The Tropical Rainfall Potential (TRaP) Technique. Part II: Validation

RALPH FERRARO,* PAUL PELLEGRINO,+ MICHAEL TURK,* WANCHUN CHEN,+ SHUANG QIU,+ ROBERT KULIGOWSKI,* SHELDON KUSSELSON,* ANTONIO IRVING,* STAN KIDDER,# AND JOHN KNAFF#

*NOAA/NESDIS, Camp Springs, Maryland
+QSS Group, Inc., Lanham, Maryland
#CIRA, Fort Collins, Colorado

(Manuscript received 6 February 2004, in final form 10 January 2005)

ABSTRACT

Satellite analysts at the Satellite Services Division (SSD) of the National Environmental, Satellite, Data, and Information Service (NESDIS) routinely generate 24-h rainfall potential for all tropical systems that are expected to make landfall within 24 to at most 36 h and are of tropical storm or greater strength (>65 km h^{-1}). These estimates, known as the tropical rainfall potential (TRaP), are generated in an objective manner by taking instantaneous rainfall estimates from passive microwave sensors, advecting this rainfall pattern along the predicted storm track, and accumulating rainfall over the next 24 h.

In this study, the TRaPs generated by SSD during the 2002 Atlantic hurricane season have been validated using National Centers for Environmental Prediction (NCEP) stage IV hourly rainfall estimates. An objective validation package was used to generate common statistics such as correlation, bias, root-mean-square error, etc. It was found that by changing the minimum rain-rate threshold, the results could be drastically different. It was determined that a minimum threshold of 25.4 mm day^{-1} was appropriate for use with TRaP. By stratifying the data by different criteria, it was discovered that the TRaPs generated using Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) rain rates, with its optimal set of measurement frequencies, improved spatial resolution, and advanced retrieval algorithm, produced the best results. In addition, the best results were found for TRaPs generated for storms that were between 12 and 18 h from landfall. Since the TRaP is highly dependent on the forecast track of the storm, selected TRaPs were rerun using the observed track contained in the NOAA/Tropical Prediction Center (TPC) "best track." Although some TRaPs were not significantly improved by using this best track, significant improvements were realized in some instances. Finally, as a benchmark for the usefulness of TRaP, comparisons were made to Eta Model 24-h precipitation forecasts as well as three climatological maximum rainfall methods. It was apparent that the satellite-based TRaP outperforms the Eta Model in virtually every statistical category, while the climatological methods produced maximum rainfall totals closer to the stage IV maximum amounts when compared with TRaP, although these methods are for storm totals while TRaP is for a 24-h period.

1. Introduction

Damage and deaths resulting from the direct and indirect effects of rainfall associated with landfalling tropical systems in the United States exceed those that are caused by both wind and wave damage (Rappaport 2000). Current operational numerical weather prediction (NWP) forecast models are believed not to accurately predict the rainfall associated with such systems. Satellite techniques are relied upon by operational forecasters in the National Oceanic and Atmospheric Administration's (NOAA), National Centers for Environmental Prediction (NCEP), National Weather Service (NWS) Weather Forecast Offices (WFOs), and River Forecast Centers (RFCs) to get a better assessment of the rainfall potential of systems that are approaching land, in particular, those out of range of coastal radar. Such was the motivation in the development of the NOAA's National Environmental Satellite, Data, and Information Service (NESDIS), Satellite Services Division (SSD), tropical rainfall potential (TRaP) technique, which has been run in its current experimental mode for the past few years and recently

Corresponding author address: Ralph Ferraro, CICS/ESSIC, 2207 Space and Computer Sciences Building #224, University of Maryland, College Park, College Park, MD 20742.
E-mail: Ralph.R.Ferraro@noaa.gov

became operational during the 2003 Western Hemisphere hurricane season.

TRaP is described in detail in the first part (Kidder et al. 2005, hereafter Part I) of this two-part paper series. It is the purpose of this paper, Part II, to describe the validation of TRaP for the 2002 Atlantic hurricane season. This study improves upon an original validation effort for the 2001 season (Ferraro et al. 2002) where the validation methodology was standardized and made more robust. In addition, the 2002 season offered many more opportunities with several landfalling systems over the United States than were available during the 2001 season.

Section 2 of this paper presents a brief overview of the TRaP technique. Section 3 describes in detail the validation of TRaP followed by a summary and suggestions for future work in section 4.

2. TRaP overview

Part I presents a historical evolution of TRaP from its earliest roots to its present, operational status. For completeness in this paper, a brief discussion of the current NESDIS/SSD operational TRaP is presented.

The NESDIS/SSD Areal TRaP is generated automatically for any tropical disturbance worldwide whenever a new microwave rain-rate image or a new track forecast is received. Currently, SSD uses microwave rain-rate estimates from the Advanced Microwave Sounding Unit (AMSU) on the NOAA polar-orbiting satellites, from the Special Sensor Microwave Imager (SSM/I) on the Defense Meteorological Satellite Program (DMSP) satellites, and from the Microwave Imager (TMI) on the Tropical Rainfall Measuring Mission (TRMM) satellite. The latest rain-rate image and the latest tropical cyclone center track forecast is used only if they are less than 6 h from the time of the latest track forecast and the rain-rate image, respectively. NESDIS/SSD operational analysts, working around the clock (24 h day^{-1}, 7 days $week^{-1}$, 365 days yr^{-1} provide quality assurance of the automated TRaPs and send the final product to their Internet home page (http://www.ssd.noaa.gov/PS/TROP/trap-img.html) only when a storm has wind speeds greater than 65 km h^{-1} (35 kt) and is within 24 to 36 h of landfall. The quality assurance is to guarantee that the final TRaP includes a full (better than 75% coverage) rain-rate image over the storm. This quality assurance results in only a portion of all the automated TRaPs generated at NESDIS becoming operational.

3. 2002 validation

Selected for the 2002 validation study were a total of 42 operational TRaPs that were generated for five U.S. landfalling tropical cyclones. Although the automated TRaP program operationally generated many more TRaP products, this study focuses only on those storms that affected the United States, were within 24 to at most 36 h of landfall, and passed the quality assurance performed by the satellite analyst. These are summarized in Table 1. As one can see, 16 were generated from SSM/I, 11 from AMSU, and 15 from TMI. In contrast to the Ferraro et al. (2002) validation study for the 2001 hurricane season, the validation strategy for the 2002 storms was much more objective in nature because of the uniformity of the data sources available. First, NCEP stage IV analysis fields were used as the ground reference datasets (Fulton et al. 1998). These hourly rain estimates were accumulated to best match the 24-h period of the TRaP. Second, a single validation package was used that objectively computes a number of statistical parameters between the two rain fields (Ebert et al. 2003). Third, since the TRaP software was run in a quasi-operational mode for the 2002 hurricane season, all input and output data were archived, so it was simple to rerun any TRaPs for the variety of analyses that were employed in this study. As an example, Fig. 1 presents a four-panel sequence that shows the input satellite rain field, the derived TRaP, and then the output from the statistical comparison versus the stage IV hourly rain field. It should be noted that the original stage IV estimates are on a higher spatial resolution grid (~4 km grid) than the TRaP estimates (~25 km grid), but the validation is performed at the higher spatial resolution (i.e., the TRaP estimates are bilinearly interpolated to a finer grid). This "upscaling" of the TRaP was done in order to try to identify its performance at the highest rainfall values: those that can contribute to life-threatening flash flooding.

