OPERATIONAL PREDICTION SYSTEM NOTES

The National Hurricane Center Tropical Cyclone Model Guidance Suite

Mark DeMaria,[a] James L. Franklin,[b] Rachel Zelinsky,[c] David A. Zelinsky,[c] Matthew J. Onderlinde,[c]
John A. Knaff,[d] Stephanie N. Stevenson,[c] John Kaplan,[e] Kate D. Musgrave,[a] Galina Chirokova,[a]
and Charles R. Sampson[f]

[a] *CIRA/Colorado State University, Fort Collins, Colorado*
[b] *Systems Research Group, NOAA/National Hurricane Center, Miami, Florida*
[c] *NOAA/National Hurricane Center, Miami, Florida*
[d] *NESDIS/STAR, Fort Collins, Colorado*
[e] *NOAA/AOML/Hurricane Research Division, Miami, Florida*
[f] *Naval Research Laboratory, Monterey, California*

ABSTRACT: The National Hurricane Center (NHC) uses a variety of guidance models for its operational tropical cy-clone track, intensity, and wind structure forecasts, and as baselines for the evaluation of forecast skill. A set of the simpler models, collectively known as the NHC guidance suite, is maintained by NHC. The models comprising the guidance suite are briefly described and evaluated, with details provided for those that have not been documented previously. Decay-SHIFOR is a modified version of the Statistical Hurricane Intensity Forecast (SHIFOR) model that includes decay over land; this modification improves the SHIFOR forecasts through about 96 h. T-CLIPER, a climatology and persistence model that predicts track and intensity using a trajectory approach, has error characteristics similar to those of CLIPER and D-SHIFOR but can be run to any forecast length. The Trajectory and Beta model (TAB), another trajectory track model, applies a gridpoint spatial filter to smooth winds from the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) model. TAB model errors were 10%–15% lower than those of the Beta and Advection model (BAM), the model it replaced in 2017. Optimizing TAB's vertical weights shows that the lower troposphere's environmental flow provides a better match to observed tropical cyclone motion than does the upper troposphere's, and that the optimal steer-ing layer is shallower for higher-latitude and weaker tropical cyclones. The advantages and disadvantages of the D-SHIFOR, T-CLIPER, and TAB models relative to their earlier counterparts are discussed.

SIGNIFICANCE STATEMENT: This paper provides a comprehensive summary and evaluation of a set of simpler forecast models used as guidance for NHC's operational tropical cyclone forecasts, and as baselines for the evaluation of forecast skill; these include newer techniques that extend forecasts to 7 days and beyond.

KEYWORDS: Hurricanes/typhoons; Forecast verification/skill; Operational forecasting; Statistical forecasting

---

1. Introduction

The National Hurricane Center (NHC) and Central Pacific Hurricane Center (CPHC) use a variety of models as guidance for their operational tropical cyclone (TC) track, intensity, and wind structure forecasts, and as baselines for evaluation of fore-cast skill. These range in complexity from simple statistical track and intensity models based on climatology and persistence, to nested ocean–atmosphere–coupled dynamical models such as the Hurricane Weather Research and Forecasting (HWRF) Model (Tallapragada et al. 2014). The more complex dynami-cal modeling systems are supported by other agencies such as the National Centers for Environmental Prediction (NCEP)

Environmental Modeling Center, while the simpler models are maintained and updated by NHC; this latter set of simpler models is collectively known as the NHC guidance suite.

In this paper we describe the models comprising the 2021 version of the NHC guidance suite, with particular emphasis on those models that have not been documented previously, such as Decay-SHIFOR, a modified version of the Statistical Hurricane Intensity Forecast (SHIFOR) model that includes decay over land, and a new Trajectory Climatology and Persistence model (T-CLIPER) that predicts track and inten-sity using a trajectory approach.

One traditional member of the NHC guidance suite was the Beta and Advection Model (BAM), a simple trajectory track model based on a theoretical development by Holland (1983) and adapted for operational use by Marks (1992). In the BAM, a TC follows a trajectory from the NCEP Global Fore-casting System (GFS), modified by a poleward and westward

---

TABLE 1. Summary of models comprising the 2021 NHC guidance suite. For the WCOSS type, "P" indicates production and "QP" indicates quasi-production. Intensity refers to the maximum sustained (1-min mean) wind. Bold font indicates models discussed in detail in this paper.

Model	ATCF tech ID	WCOSS type	Forecast parameters	Model type
Extrapolation	XTRP	P	Track	Statistical
CLIPER5	CLP5 and OCD5	P	Track	Statistical
SHIFOR5	SHF5	P	Intensity	Statistical
D-SHIFOR	OCD5	P	Intensity	Statistical
T-CLIPER	TCLP	P	Track, intensity	Statistical
TAB (GFS)	TABS, TABM, TABD	P	Track	Dynamical
TAB (ECMWF)	TBSE, TBME, TBDE	P	Track	Dynamical
SHIPS	SHIP	P	Intensity	Statistical–dynamical
D-SHIPS	DSHP	P	Intensity	Statistical–dynamical
LGEM	LGEM	P	Intensity	Statistical–dynamical
DSWR	DSWR	QP	Wind radii	Statistical–dynamical
SHIPS-RII	RIOB, RIOD, RIOL, RIOC	P	Probability of rapid intensification	Statistical–dynamical
DTOPS	DTOP	P	Probability of rapid intensification	Statistical–dynamical
HCCA	HCCA	QP	Track, intensity	Statistical–dynamical
NNIB	NNIB	P	Intensity	Statistical–dynamical
NNIC	NNIC	P	Intensity	Machine learning

displacement to account for beta drift. The steering flow that defined the trajectories was calculated by smoothing the global model wind fields using a T25 spherical harmonic truncation and then recreating corresponding horizontal wind vectors. The smoothed winds were then averaged in the vertical. Shallow, medium, and deep versions of BAM were run with vertical averages over 850–700 hPa (BAMS), 850–400 hPa (BAMM), and 850–200 hPa (BAMD), respectively. Although the BAM models were not as skillful as the global and regional dynamical models, they still provided useful information about shear in the steering flow and required trivial computational resources. The BAMM model also provided a convenient alternative track for certain statistical intensity models (e.g., SHIPS) when an official forecast from NHC was unavailable.

The dynamical core of the NCEP GFS was converted from a spectral model to the finite volume model in 2019 (https://www.gfdl.noaa.gov/fv3/, Harris et al. 2021), and with that transition the spectral coefficient files were no longer available as drivers for BAM. This motivated the development of the Trajectory and Beta model (TAB), a modified version of the BAM that uses a gridpoint smoother to determine the steering flow. The TAB and BAM models were run in parallel in 2016, and BAM was replaced by TAB in 2017.

NHC also runs two statistical postprocessing models as guidance for its operational TC genesis forecasts (Dunion et al. 2019; Halperin et al. 2017). These techniques are still considered experimental but may be added to the operational guidance suite in the future.

In section 2 we review the models currently comprising the NHC guidance suite. Section 3 describes the D-SHIFOR, T-CLIPER, and TAB model developments, including the adaptation of T-CLIPER for the western North Pacific, Indian Ocean, and Southern Hemisphere for use by the Joint Typhoon Warning Center (JTWC). Verification results for D-SHIFOR, T-CLIPER, and TAB are presented in section 4, along with the evaluation of a method for determining an optimal steering layer for TAB. Conclusions are provided in section 5.

2. The NHC guidance suite

NHC has a long history of developing and running statistical guidance models (DeMaria and Gross 2003). In 2014, NHC's models were transitioned to the Weather and Climate Operational Supercomputer System (WCOSS), and NCEP Central Operations (NCO) assumed responsibility for running the real-time NHC guidance suite. Prior to that time the models were run by NHC on a variety of computer systems. The suite is run for all TCs in NHC's areas of responsibility (AORs) in the North Atlantic and eastern North Pacific to 140°W, as well as for TCs in the CPHC AOR (the North Pacific from 140°W to the international date line). Real-time model outputs are copied from WCOSS to the Automated Tropical Cyclone Forecast (ATCF) system (Sampson and Schrader 2000) for operational use by NHC and CPHC forecasters. There are two types of runs on WCOSS, production and quasi-production. Production runs are managed by NCO and updates are generally made only once per year, although emergency bug fixes are possible through a request for change (RFC) process. Quasi-production runs are managed by NHC and are used to evaluate models before they are moved to production. Both types of runs are available to forecasters in real time.

Table 1 summarizes the models comprising the 2021 guidance suite. In the ATCF, each model has a unique four-character identifier; these are denoted by "Tech ID" in Table 1 and for brevity will be used to identify models in some of the figures and tables below (a full list of acronyms and their expansions are given in the appendix). The ATCF supports several different file types for each storm: deterministic forecasts appear in what is known as the a-deck, probabilistic forecasts reside in the e-deck, while a TC's official history (the "best track") used for verification is in the storm's b-deck.

The models in Table 1 are listed in rough order of increasing complexity. XTRP is simply an extrapolation of the current observed storm motion through 7 days to provide a reference for track changes. The Climatology and Persistence model (CLIPER)

forecasts TC track using the initial latitude, longitude, maximum sustained wind (intensity), the time tendencies of position and intensity, and the Julian day. The original version (Neumann 1972) provided forecasts to 72 h and was later updated to provide 120-h forecasts (CLIPER5, Aberson 1998). CLIPER was first developed as an explicit forecast tool, but its direct use diminished by the 1980s as more accurate track prediction models became available (DeMaria and Gross 2003). However, CLIPER5 remains useful as a method to account for annual forecast difficulty (through its mean track errors, Cangialosi 2021) and as a benchmark to measure track forecast skill.

The Statistical Hurricane Intensity Forecast model (SHIFOR) was developed using an approach analogous to CLIPER's to predict a TC's maximum sustained winds out to 72 h (Jarvinen and Neumann 1979). An updated version that extended the forecasts to 120 h (SHIFOR5) was developed by Knaff et al. (2003). An important limitation of SHIFOR5, however, was that it presumed a TC would remain over water. A more useful version, known as Decay-SHIFOR (D-SHIFOR), was developed in 2007; this version assumes a forecast track given by CLIPER5 and applies a climatological intensity decay rate for the overland portions of that track. For convenience, the 120-h D-SHIFOR intensity forecasts are combined with the CLIPER5 track forecasts into a single aid in the ATCF designated OCD5. D-SHIFOR is described in more detail in sections 3 and 4, along with the next two models in Table 1, T-CLIPER and TAB.

The remaining models in Table 1 are statistical–dynamical models, meaning that they use output from dynamical models as predictors in statistical algorithms. The Statistical Hurricane Intensity Prediction Scheme (SHIPS) and Logistic Growth Equation Model (LGEM) are described by DeMaria et al. (2005) and DeMaria (2009), respectively. D-SHIPS is a version of SHIPS that accounts for the decrease in maximum wind when a TC moves over land.

SHIPS and LGEM have been updated several times since the models were described by DeMaria (2010). In 2011, the 700–850-hPa layer-mean temperature advection was added as a predictor to reflect the impact of baroclinic interactions on forecasts at higher latitudes. Klein et al. (2000) had shown that interaction of a TC with preexisting baroclinic zones provides an alternate energy source for the system during extratropical transition. In SHIPS and LGEM, the temperature advection is calculated from the 0–600-km-averaged TC environmental horizontal winds, and the geostrophic thermal wind is used to estimate the horizontal temperature gradient.

Another change to SHIPS and LGEM was made in 2016; prior to that year the vortex-removal procedure included winds but not temperature, which tended to cause a low intensity forecast bias as the global model representation of the upper-level warm core of the TC improved. This low bias was caused by the contribution of the vortex warm anomaly to the environmental upper-level temperature predictor, which artificially increased the vertical stability. To account for that, the 200-hPa temperature anomaly associated with the symmetric vortex averaged from a radius of 0–1000 km was added as a new predictor. The starting point for the calculation of the temperature anomaly is the symmetric tangential wind at each pressure level, which is calculated from 0- to 1000-km radius. Assuming the symmetric tangential wind is in

gradient and hydrostatic balance, the gradient thermal wind equation provides an estimate of the radial temperature gradient associated with the symmetric vortex. That gradient at 200 hPa is integrated inward from 1000 to 0 km to provide the temperature anomaly associated with the vortex as a function of radius. By including a measure of the TC's own upper-level temperature anomaly as an explicit predictor, the negative impact of that anomaly on the inferred vertical stability can be overcome.

