Combining Monte Carlo and Ensemble Probabilities in Tropical Cyclone Forecasts near Landfall

Xin LI[1], Zhaoxia PU[1*], and Zhiqiu GAO[2]

1 *Department of Atmospheric Sciences, University of Utah, Salt Lake City, UT* 84112, *USA*
2 *Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing* 100029, *China*

(Received July 12, 2020; in final form April 5, 2021)

ABSTRACT

The Monte Carlo probability (MCP) model, which has been used for official tropical cyclone (TC) warnings to the public by the United States' National Hurricane Center (NHC), can estimate the probability of wind speed in the vicinity of a TC during the forecast period. It has been successful in the operational environment for many years. However, due to its strong dependence on a given forecast track (e.g., forecast from the NCEP Global Forecast System), the MCP model may generate a poor probability map for TCs near landfall. In this study, we proposed and tested a modified MCP method for TC forecasts near landfall. We first adjusted the MCP model by adding limits to the direction angle and motion distance to deal with the substantial change in TC moving direction and the low wind speeds during landfall. Then, we combined ensemble probability maps generated from ECMWF, United Kingdom MetOffice (UKMO), and NCEP ensemble forecasts, obtained from The International Grand Global Ensemble (TIGGE), into the MCP model to configure a modified MCP model. Wind speed probability maps for the 0–120-h forecast from both the original and modified MCP models are compared. It is found that the modified MCP model can provide a better wind speed probability map during landfall, especially at wind speeds of 20–64 kt near TC landfall. The results from this study prove the benefits of combining the MCP model with ensemble forecasting in potential applications for improved TC forecasts.

Key words: tropical cyclone, Monte Carlo model, ensemble forecasting, wind speed probability

Citation: Li, X., Z. X. Pu, and Z. Q. Gao, 2021: Combining Monte Carlo and ensemble probabilities in tropical cyclone forecasts near landfall. *J. Meteor. Res.*, **35**(4), 607–622, doi: 10.1007/s13351-021-0128-9.

1. Introduction

Accurate forecasts of tropical cyclones (TCs), especially landfalling hurricanes, are crucial for reducing the damage they cause (Smith, 2017). Despite progress in numerical weather prediction (Bauer et al., 2015) and statistical forecast (DeMaria et al., 2009), forecast error mitigation is difficult due to the complexity of TC evolution and the lack of observations (Rappaport et al., 2009; Emanuel and Zhang, 2016). To overcome forecast uncertainty and provide early warning for TCs, a TC probability program was implemented in 1983 (Sheets, 1985) with support from the National Hurricane Center (NHC). With this program, 10-yr samples of official forecast errors are composited relative to the forecast position at the discrete forecast verification times of 12, 24, 48, and 72 h (up to 120 h after 2003). These provide information on the likelihood of a storm of interest moving over a given area. Following this effort, DeMaria et al. (2009) developed a Monte Carlo probability (MCP) model to generate a TC wind probability map at individual locations during a 0–120-h forecast period. In this wind probability model, random samples from the past five years of forecast errors [usually with a sample size of 1000, as suggested by DeMaria et al. (2009)] are required to generate cyclone track, intensity, and corresponding radii realizations [with the simple climatology and persistence model (radii-CLIPER); Knaff et al., 2007] based on an official forecast track from NHC. These generated realizations can then be used to produce a probability map in a single forecast result by counting the number of realizations for which the point is inside the radius of the wind

The first two authors, X. Li and Z. X. Pu, are supported by the US National Science Foundation (ECCS-1839833 and OAC-2004658).

*Corresponding author: Zhaoxia.Pu@utah.edu

speed threshold of interest (e.g., 34, 50, and 64 kt, where 1 kt = 0.5144 m s^{-1}).

To narrow the forecast track error distribution, after 2010, the NHC introduced the Goerss predicted consensus error (GPCE; Goerss, 2007) into the MCP model to generate a more relevant wind probability map. The different GPCE selections at different forecast times lead to a slight improvement in MCP model forecasts in most cases (DeMaria et al., 2013). Meanwhile, because of well-distributed historical errors with a mean value of about zero (DeMaria et al., 2009), the MCP model generates track realizations that are always distributed along the given forecast track. When there are strong spreads between the forecast and best track [i.e., the poststorm best estimates of the cyclone track and intensity, introduced by Jarvinen et al. (1984) and Knapp et al. (2010)], the 1000 generated track realizations can be identified as inaccurate realizations that lead to poor TC realizations and result in poor wind probability forecasts. In order to reduce the forecast errors, more forecast tracks and errors should be considered and added to the MCP model.

Besides the MCP model, major operational numerical weather prediction centers, such as the United States' NCEP, the ECMWF, and the United Kingdom Met Office (UKMO), all provide reliable TC ensemble forecasts. While these ensembles can improve probabilistic forecasts for severe weather events (Froude, 2010, 2011; Frame et al., 2011; Swinbank et al., 2016), wind probability maps (Matsueda and Nakazawa, 2015) for hurricane forecasts can also be generated by combining ensemble forecasts from multiple sources (e.g., from ECMWF, UKMO, and NCEP).

In this study, we develop a modified MCP model by combining probability maps from both MCP and multimodel ensemble forecasts to improve TC probability forecasts near landfall. Based on the MCP model using the NCEP Global Forecast System (GFS) forecast as the official forecast track, we first evaluate and adjust the MCP method by limiting the unrealistic TC moving speeds and directions to be more suitable for TCs near landfall. Then, we combine it with ensemble forecast maps from ECMWF, UKMO, and NCEP to further mitigate the unrealistic TC moving directions to improve hurricane 0–120-h cumulative wind speed probability forecasts. It is our purpose to explore a feasible method of combining MCP and ensemble forecasts for accurate TC wind forecasts near landfall. Hurricanes Sandy (2012), Irma (2017), Harvey (2017), and major hurricanes during the 2018 season over the Atlantic Ocean are used as cases for this study. The MCP model, ensemble forecasts, and modifications of the MCP model are described in Section 2. Probability forecast results and statistical evaluation of cases are provided in Section 3. Section 4 includes an evaluation based on major hurricanes during the 2018 season. Concluding remarks are presented in Section 5.

2. Wind speed probability forecasts and evaluation

2.1 *MCP model*

The MCP model estimates the probability of wind speed at individual locations during a TC forecast period (e.g., 0–120 h) based on randomly sampled errors from previous years (historical) of official track errors by adding these errors to the official deterministic forecast positions. The sample error distributions are calculated by comparing the official forecast positions to the NHC "best track" positions, which are the poststorm best estimates of the cyclone track and intensity from observations. Therefore, with the MCP model, random samples of track and intensity forecast errors from operational forecasts, such as those from the NCEP GFS, in the previous five years are added to a single, deterministic forecast (e.g., official GFS forecast) to generate 1000 realizations of cyclone track, intensity, and corresponding radii. The model produces a probability map by counting the number of realizations for which the points are inside the radius of the wind speed threshold of interest (e.g., 34, 50, and 64 kt). Commonly, the NHC generates 0–120-h official wind probability maps based on its official cyclone forecasts (DeMaria et al., 2009, 2013).

2.1.1 *Track realizations*

TC track errors are defined as the great-circle distance from the forecast track to the best-track positions. According to DeMaria et al. (2009), track errors can be decomposed into along-track (AT; positive when the forecast position is ahead of the best-track position) and cross-track (CT; positive when the forecast position is to the right of the best-track position) errors. Since these AT and CT errors (i.e., ATE and CTE) at a certain time can be affected by forecast errors at earlier times (DeMaria et al., 2009), directly adding these randomly sampled past ATE and CTE to a single forecast can lead to unrealistic tracks. Therefore, a serial correlation is used to eliminate this problem and generate realistic track forecasts. DeMaria et al. (2009) used a linear fit method to determine the influence of an error at a previous time on ATE and CTE:

$$\begin{aligned} \mathrm{ATE}_t &= a_t \mathrm{ATE}_{t-12} + b_t, \\ \mathrm{CTE}_t &= c_t \mathrm{CTE}_{t-12} + d_t, \end{aligned} \tag{1}$$

where a, b, c, and d are fitting coefficients determined by the input historical ATE and CTE, and the subscript t is the forecast time. Then, based on Eq. (1), the historical ATE (CTE) can be divided into two parts: the influence of previous errors (i.e., calculated errors), which is calculated from Eq. (1) with coefficients a, b, c, and d; and the residuals, which are the historical ATE (CTE) minus the calculated values from Eq. (1). For $t = 12$, the coefficients a and c are equal to zero, and the ATE (CTE) calculated from Eq. (1) is equal to coefficient b (d), which is the mean of the historical ATE (CTE) at this time. The residuals are generated from the differences between the historical ATE (CTE) and the coefficient b (d). From $t = 24$ to $t = 120$, the residuals are generated from the differences between the historical ATE (CTE) and calculated errors from Eq. (1). These residuals at each time are the samples of the following MCP realizations.

After the historical ATE (CTE) is transformed into residuals as discussed above, the residuals are randomly sampled at each forecast time to generate the final ATE (CTE) of MCP and make the MCP forecast. At $t = 12$, The residuals are randomly sampled and added to Eq. (1) with coefficients b_{12} and d_{12} as the sample mean AT and CT track error biases at 12 h (as the error is equal to zero at $t = 0$; $a_{12} = c_{12} = 0$). Then these values of ATE (CTE) are input into Eq. (1) to obtain the calculated errors and added to the randomly sampled residuals at 24 h to generate the final ATE (CTE) at 24 h. Following a similar step, the final ATE (CTE) at 24 h is again input into Eq. (1) and combined with the residuals to generate the final ATE (CTE) at the subsequent forecast time. Finally, with the forecast errors from the past five years, a total of 1000 × 10 final ATE (CTE) are generated. These final errors are added to a single forecast track, thus producing 1000 hurricane tracks for the 120-h forecast.

