Probabilistic Multiple Linear Regression Modeling for Tropical Cyclone Intensity

CHIA-YING LEE

International Research Institute of Climate and Society, Columbia University, Palisades, New York

MICHAEL K. TIPPETT

*Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York,
and Department of Meteorology, King Abdulaziz University, Jeddah, Saudi Arabia*

SUZANA J. CAMARGO

Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York

ADAM H. SOBEL

*Department of Applied Physics and Applied Mathematics, Columbia University, New York, and
Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York*

(Manuscript received 28 May 2014, in final form 31 October 2014)

ABSTRACT

The authors describe the development and verification of a statistical model relating tropical cyclone (TC) intensity to the local large-scale environment. A multiple linear regression framework is used to estimate the expected intensity of a tropical cyclone given the environmental and storm conditions. The uncertainty of the estimate is constructed from the empirical distribution of model errors. NCEP–NCAR reanalysis fields and historical hurricane data from 1981 to 1999 are used for model development, and data from 2000 to 2012 are used to evaluate model performance. Seven predictors are selected: initial storm intensity, the change of storm intensity over the past 12 h, the storm translation speed, the difference between initial storm intensity and its corresponding potential intensity, deep-layer (850–200 hPa) vertical shear, atmospheric stability, and 200-hPa divergence. The system developed here models storm intensity changes in response to changes in the surrounding environment with skill comparable to existing operational forecast tools. Since one application of such a model is to predict changes in TC activity in response to natural or anthropogenic climate change, the authors examine the performance of the model using data that is most readily available from global climate models, that is, monthly averages. It is found that statistical models based on monthly data (as opposed to daily) with only a few essential predictors, for example, the difference between storm intensity and potential intensity, perform nearly as well at short leads as when daily predictors are used.

1. Introduction

The intensities of tropical cyclones (TCs) depend on their surrounding environments. The question of what distribution of storm intensities is consistent with a given environment is important both for short-range intensity forecasts and for long-term climate projections. In long-term climate projections, environmental factors that

influence TC intensity, such as the large-scale circulation and atmospheric moisture, are also expected to change. The issue of how these changes could influence the climatology of TC occurrence and intensity has been the subject of a number of recent studies, both from observational (Emanuel 2005; Webster et al. 2005) and from global climate modeling (Zhao et al. 2009; Bender et al. 2010; Knutson et al. 2010; Camargo 2013; Chen and Lin 2013; Emanuel 2013) perspectives. However, it remains difficult to confidently assess future changes in storm intensity. One of the primary reasons is that present climate models have horizontal resolutions that are too coarse to resolve the inner-core convective structure in

Corresponding author address: Chia-Ying Lee, International Research Institute for Climate and Society, Columbia University, 61 Route 9W, 202 Monell, Palisades, NY 10964.
E-mail: clee@iri.columbia.edu

storms. This deficiency prevents these climate models from simulating major storms at their observed intensities and providing a reasonable distribution of storm intensity, even in the current climate.

Downscaling is a strategy for connecting global-scale predictions and quantities not resolved by the global model, in our case, TC intensity. Dynamical downscaling with comprehensive models has been shown to be a very useful tool for understanding climate projections of TC activity (e.g., Knutson et al. 2007, 2008; Bender et al. 2010; etc.), but is both computationally expensive and limited in the range of intensities that can be simulated. Statistical downscaling is much cheaper computationally. This method has been used for seasonal prediction of North Atlantic hurricane activity by Vecchi et al. (2011), who developed a basinwide hurricane frequency forecast system based on a Poisson regression with two predictors: Atlantic main development region sea surface temperature (SST) and global tropical-mean SST from a suite of high-resolution global models. Their results from retrospective forecasts indicated that their statistical model can provide a skillful seasonal prediction. The hybrid statistical–dynamical downscaling system devised by Emanuel (2006), using an idealized dynamical model, is relatively inexpensive and provides comprehensive predictions, including information about the most extreme storms, but hinges on that particular idealized model (such that other comparable approaches are desirable if only for estimating model uncertainty) and is not publicly available for use by other investigators.

As there is always inherent uncertainty in TC projections and predictions, a probabilistic representation of this uncertainty should be provided. There are two general approaches for constructing probabilistic forecasts: ensemble and statistical approaches. In the ensemble approach, a set of forecasts are conducted from either parallel multiple dynamical or statistical models (Krishnamurti et al. 1999; Puri et al. 2001; Weber 2005), or a single numerical model under various initial conditions (Zhang and Krishnamurti 1999), physical parameterization schemes, or stochastic perturbations (Chen et al. 2014). In the statistical approach, the uncertainties are based on the probability density function (PDF) of the past errors or on forecasts from a deterministic numerical or statistical model (DeMaria et al. 2009). Vecchi et al. (2011) used both ensemble and statistical approaches to estimate the uncertainties.

From the perspective of short-range TC intensity forecasting, studies from DeMaria and Kaplan (1994, 1999), DeMaria et al. (2005, 2007), and DeMaria (2009) show that statistical models [e.g., Statistical Hurricane Intensity Prediction Scheme (SHIPS); Logistic Growth Equation Model (LGEM)] with the environmental

parameters from global numerical weather prediction models, in addition to climatological and persistence predictors, are capable of predicting storm intensity with some skill. These forecast models are essentially a form of statistical downscaling applied to individual storms in the present climate. The results of these studies suggest that statistical downscaling provides an alternative option for climate projections of the TC intensity distribution. A key component of this approach would be a statistical model to relate the global climate model environment to TC intensity. In other words, a statistical TC intensity model that responds correctly to changes in the surrounding (local) environment in the current climate could be applied to projections of future environmental conditions to produce projections of future TC intensity.

The National Hurricane Center (NHC) started issuing short-range wind speed probabilistic forecasts in 2006 based on a statistical approach (DeMaria et al. 2009, 2013). A Monte Carlo method generates 1000 forecast tracks based on random samples of track errors from the past 5 years. A similar method is applied to account for the uncertainties in storm intensity and structure. Emanuel et al. (2006) further used synthetic TC tracks generated from an algorithm with both deterministic and random components together with an idealized axisymmetric dynamical hurricane intensity model (run along the predicted track many times) to generate a set of intensity probabilistic forecasts for a specific geographic location. This approach has now been used to study tropical cyclone risk for a number of locations and from a number of perspectives (e.g., Lin et al. 2010).

Our ultimate goal is to understand the tropical cyclone intensity distribution in the projected climate scenarios (from global climate models) from a statistical perspective. The present study represents an initial step of our approach: the development of a probabilistic statistical model for the TC intensity distribution as a function of environmental variables. Following the example of DeMaria et al.'s work, we develop a SHIPS-like multiple linear regression model from global reanalysis fields derived from observations of the current climate. Different from SHIPS, we would like to select the smallest number of predictors possible, both for the sake of model parsimony and to minimize the number of required global climate model fields. Because our goal is to make projections of TC intensity under future climate change scenarios (although such projections are not yet attempted in this study), we avoid predictors whose utility is limited to current climate. SST is the most obvious example; the absolute value of SST may be a useful predictor for storm intensity in the present climate, but we expect the relationship of SST to TC intensity to

change in future climates (e.g., Vecchi and Soden 2007; Johnson and Xie 2010; Ramsay and Sobel 2011; Emanuel and Sobel 2013). We also determine whether monthly data can be used instead of daily, as this greatly reduces the data requirement for climate studies. Utilizing monthly data also requires a lower level of fidelity from global climate models.

Again, while the model we are developing is SHIPS-like, our goal is not a forecast model for real-time operational use, but a downscaling model for climate research. However, many of the development steps are the same, for example, parameter estimation, predictor selection, skill assessment, and uncertainty quantification. Detailed descriptions of data and the development of the multiple linear regression forecast model (MLR) will be given in sections 2 and 3, respectively. Examples of MLR forecasts and their verification will be shown in section 4. The dependence of MLR performance (errors) on predictors will also be discussed. In section 5, we will demonstrate a skillful and reliable probabilistic forecast from the MLR that is conducted with the past error distribution. In section 6, we assess the sensitivity of MLR development and its performance to the type of reanalysis data, including the possibility of using monthly averaged output. We summarize our findings in section 7.

2. Data

The best track dataset known as the revised Atlantic hurricane database (HURDAT2), produced by NHC for North Atlantic TCs, provides historical storm information, which includes storm location, maximum wind speed, and minimum sea level pressure (Jarvinen et al. 1984; Landsea and Franklin 2013). For synoptic conditions, we use the $2.5^\circ \times 2.5^\circ$ daily and monthly National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis dataset (Kalnay et al. 1996), and the $2.5^\circ \times 2.5^\circ$ 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) daily data (Uppala et al. 2005). In addition to atmospheric fields, we obtain upper-ocean structure from the $\sim 1.8^\circ \times 1.8^\circ$ NOAA/NCEP Environmental Model Center (EMC)/Climate Modeling Branch (CMB) Global Ocean Data Assimilation System (GODAS; Behringer and Xue 2004). We develop and test the MLR primary with the daily NCEP–NCAR reanalysis data. We utilize the data from 1981 to 1999 for model training—both predictor selection and coefficient estimation. The data for the period 2000–12 are used for testing model performance. When utilizing the daily ERA-40 data (for the sensitivity tests), the data from 1981 to 1992

are used for model training, and the data from 1993 to 2001 are for testing model performance.

