P8.4 VALIDATION OF HYDRO-ESTIMATOR AND NEXRAD OVER PUERTO RICO

Nazario D. Ramirez-Beltran[1*], Robert J. Kuligowski[2], Eric Harmsen[1], Sandra Cruz-Pol[1], Joan M. Castro[1], and Israel Matos[3]

University of Puerto Rico, Mayaguez Puerto Rico[1]
NOAA/NESDIS Center for Satellite Applications and Research (STAR)[2]
National Weather Service, San Juan, Puerto Rico[3]

1. INTRODUCTION

Estimation of rainfall amounts is critical for protecting human lives and infrastructure, particularly in the case of heavy rainfall that triggers flash floods or landslides. In Puerto Rico (PR) during the last 5 years, five severe storms seriously impacted human lives and the economy. PR has extremely diverse terrain, and during the rainy season severe rainstorms can develop due to complex orographic attributes. Easterly winds come from the eastern Atlantic almost all year and play an important role bringing humidity into the island and stimulating orographic rainfall over the mountains of PR. Cold fronts dominate the weather pattern during wintertime. Tropical waves occur during the rainy season and frequently generate large amounts of rainfall in the Caribbean basin. These tropical waves are typically the precursor of tropical storms and hurricanes from June to November.

For these types of events, estimates of rainfall from instruments on geostationary platforms such as the Geostationary Operational Environmental Satellites (GOES) are preferred over microwave-based estimates of rainfall from low-Earth-orbiting platforms because of the rapid refresh (every 15 minutes over the CONUS and nearby regions) and very short data latency times of GOES data relative to low-Earth orbit data. Numerous algorithms have been developed to estimate precipitation from GOES-based satellite data. The current generation of algorithms produced at NESDIS are the Hydro-Estimator (HE; Scofield and Kuligowski 2003), GOES Multi-Spectral Rainfall Algorithm (GMSRA; Ba and Gruber 2001), and the Self-Calibrating Multivariate Precipitation Retrieval (SCaMPR; Kuligowski, 2002). The HE relies on GOES data from the infrared (IR) window channel with a fixed relationship to rainfall rates; similarly, Palmeira et al. (2004) presented a self-consistent algorithm for rainfall estimation based on GOES data plus lightning data in Brazil. The GMSRA uses additional data from three other GOES channels and updates its calibration in real time based on matches with radar rain rates. SCaMPR calibrates GOES IR parameters against passive microwave rain rates, which is an approach similar to Kidd et al. (2003) and the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Network (PERSIANN; Sorooshian et al. 2000) algorithm. Another algorithm called the CPC Morphing Algorithm (CMORPH; Joyce et al. 2004) also combines IR data and microwave rain rates, but uses the IR data as the basis for interpolating the microwave rain rates in time between low-Earth orbit satellite overpasses.

*Corresponding author address: Nazario D. Ramirez, University of Puerto Rico Department of Industrial Engineering, Mayaguez PR 00680 e-mail: Nazario@ece.uprm.edu

The HE, which will be the focus of this paper, also uses information from numerical whether prediction models to estimate rain rate (Scofield and Kuligowski 2003). Rainfall rates are adjusted upward or downward for moist or dry environments as indicated by National Centers for Environmental Prediction (NCEP) North American Model (NAM) or Global Forecast System (GFS) total column precipitable water and mean-layer relative humidity for the lowest third of the model vertical domain. Another adjustment enhances rainfall rates in regions where the convective equilibrium level temperature is relatively high; i.e., regions where very cold cloud tops are not thermodynamically possible but where strong updrafts and heavy rainfall can still occur. Finally, low-level winds and digital topography are combined to produce enhancements of rainfall rates in upslope regions and reductions in downslope regions, using a technique described in Vicente et al. (2001).

The HE has been the operational satellite rainfall algorithm of the National Environmental Satellite, Data, and Information Service (NESDIS) since 2000 and produces rainfall estimates at the full spatial and temporal resolution of GOES over the CONUS and surrounding regions, including

PR. However, validation of the Hydro-Estimator has generally focused on the CONUS (e.g., Kuligowski and Scofield 2003; Ebert et al. 2007) and has not been performed over Puerto Rico, and given the differences in topography and climate of Puerto Rico relative to the CONUS, previous validation efforts may not necessarily be informative to users in PR. Furthermore, validation of the HE over PR may illuminate opportunities to enhance the algorithm for application over PR.