To improve the performance of SHIPS and LGEM for TCs undergoing rapid intensification (RI), in 2016 the probability of RI for the 30 kt $(24\,\text{h})^{-1}$ $(1\,\text{kt} \approx 0.51\,\text{m s}^{-1})$ threshold was added as a predictor, based on a stand-alone version of the SHIPS-RII (RIOD, see below). This predictor was replaced by the RI probability for the 55 kt $(48\,\text{h})^{-1}$ threshold in 2018, and by the 35 kt $(36\,\text{h})^{-1}$ probability in 2019. The performance of SHIPS and LGEM for RI cases has improved since the RI predictors were added (DeMaria et al. 2021).

In 2014 the ocean heat content (OHC) estimated from satellite altimetry (Mainelli et al. 2008) was replaced by the Navy's Coupled Ocean Data Assimilation (NCODA; Cummings 2005) OHC, which is calculated at NHC from the full three-dimensional temperature fields obtained from the Fleet Numerical Meteorology and Oceanography Center. In 2017, the weekly Reynolds sea surface temperature (SST) analyses (Reynolds and Smith 1994) were replaced by the daily Reynolds SST analyses (Banzon et al. 2020). During the 2018 season, the daily Reynolds SSTs had unrealistic cold anomalies in parts of the eastern Pacific, so the weekly Reynolds SSTs were again used for the second half of that season. The weekly Reynolds SSTs also showed unrealistic anomalies in parts of the Atlantic in the latter part of 2018, so the NCODA SSTs were used beginning in 2019, a change that also made the surface and subsurface ocean inputs consistent.

Perhaps most significantly, in 2020 the SHIPS and LGEM forecasts were extended from 120 to 168 h (7 days). Retrospective runs from 2013 to 2019 showed that beyond 5 days, the 144- and 168-h forecasts were skillful relative to T-CLIPER.

DSWR (D-SHIPS Wind Radii) is a statistical–dynamical model for predicting the radial extent of 34-, 50- and 64-kt winds outward from the TC's center using a subset of the SHIPS predictors and the D-SHIPS intensity forecast. Details can be found in Knaff et al. (2017).

The guidance suite includes four models that use a subset of predictors from SHIPS to estimate the probability of rapid intensification in the next 12–72 h: the RI Operational Discriminate Analysis (RIOD), RI Operational Bayesian (RIOB), RI Operational Logistic Regression (RIOL), and the RI Operational Consensus (RIOC). A different approach is taken by the Deterministic to Probabilistic Statistical (DTOPS) model, which estimates RI probabilities using deterministic intensity model forecasts as input. These RI models are described in DeMaria et al. (2021).

The Hurricane Forecast Improvement Project (HFIP) Corrected Consensus Approach (HCCA) is a consensus track and intensity model that uses optimized weights of several input models, where the weights are determined by multiple linear regression with input from the previous few hurricane seasons (Simon et al. 2018).

The Neural Network Intensity Consensus (NNIC) model is NHC's first attempt at using machine learning techniques to forecast TC intensity change. It uses input from four deterministic models and a few storm environment parameters, such as vertical shear, to predict intensity changes using a three-layer neural network. The Neural Network Intensity Baseline (NNIB) is just the equally weighted average of the four models used as input to NNIC. Details on NNIB and NNIC can be found in DeMaria (2021).

3. Model development

D-SHIFOR, T-CLIPER, and TAB are not well documented, so further details are presented here. In the discussion that follows, ATL refers to the North Atlantic basin while EP/CP refers to a combination of the eastern North Pacific and central North Pacific basins.

a. D-SHIFOR

As noted in section 1, the 72-h versions of the CLIPER and SHIFOR models have been run operationally at NHC for several decades, and the 120-h versions (CLIPER5 and SHIFOR5) have been run since 2003. However, SHIFOR5 has limited utility as a skill baseline for intensity because it does not include the effects of land. To address that limitation, the methodology developed to include land effects in SHIPS was applied to SHIFOR5, as described below.

In SHIPS, an initial forecast is obtained that ignores the presence of land. A postprocessing step is then applied, in which the most recent NHC official forecast track is linearly interpolated to 1-h intervals, and the points for which the TC center is over land are identified. The inland decay models developed by Kaplan and DeMaria (1995) for TCs south of 35°N, and by Kaplan and DeMaria (2001) for TCs north of 40°N, are applied to reduce the forecast intensity for the portion of the track over land. (For locations from 35° to 40°N the decay rate is linearly interpolated from the two decay models.) When the center is over land, but some portion of the area within 111 km (1° latitude) of the center is still over water, the decay rate is reduced by the fraction of that area that is over water, as described by DeMaria et al. (2006). If the TC moves back over water, the time tendency of the intensity forecast for the remainder of the track is set to that from the SHIPS forecast without the land interaction.

A similar procedure was applied to the SHIFOR5 intensity forecast to produce D-SHIFOR, a climatology/persistence model that accounts for TC decay over land. In this case, however, the CLIPER5 track forecast was used instead of the NHC official forecast to determine which parts of the track are over land.

It should be noted that climatology/persistence models can generally be run in one of two modes, depending on the application. In forecast applications and for the remainder of this paper, operational estimates of storm parameters (e.g., motion, intensity, etc.) have been used. For certain applications, such as long-term historical analyses when operational estimates may not be available, it is possible to run climatology/persistence models using storm positions and intensities taken from the final best track determined from all available observations after the storm is over. Because best track data are more accurate than operational data, errors from climatology/persistence models are generally smaller when run in best track mode.

b. T-CLIPER

The T-CLIPER model development required a climatology of TC tracks and intensities, which were obtained from NHC best tracks (Landsea and Franklin 2013) for the ATL sample and from the combined NHC and CPHC best tracks for the EP/CP sample. The best track data for the initial development of T-CLIPER included all TC cases from 1982 to 2011. To ensure adequate coverage over areas where TCs are rare, the entirety of a storm's best track was used, including any data from a storm's extratropical stage. Unnamed depressions (i.e., those that did not go on to become tropical storms) were included in the developmental sample only from 1989 to 2011. Unnamed depressions in the best tracks prior to 1989 were not included because the procedures for identifying these cases were very different before that time, resulting in a much larger number of unnamed depressions, many of which were short lived. To increase coverage outside of hurricane season, all TC cases from December through April were included back to 1946 (the start of the aircraft reconnaissance era). In a subsequent update of T-CLIPER for the 2020 season, best track data from 2012 to 2018 were added to the developmental sample. For the JTWC version of T-CLIPER, the developmental data were the best tracks from 1982 to 2015.

Part of the motivation for the development of T-CLIPER was the need for a 7-day skill baseline to support evaluations of NHC's experimental 7-day forecasts and longer-range dynamical models. Extending the CLIPER5 and SHIFOR5 models from 120 to 168 h was impractical because the models' multiple-regression method required TC track and intensity changes over the full forecast interval, and 7-day lifetimes were too scarce. Figure 1 shows that only about 27% (21%) of the ATL (EP/CP) TC best track points still had an ongoing TC or post-tropical cyclone in the best track 7 days later, and only about 13% (8%) did so 10 days later. Such a sample was simply too small to yield a robust result.

To overcome the sample size problem, T-CLIPER was developed using a trajectory approach based on the following equations:

$$\frac{d\lambda}{dt} = \frac{1}{a}\cos\theta(w_{up}u_p + w_{us}u_s + \cos\theta\beta_x), \quad (1a)$$

$$\frac{d\theta}{dt} = \frac{1}{a}(w_{vp}v_p + w_{vs}v_s + \cos\theta\beta_y), \quad (1b)$$

where λ is longitude; θ is latitude; t is time; a is Earth's radius; u_p and v_p are the zonal and meridional components of the initial storm motion vector, respectively; u_s and v_s are the zonal and meridional components of the steering flow, respectively; β_x and β_y are the eastward and poleward contributions to storm motion due to the beta effect, respectively; and w_{up}, w_{us}, w_{vp}, and w_{vs} are time-dependent weights. Given an initial latitude and longitude, Eqs. (1a) and (1b) were solved numerically with a

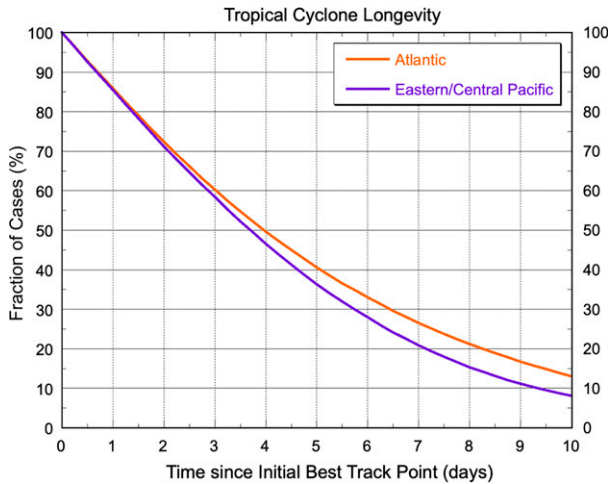

FIG. 1. The percentage of TC best track points that were still on-going as a tropical or post-tropical cyclone at the indicated times into the future. The sample comprises best track data, from 1982 to 2018 for the in-season months of May–November, and 1946–2018 for the out-season months December–April. The ATL (EP/CP) samples consists of 16 104 (20 296) best track points.

second-order Adams–Bashforth time differencing scheme with a 1-h time step. T-CLIPER forecasts were run to 10 days or for as long as the TC center remained in the computational domain (0°–60°N, 170°–10°E).

The components of the initial storm motion vector (u_p and v_p) were determined from the initial and 12-h-old storm positions and were held constant during the time integration. The components of the steering flow (u_s and v_s) represent the climatological motion vector as a function of latitude and longitude and were determined from the best track positions. To obtain the climatological motion vectors, single-pass Barnes (1964) objective analyses of storm motion components were performed over the computational domain for each month from May to December using the combined best track data from the Atlantic, eastern Pacific, and central Pacific basins. Due to the lack of TCs in the off-season, all motion vectors from January to April were combined to form a single climatology for those months. The Barnes analysis used an *e*-folding radius of 1500 km and an influence radius of 4000 km to provide the storm motion fields on the latitude/longitude grid. The beta terms in Eq. (1) were set to zero because the initial and climatological motion vectors were determined from observed storm motions, which are presumed to already include any beta drift. The monthly means were linearly interpolated to the day of the forecast for the T-CLIPER forecasts.

The weights in Eq. (1) determine the relative importance of the persistence and climatological motion vector components. To estimate an optimal weighting, a lag correlation analysis was performed for the combined basin developmental sample. Figure 2 shows that the correlation between the initial and time-lagged motion vector components drops off quickly at first, and then more slowly after a day or two, suggestive of an exponential decay behavior. The figure also shows that the correlation of the zonal component decreases more slowly
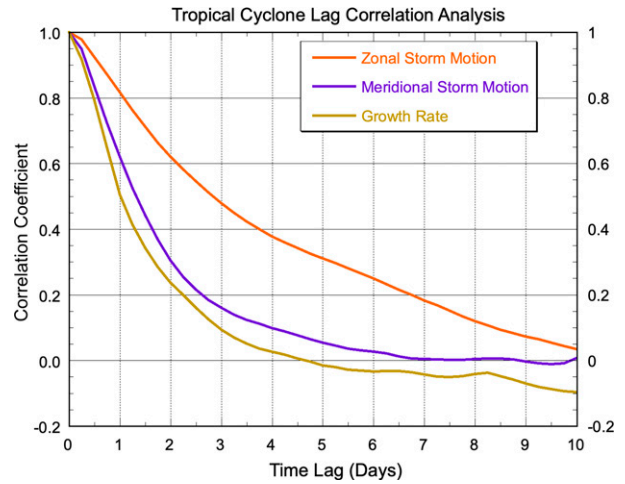

FIG. 2. Lag correlation of the zonal and meridional components of the storm motion vector and the maximum-wind growth rate from the combined ATL and EP/CP best track data used for Fig. 1.

than the meridional component. Based on the correlation analysis, the persistence weights were assumed to be given by

$$w_{up} = A_u \exp\!\left(-\frac{t}{t_{\text{ef}u}}\right), \qquad (2a)$$

$$w_{vp} = A_v \exp\!\left(-\frac{t}{t_{\text{ef}v}}\right), \qquad (2b)$$

where $t_{\text{ef}u}$ and $t_{\text{ef}v}$ are the *e*-folding times of the persistence weighting functions, and A_u and A_v are factors between 0 and 1 that determine the initial weight of the initial motion vector. Once the persistence weights are calculated, the weights on the steering wind components are given by

$$w_{us} = 1 - w_{up}, \qquad (3a)$$

$$w_{vs} = 1 - w_{vp}. \qquad (3b)$$

It was assumed that the short-term TC motion would follow the current motion vector, so A_u and A_v were set to 1.

