A Description of the Real-Time HFIP Corrected Consensus Approach (HCCA) for Tropical Cyclone Track and Intensity Guidance

ANU SIMON

Cyberdata Technologies, Inc., Herndon, Virginia, and National Hurricane Center, Miami, Florida

ANDREW B. PENNY

Systems Research Group, Inc., Colorado Springs, Colorado, and National Hurricane Center, Miami, Florida

MARK DEMARIA, JAMES L. FRANKLIN, RICHARD J. PASCH,
EDWARD N. RAPPAPORT, AND DAVID A. ZELINSKY

National Hurricane Center, Miami, Florida

(Manuscript received 23 May 2017, in final form 31 October 2017)

ABSTRACT

This study discusses the development of the Hurricane Forecast Improvement Program (HFIP) Corrected Consensus Approach (HCCA) for tropical cyclone track and intensity forecasts. The HCCA technique relies on the forecasts of separate input models for both track and intensity and assigns unequal weighting coefficients based on a set of training forecasts. The HCCA track and intensity forecasts for 2015 were competitive with some of the best-performing operational guidance at the National Hurricane Center (NHC); HCCA was the most skillful model for Atlantic track forecasts through 48 h. Average track input model coefficients for the 2015 forecasts in both the Atlantic and eastern North Pacific basins were largest for the European Centre for Medium-Range Weather Forecasts (ECMWF) deterministic model and the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) ensemble mean, but the relative magnitudes of the intensity coefficients were more varied. Input model sensitivity experiments conducted using retrospective HCCA forecasts from 2011 to 2015 indicate that the ECMWF deterministic model had the largest positive impact on the skill of the HCCA track forecasts in both basins. The most important input models for HCCA intensity forecasts are the Hurricane Weather Research and Forecasting (HWRF) Model and the Coupled Ocean–Atmosphere Mesoscale Prediction System–Tropical Cyclone (COAMPS-TC) model initialized from the GFS. Several updates were incorporated into the HCCA formulation prior to the 2016 season. Verification results indicate HCCA continued to be a skillful model, especially for short-range (12–48 h) track forecasts in both basins.

1. Introduction

Initiated in 2009, the National Oceanic and Atmospheric Administration (NOAA)-funded Hurricane Forecast Improvement Program (HFIP) goals are to reduce the average track and intensity guidance forecast errors for tropical cyclones (TCs) by 20% in 5 years and by 50% in 10 years for 1–5-day forecasts (Gall et al. 2013; Gopalakrishnan et al. 2016). In general, both deterministic and ensemble numerical weather prediction (NWP) guidance for TCs have shown steady improvements during HFIP (Cangialosi and Franklin 2016). To extend these improvements in NWP skill,

multimodel consensus postprocessing techniques can also be applied.

Multimodel consensus forecast guidance is widely used in many operational weather forecasting centers. The National Hurricane Center (NHC) relies on various consensus aids to help improve TC track and intensity forecasts. Goerss (2000) found that a simple, equally weighted average of several models consistently outperforms each of the individual member models that make up the consensus. Bias correction can also be applied to remove the systematic error of each model as determined from past performance (e.g., Krishnamurti et al. 2010). One disadvantage of an equally weighted consensus is that poorly performing models receive the same weight as superior models. The multimodel

Corresponding author: Anu Simon, anu.simon@noaa.gov

DOI: 10.1175/WAF-D-17-0068.1

For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy \(www.ametsoc.org/PUBSReuseLicenses\)](http://www.ametsoc.org/PUBSReuseLicenses).

superensemble technique (Krishnamurti et al. 1999, 2000a,b, 2001; Krishnamurti 2003) addresses this shortcoming of the equally weighted consensus by applying unequal weights to bias-corrected member model output. The superensemble technique is now being widely used in various operational centers for forecasting precipitation (e.g., Cartwright and Krishnamurti 2007; Cane and Milelli 2010) and TC track and intensity (e.g., Williford et al. 2003; Krishnamurti et al. 2011).

This study focuses on the development and implementation of the HFIP Corrected Consensus Approach (HCCA) for TC track and intensity (maximum 10-m wind speed) guidance at NHC. As a Regional Specialized Meteorological Center (RSMC), NHC is responsible for issuing TC forecasts throughout the North Atlantic basin and for systems east of 140°W in the eastern North Pacific. Although HCCA is based on the Florida State Super-Ensemble (FSSE) methodology (Williford et al. 2003), which is also used for forecasting TC track and intensity, it differs in several ways: (i) HCCA uses different input models, (ii) input model weighting coefficients are derived from different sets of training forecasts, and (iii) beginning in 2016, modifications were made to the HCCA technique for computing medium- to long-range intensity forecasts (see section 6 for more details). Although FSSE forecasts are provided by WeatherPredict Consulting Inc. for operational use at NHC, HCCA was developed as an “in house” guidance product to address specific operational constraints at NHC and to provide the forecasters with greater flexibility and transparency in the choice of input models so that potential improvements to the technique could be developed, tested, and implemented. For example, while FSSE forecasts are at present only provided to NHC for systems of tropical storm strength or greater, HCCA forecasts are generated for systems below tropical storm strength, including “invests” that have yet to develop a well-defined low-level circulation and/or have a lack of organized deep convection.

The remainder of this paper is outlined as follows: Section 2 describes the methodology of HCCA, section 3 provides verification results for the 2015 Atlantic and eastern North Pacific seasons, section 4 examines the impact of the input models, section 5 assesses the performance of HCCA intensity forecasts during rapid intensification, section 6 outlines changes to the HCCA configuration for the 2016 season and provides preliminary verification of the 2016 HCCA forecasts, and section 7 provides a summary.

2. Methodology

As mentioned above, HCCA is similar to the superensemble technique used for the FSSE forecasts

(Williford et al. 2003). This methodology employs an “increment” approach such that statistics are derived from the 12-h change (increment) in the latitude, longitude, or intensity of several input models (Krishnamurti et al. 2011). For example, at a particular forecast hour t , the HCCA latitude is computed by expanding Eq. (1) of Krishnamurti et al. (2001) as

$$\begin{aligned} \text{HCCA}_t = & \text{HCCA}_{t-1} \\ & + \sum_{i=1}^N a_i \left[(F_{i,t} - F_{i,t-1}) - \frac{1}{K} \sum_{k=1}^K (T_{i,k,t} - T_{i,k,t-1}) \right] \\ & + \frac{1}{K} \sum_{k=1}^K (O_{k,t} - O_{k,t-1}), \end{aligned} \quad (1)$$

where HCCA_{t-1} is the HCCA forecast value at hour $t - 12$ for latitude, N is the number of input models, a_i is the weighting coefficient for model i , $(F_{i,t} - F_{i,t-1})$ is the current forecast increment value for model i , $(1/K) \sum_{k=1}^K (T_{i,k,t} - T_{i,k,t-1})$ is the mean increment value for model i from a set of K training set forecasts, and $(1/K) \sum_{k=1}^K (O_{k,t} - O_{k,t-1})$ is the mean increment of the verifying analysis (i.e., best track) corresponding to the same set of training forecasts. A similar equation is also solved for longitude and intensity.

