

Understanding Extreme Typhoon Quantitative Precipitation Forecasts (QPFs) in Taiwan by the 2.5-km CReSS Model: An Update

Chung-Chieh Wang (王重傑) and Chien-Chang Huang (黃建彰)
Department of Earth Sciences, National Taiwan Normal University, Taipei, Taiwan

Abstract

A strong dependency of model performance in quantitative precipitation forecasts (QPFs) on rainfall amount in categorical statistics such as the threat score (TS), i.e., the better model performs when there is more rain, is recently demonstrated by the first author through real-time forecasts by the 2.5-km Cloud-Resolving Storm Simulator (CReSS) for 15 typhoons in Taiwan in 2010-2012 (Wang, 2015). For typhoon QPFs in Taiwan, this dependency is due not only to the positive correlation between scores and rain-area sizes but also to the model's ability to properly handle (within 72 h) the processes leading to more rain, as these factors are mostly at least meso- α scale or linked to the island's topography. Thus, the common perception that models (cloud-resolving models in particular) perform poorly for extreme rainfall events is incorrect.

Due to this dependency, the performance of model QPFs for extreme events can be understood only when forecasts targeted for large events (which, by definition, are infrequent) are evaluated. To include events of various magnitudes, contingency tables should be combined into one before the scores are computed, so that each point is weighted the same. Following the above method, the statistics are updated through the 2013 season here. For the most-rainy 24-h of the top-8 typhoons in 2010-2013 (roughly top 5% of all sample, with high hazard potential), the day-1 (0-24 h) QPFs by CReSS have TS of 0.78, 0.75, 0.66, 0.56, 0.37, and 0.24 at thresholds of 25, 50, 130, 200, 350, and 500 mm (per 24 h), respectively. The day-2 (24-48 h) QPFs have TS of 0.82, 0.79, 0.65, 0.56, 0.37, and 0.20 at the same thresholds, almost identical to the day-1 scores. The day-3 (48-72 h) QPFs have TS of 0.68, 0.61, 0.46, 0.40, 0.28, and 0.10, respectively. Even at the extreme threshold of 750 mm, the TS at day-1 to day-3 QPFs are 0.12, 0.15, and 0.07 respectively, and all at least close to 0.1, indicating certain ability in QPFs. The above scores indicate superior performance for typhoon extreme rainfall even 2-2.5 days in advance, when a cloud-resolving model like CReSS is employed.

Key words: Quantitative precipitation forecast (QPF), model verification, typhoon, Taiwan, cloud-resolving model, CReSS

1. Introduction

The quantitative precipitation forecast (QPF) for heavy to extreme rainfall events, due to their hazardous nature, is one of the most important and challenging task in modern numerical weather forecasting. This is especially true in Taiwan, as most of its weather hazards come from heavy rainfall brought by tropical cyclones (TCs) in summer and mesoscale convective systems (MCSs) in the Mei-yu season. To make better QPFs, (1) the improvement in models (e.g., basic configuration and data assimilation), (2) the adoption of more effective forecast strategy (e.g., Wang et al., 2015), and (3) better understanding of verification tools for proper evaluation of model QPFs are all crucial in the process. Of course, improvements in observing system, data quality, and global models will also provide better initial and boundary conditions (IC/BCs) for regional models that produce QPFs, but this will be a more gradual and long-term process by comparison.

One key aspect to improve QPFs in Taiwan is believed to be in the configuration of the models, where high resolution is required to better resolve the basic structure (e.g., updrafts and downdrafts) of deep convection and the island's complex topography. Thus, in forecast experiments performed by the first author using the Cloud-Resolving Storm Simulator (CReSS, Tsuboki and Sakakibara, 2007), a 2.5-km grid size has been used since 2010, with a domain size of 1080×900 km² (432×360 grid points) and 40 levels. This domain has been enlarged to 1500×1200 km² (600×480) since 2012, and the model has been the only "cloud-resolving" member participating in the QPF ensemble experiment of the Taiwan Typhoon and Flood Research Institute (TTFRI). In Wang (2015), the performance of the day-1 to day-3 QPFs by this 2.5-km CReSS for all 15 typhoons to hit Taiwan during 2010-2012 is evaluated, and the study demonstrated (1) the fundamental property of "the more rain, the better the model performs" in traditional categorical statistics, and (2) the superior performance of the 2.5-km CReSS

in QPFs for large events once proper classification is introduced, according to the above-mentioned property.

