Benchmark rainfall verification of landfall tropical cyclone forecasts by operational ACCESS-TC over China

Zifeng Yu1,2 | Ying Jun Chen3 | Beth Ebert3 | Noel E. Davidson3 | Yi Xiao3 | Hui Yu1 | Yihong Duan2

1Shanghai Typhoon Institute (STI) of China Meteorological Administration (CMA), Shanghai, China
2Chinese Academy of Meteorological Sciences, China Meteorological Administration (CMA), Beijing, China
3Research and Development Branch, Bureau of Meteorology, Melbourne, Australia

Correspondence
Zifeng Yu, Shanghai Typhoon Institute (STI), 166 Puxi Road, Xujiahui, Shanghai 200030, China.
Email: zifengyu@yahoo.com, yuzf@typhoon.org.cn

Funding information
Key Program for International S&T Cooperation Projects of China, Grant/Award Numbers: 2017YFE0107700, GASI-IPOVAI-04; National Natural Science Foundation of China, Grant/Award Numbers: 41875080, 41775064, 41475058; Scientific Research Program of Shanghai Science and Technology Commission, Grant/Award Number: 19dz1200101

Abstract
Results from object-based verification of rainfall forecasts for landfalling tropical cyclones (TCs) over China during the period 2012–2015 are presented. The sample consists of 25 landfall events and 133 operational numerical forecasts from the TC version of the Australian Community Climate and Earth System Simulator. Mean equitable threat scores, probabilities of detection and false alarm ratios for the 30 mm isohyet for the unadjusted forecasts at 0–6 hr (essentially the initialization) are (0.23, 0.55, 0.65), while the performance measures of 24 hr forecast accumulations are the best for the 0–24 hr forecast (0.37, 0.67, 0.40) and then worsen to (0.16, 0.38, 0.66) for the 48–72 hr forecast. Forecast ability also decreases with the increase in rainfall amount. The contiguous rain area (CRA) verification method is used to diagnose the source of systematic errors from the displacement, rotation, volume and pattern of the forecasted rain fields. Results show that the errors are mostly from rainfall patterns, followed by displacement errors, particularly for very heavy rain. After application of the displacement and rotation adjustments of the CRA method, averaged errors improve by about 15%. Results suggest that rainfall prediction will continue to improve with improved track prediction, but more work is needed on model initialization and the prediction of TC structure. The study has uncertainty related to the limited sample size, which could cause large variability, particularly for heavy rainfall at 6 and 72 hr. However, the results still represent a useful benchmark for future verification of landfalling TCs.

KEYWORDS
tropical cyclone, typhoon, precipitation forecast, rainfall evaluation

1 | INTRODUCTION

Tropical cyclone (TC) track forecasts have greatly improved in the past few decades, while TC intensity and rainfall forecasts have shown only minor improvements (Gall et al., 2013; Lei et al., 2016; Leroux et al., 2018). Wang et al. (2012) used threat score (TS) and bias to perform verification of TC rainfall predictions from the global models of the China Meteorological Administration (CMA) and the Japan Meteorological Agency (JMA) in 2009 and found that the TSs of 24 hr rainfall forecasts of > 25 mm are 0.218 and 0.196 for the JMA and CMA, respectively. Since then,
been rather limited, except for Marchok et al. (2007). For landfalling TCs, the verification of rainfall forecasts needs not only conventional methods but also the application of some new methods, which might provide more insights into systematic errors to improve the forecasts. Sometimes even though the track during landfall may be well predicted, the exact rainfall distribution and intensity are not. Thus, in order to improve rainfall forecasts, the first step is to document the current forecast performance (benchmark verification) and to try to isolate any systematic errors that numerical systems may have. Particularly during recent years, there have been marked improvements in the resolution of numerical systems, data assimilation and specific methods of vortex initialization. Therefore, it is valuable to perform the verification for a current operational TC-forecast system.

In an early verification study, Marchok et al. (2007) proposed validation schemes for the TC quantitative precipitation forecasts of US landfalling TCs during the period 1998–2004. They tried to take advantage of the TC rainfall characteristics by evaluating the rainfall forecasting skill with three metrics: the ability to match observed rainfall patterns, the ability to match the mean observed rainfall and the ability to produce extreme rainfall amounts often observed in TCs. However, to date “object” verification methods have not been extensively used in TC rainfall validation. Therefore, the objective of the present study is to give a quantitative verification of a numerical forecast system for a large number of landfall-TC rainfall forecasts. This will act as a benchmark to measure any future improvements.

Till now, some conventional verification scores including equitable threat score (ETS), false alarm ratio (FAR) and probability of detection (POD) have often been used to analyze quantitative precipitation estimates or forecast quality (Yu et al., 2009). Though these traditional performance measures have been extensively used in assessing forecasts, and are still valuable, they have limited diagnostic ability to provide insights and possible reasons for the source of errors in forecasts. They simply show users whether or not the forecast is good. Therefore, if forecasters and model developers want to obtain more information, such as rain volume and displacement errors, it is difficult to extract such information from those traditional verification results. As a result, two categories of new verification methods have been proposed, including filtering methods (e.g. scale separation) and object-based methods to evaluate displacement of rainfall features (Gilleland et al., 2009). The object-based methods focus on multiple attributes (e.g. shape, volume) of each object within a forecast or observation field. Precipitation is a common variable that lends itself to object-based methods.

Rain “objects” can be treated as contiguous regions where the rain rate or accumulation exceeds a specified threshold. Over recent years, object-oriented methods, including the contiguous rain area (CRA) method (Ebert and McBride, 2000; Chen et al., 2018), the method for object-based diagnostic evaluation (MODE; Davis et al., 2006, 2009) and the structure–amplitude–location (SAL; Wernli et al., 2008) method have been proposed. What follows will show the results from the application of the CRA method.

