

A Simple Bias and Uncertainty Scheme for Tropical Cyclone Intensity Change Forecasts

BENJAMIN C. TRABING,^a K. D. MUSGRAVE,^a M. DEMARIA,^a AND E. BLAKE^b

^a *CIRA/CSU, Fort Collins, Colorado*

^b *NOAA/NWS/National Hurricane Center, Miami, Florida*

(Manuscript received 19 April 2022, in final form 9 August 2022)

ABSTRACT: To better forecast tropical cyclone (TC) intensity change and understand forecast uncertainty, it is critical to recognize the inherent limitations of forecast models. The distributions of intensity change for statistical–dynamical models are too narrow, and some intensity change forecasts are shown to have larger errors and biases than others. The Intensity Bias and Uncertainty Scheme (IBUS) is developed in an intensity change framework, which estimates the bias and the standard deviation of intensity forecast errors. The IBUS is developed and applied to the Decay Statistical Hurricane Intensity Prediction Scheme (DSHP), the Logistic Growth Equation Model (LGEM), and official National Hurricane Center (NHC) forecasts (OFCL) separately. The analysis uses DSHP, LGEM, and OFCL forecasts from 2010 to 2019 in both the Atlantic and east Pacific basins. Each IBUS contains both a bias correction and forecast uncertainty estimate that is tested on the training dataset and evaluated on the 2020 season. The IBUS is able to reduce intensity biases and improve forecast errors beyond 120 h in each model and basin relative to the original forecasts. The IBUS is also able to communicate forecast uncertainty that explains ~7%–11% of forecast variance at 48 h for DSHP and LGEM in the Atlantic. Better performance is found in the east Pacific at 96 h where the IBUS explains up to 30% of the errors in DSHP and 14% of the errors for LGEM. The IBUS for OFCL explains 9%–13% of the 48-h forecast uncertainty in the Atlantic and east Pacific with up to 30% variance explained for east Pacific forecasts at 96 h. IBUS for OFCL has the capability to provide intensity forecast uncertainty similar to the “cone of uncertainty” for track forecasts.

KEYWORDS: Hurricanes/typhoons; Bias; Uncertainty; Forecast verification/skill; Statistical forecasting

1. Introduction

Forecasting tropical cyclone (TC) intensity change remains a challenge for forecasters because of the complex interactions between the convective, mesoscale, and synoptic scales in addition to the reliance on an accurate track forecast (Kaplan et al. 2010). Although track forecasts have been clearly improving over the last several decades, intensity forecasts have been improving at a lower rate (DeMaria et al. 2014). Since 2010, Cangialosi et al. (2020) found that the skill of intensity guidance has increased due to improvements in the numerical weather prediction (NWP) models stemming from the Hurricane Forecast Improvement Project (HFIP; Gall et al. 2013), consensus aids, rapid intensification (RI) aids, and the ability of forecasters to synthesize the data. In addition to improving hurricane track and intensity forecasts, estimates of uncertainty of those parameters also needs to be improved in order to communicate risk (Marks et al. 2019). The National Hurricane Center (NHC) uses the “cone of uncertainty,” which is a static estimate of uncertainty using the 67th percentile of track errors from the past five years, to provide an empirical estimate of track errors. At this time, no such uncertainty product is provided for intensity forecasts by NHC. Although ensembles have become an important source for uncertainty estimation, estimating uncertainty in deterministic models is inherently more difficult. It is the goal of this manuscript to offer a simple

technique to both improve biases in deterministic intensity forecasts and assess the forecast dependent uncertainty.

When making track forecasts, forecasters at NHC utilize the prediction of consensus tropical cyclone track errors provided by the Goerss predicted consensus error (GPCE; Goerss 2007). GPCE was expanded for intensity forecasts to provide estimates for the lower and upper bounds for intensity forecasts to be correct at the 67th percentile, which is also utilized by forecasters in real time (Goerss and Sampson 2014). The Monte Carlo Wind Speed Probability (MC-WSP) model uses the GPCE consensus track spread and the NHC forecast to estimate the probability that 34-, 50-, and 64-kt winds ($1 \text{ kt} \approx 0.51 \text{ m s}^{-1}$) will occur, which is able to capture some estimate of forecast uncertainty at those wind speed thresholds (DeMaria et al. 2013). In contrast to the track, the MC-WSP model uses only climatological intensity errors and the NHC forecast to provide uncertainty for the intensity forecast. Bhatia and Nolan (2013) found that individual forecast model errors have a dependence on select predictors, which led to the development of the Prediction of Intensity Model Error (PRIME) model (Bhatia and Nolan 2015). Bhatia et al. (2017) showed that PRIME was able to skillfully predict model error using a stepwise multiple linear regression framework to improve consensus forecasts. PRIME was never put into operations and is now defunct due to its dependence on the Geophysical Fluid Dynamics Laboratory (GFDL) model, which was retired in 2017. Although GPCE and PRIME predict consensus model errors to improve forecaster guidance, GPCE and PRIME do not produce error estimates for NHC forecasts that can potentially

Corresponding author: Benjamin Trabling, btrabling@colostate.edu

TABLE 1. The sample size of forecasts in the Atlantic and east Pacific basins from 2010 to 2019 used in the development of the IBUS.

Basin	Model	Forecast hour													
		12	24	36	48	60	72	84	96	108	120	132	144	156	168
Atlantic	DSHP	2785	2468	2186	1921	1673	1474	1286	1139	1009	902	702	609	521	464
	LGEM	2785	2468	2186	1921	1673	1474	1286	1139	1009	902	702	609	521	464
	OFCL	2973	2644	2340	2057	163	1577	0	1222	0	968	0	407	0	307
East Pacific	DSHP	3126	2755	2421	2124	1850	1610	1385	1184	992	837	629	514	413	341
	LGEM	3126	2755	2421	2124	1850	1610	1385	1184	992	837	629	514	413	341
	OFCL	3389	3013	2658	2338	126	1795	0	1345	0	971	0	522	0	383

be used to communicate risk or NHC forecast uncertainty to the public and private interests.

To improve forecasts of intensity, we must first understand the inherent biases of intensity forecasts to characterize where improvements can be made. [Trabing and Bell \(2020\)](#) showed that rapid intensity changes correspond to the tails of the intensity error distributions, and showed that larger NHC forecast errors can be expected in thermodynamic environments favorable for intensification. [Bhatia and Nolan \(2013\)](#) demonstrated that the high variance in intensity forecast performance between different storms, models, and days is often dependent on TC attributes and synoptic conditions. [Trabing and Bell \(2020\)](#) similarly showed that large forecast errors often occur in favorable thermodynamic and kinematic environments for intensification contributing to larger biases for RI forecasts compared to rapid weakening (RW) forecasts. The dependence of forecast errors and biases on the intensity change suggests that forecast uncertainty can be estimated based on intensity change forecasts. In both PRIME and GPCE, the initial intensity and the forecast intensity are key predictors in the forecast error and warrants the investigation into whether intensity change forecasts and their corresponding error characteristics can provide simple estimates of forecast uncertainty.

In this manuscript, we will develop a simple Intensity Bias and Uncertainty Scheme (IBUS) that can be implemented in real time. The IBUS will both quantify inherent deterministic forecast biases to improve forecasts and evaluate forecast errors to communicate forecast uncertainty. The following section will detail the data used in this study. [Section 3](#) will explore the intensity change forecast distributions to motivate the techniques used to create the IBUS. In [section 4](#) we will test and evaluate the IBUS for two statistical–dynamical models. [Section 5](#) will explore the use of IBUS on NHC forecasts. A summary and discussion of potential uses and limitations for IBUS will be offered in [section 6](#).

2. Methodology

a. Data

In this work we will utilize two operational statistical–dynamical models: the Statistical Hurricane Intensity Prediction Scheme (SHIPS; [DeMaria and Kaplan 1994, 1999](#); [DeMaria et al. 2005](#)) and the Logistic Growth Equation Model (LGEM; [DeMaria 2009](#)). SHIPS uses multilinear regression to forecast

intensity change and includes both oceanic and atmospheric predictors. The decay SHIPS (DSHP) model includes an inland wind decay model to account for landfall and will be the version used in this study. The LGEM is a simplified dynamical intensity model that uses select SHIPS predictors. Both models provide intensity forecasts at 6-h intervals with the latest model runs being available to forecasters in real time.

DSHP and LGEM are statistical–dynamical models that are used in the intensity consensus (ICON), the intensity variable consensus (IVCN), and the HFIP corrected consensus approach (HCCA; [Simon et al. 2018](#)). IVCN has equal weights between the consensus members and has a two member minimum, meaning that on occasion DSHP and LGEM are the only intensity models included. The advantage of using DSHP and LGEM for this analysis is that the new versions of the model can be run over long periods of time without consuming significant resources. Because of this, DSHP and LGEM intensity change forecasts between 2010 and 2019 can be examined using the 2020 configuration of the models, which prevents year-to-year variations in the model formulation from affecting the characteristics of the model error. We utilize DSHP and LGEM in both the Atlantic and east Pacific basins from 2010 to 2019 to create the IBUS and test the data on the TCs from the 2020 season which are available from the best tracks in the Automated Tropical Cyclone Forecasting System (ATCF; [Sampson and Schrader 2000](#)). [Table 1](#) shows the sample size of DSHP and LGEM forecasts from 2010 to 2019 in each basin.

We evaluate DSHP and LGEM in an intensity change framework in which the intensity forecast over 12–168 h is subtracted from the initial intensity. The initial intensity is defined to be the intensity from the model initialization, and in real time will not always be the same as the final intensity found in the best track dataset. The intensity change will be calculated and evaluated at 12-h intervals against the best track dataset for the 2010–20 season in both the Atlantic and east Pacific basins. Central Pacific TCs are not included in the sample.