Validation of the rainfall retrieval algorithm consists of comparing the rainfall estimates with observations (rain gauges in this study) over the same time and space. The accuracy of rainfall estimates can be measured by decomposing the rainfall process as sequences of discrete and continous random variables; i.e., the presence or absence of rainfall events (discrete variable) and the amount of rainfall (continuous variable). The occurrence of rainfall events in a given area and at a particular time follows a Bernoulli process and consequently the estimation accuracy of rainfall events can be conducted by analyzing the contingency table. The typical scores that measure the accuracy of categorical forecasts are: hit rate (H), probability of detection (POD), false-alarm rate (FAR), and discrete bias (DB). The continuous validation strategy consists of comparing the amount of rainfall that occurred at specific area in a particular time and the continuous measurements of accuracy are: mean absolute error (MAE), root mean squared error (RMSE), and continuous bias (CB).

The second section of this paper describes the data collection process and source of information. The third section describes the conventional statistical techniques to perform validation. The fourth section presents validation results during heavy storms over PR, and includes a comparison for rain gauges versus HE and rain gauges versus NEXRAD. The fifth section presents some conclusions.

2. DATA COLLECTION

Puerto Rico has a rain gauge network that collects rainfall measurements every 15 minutes and includes 125 rain gauges with data available since January 2000. The rain gauge data are used to perform validation of the HE and the NEXRAD. Since this an ongoing project we present a preliminary data set used for validation and include only one storm that seriously impacted PR. In a near future, a more complete validation will be published.

NEXRAD data over Puerto Rico comes from a WSR-88D unit located in Cayey (18.12^{o}N, 66.08^{o}W, 886.63 m elevation). The radar frequency is 2.7 GHz and the maximum horizontal coverage is 462.5 km, and the radar scans the entire island every 6 minutes. The National Oceanic and Atmospheric Administration (NOAA) National Server Storms Laboratory (NSSL) conducted a significant effort to make possible an affordable nationwide operational capture, distribution, and archiving of Level II NEXRAD data (Kelleher et al. 2007). Unfortunately, for Puerto Rico the Level II data are available only until 2002 (NCDC, 2005a). The NWS did resume archiving level II data for PR during the summer of 2007. On the other hand, Level III data for PR are available continuously since 2000 (NCDC, 2005b), so the Level III data were selected to perform validation since the most recent and catastrophic floods over PR occurred after 2002. The scanning angle for reflectivity data was selected as 0.5 degrees for this research in order to avoid beam overshoot over western PR. Figure 1 shows the location of the radar and the spatial distribution of rain gauges, the black dot indicates the location of the NEXRAD and the red starts show the location of the rain gauges.

As mentioned in the Introduction, the HE uses satellite IR window (10.7-µm) data and numerical whether prediction data to estimate rainfall over the CONUS and PR every 15 minutes at 4 km spatial resolution, and they are available for the entire period of interest. In order to ensure consistency among these data sets during the comparison, both the NEXRAD and HE rain rates were aggregated in time over the corresponding 15-minute accumulation period of the gauges.

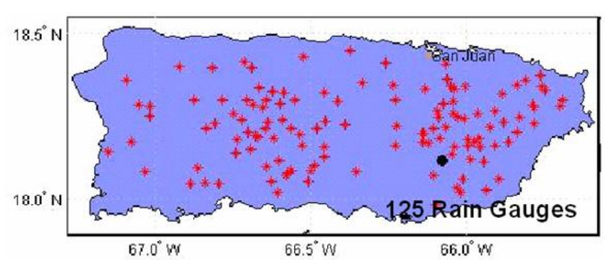

FIG. 1. Location of rain gauges (red stars) and NEXRAD (black dot) in PR.

3. VALIDATION TECHNIQUES

Validation of the rainfall retrieval algorithm consists of comparing the rainfall estimates with observations over the same time and space. The accuracy of rainfall estimates can be measured by decomposing the rainfall process into sequences of discrete and continuous random variables, i.e., the presence or absence of rainfall events and the amounts of rainfall. The occurrence of rainfall events in a given area and at a particular time follows a Bernoulli process and consequently the estimation accuracy of rainfall events can be conducted by analyzing contingency tables (Wilks 1995). Table 1 shows the classical two-way contingency table.