With improvements in TC track forecasts, both TC forecasters and researchers need to understand different aspects of rain forecast quality more deeply, including systematic rain error sources from the location, intensity and spatial pattern of the forecasted rain fields. As one of the early object-based verification methods, the CRA method uses pattern-matching techniques to determine location errors, as well as rain intensity and spatial-pattern errors. By using the CRA method, total rainfall errors can be decomposed into components due to location, volume and pattern errors. Their relative magnitudes can hint at the sources of model error. This is a useful property for model developers to use to diagnose the error source and thus possibly improve numerical weather prediction (NWP) systems. The CRA method has been applied in many studies, but it has not been extensively used for landfalling TC forecasts. Ebert et al. (2011) used CRA verification to evaluate satellite-based rainfall nowcasts of 16 hurricanes that made landfall in the United States between 2004 and 2008. Chen et al. (2018) chose 15 TC cases over the ocean to perform CRA verification against TRMM 3B42 data, but Chen et al. (2013) found TRMM 3B42 data would underestimate heavy rainfall (> 100 mm day$^{-1}$) with an average of 30% over the ocean and 70% over terrains. Therefore, to avoid the underestimation issue of TC rainfall in the satellite-retrieved (including TRMM 3B42) data, the present study will use gauge-satellite-merged rain data. In particular, it will focus on landfalling TC rainfall forecasts over China and explore and document the useful information extracted by the CRA method for the verification. Note that the rain-gauge network around coastal China is rather dense, so it should provide very reliable data for the verification.

ACCESS-TC (Davidson et al., 2014) is the TC version of the Australian Community Climate and Earth System Simulator, which has been operating over the West Pacific and eastern Indian Oceans since 2012. ACCESS-TC model forecasts will be evaluated as a demonstration of rainfall forecast ability for landfalling TCs. Details of the ACCESS-TC model, the observational rain data and a description of the CRA method are given in Section 2. Traditional and CRA verification results are described in Sections 3 and 4. Two landfalling TC rainfall cases are analysed and
illustrated in detail in Section 5. The conclusion and discussion are in Section 6.

2 | DATA AND VERIFICATION METHOD

2.1 | ACCESS-TC model

Rainfall forecasts from ACCESS-TC are used in the present study. The ACCESS-TC model was adapted for operational and research applications on TCs (Davidson et al., 2014), with a resolution of 0.11° and 50 vertical levels. The model domain is nested and relocatable with five cycles of four-dimensional (4D) variational data assimilation (VAR) over 24 hr. It makes 72 hr forecasts with a new run every 12 hr. To build physically based and synthetic inner-core structures, significant efforts have been made by using historical dropsonde data and surface analyses over the Atlantic. Vortex specification is applied to filter the analysed circulation, reconstruct the inner core of the TC in question, locate the TC centre to the observed location and then merge the circulation with the large-scale analysis at the outer radii. The construction is based on location, central pressure and storm size estimates. The 4D VAR then builds a balanced and intense 3D TC vortex with a well-developed secondary circulation and a maximum wind at the approximate maximum wind radius, by using all conventional observations and only synthetic surface pressure observations from the idealized vortex, to adjust the initial location and structure of the TC (Davidson et al., 2014).

2.2 | Best track data

The intensities and locations of TCs during the period 2012–2015 are extracted from the best-track data of the Shanghai Typhoon Institute (STI) of the CMA. There are in total 25 TCs, which were at least tropical storm

| 25 landfalling TCs | ID_china | Year | Month | date_start | hour_start | date_end | hour_end |
|--------------------|---------|------|-------|------------|-----------|----------|---------|
| Bebinca 2013       | 1305    | 2013 | 06    | 21         | 00        | 24       | 00      |
| Jebi 2013          | 1309    | 2013 | 08    | 01         | 00        | 03       | 00      |
| Doksurí 2012       | 1206    | 2012 | 06    | 29         | 00        | 29       | 12      |
| Vicente 2012       | 1208    | 2012 | 07    | 23         | 00        | 24       | 12      |
| Usagi 2013         | 1311    | 2013 | 08    | 13         | 00        | 14       | 12      |
| Tembin 2012        | 1214    | 2012 | 08    | 23         | 00        | 30       | 00      |
| Soulik 2013        | 1307    | 2013 | 07    | 12         | 00        | 13       | 12      |
| Cimaron 2013       | 1308    | 2013 | 07    | 17         | 00        | 18       | 12      |
| Trami 2013         | 1312    | 2013 | 08    | 21         | 00        | 22       | 12      |
| Fitow 2013         | 1323    | 2013 | 10    | 06         | 00        | 07       | 00      |
| Haikui 2012        | 1211    | 2012 | 08    | 07         | 00        | 09       | 12      |
| Rumba 2013         | 1306    | 2013 | 07    | 01         | 00        | 02       | 00      |
| Saola 2012         | 1209    | 2012 | 08    | 01         | 00        | 03       | 00      |
| Kalmegi 2014       | 1415    | 2014 | 09    | 15         | 00        | 17       | 00      |
| Matmo 2014         | 1410    | 2014 | 07    | 22         | 00        | 25       | 00      |
| Fung-wong 2014     | 1416    | 2014 | 09    | 20         | 00        | 23       | 12      |
| Kai-Tak 2012       | 1213    | 2012 | 08    | 16         | 00        | 18       | 00      |
| Hagibis 2014       | 1407    | 2014 | 06    | 14         | 12        | 17       | 12      |
| Rammason 2014      | 1409    | 2014 | 07    | 17         | 00        | 19       | 12      |
| Kujira 2015        | 1508    | 2015 | 06    | 21         | 00        | 24       | 12      |
| Linfa 2015         | 1510    | 2015 | 07    | 08         | 00        | 09       | 12      |
| Soudelor 2015      | 1513    | 2015 | 08    | 07         | 00        | 09       | 00      |
| Dajuan 2015        | 1521    | 2015 | 09    | 28         | 00        | 29       | 00      |
| Mujigae 2015       | 1522    | 2015 | 10    | 03         | 00        | 04       | 12      |
mean square error (MSE) of the original forecast can be.

In order to make the forecast field match the observed field, the

Finally, by adjusting via displacement and rotation, in

order to perform the verification.

2.3 | Gauge-satellite-merged rain data

The National Meteorological Information Center (NMIC) of

the CMA has generated a gauge-CMORPH (Climate Precipita-

tion Center Morphing)-satellite-merged rain product over

China over recent years (Shen et al., 2010, 2014), which is

used for the reference analysis in the present study. The

CMORPH precipitation product is known to provide quite

reasonable 6 and 24 hr rainfall distributions, but it can

underestimate heavy rainfall and overestimate very light

rainfall (Yu et al., 2009). To improve the accuracy of rain

estimates, the gauge-satellite-merged rain data are produced

by using hourly rain gauge data over China and the

CMORPH precipitation product (Shen et al., 2010). The

improved rain product has shown greatly improved quality

over land at hourly intervals, with a 0.1° resolution and

much smaller bias (Shen et al., 2014). It can also better cap-

ture some varying features of hourly precipitation in severe

weather systems. Therefore, it is a high-quality reference to

verify the ACCESS model rainfall forecasts during the TC

landfall. The model forecasts are interpolated onto the rain-

fall analysis grid of 0.1° in order to perform the verification.

