

# Experimental Dynamical Seasonal Forecasts of Tropical Cyclone Activity at IRI

SUZANA J. CAMARGO AND ANTHONY G. BARNSTON

*International Research Institute for Climate and Society, The Earth Institute at Columbia University, Palisades, New York*

(Manuscript received 7 December 2007, in final form 15 September 2008)

## ABSTRACT

The International Research Institute for Climate and Society (IRI) has been issuing experimental seasonal tropical cyclone activity forecasts for several ocean basins since early 2003. In this paper the method used to obtain these forecasts is described and the forecast performance is evaluated. The forecasts are based on tropical cyclone-like features detected and tracked in a low-resolution climate model, namely ECHAM4.5. The simulation skill of the model using historical observed sea surface temperatures (SSTs) over several decades, as well as with SST anomalies persisted from the previous month's observations, is discussed. These simulation skills are compared with skills of purely statistically based hindcasts using as predictors recently observed SSTs. For the recent 6-yr period during which real-time forecasts have been made, the skill of the raw model output is compared with that of the subjectively modified probabilistic forecasts actually issued.

Despite variations from one basin to another, the levels of hindcast skill for the dynamical and statistical forecast approaches are found, overall, to be approximately equivalent at fairly modest but statistically significant levels. The dynamical forecasts require statistical postprocessing (calibration) to be competitive with, and in some circumstances superior to, the statistical models. Skill levels decrease only slowly with increasing lead time up to 2–3 months. During the recent period of real-time forecasts, the issued forecasts have had higher probabilistic skill than the raw model output, due to the forecasters' subjective elimination of the "overconfidence" bias in the model's forecasts. Prospects for the future improvement of dynamical tropical cyclone prediction are considered.

## 1. Introduction

Tropical cyclones (TCs; see the appendix for a list of the acronyms used in this paper) are one of the most devastating types of natural disasters. Seasonal forecasts of TC activity could help the preparedness of coastal populations for an upcoming TC season and reduce economical and human losses.

Currently, many institutions issue operational seasonal TC forecasts for various regions. In most cases, these are statistical forecasts, such as the Atlantic hurricane outlooks produced by NOAA (information online at <http://www.cpc.noaa.gov/products/outlooks/hurricane.shtml>), and Colorado State University (Gray et al. 1993; Klotzbach 2007a), the typhoon activity forecasts of the City University of Hong Kong (Chan et al. 1998, 2001), and Tropical Storm Risk (Saunders and Lea 2004). A review of TC

seasonal forecasts is found in Camargo et al. (2007a), and the skill levels of some of them were discussed in Owens and Landsea (2003).

Since April 2003 the International Research Institute for Climate and Society (IRI) has been issuing experimental dynamical seasonal forecasts for five ocean basins (information online at <http://portal.iri.columbia.edu/forecasts>). In this paper, we describe how these forecasts are produced and discuss their skills when the atmospheric general circulation model (AGCM) is forced by predicted sea surface temperature (SST) in a two-tiered prediction system.

The possible use of dynamical climate models for forecasting seasonal TC activity has been explored by various authors (e.g., Bengtsson et al. 1982). Although the low horizontal resolution of climate general circulation models of the early 2000s is not adequate to realistically reproduce the structure and behavior of individual cyclones, such models are capable of forecasting with some skill several aspects of the general level of TC activity over the course of a season (Bengtsson 2001; Camargo et al. 2005). Dynamical TC forecasts can serve

---

*Corresponding author address:* Suzana Camargo, Lamont-Doherty Earth Observatory, The Earth Institute at Columbia University, 61 Rte. 9W, P.O. Box 1000, Palisades, NY 10964-8000.  
E-mail: [suzana@ldeo.columbia.edu](mailto:suzana@ldeo.columbia.edu)

specific applications, for example, TC landfall activity over Mozambique (Vitart et al. 2003). The level of performance of dynamical TC forecasts depends on many factors, including the model used (Camargo et al. 2005), the model resolution (Bengtsson et al. 1995), and the inherent predictability of the large-scale circulation regimes (Vitart and Anderson 2001), including those related to El Niño–Southern Oscillation (ENSO) (Wu and Lau 1992; Vitart et al. 1999).

In addition to IRI's dynamically based experimental TC forecasts, such forecasts are also produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) (Vitart 2006), the Met Office, and the European Seasonal to Interannual Prediction (EUROSIP) superensemble of ECMWF, Met Office, and Météo-France coupled models (Vitart et al. 2007). An important consideration is the dynamical design used to produce the forecasts. The European dynamical TC forecasts are produced using fully coupled atmosphere–ocean models (Vitart and Stockdale 2001; Vitart 2006). At IRI, a two-tiered (Bengtsson et al. 1993), multimodel (Rajagopalan et al. 2002; Robertson et al. 2004) procedure is used to produce temperature and precipitation forecasts once an SST forecast (or set of them) is first established (Mason et al. 1999; Goddard et al. 2003; Barnston et al. 2003, 2005). The IRI experimental TC forecasts use a subset of the IRI two-tier forecast system, in that only a single AGCM is used, compared with several AGCMs for surface climate. As described below, more than one SST forcing scenario is used.

TCs in low-resolution models have many characteristics comparable to those observed, but at much lower intensity and larger spatial scale (Bengtsson et al. 1995; Vitart et al. 1997). The climatology, structure, and interannual variability of model TCs have been examined (Bengtsson et al. 1982, 1995; Vitart et al. 1997; Camargo and Sobel 2004). A successful aspect of this work has been that, over the course of a TC season in a statistical sense, the spatial and temporal distributions, as well as the interannual anomalies of the number and total energy content, of model TCs roughly follow those of observed TCs (Vitart et al. 1997; Camargo et al. 2005). There have been two general methods in which climate models are used to forecast TC activity. One method is to analyze large-scale variables known to affect TC activity (Ryan et al. 1992; Thorncroft and Pytharoulis 2001; Camargo et al. 2007c). Another approach, and the one used here, is to detect and track the cyclonelike structures in climate models (Manabe et al. 1970; Broccoli and Manabe 1990; Wu and Lau 1992), coupled ocean–atmosphere models (Matsuura et al. 2003; Vitart and Stockdale 2001), and regional climate models (Landman et al. 2005; Knutson et al. 2007). These methods have also been used

TABLE 1. Tropical Pacific SST forecast types used in this study. The concurrent, OSST data, for nonforecast simulations (lead time less than zero), S denotes the Reynolds version 2 data (Reynolds et al. 2002); the real-time persisted SST (FSST<sub>p</sub>) and hindcast persisted SST (HSST<sub>p</sub>) are undamped and then damped with persisted anomalies initialized from anomalies observed the previous month. The evolving anomalous SST (FSST<sub>e</sub>) is from one or more of the following models: the NCEP coupled model (Ji et al. 1998), the CFS (Saha et al. 2006), the LDEO-5 (Chen et al. 2004), and/or the statistical constructed analog (CA) (van den Dool 1994, 2007, chapter 7). More details about the FSST<sub>e</sub> are provided in Camargo and Barnston (2008).

