	Φ_1	Fractional area of the scene, where α_1 exceeds 0.5
10	Φ_2	Φ_2	Fractional area of the scene, where NDVI exceeds -0.25
11	Φ_3	Φ_3	Fractional area of the scene, where χ exceeds 0.5, $\chi = 1 + (\alpha_1 + \mathrm{NDVI} - \alpha_3)$
12	TMEAN	TMEAN	Mean cloud top temperature
13	TMIN	TMIN	Minimum of the temperature in 32 × 32 VIRS scene
14	ASYM	Asymmetry	Textural asymmetry in 32 × 32 VIRS scene, computed using channel-1
15	HOM	Homogeneity	Textural homogeneity in 32 × 32 VIRS scene, computed using channel-1

lines) surrounding the mid point. This mean rain rate is then considered as the rain rate over the VIRS scene. Different parameters were computed from the five channels of VIRS for the cloud classification. A detailed description of these parameters is presented in the next section.

3. Selection of physical and textural parameters from VIRS

For the VIS/IR based cloud classification, various useful input parameters have been suggested by earlier researchers like Miller and Emery (1997) (referred to as ME97 here onwards), Garand (1988), Gu *et al*. (1989), Ebert (1987), etc. In the present study, TRMM observations give us an advantage, that an optimum selection of limited number of parameters can be made by testing the correlation of individual parameters with the collocated rainfall observations. Initially we started with the parameters suggested by ME97, but found that in case of some parameters, certain functions of the parameters were better correlated with the coincident rainfall than the parameters themselves. This led us to formulate a new set of 15 parameters, which were used as input for cloud classification. These parameters include total cloud fraction (TCF), cloud albedo in first three channels (α_1, α_2 and α_3 respectively), normalized difference vegetation index (NDVI), mean cloud top temperature (TMEAN), minimum cloud top temperature (TMIN), difference of squares of channel-4 and channel-5 brightness temperatures ($\delta^2_{4\text{-}5}$), cloud connectivity (CC), and background connectivity (BC), etc. Parameters based on the textural analysis of cloud features include contrast (CON), asymmetry (ASYM), and homogeneity (HOM).

TABLE 2

Result of classification

Class No.	Number of 32 × 32 scenes	Number of raining scenes	Probability of rain	% of total rain received	Mean TMI rain rate
1	753	454	60.29	9.08	0.76
2	836	740	88.52	35.63	2.67
3	635	253	39.84	2.95	0.29
4	452	311	68.81	5.52	0.77
5	1060	90	8.49	0.25	0.01
6	1144	190	16.61	0.47	0.03
7	854	338	39.58	3.26	0.24
8	647	97	14.99	0.39	0.04
9	690	76	11.01	0.26	0.02
10	439	339	77.22	11.00	1.57
11	626	61	9.74	0.12	0.01
12	498	191	38.35	1.01	0.13
13	428	311	72.66	8.34	1.22
14	691	51	7.38	0.06	0.01
15	1233	136	11.03	0.30	0.02
16	574	261	45.47	1.75	0.19
17	521	409	78.5	10.30	1.24
18	522	87	16.67	0.54	0.07
19	631	39	6.18	0.10	0.01
20	370	178	48.11	1.71	0.29
21	718	396	55.15	3.79	0.33
22	794	326	41.06	2.99	0.24
23	691	58	8.39	0.18	0.02
24	993	17	1.71	0.01	0.00

Each parameter is generated for a 32 × 32 pixel region of VIRS. This corresponds to a region approximately 70 km on a side. Miller and Emery (1997) used 32 × 32 region of AVHRR pixels, which corresponded to regions about 32 km on a side, while Garand (1988) used regions 128 km on a side, but at lower pixel resolution. ME97 have argued that a smaller region size, *e.g.* 32 km, is preferable for resolving the features associated with the convective activity. However, considering the coarser resolution of VIRS (2.11 km at Nadir) the selection of a smaller region would have meant some degree of compromise on image statistics, which needs a large number of pixels.