The three sources of satellite data all have different sensor configurations: footprint size [field of view (FOV)], sampling rate across the scan, scanning geometry, etc. We attempt to minimize the differences by mapping them, with bilinear interpolation, to a common 25-km grid at NOAA/NESDIS, where they are generated operationally for a number of applications,

TABLE 1. Summary of the number of TRaPs generated for each of the 2002 Atlantic tropical systems used in this analysis.

2002 storm name	Dates	SSM/I	AMSU	TMI
Bertha	4–5 Aug	3	0	1
Fay	6–7 Sep	4	0	4
Hanna	13–14 Sep	4	2	5
Isidore	25–26 Sep	1	7	1
Lili	2–3 Oct	4	2	4
Total		16	11	15

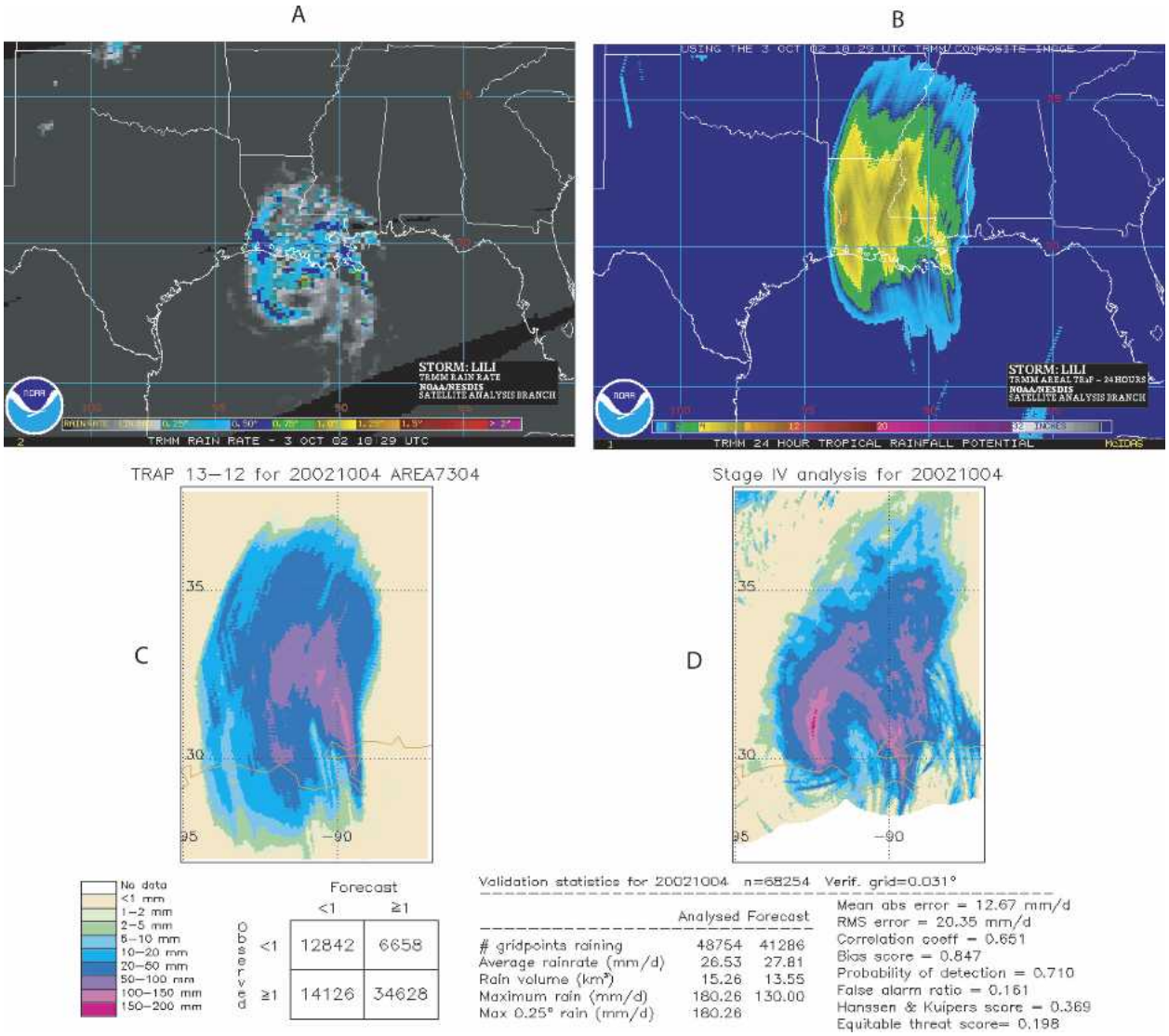

FIG. 1. (b) An example of an operational TRaP (in.) generated for Hurricane Lili using the TRMM overpass from (a) 1029 UTC on 3 Oct 2002 (in. h^{-1}). The output from the images (c) of TRaP (mm) and (d) of stage IV rainfall (mm); (c) and (d) also include the statistical parameters generated.

including TRaP. However, this does come with risks, the so-called "representativeness error" as described by Tustison et al. (2001), which can sometimes cause misleading statistical results. Since the comparisons are done for 24-h rainfall, the spatial correlation length of the rainfall for large precipitation systems like hurricanes should tend to reduce this effect. Studying the impact of these sensor characteristics would form the basis for another study and is beyond the scope of this paper. However, this concept is elaborated upon in the last section.

The statistical results of the 42 TRaPs were compiled and stratified by a number of criteria, including by storm, by the number of hours prior to landfall (i.e., 6, 12, 18, or 24 h), and by the instrument from which the input rain fields were obtained (i.e., SSM/I, AMSU, or TMI). In addition, statistics were computed using only those points with observed rainfall that exceeded a minimum threshold value. Several thresholds were used and it was determined that two such thresholds would be presented: 1 and 25.4 mm day^{-1}. The latter of the two was deemed to be more important to the application of TRaP, namely, for use in heavy rainfall prediction. The interpretation of the results from the two thresholds changes since statistical tools such as correlation are highly dependent upon the data range. Because the number of samples is relatively small, it was not practical to apply significance testing to the

TABLE 2. Statistical summary (mean value for all TRaPs for each storm) of the TRaP compared to the stage IV analysis for a 1 mm day^{-1} minimum threshold. The rain rate and rain volume (area of rain times the rain magnitude) represent the TRaP value normalized by the stage IV mean. MAE is mean absolute error, rms is root-mean-square error (normalized by the stage IV mean), *R* is correlation coefficient, POD is probability of detection, FAR is false alarm ratio, and ETS is equitable threat score.