To test the application of the IBUS to NHC forecasts, we will use the official NHC forecasts from 2010 to 2019. The NHC forecasts are in intervals of 5 kt and are available at 12-h intervals through 72 h and at 24-h intervals beyond 72–168 h. The forecasts beyond 120 h are experimental and not provided to the public but are included here for completeness and potential future applicability. The 60-h forecast was implemented after the 2019 season, and will not be included in the analysis because of the small sample size shown in [Table 1](#). The sample size for forecasts beyond 120 h is also significantly smaller than

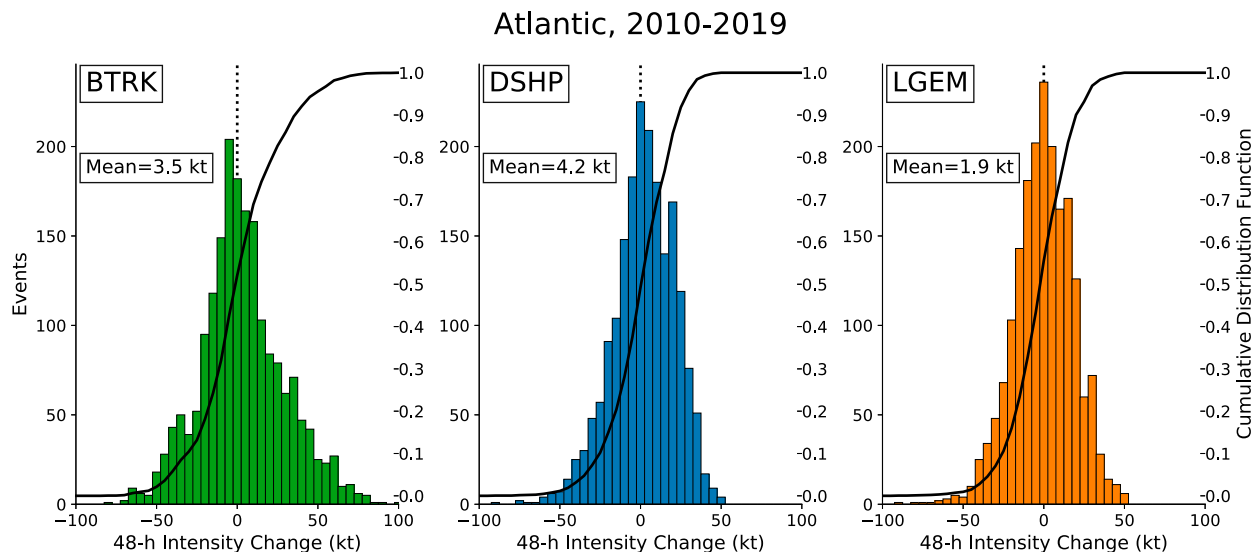


FIG. 1. The number of occurrences of 48-h intensity change values from 2010 to 2019 in the Atlantic basin for the (left) best track, (center) DSHP, and (right) LGEM. The intensity change is binned at 5-kt intervals, and the right ordinate denotes the cumulative distribution function. The mean intensity change values for the best track, DSHP, and LGEM are also denoted.

those for shorter forecast lengths because of verification constraints and the average duration of TCs. The average duration of TCs in the east Pacific is shorter compared to the Atlantic leading to fewer samples at long lead times (Table 1). As in the case for LGEM and DSHP, the IBUS will be evaluated on TCs from the 2020 season. Select cases are examined within the 2020 season to give more insight into the behavior of the IBUS.

b. Forecast verification

The verification of the IBUS will follow the official NHC forecast verification rules (Cangialosi and Franklin 2014). Forecasts are only included in the analysis and verified if the system is a tropical or subtropical cyclone at both the forecast initialization time and the verifying time. A tropical cyclone that decays or transitions to extratropical within the forecast period is not included. In addition, all other stages of development including tropical wave, or remnant low are excluded. Tropical depressions are included in the analysis. A homogeneous sample of DSHP and LGEM forecasts is used for the development dataset while the OFCL forecasts are evaluated independently.

The forecast verification will be measured in terms of the mean absolute error (MAE) of the forecasts. The error is the absolute value of the difference between the intensity change forecast and the actual intensity change and the MAE is defined as the average of the error over multiple forecasts. The bias of the forecast is the mean of the difference between the intensity change forecast and the true intensity change. A positive bias for intensity change forecasts means that the model or forecaster expected the storm intensity to be stronger than it actually was. Because intensity change can be negative, a positive bias for an intensity change forecast can be either an overestimated intensification or an underestimated weakening.

In this work we will use the standard deviation of errors (STDE; σ) to estimate the uncertainty in a forecast. Previous

studies, such as Bhatia and Nolan (2015), used the MAE to estimate the uncertainty of a forecast. Both the STDE and MAE will provide similar results and the two variables are highly correlated (not shown). The STDE is equitable to the root-mean-squared error for an unbiased forecast error distribution. The advantage of using the STDE over MAE is that if we assume that the forecast error distribution is normal, we can apply the empirical 68–95–99.7 rule. The rule states that ~68% of the data lies within one standard deviation of the mean while 95% of the data lies within two standard deviations of the mean. Using the STDE thus could give forecasters intensity bounds when making forecasts similar to that provided by GPCE. A major assumption here is that the underlying data are normal, which is often but not always the case as will be shown later.

When forecasting intensity change, it is critical that we also address the role that land interactions and landfalls have on the intensity change. In a sensitivity test, it was found that including landfall intensity change forecasts improved the relationship between forecast errors and the STDE but slightly degraded the utility of the bias correction on intensity forecasts. Because of the offsetting influence and in order to capture the entire error distribution, all intensity change forecasts that may have land impacts, including island and major landmasses, are included in the sample. The authors note that the bias correction skill improves by 1%–3% when landfall forecasts are removed (not shown).

3. The Intensity Bias and Uncertainty Scheme

a. Intensity change distributions

It is first helpful to understand the distribution of intensity change forecasts to further motivate the development of a bias correction and uncertainty scheme. The intensity change distributions for 48-h forecasts from DSHP and LGEM are

compared with the best track dataset intensity change distributions in Fig. 1. The best track intensity change distribution is approximately Gaussian with a slightly positive mean despite a negative median. The discretization of the best track dataset to 5-kt intervals contributes to the slight offset between the median and mean but the distribution is overall slightly skewed. The forecast intensity change distribution for DSHP and LGEM is also essentially Gaussian, but slightly skewed. The number of both large positive and negative intensity changes in the best track are infrequent as expected; however, DSHP and LGEM tend to capture the relative frequency of large weakening events but significantly underestimates the number of large intensification events. Over 48-h forecasts, RI is commonly defined as an increase in winds exceeding 55 kt (Kaplan et al. 2010). Neither DSHP nor LGEM are able to predict RI at the 48-h threshold over the 10-yr dataset, which is a known inherent limitation of statistical-dynamical forecast models.

The underestimation of RI is a common limitation for statistical-dynamical models because the events are rare. The statistical models are designed to reduce forecast errors and the penalty for RI false alarms outweighs the benefit of forecasting RI as too many false alarms lowers users confidence. Capturing RI is also difficult for global forecast models because of grid resolution which cannot explicitly resolve the TC inner core or the strongest winds found there. Forecasts of intensity change within these models when RI occurs are nearly exclusively biased too low; however, not all intensity forecasts are biased. In Fig. 2 we show the normalized distributions of 48-h intensity change events conditioned on specific intensity forecasts. Figure 2 shows that when DSHP forecasts a zero change intensity (from -2 to 2 kt in 48 h), the most common intensity change that occurs is a decrease of 5 kt, although the mean is very close to zero. For larger intensification rates predicted by DSHP between 10 and 20 kt, the peak of the distributions remain at lower intensity changes of 0 or 5 kt. LGEM shows similar behavior for weakening forecasts with the tendency to over predict the weakening rates. Overall, Fig. 2 shows that for increasing magnitudes of predicted intensity change (both positive and negative), the distributions of intensity change vary and can be treated as having different error characteristics. Treating similar intensity change forecasts as distinct from other forecasts will form the basis of our intensity change bias correction and uncertainty scheme.

b. IBUS creation

Figure 3 shows the bias, MAE, and the STDE for 12-hourly intensity forecasts of DSHP and LGEM from 2010 to 2019. While the bias shows that DSHP tends to have a slight positive intensity bias over the first 5 days in both basins, this information when applied in real time may not improve forecast errors. In addition, bulk estimates of the MAE or STDE over multiple storms/years, while helpful in comparing model performance, do not help forecasters in determining potential errors for any individual forecast. Figure 2 showed that model biases, as a function of intensity change, could add additional

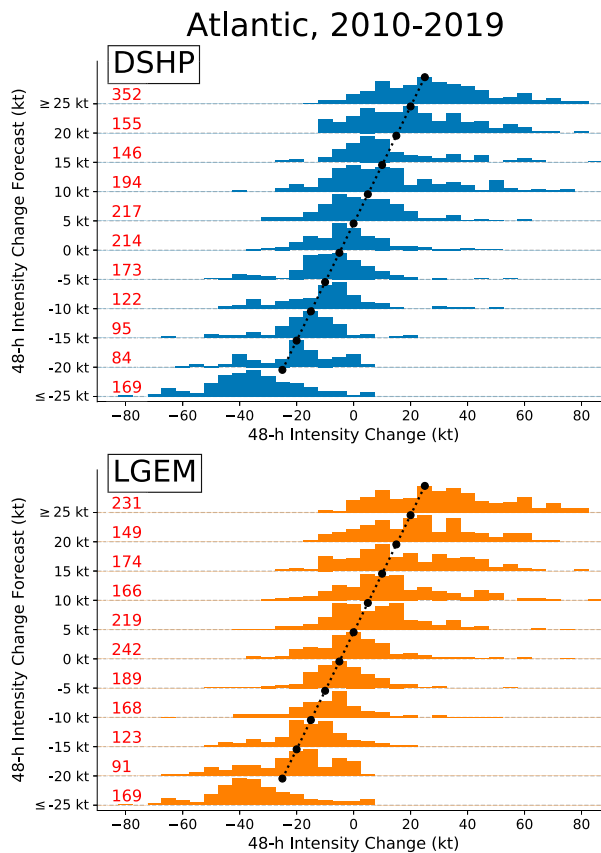


FIG. 2. Normalized distributions of 48-h intensity change are stacked corresponding to when forecasts of (top) DSHP and (bottom) LGEM were made given the intensity change values on the ordinate. The intensity change distribution is binned every 5 kt, and the black dots denote the mean intensity change value for an unbiased forecast. The red text denotes the total number of 48-h forecasts with those intensity change values.

value to forecasters by estimating when larger model biases and larger model errors are expected to occur.

