Statistical–Dynamical Predictions of Seasonal North Atlantic Hurricane Activity

GABRIEL A. VECCHI,* MING ZHAO,*,+ HUI WANG,#,@ GABRIELE VILLARINI,&,** ANTHONY ROSATI,*
ARUN KUMAR,# ISAAC M. HELD,* AND RICHARD GUDGEL*

*NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey
+University Corporation for Atmospheric Research, Boulder, Colorado
#NOAA/Climate Prediction Center, Camp Springs, Maryland
@Wyle Information Systems, McLean, Virginia
&Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey
**Willis Research Network, London, United Kingdom

ABSTRACT

Skillfully predicting North Atlantic hurricane activity months in advance is of potential societal significance and a useful test of our understanding of the factors controlling hurricane activity. In this paper, a statistical–dynamical hurricane forecasting system, based on a statistical hurricane model, with explicit uncertainty estimates, and built from a suite of high-resolution global atmospheric dynamical model integrations spanning a broad range of climate states is described. The statistical model uses two climate predictors: the sea surface temperature (SST) in the tropical North Atlantic and SST averaged over the global tropics. The choice of predictors is motivated by physical considerations, as well as the results of high-resolution hurricane modeling and statistical modeling of the observed record. The statistical hurricane model is applied to a suite of initialized dynamical global climate model forecasts of SST to predict North Atlantic hurricane frequency, which peaks during the August–October season, from different starting dates. Retrospective forecasts of the 1982–2009 period indicate that skillful predictions can be made from as early as November of the previous year; that is, skillful forecasts for the coming North Atlantic hurricane season could be made as the current one is closing. Based on forecasts initialized between November 2009 and March 2010, the model system predicts that the upcoming 2010 North Atlantic hurricane season will likely be more active than the 1982–2009 climatology, with the forecasts initialized in March 2010 predicting an expected hurricane count of eight and a 50% probability of counts between six (the 1966–2009 median) and nine.

1. Introduction

Substantial attention has been placed in developing forecast systems for the seasonal activity of North Atlantic hurricanes and tropical storms, with various methods developed over the years [see review in Camargo et al. (2007)]. Seasonal basin-wide activity forecast systems have been developed that show skill in retrospective forecasts over many years starting from April through July for the hurricane season that peaks in August–October (e.g., Gray 1984; Elsner and Jagger 2006; Vitart 2006; Vitart et al. 2007; Klotzbach and Gray 2009; LaRow et al. 2010; Wang et al. 2009; Zhao et al. 2010). In general, one can view the seasonal hurricane prediction problem as a two-step process: (i) predicting the state of the future climate system on large scales (the climate prediction) and (ii) predicting the response of seasonal basin-wide hurricane frequency to that future climate state (the hurricane prediction). Sometimes the two steps occur within a single process, such as when dynamical coupled climate models are used to predict the state of future climate, and the response of the hurricane-like vortices in the models is used to estimate future hurricane activity (e.g., Vitart 2006; Vitart et al. 2007), or when a statistical relationship between conditions prior to the hurricane season and the future season's hurricane activity is used (e.g., Gray 1984; Elsner and Jagger 2006; Klotzbach and Gray 2009). Since both the evolution of the climate system and the response of the hurricane activity to climate are chaotic processes, these forecasts are necessarily probabilistic.

The methods developed to date have been shown to have skill at predicting basin-wide tropical cyclone statistics

in the North Atlantic from the summer (Vitart 2006; Klotzbach and Gray 2009; LaRow et al. 2010; Zhao et al. 2010), and as early as spring (Vitart et al. 2007; Wang et al. 2009). Here, we develop a statistical–dynamical forecast system to extend the lead times of the forecasts to the winter prior to the hurricane season, with explicit uncertainty estimates.

Our approach is to build on recent results with the National Oceanic and Atmospheric Administration/ Geophysical Fluid Dynamics Laboratory (NOAA/GFDL) High-Resolution Atmospheric Model (HiRAM-C180; Zhao et al. 2009, 2010), a model that is able to recover much of the observed year-to-year variations in North Atlantic hurricane frequency when forced with observed sea surface temperatures (SSTs). We take advantage of a variety of experiments that have been performed to date with this dynamical model across a broad range of climates to develop a statistical emulation of the sensitivity of hurricane frequency to SST in this dynamical model. Training our statistical model on the hurricane frequency sensitivity of the HiRAM-C180 dynamical model allows us to address several issues that emerge when constructing statistical models from observations: (i) since we know the climate forcing in the model, we can be more confident of the causal connection between predictors and predictands; (ii) training on the sensitivity of hurricanes to a wide range of climate states allows us to better describe the relation between the predictors and the predictand; (iii) the dynamical model ensemble allows us to separate the forced and stochastic components of the hurricane frequency, which helps us avoid "overfitting" the statistical model; and (iv) the influence uncertainties arising from changes in observing and recording practices on SST (e.g., Vecchi and Soden 2007; Villarini et al. 2010) and hurricane frequency (e.g., Vecchi and Knutson 2008; Landsea et al. 2010) do not influence a statistical model built from the hurricane sensitivity of a dynamical model. An obvious disadvantage of training the statistical emulator on the sensitivity of the HiRAM-C180 dynamical model is that the sensitivity of hurricanes to large-scale conditions in that model may be incorrect; however, the strong agreement between the statistics of the observed and HiRAM-C180 hurricane frequencies (Zhao et al. 2009, 2010) and the retrospective skill of the forecast system that we show here gives us confidence to proceed with this methodology.

The statistical formulation of the emulator is based on the results of Villarini et al. (2010) and is motivated by physical considerations. The emulator is applied to dynamical predictions of SST from two initialized coupled modeling prediction systems in order to make predictions of Atlantic hurricane frequency. We use retrospective forecasts over the period 1982–2009 to assess the skill of our method, and we make forecasts for the upcoming (2010) hurricane season.

2. Methods

a. High-Resolution Atmospheric Model

Our hybrid forecasting technique is built on the results from a dynamical hurricane climate model recently developed at NOAA/GFDL, the High-Resolution Atmospheric Model, specifically the HiRAM-C180 [cubed sphere dynamical core (Putman and Lin 2007) with 180×180 grid points on each face of the cube, resulting in grid sizes ranging from 43.5 to 61.6 km]. When forced with the observed SSTs of Rayner et al. (2003), the HiRAM-C180 model simulates hurricane statistics that compare well with the observations for the period 1981–2008 (Zhao et al. 2009, 2010). The model recovers the overall geographical distribution of storm genesis locations, as well as the seasonal cycle and the interannual variability of hurricane frequencies for the North Atlantic and the eastern and western Pacific.

The quality of our dynamical model forced with the observed SSTs has encouraged us to use this model as a tool to explore issues related to seasonal hurricane predictions. As a first step, Zhao et al. (2010) used HiRAM-C180 to pursue retrospective predictions of seasonal hurricane frequency in the North Atlantic and eastern Pacific by simply persisting SST anomalies from June throughout the hurricane season (these are effectively persistence forecasts initialized in July). Using an ensemble of five realizations for each year between 1982 and 2008, the correlation of the model mean with observations of basin-wide hurricane frequency is 0.69 in the North Atlantic. Furthermore, Zhao et al. (2010) found that a significant part of the degradation in skill [compared to a model forced with observed SSTs (correlation = 0.78) during the hurricane season] can be explained by the change from June through the hurricane season in one parameter, the difference between the SST in the main development region and the tropical mean SST. This indicates that the quality of seasonal forecasts based on a coupled atmosphere–ocean model will depend in large part on the model's ability to predict the evolution of the difference between main development region SSTs and tropical mean SSTs.

In Zhao et al. (2009), the same dynamical model was also used to simulate the hurricane response to four different SST anomalies generated by coupled models in the Third Coupled Model Intercomparison Project (CMIP3; Meehl et al. 2007) archive, for the late twenty-first century based on the Intergovernmental Panel on

Climate Change's (IPCC) Special Report on Emissions Scenarios's (SRES) A1B scenario. The SST anomalies were obtained from single realizations of three models [GFDL's Coupled Model version 2.1 (CM2.1), the Met Office's (UKMO) third climate configuration of the Met Office Unified Model (HadCM3), and the Max Planck Institute's (MPI) ECHAM5], and from the ensemble mean for the simulations for 18 models.

