Ensemble Tropical Rainfall Potential (eTRaP) Forecasts

ELIZABETH E. EBERT

Centre for Australian Weather and Climate Research, Melbourne, Victoria, Australia

MICHAEL TURK, SHELDON J. KUSSELSON, JIANBIN YANG, AND MATTHEW SEYBOLD

NOAA/NESDIS/OSDPD, Camp Springs, Maryland

PETER R. KEEHN AND ROBERT J. KULIGOWSKI

NOAA/NESDIS/STAR, Camp Springs, Maryland

(Manuscript received 12 May 2010, in final form 26 October 2010)

ABSTRACT

Ensemble tropical rainfall potential (eTRaP) has been developed to improve short-range forecasts of heavy rainfall in tropical cyclones. Evolving from the tropical rainfall potential (TRaP), a 24-h rain forecast based on estimated rain rates from microwave sensors aboard polar-orbiting satellites, eTRaP combines all single-pass TRaPs generated within ±3 h of 0000, 0600, 1200, and 1800 UTC to form a simple ensemble. This approach addresses uncertainties in satellite-derived rain rates and spatial rain structures by using estimates from different sensors observing the cyclone at different times. Quantitative precipitation forecasts (QPFs) are produced from the ensemble mean field using a probability matching approach to recalibrate the rain-rate distribution against the ensemble members (e.g., input TRaP forecasts) themselves. ETRaPs also provide probabilistic forecasts of heavy rain, which are potentially of enormous benefit to decision makers. Verification of eTRaP forecasts for 16 Atlantic hurricanes making landfall in the United States between 2004 and 2008 shows that the eTRaP rain amounts are more accurate than single-sensor TRaPs. The probabilistic forecasts have useful skill, but the probabilities should be interpreted within a spatial context. A novel concept of a ''radius of uncertainty'' compensates for the influence of location error in the probability forecasts. The eTRaPs are produced in near–real time for all named tropical storms and cyclones around the globe. They can be viewed online (http://www.ssd.noaa.gov/PS/TROP/etrap.html) and are available in digital form to users.

1. Introduction

Tropical storms and cyclones are among the most damaging natural hazards worldwide. In addition to strong winds and storm surge, heavy rainfall in these storms can lead to dangerous inland flooding. Freshwater floods associated with hurricanes were responsible for more than 300 deaths in the United States during the period 1970–99, including 50 deaths related to Hurricane Floyd in 1999 (Rappaport 2000).

To address the need for cyclone-related heavy rain information, the National Environmental Satellite, Data, and Information Service (NESDIS) has been producing operational areal tropical rainfall potential (TRaP) forecasts of rainfall for landfalling tropical cyclones since the early 2000s. TRaP forecasts (called TRaPs in this paper) are essentially 24-h extrapolation forecasts of satellite-estimated rain rates that give the expected location and intensity of the rain maximum as well as the spatial rainfall pattern (Kidder et al. 2005). TRaPs are derived from rain-rates estimate from passive microwave sensors on several polar-orbiting satellites, and include the Advanced Microwave Sounder Unit (AMSU), Special Sensor Microwave Imager (SSM/I), and the Tropical Rainfall Monitoring Mission (TRMM) Microwave Imager (TMI). In 2008 TRaP forecasts from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) became available. Experimental TRaPs from the operational NESDIS Hydro-Estimator (H-E; Scofield and Kuligowski 2003), which generates rainfall estimates from infrared data from geostationary satellites, have been made for U.S. hurricanes since 2004.

Corresponding author address: Elizabeth E. Ebert, CAWCR, Bureau of Meteorology, Melbourne, VIC 3001, Australia.
E-mail: e.ebert@bom.gov.au

DOI: 10.1175/2010WAF2222443.1

TRaP forecasts are conceptually quite simple. To produce an areal TRaP, a satellite ''snapshot'' of instantaneous rain rates is propagated forward in time following the predicted path of the cyclone using track forecasts made at operational tropical cyclone warning centers in the region under threat. Every 15 min a new position is calculated and the spatial rain rates are applied over a rectangular grid of approximately 4-km resolution; the 15-min accumulations are summed over a period of 24 h (Kidder et al. 2005). Three basic assumptions are made in the calculation of TRaP forecasts: (a) the satellite rain-rate estimates are accurate, (b) the forecasts of cyclone track are accurate, and (c) the rain rates over a 24-h period can be approximated as steady state following the cyclone path. Errors in TRaP rainfall predictions can be attributed to flaws in one or more of these assumptions.

Studies by Ferraro et al. (2005) and Ebert et al. (2005) on the accuracy of 24-h TRaP forecasts over the United States and Australia, respectively, have shown that in general the TRaPs give reasonable estimates of both the maximum rainfall accumulation and its spatial distribution but underestimate the total rain volume by about one-third in both regions. The overall accuracy was similar to that of mesoscale models at that time. The results from these validation studies suggest that the errors in TRaP forecasts are mainly related to errors in satellite rain rates and the assumption of steady-state rainfall, and less so to errors in operational track predictions. While there is some systematic error in the TRaPs (e.g., underestimation of rain volume), the TRaP performance varies greatly from storm to storm and, indeed, among different TRaPs for a single storm. This means that it is difficult to estimate a priori the accuracy of a particular TRaP forecast.

One way to reduce the random error is to average several forecasts together in a ''poor man's ensemble.'' Numerous studies have shown that the combination of independent forecasts yields a consensus forecast that is usually more accurate than the individual components (e.g., Clemen 1989; Woodcock and Engel 2005). Averaging has the effect of smoothing the rain field, with the associated advantages and disadvantages. The mean field is less likely to produce very large errors when compared to the observations; however, the averaging damps the high rain intensities, which were the original motivation for making the TRaP forecasts. A more intelligent approach would be to retain information on the distribution of forecasts within the ensemble, making use of the uncertainty (variability) among the TRaP forecasts composing the ensemble. One can generate probabilistic forecasts of rain exceeding certain critical thresholds in locations of interest, an approach that is very amenable to risk management and mitigation strategies. Kidder et al. (2005) and Ebert et al. (2005) both suggested ensemble TRaP as a possible way forward.

In recent years 6-h TRaP rainfall accumulations have been produced and archived as part of the operational processing of 24-h TRaPs. These provide useful short-period forecasts that can be used to generate time series of predicted rain evolution at locations of interest. These short-period forecasts can also be combined in multiple permutations to make an ensemble of TRaP forecasts for 6- and 24-h accumulations.

In principle, an ensemble TRaP, henceforth called an eTRaP, can be made up of forecasts using observations from several microwave sensors, initialized at several observation times, using several different track forecasts. The diversity among the ensemble members helps to reduce the large (unknown) errors associated with a single-sensor, single-track TRaP. The large number of perturbations leads to ensembles with many members, allowing probability forecasts to be issued.

The next section describes the methodology for generating eTRaPs. We demonstrate eTRaP forecasts of heavy rain associated with the landfall of Hurricane Rita during September 2005. The accuracy of the ensemble TRaP QPFs is quantified and compared with the accuracy of individual single-sensor TRaPs and statistical rainfall climatology and persistence model (R-CLIPER) forecasts (Tuleya et al. 2007) for heavy rain forecasts in 16 Atlantic tropical storms and hurricanes making landfall in the United States during 2004–08. The skill of the probabilistic forecasts is also assessed. The paper concludes with suggestions for future work.