To improve both the inherent biases in the forecast models and to gauge uncertainty in the model forecasts, we create the IBUS. The goal is for IBUS to be a type of lookup table based solely on data available in real time to quickly estimate expected model bias and uncertainty from retrospective data. To do this, we use the intensity change statistics from 2010 to 2019 for DSHP and LGEM. We separate each intensity change forecast into bins based on the forecast hour and intensity change forecast and then calculate the bias, MAE and STDE for each of those bins. Within each bin is a distribution, such as that shown in Fig. 2, which we can treat independently for each forecast hour and bin. The intensity change forecast values are rounded to the nearest 5-kt forecast interval, which is used by NHC. For example, forecasts between 8 and 17 kt by DSHP would translate to the 10–15-kt forecast bin in Fig. 4. The intensity change bins are varied such that 10-kt bins are used everywhere except between -5 and 5 kt where 5-kt bin sizes are used. This technique is used for each model in each

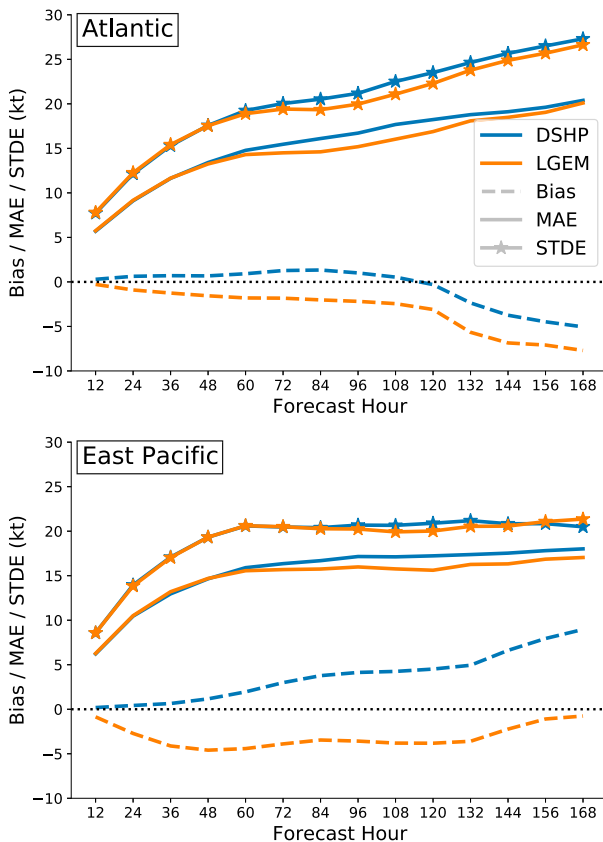


FIG. 3. The bias (dashed), mean absolute error (solid), and standard deviation of errors (stars) for DSHP and LGEM forecasts from 2010 to 2019 in the (top) Atlantic and (bottom) east Pacific. The dotted black line indicates zero bias.

basin separately to account for differences in error statistics between TCs in different basins.

Because a dependence on previous forecasts is inherently limited by previous events, we cannot bias correct a forecast value that has not been predicted in the past. Sample size also becomes an issue for rare forecasts, such as the top few percentiles of the intensity change distribution. To account for both limitations, extreme forecasts are binned in the top and bottom 3rd percentiles and applied to any intensity change that exceeds those percentiles. For example, DSHP has never forecast a 55-kt intensification in 48 h but should it or larger intensification rates occur, we assume that the bias would be similar to a 45-kt intensification which is in the 97th percentile. To help with bins that may have an under representative sample of values and to reduce noise, we apply a 2D Gaussian filter to the bias correction field. After smoothing the fields, the result is Fig. 4, which shows the bias in shading and the STDE in text within each bin. To get an unbiased forecast, we would subtract the shaded values in Fig. 4 from the intensity change forecast. The STDE in the text would be the uncertainty of the forecast given the intensity change forecast and forecast hour. It is important to note here that although Fig. 4 shows forecasts for DSHP and LGEM at extreme intensity

change rates (80 kt in 24 h), these values are caused by the binning of the top 3rd percentile and smoothing by the Gaussian filter and have never been forecast by either model. We will explore in section 3c how to interpret the IBUS and some basic insights gained.

c. Understanding the IBUS

The IBUS for LGEM and DSHP shown in Fig. 4 confirms that different intensity change forecasts have different biases and uncertainty. In the Atlantic, DSHP tends to have a positive bias for positive intensity change forecasts and small weakening forecasts, while large weakening forecasts tended to have a negative bias through 96 h. A more complicated pattern of bias beyond 96 h is present in DSHP as 80-kt increases over 168 h are negatively biased, a 40-kt increase over 168 h is positively biased, and no change in intensity over 168 h is negatively biased. The bias in the east Pacific for DSHP is quite different than the Atlantic, with a near positive bias everywhere except for short term RI events. The bias distribution is also distinctly different in LGEM between the Atlantic and east Pacific basins.

To interpret the distribution of biases we need to consider the sign of the bias and the sign of the intensity change. Considering both signs allows us to determine whether or not the intensity change forecast distribution is narrowed or widened. A positive bias for positive intensity change forecast means that the amplitude of intensity change should be reduced closer to zero in order to improve the model bias. A negative bias for a positive intensity change forecast means that the amplitude of intensity change should be increased which would expand the intensity change distribution. In the east Pacific, both DSHP and LGEM have a negative bias for 12–24-h forecasts of intensity change > 20 kt indicating that intensity change forecasts within those ranges should be increased (to potentially RI) in order to improve the model bias, which would help to improve the inherent limitations of the models shown in Fig. 1. In contrast, the same 12–24-h forecasts in the Atlantic for both DSHP and LGEM have a positive bias meaning that the models are predicting the large intensity changes too often compared to observations with too many false alarms for RI, so the intensity change forecasts should be scaled down.

In addition to the bias correction, the STDE shown in Fig. 4 is capable of communicating the uncertainty of intensity change forecasts assuming a normal distribution of model errors. The STDE shows a systematic increase from negative intensity change forecasts to positive intensity change forecasts, with the largest STDE occurring when large intensification events are predicted over the first 5 days. The STDE objectively communicates that RI events are more difficult to forecast than RW events and often have larger errors (Na et al. 2018; Trabing and Bell 2020). For example a DSHP forecast of 50 kt in 48 h in the Atlantic has an STDE of 15 kt compared to a 10-kt STDE for -50 kt in 48-h intensity change. For forecasts in the 6–7-day range the STDE can be more variable, similar to the bias, and shows that a weakening Atlantic hurricane (from -10 to -30 kt) in 6–7 days has more

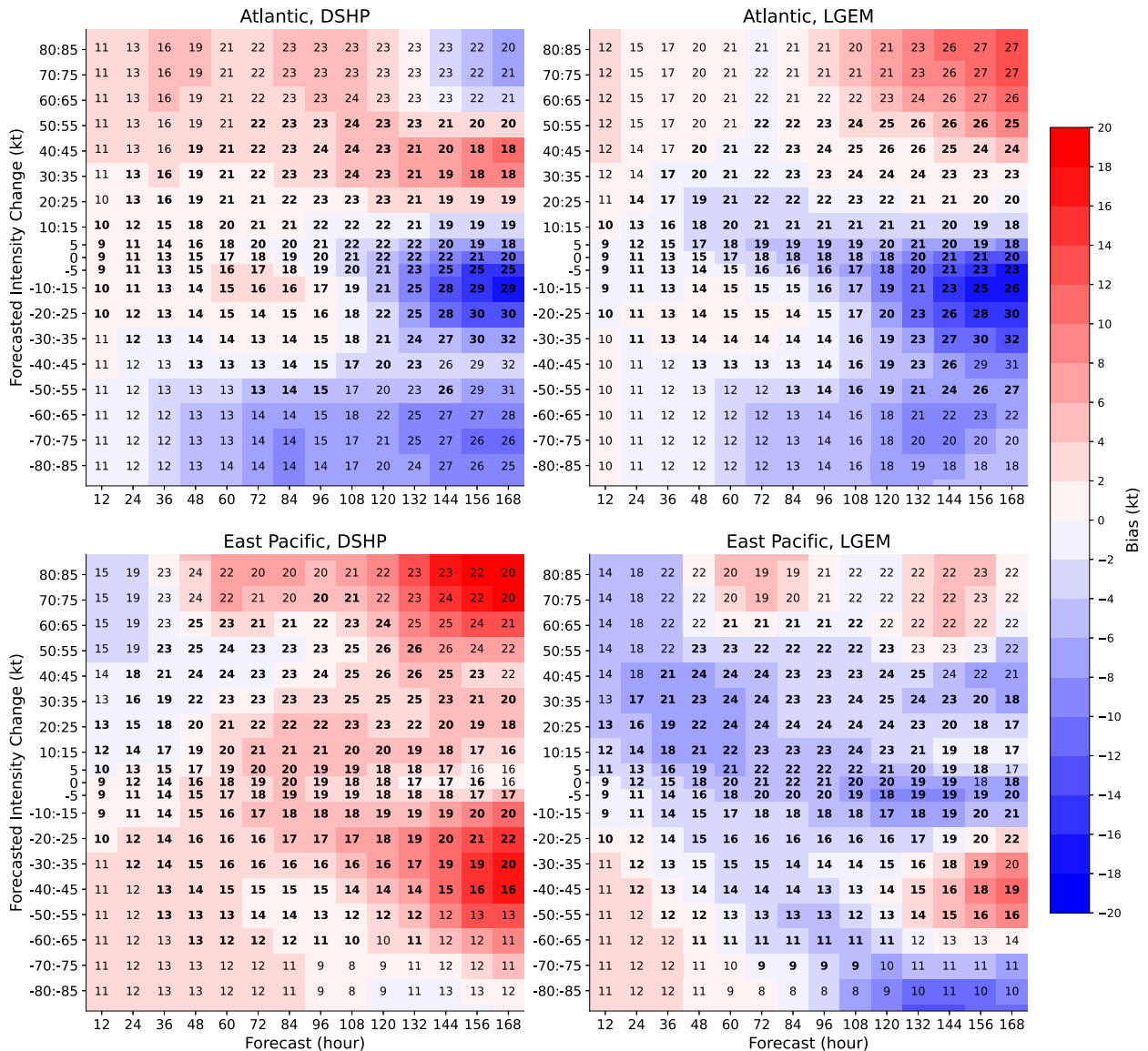


FIG. 4. Distribution of bias (shaded) and standard deviation of the intensity change errors (numbers; kt) for (left) DSHP and (right) LGEM conditioned on intensity change forecasts. The bias and STDE is shown for the (top) Atlantic and (bottom) east Pacific. Bias and STDE are shown for forecast hours at 12-h intervals with variable bin sizes of 5–10 kt for the intensity change forecasts. Bins that have a sample size greater than 20 have bold numbers.

uncertainty than if a slight strengthening was predicted. Although the STDE will provide a slightly larger estimate of expected errors compared to using MAE (e.g., Fig. 3), it is also important that the uncertainty estimate is large enough to encapsulate a high percentage of the forecasts to be useful.

