UC Irvine
UC Irvine Previously Published Works

Title

Self-organizing nonlinear output (SONO): A neural network suitable for cloud patch–based rainfall estimation at small scales

Permalink

https://escholarship.org/uc/item/1q95s8d3

Journal

Water Resources Research, 41(3)

ISSN

0043-1397

Authors

Hong, Yang
Hsu, Kuo-lin
Sorooshian, Soroosh
et al.

Publication Date

2005-03-01

DOI

10.1029/2004wr003142

Copyright Information

Peer reviewed

Self-organizing nonlinear output (SONO): A neural network suitable for cloud patch−based rainfall estimation at small scales

Yang Hong,[1] Kuo-lin Hsu, Soroosh Sorooshian, and Xiaogang Gao

Center for Hydrometeorology and Remote Sensing, Department of Civil and Environmental Engineering, University of California, Irvine, California, USA

[1] Accurate measurement of rainfall distribution at various spatial and temporal scales is crucial for hydrological modeling and water resources management. In the literature of satellite rainfall estimation, many efforts have been made to calibrate a statistical relationship (including threshold, linear, or nonlinear) between cloud infrared (IR) brightness temperatures and surface rain rates (RR). In this study, an automated neural network 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 geostationary scales (i.e., 4 km and every 30 min). Instead of calibrating only one IR-RR function for all clouds the SONO classifies varied cloud patches into different clusters and then searches a nonlinear IR-RR mapping function for each cluster. This designed feature enables SONO to generate various rain rates at a given brightness temperature and variable rain/no-rain IR thresholds for different cloud types, which overcomes the one-to-one mapping limitation of a single statistical IR-RR function for the full spectrum of cloud-rainfall conditions. In addition, the computational and modeling strengths of neural network enable SONO to cope with the nonlinearity of cloud-rainfall relationships by fusing multisource data sets. Evaluated at various temporal and spatial scales, SONO shows improvements of estimation accuracy, both in rain intensity and in detection of rain/no-rain pixels. Further examination of the SONO adaptability demonstrates its potentiality as an operational satellite rainfall estimation system that uses the passive microwave rainfall observations from low-orbiting satellites to adjust the IR-based rainfall estimates at the resolution of geostationary satellites.

1. Introduction

[2] Accurate measurement of precipitation information is critical in understanding the small- and large-scale components within the hydrologic cycle and in understanding the balance of Earth's energy and water cycle exchange that drives most hydrological modeling and water resources applications [*Sorooshian*, 2004]. The distribution of precipitation at various spatial and temporal scales is also crucial for water management for agriculture, electrical power, and drought and flood control. Rainfall measurement from satellites has been an active field of study for decades. Both visible/infrared (IR) and microwave (MW) radiometers are used to infer precipitation. In the algorithm intercomparison experiment, *Ebert and Manton* [1998] summarized that the MW-based algorithms give better instantaneous estimates; however, IR-based algorithms give better long-term estimates, such as monthly rainfall, from the sensors that provide better temporal samples. Naturally, several combined IR/MW rainfall estimation algorithms also have been developed in order to make use of the individual strengths of the IR and MW data [*Kidd et al.*, 2003].

[3] Satellite-based rainfall estimation approach is capable of generating precipitation at or near global coverage, which is suitable as a supplement of the ground observations in providing precipitation product for wide applications [*Adler et al.*, 2003]. The World Climate Research Programme (WCRP) established the Global Precipitation Climatology Project (GPCP) that has been succeeding in producing precipitation data of $2.5° \times 2.5°$ (latitude-longitude) monthly accumulations of estimates over the last two decades. Since 1996, the daily $1°$ gridded rainfall product using the Geostationary Precipitation Index (GPI) has been archived under the auspices of the GEWEX (Global Energy and Water cycle Exchange). *Sorooshian et al.* [2000] have been providing global tropical $(35°S−35°N)$ 6-hour $1° \times 1°$ rainfall data from the year 2000 to the present. Currently, numerous applications of hydrology and water resource management impose a growing need of precipitation measurements at shorter sampling frequencies (subdaily or 3-hour) and higher spatial scale (25 km or down to geostationary satellite pixel resolution, 4 km).

[4] Because satellites measure rainfall as an integral of space at a point in time, the sampling frequency (near $4 \times 4 \ km^2$ and 30-min sampling interval) of infrared

---

[1]Formerly at Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona, USA.

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 (RR) observation either from the ground (gauge/radar) or from space (microwave imager). This statistical relationship, hereinafter referred to as IR-RR, is then applied at the temporal resolution of IR data from geostationary satellites. According to the types of fitting functions, the IR-RR relationships may be broken down further into (1) threshold-based, (2) linear fitting, and (3) nonlinear fitting approaches, as described in equations (1)–(3), respectively.

[5] GPI (geostationary precipitation index) [*Arkin and Meisner*, 1987] is a typical threshold algorithm, which derives rainfall (RR) over large grid cells (e.g., 2.5°):

$$RR = F_c \times G \times IR_T \qquad (1)$$

where F_c is the fractional coverage of cloud, G is the fixed coefficient of rain rate (3 mm hr^{-1}), and IR_T is the rain/no-rain threshold 235K.

[6] G rain rates are assigned to pixels colder than IR_T. Although GPI provides a useful benchmark to assess other algorithms, it is essentially an area-time integral approach to rainfall estimation [*Atlas and Bell*, 1992], adjusted GPI (AGPI) [*Adler et al.*, 1994], and universal AGPI (UAGPI) [*Xu et al.*, 1999a] employ the similar GPI idea but with threshold temperatures (IR$_T$) and rain rates (G) adjustable monthly.

[7] In the linear fitting approach, *Miller et al.* [2001] suggested equation (2) to fit the IR-RR relationship:

$$RR = a \times IR + b, \qquad (2)$$

where a and b are parameters. *Vicente et al.* [1998] calibrated IR and rainfall data from the Amazon plain using a two-parameter exponential IR-RR function (see equation (3)) to estimate rainfall with postadjustment from other factors such as soil moisture. However, the parameters v_1 and v_2 are fixed once after the initialization.

$$\begin{aligned} RR &= v_1 \times \exp(v_2 \times IR) \\ &= 1.1183 \times 10^{11} \times \exp(-0.036382 \times IR) \end{aligned} \qquad (3)$$

[8] The above equations assume that colder IR temperatures are always associated with higher rain rates. Although this assumption is widely used and has been demonstrated to be quite effective at climatic scale, e.g., monthly 2.5°, however, nonlinear relationships generally exist between cloud IR data and surface rain rates at higher resolutions [*Arkin and Meisner*, 1987; *Adler et al.*, 1994]. Particularly, this assumption may not be able to overcome the contamination by very cold anvil cirrus clouds and has difficulty in detecting relatively warm cloud rainfall. In addition, these equations fail to represent the variation of cloud-precipitation relationships because they only statistically determine one single set of parameters for the IR-RR functions and then apply to whole study region for subsequent time period. Although enjoying the benefits of simplicity and low computational cost, they provide only a climatologic calibration and may not respond adequately
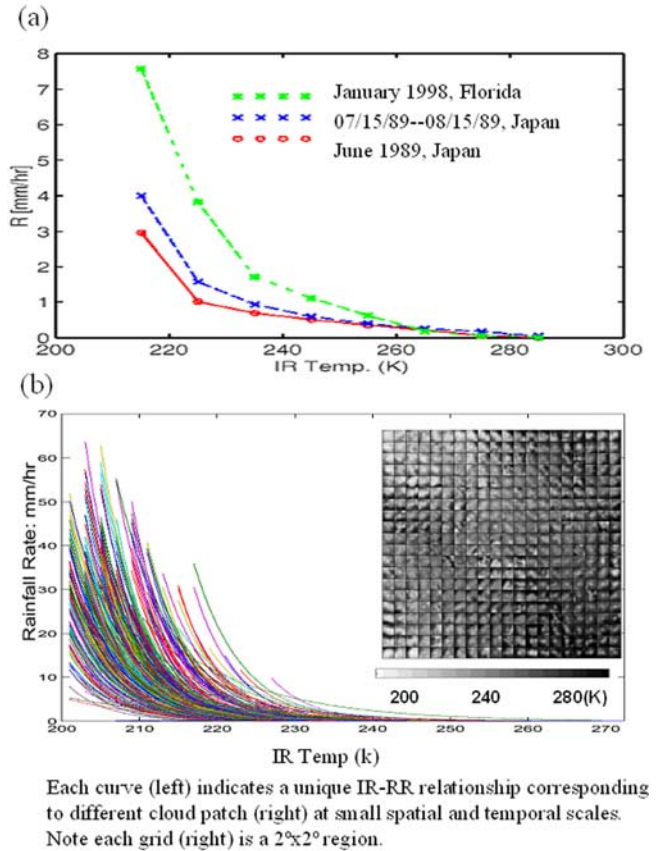
Figure 1. Diverse cloud IR temperature and rain rate relationship (IR$_{Tb}$-RR). (a) Monthly averaged IR-RR curves in three case studies (source *Hsu et al.* [1997]). (b) Cloud patch−based IR-RR relationships, which vary significantly from one cloud patch to another at high resolution, such as hourly and 4 km.

to hydrometeorological variations. As a result, considerable temporal and spatial integration, e.g., monthly 2.5°, is conducted to reduce the estimation error. Rainfall estimation at smaller scales remains a challenge.