SST type	Period	Ensembles	Lead	Pacific SSTs
OSST	Jan 1950–Dec 2005	24	S	Reynolds
OSST <sub>r</sub>	Jan 1970–Dec 2005	24	S	Reynolds
HSST <sub>p</sub>	Jan 1968–May 2003	12	4	Persisted
FSST <sub>p</sub>	Aug 2001–Aug 2007	24	4	Persisted
FSST <sub>e</sub>	Aug 2001–Aug 2007	24	6	NCEP, CFS, LDEO-5, CA

in studies of possible changes in TC intensity due to global climate change using AGCMs (Bengtsson et al. 1996; Royer et al. 1998; Bengtsson et al. 2007a,b) and regional climate models (Walsh and Ryan 2000; Walsh et al. 2004).

In section 2 we describe how the real-time seasonal tropical forecasts are produced at IRI. The model's performance over a multidecadal hindcast period and over the recent 6-yr period of real-time forecasting is discussed in section 3. A comparison of the AGCM performance result with that of simple SST-based statistical forecasts is shown in section 4. The conclusions are given in section 5.

## 2. Description of the real-time forecasts

The IRI climate forecast system (Mason et al. 1999) is two-tiered: SSTs are first forecasted, and then each of a set of atmospheric models is forced with several tropical SST forecast scenarios. Many ensemble members of atmospheric response are produced from each model forced with the SST scenarios. For the TC seasonal forecasts, just one atmospheric model is used: ECHAM4.5, which is run on a monthly basis. Six-hourly output data are used, as this fine temporal resolution makes possible the detection of the needed TC characteristics. The ECHAM4.5 was developed at the Max Planck Institute for Meteorology in Hamburg, Germany (Roeckner et al. 1996), and has been studied extensively for various aspects of seasonal TC activity (Camargo and Zebiak 2002; Camargo et al. 2005, 2007c).

The integrations of the ECHAM4.5 model are subject to differing tropical SST forcing scenarios (Table 1). In all of the scenarios, the extratropical SST forecasts

consist simply of the damped persistence of the anomalies from the previous month's observation (added to the forecast season's climatology), with an anomaly  $e$ -folding time of 3 months (Mason et al. 1999). In the tropics, multimodel, mainly dynamical, SST forecasts are used for the Pacific, while statistical and dynamical forecasts are combined for the Indian and Atlantic Oceans. Statistical forecasts play the greatest role in the tropical Atlantic. The models contributing to the tropical SST forecasts, particularly for the Pacific, have changed during our study period as forecast-producing centers have introduced newer, more advanced prediction systems. In the non-Pacific tropical basins during seasons having near-zero apparent SST forecast predictive skill, damped persisted SST anomalies are used, but at a lower damping rate than that used in the extratropics. (No damping occurs in the first 3 months, followed by linear damping that reaches zero by month 8.) However, for seasons in which SST predictive skill is found beyond that of damped persistence, CCA models are used in the Indian Ocean (Mason et al. 1999) and the tropical Atlantic Ocean (Repelli and Nobre 2004).

Globally undamped anomalous SST persisted from the previous month, applied to the climatology of the months being forecast, is used as an additional SST forcing scenario (called FSST<sub>p</sub>). In this case the 24 ensemble members of ECHAM4.5 are integrated using persisted SST anomalies out to 5 months beyond the previous month. For example, for a mid-January forecast, the model is forced from January to May using undamped persisted SST anomalies from December globally.<sup>1</sup>

In the case of the nonpersisted, evolving forecasted SST anomalies (denoted by FSST<sub>e</sub>), the AGCM is run out to 7 months beyond the previous month's observed SST (e.g., for a mid-January forecast, observed SST exists for December, and the model is forced from January to July with evolving SST predictions). Several versions of the forecasted SST anomalies have been used since 2001. These are described in detail in Camargo and Barnston (2008).<sup>2</sup>

The ECHAM4.5 was also forced with the actual observed SSTs (OSSTs; Reynolds et al. 2002) prescribed

during the period from 1950 to the present. These AMIP-type runs provide estimates of the upper limit of the skill of the model in forecasting TC activity, as discussed in previous studies (Camargo and Zebiak 2002; Camargo et al. 2005). The skill levels presented below are broken out into three SST forcing types: 1) FSST (for real-time forecasts, comprising FSST<sub>p</sub> and FSST<sub>e</sub>), 2) HSST<sub>p</sub> (long-term hindcast anomaly persisted SST), and 3) OSST (long-term observed SST for AMIP-type AGCM simulations).

For any type of SST forcing, we analyze the output of the AGCM for TC activity. To define and track TCs in the models, we used objective algorithms (Camargo and Zebiak 2002) based in large part on prior studies (Vitart et al. 1997; Bengtsson et al. 1995). The algorithm has two parts: detection and tracking. In the detection part, storms that meet environmental and duration criteria are identified. A model TC is identified when chosen dynamical and thermodynamical variables exceed thresholds calibrated to the observed tropical storm climatology.<sup>3</sup> Most studies (Bengtsson et al. 1982; Vitart et al. 1997) use a single set of threshold criteria globally. However, to take into account model biases and deficiencies, we use basin- and model-dependent threshold criteria, based on analyses of the correspondence between the modeled and observed climatologies (Camargo and Zebiak 2002). Thus, we use a threshold exclusive to ECHAM4.5. Once detected, the TC tracks are obtained from the vorticity centroid, defining the center of the TC, using relaxed criteria appropriate for the weak model storms. The detection and tracking algorithms have been applied to regional climate models (Landman et al. 2005; Camargo et al. 2007b) and to multiple AGCMs (Camargo and Zebiak 2002; Camargo et al. 2005).

Following detection and tracking, we count the number of TCs (NTC) and compute the model accumulated cyclone energy (ACE) index (Bell et al. 2000) over a TC season. ACE is defined as the sum of the squares of the wind speeds in the TCs active in the model at each 6-h interval. For the observed ACE, only TCs of tropical storm intensity or greater are included.

The model ACE and NTC results are then corrected for bias, based on the historical model and observed distributions of NTC and ACE over the 1971–2000 period, on a per basin basis. Corrections yield matching

<sup>1</sup> The FSST<sub>p</sub> runs have been produced in real time from August 2001 to the present and are available also in hindcast mode (HSST<sub>p</sub>) over the period January 1968–May 2003, with 12 ensemble members (Table 1).

<sup>2</sup> The tropical Pacific SST has been based on one or more of the following: the NCEP coupled ENSO prediction model (Ji et al. 1998), the NCEP Climate Forecast System (NCEP-CFS) (Saha et al. 2006), the Lamont-Doherty Earth Observatory intermediate model, version 5 (LDEO-5), (Chen et al. 2004), and the statistical constructed analog (CA) model (van den Dool 1994, 2007, chapter 7).