We only consider storms that reach at least tropical storm strength (>34 kt; $1 \text{ kt} = 0.5144 \text{ m s}^{-1}$) during their lifetimes. As mentioned in DeMaria and Kaplan (1994), the statistical properties of storms over land are different from those over the ocean, so we only use the data for the period when storms are not over land. With these restrictions, there are more than 4000 cases of 12-hourly forecasts in both the training and testing datasets. The number of forecasts decreases with increasing lead time. Here we consider 12-hourly lead times from 12 to 120 h. There are only about 2000 cases left for both training and testing for the 120-h lead forecast. Additionally, the SHIPS TC forecasts from the Automated Tropical Cyclone Forecasting System (ATCF; Sampson and Schrader 2000) and the forecasts from the SHIPS development system (M. DeMaria 2014, personal communication) are used for evaluating the MLR performance.

3. Multiple linear regression model development

We begin by describing the initial pool of potential predictors for the MLR. We then choose a subset of those predictors for inclusion in the MLR using a forward selection (stepwise regression) procedure. The next step is to use a fitting procedure on the training data to derive the MLR, which consists of 10 linear equations for forecasts with 12-hourly lead times from 12 to 120 h.

a. Initial pool of predictors

Four variables are chosen that represent the storm itself: initial maximum wind speed (V_0), minimum sea level pressure (MSLP), change of storm maximum wind speed in the previous 12 h (dV/dt), and the storm translation speed (trSpeed). The environmental (synoptic) conditions are represented by the deep-layer vertical wind shear, moisture, environmental stability, outflow temperature, upper-tropospheric divergence, SST, and upper-ocean thermal structures.

The most important synoptic predictor in the SHIPS includes the maximum potential intensity (PI), which is derived empirically (DeMaria and Kaplan 1994). Here, we use Emanuel's PI definition developed in Emanuel (1988) and modified in Emanuel (1995) and Bister and Emanuel (2002). The PI is a function of the ratio of the exchange coefficient for enthalpy to that for momentum, the ratio of outflow temperature to SST, and the vertical integral of air temperature along iso-entropy surface at the radius of maximum wind speed. In Camargo et al. (2007, 2009), the importance of PI for predicting tropical cyclogenesis was shown using PI computed both from daily and from monthly NCEP–NCAR reanalysis data.

We use here PI from Camargo's dataset extended to 2012. PI is first averaged over a disk extending 500 km from the storm center and then averaged over the forecast interval (e.g., 12, 24, . . . , 120 h) along the observed storm track. PI enters the MLR as the difference of the mean PI and initial intensity V_0 , denoted dPI_V_0 .

We do not use SST as a predictor, nor any function of SST that includes specific threshold values based on observations of the historical relationship between SST and TC intensity. The relationship of SST to TC intensity is expected to change with the mean climate state (e.g., Vecchi and Soden 2007; Ramsay and Sobel 2011; Johnson and Xie 2010; Emanuel and Sobel 2013). The relative SST, or difference between the local SST and an appropriately defined tropical mean, may be a better predictor than absolute SST, but even that is limited in that it cannot (by definition) capture intensity changes that are related to global-mean changes (e.g., Emanuel and Sobel 2013). PI is a theoretically derived quantity whose formulation does not include any explicit tuning to the present climate, and thus in principle should be able to capture intensity changes that result from global climate changes. Emanuel's PI theory has been critiqued on theoretical grounds (e.g., Smith et al. 2008) but appears at this point to be as good as any existent theory (e.g., Bryan and Rotunno 2009). Our use of PI to the exclusion of any SST parameter thus appears to be a reasonable choice for our purpose. If advances in the theoretical understanding of TC intensity and its relationship to the environment are made in the future, our model can be easily rederived to use any improved PI parameter that may be developed.

Strong vertical wind shear tends to have a negative effect on TC intensification (DeMaria 1996; Tang and Emanuel 2012), and is another key predictor in the SHIPS (DeMaria et al. 2005). In this study, the deep-layer mean vertical wind shear (SHRD) is defined as the magnitude of the difference between the mean wind vectors at 200 and 850 hPa averaged over an annulus extending 200–800 km around the storm center (Chen et al. 2006). Moisture is another factor known to affect TC intensity. Kaplan and DeMaria (2003) found that rapidly intensifying TCs are often embedded in a high moisture environment. In the SHIPS, four variables representing relative humidity at various levels (1000, 850–700, 700–500, and 500–300 hPa) have been used in the past. Until 2013, only the 700–500-hPa mean relative humidity was used.¹ For the MLR, we initially consider two moisture variables: the 500–300-hPa relative

humidity (rhMid) and the column-integrated relative humidity (rhCol; Bretherton et al. 2004). The latter accounts for the total column water vapor from the top of the boundary layer to 200 hPa. In practice, it is most sensitive to the relative humidity of the lower troposphere, since the specific and saturation humidities are column-integrated first before taking the ratio, and most of the moisture is in the lower troposphere. High values in rhMid and rhCol favor TC development. For the environmental stability, we use the conditional instability ($d\theta_{es}/dz$), defined as the vertical gradient of saturated equivalent potential temperature ($d\theta_{es}$) averaged from boundary layer to 300 hPa. SHIPS considers convective instability (using equivalent potential temperature θ_e) instead of conditional instability, which is usually related to the lifting of an entire layer. The upper-tropospheric divergence is usually used to estimate the large-scale forcing, and we use the divergence field at 200 hPa (div200) for this purpose. While there has been some suggestion that convective available potential energy (CAPE) is not a good predictor for deep tropical convection (Zipser 2003), CAPE is included in our initial predictor pool. The last atmospheric environmental predictor is the temperature at 200 hPa (T200), representing the outflow temperature.

Similar to the environmental predictors calculated in the SHIPS, rhMid, rhCol, $d\theta_{es}$, CAPE, and T200 are first averaged over an annulus centered at the storm location at forecast time with an inner radius of 200 km and an outer radius of 800 km. Here div200 is averaged over a circle with radius of 1000 km. To account for the time variation of these variables, they are also averaged over the forecast interval and along the storm track, as described for dPI_V_0 . As we are using the best track data here, we do not apply a vortex removal scheme as used in the SHIPS (DeMaria 2010) to account for differences in the instantaneous Global Forecast System (GFS) forecast fields and the NHC forecast of the storm center.

In 2004, the SHIPS started incorporating predictors associated with the upper ocean (ocean heat content, ocean depth of 20° and 26°C) from satellite altimetry data to account for the oceanic negative impact on the storm (DeMaria et al. 2005). Here, instead of upper-ocean heat content, we use another air–sea interaction quantity, the ocean temperature averaged over the top 100 m (OT100; Price 2009), which represents the lowest SST that can be induced by a storm. As daily oceanic data are not readily available, monthly data from GODAS (Behringer and Xue 2004) are used here.

Our initial pool of predictors contains 13 variables: V_0 , MSLP, $dVdt$, trSpeed, dPI_V_0 , SHRD, rhMid and rhCol, $d\theta_{es}$, CAPE, div200 and T200, and OT100.

¹ http://rammb.cira.colostate.edu/research/tropical_cyclones/ships/docs/SHIPS_predictor_file_2013.doc.

FIG. 1. RMSE as a function of number of predictors used in the MLR for forecast lead time: (a) 12, (b) 24, (c) 48, (d) 72, (e) 96, and (f) 120 h. The predictors are listed on the right of each panel in the order that they are selected by the forward selection procedure. The order of the predictor selection is stable above the red line.

The change of the maximum wind speed over different forecast intervals (e.g., 12, 24, 48, 72, 96, and 120 h) is the predictand. The assumption of a linear relationship between intensity change and each predictor was tested using the same method as in [Tippett et al. \(2011, see their Fig. 12\)](#) and found (not shown) to be adequate.

b. Forward selection procedure

In the forward selection procedure, the initial model contains no predictors other than the constant term. The reduction obtained in root-mean-square error (RMSE) from adding a single variable is computed for each

variable, and the variable that most reduces the RMSE is added. This procedure is repeated until all the variables are added. The RMSE is calculated with 10-fold cross validation; the training data are divided into 10 random subsets, 9 are used to estimate regression coefficients and the tenth is used to compute an estimate of the RMSE. Here, random partitioning of the data is done 10 times, providing 100 estimates of the RMSE, similarly to [Tippett et al. \(2012\)](#). The mean RMSE as a function of the number of predictors for all lead times are shown in [Fig. 1](#). The predictors are listed on the right of each panel in [Fig. 1](#) in the order in which they enter

the forward selection procedure. A reasonable rule for the number of predictors to include in the MLR is to stop including predictors at the point where adding additional predictors does not significantly decrease the mean RMSE.

To estimate this stopping point robustly, we apply the forward selection procedure on the training data 10 times. While the order of the first few predictors is robust, the order of the remaining predictors is not and varies with the particular partition of the data. For example, for the 12-h forecast, the order of the 3rd–13th predictors varies randomly each time we run the forward selection procedure, indicating little difference in the utility of the remaining predictors. Hence, we add the second stopping rule at the point (red lines in Fig. 1) when the order of the selected predictors is stable for at least 7 out of 10 times that the forward selection procedure is run.