Overall the amplitude of the model biases and STDE increases with forecast length as larger intensity change events are possible. The biases and STDE also tend to increase in magnitude because of the smaller sample sizes within those bins, although we have mitigated that issue by smoothing the resulting fields. The LGEM forecast biases in the east Pacific are somewhat unique in that the overall biases and STDE at

longer forecast lengths are not as large as for the other models. Even with the 10-yr dataset the sample size for intensity forecasts extending out 7 days is limited. It should also be noted that interpreting 6–7-day intensity change distributions is complicated by the fact that there is little dependence on the intensity change over the previous 5 days into the potential biases at those times. For example a 0-kt intensity change over 7 days could indicate a steady-state storm or encompass multiple RI and RW events within that time frame. In addition, those biases will also have a stronger dependence on the track error distribution for the model which can be large at days 6–7.

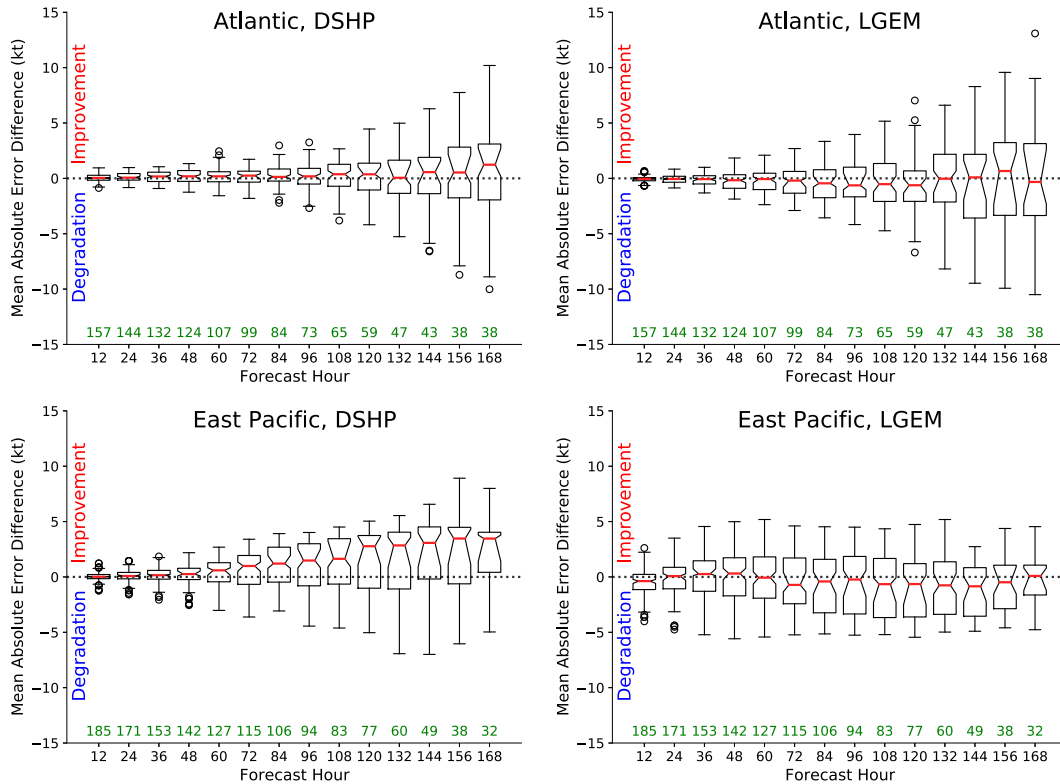


FIG. 5. Leave-one-out validation for (left) DSHP and (right) LGEM forecasts for all (top) Atlantic and (bottom) east Pacific TCs in the 2010–19 dataset. The bias correction model is re-derived without the storm that the model is being tested on and then applied to all forecasts for that individual storm. The box-and-whisker plots are of the mean error for all valid forecast hours from each individual storm and the bias-corrected model. Positive error difference indicates reduced errors in the bias-corrected forecast compared to the original model forecast. The red lines are the median and the notches indicate the confidence intervals for the median at the 95th percentile using the bootstrap method with 100 000 iterations. Circles indicate outliers that are beyond 1.5 times the interquartile range from the median. The green text indicates the sample size.

4. Testing the IBUS

Now we will test the bias correction to determine if it could potentially improve forecasts of intensity change and definitively evaluate whether the STDE is able to assess forecast dependent uncertainty.

a. Testing the bias correction

The IBUS that was developed and shown in the previous section is now rigorously tested to determine its potential use. First we conduct a leave-one-out validation on individual TCs in the training dataset. We first remove all forecasts from a single TC from the training database and then recalculate the bias and STDE to create the IBUS. The intensity forecasts for each forecast cycle are then bias corrected and the MAE is calculated for both the original forecast and the bias corrected forecast at each forecast hour. Note that each TC has a different number of forecast cycles and verifying hours. The process is then repeated for each TC within the 2010–19 sample to get an adequate sample. We should note that a leave-one-out validation was also attempted on a yearly basis, which yielded inconclusive results because of large year-to-year

variability in the number and types of TCs. In 2017 there were several long-lived TCs such as Maria and Irma that had extensive intensification periods that when excluded from the database together and tested on cause severe degradation to the forecast skill.

The performance of DSHP and LGEM in both basins in the leave-one-out validation on individual storms is shown in Fig. 5. The box-and-whisker plot of the MAE difference between the bias corrected and original model forecasts is shown such that positive values indicate reduced errors in the bias corrected model. The red line is the median MAE and the notches indicate the confidence intervals for the median at the 95th percentile using the bootstrap method. The bias correction overall had a neutral to positive impact on the forecast errors of the models with a wide range of potential improvements in some TCs and degradation in others. On average the bias correction showed positive impacts in both basins for DSHP with small but statistically significant improvements to Atlantic forecasts from 24 to 72 h and larger statistically significant improvements for forecasts beyond 48 h in the east Pacific. The bias correction did not perform as well for LGEM with neutral to slightly negative impacts

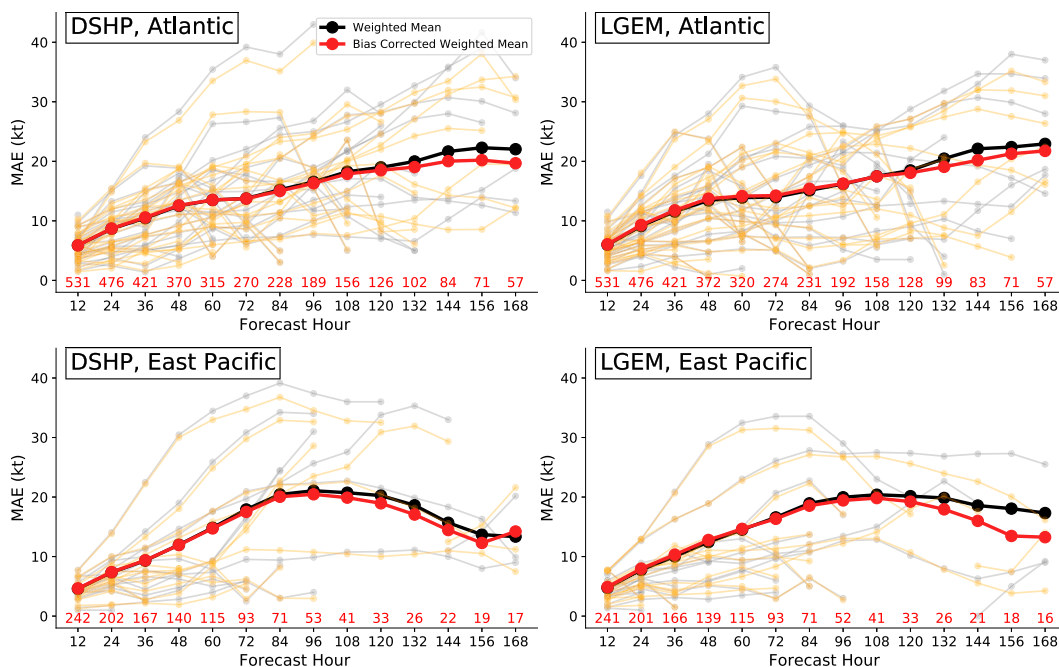


FIG. 6. Mean absolute error (MAE) with forecast hour for each TC in 2020. The MAE for each individual TC is shown with the gray lines (operational forecast) with the light orange being the bias-corrected forecast errors. The weighted mean for the original forecasts is shown by the black line, and the weighted mean for the bias-corrected forecasts is shown by the red line. The sample size is indicated by red text.

in both basins; however, none of the degradation is statistically significant at the 95% confidence level. Because the bias correction has overall small biases in the short term, the range of MAE differences gets larger with forecast length for each case excluding LGEM in the east Pacific.

To further evaluate the bias correction component of the IBUS, we now verify the IBUS for each model on the 2020 season using the 2010–19 training dataset. Figure 6 shows the MAE for the original forecasts and the bias corrected forecasts for each TC and the weighted mean. In the Atlantic, the bias correction was able to reduce the intensity errors for 5–7-day forecasts of both DSHP and LGEM compared to the operational models. By comparing the cases, we can determine that improvements caused by the bias correction were largest in cases that had overall large forecast errors. In the Atlantic, some of the largest average forecast errors came from Tropical Storm Gonzalo, Hurricane Paulette, and Tropical Storm Rene (AL07, AL17, and AL18) which were all improved by the bias correction. In contrast, the bias correction degraded forecasts for Tropical Storm Cristobal (AL03) and Hurricane Laura (AL13), where multiple interactions with land and landfall events occurred. Although landfall forecasts were included in the sample, the potential for changes to the intensity such as suggesting an increase in intensity overland, can lead to errors in the bias correction which will be incorporated into future work.

