

Paper:

A Data Mining Approach to Rainfall Intensity Classification Using TRMM/TMI Data

Shan-Tai Chen*, Shung-Lin Dou*, and Wann-Jin Chen**

*Dept. of Computer Science, Chung Cheng Institute of Technology, National Defense University
No.190, Sanyuan 1st St., Tashi, Taoyuan, Taiwan, R.O.C.
E-mail: stchen@ccit.edu.tw

**Dept. of Environment Information Engineering, Chung Cheng Institute of Technology, National Defense University
[Received June 7, 2008; accepted August 20, 2008]

The systematic approach we propose for classifying oceanic rainfall intensity during the typhoon season consists of two major steps – 1) identifying the rain areas and 2) classifying rainfall intensity into *normal* and *heavy* for these areas. The heterogeneous hierarchical classifier (HHC), an ensemble model we developed for accurately identifying heavy rainfall events, consists of a set of base classifiers. The base classifiers are independently constructed through heterogeneous data mining approaches such as artificial neural networks, decision trees, and self-organizing maps. The meteorological satellite Tropical Rainfall Measuring Mission (TRMM) microwave imager (TMI) data from 2000 to 2005 are used to create the classification models. TRMM precipitation radar (PR) data and rain gauge data from Automatic Rainfall and Meteorological Telemetry System (ARMTS) measurement are used as ground truth data to evaluate models. Two thirds of the dataset is used for model training and one third for testing. Experimental results show that the proposed model classifies rainfall intensity highly accurately and outperforms previously published methods.

Keywords: data mining, classification, rainfall intensity, TRMM, microwave

1. Introduction

Heavy typhoon season rainfall causing landslides, floods, and other weather-related disasters is a critical problem in Taiwan, among other areas, making it important to predict areas of heavy rainfall to minimize potential weather-related disasters.

Among the extensive heavy-rainfall research done, Alexiuk [1] used meteorological volumetric radar data to detect thunderstorms and hybrid strategies to classify storm events into wind, heavy rain, tornados, and hail. Using cluster analysis, Martinez [2] classified heavy-rain events into eight atmospheric patterns and further determined the relationship between heavy-rain regions and the atmospheric patterns. Using self-organizing maps

(SOMs), Nishiyama [3] identified a typical synoptic pattern causing heavy rainfall during the rainy season.

Satellite observations provide useful information, e.g., visible image data, infrared image data, and microwave channel data, for estimating surface rainfall. They also provide more frequent observations than conventional rain gauge measurement and cover a wider area than radar data. Lee [4] proposed pattern recognition using visible and infrared satellite images to classify rain rates into *none*, *light*, and *heavy*, achieving an accuracy exceeding 70%. Using visible and infrared samples from VHRR images, Parvathi [5] took the maximum likelihood decision rule and classified rain rate events into four categories for an accuracy exceeding 80%. Limitations associated with such data, however, are that 1) visible satellite images are available only during daytime and 2) infrared observations cannot provide information on vertical cloud structure.

Passive microwave radiometers physically sense raindrops and hydrometeors within precipitating clouds, so microwave channel data is widely used in heavy oceanic rainfall estimation. The TRMM/TMI [6] satellite provides passive microwave and infrared data for monitoring and studying tropical rainfall, and has been used in different precipitation-related applications [7, 8]. Several types of rain-area identification have been developed using TRMM/TMI data, including a scattering index [9], threshold check [10], and rain flag [11]. The rain flag is also used to identify heavy-rain events, classifying weather systems into four categories – no rain, uncertain, rain, and heavy rain.

In this paper, a new algorithm is developed for determining rainfall intensity during the typhoon season using TRMM/TMI data. As in [4], we classifies events into one of three rain-rate categories – none, normal, and heavy. We constructed two ensemble models – the neural committee classifier (NCC) [12] and the heterogeneous hierarchical classifier (HHC) – for rain-area identification and rainfall intensity retrieval. Experimental results show that our algorithm is highly accurate. Our model has been applied to determining rainfall intensity for an actual typhoon. Results show relatively high agreement between the Goddard profiling algorithm (GPROF) [13] and ours.

This paper is organized as follows: Section 2 presents



Table 1. Nine TRMM/TMI microwave channels [15].

V: Vertical; H: Horizontal.