[9] Condensed water vapor appears in the atmosphere as a variety of types of clouds. The cloud-precipitation mechanism is determined by different dynamical and thermodynamical processes, which are highly time-dependent and space-variant. Figure 1a shows the IR-RR relationships are seasonally and regionally dependent [*Hsu et al.*, 1997]. As can be seen from Figure 1b, the IR-RR relationships can vary significantly from one cloud patch to another at small scale, such as hourly and 4-km grids. Previous studies [*Griffith et al.*, 1978; *Woodley et al.*, 1980; *Adler and Negri*, 1988; *Xu et al.*, 1999b] suggested that the development of a cloud patch−based algorithm is desired to capture the high variation of precipitation distribution. Classification of cloud patches into a number of groups would improve discrimination of different precipitation-generating systems, as suggested in Figure 1. Adopting the ideas from the above studies, an artificial neural network, self-organizing nonlinear output (SONO), is proposed to account for the high variability in cloud-rainfall processes. The unique feature of the SONO is that it searches multiple IR-RR mapping relationships for different clouds based on a natural seg-
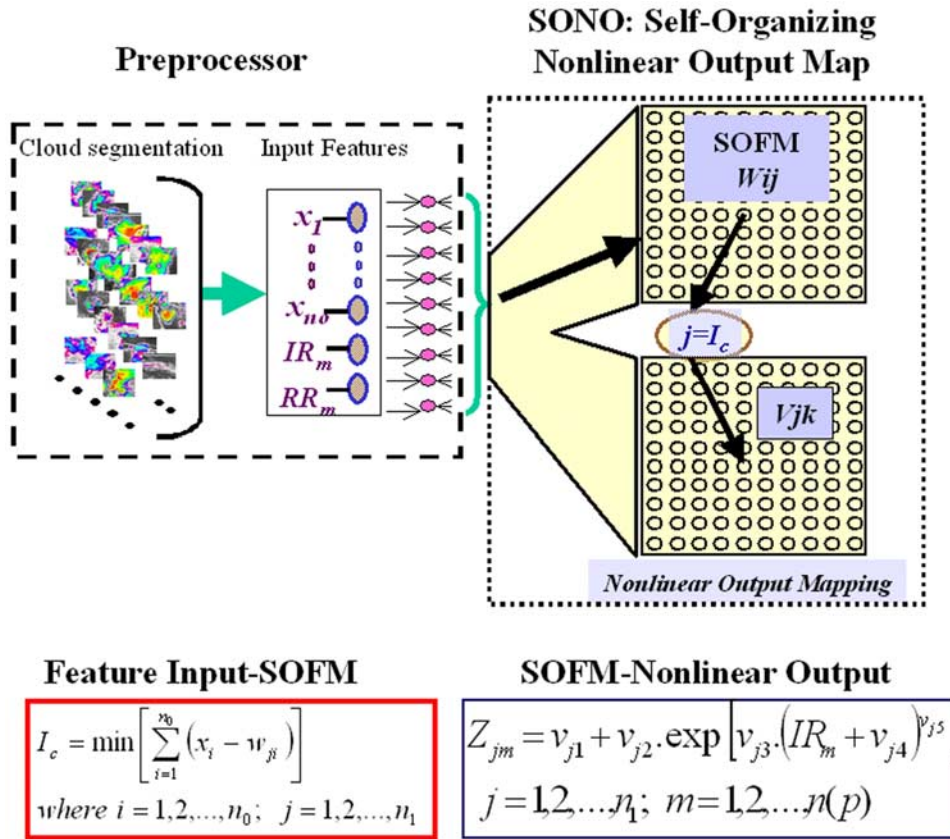
Preprocessor

SONO: Self-Organizing Nonlinear Output Map

Cloud segmentation   Input Features

x_1

x_{n0}

IR_m

RR_m

SOFM W_{ij}

$j = I_c$

V_{jk}

Nonlinear Output Mapping

Feature Input-SOFM

$$I_c = \min\left[\sum_{i=1}^{n_0}\left(x_i - w_{ji}\right)\right]$$

$$\text{where } i = 1,2,\ldots,n_0; \quad j = 1,2,\ldots,n_1$$

SOFM-Nonlinear Output

$$Z_{jm} = v_{j1} + v_{j2}.\exp\left[v_{j3}.\left(IR_m + v_{j4}\right)^{v_{j5}}\right]$$

$$j = 1,2,\ldots n_1; \quad m = 1,2,\ldots n(p)$$

Figure 2. Preprocessor and self-organizing nonlinear output (SONO).

$T_b = 220k$

$T_b = 235k$

$T_b = 253k$

$T_b = 220k$

$T_b = 235k$

$T_b = 253k$
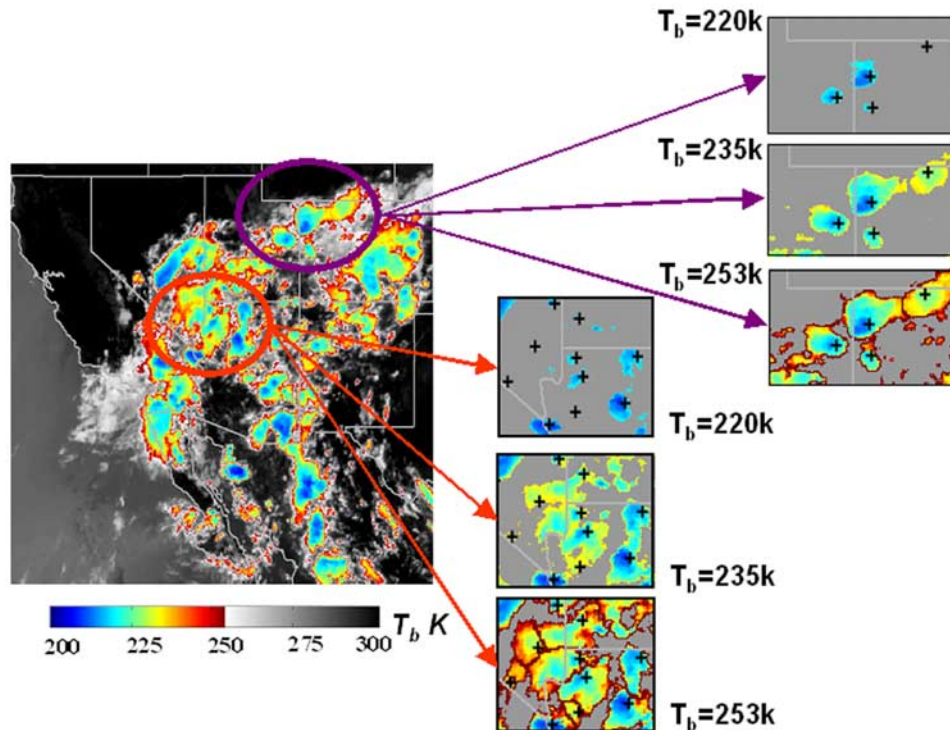
200   225   250   275   300 $T_b\, K$

Figure 3. The segmented cloud patches from GOES infrared imagery (0000 UTC, 9 July 1999) using ITT segmentation algorithm.

Table 1. Input Features Extracted From Cloud Patches

Category	Features
Coldness	minimum temperature of a cloud patch (IR_{min}) and mean temperature of a cloud patch (IR_{mean})
Geometric	cloud patch size (size) and cloud patch shape index (SI)
Texture	standard deviation of cloud patch (STD), mean value of local standard deviation of cloud patch (MSTD (5×5)), standard deviation of local standard deviation (STD_{std} (5×5)), and gradient of cloud top brightness temperature (TOPG)

mentation/classification of clouds (i.e., a cloud patch–based manner), which overcomes the limitation of some other techniques (e.g., equations (1)–(3)) that only determine a single IR/rainfall function to an arbitrary calibration domain (in space and time). Therefore SONO 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 enable SONO to cope with the nonlinearity of cloud-rainfall relationships by fusing multisource data sets [*Tapiador et al.*, 2002].

[10] The scope of this paper is organized as follows. The development of SONO, which includes segmentation and classification of cloud image and IR-RR function mapping, is described in section 2. Model calibration is presented in section 3, and model validation and comparison are given in section 4. Finally, the summary and future development for SONO are discussed in section 5.

2. Principle of the Self-Organizing Nonlinear Output

[11] 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]. As shown in Figure 2, SONO has an input preprocessor, which segments IR images and extracts cloud features as input variables. Then, SONO performs 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 IR-RR relationship. The approximator in the nonlinear output layer calibrates a nonlinear IR-RR function for the cloud patch cluster, which is turned "on" by SOFM. Therefore SONO consists of a number of IR-RR functions, and each function addresses one cluster of clouds that contains similar features.