The weighting coefficients are chosen to minimize the sum of the squared error over K training forecasts using multiple linear regression. Error is defined as the difference between the best-track increment and the sum of the second and third terms on the right-hand side of Eq. (1). Coefficients can be positive or negative and are not constrained to add up to one. Although the best-performing models generally receive the largest weights, negative coefficients do not necessarily indicate that a particular model is inferior to the other input models. Separate weighting coefficients are computed for the latitude, longitude, and intensity forecasts.

Model guidance at NHC is defined as either being “late” or “early,” depending on its availability in a timely manner for the operational forecast cycle (see NHC 2017). After initialization, dynamical models (both global and regional) require several hours of computation time and are generally not yet available to the forecasters before they must issue their forecasts 3 h after synoptic time. To mitigate this timeliness issue, the late model track and/or intensity forecasts from the previous forecast cycle are smoothed and shifted to match the current position and intensity of the tropical cyclone. These forecasts are referred to as “interpolated” forecasts, and model identifiers ending in “1” or “2” indicate the forecast was interpolated from the previous 6- or 12-h forecast, respectively.

Not surprisingly, an appropriate choice of input models is critical for producing skillful HCCA forecasts.

TABLE 1. List of track and intensity models referred to in this paper. The HCCA input models used to derive the track and intensity coefficients are indicated in the right-most column. Interpolation (h) refers to the operational timeliness of the models. For example, for AEMI/2, AEMI refers to the 6-h-old interpolated model, and AEM2 refers to the 12-h-old interpolated model.

Model identifier	Description	Interpolation (h)	HCCA input
AEMI/2	National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) ensemble mean	6/12	Track
AVNI/2	NCEP GFS deterministic model	6/12	Track/intensity
BEST	NHC postanalysis “best track”	—	—
CTCI/2	COAMPS-TC model with GFS initial and lateral boundary conditions	6/12	Intensity (starting in 2016)
DSHP	Decay Statistical Hurricane Intensity Prediction Scheme	—	Intensity
EGRI/2 ^a	Met Office deterministic model	6/12	Track
EMNI/2 ^a	ECMWF ensemble mean	12/18	Track (starting in 2016)
EMXI/2 ^a	ECMWF deterministic model	6/12	Track
GHMI/2	Geophysical Fluid Dynamics Laboratory (GFDL) Hurricane Model	6/12	Track/intensity
HWFI/2	Hurricane Weather Research and Forecasting Model (HWRF)	6/12	Track/intensity
IVCN	Intensity variable consensus (equally weighted; two-member min; CTCI/2 ^b + DSHP + GHMI/2 + HWFI/2 + LGEM)	—	—
LGEM	Statistical Hurricane Intensity Prediction Scheme Logistical Growth Equation Model	—	Intensity
OFCL	NHC official forecast	—	—
TVCN	Track variable consensus (equally weighted; two-member min; AVNI/2 + CTCI/2 ^b + EGRI/2 + EMXI/2 + GHMI/2 + HWFI/2)	—	—

^a The model is only initialized twice per day.

^b CTCI/2 was added to TVCN/IVCN beginning in 2016.

Input models for HCCA were selected from various dynamical/statistical models that are available in a timely manner for operations by assessing their impact on the skill of the HCCA forecasts. Initial consideration was given to the input models used by NHC operationally to form its equally weighted variable consensus products for track (TVCN) and intensity (IVCN). For the 2015 HCCA configuration, six input models were selected for track (latitude and longitude) and five input models were used for intensity. A complete description of the HCCA input models and other forecast guidance referred to in this paper is given in Table 1. At present, HCCA uses identical input models for the Atlantic and eastern North Pacific forecasts.

The proper selection of training forecasts that are used to derive the input model coefficients is also very important for successful HCCA forecasts. Training forecasts are kept separate for the Atlantic and eastern North Pacific basins. Retrospective forecasts using the most current configuration of the input models are used whenever possible for training forecasts (and for testing and evaluation purposes). However, because models are generally upgraded once per year, retrospective forecasts are generally limited in number (if available at all), so it is necessary to supplement the training set with forecasts from previous model configurations. For the 2015 and 2016 operational HCCA forecasts, forecasts were included in the training set that extended back to 2011. In addition, at longer lead times, it becomes increasingly likely that modeled storms will

dissipate, which decreases the number of input model forecasts available for the training set at longer forecast lead times.

Training set forecasts are only included if the initial and verifying classifications in the best track are tropical depression (TD), tropical storm (TS), subtropical depression (SD), subtropical storm (SS), or hurricane (HU); classifications of low (LO), extratropical (EX), disturbance (DB), and wave (WV) are excluded. Training forecasts for the eastern North Pacific basin include all forecasts from storms originating from that basin, even if the storm subsequently moves west of 140°W and out of NHC’s area of responsibility.

An important aspect of the training set is that its length is not fixed during the season; following Krishnamurti et al. (2011), after a storm dissipates and if no other storms are present in the basin, forecasts from that storm are added to the training set. Updating the training set with forecasts from the current season helps to take into consideration the performance of the input member models during the season.

To produce an operational HCCA forecast, a forecast-specific training set is created in real time based on forecast model availability; cases are selected from the training set that match the set of input models and forecast hours (i.e., training cases are excluded if one of the input models for the current forecast is not present in the training case). At present, training cases are not stratified by location within a basin, or by time of year, etc. In the event that only a 12-h interpolated forecast is

FIG. 1. Track skill (%) relative to climatology and persistence (see text) for the 2015 operational track forecasts of HCCA, HCCA track input models, NHC OFCL, and TVCN for the (a) Atlantic and (b) eastern North Pacific basins.

available operationally, the training set is composed of a combination of 6- and 12-h interpolated forecasts. Since there are only a limited number of 12-h interpolated training forecasts from models that are initialized four times per day, using a blend of 6- and 12-h interpolated forecasts for the training ensures there will be a sufficient number of training cases present when only a 12-h interpolated forecast is available in operations. For the input track models EGRI/EMXI (see Table 1) that are only initialized twice per day, the number of 12-h interpolated forecasts (EGR2/EMX2) available in the training set is roughly half the total number of forecasts.

During initial testing, it was found that a minimum of three input models is needed to ensure a stable HCCA forecast; if fewer than three input models are present, a HCCA forecast will not be computed. A HCCA intensity forecast is not generated if a track forecast cannot be computed. A track forecast will be generated, however, even if the intensity forecast cannot be computed; HCCA intensity forecasts are truncated once the computed intensity is less than 15 kt (where $1 \text{ kt} = 0.51 \text{ ms}^{-1}$).

3. Operational HCCA forecasts for 2015

During the 2015 hurricane season, HCCA was run in real time quasi operationally at NHC. Track and intensity verification statistics are provided in Figs. 1 and 2 for a homogeneous sample of 2015 forecasts from HCCA, the HCCA input models, and the two-member minimum equally weighted variable consensus for track (TVCN) and intensity (IVCN). For 2015, TVCN was composed of AVNI, EGRI, EMXI, GHMI, and HWFI, and IVCN consisted of DSHP, GHMI, HWFI, and LGEM [see Table 1 and NHC (2017) for a more complete description of the model definitions]. The NHC official forecasts (OFCL), which are issued in real time by NHC forecasters and are based on a broad array of numerical guidance and observational indicators, are also included in the verification to help assess model performance. Verification was computed using the Model Evaluation Tools Tropical Cyclone (MET-TC) verification software provided by the Development Testbed Center (2015) and is shown in terms of skill relative to the climatology and persistence statistical models

FIG. 2. Intensity skill (%) relative to climatology and persistence (see text) for the 2015 operational intensity forecasts of HCCA, HCCA intensity input models, NHC OFCL, and TVCN for the (a) Atlantic and (b) eastern North Pacific basins.