Channel	1	2	3	4	5	6	7	8	9
Frequency (GHz)	10.65	10.65	19.4	19.4	21.3	37	37	85.5	85.5
Polarization	V	H	V	H	V	V	H	V	H

a brief background and Section 3 details the HHC concept and the algorithm for constructing it. Section 4 provides the framework for our approach to rain-rate classification and Section 5 reviews experimental results. Section 6 gives conclusions.

2. Background

We begin by reviewing the background to this work.

2.1. Data Source

Datasets used for training and testing in this research include TRMM/TMI, TRMM/PR, and rain gauge data, which are introduced as follows:

TRMM Data

The international Tropical Rainfall Measuring Mission (TRMM) [6], jointly sponsored by the US National Aeronautics and Space Administration (NASA) and the Japan Aerospace Exploration Agency (JAXA), provides passive microwaves with nine channels (**Table 1**), and infrared data for monitoring and studying tropical rainfall. The TRMM satellite, with a period of 91.5 minutes and launched in November 1997, provides satellite observations from 40°N to 40°S. TMI level-1B11 brightness temperature (Tb) data and TRMM/PR rainfall estimation data we have used here are downloaded from the NASA website [14].

Island rain gauge data

Rain gauge data used in this study is from the Japan Meteorological Agency (JMA) and the data range period is from July to October 1998-2005. **Table 2** shows the 11 islands providing rain gauge data (mm/10-min) we used here. Rain gauge data, with 60-minute accumulation rainfall, is used as ground truth for evaluating results. The data is accumulated from 30 minutes before and after the time when the satellite passed over a rain gauge [15–17].

We define a TMI pixel as a rain area if accumulated rainfall exceeds 1 mm/hr and as a heavy-rain area if it exceeds 15 mm/hr.

2.2. Rain Flag Algorithm

The rain flag (RF) [11], developed by Goodberlet et al., retrieves oceanic wind speed through microwave observation. Depending on variations in the Tbs of 19H, 37V, and 37H GHz, weather systems are classified into four types, i.e., RF = 0, 1, 2, and 3, which denote the *no-rain*,

Table 2. Island rain gauge station sites [15].

NO	Station	Code of Station	North Lat., East Lon.	Altitude(m)
1.	IRABU	93011	24.82°N, 125.17°E	10
2.	MIYAKOJIMA	93041	24.79°N, 125.27°E	40
3.	GUSUKUBE	93051	24.74°N, 125.41°E	55
4.	TARAMA	93061	24.66°N, 124.69°E	16
5.	IBARUMA	94001	24.50°N, 124.28°E	15
6.	KABIRA	94036	24.46°N, 124.14°E	7
7.	YONAGUNIJIMA	94017	24.46°N, 123.01°E	30
8.	IRIOMOTEJIMA	94061	24.38°N, 123.74°E	9
9.	ISHIGAKIJIMA	94081	24.33°N, 124.16°E	6
10.	OOHARA	94101	24.26°N, 123.87°E	28
11.	HATERUMA	94116	24.05°N, 123.76°E	32

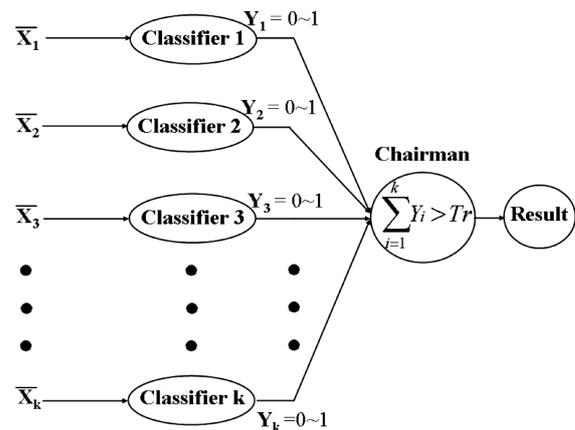


Fig. 1. A logical view of NCC [12]. It conducts classification by aggregating predictions made by individual weak classifiers.

uncertain, rain, and heavy rain types, respectively. Classification rules for the four are as follows:

$$0: Tb_{37V} - Tb_{37H} > 50K \text{ and } Tb_{19.4H} < 165K \quad (1)$$

$$1: 37K \leq Tb_{37V} - Tb_{37H} < 50K \text{ or } Tb_{19.4H} \geq 165K \quad (2)$$

$$2: 30K \leq Tb_{37V} - Tb_{37H} < 37K \quad (3)$$

$$3: Tb_{37V} - Tb_{37H} < 30K. \quad (4)$$

We classify a TMI pixel as no rain for RF=0, normal for RF=2, or heavy for RF=3. Pixels classified as RF=1 cannot be classified by RF, so we omit them from our experiments.