2002 storm name	Rain rate (TRaP/ stage IV)	Rain volume (TRaP/ stage IV)	TRaP max (mm day^{-1})	Stage IV max (mm day^{-1})	MAE (mm day^{-1})	Rms	*R*	Bias score	POD	FAR	ETS
Bertha	0.33	0.32	66.8	330.1	25.2	1.16	0.22	0.63	0.49	0.22	0.10
Fay	1.28	0.88	220.3	359.1	20.6	1.26	0.54	0.69	0.57	0.16	0.15
Hanna	0.53	0.46	88.9	473.8	22.8	1.06	0.47	0.82	0.76	0.07	0.28
Isidore	0.47	0.49	87.0	340.4	29.0	0.95	0.40	1.02	0.91	0.11	0.24
Lili	0.72	0.70	92.5	225.1	17.8	0.80	0.62	0.85	0.76	0.11	0.33
Average	0.64	0.54	112.5	352.7	22.9	1.06	0.47	0.82	0.72	0.12	0.24

results. However, graphical representation of many of the results to follow is shown through the use of "box and whisker" charts (Tukey 1977) where the range of the data is presented along with the median and lower and upper quartiles, so some general conclusions can still be made.

a. 24-h rainfall threshold results

The statistics generated by the validation package include a variety of measures of fit including correlation, root-mean-square (rms) error, bias, etc. (Information regarding statistical parameters used in the validation of satellite-based rainfall estimates can be found online at http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_page.html.)

These were computed for 24-h rainfall based on TRaP and stage IV. The geographic region of validation for each TRaP was subjectively determined so that rainfall associated with other precipitation systems, not directly related to the tropical system, would be minimal. Thus, the statistics would best represent TRaP performance. Shown in Tables 2 and 3 are the results for 1 and 25.4 mm day^{-1} minimum thresholds, respectively, as a function of tropical cyclone. It should also be mentioned here that image comparison is also a validation method and that it is part of the standard package of information that is illustrated in Fig. 1. However, there is no way to quantify such results in an objective manner.

Overall, it was seen that for all storms except Fay, the TRaP underestimates both the area-wide rain rate and volume, for both thresholds. Using the 25.4 mm day^{-1} threshold generally brings the two totals in closer agreement, apparently due to the removal of lighter rain rates that the passive microwave satellite estimates may not be able to retrieve. It is also obvious that the TRaP grossly underestimates the maximum rainfall, but this is at least partly due to the differing spatial resolution of the two data types, in particular, to the coarse resolution of the passive microwave sensors used as input into TRaP (which can range from 5 to 50 km depending upon the sensor and nature of the retrieval) versus the higher resolution of the stage IV data. A more fair overall statistical comparison would have been to average the stage IV up to the 25-km TRaP grid as suggested in the study by Tustison et al. (2001); however, as previously mentioned, one of our purposes is to determine the TraP's ability to capture the extreme rain rates.

The interpretation of the other statistical parameters can vary widely between the two thresholds used, and this has always been an issue, in general, when trying to validate satellite-based rainfall estimates (e.g., Ebert et al. 1996). The correlation coefficient—perhaps the most

TABLE 3. As in Table 2, but for a 25.4 mm day^{-1} minimum threshold.

2002 storm name	Rain rate (TRaP/ stage IV)	Rain volume (TRaP/ stage IV)	TRaP max (mm day^{-1})	Stage IV max (mm day^{-1})	MAE (mm day^{-1})	Rms	*R*	Bias score	POD	FAR	ETS
Bertha	0.57	0.79	66.8	330.1	42.9	0.80	−0.07	0.33	0.21	0.25	0.12
Fay	1.37	1.33	220.3	359.1	58.0	1.00	0.29	0.67	0.46	0.29	0.28
Hanna	0.53	0.64	88.8	473.8	45.7	0.77	0.17	0.52	0.38	0.22	0.20
Isidore	0.50	0.49	87.0	340.4	41.2	0.74	0.06	0.74	0.57	0.23	0.25
Lili	0.67	0.74	92.5	225.1	32.9	0.78	0.07	0.78	0.57	0.25	0.33
Average	0.71	0.65	112.5	352.7	44.1	0.78	0.13	0.64	0.47	0.25	0.25

commonly used statistical measure of "goodness of fit" between two datasets—is much higher for the 1 mm day^{-1} threshold than for the 25.4 mm day^{-1} threshold. The explanation is quite simple: the removal of low end rainfall values that dominate the dataset (refer to Fig. 1) and usually exhibit less scatter around the perfect fit line significantly impacts the correlation. Hence, this parameter alone is probably not appropriate in determining the usefulness of TRaP at the higher rain threshold.

Other parameters such as the mean absolute error (MAE), rms, bias score (BS), probability of detection (POD), false alarm rate (FAR), and equitable threat score (ETS) are also useful. The MAE measures the mean of the absolute value of the difference between the TRaP and stage IV; the lower this value, the better the TRaP. The MAE is essentially twice as high with the 25.4 mm day^{-1} threshold as with the 1 mm day^{-1} threshold and, in general, is fairly consistent from storm to storm. The MAE is also higher at the 25.4 mm day^{-1} threshold because regions of light precipitation (which have lower errors) have been eliminated. The rms shown here is normalized by the stage IV mean in order to place some perspective on the magnitude of the error. The rms is significantly lower for all instances at the 25.4 mm day^{-1} threshold than for the 1 mm day^{-1} threshold and is approximately 80% of the observed mean rainfall.

