VALIDATION OF THE ENSEMBLE TROPICAL RAINFALL POTENTIAL (e-TRaP) FOR LANDFALLING TROPICAL CYCLONES

Elizabeth E. Ebert*
Bureau of Meteorology Research Centre (BMRC), Melbourne, VIC, Australia

Anthony E. Salemi*, Michael Turk[†], Michelle Spampata*, and Sheldon Kusselson[†]
[†]Satellite Services Division (NOAA/NESDIS), Washington, D.C., *Riverside Technology, inc.

1. INTRODUCTION

NESDIS has been producing operational areal Tropical Rainfall Potential (TRaP) forecasts of rainfall for landfalling tropical cyclones since the early 2000's. TRaP forecasts are 24-hour precipitation forecasts based on along-track extrapolation of satellite-estimated rain rates. These are derived from passive microwave sensors onboard NOAA’s Advanced Microwave Sounder Unit, Defense Meteorological Satellite Program’s (DMSP) Special Sensor Microwave Imager (SSM/I), and NASA's Tropical Rainfall Monitoring Mission’s Microwave Imager (TMI) and Advanced Microwave Scanning Radiometer (AMSR-E)**.** Experimental TRaPs from the operational NESDIS Hydro-Estimator (H-E), which bases rainfall estimates on infrared data from geostationary satellites, have been made for US hurricanes starting in 2004.

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 minutes a new position is calculated and the spatial rain rates applied over a rectangular grid of approximately 4 km resolution; the 15-minute accumulations are summed over a period of 24 hours (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 US and Australia, respectively, have shown that in general the TRaPs do a reasonable job of estimating both the maximum rainfall accumulation and its spatial distribution but underestimate the total rain volume by about 1/3 in both regions. The overall accuracy is similar to that of regional NWP models, and depends to some extent on the sensor being used, with AMSU and TRMM-derived TRaPs tending to perform better than SSM/I TRaPs. Kuligowski (2006, personal communication) compared the performance of H-E TRaPs to that of passive microwave TRaPs and found both data sources provided forecasts of similar quality.

These validation studies have suggested that the errors in TRaP forecasts are more likely to be related to errors in satellite rain rates and the assumption of steady state rainfall than to errors in operational track predictions. While there is some systematic error in the TRaPs (e.g., underestimation of rain volume), the variation in TRaP performance from storm to storm, and indeed among different TRaPs for a single storm, is very large. This large random error component 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 kind of poor man's ensemble. This has the effect of smoothing the rain field, with 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 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 comprising the ensemble. One can generate probabilistic forecasts of rain exceeding certain critical thresholds in locations of interest, an approach 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

* *Corresponding author address:* Elizabeth E. Ebert, CAWCR, GPO Box 1289, Melbourne, VIC 3001. AUSTRALIA: e-mail: e.ebert@bom.gov.au

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 h, 12 h, and 24 h accumulations.

In principle an ensemble TRaP (abbreviated 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 with good precision and reliability.

2. GENERATION OF ENSEMBLE TRAP FORECASTS

Figure 1 illustrates schematically how eTRaPs are generated. eTRaP products include both deterministic and probabilistic quantitative precipitation forecasts (QPFs and PQPFs) generated from weighted ensemble members, where the weights indicate the expected relative accuracy. TRaPs are assigned to the nearest synoptic time of 00, 06, 12, or 18 UTC, which means that they are at most 3 hours offset in time. All forecasts were remapped onto a regular 0.25° latitude/longitude grid prior to combination.

6 h TRaP rainfall accumulations, or segments, can be combined into 12 h and 24 h accumulations as needed. The number of ensemble members in the eTRaP is the number of permutations possible for combining the various segments of the forecast. For example, if there are five 6-hour TRaPs available for the first six hours of a 24 h forecast, four TRaPs for the second six hour segment, four TRaPs for the third segment, and two TRaPs for the fourth segment, then the number of ensemble members comprising the 24 h forecast is 5 x 4 x 4 x 2 = 160.

Every 6 h TRaP contributing to the ensemble is weighted according to its expected accuracy. The weight assigned to the *i*th TRaP forecast, w_i, is the product of its sensor weight and its forecast latency weight, $w_i = w_{sensor} \times w_{latency}$.

The sensor weights are based on the validation results of Ferraro et al. (2005) and Ebert et al. (2005), who investigated the dependence of TRaP performance on the satellite sensor used to derive the TRaP during the 2002 Atlantic hurricane season and the 2003-04 Australian tropical cyclone season. The relative weights are proportional to the inverse of the mean squared error (MSE) for TRaPs from each sensor.

$$w_{sensor} = \left[\frac{MSE_{sensor}}{\overline{MSE}} \right]^{-1} \qquad (1)$$

where the sensor is either AMSU, TRMM, or SSM/I. $\overline{MSE}$ is the average MSE over all sensors. This approach yielded sensor weights of 1.3, 1.0, and 0.7 for AMSU, TRMM, and SSM/I-based TRaPs, respectively. The accuracy of AMSRE TRaPs has not been determined yet, so a value of 1.0 was assigned.

The weights for forecast latency were subjectively assigned, with the most recent 6 h segments receiving the most weight and the oldest 6 h segments receiving the least (Table 1). This reflects the expectation that steady state rainfall is a more valid assumption early in the forecast period than later.

Table 1. Forecast latency weights, $w_{latency}$, used in computing eTRaP.

Forecast latency	Weight
0 h	1.0
6 h	0.7
12 h	0.4
18 h	0.1

In the case of 12h and 24 h eTRaP, the rain accumulation for each ensemble member is simply the sum of the rain in its 6 h TRaP segments. Its weight is the sum of the weights for the segments.

In many cases a large number of 6 h TRaPs 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 give useful information on track uncertainty. If, after this step is taken, the number of potential ensemble members still exceeds 1000, the permutations are randomly culled to reduce the number to about 1000, to speed computing time.