In this paper, after section 2 which describes our data and methodology, the main results of Wang (2015) is reviewed to demonstrate the fundamental property of “the more rain, the better the model performs” and discuss its significance, particularly in terms of how to properly evaluate model QPFs for extreme rainfall events. Then, some forecast examples are provided and the verification results of the 2.5-km CReSS are updated through the 2013 season with the addition of six more typhoons.

2. Data and methodology

The 2.5-km CReSS model evaluated in this paper, as described briefly above, is the same as that in Wang (2015), so the readers are referred to that paper and the detailed configuration of the model is not repeated here. To remedy the issue of “double penalty” in high-resolution models (e.g., Ebert and McBride, 2000; Davis et al., 2006), we choose to verify the 24-h QPFs, for day 1 (0-24 h), day 2 (24-48 h), or day 3 (48-72 h). For this purpose, hourly rainfall data from more than 400 automated rain gauges over Taiwan are used. The model QPFs are first interpolated onto these observation sites, and then the widely-used categorical statistics is computed as the following.

The objective verification method based on the 2×2 contingency table (i.e., categorical measure), such as the threat score (TS) is calculated for 24-h QPFs (days 1-3) from runs starting at 0000 and 1200 UTC, across a threshold range up to 1000 mm. At any given threshold, the TS is defined as $TS = H/(H + M + FA)$, where H, M, and FA are the counts of hits (observed and predicted), misses (observed but not predicted), and false alarms (predicted but not observed), respectively, among all verification points (N). Thus, TS has a value between 0 and 1, and the higher the better. Another measure shown in this paper is the bias score (BS), which is defined as $BS = (H + FA)/(H + M)$ and measures whether the model over-predict ($BS > 1$) or under-predict ($BS < 1$) the rain area (e.g., Schaefer, 1990; Wilks, 1995).

3. The dependency of QPF skill on rainfall

Figure 1 shows the main results of Wang (2015), the TS of day-1 to day-3 QPFs for the 15 typhoons in 2010-2012 when all 24-h segments (sample size $n = 99$, cf. Table 1) are classified based on observed rainfall amount. The classification scheme is as follows: The segment belongs to A, B, or C group if at least 50 sites (roughly 1/8) have accumulative rainfall amount of ≥ 100 , 50, or 25 mm, respectively. For segments that cannot meet the criteria of C, they are in D group. Thus, A-D groups include rainfall events from large to small, and any segment belongs to just one group (exclusive). From group A, the top-5 events are picked and

represent roughly the top 5% most rainy, and thus the most hazardous, 24-h periods (red bold letters in Table 1). In each group and at each range (day 1, 2, or 3), the scores are first computed for all segments and then their mean values are obtained (as many researchers often do). As shown in Fig. 1, the mean TS values for the top-5 periods are higher than those for group A, and those for A are higher than group B, and then C and D. For the top-5 cases, the 0-24-h (day-1) QPFs by CReSS have mean TS of 0.67, 0.58, 0.51, and 0.32 at thresholds of 50, 130, 200, and 350 mm, the 24-48-h (day-2) QPFs yield mean TS of 0.73, 0.57, 0.42, and 0.17, and the 48-72-h (day-3) QPFs yield 0.57, 0.37, 0.33, and 0.22, respectively. For the biggest event in Morakot (2009), the scores from real-time forecasts are even higher, and those of day-2 QPFs reach 0.87, 0.69, 0.50, and 0.38 at heavy to extreme thresholds of 200, 350, 500, and 1000 mm, respectively (Fig. 1b, also Wang, 2014). While this particular forecast for Morakot (Fig. 2; cf. Wang et al., 2013) produced peak 48-h total rainfall (7-8 August) very close to the observation (about 2200 mm), other models at the time struggled to reach 1400 mm in 3- or 4-day total (e.g., Hendricks et al. 2011). Thus, the evidence points strongly toward a cloud-resolving setting to improve QPFs, as attested by Wang (2015). For the smaller events in groups B, C, and D, they are less and less significant. Toward higher thresholds, the points involved to compute TS for each segment also tend to be fewer due to their smaller rain-area size (Fig. 1d), and it is not appropriate to average TS for all periods without proper stratification. In Fig. 1, the property of “the more rain, the higher the score” in categorical statistics is well demonstrated. Thus, it is a misperception that the models (cloud-resolving models in particular) have little ability to predict extreme rainfall events. Due to this dependency property, the ability of the models in extreme events can only be understood through proper classification (to evaluate only such events, not those with smaller magnitudes).