With the above formulation, the weights in Eq. (1) can be calculated at any forecast time once the two *e*-folding values in Eq. (2) are specified. These were determined by running the T-CLIPER forecasts for the developmental sample for a range of *e*-folding times and choosing the values that minimized the sample-mean track errors averaged through 10 days. That analysis resulted in values of 76 and 40 h for $t_{\text{ef}u}$ and $t_{\text{ef}v}$, respectively. The larger value for the zonal component is consistent with the slower decrease of the correlation with time for the zonal component in Fig. 2.

Even with the 30-yr hurricane season sample and 66-yr off-season sample, there were still some areas of the analysis domain with sparse coverage. As noted above, T-CLIPER was updated for the 2020 season by adding cases from 2012 to 2018; for this update the weighting parameters were also reevaluated. The added years improved the data coverage, but there were still portions of the analysis domain with

TABLE 2. The parameters used in the T-CLIPER and TAB models. N/A indicates a parameter is not applicable to that model. The first two T-CLIPER lines are for the NHC/CPHC versions, and the second two T-CLIPER lines are for the 2018 JTWC Northern and Southern Hemisphere versions.

Model	A_{pu}	A_{pv}	t_{efu} (h)	t_{efv} (h)	$t_{ef\kappa}$ (h)	β_x (m s^{-1})	β_y (m s^{-1})	R_{avg} (km)
T-CLIPER (2013)	1.0	1.0	76	40	8	0	0	N/A
T-CLIPER (2020)	1.0	1.0	48	40	12	0	0	N/A
T-CLIPER (J-North)	1.0	1.0	70	46	31	0	0	N/A
T-CLIPER (J-South)	1.0	1.0	63	42	32	0	0	N/A
TAB (2016)	0.85	0.85	6	6	N/A	−0.33	0.25	350
TAB (2018)	0.85	0.85	6	6	N/A	−0.33	0.25	400
TAB (2019)	0.85	0.85	6	6	N/A	−0.33	0.25	600

sparse data, and the 4000-km influence radius of the Barnes analysis was still needed to obtain a smooth motion vector field. T-CLIPER parameters for the original and updated versions of the model can be found in Table 2.

In principle, the approach used for the track component of T-CLIPER could also be applied to intensity, by calculating the climatological and persistence components of the maximum wind tendency dV/dt and integrating that with time. However, a TC's intensity is bounded by zero and a maximum potential intensity (MPI) determined by the thermodynamic environment (e.g., Bister and Emanuel 2002). Simply integrating the climatological maximum wind tendency could result in negative intensity values or values greater than the MPI. Also, the model would be more useful if decay over land were included. For these reasons, the predictive equation for the LGEM (DeMaria 2009) was adapted for the intensity portion of T-CLIPER, where the LGEM model parameters were determined from climatology and persistence.

The prediction equation for LGEM is given by

$$\frac{dV}{dt} = \kappa V - \beta \left(\frac{V}{V_{mpi}}\right)^n V, \qquad (4)$$

where t is time, V is the maximum wind, V_{mpi} is the maximum potential intensity, β and n are empirical constants that determine the rate at which the solution relaxes toward V_{mpi}, and κ is the maximum-wind growth rate. This equation bounds the solution for V to lie between zero and a fraction of V_{mpi} so that the T-CLIPER maximum-wind prediction cannot become negative or unrealistically large. The constants β and n in Eq. (4) used for T-CLIPER were set to the same values used in the operational version of LGEM (1/24 h^{-1} and 2.5, respectively). Theoretical studies (e.g., Bister and Emanuel 2002) show that V_{mpi} is a function of the SST and the temperature and moisture structure in the storm environment. However, the operational LGEM uses an empirical formula that only depends on SST and the storm translation speed to calculate V_{mpi}. That same approach was used for T-CLIPER, except that a climatological SST was used in place of the observed SST. The time tendency dV/dt and V_{mpi} were calculated for each point in the best track developmental dataset, and with β and n specified, the value of κ was calculated from Eq. (4). The Barnes analysis described previously was used to provide climatological monthly gridded values of κ.

The climatological SST for T-CLIPER was derived from the global Reynolds weekly SST analyses version 2 (Reynolds and Smith 1994), averaged to provide monthly means. For the initial version of T-CLIPER, SST data from 1982 to 2011 on a 1° latitude–longitude grid were used to calculate the monthly means; SST analyses through 2018 were added to the climatological fields for the 2020 model update. The climatological values are linearly interpolated to the position and day of a particular storm case for T-CLIPER.

For the T-CLIPER intensity prediction, κ was assumed to be given by a weighted mean of the persistence and climatological components, similar to that for the motion vector in Eq. (2), using

$$\kappa = w_{\kappa p} \kappa_p + w_{\kappa c} \kappa_c, \qquad (5)$$

where κ_p is the initial growth rate value determined from the initial and 12-h-old intensity, κ_c is the climatological growth rate, and $w_{\kappa p}$ and $w_{\kappa c}$ are time-dependent weights. Figure 2 shows the time-lag analysis for κ; the correlation has an exponential decay behavior for the first day or two, leading to a persistence weight given by

$$w_{\kappa p} = \exp\left(-\frac{t}{t_{ef\kappa}}\right), \qquad (6)$$

where $t_{ef\kappa}$ is the e-folding time of the initial growth rate, and a climatology weight given by

$$w_{\kappa c} = 1 - w_{\kappa p}. \qquad (7)$$

The e-folding time in Eq. (6) was determined empirically for the developmental sample by integrating Eq. (5) to 240 h along the portion of the T-CLIPER tracks that were over water and choosing the value that minimized the 240-h intensity error. An e-folding time $t_{ef\kappa}$ of 8 h provided the most accurate T-CLIPER intensity forecasts for the developmental sample. When the T-CLIPER track crossed land, the empirical inland decay model described by DeMaria et al. (2006) was used to determine the intensity tendency dV/dt.

The behavior of T-CLIPER for track and intensity can be understood from Eqs. (1) and (4). The short-term forecast motion and intensification rate will be governed mainly by persistence. In the longer term (after 48–72 h when t exceeds the e-folding times in the weights) the motion will be determined by the climatological motion vector analysis. When the
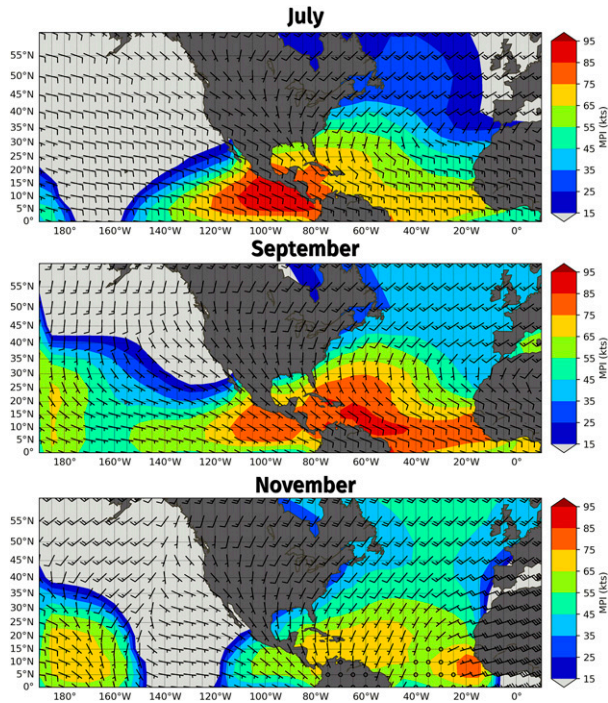
FIG. 3. The climatological storm motion vector field (kt; full barb equals 10 kt) and steady-state intensity (shading; kt) for (top) July, (middle) September, and (bottom) November used in the 2020 version of the T-CLIPER model.

growth rate κ and V_{mpi} vary slowly, as is the case when they are determined from climatological analyses, Eq. (4) has a steady-state solution for V given by

$$V_s = V_{mpi}\left(\frac{\kappa}{\beta}\right)^{1/n}. \qquad (8)$$

Once the influence of persistence becomes small (when t exceeds the e-folding time of 8 h), the intensity will approach the steady-state intensity given by Eq. (8). Figure 3 shows examples of the climatological motion vectors and steady-state maximum wind for early, peak, and late months of the hurricane season. The seasonal variations can be seen in Fig. 3, with Atlantic TCs much more likely to recurve late in the season than at other times, and peak steady-state intensities occurring in July for the eastern Pacific and in September for the tropical Atlantic.

In 2016, T-CLIPER was adapted to JTWC's AORs, which comprise the western North Pacific, Indian Ocean, and

Southern Hemisphere. Two computational domains were used for JTWC; the J-North domain covers 0°–70°N, 30°E–140°W and the J-South domain covers 0°N–50°S, 20°E–140°W. The analysis of the climatological motion vectors, growth rates, and the persistence time weights largely followed that for the NHC basins but with a few modifications. A two-pass Barnes analysis was performed instead of a one-pass analysis, with a large radius of influence and e-folding radius on the first pass and smaller values on the second pass; this was necessary to ensure coverage over the large JTWC domains. The radii of influence and e-folding distances for trajectory and growth-rate components are provided in Table 3. It was also found that better performance was obtained by modifying the n and β parameters in the LGEM Eq. (8), based on verifications of real-time runs from 2016 to 2017. The n parameter was reduced to 2.1 for both the J-North and J-South domains (compared with 2.5 for NHC), and the β parameter was increased to 1/18 h^{-1} for the J-South domain. The zonal motion was also increased by 5% in the J-North domain to correct a slow bias, and the e-folding persistence time weights were adjusted. The 2018 JTWC T-CLIPER parameter values are shown in Table 2, and only that version will be considered in the remainder of the paper. Figure 4 shows examples of the climatological motion vectors and growth rate for the J-North and J-South domains. This figure shows the scale of the spatial variability of the climatological motion vector and intensity growth-rate fields.

c. TAB

Equation (1), used for T-CLIPER, is also used for the TAB model, but where the steering vector components u_s and v_s are determined by horizontally and vertically averaged winds from a parent dynamical model, the beta-drift terms are not assumed to be zero, and the initial persistence weights A_u and A_v are not assumed to be equal to 1. The horizontal area for averaging the steering components is a circle of radius R_{avg}. Thus, the use of Eq. (1) for the TAB model requires the estimation of the seven parameters R_{avg}, β_x, β_y, t_{efu}, t_{efv}, A_u, and A_v.

The steering winds for TAB are obtained from GFS or European Centre for Medium-Range Weather Forecasts (ECMWF) horizontal wind fields with 1° latitude–longitude resolution at mandatory pressure levels. Although higher horizontal and vertical resolution data are available, the smoothing applied by the TAB model makes higher-resolution inputs unnecessary.

The specific values of the seven TAB parameters were chosen by minimizing the average track errors out to 7 days for

TABLE 3. Radii of influence (RoI) and e-folding distances (EFD) used for the two-pass Barnes analysis of climatological trajectory and growth rate components for the J-North and J-South T-CLIPER model developments.