For 12-h forecasts (Fig. 1a), a substantial decrease in the RMSE occurs when the number of the predictors increases from one ($dVdt$) to two ($dVdt$ and dPI_{V_0}). Additional predictors result in little reduction of the RMSE, indicating that the intensification trend and the difference between PI and the initial intensity are the most important predictors for short-lead forecasts. The failure of other environmental predictors to reduce RMSE may indicate that the predictable impact of these environmental parameters, such as shear, is implicitly included in the previous 12-h intensity changes ($dVdt$). For a 24-h lead time, dPI_{V_0} is the most important predictor, followed by $dVdt$ and SHRD. Therefore, for forecast times shorter than 1 day, three predictors, $dVdt$, dPI_{V_0} , and SHRD, are adequate, and other predictors do not substantially reduce the RMSE.

The number of predictors required in the MLR (before RMSE stops decreasing) increases with the forecasting lead time. For the 72-h forecast, we need six predictors before the RMSE stops decreasing sharply, and the first seven predictors selected have a fixed order. These predictors are dPI_{V_0} , $trSpeed$, SHRD, V_0 , $dThetaEs$, $dVdt$, and $div200$. For the 96- and 120-h forecasts, however, the sharp decrease of RMSE happens at four, and only five predictors are selected in a fixed order each time we conduct the selection procedure. A possible reason could be that we need predictors other than those we have to further improve the 96- and 120-h intensity forecasts. Another possible reason is that there is a natural limitation of the linearity assumption for long-forecast intervals; that is, we cannot continue to improve the forecast from a multiple linear regression model with additional predictors, and a more sophisticated model is needed.

Overall we see that no more than seven predictors (dPI_{V_0} , $dVdt$, $trSpeed$, V_0 , SHRD, $div200$, and $dThetaEs$) are needed to achieve most of the possible reduction in RMSE. For simplicity and consistency, these seven predictors are used to develop MLR at all lead times. The absence of any moisture parameters is noteworthy and implies that these moisture variables do not add independent information. However, this result does not mean that moisture has no role. For instance, the impact of dry air is greatest when it reaches the core (Braun et al. 2012), which requires shear–dry air interaction (e.g., Ge and Li 2013; Tao and Zhang 2014). To capture such effect would require some other combination of shear and moisture parameters. Another possible reason is the use of the daily NCEP–NCAR reanalysis data. P. Klotzbach (2014, personal communication) found that the correlation between TC activity on the seasonal level and moisture calculated from ERA-Interim is much stronger than that calculated from NCEP–NCAR data. The sensitivity of parameter selection to the reanalysis dataset will be further addressed in detail later (section 6).

OT100, the mean upper-ocean temperature, is also not chosen. The analogous parameter in SHIPS, ocean heat content (OHC), yields 4%–8% improvement in the forecast error only when a threshold (50 kJ cm^{-2}) is applied (DeMaria et al. 2005). It is possible that we need to use some kind of threshold when including OT100. Nevertheless, the oceanic negative feedback also depends on $trSpeed$, which is selected here. Although T200 does not pass the selection test, the upper-tropospheric temperature is included in the PI calculation. CAPE is eliminated as well, which is not too surprising since it has been noted that it might not be a useful predictor for tropical convection (Zipser 2003). However, as we will show later in section 6, the utility of CAPE seems to depend on the choice of reanalysis dataset.

The signs of the coefficients (Fig. 2) are consistent with the physical relationships between predictors and TC intensity change, though such a result is not guaranteed for correlated predictors. Storms are more likely to follow their historical development and therefore an intensifying storm would continue intensifying, yielding a positive coefficient of $dVdt$. Increasing storm speed reduces the negative oceanic feedback, and the coefficient of $trSpeed$ is positive. There is a greater potential for a storm far from its PI value to continue intensifying, and the coefficient of dPI_{V_0} is positive. Stronger shear prevents storms from strengthening, and the coefficient of SHRD is negative. The stronger the divergence and the more unstable the atmosphere is, the better chance a system has to develop. Therefore the coefficients of $div200$ and $dThetaEs$ are positive and negative, respectively.

FIG. 2. Coefficients of predictors in the MLR (red) and the MLR_Monthly (blue).

4. MLR intensity forecasts and errors

a. Examples: Rita (2005), Earl (2010), Irene (2011), and Isaac (2012)

After reducing the number of predictors from 13 to 7, we test the MLR with the independent data from the period 2000–12. As examples, we show the intensity predictions for four storms, Hurricanes Rita (2005), Earl (2010), Irene (2011), and Isaac (2012); their tracks

are shown in Fig. 3. These storms encompassed a wide range of intensities, and have tracks that are typical for North Atlantic storms. Rita was a category 5 hurricane, Earl was a strong category 4, Irene was a category 3, and Isaac was a weak category 1. Both Rita and Earl underwent rapid intensification. Rita and Isaac passed over the Gulf of Mexico while Earl and Irene recurved and moved toward the northeastern United States.

FIG. 3. Tracks for Hurricanes Rita (2005), Earl (2010), Irene (2011), and Isaac (2012). The intensity predictions for each storm are shown in Fig. 4.

The MLR creates 5-day forecasts at 12-hourly intervals, and we apply it every 12 h during the lifetime of the storms (red lines in Fig. 4). The MLR overpredicts the intensity change (Isaac 18 to 21 August; Earl 25 to 28 August) while Isaac and Earl were in their early stage [tropical storm (TS) intensity]. Aside from failing to keep a storm in its TS stage, it is not capable of capturing rapid intensification, either. As addressed in DeMaria

and Kaplan (1994), this is the limitation of a simple linear regression model. For a case where the storm strengthens gradually (Irene), the MLR does a fair job (Fig. 4d). In the decaying period, when the intensities drop rapidly, it again has difficulty (after 22 September in Rita; after 3 September in Earl).

b. Verification of the MLR

To evaluate the overall performance of the MLR, we calculate the mean absolute error (MAE) and the RMSE against the best track data from all the testing cases (red solid line in Fig. 5a). Conventionally, for any TC intensity prediction model to be considered skillful, the model has to provide forecasts that are better than the Statistical Hurricane Intensity Forecast model (SHIFOR; Knaff et al. 2003), which predicts storm intensity changes based on climatology and storm persistence (gray dashed lines in Fig. 5a). However, the specific data used by the MLR (e.g., reanalysis rather than forecast fields) give an advantage to the MLR over any operational forecast, including SHIFOR. Hence, another baseline model, the persistence model whose forecast intensity at all leads is simply the initial intensity from best track data (gray solid line in Fig. 5a), is used here. While using the persistence model as a baseline mode is not what is usually done for a short-range TC intensity forecasts, it is not unusual for evaluating the probabilistic forecasts, which will be discussed soon in the next section (section 5). To ensure both the fairness and consistency, throughout this study, we are defining skill

FIG. 4. Five-day intensity forecasts made every 12 h by the MLR (red) of (a) Rita, (b) Isaac, (c) Earl, and (d) Irene up to landfall. The black line shows the best track intensity of each storm.

FIG. 5. (a) MAE of intensity predictions from the persistence (gray solid line), the SHIPS forecast (black solid line), the SHIPS-dependent (black dashed line), the MLR forecast (red solid line), and the MLR-dependent (red dashed line) models. (b) As in (a), but for RMSE. (c),(d) As in (a) and (b), but for different sets of experiments, which are described in Table 1.

with respect to the persistence model. The MLR has skill by this criterion at all lead times.

Both the MAE and RMSE of the MLR increase with lead time from 12- to 96-h forecasts, and they do not continue to grow from 96 to 120 h. In theory, we expect the error to increase with forecast time, but the error cannot grow indefinitely. In other words, there should be an upper bound on the forecast error, and we expect the error to level off after a certain lead time. However, although not significant, there is a small decline of the error from 96 to 120 h in the MLR, which could be due to the small sample size.

Errors from the SHIPS forecast model from 2000 to 2012 are also shown. Compared to the SHIPS, the MLR has smaller MAE and RMSE for the 12-h forecasts, and the 12-h SHIPS MAE is about 0.5 kt larger than that of the baseline model. This behavior is due to the fact that the persistence model (as well as the MLR) is using the storm track and initial intensity from best track data rather than the NHC intensity forecast (provided by ATCF), which is the best information available for the

SHIPS to use. Since all the models are verified against the best track data, it is reasonable that our baseline model and the MLR have smaller errors than does the SHIPS (as addressed earlier, the MLR has an unfair advantage over all operational forecast models).

To conduct a more fair comparison, errors from the SHIPS development model with the dependent data from 2000 to 2012 are shown (the SHIPS-dependent data, black dashed line in Fig. 5a; M. DeMaria 2014, personal communication), as well as the errors from the MLR with dependent data (MLR-dependent data, red dashed line in Fig. 5a). The SHIPS development model is trained with best track data and the GFS analysis fields. The SHIPS-dependent data have mean errors that are about 1 kt smaller than those in the MLR-dependent data for 24-h forecast. The difference in errors between these two models increases with time. At the 120-h lead time, it is more than 5 kt. A perfectly consistent comparison between the dependent versions of the MLR and the SHIPS is not possible. The SHIPS-dependent data use the GFS analysis fields, while the MLR-dependent

TABLE 1. Descriptions of experiments.