<sup>3</sup> A model TC needs to exceed simultaneously thresholds for low-level vorticity (850 hPa), surface wind speed, and vertically integrated local temperature anomaly for at least 2 days, and must also have a relative local minimum of sea level pressure, local maximum of temperature anomalies in various levels, and mean wind speed at 850 hPa larger than at 300 hPa.

values within a percentile reference framework (i.e., a correspondence is achieved nonparametrically). Using 1971–2000 as the climatological base period, tercile boundaries for model and observed NTCs and ACEs are then defined, since the forecasts are probabilistic with respect to tercile-based categories of the climatology (below, near, and above normal).<sup>4</sup>

For each of the SST forcing designs, we count the number of ensemble members having their NTCs and ACEs in a given ocean basin in the below normal, normal, and above normal categories, and divide by the total number of ensembles. These constitute the “raw,” objective probability forecasts. In a final stage of forecast production, the IRI forecasters examine and discuss these objective forecasts and develop subjective final forecasts that are posted on the IRI Web site. The most typical difference between the raw and the subjective forecasts is that the latter have weaker probabilistic deviations from climatology, given the knowledge that the models are usually too “confident.” The overconfidence of the model may be associated with too narrow an ensemble spread, too strong a model signal (deviation of ensemble mean from climatology), or both of these. The subjective modification is intended to increase the probabilistic reliability of the predictions. The issues of model overconfidence, calibration to correct it, and probabilistic reliability will be discussed in more detail in section 3b. Another consideration in the subjective modification is the degree of agreement among the forecasts, in which less agreement would suggest greater uncertainty and thus more caution with respect to the amount of deviation from the climatological probabilities.

The raw objective forecasts are available starting from August 2001. The first subjective forecast for the western North Pacific basin was produced in real time in April 2003. However, subjective hindcasts were also produced for August 2001–April 2003 without knowledge of the observed result, making for 6 yr of experimental forecasts.

For each ocean basin, forecasts are produced only for the peak TC season, from certain initial months prior to that season (Table 2), and updated monthly until the first month of the peak season.<sup>5</sup> The lead time of this

TABLE 2. Ocean basins in which IRI experimental TC forecasts are issued: eastern North Pacific (ENP), western North Pacific (WNP), North Atlantic (ATL), Australia (AUS), and South Pacific (SP). Also shown are the month and year of the first issued forecast; the seasons for which TC forecasts are issued (JJAS, ASO, JASO, JFM, and DJFM); months in which the forecasts are issued; and variables forecasted—NTC and/or ACE.

Basin	First forecast	Season	Months forecasts are issued	Variables
ENP	Mar 2004	JJAS	Mar, Apr, May, Jun	NTC, ACE
WNP	Apr 2003	JASO	Apr, May, Jun, Jul	NTC, ACE
ATL	Jun 2003	ASO	Apr, May, Jun, Jul, Aug	NTC, ACE
AUS	Sep 2003	JFM	Sep, Oct, Nov, Dec, Jan	NTC
SP	Sep 2003	DJFM	Sep, Oct, Nov, Dec	NTC

latest forecast is defined as being zero, and the lead times of earlier forecasts are defined by the number of months earlier that they are issued.

The basins in which forecasts are issued are shown in Fig. 1, and the numbers of years available for each SST scenario and basin are indicated in Table 3. In the Southern Hemisphere (South Pacific and Australian regions), only forecasts for NTC are produced, while in the Northern Hemisphere basins both NTC and ACE forecasts are issued. ACE is omitted for the Southern Hemisphere because ACE is more sensitive to data quality than NTC, and the observed TC data from the Southern Hemisphere are known to be of somewhat questionable quality, particularly in the earlier half of the study period (e.g., Chu et al. (2002); Buckley et al. (2003); Landsea et al. (2006); Trewin (2008); Harper et al. (2008)).

The observed TC data used to correct historical model biases and for verification of the model forecasts is the best-track data from the National Hurricane Center (Atlantic and eastern North Pacific; information online at <http://www.nhc.noaa.gov>) and the Joint Typhoon Warning Center (western North Pacific and Southern Hemisphere; information online at <https://metocph.nmci.navy.mil/jtwc.php>).

### 3. Performance in hindcasts and real-time forecasts

NTC or ACE historical simulation and real-time predictive skill results are computed for each ocean basin for their respective peak TC seasons. Both deterministic and probabilistic skills are examined.

#### a. Deterministic skills

Temporal anomaly correlation skills are shown in Table 4 for NTC by lead time, for each type of SST forcing, and likewise for ACE in Table 5. The simulation

<sup>4</sup> For the South Pacific, a temporal aspect of bias correction is used: the TC forecast season is DJFM, but the model output for NDJF is used to forecast DJFM, including the bias correction. It is found that hindcast skill levels are appreciably higher with this 1-month offset, which we consider a temporal aspect of the bias correction.

<sup>5</sup> The data available for the forecast released during the first month of the TC peak season cover only through the end of the previous month.

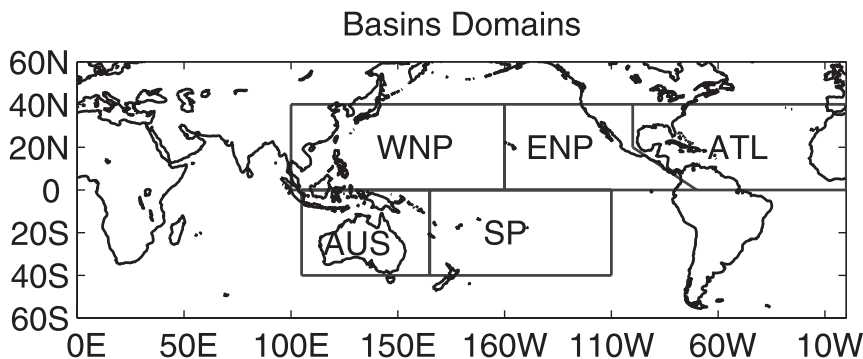


FIG. 1. Definition of the ocean basin domains used in this study: Australian (AUS), 105°–165°E; South Pacific (SP), 165°E–110°W; western North Pacific (WNP), 100°E–160°W; eastern North Pacific (ENP), 160°–100°W; and Atlantic (ATL), 100°W–0°. All latitude boundaries are along the equator and 40°N or 40°S. Note the unique boundary paralleling Central America for the ENP and ATL basins.

skills are shown both for the full period of 1950–2005 and for 1970–2005, during which time the TC data are known to be of higher quality. The correlations for the real-time predictions are uncentered.<sup>6</sup> Simulation skills (OSST) are seen to be at statistically significant levels for most of the ocean basins. Skills for the longer period (OSST<sub>r</sub>) tend to exceed those for 1970–2005, due both to better average data quality and the greater ENSO variability following 1970. Consistent with Camargo et al. (2005), the highest skill results occur in the Atlantic basin with correlations of roughly 0.50, with more modest skill levels in the other basins. Skill levels for zero-lead forecasts using SST anomalies persisted from those of the most recent month (HSST<sub>p</sub>, lead 0), as expected, are usually lower than those of observed simultaneous SSTs. For the three Northern Hemisphere basins, simulation skills are higher for ACE than for NTC, as noted also in Camargo et al. (2005). This may be related to the continuous nature of ACE as opposed to the discrete, more nonparametric, character of NTC.