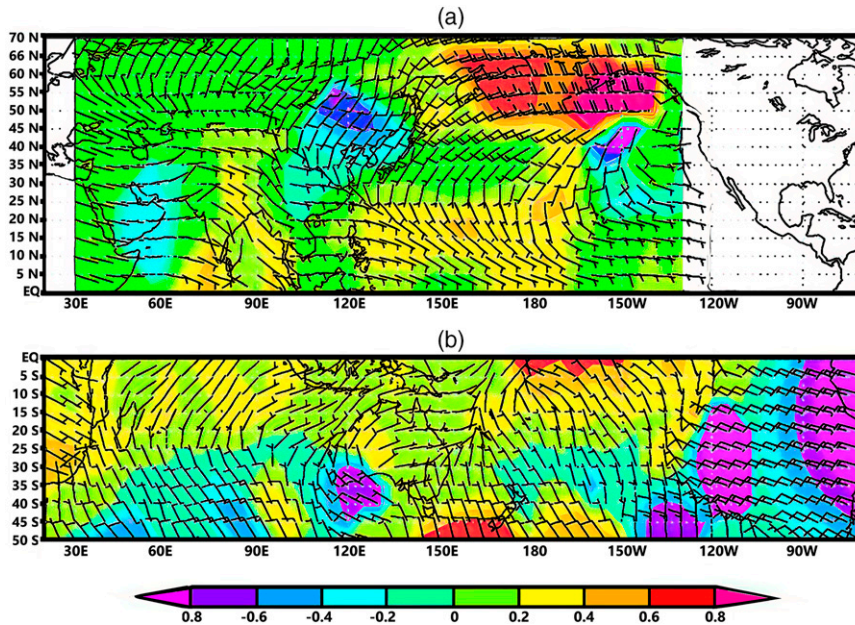| Model	First RoI (km)	Second RoI (km)	First EFD (km)	Second EFD (km)
J-North trajectory	6250	1875	3375	1250
J-South trajectory	12 500	1875	2500	675
J-North growth rate	3750	1875	750	375
J-South growth rate	12 500	1875	2500	375

FIG. 4. Climatological motion vectors (kt; full barb equals 10 kt) and growth rates (shading; day^{-1}) for (top) September in the J-North domain and (bottom) in February for the J-South domain. Note that the J-North plot was extended to the east of the J-North domain for consistency with the J-South plot.

the ATL and EP/CP forecasts from 2010 to 2015. Three versions of TAB were run using the same vertical layer averages as for the BAM model described earlier. These are designated as TABS, TABM, and TABD, where S, M, and D designate shallow, medium, and deep, respectively. The ranges of parameter values to test were determined from physical reasoning. Some additional guidance on the range for R_{avg} was determined by comparing the T25 spectral truncation used in the BAM model to a gridpoint smoother based on an area average used in the TAB model. Sardeshmukh and Hoskins (1984) showed that a triangular truncation results in uniform resolution on the sphere, which is consistent with a circular averaging area for TAB. With 25 waves at the equator, that corresponds to a wavelength of 1600 km. A simple horizontal average can be interpreted as a low-pass filter, and it can be shown that a filter with a half-power wavelength of 1600 km corresponds to a horizontal average over an area with a radius $R \sim 430$ km. That provided some guidance on how to choose the radius of the averaging area to roughly correspond to a T25 truncation. Several values around 430 km were tested and it was found that of all the model parameters, the track errors were most sensitive to R_{avg}, and the optimal value varied somewhat across the two basins and the three vertical layers. However, the differences were not large enough to justify the complication of using varying values of R_{avg}.

Table 2 shows the final parameters used for the 2016 operational version of TAB, based on the optimizations from the 2010 to 2015 developmental sample. The optimal value of R_{avg} (350 km) is close to the 430-km estimate based on the correspondence between an area average and spectral truncation. The e-folding times of the persistence weights are much

smaller for TAB than for T-CLIPER—an expected result since global model forecasts provide a much more accurate steering flow than climatology. The beta terms indicate a poleward and westward drift, which is consistent with beta-drift theory (e.g., Holland 1983). However, the diagnosed optimal beta drift is only ~0.4 m s^{-1}, which is much less than the 1–4 m s^{-1} found in idealized modeling studies as summarized by Shan and Yu (2020). This too is unsurprising, because the beta effect modifies the storm environment and so would be largely already captured in the steering flow from the GFS or ECMWF, minimizing the need for an explicit correction.

Operational forecasts from 2016 showed that TAB forecast tracks sometimes had unrealistic cyclonic oscillations with periods of about 12–48 h, due to inadequate removal of the global model representation of the TC. This was especially prominent with TABS. To help with that problem, R_{avg} was increased to 400 km in 2018. For the 2019 version of TAB, a more systematic analysis was performed, examining the distance traveled in the forecast compared with that in the observed tracks (since the spurious cyclonic oscillations result in a high bias in distance traveled). Optimization using the 2010–18 cases showed that an R_{avg} value of 600 km greatly reduces the oscillations with only a small increase in forecast error. That value has been used in all three TAB models since 2019.

d. Optimal steering layer analysis

As part of the SHIPS intensity model, a method was developed to identify vertical weights of the horizontal winds from global models that provided the best match to the observed storm motion vector. It was found that TCs steered by deep

layers were more likely to intensify (DeMaria 2010), so the center of mass of the optimal steering layer was added as a predictor to SHIPS and LGEM. A similar analysis was performed to provide insight into which version of TAB is the most appropriate for a given forecast situation and the physical factors that determine the steering layer, such as TC intensity and tropopause height.

The steering flow is determined by weighting area-averaged horizontal winds given by u_i, v_i at N vertical levels. The optimal weights w_i are found by minimizing the error function E given by

$$E = \left[\left(\sum_{i=1}^{i=N} w_i u_i\right) - c_x\right]^2 + \left[\left(\sum_{i=1}^{i=N} w_i v_i\right) - c_y\right]^2$$
$$+ \alpha\left(\sum_{i=1}^{i=N} \frac{w_i - m_i}{m_i}\right)^2, \tag{9}$$

subject to the constraint

$$\sum_{i=1}^{i=N} w_i = 1. \tag{10}$$

In Eq. (9), c_x and c_y are the eastward and northward components of the storm-motion vector, α is a parameter that controls the weight of a constraint term that ensures a unique solution for the optimal weights, and m_i are the weights based on a mass-weighted average. The first two terms on the right side of Eq. (9) measure the difference between the vertical average of the steering flow and the observed storm motion, and the third is a penalty term that measures how far the optimal weights are from the mass weights. The optimal weights are found by using Eq. (10) to eliminate one of the weights in Eq. (9), and then setting the derivative of Eq. (9) with respect to the remaining weights w_i to zero. That gives a symmetric linear system that can be solved for w_i from $i = 1$ to $N - 1$, with the last weight determined from Eq. (10). The penalty term in Eq. (9) is needed to prevent the linear system from being singular. For example, the matrix would be singular if the winds at all levels were the same, in which case there is no unique solution for the weights. From experimentation it was found that α of 0.4 kt^{-2} gives reasonable results. The optimal weights were not too sensitive to the choice of α because the wind profile usually has enough vertical variability to provide a unique solution.

Once the optimal weights are determined, the center of mass of the optimal steering layer [pressure of the steering layer (PSL$_O$)] is given by

$$\text{PSL}_o = \sum_{i=1}^{i=N} w_i P_i, \tag{11}$$

where the P_i are the pressures of each vertical level. With the mass-weighted average, the pressure of the steering layer (PSL$_M$) is just the average of the pressure at the top and bottom levels included in the vertical average. The PSL$_M$ values for TABS, TABM, and TABD are 775, 625, and 525 hPa, respectively. Comparing PSL$_O$ to the PSL$_M$ values for the three

TABLE 4. Years included in the verification of the D-SHIFOR, T-CLIPER, and TAB models for the ATL, EP/CP, and JTWC AORs. The JTWC AORs comprise the western North Pacific, Indian Ocean, and Southern Hemisphere.

Model	ATL	EP/CP	JTWC AORs
D-SHIFOR	2001–20	2001–20	
T-CLIPER	2012–20	2012–20	2018–20
TAB (GFS)	2016–20	2016–20	
TAB (ECMWF)	2018–20	2018–20	

TAB models provides guidance on which TAB model is best. Examples are presented in section 4c.

4. Verification results

Table 4 shows the years included in the D-SHIFOR, T-CLIPER, and TAB verifications presented in this section. D-SHIFOR was first run operationally in 2013, but its forecasts had been generated poststorm for use as a skill baseline by NHC beginning in 2006. During the preparation of this paper, it was discovered that the operational D-SHIFOR model from 2013 to 2019 had a bug for forecasts crossing the international date line; therefore, all the runs evaluated here were regenerated using the 2020 operational version of D-SHIFOR. The regenerated sample starts in 2001, the first year NHC made experimental 5-day forecasts, although it should be noted that the years 2001–04 are part of the CLIPER5 model's developmental dataset. The T-CLIPER and TAB forecasts evaluated here are the operationally available runs.

We follow here the standard NHC verification rules, in which the track and intensity forecasts are compared with the final NHC or CPHC best track, including only those cases when the system was classified as a tropical or subtropical cyclone at both the initial and verifying times. When comparing errors between models, a standard two-tailed statistical test is applied to determine statistical significance. Serial correlation is accounted for using the method described in Franklin and DeMaria (1992), but with 18 h used as the time between independent observations rather than 30 h; that change was made based on NHC's more recent unpublished assessment of the serial correlation between forecasts. Unless otherwise indicated, the 95% level was used as the threshold for statistical significance.

a. D-SHIFOR

Figure 5 shows the D-SHIFOR mean absolute errors for the ATL and EP/CP samples. The intensity errors increase at a fairly steady rate with forecast time through about 60 h and then begin to level off. The errors for the EP/CP are a little larger than for the ATL except at 120 h, but the behavior with time is very similar. Figure 5 also shows the percentage of cases when the D-SHIFOR forecasts differed from the SHIFOR5 forecasts, which occurs when the CLIPER5 track crosses land during the forecast period. By 120 h about 28% of the Atlantic cases included a land decay, but only about 9% of the Pacific cases did so. This is not surprising since
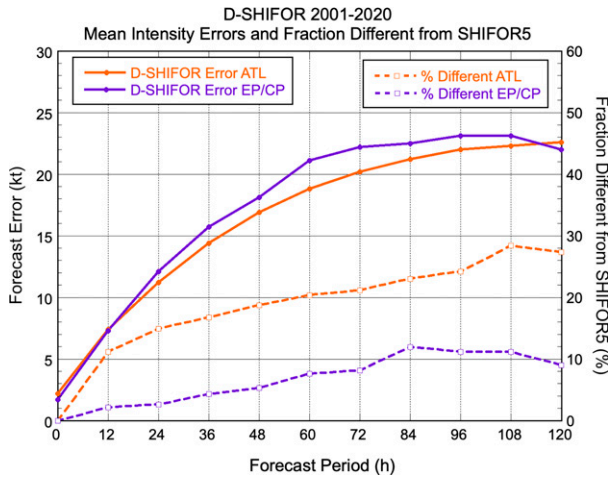
FIG. 5. Mean D-SHIFOR intensity errors for the ATL and EP/CP samples (solid lines) and the percentage of the cases where the D-SHIFOR and SHIFOR5 forecasts were different (dashed lines). The samples consist of cases from 2001 to 2020.

most eastern Pacific TCs move away from the Mexican coast, and in the central Pacific the Hawaiian Islands do not pose a very large target for the few TCs that do not dissipate before reaching them.

Figure 6 shows the improvement of D-SHIFOR over SHIFOR5 for the total sample and for the subsamples that only include cases when the forecasts are different (i.e., whenever the land-decay adjustment is invoked for D-SHIFOR). For the total sample, the inclusion of the inland decay based on the CLIPER5 track produces a modest error reduction for the ATL through 72 h and little change in the EP/CP. By 120 h, the inclusion of land effects slightly degrades the intensity forecasts, likely because the inaccuracy of the longer-term CLIPER5 track forecasts tends to poorly represent when a land decay correction is appropriate. When the D-SHIFOR and SHIFOR5 forecasts are different, however, Fig. 6 shows that the D-SHIFOR forecasts are much improved over SHIFOR5, through at least 72 h. It is also seen that improvements due to land decay are larger in the ATL than the EP/CP. This is probably because the EP/CP results are more sensitive to small errors in CLIPER5 tracks (because the Hawaiian Islands are small targets, and eastern Pacific TCs sometimes move close and parallel to the west coast of Mexico). For the subsamples, the differences between D-SHIFOR and SHIFOR5 are statistically significant at the 95% level at 12–72 h for the ATL and EP/CP, but the degradations at 120 h are not statistically significant. Collectively, these results indicate that D-SHIFOR is a better baseline than SHIFOR5 for assessing intensity forecast skill.

b. T-CLIPER

T-CLIPER was developed using data through 2011 so our verification period begins with 2012. There have been two model updates since 2012; the treatment of persistence in 2013, and an update of the developmental sample in 2020. To obtain a consistent and independent dataset for evaluation
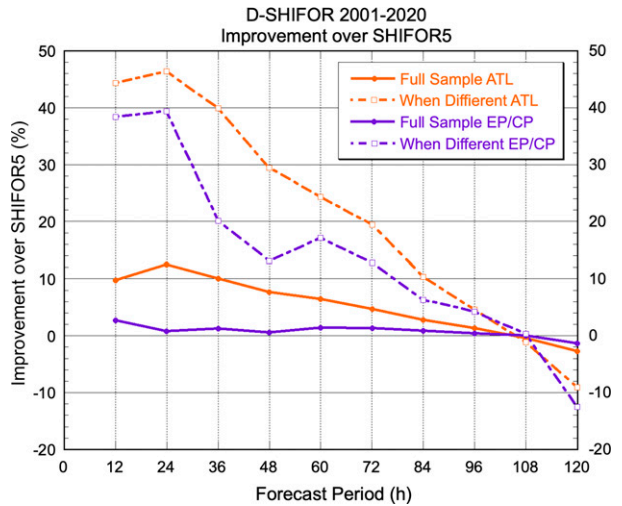

FIG. 6. Improvement of D-SHIFOR over SHIFOR5 for total 2001–20 ATL and EP/CP samples (solid lines) and for the subsamples that include only those cases when the D-SHIFOR and SHIFOR5 forecasts are different (dashed lines).

here, the 2013 version of T-CLIPER was rerun for the period 2012–20.