The next sets of statistical measures indicate the frequency of occurrence of various rain forecasts. The BS, which is the ratio of the forecasts above a particular rain threshold to observations above the same threshold and can range from zero to infinity, indicates whether the TRaP has a tendency to under-forecast (BS < 1) or overforecast (BS > 1) rainfall at a particular threshold. A BS equal to one indicates an unbiased forecast. The BS is worse for the 25.4 mm day^{-1} threshold than the 1 mm day^{-1} threshold, indicating that either the number of false alarms decreases (a positive indicator) or that the number of misses increases (a negative indicator). Inspection of the images indicates that the decrease in the BS arises from the decrease in false alarms at the lower rainfall values. The POD is the ratio of the number of correct forecasts (i.e., values above the rainfall threshold) to the number of observations above the same threshold, that is, the fraction of observations above the threshold that were correctly predicted to be above that threshold. The POD ranges from zero to one, with POD = 1 indicating that all of the observations above the threshold were correctly predicted to be above that threshold. The POD should be used in conjunction with the FAR, which is the ratio of the number of false alarms (i.e., values above the rainfall threshold that did not occur) to the number of forecasts above the threshold, that is, the fraction of forecasts above the threshold that corresponded to values below the threshold. Like with POD, FAR can range from zero to one, with FAR = 0 indicating that all of the forecasts above the threshold were correct. Using these two parameters together is needed since a forecast of no rain everywhere would yield a FAR = 0, but also a POD = 0; conversely, rain forecasted everywhere would yield a POD = 1, but a FAR = 1. Thus, examination of the two together eliminates any misleading interpretation of the forecast performance. As can be seen from Tables 2 and 3, both the POD and FAR are worse as the rain threshold is increased from 1 to 25.4 mm day^{-1}. Overall, over 70% of the rainfall above 1 mm day^{-1} is detected by TRaP, while approximately 10% of the TRaP forecasts are false alarms; nearly 50% of the rainfall above 25.4 mm day^{-1} is detected and 25% is incorrectly forecasted by TRaP.

A final statistical measure that is widely used in assessing forecasting skill and rainfall validation is the ETS (Schaefer 1990). The ETS ranges in value from $-1/3$ to 1, with an ETS = 0 indicating no skill and ETS = 1 being a perfect score. The ETS is sensitive to correct forecasts and penalizes for both misses and false alarms. It also accounts for a climatological event frequency. On average, the ETS values are essentially the same for both rainfall thresholds. However, there is noticeable difference for two storms: Fay and Hanna. For Fay, the 25.4 mm day^{-1} threshold forecast was superior to the 1 mm day^{-1} forecast, while for Hanna, the opposite was true. Fay was a short-lived storm that had erratic movement while Hanna was in a strong sheer environment; these factors may have contributed to the performance of the ETS at the two rainfall thresholds.

b. Intersatellite comparisons

The first stratification of the 42 TRaPs (Table 1) that was performed was sorting the TRaPs by sensor type. Overall, there were 11 TRaPs generated from AMSU, 16 from SSM/I, and 15 from TMI. Table 4 and Fig. 2 summarize these results using a 25.4 mm day^{-1} threshold. As can be seen, the SSM/I TRaPs were generally the poorest, while on average, TMI performed better than AMSU, although the two were comparable for a number of categories. Some possible explanations for this are as follows. The SSM/I operational algorithm (Colton and Poe 1994) has been virtually unchanged for a number of years and is most recently described by Ferraro (1997). A number of deficiencies have been found in that algorithm (e.g., McCollum et al. 2002), although the algorithm is still instrumental in a number

TABLE 4. As in Table 3, but as a function of sensor.

Sensor (cases)	Rain rate (TRaP/ stage IV)	Rain volume (TRaP/ stage IV)	TRaP max (mm day^{-1})	Stage IV max (mm day^{-1})	MAE (mm day^{-1})	Rms	R	Bias score	POD	FAR	ETS
AMSU (11)	0.59	0.55	90.1	351.8	37.0	0.70	0.10	0.79	0.60	0.24	0.30
SSMI (16)	0.71	0.65	106.1	345.0	49.8	0.81	0.16	0.36	0.29	0.18	0.16
TMI (15)	0.84	0.77	140.5	355.9	42.7	0.78	0.11	0.83	0.56	0.32	0.31

of applications (e.g., Xie et al. 2003). Thus it is not surprising that it performed more poorly than TRaPs using the other two sources of microwave information.

The TRMM TMI rain algorithm [i.e., the Goddard profiling algorithm (GPROF)]—in particular, the oceanic component—provides the best passive microwave rain-rate estimate from all available sensors because of the set of frequencies that it observes (nine measurements between 10.7 and 85.5 GHz), plus its overall better spatial resolution (i.e., as low as 5 km at 85 GHz; Kummerow et al. 2001). Thus, it is not surprising that it performed at least as well as the other algorithms. Interestingly, the AMSU-based TRaPs performed competitively, which is a testimony to the success of the high-frequency retrieval algorithm developed for a sensor that was not specifically designed for quantitative precipitation estimation (Weng et al. 2003). In the near future, a version of GPROF developed for use with SSM/I will be available to users that should improve its performance with TRaP.