Since the upwelling radiance in channel-3 (3.7 μm) receives contributions from reflected solar radiance as well as from thermal emission, the albedo of this channel is estimated by subtracting the channel 3 emissive radiance for an object at the temperature given by channel-4 from the measured channel radiance. We denote this albedo as α_3. Using the combinations of α_1, α_3 and NDVI, we introduced following features :

$$\chi = 1 + (\alpha_1 + \text{NDVI} - \alpha_3)$$

and also three fractional quantities ϕ_1, ϕ_2 and ϕ_3. The first fraction, ϕ_1 denotes the fraction of the VIRS scene where α_1 exceeds 0.5, therefore this fraction denotes the population of highly reflective and thick clouds. Similarly ϕ_2 is the fraction of VIRS scene where NDVI > -0.25, and ϕ_3 is the fraction where $\chi \geq 0.3$. The parameters χ, ϕ_1, ϕ_2 and ϕ_3 were formulated intuitively after several experiments, based on their correlations with the collocated TMI observed rain-rates. The 15 parameters were then used as input features in K-Mean algorithm (Hartigan and Wong, 1979) for classification.

TABLE 3

Feature mean for selected raining classes

NC	CON	HOM	ASYM	CC	BC	T MEAN (K)	δ^2_{4-5} (K^2)	α_1	NDVI	χ	TCF	TMI N (K)	ϕ_1	ϕ_2	ϕ_3
1	162.32	1.70	7.7	1.0	0.93	234.3	1173	0.74	0.21	0.51	0.95	212.5	0.16	0.07	0.31
2	45.23	3.15	14.4	1.0	0.97	221.7	1022	0.87	0.17	0.72	1.00	208.5	0.95	0.90	0.99
4	562.03	1.26	4.0	0.9	0.58	245.4	965	0.73	0.22	0.58	0.61	222.0	0.01	0.01	0.06
10	116.28	2.08	9.1	1.0	0.91	229.3	1033	0.79	0.18	0.63	0.99	210.5	0.46	0.44	0.77
13	239.81	1.35	5.8	1.0	0.39	247.5	932	0.76	0.21	0.65	0.96	222.4	0.20	0.10	0.59
17	86.82	2.24	10.1	1.0	0.98	240.0	1003	0.80	0.23	0.64	1.00	223.8	0.83	0.08	0.91

TABLE 4

Centroid means for the selected classes obtained from the classification algorithm

NC	CON	HOM	ASYM	CC	BC	TMEAN	δ^2_{4-5}	α_1	NDVI	χ	TCF	TMIN	ϕ_1	ϕ_2	ϕ_3
1	0.11	0.08	0.06	1.00	0.93	0.44	0.26	0.70	0.78	0.52	0.94	0.29	0.16	0.07	0.31
2	0.03	0.17	0.13	1.00	0.97	0.29	0.22	0.85	0.90	0.69	1.00	0.25	0.95	0.90	0.99
4	0.40	0.05	0.02	0.91	0.56	0.69	0.21	0.69	0.76	0.58	0.52	0.39	0.01	0.01	0.06
10	0.08	0.10	0.07	1.00	0.91	0.38	0.22	0.76	0.86	0.62	0.99	0.27	0.46	0.44	0.77
13	0.17	0.06	0.04	1.00	0.36	0.60	0.20	0.72	0.79	0.64	0.95	0.39	0.20	0.10	0.59
17	0.06	0.11	0.09	1.00	0.97	0.51	0.22	0.77	0.73	0.62	1.00	0.41	0.83	0.08	0.91

A brief description of different features and notations used for them in forthcoming text is given in Table 1.