Since the Zhao et al. (2009) study, we have further pursued five additional SST warming experiments using SSTs from different coupled models [GFDL's CM2.0, UKMO's Hadley Centre Global Environmental Model version 1 (HADGEM1), the Canadian Centre for Climate Modelling and Analysis's (CCCma) Coupled General Circulation Model (CGCM), the Meteorological Research Institute's (MRI) Coupled General Circulation Model version 2.3.2 (CGCM2.3.2), and the Center for Climate Systems Research's (CCSR) high-resolution version of the Model for Interdisciplinary Research on Climate (MIROC-HI)]. As described in Zhao et al. (2009), we generated a control simulation by prescribing the climatological SST (seasonally varying with no interannual variability) using time-averaged (1982–2005) Hadley Centre Global Sea Ice Coverage and Sea Surface Temperature (HadISST) data (Rayner et al. 2003). We then added the SST warming anomalies (also seasonally varying with no interannual variability) projected by the coupled model to the climatological SST to pursue the global warming experiments. Ten-year integrations were carried out for both the control and the perturbation experiments.

b. Statistical hurricane model

Ideally, we would like to use as our hurricane forecast tool a global dynamical modeling system like that of Zhao et al. (2009, 2010), which exhibits both substantial skill in retrospective hurricane frequency simulations with prescribed SSTs and represents the response of hurricanes across a broad range of possible climate states. However, the computational expense of running this high-resolution model over the hundreds of retrospective forecast experiments needed to assess the retrospective skill of the model is prohibitive. Our approach in this work is to take advantage of the hundreds of years of experiments with the HiRAM-C180 model across a broad range of climates to develop a statistical emulation of this dynamical model, which will have the advantage of being computationally inexpensive, allowing us to perform the hundreds of retrospective forecasts needed. In addition, we will train this statistical emulator on both the historical HiRAM-C180 experiments as well as a suite of experiments exploring substantially altered climate states, in order to build a statistical system that is more likely to be robust across different possible climates, and not optimized to retrospectively predicting past climate. An additional advantage is that once it is developed, the statistical model can be applied to a wide range of SST forecast products, for example, by applying it to seasonal SST forecasts from various operational centers.

To build our statistical emulation of HiRAM-C180, we must select a set of potential climate-state covariates to serve as predictors. We focus on covariates based on SST since the results of Zhao et al. (2009) strongly suggest that a large fraction of the climate information for hurricane activity is contained in the monthly evolution of global patterns of SST. For this work, we wish to build as parsimonious a statistical model of basin-wide hurricane frequency as possible, but one that is still able to describe the variability exhibited by the data. For this reason and building on the results of various recent studies (Vecchi and Soden 2007; Swanson 2008; Vecchi et al. 2008; Knutson et al. 2008; Zhao et al. 2009, 2010; Villarini et al. 2010), we limit our predictors to two: Atlantic main development region SST (SST_{MDR}) and global tropical-mean SST (SST_{TROP}). We choose as one possible predictor SST_{MDR} (SST averaged over 10°–25°N, 80°–20°W, and averaged over August–October, the peak months of the hurricane season), since we expect hurricane activity in the Atlantic to depend partly on the evolution of SST local to hurricane development (e.g., Emanuel 2005; Mann and Emanuel 2006; Vecchi and Soden 2007; Swanson 2008; Knutson et al. 2008; Zhao et al. 2009; Villarini et al. 2010). In addition, we choose as another possible covariate SST_{TROP} (SST averaged 30°S–30°N and during the period August–October) because (i) there is a substantial body of evidence that tropical-mean SST changes can influence large-scale climatic conditions in the North Atlantic that affect hurricane activity, such as wind shear (Latif et al. 2007) and upper-tropospheric temperature (Sobel et al. 2002), that in turn influences hurricane potential intensity (Vecchi and Soden 2007) and other measures of thermodynamic instability (e.g., Shen et al. 2000; Tang and Neelin 2004), (ii) tropical-mean SST appears to be a useful covariate in the description of the historical changes in other measures of basin-wide tropical cyclone activity in the North Atlantic (e.g., Swanson 2008; Vecchi et al. 2008; Villarini et al. 2010), and (iii) the differences between tropical Atlantic and tropical-mean SSTs are a good indicator of the response of hurricane frequency to changes in climate in high-resolution atmospheric models (e.g., Knutson et al. 2008; Zhao et al. 2009, 2010; Villarini et al. 2011). Based on the studies cited above, we expect a priori that SST_{MDR} will emerge as a positively correlated predictor of hurricane frequency, while SST_{TROP} should emerge as a negatively correlated predictor.
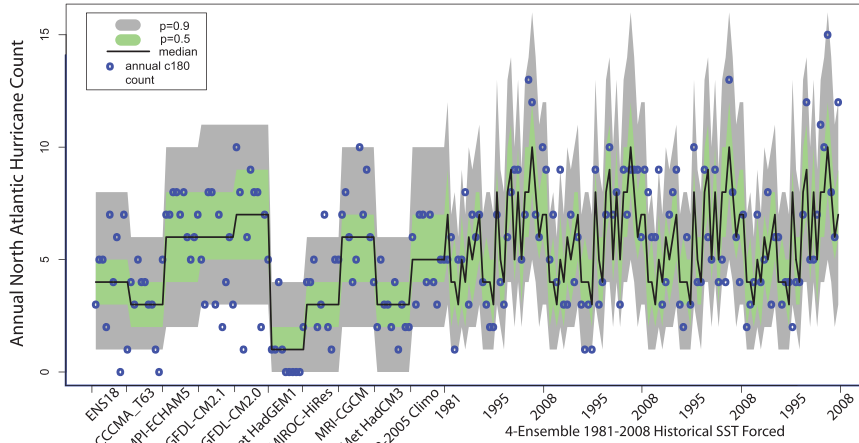
FIG. 1. Construction and evaluation of the statistical hurricane frequency model, showing the fit of the Poisson regression model of hurricane counts for the North Atlantic basin to the training set of HiRAM-C180 high-resolution atmospheric model hurricane counts: blue dots show the counts from the HiRAM-C180 model, the black line shows the median of the statistical model fit, and the green and gray shadings show the 50% and 90%, respectively, confidence intervals on the fit. The parsimonious statistical model with only two covariates is able to capture the behavior of basin-wide hurricane counts across a broad range of climates in this dynamical atmospheric model.

To emulate the response of hurricane frequency to changes in SST in HiRAM-C180, we use a Poisson regression model in which the rate of occurrence λ is a function of both SST_{MDR} and SST_{TROP}. The statistical model is built from 212 yr of model integration from HiRAM-C180 (Fig. 1), which include ten 10-yr climate change experiments [each of the ten experiments explores the hurricane response to a different projection of twenty-first-century SST from the IPCC Fourth Assessment Report (AR4)–CMIP3 archive] and four historical ensembles in which the model was forced with the observed SSTs (Rayner et al. 2003) over the period 1981–2008 (Zhao et al. 2009, 2010). We then build a Poisson regression model [using the Generalized Additive Models for Location Scale and Shape (gamlss) R package; Stasinopoulos et al. (2009)] for the HiRAM-C180 hurricane counts to the corresponding SST_{MDR} and SST_{TROP}. Our training was performed on anomalies relative to the 1982–2005 average. We also performed sensitivity experiments by training the statistical model separately on the four-member historical experiment dataset of HiRAM-C180 and on the suite of climate change experiments. For all of the training experiments, both tropical Atlantic and mean tropical SSTs are retained as significant predictors; the coefficients of these two covariates have similar magnitudes but opposite signs (SST_{MDR} acting as a positive predictor and SST_{TROP} as a negative one, conforming to our prior expectation). More specifically, in each of the training experiments the

magnitude of the influence of SST_{TROP} is always larger than that of SST_{MDR}, indicating a tendency of uniform tropical-mean warming ($SST_{MDR} = SST_{TROP}$) to reduce Atlantic hurricane frequency, and consistent with the sensitivity of historical North Atlantic tropical-storm frequency (Villarini et al. 2010). We retain as our "best" emulator the model fitted to the entire record, since it spans the broadest range of possible climate conditions, giving us more confidence as to its applicability to novel situations. When trained on the entire suite of HiRAM-C180 experiments (exploring both historical and climate change conditions), the rate of occurrence λ of the Poisson regression model can be described as a linear function of both SST_{MDR} and SST_{TROP} (via a logarithmic link function) as follows:

$$\lambda = e^{1.707+1.388SST_{MDR}-1.521SST_{TROP}}, \tag{1}$$

where SST_{MDR} and SST_{TROP} are anomalies in the regional SST indices relative to the 1982–2005 average.

