CLOUD PATCH-BASED RAINFALL ESTIMATION USING A SATELLITE IMAGE CLASSIFICATION APPROACH

Yang Hong, Kuo-lin Hsu, and Soroosh Sorooshian

Center for Hydrometeorology and Remote Sensing,
Department of Civil and Environmental Engineering
University of California at Irvine
E-4130 Engineering Gateway, Irvine, CA 92697-2175

ABSTRACT

The relationship between cloud and surface rain rate varies significantly from cloud patch to cloud patch. Therefore, a rainfall estimation model characterized by significant transience, heterogeneity, and variability is needed to associate rainfall with the extremely complex and still imperfectly understood precipitating processes to produce higher quality estimates. We have responded to this by developing a high-resolution precipitation estimation algorithm dubbed "CCS" (Cloud Classification System) at UC Irvine. The CCS uses computer image processing and pattern recognition techniques to develop a patch-based cloud classification and rainfall estimation system based on co-registered passive microwave and infrared images from Low Earth-orbiting and Geostationary satellites. Unlike the region-based approach, which establishes only one T_b-R function for all clouds, this technique classifies various patches into different clusters and then searches the best-matched nonlinear T_b-R mapping function for each patch. Therefore, CCS jumps out the deadlock of the assumption that colder cloud pixel must produce higher rain rates than warmer cloud pixel, which is not all-time-true but popularly used by some other statistical regression or histogram matching approaches. This design feature enables CCS to generate various rain rates at a given brightness temperature and variable rain/no-rain IR thresholds for different cloud patches, which overcomes the one-to-one mapping limitation of a single statistical T_b-R function for the full spectrum of cloud-rainfall conditions. In addition, the computational and modeling strengths of neural network enable CCS to cope with the nonlinearity of cloud-rainfall relationships by fusing multi-source data sets. We are operating this system with the goal to produce data at spatio-temporal resolution suitable for basin scale hydrological research and applications, along with the goal to provide high-quality precipitation analysis for GEWEX CEOP sites.

1. INTRODUCTION

Because satellites measure rainfall as an integral of space at a point in time, the sampling frequency (near 4 x 4 km^2 and 30-minute sampling interval) of infrared imageries from geostationary satellites provides an attractive option for those applications where high sampling frequency is required. In the literature, several efforts have been made to calibrate a statistical relationship between cloud IR brightness temperatures and surface rain rate (R) observation either from the ground or space measurements. This statistical relationship, hereinafter referred to as T_b-R, is then applied at the temporal resolution of IR data from geostationary satellites. According to the types of fitting functions, the T_b-R relationships may be broken down further into (1) threshold-based, (2) linear fitting, and (3) nonlinear fitting approaches. Other than curve fitting approaches, histogram-matching (HM) technique is also used to evaluate the relationship between the two datasets where a regressive line would not be meaningful (Adler et al. 2001). Todd et al. (2001), Huffman et al. (2001), and Kidd et al. (2003) applied HM technique to generate rainfall products. These techniques share one feature in common: the relationship between IR T_b and (microwave or radar-derived) rain rates is established over a certain region with monthly or pentad adjustment. We might call them **region-based** T_b-R mapping fashion. We argue that these techniques fail to represent the variation of cloud-precipitation relationships because they only statistically determine one single T_b-R function and then apply to whole study region without discrimination of the innumerous clouds. Although enjoying the benefits of simplicity and low computational cost, they provide only a climatologic calibration and may not respond adequately to hydrometeorological variations, e.g., diurnal variation. As a result, considerable temporal and spatial integration, e.g., daily 1.0^o, is conducted to reduce the estimation error.

The cloud-precipitation mechanism is determined by different dynamical and thermodynamical processes, which are highly time-dependent and space-variant. As can be seen from Figure 1, the T_b-R relationships can vary significantly from one cloud patch to another at small scale. Previous studies (Griffith et al. 1978; Woodley et al. 1980; Adler and Negri 1988; Xu et al. 1999) suggested that classification of cloud patches into a number of groups would improve discrimination of different precipitation-generating systems.