(CLIPER5; Neumann 1972, Aberson 1998) for track forecasts, and Decay-SHIFOR5, which is a version of the Statistical Hurricane Intensity Forecast model (Knaff et al. 2003), for intensity forecasts. For reference, track and intensity error values are also shown in Tables 2 and 3 from a separate homogeneous verification that only includes HCCA, OFCL, and TVCN/IVCN.

Similar to the procedures used for NHC postseason verification reports (e.g., Cangialosi and Franklin 2016), only forecasts with an initial and verifying classification of TD, SD, TS, SS, or HU are included in the verification. In addition, no differentiation is made between 6- and 12-h interpolated forecasts (i.e., the EMXI and EMX2 forecasts are combined in the verification).

TABLE 2. Track error (n mi) for the 2015 Atlantic and eastern North Pacific seasons from a homogeneous sample of NHC OFCL, TVCN, and HCCA forecasts. The lowest errors at each forecast hour are set in boldface, and N indicates the number of homogeneous verifying forecasts included in the comparison.

Forecast hour	2015 Atlantic track error (n mi)				2015 eastern North Pacific track error (n mi)			
	N	OFCL	TVCN	HCCA	N	OFCL	TVCN	HCCA
12	202	25.4	25.6	23.8	331	26.0	25.3	24.6
24	182	41.1	40.7	37.1^a	300	38.7	38.3	37.8
36	159	57.1	55.8	49.3^b	272	49.3	48.8	49.0
48	136	77.4	75.8	66.3^a	246	59.4	59.3	60.6
72	96	125.2	126.9	116.6	200	83.2	81.5	87.7
96	67	188.0	203.7	175.3	158	115.1	117.0	119.5
120	49	241.4	270.8	196.6^b	127	153.8	160.3	162.8

^a The difference between HCCA and the equally weighted consensus is significant at the 90% level using a two-sided Student's t test.

^b The difference between HCCA and the equally weighted consensus is significant at the 95% level using a two-sided Student's t test.

TABLE 3. As in Table 2, but for the intensity error (kt).

Forecast hour	2015 Atlantic intensity error (kt)				2015 eastern North Pacific intensity error (kt)			
	<i>N</i>	OFCL	IVCN	HCCA	<i>N</i>	OFCL	IVCN	HCCA
12	202	5.0	5.8	6.0	331	7.5	8.3	7.9
24	182	7.4	8.2	8.4	300	12.0	13.1	11.6^a
36	159	9.9	10.0	10.3	271	14.8	16.1	14.5
48	136	11.4	11.4	11.8	244	15.9	17.8	16.1
72	96	15.6	14.7	13.6	197	16.4	17.8	16.0
96	67	19.6	20.0	16.0^a	156	15.1	16.2	16.2
120	49	20.4	27.1	16.8^b	124	15.4	16.2	16.5

^a The difference between HCCA and the equally weighted consensus is significant at the 90% level using a two-sided Student's *t* test.

^b The difference between HCCA and the equally weighted consensus is significant at the 95% level using a two-sided Student's *t* test.

Only eastern North Pacific forecasts with an initial (0-h forecast) position east of 140°W are included in the verification.

a. Track verification for 2015 forecasts

For the 2015 Atlantic track forecasts (Fig. 1a), HCCA had the most skillful forecasts from 12 to 48 h. Beyond 48 h, EMXI was most skillful. The forecasts of GHMI and HWFI were the least skillful, while the performance of AEMI, AVNI, and EGRI was slightly better than GHMI and HWFI. The TVCN forecasts were slightly less skillful than OFCL following 72 h. At longer lead times (following 72 h), the track error of the GHMI and HWFI forecasts was 250 n mi (where 1 n mi = 1.852 km) greater than the EMXI error (not shown). Much of this error can be attributed to the poor track forecasts by GHMI/HWFI for Hurricane Joaquin (2015), which composed a large percentage of the verifying forecasts at longer lead times (roughly 35% of the 120-h forecasts) for the 2015 Atlantic season.

It is worth noting that while the increase in skill over OFCL and TVCN for the 2015 HCCA Atlantic track forecasts is significant (Table 2), the improvement relative to the OFCL forecasts is generally less than that realized by the initial real-time multimodel super-ensemble forecasts of Williford et al. (2003) for the 1999 Atlantic season. This is undoubtedly a result of improvements to the forecast guidance suite and is due to the fact that consensus aids have become increasingly important in the forecast process since then.

The eastern North Pacific basin was far more active than the Atlantic in 2015; there were more than double the number of cases from 72 to 120 h. Perhaps partly because of the larger number of cases, the performance of the eastern North Pacific forecasts is much more tightly clustered for track (Fig. 1b). In general, the forecasts of OFCL and TVCN were the most skillful. The HCCA skill was most similar to that of OFCL and TVCN through 48 h and performed slightly better than

EMXI, except at 120 h. The skill levels of the AVNI, HWFI, and AEMI forecasts were similar and lagged behind the skill of the EMXI, OFCL, TVCN, and HCCA forecasts through 96 h. At 120 h, AEMI was the most skillful of these three, followed by AVNI. At all lead times, GHMI had the worst-performing track forecasts; the 120-h track error of GHMI was close to 260 n mi, which is 100 n mi more than the error of AEMI and EMXI (not shown), which were the best-performing HCCA input models.

b. Intensity verification for 2015 forecasts

For the 2015 Atlantic intensity forecasts (Fig. 2a), the best-performing forecasts were those of the OFCL (12–36 h), HWFI (48–72 h), and HCCA (following 72 h). The skill of the DSHP, LGEM, and IVCN forecasts was very similar at all lead times. The performance of the AVNI and GHMI forecasts was clearly inferior, except the skill of the AVNI forecasts increased at longer lead times such that it had the third-best intensity skill at 120 h, behind the HCCA and OFCL forecasts. Following 72 h, the skill of GHMI was considerably less than the other input models used for HCCA intensity forecasts; the skill of the 120-h intensity forecasts of GHMI is close to –35% relative to persistence and climatology and was likely responsible for the decline in the skill of the IVCN at 120 h.

For the eastern North Pacific (Fig. 2b), the differences between the skill of the intensity forecasts is far more consistent with forecast hour compared to the Atlantic basin (Fig. 2a). Similar to the track forecasts, there was more than twice the number of cases compared to the Atlantic following 48 h. Perhaps because of the large number of cases that experienced rapid intensification, the intensity error for the eastern North Pacific grew much more quickly with forecast hour prior to 48 h compared to the Atlantic (Table 3). The results of OFCL and HCCA were very similar throughout all forecast lead times. The OFCL had the most skillful

FIG. 3. Average HCCA input model coefficients for (a) latitude, (b) longitude, and (c) intensity as a function of forecast hour for the 2015 Atlantic season.

forecasts at 12, 48, and from 96 to 120 h (Table 3), while HCCA was most skillful at 24, 36, and 72 h, outperforming IVCN by a considerable margin from 12 to 72 h. The intensity forecasts of GHMI and AVNI were the least skillful for the eastern North Pacific. Despite the relatively poor performance of GHMI, IVCN performed quite well, having the third-highest skill through 72 h and second-highest skill at 96 and 120 h.

c. Average HCCA input model coefficients for 2015 forecasts

The contributions that each of the input models made to the 2015 operational HCCA forecasts are examined by comparing the average latitude, longitude,

and intensity input model coefficients for the Atlantic (Fig. 3) and eastern North Pacific (Fig. 4) as a function of forecast hour.