As can be seen in Fig. 1, the fitted statistical model is able to reproduce well the variability exhibited by the basin-wide hurricane counts under different conditions (both in the HiRAM-C180 response to historical SST changes and to large climate change projections), providing supporting evidence of its robustness. Evaluation of the model fit was performed by visual examination of the residual plots (e.g., qq-plots and worm plots), as well as computing residuals' statistics, such as mean, variance,
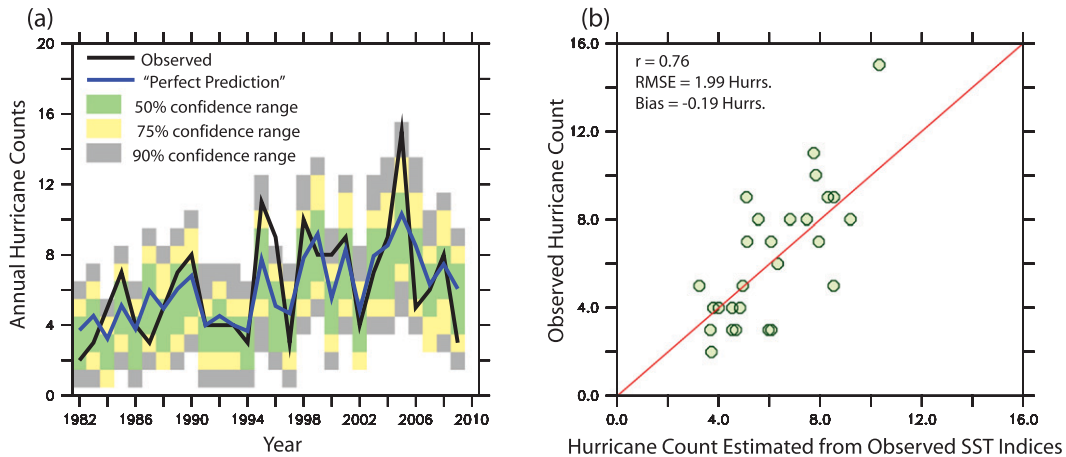
FIG. 2. Results of applying the statistical emulator of the sensitivity of HiRAM-C180 hurricane frequency to the observed SST data of Rayner et al. (2003). (a) A time series plot of the observed hurricane counts (black line) and the expected counts from applying the statistical model to the observed SSTs (blue line); the 90%, 75%, and 50% confidence intervals of the statistical fit are shown by the gray, yellow, and green shadings, respectively. (b) A scatterplot of the observed hurricane counts vs the mean hurricane counts estimated by the statistical model from the observed SSTs.

skewness, kurtosis, and a Filliben correlation coefficient [consult Villarini et al. (2010) for a more extensive discussion about model fitting and evaluation]; these assessments of the model's residuals supported the model's selection.

To test the relevance of this statistical emulator of a global dynamical model to the observed record, we apply the statistical hurricane frequency model to the observed August–October SST indices (Fig. 2), which we refer to as a perfect prediction, as it indicates the skill of the emulator if we were able to perfectly predict the predictors. In this perfect prediction mode, the statistical emulator recovers much of the observed variability in hurricane activity, with a correlation coefficient of 0.76 and a root-mean-square error (RMSE) of 1.99 hurricanes over the period 1982–2009. These values compare very well to the 1982–2008 correlation of 0.79 and RMSE of 1.86 hurricanes from the full HiRAM-C180 AGCM when forced with the observed monthly varying SST field (Zhao et al. 2009, 2010). Therefore, despite its parsimony (only two predictors), our simple hurricane statistical model performs very well, and there is little statistical justification for adding predictors to our emulator since the correlation to the observed record is so similar to that of the full high-resolution dynamical model we are seeking to emulate.

c. Forecasts of sea surface temperature

We explore two initialized forecast systems built on coupled ocean–atmosphere models: the CM2.1 experimental seasonal-to-interannual (S-I) prediction system and the National Centers for Environmental Prediction (NCEP) Climate Forecast System (CFS).

The NOAA/GFDL experimental S-I prediction system is built on the CM2.1 coupled climate modeling system (Delworth et al. 2006), which was also used to provide data for the recent AR4. The GFDL–CM2.1 retrospective forecasts consist of a set of retrospective predictions initialized over the period November 1981– March 2010, each with a 10-member ensemble initiated from the first day of every month (November–August) with an integration of 12 months. Thus, for each of 28 August–October seasons (1982–2009) we have 100 retrospective integrations (10 ensemble members initialized in each of the 10 months). In addition, we have 10-member forecasts for the upcoming 2010 hurricane season initialized from 1 November 2009 to 1 March 2010.

The state of the ocean and atmosphere for each of the GFDL–CM2.1 forecast experiments is set using a state-of-the-art coupled (ocean–atmosphere) ensemble Kalman filter (EnKF) assimilation system, which incorporates the observed oceanic and atmospheric states available prior to the initiation of the forecast run (Zhang et al. 2007). The initial conditions for the 10 ensemble members for each forecast experiment are selected from the ensemble members of the EnKF initialization scheme. The 10 ensemble members of each forecast experiment were used to calculate the ensemble mean climatology, with each start month having its own separate climatology to correct for systematic model drift. For each of the forecast initialization times (1 November– 1 August) we computed 1 August(0)–31October(0) SST anomaly (relative to each forecast system's 1982– 2005 climatology) indices for the Atlantic MDR and tropics.

We also use SST forecasts from the current NCEP operational Climate Forecast systems (CFS) model (Saha et al. 2006). These are the standard initialized forecasts with the CFS system available from NCEP. The CFS is a fully coupled ocean–land–atmosphere dynamical seasonal prediction system, and forecasts are initialized from the observed estimates of the ocean, atmosphere, and land conditions. Various elements of the CFS consist of the NCEP Global Forecast Systems version 2003 (Moorthi et al. 2001) and the GFDL Modular Ocean Model version 3 (Pacanowski and Griffies 1998). Initial conditions for the CFS are taken from the NCEP/Department of Energy (DOE) Reanalysis-2 (R2) for the atmosphere and land, and from the NCEP Global Ocean Data Assimilation System (GODAS) for the ocean (Behringer et al. 1998; Behringer and Xue 2004).

For each month (January–July) in the NCEP–CFS forecast period (1982–2009), 15 forecast members were produced by the CFS for the nine following (target) months. The 15 forecast members are in a group of three sets with five members each. The three sets of forecast runs are initiated from 0000 UTC around the 1st, 11th, and 21st calendar day of the month. The five members in each set of the forecast runs are from five observed atmospheric and land initial conditions that are 1 day apart, but share the same oceanic initial conditions.