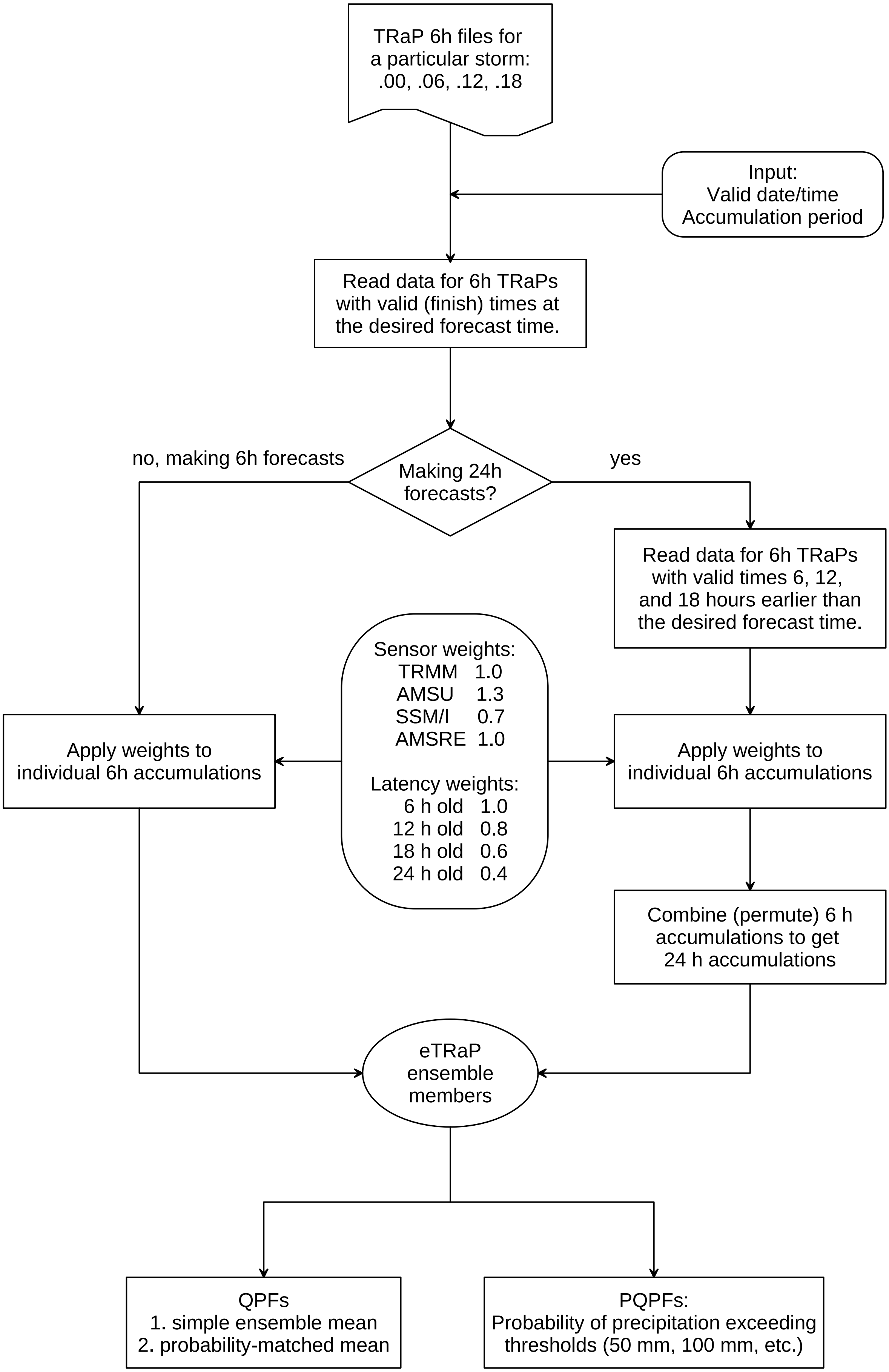

Figure 1. Steps in the generation of 24 h ensemble TRaP forecasts.

The ensemble mean is simply the weighted average of the ensemble members at every grid box in the domain. It is preferable to use the probability-matched ensemble mean, which has the same relative spatial distribution of rain as the ensemble mean but the intensity distribution is transformed using probability (histogram) matching to have the same intensity distribution as the full ensemble. The purpose of this transformation is to remove the excess light rain caused by the averaging process, and to restore the heavy rain accumulations that may have been lost during averaging. For heavy rain events this can be an important correction (Ebert, 2001). The weights are applied during the histogram matching as well as during the calculation of the ensemble mean.

Probabilistic forecasts are also weighted to give greater influence to the ensemble members with greater expected accuracy:

$$P(R \geq R_T) = \frac{\sum_{i=1}^{n} w_i I_i}{\sum_{i=1}^{n} w_i}, \quad I = \begin{cases} 0 & R_i < R_T \\ 1 & R_i \geq R_T \end{cases} \qquad (2)$$

where R_i is the areal rainfall of the ith (possibly summed) TRaP forecast and R_T is a threshold rain amount. The thresholds chosen for computing probabilistic forecasts are 25, 50, 75, and 100 mm for 6 h and 12 h forecasts, and 50, 100, 150, and 200 mm for 24 h forecasts.

The eTRaP products are displayed as maps of rain amount and probability. It would be possible to issue meteograms showing the expected time evolution of the rainfall at a particular location, but these would have limited usefulness since eTRaPs forecasts have a maximum lead time of 24 h with 6 h time steps.

3. ENSEMBLE TRAPS FOR HURRICANE RITA 24 SEPTEMBER 2005

Hurricane Rita was one of the most intense tropical cyclones ever observed 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 on 24 September 2005.