2.3. Neural Committee Classifier

The neural committee classifier (NCC) [12], an ensemble classifier for rain-area identification, consists of several weak classifiers (WC) of neural networks (**Fig. 1**). X_i and Y_i are the input vector and output value for each WC i . Each WC is a multilayer neural network trained independently by the back propagation algorithm using different TRMM/TMI microwave channel data. The chairman function, which judges whether a TMI pixel is a rain area

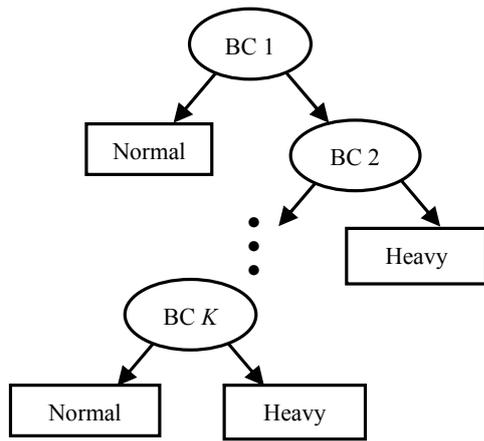


Fig. 2. General HCC architecture with individual internal nodes as base classifiers.

```

HHC_Construction (T, B)
Input: set of training events T; set of BCs B
Output: HHC
1: If stop_condition(T, B) == False
2:   root = CreateNode();
3:   root.bc = Find_best_BC(T, B);
4:   (Tnormal, Theavy) = Split(T, root.bc);
5:   Child_1 = HHC_Construction (Tnormal, B);
6:   Child_2 = HHC_Construction (Theavy, B);
7:   Add Child_1 and Child_2 as two descendents of root.
8: Else
9:   leaf = CreateNode();
10:  leaf.label = Classification(T);
11:  Return leaf;
12: End If
13: Return root;
    
```

Fig. 3. Skeleton algorithm for HHC construction.

by accumulating output from all weak classifiers, is as follows:

$$\sum_{i=1}^k Y_i > Tr \quad \dots \quad (5)$$

where Y_i is the output of WC i and Tr is a threshold.

3. Heterogeneous Hierarchical Classifier

The heterogeneous hierarchical classifier (HHC) is developed for classifying rain intensity, i.e., heavy or normal. In the general HHC architecture (Fig. 2), individual internal nodes are *base* classifiers (BCs).

BCs are selected from a pool of classifiers that are independently constructed by heterogeneous data mining approaches such as artificial neural networks, decision tree learning, and SOMs.

Input to the skeleton algorithm for building the HHC model (Fig. 3) consists of a set of training events T and a set of BCs B . The algorithm recursively selects the best BC from B to split T and expand HHC leaf nodes until a

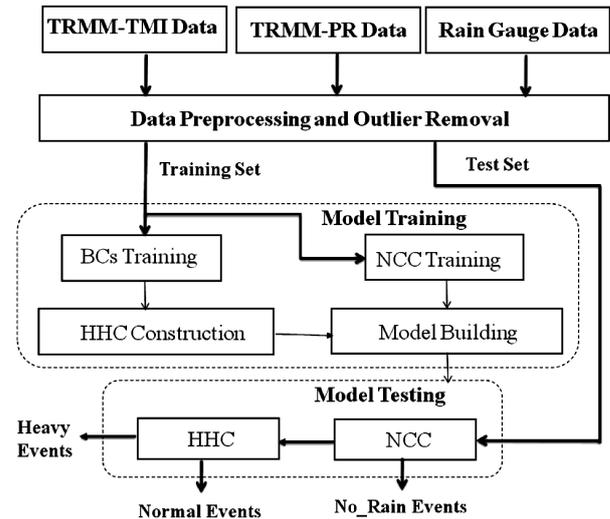


Fig. 4. Data mining framework for rainfall intensity retrieval.

stop condition is met.