4. Updated results through 2013

In this section, the results of Wang (2015) are updated through the 2013 season, with the addition of six typhoons. Using the same classification scheme, the typhoons and their grouping results are listed in Table 2, with a total of 153 24-h segments since 2010. Another three segments (from Soulik, Trami, and Kong-Rey) are selected into the top category, making it the top-8 group which represent about top 5% of sample still (Table 2). Here, before the overall results are shown, a specific forecast example is given for Soulik. In the forecast starting at 12Z 10 July 2013, the 2.5-km CReSS captured very well the track and rainfall structure of this intense storm (Fig. 3), and subsequently produced very high-quality QPF on day 3 with TS decreasing slowly from a perfect 1.0 at low thresholds to about 0.38 at the extreme threshold of 750 mm (Fig. 4). Such an impressive forecast on day 3

(48-72 h) is due exactly to the model's high resolution and large fine domain size.

As shown above and in Wang (2015), it is not appropriate to compute the scores first then make the averages, because the values from smaller rain areas (i.e., fewer points) would be weighted more than they should. Thus, in the following update, the contingency tables for each segment (in the same group) are first combined into one table before the scores are computed. This way, each point has the same weight, regardless which category (H, M, FA, or correct negatives) it falls in. Following this method, the statistics are updated through the 2013 season and shown in Fig. 5. For the most-rainy 24-h of the top-8 typhoons in 2010-2013 (roughly top 5% of all sample), the day-1 (0-24 h) QPFs by CReSS have TS of 0.78, 0.75, 0.66, 0.56, 0.37, and 0.24 at thresholds of 25, 50, 130, 200, 350, and 500 mm (per 24 h), respectively. The day-2 (24-48 h) QPFs have TS of 0.82, 0.79, 0.65, 0.56, 0.37, and 0.20 at the same thresholds, almost identical to the day-1 scores. The day-3 (48-72 h) QPFs have TS of 0.68, 0.61, 0.46, 0.40, 0.28, and 0.10, respectively. Even at the extreme threshold of 750 mm, the TS for day-1 to day-3 QPFs are 0.12, 0.15, and 0.07 respectively, and all at least close to 0.1, indicating certain ability in QPFs. Therefore, with TS ranging at 0.28-0.66, the 2.5-km CReSS is not only skillful at heavy rainfall thresholds (130-350 mm) within 72 h, but also has skill at extreme thresholds up to 750 mm even at day 2-3. The above scores indicate superior performance for typhoon extreme rainfall even 2-2.5 days in advance, when a cloud-resolving model like CReSS is employed. In short, the future in cloud-resolving model QPFs for extreme rainfall is bright!

5. Conclusion

A strong dependency of model performance in quantitative precipitation forecasts (QPFs) on rainfall amount in categorical statistics such as the threat score (TS), i.e., the better model performs when there is more rain, is recently demonstrated in Wang (2015). For typhoon QPFs in Taiwan, this dependency is due not only to the positive correlation between scores and rain-area sizes but also to the model's ability to properly handle (within 72 h) the processes leading to more rain. Thus, the common perception that models (cloud-resolving models in particular) perform poorly for extreme rainfall events is incorrect.