Figure 2 shows the 6 h TRaP segments available within ±3 hours of 0000 UTC on 24 September 2005, that were used to construct

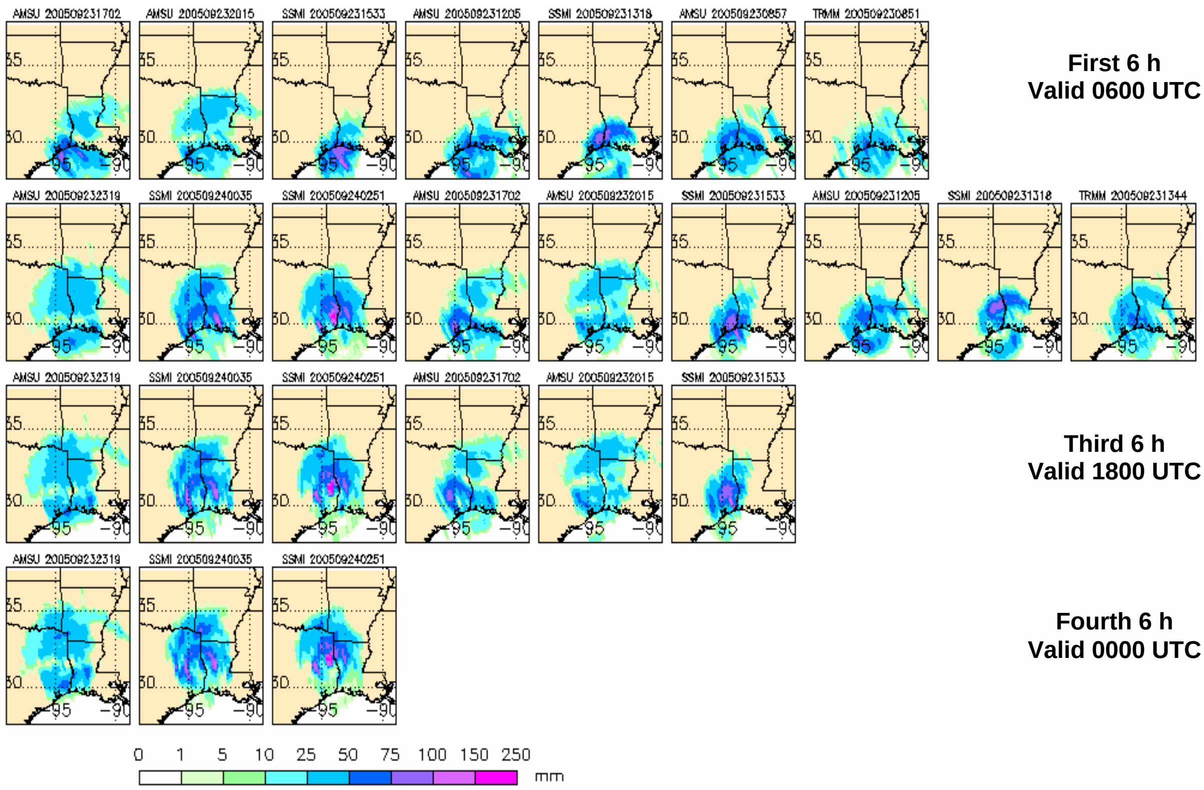

Figure 2. 6 h TRaP segments available within 3 hours of 0000 UTC on 24 September 2005 for Hurricane Rita, from which can be generated 6 h, 12 h, and 24 h eTRaP forecasts valid at 0600 UTC 24 September 2005, 1200 UTC 24 September 2005, and 00 UTC 25 September 2005, respectively.

ensemble TRaPs for the subsequent 24 h period. For the first six hour period seven TRaP segments were available, ranging from the most recent to nominally eighteen hours old. The second six hour period had nine segments with latencies of up to twelve hours. The third period had six segments, and the final six hour period had only three segments since only the most recent forecasts would have been available. Although 1134 permutations were potentially available to generate 24 h ensemble forecasts, the ensemble size was culled to 1000 members.

One thing to note 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 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 3 to 5 show 6 h, 12 h, and 24 h eTRaP forecasts made using the 6 h TRaP segments shown in Fig. 2. Corresponding observed rainfall accumulations from the National Centers for Environmental Prediction's Stage IV radar-rain gauge analyses (Lin and Mitchell, 2005), also remapped to a 0.25° grid, are shown in Figure 6.

The 6 h probability matched ensemble mean (PM QPF) 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. 3). These correspond quite well to the observed structure of the rainfall as shown in the Stage IV data (Fig. 6a). However, the predicted maximum rainfall of 130 mm was well under the observed value of 200 mm. The probability maps showed >50% chance of exceeding 50 mm along the Texas-Louisiana border, but did not predict any precipitation accumulation (on land) exceeding