Figure 7 shows the T-CLIPER track errors through 120 h and the skill relative to CLIPER5 for the 2012–20 sample. The T-CLIPER errors are larger for the ATL than the EP/CP at all forecast times, which is also true for CLIPER5 track errors (not shown). For the EP/CP, the T-CLIPER mean track errors are within about 1% of those from CLIPER5, and none of the differences are statistically significant at the 95% level. For the AL, the T-CLIPER errors are about 5% larger than those from CLIPER5, and the differences through 72 h are statistically significant. If T-CLIPER were used as a replacement for CLIPER5 as a track skill baseline, there would be an apparent (artificial) skill improvement in tested forecast
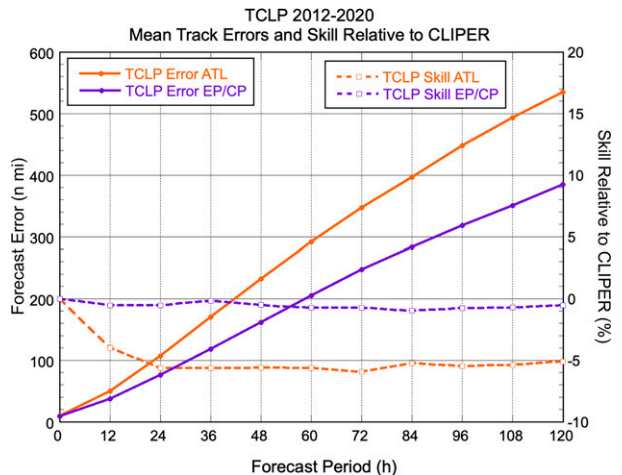

FIG. 7. T-CLIPER track forecast errors for 2012–20 (solid lines) and the skill (percent improvement relative to CLIPER5, dashed lines) for the ATL and EP/CP samples.
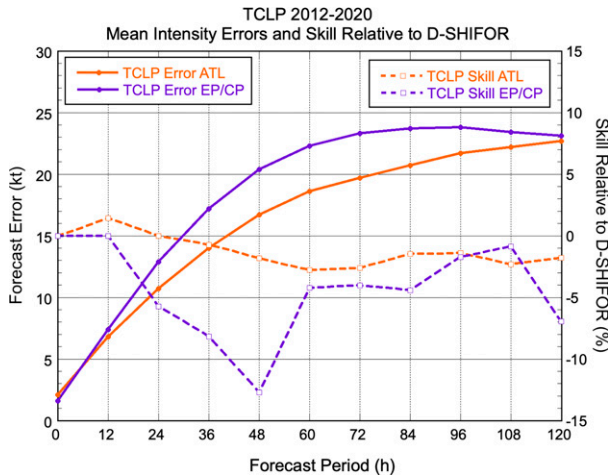
FIG. 8. T-CLIPER intensity forecast errors for 2012–20 (solid lines) and the skill (percent improvement relative to D-SHIFOR, dashed lines) for the ATL and EP/CP samples.

models or methods of a few percent just due to the slightly larger errors of the T-CLIPER model.

Figure 8 shows the mean T-CLIPER intensity errors and T-CLIPER skill relative to D-SHIFOR. Comparison with Fig. 5 shows that T-CLIPER behaves similarly to D-SHIFOR, in that EP/CP errors are larger than ATL errors initially but converge by 120 h. T-CLIPER intensity errors also level off with time, although perhaps not quite as quickly as with D-SHIFOR. The T-CLIPER skill relative to D-SHIFOR is less than ±3% for the ATL sample, and none of the differences between the two models were statistically significant. For EP/CP, T-CLIPER errors are about 13% larger than D-SHIFOR at 48 h and by smaller amounts at other forecast times. The differences between the T-CLIPER and D-SHIFOR errors for the EP/CP sample are statistically significant at 24–48 h. As noted above for track, if D-SHIFOR were replaced by T-CLIPER as the intensity skill baseline, there would be an apparent increase in forecast skill in tested forecast models or methods just due to the larger T-CLIPER errors, especially for shorter-range eastern Pacific forecasts.

T-CLIPER was updated for the 2020 season by adding cases from 2012 to 2018 to the developmental data, and reanalyzing the climatological motion and intensity growthrate fields. Table 2 shows that with the larger dataset, the e-folding time of the zonal component of motion decreased, while the e-folding time of the intensity growth rate increased. Results of dependent tests for cases from 2013 to 2019 showed that the new version did improve the performance relative to CLIPER5 and D-SHIFOR by a few percent. A comparison of the 2013 and 2020 versions of T-CLIPER for the 2020 season showed that the newer version improved track and intensity forecasts at most times, although the sample sizes were not large enough to show statistical significance. Several more years of independent cases will be needed to fully evaluate the performance of the updated version.

NHC has traditionally used CLIPER5 and D-SHIFOR as skill benchmarks for the evaluation of the more complex models and the official NHC forecasts (e.g., Cangialosi 2021). For example, a season with more storms in the deep tropics, where the environmental steering flow is relatively uniform, is typically associated with lower NHC official track forecast errors than a season that features TCs in the more complicated environments of the midlatitudes. Seasons that are active in the deep tropics also tend to be associated with relatively low CLIPER5 errors. To the extent that erratic or aclimatological behavior is difficult to anticipate by a forecaster or a model, the seasonal variations in CLIPER5 errors offer a convenient, if imperfect, way to normalize seasonal errors for forecast difficulty. This in turn is helpful to elucidate longer-term trends in forecast accuracy.

To determine if T-CLIPER can similarly be used as a measure of forecast difficulty, the annual average intensity and track errors were correlated with those from the NHC official forecasts and compared with the same correlations for D-SHIFOR and CLIPER5. Figure 9 shows that correlations with the NHC official forecast errors are positive at most forecast times in both basins, indicating that for this limited period the baseline model errors are indicators of annual track and intensity forecast difficulty (at least as experienced by NHC forecasters). There is some variability in the strength of the correlation for T-CLIPER compared with D-SHIFOR or CLIPER5, but overall, the results are comparable. It is interesting that the correlations are higher at the early and later forecast times, with minima near 48–72 h.

The cases with very low or negative correlations in Fig. 9 were examined in more detail and tended to be caused by an outlier point in a single year out of the nine years in each sample. To reduce the noise in the correlation analysis due to the small sample sizes, the data from all five forecast time periods (24, 48, 72, 96, and 120 h) were combined. Because the OFCL and baseline model errors tend to increase with forecast time, the correlation will be artificially inflated when the times are combined. To correct for that problem, the forecast errors for each model at each forecast time were standardized by subtracting the mean and dividing by the standard deviation before the forecast times were combined. The correlation coefficients between the OFCL and CLIPER5 or T-CLIPER for track and OFCL and D-SHIFOR or T-CLIPER for intensity were all positive for both basins with the combined samples. The correlation coefficients ranged from 0.22 to 0.72 and were fairly similar for OFCL versus CLIPER5 or T-CLIPER and OFCL versus D-SHIFOR or T-CLIPER. This result indicates that T-CLIPER can be used to assess forecast difficulty in a manner similar to CLIPER5 or D-SHIFOR.

NHC began experimental in-house 7-day track and intensity forecasts in 2012 and continued those through 2022, with a one-year break in 2017. Figure 9 shows that correlations of the NHC extended-range errors against T-CLIPER errors are comparable to those through 120 h for both basins. These results indicate that T-CLIPER accounts for annual forecast difficulty in a way that is similar to D-SHIFOR and CLIPER5 through 120 h, and that those relationships hold through 168 h. Thus, even though the average T-CLIPER errors are a

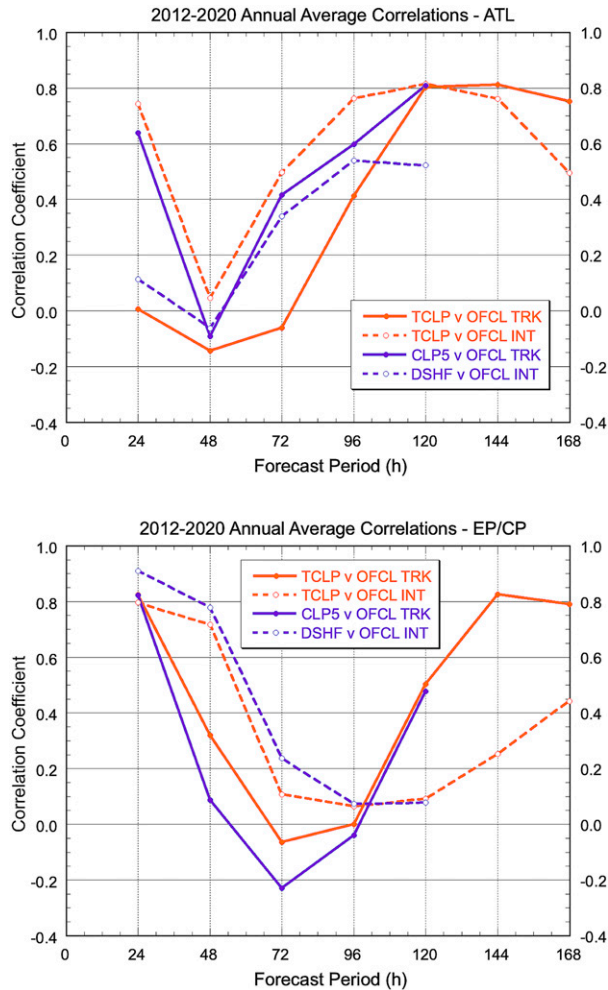FIG. 10. T-CLIPER track (solid lines) and intensity (dashed lines) forecast errors through 240 h.

FIG. 9. Correlations of the 2012–20 annual-average NHC OFCL forecast errors against errors from T-CLIPER, CLIPER5, and D-SHIFOR, for the (top) ATL and (bottom) EP/CP samples. Track error correlations are shown with solid lines, and intensity error correlations are shown in dashed lines. Correlations against T-CLIPER are in orange, while correlations against CLIPER5 and D-SHIFOR are in purple.

little larger than those of D-SHIFOR and CLIPER5 at some forecast times as described above, the availability of a consistent measure of forecast difficulty for time periods beyond 5 days is a significant advantage.

To get a feel for whether T-CLIPER could be used as a skill baseline at even longer lead times, Fig. 10 shows the T-CLIPER track and intensity errors through 240 h. The error evolution from 168 to 240 h is generally consistent with the errors at earlier leads, suggesting that it could be used as a baseline for extended-range forecast models. The track errors continue to increase at a roughly linear rate through 240 h, although the ATL sample becomes somewhat noisy by that time. The EP/CP intensity errors peak at 96 h and then slowly decrease after that. The ATL intensity errors start to level off after 48 h but then increase at a slightly faster rate around
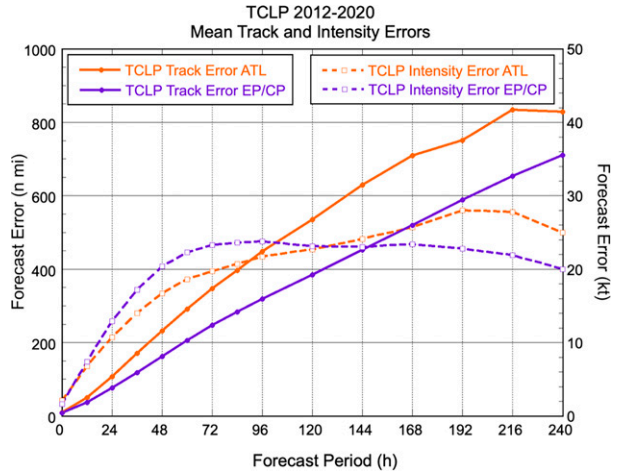
144 h, with the peak errors at ~200 h. The difference in the intensity error behavior between the ATL and the EP/CP samples may be related to TCs that intensify after recurvature, which is much more common in the Atlantic.

