Ten-year Performance of HWRF Model in RI Forecasts -- A New Metric

Weiguo Wang, Bin Liu, Zhan Zhang, Lin Zhu, Avichal Mehra* and Vijay Tallapragada*
IMSG@EMC/NCEP/NWS,*EMC/NCEP/NWS, College Park, MD 20740
Email: Weiguo.Wang@noaa.gov

1. Introduction

Forecasts of rapid intensification (RI) of tropical cyclones (TC) are still a challenge, in spite of improvements in track and intensity forecasts in the past decade. RI is a scenario where the intensity of a TC increases dramatically in a very short period of time. In practice, RI is defined as an increase in the maximum sustained winds of a TC equal to or greater than 30 knots (55 km/h) in a 24-hour period (Kaplan and DeMaria, 2003). Improving the ability of NCEP hurricane models to forecast RI events is a top priority for EMC developers. Currently, the probability of detection (POD) and false alarm ratio (FAR) are routinely used to measure the performance of RI forecasts. POD of RI is quantified as a percentage of the total number of observed individual RI events which are correctly forecasted, while FAR of RI is a percentage of RI forecasts that were not RI events based on observations. While this method is effective in assessing the overall RI forecast performance of a model, it is not straightforward in revealing how well individual forecasts over a period of time (e.g., typically 5 days for mesoscale models) perform in capturing RI events. In other words, a POD may not be able to reflect how many 5-day forecasts successfully capture RI events. This is because there may be multiple RI events during a 5-day period. To this end, we proposed a new metric, which is based on the total number of RI events forecasted during the whole integration time in a model. This gives modelers a direct assessment of the number or percentage (i.e., success rate) of 5-day forecast cycles capturing some or all of RI events. With the new metric, we calculated success rates of RI forecasts in a 5-day period by NCEP HWRF in the past decade, showing the model is improving RI forecasts.

2. Methodology

The question we would like to answer is how many 5-day model integrations (cycles) can successfully capture one or more observed RI events based on best-track data. For a threshold of wind speed increase, RI events can be identified as binary (yes or no) every six hours during a 5-day integration period. The same procedure is applied to observational data (e.g., NHC best-track data). Then one can compare the results from the model with observational data, and determine how many observed RI events have been captured by the 5-day forecast. To illustrate the method, Figure 1 presents an example of a 5-day time series of the maximum 10-m wind speed of Hurricane Lorenzo (2019) forecasted by the operational HWRF model initialized at 18 UTC, September 24, 2019. RI events (with the threshold of 30 kt) are identified every 6 hours, denoted by blue crosses for observations and triangles for HWRF. The best-track data suggests RI cases occurred at 10 lead times (hours). RI cases predicted by HWRF occurred at 4 lead times. Triangles in red indicate that the occurrence times of RI events simulated by HWRF exactly match those of observations. In this example, RI cases at the 36th, 42nd, and 48th hours are successfully captured by HWRF. HWRF produced one false alarm prediction at the 54th hour and failed to capture RI at seven lead times (hours).

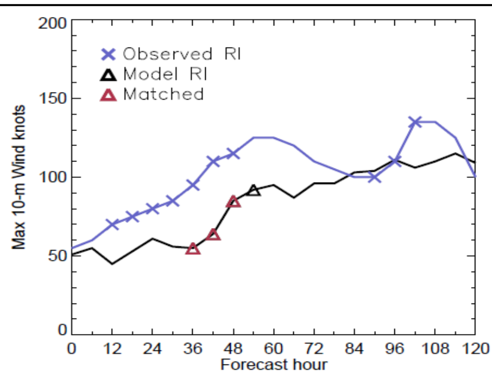

FIG 1 Time series of maximum 10-m wind speed of 5-day periods from a HWRF forecast (black line) and from observation data (blue line).

Given the uncertainties in numerical models and errors in observations, multiple criteria or thresholds can be used to determine whether a 5-day forecast by HWRF is a success or not. Three criteria were tested, depending on what percentage of observed RI cases have been produced in a 5-day forecast by HWRF. The first requires that all observed RI cases must be forecasted by HWRF with matching RI occurrence times. This is a very strict criterion, particularly for multiple RI events occurring in a 5-day forecast period. The second is somewhat relaxed, and requires that half of observed RI cases are produced (with time matching). The third is most relaxed criterion, requiring that at least one observed RI case is produced (with time matching). A false alarm RI cycle is defined as a 5-day forecast during which HWRF predicts RI events but there are no observed RI events in that period. A success rate is a percentage of total 5-day periods with observed RI events successfully produced by HWRF. In addition, due to the fact that observation error in intensity can be greater than 5 knots, we used 20 and 30 knots as the thresholds of the increase in wind speed when determining RI cases for comparisons.