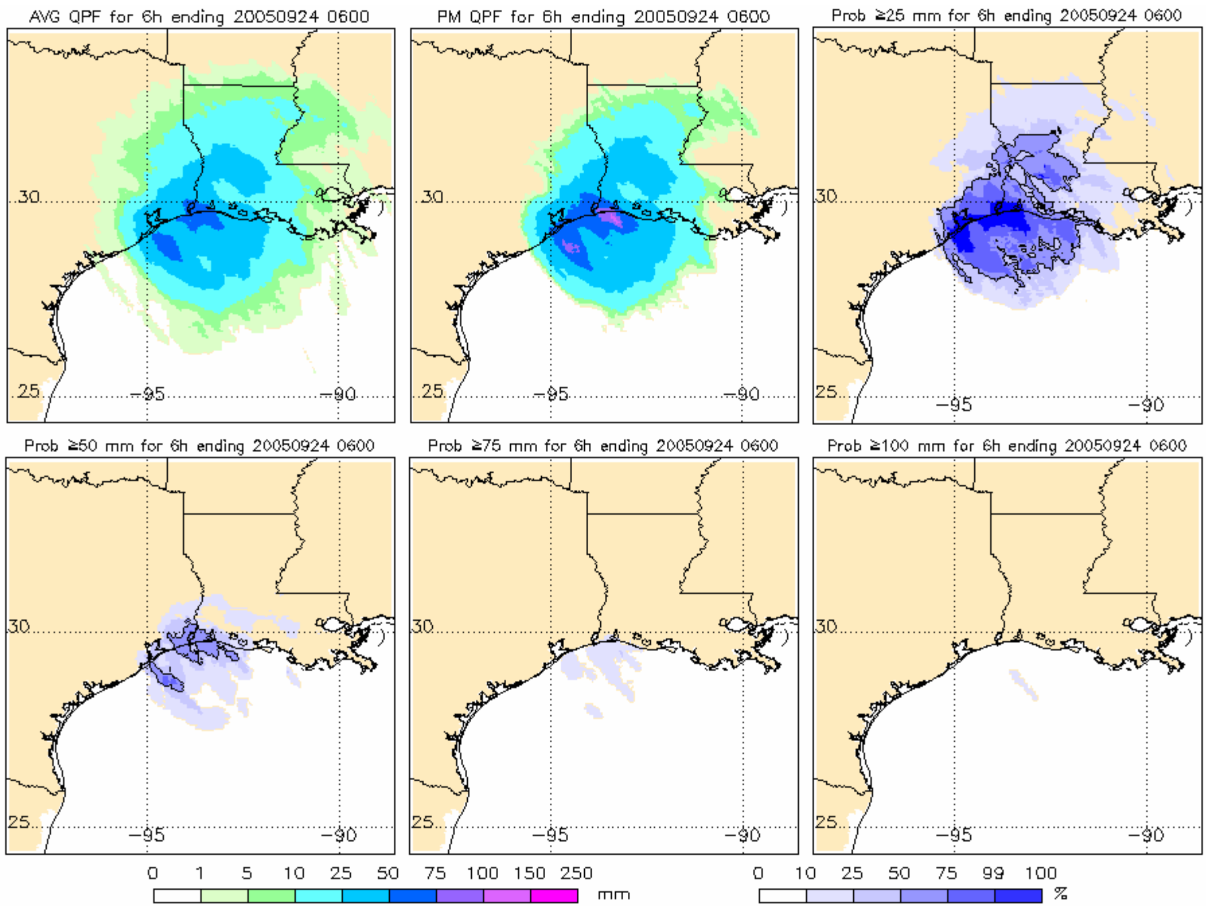

Figure 3. 6 h eTRaP forecast for rain in Hurricane Rita ending at 0600 UTC on 24 September 2005. Nine TRaPs contributed to this ensemble (see Fig. 2). The first two panels show the simple ensemble mean and probability matched ensemble mean QPFs, while the remaining four panels show probabilities of 6 h precipitation exceeding 25, 50, 75, and 100 mm.

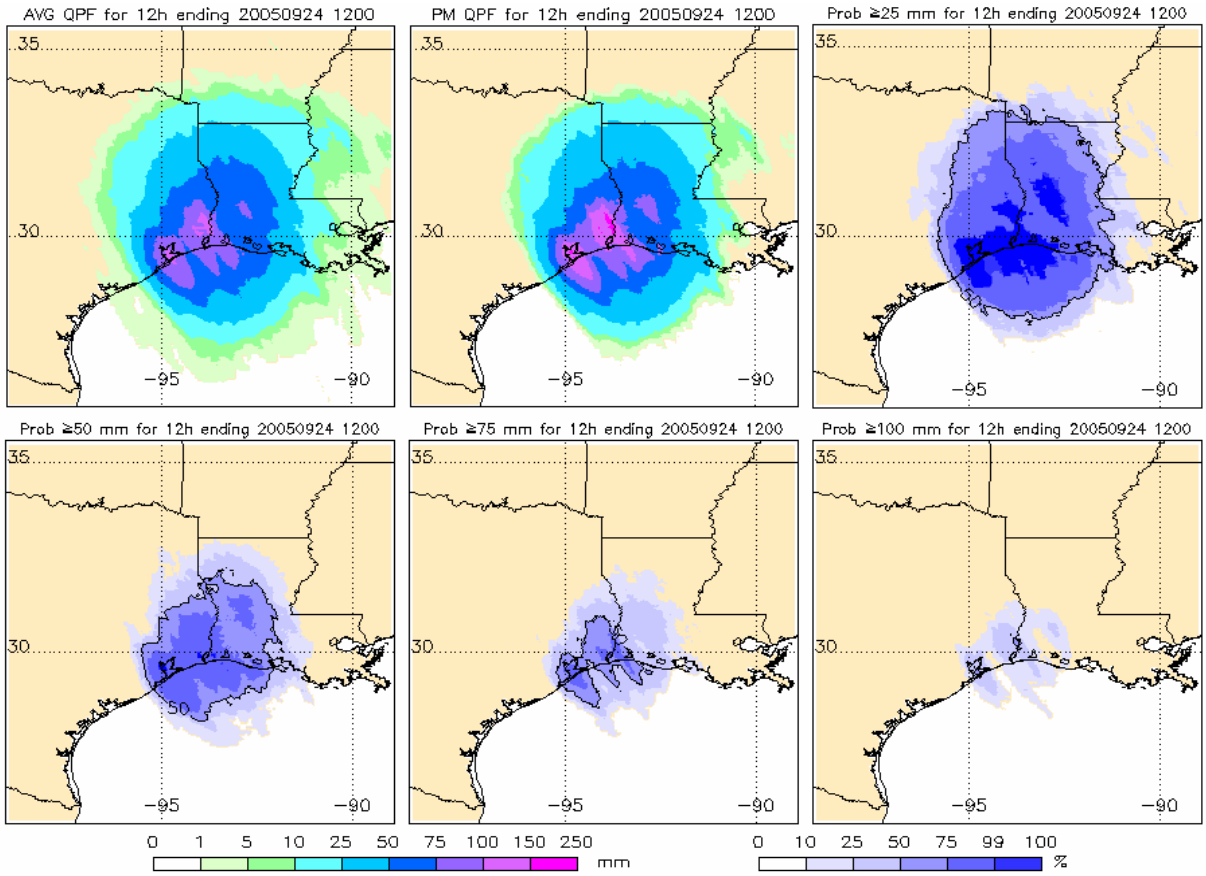

Figure 4. As in Fig. 3, for the 12 h eTRaP forecast valid at 1200 UTC on 24 September 2005 (130 members).