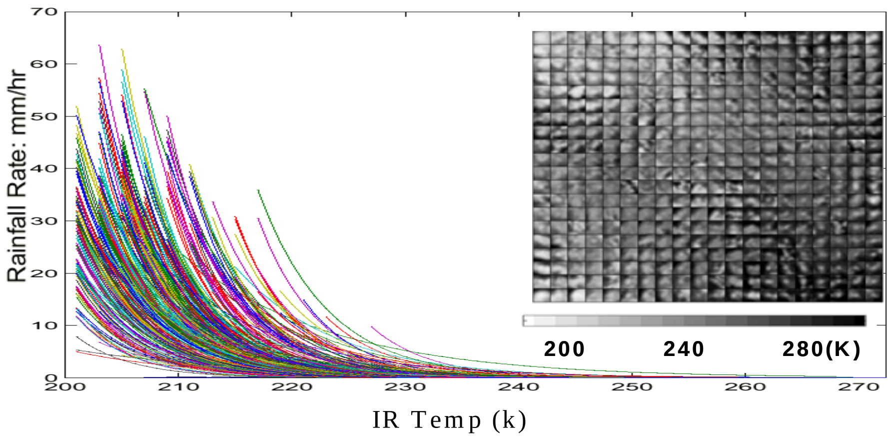


Figure 1. Each curve (left) indicates a unique T_b-R relationship corresponding to different cloud patch (right) at small spatial and temporal scales. Note each grid (right) is a 2^ox2^o region.

Unlike region-based T_b-R combined approaches, this paper outlines a **cloud patch-based** T_b-R combined technique to establish different cloud-precipitation relationships for respective cloud groups clustered by pattern recognition algorithm according to cloud patches' physical-statistical features. The concept behind this is that cloud patches with variable geophysical, geometric, and textural features would exhibit different precipitation generation processes. The unique feature of the cloud classification technique is that it searches the best T_b-R mapping relationships for different clouds based on a natural segmentation/classification of clouds (i.e., a cloud-patch based manner), which overcomes the limitation of some other techniques that only determine a single IR/rainfall function to an arbitrary calibration domain (in space and time). Therefore, this technique can generate a variety of rain rates and rainfall amounts for different cloud types such as convective Cumulonimbus, high-level Cirrus, and low-level stratiform clouds etc. In addition, the computational and modeling strengths of neural network may cope with the nonlinearity of cloud-rainfall relationships by fusing multi-source data sets (Tapiador et al. 2002).

2. METHODOLOGY

This algorithm establishes different T_b-R relationships, calibrated by co-located cloud images and microwave or radar rainfall data, for every classified cloud group by implementing a sequence of four steps (Figure 2): (1) separating cloud images into distinctive cloud patches, (2) extracting cloud features including coldness, geometry, and texture, (3) clustering cloud patches into well-organized subgroups, and (4) calibrating cloud-top temperature and rainfall (T_b-R) relationships for the classified cloud groups using microwave or gauge-corrected radar rainfall data. An automated network, SONO (Self-organizing Nonlinear Output; Hong et al. 2005a) model is developed to carry out the step 3-4 functions.

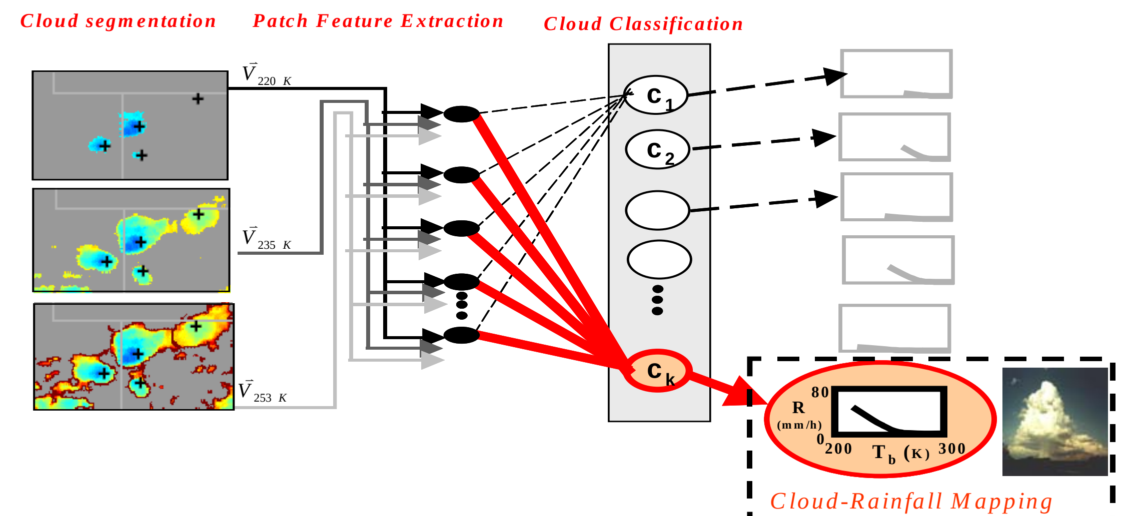
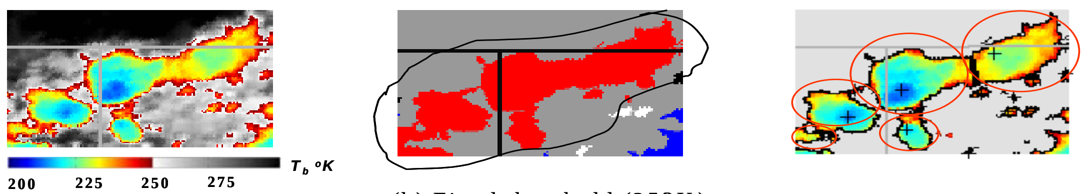