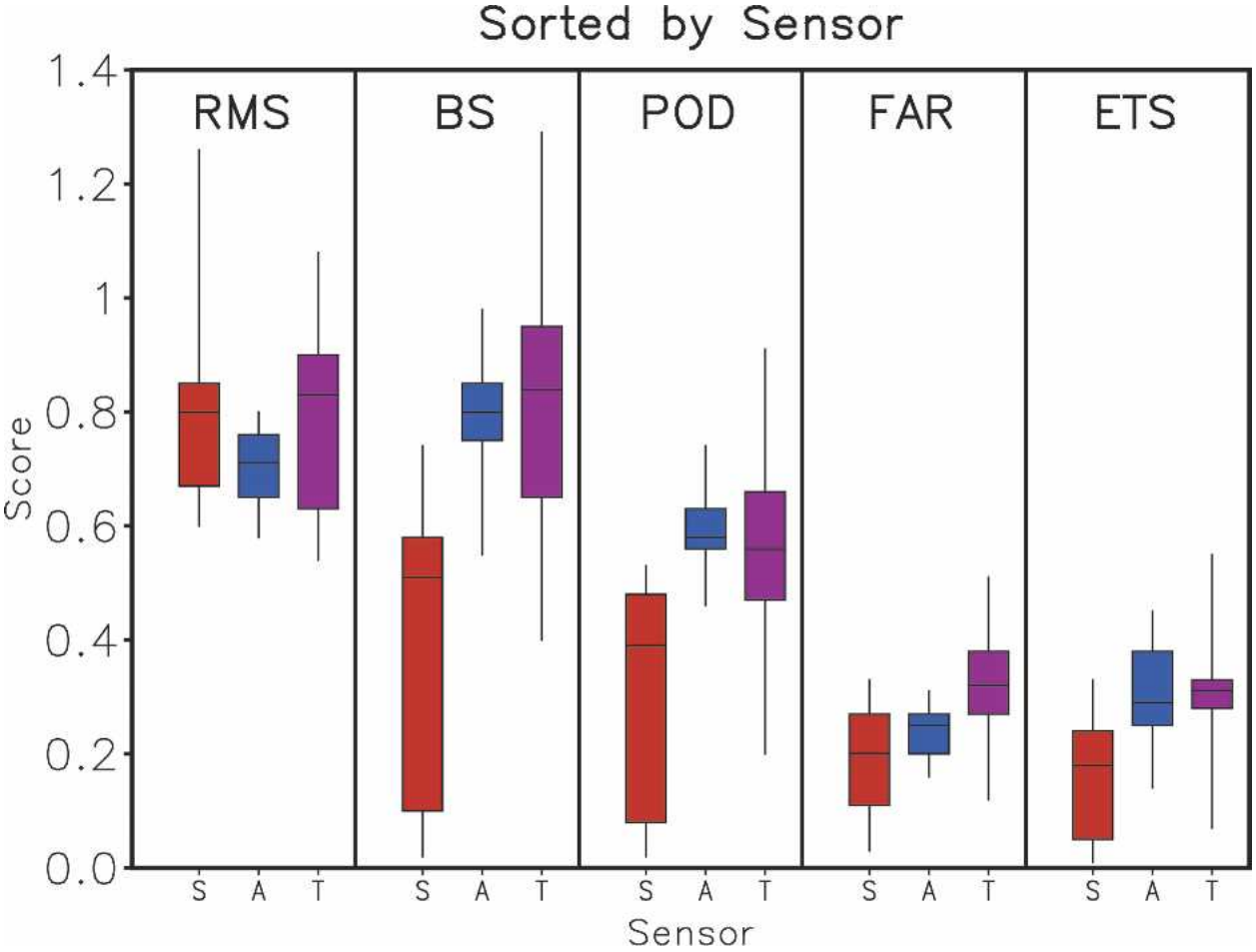

FIG. 2. Box and whisker plots comparing the rms (mm day^{-1}), BS, POD, FAR, and ETS parameters as a function of sensor type (S = SSM/I, A = AMSU, and T = TMI). A 25.4 mm day^{-1} threshold was used with the stage IV estimates being the reference dataset.

TABLE 5. As in Table 4, but as a function of time before landfall (using 25.4 mm day^{-1} minimum threshold).

Time before landfall (cases)	Rain rate (TRaP/ stage IV)	Rain volume (TRaP/ stage IV)	TRaP max (mm day^{-1})	Stage IV max (mm day^{-1})	MAE (mm day^{-1})	Rms	*R*	Bias score	POD	FAR	ETS
6 (6)	0.64	0.62	97.3	362.2	34.3	0.68	0.16	0.67	0.48	0.28	0.25
12 (12)	1.00	0.82	150.0	288.2	45.4	0.80	0.23	0.70	0.51	0.24	0.29
18 (17)	0.65	0.59	105.7	389.5	47.4	0.79	0.10	0.67	0.50	0.22	0.28
24 (7)	0.50	0.54	88.4	359.5	40.7	0.65	−0.03	0.48	0.34	0.30	0.15

c. *Time before landfall*

The second set of data stratification applied to the 42 TRaPs was to examine the TRaPs by the time before landfall. These were sorted in groups of 6-hourly periods prior to landfall. Table 5 and Fig. 3 summarize these results using a 25.4 mm day^{-1} threshold. First examining Table 5, it is apparent that the rain rate, volume, and maximum value are superior for the TRaPs generated 12 h prior to landfall. Overall, the TRaPs generated between 12 and 18 h of landfall show similar statistical values and outperformed the TRaPs generated at the other two times (6 and 24 h). Figure 3 is a little less convincing than the mean values shown in Table 5 in the sense that the 6-h TRaP can be competitive at times (note the BS and ETS). Nonetheless, based on both sources of information, we conclude that on average the best TRaPs are those that are generated between 12 and 18 h of landfall. However, from Table 5, it is noted that for any given TRaP at the other times,

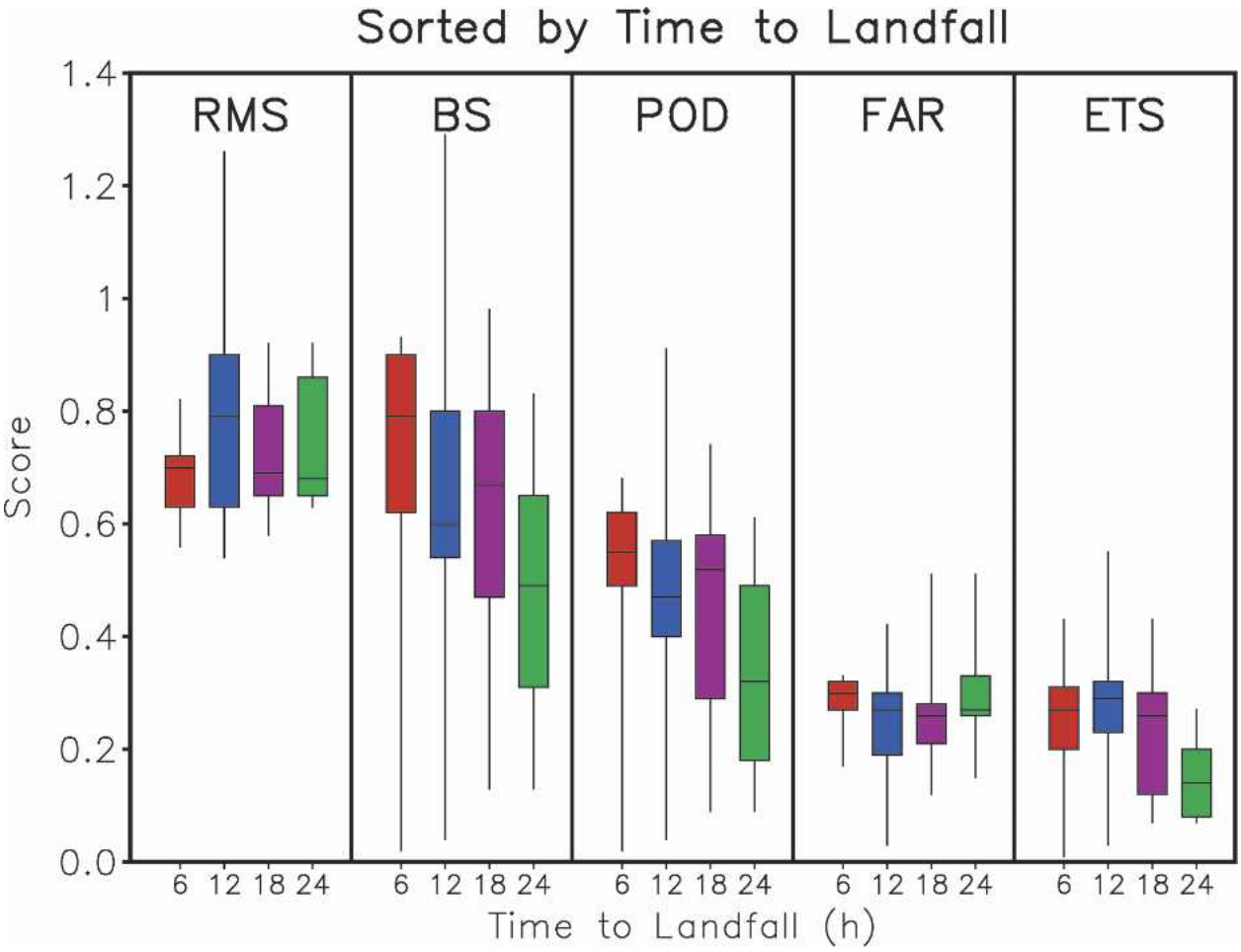

FIG. 3. As in Fig. 2, but as a function of the time before landfall.