The algorithm first chooses the BC having the highest classification accuracy over training events T as the HHC root (line 3). It then uses the BC to classify rain events into either normal or heavy (line 4). Line 1 determines whether the stop condition is met, e.g., classification accuracy for a class achieves a given threshold. If so (lines 8-11), it labels the class as normal (or heavy); otherwise, it chooses other BCs to recursively continue classification (lines 1-7) until all data events are labeled.

4. Data Mining Framework

Our data mining framework for rain-rate classification consists of four phases – data collection, data preprocessing, model training, and model testing (Fig. 4), introduced below.

4.1. Data Collection

The research domain ranges from latitude 15°N to 30°N and longitude 110°E to 130°E. We used TRMM/TMI, TRMM/PR, and rain gauge data from the 1998 to 2005 typhoon seasons.

4.2. Data Preprocessing

The goal of data processing is to make data more suitable for data mining and to build more accurate classification models. This is done by data cleaning, integration, and outlier removal.

Beam filling error [18] is caused by variations in rainfall inhomogeneity within the relatively coarse resolution of today’s sensors. To avoid the errors, we apply TRMM/VIRS data in data collection, i.e., only rain events with full homogeneous clouds observed in TRMM/VIRS are collected. Similarly, in no-rain event collection, only events with no clouds observed in TRMM/VIRS are collected.

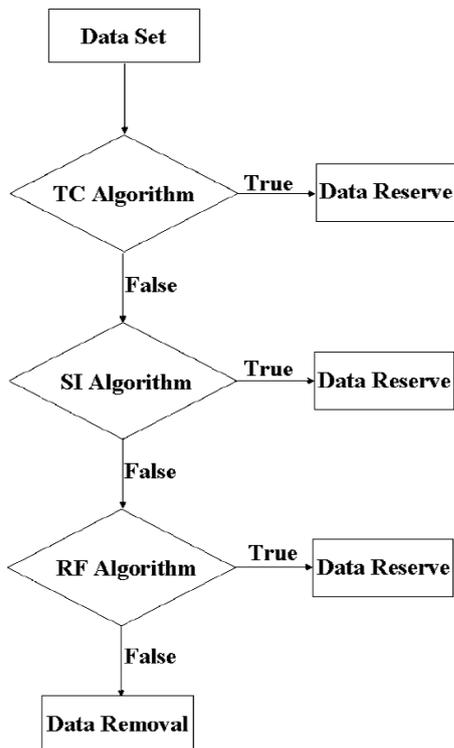


Fig. 5. Outlier-removal algorithm flowchart [12].

In outlier removal (Fig. 5), the previous rain-area identification methods – scattering index [9], threshold check [10], and rain flag [11] – are used to filter events in the dataset. An event is removed if its ground truth is inconsistent with all results generated by these three methods.

4.3. Model Training

The framework (Fig. 4) contains two major phases conducted on classifiers NCC and HHC, which are trained independently, then combined to build a rain intensity classification model. The details for training and constructing the classifiers refer to Sections 2 and 3. Note that training data for the NCC includes all of the training set, i.e., no-rain, normal, and heavy events, but only normal and heavy events for training the HHC.

4.4. Model Testing

Once a model has been constructed, a test event is classified starting from the NCC that classifies the event into no-rain or rain. The event is labeled if it is classified as no-rain by the NCC; otherwise, the event passes to the HHC. The root node of the HHC, e.g., BC1, classifies the rain event as either normal or heavy, then recursively follows the appropriate branch based on the classification outcome. The class label associated with each leaf node is then assigned to events.

To further verify the model, we apply it to an actual typhoon and compare results to GPROF [13] as detailed in Section 5.