Due to this dependency, the performance of model QPFs for extreme events can be understood only when forecasts targeted for large events (which, by definition, are infrequent) are evaluated. To include events of various magnitudes, contingency tables should be combined into one before the scores are computed, so that each point is weighted the same. Following the above method, the statistics are updated through the 2013 season in this paper with a total of four typhoon seasons (21 cases). For the most-rainy 24-h of the top-8 typhoons in 2010-2013 (roughly top 5% of all sample,

with high hazard potential), the day-1 (0-24 h) QPFs by the 2.5-km CReSS have TS of 0.78, 0.75, 0.66, 0.56, 0.37, and 0.24 at thresholds of 25, 50, 130, 200, 350, and 500 mm (per 24 h), respectively. The day-2 (24-48 h) QPFs have TS of 0.82, 0.79, 0.65, 0.56, 0.37, and 0.20 at the same thresholds, almost identical to the day-1 scores. The day-3 (48-72 h) QPFs have TS of 0.68, 0.61, 0.46, 0.40, 0.28, and 0.10, respectively. Even at the extreme threshold of 750 mm, the TS at day-1 to day-3 QPFs are 0.12, 0.15, and 0.07 respectively, and all at least close to 0.1, indicating certain ability in QPFs. The above scores indicate superior performance for typhoon extreme rainfall even 2-2.5 days in advance, when a cloud-resolving model like CReSS is employed. Such high scores are obtained in real time without any additional data assimilation, vortex bogus, or relocation procedure. Hence, the importance of a better model configuration (cloud-resolving capability and a larger fine domain) in the improvement of heavy to extreme rainfall prediction is again demonstrated, and it is imperative that the operational agencies to move toward this direction.

In Wang et al. (2015), we also proposed a feasible and effective method to perform ensemble QPFs through a time-lagged approach. Using such a method, a high resolution of 2.5 km, a large domain of $1860 \times 1360 \text{ km}^2$, and an extended range out to 8 days can all be achieved simultaneously using computational resources comparable to a typical multi-member ensemble currently in use. By combining the strengths of high resolution (for QPF) and longer lead time (for hazard preparation) in an innovative way, the system can provide a wide range of rainfall scenarios in Taiwan early on at longer lead time (at about 4-8 days), each highly realistic for its track, for advanced preparation for the worst case. Then, at shorter ranges within 3 days, the authority can make adjustments toward the more-likely scenario as the typhoon approaches and the forecast uncertainty reduces. With improved resolution and forecast strategy, the above method will enhance greatly our ability to predict heavy to extreme typhoon rainfall events in Taiwan.

References

- Davis, C., B. Brown, and R. Bullock, 2006: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772-1784.
- Ebert, E. E., and J. L. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *J. Hydrol.*, **239**, 179-202.
- Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting*, **5**, 570-575.
- Tsuboki, K., and A. Sakakibara, 2007: *Numerical Prediction of High-Impact Weather Systems: The Textbook for the Seventeenth IHP Training Course in 2007*. Hydrospheric Atmospheric Research Center, Nagoya University, and UNESCO, 273 pp.

Wang, C.-C., 2014: On the calculation and correction of equitable threat score for model quantitative precipitation forecasts for small verification areas: The example of Taiwan. *Wea. Forecasting*, **29**, 788-798.

Wang, C.-C., 2015: The more rain, the better the model performs---The dependency of quantitative precipitation forecast skill on rainfall amount for typhoons in Taiwan. *Mon. Wea. Rev.*, **143**, 1723-1748.

Wang, C.-C., H.-C. Kuo, T.-C. Yeh, C.-H. Chung, Y.-H. Chen, S.-Y. Huang, Y.-W. Wang, and C.-H.

Liu, 2013: High-resolution quantitative precipitation forecasts and simulations by the Cloud-Resolving Storm Simulator (CReSS) for Typhoon Morakot (2009). *J. Hydrol.*, **506**, 26-41.

Wang, C.-C., S.-Y. Huang, S.-H. Chen, C.-S. Chang, and K. Tsuboki, 2015: Cloud-resolving typhoon rainfall ensemble forecasts for Taiwan with large domain and extended range through time-lagged approach. *Wea. Forecasting* (under second review)

Wilks, D. S., 1995: *Statistical methods in the atmospheric sciences*. Academic Press, 467 pp.

Table 1 The 15 typhoons included in this study (and Wang, 2015) in 2010-2012, their case periods (for verification), and classification results (A to D). Bold face indicates most rainy period of each case, and red color marks those of the top-5 cases.