For comparing the various retrospective forecasts, we use the convention that a Month-M forecast should include data no more recent than the first week of month M in its initialization, so that we are assessing the ability to issue a hurricane forecast early in month M. Thus, the GFDL–CM2.1 Month-M forecasts compose the 10-member ensemble initialized on the 1st day of that month, the NCEP–CFS Month-M forecasts compose the 15-member ensemble made up of the 10 members initialized on the 11th and 21st of month $M - 1$ and the five members initialized on the 1st of month M, and for SST persistence the Month-M forecasts are those in which SST anomalies from month $M - 1$ are persisted.

d. Seasonal hurricane forecasts and uncertainty estimates

To perform our hybrid statistical–dynamical forecasts of hurricane frequency, we take the dynamical predictions of SST_{MDR} and SST_{TROP} and use them as input for our statistical emulation of the HiRAM-C180 high-resolution dynamical atmospheric model. In our seasonal forecasts of hurricane frequency we wish to explicitly compute uncertainty estimates arising from two sources that we can currently quantify: (i) given an initial state of the climate system, the forecasts of the SST indices are uncertain, with variations arising due to the chaotic nature of the global climate system, and (ii) given a particular realization of SST the forecasts of hurricane frequency are uncertain, due to variations in hurricane frequency that are not constrained by our SST indices. We can estimate the uncertainty arising from the first of these sources by exploring the ensemble suite of the initialized forecasts, and we can estimate the uncertainty arising from the second source through the parametric probability estimates in the Poisson model of hurricane frequency given a rate parameter (λ) (Rice 1995, p. 41):

$$p(\text{Count} = k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \ (k = 0, 1, 2 \ldots). \quad (2)$$

To compute our predictions of hurricane activity, we first, for each initialization, take the predicted August–October SST_{MDR} and SST_{TROP} from each of the ensemble members (anomalies computed from the forecast system 1982–2005 climatology). We then apply these indices to the fit to obtain the predicted rate parameter λ from our Poisson fit emulator of the HiRAM-C180 model [Eq. (1)], and from Eq. (2) we can compute the probability density function (PDF) of predicted hurricane counts for each of the ensemble members. We then convolve the PDFs from the suite of forecast ensemble members in order to arrive at an ensemble PDF for each month of initialized forecasts:

$$p(\text{Count} = k|\text{Initialization}) = \frac{1}{N}\sum_{i=1}^{N} p(\text{Count} = k|\lambda_i),$$

$$(3)$$

where λ_i is the rate parameter computed from the ith ensemble member for a given initialization month's forecast and N is the number of ensemble members available. With this PDF, we can compute an expected value (mean), as well as various confidence intervals and exceedance probabilities in the counts for each year and for each forecast month.

We use two ensemble-averaging techniques to generate our predictions of annual hurricane frequency. The first technique is a simple ensemble average, in which we take the ensemble members that are initialized using data from the month just prior to a given month as the total ensemble set. The second technique is the lagged superensemble, which is motivated by a desire to increase the ensemble size from 10 to 15, and results from the constraints on the size of the ensemble that have been run. In this lagged superensemble we build a 30-member ensemble set consisting of the three 10-member ensembles initialized that particular month and the two preceding months for GFDL–CM2.1, and of the two 15-member

ensembles initialized by the particular month and that preceding it for NCEP–GFS. The assumption in generating the lagged superensemble is that the error growth over the period between the initialization of the oldest ensemble set and the newest is smaller than the enhanced fidelity of additional ensemble members. This assumption is justified a posteriori by the examination of the patterns of behavior of our retrospective forecasts presented below. We can compute a two-model ensemble by using as the set of SST forecasts the total ensemble from NCEP–CFS and GFDL–CM2.1, which for monthly forecasts is 25 and for lagged ensembles is 60.

3. Retrospective forecast results

In this section we explore the retrospective skill of our basin-wide hurricane frequency forecast system. To assess the skill of the predictions of hurricane frequency based on the SST forecasts from GFDL–CM2.1 and NCEP–CFS, we compare them against the persistence of anomalies in SST_{MDR} and SST_{TROP} from the Rayner et al. (2003) dataset. As shown in Zhao et al. (2010), we expect a sharp dropoff in the skill of persistence forecasts across late boreal spring. Our hope is that, by initializing the coupled ocean–atmosphere system from observations and representing the dynamical and thermodynamical processes that control the evolution of the climate system, the dynamical forecast systems could extend the forecast window into boreal winter.

In these retrospective forecasts, we perform forecasts of hurricane activity in past years attempting to keep out information that would not have been available at the time of the forecast. It must be acknowledged that the construction of our hurricane prediction scheme is not entirely independent of the historical hurricane record (albeit in an indirect manner), as we decided to build a statistical emulation of the HiRAM-C180 atmospheric model because it is able to recover much of the historical hurricane frequency when forced with historical SSTs (Zhao et al. 2009, 2010). However, the HiRAM model was not optimized to recover the year-to-year variations of hurricane frequency (Zhao et al. 2009), and therefore the observed hurricane variations did not go into its construction. Moreover, in building the statistical emulator we did not use the observed hurricane record (which we seek to retrospectively forecast), but used only the HiRAM-C180 response to observed SST as well as to a suite of global climate change simulations. In doing so, we have attempted to minimize the direct influence of the observed hurricane record in our retrospective forecasts.

Figure 3 shows the retrospective skill exhibited by the GFDL–CM2.1 and NCEP–CFS systems in forecasting the SST indices that are used in the statistical hurricane model over 1982–2009. GFDL–CM2.1 exhibits correlations greater than 0.6 for both indices at all leads explored here, while the correlation coefficient for both indices in NCEP–CFS drops relatively more rapidly than for GFDL–CM2.1. Part of the drop in the correlations of the NCEP–CFS forecast of each index is due to a failure of NCEP–CFS to capture much of the 1982–2009 trend in both SST indices. The inability of NCEP–CFS to capture the observed warming in these SST indices is likely related to the use of invariant greenhouse and aerosol forcing in the model (Cai et al. 2009).

Our hurricane forecasts depend on a particular combination of these two indices, with the difference between the indices, roughly speaking, controlling the hurricane activity. To the extent that errors in the predictions for the individual indices arise from the spatially uniform component of SST trends, they will not be closely related to the hurricane forecast skill. In fact, as described below, the NCEP–CFS forecasts provide better short-lead-time seasonal hurricane forecasts than does the GFDL system.

Figure 4 illustrates the retrospective skill in hurricane counts of the forecast system in two initialization months (January and March) and using the two ensemble averaging techniques (simple ensemble and lagged superensemble). Since the hurricane season peaks in August–October, these can be considered 7- or 5-month-lead forecasts. Since January-initialized forecasts of August–October are not available from NCEP–CFS, the top panels Fig. 4 show the confidence intervals from the GFDL–CM2.1-based forecast, while the bottom panels show the confidence intervals from the two-model (NCEP–CFS + GFDL–CM2.1) forecasts. In each of the panels in Fig. 4 the various ensemble averages of the forecast systems, though not perfect, exhibit some visual similarity to the observed hurricane record. For the retrospective forecasts shown in Fig. 4, the uncertainty estimates seem conservative, with nominally fewer than 25% and 10% of the cases outside the 75% and 90% confidence intervals, yet this difference is not statistically significant, and over the entire forecast suite there is no systematic bias in the uncertainty estimate. Measures of correlation and RMSE bear out the visual similarity of the observed and forecast time series, with correlations between 0.5 and 0.6 and RMSEs of 2.4 and 2.6 hurricanes per year.