2.1.2 *Intensity realizations*

Similar to track realizations, TC intensity errors in the previous five years are required as random samples for intensity realizations. In the MCP model, intensity is represented by the maximum wind speed at 10-m height. Therefore, the intensity error is the maximum wind speed error (SE). Meanwhile, the SE at any given time can be influenced not only by earlier SE but also by forecast cyclone maximum wind speed (SM) and the distance from the cyclone center to land (D, which is positive when the cyclone is over the ocean) at a given time. Considering the small influence of land when a cyclone is sufficiently far away from land, DeMaria et al. (2009) set $D = 500$ km when $D > 500$ km. Therefore, a linear relationship is assumed between the SE 12 h previous and the forecasts of SM, D, and SE at time t (DeMaria et al., 2009), thus:

$$\mathrm{SE}_t = e_t \mathrm{SE}_{t-12} + f_t \mathrm{SM}_t + g_t D_t + h_t, \quad (2)$$

where e, f, g, and h are least squares fitting coefficients, determined by the historical SE, SM, and D, and the subscript t is the forecast time. The historical SE at 12–120 h is transformed into a residual SE by subtracting the calculated SE (which represents the influence from the error in the previous 12 h) using Eq. (2). Similar to the track realizations, at $t = 12$, the final SE is equal to the sum of the residual and the coefficient h. The final SE at $t = 12$ is then input into Eq. (2) to calculate the SE at $t = 24$. Then, this calculated SE is combined with randomly sampled residual SE to generate the final SE at $t = 24$. Similar steps continue and generate the final SE at subsequent forecast times, and these final SEs are then added to the official (deterministic) forecast intensity to produce 1000 hurricane intensity realizations for the 120-h forecast at 12-h intervals.

Furthermore, when the underlying surface, either ocean or land, is different in the realization than in the forecast, the difference can lead to an incorrect SE, and consequently to poor intensity realizations. Therefore, two methods are used to modify these poor intensity realizations produced by the mismatched underlying surfaces between realizations and forecasts. When the deterministic forecast position is inland and the MCP realizations are over the ocean, SM_t in Eq. (2) will be replaced by the value over the ocean at the nearest forecast time; when the forecast position is over the ocean and the MCP realizations are inland, SM_t will be replaced by the value from an empirical inland wind decay model, as suggested by Kaplan and DeMaria (1995):

$$\mathrm{SM}_t = 27 + (\mathrm{SM}_0 - 27)\mathrm{e}^{-0.115t}, \quad (3)$$

where SM_0 is the MCP forecast intensity at the first time when the TC moves inland.

2.1.3 *Radius realizations*

The radii-CLIPER model suggested by Knaff et al. (2007) is used to generate radius realizations in the MCP model:

$$V(r,\theta) = (V_\mathrm{m} - A)\left(\frac{r_\mathrm{m}}{r}\right)^x + A\cos(\theta - \theta_0) \quad \text{for } r \geqslant r_\mathrm{m}, \quad (4)$$

where V is the wind speed, r is the radial distance from the cyclone center, θ is the azimuth measured counterclockwise relative to 90° to the right of the cyclone's motion direction, V_m is the maximum wind speed, r_m is the radius of the maximum wind, A is an asymmetry factor, and θ_0 is a constant that allows the maximum wind speed to be located at an azimuth other than 90° to the right of the motion direction. The values of A (= 1.06) and θ_0 (=

17) are both suggested by Knaff et al. (2007).

The cyclone size parameter x and r_m can be estimated by a simple climatic model (Knaff et al., 2007):

$$\begin{aligned} x_c &= 0.1147 + 0.0055V_m + 0.001(\varphi - 25^\circ), \\ r_{mc} &= 36.1 - 0.0492V_m + 0.574(\varphi - 25^\circ), \end{aligned} \tag{5}$$

where φ is the latitude and the subscript c denotes a climatic estimate. With historical V, r, V_m, r_m, and θ, the best fit x for each historical hurricane forecast record can be produced with Eq. (4). After that, the best fit x is transformed to residual x by subtracting the climatic-estimated x_c from Eq. (5). Then, similar to the track and intensity realizations, the final x for each forecast time can be produced by adding the randomly sampled residual x to the climatic-estimated x_c in Eq. (5). Meanwhile, since the climatic-estimated x_c is often unequal to the real x at t = 0, when there is no residual x for realizations, this mismatch between climatic-estimated x_c and real x may transmit to subsequent realizations. Therefore, as suggested by DeMaria et al. (2009), exponential decay with an e-folding time of 32 h is used in radius realizations for each forecast time to cover the mismatch of climatic-estimated x_c and simulated x at the forecast start time. Finally, after these transformations and modifications, the final x for MCP is produced with climatic-estimated x, randomly sampled residual x, and the start-time x error. The azimuth-dependent wind profile for each realization is produced through Eq. (4) with this final x, hurricane maximum wind speed V_m, maximum wind radius r_m, and wind speed thresholds of interest (e.g., 34, 50, and 64 kt).

2.2 *Adjustments to the MCP model*

In this study, we use the NCEP GFS deterministic forecasts as official forecasts in the MCP model to generate the 0–120-h probability forecasts near and during landfalls of three major hurricanes: Hurricane Sandy, as an extratropical cyclone during its landfall from 0000 UTC 27 to 0000 UTC 29 October 2012; Hurricane Harvey from 0000 UTC 25 to 0000 UTC 29 August 2017; and Hurricane Irma from 0000 UTC 8 to 1200 UTC 10 September 2017. The GFS forecasts from the previous five years for each case, provided by the Automated Tropical Cyclone Forecast (ATCF) database, are used to provide the historical forecast errors for the MCP model.

For track realizations, Table 1 shows the coefficients a, b, c, and d in Eq. (1) obtained from the previous five years, namely, 2012–2016 TC data for Hurricanes Harvey and Irma (2017). A determination coefficient (R^2) is also defined to measure the correlations between ATE_t (CTE_t) and ATE_{t+12} (CTE_{t+12}). We found that there are considerable small determination coefficients (R^2) at some lead times, for example, $R^2 < 0.5$ at 24, 84, and 108 h in our case. This indicates that there might be some track data that fail to satisfy Eq. (1). Namely, in some cases, early ATE (CTE) may not affect the following ATE (CTE). For instance, when a TC undergoes a large change in moving direction during 24 h, the rapid change in its track means that there is no significant relationship between ATE (CTE) in the first 12 h and those in the subsequent 12 h. Considering this problem, we divide the TC tracks into two groups based on direction changes ($\mathrm{d}\gamma$) between a certain time t and t + 12, namely, the cases of $\mathrm{d}\gamma \leqslant 45^\circ$ or $\mathrm{d}\gamma > 45^\circ$.

Table 2 shows the parameters obtained from the TC data during 2012–2016 for cases of $\mathrm{d}\gamma \leqslant 45^\circ$ and $\mathrm{d}\gamma > 45^\circ$. The determination coefficient for $\mathrm{d}\gamma \leqslant 45^\circ$ is larger than that in Table 1, while it is close to zero for $\mathrm{d}\gamma > 45^\circ$. The results here confirm that Eq. (1) cannot be well applied in the track realizations when the subsequent forecast track has large directional changes. For instance, there are 21 times with $\mathrm{d}\gamma > 45^\circ$ from 5 × 120 h GFS track forecasts, and the track realizations using Eq. (1) based on these GFS tracks will lead to poor wind speed probability forecasts. Therefore, we made an adjustment to the MCP model to ensure that the final ATE and CTE are randomly sampled directly from the past five years of GFS ATE and CTE without applying Eq. (1) when $\mathrm{d}\gamma > 45^\circ$.

Additionally, the moving speed of a TC at a certain time is usually related to its speed at an earlier time, and the forward speed will not increase quickly during a short period. Thus, some randomly sampled ATE and CTE could lead to track realizations with an unrealistic forward speed. To mitigate these unrealistic track realizations, we introduce a coefficient r_{dis}, defined as the motion distance ratio, namely:

$$r_{dis} = \mathrm{dis}_{t+12}/\mathrm{dis}_t, \tag{6}$$

where dis_t and dis_{t+12} are the distance traveled between forecast intervals $t - 12$ and t, and between t and $t + 12$, respectively; r_{dis} denotes the ratio of TC motion distance

Table 1. Coefficients in Eq. (1) based on the 2012–2016 TC database

Time (h)	Along track			Cross track		
	a	b	R^2	c	d	R^2
12	0	5.52	0.00	0	−2.37	0.00
24	1.05	−1.68	0.40	0.95	−0.81	0.38
36	1.11	−2.69	0.56	0.96	−1.85	0.49
48	1.01	−2.17	0.54	0.88	−1.50	0.46
60	1.00	3.49	0.52	0.84	−1.50	0.47
72	1.14	−4.84	0.83	1.14	−3.46	0.78
84	1.15	4.26	0.69	0.76	13.94	0.39
96	1.17	−2.78	0.89	1.21	−1.98	0.86
108	0.96	10.49	0.55	0.59	4.04	0.26
120	1.16	3.94	0.92	1.11	−1.24	0.89

Table 2. As in Table 1, but for different direction change angles. The "–" denotes the null value due to nonexistent cases

Time (h)	Along track [d$\gamma \leq 45°$ (d$\gamma > 45°$)]			Cross track [d$\gamma \leq 45°$ (d$\gamma > 45°$)]		
	a	b	R^2	c	d	R^2
12	0.00 (0.00)	5.52 (5.52)	0.00 (0.00)	0.00 (0.00)	−2.37 (−2.37)	0.00 (0.00)
24	1.15 (0.36)	−1.01 (−3.88)	0.48 (0.04)	1.08 (0.29)	−1.10 (0.03)	0.47 (0.05)
36	1.17 (0.24)	−2.57 (1.93)	0.67 (0.01)	1.12 (0.23)	−1.42 (−5.61)	0.59 (0.08)
48	1.21 (0.06)	−4.46 (19.36)	0.75 (0.00)	1.05 (−0.04)	−2.83 (13.40)	0.68 (0.00)
60	1.15 (–)	0.07 (–)	0.72 (–)	1.03 (–)	1.88 (–)	0.65 (–)
72	1.14 (0.10)	−4.84 (27.91)	0.83 (0.00)	1.14 (0.31)	−3.46 (−14.32)	0.78 (0.09)
84	1.23 (–)	3.04 (–)	0.82 (–)	1.04 (–)	9.90 (–)	0.70 (–)
96	1.17 (0.56)	−2.78 (5.38)	0.89 (0.13)	1.21 (0.03)	−1.98 (15.76)	0.86 (0.00)
108	1.19 (–)	7.96 (–)	0.83 (–)	0.98 (–)	0.89 (–)	0.69 (–)
120	1.16 (0.18)	3.94 (45.07)	0.92 (0.02)	1.11 (−0.09)	−1.24 (−0.52)	0.89 (0.01)

dis_{t+12} and dis_t. Based on data from the previous five years, we found that r_{dis} has a maximum and changes with dis_t. The original MCP model generates some unrealistic track realizations with r_{dis} over the maximum in historical data. Therefore, in the adjusted MCP model, a maximum r_{dis} limit is introduced into track realizations to eliminate these unrealistic tracks. For each forecast time, a series of maximum r_{dis} in the past five years is determined at different dis_t. The archived best track can be accurate only at 0.1° for latitude and longitude, namely, 0.1° of the great circle for the cyclone around the equator. Therefore, the dis_t interval for which the maximum r_{dis} is determined is set as 0.1° of the great circle on the earth (11.13 km). Then for each MCP forecast, we calculate the forecast r_{dis} and compare it with the corresponding threshold of the maximum r_{dis} to eliminate excessive points. For example, for dis_{12} of 1.5°, the realizations with dis_{24} of 6.0° are eliminated by the maximum r_{dis} of 2.4.