Typhoon	Verification period	Classification
Lionrock	8/28 00Z-9/3 12Z	CC[CBAB]*CBACDD
Namtheun	8/29 00Z-8/31 12Z	CBAB
Meranti	9/6 00Z-9/11 12Z	DDDDCBBCDC
Fanapi	9/16 00Z-9/21 12Z	DDDDBAACD
Megi	10/19 00Z-10/24 12Z	BBBAABCDD
Aere	5/9 00Z-5/10 00Z	D
Songda	5/26 00Z-5/29 12Z	CCBCDD
Meari	6/24 00Z-6/26 12Z	BABC
Muifa	8/6 00Z-8/7 12Z	DD
Nanmadol	8/27 00Z-9/1 12Z	BAAAAACCC
Talim	6/19 00Z-6/22 12Z	BAABCC
Doksuri	6/28 00Z-6/30 12Z	CCDD

Table 1 (Continued)

Typhoon	Verification period	Classification
Saola	7/30 00Z-8/3 12Z	BAAAAAA
Tembin	8/22 00Z-8/28 12Z	CCBAABCDDDBB
Jelawat	9/27 00Z-9/29 12Z	DCCD
Total	99 segments	A-D: 26, 21, 26, 26

*Brackets indicate repeated periods (counted just once).

Table 2 As in Table 1, except for the six additional typhoons in 2013. Red color marks the three periods included in top-8 group, and the bottom row shows the result of all typhoon cases in 2010-2013.

Typhoon	Verification period	Classification
Soulik	7/11 00Z-7/16 12Z	DDAAACCCCD
Cimaron	7/17 00Z-7/20 12Z	CCDDDD
Trami	8/19 00Z-8/24 12Z	DBAAAABBB
Kong-Rey	8/27 00Z-8/31 12Z	DDAAAAAA
Usagi	9/19 00Z-9/24 12Z	DDBAAABCDD
Fitow	10/4 00Z-10/9 12Z	DCCBCDDDDDD
Total	153 segments	A-D: 43, 28, 36, 46