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 12 h and 24 h eTRaP forecasts showed excellent placement of the rain maximum, but did not capture the observed heavy rain in eastern Louisiana. The 12 h PM QPF underestimated the rain maximum by about 30% while the 24 h forecast captured the total rain amount quite accurately. The probability contours were smoother than for the 6 h case due to the much larger ensemble size. The non-zero probability of rain exceeding 100 mm corresponded well to the areas in which rain greater than 100 mm was observed, and in the 24 h forecast a region of greater than 25% probability of exceeding 200 mm lay directly over the observed elongated rain band in eastern Texas.

This example shows that ensemble TRaP would have provided very useful deterministic and probabilistic guidance for heavy rainfall in Hurricane Rita. In the next section we perform a quantitative verification of eTRaP forecasts using a much larger sample of storms.

4. PERFORMANCE OF ETRAP FOR ATLANTIC STORMS DURING 2004-2008

To obtain a more quantitative evaluation of ensemble TRaP and compare its performance to that of (single-sensor) TRaP, forecasts for eighteen Atlantic tropical storms and hurricanes were verified against Stage IV observations. The storms and relevant dates are given in Table 2. The forecasts were verified within a track-following domain of size 10° x 10° centered on the forecast track position, and was done on the grid of the Stage IV product, about 4 km. In order for the verification of a forecast to be counted a minimum of 100 valid points for 6 h accumulations (400 for 24 h accumulations) were required in the observed field. The aggregate statistics were computed by weighting each verification according to its sample size. The large number of samples and independent storms suggests that the verification results should be robust.

Table 2. Atlantic tropical storms and hurricanes for which TRaP and eTRaP forecasts were verified against Stage IV radar-gauge analyses. 423 eTRaP and 911 TRaP forecasts were verified for 6 h accumulation periods; 145 eTRaP and 343 TRaP forecasts were verified for 24 h accumulation periods.

Table 2

Storm	**Month**
Bonnie	Aug 2004
Charley	Aug 2004
Frances	Sept 2004
Ivan	Sept 2004
Arlene	June 2005
Cindy	July 2005
Katrina	Aug 2005
Rita	Sept 2005
Barry	June 2007
Erin	Aug 2007
Gabrielle	Sept 2007
Humberto	Sept 2007
Dolly	July 2008
Edouard	Aug 2008
Fay	Aug 2008
Gustav	Aug-Sept 2008
Hanna	Sept 2008
Ike	Sept 2008

Table 3 shows some average performance statistics for 6 h and 24 h rainfall accumulations predicted by eTRaP and TRaP. The 6 h results were combined for all lead times (6, 12, 18, and 24 h). For both 6 h and 24 h accumulations the eTRaP maximum rain was closer to the observed value than the TRaP maximum rain. The rain volumes were too low by about 10% for 6 h accumulations, and nearly perfect for 24 h accumulations. The root mean square error (RMSE) was similar for the two products for 6 h accumulations, while for 24 h accumulations the eTRaP RMSE was more than 10% lower than that for TRaP, due to the smoothing effect of the ensemble averaging of many members. The correlation coefficients for eTRaP were higher than for TRaP for 6 h accumulations, but not significantly so for 24 h accumulations.

Categorical statistics were also computed for a number of rain thresholds to evaluate how well eTRaP and TRaP QPFs predicted the occurrence of rain of various intensities. Figure 7 shows the performance of eTRaPs and TRaPs for rain exceeding thresholds ranging from 12.7 mm (1/2 inch) to 228 mm.

The frequency bias, which measures the ratio of the forecast to observed rain frequency, was close to the perfect value of 1 for rain exceeding the lighter thresholds (12.7-25 mm) for both 6 h and 24 h accumulations. As the