The T-CLIPER model for the JTWC basins was first run in real-time in 2016, but a significant change was implemented for 2018. Therefore, the T-CLIPER verification for the JTWC basins is restricted to 2018–20 using the real-time runs. J-North and J-South T-CLIPER forecast tracks are compared with the C120 model, the 120-h track climatology/persistence model for the western North Pacific (Aberson and Sampson 2003), and CLIP, the 72-h climatology/persistence model based on Neumann (1972) for the Southern Hemisphere. Similarly, J-North and J-South T-CLIPER intensity forecasts are compared with ST5D, the (Knaff et al. 2003) 120-h implementation of SHIFOR.

Figure 11 (top panel) shows the track verification results for the JTWC versions of T-CLIPER. The track errors are very similar for the J-North and J-South domains. The T-CLIPER track skill for J-North relative to the C120 model is close to zero, indicating that the newer model is a suitable replacement. The T-CLIPER track skill for the J-South domain is 20%–40% better than the much older CLIPER (CLIP) model through 72 h. Thus, the differences in T-CLIPER track skill between the two domains is an artifact of evaluating the skill relative to two different baselines. If T-CLIPER replaced CLIPER as a baseline for track skill for the J-South domain, there would be a considerable (apparent) loss of skill in tested forecast models or methods due to the smaller T-CLIPER track errors.

Figure 11 (bottom panel) shows the intensity verification results for the JTWC versions of T-CLIPER. Model errors in the J-North domain are larger than in J-South, and the errors for both domains level off after about 72 h. For the J-North domain, the T-CLIPER intensity errors relative to ST5D are close to zero at all forecast times, indicating that it could be used as a replacement baseline model for measuring intensity
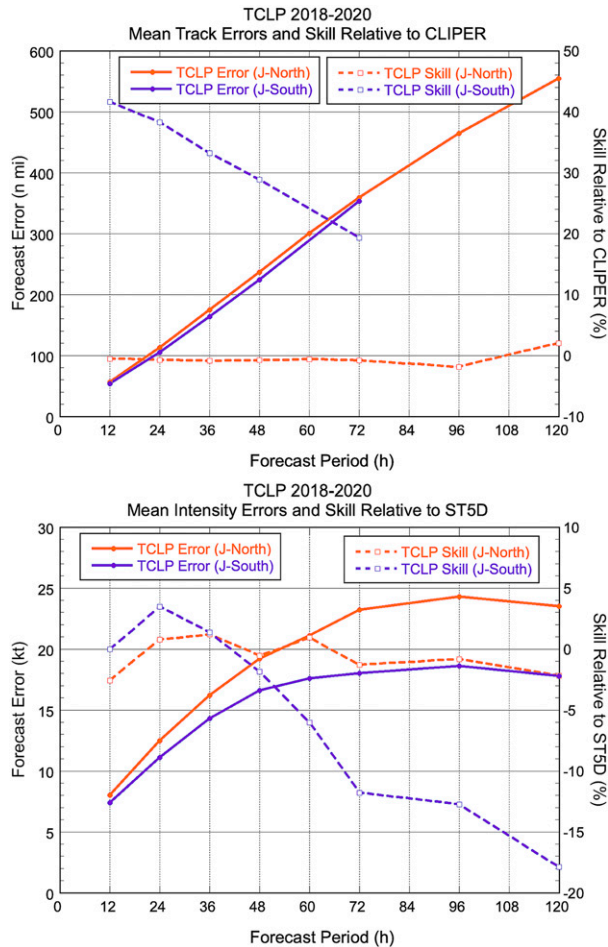
FIG. 11. (top) Track and (bottom) intensity verification of the J-North and J-South versions of T-CLIPER developed for the JTWC. Solid lines are the mean T-CLIPER errors and dashed lines show the percent improvement of T-CLIPER relative to JTWC's prior baseline models for track (C120 for J-North and CLIP for J-South) and intensity (ST5D). Note that CLIP only runs through 72 h, so the homogeneous comparison with T-CLIPER only extends to that time period.

forecast skill with little impact. In contrast, T-CLIPER intensity errors for the J-South domain are comparable to or a little smaller than ST5D through 36 h, but are larger by as much as 18% by 120 h. This indicates that there would be an apparent increase in skill in the longer-range forecasts if T-CLIPER replaced ST5D as the intensity skill baseline for the J-South domain.

The differences in the performance of T-CLIPER relative to ST5D in the J-South domain compared with the J-North domain in Fig. 11 (bottom panel) for the longer-range forecasts may be due to the interaction with strong SST gradients and land interaction. The SST gradients in the north Indian Ocean and western part of the western North Pacific basin where most of the TCs recurve, are much weaker than in the J-South domain. Thus, T-CLIPER track errors in the J-South domain would lead to a larger error in the estimate of the
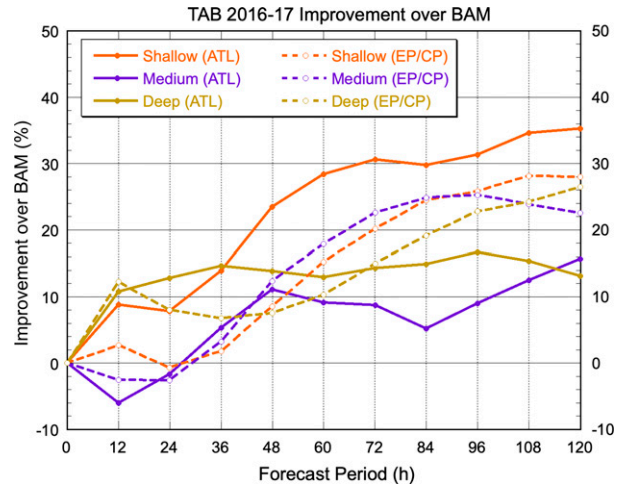

FIG. 12. The track forecast improvement of the TAB model over the BAM model for the three versions of each (shallow, medium, and deep) for the 2016 season and the first TC of 2017. The TAB errors for each version are relative to the corresponding version of BAM (e.g., TABS is compared to BAMS, etc.).

climatological growth rate than in the J-North domain. At the longer lead times, the use of the very inaccurate T-CLIPER tracks to estimate the J-South growth rate may be worse than only including the initial position as a predictor as in ST5D. This result is similar to the difference between D-SHIFOR and SHIFOR in Fig. 6, where the use of a CLIPER track in D-SHIFOR to determine the interaction with land degrades the intensity forecast after ~108 h compared with not including the CLIPER track forecast information in SHIFOR.

Beyond 120 h (not shown), T-CLIPER track errors continue to increase at a linear rate to about 700 and 800 n mi (1 n mi = 1.852 km) by 168 h in the J-North and J-South domains, respectively. T-CLIPER intensity errors for the J-North domain decrease slightly to about 20 kt by 168 h, while the J-South intensity errors remain roughly constant from 120 to 168 h.

c. TAB

The TAB model driven with GFS input was first run as part of the guidance suite in 2016. Here we report on the verification of the real-time TAB forecasts as recorded in the ATCF a-decks for the period 2016–20.

A comparison of TAB track forecasts with BAM is shown in Fig. 12. Because the BAM model was discontinued early in the 2017 season, there is just a little over one full season of direct comparisons between the two models. Figure 12 shows the percent improvement of TAB relative to BAM for the deep, medium, and shallow versions of each model for the ATL and EP/CP samples. TAB forecasts show improvement over BAM at almost all forecast times in both basins, with improvements as high as 35%, a result consistent with the 2010–15 cases that were used to select the TAB parameters. Because the samples in Fig. 12 are small, only the TABS

2016-20 Average Track Forecast Error
Atlantic Basin

2016-20 Average Track Forecast Error
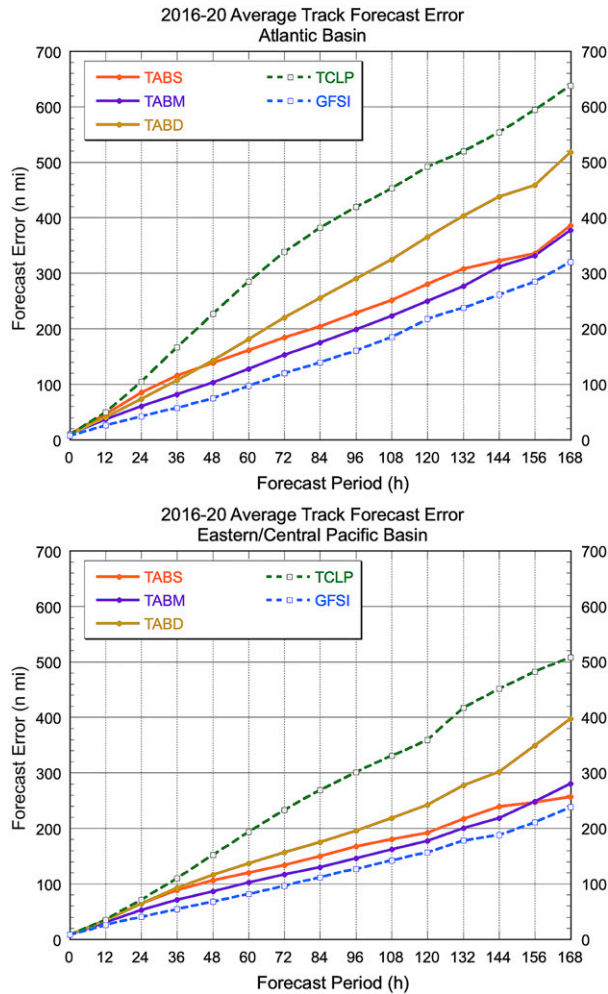### Eastern/Central Pacific Basin

FIG. 13. TAB model track errors for the 2016–20 operational forecasts for the (top) ATL and (bottom) EP/CP samples. The T-CLIPER track errors and those from the GFS (GFSI) are also included for reference.

improvements for the ATL sample at 48–72 h are statistically significant at the 95% level. However, the differences at several other forecast times for TABD and TABM satisfy the 90% significance level. These results suggest that TAB likely could replace BAM with no loss of accuracy. Although not shown, the TAB model also demonstrates a similar divergence of tracks among its three versions in sheared environments as typically occurs with the three versions of BAM.

Figure 13 shows TAB errors through 168 h for operational runs for the period 2016–20. T-CLIPER and operationally available GFS track errors (GFSI) are also shown for reference. ATL errors are larger than those for EP/CP with each version of TAB, but the relative performances are similar across the basins. TABD errors are much larger than those of TABM and TABS, with TABM being the best of the three at nearly all forecast times. All versions of TAB perform much better than T-CLIPER, indicating positive forecast skill. As a group, the TAB errors are larger than the GFSI errors, which

is to be expected given the simplicity of the TAB model and its assumption of a fixed steering layer throughout the forecast. Surprisingly, however, the TABM errors are not much larger than GFSI on average, which confirms that the substitution of TABM as input for the statistical intensity models when the NHC official forecast is not available is a reasonable choice. The following section on the analysis of steering layers offers some insight on the relative performance of the three TAB models.

Versions of TAB driven by ECMWF wind fields (Table 1) were added in 2018. Operational ECMWF-based TAB forecasts for the 2018–20 seasons were also verified and compared with the GFS versions. The ECMWF TAB errors (not shown) were generally similar to those from the GFS version, with the medium version (TBME) being the best model at most forecast times in both basins, and all versions having skill relative to T-CLIPER. The primary difference between the GFS and ECMWF versions of TAB was that TBDE had lower errors than TABD in both basins, especially at the longer ranges where the TBDE errors were up to 30% smaller than the TABD errors by day 7. However, the errors of the medium versions of TAB for the GFS and ECMWF were comparable (within 15% or less of each other), with the TABM being better for the EP/CP sample and TBME being better for the ATL sample.

d. Steering layer analysis

As mentioned above, one of the purposes of TAB is to provide forecasters with an estimate of the vertical shear in the steering flow by comparing the tracks using three different layers. Having versions with different vertical layers also provides forecasters with guidance on how the track might differ depending on the depth of the TC's circulation.

To provide more insight into how various factors might affect TC steering, Eqs. (9) and (10) were used to find the optimal weights for the ATL and EP/CP samples for all cases in the SHIPS developmental database for 1982–2020; the results were then stratified by TC intensity, latitude, and basin. The environmental wind components at the mandatory pressure levels from 1000 to 100 hPa, u_i and v_i, were determined from the GFS analyses averaged from 0 to 500 km from the TC center, and the storm motion vectors were computed from best track data. The calculation did not include the 925-hPa level since the SHIPS GFS database does not include that level for some of the early years.