With the mentioned adjustment, the final ATE (CTE) is determined through the MCP model and added to the GFS deterministic track forecasts to generate the 0–120-h track realizations for each hurricane with time intervals of 12 h. For the intensity realizations, there is no adjustment to the MCP model. The final SE is determined by the MCP model with GFS forecast data from the previous five years. It is then combined with the GFS deterministic track, intensity forecasts, and track realizations to produce the intensity realizations.

For radius realizations, there are negative x residuals in the GFS data from the past five years that can lead to a small final x and large unrealistic radii. To eliminate these excessive radii, in the modified model, radius realizations are limited by maximum radii of 34, 50, and 64 kt at different forecast times based on cyclone samples from the past five years. Furthermore, due to the underlying surface friction of land, the wind speeds of major hurricanes such as Sandy, Irma, and Harvey decrease very quickly during landfall. Data for standard radii of 34, 50, and 64 kt are not available, resulting in a lack of data for evaluation. To better assess the MCP forecast, we also forecast the impact radii for lower wind speeds, such as 20 and 27 kt. Similar to 34–64 kt, the radii for 20 and 27 kt are limited by the maximum radii that are determined by the regression with radii data for 34, 50, and 64 kt. With this correction, the final x is applied to the GFS track and intensity realizations to generate radius realizations for each hurricane. For convenience, the adjusted MCP model is referred to hereafter as the AMCP model.

2.3 *Combined AMCP and ensemble probability: A modified AMCP*

Considering that hurricanes can move several hundred kilometers within 12 h, directly assessing wind speed probability with 10 AMCP forecasts (12–120 h with 12-h intervals) could miss the influence of cyclone motion during the 12-h time period. For instance, the radii for 64-kt wind usually measure less than 40 km from the cyclone center. If we directly use the 12-h interval realizations, such as 12–24 h, to calculate the probabilities of 64-kt wind for any point outside the radius of 40 km, the probability will be zero and these points will be omitted in the calculation. Therefore, to reduce this omission, DeMaria et al. (2009) linearly interpolated the realizations into 2-h intervals, during which the cyclones, especially after landfall, usually do not move far. Following DeMaria et al. (2009), the realizations here including track, intensity, and radii are linearly interpolated at 2-h intervals within the realizations at 12-h intervals. For instance, the track, intensity, and radii at 12 and 24 h are used to linearly interpolate the track, intensity, and radii at 14, 16, 18, 20, and 22 h. Then, AMCP at any given location and for any forecast interval to 120 h can be estimated by counting the number of realizations for which the point is inside the radius of the wind speed threshold of interest, relative to the total number of realizations. For example, the wind probability for 34 kt at a given

location from 0 to 120 h is the proportion of the number of realizations in the area covered by the radius of 34 kt during 0–120 h, to the total number of realizations.

Because MCP and AMCP are based on official forecasts (e.g., GFS) with random samples, a poor track forecast may introduce large errors to wind speed realizations. Therefore, additional wind speed probabilistic forecasts are needed to supply MCP or AMCP forecasts and generate a combined wind speed probability forecast. Given the availability of ensemble forecasting at large operational centers and its usefulness in improving operational TC forecasts, it seems natural to combine AMCP with ensemble members for improved forecasts. In this study, we attempt to evaluate whether including ensemble probability will improve MCP forecasts.

Major operational centers make probability forecasts through their ensemble forecast systems with multiple forecasts from numerical weather prediction models with either different initial conditions or varying numerical representations of the model physics (Toth and Kalnay, 1997; Palmer et al., 2009; Barkmeijer et al., 2013; Buizza, 2014). In this study, we use wind speed probability maps from ensemble forecasts, obtained from The International Grand Global Ensemble (TIGGE; https://confluence.ecmwf.int/display/TIGGE), for three major operational centers: ECMWF (with 50 ensemble members), UKMO (with 23 members for Hurricane Sandy and 17 members for Hurricanes Irma and Harvey), and NCEP (with 20 ensemble members). While these ensemble forecasts of TCs provide the uncertainties in the forecasts (Swinbank et al., 2016), we can also make probability maps by counting the occurrence percentage of a certain forecast variable among all ensemble forecasts. The ensemble wind speed probabilities are calculated similar to the method used in the AMCP model, namely, by counting the total number of ensemble members for each of the forecasts that exceed the prescribed threshold (20–64 kt) and dividing by the total number of ensemble forecasts.

By combining the probability maps from both AMCP and the ensemble forecasts as in the following, a modified AMCP ($\mathrm{AMCP_{mf}}$) is produced:

$$\mathrm{AMCP_{mf}} = \begin{cases} 0.5\mathrm{AMCP} + 0.5/3(P_{\mathrm{NCEP}} + P_{\mathrm{ECMWF}} + P_{\mathrm{UKMO}}) & \text{over sea} \\ 0.1\mathrm{AMCP} + 0.9/3(P_{\mathrm{NCEP}} + P_{\mathrm{ECMWF}} + P_{\mathrm{UKMO}}) & \text{inland} \end{cases}, \quad (7)$$

where P_{ECMWF}, P_{UKMO}, and P_{NCEP} are wind speed probabilities of the ECMWF, UKMO, and NCEP ensemble forecasts, respectively. With TIGGE, the ECMWF, UKMO, and NCEP ensemble forecast data are all available at a grid size of 0.5° × 0.5°. Therefore, the $\mathrm{AMCP_{mf}}$ probability result is also calculated at a grid size of 0.5° × 0.5° for easy processing in Eq. (7).

As shown in Section 3, AMCP overestimates wind speed probability over land (Splitt et al., 2010). Based on tests of various weights for the ensemble and AMCP combinations shown in Table 3, we use a larger weight coefficient (0.9, which leads to a multiplicative bias that is closer to 1) for the ensemble forecast probability results in Eq. (7). Over the ocean, we use weight coefficients of 0.5 for both AMCP and ensemble forecasts to compensate the confidence for each. For consistence, the NCEP GFS final analysis (FNL) data, which are publicly available at a grid size of 0.5° × 0.5°, are used to obtain the analysis (or "truth") of wind speed frequency by counting the frequency of wind speeds greater than a certain threshold of interest at a given location for the forecast time period. For instance, the frequency of wind speed of 34 kt for a given location is the number of impacted times with wind speed over 34 kt divided by the total number of analysis times of FNL. Namely, for a certain time that has wind over 34 kt over a given area, the frequency should be 1/60 (120-h analysis interpolated into 2-h intervals). The cyclone center is determined by the minimum surface pressure in both ensemble products and NCEP FNL. The ensemble mean track is generated by averaging the cyclone centers from all

Table 3. The multiplicative bias in Eq. (8) of the 120-h and 34-kt inland wind probability forecast with different inland weight coefficients of the ensemble products in Eq. (7)

Lead time (h)	Inland weight coefficient of ensemble products								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
12	1.00	1.15	1.31	1.47	1.63	1.79	1.95	2.11	2.27
24	3.81	5.43	7.05	8.67	10.29	11.92	13.54	15.16	16.78
36	4.15	5.69	7.22	8.76	10.30	11.84	13.37	14.91	16.45
48	4.19	5.67	7.16	8.64	10.13	11.61	13.10	14.58	16.07
60	5.03	6.61	8.20	9.79	11.38	12.96	14.55	16.14	17.72
72	3.87	4.98	6.09	7.19	8.30	9.40	10.51	11.62	12.72
84	3.93	5.21	6.48	7.76	9.04	10.31	11.59	12.87	14.14
96	1.99	2.88	3.77	4.66	5.55	6.45	7.34	8.23	9.12
108	2.17	3.23	4.28	5.34	6.39	7.45	8.50	9.56	10.61
120	2.62	4.61	6.59	8.57	10.55	12.54	14.52	16.50	18.48

members. The $AMCP_{mf}$ track is the average of the AMCP track and the mean ensemble track. Similar to the $AMCP_{mf}$ model, the track, intensity, and radii from ensemble forecasts and FNL analysis are also linearly interpolated at 2-h intervals from their 6-h data.

2.4 *Verification method*

Because the change in forecast wind speed probability from MCP to AMCP is small (commonly less than 2%), we emphasize the comparison between AMCP and $AMCP_{mf}$ in the evaluation. Statistical methods are applied. First, a multiplicative bias is assessed:

$$\text{Bias} = \frac{\frac{1}{N}\sum_{i=1}^{N} F_i}{\frac{1}{N}\sum_{i=1}^{N} O_i}, \tag{8}$$

where F_i is the forecasted probability, O_i is the frequency from the FNL analysis, and N is the total number of forecasted/observed points over the entire domain.

To assess the accuracy of the above binary deterministic forecast, namely, whether the specific wind speed (e.g., 20 or 34 kt) of a hurricane does or does not affect a certain area based on the AMCP/$AMCP_{mf}$ probability forecasts, the threat score (TS) is calculated:

$$\text{TS} = \frac{H}{H + M + \text{FA}}, \tag{9}$$

where H is the number of hits (forecasted and observed), M is the number of misses (not forecasted but observed), and FA is the number of false alarms (forecasted but not observed). To determine the TS, a threat probability (TP) is defined based on the probability map of whether an event is forecasted with accuracy varying from 1% to 100%. The forecasted event is defined as a location with the wind speed probability over the given TP. Otherwise, it is not a forecasted event. With this definition of TP and observations, the hits, misses, and false alarms are generated. From the perspective of the TS, a larger maximum implies a better deterministic forecast. Thus, TS and TP are used to assess the improvement in $AMCP_{mf}$. To assess the coverage of the wind speed warning map, the missing score (MS) and false score (FS) are calculated:

$$\text{MS} = \frac{M}{H + M + \text{FA}}, \tag{10}$$

$$\text{FS} = \frac{\text{FA}}{H + M + \text{FA}}. \tag{11}$$

The increased MS (FS) indicates an increasing number of missed alarms (false alarms). Since missed alarms and false alarms are also critical for public disaster warnings, the MS and FS are used to assess the improvement in $AMCP_{mf}$ from ensemble forecasts, which likely lead to large MS because of the limited ensemble members. Furthermore, a bootstrap method (Efron and Tibshirani, 1994) is used to evaluate the significance of the modification from AMCP to $AMCP_{mf}$.