TABLE 1. Sample contingency table.

		Observed rainfall (Rain gauge)	
		Yes	*No*
Estimated rainfall (HE or NEXRAD)	*Yes*	*a*	*b*
	No	*c*	*d*

It is assumed that the values provided by the rain gauges are the “ground truth” while the HE and the NEXRAD provide estimated rainfall values. The variable *a* in the contingency table is the number of times that the rain gauge identifies a rainfall event and the estimator also correctly identifies a rainfall event at the same time and space. The variable *d* represents the number of times the rain gauge does not observe a rainfall event and the estimator correctly determines that there is no rainfall event. The variable *b* indicates the number of times the rain gauge does not observe a rainfall event but the estimator incorrectly indicates that there is a rainfall event. The variable *c* shows the number of times that the rain gauge detects a rainfall event but the estimator incorrectly does not detect the rainfall event. The typical scores that measure the accuracy of categorical estimation are:

$$H = \frac{a+d}{n_o}, \quad \text{where} \quad n_o = a+b+c+d \tag{1}$$

$$POD = \frac{a}{a+c} \tag{2}$$

$$FAR = \frac{b}{a+b} \tag{3}$$

$$DB = \frac{a+b}{a+c} \tag{4}$$

where *H* is the hit rate, *POD* is the probability of detection, *FAR* is the false-alarm rate, and *DB* is the discrete bias. Hit rate is the fraction of the n_o estimating occasions when the categorical estimation correctly determines the occurrence of rainfall event or nonevent. Probability of detection is the likelihood that the event would be estimated, given that it occurred. The false-alarm rate is the proportion of estimated rainfall events that fail to materialize. Bias is the ratio of the number of estimated rainfall events to the number of observed events (Wilks 1995).

The continuous validation strategy consists of comparing the amount of rainfall that occurred with the estimated amount of rainfall at specific area in a particular time and the continuous accuracy scores are:

$$e_{ij} = y_{ij} - \hat{y}_{ij} \quad i = 1,\ldots,n \quad \text{and} \quad j = 1,\ldots,m \tag{5}$$

$$MAE = \frac{1}{n\ m}\sum_{i=1}^{n}\sum_{j=1}^{m}\left|e_{ij}\right| \tag{6}$$

$$RMSE = \sqrt{\frac{1}{n\ m}\sum_{i=1}^{n}\sum_{j-1}^{m}e_i^2} \tag{7}$$

$$CB = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\hat{y}_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{m}y_{ij}} \tag{8}$$

where *y* and *ŷ* are the observed and estimated amount of rainfall. The *i* and *j* subscripts represent time and space, respectively. The constant *n* is the total number of time intervals for a given storm, and m is the number of rain gauges that are collecting rain during a storm. The error *e* is the deviation between the observed and estimated amount of rainfall at a particular time and space and was computed only when at least one of *y* or *ŷ* is greater than zero. *MAE* is the mean absolute error, *RMSE* the root mean squared error, and *CB* is the continuous bias.

4. VALIDATION RESULTS

4.1. Discrete validation

A contingency table was computed for each rain gauge during the storm and the scores of those tables were summarized to create contingency tables for the HE and NEXRAD which are shown in Table 2 while the associated scores are shown in Table 3. The bias of the HE was 0.50, indicating that the HE heavily underestimates the number of times that rainfall events have occurred. Meanwhile, the NEXRAD had a bias of 1.07, which is quite close to the ideal value of unity. The hit rates of the HE and NEXRAD were 0.70 and 0.81 respectively, indicating that the HE has a lower percentage of correct rain / no rain estimates than does the NEXRAD. As would be expected given the strong dry bias of the HE, the HE correctly detected a much smaller percentage of the observed rainfall events (30%) than did NEXRAD (80%) for these events. Surprisingly, even though the HE indicated rainfall less than half the time that NEXRAD did, the false alarm rate was actually higher for the HE (39%) than for NEXRAD (32%).

TABLE 2. Contingency tables for (a) the Hydro-Estimator and (b) NEXRAD.

a)

		Rain Gauge	
		Yes	No
Hydro-Estimator	Yes	1105	699
	No	2603	6708

b)

		Rain Gauge	
		Yes	No
NEXRAD	Yes	2986	1392
	No	730	6124

TABLE 3. Discrete validation scores for the Hydro-Estimator.