The EMXI Atlantic latitude coefficients (Fig. 3a) are the largest coefficients between 12 and 60 h, and from 96 to 120 h. While the AEMI coefficients are relatively small at early forecast hours, they increase following 36 h and are the largest coefficients at 72 and 84 h. On the other hand, the AVNI latitude coefficients are initially larger than those for AEMI, but decrease with forecast hour and become negative at longer lead times. Negative coefficients mean that the portion of the total HCCA increment that results from AVNI is of opposite sign to the actual AVNI increment

FIG. 4. As in Fig. 3, but for the eastern North Pacific.

(e.g., a northward AVNI track between 84 and 96 h would result in a southward increment applied to HCCA). The latitude coefficients for the other track models (GHMI, EGRI, and HWFI) were relatively small at all forecast hours.

The latitude coefficients for the eastern North Pacific (Fig. 4a) share similar characteristics with those from the Atlantic (Fig. 3a). The EMXI latitude coefficients are the largest until 96 h, while the AEMI coefficients are the largest at 108 and 120 h. Whereas the Atlantic EMXI and AEMI latitude coefficients (Fig. 3a) are considerably larger than those of the other input models at longer lead times, except for the AVNI coefficients, there is less difference between the eastern North Pacific latitude coefficients.

The longitude input model coefficients for the Atlantic (Fig. 3b) are quite similar at early forecast hours between models. However, at longer lead times, the longitude coefficients of EMXI and AEMI are quite large (~ 0.9), while the longitude coefficients of other track models are near or below zero, indicating that the majority of the HCCA increment is made up of the AEMI and EMXI increments at longer lead times.

Apart from an increase in the AEMI coefficients and a decline in the AVNI and GHMI coefficients at longer lead times, the longitude coefficients for the eastern North Pacific (Fig. 4b) remain fairly constant at all forecast hours. The EMXI longitude coefficients are the largest by a considerable margin through 96 h, and the AEMI longitude coefficients are largest at 108 and 120 h.

Intensity coefficients for the Atlantic (Fig. 3c) and eastern North Pacific (Fig. 4c) HCCA forecasts are far more variable in comparison to the latitude and longitude coefficients, since no model has coefficients that are consistently larger than those of the other models at all forecast hours. For the Atlantic (Fig. 3c), the largest intensity coefficients generally correspond to those from GHMI, AVNI, LGEM, and HWFI. The DSHP coefficients are noticeably smaller than those of the other models and are negative at several forecast hours. There is a bit more spread in the values of the intensity coefficients for the eastern North Pacific forecasts (Fig. 4c). Between 36 and 72 h, the intensity coefficients for HWFI are largest (~ 0.6). At longer lead times, the AVNI, LGEM, and DSHP coefficients are largest, as the HWFI intensity coefficients decrease following 48 h.

An interesting characteristic of the coefficients is that several sets of input model coefficients appear to have an inverse relationship with forecast hour. The increase (decrease) in AEMI (AVNI) track coefficients (Figs. 3a,b and 4a,b) with forecast hour is a prominent feature. While there might be reason to believe this behavior is due to the relative performance of these models (e.g., track errors for the AVNI are generally less than for AEMI at shorter lead times, while AEMI tends to outperform AVNI at longer lead times), in fact this behavior indicates that the error of these input models is highly correlated. When derived individually, the weighting coefficients for all of the input models are positive at all forecast hours (not shown), but when the coefficients are computed using multiple linear regression, the coefficients of correlated input models tend to offset each other somewhat, such that one coefficient is positive and the other is negative. In this way, the technique uses the differences between the two highly correlated models as a predictor. A similar pattern is evident when comparing the LGEM and DSHP intensity coefficients (Figs. 3c and 4c). The intensity coefficients of DSHP are negative between 60 and 84 h, while the LGEM coefficients are relatively large and positive during this period. DSHP and LGEM are statistical intensity models that share many characteristics and, along with many of the HCCA input models (i.e., HWFI, GHMI, and CTCI), are based on, or rely on, the GFS model to some extent.

It is worth mentioning that, although the coefficients for a particular model may be small, at times its forecast increment may be considerably larger than those of the other input models. This can result in a large percentage of the total HCCA increment being derived from a model with a small coefficient. Determining a method to account for the size of the current input forecast increments is an area of future research.

FIG. 5. Track forecasts for Hurricane Joaquin (2015) initialized at 1200 UTC 30 Sep for HCCA, HCCA track input models, HCCA computed without EMX2 as an input model (HEMX; red dashed line), NHC OFCL, and TVCN. Also shown are the verifying NHC best-track positions (BEST; solid black line). All colored markers indicate track positions at 12-h intervals, except for the 48–120-h OFCL forecast positions, which are shown at 24-h intervals.

d. Example of an operational HCCA track forecast for Hurricane Joaquin (2015)

To illustrate the impact of the unequal weighting coefficients, a 5-day track forecast for Hurricane Joaquin initialized at 1200 UTC 30 September 2015 is shown in Fig. 5. The track forecast of EMX2 (12 h interpolated; pink dashed line) lies closest to the verifying NHC best-track positions, while the rest of the HCCA track input models (AEMI, AVNI, EGR2, GHMI, and HWFI) forecast Joaquin to move northward and make landfall along the U.S. East Coast. An additional HCCA forecast without EMX2 as an input model (HEMX; dashed red line) was also computed, and the resulting track is even farther west than the equally weighted variable consensus (TVCN; brown line). This clearly demonstrates the positive impact that EMX2 had on the HCCA forecast for this case. Since the input track model coefficients of EMX2 were among the largest at all lead times for this forecast (not shown), the EMX2 track forecast heavily affects the HCCA track, initially bringing it farther south into the Bahamas and shifting

FIG. 6. Atlantic track skill relative to the 2016 HCCA configuration for (a) retrospective HCCA forecasts from 2011–15 that had an input member model excluded (see legend) and (b) individual input model forecasts (retrospective forecasts when available). To remove extraneous HCCA forecasts from (a) that were not impacted by the excluded input model, forecasts were only included in the sample if the input member model was present. (c) The number of forecasts in each of the nonhomogeneous samples.

the long-range track well east of the other input guidance. Overall, EMXI/2 was the best-performing track model for Hurricane Joaquin (Berg 2016), and because of the unequal weighting coefficients, HCCA outperformed TVCN (not shown), having the least error during early forecast hours and only trailing behind the performance of EMXI/2 beyond 48 h.