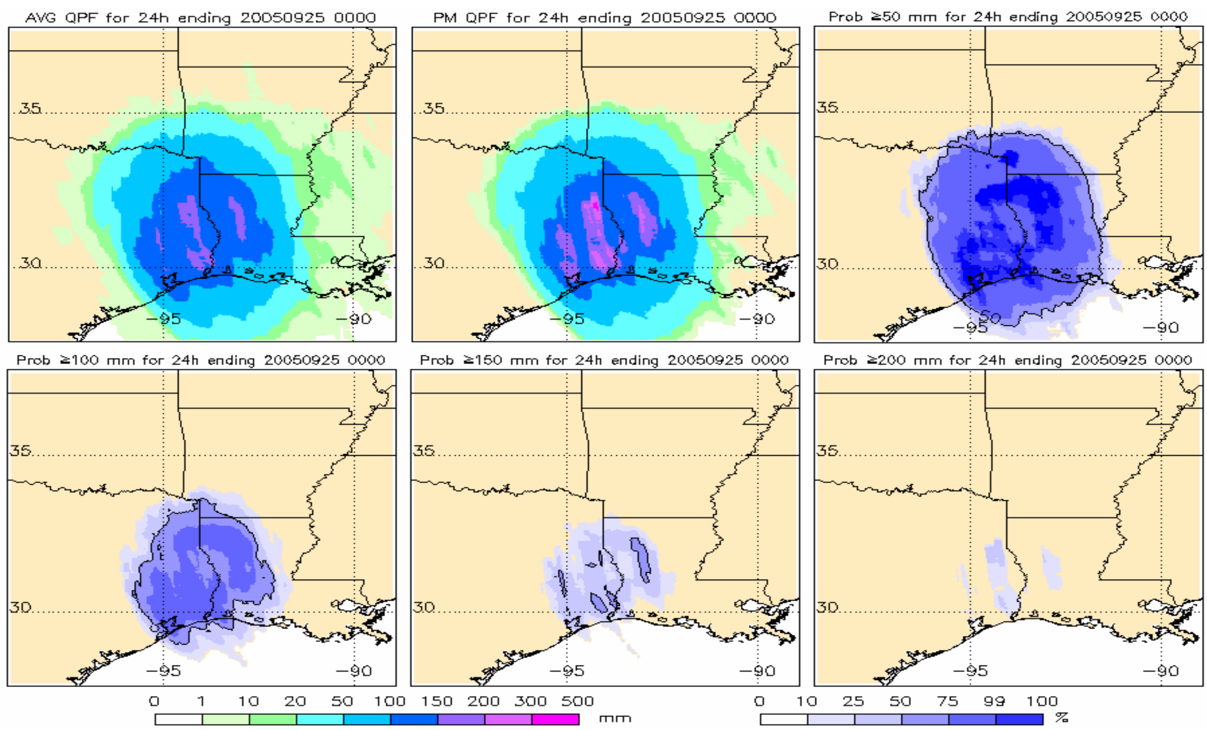

Figure 5. As in Fig. 3, for the 24 h eTRaP forecast valid at 0000 UTC on 25 September 2005 (1000 members). Note that the probability contours and color scale differs from those used in Figs. 2-4.

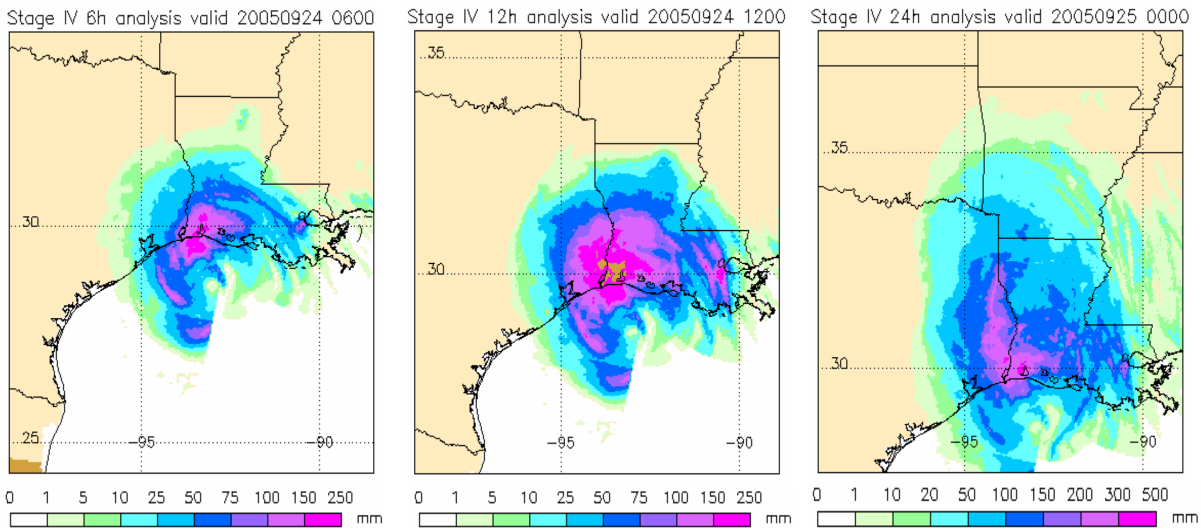

Figure 6. Stage IV rainfall accumulation in Hurricane Rita: (a) 6 h ending 0600 UTC 24 September 2005, (b) 12 h ending at 1200 UTC 24 September 2005, and (c) 24 h ending at 0000 UTC on 25 September 2005. The third image uses a different color scale than the first two.

rain intensity increased, the frequency bias for 6 h rainfall decreased to nearly 0.5 for eTRaP and 0.6 for TRaP for a 100 mm threshold, then increased again for higher thresholds. This suggests that the 6 h TRaPs, and the eTRaPs which are constructed from them, do not adequately capture the short-term development of heavy rain areas that are observed. This is corroborated by the small correlation coefficients in Table 3, and low values of the equitable threat score (ETS), which is a commonly used statistic to evaluate the correct placement of rain areas. The low scores are also partly related to the use of a very fine verification grid. The correlation, RMSE, and ETS all improve when the eTRaP and TRaP forecasts are verified on a coarser grid (not shown).

Table 3

	6 h accum. Obs max=164 mm		24 h accum. Obs max=300 mm	
	eTRaP	TRaP	eTRaP	TRaP
Predicted max rain (mm)	130	94	246	162
RMSE (mm)	28	29	46	53
Correlation coefficient	0.13	0.07	0.21	0.20
Volume ratio	0.89	0.89	0.97	0.99

Table 3. Average performance of eTRaP and TRaP for eighteen Atlantic storms and hurricanes in Table 2. The verification grid scale was 4 km.

Observed 24 h accumulated rainfall patterns are smoother than short period accumulations due to the integrating effect of the longer accumulation period, and both eTRaP and TRaP were better able to capture the observed spatial structure. The ETS values for 24 h rain were much higher than for 6 h rainfall, scoring over 0.4 for eTRaP QPFs of rain exceeding 12.7 mm. The frequency bias of eTRaP 24 h rainfall was close to one for all but the heaviest rain, an improvement on TRaP 24 h forecasts which predict too little rain coverage.