Among the years that were problematic for the system to retrospectively predict was the incredibly active season of 2005. None of the January or March retrospective forecasts were able to foretell in their mean estimate that 2005 should have been the most active year over 1982–2005, much less its extreme value that lay outside the 90% confidence interval for all forecasts (Fig. 4).
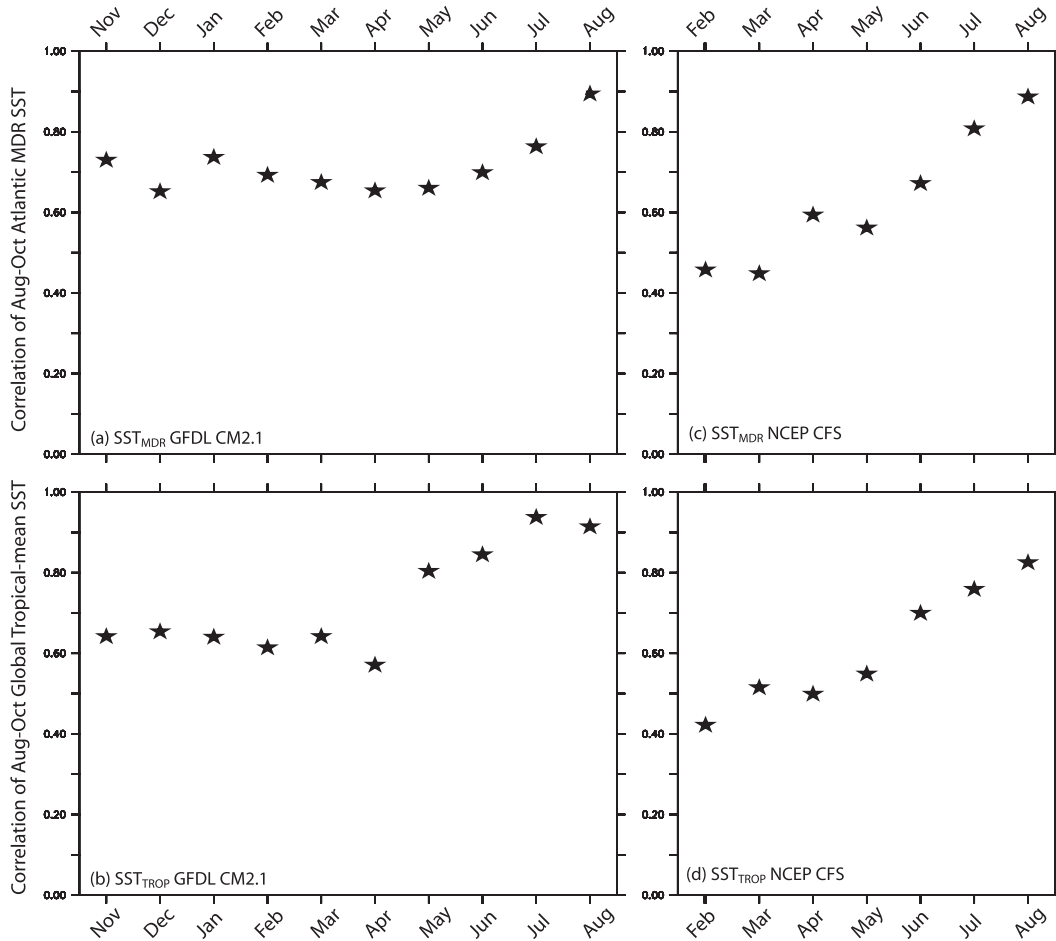
FIG. 3. Retrospective prediction skill by two dynamical forecast systems of the SST indices in the hurricane count statistical model: (a),(c) August–October Atlantic MDR SST and (b),(d) August–October global tropical-mean SSTs. Dates indicated along the horizontal axes identify the initialization month of the forecasts. Panels (a) and (b) are computed using the retrospective predictions from the GFDL–CM2.1 system, while (c) and (d) use NCEP–CFS. The black stars indicate the 1982–2009 correlation of the ensemble-mean-predicted SST index with the observed SST index.

However, with "perfect" SST values our statistical emulator is able to recover that 2005 should have been the most active year over 1982–2009 (Fig. 2), suggesting that predicting the evolution of our SST indices to the values leading to an extreme hurricane season proved difficult for the dynamical SST forecast systems. Furthermore, the observed counts for 2005 were outside the 90% confidence interval of the statistical emulator even with perfect SST (Fig. 2), yet the observed value of the 2005 hurricane frequency was within the spread of the four-member ensemble from the dynamical HiRAM-C180 model (Zhao et al. 2009) forced using monthly fields of SST. This suggests that either nonlinear dynamics captured by the HiRAM-C180 framework but absent in our simplified statistical model or the response to details in SST beyond our two indices were important to the

extreme 2005 season; that is, it may be that the statistical emulator may have been inadequate for predicting 2005. However, the observed hurricane frequency for 2005 is only one of a couple of years from 28 that is outside the 90% confidence interval in Fig. 2, suggesting that its extreme hurricane frequency could be understood as a stochastic enhancement to an active season. As noted in section 2b, we cannot currently justify additional predictors since the statistical model explains as much variance in hurricane frequency as the dynamical model indicates should be explainable. Further analyses should explore the key mechanisms that led to the extreme 2005 season.

We can use the correlation coefficients and RMSEs from Zhao et al. (2009) and from applying the statistical hurricane emulator to observed SSTs ($r = 0.78$ and 0.76,
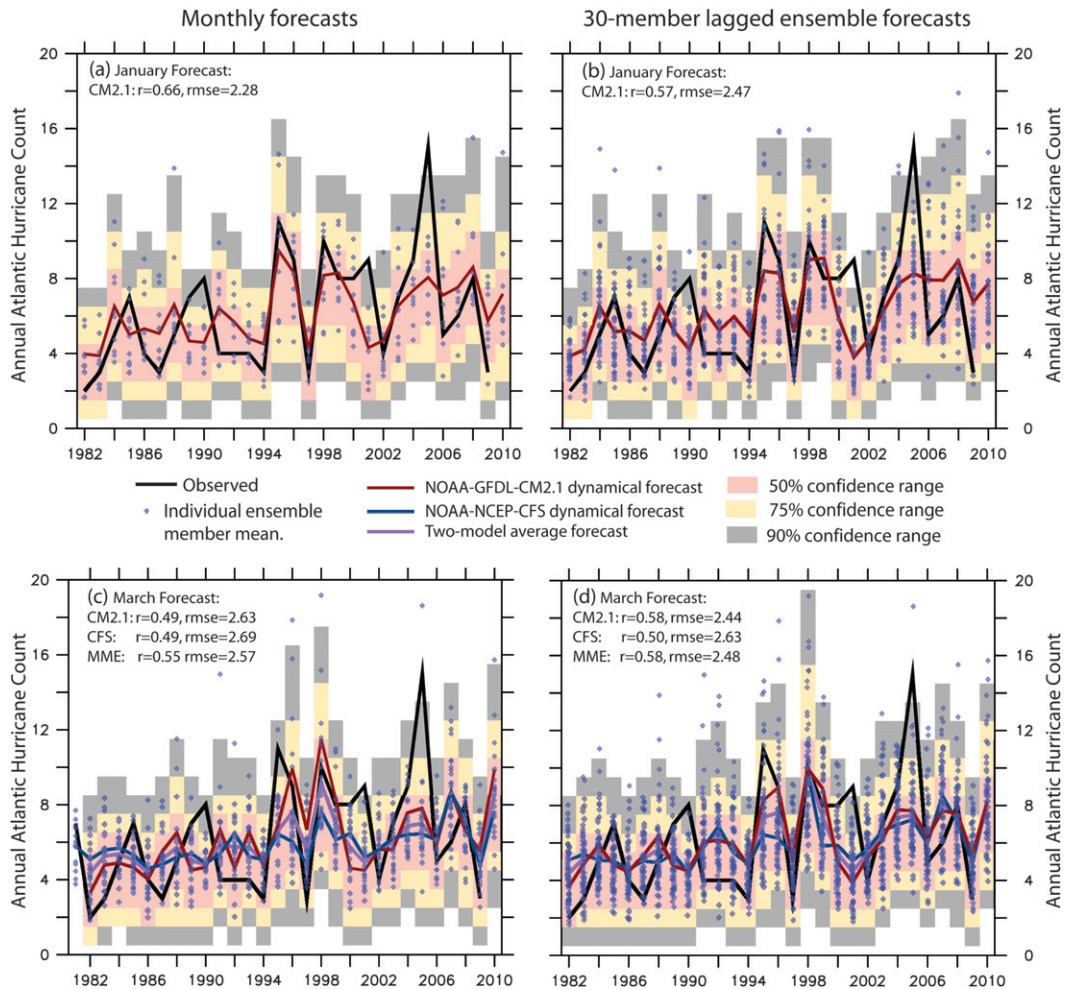
FIG. 4. Retrospective 1982–2009 forecasts of North Atlantic hurricane frequency for (a),(b) January and (c),(d) March, for both the (a),(c) simple monthly ensemble and (b),(d) 30-member lagged ensemble. These plots illustrate the potential for long-lead hurricane frequency forecasts from as early as January. In all the panels, the black lines show the observed North Atlantic hurricane counts. The colored lines show the ensemble-averaged forecasts for the mean count using SST predictions from the GFDL–CM2.1 system (red; CM2.1), the NCEP–CFS system (blue; CFS), and the two-system average (violet; MME). The blue dots show the forecasts of the mean count from the individual ensemble members. Color shading indicates estimates of uncertainty in the forecasts (the gray shading shows the 90% confidence range, the yellow shading shows the 75% confidence range, and the pink shading shows the 50% confidence range). Confidence ranges are based on the GFDL–CM2.1 system in the top panels, and on the two-system ensemble in the bottom panels.

and RMSE = 1.91 and 1.99 hurricanes per year, respectively) as estimates of the explainable variance in hurricane counts from perfect knowledge of SSTs. Within this context, our retrospective forecasts initialized in January and March explain around half of this explainable variance, with forecasts initialized in December and November exhibiting slightly lower skill (Fig. 5).