A reference forecast more difficult to beat than a random or a climatology forecast is that of simple persistence of observed TC observation from the previous year. The correlation score for such a reference forecast is just the 1-yr autocorrelation coefficient over the 1971–2005 base period, and is shown at the bottom of Tables 4

and 5. The persistence correlation scores are lower than those of the AGCM's forecast using observed or persisted SST, with the one exception of the NTC forecasts in the northwestern Pacific.

Real-time predictive verification skill levels (FSST in Tables 4 and 5) over the basins not only have lower expected values than those using simultaneous observed SST due to the imperfection of the predicted SST forcing, but also much greater sampling errors given only six to seven cases per lead time per basin (Table 3). These skills range from near or below zero for the western North Pacific NTC to approximately 0.5 for the three shortest leads for the eastern North Pacific ACE. For all basins collectively and for NTC and ACE together, the skill results approximate those of HSST<sub>p</sub>, with individual differences likely due foremost to sampling variability. Consistent with the small sample problem, the correlations for FSST for all of the basin–lead time combinations are statistically nonsignificant, as nearly 0.8 is required for significance.

A look at the possible impact of differing SST forcing scenarios and lead times on the real-time forecast skills is more meaningful when results for all oceans basins are combined, lessening the sampling problem. Basin-combined skill results by lead time and SST forcing type are shown in Table 6 for NTC and ACE. Table 6 shows NTC results for Northern Hemisphere basins only, allowing a direct comparison between NTCs and ACEs. The results show higher skill levels for forecasts of ACE than NTC, and only a very weak tendency for decreasing skill with increasing lead time. This is summarized further in the bottom row of Table 6, showing the results for NTC and ACE combined.

Skill levels were evaluated using additional deterministic verification measures: the Spearman rank correlation,

<sup>6</sup> In computing the correlation skill for forecasts for much shorter periods than the climatological base period, the subperiod means are not removed and are not used for computing the standard deviation terms. Instead, the longer base period means are used. This is done so that, for example, if in the subperiod the forecasts and observations have small-amplitude out-of-phase variations but both are generally on the same side of the longer period mean, a positive correlation would result, and we believe justifiably.

TABLE 3. Number of years for each lead and SST type. Here, S denotes simulations, with a negative lead time.

SST type	ENP					WNP					ATL						
	Lead	3	2	1	0	S	3	2	1	0	S	4	3	2	1	0	S
FSST	6	6	6	6	—	6	6	6	6	—	6	6	6	6	7	—	—
HSST <sub>p</sub>	—	—	35	35	—	—	—	35	35	—	—	—	35	35	35	—	—
OSST	—	—	—	—	56	—	—	—	—	56	—	—	—	—	—	—	56
OSST <sub>r</sub>	—	—	—	—	36	—	—	—	—	36	—	—	—	—	—	—	36

SST type	AUS					SP						
	Lead	4	3	2	1	0	S	3	2	1	0	S
FSST	6	6	6	6	6	—	6	6	6	6	—	—
HSST <sub>p</sub>	—	—	34	34	35	—	—	—	34	34	—	—
OSST	—	—	—	—	—	56	—	—	—	—	55	—
OSST <sub>r</sub>	—	—	—	—	—	36	—	—	—	—	35	—

the Heidke skill score, and the mean squared error skill score (MSESS). Table 7 provides an example of the four scores together, for ACE in the northwestern Pacific basin. The rank correlation and Heidke skill scores are roughly consistent with the correlation skill, allowing for expected scaling differences where the Heidke is roughly one-half of the correlation (Barnston 1992). The MSESS, however, which uses the 1971–2000 climatology as the zero-skill reference forecast, is comparatively unfavorable: some of the cases having positive correlation and Heidke skills have negative MSESS results. This outcome is attributable to a marked tendency of the model forecasts toward too great a departure from climatological forecasts, given the degree of inherent uncertainty and thus the relatively modest level of true predictability. Such “overconfidence” in the model forecasts, which can be adjusted for statistically, will be discussed in more detail below within the context of

probabilistic verification, where a detrimental effect on scores comparable to that seen in MSESS will become apparent.

*b. Probabilistic skills*

The TC forecasts were verified probabilistically using the ranked probability skill score (RPSS), likelihood skill score, and, for the real-time forecasts, the relative operating characteristics (ROC) score.

RPSS (Epstein 1969; Goddard et al. 2003) measures the sum of the squared errors between categorical forecast probabilities and the observed categorical probabilities, cumulative over categories, relative to a reference (or standard baseline) forecast—here, the climatology forecast of 1/3 probability for each category. The observed probabilities are 1 for the observed category and 0 for the other categories.

TABLE 4. Correlations ( $\times 10^2$ ) with observations for NTC, per basin, by lead time and SST forecast scenario. An “S” denotes simulations, whose lead times are negative. “Pers” denotes 1-yr simple persistence, with a lead potentially longer than 4 months, but shown in the column of the longest lead. Statistically significant skills are shown in boldface.

SST type	Eastern Pacific					Western Pacific					Atlantic						
	Lead	3	2	1	0	S	3	2	1	0	S	4	3	2	1	0	S
FSST	53	58	58	46	—	—84	—82	—68	2	—	13	48	65	35	11	—	—
HSST <sub>p</sub>	—	—	32	<b>50</b>	—	—	—	11	7	—	—	—	7	<b>33</b>	<b>36</b>	—	—
OSST	—	—	—	—	<b>32</b>	—	—	—	—	21	—	—	—	—	—	—	<b>40</b>
OSST <sub>r</sub>	—	—	—	—	<b>37</b>	—	—	—	—	28	—	—	—	—	—	—	<b>55</b>
Pers	9	—	—	—	—	<b>33</b>	—	—	—	—	14	—	—	—	—	—	—

SST type	Australia					South Pacific						
	Lead	4	3	2	1	0	S	3	2	1	0	S
FSST	52	40	7	—1	21	—	44	72	63	34	—	—
HSST <sub>p</sub>	—	—	—4	26	16	—	—	—	<b>42</b>	<b>42</b>	—	—
OSST	—	—	—	—	—	<b>38</b>	—	—	—	—	<b>43</b>	—
OSST <sub>r</sub>	—	—	—	—	—	<b>40</b>	—	—	—	—	<b>42</b>	—
Pers	11	—	—	—	—	—	—26	—	—	—	—	—

TABLE 5. As in Table 4 (correlations) but for ACE.