	Hydro-Estimator	*NEXRAD*
Discrete Bias	*0.50*	*1.07*
Hit Rate	*0.70*	*0.81*
Probability of Detection	*0.30*	*0.80*
False Alarm Rate	*0.39*	*0.32*

4.2. Continuous validation

The accumulated rainfall across the island was computed to perform comparison between the observed and the estimated rainfall:

$$Y_i = \sum_{j=1}^{m} y_{ij} \quad i = 1, 2, \cdots, n \tag{9}$$

where Y_i is the total rainfall recorded by all 125 rain gauges across the island or the closest HE or radar pixels at the i^{th} time. Figure 2 shows the time series of comparisons of the HE and NEXRAD against the corresponding gauge rainfall during the storm that occurred in PR on 17 April 2003. Figure 3 contains scatterplots comparing the observed and estimated 15-minute accumulations of rainfall for all of the gauges in the network for both the H-E and NEXRAD, and the underestimation by the HE is clear, though it is not necessarily apparent whether the underestimation is simply because of the detection problems of the HE or if it is also the result of underestimation of rainfall amounts by the HE. A comparison of the values for individual gauges in makes it clear that the HE is underestimating in terms of amounts as well as detection.

This is also reflected in the continuous validation scores in Table 4. The continuous bias of the HE is even stronger than the discrete bias, with the HE estimating only 26% of the total observed rainfall. By comparison, the radar data are nearly without bias. As a result, the mean absolute error of the HE is also higher than that of NEXRAD by roughly a third. However, the RMSE is surprisingly lower for the HE than for the NEXRAD. This may be because of the tendency of the latter statistic to penalize larger errors more than smaller ones, and thus more penalize the NEXRAD for the occasional overestimation of light rainfall that does not occur in the HE (Fig. 4).

TABLE 4. Continuous validation scores for the Hydro-Estimator.

	Hydro -Estimator	*NEX RAD*
Continuous Bias	*0.26*	*1.01*
Mean Absolute Error (mm)	*1.33*	*0.99*
Root Mean Square Error (mm)	*2.73*	*3.76*

Figure 5 compares the spatial distribution of rain gauges, the HE, and NEXRAD for 14:30 UTC 17 April 2003 to illustrate the differences among these fields. Again, the HE shows a significant underestimation of the magnitude of rainfall, especially over eastern PR. In this particular case the NEXRAD also fails to capture the heaviest rainfall over eastern PR, but to a much lesser degree than the HE. There are also significant differences in the shape of the rainfall area depicted over south central PR by the HE compared to the gauges and NEXRAD, with the HE showing light rain much farther to the north and west than the radar and gauges. This presumably is due to the presence of relatively cold but non-raining cirrus clouds.