TABLE 6. As in Table 3, but for the Eta Model forecasts (using a 25.4 mm day^{-1} minimum threshold).

2002 storm name	Rain rate (Eta/ stage IV)	Rain volume (Eta/ stage IV)	Eta max (mm day^{-1})	Stage IV max (mm day^{-1})	MAE (mm day^{-1})	Rms	R	Bias score	POD	FAR	ETS
Bertha	0.46	0.48	40.1	201.7	50.7	1.11	−0.10	0.34	0.15	0.48	0.06
Fay	0.64	0.64	46.6	165.1	46.6	2.14	−0.01	0.89	0.08	0.92	−0.05
Hanna	0.43	0.43	31.4	419.5	23.0	0.37	0.14	0.06	0.05	0.17	0.03
Isidore	0.48	0.48	54.2	240.7	44.4	0.86	0.20	0.27	0.14	0.51	0.01
Lili	0.56	0.57	53.8	187.8	25.80	0.57	−0.20	0.28	0.12	0.52	0.04
Average	0.51	0.51	45.2	242.9	50.31	0.95	0.01	0.37	0.11	0.52	0.02

they too can be useful. Apparently, the satellite-derived rain rates associated with the approaching storm are most valid between 12 and 18 h prior to landfall. Also, storms that are within 6 h of landfall already have part of their circulation over land and are usually beginning to decay; thus, the extrapolation of their rain fields for the next 24 h is no longer a valid assumption. Finally, it should be pointed out that coastal retrievals are problematic in the passive microwave (e.g., Bennartz 1999), so this may also contribute to these findings.

d. Comparisons versus Eta Model forecasts

Without the TRaP or other "value added" forecasts, the only other readily available means that a forecaster has to predict both intensity and areal rainfall from an impending landfalling tropical system is NWP model forecasts. For the United States, the most commonly used models are the Eta (Black 1994) and Global Forecast System [GFS, previously Aviation (AVN)] models (Kanamitsu et al. 1991) although there are other models, such as the Geophysical Fluid Dynamics Laboratory (GFDL; Kurihara et al. 1998), that are more appropriate for tropical systems. To assess the value of the TRaP, it is worthwhile to make comparisons to the Eta Model forecasts to serve as a benchmark. As such, the 24-h rainfall forecasts made at the times closest to the 42 TRaPs were obtained and subjected to the same statistical analysis. However, since the archived Eta Model data came on a 50-km grid, a fair comparison between the TRaP and Eta as if they were put on the same grid (i.e., we linearly averaged the stage IV data up to the Eta grid size) is not exact. However, it was felt that upscaling the Eta Model rainfall fields to match that of the stage IV would be inappropriate because of the large differences in the spatial resolution (Tustison et al. 2001). Nonetheless, the objective here is to see whether there is a clear distinction between the values of the TRaP versus the Eta, not a direct comparison in terms of actual statistical measures.

Since the Eta forecasts are made at either 0000 or 1200 UTC, the closest 24-h rainfall to the time of the TRaP was used in the comparison. TRaPs that were generated more than ±6 h from this nominal model's forecast times were not included in this analysis in order to insure that there was a reasonable comparison between the two rainfall estimates. Roughly 90% of the TRaPs generated were within ±3 h of the forecast model times.

Table 6 and Fig. 4 summarize the statistical parameters from the Eta Model forecasts for 35 forecasts (out of the 42 original TRaP forecasts) using the 25.4 mm day^{-1} threshold. It is obvious that the Eta Model 24-h forecasts were significantly worse than those from the satellite-based TRaP in virtually every statistical category. The Eta Model performance is not surprising since it is not well suited for tropical systems. The low ETS values reported here are similar in magnitude to those found by Gallus (2002) using a 25.4 mm day^{-1} threshold in validating warm-season Eta Model quantitative precipitation forecasts on a 30-km grid scale.

e. Comparison to climatological schemes

Prior to the development of TRaP, various researchers established several "rules of thumb" regarding the maximum rainfall associated with landfalling tropical cyclones. Recently, Pfost (2000) discussed and compared such techniques, all of which include a simple relationship between rainfall and storm speed. It was felt that some of these should be used in this study to compare to the magnitude of the maximum rainfall from TRaP. It should be noted that the obvious shortcoming of such techniques is predicting the location and distribution of the heavy rainfall. To generate these estimates, the storm speed based upon the Tropical Prediction Center (TPC) advisory, closest to the time of the TRaP, was used. Table 7 shows the results. It is quite apparent that, in general, the climatological schemes do in fact offer maximum rainfall totals much closer to the maximums observed, although the location of this rainfall is not provided. However, it should be noted that since the microwave rain rates represent relatively large areas on the surface (anywhere from 5