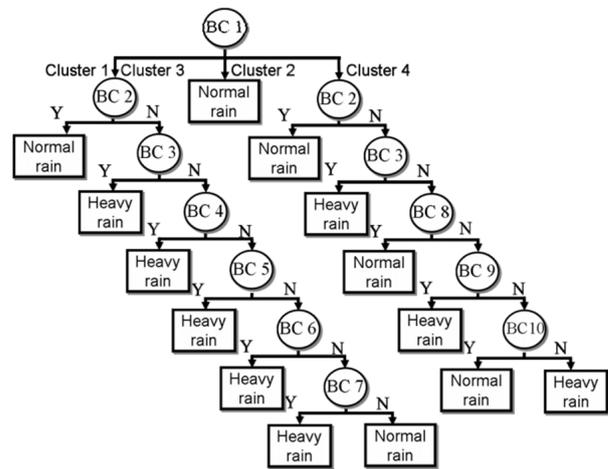


Fig. 6. HHC architecture for rainfall intensity classification.

Table 3. HCC base classifiers.

Base classifier (BC)	Classification rule (Tb)	Normal (< 15mm/hr)	Heavy (≥ 15mm/hr)	Total
1	SOM cluster 1	125	75	200
	2	131	1	132
	3	21	80	101
	4	123	44	167
2	19.4H < 211.8	116	0	116
3	19.4H ≥ 265.7	5	50	55
4	19V-37V ≥ 13.8	29	9	38
5	19V-37V ≥ 4.7 and 37V-37H < 4.4	25	10	35
6	13.8 > 9V-37V ≥ 9.2 and 37V-37H < 4.4	3	36	39
7	13.8 > 19V-37V ≥ 4.6 and 37V-37H ≥ 4.4	4	17	21
8	4.7 > 19V-37V or 19V-37V ≥ 13.8 and 37V-37H ≥ 4.4	180	55	235
9	4.7 > 19V-37V ≥ 3.6 and 37V-37H < 4.4 and 10.65H < 160.9	19	4	23
10	5.7 > 19V-37V ≥ 4.7 and 37V-37H < 4.4	18	3	21

5. Experimental Results

We used meteorological satellite TRMM-TMI data from 2000 to 2005 to create classification models and TRMM-PR data and rain gauge data as the ground truth to evaluate models. Two thirds of the dataset is used for model training and one third for testing.

The HHC architecture for rainfall intensity classification (Fig. 6) is built using the algorithm in Section 3. In experiments, we chose classifiers with accuracy exceeding 75% as BCs (Table 3).

Results show that our approach is able to achieve high

Table 4. Confusion matrix for rainfall intensity classification obtained by our model.

Actual Class	Predicted Class			Accuracy (%)
	No Rain	Normal	Heavy	
No Rain	92	2	1	96.8
Normal	0	159	41	79.5
Heavy	0	25	75	75
Overall accuracy: 82.5				

Table 5. The Confusion matrix for rainfall intensity classification using the rain flag method.

Actual class	Predicted Class			Accuracy (%)
	No Rain	Normal	Heavy	
No rain	56	21	1	71.8
Normal	0	18	182	9
Heavy	0	0	100	100
Overall accuracy: 46				

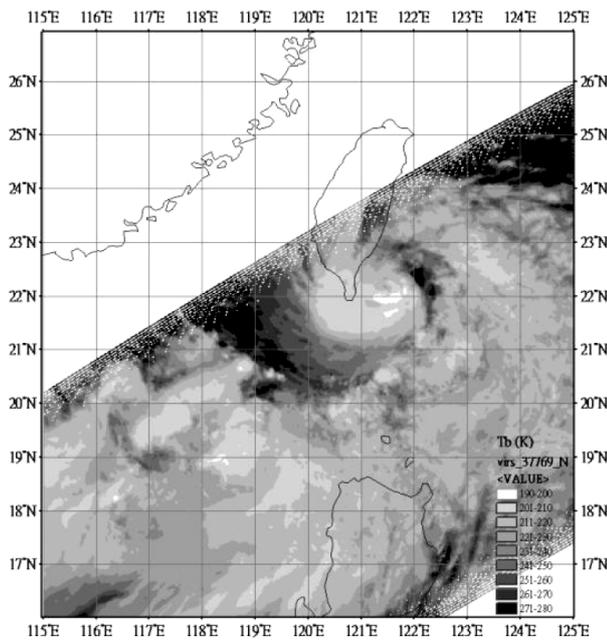


Fig. 7. Mindlle typhoon viewed by the TRMM/VIRS.