2.4 | CRA verification method

The CRA verification method developed by Ebert and

McBride (2000) is used to evaluate systematic errors in the

prediction of the landfalling TC rainfall in the present study.

It is one of the earliest object-oriented methods to compare

objectively the properties of matched forecast and observed

rainfall systems. It decomposes the rainfall forecast and

observational distribution into entities to make the compari-

son. The CRA is defined as a region bounded by a user-

specified isohyet (e.g. a rain contour of 30 mm) in the

forecasts and observations. The forecast is horizontally sup-

erimposed over the observations and then rotated around the

centroid of the entity until a best-fit criterion is met (Moise

and Delage, 2011), which can be the minimum squared error

(Ebert and McBride, 2000), maximum overlap (Ebert et al.,

2004) or maximum correlation co-efficient (CC) (Grams

et al., 2006). Ebert and Gallus (2009) tested all the matching

approaches and found they gave similar location errors in

most conditions. The best-fit criterion used in the present

study is the minimum squared error (Ebert and McBride,

2000). Location error is the vector displacement of the fore-

cast. Finally, by adjusting via displacement and rotation, in

order to make the forecast field match the observed field, the

mean square error (MSE) of the original forecast can be
decomposed into displacement (D), rotation (R), volume

(V) and pattern (P) error components:

\[
MSE_{\text{total}} = MSE_{\text{displacement}} + MSE_{\text{rotation}} + MSE_{\text{volume}} + MSE_{\text{pattern}}
\] (1)

Based on the minimum squared error best-fit criterion, the error components can be calculated as follows:

\[
\begin{align*}
MSE_{\text{displacement}} &= MSE_{\text{total}} - MSE_{\text{ShiftedOnly}} \\
MSE_{\text{rotation}} &= MSE_{\text{ShiftedOnly}} - MSE_{\text{shifted + rotated}} \\
MSE_{\text{volume}} &= (\tilde{F} - \tilde{X})^2 \\
MSE_{\text{pattern}} &= MSE_{\text{shifted + rotated}} - MSE_{\text{volume}},
\end{align*}
\] (2)

where \(F\) and \(X\) are the means of the forecast and observation values, respectively.

Therefore, based on the CRA method, 30/50/100/250 mm rain thresholds are used to target increasingly heavy rainfall for verification. By evaluating the ACCESS-TC model, it is demonstrated that the CRA can help guide model developers and forecast users towards understanding the systematic errors in the ACCESS-TC, which the present authors believe is a representative operational numerical forecast system. As shown in Figure 1, mean track and intensity errors for the landfalling TC sample over China during the period 2012–2015 have been calculated as real-time operational results from the Australian National Meteorological and Oceanographic Centre. At 48 hr, the mean track error is 175.5 km, and the mean errors of the maximum wind speed and central sea-level pressure are 10.0 kt and 7.1 hPa, respectively. By comparison, the 48 hr track and intensity errors of the US National Hurricane Center (NHC) official forecasts are 143.3 km (i.e. 77.4 n-mi) and 11.4 kt, respectively, in 2015 (data not shown; see tables 2 and 4 on the NHC website, https://www.nhc.noaa.gov/verification/pdfs/Verification_2015.pdf). Therefore, it shows that the ACCESS-TC predictions are competitive with the skill of other international forecast systems, particularly when it is realized that reconnaissance is not so comprehen-
sive over the northwest Pacific as it is over the Atlantic and eastern Pacific. Some later studies, such as Lei et al. (2016), also provide useful comparisons. With the somewhat limited resolution of the operational ACCESS-TC, the minimum central pressure error (or bias) is generally negative (Figure 1d), which means that the forecast TC intensities are generally weaker than the estimates from the best track data. However, in general, the objective verification of track and intensity forecasts from ACCESS-TC show an encouraging and slightly worse performance than TC forecasts from the ECMWF, the UK Meteorological Office and N'EP's Global Forecast System.
3 | TRADITIONAL AND CRA VERIFICATION RESULTS

Figure 2 shows both the traditional and the CRA rainfall verification results of TC Vicente (2012) making landfall in China. The traditional categorical performance measures including ETS, POD, FAR and extremal dependence index (EDI; Ferro and Stephenson, 2011) are given (their definitions are shown in Appendix Table A1), with a threshold of 30 mm. Since the rain forecast errors are decomposed by the CRA method into the four contributions of rotation, displacement, volume and pattern errors, the sources of errors can be detected by the CRA analysis. The ETS score of 0.36 and visual examination of the observed and forecast rain distributions indicate that this is a reasonably skilful forecast, based on comparisons with other verification results for other rain events, using, for example, a 50 mm rainfall threshold (data not shown). When the forecast rain field is adjusted via shifting and rotating to best fit the observed rain, the CC increases from 0.25 to 0.45. However, the ETS score does not change. The errors are mostly related to a pattern error, and to a lesser extent displacement error. As will be discussed below, these results are indicative of the systematic errors present in the ACCESS-TC forecasts.

Table 2 shows the total sample sizes of forecasts for 30, 50, 100 and 250 mm rainfall thresholds and for different forecast lead times (0–6, 0–24, 24–48, 48–72 and 0–72 hr). While the 0–24 and 0–72 hr-accumulated rainfall forecasts are < 100 mm, the sample sizes are relatively large (near 130), whereas for the 250 mm rainfall forecasts, the sample numbers decrease to < 100. Especially for the shorter (0–6 hr) and longer (24–48 and 48–72 hr) lead times, the sample numbers for the 250 mm rainfall decrease to < 50. Therefore, note the relatively limited forecast sample sizes that might cause some uncertainties in the verification results.
for extremely heavy rainfall (250 mm), as well as shorter (0–6 hr) and longer (48–72 hr) lead-time forecasts.