The eTRaP probabilities of precipitation exceeding various thresholds were verified using reliability diagrams, which measure bias in the predicted probabilities, and relative operating characteristic (ROC) diagrams, which measure the ability of the forecast to discriminate between events and non-events. Figure 8 shows that the forecast probabilities for 6 h and 24 h accumulations were too high (overconfident), with better performance for lighter thresholds and long accumulations, and poorer performance for heavier thresholds and short accumulations. The departure of the reliability curves from the diagonal means that the probabilities should not be taken at face value, and that some calibration should be applied to improve their reliability.

The ROC diagrams confirm that the eTRaP precipitation probabilities have useful skill (curve located well to the left of the diagonal). The verification only resolved probabilities to the nearest 0.1, leading to insufficient sampling along the ROC curve, especially for the higher thresholds; future verification will resolve the probability forecasts into finer intervals to correct this deficiency in the ROC evaluation.

5. DISCUSSION AND CONCLUSIONS

Ensemble TRaP provides predictions of 6 h, 12 h and 24 h rainfall amount and probabilities of exceeding various thresholds, for rain in landfalling tropical cyclones. The eTRaP QPFs are more accurate than single-sensor TRaP forecasts in all important aspects: maximum rainfall amount, spatial pattern, RMSE, rain intensity distribution, and location. Importantly, eTRaP offers the possibility to provide probabilistic forecasts for decision makers.

Many improvements can be made to eTRaP. For example, it would be possible to apply a bias correction step to the TRaPs to remove expected systematic errors before generating the eTRaP forecasts. Evaluation of the past performance of TRaPs during the 2002 Atlantic hurricane season and the 2003-04 Australian tropical cyclone season suggests that the microwave TRaPs underestimated the total rain volume by an average of 1/3 (Ferraro et al. 2005, Ebert et al. 2005). This value was approximately the same for all sensors. However, Kuligowski et al.'s (2006) validation results suggested that the microwave TRaPs had little mean bias for the 2004 and 2005 Atlantic hurricane seasons, and results of the 4-year validation support Kuligowski's results.

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 resolution of geostationary data are much greater than for passive microwave data, offering more detailed rainfall estimates and potentially very large ensembles. However, this increased detail can also lead to noisy TRaPs with small regions of unrealistically high rain rates. Spatial and temporal averaging of the H-E rainfall estimates prior to extrapolation can reduce some of this unwanted variability. In this section we test the addition of H-E TRaPs to the microwave TRaPs used to generate the ensemble forecasts.

Ensemble TRaP could benefit from adding R-CLIPER and/or NWP models to the ensemble. NWP has the advantage that its forecasts extend out much longer than 24 h. It might be advantageous to blend eTRaP into longer range model forecasts in order to make time series products for locations at risk.

Further improvements to TRaP will be investigated in the next few years. These include modification of TRaP land-based rainfall to account for orographic enhancement (Vicente 2002), and inclusion of rotation in the TRaP extrapolation forecast (Liu et al. 2008).

References

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.

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

Ferraro, R., P. Pellegrino, M. Turk, W. Chen, S. Qiu, R. Kuligowski, S. Kusselson, A. Irving, S. Kidder and J. Knaff, 2005: The Tropical Rainfall Potential (TRaP) technique. Part 2: Validation. *Wea. Forecasting*, **20**, 465-475.

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 1: Description and examples. *Wea. Forecasting*, **20**, 456-464.

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. *2nd Intl. Symp. Quantitative Precipitation Forecasting and Hydrology, Boulder, CO, 4-8 June 2006.*

Lin, Y. and K. Mitchell, 2005: The NCEP Stage II/IV hourly precipitation analyses: development and applications. *Proceedings 19th Conf. Hydrology, Amer. Met. Soc., San Diego, CA, 9-13 January 2005.*

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*, **22**, 259-269.

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.

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 Sensing*, **23**, 221-230.