2.1. SONO Preprocessor: Cloud Patch Segmentation and Feature Extraction

[12] Segmentation of IR imagery is a preprocess 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.*, 2004], 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 [*Adams and Bischof*, 1994], and, meanwhile, eliminates manual selection of seeds through hierarchical thresholding. Examples of cloud segmentation using the ITT method are shown in Figure 3. Given a snapshot of satellite IR images, ITT first locates the minimum temperature (IR_{min}) 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.

[13] 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 interrelationships (i.e., correlation and covariance) among the features help to determine the importance of the features in discriminating classes. A final set of eight features was chosen (Table 1). 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 formulas for computation of each feature shown in Table 1 are detailed in Appendix A.

[14] Figure 1b also singles out several cloud patches (right) delineated by ITT, and each patch corresponds to a variable cloud-precipitating mapping curve (left) because of their different features in terms of brightness temperature, size, and texture. Therefore cloud patch segmentation and feature extraction prepare the input to the SONO, a cloud patch–based rainfall estimation system.

2.2. SONO Network

[15] Given a cloud patch, p, the preprocessor provides three sets of input vectors for SONO:

$$p = ([x_i], [IR_m], [RR_m]), \qquad I = 1, \ldots, n_0; m = 1, \ldots, N(p) \tag{4}$$

where $[x_i] = (x_1, x_2, \ldots, x_{n0})$ is the feature vector listed in Table 1; n_0 is the number of input features; $[IR_m]$ is

the IR brightness temperature vector within the patch; [RR$_m$], if available, is the corresponding observation of rainfall vector; and $N(p)$ is the number of pixels of the cloud patch, p.

[16] SONO consists of two major components: SOFM classification layer and nonlinear regression mapping layer (see Figure 2). The classification layer compares n_1 units or groups: $\mathbf{y} = [y_1, y_2, \ldots, y_{n1}]$; the weight matrix w_{ij} is a set of parameters connecting from the input node x_i to the classification units y_j. The mapping layer, $\mathbf{z} = [z_1, z_2, \ldots, z_{n1}]$ contains n_1 nonlinear fitting functions, where v_{jk} is the kth parameters of a nonlinear function connecting from unit y_j to output unit z_j. The SOFM classification proceeds based on the similarity of the cloud patch features $\mathbf{x}$ and the connection weights $\mathbf{w}$. The training of SOFM projects input patterns of many variables into an organized cloud classification map. Details of the training procedure are described by *Kohonen* [2001] and *Hsu et al.* [1999]. A brief summary of the training algorithm is listed below.

[17] 1. Prepare the input $\mathbf{x}$ and initialize the weights w_{ij} as normalized random numbers.

[18] 2. Determine the winner node that has minimum distance between $\mathbf{x}$ and weight $\mathbf{w}$:

$$j^* = \arg_j \left(\min \| x_{ij} - w_{ij} \| \right) \qquad (5)$$

[19] 3. Update weight for the neighborhood (radius r) nodes of j^* with learn rate α:

$$w_{ij} = w_{ij} + \alpha \left(x_{ij} - w_{ij} \right) \qquad (6)$$

[20] 4. Terminate if the w_{ij} is converged or reduce r and α and proceed to step 3.

[21] As a result of SOFM classification, a data set of IR-RR pairs is classified to the unit j, and the data pairs are sorted based on the probability matching method (PMM) [*Atlas et al.*, 1990]. Theoretical and experimental evidence exists that higher rain rates are associated with lower IR brightness temperatures [*Arkin and Meisner*, 1987]. Thus the PMM matches the cumulative distribution functions (CDF) of the data pairs of IR-RR, as follows:

$$\int_0^{RR_m} P(RR)dRR = \int_{IR_{min}}^{IR_{max}} P(IR)dIR - \int_{IR_{min}}^{IR_m} P(IR)dIR$$

$$= 1 - \int_{IR_{min}}^{IR_m} P(IR)dIR \qquad (7)$$

where the $P(.)$ is the probability distribution function, and the value of IR and RR is in the range [0 RR$_{max}$] and [IR$_{min}$ IR$_{max}$], respectively.
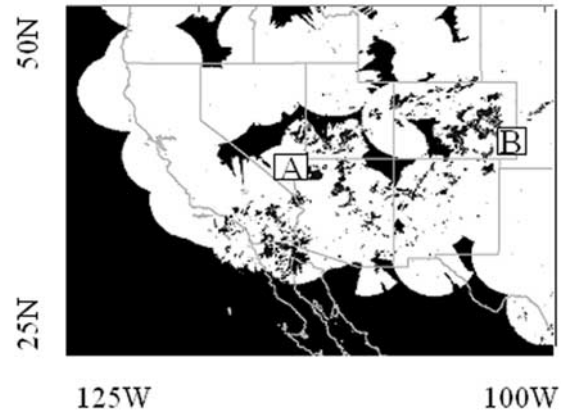
Figure 4. The space domain (25°−50°N, and 125°−100°W) for SONO calibration. Note that the white area indicates WSR-88D radar coverage, and the locations of two storm events, A and B, are chosen for the case study of SONO adaptability.

[22] The use of power law mapping between IR and RR was suggested by the study of *Gagin et al.* [1985] and implemented by *Martin et al.* [1990] and *Goodman et al.* [1994]. Here the nonlinear IR-RR fitting function is

$$\begin{aligned} RR_{jm} &= f\left(v_{jk}, IR_m\right) \\ &= v_{j1} + v_{j2} \times \exp\left[v_{j3} \times \left(IR_m + v_{j4}\right)^{v_{j5}}\right], \\ m &= 1, 2, \ldots, N(p) \end{aligned} \qquad (8)$$

The above function consists of parameters $[v_{jk}], k = 1, 2, \ldots, 5$. A look-up table is also an alternative to representing the nonlinear relationships of IR-RR. However, considering the network size of SONO, 20×20, it is not convenient to store all the data tables. In SONO, only five parameters are needed to store for each cloud cluster. In addition, it is much easier to adjust the IR-RR relationships when dealing with parameterized functions than look-up tables.

[23] Compared to the autoestimator (AE), the essential difference is that SONO possesses a large number (n_1) of IR-RR mapping relationships corresponding to different clouds, while AE applies equation (3) into all situations though corrected by other factors such as soil moisture.

3. Data and SONO Calibration

[24] The study area in this paper is shown in Figure 4. The temporal domain for calibration is June 1999, and July 2002 is selected as the validation period. Data used in this study are listed in Table 2. The brightness temperature images of IR channel (10.7 μm) are stored at the Climate

Table 2. GOES IR, WSR-88D, and TMI Data Used in This Study[a]

Data	Description
GOES IR	long-wave infrared (10.7 μm) data provided by GOES 8 and GOES 10 at 30-min 0.04° × 0.04° scale
WSR-88D radar network	hourly 0.04° × 0.04° rainfall data collected by the WSR-88D radar network operated by the National Centers for Environmental Prediction at NOAA
TRMM 2A12	instantaneous rainfall rate based upon the TRMM Microwave Imager (TMI)

[a]Spatial domain is 25°−50°N, 125°W−100°W. Calibration (validation) domain is June 1999 (July 2002).
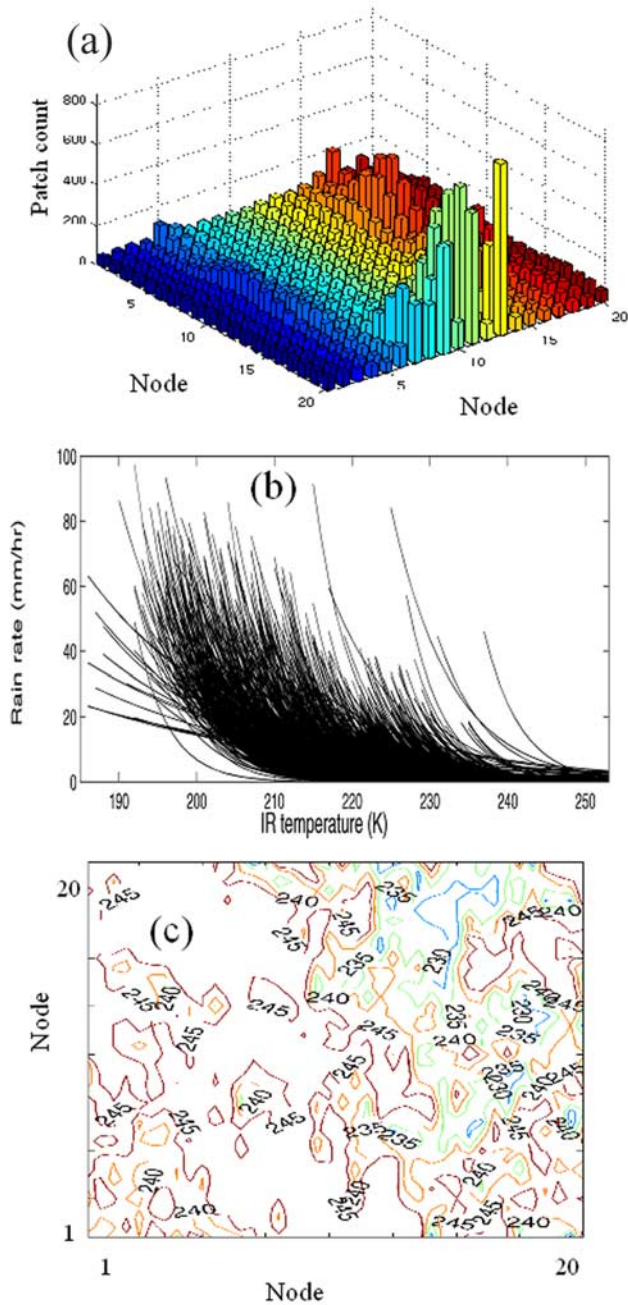
Figure 5. SONO training results: (a) number of cloud patch samples classified into each of 20 × 20 nodes after SOFM classification (minimum 51, maximum 855, and average 192), (b) the 20 × 20 IR-RR curves plotted at an IR-RR plane, and (c) the variation of rain/no-rain thresholds derived from the 20 × 20 IR-RR curves.