3. Hurricane case studies

3.1 *Probability forecasts*

In this section, probability products for Hurricanes Irma, Harvey, and Sandy are discussed to characterize the performance of $AMCP_{mf}$ compared with AMCP. Figures 1a–d illustrate the 34-kt cumulative wind speed probability forecast generated by AMCP, mean ensemble forecasts (global average forecasts from ECMWF, UKMO, and NCEP probability forecasts), $AMCP_{mf}$, and FNL analysis, respectively, for Hurricane Irma for 120-h forecasts at 0000 UTC 9 September 2017. Figures 1e–g show the corresponding 120-h and 34-kt probability differences between the forecasts and FNL analysis cumulative wind speed probability map for Hurricane Irma at 0000 UTC 9 September 2017.

From the probability map, there is a significant decrease in wind speed probability in $AMCP_{mf}$ inland that is closer to FNL, compared with AMCP. The results imply that the combination with the ensemble forecasts will improve wind speed probability forecasts, especially for TCs inland. To qualify this improvement, Figs. 1e–g also give the probability differences between the forecasts and FNL. Since there are areas with zero probabilities for the forecast maps from each probability forecast model, directly subtracting the FNL analysis from these probability maps will generate a lot of invalid points for the assessment. Therefore, these points (or locations) with zero probabilities in all maps, including AMCP, the ensemble, and $AMCP_{mf}$, are ignored during the comparison in this section. After eliminating these zero points, the mean (standard deviation) probability difference is 1.34% (3.32%) for AMCP, 0.72% (2.26%) for the ensemble, and 0.66% (1.94%) for $AMCP_{mf}$. From the perspective of the mean and standard deviation probability difference, a mean and standard deviation probability difference close to zero implies a small forecast bias. It is obvious that after modification, $AMCP_{mf}$ provides a better wind probability forecast with minimum bias. For forecasts over the ocean, the mean (standard deviation) probability difference is 0.37% (1.96%) for AMCP, 0.66% (2.07%) for the ensemble, and 0.51% (1.72%) for $AMCP_{mf}$. For inland forecasts, the mean (standard deviation) probability difference is 0.98% (2.82%) for AMCP, 0.06% (0.95%) for the ensemble, and 0.15% (0.98%) for $AMCP_{mf}$. With a lower mean probability difference, we found that AM-

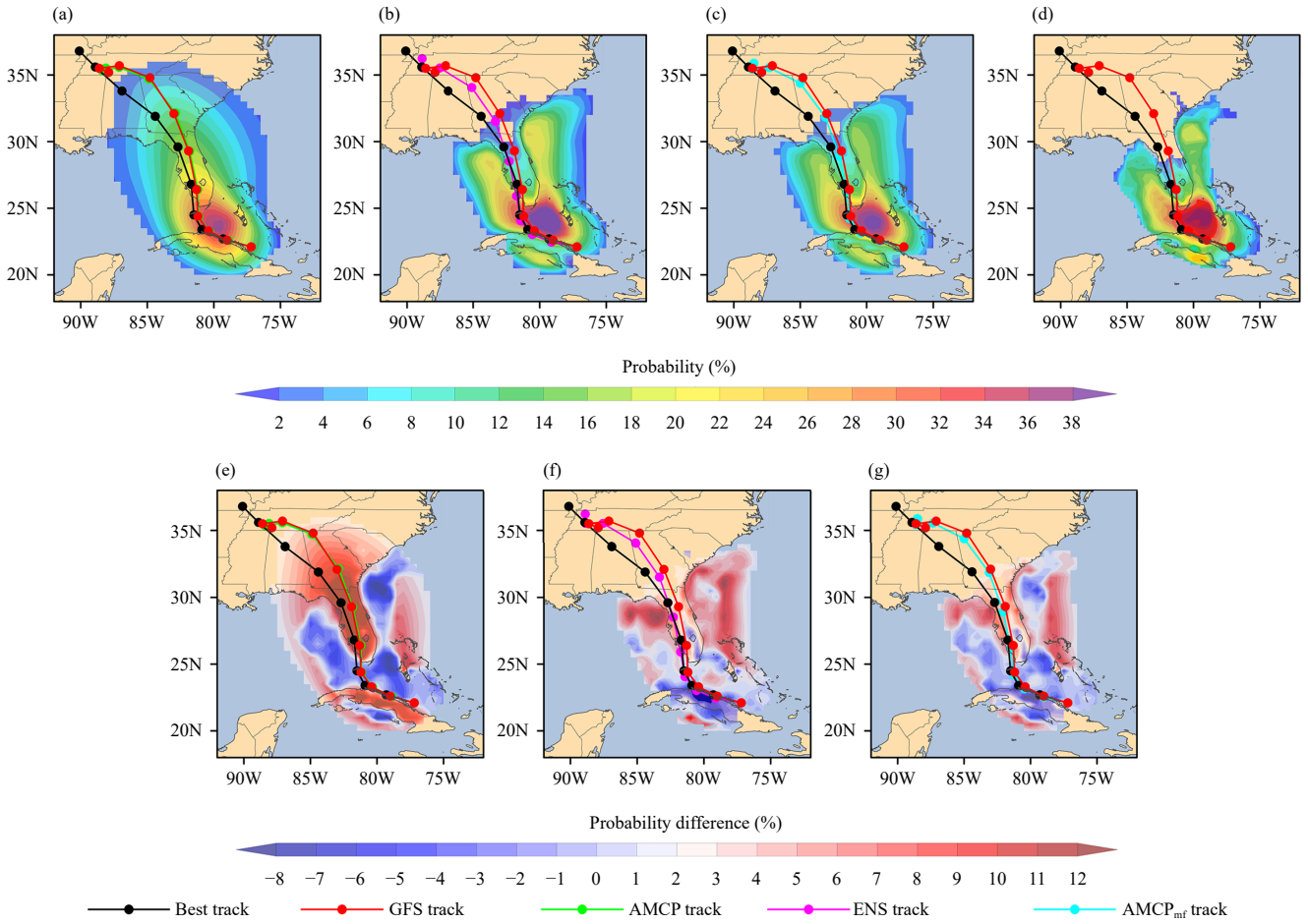

Fig. 1. The 34-kt cumulative wind speed probability for Hurricane Irma from (a) AMCP, (b) the ensemble mean (ENS), (c) $AMCP_{mf}$, and (d) FNL analysis for 120 h at 0000 UTC 9 September 2017. (e–g) The 34-kt cumulative wind speed probability difference for Hurricane Irma between AMCP, the mean ensemble forecast, and $AMCP_{mf}$, respectively, and FNL analysis for the 120-h forecast at 0000 UTC 9 September 2017. The red, green, magenta, cyan, and black lines (points) represent the hurricane track (positions every 12 h; there are only 10 position points for the best track because of the dissipation of Irma) from GFS, AMCP, the mean ensemble model, $AMCP_{mf}$, and NHC best track, respectively.

CP provides a better probability forecast over the ocean compared to the ensemble. For the probabilities inland, AMCP overestimates the probabilities despite the wind speed decay model contained in the AMCP model (Kaplan and DeMaria, 1995), and the ensemble provides a better forecast inland. During this forecast period, a single model, either AMCP or the ensemble, is not good enough. In contrast, in combination they provide better probability forecasts compared to a single model.

Figures 2a–c illustrate the 27-kt cumulative wind speed probability forecast generated by AMCP, mean ensemble forecasts (global average forecasts from ECMWF, UKMO, and NCEP probability forecasts), and $AMCP_{mf}$, respectively, for Hurricane Harvey for 120-h forecasts at 0000 UTC 27 August 2017. Figure 2d shows the corresponding 120-h and 27-kt FNL analysis cumulative wind speed probability map for Hurricane Harvey at 0000 UTC 27 August 2017. Similarly, the false forecast of AMCP inland is significantly reduced after combined with the ensemble forecast. To qualify this improvement, Figs. 2e–g show the probability difference between the forecast and FNL. Similar to the analysis with Hurricane Irma, after ignoring the zero probability points as discussed above, the mean (standard deviation) probability difference is 5.73% (11.67%) for AMCP, 0.17% (1.27%) for the ensemble, and 1.18% (3.24%) for $AMCP_{mf}$. While the GFS track moves northwest inland and the best track moves northeast, AMCP produces a poor track realization and a poor wind speed probability forecast inland with a large bias. After modification, $AMCP_{mf}$ provides a better wind probability forecast with a lower bias compared to AMCP. This indicates that combining AMCP with the ensemble forecasts improves the forecast ability for Hurricane Harvey. For forecasts over the ocean, the mean (standard deviation) probability difference is 1.30% (5.26%) for AMCP, 0.15% (1.24%) for the en-

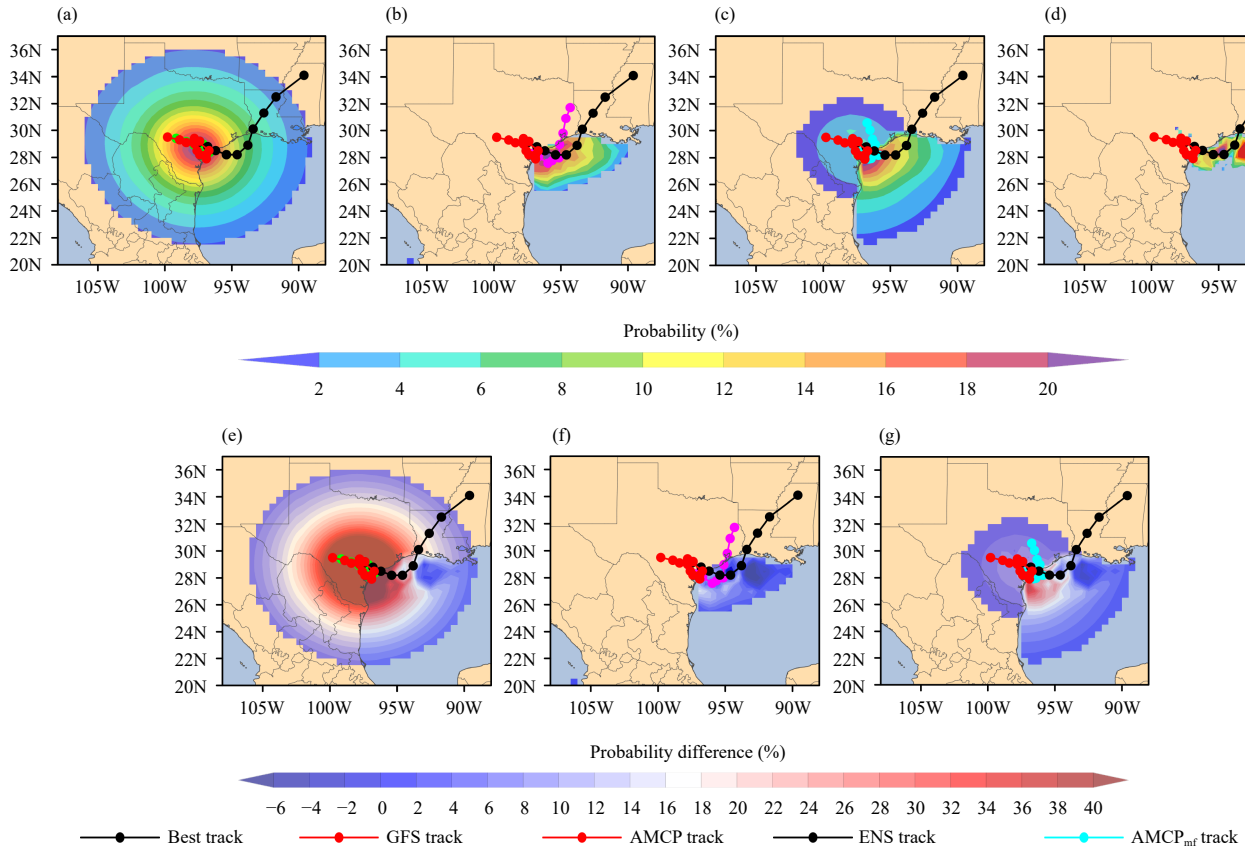

Fig. 2. As in Fig. 1, but for 27-kt cumulative wind speed probability and probability difference for Hurricane Harvey during 120 h at 0000 UTC 27 August 2017.