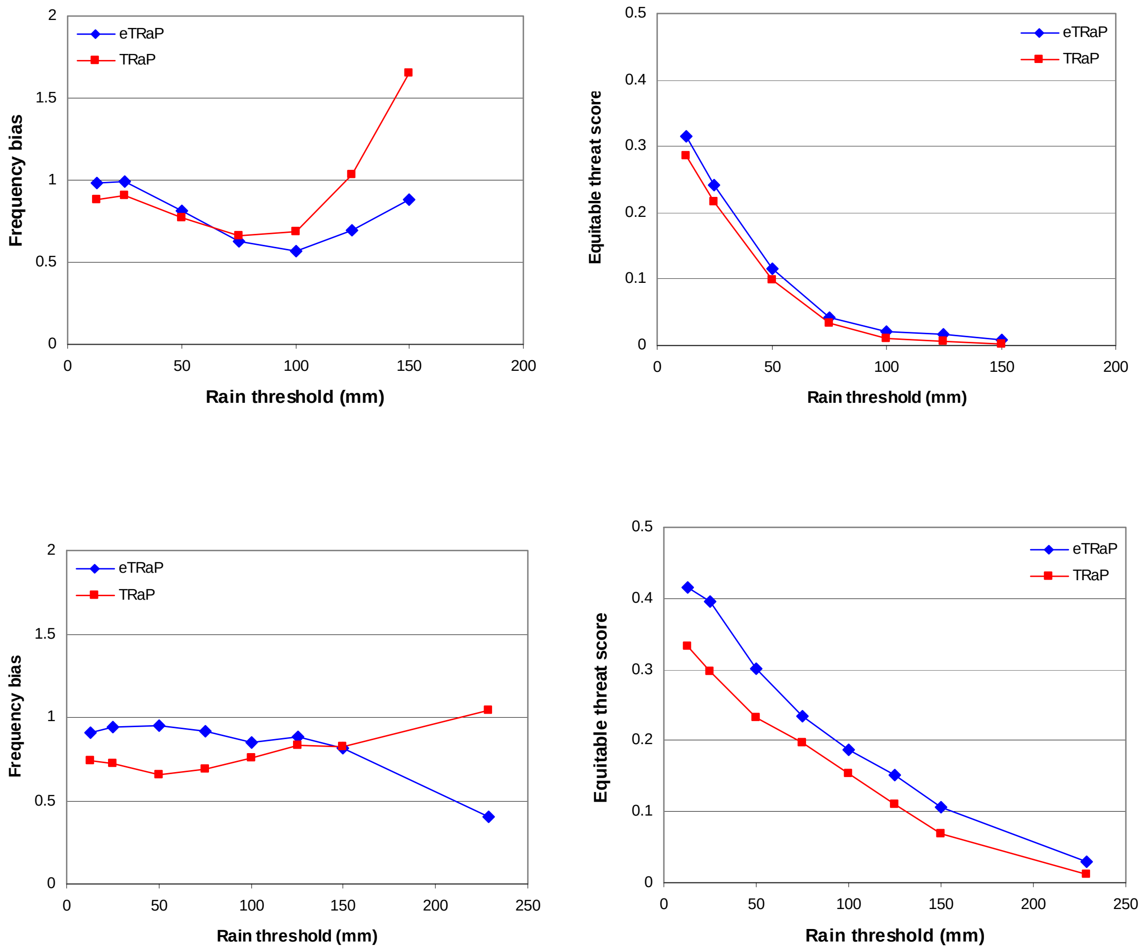

Figure 7. Verification of 6 h accumulated precipitation (all lead times) (top row) and 24 h accumulated precipitation (bottom row) predicted by eTRaP and TRaP for the storms listed in Table 2. The eTRaP QPFs are the probability-matched ensemble mean fields.

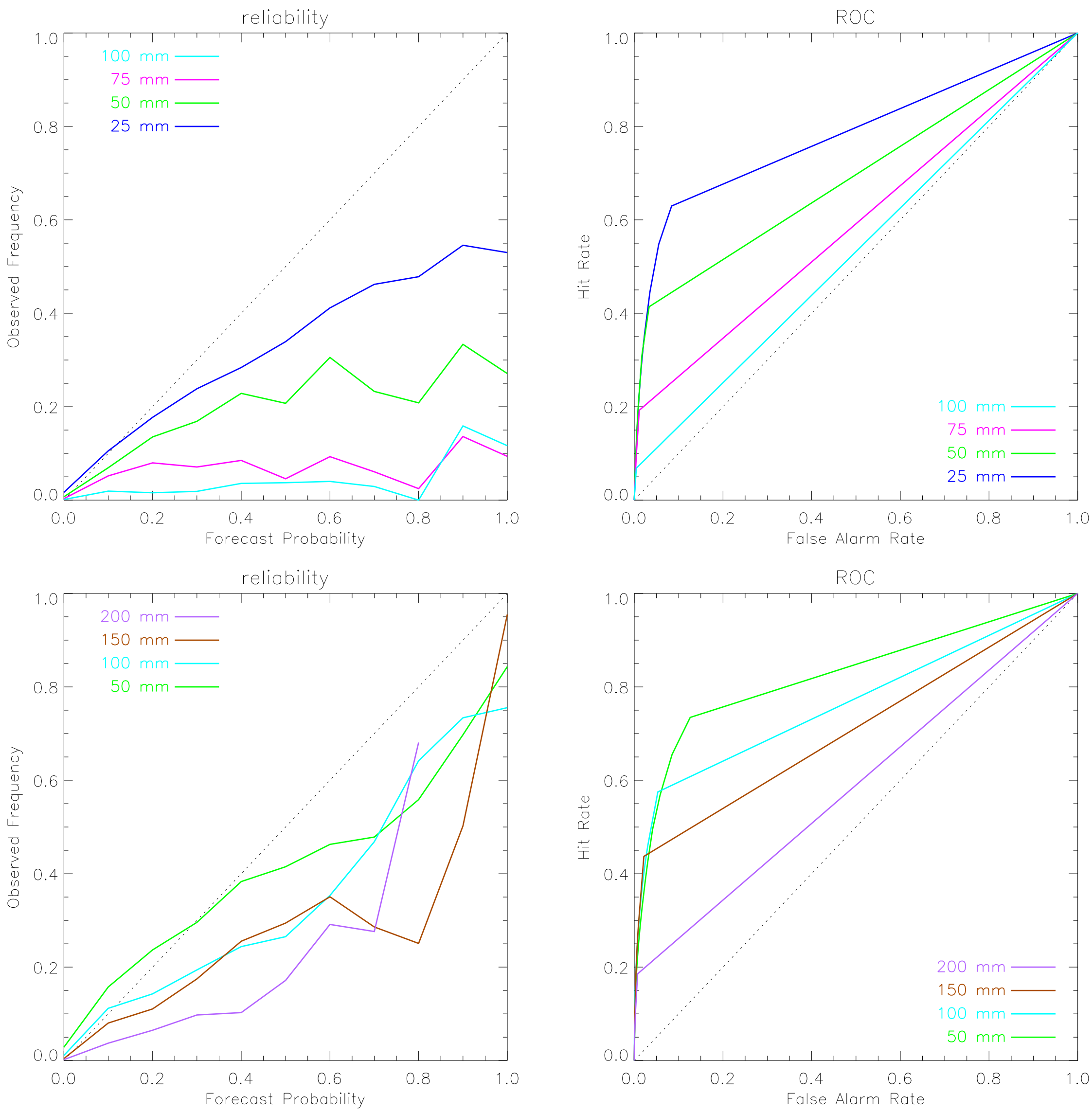

Figure 8. Verification of eTRaP probability forecasts for 6 h accumulations (all lead times) (top row) and 24 h accumulations (bottom row) for the storms listed in Table 2.