Figure 2. The cloud classification model structure

2.1 Segmentation of satellite infrared cloud images

Segmentation of satellite IR imagery is a preprocessing step for cloud analyses such as cloud-feature extraction, cloud-type classification, and wind detection. The proposed segmentation method, Incremental Temperature Threshold (ITT; Hong et al. 2003), falls under the category of hybrid segmentation approaches because it combines the hierarchical thresholding and Seeded Region Growing (SRG; Adams and Bischof 1994). As a hybrid method, ITT retains the advantage of SRG—fast execution and robust segmentation and, meanwhile, eliminates manual selection of seeds through hierarchical thresholding. Examples of cloud segmentation using the ITT method are shown in Figure 3.



(a) Infrared cloud image

(b) Fixed threshold (253K) segmentation of cloud image

(c) ITT segmentation

Figure 3. Cloud IR segmentation using fixed threshold and ITT method.

A cloud image is shown in Figure 3a and cloud-patch segmentation of this image using a fixed threshold (253K) is shown in Figure 3b. Note that even though the cloud image contains several convective cells, the single threshold cannot separate them effectively. Given the same IR image, ITT first locates the minimum temperature as seeds (illustrated by the cross marker), and then starts to iteratively expand each seed's area one neighborhood size at a time until touching neighboring clouds or temperature threshold that delineates clouds from the clear sky. Along with the expanding process, new cloud patch with lower altitude might be identified.

2.2. Extraction of cloud-patch features

An empirical-statistical analysis was conducted to investigate different sets of feature combinations in terms of three criteria: precipitation relevance, classification impact, and computation efficiency, in the order of decreasing importance. Additionally, the inter-relationships (i.e., correlation and covariance) among the features help to determine the importance of the features in discriminating alternative clusters. The characteristics of cloud patches, relevant to precipitation, are grouped into three categories: coldness, geometry, and texture. The first category is generally associated with the geophysical variables—cloud brightness temperature; the second one is derived from the geometric properties of cloud patches; and the third category is the texture variation of cloud-brightness temperature. Statistical analyses found that the first category is mostly relevant to the rainfall intensity in a manner of negative correlation and that the size in the second category is positively correlated to rainfall volume. Although the features in the third category are not necessarily directly related to rain rate or rainfall volume, they do improve the discrimination of cloud clusters. The fourth category is the geo-location (latitude and longitude) of center of the cloud patch and the averaged altitude of ground surface shadowed by the cloud patch.

2.3 SONO: cloud patch classification and rainfall mapping

An automated neural network suitable for cloud-patch based rainfall estimation, entitled Self-Organizing Nonlinear Output (SONO) model, is developed to account for the high variability of cloud-rainfall processes at small scales (Hong et al. 2005a). The SONO model is a modification to the Self-Organizing Linear Output (SOLO) model, designed for efficient and effective estimation of network parameters and output (Hsu et al. 2002). In this study, SONO executes cloud classification and T_b-R mapping by performing the two basic functions of a "switchboard" and an "approximator". First, the SOFM (Self-Organizing Feature Map) functions as a "switchboard" to switch "on" or "off" the nodes in the nonlinear output layer, i.e., the SOFM classifies cloud patches into a number of clusters and determines to which node (cluster) in the nonlinear output layer it must be routed for approximation of the T_b-R relationship. The "approximator" in the nonlinear output layer calibrates a nonlinear T_b-R function for the cloud-patch cluster, which is turned "on" by SOFM. Therefore, SONO consists of a number of T_b-R functions, and each function addresses one cluster of clouds that contains similar features.