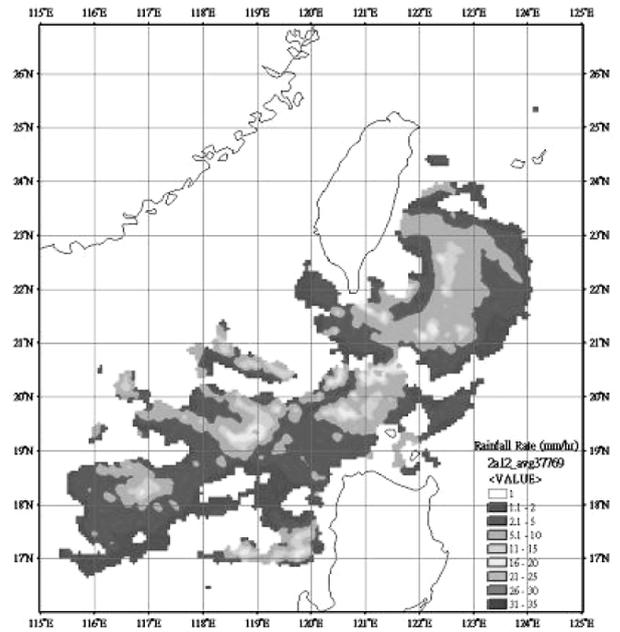


Fig. 8. Rainfall estimated by the GPROF algorithm [17].

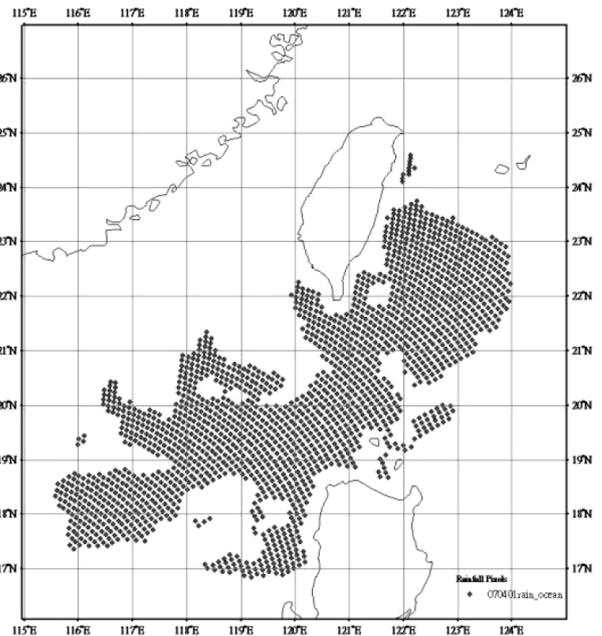


Fig. 9. Rain area recognized by the NCC [12].

accuracy for rainfall intensity classification. **Tables 4, 5** list confusion matrices for the proposed model and rain flag approach. Although the RF can identify all heavy events, it overestimates rainfall intensity for datasets in this research domain range. Our model’s overall accuracy is 82.5%, which is superior to the RF’s 46% [18], Lee’s 70% [4], and Parvathi’s 80% [5].

We have applied our approach to classifying rainfall intensity for an actual typhoon case. **Fig. 7** shows the image of Mindlle typhoon from TRMM/VIRS (visible and infrared scanner). **Fig. 8** is the result derived by the Goddard profiling algorithm (GPROF), **Fig. 9** by the NCC, and **Fig. 10** by the HCC. It demonstrates relatively high agreement between GPROF and our algorithm.

6. Conclusions

Using a data mining approach, we have presented an efficient, near-real-time framework to retrieve rainfall intensity using microwave data from the TRMM satellite. In this framework, we used association rule analysis to determine important dataset attributes. We have also developed two ensemble models – the neural committee classifier (NCC) and the heterogeneous hierarchical classifier (HHC) – to improve classification accuracy. Experimental results show that our model outperforms previous rainfall intensity retrieval methods. Due to the interpretability and