The retrospective hurricane forecast skill as a function of initialization month is compared to a variety of other estimates of predictability in Fig. 5. As one approaches the hurricane season (forecasts initialized in June and later), the hurricane forecasts using GFDL–CM2.1 forecasts of SST do not outperform the persistence of SST anomalies [either with our statistical emulator or in the results of Zhao et al. (2010)]. In contrast, the short-lead (July and August) hurricane forecasts using the NCEP–CFS system reach our estimates of potential predictability, with May and June forecasts approaching that measure. For longer lead times (April and earlier), the dynamical predictions of SST from both dynamical systems lead to substantial increases in hurricane forecast skill over those with persisted SST anomalies. All of the
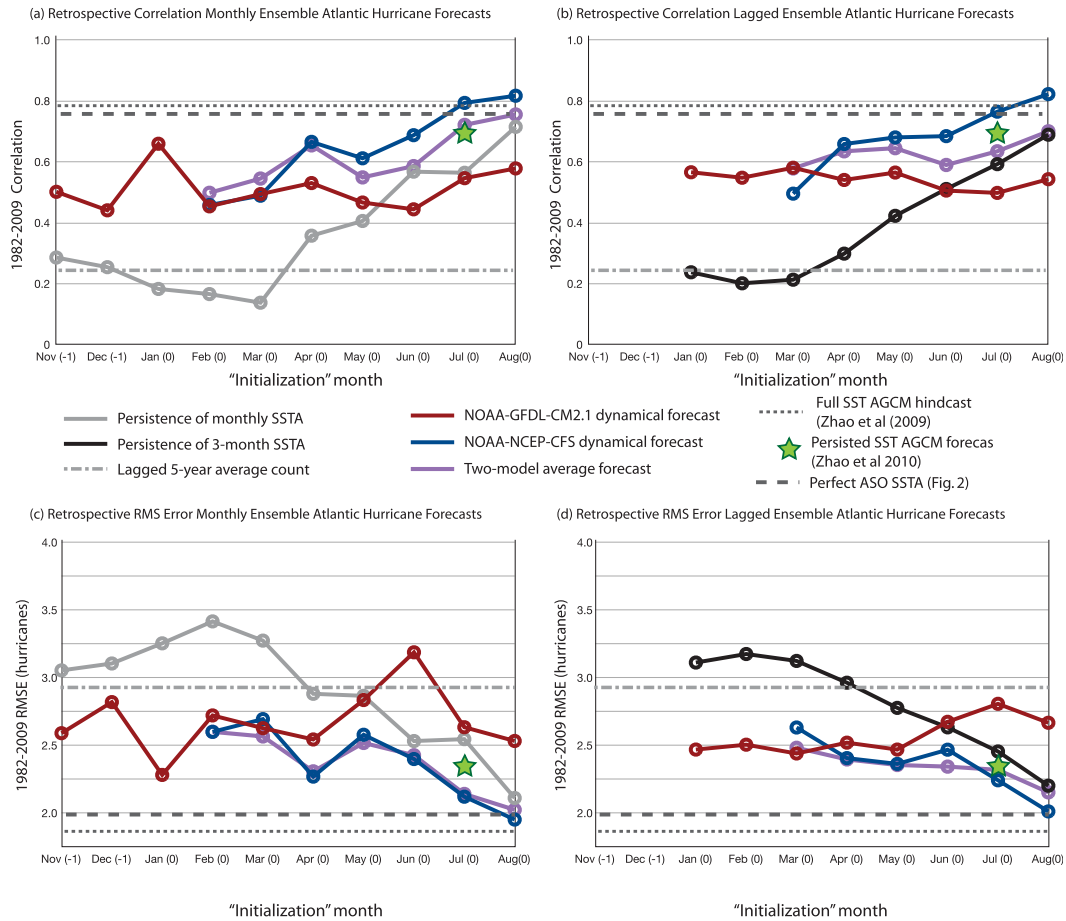
FIG. 5. Performance of the dynamical–statistical S-I North Atlantic hurricane frequency forecast system as a function of forecast lead time. The system skill is evaluated using (a),(b) correlation and (c),(d) RMSE, for both (a),(c) the simple monthly ensembles and (b),(d) the lagged ensemble schemes. The colored lines show the skill of the statistical–dynamical system using GFDL–CM2.1 (red), NCEP–CFS (blue), and the two-model average (violet). The skill of the dynamical–statistical scheme is compared with two measures of "null skill": the WMO-suggested standard (WMO 2008) of a lagged 5-yr average for the evaluation of hurricane forecast skill (gray dashed–dotted) and the statistical hurricane frequency model applied to the persistence of SST anomalies (solid gray and black lines, for persistence of monthly and 3-month SST anomalies). Two measures of "potential skill" are shown: the performance of the statistical hurricane model using observed August–October SSTs (black dashed line) and the performance of the HiRAM-C180 dynamical atmospheric model using observed monthly Rayner et al. (2003) SST fields from 1982 to 2008. For reference, we show the Zhao et al. (2010) persisted June SST anomaly skill using the HiRAM-C180 dynamical atmospheric model (star symbol). The NCEP–CFS and two-model average schemes outperform the persistence forecasts for all months, and the GFDL–CM2.1 forecasts outperform persistence for all months prior to June.

schemes, including the persisted SST anomaly schemes, outperform the trailing 5-yr average [gray dashed–dotted line in Fig. 5; a skill metric suggested by the World Meteorological Organization (WMO 2008)] and the persistence of the previous season's hurricane frequency (correlation of 0.14 and RMSE of 3.83 hurricanes per year).

In Fig. 5 one can see that the retrospective skill levels (as measured by correlation and RMSE) of the 10- and 15-member simple ensemble forecasts (left panels) exhibit substantial variability from month to month; for example, the correlation for GFDL–CM2.1 reaches 0.66

in January but drops to 0.45 in the February initializations before rising again to 0.49 for its March initialization. On the other hand, the lagged superensemble predictions (Fig. 5, right panels) exhibit not only stable skill in time, as one would expect from its inherent time smoothing, but also a consistently greater level of skill than the individual simple ensembles (except for January in GFDL–CM2.1 and the forecasts from boreal summer in NCEP–CFS). This difference in the behavior of the two types of ensemble strategies suggests that the forecast systems (particularly GFDL–CM2.1) could benefit from a larger ensemble set

(e.g., 30 members instead of 10), and that—as was assumed above—over the time between the newest and oldest of the members of the superensemble, the impacts of error growth are smaller than the beneficial effects from additional ensemble members. Because of this, we view the lagged superensemble system as likely being superior to the simple ensemble particularly at longer leads, though perhaps a larger ensemble set would be preferable to either.