In each classified cloud-patch group, the T_b-R pixel pairs are first redistributed using the Probability Matching Method (PMM; Atlas et al. 1990). This method matches histograms of T_b and R observations such that the proportion of the R distribution above a given rain rate is equal to the proportion of the T_b distribution below the associated T_b threshold value. The redistributed pixels are fitted with a nonlinear exponential function for each patch group. Given classified patch group j, the T_b-R relationship is specified as:

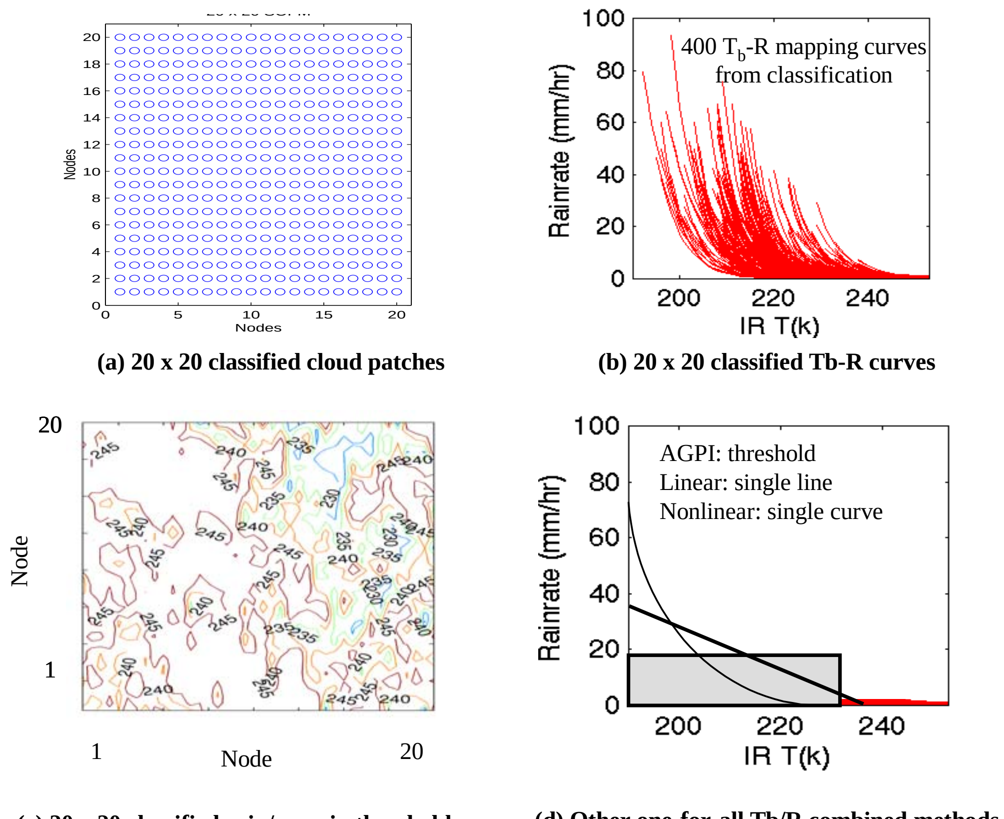
$$R^j = v^j{}_1 + v^j{}_2 \cdot \exp[v^j{}_3 \cdot (T_b + v^j{}_4)^{v^j{}_5}] \tag{1}$$

Where R is the rainfall rate (mm hr^{-1}), T_b is the cloud top brightness temperature (K), and v^j_1, v^j_2, v^j_3, v^j_4, *and* v^j_5 are parameters with respect to patch group j. The parameters of the T_b-R functions in each patch group are calibrated using a large amount of GOES infrared images and their co-registered microwave or radar rainfall. This SONO is an automated network system, state variables and model parameters of which can be recursively adjusted/re-calibrated whenever accumulation of new observations from ground or space is sufficient (e.g. pentad fashion) during later operation.