4. Results

Table 2 shows the types of classes obtained by applying the clustering algorithm on the training data set. This table contains number of scenes, number of raining scenes, raining probability, percentage of rain received and rain rate in each class. It can be seen from Table 2 that different classes are associated with mean rain rate varying from very low (0.1 mm/h) to moderate (~ 2.67 mm/h) ones. It can be seen that there are six classes (1, 2, 4, 10, 13, 17) each of which has probability of rain greater than 60% and cumulatively these accounts for 80% of the total rainfall received. These classes have rain rates between 0.5-2 mm/h. There are few other classes (3,7,12,16,20,21,22) with probability of precipitation between 30-60% and rain rates between 0.1 – 0.5 mm/h. The remaining classes in Table 2 are those with low rain rates (< 0.1 mm/h) and probabilities of precipitation (< 30 %). It can be seen here that class 2 has the highest rain rate (2.67 mm/h). Table 3 lists the feature means for the above-mentioned six classes with prominent rainfall. Some distinct features associated with class-2 are; very low cloud top temperature (~ 208.5 K), total cloud cover, with high connectivity. The features α_1 and χ have high values for this class while NDVI for this class is smallest among all classes. This class can also be distinguished by very high values of features ϕ_1, ϕ_2 and ϕ_3 as compared with other classes. The probability of precipitation in this particular class is 88.5%, which is fairly high, and the class produced 35% of the total rainfall. It can be seen that the feature means in a particular class have a distinct value that distinguishes them from other classes. Thus the K-Mean Clustering Algorithm is able to separate out the sample data scenes into various meaningful classes.

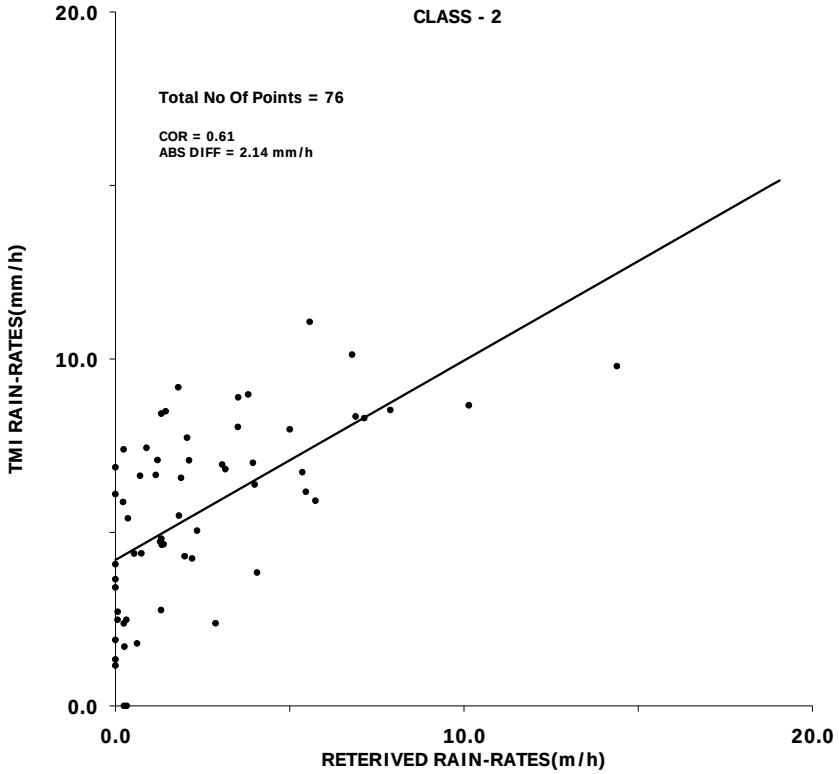

Fig. 2. Comparison of VIRS retrieved rain rates with collocated TMI rain rates for different classes