2. Generation of ensemble TRaP forecasts

Ensemble TRaP products include both deterministic and probabilistic quantitative precipitation forecasts (QPFs and PQPFs) displayed as maps of rain amount and probability. To generate the eTRaPs, the input TRaPs are first assigned to the nearest synoptic time of 0000, 0600, 1200, or 1800 UTC, which means that they are at most 3 h offset from the synoptic time they represent. Prior to combination, all TRaP forecasts are remapped onto a common 800×800 pixel domain (roughly 30° square) centered on the forecast position of the storm midway through the 24-h period.

In principle, it would be possible to create an ensemble from only the most recent TRaP forecasts, since these are normally more accurate than older forecasts. However, in practice the number of TRaPs available within the most recent 6-h period is quite small, typically between one and four, and would not lead to a very well-defined probability distribution. We therefore adopt the strategy of combining newer TRaPs with older ones.

Each 24-h TRaP consists of four 6-h TRaP rainfall accumulations (called TRaP segments in this paper)

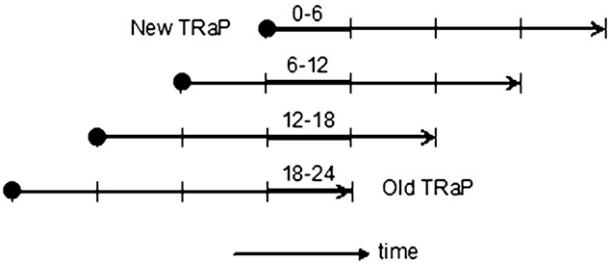

FIG. 1. Schematic showing four TRaPs of different ages. The filled circles indicate the start times of each TRaP, and the dark bars show the latency of each TRaP for a particular 6-h period being forecast.

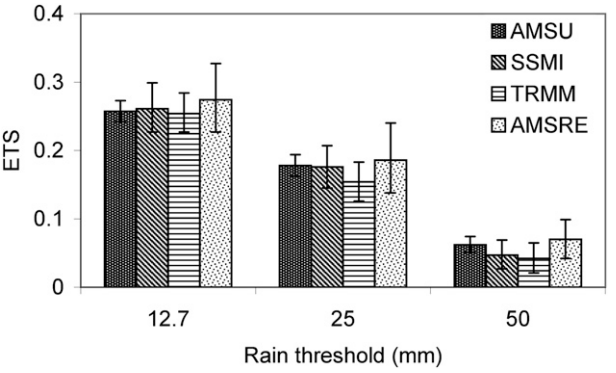

FIG. 2. The ETSs for 6-h rainfall exceeding 12.7, 25, and 50 mm, for TRaPs based on AMSU, SSM/I, TRMM, and AMSR-E data during 2007–08. The 95% confidence intervals are shown.

corresponding to the first, second, third, and fourth 6-h periods. As shown in Fig. 1, for a particular 6-h period of time there may be some TRaP segments starting at that time, and some that start 6, 12, or 18 h earlier. The number of ensemble members in an eTRaP is the number of permutations possible for combining the various segments of the forecast. For example, if there are five 6-h TRaP segments available for the first 6 h of a 24-h forecast, four TRaP segments for the second 6-h period, four TRaP segments for the third period, and two TRaP segments for the fourth period, then the number of ensemble members composing the 24-h forecast is $5 \times 4 \times 4 \times 2 = 160$.

In many cases a large number of 6-h TRaP segments are available for generating an ensemble, such that more than a thousand ensemble members are possible. Some culling procedures are invoked to keep the ensemble to a manageable size. If more than one TRaP is issued from a given satellite overpass, then only the latest TRaP is included in the ensemble. An exception is when the TRaPs were issued by different operational centers, in which case both are retained because the different track forecasts may give useful information on track uncertainty. If the number of potential ensemble members exceeds 200, the permutations are randomly culled to reduce the number to 200 to reduce processing time.

Every 6-h TRaP segment contributing to the ensemble is weighted according to its expected accuracy, which depends on the age (latency) of the TRaP segment and the sensor from which the rainfall is derived. The overall weight assigned to the ith TRaP segment, $w_{i,6}$, is the product of its sensor weight and its forecast latency weight:

$$w_{i,6} = w_{\text{sensor}} \times w_{\text{latency}}. \quad (1)$$

The weights were determined by verifying the 6-h TRaPs against 6-h gridded observations derived from National Centers for Environmental Prediction's hourly stage IV radar–rain gauge analyses over the United States (Lin and Mitchell 2005). The stage IV analyses were remapped to the 4-km Mercator grid of the TRaP forecasts and subset to a domain of dimensions 400×400 (roughly 15° square) centered over the storm's position as determined by the National Hurricane Center (NHC). Stage IV rainfall values from grid boxes located more than 50 km offshore were not used in the validation due to potential radar range effects and the absence of gauge data for local bias correction. To generate weights for combination, the inverse of the mean square error is frequently used (e.g., Woodcock and Greenslade 2007). For eTRaPs we wish to emphasize the correct prediction of heavy rain location and amount, and so we based the weights on the values of the equitable threat score (ETS; Jolliffe and Stephenson 2003). The ETS is a metric frequently used in precipitation verification, and measures the fraction of observed and/or forecast occurrences of rain exceeding a particular threshold that were correctly predicted, adjusted for hits due to random chance.

Figure 2 shows the ETS for 6-h rain exceeding three different rain thresholds, for the AMSU-, SSM/I-, TRMM-, and AMSR-E-based TRaPs during 2007–08. This 2-yr period was chosen because the AMSU precipitation algorithm was updated in May 2007 (Vila et al. 2007), and we wish to have weights that reflect the current precipitation algorithms. Figure 2 shows better performance for lighter rain thresholds than for the heaviest one, but no significant difference in performance among the TRaPs from different sensors. Therefore, we set $w_{\text{sensor}} = 1$ for all sensors. This is in contrast to the earlier results of Ferraro et al. (2005) and Ebert et al. (2005), who found that 24-h TRaPs based on AMSU data were more accurate than those from SSM/I and TRMM.

The weights for forecast latency were based on the ETS values shown in Fig. 3, where the more recent 6-h TRaP segments were clearly more accurate than the older segments. This confirms our expectation that steady-state rainfall is a more valid assumption early in the forecast

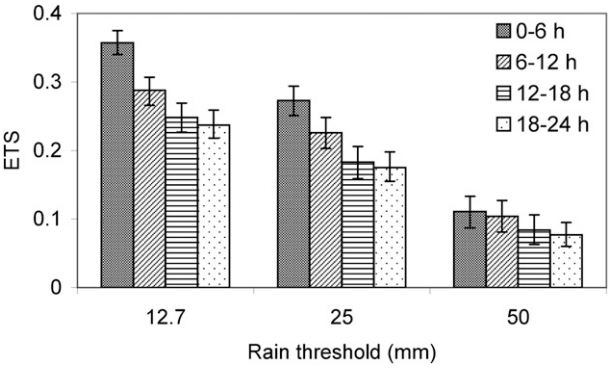

FIG. 3. As in Fig. 2, but for TRaPs with latencies of 0–6, 6–12, 12–18, and 18–24 h during 2004–08.

period than later. Based on these results, the latency weight w_{latency} was approximated as a linearly decreasing function of time, with values of 1.00 for 0–6-h latency, 0.85 for 6–12-h latency, 0.70 for 12–18-h latency, and 0.55 for 18–24-h latency.