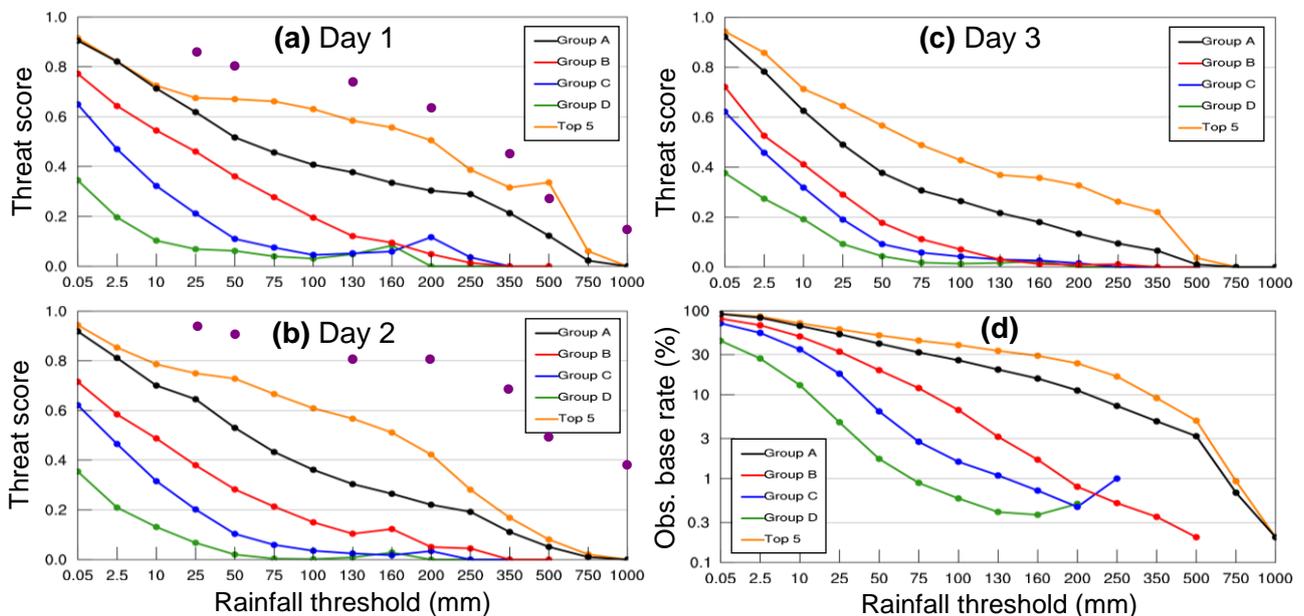


Fig. 1 Mean TS of 24-h QPFs for groups A-D (most to least rain, each about 1/4) and top-5 cases (top 20% of A, or top 5% of all sample) at (a) day 1 (0-24 h), (b) day 2 (24-48 h), and (c) day 3 (48-72 h) across a threshold range from 0.05 to 1000 mm (per 24 h) by the 2.5-km CReSS, averaged among all segments (commonly used), and the TS for Morakot (00-24Z 8 Aug, >1500 mm; purple dots) from single CReSS forecasts (days 1-2 only). (d) Observed base rate [i.e., (H + M)/N = O/N, %] in groups A-D and top-5 cases in logarithmic scale across the threshold.

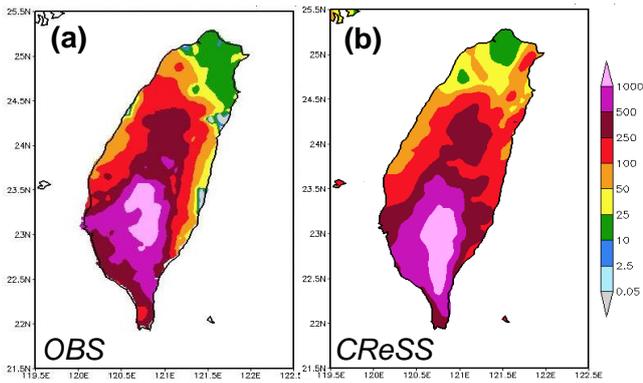


Fig. 2 (a) Observed 24-h accumulative rainfall distribution (mm) on 8 August 2009 (00-24 UTC) during Typhoon Morakot, and (b) predicted 24-h rainfall valid for 8 August by CReSS in real time, starting from 00 UTC 7 August (day-2 QPF). The TS values of this forecast at selected thresholds are shown in Fig. 1b.