Experiment	Description
MLR	An MLR model with 7 selected predictors calculated from the NHC best track data and the daily NCEP–NCAR reanalysis fields.
dPI_ V_0	A linear regression model with only dPI_ V_0 as a predictor calculated from the NHC best track data and the daily NCEP–NCAR reanalysis fields.
dPI_ V_0 +dVdt	Same as dPI_ V_0 , but with one more predictor: dVdt.
...+SHRD	Same as dPI_ V_0 +dVdt, but with one more predictor: SHRD.
MLR_Monthly	Same as the MLR, but with the environmental predictors calculated from the monthly NCEP–NCAR reanalysis fields.
MLR_EC	Same as the MLR, but with environmental predictors calculated from the daily ERA-40 reanalysis fields.
MLR_1	A probabilistic intensity forecast model consists of the PDF with the mean value given by the MLR and spread given by the lead-time-dependent error distribution.
MLR_2	Same as the MLR_1, but with spread given by the lead-time and V_0 -dependent error distribution.
MLR_Monthly_2	Same as MLR_2, but with the mean value given by the MLR_Monthly.
MLR_EC_2	Same as MLR_2, but with mean value given by the MLR_EC.

data use the NCEP–NCAR reanalysis fields; the SHIPS-dependent data include not only storms that reached tropical storm strength in their lifetime (the cases that the MLR-dependent data are using), but also tropical depressions that never reach tropical storm intensity. Although the above discrepancies can result in different forecast errors, their contribution is small (not shown). A likely cause for the difference in the forecast errors between the MLR-dependent data and the SHIPS-dependent data may be that the SHIPS includes many more environmental parameters than the MLR does (until 2013, SHIPS had 24 predictors; Schumacher et al. 2013). We notice that the difference in errors between SHIPS and SHIPS-dependent data are much larger than that between MLR and MLR-dependent data. This result suggests that while having more predictors can significantly reduce the errors with a dependent dataset, it does not necessarily improve the model in the forecasting mode.

We conduct experiments to see how varying the numbers of predictors impacts the forecasting errors (Table 1). Figures 5c and 5d show the errors from three of these experiments. When using dPI_ V_0 (dPI_ V_0 : blue line) as the only predictor, we find (perhaps surprisingly) that the intensity errors, both MAE and RMSE, are not too far from those obtained with the full MLR. Adding dVdt (dPI_ V_0 +dVdt: solid sky-blue line) as the second predictor results in a small improvement. These results suggest that with only one or two predictors, the regression model still has some degree of skill. If we continue adding predictors, we continue to see an improvement in the MLR performance, such as adding SHRD (...+SHRD: dashed sky-blue line). Using all 13 predictors from the initial pool of predictors (MLR_13: orange line), however, results in larger errors for 24- to 60-h forecasts, an indication of overfitting. The experiments with different combination of predictors suggest

that a multiple linear regression model with even just a few of the most essential predictors has substantial skill.

c. Dependence of forecast errors on storm characteristics and synoptic conditions

Errors in intensity predictions (from both dynamic and statistical models) are related to conditions of the storm itself and that of the environment (i.e., the predictors), such as initial intensity, translation speed, latitude, current intensity, PI, and wind shear (Bhatia and Nolan 2013). The relationship between errors and predictors can differ among forecast times, and is not linear. Understanding the relationship between errors and predictors can help us to better interpret and utilize forecast results.

We stratify the errors by predictor value and forecast time (not shown). The weaker the initial storm intensity, the larger the error can be, especially for longer lead times. Strong storms are closer to their PI, and there is less potential for them to intensify, let alone undergo rapid intensification. Strong storms are also more resilient to shear, and are unlikely to undergo rapid weakening. Therefore, there is an upper bound on errors from both intensifying and decaying storms. On the other hand, we can expect that the MLR might have larger errors while predicting weak storms. It is also observed that the largest errors occur when dVdt is large (strong storm intensification in the previous 12 h), which is believed to be related to rapid intensification processes that are unresolved in the MLR. Larger errors also occur as storms move faster. Such situations usually occur when storms recurve in the midlatitudes. Another situation with possible larger errors occurs when the storms are in an environment that is favorable for rapid intensification (large PI and large positive dPI_ V_0). Forecast errors are also sensitive to shear and atmospheric

FIG. 6. A sparkline chart of the standard deviation (σ) of forecast errors for 24-h forecasts. The σ is calculated with data that have been binned based on the magnitude of each individual predictor, and therefore the σ of errors is a function of the predictor here. The unit of the x axis is one standard deviation of the individual predictor.

stability, and the largest errors occur in unstable environments with moderate shear.

To understand the relationship between the predictors and forecast errors, in Fig. 6 we show a sparkline chart of the standard deviation (σ) of forecast errors as a function of the individual predictor for 24-h lead time. The larger the standard deviation's variability over the individual predictor's magnitude range, the more sensitive is the error to the predictor. Among all predictors, forecast errors seem to be most sensitive to the initial storm intensity, somewhat less to PI and dPI_ V_0 , and not very sensitive to the remaining predictors.

5. Probabilistic intensity prediction

a. Forecast distribution

As described in the introduction, accounting for the uncertainty of the MLR forecast increases its value and utility, because we want to know the distribution of storm intensities. The simplest and most direct estimate of forecast uncertainty is given by the empirical distribution of errors presented in the training process (using all the training data). This approach is feasible because

of the size of the training dataset used here. Although we recognize that rare events may be poorly sampled and may benefit from parametric descriptions, we only use the empirical distribution of errors here. The error distribution depends on forecast lead time, and Fig. 7a shows the increasing spread of the error PDF as the forecast lead time increases from 12 to 72 h. There is no significant change in the error PDF between 96 and 120 forecast hours; the errors appear to have saturated at that point. The errors are close to being normally distributed for all lead times. Figure 7b is the cumulative distribution function (CDF) and Fig. 7c compares the distribution of the MLR errors for a 24-h forecast to the standard normal distribution. Except for the extreme points, most of the points fit well within the theoretical normal distribution profile. All the large negative errors (< -35 kt) are from the cases with rapid intensification. For instance, the largest negative error in Fig. 7b is about -86 kt, and corresponds to a forecast of Hurricane Emily in 1987 when Emily underwent rapid intensification. Our probabilistic intensity forecasts consist of the PDF with mean value given by the MLR and spread given by the lead-time-dependent error distribution (Fig. 8).

The analyses in section 4c suggest that the forecast error and its distribution are sensitive to the predictors, and are most sensitive to V_0 (initial storm intensity). Strictly speaking, such dependence is beyond the linear regression framework, in which the error variance is assumed independent of the predictors. However, regression-based models with varying spread are common in weather forecasting (Gneiting et al. 2005; Wilks and Hamill 2007). Here, we examine the sensitivity of forecast error distribution to V_0 . We illustrate the dependence of the error distribution on initial wind speed in the case of the 24-h forecast error (Fig. 9). The standard deviation of all 24-h forecast errors from all training data is 12 kt (dashed line in Fig. 9a, which is calculated from the blue line in Fig. 7a). However, binning 24-h forecast errors based on the corresponding values of V_0 shows that the error standard deviation increases, roughly linearly, as V_0 increases (black diamond line in Fig. 9c). Now, if we do not account for the fact that the distribution of forecast error is a function of V_0 , the joint CDF of errors between lead times and V_0 would look like Fig. 9b. If we do, the joint CDF would look like Fig. 9c. To understand how accounting for the dependence of error distribution to V_0 changes the results, we develop two probabilistic models. We name the probabilistic model with error distribution as a function of lead time only MLR_1, while the one utilizing error distribution as a function of both lead time and V_0 is named MLR_2 (Table 1).

FIG. 7. (a) PDF of the MLR forecast errors from all the training data. (b) As in (a), but for CDF. In (a) and (b), colors indicate various forecast times as listed on the legend. (c) Quantile–quantile plot showing the fit of the MLR errors for 24-h forecast (from training data) to the normal distribution. The quantiles of the MLR errors are shown on the y axis while the quantiles of the normal distribution are on the x axis. The unit of the x axis is one standard deviation of MLR errors.

In addition to MLR_1 and MLR_2, we construct another probabilistic model as a baseline reference model from the persistence model (Fig. 5a) and its error distribution (Fig. 8b). Figure 8b indicates that the mean of the distribution is near zero at any given lead time and that the spread increases with lead time. The spread here is larger than the spread in Fig. 8a. This is consistent with results in Fig. 5b: the persistence model error standard

deviation is larger than that in the MLR. We call this model the probabilistic persistence model. When making a probabilistic forecast with the persistence model, we center the distribution in Fig. 8b on the initial intensity.

The MLR_1 and MLR_2 are considered skillful to the extent to which their probabilistic performance is better than that of the probabilistic persistence model.

FIG. 8. Joint CDF (%) of the magnitude of errors and lead time from (a) the MLR_1, and (b) the probabilistic persistence model based on training data. In both (a) and (b), the white dashed lines indicate where the mean error is, and the black solid and dashed lines show where the highest 10% and 50% probabilities are.

FIG. 9. (a) Standard deviation (σ) of error distribution for 24-h intensity predictions from the training data. Black line is calculated with all the errors (MLR_1), while black diamond line is calculated with errors that have been binned based on the initial storm intensity (V_0) (MLR_2). (b) Joint CDF (%) from MLR_1. Note that the CDF is not a function of V_0 . (c) As in (b), but now the CDF is calculated from MLR_2 and is also a function of V_0 .

