Weighted Analog Technique for Intensity and Intensity Spread Predictions of Atlantic Tropical Cyclones

HSIAO-CHUNG TSAI

Department of Water Resources and Environmental Engineering, Tamkang University, New Taipei City, Taiwan

RUSSELL L. ELSBERRY

Trauma, Health, Hazards Center, University of Colorado Colorado Springs, Colorado Springs, Colorado, and Department of Meteorology, Naval Postgraduate School, Monterey, California

(Manuscript received 2 March 2015, in final form 1 July 2015)

ABSTRACT

A situation-dependent intensity and intensity spread prediction technique for the Atlantic called the Weighted Analog Intensity Atlantic (WAIA) is developed using the same procedures as for a similar technique for the western North Pacific that is operational at the Joint Typhoon Warning Center. These simple techniques are based on rankings of the 10 best historical track analogs to match the official track forecast and current intensity. A key step is the development of a bias correction to eliminate an overforecast bias. The second key step is a calibration of the original intensity spread among the 10 analogs to achieve a probability of detection of about 68% at all forecast intervals, which it is proposed would be an appropriate intensity spread for the National Hurricane Center (NHC) official intensity forecasts. The advantages of WAIA as an operational intensity forecast product for Atlantic tropical cyclones are described in terms of mean absolute errors, sample-mean biases, and geographic distributions of WAIA versus various guidance products available at NHC. Specific attention is given to the four guidance products that are included in the intensity consensus (ICON) technique that is the most skillful of all the products. Evidence is given that WAIA would be an independent, and more likely skillful at longer forecast intervals, technique to include in ICON. Consequently, WAIA would likely lead to improved NHC intensity forecasts at 4–5-day intervals.

1. Introduction

Contrary to what is often reported in the literature, DeMaria et al. (2014) have demonstrated that tropical cyclone intensity forecast techniques and models have improved at 48 h over the past two decades at a rate that is statistically significant. This conclusion is particularly true in the Atlantic where skillful statistical–dynamical and dynamical models are available to National Hurricane Center (NHC) forecasters. Of course, the Atlantic basin also has the advantage of aircraft reconnaissance and excellent remote sensing observations that are not available in some other basins. As has been the case for track forecasting, an intensity consensus (ICON; see the appendix for acronym expansions) of four skillful intensity guidance products provides more accurate intensity forecasts than do the individual guidance products over a sufficiently large sample.

The purpose of this article is to introduce another skillful intensity guidance product for the Atlantic that is based on a selection of historical analogs with similar tracks and current intensities and that, thus, can be produced on a desktop computer in a few seconds. However, the primary usefulness of this new technique is to provide an estimate of the uncertainty in the intensity, which can be determined based on the intensity spread among these historical analogs. Goerss and Sampson (2014) have defined a measure of forecast intensity uncertainty in terms of confidence intervals within which the verifying intensity will occur 67% of the time. Their technique uses various predictors including the initial intensity and the spread from another NHC intensity consensus model [intensity variable consensus (IVCN)]. While the Goerss and Sampson technique is predicting the IVCN forecast errors, these

---

Corresponding author address: R. L. Elsberry, Trauma, Health, Hazards Center, University of Colorado Colorado Springs, 1420 Austin Bluff, Colorado Springs, CO 80919.
E-mail: elsberrylr@comcast.net

DOI: 10.1175/WAF-D-15-0030.1

errors are likely to be representative of the NHC intensity forecast errors.

Availability of this new intensity spread estimate would give NHC forecasters another technique for providing intensity uncertainty information. Such an intensity spread product would be useful for the disaster management community in preparing for the threat of an Atlantic tropical cyclone.

This Weighted Analog Intensity Atlantic (WAIA) technique follows from the technique developed by Tsai and Elsberry (2014) for western North Pacific tropical cyclones [Weighted Analog Intensity Pacific (WAIP)]. In the WAIP, 10 historical track analogs are matched with the Joint Typhoon Warning Center (JTWC) official track forecast. Tsai and Elsberry (2014) have demonstrated that the mean absolute errors (MAEs) for WAIP at 120 h were 5 knots (kt; where 1 kt = $0.51\,\mathrm{m\,s^{-1}}$) better (20%) than the JTWC official intensity errors. In addition, situation-dependent intensity spread guidance was generated that included about 68% of the verifying intensities at all forecast intervals to 120 h. Finally, Tsai and Elsberry provided examples of the WAIP intensity spread guidance to illustrate how the JTWC forecaster might use this information for potential landfall and intensity bifurcation (two mode) situations. This WAIP is now operational at JTWC.

A brief description of WAIA and an evaluation of its accuracy compared to the intensity guidance methods available at NHC will be provided in section 2. In addition, the methodology for providing intensity spread guidance for Atlantic tropical cyclones and a demonstration that this uncertainty measure includes about 68% of the verifying intensities are provided in section 2. The performance of WAIA relative to other intensity forecast guidance available at NHC is summarized in section 3. Concluding remarks are given in section 4.

2. Weighted Analog Intensity Atlantic prediction technique

The basic premise of WAIA is that the track of the tropical cyclone is a primary determinant of the intensity on time scales of 3–5 days or longer, and the spread among the 10 best historical analogs provides a measure of the range of environmental influences that may be occurring along that track. It is of course also important to consider at what stage in the life cycle the tropical cyclone is at, that is, is the tropical cyclone still a tropical depression or has it already become a tropical storm or a hurricane? A challenge for 5-day intensity forecasts in the Atlantic is whether (and when) an African easterly wave will develop in the eastern Atlantic or farther west over the warm ocean in the western Atlantic or Caribbean. Another challenge is whether (and when) landfall will occur with rapid decay. For recurvature situations, maximum intensity is expected to occur at or shortly after the time of recurvature because tropical cyclones tend to weaken in the postrecurvature phase in response to both the increasing vertical wind shear associated with the midlatitude trough and the decreasing sea surface temperatures to the north.

As in WAIP (Tsai and Elsberry 2014), the 10 best historical analogs for the Atlantic are selected from the NHC best tracks within ±30 days of the current date and ranked according to the average Euclidean track distance d_T between the target storm and the candidate analog storm plus the initial magnitude of the intensity difference d_V between the target and analog storms. While the 10 best historical track analogs are sought from the NHC files from 1945 to 2009, the NHC official (OFCL) track forecasts from a training dataset during the 2006–09 seasons are matched rather than the NHC best tracks. This use of the NHC official track forecasts is consistent with the statistical–dynamical intensity forecasts available at the NHC (DeMaria et al. 2014).

Following the procedure in WAIP (Tsai and Elsberry 2014), the average track distance between the NHC track forecast and the candidate analog storm is calculated with a linearly varying weighting factor from 1.0 at the initial time to 2.0 at 72 h and then a constant weighting factor of 2.0 in the 72–120-h interval. A range of weighting factors for d_T and the initial intensity difference was tested for the sample of NHC official forecast tracks in the development of WAIA (Fig. 1). In contrast to the development of WAIP in which the weighting factors of 0.8 for d_T and 0.2 for d_V are used to rank of the potential analogs, equal weighting factors (0.5) were found to be most appropriate for WAIA. This equal weighting may be attributed to the similarity of many tracks in the Atlantic (especially for those tropical cyclones that form in the main development region) that can have different intensities depending on where along those tracks the formation has occurred or will occur.