4. Input model sensitivity experiments

To assess the relative impacts of the individual input member models, additional sets of retrospective HCCA forecasts from 2011 to 2015 were computed while sequentially excluding each of the input models (Figs. 6–10). Input model exclusion tests are also performed during the hurricane season to evaluate the in-season model impacts on HCCA. Retrospective forecasts from the input models are used whenever possible. When retrospective forecasts are not available, forecasts from the most-recent model configuration are used. A “cross validation” technique is used for retrospective forecasts such that all training forecasts corresponding to the current storm are withheld from the training set. The European Centre for

Medium-Range Weather Forecasts (ECMWF) ensemble mean (EMNI) and the Coupled Ocean–Atmosphere Mesoscale Prediction System–Tropical Cyclone (COAMPS-TC) model (CTCI) are included in the comparison for track and intensity forecasts, respectively, since these models were added as input models prior to the 2016 season (see section 6). The solid lines in Figs. 6a, 7a, 9a, and 10a indicate the skill of the HCCA forecasts when a particular model is excluded as an input model. To remove the influence of extraneous forecasts, HCCA forecasts are only included in the verification when the input model being examined was available. The skill is measured relative to the HCCA configuration using all available input models. For example, the pink solid line in Fig. 6a (“no EMXI”) displays the track skill of HCCA forecasts computed without EMXI. The 6- and 12-h interpolated forecasts are combined for the verification of the input model sensitivity experiments. The numbers of forecasts included in the verification for the different input models are indicated in Figs. 6c, 7c, 9c, and 10c. For these nonhomogeneous comparisons, negative (positive) skill indicates that excluding a particular model degrades (improves) the HCCA forecasts.

FIG. 7. As in Fig. 6, but for eastern North Pacific track skill.

As is evident from Fig. 6a, the skill of the Atlantic HCCA track forecasts is degraded drastically when EMXI is excluded. The skill drops from about -3% at 24 h to less than -10% at 120 h without EMXI. The change in skill for all other input models is within a range from -3% to $+3\%$, which indicates that EMXI is the single most important input model for Atlantic HCCA track forecasts. Figure 6a also indicates that HWFI has a positive impact on HCCA track forecasts at early forecast hours, but has a negative impact at longer lead times. Including AVNI and AEMI slightly improves the 12–120-h track skill, while EGRI slightly degrades the Atlantic HCCA track forecasts from 36 to 120 h.

Figures 6b, 7b, 9b, and 10b display the skill of the individual input models (dashed colored lines) relative to HCCA forecasts using the 2016 configuration. For the Atlantic track forecasts (Fig. 6b), all of the individual input models have negative skill compared to HCCA. As expected, the three best models are AVNI, AEMI, and EMXI, while GHMI and EGRI have the lowest skill relative to HCCA.

The input model sensitivities for eastern North Pacific forecasts (Fig. 7a) indicate that, not surprisingly, the impacts differ compared to the Atlantic track forecasts (Fig. 6a). Whereas the change in skill for eastern North Pacific HCCA track forecasts (Fig. 7a) ranges from about -4% to $+1\%$, the change in skill for the Atlantic

forecasts (Fig. 6a) is much greater (from -10% to $+3\%$). This indicates that there is more variability in the impact of the input models for the Atlantic forecasts compared to the eastern North Pacific forecasts. While the impact of EMXI is positive for track forecasts in both basins, the relative importance of EGRI (Fig. 7a) is markedly different. Unlike for the Atlantic forecasts, the inclusion of EGRI in the eastern North Pacific track forecasts has a positive impact from 24 to 120 h. In fact, the drop in skill for 72-h forecasts that exclude EGRI is similar in magnitude to that when EMXI is excluded. This difference between basins is interesting, especially considering that the performance of EGRI is relatively poor for both the Atlantic (Fig. 6b) and eastern North Pacific (Fig. 7b) forecasts in this sample. An examination of the 48-, 72-, and 120-h track biases for HCCA and the HCCA input models (Fig. 8) reveals the biases for the eastern North Pacific forecasts (Fig. 8b) are more consistent with forecast hour compared to the Atlantic forecasts (Fig. 8a), which may suggest that a greater percentage of the EGRI error is systematic and correctable for bias. However, the positive impact of EGRI in the eastern North Pacific despite its lack of skill could also be the result of a greater amount of random error being canceled out from the other input model forecasts. In contrast to EGRI, AEMI degrades the eastern North Pacific HCCA track forecasts (Fig. 7a) after 48 h but improves the Atlantic forecasts (Fig. 6a).

FIG. 8. Track bias shown as a direction ($^{\circ}$ from north) and distance (n mi) for the 48-, 72-, and 120-h HCCA and HCCA input model forecasts from 2011 to 2015 for the (a) Atlantic and (b) eastern North Pacific basins. Marker shapes differentiate forecast hours and colors correspond to different models (see legends). The number of forecasts included in the homogeneous sample is listed in the forecast hour legend. Note that the distance scales in (a) and (b) are different.

FIG. 9. As in Fig. 6, but for Atlantic intensity skill.

Results for the input model sensitivity experiments for Atlantic and eastern North Pacific intensity forecasts are shown in Figs. 9 and 10, respectively. Excluding HWFI degrades the Atlantic HCCA intensity forecasts (Fig. 9a) throughout the forecast period, and excluding CTCI degrades the forecasts from 72 h onward. This indicates that HWFI and CTCI are the most important input models for the Atlantic HCCA intensity forecasts, while the other intensity input models slightly degrade the forecasts. It should be noted that, despite these degradations, including these models provides greater operational stability, especially when one of the high-impact model forecasts is not available. The skill of the individual input model intensity forecasts relative to the 2016 HCCA configuration (Fig. 9b) matches the exclusion pattern (Fig. 9a), as the HWFI and CTCI forecasts have the greatest skill. Similar to the Atlantic forecasts, excluding HWFI degrades the HCCA intensity skill for the eastern North Pacific forecasts at all forecast hours (Fig. 10a). Excluding CTCI only degrades the skill following 96 h. In general, the relative contributions of the intensity input models are similar for both basins (Figs. 9a and 10a), which is different from the sensitivity experiments involving track skill (Figs. 6a and 7a). This agrees with the fact that there is less variability in the individual performances of the intensity input models between basins (Figs. 9b and 10b).

5. Rapid intensification

One of the most challenging aspects of intensity forecasting is to reliably forecast rapid intensification (RI), which has been defined as a 30-kt increase over a 24-h period (Kaplan and DeMaria 2003). The intensity error for forecasts encompassing RI events is generally quite large. In fact, even having the ability to forecast large intensity changes can negatively affect the overall verification statistics of a model, as the mean absolute error for RI forecasts that do not verify in reality can be large.

The ability of HCCA to produce a forecast outside of the input models' forecast envelope appears to give it an advantage for capturing RI forecasts over the equally weighted variable consensus (IVCN). For example, the intensity forecast for Hurricane Linda (2015) initialized at 0000 UTC 6 September (Fig. 11) reveals that the HCCA forecast values quickly increase and remain well above the other input model values from 12 to 84 h, although the results are still far below the verifying intensity according to the NHC best track. This characteristic of the HCCA forecasts appears to be a result of applying a bias correction to the intensity trend and using unequal weighting coefficients with the increment technique. To examine this further, the intensity forecast of the bias-corrected ensemble mean (BCEM) is

FIG. 10. As in Fig. 6, but for eastern North Pacific intensity skill.

also shown in Fig. 11. Except for the use of equal weighting coefficients, the BCEM is similar to HCCA; the input model forecasts are bias corrected for the intensity trend (increment), and the BCEM forecast values are computed by adding the mean of the input model increments to the previous 12-h BCEM forecast value. The fact that the BCEM forecast (Fig. 11) is closer to the middle of the guidance envelope suggests that, at least for this forecast, using unequal weighting

coefficients with the increment technique (i.e., the current forecast value depends on the forecast value at $t - 12$) is primarily responsible for the HCCA intensity forecast being outside the range of the input model forecasts.