The fundamental difference is that the probabilistic persistence distribution is conditioned only on the current intensity and historical changes in storm intensity, while the MLR_1 and MLR_2 distributions are

additionally conditioned on the current and forecast environment.

b. Probabilistic forecasts and verification

Probabilistic forecasts (persistence, MLR_1, and MLR_2) are made for all the cases in the testing dataset. Figure 10 shows probabilistic forecasts for Hurricane Earl initialized at 0000 UTC 28 August 2010, a period during which Earl underwent rapid intensification (36–60 forecast hours). For the 12–36-h lead times, the observed storm intensity falls into the 25–75 percentile region in MLR_1 and MLR_2 forecast, while in the persistence model it does not. Once Earl started to undergo rapid intensification, the best track data fall beyond the highest 50 percentile region in all models, although they are located at higher probabilities in MLR_1 and MLR_2 than in the persistence model. Qualitatively, the case in Figs. 10a–c exhibits the advantage in using MLR_1 and MLR_2 instead of the persistence model for probabilistic forecast. Knowing current and forecast environmental conditions helps probabilistic intensity prediction.

We use the rank probability skill score (RPSS) to verify quantitatively whether the MLR_1 and MLR_2 have more skill than the persistence model. Rank probability score (RPS) is a squared-error score with respect to the observational probability, which is 1 if the forecast event occurs, and 0 if the event does not occur. The cumulative forecasts and observations, denoted as P_m and O_m , are

$$P_m = \sum_{j=1}^m p_j, \quad m = 1, \dots, J, \quad \text{and} \quad (1)$$

$$O_m = \sum_{j=1}^m o_j, \quad m = 1, \dots, J, \quad (2)$$

where p_j is the forecast probability of the storm intensity falling in the j th bin, the observed probability o_j is 1 if the observations fall in the j th bin and zero otherwise, and J is the total number of bins. Here, we stratify the probabilities by the corresponding $V_{\max}$ every 5 kt from 0 to 200 kt. The RPS is the sum of the squared differences between the cumulative probabilities P_m and O_m :

$$\text{RPS} = \sum_{m=1}^J (P_m - O_m)^2. \quad (3)$$

RPS is oriented so that smaller values indicate better forecasts. A correct forecast with no uncertainty has a RPS of 0. The RPSS compares the average RPS to that of the persistence forecast:

FIG. 10. (a) Intensity probabilistic prediction in CDF for Hurricane Earl (2010) based on the probabilistic persistence model (Fig. 8b). (b) As in (a), but the CDF is based on the MLR_1 (Figs. 8a and 9b). (c) As in (b), but the CDF is based on the MLR_2 (Fig. 9c). The initial time for (a)–(c) is 0000 UTC 28 Aug 2010. The white solid lines indicate the observed maximum wind speed, while the white dashed lines are the MLR and the persistence model predicted intensity, which is also where the mean error is in Fig. 8. Black solid and dashed lines show where the highest 10% and 50% probabilities are. (d) RPSS of the two MLR probabilistic predictions relative to those from the probabilistic persistence model for all forecasts in Earl's life cycle.

$$\text{RPSS} = 1 - \frac{\overline{\text{RPS}}}{\overline{\text{RPS}}_{\text{persistence}}}. \quad (4)$$

If the probabilistic model is skillful, RPSS is positive, and the larger the value is, the more skill the model has. The RPS and RPSS can be computed for each forecast. Figure 10d shows the RPSS for all of the forecasts of Earl. Compared to the persistence model, the skill of MLR_1 and MLR_2 increases with the length of forecast interval, reflecting the fact that the value of knowing the environment increases as lead time increases. As the forecast interval becomes longer, the environmental conditions become more crucial for intensity prediction. Hence, the MLR_1 and MLR_2 forecasts show greater skill compared to the persistence model with increasing lead time. For all the testing cases (Fig. 11), RPSS for the MLR_1 and MLR_2 again increases with time. The modest advantage in RPSS that the MLR_2 shows over the MLR_1 indicates that considering the sensitivity of forecast errors to initial storm intensity does improve probabilistic intensity predictions.

The performance of probabilistic forecasts of particular intensity categories (TS, hurricane intensity categories 1 and 2 and 3–5) can also be verified. Figure 12

shows the reliability diagrams of MLR_1, MLR_2, and the persistence model for categorized cases from 2008 to 2012 at all lead times. Forecast probabilities at each forecast time for each intensity category are grouped into bins at 0.1 (10%) intervals. For reference we show the reliability from the NHC official probabilistic forecast (OFCL). We emphasize again that because we use

FIG. 11. RPSS of the two MLR probabilistic predictions relative to those from the probabilistic persistence model for all testing cases.

FIG. 12. Reliability diagrams showing observed frequency as a function of forecast probabilities for (a) tropical storms (TSs), (c) category 1 and 2 hurricanes, and (e) category 3–5 hurricanes, respectively. Colors indicate forecasts from MLR_1 (pink), MLR_2 (red), persistence (gray), and NHC (OFCL, black) from 2008 to 2012. (b), (d), (f) The sample size used in (a), (c), and (e) for each forecast.

best track data and reanalysis fields, this comparison does not indicate the true difference in reliability between MLR and OFCL given equivalent input data; we show it nonetheless because no exactly appropriate reference is available.

The probabilistic forecasts of the occurrence of TS and hurricane categories 1 and 2 in the Saffir–Simpson scale by the two MLR models, and persistence model all show a fair reliability as the observed frequency is close to the forecast probability (Figs. 12a,c). The MLR_1, MLR_2, and the persistence model forecasts of TS

occurrence show almost perfect reliability for high probabilities (>0.6 bin) and underforecasting for low probabilities. MLR_1 and MLR_2 slightly overpredict occurrence of categories 1 and 2 hurricanes reliably for forecasts probabilities of 0–0.3 (0%–30%), while the persistence model underforecasts it. For higher probabilities (>0.3 bin), they all overforecast the occurrence.

For major hurricanes (categories 3–5), the reliability diagram curves are very different among these three forecast models. MLR_1 and MLR_2 are still fairly reliable. The persistence model, however, is substantially

less reliable, overforecasting the occurrence of major hurricanes. For the persistence model, the probabilistic forecast always maintains relatively higher probabilities close to the initial storm intensity, as shown in Fig. 10a. Hence, the persistence model could easily overforecast decaying major storms. Compared to the MLR_1, MLR_2 is more reliable (the red line in Fig. 12 is closer to the one-to-one line than the pink line).

6. Sensitivity of MLR performance to the reanalysis data

NCEP–NCAR daily reanalysis dataset is initially chosen for use in this study because of its temporal coverage and similarity to the GFS analysis fields used for the SHIPS development system. However, a TC and its surrounding environment may be presented differently in other global reanalysis datasets (Schenkel and Hart 2012). This is expected to affect the relationships between TC intensity change and the environmental predictors. The sensitivity of our findings to the choice of reanalysis dataset is explored here.

Given our interest in the variations of environment and TC intensity on long time scales, an interesting question is to what extent the MLR can predict intensity changes given monthly rather than daily data. In an application where global climate model output is used, for example, being able to use monthly data is a great advantage. Furthermore, we expect that global climate models may simulate time-averaged fields better than they simulate synoptic variability. Hence, in the first part of this section, we use monthly NCEP–NCAR reanalysis data to calculate environmental predictors to assess the ideal of using low-frequency reanalysis data.

In section 3b, we noted that one of the possible reasons why moisture variables are not selected as predictors is because of the specific data (daily NCEP–NCAR) used. Analysis from P. Klotzbach (2014, personal communication) found that the correlation between hurricane activity and atmospheric moisture is positive (0.46) using the ERA-Interim reanalysis, while it is much weaker using the NCEP–NCAR reanalysis. Despite the fact that ERA-Interim is the latest and improved reanalysis data from ECMWF, daily PI from ERA-Interim is not currently available, and the calculation is time consuming. On the other hand, PI from an older version of reanalysis data from ECMWF, daily ERA-40, is available from Camargo et al. (2009). In the second part of this section, we will therefore use ERA-40 for the calculation of environmental predictors to understand how sensitive the MLR is to global reanalysis models.

a. Monthly versus daily NCEP–NCAR

With the assumption that monthly averaged fields represents conditions on the 15th of each month, the monthly reanalysis fields are first linearly interpolated to the day of the forecast time. The same method has been used for studying TC statistics (Emanuel 2000), and TC risk assessment (Emanuel et al. 2006). Interpolating data onto the day instead of holding the data constant throughout a month avoids any discontinuity when a storm lives across two months. Then, the predictors are linearly interpolated to the storm location at the forecast time, in the same manner as the daily data.

The choice of environmental predictors is expected to be sensitive to the characteristics of the reanalysis data, which would be different for monthly and daily data. If we apply the forward selection procedure with the monthly predictors, div200 and dThetaEs are left out and rhMid is included in addition to dPI_V₀, and SHRD (not shown). However, to stay focused on testing the ability of the MLR and its probabilistic model using low-frequency data, we use the same predictors (dPI_V₀, dVdt, trSpeed, V₀, SHRD, div200, dThetaEs) obtained from forward selection with the daily data (section 3b).

Now, with the monthly NCEP–NCAR reanalysis data and the seven predictors listed above, a monthly linear regression model (called MLR_Monthly) is trained and tested. The forecast errors of the MLR_Monthly are comparable to those of the MLR (Figs. 13a,b). There is almost no discrepancy between the MAE (and RMSE) of the MLR and the MLR_Monthly for lead times less than 24 h. For longer lead times (>24 h), the advantage of using daily data is clear, but not significant. The maximum difference in forecast errors between monthly and daily is less than 2 kt. Such a good performance of the MLR_Monthly on one hand suggests the difficulty of TC intensity prediction: there is no dramatic improvement from using more frequent data, which are assumed to provide a more accurate description of TC environment. In addition, there might be some usable relation between the daily and monthly predictors. A similar finding was also noted in Emanuel et al. (2004). With an idealized coupled hurricane model in which the pres-torm PI is one of the inputs, they found a slight, but statistically insignificant, decrease in the forecast errors (at the time of storm's peak intensity) when daily PI, as opposed to climatological monthly mean PI, is used (see their Fig. 16).