An ensemble TRaP forecast for 6-h rain accumulation is made up of the set of 6-h TRaP segments valid at the time of interest, with weights w_i equal to the values of $w_{i,6}$. For a 24-h ensemble TRaP forecast, the rain accumulation for each ensemble member is simply the sum of the rain in each of its four 6-h TRaP segments, and its corresponding weight w_i is the sum of the weights $w_{i,6}$ for the 4 segments.

The (weighted) ensemble mean forecast for rain accumulation is computed as

$$R_{\text{ensmean}} = \frac{\sum_{i=1}^{n} w_i R_i}{\sum_{i=1}^{n} w_i}, \tag{2}$$

where R_i is the areal rainfall of the ith (possibly summed) TRaP ensemble member and n is the number of ensemble members. A probability-matching transformation is then applied to the ensemble mean (e.g., Ebert 2001) to obtain the final eTRaP rain accumulation, R. This approach assumes that the spatial distribution of precipitation is best represented by the field of ensemble mean values, but the intensity distribution is better represented by the original rain rates from the full set of ensemble members. Probability, or histogram, matching replaces the intensity distribution in the ensemble mean by that of the full ensemble. The purpose of this transformation is to restore the heavy rain intensities that may have been lost during averaging, as well as to trim areas of excess light rain caused by the averaging process. For heavy rain events, probability matching provides an important correction to the ensemble mean (Ebert 2001). The weights are applied during the histogram matching as well as during the calculation of the ensemble mean.

Areal forecasts for the probability of rain exceeding a given threshold are computed as the weighted fraction of individual rain forecasts exceeding the threshold:

$$P(R \geq R_T) = \frac{\sum_{i=1}^{n} w_i I_i}{\sum_{i=1}^{n} w_i}, \qquad I_i = \begin{cases} 0 & R_i < R_T \\ 1 & R_i \geq R_T \end{cases}, \tag{3}$$

where R_T is a threshold rain amount. The thresholds chosen for computing probabilistic forecasts are 25, 50, 75, and 100 mm for 6-h forecasts, and 50, 100, 150, and 200 mm for 24-h forecasts. A sample eTRaP forecast is shown in the next section.

ETRaP quantitative and probabilistic precipitation forecasts are available on the Internet (http://www.ssd.noaa.gov/PS/TROP/etrap.html). Forecasts are generated for every tropical storm and cyclone around the globe for which operational bulletins are issued, with updates issued every 6 h at about 3¼ h past the nominal start time. Figure 4 shows an example of an online eTRaP forecast for Typhoon Morakot, which struck Taiwan in August 2009.

3. Example of ensemble TRaP for Hurricane Rita, 24 September 2005

Hurricane Rita was a category 5 hurricane that rapidly intensified in the Gulf of Mexico and caused billions of dollars of damage to communities along the Gulf coast. It made landfall near the Texas–Louisiana border at around 0730 UTC 24 September 2005.

Figure 5 shows the 6-h TRaP segments available within ±3 h of 0000 UTC 24 September 2005, which were used to construct ensemble TRaPs for the subsequent 24-h period. For the first 6-h period 11 TRaP segments were available, ranging from the most recent to nominally 18 h old. The second 6-h period had nine segments with latencies of up to 12 h. The third period had six segments, and the final 6-h period had only three segments since only the most recent TRaP forecasts would have been available. Although 1782 permutations were potentially available to generate 24-h ensemble forecasts, the ensemble size was culled to 200 members.

One thing to note in Fig. 5 is the variability among TRaP estimates for any give period, depending on the latency of the forecast and the satellite sensor used to produce it. For example, the AMSU-based TRaPs tended to have broader rain areas with lower maximum rain accumulations than those based on SSM/I or TRMM data. It would be difficult

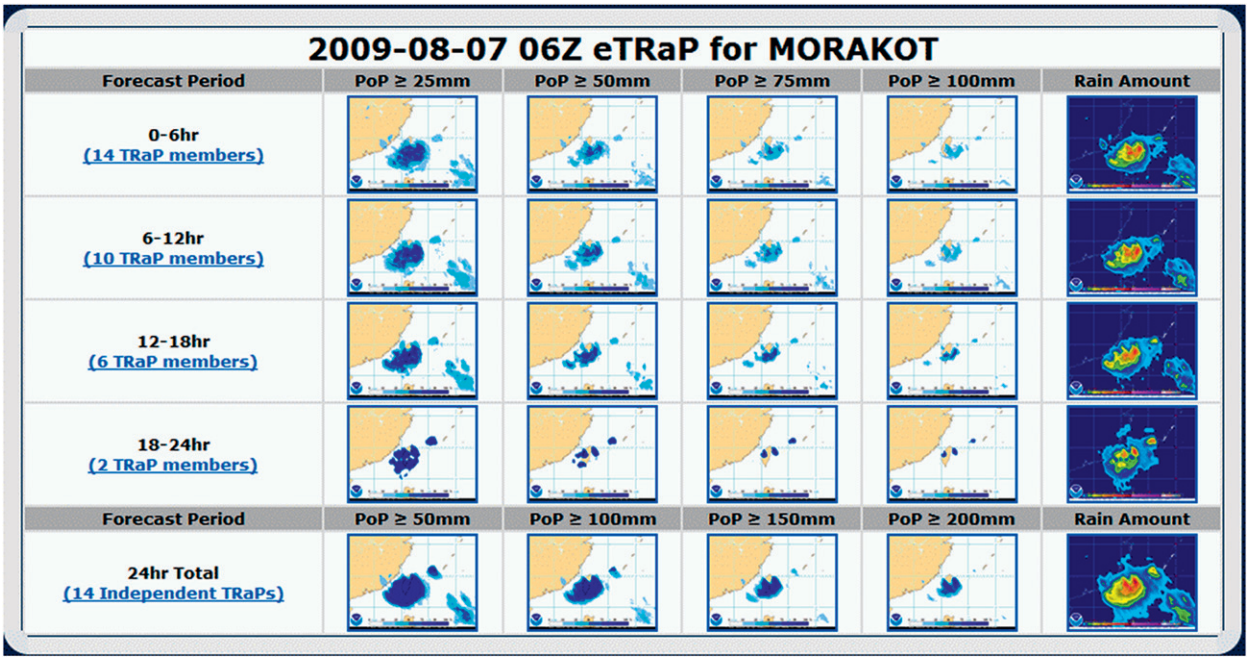

FIG. 4. The ETRaP Web display for Typhoon Morakot on 7 Aug 2009.

to guess in advance which of these many forecasts was likely to be most accurate. The ensemble approach uses a consensus approach to produce a "best guess" quantitative forecast, and represents the forecast uncertainty in the form of probabilities.

Figures 6 and 7 show 6- and 24-h eTRaP forecasts, respectively, made using the 6-h TRaP segments shown in Fig. 5. Corresponding observed rainfall accumulations from the stage IV radar–rain gauge analyses are shown in Fig. 8.