In the east Pacific, Fig. 6 shows that the bias correction performed just as well as in the Atlantic despite the lower sample size. The bias-corrected DSHP and LGEM forecasts had

reduced average errors for forecasts beyond 72 h, with the exception of the 168-h forecasts of DSHP. The bias correction had an overall small effect on the average short term forecast errors where intensity forecasts have already low errors. Similar to the Atlantic, the bias correction was able to successfully reduce the average forecast errors for cases in which large intensity errors were present. DSHP had large forecast errors for Tropical Storms Cristina, Fausto, and Norbert (EP05, EP12, and EP19) while LGEM struggled the most with Tropical Storm Cristina, Tropical Storm Fausto, and Hurricane Douglas (EP05, EP12, and EP08). In each of the cases for both models, the bias correction was able to reduce the MAE of the intensity forecasts at nearly all forecast hours. However, we would like to emphasize that the bias correction does not improve all of the forecasts, and small degradation is found in cases with already low forecast errors such as for Tropical Storm Karina (EP16).

Overall, the application of the bias correction appears to have a positive impact on the MAE of the individual models in both basins, with the most benefit at longer forecast times. The simple scheme is only a function of the intensity change forecast and forecast hour which can be quickly calculated from the ATCF and applied in real time, or taken into consideration by forecasters upon examining Fig. 4.

b. Testing the uncertainty estimation

Now we will examine whether the STDE over each forecast hour can communicate uncertainty in the model forecast. To objectively determine the value of the STDE we examine the

correlation between the STDE from the predicted intensity change to the absolute error of the forecast. To examine the correlation, we need to have a large training and testing sample, so we perform this analysis on all TCs in the 2020 dataset using the 2010–19 IBUS.

Figure 7 shows the correlation between the absolute error and the STDE for DSHP and LGEM in the Atlantic and east Pacific basins. The correlation between STDE and absolute forecast errors is positive over the first 120 h for each model and in both basins. The positive correlation means that larger errors are found when the STDE is larger, which is the expected result (Trabing and Bell 2020). The positive correlation indicates that the STDE can communicate the desired uncertainty but with varying degrees of success. DSHP and LGEM in the Atlantic have statistically significant correlations over the first 120 and 84 h, respectively, which are strongest for 36–48-h forecasts and decreases with increasing forecast length. The variance explained for DSHP and LGEM for 48-h forecasts is 7% and 11%, respectively. The small variance explained is not surprising, given the small range of potential STDE values shown in Fig. 4. This variance is the amount explained solely by the intensity change forecast within the model and exceeds the 5%–6% error variance explained for 12–48-h forecasts using the GPCE in 2012 (Goerss and Sampson 2014). In the east Pacific, the error correlation tended to increase with forecast hour, reaching to correlations just below 0.5 at 120 h. Similar increases in error variance with forecast length was found for east Pacific TCs in Goerss and Sampson (2014), which suggests that east Pacific forecast errors are easier to explain in terms of intensity change compared to those in the Atlantic. The variance explained by the STDE is lower in the east Pacific over the first 48 h compared to the Atlantic with 4% for DSHP and 8% for LGEM.

For forecasts beyond 120 h, the correlation between STDE and absolute errors is weaker and reverses sign for some forecast hours. DSHP in the east Pacific has a negative correlation between the error and STDE which means that DSHP had reduced errors for those forecasts compared to the errors that had occurred in the 2010–19 dataset. The reason for the negative correlation is partially due to the low sample size of forecasts in the east Pacific in 2020, but it also indicates that the sample size of the 2010–19 dataset for day 5–7 forecasts is not large enough in some intensity change forecast bins to be characterized as normal. Another contributor to the low correlations beyond day 5 is that track errors will play more of a role in the uncertainty forecasts which have not been considered in this framework. The negative correlations suggest that at long range, the intensity uncertainty is more related to other factors including track than the intensity change forecast alone.

Despite the small variance explained between absolute error and STDE, the results are similar in magnitude to the performance of GPCE which predicts the errors for IVCN (Goerss and Sampson 2014). The simple uncertainty estimates are derived from past performance but adds information beyond the static estimates of MAE or STDE over the course of a season. While the bias correction showed the most promise for forecasts with longer lead times, the

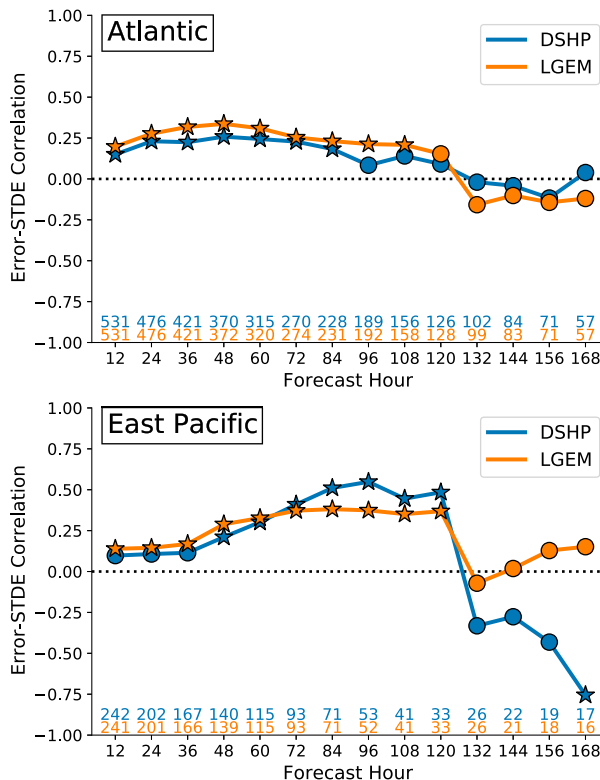


FIG. 7. The correlation between the mean absolute error of 2020 forecasts and the standard deviation of errors derived from the intensity change distributions. Stars indicate a statistically significant correlation from zero at the 95% confidence level. Sample size for the corresponding model is shown in the colored text.

uncertainty estimates appear to perform better over shorter forecast lengths in the Atlantic.

c. IBUS example

As stated earlier, the IBUS is based solely on previous forecasts of intensity change biases and is essentially a lookup table that can be implemented in seconds. Figure 8 shows two examples of how IBUS could be implemented on DSHP and LGEM and potentially be used for a consensus. The first forecast for Hurricane Laura (AL132020) was while land interactions and track uncertainty near the islands in the Caribbean were causing forecasts to agree on a slight weakening of the storm which resulted in low biases for DSHP, LGEM, and the average. The shading and black dashed lines are the STDE which helps to show the uncertainty in the forecast of which the actual intensity was located just above the uncertainty estimates until RI began. The larger shading (blue and orange) near the time of RI helps to communicate that the intensity of Laura at that time is more uncertain. The intensification of Hurricane Laura to peak intensity at the 96-h forecast was in the 99th percentile of intensity change forecasts for both statistical-dynamical models. The forecasts of intensity after landfall of Laura have the most uncertainty, although in actuality that

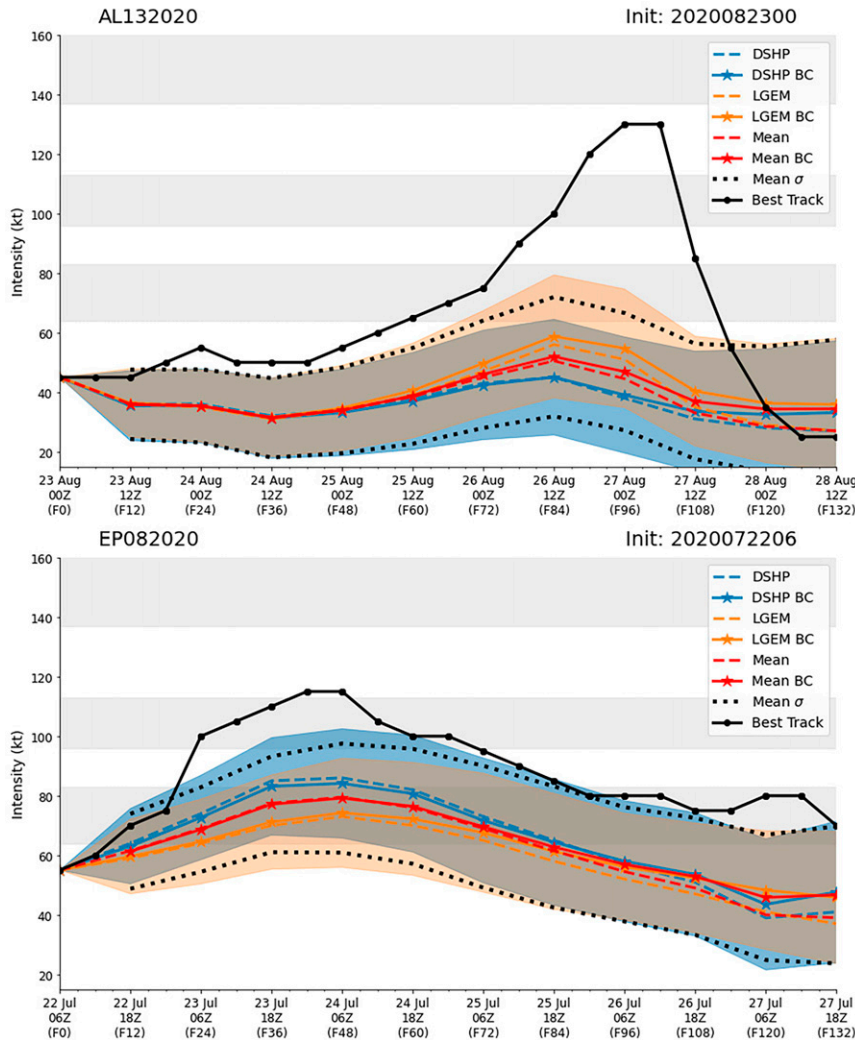


FIG. 8. (top) Intensity forecasts for AL132020 (Laura) initialized at 0000 UTC 23 Aug and (bottom) forecasts for EP082020 (Douglas) initialized at 0600 UTC 22 Jul. The black line is the 6-h preliminary best track intensity. The forecasts for DSHP, LGEM, and the mean are shown in dashed lines with the bias-corrected versions indicated by the solid lines with stars. The shading for DSHP (blue) and LGEM (orange) shows \pm STDE with the bias-corrected ensemble STDE shown using the black dotted line. Note that the dark gray shading is where the DSHP and LGEM STDE overlap. The Saffir-Simpson wind speed intensity categories are denoted in alternating gray-white shading. The forecasts are shown through 132 h.

intensity uncertainty is likely falsely characterized at the end of the forecast because of the land interactions. On average, we would expect lower forecast errors to occur when a TC moves over land and weakens as a function of its peak winds (Kaplan and DeMaria 1995). The IBUS does not include the effects of land and in this form cannot account for uncertainties in the timing of landfall which will heavily affect the intensity forecast. As such the bias correction tended to indicate a stronger tropical cyclone post landfall at 132 h.