3. MODEL TRAINING AND CASE STUDY

Three data sets are used to conduct model initialization and case study. The brightness temperature images of IR channel (10.7*um*) are stored at the Climate Prediction Center (Janowiak et al. 2001), and the nationwide radar network—Weather Surveillance Radar-1988 Doppler (WSR-88D)—is used as reference data in the model calibration and validation. In addition to the radar rainfall, low-orbital satellite rainfall data derived from the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) instrument are also used in the model adjustment and evaluation.

Choosing the network size of the CCS is a tradeoff between efficiency and accuracy of model fitting; the decision is usually made through progressively increasing the classification network size until a stable fitting criterion is reached. In this study, the SOFM classification layer is determined as a 20 x 20 matrix (Figure 4a), which classifies cloud patches into 400 groups. The minimum (maximum) count of cloud-patch samples held by a classification unit is 51 (855), while the average count is 192.



(a) 20 x 20 classified cloud patches

(b) 20 x 20 classified Tb-R curves

(c) 20 x 20 classified rain/no-rain thresholds

(d) Other one-for-all Tb/R combined methods

Figure 4. Model Training and Classified Curve fitting

Note that both the SOFM classification layer and nonlinear mapping layer consist of the same arrangement of units in a 2-D coordinate. Therefore, a matrix (20 x 20) of T_b-R relationships is determined according to Equation (1), and all the 400 curves are plotted on an T_b-R plane (Figure 4b). Long and flat curves indicate cirrus clouds where cloud-top temperatures are usually cold but produce little or no rain. Steep curves represent convective clouds that are capable of producing significant rainfall. In particular, curves with short temperature range and steep slope are relevant to convective cloud in its early stage and often go largely unnoticed by many rainfall estimation techniques. Notably, this designed feature enables CCS to generate varied rain rates at a given brightness temperature for different cloud types, which overcomes the one-to-one mapping limitation of a single T_b-R function for the full spectrum of cloud-rainfall conditions.

It is worth mentioning that this algorithm is capable of generating variable rain/no-rain IR thresholds (IR_T) for different cloud groups. The appropriate IR_T is derived for each cloud group by matching a small rain rate (i.e., 0.1 mm hr^{-1}) according to the fitting function. The IR_T varies from 220K to 245K with respect to the 20 x 20 classified cloud groups, as shown in Figure 4c. Clearly, the large variation shown in Figures 4b-c cannot be well represented by a single T_b-R mapping relationship as shown in Figure 4d.

During the evolution of a raining cloud, the T_b-R distribution varies significantly from the initial stage to the dissipated stage of the cloud patch. Figure 5 illustrates the evolution of a convective cloud patch from beginning (small-warm), pre-mature (midsize-cold), mature (maximum size-coldest), dissipating (midsize-cold), and to disappearing life stages. Note that the corresponding T_b-R functions also change, in sequence for each life stage, respectively. This clearly shows CCS possesses temporal correspondence between cloud and precipitation at every life stage of the cloud patch, which region-based T_b-R approaches would not capture.