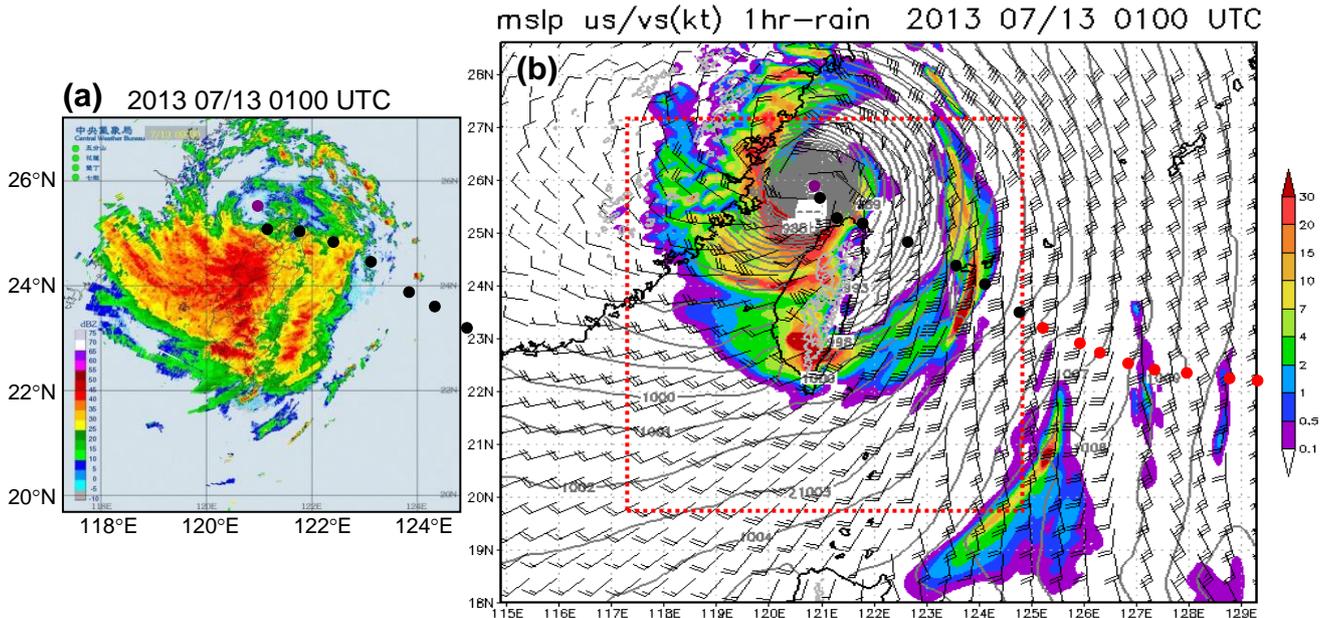


Fig. 3 (a) Radar reflectivity composite (dBZ) at 01 UTC 13 July 2013 (during Soulik) and (b) 2.5-km CReSS forecast (at 12Z 10 July) of sea-level pressure (hPa), surface wind (kt), and hourly rainfall (mm, color) valid at the same time ($t = 61$ h). Purple, black, and red dots depict typhoon center at the time of (a),(b), and every 3 h inside, and outside the domain of (a), respectively. In (b), the domain of (a) is also shown (red dotted box).

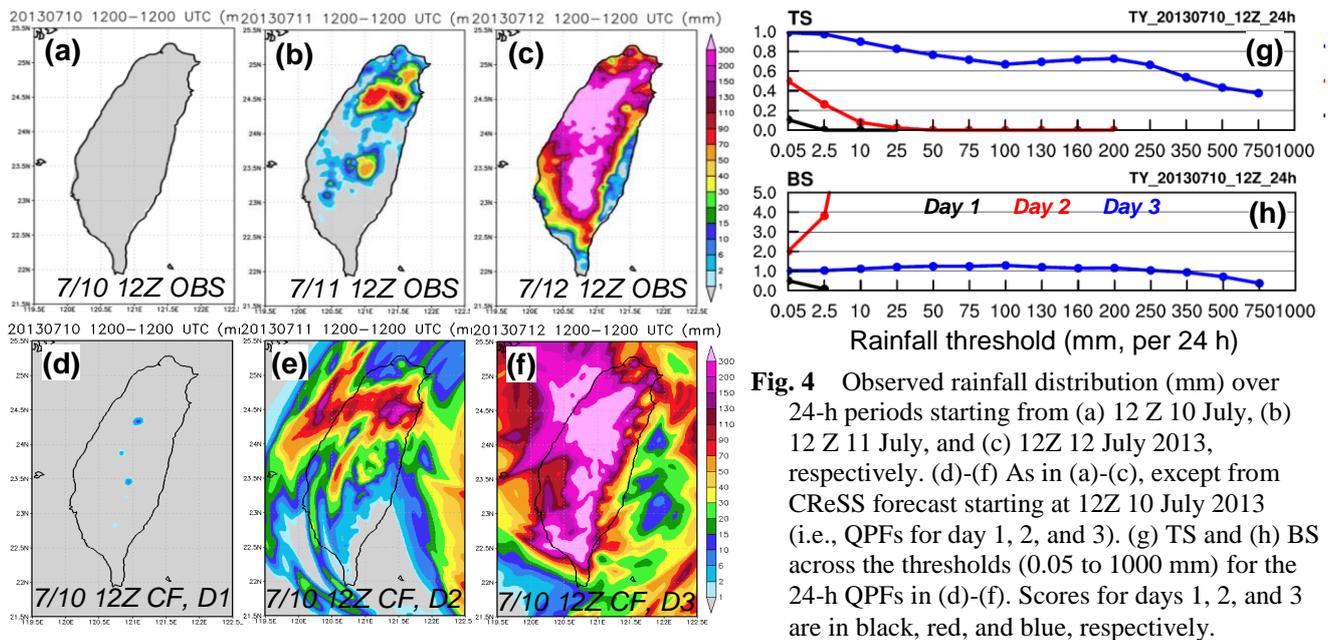


Fig. 4 Observed rainfall distribution (mm) over 24-h periods starting from (a) 12 Z 10 July, (b) 12 Z 11 July, and (c) 12Z 12 July 2013, respectively. (d)-(f) As in (a)-(c), except from CReSS forecast starting at 12Z 10 July 2013 (i.e., QPFs for day 1, 2, and 3). (g) TS and (h) BS across the thresholds (0.05 to 1000 mm) for the 24-h QPFs in (d)-(f). Scores for days 1, 2, and 3 are in black, red, and blue, respectively.