The forecasts of Hurricane Douglas (EP082020) for LGEM and DSHP tended to underestimate the intensity. The bias corrections to the statistical-dynamical models were opposite

early on such that the IBUS for DSHP would have reduced the intensity forecast but for LGEM the intensity would have increased which had an overall negligible effect on the mean though 72 h. Beyond 72 h the biases between DSHP and LGEM were similar such that the overall forecast was improved. For this forecast of Hurricane Douglas, the STDE did not extend far enough to capture the actual intensity evolution for most of the forecast which is due to the models not capturing the extent of RI that Douglas underwent over the first 24 h of the forecast. Although the RI was significantly under forecast, the upper bound provided by the STDE nearly matched the actual intensity for the forecast between 60 and 108 h

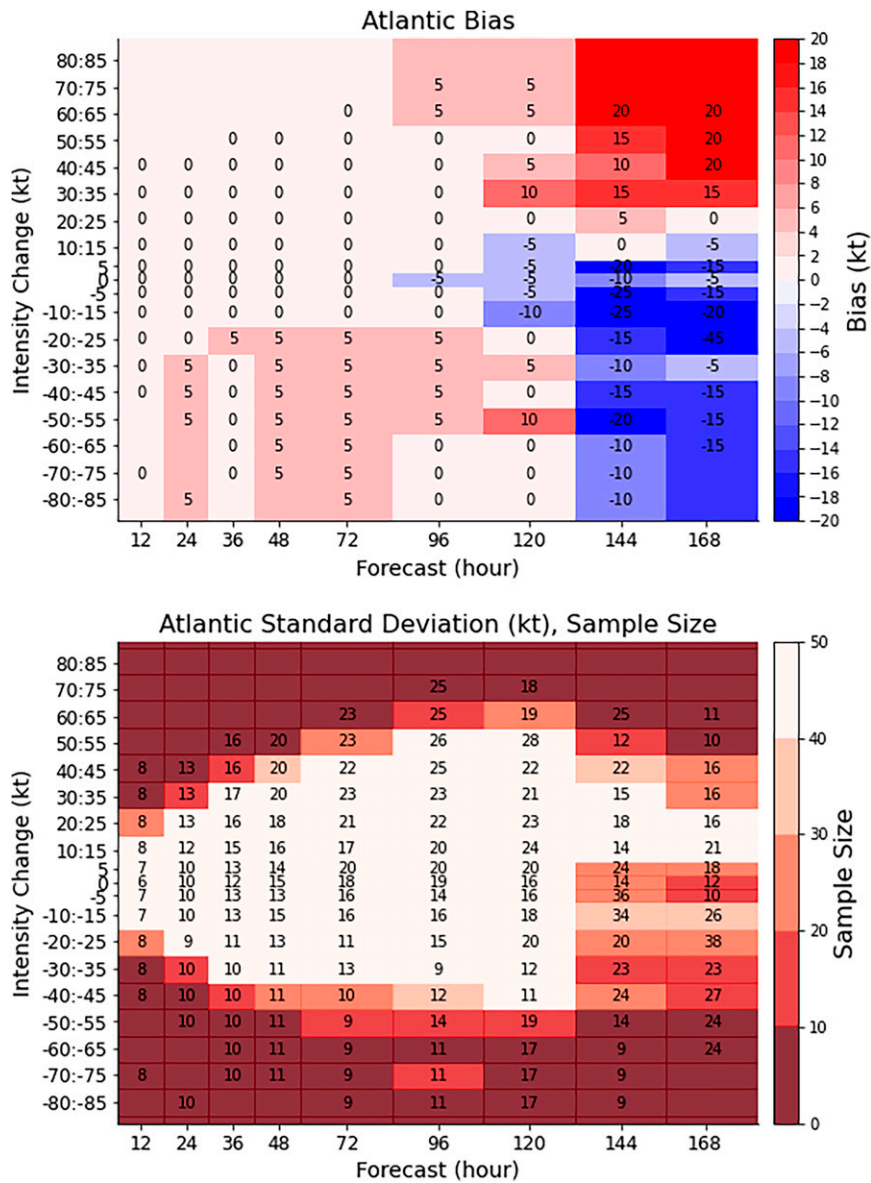


FIG. 9. (top) Distribution of bias for official NHC forecasts of Atlantic hurricanes from 2010 to 2019. The number listed is the bias rounded to the nearest 5-kt interval for each intensity change forecast and forecast hour. (bottom) The number of forecasts (shaded) and STDE (numbers; kt) for each intensity change and forecast hour bin. Forecast hours are shown for operational intensity change forecasts at 12-h intervals extended to 24 h beyond 3 days. Variable bin sizes of 5–10 kt for the intensity change forecast are utilized similar to Fig. 4. Biases and STDEs are not shown for bins that have no forecasts within the 10-yr sample.

after rounding to 5-kt intervals. Both the bias and uncertainty estimate in this case adds value to understanding the forecast and its limitations.

5. Application to NHC forecasts

We have shown that the IBUS can be used for statistical-dynamical models, but this framework is also applicable to dynamical models with a large enough dataset and to forecasts

from the NHC. We will now explore what the IBUS would look like for NHC forecasts using the data available from 2010 to 2019. Because NHC forecasts only in 5-kt intervals and only for a limited number of forecast hours, we apply all the same steps as before but we do not apply any smoothing to the bias or STDE fields because of the differences in forecast hour resolution. The biases are then rounded to the nearest 5-kt interval consistent with the intervals used by NHC forecasters.

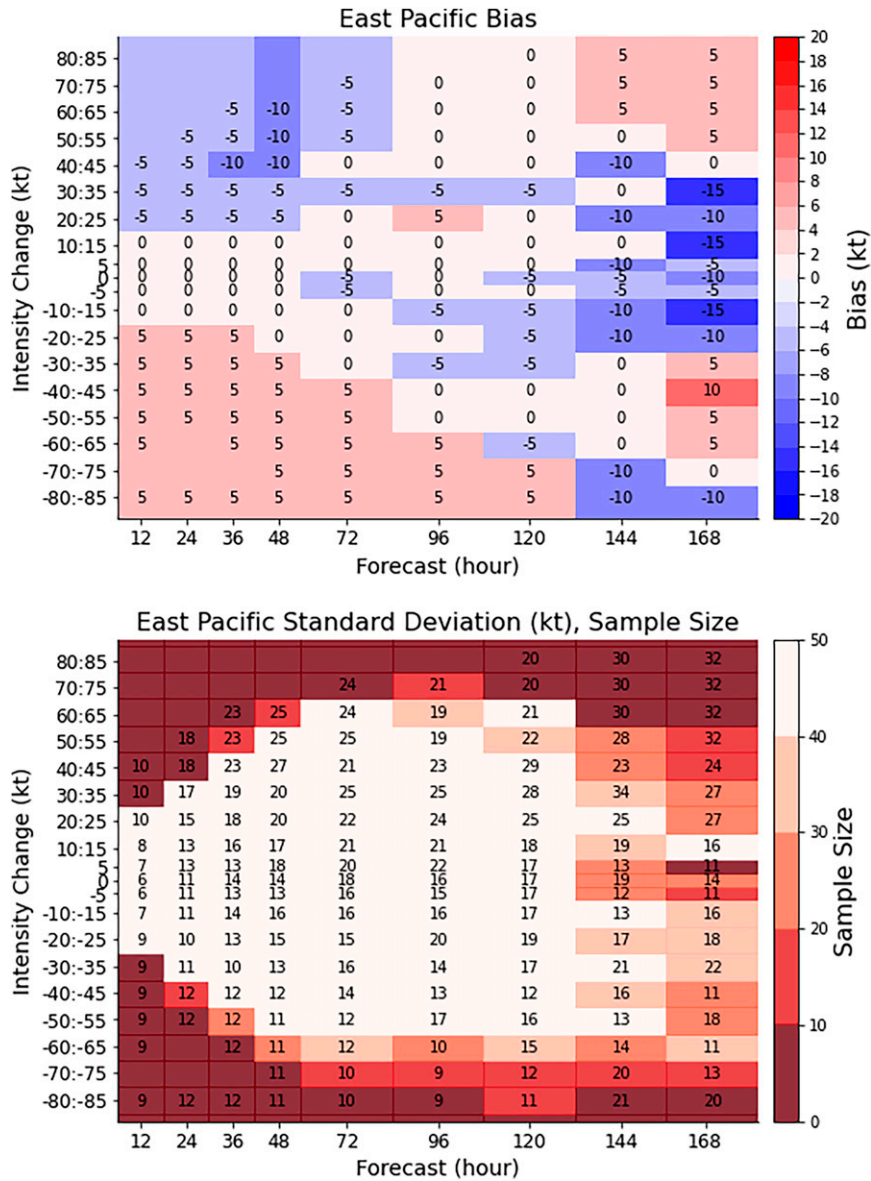


FIG. 10. As in Fig. 9, but for the east Pacific basin.