As in Tsai and Elsberry (2014), the final ranking of the candidate analogs is according to

$$\mathrm{Rank}_{\mathrm{analog}} = W_T\mathrm{Rank}_T + W_V\mathrm{Rank}_V, \qquad (1)$$

where $W_T + W_V = 1.0$. The $\mathrm{Rank}_{\mathrm{analog}}$ is then sorted in ascending order to select the 10 best analogs. The choice of only 10 analogs is made because it is difficult to find a larger number when the ranking is by similarity with the NHC track forecast, initial intensity difference, and within ±30 days. A second conditioning on the initial intensity V_0 in the WAIP technique for subsamples of

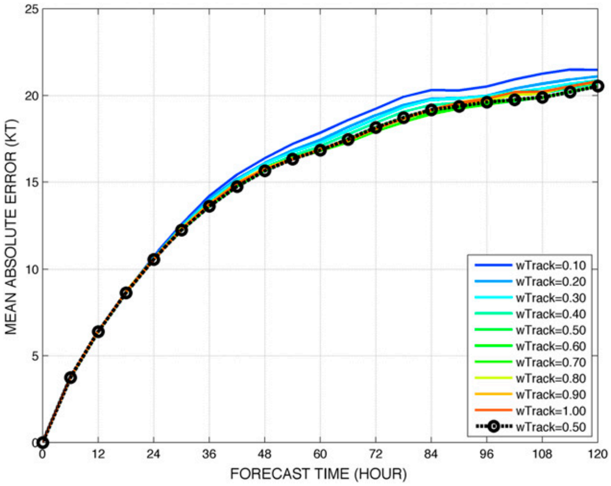

FIG. 1. Test of the optimum weighting factor for the ranking of the track analogs in Eq. (1) using a range of values from 0.1 to 1.0 (see colors in legend) in terms of the MAE in intensity (kt). An optimum value of 0.5 is indicated by the circles with a black dashed line.

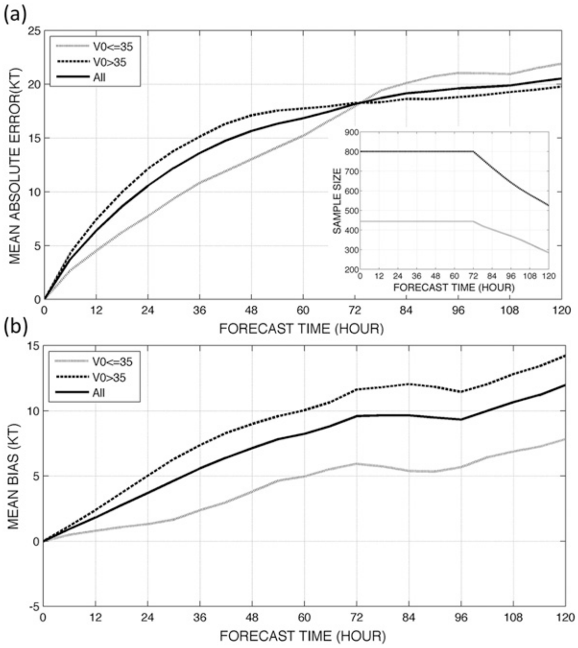

FIG. 2. (a) MAE and (b) mean bias for the WAIA hindcast intensity (kt) subsamples of initial intensities $V_0 \le 35$ and $V_0 > 35$ kt and all samples (see lines in legend) prior to the bias correction and calibration steps. The inset in (a) indicates the sample sizes for the $V_0 \le 35$ (gray dotted line) and $V_0 > 35$ kt (black dotted line) subsamples. Note that only storms that lasted at least 72 h are included in the sample.

$V_0 \le 35$ and $V_0 > 35$ kt was also applied in WAIA. The mean absolute errors for the $V_0 \le 35$ kt sample increase at a much slower rate with forecast interval up to 72 h than do the MAEs for the $V_0 > 35$ kt sample (Fig. 2a). In addition, these two samples have different mean bias characteristics (Fig. 2b) with an overestimation of the intensity, especially for the $V_0 > 35$ kt sample.

Another special feature of WAIP that is also applied in the WAIA technique is to give greater weight to those analog tracks that most closely match the NHC track forecast (e.g., match the path leading to the landfall position and timing). The weighted mean intensity V_w of the n analogs at each time t is

$$V_w = \sum_{i=1}^{n} (w_i V_i) \Big/ \sum_{i=1}^{n} w_i, \tag{2}$$

where V_i is the intensity of the ith track analog and $w_i = (1/d_{T,i})/\sum_{i=1}^{n}(1/d_{T,i})$.

Tsai and Elsberry (2014) developed intensity spread guidance at each forecast interval to 120 h by utilizing the 10 intensities from the 10 best historical track analogs to the JTWC official track forecasts. First, an intensity bias correction was developed to reduce the bias in the WAIP forecasts arising from the weighted average of the 10 best analogs. The second step was to calibrate the spread among the 10 intensity estimates from the track analogs such that at each 12 h the probability of detection (POD) is at least about 68.26%.

The first step in reducing the mean intensity bias in Fig. 2b was to develop a training set by randomly selecting 70% of the ~1300 WAIA forecasts during the 2006–09 seasons, and then retain the other 30% of these forecasts for an independent verification. A linear regression was used at each 12 h to correct for the bias in the intensity:

$$V_m = a'X + b', \tag{3}$$

where V_m is the bias-corrected intensity, X represents various predictors, and a' and b' are the regression coefficients. In addition to the weighted average of the 10 intensities from the analogs, the latitude, longitude, initial intensity, spread of the initial intensities, and spread among the 10 analog tracks are also provided as potential predictors. The success of this bias correction is indicated in Fig. 3b with the WAIA mean intensity bias for the independent sample being reduced to less than 1 kt over all forecast intervals. Whereas the MAEs for the original WAIA forecasts (Fig. 3a, gray dashed line) increased steadily to more than 20 kt at 120 h, the MAEs in the independent sample after application of the bias correction (Fig. 3a, circles with a solid black line) increased more slowly with forecast interval and had a maximum of about 16 kt at 120 h. This reduction in the intensity forecast biases of up to 10 kt in the original WAIA forecasts to near-zero biases will be shown in

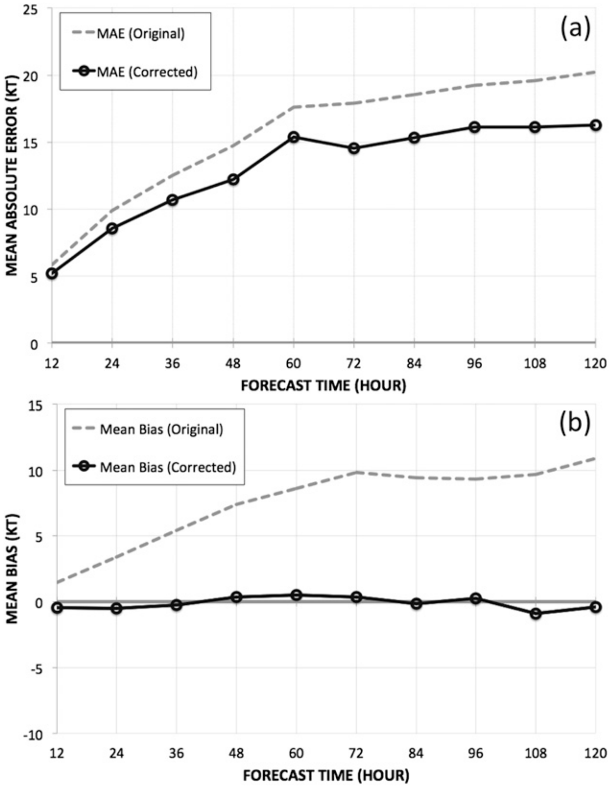

FIG. 3. (a) MAE and (b) mean bias for the WAIA hindcast intensity (kt) independent sample before (dashed line corresponding to all samples in Fig. 2b) and after (circles with solid line) bias correction is applied.

section 3 to be a factor in the improvement of WAIA over other intensity forecast techniques at NHC.