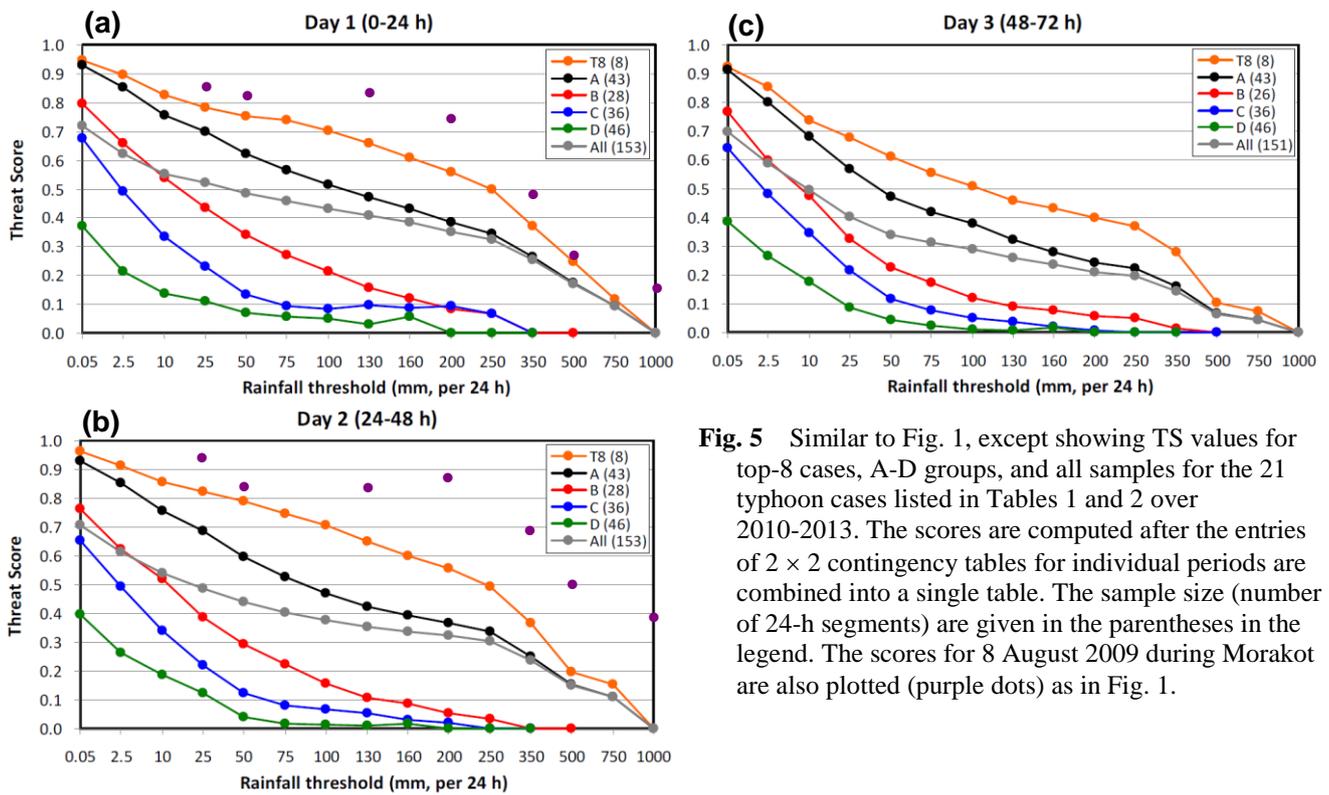


Fig. 5 Similar to Fig. 1, except showing TS values for top-8 cases, A-D groups, and all samples for the 21 typhoon cases listed in Tables 1 and 2 over 2010-2013. The scores are computed after the entries of 2×2 contingency tables for individual periods are combined into a single table. The sample size (number of 24-h segments) are given in the parentheses in the legend. The scores for 8 August 2009 during Morakot are also plotted (purple dots) as in Fig. 1.