semble, and 0.72% (3.12%) for AMCP$_{mf}$. For inland forecasts, the mean (standard deviation) probability difference is 4.43% (10.96%) for AMCP, 0.01% (0.27%) for the ensemble, and 0.46% (1.18%) for AMCP$_{mf}$. We found that the ensemble can improve not only inland forecasts but also forecasts over the ocean. Considering that there is a larger error for the GFS forecast track, the ensemble model can modify the AMCP forecast based on that poor track from the GFS.

Figures 3a–c show the 34-kt cumulative wind speed probability maps for Hurricane Sandy from AMCP, mean ensemble forecasts, and AMCP$_{mf}$, respectively, for the 108-h forecast at 0000 UTC 28 October 2012. Figure 3d shows the 108-h and 34-kt FNL analysis cumulative wind speed probability map for Hurricane Sandy at 0000 UTC 28 October 2012. We found that the combination not only eliminates the false inland forecast of AMCP but also improves the forecast over the ocean, as AMCP$_{mf}$ provides better location forecasts of maximum probability. Figures 3e–g show the differences between the forecasts and FNL. After ignoring the zero points, the mean (standard deviation) probability difference is −3.29% (8.54%) for AMCP, 1.75% (4.36%) for the ensemble, and −0.97% (4.77%) for AMCP$_{mf}$. For forecasts over the ocean, the mean (standard deviation) probability difference is −3.96% (7.90%) for AMCP, 1.59% (4.36%) for the ensemble, and −1.19% (4.65%) for AMCP$_{mf}$. For inland forecasts, the mean (standard deviation) probability difference is 0.67% (2.32%) for AMCP, 0.17% (0.78%) for the ensemble, and 0.22% (0.80%) for AMCP$_{mf}$. Compared to AMCP, similar to Hurricanes Irma and Harvey, the AMCP$_{mf}$ probability forecast is also improved by the ensemble forecast with lower bias. Meanwhile, similar to Fig. 1, AMCP$_{mf}$ provides the best wind probability forecast with minimum bias compared to AMCP and the ensemble. This again confirms that this combination (AMCP$_{mf}$) performs better than either single model, namely, AMCP or the ensemble, in some cases.

Based on the probability maps above, we can conclude that the AMCP model may produce poor forecasts with a poor essential cyclone track. The ensemble forecast results are useful in modifying the AMCP results, not only for track direction but also for inland probability distribution. To further explore the characteristics of the AMCP, AMCP$_{mf}$, and ensemble forecast wind speed probability, Fig. 4 shows three scatter diagrams that in-

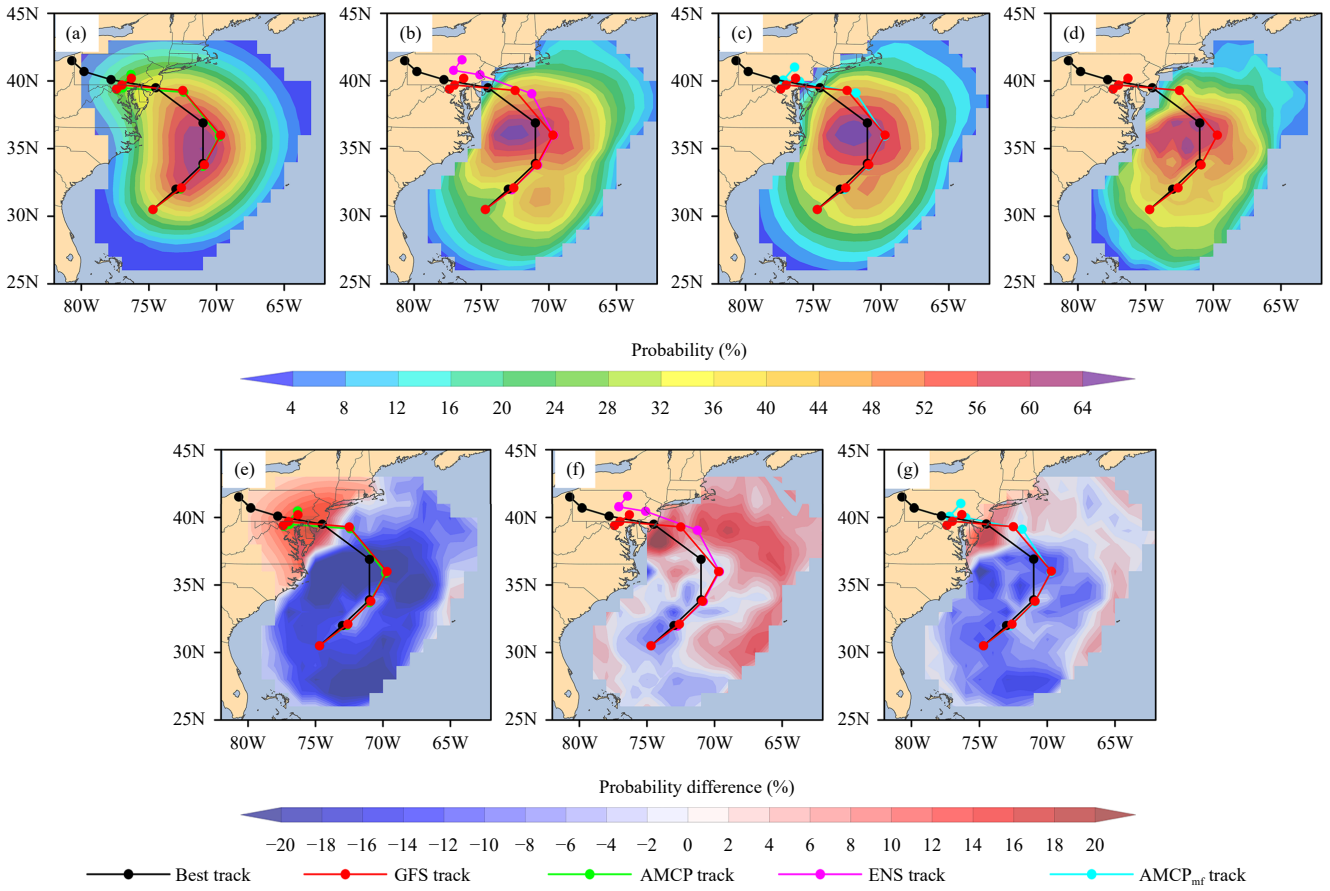

Fig. 3. As in Fig. 1, but for 34-kt cumulative wind speed probability and probability difference for Hurricane Sandy during 108 h at 0000 UTC 28 October 2012.

dicate the relationship of AMCP, ensemble forecasts, and AMCP$_{mf}$ to the FNL analysis for the 34-kt and 120-h (108 h for Sandy) cumulative probability forecasts for Hurricanes Irma, Harvey, and Sandy. Considering the small intercept (< 1%), we focus only on the slope of linear fitting in this section. From the perspective of the slope and the determination coefficient (R^2) in fitting, a slope close to 1 implies a better forecast and a larger R^2 implies a small dispersion for the wind speed probability forecast. For Hurricane Irma, the slope (R^2) for AMCP, the ensemble, and AMCP$_{mf}$ is 0.83 (0.75), 1.66 (0.89), and 1.29 (0.89), respectively. Compared to the FNL analysis, AMCP overestimates while the ensemble underestimates the wind speed probability. After modification, AMCP$_{mf}$ produces a more concentrated probability distribution with increased R^2. For Hurricane Harvey, the slope (R^2) for AMCP, the ensemble, and AMCP$_{mf}$ is 0.14 (0.22), 0.74 (0.63), and 0.5 (0.53), respectively. Compared to the FNL analysis, both AMCP and the ensemble underestimate the wind speed probability. Similar to Hurricane Irma, after modification, AMCP$_{mf}$ produces a more concentrated probability distribution with increased R^2 and a lower bias with slope closer to 1. For Hurricane Sandy, the slope (R^2) for AMCP, the ensemble, and AMCP$_{mf}$ is 1.47 (0.70), 0.96 (0.94), and 1.27 (0.93), respectively. Again, combined with the ensemble, AMCP$_{mf}$ provides a more concentrated probability forecast and is more likely to distribute along the "$x = y$" line. Through the above analysis, it can be concluded that combining AMCP with the ensemble can decrease the dispersion of the AMCP probability forecast. In some cases, AMCP$_{mf}$ is better than the ensemble forecast, proving that the combination of AMCP and ensemble forecasts can provide a better forecast than any of these single models. To further clarify the reliability of AMCP$_{mf}$, statistical evaluation is provided in the next section.