Following Tsai and Elsberry (2014), the objective of the intensity spread calibration for WAIA is to achieve as closely as possible the goal of a plus or minus one standard deviation (68.26%) POD since the intensities are assumed to be normally distributed about the mean. That is, the $P[|z| \leq 1.0]$, where z is the normal distribution z score $(x - \mu)/s$. Here, x is the variable, μ is the mean, and s is the standard deviation. In the WAIA technique, the raw intensity spreads σ are also weighted:

$$\sigma = \sqrt{\sum_{i=1}^{n}[w_i(V_i - V_w)^2] \Big/ \sum_{i=1}^{n} w_i}, \qquad (4)$$

where $i = 1 - n$ analogs and $w_i = (1/d_{T,i})/\sum_{i=1}^{n}(1/d_{T,i})$. The intensity spread calibration at each time is

$$\sigma' = |a\sigma + b|, \qquad (5)$$

where σ' is the calibrated intensity spread, σ is the original spread, and a and b are the calibration factors to be determined. The calibration factor a is constrained as $0.5 < a < 1.5$ to avoid excessive reductions within the overdetermined region or excessive amplifications in the underdetermined region. The calibration factor b is constrained to $-0.5\overline{\sigma} \leq b \leq 0.5\overline{\sigma}$, where $\overline{\sigma}$ is the overall sample-mean forecast spread at forecast interval t_i to ensure realistic values.

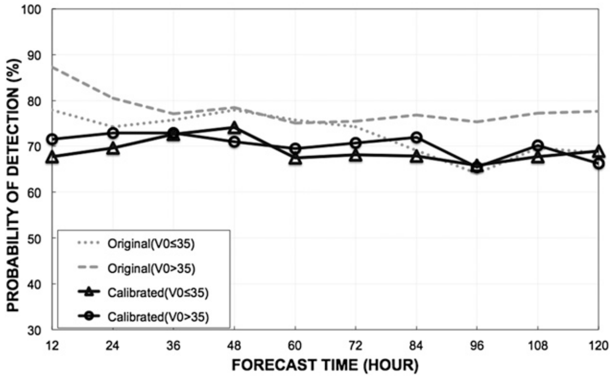

FIG. 4. POD as a function of forecast interval that the verifying intensity will be within the original noncalibrated (dashed lines) and calibrated (solid lines) spreads of the intensities of the 10 best-track analogs in the WAIA technique for subsamples with initial intensities $V_0 \leq 35$ and $V_0 > 35$ kt (see lines in legend). These PODs are only shown from the 12-h forecast interval and the calibrated PODs are for the independent dataset.

A cost function is calculated from the training sample:

$$J_t = J_{1t} + J_{2t} + J_{3t}, \qquad (6)$$

where t is the forecast interval; J_{1t} is the probability of having POD $\geq$ 68.26%; J_{2t} is the correlation coefficient r between σ_i' and E_i, which is the intensity forecast error defined as forecast minus observed; and J_{3t} is the penalty term, which is the probability of having the $L_{\text{ratio}} = \sigma_i'/E_i \geq 2.0$. Given the different units in the three cost functions, each of them is normalized by dividing by their standard deviation to define a modified cost function J' as

$$J_t' = J_{1t}' + J_{2t}' + J_{3t}'. \qquad (7)$$

This modified cost function is then minimized to obtain the calibration factors a and b at each 12-h forecast interval.

As expected for the training dataset (not shown), the calibration procedure successfully adjusts to about 68% of the over- and underdetermined raw intensity spreads for all of the samples and for the $V_0 \leq 35$ and $V_0 > 35$ kt subsamples. For the independent dataset (30% of the 1284 cases), the POD for the noncalibrated intensity spread for the $V_0 > 35$ kt subsample (Fig. 4, long dashed line) is overdetermined (intensity spread is too large) at all forecast intervals. After calibration (Fig. 4, circles

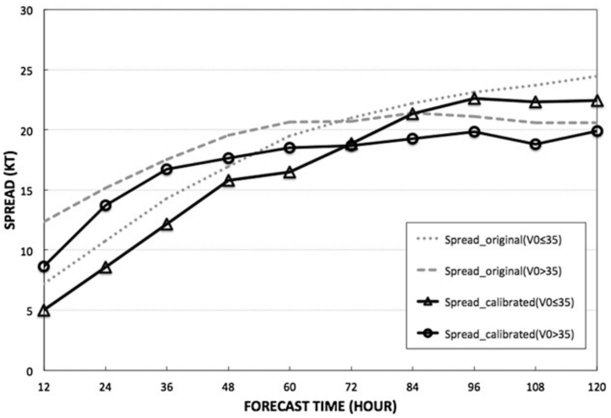

FIG. 5. Noncalibrated (dashed lines) and calibrated (solid lines) spread of the intensities for the independent sample of WAIA intensity forecasts as a function of forecast interval for the $V_0 \leq 35$ and $V_0 > 35$ kt subsamples.

with a solid black line), the POD for the $V_0 > 35$ kt subsample is still overdetermined (>70%) through 84 h. The original $V_0 \leq 35$ kt subsample (Fig. 4, short dashed line) is also overdetermined (>70%) through 72 h. After calibration, the POD for the $V_0 \leq 35$ kt subsample is close to the desired 68% throughout the 120-h forecast interval. Therefore, the calibration procedure is largely successful in reducing the POD for the overdetermined intensity spreads, as well as increasing the POD for the underdetermined intensity spreads.

As indicated in Fig. 5, the "raw" intensity spread at each 12 h among the 10 best historical analogs for the $V_0 > 35$ kt subsample of the independent sample (Fig. 5, long dashed line) starts at ±12.5 kt already at 12 h, and then increases to ±20 kt at 48 h and beyond. For the $V_0 \leq 35$ kt subsample of the independent sample (Fig. 5, short dashed line), the non-bias-corrected and non-calibrated intensity spread starts at ±7.5 kt, but then increases to ±20 kt by 64 h and continues to increase to ±25 kt at 120 h. The growth in these intensity spread tendencies with time is somewhat analogous to the growth in the MAEs for these subsamples in Fig. 2a. For the $V_0 \leq 35$ kt sample, the intensity spread at the early forecast intervals is not much larger than the discretization interval for intensity estimates (5 kt). However, the storms in this subsample may remain at the tropical depression stage for the entire 120-h forecast interval, or experience one or more rapid intensifications (defined here as 30 kt day^{-1}) to an intense hurricane, so the raw intensity spread increases rapidly to large values (Fig. 5, short dashed line). For the $V_0 >$ 35 kt (tropical storm or greater) subsample, the likelihood of the analogs having a standard intensification and decay cycle is greater, so the raw intensity spread tends to level off after early growth (Fig. 5, long dashed line).

After applying a bias correction and a calibration procedure to the $V_0 \leq 35$ kt subsample, the intensity spread at 12 h is indeed equal to the 5-kt discretization interval (Fig. 5, triangles with a solid black line). However, a subsequent rapid increase in intensity spread (uncertainty) is still to be expected owing to the wider range of possibilities from no intensification to rapid intensification. The success of the intensity bias correction and calibration is also evident for the $V_0 > 35$ kt subsample (Fig. 5, circles with a solid black line), as the 12-h intensity spread is reduced to ±9 kt and a consistent reduction of ±2 kt over the 48–120-h forecast interval.

Since these calibrated intensity spreads are quite successful in representing the uncertainty in the situation-dependent WAIA technique, they are proposed to provide first-order intensity spread guidance for the POD for the NHC official intensity forecast. Given the limited skill of the present official intensity forecasts, it is advocated that this WAIA and the Goerss and Sampson (2014) intensity spread guidance be added to the NHC warnings so that too much focus is not given to the single line representing the intensity forecast. Rather, the intensity spread guidance can provide useful uncertainty information for the forecasters, decision-makers, and informed members of the public.

3. Comparisons of WAIA with other intensity forecast guidance

DeMaria et al. (2014) discuss the origins of the various intensity forecast guidance products available at the NHC. In this section, the WAIA intensity forecast errors will be compared with the errors for the same guidance products that DeMaria et al. discussed. Sample sizes for these homogeneous comparisons during the 2010–13 Atlantic seasons are listed in Table 1. Although the sample sizes decrease with increasing forecast intervals, the minimum number of cases is 499 at 120 h.