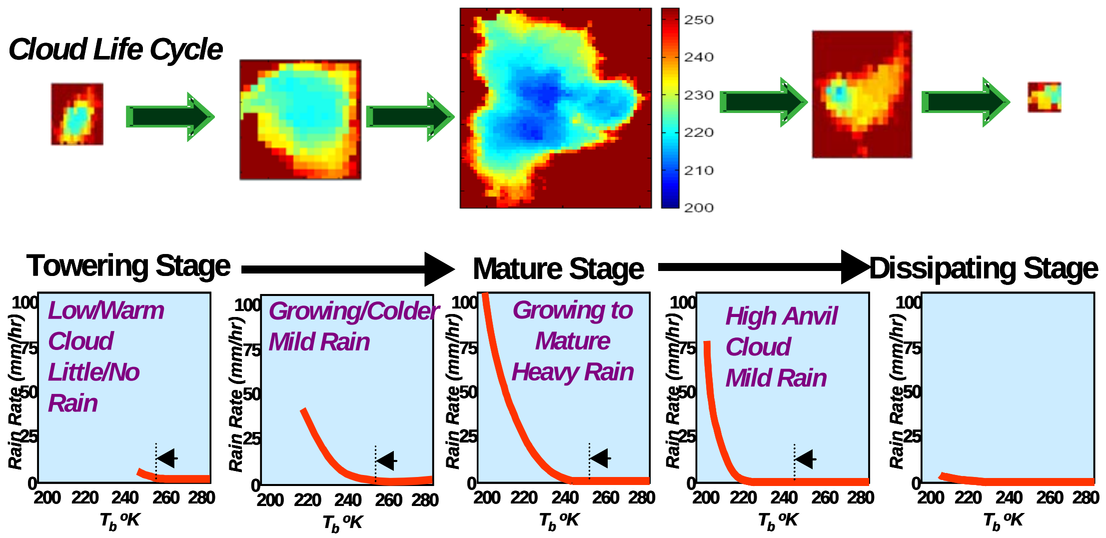


Figure 5.Cloud life cycle and its variable Tb-R histogram matching.

Hourly rainfall on 8-9 July 1999 were simulated over a relatively large area (25^o-45^oN; 100^o-130^oW) and then accumulated to daily rainfall for model evaluation. A comparison of CCS with PERSIANN results, after re-mapping at a daily 0.25^ox 0.25^o scale, as shown in Figure 6, indicates that PERSIANN estimates demonstrates lower spatial correlation than CCS.

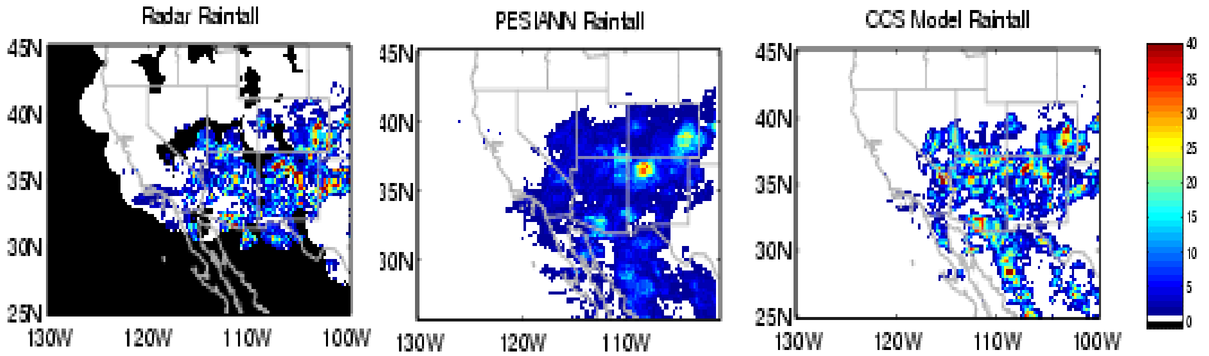


Figure 6. Comparison of daily rainfall from radar, PERSIANN, an d CCS daily estimates at 0.25^ox0.25^o on July 9, 1999.(Source: Hong et al. 2005b)

4. CONCLUSION AND FUTURE WORK

Reliable product of precipitation at high spatial and temporal resolution would be invaluable for hydrology-related researches and applications. Within this decade, the deployment of Global Precipitation Measurement constellation of low-altitude orbiting satellites will increase to 90% of the global that will be sampled with a return interval of less than 3-hours. The use of this resource in conjunction with the existing suite of geostationary satellites is expected to result in significant improvements in scale and accuracy of precipitation estimates. We have responded to this by developing a new high-resolution precipitation estimation algorithm dubbed “CCS” (Cloud Classification System) at UC Irvine. The CCS uses computer image processing and pattern recognition techniques to develop a patch-based cloud classification and rainfall estimation system based on co-registered passive microwave and infrared images from Low Earth-orbiting and Geostationary satellites.

From the beginning of this analysis, it was anticipated that the relationship between cloud-top temperature and surface rain rate varies significantly from cloud patch to cloud patch. Therefore, a rainfall estimation model characterized by significant transience, heterogeneity, and variability is needed to associate rainfall with the extremely complex and still imperfectly understood precipitating processes to produce higher quality estimates. Unlike region-based approach, which calibrates only one T_b-R function for all clouds, this technique classifies varied patches into different clusters and then searches the best-matched nonlinear T_b-R mapping function for each patch. This designed feature enables CCS to generate various rain rates at a given brightness temperature and variable rain/no-rain IR thresholds for different cloud patches, which overcomes the one-to-one mapping limitation of a single statistical T_b-R function for the full spectrum of cloud-rainfall conditions. In addition, the computational and modeling strengths of neural network enable CCS to cope with the nonlinearity of cloud-rainfall relationships by fusing multi-source data sets.