Based on the entire forecast sample, results from traditional verification, including the ETS, POD, FAR and EDI for the rainfall thresholds of 30, 50, 100 and 250 mm and different lead-time forecasts, are shown in Figure 3. Generally, the 0–24 hr-accumulated rainfall forecasts have the highest performance. Means of the ETS, POD and FAR for the 24 hr 30 mm rainfall forecasts are 0.37, 0.67 and 0.40, respectively. However, for 6 and 72 hr 30 mm forecasts, the performance measures worsen to 0.23, 0.55, 0.65, and 0.16, 0.38, 0.66, respectively. In addition, the average ETSs of the 24 hr rainfall forecasts decrease with increasing rainfall. For 100 mm rainfall, the mean ETS of the 24 hr forecasts reduces to 0.21, and then decreases quickly to 0.05 when the 24 hr rainfall amount increases to 250 mm. However, the 0–72 hr-accumulated rainfall has comparable mean ETSs with the 24 hr rain forecasts, while the mean POD, FAR and EDI show similar variations. Therefore, on average, the 24 hr 30 mm rainfall forecasts show the best performance compared with other lead times and thresholds.
As discussed above and from Table 2, since the sample numbers are relatively larger for the 0–24 hr and < 100 mm for the 0–72 hr-accumulated rainfall forecasts, the results represent reasonable for normal distributions. However, for some results, such as 0–6 hr 100 mm and 48–72 hr 250 mm rainfall forecasts, the boxes are somewhat small or large, or the medians are skewed (not normally distributed; Figure 3). Thus, the uncertainty of the results should be noted, especially for the 0–6 and 48–72 hr 250 mm rainfall forecasts, since they are likely influenced by the limited sample sizes (Table 2). In addition, there are some outliers even for the forecasts with the same rain threshold and the same lead time across the cases. For example, the maximum ETS for the 24 hr 30 mm rainfall forecasts is about 0.65, while the minimum is nearly 0.1, which indicates that there is uncertainty for the rainfall forecasts of different TC samples as well. The source of the variability remains unclear and deserves further investigation.

| Lead time (hr) | 30 mm | 50 mm | 100 mm | 250 mm |
|---------------|-------|-------|--------|--------|
| 0–6           | 131   | 125   | 50     | 1      |
| 0–24          | 133   | 133   | 128    | 42     |
| 24–48         | 126   | 124   | 113    | 32     |
| 48–72 hr      | 82    | 82    | 71     | 15     |
| 0–72          | 133   | 133   | 133    | 90     |

**Figure 3** Conventional verification results of equitable threat score (ETS), probability of detection (POD), false alarm ratio (FAR) and extremal dependence index (EDI) for four rainfall thresholds of 30, 50, 100 and 250 mm (i.e. every four columns from left to right, except for 0–6 hr forecasts), respectively, for accumulated rain forecast at lead times of 0–6, 0–24, 24–48, 48–72 and 0–72 hr (i.e. “6”, “24”, “48”, “72” and “720” shown on the x-axis, respectively). The y-axis shows the performance measure of the forecasts. The box-and-whisker plot shows medians, upper and lower quartiles with black lines, and outliers with black dots.
The paper now presents comparisons with other rainfall forecast-verification results. A comparison made is against Wang et al. (2012) for landfalling TCs over China. For the JMA global model rainfall verification, the mean ETSs are 0.22, 0.19 and 0.15 for the 24, 48 and 72 hr forecasts for the 25 mm-threshold rainfall. These values are much lower than the mean ETSs of 0.37, 0.25 and 0.16 for the 24, 48 and 72 hr forecasts of the 30 mm-threshold rainfall reported in the present study. The present authors also checked another published work of more general rainfall forecasts, such as the Meiyu season rainfall of Wang et al. (2017). The mean ETSs reported are about 0.2 for the 24 hr 25 mm-threshold rainfall forecasts. Since those results were obtained based on different rainfall samples or rain events, one cannot focus too much on the direct comparisons of the performance measure scores. However, there is the advantage of slightly higher resolution and more sophisticated vortex initialization procedures. Therefore, a fairly reasonable ability of landfalling TC rainfall forecasts is demonstrated in the present study.

With regard to the verification reported here, the noteworthy points are as follows. First, the short-period 0–6 hr rainfall accumulations are difficult to forecast since this period is short and the initialization cannot guarantee the vortex structure, location and intensity of rain bands and convective systems. The latter are defined in the 4D VAR initialization and there is only limited evidence (Davidson et al., 2014) that these smaller scale rainfall features are adequately initialized. Second, just as the skill of track and intensity forecasts diminishes with increasing forecast lead time, the rain forecasting ability also decreases. However, what is the origin of the forecast rain errors? They may be connected to the resolution of the system and also to the initialization and prediction of the vortex structure, as well as

**Figure 4** Rain event verification for 30, 50, 100 and 250 mm rainfall thresholds with different forecast lead times of 0–6, 0–24, 24–48, 48–72 and 0–72 hr, respectively. “720 hr” means the accumulated rainfall between 0 and 72 hr. Numbers in parentheses show the fraction and percentage of each category. Darker colours indicate higher percentages, showing which categories are dominant.
the parameterized moist processes. Third, the verification results for 30 mm rainfall are quite encouraging. Rain amounts > 30 mm tend to occur in smaller scale regions and thus are more difficult to forecast. In addition, since the physical processes leading to extremely heavy rain are not well understood, it is very difficult to determine if such processes can be represented by the model.

4 | CRA VERIFICATION RESULTS FOR LANDFALLING TC RAINFALL

By treating TC forecast rainfall as objects, the rainfall can be classified by a hit and false alarm, and so on, according to whether its location and intensity are accurately considered (Ebert and McBride, 2000). The specific criteria that defines a “hit” (or “well predicted”) depends on the user’s needs. In the present study, the subjective criteria include a “close” rain centre (location error < 100 km) and an “approximate” rain volume (forecast volume error within 25% of the observed volume). Figure 4 shows the frequencies of six kinds of rain events (represented as overestimate, false alarm, hit, missed location, underestimate and missed event) for 30, 50, 100 and 250 mm rainfall thresholds and different forecast lead times (0–6, 0–24, 24–48, 48–72 and 0–72 hr). Less, approximate and more shown on the x-axes represent the forecasted rain amount errors. Far and close shown on the y-axes represent the forecasted rain centre location errors. Compared with the observed, if the forecast has both an approximate rain amount and a close rain centre, the rain forecast event is considered a “hit”. If the forecast has both a larger rain amount and a distant rain centre, the rain forecast event is counted as a “false alarm”. If the forecast has both a much-reduced rain amount and a distant rain centre, the rain forecast event is counted as a “missed event”.