The MAE comparisons for these homogeneous samples are summarized in Table 2, with positive (negative) values indicating that the guidance product has a larger (smaller) MAE than for WAIA. Statistically significant differences at the 5% level are indicated by asterisks, and where WAIA (other guidance) is the more accurate technique, the value is highlighted by positive (negative) values.

a. Comparison of WAIA with SHF5

The Statistical Hurricane Intensity Forecast model (SHIFOR) 5-day version (SHF5) is given in row 1 of Tables 1–3. This model by Knaff et al. (2003) uses only

TABLE 1. Sample sizes for the homogeneous comparisons of the WAIA technique with various intensity forecast guidance techniques available at NHC during the 2010–13 Atlantic seasons. Definitions of the four-letter acronyms for the various guidance techniques and the NHC OFCL intensity forecasts are given in the text.

Guidance ID	Forecast interval					
	24 h	48 h	72 h	96 h	108 h	120 h
SHF5	1140	1060	911	727	642	588
SHIPS	1154	1068	903	716	641	575
DSHP	1154	1068	903	716	641	574
LGEM	1140	1070	939	757	665	609
GHMI	1113	1027	870	677	587	521
HWFI	1128	1042	887	706	622	560
ICON	1125	1029	854	658	571	499
OFCL	1157	1075	924	737	651	597

TABLE 2. As in Table 1, but for the guidance technique minus the WAIA technique MAEs (kt). Cases where WAIA (other guidance) is the more accurate technique are highlighted by positive (negative) values.

Guidance ID	Forecast interval					
	24 h	48 h	72 h	96 h	108 h	120 h
SHF5	0.10	1.26	3.01*	3.66*	3.59*	4.65*
SHIPS	−0.76	0.42	1.70*	1.34*	1.38*	1.39*
DSHP	−1.10	−1.12	−0.63	−0.27	−0.02	0.06
LGEM	−0.94	−0.96	−0.45	−0.63	0.24	0.67
GHMI	−0.84	−0.85	0.17	1.95*	2.51*	3.21*
HWFI	−1.35*	−1.68*	−0.05	1.37*	2.70*	3.79*
ICON	−2.29*	−3.38*	−2.66*	−2.35*	−1.81	−1.40
OFCL	−2.24*	−2.41*	−2.42*	−1.83*	−1.38	−1.11

* Statistically significant difference at the 5% level.

climatology and persistence variables available at the initial time and, thus, has been used as a measure of intensity forecast skill. That is, if the MAEs for an intensity guidance product are smaller than for SHF5, then that guidance is said to have skill. Another application of such a skill measure is that if the seasonal-mean SHF5 intensity forecast errors are larger than the long-term mean errors, the storms in that season are considered to have been more difficult to forecast. The considerably larger SHF5 errors at all forecast intervals for the western North Pacific (DeMaria et al. 2014, their Fig. 4b) than for the Atlantic (DeMaria et al. 2014, their Fig. 4a) suggest that it is more difficult to forecast intensity in the western North Pacific than in the Atlantic (assuming equivalent climatological files and capability to prepare persistence of past motion forecasts). The fact that the WAIP intensity forecast errors in Tsai and Elsberry (2014) are consistently larger than the WAIA forecast errors at all forecast intervals in this study also supports the conclusion that intensity forecasting is more difficult in the western North Pacific.

Elsberry and Tsai (2014) and Tsai and Elsberry (2014) suggest that an alternate intensity skill metric might also include knowledge of the official track forecast since that information is also available at the time of the intensity forecast. Specifically, Tsai and Elsberry (2014) propose that WAIP, which matches the 10 historical analogs in the western North Pacific to the JTWC official track forecast, could reveal the additional intensity skill beyond that available from knowledge of the official track forecasts. For example, all of the SHF5–WAIA differences in MAE are positive with statistically significant differences from 72 to 120 h. If WAIA was used as the alternate intensity skill measure, the guidance product intensity MAE at 120 h would need to have at least a 4.65 kt smaller MAE than for SHF5 to have skill beyond knowledge gained from the official track forecast.

One of the explanations for the smaller MAEs for WAIA compared to SHF5 is the bias correction for WAIA that reduced the mean biases in Fig. 2b to near-zero values (see Fig. 3a). Note in Table 3 (row 1) that SHF5 has larger mean biases than for WAIA. These mean bias differences are statistically significant for the 48-h forecast interval and beyond, and the SHF5 mean biases at 72–120 h are particularly large compared to the near-zero WAIA mean biases. It is highly likely that these large mean biases are a factor in the larger statistically significant MAEs for SHF5 compared to WAIA (Table 2).

Another display to illustrate the advantage of the WAIA relative to another intensity guidance product is by geographic areas. The average intensity forecast improvements for WAIA relative to SHF5 in 10° × 10° latitude–longitude boxes are shown in Fig. 6. Note that WAIA has skill relative to SHF5 at 72 h (Fig. 6a) of 15–20 kt for initial positions in the eastern Gulf of Mexico and following landfall over the southeastern United

TABLE 3. As in Table 2, but for the WAIA technique vs the guidance difference mean biases (kt).

Guidance ID	Forecast interval					
	24 h	48 h	72 h	96 h	108 h	120 h
SHF5	1.57	4.36*	8.47*	9.82*	11.13*	10.82*
SHIPS	2.47*	6.11*	10.06*	9.39*	8.71*	6.90*
DSHP	0.90	2.93*	6.24*	5.12*	4.12*	2.11
LGEM	−0.43	0.86	3.77*	3.45*	3.72*	2.59
GHMI	−1.06*	0.81	6.48*	9.46*	10.23*	10.91*
HWFI	−1.72*	0.43	5.28*	6.26*	6.77*	7.18*
ICON	−0.29	1.60	6.15*	7.14*	7.07*	6.82*
OFCL	0.71	2.07	4.75*	5.30*	5.93*	5.10*

* Statistically significant difference at the 5% level.

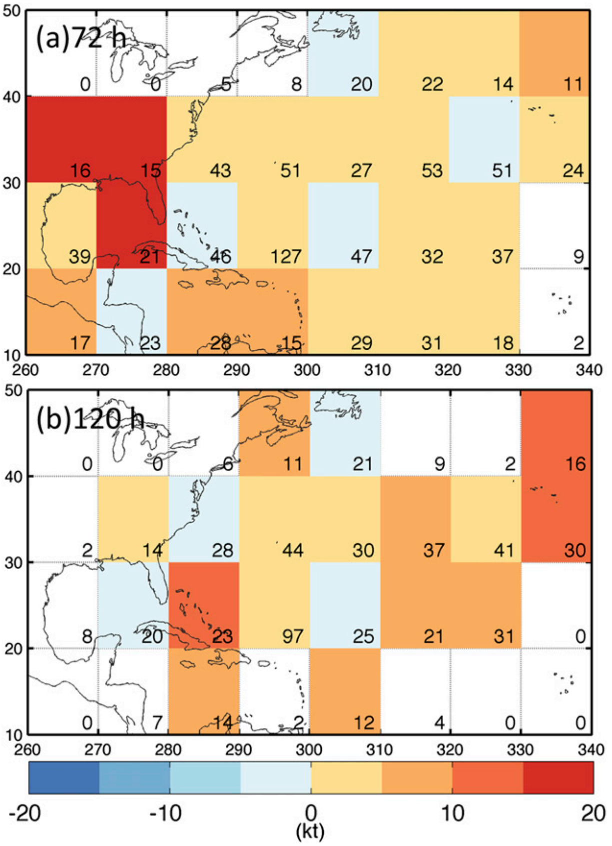

FIG. 6. Average intensity forecast improvements (kt; scale below) from the WAIA technique relative to the SHF5 technique for storms during 2010–13 with forecast tracks through 10° × 10° lat–lon boxes at (a) 72 and (b) 120 h. The sample sizes in each box are indicated, and the white boxes with <10 cases are not compared.