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 evaluation.

[25] Choosing the network size of the SONO 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 × 20 matrix (Figure 5a), 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. Statistically speaking, sufficient samples are contained for mapping IR-RR relationships in each classification unit.

[26] 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 × 20) of IR-RR relationships is calibrated according to equation (8), and all the 400 curves are plotted on an IR-RR plane (Figure 5b). 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 SONO 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 IR-RR function for the full spectrum of cloud-rainfall conditions.

[27] In previous studies, a constant temperature threshold ($IR_T = 235K$) is used in GPI [*Arkin and Meisner*, 1987]. *Adler et al.* [1994] and *Xu et al.* [1999b] later found that different IR_T are required for changing hydrolometeorological conditions. It is worth mentioning that the SONO 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 × 20 classified cloud groups, as shown in Figure 5c. Clearly, the large variation shown in Figures 5b and 5c cannot be well represented by a single IR-RR mapping curve.

4. Case Study: Application of SONO to Rainfall Estimation

[28] Validation of SONO was conducted at subdaily, daily, and monthly temporal resolutions for spatial scales such as 0.04°, 0.12°, 0.5°, and 1.0°. The validation data sets and domain (in space and time) are shown in Table 2. The statistical measures and definition used here were taken from the results of the 3rd Algorithm Intercomparison Project of the Global Precipitation Climatology Project (GPCP) [*Ebert*, 1996].

4.1. Comparison With Other Rainfall Estimation Algorithms

[29] Three types of fitting functions, threshold, linear, and nonlinear, are usually used to fit the pixel-based IR-RR relationships. In the following case study, both a threshold (UAGPI) and a nonlinear exponential fitting curve are used in the comparison. UAGPI [*Xu et al.*, 1999a] assigns a constant rain rate G to the pixels colder than certain threshold IR_T. The IR_T and the G are recalibrated at a monthly scale. The exponential function used equation (3), which approximates a power law curve to map the
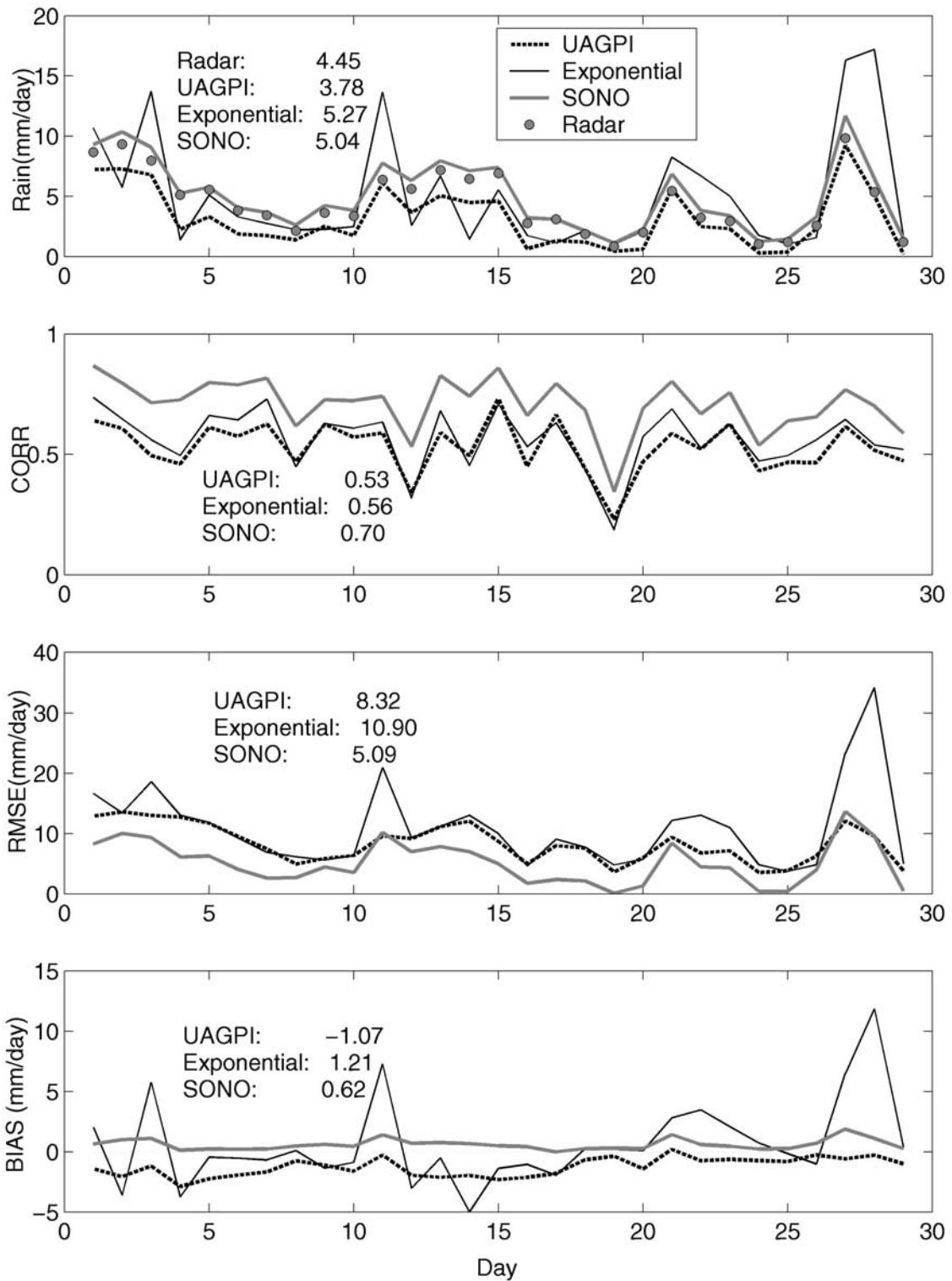
Figure 6. Plots of the time series of daily rainfall and their daily comparison statistics (Rain, CORR, RMSE, BIAS, POD, FAR, CSI, and SKILL Score) over 30°–40°N, 105°–115°W in the southwest United States from radar, UAGPI, exponential curve, and SONO in July 2002.

IR-RR relationship for all IR cloud images. SONO, however, is a patch-based rainfall estimation algorithm, which assigns a unique IR-RR relationship from a set of 400 curves to a cloud patch.