Figure 4 shows that, overall, the 24 hr forecast rain events have the highest hit rate of 29%, and the 72 hr-accumulated rainfall forecast has a quite comparable hit rate of 23%. Even for 50 mm rain, the 72 hr-accumulated rainfall forecast has a little higher hit rate (20%) than that for the 24 hr rainfall forecast (19%), while the 24 hr rainfall forecast has a higher rate of overestimation (48%) and underestimation (2%).

**FIGURE 5** Contiguous rain area (CRA) error decomposition of rain events for 30, 50, 100 and 250 mm rainfall thresholds with different forecast accumulation periods of 0–6, 0–24, 24–48, 48–72 and 0–72 hr. “720 hr” means the accumulated rainfall between 0 and 72 hr. R, D, V and P represent rain errors from rotation (R), displacement (D), volume (V) and pattern (P), respectively. The box-and-whisker plot shows medians with short red lines, upper and lower quartiles with black lines, and outliers with red crosses.
results indicate that the 72 hr-accumulated rain could offset some errors brought by longer lead-time forecasts (i.e. 24–48 and 48–72 hr). The longer averaging reduces some of the timing errors. For longer and shorter lead-time forecasts, the hit rate decreases, which is understandable based on the discussion above. In addition, with the increase of rain amount to 250 mm, the hit rate decreases rapidly to 5%, and the false alarm rate increases to 48% for the 24 hr forecasts.

The rain error sources for different rain thresholds and different lead-time forecasts are further analysed by the CRA error decomposition method and shown via box-and-whisker plots (showing medians with shorter red lines, upper and lower quartiles with black lines, and outliers with red crosses) in Figure 5. For the 24 hr 30 mm rainfall forecasts, the maximum forecast error is from the pattern error (P) on average (> 50%), which suggests that the spatial distribution of the rain is worst-predicted. This result in turn suggests that the initial vortex structure may not be accurately defined in the initialization or predicted well by the model. With the increase of forecast lead time and rain amount, the error source contribution from the displacement error (D) increases. Especially for the 250 mm rainfall predicted by 24, 48 and 72 hr forecasts, the averaged maximum error is contributed by the displacement of the rain centre. This result could be interpreted as mainly related to the forecast track and possibly intensity errors. However, for all rain forecasts, the mean error contributions from both rotation (R) and volume (V) are little changed and remain small. It is possible that since the rotation errors are small and the convective asymmetries are in the correct sector of the storm, the model generally represents the interaction between the environmental wind shear and the TC quite well. However, because of errors in the vortex structure and limitations with resolution, the rainfall patterns may be in error.

**FIGURE 6**  As for Figure 3, but for new results of equitable threat score (ETS), probability of detection (POD), false alarm ratio (FAR) and extremal dependence index (EDI), respectively, after adjustments based on the contiguous rain area (CRA) method.
Since the forecast rain of a TC is treated as an entity, the rain errors from D and R can be detected and rectified for the TC. In this way, the systematic errors in forecast rain can be further determined. As described in Section 2.4, the forecast rain object can be moved and rotated by displacement and rotation adjustment of the CRA analysis method so that it matches the observed rain area best, thereby minimizing the forecast errors. After such adjustment, the 24 hr rain forecasts still have the best performance measures on average (Figure 6).

However, the most notable difference is that the 48 and 72 hr rainfall forecasts have greatly improved performance overall compared with the original unshifted forecasts. The means of the ETS for the 48 and 72 hr 30 mm rainfall forecasts are improved from 0.25 and 0.16 to 0.33 and 0.29, respectively. In addition, the mean ETSs for the 48 and 72 hr 250 mm rainfall forecasts have been greatly improved both from near 0 to about 0.16. However, since the 72 hr 250 mm forecast samples are very limited, there is large uncertainty and thus further analysis is still needed in future.

As seen in Figure 7, the 250 mm rainfall forecasts are the most improved, compared with the other rainfall thresholds. Since rain centres are mainly located near TC centres (Yu et al., 2015; 2017), TC track errors would directly affect the locations of the rain centre. Through the shifting and rotating adjustment of the CRA, both the longer lead-time forecasts and larger rain-amount forecasts are greatly improved, which could indicate that track errors may mostly contribute to the rain centre errors, especially for extremely heavy rain centre locations.

**FIGURE 7** As for Figure 6, but for differences between the original and new results of (a) equitable threat score (ETS), (b) probability of detection (POD), (c) false alarm ratio (FAR) and (d) extremal dependence index (EDI) for 30, 50, 100 and 250 mm rainfall thresholds, respectively, after adjustments based on contiguous rain area (CRA) method.
The fact that, except for the special case of the 6 hr forecast, the averaged scores are roughly constant and near to the 0–24 hr forecast scores after adjustment (Figure 6) suggests that the performance of rain forecasts will be improved with improved track predictions. Besides, it also indicates that further improvements of the 0–6, 0–24 and 0–72 hr rain forecasts still need additional developments in the overall forecast system, such as resolution, vortex structure initialization and parameterizations. Figure 7 shows that from the 24 to the 72 hr forecasts, the 72 hr rain forecast is the most improved, reinforcing the view that errors in rain centre location may be related to track errors in the longer lead-time forecasts. This is also supported by examining the CC for original forecasts and the CC difference between the original and new results after shifting and rotating (Figure 8), which show that the mean CC is greatly improved for all forecast lead times.

However, the three-day accumulated rainfall forecast is the least improved based on the displacement and rotation adjustment method (Figure 7), probably because the rain pattern error is the largest source of error even for extremely heavy rainfall of 250 mm accumulated for the three day forecasts (Figure 5). Therefore, to improve the rainfall forecasts, both the TC track (associated with the rain centre location) and TC structure (associated with the rain pattern) are essential for more accurate rainfall predictions.