Figure 9 shows the bias, STDE, and the sample size as a function of forecast hour and forecast intensity change in the Atlantic. The biases show that NHC forecasters tend to have a zero bias on most 12–72-h forecasts. OFCL forecasts of RI in the Atlantic show a near zero bias over the 2010–19 time frame, although a negative bias is present if the sample is limited to the 2015–19 time frame (not shown). As was also shown in Trabling and Bell (2020), RW forecasts have a positive bias meaning that the amount of intensity change is under predicted. Beyond 72 h that bias appears reversed, with large intensification forecasts having a positive bias and large weakening forecasts being negatively biased. Figure 10 shows the bias in the east Pacific which overall has similar biases but with increased amplitudes compared to the Atlantic through

120 h. There are overall more large weakening events in the east Pacific compared to the Atlantic where RW and decay is less common and extratropical transitions, which are not included in the verification, are common (Jones et al. 2003; Wood and Ritchie 2015). The negative bias for RI events over 12–72 h is more pronounced in the east Pacific.

The STDE is indicated by the numbers in the bottom panels of Figs. 9b and 10b for the Atlantic and east Pacific. For forecast lengths of less than 72 h, the STDE is larger for intensification forecasts compared to weakening forecasts which is similar to what was shown for DSHP and LGEM. There are variations in the STDE between the basins such as a larger STDE for short term RI forecasts in the east Pacific compared to the Atlantic. Despite overall smaller biases in the east

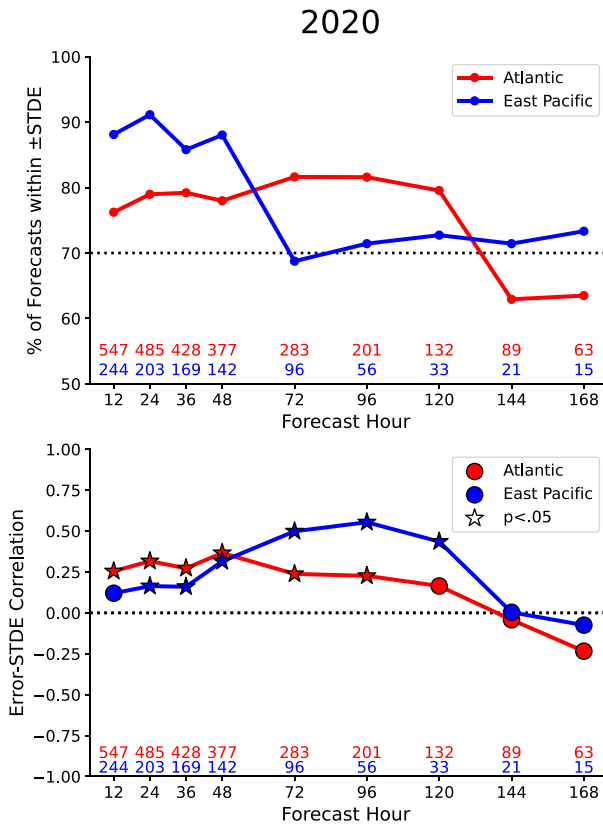


FIG. 11. (top) The percentage of intensity forecasts within ± 1 STDE of the intensity change forecast for all forecasts in 2020 in each basin based on the IBUS computed from 2010 to 2019. The numbers indicate the number of forecasts within the Atlantic (red) and east Pacific (blue) basins. (bottom) The correlation between the STDE and OFCL MAE in the 2020 season for each basin. Stars indicate a statistically significant correlation from zero at the 95% confidence level.

Pacific, there appears to be a larger STDE in the east Pacific. One potential reason for the larger uncertainty in the east Pacific is the more common RI and RW events found there (Kaplan et al. 2010; Hendricks et al. 2010; Trabing and Bell 2020).

The IBUS for NHC forecasts in the Atlantic and east Pacific are now evaluated on the 2020 season, although we only show the results of the uncertainty estimation in Fig. 11. The bias correction shows mixed results due to the fact that OFCL forecasts are issued in 5-kt intervals (not shown). The absolute error to STDE correlation is found to be positive in the Atlantic and east Pacific basins through 120 h, consistent with the results for DSHP and LGEM (Fig. 7). Over the forecasts through 48 h in the Atlantic, the correlation is statistically significant at the 95% level but only explains ~6%–13% of the variance in the forecasts. The 24–48-h forecasts in the east Pacific show a small but statistically significant relationship between the STDE and the absolute errors explaining up to 9% variance at 48 h. The correlation is larger for 48–120-h forecasts in the east Pacific which have a smaller sample size. The variance explained in the 72–96-h forecasts by the IBUS in the east Pacific is 24%–30%.

In addition to the error correlation, we also calculate the percentage of NHC forecasts where the intensity exceeded ± 1 STDE from the forecast intensity, shown in Fig. 11. It is important that the STDE be both large enough to capture uncertainty but not too large such that every forecast be contained within the bounds. Empirically we would expect that given a large enough sample size, that roughly 68% of forecasts would lie within one standard deviation. Overall, the number of forecasts within that range is ~80% over 120 h in the Atlantic dropping to ~65% beyond 120 h. In the east Pacific there were too many forecasts within the STDE range (~90%) over the first 48 h which is one reason for the low absolute error–STDE correlation found there. The percentage slightly increased in the east Pacific with forecast length after 72 h from 69% to just above 70% at 168 h. Given the fairly robust sample size in the Atlantic in 2020, we can state that approximately 4 out of 5 forecast intensities through 120 h would lie within the bounds estimated by IBUS, which is slightly larger than the a priori estimates from GPCE and could be valuable to recipients of NHC forecasts.

In real time, the IBUS for NHC intensity forecasts could be applied after forecasts are made. Although the bias correction could be applied to the forecasts, it would be preferable to relay the potential biases to the forecaster for consideration and only apply the uncertainty scheme. Figure 12 shows two NHC forecasts with the bias correction applied with the uncertainty scheme for both the bias corrected and official forecasts. The forecasts are at the same times shown earlier for LGEM and DSHP forecasts of Hurricane Laura and Hurricane Douglas in Fig. 8. In both cases, the STDE provides a good estimate of the range of potential intensities of the TCs. The STDE also well communicates that the largest uncertainty will occur where the intensification of Hurricane Laura is expected on 27 August and near and after the time of peak intensity in Hurricane Douglas from 0600 UTC 24–25 July. The forecast for Hurricane Laura did not have a bias; however, the bias correction if applied to the Hurricane Douglas forecast would have amplified the intensification and lead to an overall reduction of errors by the NHC forecasters.

In summary, we have shown that an IBUS can be created for the NHC forecast. The IBUS shows that intensification forecasts up to 72 h have the most uncertainty and are generally negatively biased, while weakening forecasts have lower uncertainty and are positively biased. Using this simple approach, we can explain a small amount of the variance in the intensity forecast errors that encompasses roughly 75% of the forecast intensity. The uncertainty estimation can be quickly run in real time and could provide meaningful, state dependent uncertainty forecasts for intensity change that are based on NHC forecast skill. Because the IBUS relies on past performance, better forecasts can reduce the STDE such that as intensity forecasts improve with time, so too will the uncertainty estimates.

6. Conclusions

In this manuscript, we have documented how a bias correction and uncertainty scheme can be created for deterministic forecasts. Because some intensity change forecasts are known

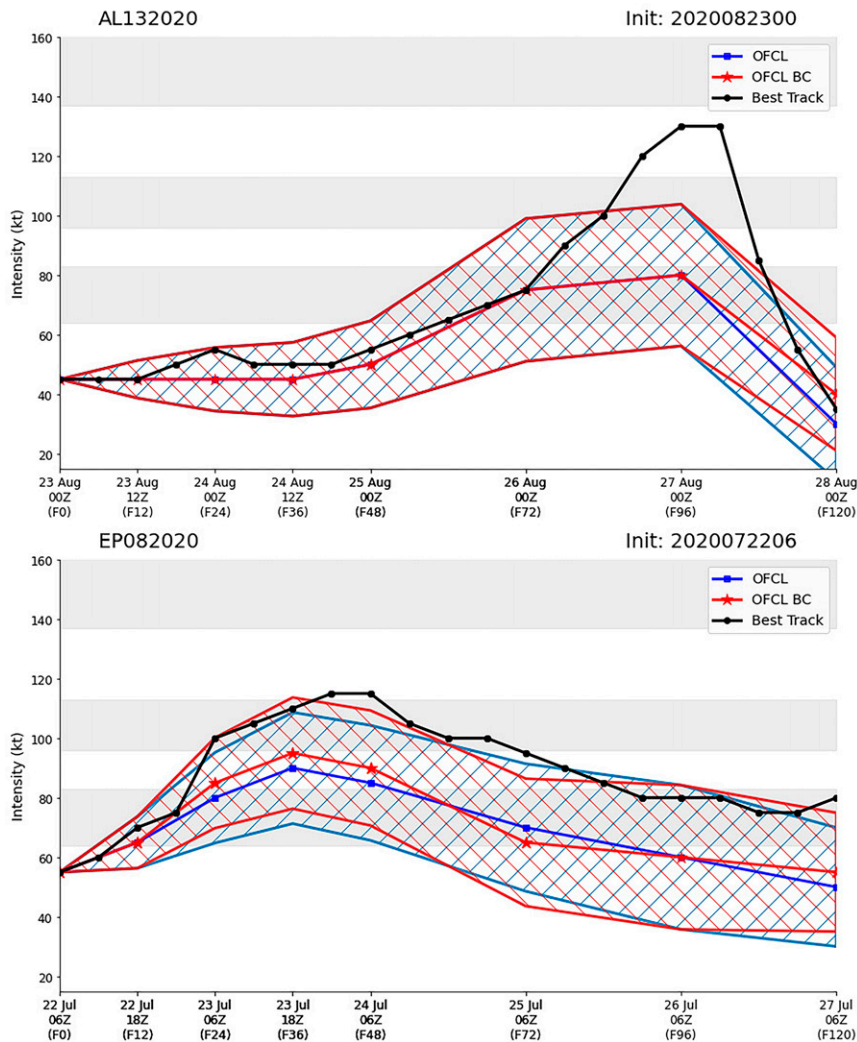


FIG. 12. Same intensity forecast times as Fig. 8, but with the official NHC forecasts. The STDE from IBUS is applied to both the bias-corrected intensity forecast (red) and the actual forecast (blue) where the hatching indicates the STDE for the bias-corrected and original forecast.

to be more accurate than others (Na et al. 2018; Trabing and Bell 2020), an Intensity Bias and Uncertainty Scheme (IBUS) can be created based solely on intensity change forecasts from the initial to the verifying time. The IBUS is created by binning intensity change forecasts over discrete forecast hours and evaluating the bias and the standard deviation of the forecast errors. The IBUS essentially gives a historical perspective (relative to the training dataset) of how well similar forecasts of intensity change performed in the past to forecast the bias and uncertainty.