4.1.1. Daily Time Series and Monthly Rainfall (July 2002 Over the Texas Plain)

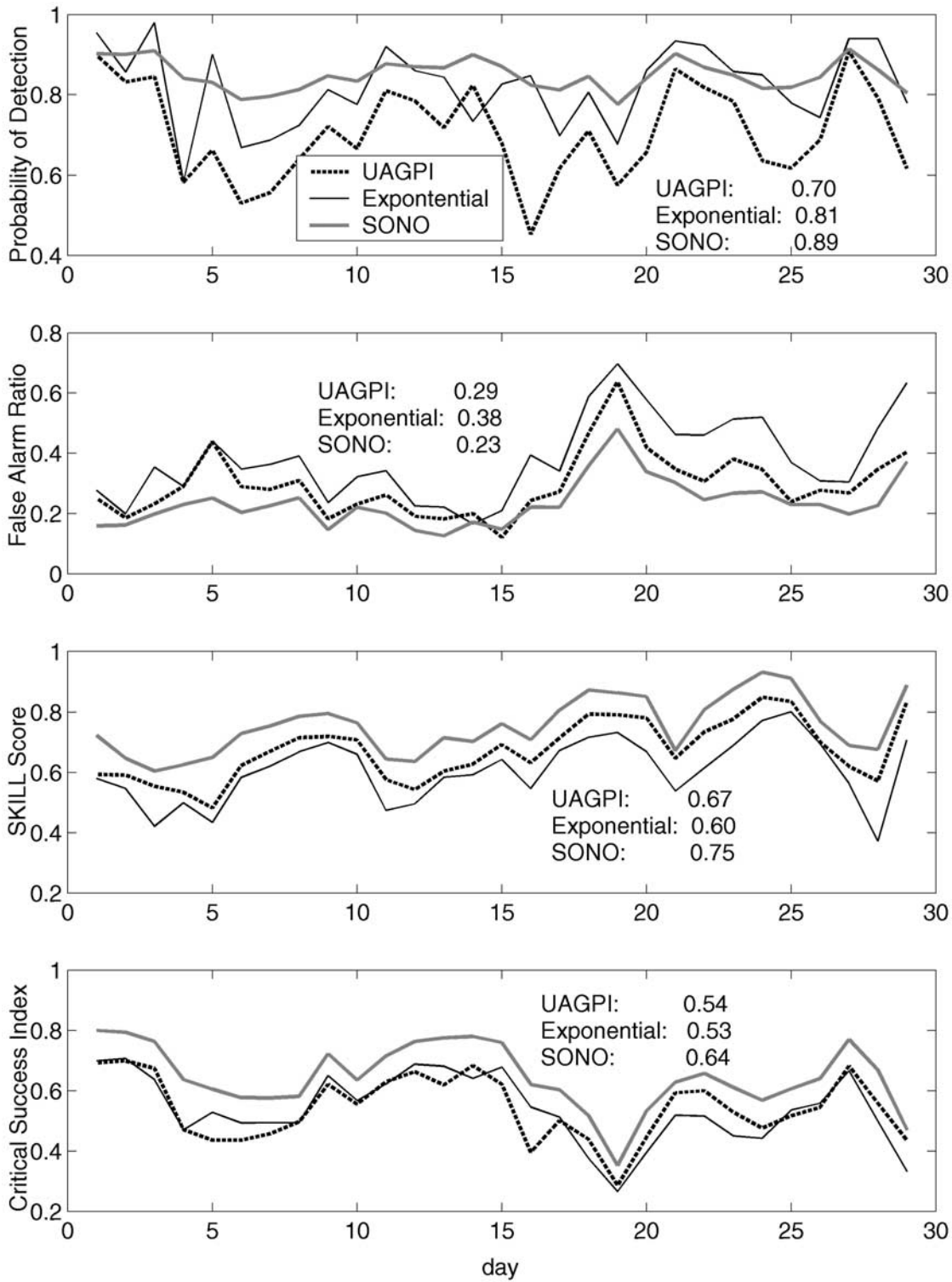[30] To assess the stability of SONO performance, a monthly (July 2002) rainfall estimation study was

Figure 6. (continued)

conducted at a $10° \times 10°$ region located in the central Texas plain ($30°–40°$N, $105°–115°$W), where high-quality ground observation data from the WSR-88D network are available for comparison [*Maddox et al.*, 2002]. SONO rainfall data are estimated on an hourly basis and then accumulated to daily and monthly scales.

[31] The statistical comparison of the daily rainfall time series derived from UAGPI, the exponential function, and the SONO against WSR-88D radar observation, including the averaged statistical information, are shown in Figure 6. The scatterplots of the monthly rainfall total at the $0.04°$ grid scale for the three algorithms are given in Figure 7. In both daily and monthly scales, the exponential
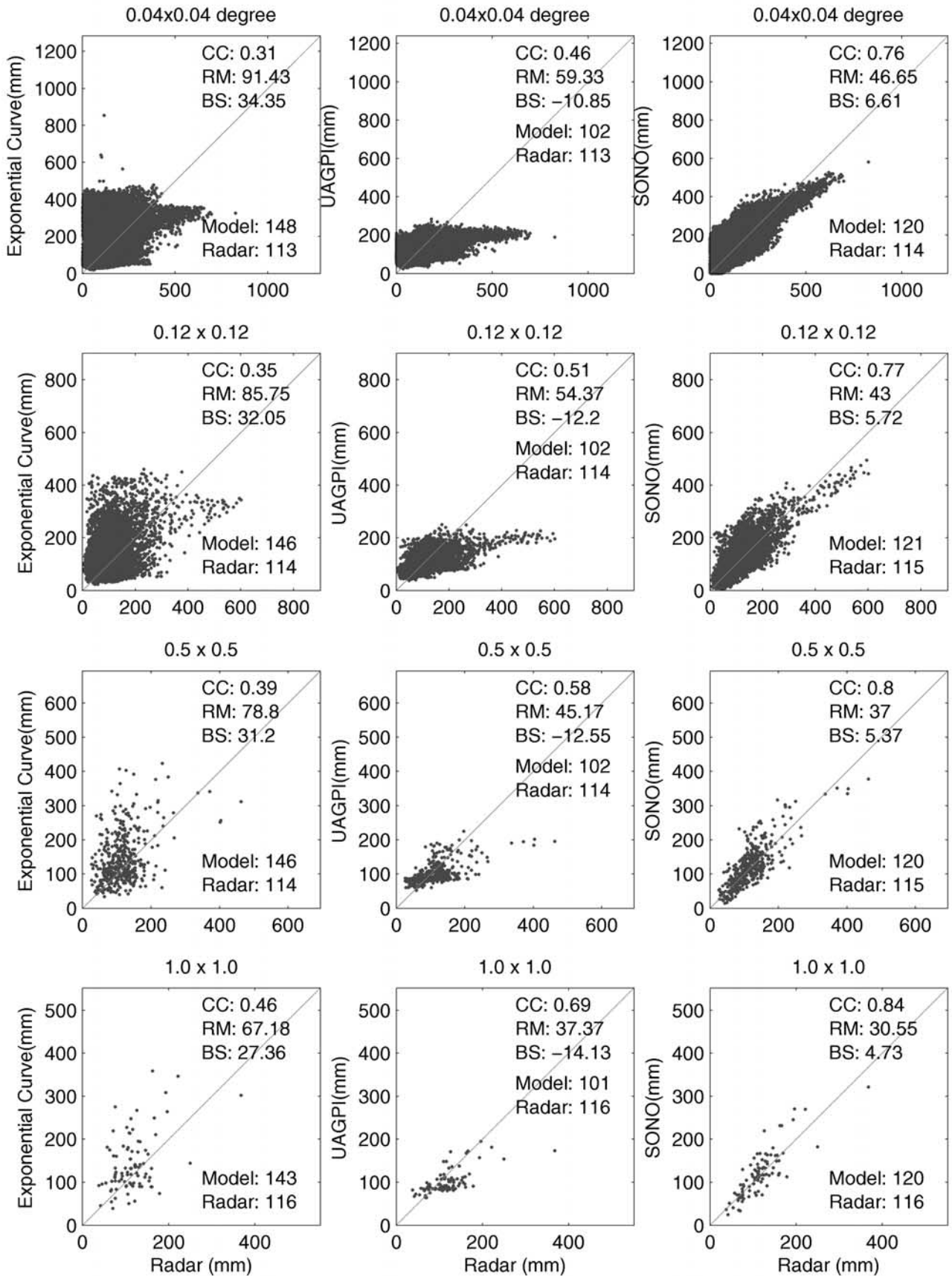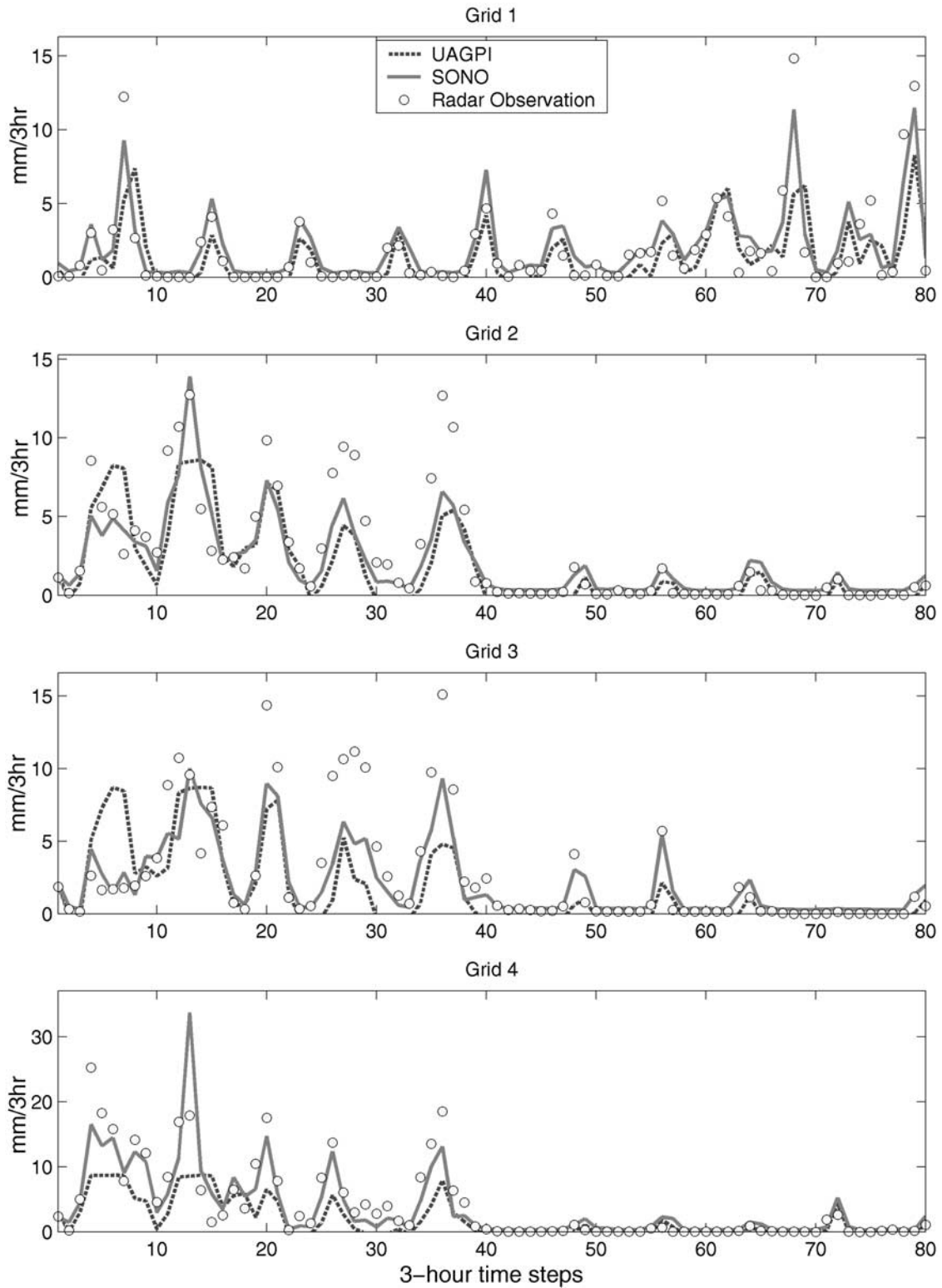
Figure 7. Scatterplots of July 2002 monthly rainfall total derived from exponential function, UAGPI, and SONO versus radar at 30°–40°N, 105°–115°W.
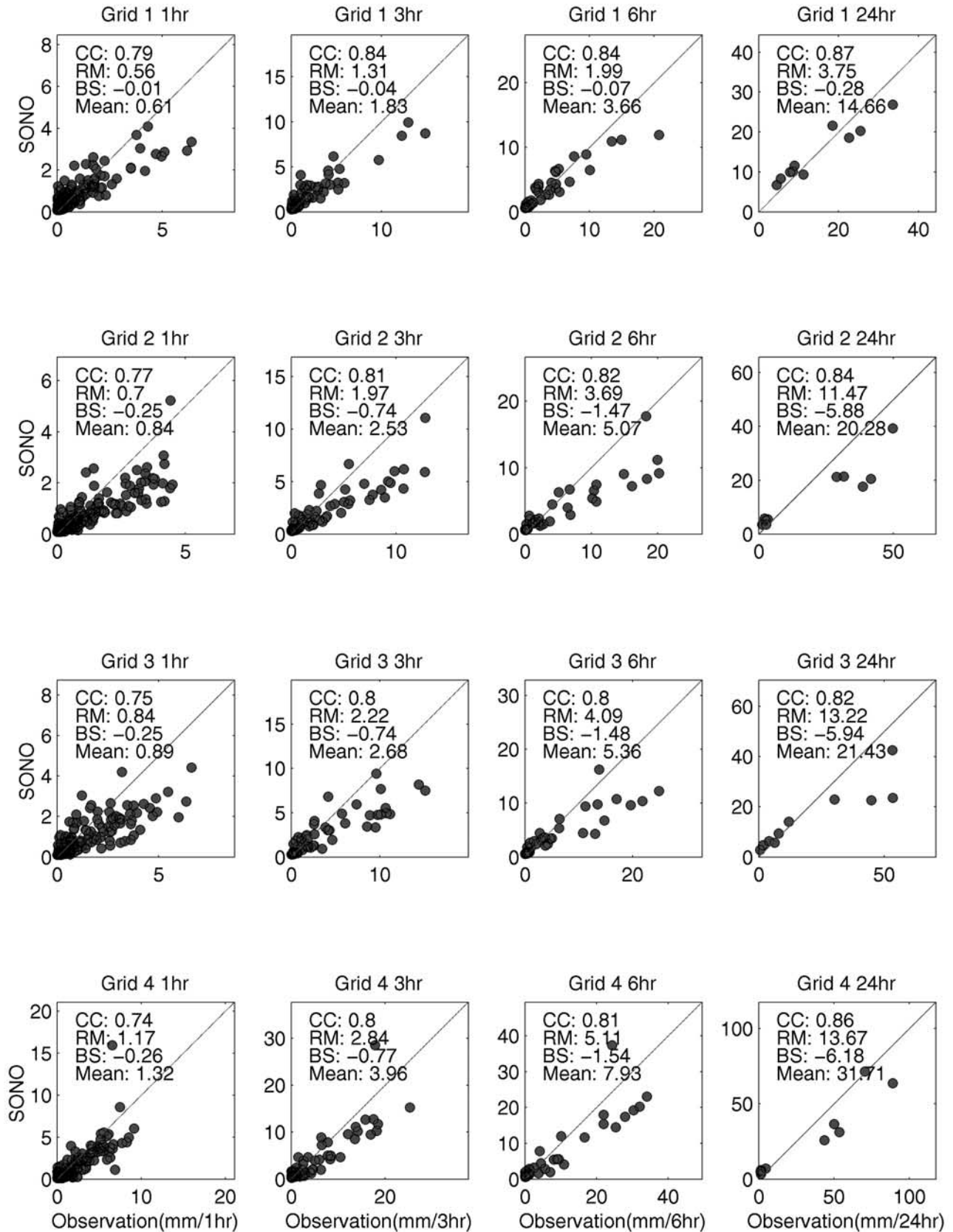
Figure 8. Scatterplots of time series of hourly, 3-hour, 6-hour, and 24-hour rainfall at four $1° \times 1°$ grids for 1–10 July 2002 located in Texas, $28°–30°N$, $102°–100°W$.