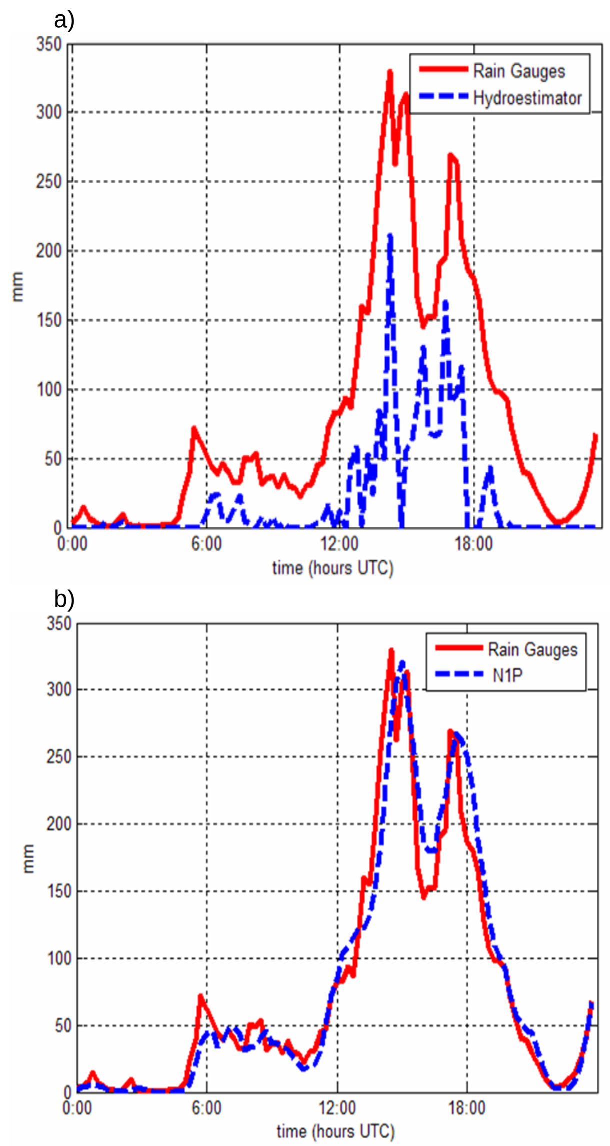

FIG. 2. Time series of accumulated 15-minute rainfall for all 125 gauges across PR along with the corresponding (a) HE and (b) NEXRAD pixels for 17 April 2003.

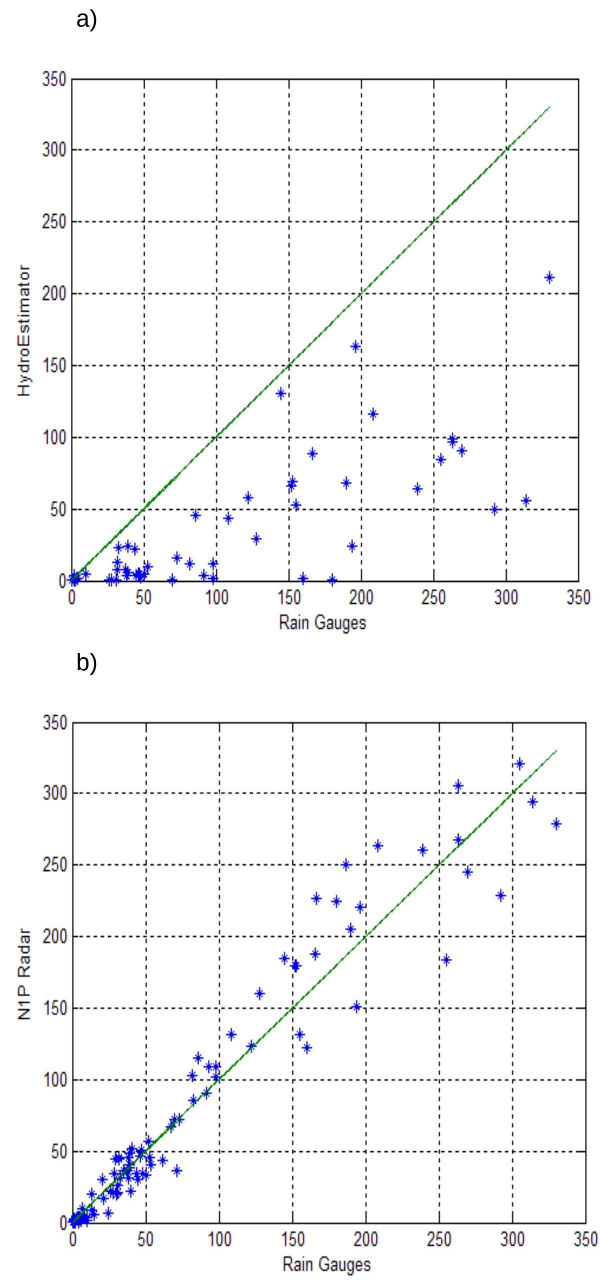

FIG. 3. Comparison of observed total rainfall for all 125 gauges across PR with the corresponding (a) HE and (b) NEXRAD pixels for 17 April 2003.