SST type	Eastern Pacific					Western Pacific					Atlantic						
	Lead	3	2	1	0	S	3	2	1	0	S	4	3	2	1	0	S
FSST	—	−16	47	48	62	—	69	2	21	36	—	−4	25	17	61	14	—
HSST <sub>p</sub>	—	—	—	20	23	—	—	—	6	23	—	—	—	18	<b>43</b>	<b>37</b>	—
OSST	—	—	—	—	—	7	—	—	—	—	26	—	—	—	—	—	<b>57</b>
OSST <sub>r</sub>	—	—	—	—	—	<b>45</b>	—	—	—	—	<b>33</b>	—	—	—	—	—	<b>60</b>
Pers	21	—	—	—	—	—	20	—	—	—	—	<b>35</b>	—	—	—	—	—

Verifications using the RPSS are shown for NTC and ACE in Tables 8 and 9. These skills are mainly near or below zero. This poor result can be attributed to the lack of probabilistic reliability of the model ensemble-based TC predictions as is seen in many predictions made by individual AGCMs—not just for TC activity but for most climate variables (Anderson 1996; Barnston et al. 2003; Wilks 2006). Climate predictions by AGCMs have model-specific systematic biases, and their uncorrected probabilities tend to deviate too strongly from climatological probabilities due to too small an ensemble spread and/or too large a mean shift from climatology. This problem leads to comparably poor probability forecasts, despite positive correlation skills for the ensemble means of the same forecast sets. Positive correlations, but negative probabilistic verification, are symptomatic of poorly

calibrated probability forecasts—a condition that can be remedied using objective statistical correction procedures.

Probabilistic persistence may be a more competitive simple reference forecast than forecasts of climatological probabilities. Based on the weak but generally positive year-to-year autocorrelations shown in Tables 4 and 5, we designed the persistence probabilistic forecasts to be 0.4 for the tercile-based category observed the previous year, and 0.3 for the other two categories. Resulting RPSSs are shown at the bottom of Tables 8 and 9. These weakly persistent probabilistic forecasts often have better RPSS scores than those of the AGCM forced with persisted SSTs (HSST<sub>p</sub>), and sometimes as good as or better than those forced with observed SSTs. Rather than showing that use of the AGCM with observed or predicted SSTs is unsuccessful, this outcome again shows that probabilities that deviate only mildly from climatological probabilities, even if derived from something as simple as the TC activity of the previous year, fare better under calibration-sensitive probabilistic verification

TABLE 6. Correlations ( $\times 10^2$ ) for all basins combined, by lead time and SST forecast scenario. The S denotes simulations, whose lead times are negative. Statistically significant skills are shown in boldface. The sample size is doubled for significance evaluations for all basins combined, and increased by 60% for the Northern Hemisphere basins combined, relative to the single-basin sample.

Lead	4	3	2	1	0	S
NTC all basins						
FSST	38	18	35	28	27	—
HSST <sub>p</sub>	—	—	1	<b>25</b>	<b>27</b>	—
OSST	—	—	—	—	—	<b>25</b>
OSST <sub>r</sub>	—	—	—	—	—	<b>46</b>
NTC Northern Hemisphere						
FSST	13	−25	17	−3	27	—
HSST <sub>p</sub>	—	—	7	23	<b>28</b>	—
OSST	—	—	—	—	—	<b>31</b>
OSST <sub>r</sub>	—	—	—	—	—	<b>42</b>
ACE Northern Hemisphere						
FSST	−4	41	29	46	41	—
HSST <sub>p</sub>	—	—	17	<b>28</b>	<b>29</b>	—
OSST	—	—	—	—	—	<b>39</b>
OSST <sub>r</sub>	—	—	—	—	—	<b>48</b>
NTC and ACE all basins						
FSST	31	29	33	36	33	—

TABLE 7. Comparison of four skill measures ( $\times 10^2$ ) for FSST, HSST<sub>p</sub>, and OSST and OSST, for ACE in the western North Pacific basin. Statistically significant skills are shown in boldface.

Western North Pacific ACE		Lead		
SST type	Score	1	0	S
FSST	Correlation	11	34	—
FSST	Rank correlation	44	26	—
FSST	Heidke skill score	3	7	—
FSST	MSESS	−63	−37	—
HSST <sub>p</sub>	Correlation	6	23	—
HSST <sub>p</sub>	Rank correlation	−5	<b>34</b>	—
HSST <sub>p</sub>	Heidke skill score	1	12	—
HSST <sub>p</sub>	MSESS	−61	−27	—
OSST	Correlation	—	—	26
OSST	Rank correlation	—	—	27
OSST	Heidke skill score	—	—	8
OSST	MSESS	—	—	4
OSST <sub>r</sub>	Correlation	—	—	33
OSST <sub>r</sub>	Rank correlation	—	—	<b>39</b>
OSST <sub>r</sub>	Heidke skill score	—	—	11
OSST <sub>r</sub>	MSESS	—	—	10

TABLE 8. Ranked probability skill scores ( $\times 10^2$ ) for NTC, per basin, by lead time and SST forecast scenario. Here, “S” denotes simulations, whose lead times are negative. “Pers” denotes 1-yr weak probabilistic persistence (see text), with a lead potentially longer than 4 months, but shown in the column of the longest lead. Statistically significant skills are shown in boldface.

SST type Lead	Eastern Pacific					Western Pacific					Atlantic					
	3	2	1	0	S	3	2	1	0	S	4	3	2	1	0	S
FSST	-12	8	7	-60	—	-76	-52	-63	-34	—	-51	-8	3	-13	-61	—
HSST <sub>p</sub>	—	—	-92	-80	—	—	—	-35	-36	—	—	—	-107	-70	-68	—
OSST	—	—	—	—	-18	—	—	—	—	-5	—	—	—	—	—	3
OSST <sub>r</sub>	—	—	—	—	-24	—	—	—	—	3	—	—	—	—	—	-21
Issued	13	13	9	8	—	8	0	1	15	—	-5	6	11	15	-4	—
Pers	-2	—	—	—	—	-2	—	—	—	—	-2	—	—	—	—	—
SST type Lead	Australia					South Pacific										
	4	3	2	1	0	S	3	2	1	0	S					
FSST	-24	-36	-15	-74	-68	—	1	-27	-38	-27	—					
HSST <sub>p</sub>	—	—	-40	-34	-51	—	—	—	-28	-42	—					
OSST	—	—	—	—	—	-32	—	—	—	—	-93					
OSST <sub>r</sub>	—	—	—	—	—	-21	—	—	—	—	-88					
Issued	5	5	15	2	5	—	10	15	10	3	—					
Pers	1	—	—	—	—	—	-1	—	—	—	—					

measures (here, RPSS) than the higher-amplitude probability shifts from climatology typically produced by today’s AGCMs without proper statistical calibration.

The probability forecasts actually issued by IRI begin with the “raw” AGCM probabilities, modified to what the forecasters judge to have better probabilistic reliability. This nearly universally involves damping the amplitude of the model’s deviation from climatological probabilities. A typical adjustment might be to modify the model’s predicted probabilities of 5%, 10%, and 85% to 20%, 30%, and 50% for the below, near, and above normal categories, respectively. A less common adjustment is that of “rounding out” a bimodal probability forecast such as 35%, 5%, and 60% to a more Gaussian distribution such as 25%, 30%, and 45%.<sup>7</sup>

<sup>7</sup> It is expected that NTC and ACE are composed of enough independent individual TC events over the season that the central limit theorem would result in a smoother, and usually unimodal, forecast probability distribution.