Figure 9. Time series of 3-hour rainfall at four $1° \times 1°$ grids for $1-10$ July 2002 located in Texas, $28°-30°$N, $102°-100°$W.

function catches the high rain rate events, but markedly overestimates rainfall, resulting in large root-mean-square error (RMSE) and bias, while the threshold algorithm, UAGPI, fails to catch high rainfall but shows better performance than the exponential function. However, SONO out-

performs these two single-mapping algorithms in terms of the overall statistics at both pixel and accumulated scales.

4.1.2. Time Series of Subdaily Rainfall

[32] Statistical comparison of SONO with UAGPI is conducted at subdaily timescales for four grids ($1° \times 1°$)

Table 3. Statistical Comparison of Time Series Averaged $1° \times 1°$ SONO Estimates Versus WSR-88D Radar Located in the Texas Region $(29°-30°\text{N}, 101°-100°\text{W})$ at 1-Hour, 3-Hour, 6-Hour, and Daily Intervals From 1 to 10 July 2002

Time Interval	Radar Mean, mm	RMSE, mm		CORR		Bias, mm		CSI	
		UAGPI	SONO	UAGPI	SONO	UAGPI	SONO	UAGPI	SONO
Hour	1.32	1.51	1.39	0.69	0.79	−0.62	−0.18	0.62	0.78
3 hours	3.96	4.23	3.12	0.77	0.85	−1.86	−0.55	0.77	0.95
6 hours	7.93	7.56	5.42	0.80	0.87	−3.72	−1.10	0.89	0.98
Daily	31.71	23.21	12.97	0.89	0.93	−14.93	−4.40	1.0	1.0

located in Texas from 1 to 10 July 2002. The model performance improved as the temporal resolution changed from high (hourly) to low (daily) for the four grids, as displayed in Figure 8. SONO shows a better fit of 3-hour rainfall time series than the UAGPI (see Figure 9). The comparison statistics averaged from the four grids are listed in Table 3. SONO has a higher correlation coefficient (CORR) than the UAGPI, especially at timescales less than 6 hours (Table 3). At larger timescales, i.e., daily, both UAGPI and SONO perform similarly at CORR and Critical Success Index (CSI), but the Bias and RMSE are more favorable for SONO.

4.2. Adaptability of the SONO Model

[33] The ability to update network parameters from new data sources is necessary for the SONO to track the temporal and spatial variations in the rainfall distribution over extended regions and time periods. The parameter values (w_{ij} and v_{jk}) of SONO were initially trained using 1-month (June 1999) coregistered IR and WSR-88D radar rainfall data over the region of $25°-50°\text{N}, 125°-100°\text{W}$. Afterward, the SONO can operate in two different modes: fixed and adaptive mode. In the fixed mode, the SONO generates rainfall estimates with the fixed parameters w_{ij} and v_{jk}. In the adaptive mode, the fixed mode was used as an "initial configuration", and only the parameters v_{jk} were adjusted with retraining data set before computing the rainfall estimates. The parameters are incrementally adjusted in the directions of the negative gradients of their matched error, i.e., $\eta.[t_{obs} - z_{sim}]$, where t_{obs} is observed rainfall such as microwave estimates or ground truth data, z_{sim} is the simulated SONO rainfall before adjustment, and η is the learning rate chosen between 0 and 1 (e.g., 0.05). Because the parameter updating is quite straightforward, the formula is ignored here. For more detail, please refer to *Hsu et al.* [1999]. As a result, the fixed mode of SONO generates rainfall estimates with the existing network parameters, while the adaptive mode estimates rainfall with the adjusted parameters through its recursive updating procedure.