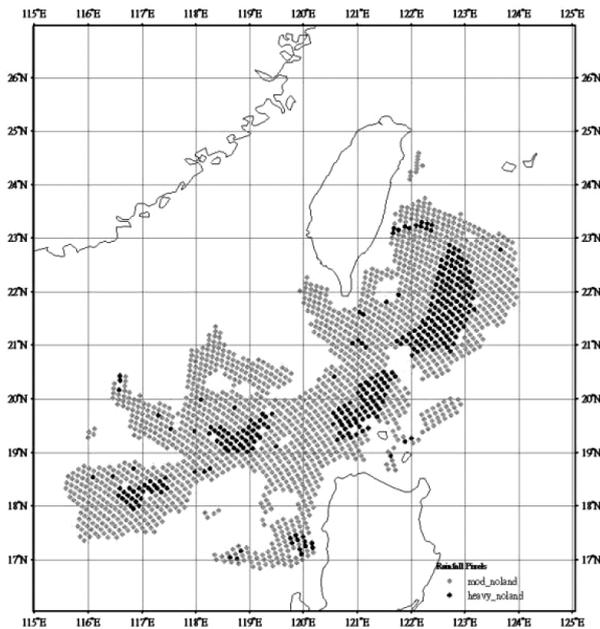


Fig. 10. Rainfall intensity classified by the HHC. Dark pixels denote areas of heavy rainfall.

accuracy of the results obtained and the simplicity of the calculation involved, our algorithms are helpful in retrieving real-time rainfall intensity of typhoons and provide a useful comparison to modeling results.

Acknowledgements

This research was supported in part by the National Science Council of ROC under grants NSC 97-2625-M-606-001. We thank NASA/GSFC DAAC for providing TMI and VIRS data and the Japan Meteorological Agency for providing 10-min AMEDAS rain gauge data.

References:

- [1] M. Alexiuk, N. Pizzi, and W. Pedrycz, "Classification of volumetric storm cell patterns," In Proc. of the 1999 IEEE Canadian Conf. on Electrical and Computer Engineering, pp. 1081-1085, 1999.
- [2] C. Martinez, J. Campins, A. Jansa, and A. Genoves, "Heavy rain events in the Western Mediterranean: an atmospheric pattern classification," *Advances in Sci. and Res.*, Vol.2, pp. 61-64, 2008.
- [3] K. Nishiyama, S. Endo, K. Jinno, C. B. Uvo, J. Olsson, and R. Berndtsson, "Identification of typical synoptic patterns causing heavy rainfall in the rainy season in Japan by a self-organizing map," *Atmospheric research*, Vol.83, pp. 185-120, 2007.
- [4] B. G. Lee, R. T. Chin, and D. W. Martin, "Automated Rain-Rate Classification of Satellite Images Using Statistical Pattern Recognition," *IEEE Transactions On Geoscience And Remote Sensing*, Vol. GE-23, No.3, pp. 315-324, May, 1985.
- [5] R. Parvathi, B. Manikiam, V. Jayaraman, and M. G. Chandrasekhar, "Rain-rate classification of INSAT-VHRR images through statistical methods," *Advances in Space Research*, Vol.13, Issue 5, pp. 143-148, May, 1993.
- [6] C. Kummerow, W. Barnes, T. Kozu, J. Shiue, and J. Simpson, "The Tropical Rainfall Measuring Mission (TRMM) Sensor Package," *Journal of Atmospheric and Oceanic Technology*, Vol.15, No.3, pp. 809-817, 1998.
- [7] G. J. Huffman, R. F. Adler, D. T. Bolvin, G. 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, Issue 1, pp. 38-55, February, 2007.