Part of the reason for sharply bimodal distributions is assumed to be the limited (24 member) ensemble size. A still less common case for modification, and one that does not always improve the forecast quality, is that of the forecasters’ judgment against the model forecasts, believing there is a model bias. Such doubt can pertain also to the SST forecast used to force the AGCM.

Tables 8 and 9 indicate that the actually issued forecasts have better probabilistic reliability than the forecasts of the model output. Likelihood skill scores (not shown), and especially RPSS, are mainly positive for the issued forecasts, although modest in magnitude. This implies that the probability forecasts of the AGCM are potentially useful, once they are calibrated to correct for overconfidence or an implausible distribution shape. Such calibration could be done objectively, based on the longer hindcast history, rather than subjectively by the forecasters as done to first order here.

Figure 2 shows the approximately 6-yr record of AGCM ensemble forecasts of NTC and ACE at all forecast lead times for each of the ocean basins. The

TABLE 9. As in Table 8 but for ACE.

SST type Lead	Eastern Pacific					Western Pacific					Atlantic					
	3	2	1	0	S	3	2	1	0	S	4	3	2	1	0	S
FSST	-30	9	9	-3	—	26	-8	-1	-11	—	-45	-6	-16	-6	-31	—
HSST <sub>p</sub>	—	—	-35	-24	—	—	—	-40	-9	—	—	—	-33	2	-6	—
OSST	—	—	—	—	-26	—	—	—	—	<b>6</b>	—	—	—	—	—	<b>11</b>
OSST <sub>r</sub>	—	—	—	—	-7	—	—	—	—	<b>13</b>	—	—	—	—	—	<b>16</b>
Issued	2	0	6	10	—	15	5	5	4	—	-7	10	12	15	6	—
Pers	5	—	—	—	—	1	—	—	—	—	2	—	—	—	—	—

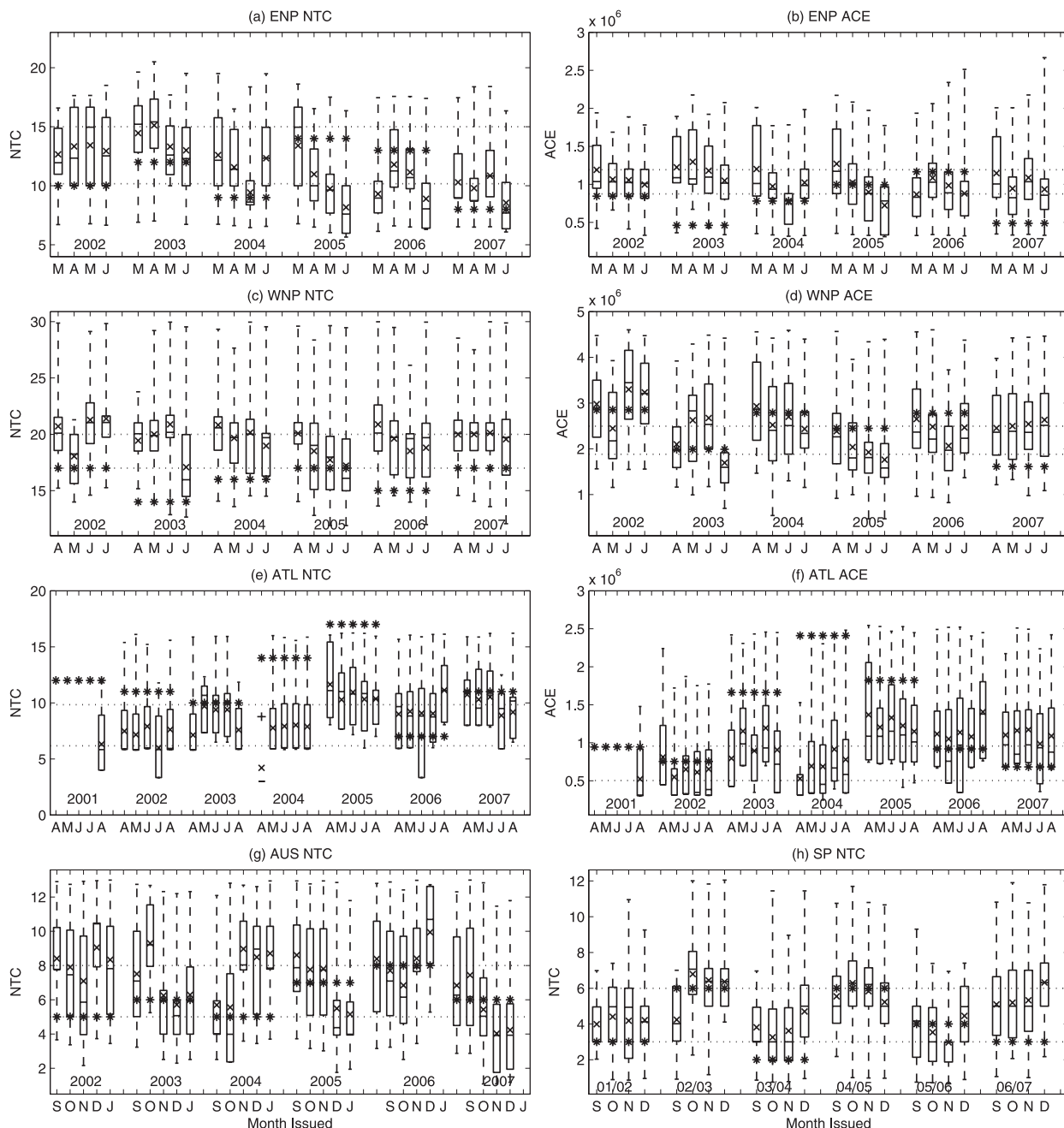


FIG. 2. Model (raw) forecasts (box-and-whiskers plots) and observations (asterisks) of NTCs and ACE for all basins and leads. The cross inside the box shows the ensemble mean, and the horizontal line shows the median. Also shown by dotted horizontal lines are the boundaries between the tercile categories. (a)–(f) The Northern Hemisphere basins, with (left) NTC and (right) ACE for (a), (b) ENP; (c), (d) ATL; and (e), (f) WNP. The NTCs in the Southern Hemisphere basins: (g) AUS and (h) SP.

vertical boxes show the interquartile range among the ensemble members, and the vertical dashed lines (“whiskers”) extend to the ensemble member forecasts outside of that range. The asterisk indicates the observation value. Favorable and unfavorable forecast outcomes can be identified, such as, respectively, the ACE forecasts for

the western North Pacific for 2002 and the ACE forecasts for the North Atlantic for 2004.