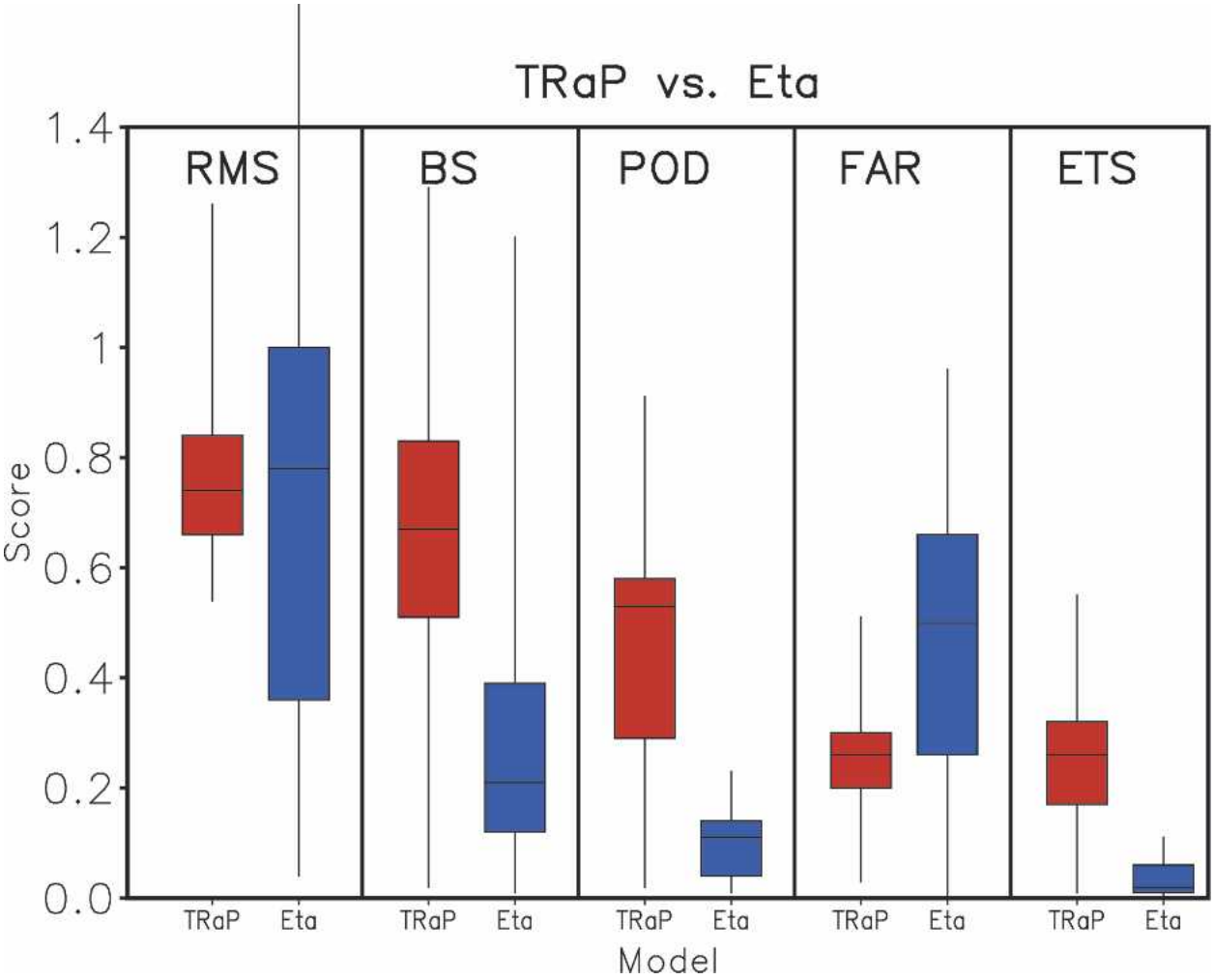

FIG. 4. As in Fig. 2, but for the TRaP and Eta Model forecasts.

to 25 km depending upon the sensor), the maximum rainfall it can produce in TRaP does not represent a point measurement like the climatological methods do. Also, the climatological schemes may include rainfall for periods in excess of 24 h, whereas the TRaPs used in this study are strictly for 24 h.

f. Storm track sensitivity

As was suggested in Ferraro et al. (2002), the accuracy of the TRaP is dependent upon the accuracy of the forecast storm speed and direction. This makes sense since the TRaP formula in its simplest form (i.e., TraP $= R_{\text{avg}}DV^{-1}$) implies that a slower forward speed would mean an increase in rainfall. For example, an error by 50% in storm speed alone would double the rainfall, and any direction error would compound the problem by putting the maximum rainfall at the wrong location. Using the TPC "best track" of a particular storm (e.g., the actual track of the storm as determined in these cases by the Tropical Prediction Center), selected TRaPs were rerun. It was found that interpolation of the track to positions every 3 h produced the best results. So in general, the results that we found were mixed. In some cases, the TRaPs were vastly im-

TABLE 7. Comparison of the storm mean maximum rainfall (mm day^{-1}) derived from stage IV, TraP, and three climatological schemes based on storm speed (x in kt; y in m h^{-1}).

Storm name	Stage IV	TRaP	$100/x$	$31.1 \times (0.915)^y$	$9.75 - 0.039x$
Bertha	330.1	66.8	423.3	425.9	241.6
Fay	359.1	220.3	473.4	450.3	242.1
Hanna	473.8	88.8	327.2	356.3	239.8
Isidore	340.4	87.0	250.0	279.1	237.5
Lili	225.1	92.5	183.6	191.3	233.9
Average	352.7	112.5	330.2	240.4	239.0

proved; others exhibited very little change. Never did we see degradation in performance. Most likely, in those cases, the original TRaP was simply poor, most likely due to the proximity of the storm to land, intensity changes, error in the satellite rain-rate retrievals, etc. (i.e., poor assumptions in the basic TRaP). This is an area for further investigation.

4. Summary and suggestions for future work

This study has focused on the validation of the NOAA/NESDIS operational TRaP product for the 2002 tropical Atlantic hurricane season. Forty-two operational TRaPs that were generated within 24 h of five landfalling storms over the continental United States were validated using an automated objectively based statistical package. A number of statistical parameters were used in the comparison. It was recognized that a number of limitations in this study exist, including the relatively small number of storms analyzed and the various spatial scales of the operational data sources being compared that could impact the interpretation of the statistics. Nonetheless, important results were found that should aid operational users of the TRaP.

It was found that considering only those data points with a minimum daily rainfall threshold of 25.4 mm day^{-1} of observed precipitation was more insightful than a 1 mm day^{-1} threshold since the main focus of the TRaP is on heavy precipitation potential leading to flooding. When this higher threshold is used, it changed some of the statistical parameters quite drastically, most notably, the correlation coefficient, due to the elimination of the low-end values.

Despite the relatively small number of TRaPs being analyzed, they were further stratified by satellite sensor type (i.e., AMSU, SSM/I, and TMI) and time before landfall to better understand their tendencies. It was found that the TMI TRaPs performed the best, although the AMSU-based TRaPs were a close second in performance. The SSM/I TRaPs performed the worst and this was attributed to a lack of upgrades in the operational algorithm since the late 1990s. It was noted that a much improved SSM/I algorithm, namely a version of GPROF, is being run in an experimental mode at NESDIS and will be available for use during the 2005 tropical season. Also, the TRaPs generated between 12 and 18 h of landfall, on average, performed better than those within 6 h or later than 18 h. However, there are instances where even those TRaPs performed comparably.

To provide some sort of reference to the value of the TRaP, we validated the performance of the Eta Model 24-h rainfall forecasts to the same stage IV datasets for the majority of the TRaP cases. It was found that the TRaP outperformed the Eta Model in virtually every statistical category. In addition, three different climatologically based techniques that estimate the maximum rainfall based on storm speed movement were examined. On average, all of these methods produced maximum rainfall forecasts closer to the stage IV than did TRaP. This may simply be due to the large footprints of the microwave rain-rate estimates, which represent area averages and not point rain rates. As noted, these climatological techniques give no information as to the location and distribution of the rainfall.

The TRaP is highly dependent on the forecast track of the storm. A limited number of TRaPs were recomputed using 3-hourly time-interpolated positions based on the actual track of the storm. Results were inconclusive due to the small sample of data examined, but never did the TRaP perform any worse.