3.2 *Statistical evaluation*

In this section, AMCP and AMCP$_{mf}$ are quantitatively compared to assess the improvement by introducing the ensemble model. First, the multiplicative bias based on cumulative wind speed probability is examined as a function of lead time, as shown in Fig. 5. According to Eq. (8), multiplicative bias > 1 (< 1) implies an overestima-

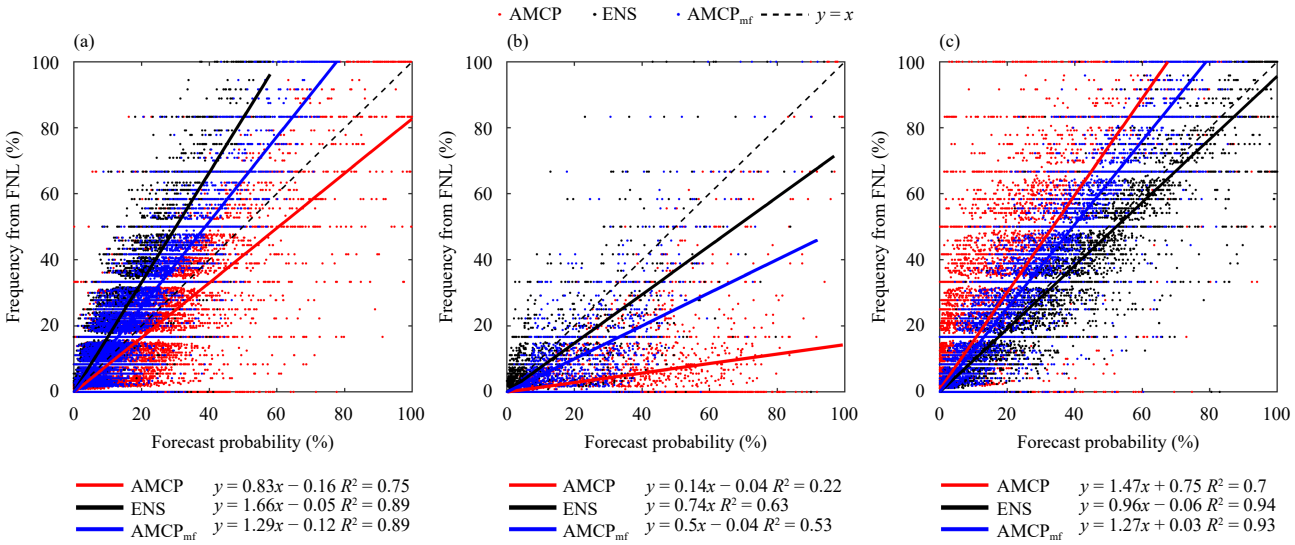

Fig. 4. Scatter diagram of AMCP (red points and lines), ensemble (black points and lines), and AMCP$_{mf}$ (blue points and lines) forecasts of 34-kt cumulative wind speed probability contrasted with the frequency from the FNL analysis for Hurricanes (a) Irma, (b) Harvey, and (c) Sandy. The fitting functions are also added.

tion (underestimation) of the averaged forecast probability. Figure 5a shows the averaged multiplicative biases of AMCP and AMCP$_{mf}$ for Hurricane Irma at different forecast cycles (0000 UTC 8 September 2017 for the 0–120-h forecast; 0000 UTC 9 September 2017 for the 0–108-h forecast; and 1200 UTC 10 September 2017 for the 0–72-h forecast; the forecast times are limited by hurricane dissipation time from the GFS forecast). According to Fig. 5a, the differences between AMCP and the analysis are significantly reduced by adding the ensemble forecasts. After modification, for the 36–72-h forecast, the multiplicative bias for each wind speed threshold is closer to 1, meaning that AMCP$_{mf}$ produces a wind speed probability closer to the analysis. During this forecast period, the hurricane is in its landfall, and the wind speed in the inland part of the hurricane does not decrease. This causes AMCP to overestimate probabilities for inland areas. At the same time, the ensemble forecast gives a more accurate wind speed and probability forecast. Therefore, the combination with the ensemble can provide better probability forecasts with reduced multiplicative bias. The significant decrease in multiplicative bias for AMCP at the 84–120-h forecast might be caused by wind speed decay through Eq. (3) after hurricane landfall (based on the GFS track, Hurricane Irma's landfall occurred at 0000 UTC 11 September 2017). The decreased wind speed leads to smaller probability and reduces the difference between AMCP and the analysis, especially for a wind speed of 64 kt, in which the multiplicative bias decreases from 2.57 (for the 72-h forecast) to 1.36 (for the 84-h forecast).

For Hurricane Harvey, the averaged multiplicative biases of AMCP and AMCP$_{mf}$ for different forecasts (0000 UTC 25–28 August 2017 for the 0–120-h forecast, and 29 August 2017 for the 0–108-h forecast; the forecast times are limited by the hurricane dissipation time from the GFS forecast) are shown in Fig. 5b. Because of the lower wind speeds of Hurricane Harvey at landfall (see Fig. 2d), only the multiplicative biases for AMCP and AMCP$_{mf}$ at 20–34 kt are displayed here. Errors in both the GFS forecast track and inland AMCP intensity realizations produce higher probabilities inland compared with the FNL analysis. These higher probabilities from the AMCP forecast lead to a very large multiplicative bias, with a maximum of about 9.66 for 20-kt wind, 26.21 for 27-kt wind, and 103.89 for 34-kt wind. After modification, all multiplicative bias decreases, and the largest multiplicative bias at 20, 27, and 34 kt is reduced to about 2.76, 6.25, and 23.00, respectively.

For Hurricane Sandy, the averaged multiplicative biases of AMCP and AMCP$_{mf}$ for different forecast cycles (0000 UTC 27 October for the 0–108-h forecast; 28 October for the 0–84-h forecast; and 29 October 2012 for the 0–60-h forecast; the forecast times are limited by the hurricane dissipation time from the GFS forecast) are shown in Fig. 5c. Considering that there are many lower probability forecasts in AMCP (see Fig. 3a), the multiplicative bias is less than 1 for wind speeds of 20, 27, and 34 kt. The probability at 50 and 64 kt has a multiplicative bias over 1 because of the lower frequency of the FNL high wind speed. That is, AMCP overestimates high wind speeds and leads to large multiplicative bias. After

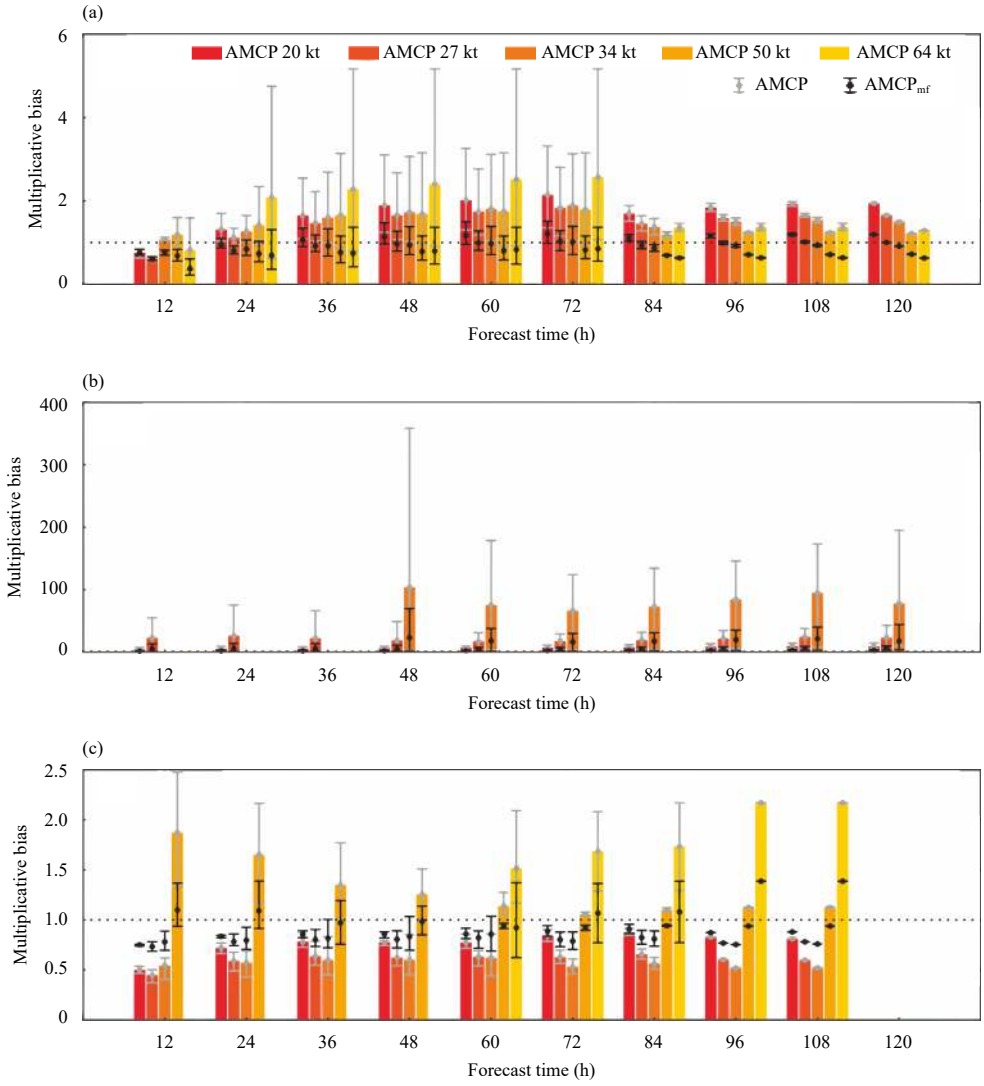

Fig. 5. Averaged multiplicative biases of AMCP (bars) and AMCP$_{mf}$ (black points) at different wind speeds for Hurricanes (a) Irma, (b) Harvey, and (c) Sandy based on cumulative probability. The error bars for AMCP and AMCP$_{mf}$ are shown as gray and black whiskers, respectively.

modification including the ensemble forecasts, the multiplicative bias increases at 20, 27, and 34 kt, while it decreases at 50 and 64 kt, leading to a better probability forecast for multiplicative bias close to 1.

Based on the cumulative forecast data depicted in Fig. 5, Fig. 6 shows the maximum TS and corresponding TP variation before and after the modification for the combined datasets of Hurricanes Irma, Harvey, and Sandy. Because the maximum TSs for Hurricanes Irma, Harvey, and Sandy are all higher in the AMCP$_{mf}$ than in the AMCP forecast, the maximum TS shows an increase of about 0.18–0.33 at 20–34 kt, as shown in Fig. 6a. This indicates an improvement in binary deterministic forecasts for the modified method. Meanwhile, the corresponding TP is lower except at 27 kt for the 12-h forecast. The larger TS and smaller TP indicate a better binary (TC affects or does not affect) forecast map for disaster warnings.