[34] Section 4.2.1 addresses the first adaptive case study with retraining data set of coregistered IR and WSR-88D

radar data for July 1999 over the same space domain as the fixed mode. The second adaptive case uses the retraining data set of coregistered IR and two swathes of TMI rainfall product (2A12) during July 2002 over two locations specified in section 4.2.2. In all cases, the adaptive SONO starts from the "existing configuration" of the fixed mode.

4.2.1. SONO Adjustment With Radar Rainfall Observations

[35] Radar rainfall observations from two storm events were used to adjust SONO (see Figure 4 for the storm locations). Storm A spanned from 1700 to 2200 UTC on 8 July 1999, and storm B spanned from 1400 to 2345 UTC on 9 July 1999. The statistical comparison between UAGPI, exponential function mapping, fixed SONO, and adaptive SONO for the two storms, respectively, are given in Table 4. The results show that both the fixed and adaptive SONO perform significantly better than the other two algorithms. In comparison with the fixed SONO, the adaptive SONO improves the CORR to approximately 0.9, proving the worth of model adjustment. The adaptive SONO demonstrates the potential to capture precipitation variation at very short-term scales.

### 4.2.2. SONO Adjustment Using TMI Rainfall Estimates

[36] Two swaths of TMI rainfall estimates and their collocated IR images are used as the retraining data in the adaptive mode. The locations are shown in Figure 10, and the results are listed in Tables 5 and 6. Both cases demonstrate high correlation and low bias between TMI and radar data. Compared with radar and TMI data, SONO gives an impressive performance with averaged CORR 0.5, CSI 0.55, POD (probability of detection) 0.70, and FAR (false alarm ratio) 0.3 at a spatial scale $0.04°$, while the performance improved along with a decrease in spatial scales.

5. Conclusion and Future Work

[37] From the beginning of this analysis, it was anticipated that different cloud patches would yield various rain intensities and rainfall amounts. In particular, the relationship between cloud top temperature and surface rain rate varies significantly from cloud patch to cloud patch. It was

Table 4. Statistical Comparison of UAGPI, Exponential Curve, Fixed SONO, and Adaptive SONO Against WSR-88D Rainfall for Storms A and B

	Storm A				Storm B			
	UAGPI	Exponential	Fixed	Adaptive	UAGPI	Exponential	Fixed	Adaptive
RMSE	1.57	1.16	0.42	0.17	2.99	1.92	1.49	0.47
CORR	0.23	0.59	0.77	0.91	0.25	0.47	0.71	0.93
BIAS	−0.89	0.68	0.24	0.13	−1.73	−1.09	−0.49	0.23
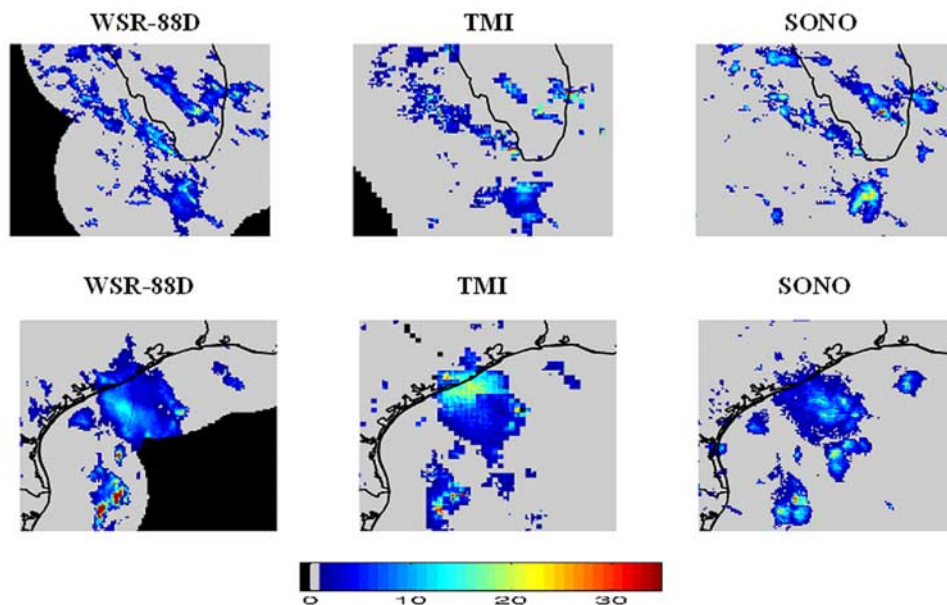
Figure 10. Comparison of instantaneous rainfall derived from adaptive SONO versus WSR-88D radar and TMI 2A12 rain rate estimates at (top) 0145 UTC, 9 July 2002, Florida, and (bottom) 0445 UTC, 16 July 2002, Texas.

also believed that precipitation estimation at small scales is difficult to achieve without the development of a cloud patch–based algorithm that identifies different IR-RR functions for the variable relationships between cloud patches and their corresponding precipitation releases. Thus how to segment cloud IR images, retrieve cloud features, classify cloud patches into groups, and calibrate different nonlinear functions for different groups of clouds was given utmost consideration throughout this analysis, which also sets this study apart from others.

[38] In this work, SONO, a cloud patch–based model for rainfall estimation at small scales, has been presented. Instead of calibrating only one statistical IR-RR function for all clouds, SONO classifies varied cloud patches into different clusters and then searches a nonlinear IR-RR mapping function for each cluster to account for the high variability of cloud-rainfall processes at geostationary scales (4 km and every 30 min). This designed feature enables SONO to generate various rain rates at a given brightness temperature and variable rain/no-rain IR thresholds for different cloud types, which overcomes the one-to-one

mapping limitation of a single IR-RR function for the full spectrum of cloud-rainfall conditions. In addition, the computational and modeling strengths of neural networks enable SONO to cope with the nonlinearity of cloud-rainfall relationships by fusing multisource data sets [*Tapiador et al.*, 2002]. Therefore SONO contains a great deal of cloud-rainfall mapping variability and shows improvements of estimation accuracy, both in rain intensity and in detection of rain/no-rain pixels. Validation also indicates that SONO has the potential to provide rainfall estimates at 3-hour 0.12° grid resolution with a relatively high correlation coefficient (~0.60). However, low correlation (0.2~0.5) exists at pixel scale 0.04°, which might partially result from the displacement of IR images from the surface rainfall due to the high altitude of geostationary satellite. In summary, SONO possesses the capabilities that (1) segment IR imagery into distinct cloud patches and extract feature information with an automatic preprocessing procedure; (2) function not only as an "analyzer" to classify large amounts of cloud patches into clusters but also as an "approximator" to calibrate different nonlinear cloud-precipitation relation-

Table 5. Statistical Comparison of Adaptive SONO Estimates Versus Radar and TMI Data at Region 22°–26°N, 79°–83°W for Case I: 0145 UTC, 9 July 2002 at Four Spatial Scales[a]

Spatial Scales	Data	CORR	RMSE	Ratio	POD	FAR	CSI
0.04° × 0.04°	SONO[R]	0.54	2.69	0.89	0.68	0.31	0.46
0.12° × 0.12°	SONO[R]	0.62	2.11	0.88	0.74	0.16	0.58
0.12° × 0.12°	SONO[T]	0.69	1.94	0.91	0.75	0.07	0.70
0.12° × 0.12°	TMI[R]	0.75	1.56	0.97	0.72	0.11	0.68
0.5° × 0.5°	SONO[R]	0.77	0.71	0.80	0.74	0.12	0.67
0.5° × 0.5°	SONO[T]	0.84	0.68	0.82	0.79	0.05	0.78
0.5° × 0.5°	TMI[R]	0.93	0.49	0.98	0.78	0.06	0.73
1.0° × 1.0°	SONO[R]	0.85	0.49	0.87	0.86	0.0	0.85
1.0° × 1.0°	SONO[T]	0.86	0.42	0.89	0.95	0.0	1.0
1.0° × 1.0°	TMI[R]	0.96	0.19	0.98	0.91	0.0	0.86

[a]SONO[R] indicates the SONO estimates compared with radar, and SONO[T] indicates those compared with the TRMM TMI rain rate. TMI[R] is the comparison between TMI and radar. The mean radar hourly rain rate at 0.12° × 0.12° is 0.92 mm hr^{-1}, and TMI is 0.90 mm hr^{-1}.