As described in Part I, there are a number of ways in which the TRaP technique itself could be extended and improved. This includes changing assumptions in the technique, improving the microwave rain retrieval schemes, in particular, to work better for tropical systems, and developing better ways to use storm track information. All of these potential improvements would add more credibility to future validation studies. Additionally, future validation studies must consider the following:

- The impact of the different spatial resolutions from the various satellite sensors and comparison datasets on the probability distribution function of the rain-rate field and on the validation statistics.
- Tustison et al. (2001) and Gallus (2002) point out that there are major difficulties in trying to compare rainfall observations on different spatial scales. Comparing radar and rain gauge data is the classic problem in this regard, but verifying model forecasts or TRaP forecasts with radar data also have "representativeness" errors. These require further study.
- Expanding the scope of the validation in space (by using ground-based rainfall estimates from other nations) and in time (using data from multiple years). These steps will improve the applicability and statistical significance of the results.
- Closer examination of the impact of the storm track forecast on the TRaP. Again, a larger sample size is required to draw conclusions on this.
- Finally, it is noted that this study did not consider other advanced tropical rainfall estimation schemes. These include the Rainfall Climatology and Persistence model (R-CLIPER) (Marks et al. 2002) and the GFDL hurricane model (DeMaria and Tuleya 2001).

It is imperative that future validation efforts include comparisons from these models.

Acknowledgments. We are deeply indebted to the comments from our anonymous reviewers and thank them for their contribution on making this a better paper. The authors would like to acknowledge the cooperation of E. Ebert (Bureau of Meteorology Research Centre, Melbourne, Australia) for making her validation software available for our use in this study. Also, we extend our thanks to C. Shih (National Center for Atmospheric Research, Boulder, Colorado) for preparing the Eta Model data. Finally, we acknowledge the support of T. Schott (NESDIS/Office of Systems Development, Suitland, Maryland) for providing the resources necessary for making TRaP a viable operational product. A portion of this study was supported and monitored by NESDIS' Center for Satellite Applications and Research of the National Oceanic and Atmospheric Administration (NOAA) under Contract Number DG133E-02-NC-0058 issued to QSS Group, Inc. The views, opinions, and findings contained in this report are those of the author(s) and should not be construed as an official National Oceanic and Atmospheric Administration or U.S. government position, policy, or decision.

REFERENCES

Bennartz, R., 1999: On the use of SSM/I measurements in coastal regions. *J. Atmos. Oceanic Technol.,* **16,** 417–431.

Black, T. L., 1994: The new NMC Mesoscale Eta Model: Description and forecast examples. *Wea. Forecasting,* **9,** 265–284.

Colton, M., and G. Poe, 1994: Shared Processing Program, Defense Meteorological Satellite Program, Special Sensor Microwave/Imager Algorithm Symposium, 8–10 June 1993. *Bull. Amer. Meteor. Soc.,* **75,** 1663–1669.

DeMaria, M., and R. E. Tuleya, 2001: Evaluation of quantitative precipitation forecasts from the GFDL hurricane model. Preprints, *Symp. on Precipitation of Extremes: Prediction, Impacts, and Responses,* Albuquerque, NM, Amer. Meteor. Soc., 340–343.

Ebert, E. E., M. J. Manton, P. A. Arkin, R. E. Allam, and A. Gruber, 1996: Results from the GPCP algorithm intercomparison programme. *Bull. Amer. Meteor. Soc.,* **77,** 2875–2887.

——, U. Damrath, W. Wergen, and M. E. Baldwin, 2003: The WGNE assessment of short-term quantitative precipitation forecasts. *Bull. Amer. Meteor. Soc.,* **84,** 481–492.

Ferraro, R. R., 1997: SSM/I derived global rainfall estimates for climatological applications. *J. Geophys. Res.,* **102,** 16 715–16 735.

——, P. Pellegrino, S. Kusselson, M. Turk, and S. Kidder, 2002: Validation of SSM/I and AMSU derived tropical rainfall potential (TRaP) during the 2001 Atlantic hurricane season. NOAA Tech. Rep. 105, 47 pp.

Fulton, R. A., J. P. Breidenbach, D. J. Seo, and D. A. Miller, 1998: The WSR-88D rainfall algorithm. *Wea. Forecasting,* **13,** 377–395.

Gallus, W. A., 2002: Impact of verification grid-box size on warm-season QPF skill measures. *Wea. Forecasting,* **17,** 1296–1302.

Kanamitsu, M., and Coauthors, 1991: Recent changes implemented into the global forecast system at NMC. *Wea. Forecasting,* **6,** 425–435.

Kidder, S. Q., and Coauthors, 2005: The tropical rainfall potential (TraP) technique. Part I: Description and examples. *Wea. Forecasting,* **20,** 456–464.

Kummerow, C., and Coauthors, 2001: The evolution of the Goddard Profiling Algorithm (GPROF) for rainfall estimation from passive microwave sensors. *J. Appl. Meteor.,* **40,** 1801–1820.

Kurihara, Y., R. E. Tuleya, and M. A. Bender, 1998: The GFDL hurricane prediction system and its performance in the 1995 hurricane season. *Mon. Wea. Rev.,* **126,** 1306–1322.

Marks, F. D., G. Kappler, and M. DeMaria, 2002: Development of a tropical cyclone rainfall climatology and persistence (R-CLIPER) model. Preprints, *25th Conf. on Hurricanes and Tropical Meteorology,* San Diego, CA, Amer. Meteor. Soc., 327–328.

McCollum, J., W. Krajewski, R. Ferraro, and M. Ba, 2002: Evaluation of biases of satellite rainfall estimation algorithms over the continental United States. *J. Appl. Meteor.,* **41,** 1065–1080.

Pfost, R. L., 2000: Operational tropical cyclone quantitative precipitation forecasting. *Natl. Wea. Dig.,* **24,** 61–66.

Rappaport, E. N., 2000: Loss of life in the United States associated with recent Atlantic tropical cyclones. *Bull. Amer. Meteor. Soc.,* **81,** 2065–2073.

Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting,* **5,** 570–575.

Tukey, J. W., 1977: Box-and-whisker plots. *Explanatory Data Analysis,* Addison-Wesley, 39–43.

Tustison, B., D. Harris, and E. Foufoula-Georgiou, 2001: Scale issues in verification of precipitation forecasts. *J. Geophys. Res.,* **106,** 11 775–11 784.

Weng, F., L. Zhao, G. Poe, R. Ferraro, X. Li, and N. Grody, 2003: Advanced Microwave Sounding Unit (AMSU) cloud and precipitation algorithms. *Radio Sci.,* **38,** 8086–8096.

Xie, P., J. Janowiak, P. Arkin, R. Adler, A. Gruber, R. Ferraro, G. Huffman, and S. Curtis, 2003: GPCP pentad precipitation analyses: An experimental dataset based on gauge observations and satellite estimates. *J. Climate,* **16,** 2197–2214.