Comparing the predictors from the monthly data to those from the daily data, we find that the monthly dPI_V₀ is larger than the daily one (Fig. 14), and that this difference is a result of larger monthly PI. But, daily and monthly PI are highly correlated. There are

FIG. 13. (a) MAE of intensity prediction from the persistence (gray), the MLR (red), the MLR_Monthly (blue), and the MLR_EC (orange). (b) As in (a), but for RMSE. (c) Reliability diagrams showing observed frequency as a function of forecast probabilities for category 3–5 hurricanes. (d) The sample size used in (c). Colors again indicate result from the MLR_2 (red), the persistence model (gray), and the MLR_Monthly_2 from 2008 to 2012, and the MLR_EC_2 (orange) from 1997 to 2001.

some positive biases for smaller values of SHRD and $d\Theta_{Es}$, but negative biases for larger values in the monthly data. For the last environmental predictor, $div200$, there is almost no relationship between monthly and daily data. This is because the monthly $div200$ is too small (and uncorrelated) compared to the daily one, which could explain why $div200$ would be left out if we applied a forward selection procedure with the monthly data. With the noticeable differences between monthly and daily predictors, the training processes of the MLR and the MLR_Monthly give them different weights (coefficients, Fig. 2). The coefficients of the three storm persistence predictors in the MLR and the MLR_Monthly are similar to each other, as well as those of $d\Theta_{Es}$. The MLR_Monthly weights the impact of dPI_{V_0} on intensity change almost twice as large as the MLR does for lead times smaller than 48 h. However, for the longer lead time, the MLR_Monthly weights dPI_{V_0} less than the MLR does. The MLR_Monthly also have smaller coefficients in SHRD and $div200$ for all lead times.

To assess which variable is the primary cause of the larger error in the MLR_Monthly, we replace each of the four environmental predictors with daily data in the MLR_Monthly separately (not shown). Our results suggest that the difference between monthly and daily values of dPI_{V_0} is the main cause for the differences in the performance of the MLR and MLR_Monthly. We examine also the performance of probabilistic forecast with the MLR_Monthly with the error distribution as function of both initial storm intensity and lead time. For tropical storms and hurricanes categories 1 and 2, there is no significant difference between the MLR_2 and the MLR_Monthly_2 (not shown). For major hurricanes (Figs. 13c,d), the MLR_2 performs better (more reliable) than the MLR_Monthly_2 does.

Results from the MLR_Monthly, and the MLR_Monthly_2 suggest that, statistically, monthly averaged reanalysis fields capture variations in environment that are important to TC intensity. In other words, monthly averaged data seem to be sufficient for predicting storm intensity changes. The improvement in both intensity

FIG. 14. Scatterplots overlaid on 2D histograms of daily (x axis) and monthly (y axis) environmental predictors used in MLR: (a) dPI_V_0 , (b) SHRD, (c) $div200$, and (d) $dThetaEs$. The black solid lines indicate the best-fit lines, while the black dashed lines are the one-to-one lines.

forecasts, and probability prediction from using higher-frequency data (daily) is modest.

b. Daily NCEP–NCAR versus daily ERA-40

Similar to [sections 3a](#) and [3b](#), we calculate all the initial pools of environmental predictors from the daily ERA-40 data, and then apply the forward selection procedure with these variables (not shown). For shorter lead times (<72 h), $dVdt$ and dPI_V_0 are the most important two predictors, followed by SHRD, and V_0 . The value of $rhCol$ becomes important with increasing lead time, and is the fifth and sixth most important predictors for 24- and 48-h forecasts, respectively. Even for a longer lead time (≥ 72 h), $rhCol$ is picked by the selection procedure. In addition to $rhCol$, CAPE seems to be

a predictor that cannot be ignored with ERA-40. The predictors $dThetaEs$ and $rhMid$ are chosen only for the 120-h forecast. The predictor $trSpeed$ is also very important for a longer lead time; it is the second most important predictor for the 72-h forecast, and the fourth and third most important predictor for the 96- and 120-h forecasts.

Generally speaking, three predictors related to storm persistence ($dVdt$, V_0 , and $trSpeed$) have a statistically significant impact on reducing forecast errors with either NCEP–NCAR or ERA-40 daily data. Among all environmental predictors, the influence of dPI_V_0 and SHRD are robust. The biggest difference in the choice of predictors with ERA-40 daily compared to NCEP–NCAR daily is that ERA-40 chooses $rhCol$ for 48–96-h

lead time. For 120-h lead time, the forward selection procedure picks another moisture variable, rhMid , with ERA-40. It leaves 200-hPa divergence out for all forecast lead times while div200 is picked for 72 and 96 h with the daily NCEP–NCAR data.

With ERA-40, we select eight predictors: dPI_{V_0} , dVdt , trSpeed , V_0 , SHRD , rhCol , CAPE , and dThetaEs . Although rhMid is picked for the 120-h forecast, we choose not to add another moisture variable here. The multiple linear regression model trained and tested with these eight predictors from the ERA-40 daily data is called the MLR_{EC} . For all lead times, forecast errors (MAE and RMSE) from the MLR_{EC} are 2–3 kt smaller than those from the MLR (Figs. 13a,b). The reliability diagram suggests an improvement of the probabilistic intensity forecast model (MLR_{EC_2}) for tropical storms and hurricane categories 1 and 2 when compared to the MLR_2 (not shown). For major hurricanes, however, the MLR_{EC_2} is less reliable for small probabilities than the MLR_2 is, but it is more reliable for moderate probabilities (Figs. 13c,d).

The comparison among the three choices of predictors using the daily NCEP–NCAR, the monthly NCEP–NCAR, and the daily ERA-40 shows that while there is an obvious dependence between the choice of predictors to the type of reanalysis data used, dPI_{V_0} and SHRD , have to be considered no matter which data are used. This is probably because the relationship between PI and SHRD , and storm evolution is less sensitive to either the frequency of the data or the different global reanalysis model. The importance of the other environmental predictors, atmospheric moisture, upper-level divergence, atmospheric stability, etc., is dataset dependent. Because there is no simple way to assess the utility of environmental variables from global climate models, predictors selection could potentially be an issue when we apply this model to the global climate model fields. Given the sensitivity of the predictor selection to the choice of reanalysis, it seems reasonable to limit the model to the most robust predictors.

7. Summary

This study describes the development of a statistical downscaling model for short-term probabilistic tropical cyclone (TC) intensity prediction. We use a multiple linear regression model (MLR) to model intensity change and the past error distribution to account for uncertainty. While such models are relevant to operational intensity forecasts, our interest is more directed at the capability of such a model to downscale TC intensity from the large-scale environment. The ability of the

MLR to relate variations in the large-scale environment to TC intensity has applications beyond short-term forecasting and can be a tool toward understanding the distribution of TC intensities in a climate undergoing natural or anthropogenic changes.

The MLR contains seven predictors selected based on their physical importance to TC development and their ability to reduce forecast errors. The predictors are the initial maximum wind speed (V_0), change of storm intensity in the past 12 h (dVdt), storm translation speed (trSpeed), the difference between PI and V_0 (dPI_{V_0}), 850–200-hPa vertical wind shear magnitude (SHRD), conditional stability (dThetaEs), and 200-hPa divergence (div200). Among these predictors, dVdt is the most important predictor for forecasts with short lead times (<24 h); dPI_{V_0} and SHRD become more important with increasing lead times. We avoid predictors whose relationship to TC intensity is likely to change in a changed global-mean climate, such as sea surface temperature (SST). Different reanalysis products (e.g., NCEP–NCAR vs ERA-40) and averaging frequency (e.g., monthly vs daily) result in different predictors being chosen because the TC intensity and its surrounding environment are represented differently. Three storm characteristic predictors (V_0 , dVdt , and trSpeed), and the two environmental predictors (dPI_{V_0} and SHRD) are included regardless of which reanalysis datasets are used. The importance of other environmental predictors on TC development depends on the reanalysis dataset.

We assess the model performance using MAE and RMSE. However, because the MLR uses data unavailable to operational forecasts (e.g., the reanalysis rather than forecast fields), a direct comparison with operational models is not appropriate. The MLR is verified against the persistence model, and it is found that the MLR has skill for all lead times. A relatively simple description of the synoptic environment is able to characterize much of the predictable component of the variability of TC intensity. The errors of a single predictor (dPI_{V_0}) model are not too far from those of the full seven-predictor model. Likewise, a model based on monthly averaged environment performs quite well, with its primary weakness being related to the lack of daily PI information.