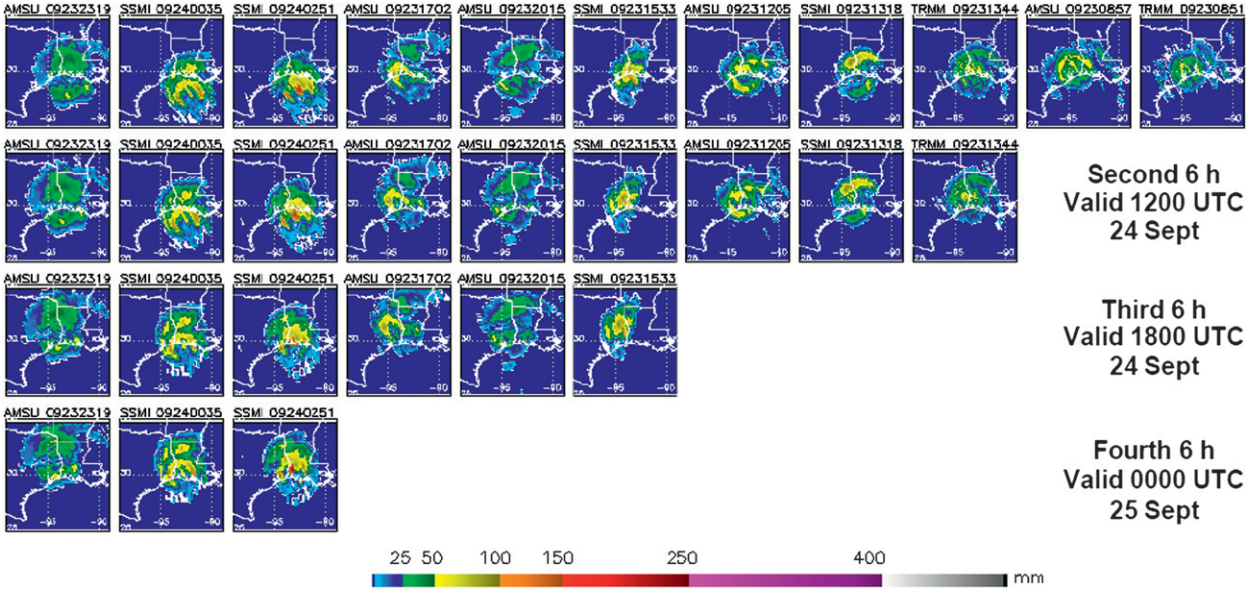

FIG. 5. The 6-h TRaP segments available within 3 h of 0000 UTC 24 Sep 2005 for Hurricane Rita, used to produce the 24-h eTRaP forecast valid at 0000 UTC 25 Sep 2005. (top row) The first 6-h period includes all TRaPs issued during the 18 h leading to (nominally) 0000 UTC 24 September, while (bottom row) the final 6-h period includes only TRaPs issued at (nominally) 0000 UTC 24 September.

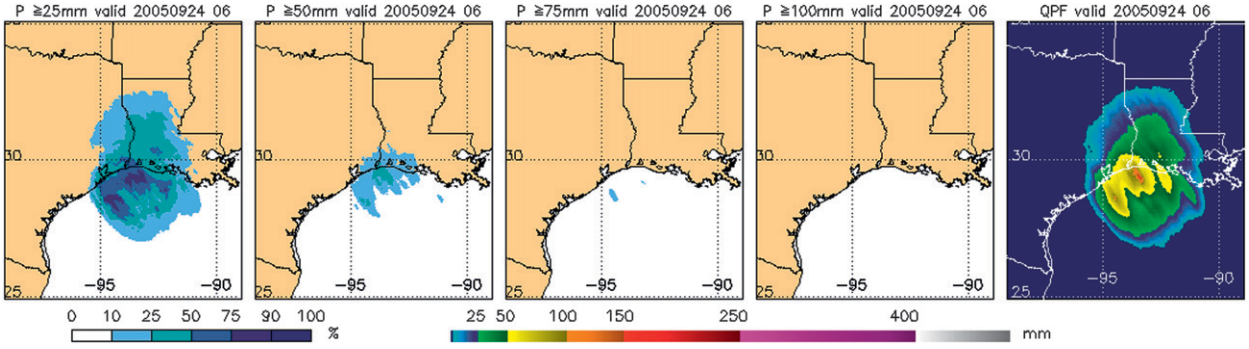

FIG. 6. The 6-h eTRaP forecast for rain in Hurricane Rita ending at 0600 UTC 24 Sep 2005. Eleven TRaPs contributed to this ensemble (top row in Fig. 5). The first four panels show the probabilities of 6-h precipitation exceeding 25, 50, 75, and 100 mm, respectively. The last panel shows the probability-matched ensemble mean 6-h accumulation.

The 6-h probability-matched ensemble mean shows a roughly circular rain region about 500 km in diameter, with the heaviest rain at landfall along the westernmost coast of Louisiana and a secondary maximum near Galveston, Texas (Fig. 6). This corresponds quite well to the observed structure of the rainfall as shown in the stage IV data (Fig. 8a). However, the predicted maximum rainfall of 130 mm was well below the observed value of 200 mm. The probability maps showed >25% chance of exceeding 50 mm along the Texas–Louisiana border, but did not predict any precipitation accumulation (on land) exceeding 100 mm. Since the TRaP assumptions of an accurate storm-track forecast and steady-state rainfall are not likely to be badly violated for such a short-range forecast, this suggests that the satellite underestimation of the rain rates was the primary cause of the error.

The 24-h eTRaP forecast showed excellent placement of the rain maximum, but did not capture the observed heavy rain in southwestern Louisiana. The total rain amount was predicted quite accurately. The probability contours were smoother than for the 6-h case due to the much larger ensemble size. The 50% probability of rain exceeding 100 mm corresponded well to the areas in which rain greater than 100 mm was observed, and a small region of greater than 10% probability of exceeding 200 mm lay directly over the observed elongated rainband in eastern Texas.

This example shows that eTRaP would have provided useful deterministic and probabilistic guidance for heavy rainfall during Hurricane Rita. The next section describes a quantitative verification of eTRaP forecasts using a much larger sample of storms.

4. Performance of eTRaP for Atlantic storms during 2004–08

a. Forecast accuracy

To obtain a more quantitative evaluation of ensemble TRaP, 6- and 24-h eTRaP forecasts for 16 Atlantic tropical storms and hurricanes were verified against stage IV observations mapped onto the 4-km eTRaP grid. The storms and relevant dates are given in Table 1. A total of 423 eTRaP forecasts were made for 6-h accumulations, and 145 eTRaP forecasts were made for 24-h accumulations.

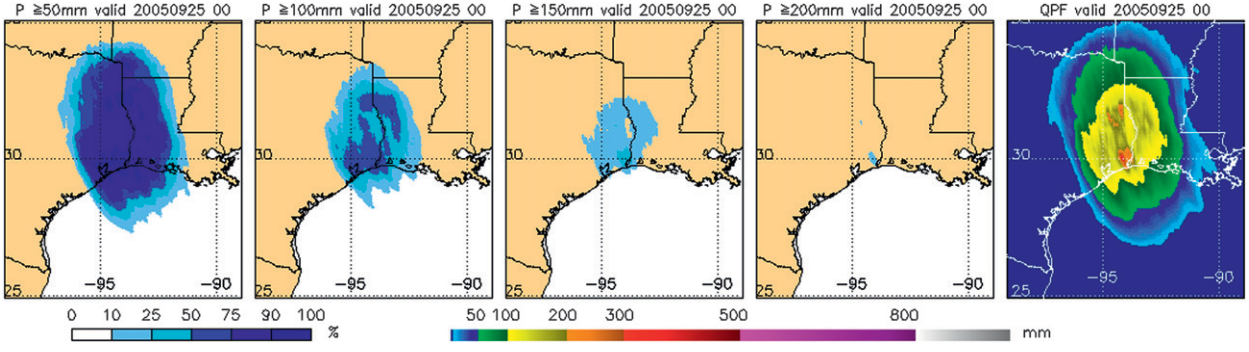

FIG. 7. The 24-h eTRaP forecast for rain during Hurricane Rita ending at 0000 UTC 25 Sep 2005. The color scale differs from that used in Fig. 6. The first 4 panels show the probabilities of 24-h precipitation exceeding 50, 100, 150, and 200 mm, respectively. The last panel shows the probability-matched ensemble mean 24-h accumulation.