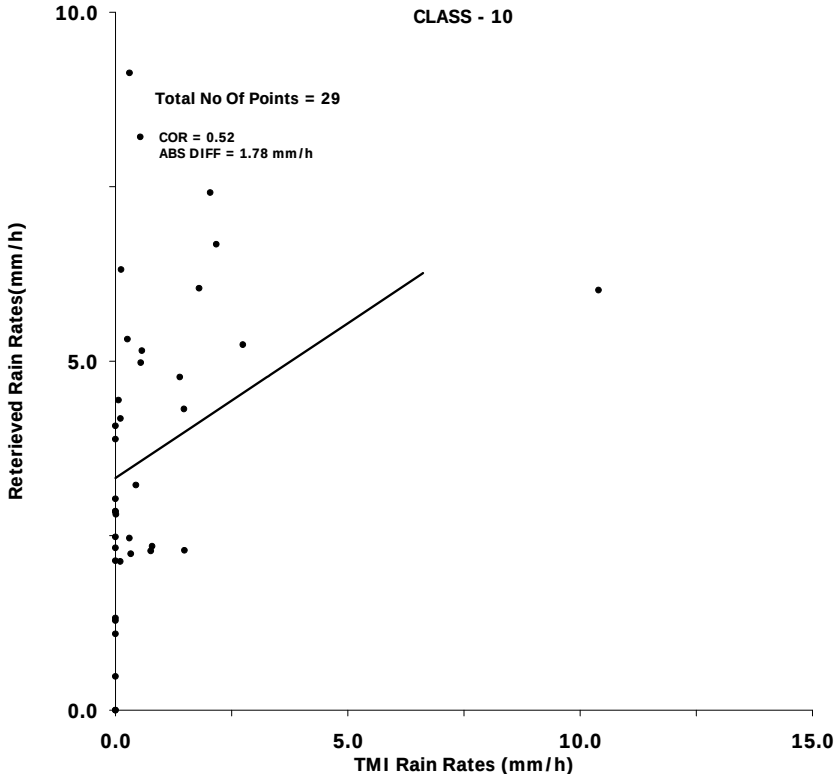

Fig. 3. Comparison of VIRS retrieved rain rates with collocated TMI rain rates for different classes

Our next step was to assess how the cloud features based on visible/IR observations are related to rainfall rates. Multiple linear regression analysis was performed separately for each class to explain average TMI rain rates in terms of cloud features observed by VIRS observations. Our analysis shows that for each class, a different set of features was necessary to explain maximum variance of observed rainfall rates. Table 4 provides a summary of regression analysis for six prominent classes. It can be seen that the explained variance for the six classes lies in the range 49-63% and the standard error is between 0.79-1.93 mm/h. The highest variance is explained by class – 4 (62.8%). The highest standard error is for class-2 (1.93). This may be because class-2 is the maximum raining class with very high rain rates at some scenes. So the error of estimation is high there.

Verification with the new data set

For testing the accuracy of our method we used three (503 scenes of 32 × 32 pixels) orbit data of VIRS (1B01 TRMM product) belonging to period July 2000 along with the collocated instantaneous TMI rain rates (2A12 product). The data was separated into 24 classes using centroid means of the training data set. Table 5 lists the results of classification for the verification data set. It can be seen that class 2; the most raining class, with mean rain rate ~ 2.77 mm/h and 76 out of 503 total scenes belonging to this class has 85.53% raining probability and it contributed 70.4% of the total rain. Besides this, class 10 also had significant contribution (12.5%) of total rain with a high mean rain rate (1.30 mm/h). Other raining classes were 1, 4, 13 and 17 with less number of raining points. It can also be seen that the non raining classes (class –24) has been very well separated out with only one raining point out of 64 total points.

Using the regression coefficients generated by training data set we estimated rain rates for the above five classes (1, 2, 4, 10 and 13) separately and the scatter plots of TMI instantaneous rain rates *vs*. the rain rates estimated by our method are shown in the Figs. (2-3). Fig. 3 shows the scatter plot for class-1 that contained 28 points. The correlation value for this class is 52% and the mean absolute difference is 1.78 mm/h. For class–2 (Fig. 2) the results are quite encouraging. Except for few heavy raining points the estimated rain rates are very well in agreement with TMI observations. The correlation for this class is 61% and the mean absolute difference is 2.14 mm/h.