States, as well as skill of 5–10 kt over the eastern and central Caribbean. This advantage of WAIA for potential landfall situations is expected because of the additional weight that is given to the historical analogs matching the 72–120-h portion of the official track forecast (see description in section 2). While most of the rest of the Atlantic has WAIA intensity forecasts at 72 h that are slightly (0–5 kt) more accurate than for SHF5, there are also scattered areas in which this simple climatology and persistence forecast is somewhat more accurate. At 120 h (Fig. 6b), the number of boxes that have at least 10 comparisons is reduced. As at 72 h, the WAIA forecasts at 120 h are generally more accurate than for SHF5, with some regions having 10–15-kt improvements. The large WAIA intensity forecast improvements relative to SHF5 in the eastern Atlantic are likely due to the knowledge of recurvature via the NHC official track forecast that would not have been evident from a persistence-of-past-motion predictor in SHF5.

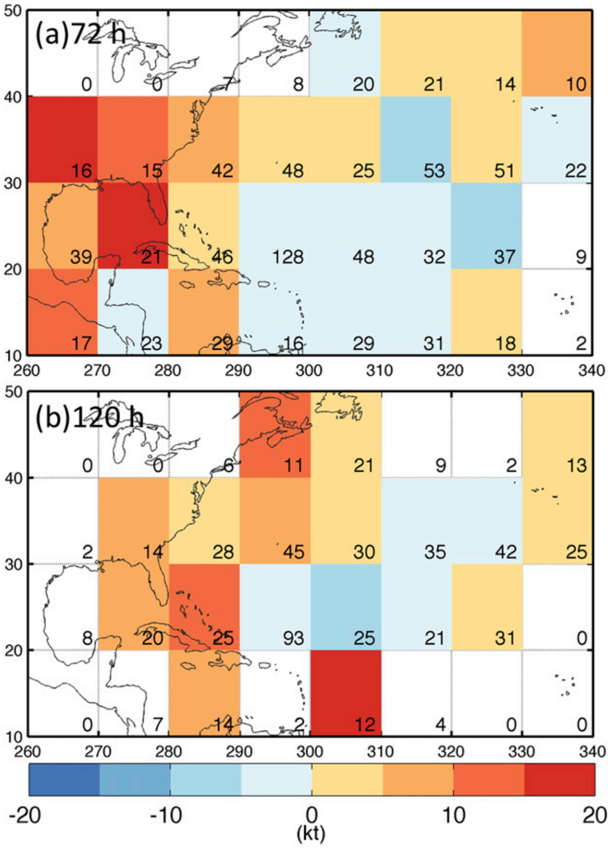

FIG. 7. As in Fig. 6, but for the WAIA technique relative to SHIPS.

b. *Comparison of WAIA with SHIPS and DSHP*

The Statistical Hurricane Intensity Prediction Scheme (SHIPS) is a statistical–dynamical technique in which predictors of climatology, persistence, global model forecast fields, and satellite data are provided along the NHC official track forecast. As indicated in Table 2 (row 2), the MAE at 24 h for SHIPS is 0.76 kt better than for WAIA, but at 48 h and longer forecast intervals WAIA is more accurate. The improvements of WAIA over SHIPS at 72–120 h are statistically significant at the 5% level. One of the reasons that SHIPS is less accurate than WAIA is likely because of the overforecasting bias relative to WAIA (which has a bias correction), with particularly large biases in the 72–120-h forecast intervals (Table 3, row 2). Note in Fig. 7a that the WAIA average intensity forecast improvements at 72 h over SHIPS are very large (as much as 15–20 kt) in the western part of the domain where landfalls may be expected. In contrast the SHIPS predictors add value over WAIA for most of the Atlantic Ocean, with some exceptions in the northeastern Atlantic where WAIA is more accurate. At 120 h (Fig. 7b), WAIA is more accurate than SHIPS

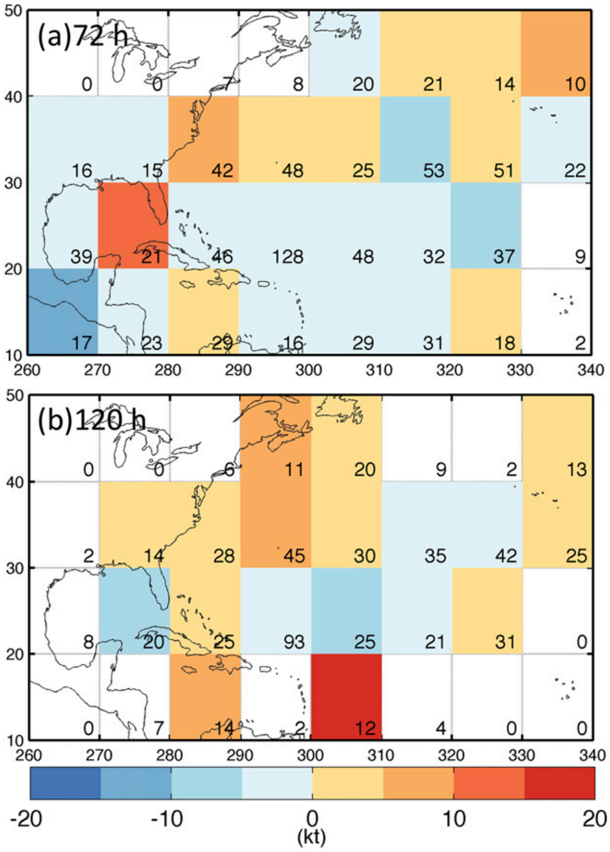

FIG. 8. As in Fig. 6, but for the WAIA technique relative to DSHP.

over almost the entire Atlantic, Caribbean, and Gulf of Mexico, and particularly in all boxes in which landfall is likely to occur. However, the SHIPS predictors add a value of 5–10 kt over WAIA in the 20°–30°N, 50°–60°W box.

The Decay-SHIPS (DSHP) version adds an explicit treatment of the effects following landfall at the location and timing in the NHC official track forecast, but otherwise is identical to SHIPS over the ocean. This additional explicit treatment of land in the DSHP version of SHIPS leads to MAEs for DSHP that are slightly more accurate relative to WAIA, although none of the differences are statistically significant (Table 2, row 3). DSHP has an overforecast bias (Table 3, row 3) relative to the near-zero bias in WAIA (recall Fig. 3a). This DSHP overforecast bias is statistically significant in the 48–108-h forecast interval. The success of the land treatment in DSHP is particularly evident at 72 h in Fig. 8a compared to Fig. 7a for SHIPS. That is, the large advantage of WAIA over SHIPS in Fig. 7a in the western domain does not exist in DSHP, which is then more accurate than WAIA for almost all of the domain south of 20°N. The exceptions are that WAIA is 10–15 kt more accurate than DSHP for storms in the eastern Gulf of Mexico, and also 5–10 kt more accurate for storms just south of Cuba and for storms along the U.S. East Coast. While both WAIA and DSHP utilize the NHC official track forecasts, WAIA includes a track uncertainty about that track forecast that is likely providing a more accurate 72-h intensity in these potential landfall situations.

At 120 h (Fig. 8b), WAIA is more accurate than DSHP over a larger fraction of the domain, and especially north of 30°N in the western Atlantic, where the storms are recurving into the midlatitudes. The improvement of WAIA over DSHP at 72 h in the eastern Gulf of Mexico is reversed at 120 h. This reversal is attributed to the fact that nearly all of these storms will have made landfall somewhere around the Gulf of Mexico by 120 h, and the land effect treatment in DSHP is superior to the weighted intensities of historical analog tracks in the WAIA technique, especially if the historical tracks and intensities have been poorly defined after landfall.

The overall good performance of DSHP has led to its inclusion as one of the four guidance products in ICON available at NHC.

c. *Comparison of WAIA with LGEM*

The Logistic Growth Equation Model (LGEM) is a statistical–dynamical model that includes the same inputs as SHIPS, but utilizes a more sophisticated prediction equation (DeMaria et al. 2014). LGEM is also included in ICON available at NHC.

The slightly better performance of LGEM relative to WAIA in terms of MAEs (Table 2, row 4) from 24 through 96 h is similar to the better performance of DSHP relative to WAIA (Table 2, row 3). Although WAIA is slightly more accurate than LGEM at 108 and 120 h, none of these differences from 24 to 120 h is statistically significant at the 5% level. LGEM also has an overforecast bias relative to the near-zero bias of WAIA (Table 3, row 4), and this LGEM bias difference is statistically significant in the 72–108-h forecast intervals.