Currently, the CCS operational system moves toward two directions: 1) incorporating more accurate measurements from multi-spectral and multi-platform microwave-derived precipitation estimates; and 2) enhancing the existing system by sequentially on-line updating the model network and parameters (long-term memory) in time and space accordingly. The computational strength of Neural Network enables us to build up a database of cloud type-rainfall mapping relationships, which are undergoing recursive (in space and time) data assimilation and system update, allowing for flexibility in the adjustment of the cloud-precipitation mapping relationships as new ground or space-based microwave/radar measurements become available. We are operating this system with the goal to produce data at spatio-temporal resolution suitable for basin scale hydrological research and applications, along with the goal to provide high-quality precipitation analysis for GEWEX CEOP sites.

5. REFERENCES

Adams, R. and L. Bischof (1994), Seeded region growing, *IEEE Trans. Pattern Anal. Machine Intell*., *16(6)*, 641-647.

Adler, R. F. and A. J. Negri, 1988: A satellite infrared technique to estimate tropical convective and stratiform rainfall. *J. Appl. Meteor*., **27**, 30-51.

Adler, R. F., C. Kidd, G. Petty, M. Morrisey, and M. H. Goodman, 2001: Intercomparison of global precipitation products: The Third Precipitation Intercomparison Projects (PIP-3). *Bull. Amer. Meteor. Soc.*, **82,** 1377-1396

Atlas, D., D. Rosenfeld, and D.B. Wolff (1990), Climatologically tuned reflectivity-rain rate relations and links to area-time integrals, *J. Appl. Meteor.*, *29*, 1120-1135.

Griffith, C. G., W. L. Woodley, P. G. Grube, D. W. Martin, J. Stout., and D. N. Sikdar, 1978: Rain estimation from geosynchronous satellite imagery—visible and infrared studies. *Mon. Wea. Rev.,* **106**, 1153-1171.

Hsu, K., H.V. Gupta, X. Gao, S. Sorooshian, and B. Imam (2002), Self-Organizing Linear Output (SOLO): an artificial neural network suitable for hydrologic modeling and analysis, *Water Resour. Res., 38(12)*, 1302-1318.

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

Janowiak, J.E., R.J. Joyce, and Y. Yarosh (2001), A real-time global half-hourly pixel resolution IR dataset and its applications, *Bull. Amer. Meteor. Soc.*, *82*, 205-217.

Kidd, C., D. R. Kniveton, M. C. Todd, T. J. Bellerby, 2003: Satellite Rainfall Estimation Using Combined Passive Microwave and Infrared Algorithms, *J. of Hydrometeor.,* **4(6)**, 1088–1104

Hong, Y., K. Hsu, and S. Sorooshian (2003), An automatic segmentation algorithm for cloud infrared satellite images: incremental temperature threshold technique, Technical Report, 03-020, 55 p., Dept. of Hydrology and Water Resources, University of Arizona, Tucson, Arizona.

Hong, Y., K. Hsu, S. Sorooshian, and X. Gao, 2005a Self-organizing nonlinear output (SONO): a neural network suitable for cloud patch-based distributed rainfall estimation from remotely sensed information, *Water Resource Research, in Press.*

Hong, Y., K. Hsu, X. Gao, and S. Sorooshian, *2005b,* Precipitation Estimation from Remotely Sensed Imagery Using Artificial Neural Network - Cloud Classification System, *Journal of Applied Meteorology,* Vol. 43, No. 12, pp. 1834–1853

Tapiador, F.J., C. Kidd, V. Levizzani, and F.S. Maraano (2002), A neural network PMW/IR combined procedure for short term/small area rainfall estimates, Proceedings of 1st Workshop of International Precipitation Working Group,

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

Woodley, W. L., C. G. Griffith, J. S. Griffin, and S. C. Stroomatt, 1980: The inference of GATE convective rainfall from SMS-1 imagery. *J. Appl. Meteor*., **19**, 338-408.

Xu, L., S. Sorooshian, X. Gao, H. Gupta, 1999: A Cloud-Patch Technique for Identification and Removal of No-Rain Clouds from Satellite Infrared Imagery**.** *J. Appl. Meteor.,* **38**, 1170–1181