Figure 3 shows the same forecasts, except probabilistically for each of the tercile-based categories, both for the AGCM’s forecasts (crisscross symbols) and for the subjectively modified publicly issued forecasts (circle

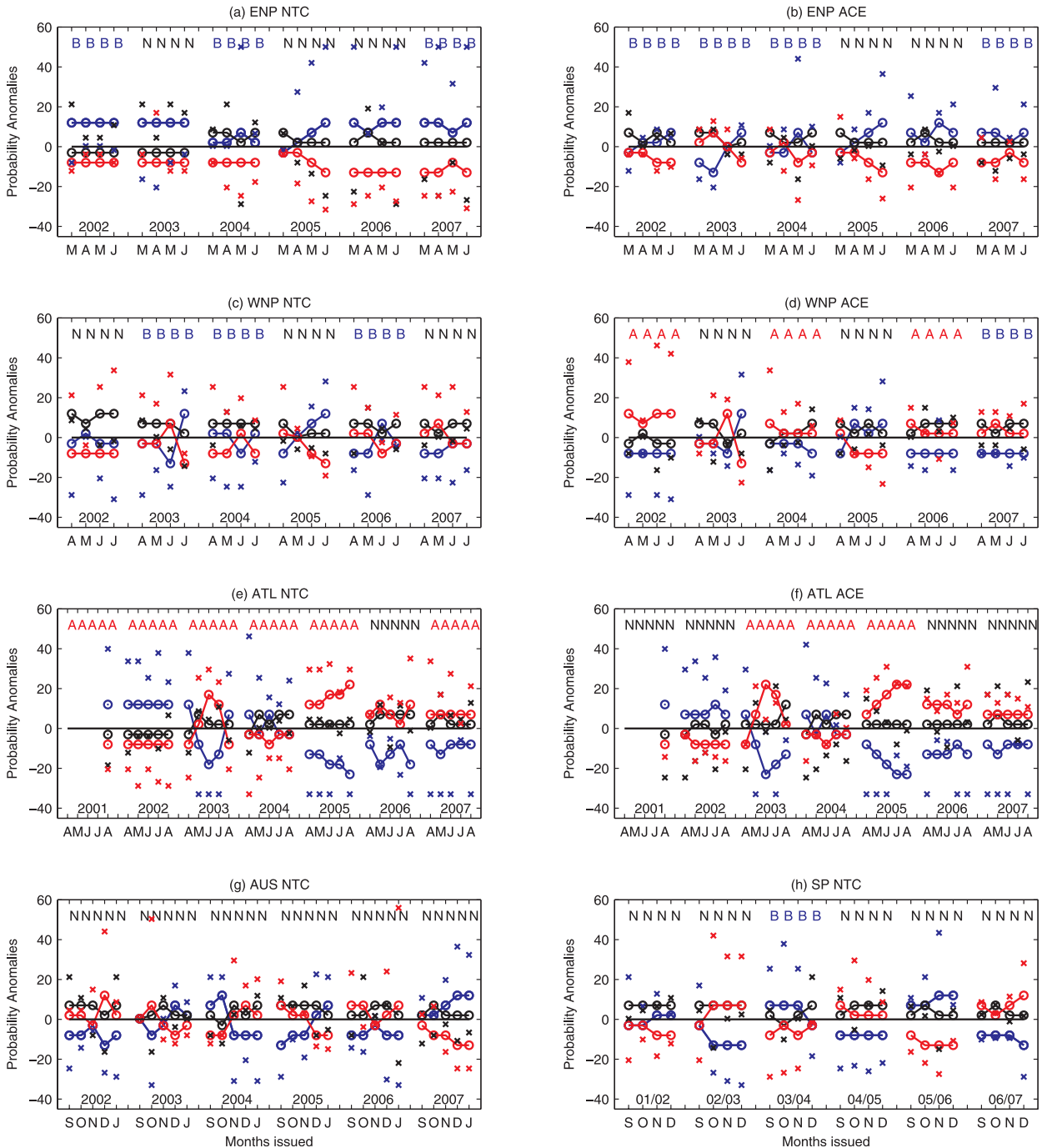


FIG. 3. Issued (circles) and modeled (crisscrosses) probability anomalies (difference of probability from 33.3% climatological probability values,  $\times 100$ ) for all leads and years in each basin. The above (below) normal category probability anomalies are given in red (blue), and the near-normal anomalies in black. The observed category is shown near the top: below normal (B), near normal (N), and above normal (A). (a)–(h) Basins are arranged as in Fig. 2.

symbols connected by lines). The AGCM’s probability forecasts often deviate by large amounts from climatology, while the issued forecasts remain closer to climatology. Figure 4 shows the RPSSs of these probability

forecasts in the same format. The AGCM’s probability forecasts result in highly variable skill (including both strongly negative and positive cases), leading to a somewhat negative overall skill. The issued forecasts,

while never reaching positive magnitudes as great as those of some of the AGCM forecasts, also avoid negative overall skill levels of more than small magnitude.<sup>8</sup> Hence, the humanly modified TC forecasts have a higher average probabilistic skill level using RPSS.

The “overconfidence” of the AGCM forecasts is shown in more concrete terms in a reliability (or attributes) diagram (Hsu and Murphy 1986) in Fig. 5. Here, the correspondence of the forecast probabilities with the observed relative frequency of occurrence is shown for the above normal and below normal categories. When the forecast probabilities closely match the observed relative frequencies, as would be desired, the lines approximate the dotted 45° line. Figures 5a and 5b show, for the 6-yr period of forecasts, the reliabilities for the issued forecasts and for the AGCM’s forecasts prior to subjective modification, respectively. Despite the “jumpy” lines due to the small sample sizes, the lines for the issued forecasts are seen to have slopes roughly resembling the 45° line, indicating favorable reliability, while the lines for the AGCM’s forecasts have a less obvious upward slope. The AGCM’s forecast probabilities for the above or below normal categories of TC activity deviate from the climatological probabilities of 1/3 by much greater amounts than do their corresponding observed relative frequencies (see bottom inset in the panels of Fig. 5), resulting in low probabilistic forecast skill. The issued forecasts’ deviations from climatological probabilities are limited by the forecasters according to the perceived level of uncertainty, and within the restricted probability ranges an approximate correspondence to the observed relative frequencies is achieved. The more reliable issued forecasts carry appropriately limited utility as represented by the lack of forecast sharpness—that is, that the forecast probabilities rarely deviate appreciably from climatology, and from one another.

The bottom panels in Fig. 5 show reliabilities for the longer historical period of AGCM hindcasts using prescribed observed SSTs (OSST<sub>p</sub>; Fig. 5c) and the persisted SST anomaly (HSST<sub>p</sub>; Fig. 5d). Here, the lines are smoother due to the larger sample sizes. Both diagrams show forecasts having some informational value, as the lines have positive slopes, but the slopes are considerably

shallower than the 45° line, indicating forecast overconfidence. The slopes for forecasts using observed SSTs are slightly steeper than those for forecasts using the persisted SST anomaly, as would be expected with the higher skill realized in forecasts forced by the actually observed lower boundary conditions.