The two-model ensemble forecast performs better than either model for forecasts initialized in February and March (late boreal winter; Figs. 4 and 5). As we approach the verification period, the skill exhibited by NCEP–CFS is sufficiently higher than that of GFDL–CM2.1 that the two-model ensemble performs worse than NCEP–CFS alone. These results would suggest that forecasts at long leads (April and earlier) should focus on the lagged ensemble, and when possible the two-model average, while for shorter leads (June initialization and shorter) forecasts using NCEP–CFS appear to be superior. For the November–January initializations we are limited to using the GFDL–CM2.1 forecasts of SST, since NCEP–CFS is not available.

4. Discussion

We have built a statistical–dynamical North Atlantic hurricane frequency prediction scheme that exhibits skill in retrospective forecasts for 1982–2009 based on North Atlantic and global tropical SSTs during boreal winter and spring prior to the hurricane season. The system was built using a statistical emulator of a high-resolution dynamical atmospheric model and initialized forecasts of SST. The forecast system predicts the probability density function of North Atlantic basin-wide hurricane frequency; thus, uncertainty estimates are explicitly computed in the forecasts.

We have explored the predictability of seasonal North Atlantic hurricane frequency using a set of retrospective forecasts from a pair of dynamical forecasting systems. The low computational cost of the statistical emulator allows us to perform the hundreds of retrospective forecasts needed to establish its skill. In addition, this statistical framework can easily take advantage of data available from many dynamical forecast models at minimal additional cost. We find a slight indication that the skill of the long-lead (February and March initializations) seasonal forecasts is enhanced through the ensemble of GFDL–CM2.1 and NCEP–CFS forecasts of SST. This suggests that employing a ''multimodel ensemble'' approach (by combining a variety of climate forecasting systems to develop a ''consensus'' estimate) could be a useful extension of the current methodology. Multiple forecast centers around the world routinely make multi-season initialized forecasts of the state of the global climate system, from which one could easily extract the relevant predictors (tropical Atlantic and tropical-mean SSTs) to use in our hurricane prediction scheme. The simplicity and negligible computational cost of the statistical hurricane frequency emulator employed in our forecast system, along with the explicit uncertainty estimates from convolving the uncertainty in SST with the uncertainty in hurricane frequency given SST, suggest that a multimodel ensemble modification to our technique is feasible.

Since El Niño variations impact seasonal North Atlantic hurricane frequency, it is reasonable to wonder whether our retrospective forecasts are only recovering skill from successful forecasts of El Niño–La Niña. Although predicting the inactive hurricane frequency during some El Niño years (e.g., 1997) and the active frequency during some La Niña years (e.g., 1995 and 1998) contributes to the skill of the retrospective hurricane forecasts, the skill of the hurricane forecasting system exceeds the skill one would expect from El Niño alone. The correlation skill of the retrospective forecasts is higher than that between commonly used measures of El Niño–La Niña and hurricane activity [0.5–0.6 versus 0.4–0.5 in HiRAM-C180 and observations; Zhao et al. (2010)], and the El Niño–La Niña forecasts in GFDL–CM2.1 and NCEP–CFS are not perfect. Additional skill in seasonal hurricane frequency may be coming from correctly predicting the coupled atmospheric processes that control seasonally phase-locked year-to-year variations in tropical Atlantic SSTs (e.g., Vimont and Kossin 2007; Doi et al. 2010).

Some of the retrospective skill comes from correctly diagnosing that the years before 1994 were more likely to be inactive, while later years had a greater tendency of being active. The causes of that heightened activity remain a topic of discussion, with natural climate variations, anthropogenic radiative forcing, or changes in atmospheric dust loading each likely contributing, though their relative influences remains a topic of active inquiry (e.g., Mann and Emanuel 2006; Zhang and Delworth 2006, 2009; Evan et al. 2009). Both dynamical SST forecast systems (CM2.1 and CFS) capture the tendency for increased frequency after 1995, as do the various persisted SST anomaly forecasts, suggesting that the SST anomalies driving the multidecadal increase in frequency are principally seasonally invariant. Since the skill of the initialized hurricane predictions is larger than that of the persisted SST anomaly forecasts and the WMO-recommended measure of the trailing 5-yr average (WMO 2008), both of which include aspects of the multidecadal shift, we see that the initialized forecasts include skill beyond correctly diagnosing the multidecadal increase in hurricane frequency.

TABLE 1. Summary of the forecasts for the upcoming 2010 North Atlantic hurricane season initialized in March 2010. The first column indicates the source of the prediction, and the second row indicates the expected value from the 1982–2009 North Atlantic hurricane record. The second column indicates the average/expected count of North Atlantic hurricanes in the 2010 season, the third column indicates the median of the counts, the fourth and fifth columns indicate the probability of exceeding 6 hurricanes (the long-term average) and 10 hurricanes (an extremely active year), and the sixth column indicates the probability of 3 or fewer hurricanes (an extremely inactive year). Notice that the forecasts are consistently suggesting that 2010 should be an active hurricane year relative to climatological values (top two rows) and even relative to the recent period of heightened North Atlantic hurricane activity (third row).

	Mean count (hurricanes)	Median (hurricanes)	p (count >6)	p (count >10)	p (count ≤ 3)
Climatology 1982–2009	6.21	5	0.46	0.07	0.21
Climatology 1966–2009	6.14	6	0.39	0.07	0.16
Climatology 1995–2009	7.67	7.5	0.67	0.13	0.13
GFDL–CM2.1 simple ensemble	9.88	9	0.80	0.39	0.03
GFDL–CM2.1 lagged ensemble	8.27	8	0.64	0.24	0.09
NCEP–CFS simple ensemble	7.64	7	0.61	0.19	0.04
NCEP–CFS lagged ensemble	8.24	8	0.65	0.25	0.03
Two-model simple ensemble	8.54	8	0.68	0.27	0.07
Two-model lagged ensemble	8.23	8	0.75	0.25	0.08

The retrospective forecasts of the extremely active 2005 initialized in January and March were particularly problematic: although the forecast system predicted an active year, it failed to predict the extreme values that occurred. It appears that part of the failure arose from an inability of the dynamical forecast systems to correctly predict the SST indices in 2005, since the statistical emulator using observed SSTs correctly identifies 2005 as the most active year over the 1982–2009 period. The observed counts for 2005 were one of a couple of years outside the 90% confidence interval of the "perfect forecasts" (Fig. 2); it is unclear at this stage if this is an indication of a strong stochastic contribution to the extreme values in 2005 (an option we cannot exclude) or a failure of the statistical emulator in 2005. Further work should explore the elements responsible for the extreme values in 2005.

The forecasting system has been developed for basin-wide hurricane frequency, yet the results of statistical modeling of tropical storm frequency (Villarini et al. 2010) and basin-wide measures that combine intensity and frequency (Swanson 2008; Vecchi et al. 2008) suggest that the procedure developed here could be adapted to these and possibly other quantities in the North Atlantic. Further, Zhao et al. (2010) indicate that a statistical model using two SST indices can successfully describe eastern Pacific hurricane activity in the HiRAM-C180 model, suggesting that a methodology analogous to the current one may be applicable to eastern Pacific seasonal hurricane forecasts. We are currently investigating these and related topics.

For the upcoming hurricane season of 2010, this seasonal hurricane forecasting system predicts an above average number of hurricanes (see Table 1); these predictions are quite similar to those from the various predictions with this system initialized since November

2009 (e.g., see Figs. 4a and 4b). The prediction system forecasts that the upcoming season has a higher expected hurricane count than the average over the recent decades, along with a higher probability of an extreme (>10 hurricanes) number of hurricanes. This indicates that it should be more similar to the active years that have dominated the North Atlantic since 1995 than the very inactive season of 2009, and suggests that we should not conclude that (due to the relatively inactive 2006, 2007, and 2009 seasons) the multidecadal period of heightened hurricane frequency has necessarily already abated.

REFERENCES

Behringer, D. W., and Y. Xue, 2004: Evaluation of the global ocean data assimilation system at NCEP: The Pacific Ocean. Preprints, *Eighth Symp. on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface,* Seattle, WA, Amer. Meteor. Soc., 2.3. [Available online at http://ams.confex.com/ams/pdfpapers/70720.pdf.]