In this section, analyses have been made from the samples shown in Table 2. Note again that though the statistical results are useful, there are some uncertainties likely related to the limited sample sizes, especially for extremely heavy rainfall (250 mm), as well as shorter (6 hr) and longer (72 hr) lead-time forecasts. Therefore, in this aspect, the results can act as a useful benchmark or a reference for future work.

5 | DETAILED ANALYSIS OF TWO TC RAINFALL FORECASTS

5.1 | Rapidly intensifying TC Rammasun (2014)

The landfall of TC Rammasun is analysed to illustrate the model forecast performance. Figure 9 shows the distributions of the original and shifted ACCESS-TC rainfall forecasts at the lead time of 24 hr. Observed and forecast tracks for Rammasun are shown in Figure 9a. The track is well forecasted and close to the best track (dashed and solid lines). The CRA verification result shows that displacement and rotation errors are very small contributions to the total error, which is likely related to the good track forecast, while the rain pattern error is the largest error source (accounting for 90% of the total errors). This might be related to the poor forecast of the TC structure and intensity-change during landfall (Figure 10). Some recent studies (Alvey et al., 2015; Harnos and Nesbitt, 2016; Yu and Wang, 2018; Yu et al., 2017) have confirmed that, on average, TC intensity and its change are closely related to the rainfall and rainfall change over both ocean and land. Stronger TCs would exhibit the higher axisymmetry structure as well as rainfall. Figure 10 shows that Rammasun experienced a rapid intensification (RI, 24 hr intensity changes ≥ 15 m s\(^{-1}\)) before landfall. However, the forecasted central pressure gives no indication of its intensification. Though the poor intensity forecast of Rammasun may be related to the relatively coarse spatial resolution of the operational forecast system, note that the RI is difficult to forecast and not regularly captured by numerical models (Gall et al., 2013; Knaff et al., 2018).
Figure 11 shows the 6 hr comparison between the satellite rain distribution and the model forecast rain during the period from 24 hr before landfall to 24 hr after landfall. Before landfall, the observed Rammasun was still undergoing its RI (1706–1800 UTC), when there was a clear eye with very little rain and active eyewall cloud with heavy rain (Figure 11a). During landfall, the clear eye disappeared (1806 UTC). At 24 hr after landfall, the rain distribution was very asymmetric (1906 UTC). However, the forecasted rain distribution is very asymmetric before landfall (1706–1800 UTC) (Figure 11b). The circular distribution of rain around the TC centre during the RI before landfall is not predicted.

The maximum rain area is always located to the southwest of the TC centre, unlike the observations that show a much more symmetric distribution. These forecast rain errors again point to the need to improve the initialization of TC structure, and for increased resolution.

5.2 Illustrative example of TC Kalmaegi (2014)

The case of Kalmaegi is chosen to illustrate rainfall performance for skillful forecasts of both TC track and intensity during the period from 24 hr before landfall to 24 hr after

![Figure 9](image-url)
landfall (Figures 12 and 13). In the rainfall verification analysis, the average forecast rain rate is 75 mm day$^{-1}$, much higher than the observed 47 mm day$^{-1}$ (Figure 12). Also, the maximum forecast rain rate is 532 mm day$^{-1}$, more than double the observed 222 mm day$^{-1}$. The CRA verification has shown that the main error source in rainfall forecasts is the pattern error (70%) and volume error (18%), while the displacement and rotation errors are relatively small for this case. The trends in the observed and forecast TC intensity changes are quite close (Figure 13), with only relatively small differences. These results indicate that the model predicts both the track and the intensity of the TCs quite well with a reasonable forecast rain centre, but the rain rate and rain pattern are still challenging forecast problems. Interestingly the model produces significantly more rain, but the observed and predicted TC intensities (central pressure) are very similar. This apparent inconsistency requires more detailed analysis.

6 | DISCUSSION AND CONCLUSIONS

The aim of the present study is to provide benchmark verification for landfalling tropical cyclone (TC) TC rainfall forecasts from an operational forecast system. In total, 25 landfalling TCs over China between 2012 and 2015, and 133 forecasts from operational ACCESS-TC model are included in the study. The verification of the TC track and intensity for the ACCESS-TC indicates the forecasts are skilfully compared with other generally available international forecast systems.

The contiguous rain area (CRA) verification method is used to evaluate the ACCESS-TC model rain forecasts. The gauge-CMORPH (Climate Precipitation Center Morphing) merged rain data from the National Meteorological Information Center (NMIC) of the China Meteorological Administration (CMA) are used as reference rainfall data, which can avoid the problem that satellite data (such as TRMM 3B42) could underestimate the TC rainfall over land. The traditional verification methods of the equitable threat score (ETS), false alarm ratio (FAR) and probability of detection (POD) show that the 24 hr rainfall forecasts have the best scores on average, while with increases in lead times to 48 and 72 hr, forecast ability is reduced, similar to track and intensity. The 0–6 hr forecast, essentially the initialization, has only modest scores and points to the need of improving the initialization of TC structure and clouds. However, the accumulated rainfall during 0–72 hr shows comparable performance with the 24 hr rain forecasts. It would imply that for shorter lead-time forecasts, the rainfall patterns or structures may be in error spatially or temporally, but owing to the longer accumulation time, the smoothing associated with 0–72 hr accumulations does help mask some forecast timing and location errors.

The error decomposition results of the CRA method have shown that for 24 hr 30 mm rainfall forecasts, the maximum forecast error source is from pattern errors is from pattern errors overall. However, with an increase in forecast lead time and rain amount, the contribution of the displacement error to the total rain error increases. These results indicate that for the rainfall below a threshold of 100 mm, the main rain forecast errors generally come from pattern errors. However, for heavy rain > 250 mm, the largest rain errors are caused by rain centre displacements.

After displacement and rotation adjustments of the CRA method, the averaged ETSs of the rainfall are improved, which is mainly due to the improvement of rain centre locations, especially for extremely heavy rain (250 mm) and longer lead-time (48 and 72 hr) forecasts.

Therefore, compared with traditional verification methods such as ETS, the CRA verification method provides more detailed information of the TC rain forecast performance, including the main error source of numerical model forecast rainfall. In turn, this can be used to help researchers and forecasters further identify research challenges and model weaknesses in rain forecast.