To assess the skill of the RI forecasts from HCCA, the HCCA intensity input models, and OFCL, intensity error statistics were compiled for retrospective forecasts from 2011 to 2015 in which at least a 20-kt intensification

FIG. 11. Intensity forecasts for Hurricane Linda (2015) initialized at 0000 UTC 6 Sep of HCCA, HCCA intensity input models, BCEM, NHC OFCL, and IVCN. Also shown is the verifying NHC best-track intensity (BEST; solid black line).

FIG. 12. Intensity error (kt) for retrospective forecasts from 2011 to 2015 of HCCA, HCCA intensity input models, and NHC OFCL for cases in which an intensification of 20 kt or greater occurred over 24 h for the (a) Atlantic and (b) eastern North Pacific basins.

occurred over 24 h at any time during a forecast. Relaxing the strict definition of RI to include events of 20 kt or greater in 24 h increased the sample size considerably. Based on a sample from 1982 to 2016, a 20-kt increase in intensity over 24 h corresponds to the top 17.5% of Atlantic cases and the top 18.5% of eastern North Pacific cases. Only forecast lead times within the 24-h period are included in the verification. For the Atlantic basin (Fig. 12a), the HCCA intensity error remains relatively small through 72 h, and the HCCA error is the smallest at 48 h. Following 72 h, however, the HCCA intensity error increases, having the second largest intensity error at 120 h.

In contrast to the Atlantic forecasts, HCCA has the smallest intensity error for RI cases in the eastern North Pacific at almost all forecast hours (Fig. 12b) and by a wide margin between 36 and 72 h. The difference in the performance of HCCA between basins suggests there may be greater predictability for these events in the eastern North Pacific. The intensity errors might be

more systematic in the eastern North Pacific and, as a result, more effectively corrected for bias. In addition, there were far more RI events in the eastern North Pacific during the 2011–15 period, especially at shorter lead times; a larger number of training forecasts with similar intensification characteristics should presumably lead to an increase in skill for a statistical technique such as HCCA.

It is also worth examining the frequency at which RI events are forecast by HCCA, the HCCA intensity input models, and OFCL. Figure 13 compares the number of intensity forecasts with at least a 20-kt increase in intensity over 24 h for the Atlantic (Fig. 13a) and eastern North Pacific (Fig. 13b) for verifying forecasts. Not surprisingly, the number of 20-kt intensification events decreases dramatically with increasing lead time for both basins. For the Atlantic (Fig. 13a), intensification events of 20 kt or greater in 24 h are forecast by HWFI most frequently at just about all forecast lead times, which agrees with the findings of Kaplan et al. (2015).

FIG. 13. The number of cases from 2011 to 2015 in which an intensification of 20 kt or greater in 24 h was forecast by HCCA, the HCCA input intensity models, and the NHC OFCL for forecast hours 24–120 h for the (a) Atlantic and (b) eastern North Pacific. Note that the scales for the ordinate axes in (a) and (b) are different.

Except for a few forecast hours, the dynamical models tend to forecast RI events in the Atlantic more frequently than HCCA and the statistical models.

In contrast, DSHP and HCCA, both of which are statistical models, tend to forecast intensification events of 20 kt or greater in 24 h far more frequently than the dynamical models in the eastern North Pacific (Fig. 13b). This suggests the errors for such events are more systematic in the eastern North Pacific and that the environmental signals that DSHP relies on to forecast intensification are clearer. Kaplan et al. (2015) found greater predictability for RI cases in the eastern North Pacific compared to the Atlantic. The fact that OFCL forecasts such events more frequently than the intensity guidance (apart from HCCA, which was only available quasi operationally beginning in 2015) suggests that forecasters are confident in forecasting a significant intensification of a system even when the suite of intensity guidance tends to be more conservative.

6. Operational HCCA forecasts for 2016

Based upon the 2015 results and an analysis of model performance during the 2015 season, additional testing indicated that including EMNI as an input model for HCCA track forecasts yielded a slight improvement for both Atlantic and eastern North Pacific forecasts (see Figs. 6a and 7a). Therefore, EMNI was added as a track input model. Because of latency issues with EMNI, only the 12-h (EMN2) and 18-h (EMN3) interpolated forecasts are available operationally. For the purpose of this manuscript, the 12- and 18-h interpolated EMNI forecasts are referred to as EMNI and EMN2, respectively. In addition, the HCCA input model configuration for intensity was updated to include CTCI, which uses GFS initial and lateral boundary conditions. Aside from a slight degradation in the medium-range skill for eastern North Pacific forecasts, including CTCI in the HCCA intensity forecasts made a positive impact, especially at longer lead times (see Figs. 9a and 10a).

FIG. 14. Intensity skill of the increment–value blend technique (blue lines) relative to the increment technique (red lines) for retrospective HCCA forecasts from 2011 to 2015 for the (a) Atlantic and (b) eastern North Pacific.

Another change made prior to the 2016 season was the implementation of a slightly different technique for medium- and long-range intensity forecasts. As previously discussed, HCCA relies on an increment approach, where model coefficients are derived by separately minimizing the squared error in the 12-h change (increment) in latitude, longitude, and intensity. While this approach generally gives the best results for HCCA track forecasts, minimizing the squared error of the intensity values, referred to here as the “value” approach, sometimes outperforms the increment approach for intensity forecasts at longer lead times. After testing, it was found that using the increment approach from 12 to 36 h and an average of the increment and value forecasts (hereafter “increment–value blend”) from 48 to 120 h provided the best results for the HCCA intensity forecasts.

Based on retrospective HCCA intensity forecasts from 2011 to 2015 (Fig. 14), using the increment–value blend approach (blue lines) dramatically improves the

skill over the increment approach (red lines), especially at longer lead times for both Atlantic (Fig. 14a) and eastern North Pacific (Fig. 14b) forecasts. While these results suggest that correcting for biases in the intensity change and incorporating the persistence that is inherent with the increment technique is best for early forecast hours (i.e., a forecast for a particular lead time relies on the previous 12-h value), it is less optimal at longer lead times.

Verification results for track (Fig. 15) and intensity (Fig. 16) for the 2016 operational forecasts reveal several notable differences compared to the 2015 verification (Figs. 1 and 2). There were only 40 verifiable track forecasts at 120 h for the 2015 Atlantic sample (Fig. 1a), but 97 verifiable forecasts for 2016 (Fig. 15a). While EMXI was clearly the best-performing HCCA input model for 2015 at all lead times, EGRI was the most skillful through 72 h for 2016. Perhaps not surprisingly, input model sensitivity experiments for the 2016 HCCA forecasts indicate that EGRI had the largest positive

FIG. 15. As in Fig. 1, but for 2016.

impact on the Atlantic track forecasts from 12 to 72 h (not shown). The skill of EMNI was very similar to that of EMXI, especially at longer lead times. For reference, track and intensity error values for 2016 forecasts are also shown in Tables 4 and 5 from a separate homogeneous verification that only includes HCCA, OFCL, and TVCN/IVCN. The performance of the HCCA forecasts was very similar to those of TVCN and OFCL at all lead times (Table 4); the TVCN and HCCA forecasts performed slightly better than the OFCL forecasts through 72 h, while the OFCL forecasts were the most skillful at 120 h.