Since LGEM uses the same predictors as DSHP, it is not surprising that the geographical distribution of the LGEM performance relative to WAIA (Fig. 9) is very similar to the geographical distribution for DSHP in Fig. 8. At 72 h (Fig. 9a), WAIA again has the better performance in the eastern Gulf of Mexico (by 10–15 kt) and along the U.S. East Coast (by 5–10 kt). In addition, WAIA has a larger advantage over LGEM in the Caribbean, since WAIA is more accurate by 5–10 kt (vs 0–5 kt for DSHP) in the eastern region and 0–5 kt more accurate in the western region where DSHP is

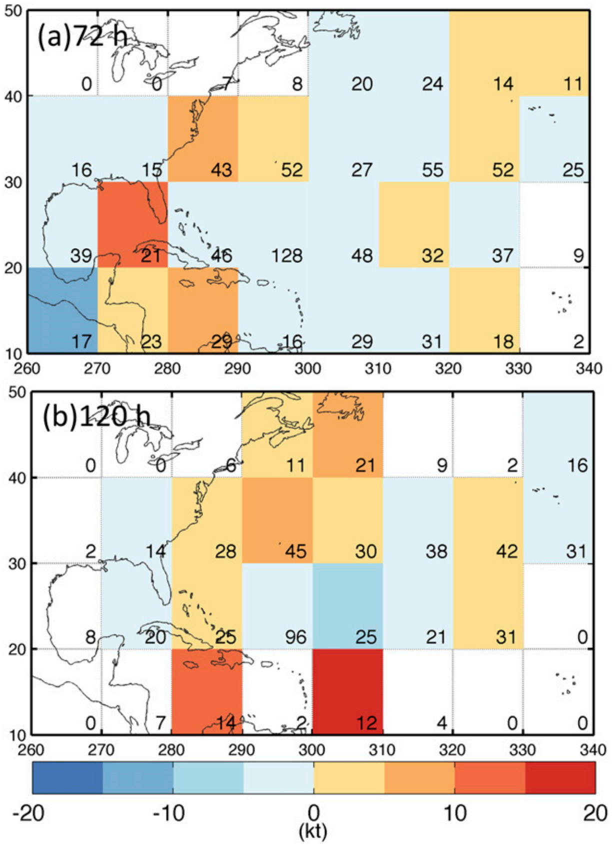

FIG. 9. As in Fig. 6, but for the WAIA technique relative to LGEM.

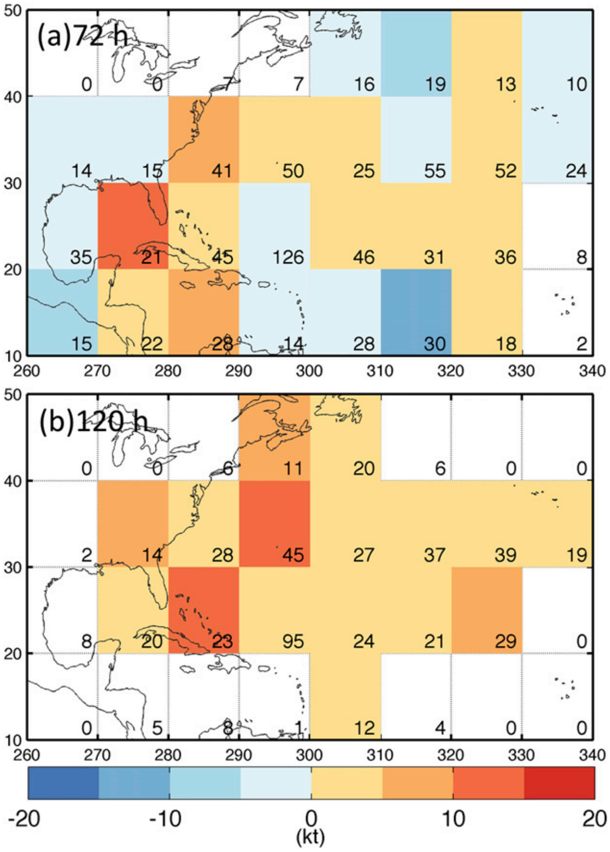

FIG. 10. As in Fig. 6, but for WAIA technique relative to GHMI.

more accurate by 0–5 kt. In contrast, LGEM has more boxes in the central North Atlantic in which it is more accurate than WAIA in the range of 0–5 kt. At 102 h (Fig. 9b), the geographic areas where LGEM is more (less) accurate relative to WAIA are quite similar to the areas where DSHP was more (less) accurate. As was the case at 72 h, WAIA has more regions in the western Atlantic and Caribbean where its performance is improved relative to LGEM than was the case for WAIA relative to DSHP. However, there is a region in the northeastern Atlantic in which the relative performance of LGEM at 120 h is better than it was for DSHP.

d. *Comparison of WAIA with GHMI*

The regional numerical model from the Geophysical Fluid Dynamics Laboratory (GFDL) that is here referred to as the GFDL hurricane model with a modified interpolator (GHMI) has provided track and intensity forecasts in the Atlantic since 1996 (DeMaria et al. 2014). The last letter ‘‘I’’ indicates the intensity is interpolated from the 6-h-old GFDL forecast by adding the difference between the current intensity and the 6-h forecast intensity through the entire 48-h forecast.

As indicated in Table 2 (row 5), GHMI has slightly smaller MAEs than WAIA at 24 and 48 h, but at longer forecast intervals WAIA is progressively more accurate than GHMI with statistically significant differences at 96–120 h. These performance characteristics are likely related to mean biases of GHMI relative to WAIA (Table 3, row 5). While GHMI has a smaller statistically significant mean bias at 24 h, it has progressively larger mean biases than WAIA from 72 through 120 h. Indeed, these overforecast biases for GHMI are among the largest for all of the intensity guidance products at NHC, and certainly degrade the longer-range performance of GHMI.

The geographical distributions of the WAIA intensity forecast improvements relative to GHMI (Fig. 10) depict the advantage of WAIA at longer ranges well. At 72 h (Fig. 10a), WAIA is more accurate by 10–15 kt in the eastern Gulf of Mexico and 5–10 kt more accurate in the eastern Caribbean and off the U.S. East Coast. In contrast, GHMI is more accurate than WAIA by 10–15 kt in the 10°–20°N, 40°–50°W box, and 5–10 kt more accurate in the 40°–50°N, 40°–50°W box. Over the remainder of the Atlantic domain, the improvement of WAIA relative to

GHMI at 72 h is in the range of 0–5 kt. By 120 h (Fig. 10b), WAIA has superior performance compared to GHMI throughout the domain with two boxes in which the differences are in the range of 10–15 kt. This advantage for WAIA at 120 h is highly likely due to the 10.91-kt overforecast bias for GHMI relative to WAIA (Table 3, row 5).

Even in view of these less desirable characteristics relative to WAIA, GHMI is one of the four guidance products included in ICON available at NHC.

e. Comparison of WAIA with HWFI

More recently, an intensity forecast from a regional, coupled hurricane–ocean numerical model called the Hurricane Weather Research and Forecasting Model (HWRF) has been provided by the National Centers for Environmental Prediction, and because of its skill it has also been included in ICON. Again, an interpolated version of HWRF (HWFI) must be used since the HWRF forecast is not received in time for consideration by the NHC forecasters prior to the warning release time. Even though HWFI is based on 6-h-old initial conditions, it has the best performance relative to WAIA at 24 and 48 h with statistically significant improvements of 1.35 and 1.68 kt, respectively (Table 2, row 6). As was the case for GHMI, this HWFI regional numerical model has larger statistically significant errors than WAIA at 96–120 h. The time evolution of the mean biases for HWFI is also similar to that of GHMI (Table 3, row 6 vs 5). That is, HWFI has a smaller statistically significant bias at 24 h, but then HWFI has increasingly large overforecast biases relative to the near-zero biases of WAIA from 72 through 120 h that are statistically significant.