An important uncertainty of the current study that should be noted is the limited sample sizes, particularly for shorter (6 hr) and longer (72 hr) lead-time forecasts of heavy rain (especially 250 mm rain). Box-and-whisker plots were used to show the variation in the results across cases, and that variability is an indicator of the uncertainty in the results. Therefore, the uncertainties can be more easily assessed. Further work is needed to consolidate the current results. The present study provides a preliminary but still useful benchmark for future verification.
Another issue is the quality of rain analyses used for verification. Since the satellite-retrieved rain data have some differences from gauge observations, there are potential uncertainties in the analyses. However, the authors do not think that this possible error source will affect the main conclusions in the study.

The TC case studies in the study have shown that rainfall structure and large rain rates continue to be challenging problems for numerical weather prediction (NWP). The results suggest that rainfall prediction will continue to be improved with improved track prediction, but resolution, initialization, prediction of TC structures and parameterization of moist processes are still limitations for prediction performance.

The present study has documented the current forecast performance of the operational ACCESS-TC for landfalling
TC rainfalls over China. It has described the main sources of forecast rain errors and highlighted where improvements could be made. This has been achieved by using the CRA verification method. Other object-based verification methods such as the method for object-based diagnostic evaluation (MODE) would be also very useful for landfalling TC rainfall verification. In future, the authors plan (1) to extend the verification to many more landfalling TC cases, to include further consideration of TC intensity, size and their changes more adequately; (2) to use the CRA results to try to understand better the reasons for good and poor forecasts; and (3) to obtain data from other NWP systems to enable a

![Image](FIGURE 12) Contiguous rain area (CRA) verification results of the 24 hr rainfall forecast for tropical cyclone (TC) Kalmaegi (2014): (a) merged analysis of rainfall, observed and forecast tracks; (b) original ACCESS-TC forecast rain distribution; (c) distribution of the shifted ACCESS-TC forecast rain based on adjustments from the CRA method; and (d) rain amount comparison between ACCESS-TC and the merged rain analysis

![Image](FIGURE 13) Observed (blue line) and ACCESS-TC forecast (red line) intensity (minimum sea-level pressure [MSLP], unit: hPa) for Kalmaegi from 24 hr before landfall to 24 hr after landfall
comparison of rainfall forecasts. In this way, the study represents a benchmark to encourage other modellers or researchers to use object-based verification methods such as the CRA to explore further model forecast performance for landfalling TCs.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the valuable suggestions made by Drs Zhian Sun, Hongyan Zhu, Yimin Ma and David Smith from the Bureau of Meteorology, Australia. Dr Baode Chen from the Shanghai Meteorological Service also provided some important opinions on this work. Thoughtful comments from the anonymous reviewers clarified many issues and greatly improved the presentation. The work was supported in part by the Key Program for International S&T Cooperation Projects of China (grant number 2017YFE0107700), the Scientific Research Program of Shanghai Science and Technology Commission (grant number 19dz1200101), the National Natural Science Foundation of China (grant numbers 41875080, 41775064 and 41475058), and in part by GASI-IPOVAI-04 and Shanghai Typhoon Innovation Team grants to the Shanghai Typhoon Institute (STI).

ORCID

Zifeng Yu https://orcid.org/0000-0002-2513-4086

REFERENCES

Alvey, G.R., III, Zawislak, J. and Zipser, E. (2015) Precipitation properties observed during tropical cyclone intensity change. Monthly Weather Review, 143, 4476–4492. https://doi.org/10.1175/MWR-D-15-0065.1.

Chen, Y., Ebert, E.E., Walsh, K.J.E. and Davidson, N.E. (2013) Evaluation of TRMM 3B42 precipitation estimates of tropical cyclone rainfall using PACRAIN data. Journal of Geophysical Research: Atmospheres, 118, 1–13. https://doi.org/10.1002/jgrd.50250.

Chen, Y., Ebert, E.E., Davidson, N. and Walsh, K. (2018) Application of contiguous rain area (CRA) methods to tropical cyclone rainfall forecast verification. Earth and Space Science, 5, 736–752. https://doi.org/10.1029/2018EA000412.

Davidson, E.N., et al. (2014) ACCESS-TC: vortex specification, 4DVAR initialization, verification, and structure diagnostics. Monthly Weather Review, 142, 1265–1289.

Davis, C., Brown, B. and Bullock, R. (2006) Object-based verification of precipitation forecasts. Part I: methods and application to mesoscale rain areas. Monthly Weather Review, 134, 1772–1784.

Davis, C., Brown, B., Bullock, R. and Halley-Gotway, J. (2009) The method for object-based diagnostic evaluation (MODE) applied to numerical forecasts from the 2005 NSSL/SPC spring program. Weather and Forecasting, 24, 1252–1267.

Ebert, E.E. and Gallus, W.A., Jr. (2009) Toward better understanding of the contiguous rain area (CRA) method for spatial forecast verification. Weather and Forecasting, 24, 1401–1415. https://doi.org/10.1175/2009WAF2222521.1.

Ebert, E.E. and McBride, J. (2000) Verification of precipitation in weather systems: determination of systematic errors. Journal of Hydrology, 239, 179–202. https://doi.org/10.1016/S0022-1694(00)00343-7.

Ebert, E.E., Turk, M., Kusselson, S.J., Yang, J., Seybold, M., Keehn, P.R. and Kuligowski, R.J. (2011) Ensemble tropical rainfall potential (eTRaP) forecasts. Weather and Forecasting, 26, 213–224.

Ebert, E.E., Wilson, L.J., Brown, B.G., Nurmi, P., Brooks, H.E., Bally, J. and Jaeneke, M. (2004) Verification of nowcasts from the WWRP Sydney 2000 forecast demonstration project. Weather and Forecasting, 19(1), 73–96.

Ferro, C.A.T. and Stephenson, D.B. (2011) Extreme dependence indices: improved verification measures for deterministic forecasts of rare binary events. Weather and Forecasting, 26, 699–713.

Gall, R., Franklin, J., Marks, F., Rappaport, E.N. and Toepfer, F. (2013) The hurricane forecast improvement project. Bulletin of the American Meteorological Society, 94, 329–343. https://doi.org/10.1175/BAMS-D-12-00071.1.

Gilleland, E., Ahijevych, D., Brown, B. and Ebert, E. (2009) Intercomparison of spatial forecast verification methods. Weather and Forecasting, 24, 1416–1430. https://doi.org/10.1175/2009WAF2222691.1.