Although the numbers of verifiable eastern North Pacific track forecasts for 2015 (Fig. 1a) and 2016 (Fig. 15a) were similar, the average track error for 2016 (Table 4) was generally less than it was for 2015 (Table 2). As in 2015, the forecasts of GHMI and EGRI in 2016 were the least skillful at all lead times, while the skill of the other HCCA input models was similar through all forecast hours. HCCA had the most skillful forecasts through 48 h, while TVCN was the most skillful from 72 to 120 h.

For the 2016 Atlantic intensity forecasts (Fig. 16a), HWFI was the most skillful of the HCCA input models from 24 to 36 h, and HWFI and CTCI outperformed the other input models at 96 and 120 h. Whereas HCCA was

more skillful than IVCN (and OFCL) at longer lead times for 2015 (Fig. 2a), the skill of the HCCA intensity forecasts for 2016 was slightly less than that of IVCN from 36 to 120 h. Relative to 2015, the performance of IVCN appears to have improved for 2016. This may have been due to the addition of CTCI beginning in 2016 (Cangialosi and Franklin 2016) or may be due to the improved performance of GHMI in 2016 relative to 2015 for Atlantic intensity forecasts (cf. Figs. 2a and 16a).

While HCCA was competitive with the OFCL eastern North Pacific intensity forecasts at early lead times for 2015 (Fig. 2b), its performance lagged behind OFCL for the 2016 eastern North Pacific intensity forecasts (Fig. 16b). The superior performance of HCCA during 2015 may have been partly due to the large number of forecasts that experienced rapid intensification, since HCCA intensity forecasts possess a high degree of skill for eastern North Pacific RI events (Fig. 12b). The skill of the 2016 eastern North Pacific intensity forecasts was very similar for HCCA and IVCN (Fig. 16b). From 48 to 120 h, the skill levels of the HWFI, OFCL, IVCN, and HCCA forecasts were very similar; HCCA had the most skillful forecasts at 48 h, and the IVCN forecasts were the most skillful at 120 h. Additional analysis (not shown) revealed that

FIG. 16. As in Fig. 2, but for 2016.

implementing the increment–value blend for the 2016 intensity forecasts made a positive impact.

7. Summary and future work

This study describes the development of the Hurricane Forecast Improvement Program (HFIP) Corrected Consensus Approach (HCCA) for tropical cyclone track and intensity forecasts that was recently implemented as an in-house operational guidance product at the National Hurricane Center.

The generation of HCCA forecasts uses the forecasts of several input models for track and intensity, in a manner similar to that used in other corrected consensus guidance

products. Unequal weighting coefficients are derived for each of the input models using multiple linear regression based on training forecasts that extend back several years.

Verification of operational forecasts for the 2015 Atlantic and eastern North Pacific seasons revealed that HCCA provides skillful guidance for both track and intensity. Compared to the HCCA input models, the equally weighted variable consensus, and the official NHC forecasts, HCCA had the largest Atlantic track skill from 12 to 48 h, the largest Atlantic intensity skill at 96 and 120 h, and the largest eastern North Pacific intensity skill from 24 to 72 h.

Average coefficients were examined for the 2015 season to assess the relative contributions of the input

TABLE 4. As in Table 2, but for 2016. Based on a two-sided Student's *t* test, differences between the HCCA and TVCN error are not considered significant.

Forecast hour	2016 Atlantic track error (n mi)				2016 eastern North Pacific track error (n mi)			
	<i>N</i>	OFCL	TVCN	HCCA	<i>N</i>	OFCL	TVCN	HCCA
12	354	24.3	23.4	23.2	373	18.5	18.4	17.8
24	321	36.5	35.0	34.8	339	28.4	28.8	27.4
36	290	47.7	45.5	45.5	304	36.9	37.0	35.6
48	258	61.8	59.0	59.2	270	45.3	45.2	43.8
72	204	88.8	90.8	87.5	206	63.2	61.5	63.5
96	161	133.1	138.0	131.4	160	85.4	81.9	85.0
120	130	166.6	185.7	178.8	121	107.0	100.1	106.1

TABLE 5. As in Table 3, but for 2016. Based on a two-sided Student's *t* test, differences between the HCCA and IVCN error are not considered significant.

Forecast hour	2016 Atlantic intensity error (kt)				2016 eastern North Pacific intensity error (kt)			
	<i>N</i>	OFCL	IVCN	HCCA	<i>N</i>	OFCL	IVCN	HCCA
12	353	5.3	6.3	6.3	373	5.9	6.9	6.6
24	320	7.6	8.2	8.3	339	8.8	9.5	9.6
36	290	9.4	9.6	9.8	304	10.4	11.2	11.3
48	258	10.5	10.7	10.9	270	11.8	12.1	11.2
72	204	11.7	12.0	12.4	204	15.6	15.1	14.9
96	161	14.7	13.5	14.0	160	16.8	15.5	15.8
120	129	15.9	15.6	16.1	121	16.4	15.3	16.1

models. For the 2015 forecasts, EMXI and AEMI consistently had the largest track (latitude and longitude) coefficients at most forecast hours. The large weighting coefficients of EMXI/2 led to improved HCCA forecasts for Hurricane Joaquin (2015), demonstrating the utility of using unequal weighting coefficients. The relative magnitudes of the input model coefficients for the intensity forecasts were more varied.

Sensitivity experiments were conducted to evaluate the impact of the input models for the 2011–15 HCCA forecasts. For Atlantic track forecasts, the decline in HCCA skill for forecasts that excluded EMXI was largest, which indicates that EMXI is the most important HCCA input model for producing skillful Atlantic track forecasts. For eastern North Pacific track forecasts, the positive impact of EMXI is largest, followed by EGRI. Input model exclusion experiments for intensity reveal that HWFI has the largest positive impact for both the Atlantic and eastern North Pacific forecasts and that CTCI makes a positive impact at longer lead times in both basins.

The fact that HCCA forecast values can exist outside the range of input model forecast values appears to be an advantage over the equally weighted consensus for intensity forecasts during rapid intensification. For eastern North Pacific intensity forecasts from 2011 to 2015, HCCA had lower error than all of the HCCA input models, the equally weighted consensus, and the NHC official forecasts from 24 to 72 h when at least a 20-kt increase in intensity was observed over 24 h. In addition, HCCA forecasts of a 20-kt increase or greater over 24 h occurred more frequently than was predicted by the input statistical guidance models (DSHP and LGEM) between 24 and 60 h and more frequently than the NHC official forecasts at 24 and 36 h.

Upon the conclusion of the 2015 season, additional testing using retrospective forecasts indicated that the addition of several input models showed promise for further reducing the HCCA forecast error. Prior to the 2016 season, the ECMWF ensemble mean (EMNI) was added as an input model for Atlantic and eastern North

Pacific track forecasts, and the COAMPS-TC initialized from GFS initial and lateral boundary conditions (CTCI) was added as an intensity model. In addition, it was found that using an average of the HCCA intensity forecasts derived from the increment and value techniques was more skillful than using either of the techniques independently. This change was also implemented prior to the 2016 season.

Verification results indicate that HCCA continued to provide skillful guidance for the 2016 season. The error in the HCCA track forecasts was the lowest of the forecasts included in the comparison from 12 to 48 h for the eastern North Pacific, and the HCCA intensity error was the lowest at 48 h for the eastern North Pacific intensity forecasts.