In addition to the deterministic MLR, we produce probabilistic information by constructing an intensity PDF with the mean given by the MLR and uncertainty given by the empirical error distribution from the training data. Two probabilistic models are developed here. In the MLR_1 , the error distribution is only a function of lead time, while in the MLR_2 it is also a function of initial storm intensity, reflecting the

dependence of forecast errors on the environment. A reference probabilistic forecast is constructed from historical storm intensity changes (independent of environment) and used to compute the rank probability score skill (RPSS). Both the MLR_1 and the MLR_2 have good skill, especially for long-lead forecasts when their advantage over the persistence forecast is greatest. The analysis of the reliability diagram further shows that the reliability among two MLR probabilistic models and the persistence model are similar to each other for TS and hurricanes of categories 1 and 2. Nevertheless, the MLR_1 and the MLR_2 are much more reliable than the persistence model for major hurricanes. Furthermore, both RPSS and reliability analyses suggest that using an environment-dependent error distribution (MLR_2) improves the forecast skill. Probabilistic intensity prediction based on the monthly averaged environment also has fairly good skill comparable to the MLR_2.

In summary, in this study we have developed a downscaling system that reliably predicts the distribution of storm intensity conditional on the surrounding local environment. We have developed and verified the system in a simulation framework using regression methods and probabilistic verification to assess resolution and reliability of the intensity distribution. Our results further suggest that a model based on monthly averages of the most essential predictors is sufficient to capture the distribution of TC intensities and could be applied to global climate model output for future risk assessment.

Acknowledgments. We thank Dr. Mark DeMaria for useful comments and sharing data from SHIPS development system. Comments from editor Dr. Patrick Harr, reviewer Dr. Phil Klotzbach, and the other anonymous reviewer have helped improve the manuscript. The research was supported by the Office of Naval Research under the research grant of MURI (N00014-12-1-0911).

REFERENCES

- Behringer, D. W., and Y. Xue, 2004: Evaluation of the global ocean data assimilation system at NCEP: The Pacific Ocean. *Eighth Symp. on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface*, Seattle, WA, Amer. Meteor. Soc., 2.3. [Available online at <https://ams.confex.com/ams/pdfpapers/70720.pdf>.]
- Bender, M. A., T. R. Knutson, R. E. Tuleya, J. J. Sirutis, G. A. Vecchi, S. T. Garner, and I. M. Held, 2010: Modeled impact of anthropogenic warming on the frequency of intense Atlantic hurricanes. *Science*, **327**, 454–458, doi:[10.1126/science.1180568](https://doi.org/10.1126/science.1180568).
- 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, doi:[10.1175/WAF-D-12-00110.1](https://doi.org/10.1175/WAF-D-12-00110.1).
- 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, doi:[10.1029/2001JD000776](https://doi.org/10.1029/2001JD000776).
- Braun, S. A., J. A. Sippel, and D. S. Nolan, 2012: The impact of dry midlevel air on hurricane intensity in idealized simulations with no mean flow. *J. Atmos. Sci.*, **69**, 236–257, doi:[10.1175/JAS-D-10-05007.1](https://doi.org/10.1175/JAS-D-10-05007.1).
- Bretherton, C. S., M. E. Peters, and L. E. Back, 2004: Relationships between water vapor path and precipitation over the tropical oceans. *J. Climate*, **17**, 1517–1528, doi:[10.1175/1520-0442\(2004\)017<1517:RBWVPA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<1517:RBWVPA>2.0.CO;2).
- Bryan, G. H., and R. Rotunno, 2009: Evaluation of an analytical model for the maximum intensity of tropical cyclones. *J. Atmos. Sci.*, **66**, 3042–3060, doi:[10.1175/2009JAS3038.1](https://doi.org/10.1175/2009JAS3038.1).
- Camargo, S. J., 2013: Global and regional aspects of tropical cyclone activity in the CMIP5 models. *J. Climate*, **26**, 9880–9902, doi:[10.1175/JCLI-D-12-00549.1](https://doi.org/10.1175/JCLI-D-12-00549.1).
- , K. A. Emanuel, and A. H. Sobel, 2007: Use of a genesis potential index to diagnose ENSO effects on tropical cyclone genesis. *J. Climate*, **20**, 4819–4834, doi:[10.1175/JCLI4282.1](https://doi.org/10.1175/JCLI4282.1).
- , M. C. Wheeler, and A. H. Sobel, 2009: Diagnosis of the MJO modulation of tropical cyclogenesis using an empirical index. *J. Atmos. Sci.*, **66**, 3061–3074, doi:[10.1175/2009JAS3101.1](https://doi.org/10.1175/2009JAS3101.1).
- Chen, J.-H., and S.-J. Lin, 2013: Seasonal predictions of tropical cyclones using a 25-km-resolution general circulation model. *J. Climate*, **26**, 380–398, doi:[10.1175/JCLI-D-12-00061.1](https://doi.org/10.1175/JCLI-D-12-00061.1).
- Chen, S. S., J. A. Knaff, and F. D. Marks, 2006: Effects of vertical wind shear and storm motion on tropical cyclone rainfall asymmetries deduced from TRMM. *Mon. Wea. Rev.*, **134**, 3190–3208, doi:[10.1175/MWR3245.1](https://doi.org/10.1175/MWR3245.1).
- , F. Judt, J. Berner, C.-Y. Lee, M. Curcic, C. Snyder, and R. Rotunno, 2014: Distinct characteristics of hurricane ensemble forecasts using physical parameterizations vs. stochastic perturbations. *26th Conf. on Weather Analysis and Forecasting/22nd Conf. on Numerical Weather Prediction*, Atlanta, GA, Amer. Meteor. Soc., J4.6. [Available online at <https://ams.confex.com/ams/94Annual/webprogram/Paper239878.html>.]
- DeMaria, M., 1996: The effect of vertical shear on tropical cyclone intensity change. *J. Atmos. Sci.*, **53**, 2076–2088, doi:[10.1175/1520-0469\(1996\)053<2076:TEOVSO>2.0.CO;2](https://doi.org/10.1175/1520-0469(1996)053<2076:TEOVSO>2.0.CO;2).
- , 2009: A simplified dynamical system for tropical cyclone intensity prediction. *Mon. Wea. Rev.*, **137**, 68–82, doi:[10.1175/2008MWR2513.1](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. [Available online at <https://ams.confex.com/ams/pdfpapers/167916.pdf>.]
- , and J. Kaplan, 1994: A Statistical Hurricane Intensity Prediction Scheme (SHIPS) for the Atlantic basin. *Wea. Forecasting*, **9**, 209–220, doi:[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, doi:[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, doi:[10.1175/WAF862.1](https://doi.org/10.1175/WAF862.1).
- , J. A. Knaff, and C. Sampson, 2007: Evaluation of long-term trends in tropical cyclone intensity forecasts. *Meteor. Atmos. Phys.*, **97**, 19–28, doi:[10.1007/s00703-006-0241-4](https://doi.org/10.1007/s00703-006-0241-4).
- , —, R. Knabb, C. Lauer, C. R. Sampson, and R. T. DeMaria, 2009: A new method for estimating tropical cyclone wind speed probabilities. *Wea. Forecasting*, **24**, 1573–1591, doi:[10.1175/2009WAF2222286.1](https://doi.org/10.1175/2009WAF2222286.1).
- , and Coauthors, 2013: Improvements to the operational tropical cyclone wind speed probability model. *Wea. Forecasting*, **28**, 586–602, doi:[10.1175/WAF-D-12-00116.1](https://doi.org/10.1175/WAF-D-12-00116.1).
- Emanuel, K. A., 1988: The maximum intensity of hurricanes. *J. Atmos. Sci.*, **45**, 1143–1155, doi:[10.1175/1520-0469\(1988\)045<1143:TMIOH>2.0.CO;2](https://doi.org/10.1175/1520-0469(1988)045<1143:TMIOH>2.0.CO;2).
- , 1995: Sensitivity of tropical cyclones to surface exchange coefficients and a revised steady-state model incorporating eye dynamics. *J. Atmos. Sci.*, **52**, 3969–3976, doi:[10.1175/1520-0469\(1995\)052<3969:SOTCTS>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<3969:SOTCTS>2.0.CO;2).
- , 2000: A statistical analysis of tropical cyclone intensity. *Mon. Wea. Rev.*, **128**, 1139–1152, doi:[10.1175/1520-0493\(2000\)128<1139:ASAOTC>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<1139:ASAOTC>2.0.CO;2).
- , 2005: Increasing destructiveness of tropical cyclones over the past 30 years. *Nature*, **436**, 686–688, doi:[10.1038/nature03906](https://doi.org/10.1038/nature03906).
- , 2006: Climate and tropical cyclone activity: A new model downscaling approach. *J. Climate*, **19**, 4797–4802, doi:[10.1175/JCLI3908.1](https://doi.org/10.1175/JCLI3908.1).
- , 2013: Downscaling CMIP5 climate models shows increased tropical cyclone activity over the 21st century. *Proc. Natl. Acad. Sci. USA*, **110**, 12 219–12 224, doi:[10.1073/pnas.1301293110](https://doi.org/10.1073/pnas.1301293110).
- , and A. H. Sobel, 2013: Response of tropical sea surface temperature, precipitation, and tropical cyclone-related variables to changes in global and local forcing. *J. Adv. Model. Earth Syst.*, **5**, 447–458, doi:[10.1002/jame.20032](https://doi.org/10.1002/jame.20032).
- , C. DesAutels, C. Holloway, and R. Korty, 2004: Environmental control of tropical cyclone intensity. *J. Atmos. Sci.*, **61**, 843–858, doi:[10.1175/1520-0469\(2004\)061<0843:ECOTCI>2.0.CO;2](https://doi.org/10.1175/1520-0469(2004)061<0843:ECOTCI>2.0.CO;2).
- , S. Ravela, E. Vivant, and C. Risi, 2006: A statistical deterministic approach to hurricane risk assessment. *Bull. Amer. Meteor. Soc.*, **87**, 299–314, doi:[10.1175/BAMS-87-3-299](https://doi.org/10.1175/BAMS-87-3-299).
- Ge, X., and T. Li, 2013: Effects of vertical shears and midlevel dry air on tropical cyclone developments. *J. Atmos. Sci.*, **70**, 3859–3875, doi:[10.1175/JAS-D-13-066.1](https://doi.org/10.1175/JAS-D-13-066.1).
- Gneiting, T., A. E. Raftery, A. H. Westveld, and T. Goldman, 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Wea. Rev.*, **133**, 1098–1118, doi:[10.1175/MWR2904.1](https://doi.org/10.1175/MWR2904.1).
- Jarvinen, B. R., C. J. Neumann, and M. A. S. Davis, 1984: A tropical cyclone data tape for the North Atlantic basin, 1886–1983: Contents, limitations, and uses. NOAA Tech. Memo. NWS NHC 22, 21 pp.
- Johnson, N. C., and S.-P. Xie, 2010: Changes in the sea surface temperature threshold for tropical convection. *Nat. Geosci.*, **3**, 842–845, doi:[10.1038/ngeo1008](https://doi.org/10.1038/ngeo1008).
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, doi:[10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).
- Kaplan, J., and M. DeMaria, 2003: Large-scale characteristics of rapidly intensifying tropical cyclones in the North Atlantic basin. *Wea. Forecasting*, **18**, 1093–1108, doi:[10.1175/1520-0434\(2003\)018<1093:LCORIT>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<1093:LCORIT>2.0.CO;2).
- Knaff, J. A., M. DeMaria, C. R. Sampson, and J. M. Gross, 2003: Statistical, 5-day tropical cyclone intensity forecasts derived from climatology and persistence. *Wea. Forecasting*, **18**, 80–92, doi:[10.1175/1520-0434\(2003\)018<0080:SDTCIF>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<0080:SDTCIF>2.0.CO;2).
- Knutson, T. R., J. J. Sirutis, S. T. Garner, I. M. Held, and R. E. Tuleya, 2007: Simulation of the recent multidecadal increase of Atlantic hurricane activity using a 18-km-grid regional model. *Bull. Amer. Meteor. Soc.*, **88**, 1549–1565, doi:[10.1175/BAMS-88-10-1549](https://doi.org/10.1175/BAMS-88-10-1549).
- , —, —, G. A. Vecchi, and I. M. Held, 2008: Simulated reduction in Atlantic hurricane frequency under twenty-first-century warming condition. *Nat. Geosci.*, **1**, 359–364, doi:[10.1038/ngeo202](https://doi.org/10.1038/ngeo202).
- , and Coauthors, 2010: Tropical cyclones and climate change. *Nat. Geosci.*, **3**, 157–163, doi:[10.1038/ngeo779](https://doi.org/10.1038/ngeo779).
- Krishnamurti, T. N., C. M. Kishtawal, T. E. LaRow, D. R. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran, 1999: Improved weather and seasonal climate forecast from multi-model superensemble. *Science*, **285**, 1548–1550, doi:[10.1126/science.285.5433.1548](https://doi.org/10.1126/science.285.5433.1548).
- 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, doi:[10.1175/MWR-D-12-00254.1](https://doi.org/10.1175/MWR-D-12-00254.1).
- Lin, N., K. A. Emanuel, J. A. Smith, and E. Vanmarcke, 2010: Risk assessment of hurricane storm surge for New York City. *J. Geophys. Res.*, **115**, D18121, doi:[10.1029/2009JD013630](https://doi.org/10.1029/2009JD013630).
- Price, J. F., 2009: Metrics of hurricane-ocean interaction: Vertically-integrated or vertically-averaged ocean temperature? *Ocean Sci.*, **5**, 351–368, doi:[10.5194/os-5-351-2009](https://doi.org/10.5194/os-5-351-2009).
- Puri, K., J. Barkmeijer, and T. Palmer, 2001: Ensemble prediction of tropical cyclones using targeted diabatic singular vectors. *Quart. J. Roy. Meteor. Soc.*, **127**, 709–731, doi:[10.1002/qj.49712757222](https://doi.org/10.1002/qj.49712757222).
- Ramsay, H. A., and A. H. Sobel, 2011: Effects of relative and absolute sea surface temperature on tropical cyclone potential intensity using a single-column model. *J. Climate*, **24**, 183–193, doi:[10.1175/2010JCLI3690.1](https://doi.org/10.1175/2010JCLI3690.1).
- Sampson, C. R., and A. J. Schrader, 2000: The Automated Tropical Cyclone Forecasting System (version 3.2). *Bull. Amer. Meteor. Soc.*, **81**, 1231–1240, doi:[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).
- Schenkel, B. A., and R. E. Hart, 2012: An examination of tropical cyclone position, intensity, and intensity life cycle within atmospheric reanalysis datasets. *J. Climate*, **25**, 3453–3474, doi:[10.1175/2011JCLI4208.1](https://doi.org/10.1175/2011JCLI4208.1).
- Schumacher, A., M. DeMaria, and J. Knaff, 2013: Summary of the new statistical-dynamical intensity forecast models for the Indian Ocean and Southern Hemisphere and resulting performance. Joint Typhoon Warning Center Final Tech. Rep., 11 pp.
- Smith, R. K., M. T. Montgomery, and S. Vogl, 2008: A critique of Emanuel's hurricane model and potential intensity theory. *Quart. J. Roy. Meteor. Soc.*, **134**, 551–561, doi:[10.1002/qj.241](https://doi.org/10.1002/qj.241).
- Tang, B., and K. A. Emanuel, 2012: A ventilation index for tropical cyclones. *Bull. Amer. Meteor. Soc.*, **93**, 1901–1912, doi:[10.1175/BAMS-D-11-00165.1](https://doi.org/10.1175/BAMS-D-11-00165.1).
- Tao, D., and F. Zhang, 2014: Effect of environmental shear, sea-surface temperature, and ambient moisture on the formation and predictability of tropical cyclones: An ensemble-mean perspective. *J. Adv. Model. Earth Syst.*, **6**, 384–404, doi:[10.1002/2014MS000314](https://doi.org/10.1002/2014MS000314).
- Tippett, M. K., S. J. Camargo, and A. H. Sobel, 2011: A Possion regression index for tropical cyclone genesis and the role of large-scale vorticity in genesis. *J. Climate*, **24**, 2335–2357, doi:[10.1175/2010JCLI3811.1](https://doi.org/10.1175/2010JCLI3811.1).