Figure 14 (left panel) shows the optimal weights for each basin and the weights for the standard mass-weighted average for the total samples. The optimal weights for ATL and EP/CP are qualitatively similar, with values lower than the mass-weighted versions from 100 to about 500 hPa and larger values from 500 to 1000 hPa. Part of the reason for the higher weights in the lower levels is because the pressure layers are thicker. Normalizing the optimal weights by the mass weights adjusts for the differences in thickness, so the weights at all levels are directly comparable, as shown in the right panel of Fig. 14. It is seen that the 850-hPa (700-hPa) level contributes the most to the ATL (EP/CP) steering and the upper levels
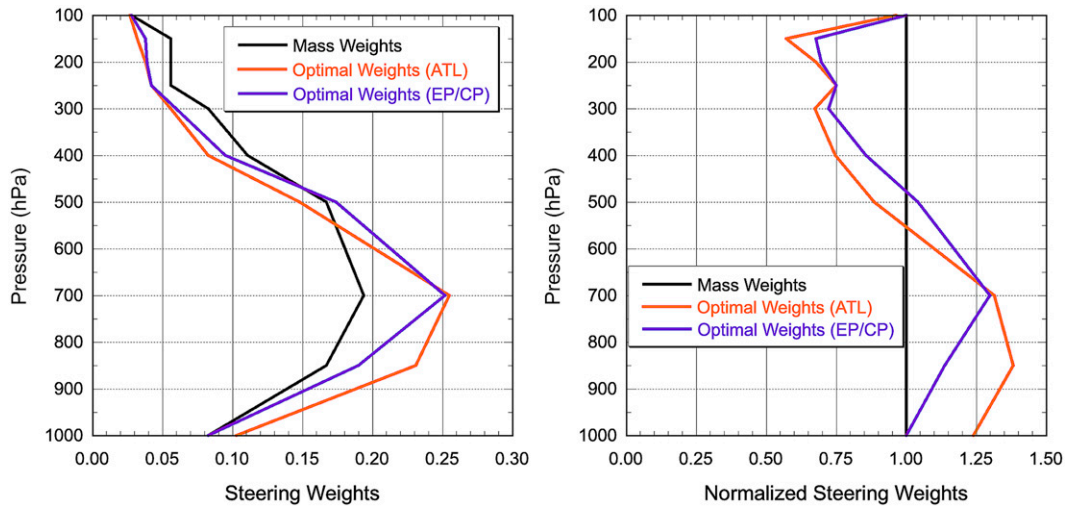
FIG. 14. Vertical weights for the 1000–100-hPa levels for a mass-weighted mean and the optimal values that best match the observed TC motion for (left) the ATL and EP/CP samples and (right) the weights normalized by the mass weights. The optimal values were determined for the average of all the cases in the SHIPS developmental database from 1982 to 2020.

contribute much less to the steering than the lower levels in both basins. These results are generally consistent with those from previous studies of TC steering levels (e.g., Velden and Leslie 1991), but the formulation in Eqs. (9) and (10) does not require calculating the steering from arbitrary levels and layers and then comparing the results. The lesser contribution to the steering from the upper levels helps to explain why TABM had lower track errors than TABD and TABS.

Equation (11) provides a way to combine the optimal weights into a single parameter to allow for easy comparison with TAB model results and compare subsets of cases stratified by physical parameters. Generally speaking, PSL_O is larger when the steering is from shallower layers. Applying Eq. (11) to the optimal steering weights shows that PSL_O is 627 and 591 hPa for the ATL and EP/CP samples, respectively. Those are closest to the PSL_M value of 625 hPa for TABM. That result is again consistent with Fig. 13, which shows that TABM had the smallest errors at most forecast times.

To examine the spatial variability of PSL_O, the Barnes objective analysis was performed for the combined ATL and EP/CP PSL_O sample. Figure 15 shows that PSL_O is closest to the TABM value over most of the domain, but there is some variability. The PSL_O value is lower in the deep tropics, indicating that the upper levels contribute more to the steering in that region, which is consistent with a higher tropopause than in higher latitudes. The environmental vertical shear is also usually less in lower latitudes, which would allow TCs to have deeper cyclonic circulations and be steered more by the higher levels. The largest PSL_O values, indicating the shallowest steering layer, are in the eastern and central Pacific north of about 20°N. The SSTs in that region are very cold, and TCs are often in the dissipating stage after they move through the strong SST gradient that normally is present from 15° to 20°N between the west coast of Mexico and Hawaii. Figure 15 suggests that an improved version of TAB could be developed if the weights
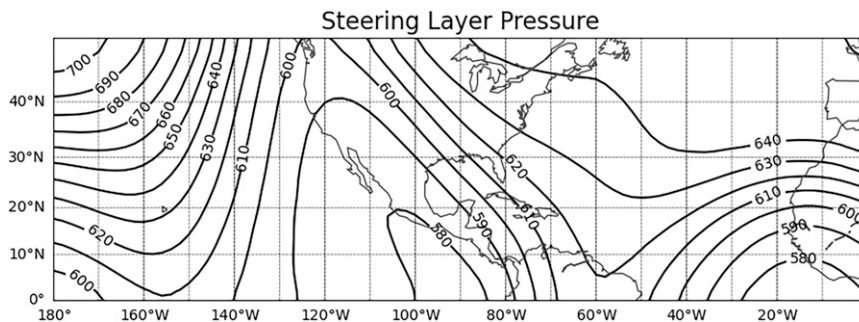

FIG. 15. Spatial variability of the average steering layer pressure for the combined 1982–2020 ATL and EP/CP samples. There was little data north of about 30°N for longitudes between about 100° and 140°W, so the values there are mostly based on extrapolation from lower latitudes.

were a function of location, but that is left as a topic for future study.

The optimal steering weights and corresponding steering layer pressures PSL_O were calculated for ATL and EP/CP subsamples stratified by maximum wind. For the EP/CP sample, PSL_O showed a fairly linear decrease from a value of 613 hPa for TCs with a maximum wind of 30 kt or less, to a value of 558 hPa for TCs with maximum winds greater than 110 kt. For the Atlantic sample, the decrease was more gradual, with an average PSL_O value of 617 hPa for TCs with maximum winds of 30 kt or less, to a value of 598 hPa for TCs with maximum winds of greater than 110 kt. Also, for the ATL sample, the average PSL_O value was nearly constant for TCs with maximum winds from 20 to 100 kt. Thus, stronger TCs tend to be steered by a deeper layer, especially in the EP/CP, but that variation is less than the variability with latitude seen in Fig. 15. The dependence on the increase in the depth of the best steering layer with increasing intensity is generally consistent with the results of Velden and Leslie (1991) for the Australian region, but the nearly constant optimal steering layer for ATL cases with maximum winds of 20–100 kt indicates that other factors, such as convective activity, may be just as important in determining the steering layer.

5. Conclusions

NHC and CPHC use a variety of models as guidance for their operational TC track, intensity, and wind structure forecasts and as baselines for evaluation of forecast skill. A subset of the simpler models, maintained and updated by NHC and collectively referred to as the guidance suite, has been described here. Three members of the guidance suite—D-SHIFOR, T-CLIPER, and TAB—were described in detail since those have not been documented elsewhere in the refereed literature.

D-SHIFOR adds a correction for landfall to the simple statistical intensity model SHIFOR using the CLIPER5 track and a climatological inland decay rate. D-SHIFOR improved on SHIFOR for forecasts through about 96 h for the ATL and EP/CP basins. By 120 h there was no improvement because of the inaccuracy of the CLIPER5 tracks for determining when the TC center is over land. Thus, D-SHIFOR is a suitable replacement for SHIFOR that can be used for TCs over land or water.

T-CLIPER was developed to extend the CLIPER5 and D-SHIFOR baseline models beyond 120 h. A trajectory approach was used, in which shorter-range forecasts are primarily determined by persistence of the initial motion or intensity tendency, while the longer-range forecasts are primarily determined from climatological motion or intensity tendencies. T-CLIPER track and intensity errors were generally similar to those from the older track and intensity baseline models, with a few exceptions. For the ATL, T-CLIPER track errors were a few percent larger than those from CLIPER5 at most times out to 120 h. For the EP/CP the T-CLIPER intensity errors were up to 13% larger than the D-SHIFOR errors through 36 h but were comparable after that time. The T-CLIPER model was updated for the 2020 season, which should help to make it

more comparable to CLIPER5 and D-SHIFOR, and it could potentially be updated again after a few more years of developmental data are available.

T-CLIPER was also developed for JTWC's AORs. It provides a 7-day baseline for track and intensity, a particularly important capability improvement for Southern Hemisphere TCs. However, in the 3-yr verification, T-CLIPER intensity errors were much larger than those from the older ST5D errors for JTWC's Southern Hemisphere forecasts. This may be due to the strong SST gradients over most of the Southern Hemisphere TC basin, which makes T-CLIPER intensity forecasts more sensitive to errors in the T-CLIPER track forecasts.

The main advantage of T-CLIPER is that it can be run to any forecast length and the error characteristics were found to be reasonable through at least 10 days. It was also shown that the correlations of annual-average T-CLIPER errors with NHC official forecast errors was similar to those of CLIPER5 and D-SHIFOR, meaning that T-CLIPER can be used as a measure of annual forecast difficulty in the same way as the older baseline models but extended beyond 120 h. The correlation of the T-CLIPER annual mean errors with the NHC forecast errors persists through 168 h, confirming its applicability as an extended range baseline.

The TAB model replaced the BAM track forecast model, both of which follow a trajectory from a global forecast model with a small correction for the beta effect. TAB uses a grid-point method for the horizontal smoothing to determine the steering flow, rather than a spherical harmonic truncation. Like BAM, three versions of TAB are run with deep, medium and shallow vertical layers used to estimate the steering flow. Both models were run in 2016 and early 2017 and results showed that TABM had track errors after about 36 h that were 20%–30% smaller than those for BAMM. These results showed that TAB was a suitable replacement for BAM. The TAB model track forecasts have considerable skill compared with T-CLIPER and require nearly trivial computation resources.

To provide insight into the vertical steering layers, a method to determine optimal vertical weights that minimize the difference between the steering estimated from GFS model analyses and the current storm motion was developed. Results show the optimal steering is obtained by weighting the lower troposphere much more than the upper troposphere. The steering layer also tends to be deeper for low-latitude TCs and stronger TCs. However, the dependence of the steering layer on the maximum wind is weak for Atlantic TCs with maximum winds below 100 kt. These results echo those from previous studies of TC steering, but with the new formulation, it is not necessary to compare results from steering calculated over arbitrary layers and the weights can deviate from mass-weighted values.

The NHC guidance suite will continue to evolve. For example, TAB could be improved by adjusting the weights based on the optimal steering layer analyses. In addition, Dong and Neumann (1986) showed that there is some advantage to including pressure levels below 850 hPa and above 200 hPa to estimate TC steering, so additional levels could be used to

improve TAB. New techniques that use advanced statistical methods such as machine learning could be used instead of simpler linear regression approaches, many of which are already being developed for intensity forecasting (e.g., Xu et al. 2021; Su et al. 2020). In addition, the simple TAB track model and the statistical–dynamical intensity models that are part of the NHC guidance suite have applications beyond short-term deterministic forecasting. For example, Shan and Yu (2020) developed a trajectory model very similar to TAB for climatological studies of global TC tracks and possible impacts of climate change and Lin et al. (2020) applied simplified track and intensity forecast models to develop a large-member TC ensemble system for estimating forecast uncertainty. Thus, the simplified models in the NHC guidance suite will continue to provide useful forecast information that complements dynamical models.

Data availability statement. Atlantic, eastern North Pacific, and central North Pacific operational model and best track data were obtained from the ATCF system a-deck and b-deck archives at ftp://ftp.nhc.noaa.gov/atcf. ATCF b-decks for the western North Pacific, Indian Ocean, and Southern Hemisphere basins were from https://www.metoc.navy.mil/jtwc/jtwc.html?best-tracks. All other datasets, including regenerated forecasts prepared for this study, are available upon request from the first author.