In an effort to continue to improve the skill of the HCCA forecasts, there are several topic areas warranting future research: (i) determination of the optimal training set length; (ii) prior to each hurricane season, continued testing to evaluate the impact resulting from changes to existing models and the introduction of additional input models; (iii) further investigation into how the collinearity among input models affects the corrected consensus; (iv) evaluation of the skill of HCCA forecasts for systems in other basins; (v) exploration of strategies for selecting a subset of the available training forecasts based on the characteristics of the current forecast to better account for model performance/bias during different forecast scenarios; and (vi) accounting for instances when the size of an input model's coefficient is disproportional to its contribution to HCCA.

Acknowledgments. Funding for this work was provided by NOAA's Hurricane Forecast Improvement Program. We thank the various modeling center teams for working to continuously improve the forecast models and for providing retrospective forecasts that were used to develop and test HCCA. We also thank Chris Landsea and three anonymous reviewers for their suggestions and comments that helped improve the manuscript.

REFERENCES

- Aberson, S., 1998: Five-day tropical cyclone track forecasts in the North Atlantic basin. *Wea. Forecasting*, **13**, 1005–1015, [https://doi.org/10.1175/1520-0434\(1998\)013<1005:FDTCTF>2.0.CO;2](https://doi.org/10.1175/1520-0434(1998)013<1005:FDTCTF>2.0.CO;2).
- Berg, R. J., 2016: Hurricane Joaquin. National Hurricane Center Tropical Cyclone Rep. AL112015, 36 pp., http://www.nhc.noaa.gov/data/tcr/AL112015_Joaquin.pdf.
- Cane, D., and M. Milelli, 2010: Multimodel SuperEnsemble technique for quantitative precipitation forecasts in Piemonte region. *Nat. Hazards Earth Syst. Sci.*, **10**, 265–273, <https://doi.org/10.5194/nhess-10-265-2010>.
- Cangialosi, J. P., and J. L. Franklin, 2016: 2015 hurricane season. National Hurricane Center Forecast Verification Rep., 69 pp., http://www.nhc.noaa.gov/verification/pdfs/Verification_2015.pdf.
- Cartwright, T. J., and T. N. Krishnamurti, 2007: Warm season mesoscale superensemble precipitation forecasts in the southeastern United States. *Wea. Forecasting*, **22**, 873–886, <https://doi.org/10.1175/WAF1023.1>.
- Development Testbed Center, 2015: MET: Version 5.1 Model Evaluation Tools users guide. DTC Rep., 316 pp., http://www.dtcenter.org/met/users/docs/users_guide/MET_Users_Guide_v5.1.pdf.
- Gall, R., J. Franklin, F. Marks, E. N. Rappaport, and F. Toepfer, 2013: The Hurricane Forecast Improvement Project. *Bull. Amer. Meteor. Soc.*, **94**, 329–343, <https://doi.org/10.1175/BAMS-D-12-00071.1>.
- Goerss, J. S., 2000: Tropical cyclone track forecasts using an ensemble of dynamical models. *Mon. Wea. Rev.*, **128**, 1187–1193, [https://doi.org/10.1175/1520-0493\(2000\)128<1187:TCTFUA>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<1187:TCTFUA>2.0.CO;2).
- Gopalakrishnan, S., and Coauthors, 2016: 2015 HFIP R&D activities summary: Recent results and operational implementation. Hurricane Forecast Improvement Project Tech. Rep. HFIP2016-1, NOAA/Hurricane Forecast Improvement Program, 44 pp., http://www.hfip.org/documents/HFIP_AnnualReport_FY2015.pdf.
- Kaplan, J., and M. DeMaria, 2003: Large-scale characteristics of rapidly intensifying tropical cyclones in the North Atlantic basin. *Wea. Forecasting*, **18**, 1093–1108, [https://doi.org/10.1175/1520-0434\(2003\)018<1093:LCORIT>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<1093:LCORIT>2.0.CO;2).
- , and Coauthors, 2015: Evaluating environmental impacts on tropical cyclone rapid intensification predictability utilizing statistical models. *Wea. Forecasting*, **30**, 1374–1396, <https://doi.org/10.1175/WAF-D-15-0032.1>.
- Knaff, J. A., M. DeMaria, B. Sampson, and J. M. Gross, 2003: Statistical, 5-day tropical cyclone intensity forecasts derived from climatology and persistence. *Wea. Forecasting*, **18**, 80–92, [https://doi.org/10.1175/1520-0434\(2003\)018<0080:SDTCIF>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<0080:SDTCIF>2.0.CO;2).
- Krishnamurti, T. N., 2003: Methods, systems and computer program products for generating weather forecasts from a multi-model superensemble. U.S. Patent 6535817 B1, filed 13 November 2000, issued 18 Mar 2003.
- , C. M. Kishtawal, T. LaRow, D. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran, 1999: Improved weather and seasonal climate forecasts from multimodel superensemble. *Science*, **285**, 1548–1550, <https://doi.org/10.1126/science.285.5433.1548>.
- , —, D. W. Shin, and C. E. Williford, 2000a: Improving tropical precipitation forecasts from a multianalysis superensemble. *J. Climate*, **13**, 4217–4227, [https://doi.org/10.1175/1520-0442\(2000\)013<4217:ITPFFA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<4217:ITPFFA>2.0.CO;2).
- , —, Z. Zhang, T. LaRow, D. Bachiochi, C. E. Williford, S. Gadgil, and S. Surendran, 2000b: Multimodel ensemble forecasts for weather and seasonal climate. *J. Climate*, **13**, 4196–4216, [https://doi.org/10.1175/1520-0442\(2000\)013<4196:MEFFWA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<4196:MEFFWA>2.0.CO;2).
- , and Coauthors, 2001: Real-time multianalysis–multimodel superensemble forecasts of precipitation using TRMM and SSM/I products. *Mon. Wea. Rev.*, **129**, 2861–2883, [https://doi.org/10.1175/1520-0493\(2001\)129<2861:RTMMSF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2861:RTMMSF>2.0.CO;2).
- , S. Pattnaik, M. K. Biswas, E. Bensman, M. Kramer, N. Surgi, and T. S. V. V. Kumar, 2010: Hurricane forecasts with a mesoscale suite of models. *Tellus*, **62A**, 633–646, <https://doi.org/10.1111/j.1600-0870.2010.00469.x>.
- , M. K. Biswas, B. P. Mackey, R. G. Ellingson, and P. H. Ruscher, 2011: Hurricane forecasts using a suite of large-scale models. *Tellus*, **63A**, 727–745, <https://doi.org/10.1111/j.1600-0870.2011.00519.x>.
- Neumann, C. J., 1972: An alternate to the Hurrell tropical cyclone forecast system. NOAA Tech. Memo. NWS SR-62, 23 pp., <https://ntrl.ntis.gov/NTRL/dashboard/searchResults/titleDetail/COM7210351.xhtml>.
- NHC, 2017: NHC track and intensity models. National Hurricane Center, <http://www.nhc.noaa.gov/aboutmodels.shtml>.
- Williford, C. E., T. N. Krishnamurti, R. C. Torres, S. Coker, Z. Christidis, and T. S. V. Kumar, 2003: Real-time multimodel superensemble forecasts of Atlantic tropical systems of 1999. *Mon. Wea. Rev.*, **131**, 1878–1894, <https://doi.org/10.1175/2571.1>.