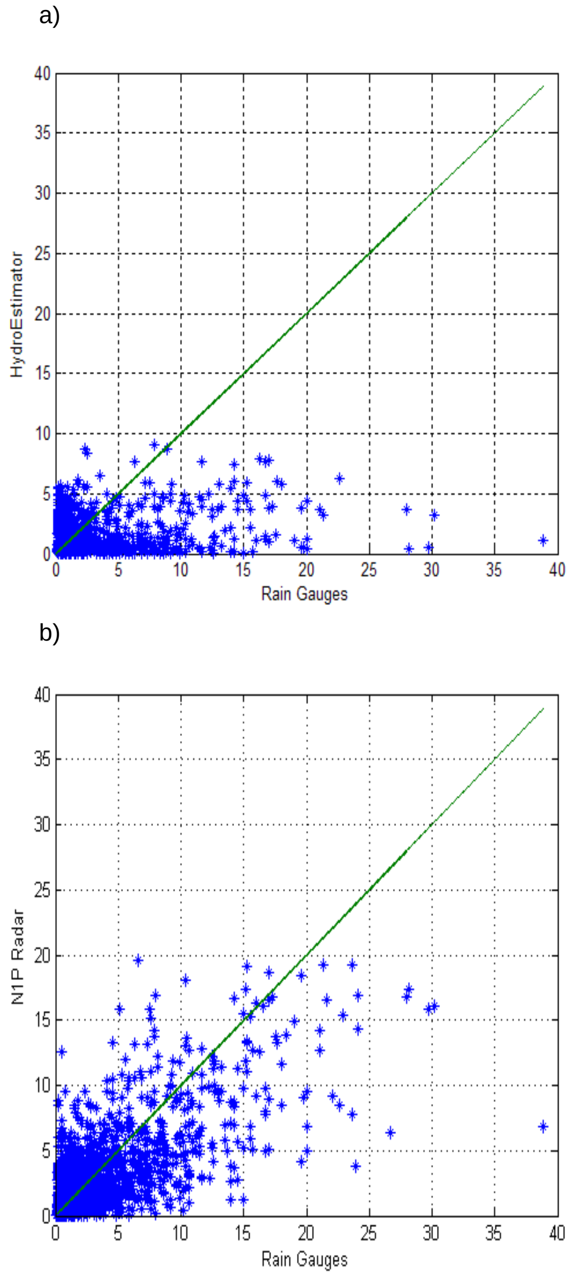

FIG. 4. Comparison of observed rainfall for each gauge in PR with the corresponding (a) HE and (b) NEXRAD pixels for 17 April 2003.

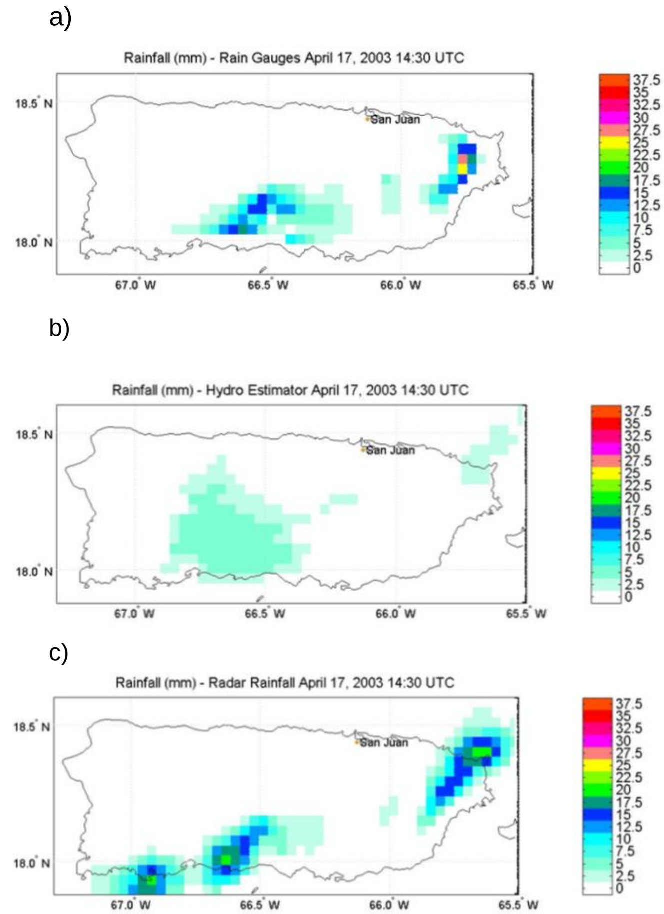

FIG. 5. Comparison of 15-minute rainfall accumulations ending 1430 UTC 17 April 2003 for (a) gauges; (b) the HE; and (c) NEXRAD.