APPENDIX

List of Acronyms and Initialisms

AOR	Area of responsibility
ATCF	Automated Tropical Cyclone Forecast System
ATL	Atlantic basin
BAM	Beta and Advection Model
BAMD	Beta and Advection Model, deep version
BAMM	Beta and Advection Model, medium version
BAMS	Beta and Advection Model, shallow version
CIRA	Cooperative Institute for Research in the Atmosphere
CLIP	ATCF identifier for CLIPER
CLIPER	Climatology and Persistence model
CLIPER5	Five-day Climatology and Persistence model
CLP5	ATCF identifier for CLIPER5
CPHC	Central Pacific Hurricane Center
DSHP	ATCF identifier for D-SHIPS
DSWR	D-SHIPS Wind Radii model
DTOP	ATCF identifier for DTOPS
DTOPS	Deterministic to Probabilistic Statistical model
D-SHIFOR	Decay-SHIFOR
D-SHIPS	Decay-SHIPS
ECMWF	European Centre for Medium-Range Weather Forecasts
EFD	*e*-folding distance
EP/CP	Eastern North Pacific and central North Pacific basins
GFS	Global Forecast System
GFSI	ATCF identifier for interpolated version of GFS
HCCA	HFIP Corrected Consensus Approach model
HFIP	Hurricane Forecast Improvement Project
HWRF	Hurricane Weather Research and Forecasting Model
JTWC	Joint Typhoon Warning Center
LGEM	Logistical Growth Equation Model
MPI	Maximum potential intensity
NCEP	National Centers for Environmental Prediction
NCO	NCEP Central Operations
NCODA	Navy's Coupled Ocean Data Assimilation
NNIB	Neural Network Intensity Baseline model
NNIC	Neural Network Intensity Consensus model
OCD5	ATCF identifier for combined CLIPER5/D-SHIFOR output
OFCL	ATCF identifier for NHC/CPHC/JTWC official forecasts
OHC	Ocean heat content
QP	Quasi-production
RFC	Request for change
RI	Rapid intensification
RIOB	RI Operational Bayesian model
RIOC	RI Operational Consensus model
RIOD	RI Operational Discriminate Analysis model
RIOL	RI Operational Logistic Regression model
RoI	Radius of influence
SHF5	ATCF identifier for SHIFOR5
SHIFOR	Statistical Hurricane Intensity Forecast model
SHIFOR5	Five-day Statistical Hurricane Intensity Forecast model
SHIP	ATCF identifier for SHIPS
SHIPS	Statistical Hurricane Intensity Prediction Scheme
SHIPS-RII	SHIPS rapid intensification index
SST	Sea surface temperature
ST5D	ATCF identifier for JTWC 5-day implementation of SHIFOR
TAB	Trajectory and Beta model
TABD	ATCF identifier for TAB, deep version
TABM	ATCF identifier for TAB, medium version
TABS	ATCF identifier for TAB, shallow version
TBDE	ATCF identifier for TAB, deep version using ECMWF data

TBME	ATCF identifier for TAB, medium version using ECMWF data
TBSE	ATCF identifier for TAB, shallow version using ECMWF data
TC	Tropical cyclone
TCLP	ATCF identifier for T-CLIPER
T-CLIPER	Trajectory Climatology and Persistence model
WCOSS	Weather and Climate Operational Supercomputer System
XTRP	ATCF identifier for a track forecast based solely on extrapolation of the current observed storm motion

REFERENCES

Aberson, S. D., 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.

——, and C. R. Sampson, 2003: On the predictability of tropical cyclone tracks in the Northwest Pacific basin. *Mon. Wea. Rev.*, **131**, 1491–1497, https://doi.org/10.1175/1520-0493(2003)131<1491:OTPOTC>2.0.CO;2.

Banzon, V., T. M. Smith, M. Steele, B. Huang, and H. M. Zhang, 2020: Improved estimation of proxy sea surface temperature in the Arctic. *J. Atmos. Oceanic Technol.*, **37**, 341–349, https://doi.org/10.1175/JTECH-D-19-0177.1.

Barnes, S. L., 1964: A technique for maximizing details in numerical weather map analysis. *J. Appl. Meteor.*, **3**, 396–409, https://doi.org/10.1175/1520-0450(1964)003<0396:ATFMDI>2.0.CO;2.

Bister, M., and K. A. Emanuel, 2002: Low frequency variability of tropical cyclone potential intensity. 1. Interannual to interdecadal variability. *J. Geophys. Res.*, **107**, 4801, https://doi.org/10.1029/2001JD000776.

Cangialosi, J. P., 2021: National Hurricane Center forecast verification report: 2020 hurricane season. NOAA/NHC, 77 pp., https://www.nhc.noaa.gov/verification/pdfs/Verification_2020.pdf.

Cummings, J. A., 2005: Operational multivariate ocean data a similation. *Quart. J. Roy. Meteor. Soc.*, **131**, 3583–3604, https://doi.org/10.1256/qj.05.105.

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

——, 2010: Tropical cyclone intensity change predictability estimates using a statistical-dynamical model. *29th Conf. on Hurricanes and Tropical Meteorology*, Tucson, AZ, Amer. Meteor. Soc., 9C.5, https://ams.confex.com/ams/29Hurricanes/techprogram/paper_167916.htm.

——, 2021: A new framework for statistical-dynamical tropical cyclone intensity forecast models. *34th Conf. on Hurricanes and Tropical Meteorology*, Online, Amer. Meteor. Soc., 8C.4, https://ams.confex.com/ams/34HURR/meetingapp.cgi/Paper/373073.

——, and J. M. Gross, 2003: Evolution of tropical cyclone forecast models. *Hurricane! Coping with Disaster: Progress and Challenges Since Galveston, 1900*, R. Simpson et al., Eds., 1st ed. Amer. Geophys. Union, 103–126.

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

——, J. A. Knaff, and J. Kaplan, 2006: On the decay of tropical cyclone winds crossing narrow landmasses. *J. Appl. Meteor.*, **45**, 491–499, https://doi.org/10.1175/JAM2351.1.

——, J. L. Franklin, M. J. Onderlinde, and J. Kaplan, 2021: Operational forecasting of tropical cyclone rapid intensification at the National Hurricane Center. *Atmosphere*, **12**, 683, https://doi.org/10.3390/atmos12060683.

Dong, K., and C. J. Neumann, 1986: The relationship between tropical cyclone motion and environmental geostrophic flows. *Mon. Wea. Rev.*, **114**, 115–122, https://doi.org/10.1175/1520-0493(1986)114<0115:TRBTCM>2.0.CO;2.

Dunion, J., J. Kaplan, and A. Schumacher, 2019: Improvement to the Tropical Cyclone Genesis Index (TCGI). Final Rep. NOAA Joint Hurricane Testbed, NA15OAR4590201, 16 pp., https://www.nhc.noaa.gov/jht/15-17reports/Dunion_201_Schumacher_202_progress_reportFINAL_rev030619.pdf.

Franklin, J. L., and M. DeMaria, 1992: The impact of omega dropwindsonde observations on barotropic hurricane track forecasts. *Mon. Wea. Rev.*, **120**, 381–391, https://doi.org/10.1175/1520-0493(1992)120<0381:TIOODO>2.0.CO;2.

Halperin, D. J., R. E. Hart, H. E. Fuelberg, and J. H. Cossuth, 2017: The development and evaluation of a statistical-dynamical tropical cyclone genesis guidance tool. *Wea. Forecasting*, **32**, 27–46, https://doi.org/10.1175/WAF-D-16-0072.1.

Harris, L., X. Chen, W. M. Putman, L. Zhou, and J. Chen, 2021: A scientific description of the GFDL finite-volume cubed-sphere dynamical core. NOAA Tech. Memo. OAR GFDL, 2021-001, 109 pp., https://doi.org/10.25923/6nhs-5897.

Holland, G. J., 1983: Tropical cyclone motion: Environmental interaction plus a beta effect. *J. Atmos. Sci.*, **40**, 328–342, https://doi.org/10.1175/1520-0469(1983)040<0328:TCMEIP>2.0.CO;2.

Jarvinen, B. R., and C. J. Neumann, 1979: Statistical forecasts of tropical cyclone intensity. NOAA Tech. Memo NWS NHC-10, 22 pp.

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

——, and ——, 2001: On the decay of tropical cyclone winds after landfall in the New England area. *J. Appl. Meteor. Climatol.*, **40**, 280–286, https://doi.org/10.1175/1520-0450(2001)040<0280:OTDOTC>2.0.CO;2.

Klein, P. M., P. A. Harr, and R. L. Elsberry, 2000: Extratropical transition of western North Pacific tropical cyclones: An overview and conceptual model of the transformation stage. *Wea. Forecasting*, **15**, 373–395, https://doi.org/10.1175/1520-0434(2000)015<0373:ETOWNP>2.0.CO;2.

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, https://doi.org/10.1175/1520-0434(2003)018<0080:SDTCIF>2.0.CO;2.

——, C. R. Sampson, and G. Chirokova, 2017: A global statistical-dynamical tropical cyclone wind radii forecast scheme. *Wea. Forecasting*, **32**, 629–644, https://doi.org/10.1175/WAF-D-16-0168.1.

Landsea, C. W., and J. L. Franklin, 2013: Atlantic hurricane database uncertainty and presentation of a new database format. *Mon. Wea. Rev.*, **141**, 3576–3592, https://doi.org/10.1175/MWR-D-12-00254.1.

Lin, J., K. Emanuel, and J. L. Vigh, 2020: Forecasts of hurricanes using large-ensemble outputs. *Wea. Forecasting*, **35**, 1713–1731, https://doi.org/10.1175/WAF-D-19-0255.1.

Mainelli, M., M. DeMaria, L. K. Shay, and G. Goni, 2008: Application of oceanic heat content estimation to operational forecasting of recent Atlantic category-5 hurricanes. *Wea. Forecasting*, **23**, 3–16, https://doi.org/10.1175/2007WAF2006111.1.

Marks, D. G., 1992: The beta and advection model for hurricane track forecasting. NOAA Tech. Memo. NWS NMC 70, 89 pp., https://repository.library.noaa.gov/view/noaa/7184.

Neumann, C. J., 1972: An alternate to the HURRAN (Hurricane Analog) tropical cyclone forecast system. NOAA Tech. Memo. NWS SR-62, 24 pp., https://repository.library.noaa.gov/view/noaa/3605.

Reynolds, R. W., and T. M. Smith, 1994: Improved global sea surface temperature analyses using optimum interpolation. *J. Climate*, **7**, 929–948, https://doi.org/10.1175/1520-0442(1994)007<0929:IGSSTA>2.0.CO;2.

Sampson, C. R., and A. J. Schrader, 2000: The Automated Tropical Cyclone Forecasting System (version 3.2). *Bull. Amer. Meteor. Soc.*, **81**, 1231–1240, https://doi.org/10.1175/1520-0477(2000)081<1231:TATCFS>2.3.CO;2.

Sardeshmukh, P. D., and B. I. Hoskins, 1984: Spatial smoothing on the sphere. *Mon. Wea. Rev.*, **112**, 2524–2529, https://doi.org/10.1175/1520-0493(1984)112<2524:SSOTS>2.0.CO;2.

Shan, K., and X. Yu, 2020: A simple trajectory model for climatological study of tropical cyclones. *J. Climate*, **33**, 7777–7786, https://doi.org/10.1175/JCLI-D-20-0285.1.

Simon, A., A. B. Penny, M. DeMaria, J. L. Franklin, R. J. Pasch, E. N. Rappaport, and D. A. Zelinsky, 2018: A description of the real-time HFIP Corrected Consensus Approach (HCCA) for tropical cyclone track and intensity guidance. *Wea. Forecasting*, **33**, 37–57, https://doi.org/10.1175/WAF-D-17-0068.1.

Su, H., L. Wu, J. H. Jiang, R. Pai, A. Lui, A. J. Zhai, P. Tavallali, and M. DeMaria, 2020: Applying satellite observations of tropical cyclone internal structures to rapid intensification forecast with machine learning. *Geophys. Res. Lett.*, **47**, e2020GL089102, https://doi.org/10.1029/2020GL089102.

Tallapragada, V., L. Bernardet, M. K. Biswas, S. Gopalakrishnan, Y. Kwon, Q. Liu, and X. Zhang, 2014: Hurricane Weather Research and Forecasting (HWRF) model: 2013 scientific documentation. HWRF Development Testbed Center Tech. Rep., 99 pp.

Velden, C. S., and L. M. Leslie, 1991: The basic relationship between tropical cyclone intensity and the depth of the environmental steering layer in the Australian region. *Wea. Forecasting*, **6**, 244–253, https://doi.org/10.1175/1520-0434(1991)006<0244:TBRBTC>2.0.CO;2.

Xu, W., K. Balaguru, A. August, N. Lalo, N. Hodas, M. DeMaria, and D. Judi, 2021: Deep learning experiments for tropical cyclone intensity forecasts. *Wea. Forecasting*, **36**, 1453–1470, https://doi.org/10.1175/WAF-D-20-0104.1.