The IBUS is developed on the 2010–19 dataset of DSHP and LGEM forecasts and then used to evaluate the 2020 Atlantic and east Pacific hurricane season. The IBUS was able to successfully communicate forecast uncertainty for the DSHP and LGEM models with 7%–11% of the variance explained in intensity errors at 48 h in the Atlantic. The bias correction was able to reduce the 2020 forecasts errors beyond 96 h in both

models with the largest improvements in errors coming from TCs with already large mean absolute errors. The model uncertainty estimates in the east Pacific showed slightly different characteristics compared to the Atlantic, which could be due to the lower number of TCs in 2020. When applied to the east Pacific, IBUS better communicated the forecast uncertainty in the 48–120-h forecasts with the variance explained reaching up to 30%.

When applied to NHC forecasts, the IBUS is able to provide uncertainty estimates for forecasts in a similar manner as the “cone of uncertainty” for track forecasts but also convey that some intensity change forecasts are more predictable than others. The IBUS for NHC forecasts is trained on the data from 2010 to 2019 and evaluated on the 2020 season. Using this simple approach, the IBUS explains a small amount of the variance in the intensity forecast errors which encompasses roughly 80% of the forecast intensities through

day 5. The IBUS provides uncertainty forecasts for intensity change that meaningfully changes based on error characteristics over the previous 5 years of NHC forecasts. Because the IBUS relies on past performance, better forecasts with time can reduce the STDE such that as intensity forecasts improve with time, so too will the uncertainty estimates. Further testing of the IBUS and applications to NHC forecasts is still needed before being recommended for operational use.

There are several advantages of using this framework. First, the IBUS can be quickly run and updated in real time. The simplicity of the scheme means that the training dataset can be continually updated during the season allowing the training dataset to constantly improve. Similar to what was employed in Krishnamurti et al. (2011) after a storm dissipates and if no other storms are present in the basin, forecasts from that storm can be added to the training set. The IBUS is also created solely on individual model performance and a homogeneous sample is not required for application to a consensus of forecast models. In addition, the uncertainty estimate for intensity in this framework provides the same messaging as the “cone of uncertainty.” When an intensity forecast is made, the intensity should fall within the STDE roughly 70% of the time and when applied to NHC forecasts is directly related to the past performance of the forecasters.

The IBUS is inherently limited by its simplicity and needs a robust sample of cases. We assume that each forecast hour and intensity change bin can be characterized as a normal distribution, which is not always the case for extreme intensity changes and for long forecast hours. The scheme itself does not discriminate between forecasts where land interactions are present and cannot add uncertainty in the cases of islands or reduce uncertainty for expected landfalls of a TC in the Gulf of Mexico. In addition, by binning the intensity changes at discrete hours, the bias and STDE is somewhat independent of other forecast hours. For example, a bias for the intensity change forecast of +10 kt in 96 h does not consider whether there was an increase of 30 kt in the first 48 h and a weakening of 20 kt in the following 48 h or if that storm was steady state for 84 h and increased by 10 kt in the last 12 h. By smoothing the fields we are able to reduce some of the noise from sample limitations and communicate the biases from one forecast hour to the next but in a limited capacity. Improving these limitations is the focus of ongoing work.

Future work will explore the application of the methodology used here to many different operational forecast models used by NHC. The development of a consensus uncertainty estimate from multiple regional forecast models is ongoing. With future advancements in computing, dynamic intensity forecast uncertainty can be provided by multimodel regional ensembles. The simplicity of the IBUS to diagnose forecast uncertainty should be used as a baseline for testing future uncertainty estimates as a single climatological value is not sufficient. Future work will also include developing an integrated measure for both track and intensity uncertainty because those errors can be dependent on each other.

Acknowledgments. This work has been funded by HFIP Award NA19OAR4320073. This work has been improved by comments from Alan Brammer, Alex Libardoni, John Knaff, and three anonymous reviewers.

Data availability statement. Python scripts can be found at <https://gitlab.com/Btrabing/IBUS>. All data are publicly available within the ATCF.

REFERENCES

- Bhatia, K. T., and D. S. Nolan, 2013: Relating the skill of tropical cyclone intensity forecasts to the synoptic environment. *Wea. Forecasting*, **28**, 961–980, <https://doi.org/10.1175/WAF-D-12-00110.1>.
- , and —, 2015: Prediction of Intensity Model Error (PRIME) for Atlantic basin tropical cyclones. *Wea. Forecasting*, **30**, 1845–1865, <https://doi.org/10.1175/WAF-D-15-0064.1>.
- , —, A. B. Schumacher, and M. DeMaria, 2017: Improving tropical cyclone intensity forecasts with PRIME. *Wea. Forecasting*, **32**, 1353–1377, <https://doi.org/10.1175/WAF-D-17-0009.1>.
- Cangialosi, J. P., and J. L. Franklin, 2014: 2013 National Hurricane Center forecast verification report. NOAA/National Hurricane Center, 84 pp., https://www.nhc.noaa.gov/verification/pdfs/Verification_2013.pdf.
- , E. Blake, M. DeMaria, A. Penny, A. Latta, E. N. Rappaport, and V. Tallapragada, 2020: Recent progress in tropical cyclone intensity forecasting at the National Hurricane Center. *Wea. Forecasting*, **35**, 1913–1922, <https://doi.org/10.1175/WAF-D-20-0059.1>.
- 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>.
- , and J. Kaplan, 1994: A Statistical Hurricane Intensity Prediction Scheme (SHIPS) for the Atlantic basin. *Wea. Forecasting*, **9**, 209–220, [https://doi.org/10.1175/1520-0434\(1994\)009<0209:ASHIPS>2.0.CO;2](https://doi.org/10.1175/1520-0434(1994)009<0209:ASHIPS>2.0.CO;2).
- , and —, 1999: An updated Statistical Hurricane Intensity Prediction Scheme (SHIPS) for the Atlantic and eastern North Pacific basins. *Wea. Forecasting*, **14**, 326–337, [https://doi.org/10.1175/1520-0434\(1999\)014<0326:AUSHIP>2.0.CO;2](https://doi.org/10.1175/1520-0434(1999)014<0326:AUSHIP>2.0.CO;2).
- , M. Mainelli, L. K. Shay, J. A. Knaff, and J. Kaplan, 2005: Further improvements to the Statistical Hurricane Intensity Prediction Scheme (SHIPS). *Wea. Forecasting*, **20**, 531–543, <https://doi.org/10.1175/WAF862.1>.
- , and Coauthors, 2013: Improvements to the operational tropical cyclone wind speed probability model. *Wea. Forecasting*, **28**, 586–602, <https://doi.org/10.1175/WAF-D-12-00116.1>.
- , C. R. Sampson, J. A. Knaff, and K. D. Musgrave, 2014: Is tropical cyclone intensity guidance improving? *Bull. Amer. Meteor. Soc.*, **95**, 387–398, <https://doi.org/10.1175/BAMS-D-12-00240.1>.
- 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., 2007: Prediction of consensus tropical cyclone track forecast error. *Mon. Wea. Rev.*, **135**, 1985–1993, <https://doi.org/10.1175/MWR3390.1>.
- , and C. R. Sampson, 2014: Prediction of consensus tropical cyclone intensity forecast error. *Wea. Forecasting*, **29**, 750–762, <https://doi.org/10.1175/WAF-D-13-00058.1>.

- Hendricks, E. A., M. S. Peng, B. Fu, and T. Li, 2010: Quantifying environmental control of tropical cyclone intensity change. *Mon. Wea. Rev.*, **138**, 3243–3271, <https://doi.org/10.1175/2010MWR3185.1>.
- Jones, S. C., and Coauthors, 2003: The extratropical transition of tropical cyclones: Forecast challenges, current understanding, and future directions. *Wea. Forecasting*, **18**, 1052–1092, [https://doi.org/10.1175/1520-0434\(2003\)018<1052:TETOTC>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<1052:TETOTC>2.0.CO;2).
- Kaplan, J., and M. DeMaria, 1995: A simple empirical model for predicting the decay of tropical cyclone winds after landfall. *J. Appl. Meteor.*, **34**, 2499–2512, [https://doi.org/10.1175/1520-0450\(1995\)034<2499:ASEMFP>2.0.CO;2](https://doi.org/10.1175/1520-0450(1995)034<2499:ASEMFP>2.0.CO;2).
- , —, and J. A. Knaff, 2010: A revised tropical cyclone rapid intensification index for the Atlantic and eastern North Pacific basins. *Wea. Forecasting*, **25**, 220–241, <https://doi.org/10.1175/2009WAF2222280.1>.
- Krishnamurti, T. N., 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>.
- Marks, F. J., N. Kurkowski, M. DeMaria, and M. Brennan, 2019: Hurricane forecast improvement project years ten to fifteen strategic plan. NOAA, 83 pp., <https://hfip.org/sites/default/files/documents/hfip-strategic-plan-20190625.pdf>.
- Na, W., J. L. McBride, X.-H. Zhang, and Y.-H. Duan, 2018: Understanding biases in tropical cyclone intensity forecast error. *Wea. Forecasting*, **33**, 129–138, <https://doi.org/10.1175/WAF-D-17-0106.1>.
- 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](https://doi.org/10.1175/1520-0477(2000)081<1231:TATCFS>2.3.CO;2).
- 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>.
- Trabing, B. C., and M. M. Bell, 2020: Understanding error distributions of hurricane intensity forecasts during rapid intensity changes. *Wea. Forecasting*, **35**, 2219–2234, <https://doi.org/10.1175/WAF-D-19-0253.1>.
- Wood, K. M., and E. A. Ritchie, 2015: A definition for rapid weakening of North Atlantic and eastern North Pacific tropical cyclones. *Geophys. Res. Lett.*, **42**, 10091–10097, <https://doi.org/10.1002/2015GL066697>.