——, M. Ji, and A. Leetmaa, 1998: An improved coupled model for ENSO prediction for ocean initialization. Part I: The ocean data assimilation system. *Mon. Wea. Rev.,* **126,** 1013–1021.

Cai, M., C.-L. Shin, H. M. van den Dool, W. Wang, S. Saha, and A. Kumar, 2009: the role of long-term trends in seasonal predictions: Implications of global warming in the NCEP CFS. *Wea. Forecasting,* **24,** 965–973.

Camargo, S. J., A. G. Barnston, P. Klotzbach, and C. W. Landsea, 2007: Seasonal tropical cyclone forecasts. *WMO Bull.*, **56**, 297–309.

Delworth, T. L., and Coauthors, 2006: GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J. Climate*, **19**, 643–674.

Doi, T., T. Tozuka, and T. Yamagata, 2010: The Atlantic meridional mode and its coupled variability with the Guinea Dome. *J. Climate*, **23**, 455–475.

Elsner, J., and T. Jagger, 2006: Prediction models for annual U.S. hurricane counts. *J. Climate*, **19**, 2935–2952.

Emanuel, K., 2005: Increasing destructiveness of tropical cyclones over the past 30 years. *Nature*, **436**, 686–688.

Evan, A. T., D. J. Vimont, R. Bennartz, J. P. Kossin, and A. K. Heidinger, 2009: The role of aerosols in the evolution of tropical North Atlantic Ocean temperature. *Science*, **324**, 778–781, doi:10.1126/science.1167404.

Gray, W. M., 1984: Atlantic seasonal hurricane frequency. Part II: Forecasting its variability. *Mon. Wea. Rev.*, **112**, 1669–1683.

Klotzbach, P., and W. Gray, 2009: Twenty-five years of Atlantic basin seasonal hurricane forecast (1984–2008). *Geophys. Res. Lett.*, **36**, L09711, doi:10.1029/2009GL037580.

Knutson, T. R., J. J. Sirutis, S. T. Garner, G. A. Vecchi, and I. M. Held, 2008: Simulated reduction in Atlantic hurricane frequency under twenty-first-century warming conditions. *Nat. Geosci.*, **1**, 359–364, doi:10.1038/ngeo202.

Landsea, C. W., G. A. Vecchi, L. Bengtsson, and T. R. Knutson, 2010: Impact of duration thresholds on Atlantic tropical cyclone counts. *J. Climate*, **23**, 2508–2519.

LaRow, T., L. Stefanova, D.-W. Shin, and S. Cocke, 2010: Seasonal Atlantic tropical cyclone hindcasting/forecasting using two sea surface temperature datasets. *Geophys. Res. Lett.*, **37**, L02804, doi:10.1029/2009GL041459.

Latif, M., N. Keenlyside, and J. Bader, 2007: Tropical sea surface temperature, vertical wind shear, and hurricane development. *Geophys. Res. Lett.*, **34**, L01710, doi:10.1029/2006GL027969.

Mann, M. E., and K. A. Emanuel, 2006: Atlantic hurricane trends linked to climate change. *Eos, Trans. Amer. Geophys. Union*, **87**, 233–244.

Meehl, G. A., and Coauthors, 2007: The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bull. Amer. Meteor. Soc.*, **88**, 1383–1394.

Moorthi, S., H.-L. Pan, and P. Caplan, 2001: Changes to the 2001 NCEP operational MRF/AVN global analysis/forecast system. NWS Tech. Procedures Bull. 484, 14 pp. [Available online at http://www.nws.noaa.gov/om/tpb/484.htm.]

Pacanowski, R. C., and S. M. Griffies, 1998: MOM 3.0 manual. NOAA/GFDL, 668 pp.

Putman, W. M., and S.-J. Lin, 2007: Finite-volume transport on various cubed-sphere grid. *J. Comput. Phys.*, **227**, 55–78.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, doi:10.1029/2002JD002670.

Rice, J. A., 1995: *Mathematical Statistics and Data Analysis*. 2nd ed. Duxbury Press, 602 pp.

Saha, S., and Coauthors, 2006: The NCEP Climate Forecast System. *J. Climate*, **19**, 3483–3517.

Shen, W., R. E. Tuleya, and I. Ginis, 2000: A sensitivity study of the thermodynamic environment on GFDL model hurricane intensity: Implications for global warming. *J. Climate*, **13**, 109–121.

Sobel, A. H., I. M. Held, and C. S. Bretherton, 2002: The ENSO signal in tropical tropospheric temperature. *J. Climate*, **15**, 2702–2706.

Stasinopoulos, D. M., R. A. Rigby, and C. Akantziliotou, cited 2009: gamlss: Generalized Additive Models for Location Scale and Shape. r package version 1.6-0. [Available online at http://www.londonmet.ac.uk/gamlss/.]

Swanson, K. L., 2008: Non-locality of Atlantic tropical cyclone intensities. *Geochem. Geophys. Geosyst.*, **9**, Q04V01, doi:10.1029/2007GC001844.

Tang, B. H., and J. D. Neelin, 2004: ENSO influence on Atlantic hurricanes via tropospheric warming. *Geophys. Res. Lett.*, **31**, L24204, doi:10.1029/2004GL021072.

Vecchi, G. A., and B. J. Soden, 2007: Effect of remote sea surface temperature change on tropical cyclone potential intensity. *Nature*, **450**, 1066–1070, doi:10.1038/nature06423.

——, and T. R. Knutson, 2008: On estimates of historical North Atlantic tropical cyclone activity. *J. Climate*, **21**, 3580–3600.

——, K. L. Swanson, and B. J. Soden, 2008: Whither hurricane activity? *Science*, **322**, 687, doi:10.1126/science.1164396.

Villarini, G., G. A. Vecchi, and J. A. Smith, 2010: Modeling of the dependence of tropical storm counts in the North Atlantic basin on climate indices. *Mon. Wea. Rev.*, **138**, 2681–2705.

——, ——, T. R. Knutson, M. Zhao, and J. A. Smith, 2011: North Atlantic tropical storm frequency response to anthropogenic forcing: Projections and sources of uncertainty. *J. Climate*, in press.

Vimont, D. J., and J. P. Kossin, 2007: The Atlantic meridional mode and hurricane activity. *Geophys. Res. Lett.*, **34**, L07709, doi:10.1029/2007GL029683.

Vitart, F., 2006: Seasonal forecasting of tropical storm frequency using a multi-model ensemble. *Quart. J. Roy. Meteor. Soc.*, **132**, 647–666.

——, and Coauthors, 2007: Dynamically-based seasonal forecasts of Atlantic tropical storm activity issued in June by EUROSIP. *Geophys. Res. Lett.*, **34**, L16815, doi:10.1029/2007GL030740.

Wang, H., J.-K. E. Schemm, A. Kumar, W. Wang, L. Long, M. Chelliah, G. D. Bell, and P. Peng, 2009: A statistical forecast model for Atlantic seasonal hurricane activity based on the NCEP dynamical seasonal forecast. *J. Climate*, **22**, 4481–4500.

WMO, 2008: Report from expert meeting to evaluate skill of tropical cyclone seasonal forecasts. World Meteorological Organization Tech. Doc. 1455, Geneva, Switzerland. 27 pp.

Zhang, R., and T. L. Delworth, 2006: Impact of Atlantic multidecadal oscillations on India/Sahel rainfall and Atlantic hurricanes. *Geophys. Res. Lett.*, **33**, L17712, doi:10.1029/2006GL026267.

——, and ——, 2009: A new method for attributing climate variations over the Atlantic hurricane basin's main development region. *Geophys. Res. Lett.*, **36**, L06701, doi:10.1029/2009GL037260.

Zhang, S., M. J. Harrison, A. Rosati, and A. T. Wittenberg, 2007: System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Mon. Wea. Rev.*, **135**, 3541–3564.

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.

——, ——, and G. A. Vecchi, 2010: Retrospective forecasts of the hurricane season using a global atmospheric model assuming persistence of SST anomalies. *Mon. Wea. Rev.*, **138**, 3858–3868.