The geographical distribution of where WAIA shows improved performance relative to HWFI (Fig. 11) is also similar to where WAIA has better performance than GHMI (Fig. 10). At 72 h (Fig. 11a), WAIA is more accurate than HWFI in the Gulf of Mexico, along the U.S. East Coast, and throughout the central Atlantic north of 20°N. In contrast, HWFI is more accurate south of 20°N except near Central America and in the eastern Atlantic. At 120 h (Fig. 11b), WAIA is more accurate than HWFI in all regions north of 30°N, with improvements of 10–15 kt in the 30°–40°N, 70°–50°W boxes. Whereas WAIA is more accurate at 120 h than GHMI throughout the Atlantic (Fig. 10b), there are four boxes between 10° and 20°N in which HWFI is more accurate than WAIA by 0–5 kt (Fig. 11b).

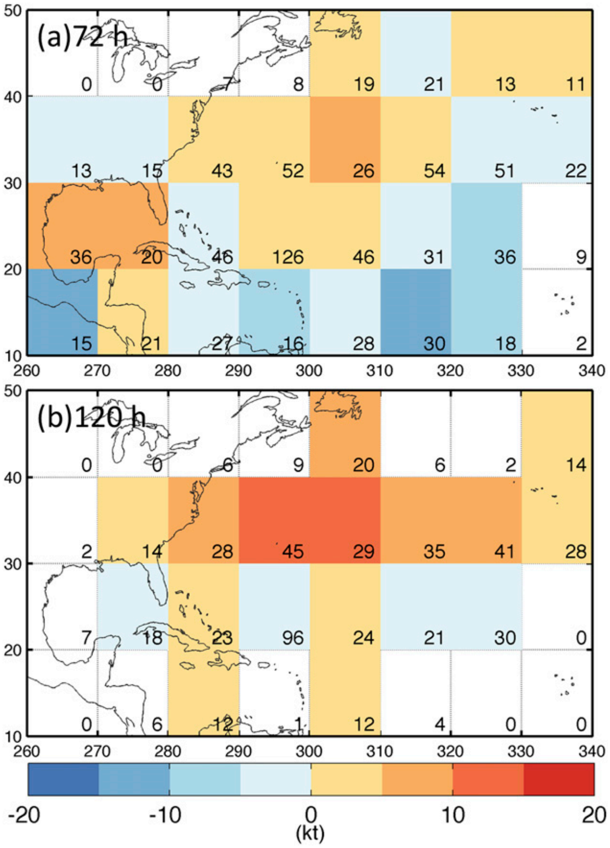

FIG. 11. As in Fig. 6, but for the WAIA technique relative to HWFI.

f. Comparison of WAIA with ICON

As noted above, ICON is a consensus of the two statistical–dynamic techniques (DSHP and LGEM) and two regional numerical models (GHMI and HWFI). Since DSHP and LGEM use the same predictors, their performance relative to WAIA is similar (sections 3b,c). Likewise, the performance characteristics of GHMI and HWFI relative to WAIA are similar (sections 3d,e). As expected for a consensus of skillful models, the performance of ICON is better than any of these four individual intensity guidance products. Relative to each of the homogeneous comparisons of WAIA with the MAEs summarized in Table 2, ICON is the most accurate with statistically significant improvements relative to WAIA from 24 through 96 h, and some small but not significant improvements at 108 and 120 h (Table 2, row 7). Indeed, the performance of ICON relative to WAIA is just slightly better than the improvement of the NHC official intensity forecasts relative to WAIA, since NHC also has statistically significant improvements relative to WAIA from 24 through 96 h (Table 2, row 8).

This improved performance of ICON relative to WAIA occurs even though ICON has larger statistically significant mean biases relative to the near-zero biases of WAIA from 72 through 120 h (Table 3, row 7). It is, of course, not surprising that ICON has large overforecast biases between 72 and 108 h when all four guidance products that are included in ICON have overforecast

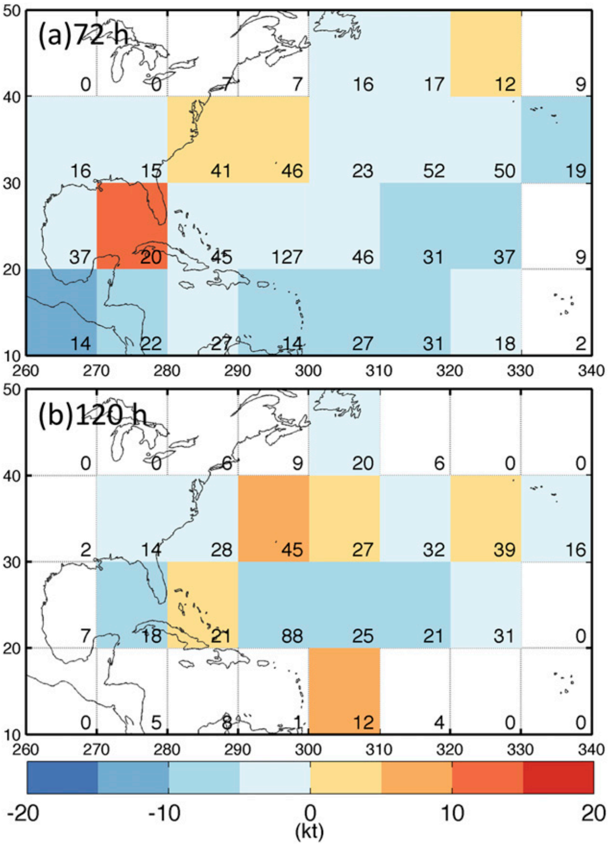

FIG. 12. As in Fig. 6, but for the WAIA technique relative to ICON.

biases (Table 3, rows 3–6). Notice that the NHC official intensity forecasts also have large statistically significant biases relative to WAIA (Table 3, row 8). Each individual NHC forecaster examines all available information and may give different weights to the guidance products (including ICON) based upon experience and past product performance. Since all of these four primary intensity guidance products have an overforecast bias, it is not surprising that the NHC official intensity forecasts also have an overforecast bias.

The geographical distribution of intensity forecast improvements of WAIA relative to ICON (Fig. 12a) at 72 h actually reveals the greater accuracy of ICON over WAIA for almost the entire Atlantic region. The only exceptions are in the eastern Gulf of Mexico where WAIA is more accurate by 15–20 kt, off the U.S. East Coast, and in the 40°–50°N, 40°–30°W box. The eastern Gulf of Mexico and the U.S. East Coast are areas where WAIA was more accurate than each of the four guidance products in ICON, so it is no surprise that WAIA is superior in those areas. However, ICON is more accurate at 72 h in the remainder of the Atlantic, and especially in the eastern Atlantic. At 120 h (Fig. 12b), WAIA is more accurate than ICON over more regions of the Atlantic. However, ICON is more accurate in the 20°–30°N band, including the eastern Gulf of Mexico where WAIA was more accurate at 72 h. As previously indicated, the storms in that area will make landfall somewhere around the Gulf of Mexico within 120 h, and the historical analog tracks and intensities used in WAIA may not be that reliable after landfall.

4. Summary and discussion

Following the development of our situation-dependent intensity and intensity spread prediction technique based on a weighted analog approach for western North Pacific tropical cyclones, a similar technique has been developed for Atlantic tropical cyclones. The basic premise is that the track and the current intensity are the primary determinants of the intensity on time scales of 3–5 days or longer, and the intensity spread among the 10 best historical analogs provides a measure of the range of environmental influences that may be occurring along the NHC official track forecast. These best historical analogs are selected from the NHC best tracks within ±30 days and are then ranked with equal weights given to the average track difference between the target storm and the candidate analog storm and the initial intensity difference. A second conditioning on initial intensities less than or greater than 35 kt demonstrates that the MAEs increase more slowly (rapidly) for the weaker (stronger) storms.