- [8] V. L. Sanderson, C. Kidd, and G. R. McGregor, "A Comparison of TRMM Microwave Techniques for Detecting the Diurnal Rainfall Cycle," *J. Hydrometeor.*, 7, pp. 687-704, 2006.
- [9] N. C. Grody, "Classification of Snow Cover and Precipitation Using the Special Sensor Microwave Imager," *J. Geophys. Res.*, Vol.96, pp. 7423-7435, 1991.
- [10] W.-J. Chen and C.-C. Li, "Oceanic Rain Rate Retrievals Using TRMM Microwave Imager Multi-Channel Brightness Temperatures During the 1998 SCSMEX," *Terrestrial, Atmospheric and Oceanic Sciences*, Vol.11, pp. 765-788, 2000.
- [11] M. A. Goodberlet, C. T. Swift, and J. C. Wilkerson, "Remote Sensing of Ocean Surface Winds with the Special Sensor Microwave/Imager," *J. Geophys. Res.*, Vol.94, C10, pp. 14547-14555, 1989.
- [12] S.-T. Chen, C.-C. Wu, W.-J. Chen, and J.-C. Hu, "Rain-Area Identification Using TRMM/TMI Data by Data Mining Approach," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol.12, No.3, pp. 243-248, 2008.
- [13] C. Kummerow, Y. Hong, W. S. Olson, S. Yang, R. F. Adler, J. McCollum, R. Ferraro, G. Petty, D. B. Shin, and T. T. Wilheit, "The evaluation of the Goddard Profiling Algorithm (GPROF) for rainfall estimation from passive microwave sensors," *J. Appl. Meteor.*, Vol.40, pp. 1801-1820, 2001.
- [14] <http://disc.sci.gsfc.nasa.gov/data/datapool/TRMM/>
- [15] J.-C. Hu, W.-J. Chen, G.-R. Liu, M.-H. Chang, and H.-P. Gang, "Quantitative rain rate over ocean using microwave observations during the typhoon season," *Weather Forecasting and Analyzing*, Vol.185, No.380, pp. 21-30, 2005.
- [16] C.-C. Li, "Retrievals and application of rainfall rate over ocean using TMI microwave observation," The thesis of Ph.D. of National Defense University, 2002.
- [17] W.-J. Chen, J.-C. Hu, G.-R. Liu, and M.-H. Chang, "Quantitative precipitation over ocean using TMI microwave observations during the typhoon season," *Journal of Atmospheric Science*, Vol.34, No.1, pp. 67-88, 2006.
- [18] C. Kummerow, "Beamfilling Error in Passive Microwave Rainfall Retrievals," *J. Appl. Meteor.*, Vol.37, pp. 356-370, 1998.



Name:
Shan-Tai Chen

Affiliation:
Assistant Professor, Chung Cheng Institute of Technology, National Defense University

Address:
No.190, Sanyuan 1st St., Tashi, Taoyuan, Taiwan, R.O.C.

Brief Biographical History:
1997-2000 Adjunct Lecturer, Department of Computer Science, Chung-Cheng Institute of Technology
2004-2005 Assistant Professor, Computer Center, Chung Cheng Institute of Technology, Taiwan, R.O.C.
2005-present Assistant Professor, Department of Computer Science, Chung Cheng Institute of Technology, National Defense University, Taiwan, R.O.C.

Main Works:
• "Strategy Optimization for Deductive Games," European Journal of Operational Research, Vol.183, pp. 757-766, 2007.

Membership in Academic Societies:
• Taiwanese Association for Artificial Intelligence (TAAI)



Name:
Wann-Jin Chen

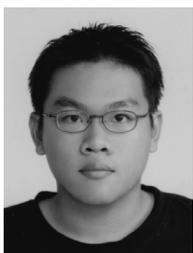
Affiliation:
Professor, Department of Environmental Information and Engineering, Chung Cheng Institute of Technology, National Defense University, Taiwan

Address:
Ta-Hsi, Tao-Yuan, Taiwan, 33510

Brief Biographical History:
2001.7-2001.12 Visiting scholar, Space Science and Engineering Center, University of Wisconsin-Madison, U.S.A.
2002.2-2005.7 Professor, Chair of Dept. of Applied Physics, Chung Cheng Institute of Technology
2005.8-2006.8 Dean, College of Science, Chung Cheng Institute of Technology
2006.11-2008.3 Dean, School of Defense Science, Chung Cheng Institute of Technology
2008.3- Professor, Dept. of Environmental Information and Engineering, Chung Cheng Insti. of Tech.

Main Works:
• "Rainfall Estimation of Mesoscale Convective System using AMSU-A data during the Mei-Yu Season," TAO, Vol.17, No.1, 2006.

Membership in Academic Societies:
• Meteorological Society of the Republic of China
• The Republic of China Society on Science and Technology of National Defense
• Chinese Geoscience Union, Republic of China



Name:
Hsiang-Lin Dou

Affiliation:
Master, Department of Computer Science, Chung Cheng Institute of Technology, National Defense University

Address:
No.190, Sanyuan 1st St., Tashi, Taoyuan, Taiwan, R.O.C.

Brief Biographical History:
2006-2008 Graduate Student, Dept. of Computer Science, Chung Cheng Institute of Technology, National Defense University R.O.C.