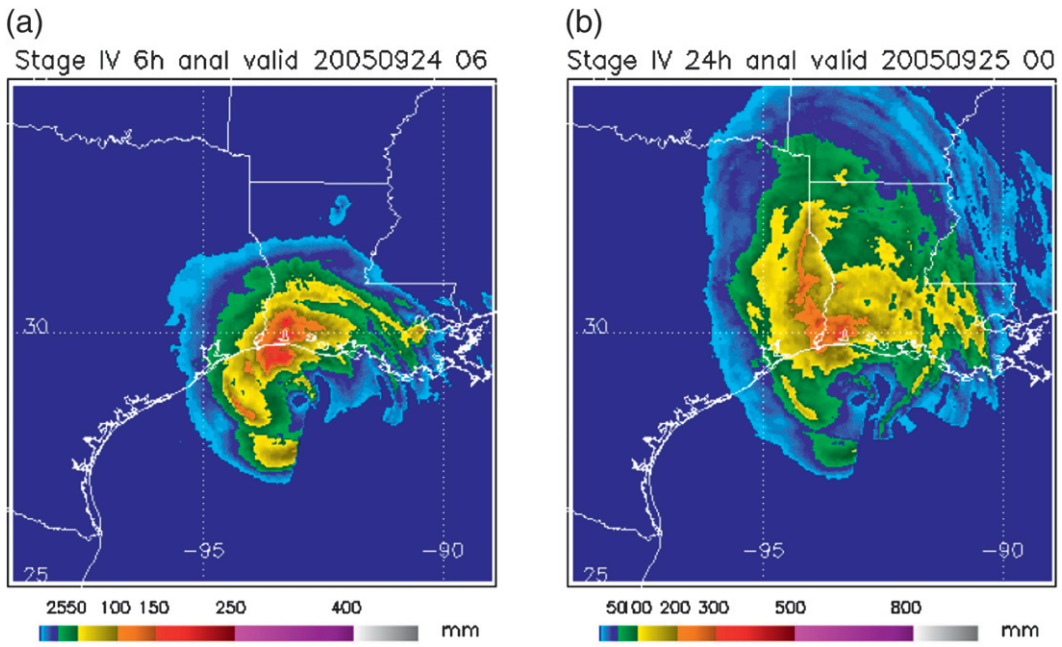

FIG. 8. Stage IV rainfall accumulation during Hurricane Rita: (a) 6 h ending 0600 UTC 24 Sep 2005 and (b) 24 h ending at 0000 UTC 25 Sep 2005. The color scales correspond to those used in Figs. 6 and 7.

As a benchmark, we compare the performance of eTRaP to that of single-orbit TRaP and 24-h forecasts from R-CLIPER (Tuleya et al. 2007; Lonfat et al. 2007), interpolated to a comparable spatial resolution of 0.25°. R-CLIPER gives a smooth and fairly conservative forecast that has been shown to have lower RMS errors than TRaP (Spampata et al. 2008), but in most regards does not perform as well as NWP (Marchok et al. 2007).

Errors in rainfall accumulation and location were computed using the contiguous rain area (CRA) verification method (Ebert and McBride 2000; Ebert et al. 2005). A rain threshold of 25 mm day^{-1} was used to delineate the areas corresponding to moderate and heavy rainfall within the storm systems. The CRA method estimates the rain location error by horizontally translating the forecast position until the greatest correlation with the observations is achieved. The location-corrected forecast is then compared to the observations using a variety of metrics. This verification includes only those grid boxes included within the contiguous rain areas before and after the forecast

TABLE 1. Atlantic tropical storms and hurricanes for which TRaP and eTRaP forecasts were verified against stage IV radar–gauge analyses.

Storm	Period verified
Tropical Storm Bonnie	1200 UTC 11 Aug–0000 UTC 13 Aug 2004
Hurricane Charley	0000 UTC 13 Aug–1800 UTC 15 Aug 2004
Hurricane Frances	0000 UTC 5 Sep–0000 UTC 8 Sep 2004
Hurricane Ivan	1200 UTC 15 Sep–0000 UTC 17 Sep 2004
Tropical Storm Arlene	0600 UTC 11 Jun–0000 UTC 13 Jun 2005
Hurricane Katrina	0000 UTC 26 Aug–0000 UTC 27 Aug 2005, 0000 UTC 29 Aug–0000 UTC 31 Aug 2005
Hurricane Rita	1200 UTC 23 Sep–0000 UTC 26 Sep 2005
Tropical Storm Barry	1200 UTC 2 Jun–1800 UTC 3 Jun 2007
Tropical Storm Erin	0600 UTC 16 Aug–1200 UTC 17 Aug 2007
Hurricane Humberto	0000 UTC 13 Sep–1800 UTC 14 Sep 2007
Hurricane Dolly	0600 UTC 23 Jul–1200 UTC 25 Jul 2008
Tropical Storm Edouard	1200 UTC 5 Aug–1800 UTC 6 Aug 2008
Tropical Storm Fay	0000 UTC 19 Aug–0000 UTC 25 Aug 2008
Hurricane Gustav	0600 UTC 1 Sep–1800 UTC 2 Sep 2008
Tropical Storm Hanna	0000 UTC 6 Sep–1800 UTC 7 Sep 2008
Hurricane Ike	0600 UTC 13 Sep–0000 UTC 15 Sep 2008

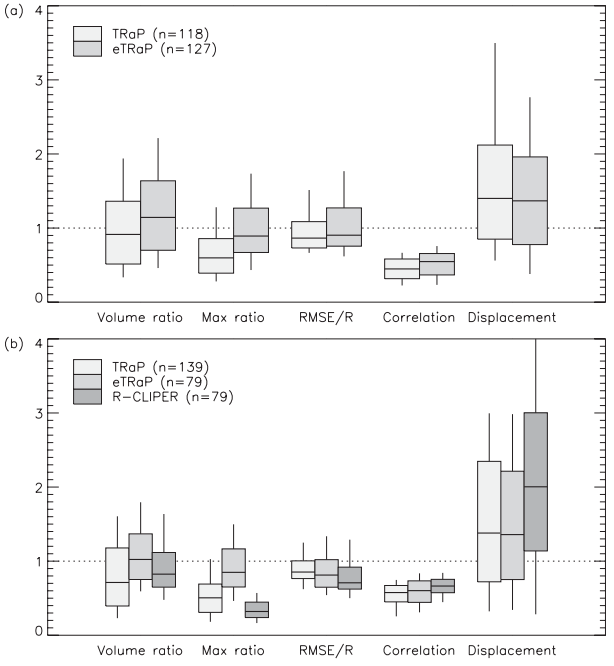

FIG. 9. Distribution of CRA verification results for the 16 Atlantic storms and hurricanes in Table 1 for (a) 6-h accumulations for TRaP and eTRaP forecasts, and (b) 24-h accumulations for TRaP, eTRaP, and R-CLIPER forecasts. The boxes indicate the 25th, 50th, and 75th percentiles of the distribution, and the vertical lines show the 10th and 90th percentiles.

translation (see Ebert and McBride 2000 for details). To ensure that the observed field was sufficiently well sampled, a requirement of at least 2000 matched grid boxes was imposed.