TABLE 5

Various classes and their characteristics using the verification data set

B	N	NR	POP	RF	RR (mm/h)
1	28	8	28.57	2.15	0.23
2	76	65	85.53	70.43	2.77
3	6	1	16.67	0.06	0.03
4	5	5	100.00	1.22	0.73
5	24	0	0.00	0.00	0.00
6	29	4	13.79	0.13	0.01
7	14	5	35.71	0.50	0.11
8	19	1	5.26	0.04	0.01
9	16	0	0.00	0.00	0.00
10	29	17	58.62	12.58	1.30
11	20	0	0.00	0.00	0.00
12	14	6	42.86	1.53	0.33
13	22	13	59.09	3.89	0.53
14	24	1	4.17	0.00	0.00
15	28	3	10.71	0.12	0.01
16	7	0	0.00	0.00	0.00
17	31	22	70.97	5.91	0.57
18	11	4	36.36	0.18	0.05
19	9	2	22.22	0.09	0.03
20	1	0	0.00	0.00	0.00
21	13	5	38.46	1.02	0.23
22	8	1	12.50	0.01	0.00
23	5	1	20.00	0.09	0.05
24	64	1	1.56	0.06	0.00

5. Conclusions

Using multifeature classification approach a rainfall estimation method is presented in this paper where only

visible and infrared satellite data is used. The training data used belonged to the month of July, which is maximum rain receiving season of the year in tropics. This method is able to classify the clouds into various raining and non - raining classes. We have also seen that the feature values are distinct for different raining and non-raining classes and the rain estimates from the obtained classes are matching well with the microwave observations which are supposed to be more direct methods of rain estimation. For the maximum raining class the method is performing very well though it did not perform satisfactorily for some classes. The reason may be less number of points in that particular class. This method can be used to estimate rain rates from geostationary satellite data, which provide information on better spatial and temporal resolution than microwave, and instantaneous rain rates over larger domain can be estimated. However, some improvements are still left do be done. One of them is that the criteria for delineating cloudy pixels from oceanic and land pixels, has to be given more thought so as to perfectly separate out cloudy regions. Another aspect, which needs more research, is the size of scene to be used for better classification.

Acknowledgements

Authors thankfully acknowledge the services by PO-DAAC and NASA/GSFC for providing valuable TRMM data for our research.

References

Arkin, P. A., 1979, "Relationship between fractional coverage of high cloud rainfall accumulation during GATE over the B-scale array", *Mon.Wea.Rev*., **23**, 541-554.

Ebert, E., 1987, "A pattern recognition technique for distinguishing surface and cloud types in the Polar Regions", *J. Climate Appl. Meteor*., **26**, 1412-1427.

Garand, L., 1988, "Automated recognition of oceanic cloud patterns. Part-I: Methodology and application to cloud climatology", *J. Climate*, **1**, 20-39.

Gu., Z. Q., Duncan, C. N., Renshaw, E., Mugglestone, M. A., Cowan, C. F. N. and Grant, P. M., 1989, "Comparison of techniques for measuring cloud textures in remotely sensed satellite meteorological image data", IEE Proc. F: Radar Signal Processing, **136**, 236-248.

Hartigan, J. A. and Wong, M. A., 1979, "Algorithm AS 136: A K-means clustering algorithm", *Applied Statistics*, **28**, 100-108.

Krishnamurti, T. N., Xue, J., Bedi, H. S., Ingles, K., Oosterohof, D., 1990, "Physical initialization for NWP over the tropics", *Tellus*, **43**AB, 53-81.

Miller, S. W. and Emery, W. J., 1997, "An automated neural network classifier for use over land and ocean surfaces", *J. Appl. Meteor*., **36**, 1346-1362.

Pal, P. K., Prakash, W. J., Thapliyal, P. K. and Kishtawal, C. M., 1999, "A technique of rainfall assimilation for dynamic extended range monsoon prediction", *Meteorl., Atmos. Phys*., **71**, 157-168.