- , A. H. Sobel, and S. J. Camargo, 2012: Association of monthly U.S. tornado occurrence with large-scale atmospheric parameters. *Geophys. Res. Lett.*, **39**, L02801, doi:[10.1029/2011GL050368](https://doi.org/10.1029/2011GL050368).
- Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012, doi:[10.1256/qj.04.176](https://doi.org/10.1256/qj.04.176).
- Vecchi, G. A., and B. J. Soden, 2007: Effect of remote sea surface temperature change on tropical cyclone potential intensity. *Natural*, **450**, 1066–1070, doi:[10.1038/nature06423](https://doi.org/10.1038/nature06423).
- , M. Zhao, G. Villarini, R. Anthony, A. Kumar, I. M. Held, and R. Gudgel, 2011: Statistical–dynamical predictions of seasonal North Atlantic hurricane activity. *Mon. Wea. Rev.*, **139**, 1070–1082, doi:[10.1175/2010MWR3499.1](https://doi.org/10.1175/2010MWR3499.1).
- Weber, H. C., 2005: Probabilistic prediction of tropical cyclones. Part II: Intensity. *Mon. Wea. Rev.*, **133**, 1853–1864, doi:[10.1175/MWR2950.1](https://doi.org/10.1175/MWR2950.1).
- Webster, P. J., G. Holland, J. A. Curry, and H.-R. Chang, 2005: Changes in tropical cyclone number, duration, and intensity in a warming environment. *Science*, **309**, 1844–1846, doi:[10.1126/science.1116448](https://doi.org/10.1126/science.1116448).
- Wilks, D. S., and T. M. Hamill, 2007: Comparison of ensemble-MOS methods using GFS reforecasts. *Mon. Wea. Rev.*, **135**, 2379–2390, doi:[10.1175/MWR3402.1](https://doi.org/10.1175/MWR3402.1).
- Zhang, Z., and T. N. Krishnamurti, 1999: A perturbation method for hurricane ensemble predictions. *Mon. Wea. Rev.*, **127**, 447–469, doi:[10.1175/1520-0493\(1999\)127<0447:APMFHE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<0447:APMFHE>2.0.CO;2).
- Zhao, M., I. M. Held, S.-J. Lin, and G. A. Vecchi, 2009: Simulations of global hurricane climatology, interannual variability, and response to global warming using a 50-km resolution GCM. *J. Climate*, **22**, 6653–6678, doi:[10.1175/2009JCLI3049.1](https://doi.org/10.1175/2009JCLI3049.1).
- Zipser, E., 2003: Some views on “hot towers” after 50 years of tropical field programs and two years of TRMM data. *Cloud Systems, Hurricanes, and the Tropical Rainfall Measuring Mission (TRMM)*, Meteor. Monogr., No. 29, Amer. Meteor. Soc., 49–57.