The benefit in QPF accuracy gained by using an ensemble can be seen in Fig. 9, which shows performance statistics for 6- and 24-h rainfall accumulations predicted by TRaP, eTRaP, and R-CLIPER for those dates and times in which all products were available. The 6-h results were combined for all lead times (6, 12, 18, and 24 h). The distribution of the results gives users an idea of the level of accuracy that can be expected from the forecasts. Importantly, the eTRaP maximum rain was closer to the observed value than the TRaP and R-CLIPER maximum rain. The eTRaP rain volumes also improved on the TRaP estimates, with median values 15% greater than observed for 6-h accumulations, and nearly unbiased for 24-h accumulations.

The median root-mean-square errors were slightly less than the magnitude of the mean observed rain for both eTRaP and TRaP. The spatial correlation coefficients for heavy rain forecasts had median values of 0.5–0.6 for the two products, with eTRaP having a slight advantage over TRaP due to the smoothing effects of the ensemble averaging of many members. The smooth R-CLIPER forecasts performed still better in terms of RMSE and correlation.

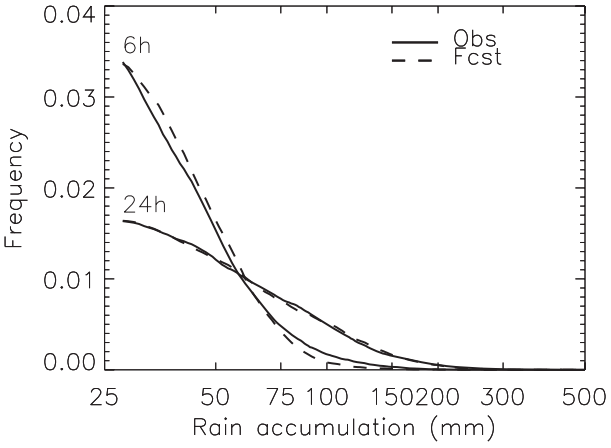

FIG. 10. Distribution of heavy rainfall for 6- and 24-h accumulation periods as forecast by eTRaP (dashed lines) and measured by the stage IV radar–gauge analysis (solid lines).

The rain location errors were typically 70–230 km, with median values of 137 (136) km for eTRaP 6-h (24 h) accumulations and 141 (138) km for TRaP 6-h (24 h) accumulations. In comparison, 24-h hurricane track forecasts made by NHC during the same period averaged about 100 km (Franklin 2010). The eTRaP rain location error can be expected to exceed the forecast track error since the satellite-based product does not account for the rotation and evolution of the rain pattern. However, by combining multiple realizations of TRaP, and therefore smoothing the final field, eTRaP reduces the rain location error slightly over that found for TRaP. The estimated location errors for R-CLIPER rain fields were substantially greater, as its smoother representation of the rain field made the pattern matching with the observations less responsive to the rain maximum.

The remainder of the verification (other than rain location error) was performed for a track-following domain of dimension 400 × 400 grid boxes centered on the best-track position of the storm as determined by NHC.

For hydrological applications it is important to check that the forecast produces a realistic rain intensity distribution (e.g., Marchok et al. 2007). Figure 10 shows the forecast and observed distributions for 6- and 24-h accumulations aggregated over the entire verification period. The forecast intensity distribution matches the observed distribution extremely well, demonstrating that the eTRaP rain amounts are not conditionally biased.

The eTRaP spatial probability forecasts for precipitation exceeding various thresholds were verified using reliability diagrams, which measure the bias in the predicted probabilities, and relative operating characteristic (ROC) diagrams, which measure the ability of the forecast to discriminate between observed events and nonevents (e.g., Jolliffe and Stephenson 2003). These diagnostics were

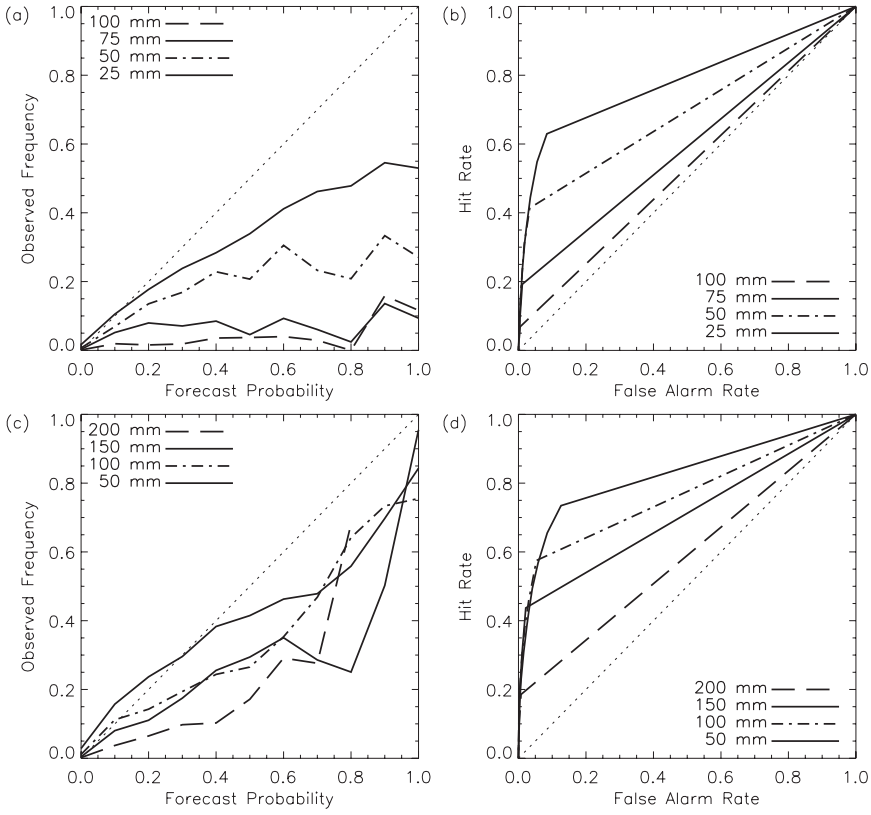

FIG. 11. (left) Reliability and (right) ROC diagrams for (a) 6- and (b) 24-h TRaP and eTRaP probabilistic precipitation forecasts for the storms listed in Table 1 for amounts of 25–100 mm. (c),(d) As in (a),(b) but for 50–200 mm.

created by matching the grid-scale forecast probabilities with the observed occurrence of rain exceeding various thresholds. The ROC diagram is constructed by converting probabilistic forecasts into binary (yes–no) forecasts using a sequence of probability thresholds, then plotting the hit rate against the false alarm rate for each of those thresholds. The results are shown in Fig. 11, where the four curves in each plot correspond to the four exceedance thresholds used in the probability forecasts.

The ROC diagrams confirm that the eTRaP precipitation probabilities have useful skill, with hit rates much greater than false alarm rates (curve located well to the left of the diagonal). The forecast skill is greater for lighter rainfall. The reliability diagrams show that the forecast probabilities for 6- and 24-h accumulations were most reliable (closest to the observed frequency) for lighter thresholds, and reliability was better for longer accumulations. For most probabilities and thresholds the grid-scale forecast probabilities were biased high; that is, they were greater than the observed frequencies.