5. CONCLUSIONS

The HE is a high resolution satellite rainfall retrieval algorithm run operationally by NESDIS that provides estimates of rainfall every 15 minutes at 4-km resolution over the CONUS and nearby areas including PR. The rain rates are primarily derived from GOES 10.7-µm brightness temperatures and then adjusted using parameters derived from a numerical weather prediction model. The HE estimator should be especially useful over regions of complex topography such as western PR because of the difficulties associate with radar in those regions such as beam block. However, for the heavy rainfall event examined in this paper, NEXRAD clearly outperformed the HE. This may be in part because of most of the rainfall occurred in the central and eastern parts of the island where the radar data would be most reliable. Specifically, the HE underestimated both the spatial extent and the amount of rainfall, whereas NEXRAD was nearly unbiased in these respects. The HE algorithm did exhibit a satisfactory hit rate, but a very low probability of detection and a false alarm rate that was surprisingly higher than that of NEXRAD despite the dry bias of the HE. A research effort is underway to improve the performance of the HE for PR; specifically, the algorithm proposed by Ramirez-Beltran at al. (2007) will be implemented to improve the HE rainfall detection and the equation that relates brightness temperatures with rain rates.

6. ACKNOWLEDGMENTS

This research has been supported by NOAA-CREST grant NA17AE1625, NOAA/NWS grant number NA06NWS468001, the NSF with a grant Number 0313747, and also by the University of Puerto Rico at Mayagüez. The authors appreciate and recognize the funding support from these institutions.

7. REFERENCES

Ba, M. B., and A. Gruber, 2000: GOES Multispectral Rainfall Algorithm (GMSRA). *J. Appl. Meteor.*, **40**, 1500-1514.

Ebert, E. E., J. E. Janowiak, and C. Kidd, 2007: Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bull. Amer. Meteor. Soc*., **88**, 47-64.

Joyce, R.J., J. E. Janowiak, P.A. Arkin, and P. Xie, 2004: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeor*., **5**, 487-503.

Kelleher, K.E., K. K. Droegemeier, J. J. Levit, C. Sinclair, D. E. Jahn, S. D. Hill, L. Mueller, G. Qualley, T. D. Crum, S. D. Smith, S. A. Del Greco, S. Lakshmivarahan, L. Miller, M. Ramamurthy, B. Domenic, and D. W. Fulker, 2007: A real-time delivery system for NEXRAD Level II data via the internet. *Bull. Amer. Meteor. Soc.*, **88**, 1045-1057.

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

Kuligowski, R.J., 2001): A Self-Calibrating real-time GOES rainfall algorithm for short-term rainfall estimates. *J. Hydrometeor.*, **3**, 112-130.

NCDC, 2005a: National Climatic Data Center: Data Documentation for DSI – 6500 – *NEXRAD Level II.* NCDC, Asheville, NC.

-----, 2005b: National Climatic Data Center: Data Documentation for DSI – 7000 – *NEXRAD Level III*. NCDC, Asheville, NC.

Palmeira, F. L. B., C. A. Morales, G. B. França, and L. Landau, 2004: Rainfall estimation using satellite data for Paraíba Do Sul Basin (Brazil). *The XX^{th} ISPRS International Society for Photogrammetry and Remote Sensing Congress*, Istanbul, Turkey - Commission 7.

Ramirez-Beltran, N.D., W.K.M. Lau, A. Winter, J.M. Castro, and N.R. Escalante, 2007: Empirical probability models to predict precipitation levels over Puerto Rico stations. *Mon. Wea. Rev*. **135**, 877-890.

Scofield, R.A., and R.J. Kuligowski, 2003: Status and outlook of operational satellite precipitation algorithms for extreme-precipitation events. *Wea. Forecasting,* **18**, 1037-1051.

Sorooshian, S., K.-L. Hsu, X. Gao, H. V. Gupta, B. Imam, and D. Braithwaite, 2000: An evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bull. Amer. Meteor. Soc*., **81**, 2035-2046.

Vicente, G. A., J. C. Davenport, and R. A. Scofield, 2002: The role of orographic and parallax corrections on real time high resolution satellite rainfall rate distribution. *Int. J. Remote Sens*., Volume 23, No. 2, pp. 221-230.

Wilks, D.S., 1995**:** *Statistical Methods in the Atmospheric Sciences: An Introduction.* Academic Press, San Diego, 467 pp.