Glems, J.S., Gallus, W.A., Jr., Koch, S.E., Wharton, L.S., Lougue, A. and Ebert, E.E. (2006) The use of a modified Ebert McBride technique to evaluate mesoscale model QPF as a function of convective system morphology during IHOP 2002. Weather and Forecasting, 21(3), 288–306. https://doi.org/10.1175/WAF918.1.

Harnos, S.D. and Nesbit, S.W. (2016) Varied pathways for simulated tropical cyclone rapid intensification. Part I: precipitation and environment. Quarterly Journal of the Royal Meteorological Society, 142, 1816–1831. https://doi.org/10.1002/qj.2780.

Knaff, A.J., Sampson, C.R. and Musgrave, K.D. (2018) An operational rapid intensification prediction aid for the Western North Pacific. Weather and Forecasting, 33, 799–811. https://doi.org/10.1175/WAF-D-18-0012.1.

Lei, X., Yu, H., Chen, G., Brown, B., Chan, S.T., and Tyagi, A. (2016) WMO typhoon landfall forecast demonstration project (WMO-TLFD) progress and future plans. Tropical Cyclone Research and Review, 5, 72–84.

Leroux, M.-D., Wood, K., Elsberry, R.L., Cayan, E., Hendricks, E.O., Kucar, M., Otto, P., Rogers, R., Sampson, C.R., and Yu, Z. (2018) Recent advances in research and forecasting of tropical cyclone track, intensity, and structure at landfall. Tropical Cyclone Research and Review, 7(2), 85–105.

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

Moise, A.F. and Delage, F.P. (2011) New climate model metrics based on object-orientated pattern matching of rainfall. Journal of Geophysical Research, 116, D12108.

Shen, Y., Xiong, A., Wang, Y. and Xie, P. (2010) Performance of high-resolution satellite precipitation product over China. Journal
of Geophysical Research, 115, D02114. https://doi.org/10.1029/2009JD012097.
Shen, Y., Zhao, P., Pan, Y. and Yu, J. (2014) A high spatiotemporal gauge-satellite merged precipitation analysis over China. *Journal of Geophysical Research*, 119, 3063–3075. https://doi.org/10.1002/2013JD020686.
Wang, C.-C., Chien, F., Paul, S., Lee, D. and Chuang, P. (2017) An evaluation of WRF rainfall forecasts in Taiwan during three Mei-Yu seasons from 2008 to 2010. *Weather and Forecasting*, 32, 1329–1351.
Wang, Y., Shen, X. and Chen, D. (2012) Verification of tropical cyclone rainfall predictions from CMA and JMA global models. *Journal of Tropical Meteorology*, 18, 537–542. https://doi.org/10.16555/j.1006-8775.2012.04.014.
Wernli, H., Paulat, M., Hagen, M. and Frei, C. (2008) SAL—A novel quality measure for the verification of quantitative precipitation forecasts. *Monthly Weather Review*, 136, 4470–4487.
Yu, Z. and Wang, Y. (2018) Rainfall distribution in landfalling tropical cyclones. In: Sallis, P. J. (Ed.) *Extreme Weather*. London, UK: IntechOpen. https://doi.org/10.5772/intechopen.75910.
Yu, Z., Wang, Y. and Xu, H. (2015) Observed rainfall asymmetry in tropical cyclones making landfall over China. *Journal of Applied Meteorology and Climatology*, 54, 117–136.
Yu, Z., Wang, Y., Xu, H., Davidson, N.E., Chen, Y., Chen, Y. and Yu, H. (2017) On the relationship between intensity and rainfall distribution in tropical cyclones making landfall over China. *Journal of Applied Meteorology and Climatology*, 56, 2883–2901.
Yu, Z., Yu, H., Chen, P., Qian, C. and Yue, C. (2009) Verification of tropical cyclone-related satellite precipitation estimates in Mainland China. *Journal of Applied Meteorology and Climatology*, 48, 2227–2241. https://doi.org/10.1175/2009JAMC2143.1.

**How to cite this article:** Yu Z, Chen YJ, Ebert B, et al. Benchmark rainfall verification of landfall tropical cyclone forecasts by operational ACCESS-TC over China. *Meteorol Appl.*, 2020;27: e1842. https://doi.org/10.1002/met.1842

**APPENDIX A**

A threshold value (e.g. 30 mm-day\(^{-1}\)) is chosen to separate rain from no-rain events. Here, \(Z\) is the number of correct predictions of rain amount below the specified threshold; \(F\) is the number of false alarms; \(M\) is the number of misses; and \(H\) is the number of correct rain forecasts or hits.

| TABLE A1 | Rain contingency table applied at each verification grid over the period of verification |
|-----------------|-----------------|-----------------|-----------------|
| **Observed**    | **Forecast**    | **No rain**     | **Rain**        |
| No rain         | \(Z\)           | \(F\)           |                 |
| Rain            | \(M\)           | \(H\)           |                 |

Several conventional categorical statistics such as equitable threat score (ETS) are also applied in the present study. The term “categorical” refers to the “yes”/“no” nature of the verification at each verification grid. Some thresholds (i.e. 30, 50, 100 and 250 mm) are considered to define the transition between a rain versus a no-rain event. At each grid, each verification time is then scored as falling under one of the four categories of the correct no-rain estimate, false alarms, misses or hits (\(Z\), \(F\), \(M\), or \(H\), as shown in Table A1).

The ETS is given by:

\[
ETS = \frac{H - CH}{H + F + M - CH},
\]

where \(CH = (H + M) \cdot (H + F)/(Z + H + M + F)\), and it is determined by assuming that the estimates are totally independent of the observations, and the estimate will match the observation only by chance. This is an unskilled estimate, which can be generated by just guessing what will happen. The ETS ranges between \(-1/3\) and 1. The minimum depends on the verification sample climatology.

In addition, false alarm ratio (FAR) is calculated as:

\[
FAR = \frac{F}{H + F}.
\]

The probability of detection (POD) is defined as:

\[
POD = \frac{H}{H + M}
\]

Finally, the extremal dependence index (EDI) is also calculated as:

\[
EDI = \frac{(\log FAR - \log HIT)}{\log FAR + \log HIT}
\]

where \(FAR = F/(H + F)\); and \(HIT\) is the hit rate.