The presence of bias in the probability forecasts means that it would be unwise for users to simply interpret the probabilities at face value. Some form of calibration, or else an informed interpretation, of these forecasts is clearly required to make the best use of the probability forecasts.

b. *Radius of uncertainty*

A common approach for calibrating probability forecasts is to reassign the forecast values to the corresponding observed frequency (e.g., a 80% probability of 6-h rain exceeding 25 mm in a particular grid box would be reassigned to 50% probability, based on the curve in Fig. 11). For some thresholds this calibration may prevent high probabilities from being predicted at all; for example, the maximum calibrated probability for 6-h rain exceeding 100 mm would be less than 20%. This type of calibration is quite appropriate for probabilities at grid scale as it reflects the uncertainty for *that particular location*. However, it may not reflect the needs of many users for information at broader scales. In fact, calibrated grid-scale probabilities could be misinterpreted as low probabilities of heavy rain *in the storm*, with potentially dangerous consequences.

We postulate that the overestimates in forecast probability are likely to be due largely to errors in the location of the predicted heavy rain (recall that location errors

were typically on the order of 100 km), rather than forecast rain rates that are too great. Forecasters understand that storm forecasts and associated heavy rain predictions are likely to contain location and timing errors, and usually interpret the forecast liberally (i.e., in the vicinity of the predicted place and time) rather than literally.

Within the context of tropical cyclones the uncertainty in location is sometimes conveyed using a ''strike probability'' to convey the likelihood of a cyclone center passing within a certain radius of a location, or a ''cone of uncertainty'' within which there is a specified likelihood of the cyclone track occurring (Broad et al. 2007). We propose a similar approach for interpreting heavy rain probability forecasts, that is, to indicate a ''radius of uncertainty'' at which the forecast probabilities are expected to be reliable. The interpretation of the forecast would then be, ''When the forecast for a particular location gives a probability *P* of exceeding a given threshold, over many such cases a fraction *P* will contain observed rain exceeding the threshold *somewhere within* ROU *km of the location of interest*,'' where ROU is the radius of uncertainty. Standard spatial forecasts have ROU = 0; that is, they make no allowance for location error. Here, ROU can be thought of as a spatial allowance or uncertainty, where we anticipate that larger allowances will be required for higher rain intensities that are more difficult to predict, and higher probabilities that are spatially more precise and therefore more prone to location error. Although approached within the context of calibration, ROU could also be treated as a validation metric, with larger values signifying poorer forecast performance.

To test this approach, the eTRaP probability forecasts were reverified using succeedingly generous spatial allowances for a hit (i.e., within one grid box of the observed rain, within two grid boxes, within three grid boxes, etc.). The value of ROU was determined to be that spatial allowance that resulted in the greatest reliability (best match of forecast probabilities with the observed frequency) and provides guidance on how the forecast should be interpreted.

The uncertainty radii for eTRaP probabilistic predictions for 6- and 24-h heavy rain accumulations are given in Table 2. The lighter rain thresholds and lower probabilities permit a stricter spatial interpretation of the forecast; for example, a forecast of 30% probability of 24 h rain exceeding 50 mm can be considered reliable at grid scale. For heavier thresholds and higher probabilities, where observed areas of heavier rain become smaller and more difficult to predict, the spatial interpretation of the probability forecasts becomes broader (e.g., within 80 km for a 50% probability of 6-h rainfall exceeding 100 mm). Until such time as calibrated eTRaPs become available, either through improvements to TRaP technology or some form of postprocessing, this radius of uncertainty approach for interpreting probabilities should provide useful information for forecast users.

TABLE 2. Radii (within) of uncertainty, ROU (to nearest 4 km), for interpreting the eTRaP probability of precipitation fields, as a function of accumulation period, rain exceedance threshold, and forecast probability. A value of 4 km indicates grid scale, 8 km indicates within ±1 grid box, etc. Blank entries indicate an inability to produce a meaningful interpretation due to insufficient sample size (<10).

Rain threshold (mm)	Forecast probability for 6-h accumulation				
	0.1	0.3	0.5	0.7	0.9
25	4 km	8 km	16 km	32 km	64 km
50	8 km	16 km	32 km	42 km	80 km
75	16 km	20 km	80 km		
100	52 km	52 km	80 km		
Rain threshold (mm)	Forecast probability for 24-h accumulation				
	0.1	0.3	0.5	0.7	0.9
50	4 km	4 km	12 km	24 km	32 km
100	4 km	12 km	20 km	28 km	32 km
150	8 km	20 km	20 km	28 km	28 km
200	16 km	32 km	24 km	28 km	

5. Discussion

Ensemble TRaP provides predictions of 6- and 24-h rainfall amounts and probabilities of exceeding various thresholds, for rain in landfalling tropical storms and cyclones. The eTRaP QPFs are more accurate than single-sensor TRaP forecasts in all important aspects: maximum rainfall amount, rain volume, spatial pattern, rain intensity distribution, and location. Importantly, eTRaP provides probabilistic forecasts for decision makers, though some spatial interpretation of the probabilities is required.

Many improvements can be made to eTRaP. Most importantly, calibration of the rain probability forecasts is clearly required. Future eTRaP development will calibrate the forecasts both at the grid scale, as is done with rain forecasts from numerical ensemble prediction systems (e.g., Eckel and Walters 1998), and also for a specified radius of influence (e.g., 40 km), as is done with severe weather hazards at the Storm Prediction Center (SPC 2010). Consultations with users of eTRaP forecasts have indicated that both approaches would be valuable.

Another enhancement is to include additional types of rainfall forecasts in the ensemble. For example, Kuligowski et al. (2006) demonstrated that TRaP could be constructed from Hydro-Estimator (H-E) rainfall estimates based on geostationary infrared observations. The spatial and temporal resolutions of the geostationary data are much finer than those for passive microwave data, offering more detailed rainfall estimates and potentially very large ensembles. Kuligowski et al. (2006) found that spatial and

temporal averaging of the H-E rainfall estimates prior to extrapolation in the TRaP process improved the quality of the forecasts. Judicious use of H-E TRaPs may not only improve the ensemble by providing independent rainfall estimates, it would provide data during time periods where few microwave TRaPs are available.

It would be possible to add numerical weather prediction (NWP) model output to the ensemble. Deterministic and ensemble model forecasts are easily available from the Global Forecast System (GFS) model and have been shown to be skillful for predicting heavy rain in tropical cyclones (Marchok et al. 2007; Mullen and Buizza 2002). NWP has the advantage that its forecasts extend out much longer than 24 h. ETRaP forecasts could be blended into model forecasts in order to make time series products for locations at risk.

R-CLIPER provides an independent rainfall estimate that is smoother and less prone to finescale errors than TRaP. Lonfat et al. (2007) describe an improved version of R-CLIPER called the Parametric Hurricane Rainfall Model (PHRaM) that includes the effects of topography and vertical shear. We plan to include R-CLIPER forecasts in future implementations of eTRaP.

Some improvements to eTRaP will involve modifications to the TRaP forecasts themselves. Topographic enhancement of TRaP land-based rainfall estimates could be included to increase the rainfall in upslope flow and reduce it in downslope flow (e.g., Vicente et al. 2002). For cyclones making landfall in mountainous regions such as Central America, some Caribbean islands, and many parts of Asia, this may be an important factor in better estimating the maximum rainfall.