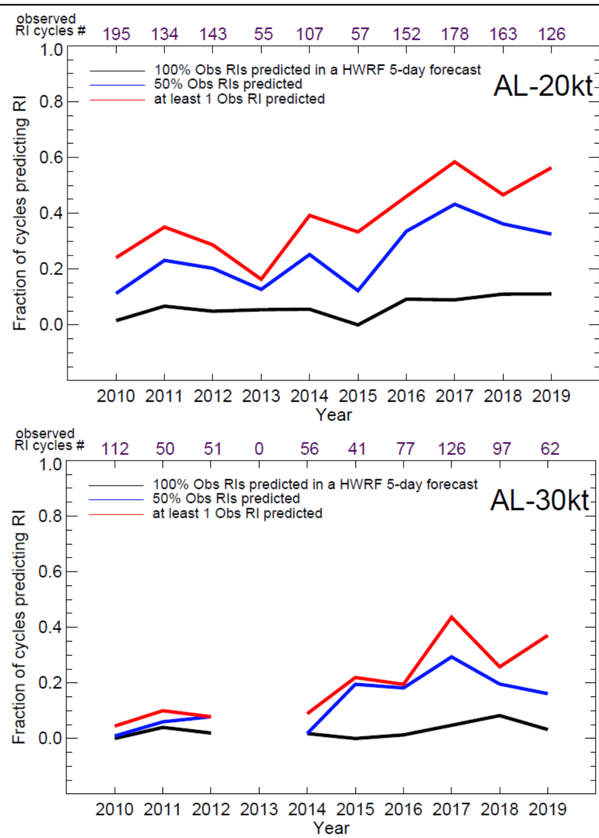

FIG. 2 Success rates of 5-day forecasts predicting observed RI events of TCs in NATL basin by operational HWRF model from 2010 to 2019. top: 20-kt as RI threshold , bottom: 30 kt.

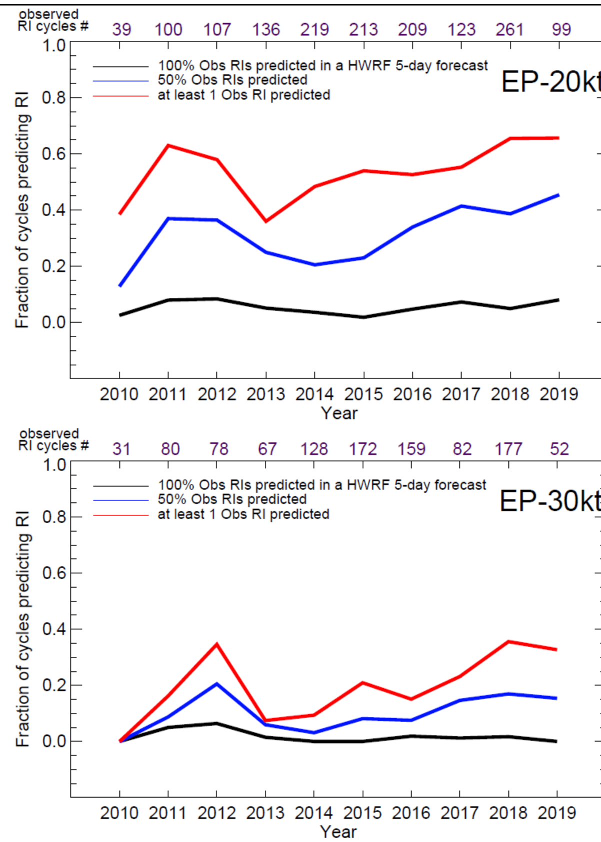

FIG. 3 as FIG. 2 except for TCs of EPAC basin.

3. Results and discussion

With the method described above, we computed success rates of 5-day forecasts for the North Atlantic (NATL) basin (Fig. 2) and East Pacific (EPAC) basin (Fig. 3) by the operational HWRF model over the past 10 years. Depending on the criterion, the fraction of the number of cycles producing observed RI events increases with years. The success rate is higher with a lower RI threshold. For the 30-kt RI threshold, as much as 40% of the cycles can predict at least one RI event during a 5-day forecast. This number increases to 60-70% for the 20-kt RI threshold. However, the success rate of cycles in predicting all observed RI events in a 5-day forecast (with the time matching requirement) is still very low, though there is an increasing trend in the NATL basin. Out of the total 5-day periods when observations did not show RI events, approximately 20% of forecasts (in the same period) predict at least one RI event, giving false alarms. This number is not changed much over the years. The improvement in success rate is attributed to yearly upgrades of the HWRF model, especially model horizontal resolution increases from 9 to 1.5 km, and in the number of vertical levels from 43 to 75 km. Tuning and calibrations of parameters in the model physics schemes, PBL and convection schemes for example, also play a key role in the improvement of the intensity and intensity change forecasts. We also calculated the intensity RMS error and bias during RI events each year for the NATL basin, showing that mean bias is significantly reduced with time from –18kt (2009) to –8 kt (2019), though RMS error curve is flat at approximately 20 kt.