Table 6. Statistical Comparison of SONO Estimates Versus Radar and TMI Data at the Region 26°–30°N, 93°–97°W for Case II: 0445 UTC, 16 July 2002 at Different Spatial Scales[a]

Spatial Scales	Data	CORR	RMSE	Ratio	POD	FAR	CSI
0.04° × 0.04°	SONO[R]	0.46	4.01	0.60	0.78	0.26	0.62
0.12° × 0.12°	SONO[R]	0.51	3.48	0.62	0.80	0.18	0.67
0.12° × 0.12°	SONO[T]	0.57	3.20	0.62	0.87	0.12	0.76
0.12° × 0.12°	TMI[R]	0.60	3.60	1.22	0.84	0.10	0.78
0.5° × 0.5°	SONO[R]	0.60	2.92	0.59	0.84	0.17	0.71
0.5° × 0.5°	SONO[T]	0.66	2.42	0.59	0.90	0.10	0.80
0.5° × 0.5°	TMI[R]	0.71	2.84	1.24	0.88	0.03	0.88
1.0° × 1.0°	SONO[R]	0.71	2.47	0.62	1.0	0.18	0.82
1.0° × 1.0°	SONO[T]	0.79	1.37	0.72	1.0	0.09	0.90
1.0° × 1.0°	TMI[R]	0.84	1.90	1.41	0.93	0.0	0.91

[a]At 0.12° pixel scale, radar mean is 1.90 mm hr^{-1}, and the TMI mean is 1.82 mm hr^{-1}; the number of points is 1044.

ships for each cloud cluster; (3) select the best approximated IR-RR function to capture the high variation of cloud-precipitation systems at small scale; (4) update the parameter of IR-RR functions to improve the transferability of SONO for extended seasons and regions; and (5) combine the strengths of multiple sensors such as IR, radar, and passive microwave and process the information in a cost-effective manner.

[39] The aim of future work is to develop SONO as an operational satellite-based system to estimate rainfall at small scale. Given the more accurate rainfall measurements of polar-orbiting imagers, such as TRMM (TMI and Precipitation Radar), Special Sensor Microwave/Imager, and Aqua microwave imager, a microwave assimilation procedure of SONO can be implemented in real time to improve the quality of rainfall estimates on the basis of near-continuous global IR satellite imagery. More thorough evaluation of SONO must await the development of such system.

Appendix A: Formulas of Cloud Feature Extraction

[40] Given a cloud patch with a size of A, a total pixel count of N, and pixel brightness temperature of $g(x, y)$, those patch features listed in Table 1 are described as follows. (1) Minimum temperature of a cloud patch ($IR_{min:}$),

$$IR_{min} = \min_{(x,y) \in A_k} \{g(x,y)\}.$$

(2) Mean temperature of a cloud patch (IR_{mean}),

$$IR_{mean} = \sum_{(x,y) \in A_k} \{g(x,y)\}/N.$$

(3) Cloud patch size (*size*),

$$Size = N * \text{pixel resolution}.$$

(4) Shape index (*SI*) is defined as the ratio of the geometric inertia momentum of a cloud patch (I) to that of a round patch with a same size (I_o):

$$SI = I/I_0,$$

where $I = \sum_{i=1}^{N} [(x_i - G_x)^2 + (y_i - G_y)^2]$, (x_i, y_i) is the coordinate of pixel I, (G_x, G_y) is the geometric center of the cloud patch, and I_0 is the geometric inertia of a round patch with the same area. (5) Standard deviation of cloud patch temperature (STD),

$$STD = \left\{ \sum_{(x,y) \in A_k} [g(x,y) - IR_{mean}]^2/(N-1) \right\}^{1/2}.$$

(6) Mean value of the local standard deviation of cloud temperature [$MSTD(5 \times 5)$],

$$MSTD(5 \times 5) = \sum_{i=1}^{N} STD(5 \times 5)_i/N$$

where $STD(5 \times 5)_i$ is the standard deviation of brightness temperature with a window size of 5 x 5 centered on pixel i. (7) Standard deviation of local standard deviation of cloud temperature [$STD_{std}(5 \times 5)$],

$$STD_{std}(5 \times 5) = \left[\sum_{i}^{N} \left(STD(5 \times 5)_i - MSTD(5 \times 5) \right)/(N-1) \right]^{1/2}.$$

(8) Gradient of cloud shooting top temperature (*TOPG*) measures the cloud patch average pixel temperature gradient around the coldest core pixel (IR_{min}). Patch pixels on a warmer threshold of 15°K above the coldest core pixel are selected. *TOPG* is calculated from the average temperature gradient from the coldest pixel to the surrounding pixels having a temperature of IR_{min} + 15°K:

$$TOPG = \sum_{i=1}^{N} \frac{15}{\mu(i)}/N,$$

where N is the number of pixels along the border of IR_{min} + 15°K, and $\mu(i)$ is the distance from the border pixel i to the coldest pixel.

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. Meteorol.*, *27*, 30–51.

Adler, R. F., G. F. Huffman, and P. R. Keehn (1994), Global tropical rain estimates from microwave-adjusted geosynchronous data, *Remote Sens. Rev.*, *11*, 125–152.

Adler, R. F., et al. (2003), The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979-present), *J. Hydrometeorol.*, *4*(6), 1147–1167.

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, *Mon. Weather Rev.*, *115*, 51–74.

Atlas, D., and T. L. Bell (1992), The relation of radar to cloud area-time integrals and implications for rain measurements from space, *Mon. Weather Rev.*, *120*, 1997–2008.

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

Ebert, E. E. (1996), Results of the 3rd Algorithm Intercomparison Project (AIP-3) of the Global Precipitation Climatology Project (GPCP), *Rep. 55*, Bur. of Meteorol. Cent., Melbourne, Vic., Australia.

Ebert, E. E., and M. J. Manton (1998), Performance of satellite rainfall estimation algorithms during TOGA COARE, *J. Atmos. Sci.*, *55*, 1537–1557.

Gagin, A., D. Rosenfeld, and R. E. Lopez (1985), The relationship between height and precipitation characteristics of summertime convective cells in south Florida, *J. Atmos. Sci.*, *42*, 84–94.

Goodman, B., D. W. Martin, W. P. Menzel, and E. C. Cutrim (1994), A non-linear algorithm for estimating 3-hourly rain rates over Amazon from GOES/VISSR observations, *Remote Sens. Rev.*, *10*, 169–177.

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. Weather Rev.*, *106*, 1153–1171.

Hong, Y., K. Hsu, X. Gao, and S. Sorooshian (2004), Precipitation estimation from remotely sensed imagery using artificial neural network—Cloud classification system, *J. Appl. Meteorol.*, *43*(12), 1834–1853.

Hsu, K., X. Gao, S. Sorooshian, and H. V. Gupta (1997), Precipitation estimation from remotely sensed information using artificial neural networks, *J. Appl. Meteorol.*, *36*, 1176–1190.

Hsu, K., H. V. Gupta, X. Gao, and S. Sorooshian (1999), Estimation of physical variables from multichannel remotely sensed imagery using a neural network: Application to rainfall estimation, *Water Resour. Res.*, *35*(5), 1605–1618.

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, doi:10.1029/2001WR000795.

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

Kidd, C., D. R. Kniveton, M. C. Todd, and T. J. Bellerby (2003), Satellite rainfall estimation using combined passive microwave and infrared algorithms, *J. Hydrometeorol.*, *4*(6), 1088–1104.

Kohonen, T. (2001), *Self-Organizing Maps*, *Springer Ser. Inf. Sci.*, vol. 30, 3rd ed., 501 pp., Springer, New York.

Maddox, R. A., J. Zhang, J. J. Gourley, and K. W. Howard (2002), Weather radar coverage over the contiguous United States, *Weather Forecast.*, *17*(4), 927–934.

Martin, D. W., B. Goodman, T. J. Schmit, and E. C. Cutrim (1990), Estimates of daily rainfall over the Amazon basin, *J. Geophys. Res.*, *95*(D10), 17,043–17,050.

Miller, S. W., P. A. Arkin, and R. Joyce (2001), A combined microwave/infrared rain rate algorithm, *Int. J. Remote Sens.*, *22*(17), 3285–3307.

Sorooshian, S. (2004), Commentary—GEWEX (Global Energy and Water cycle Experiment) at the 2 004 Joint Scientific Committee Meeting, *GEWEX News*, *14*(2), 2.

Sorooshian, S., K. Hsu, X. Gao, H. V. Gupta, B. Imam, and D. Braithwaite (2000), Evaluation of PERSIANN system satellite-based estimates of tropical rainfall, *Bull. Am. Meteorol. Soc.*, *81*, 2035–2046.

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, paper presented at 1st Workshop of International Precipitation Working Group, Madrid, Spain, 23–27 Sept.

Vicente, G. A., R. A. Scofield, and W. P. Menzel (1998), The operational GOES infrared rainfall estimation technique, *Bull. Am. Meteorol. Soc.*, *79*, 1883–1898.

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. Meteorol.*, *19*, 338–408.

Xu, L., X. Gao, S. Sorooshian, P. A. Arkin, and B. Imam (1999a), A microwave infrared threshold technique to improve the GOES precipitation index, *J. Appl. Meteorol.*, *38*, 569–579.

Xu, L., S. Sorooshian, X. Gao, and H. Gupta (1999b), A cloud-patch technique for identification and removal of no-rain clouds from satellite infrared imagery, *J. Appl. Meteorol.*, *38*, 1170–1181.

————————————

X. Gao, Y. Hong, K. Hsu, and S. Sorooshian, Center for Hydrometeorology and Remote Sensing, Department of Civil and Environmental Engineering, University of California, Irvine, CA 92612, USA. (yanghong.uci.edu; soroosh@uci.edu)