An essential step in the development of this technique was to devise an intensity bias correction at each 12-h forecast interval, which achieved a reduction in the overforecast bias of about 10 kt at 72 h and longer forecast intervals to a near-zero bias. This near-zero bias was shown in section 3 to be an advantage over all of the intensity guidance products available at NHC that have an overforecast bias.

The second essential step was to calibrate the original intensity spread to achieve a POD of about 68% at all forecast intervals, which was shown to be generally successful for an independent sample of 30% of the ~1300 cases. After the bias correction and calibration, the MAEs of WAIA at 72 (120) h were reduced by about 20%. Furthermore, the intensity spreads from WAIA were also reduced at all forecast intervals because the noncalibrated intensity spreads were overdetermined (>68%) for most of the forecast intervals for both the initial intensities less than or greater than the 35-kt subsamples. Although these intensity spreads were designed to be appropriate for the WAIA intensity forecasts, we propose that they would be a first-order intensity spread estimate for the NHC OFCL forecasts as the NHC track forecast is a primary driver for the WAIA intensity forecast. In conjunction with the Goerss and Sampson (2014) intensity forecast uncertainty estimates in terms of minimum and maximum confidence intervals, this would give NHC two methods to provide intensity uncertainty information that would be useful to forecasters, decision-makers, and informed members of the public.

Comparisons are made in section 3 with the various intensity guidance products available at NHC to illustrate some advantages of WAIA as an operational intensity forecast product for Atlantic tropical cyclones. First, WAIA has smaller statistically significant intensity forecast errors at 72–120 h relative to the SHF5 technique that has been used as an intensity skill measure. The WAIA 72-h intensity forecast improvements over SHF5 exist in almost all areas of the Atlantic, and some regions have sample-mean improvements of 15–20 kt. WAIA also has smaller statistically significant sample-mean 72–120-h MAEs than does the SHIPS product. Again, WAIA intensity forecast improvements of 15–20 kt at 72 h over the SHIPS forecasts exist in the Gulf of Mexico region where landfall is expected. DSHP, which accounts for the landfall effects along the NHC track forecast, eliminates the advantage that WAIA had over the SHIPS product. However, WAIA provides improved 120-h forecasts relative to DSHP over two-thirds of the Atlantic basin and, especially, for recurving storms north of 30°N. While LGEM uses the same predictors averaged over the prior 24 h as in DSHP, which averages these predictions over the full forecast, WAIA generally has the same improvements relative to LGEM as it did for DSHP. Even though DSHP and LGEM are not actually independent models, both are included in the ICON guidance product. Since WAIA essentially has equivalent skill as DSHP and LGEM, and is an independent model, it would be a candidate for inclusion in ICON.

The other two guidance products in ICON are the two regional numerical models: GHMI and HWFI. Both of these models have smaller MAEs than WAIA at 24 and 48 h, but have larger statistically significant MAEs than WAIA at 96–120 h. These two models also have quite large, and thus statistically significant, sample-mean biases relative to WAIA in the 72–120-h range. These large overforecast biases likely contribute to the less accurate GHMI and HWFI forecasts at 96–120 h. The geographic distribution of 120-h HWFI errors relative to WAIA clearly indicates that WAIA has superior performance for storms north of 30°N. In addition, WAIA is superior to GHMI at 120 h for the Atlantic basin.

The skill relative to WAIA of GHMI and HWFI at 24 and 48 h certainly contributes to statistically significant improvements of ICON relative to WAIA at those times. However, the poor performance of GHMI and HWFI relative to WAIA at 96–120 h likely degrades ICON's skill at later times so that by 108 and 120 h ICON is no longer better than WAIA at the 5% significance level. Thus, WAIA would again be a candidate for inclusion in ICON because of its superior intensity forecasts compared to GHNI and HWFI at longer forecast intervals. Indeed, this finding that WAIA has much smaller (near zero) intensity biases at longer forecast intervals could lead to improved NHC official intensity forecasts.

Acknowledgments. HCT and RLE are supported by the Office of Naval Research Marine Meteorology section. Mrs. Penny Jones provided excellent assistance in the manuscript preparation.

APPENDIX

Intensity Guidance Techniques

A summary of the acronyms and descriptions of intensity guidance techniques cited in this study are provided in Table A1 (see also DeMaria et al. 2014).

TABLE A1. Summary of acronyms and descriptions of intensity guidance techniques cited in this study.

Acronym	Expansion	Reference
DSHP	Decay–Statistical Hurricane Intensity Prediction Scheme	DeMaria et al. (2005)
GHMI	GFDL hurricane model with a modified interpolator	Bender et al. (2007)
HWFI	HWRF interpolated version	Gopalakrishnan et al. (2010)
ICON	Intensity consensus	Sampson et al. (2008)
IVCN	Intensity variable consensus	Goerss and Sampson (2014)
LGEM	Logistic Growth Equation Model	DeMaria (2009)
SHF5	SHIFOR 5-day version	Knaff et al. (2003)
SHIPS	Statistical Hurricane Intensity Prediction Scheme	DeMaria et al. (2005)
WAIA	Weighted Analog Intensity Atlantic	This study
WAIP	Weighted Analog Intensity Pacific	Tsai and Elsberry (2014)

REFERENCES

Bender, M. A., I. Ginis, R. Tuleya, B. Thomas, and T. Marchok, 2007: The operational GFDL coupled hurricane–ocean prediction system and a summary of its performance. *Mon. Wea. Rev.*, **135**, 3965–3989, doi:10.1175/2007MWR2032.1.

DeMaria, M., 2009: A simplified dynamical system for tropical cyclone intensity prediction. *Mon. Wea. Rev.*, **137**, 68–82, doi:10.1175/2008MWR2513.1.

——, M. Mainelli, L. K. Shay, J. A. Knaff, and J. Kaplan, 2005: Further improvements to the Statistical Hurricane Intensity Prediction Scheme (SHIPS). *Wea. Forecasting*, **20**, 531–543, doi:10.1175/WAF862.1.

——, C. R. Sampson, J. A. Knaff, and K. D. Musgrave, 2014: Is tropical cyclone intensity guidance improving? *Bull. Amer. Meteor. Soc.*, **95**, 387–398, doi:10.1175/BAMS-D-12-00240.1.

Elsberry, R. L., and H.-C. Tsai, 2014: Situation-dependent intensity skill metric and intensity spread guidance for western North Pacific tropical cyclones. *Asia-Pac. J. Atmos. Sci.*, **50**, 297–306, doi:10.1007/s13143-014-0018-5.

Goerss, J., and C. Sampson, 2014: Prediction of consensus tropical cyclone intensity forecast error. *Wea. Forecasting*, **29**, 750–762, doi:10.1175/WAF-D-13-00058.1.

Gopalakrishnan, S. G., Q. Liu, T. Marchok, D. Sheinin, N. Surgi, R. Tuleya, R. Yablonsky, and X. Zhang, 2010: Hurricane Weather and Research and Forecasting (HWRF) Model scientific documentation. Development Testbed Center, 75 pp. [Available online at http://www.dtcenter.org/HurrWRF/users/docs/scientific_documents/HWRF_final_2-2_cm.pdf.]

Knaff, J. A., M. DeMaria, C. R. Sampson, and J. M. Gross, 2003: Statistical, 5-day tropical cyclone intensity forecasts derived from climatology and persistence. *Wea. Forecasting*, **18**, 80–92, doi:10.1175/1520-0434(2003)018<0080:SDTCIF>2.0.CO;2.

Sampson, C. R., J. L. Franklin, J. A. Knaff, and M. DeMaria, 2008: Experiments with a simple tropical cyclone intensity consensus. *Wea. Forecasting*, **23**, 304–312, doi:10.1175/2007WAF2007028.1.

Tsai, H.-C., and R. L. Elsberry, 2014: Applications of situation-dependent intensity and intensity spread predictions based on a weighted analog technique. *Asia-Pac. J. Atmos. Sci.*, **50**, 507–518, doi:10.1007/s13143-014-0040-7.