To determine the significance of these changes in TS, a bootstrap *t*-test is used here. The bootstrap *t*-test is a useful method to assess the difference between two sets of small-sample data, as it randomly samples from the small sample to verify whether or not a change in the small-sample data is significant (Efron and Tibshirani, 1994). At a specific lead time from different forecast cycles, the threat score is first transformed according to Efron and Tibshirani (1994) and then randomly sampled to generate the samples for the bootstrap *t*-test. The probability of *T* (test statistic from new samples) larger than *t* (test statistic from original data) is compared to the significance level to determine the improvement. Through the bootstrap *t*-test, we found that the maximum TS improvement for forecast wind speed of 20–27 kt is significant at the 0.05 significance level at each forecast time. Some

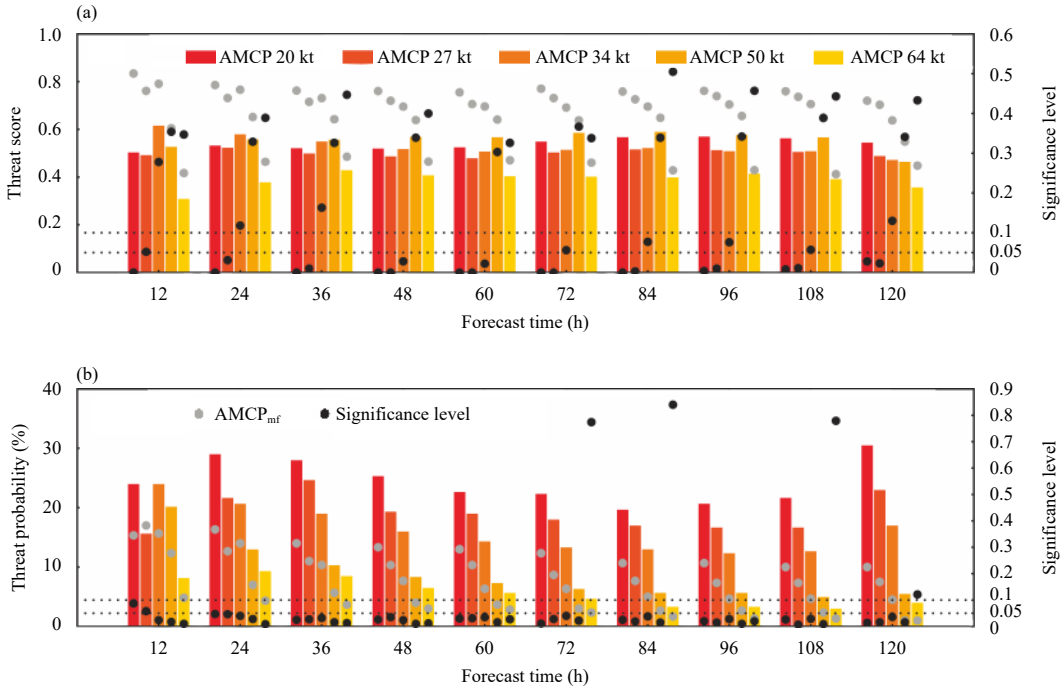

Fig. 6. (a) Maximum threat score (TS) and (b) corresponding threat probability (TP) of AMCP (bars) and $AMCP_{mf}$ (gray points) at different wind speeds for the combined cumulative forecast sets of Hurricanes Irma, Harvey, and Sandy. The black points represent the significance level from the bootstrap *t*-test for the TS and TP change between AMCP and $AMCP_{mf}$.

forecasts (at 24 and 48–120 h) at 34 kt produce significant improvements at the 0.15 significance level, and at the 0.05 significance level for some forecast times. The improvement is not significant for wind speeds of 50–64 kt at a significance level greater than 0.3. The maximum TS indicates the skill of the deterministic forecasts, that is, the binary forecast of whether or not the hurricane can affect a specific location. Therefore, the increased maximum TS at 27–34 kt implies the skill of $AMCP_{mf}$ in deterministic forecasts, namely, the coverage area for the $AMCP_{mf}$ probability map is closer to that from the FNL analysis. As indicated by the discussion in Section 3.1, the AMCP coverage area is controlled mainly by the basic forecast track. Therefore, the introduction of the ensemble model results in the AMCP model modifies the poor track forecast and improves the wind speed probability deterministic forecast at low wind speeds. For TP, the bootstrap *t*-test implies that almost all of the change in TP is significant at significance levels less than 0.1, and less than 0.05 except for the 64-kt forecast at 72, 84, and 108 h. Usually, the lower probabilities always distribute along the edge of the probability map. Therefore, the decreased TP implies that a more accurate probability map can be used to produce binary deterministic (TC affects or does not affect) forecasts. Overall, the $AMCP_{mf}$ model can provide a better deterministic forecast wind speed probability map at 20–34 kt and is more accurate at all wind thresholds.

4. Major hurricanes during the 2018 season

In this section, 120-h wind speed probability forecasts are conducted for two major landfalling hurricanes over the Atlantic Ocean in the 2018 season. Forecasts at 120 h for Hurricane Florence (first and last realization case initialized at 0000 UTC 10 and 1200 UTC 16 September 2018) and Hurricane Michael (first and last realization case initialized at 0000 UTC 8 and 0000 UTC 11 October 2018) are conducted at 0000 and 1200 UTC every day. The improvements of $AMCP_{mf}$ over the ensemble forecasts and AMCP model are quantified.

Similarly, the multiplicative bias for major hurricane cases in 2018 based on the cumulative wind speed probability is examined as a function of lead times (Fig. 7). Similar to the case study for Hurricanes Irma, Harvey, and Sandy, the multiplicative bias is significantly reduced by adding the ensemble forecasts. After modification, especially for the 36–120-h forecast, the multiplicative bias for each wind speed threshold is closer to 1, meaning that $AMCP_{mf}$ produces a wind speed probability closer to the analysis, consisting of the above case study.

Figures 8a and 8b show the TS and TP for the major hurricane cases in the 2018 season based on the cumulat-

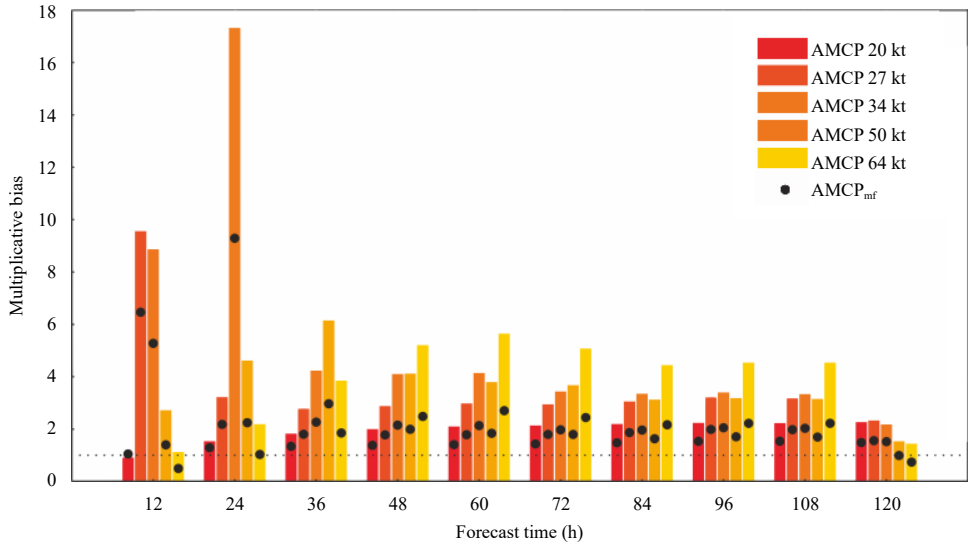

Fig. 7. As in Fig. 5, but for major hurricanes in the 2018 season. The dotted line indicates that the multiplicative bias is equal to 1.

ive wind speed probability at each forecast lead time. As in the case studies in the previous section, for the major hurricanes in the 2018 season, AMCP$_{mf}$ provides larger TS (except that of 64-kt wind speed at 120 h) and smaller TP at each lead time. The bootstrap *t*-test implies that the increased TS is significant at the 0.05 significance level (0.3 significance level) at the 36–120-h forecast for the wind speed threshold of 20–34 kt (50 kt). Simultaneously, the decreased TP at 36–120 h is significant at the 0.05 significance level except for 64-kt wind at 36 and 120 h. Significantly increased TS and/or notably decreased TP indicate a strong enhanced binary deterministic forecast for 20–64-kt wind. Overall, adding the ensemble forecast can provide a more accurate probability map that can be used to produce binary deterministic (TC affects or does not affect) forecasts.

According to the case studies, to some extent, the ensemble forecast is possibly more accurate than AMCP and AMCP$_{mf}$. However, for a public disaster warning system such as MCP/AMCP/AMCP$_{mf}$, missing alarms are more disastrous than false alarms. Considering that the MS and FS of high wind are more critical for public disaster warnings, Fig. 9 shows the false and missing scores (at the TP that generates the largest TS) of the ensemble and AMCP$_{mf}$ based on the cumulative wind speed probability at high wind (50–64 kt) thresholds. Figure 9 indicates that MS is always larger, while FS is often smaller in the ensemble forecasts. Compared to the

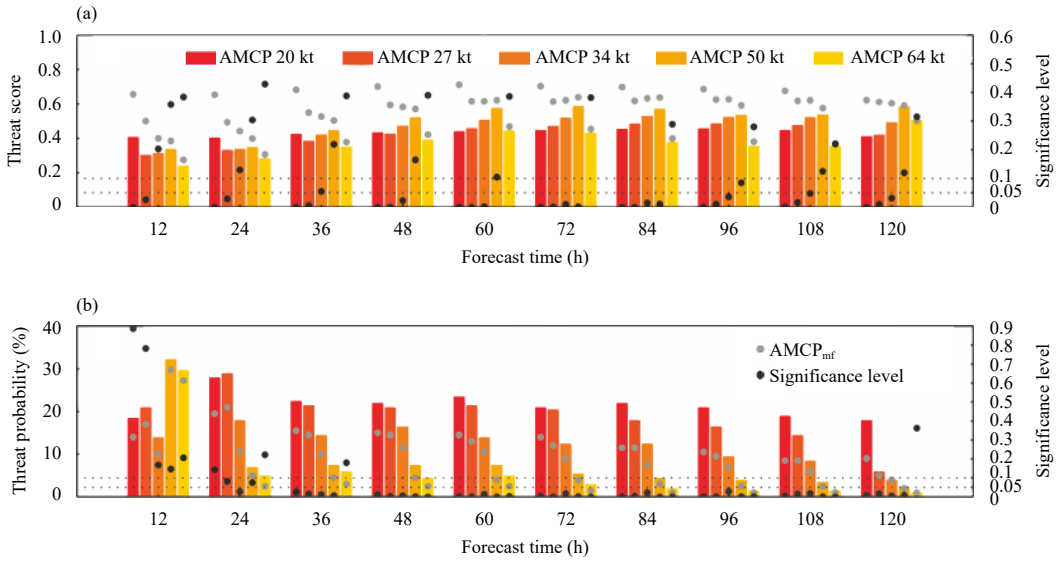

Fig. 8. As in Fig. 6, but for the combined cumulative forecast sets of the major hurricanes in the 2018 season.

strong reduced MS of $AMCP_{mf}$, especially for the wind speed of 64 kt, FS often increases slightly. The combination of AMCP and the ensemble forecasts could extend the wind speed probability map (see Figs. 1–3) and reduce MS from the ensemble forecasts. Figure 9 also indicates that the reduced missed alarms will not all become false alarms after the combination, and some will become hits. Therefore, with decreased MS and slightly increased TS, the extended wind speed probability map from $AMCP_{mf}$ can provide useful hit warning information that the ensemble forecasts miss. Due to the sample size issue, namely, the limited size of ensemble members, the probability map from the ensemble forecasts is often smaller than that from $AMCP_{mf}$. The smaller disaster warning map could lead to many missed alarms and thus result in a severe disaster. Therefore, the combination of AMCP and ensemble forecasts would be better than either one alone for providing a public disaster warning (see Fig. 2).