The existing TRaP scheme does not include storm rotation; this would be a valuable improvement and would increase the physical realism of the forecasts. Liu et al. (2008) included estimates of storm rotation from geostationary cloud drift winds the TRaP extrapolation forecast and found that this enhancement reduced the mean absolute errors by 40% compared to original TRaP forecasts for tropical cyclone rainfall over Taiwan. Low-level cyclone wind fields from AMSU and other data sources (Knaff and DeMaria 2006) will soon be available to NESDIS, from which it will be possible to estimate storm rotation rates. Alternatively, parametric vortex specification (e.g., Davidson and Weber 2000) could be used to construct a circulation to ameliorate this aspect of the steady-state assumption.

Finally, the concept of a radius of uncertainty is a novel approach for interpreting probabilistic forecasts in the presence of location error. It makes allowance for spatial uncertainty by defining a distance within which at least one observed event is expected to be recorded with the same frequency as the forecast probability. The radius of uncertainty increases with an increasing rain intensity threshold and increasing forecast probability. This dependence on both intensity and forecast probability, though not surprising, has a disadvantage relative to simple probability-reassignment-type calibration methods since it requires a new way of thinking, and a table of ROU values is required to interpret the full spatial forecast. However, a major advantage is that this type of calibration does not eliminate high probabilities; rather, it redefines them to refer to a spatial neighborhood instead of only the point of interest. This reduces the risk of users misinterpreting calibrated low probabilities to mean that no heavy rain is likely anywhere.

Acknowledgments. This work was funded by the National Oceanic and Atmospheric Administration's Satellite Product and Services Review Board. Several external TRaP users encouraged the development of ensemble TRaP. We are grateful to George Serafino and Tom Schott for helping get this project off the ground, and to Ying Lin for help with the stage IV analyses used to verify eTRaP. John McBride, Noel Davidson, and the three anonymous reviewers provided many helpful comments to improve the manuscript.

REFERENCES

Broad, K., A. Leiserowitz, J. Weinkle, and M. Steketee, 2007: Misinterpretations of the "cone of uncertainty" in Florida during the 2004 hurricane season. *Bull. Amer. Meteor. Soc.,* **88,** 651–667.

Clemen, R. T., 1989: Combining forecasts: A review and annotated bibliography. *Int. J. Forecasting,* **5,** 559–583.

Davidson, N. E., and H. C. Weber, 2000: The BMRC high-resolution tropical cyclone prediction system: TC-LAPS. *Mon. Wea. Rev.,* **128,** 1245–1265.

Ebert, E. E., 2001: Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Mon. Wea. Rev.,* **129,** 2461–2480.

——, and J. L. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *J. Hydrol.,* **239,** 179–202.

——, S. Kusselson, and M. Turk, 2005: Validation of NESDIS operational tropical rainfall potential (TRaP) forecasts for Australian tropical cyclones. *Aust. Meteor. Mag.,* **54,** 121–135.

Eckel, F. A., and M. K. Walters, 1998: Calibrated probability quantitative precipitation forecasts based on the MRF ensemble. *Wea. Forecasting,* **13,** 1132–1147.

Ferraro, R., and Coauthors, 2005: The tropical rainfall potential (TRaP) technique. Part II: Validation. *Wea. Forecasting,* **20,** 465–475.

Franklin, J. L., 2010: Tropical cyclone forecast verification at the National Hurricane Center. Preprints, *20th Conf. on Probability and Statistics in the Atmospheric Sciences,* Atlanta, GA, Amer. Meteor. Soc., 6.2. [Available online at http://ams.confex.com/ams/pdfpapers/160383pdf.]

Jolliffe, I. T., and D. B. Stephenson, 2003: *Forecast Verification. A Practitioner's Guide in Atmospheric Science*. Wiley and Sons, 240 pp.

Kidder, S. Q., S. J. Kusselson, J. A. Knaff, R. R. Ferraro, R. J. Kuligowski, and M. Turk, 2005: The tropical rainfall potential (TRaP) technique. Part I: Description and examples. *Wea. Forecasting,* **20,** 456–464.

Knaff, J. A., and M. DeMaria, 2006: A multi-platform satellite tropical cyclone wind analysis system. Preprints, *14th Conf. on Satellite Meteorology and Oceanography,* Atlanta, GA, Amer. Meteor. Soc., P4.9. [Available online at http://ams.confex.com/ams/pdfpapers/99945.pdf.]

Kuligowski, R. J., S. Kusselson, M. Turk, and E. Ebert, 2006: Evaluation of tropical rainfall potential (TRaP) forecasts for the US during 2004 and 2005. *Second Int. Symp. on Quantitative Precipitation Forecasting and Hydrology,* Boulder, CO, WMO–NCAR, P7.3.

Lin, Y., and K. Mitchell, 2005: The NCEP stage II/IV hourly precipitation analyses: Development and applications. Preprints, *19th Conf. on Hydrology,* San Diego, CA, Amer. Meteor. Soc., 1.2. [Available online at http://ams.confex.com/ams/pdfpapers/83847.pdf.]

Liu, G.-R., C.-C. Chao, and C.-Y. Ho, 2008: Applying satellite-estimated storm rotation speed to improve typhoon rainfall potential technique. *Wea. Forecasting,* **23,** 259–269.

Lonfat, M., R. Rogers, T. Marchok, and F. D. Marks Jr., 2007: A parametric model for predicting hurricane rainfall. *Mon. Wea. Rev.,* **135,** 3086–3097.

Marchok, T., R. Rogers, and R. Tuleya, 2007: Validation schemes for tropical cyclone quantitative precipitation forecasts: Evaluation of operational models for U.S. landfalling cases. *Wea. Forecasting,* **22,** 726–746.

Mullen, S. L., and R. Buizza, 2002: The impact of horizontal resolution and ensemble size on probabilistic forecasts of precipitation by the ECMWF Ensemble Prediction System. *Wea. Forecasting,* **17,** 173–191.

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

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.

Spampata, M., A. Schwartz, M. Turk, S. Kusselson, and R. Edmundson, 2008: Validation of the tropical rainfall potential (TRaP). Preprints, *22nd Conf. on Hydrology,* Amer. Meteor. Soc., JP1.3. [Available online at http://ams.confex.com/ams/pdfpapers/53370.pdf.]

SPC, cited 2010: SPC probabilistic outlook information. [Available online at http://www.spc.noaa.gov/products/outlook/probinfo.html.]

Tuleya, R. E., M. DeMaria, and R. J. Kuligowski, 2007: Evaluation of GFDL and simple statistical model rainfall forecasts for U.S. landfalling tropical cyclones. *Wea. Forecasting,* **22,** 56–70.

Vicente, G. A., J. C. Davenport, and R. A. Scofield, 2002: The role of orographic and parallax corrections on real time high resolution satellite rain rate distribution. *Int. J. Remote Sens.,* **23,** 221–230.

Vila, D., R. Ferraro, and R. Joyce, 2007: Evaluation and improvement of AMSU precipitation retrievals. *J. Geophys. Res.,* **112,** D20119, doi:10.1029/2007JD008617.

Woodcock, F., and C. Engel, 2005: Operational consensus forecasts. *Wea. Forecasting,* **20,** 101–111.

——, and D. J. M. Greenslade, 2007: Consensus of numerical model forecasts of significant wave heights. *Wea. Forecasting,* **22,** 792–803.