5. Concluding remarks

In this study, we first made a slight adjustment to the Monte Carlo hurricane wind speed probability model by adding limits to the direction angle and motion distance. Then, we introduced ensemble probability results derived from ECMWF, UKMO, and NCEP ensemble forecasts to this model to improve the predictions based on GFS model forecasts. Hurricanes Irma (2017), Harvey (2017), and Sandy (2012) and the major hurricanes in the 2018 season are used to evaluate the effectiveness of the modification.

The probability map indicates that the ensemble forecasts can improve not only inland forecasts but also forecasts over the ocean, and this combination ($AMCP_{mf}$) is better than either single model result. With 1000 random samples, the realizations generate a large number of possible tracks and cover larger areas, and this poor forecast, especially inland, can be corrected by the ensemble forecasts. Meanwhile, the AMCP model can be a good complement to the ensemble forecasts due to the limited sample size of the ensemble members.

The significant improvement in the maximum threat score, as well as the significantly decreased corresponding threat probability, implies that the $AMCP_{mf}$ model can provide a better deterministic forecast wind speed probability map at 20–64 kt. The scatter diagram shows that $AMCP_{mf}$ for all cases is close to the "$x = y$" line, which indicates a better wind speed probability forecast during the 120-h forecasts.

Overall results from this study prove the benefits of combining the MCP model with ensemble forecasting in potential applications for improved TC forecasts. For disaster warnings to the public, the increased accuracy of $AMCP_{mf}$ could provide accurate wind speed probability forecasts. The increased deterministic forecast ability means more information from $AMCP_{mf}$ can be used to provide a disaster warning. The larger threat score and

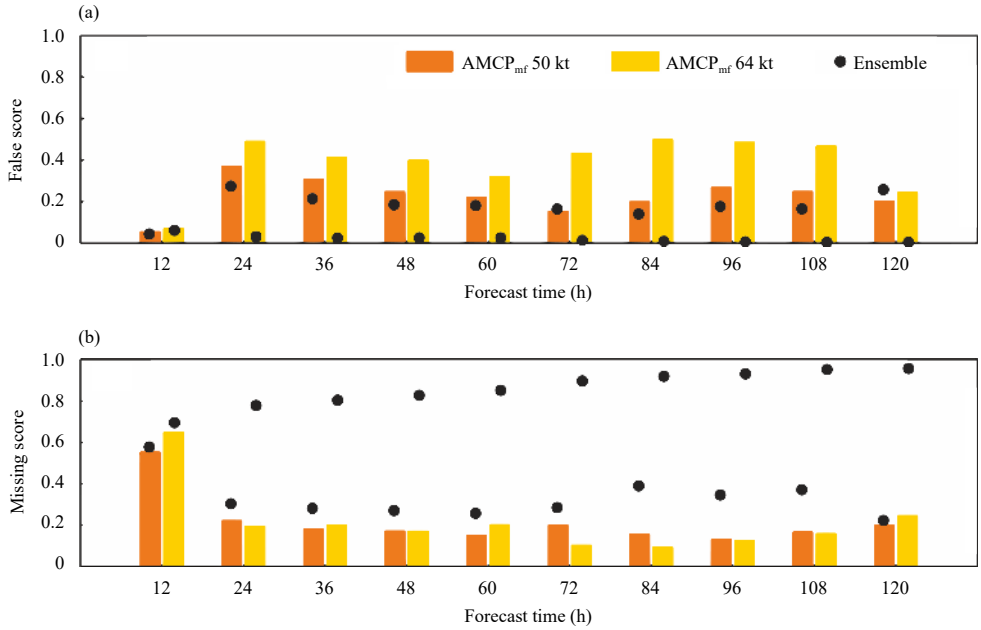

Fig. 9. (a) False score (FS) and (b) missing score (MS) of $AMCP_{mf}$ (bars) and ensemble (black points) at wind speeds of 50–64 kt for the combined cumulative forecast sets of the 2018 major hurricane season. MS and FS are collected at the threat probability corresponding to the maximum threat score in each model.

smaller threat probability imply a larger map to determine which places TCs can or cannot affect. Although the ensemble is often more accurate than $AMCP_{mf}$, missed alarms (not forecasted but observed) are more disastrous than false alarms (forecasted but not observed) for public warnings. Direct application of the ensemble forecasts would miss some early warnings for high wind, which would go against the purpose of the AMCP. Therefore, the combination of MCP and the ensemble could produce better wind speed probability forecasts.

Acknowledgments. Archived hurricane data and ensemble forecast data are obtained from the public website hosted by the University Corporation for Atmospheric Research (UCAR; https://rda.ucar.edu/) and the National Hurricane Center (https://ftp.nhc.noaa.gov/atcf/archive/).

REFERENCES

Barkmeijer, J., R. Buizza, E. Källén, et al., 2013: 20 years of ensemble prediction at ECMWF. ECMWF Newsletter, No. 134, ECMWF, Reading, 16–32, doi: 10.21957/l6nteidw.

Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525**, 47–55, doi: 10.1038/nature14956.

Buizza, R., 2014: The TIGGE Global, Medium-Range Ensembles. ECMWF Technical Memorandum, No. 739, ECMWF, Reading, 51 pp, doi: 10.21957/khygx7grk.

DeMaria, M., J. A. Knaff, R. Knabb, et al., 2009: A new method for estimating tropical cyclone wind speed probabilities. *Wea. Forecasting*, **24**, 1573–1591, doi: 10.1175/2009WAF2222286.1.

DeMaria, M., J. A. Knaff, M. J. Brennan, et al., 2013: Improvements to the operational tropical cyclone wind speed probability model. *Wea. Forecasting*, **28**, 586–602, doi: 10.1175/WAF-D-12-00116.1.

Efron, B., and R. J. Tibshirani, 1994: *An Introduction to the Bootstrap*. CRC Press, Boca Raton, 456 pp.

Emanuel, K., and F. Q. Zhang, 2016: On the predictability and error sources of tropical cyclone intensity forecasts. *J. Atmos. Sci.*, **73**, 3739–3747, doi: 10.1175/JAS-D-16-0100.1.

Frame, T. H. A., M. H. P. Ambaum, S. L. Gray, et al., 2011: Ensemble prediction of transitions of the North Atlantic eddy-driven jet. *Quart. J. Roy. Meteor. Soc.*, **137**, 1288–1297, doi: 10.1002/qj.829.

Froude, L. S. R., 2010: TIGGE: Comparison of the prediction of Northern Hemisphere extratropical cyclones by different ensemble prediction systems. *Wea. Forecasting*, **25**, 819–836, doi: 10.1175/2010WAF2222326.1.

Froude, L. S. R., 2011: TIGGE: Comparison of the prediction of Southern Hemisphere extratropical cyclones by different ensemble prediction systems. *Wea. Forecasting*, **26**, 388–398, doi: 10.1175/2010WAF2222457.1.

Goerss, J. S., 2007: Prediction of consensus tropical cyclone track forecast error. *Mon. Wea. Rev.*, **135**, 1985–1993, doi: 10.1175/MWR3390.1.

Jarvinen, B. R., C. J. Neumann, and M. A. S. Davis, 1984: A Tropical Cyclone Data Tape for the North Atlantic Basin, 1886-1983: Contents, Limitations, and Uses. NOAA Technical Memorandum NWS NHC 22, National Hurricane Center, Miami, Florida, 21 pp.

Kaplan, J., and M. DeMaria, 1995: A simple empirical model for predicting the decay of tropical cyclone winds after landfall. *J. Appl. Meteor.*, **34**, 2499–2512, doi: 10.1175/1520-0450(1995)034<2499:ASEMFP>2.0.CO;2.

Knaff, J. A., C. R. Sampson, M. DeMaria, et al., 2007: Statistical tropical cyclone wind radii prediction using climatology and persistence. *Wea. Forecasting*, **22**, 781–791, doi: 10.1175/WAF1026.1.

Knapp, K. R., M. C. Kruk, D. H. Levinson, et al., 2010: The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone data. *Bull. Amer. Meteor. Soc.*, **91**, 363–376, doi: 10.1175/2009BAMS2755.1.

Matsueda, M., and T. Nakazawa, 2015: Early warning products for severe weather events derived from operational medium-range ensemble forecasts. *Meteor. Appl.*, **22**, 213–222, doi: 10.1002/met.1444.

Palmer, T. N., R. Buizza, F. Doblas-Reyes, et al., 2009: Stochastic Parameterization and Model Uncertainty. ECMWF Technical Memorandum, No. 598, ECMWF, Reading, 42 pp, doi: 10.21957/ps8gbwbdv.

Rappaport, E. N., J. L. Franklin, L. A. Avila, et al., 2009: Advances and challenges at the National Hurricane Center. *Wea. Forecasting*, **24**, 395–419, doi: 10.1175/2008WAF2222128.1.

Sheets, R. C., 1985: The National Weather Service hurricane probability program. *Bull. Amer. Meteor. Soc.*, **66**, 4–13, doi: 10.1175/1520-0477(1985)066<0004:TNWSHP>2.0.CO;2.

Smith, J. L., 2017: Probabilistic hurricane track generation for storm surge prediction. Ph. D. dissertation, University of North Carolina at Chapel Hill Graduate School, USA, 38.

Splitt, M. E., J. A. Shafer, S. M. Lazarus, et al., 2010: Evaluation of the National Hurricane Center's tropical cyclone wind speed probability forecast product. *Wea. Forecasting*, **25**, 511–525, doi: 10.1175/2009WAF2222279.1.

Swinbank, R., M. Kyouda, P. Buchanan, et al., 2016: The TIGGE project and its achievements. *Bull. Amer. Meteor. Soc.*, **97**, 49–67, doi: 10.1175/BAMS-D-13-00191.1.

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319, doi: 10.1175/1520-0493(1997)125<3297:EFANAT>2.0.CO;2.

Tech & Copy Editor: Qi WANG