That the TC activity forecasts of the AGCM have mainly positive correlation skill is consistent with their positive slopes in Figs. 5b–d. Additionally, their mainly negative RPSS (Tables 8 and 9) is expected when the positive slopes in the reliability diagram (Fig. 5) are shallower than one-half of the ideal 45° slope = 1 line (i.e., slope < 0.5) because then the forecasts’ potential information value is more than offset by the miscalibration of the forecast probabilities (Hsu and Murphy 1986; Mason 2004). This is consistent with the deterministic TC forecasts having positive correlation skill but negative MESSs using climatology as the reference forecast, due to forecast anomalies that are stronger than warranted for the expected skill level.

The skills of the real-time probabilistic forecasts over the approximately 6-yr period are summarized in full aggregation (over basins and TC variables) in Table 10 using the RPSS, likelihood [based on the concept of maximum likelihood estimation; Aldrich (1997)], and ROC (Mason 1982) verification measures. The comparisons between the objective AGCM forecast output and the actually issued forecasts again underscore the need for the calibration of AGCM forecasts that greatly underestimates the real-world forecast uncertainty. The AGCM’s nontrivially positive scaled ROC areas for both above and below normal observed outcomes reveal their ability to provide useful information, as the ROC lacks sensitivity to calibration in a manner analogous to correlation for deterministic, continuous forecasts. In this particular set of forecasts, greater capability to discriminate the above from the below normal TC activity is suggested by the ROC skill.

### c. A favorable and an unfavorable real-time forecast

Identification of “favorable” or “unfavorable” forecasts, while straightforward when considered deterministically, is less clear when comparing an observed outcome with its corresponding probability forecast. Probabilistic forecasts implicitly contain expressions of uncertainty. The position of an observed outcome within the forecast distribution is expected to vary across cases, and many cases are required to confirm that this variation is well described by the forecast probability distributions. When an observation lies on a tail of the forecast distribution, it is impossible to determine whether this represents an unfavorable forecast or is an expected

<sup>8</sup> Because the RPSS is computed as a sum of squares of cumulative (over tercile categories) differences between forecast and observed probabilities, the lower limit of RPSS (−3.5) is farther below zero than the upper limit (+1.0) is above zero. Thus, high probabilities forecasted for an incorrect category outweigh high probabilities forecasted for the correct category, and “overconfident” forecasts result in severe penalties even when the forecasts have some positive level of informational value.

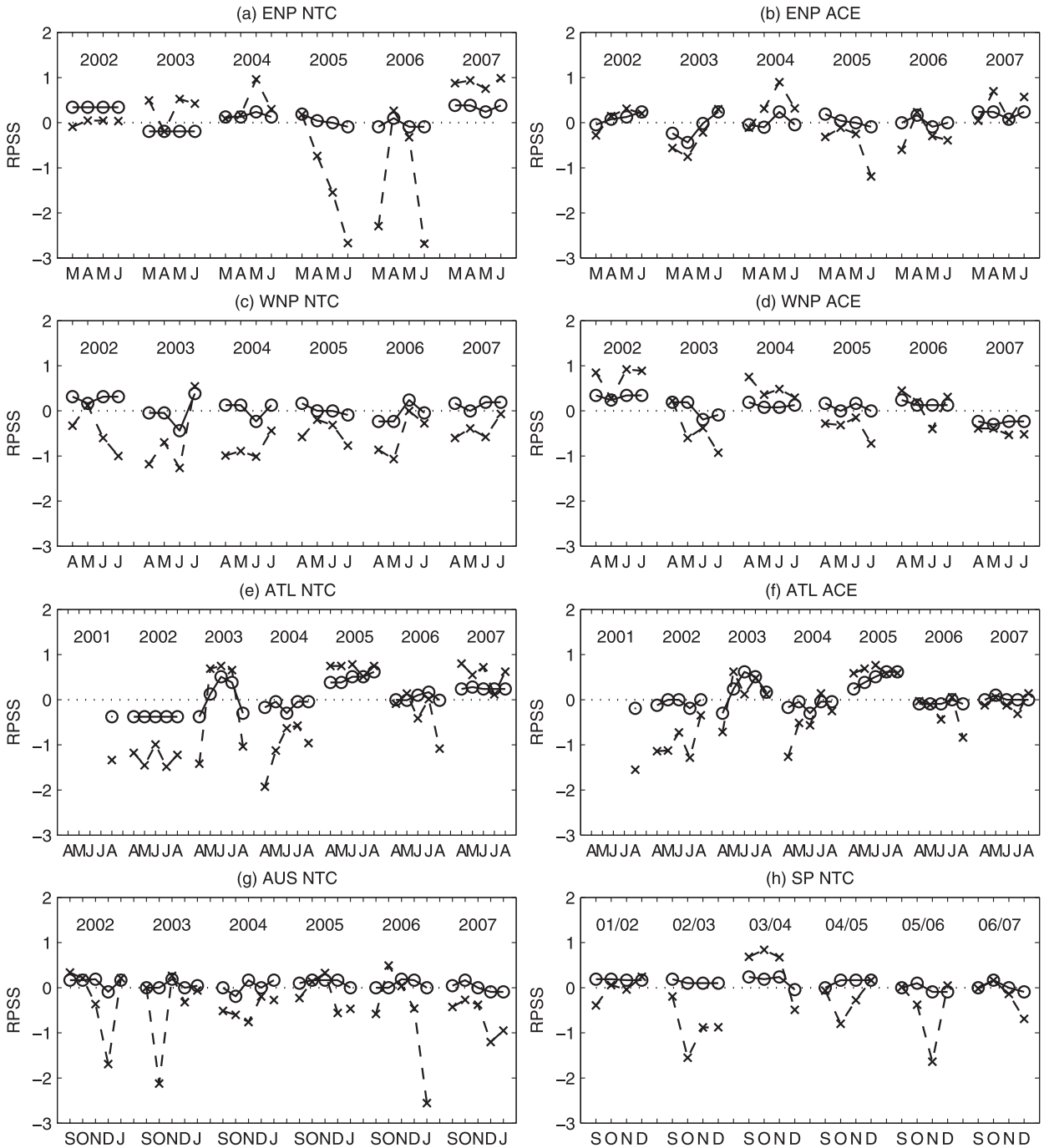


FIG. 4. RPSSs for the issued (circles) and modeled (crisscrosses) forecasts for all leads and years in each basin. (a)–(h) Basins are arranged as in Fig. 2.

rare case, without examining a large set of forecasts. The forecast distribution may be fully appropriate given the known forcing signals (Barnston et al. 2005). Here, we identify favorable and unfavorable cases in terms of the difference between the deterministic forecast (the model ensemble mean, which usually also approximates

the central tendency of the forecast probability distribution) and the corresponding observation.

A critical aspect of the SST forcing to be forecast is the ENSO state during the peak season. Figure 6 shows the IRI's forecasts of the seasonal Niño-3.4 index at 2-month lead time (e.g., a forecast for ASO SST issued

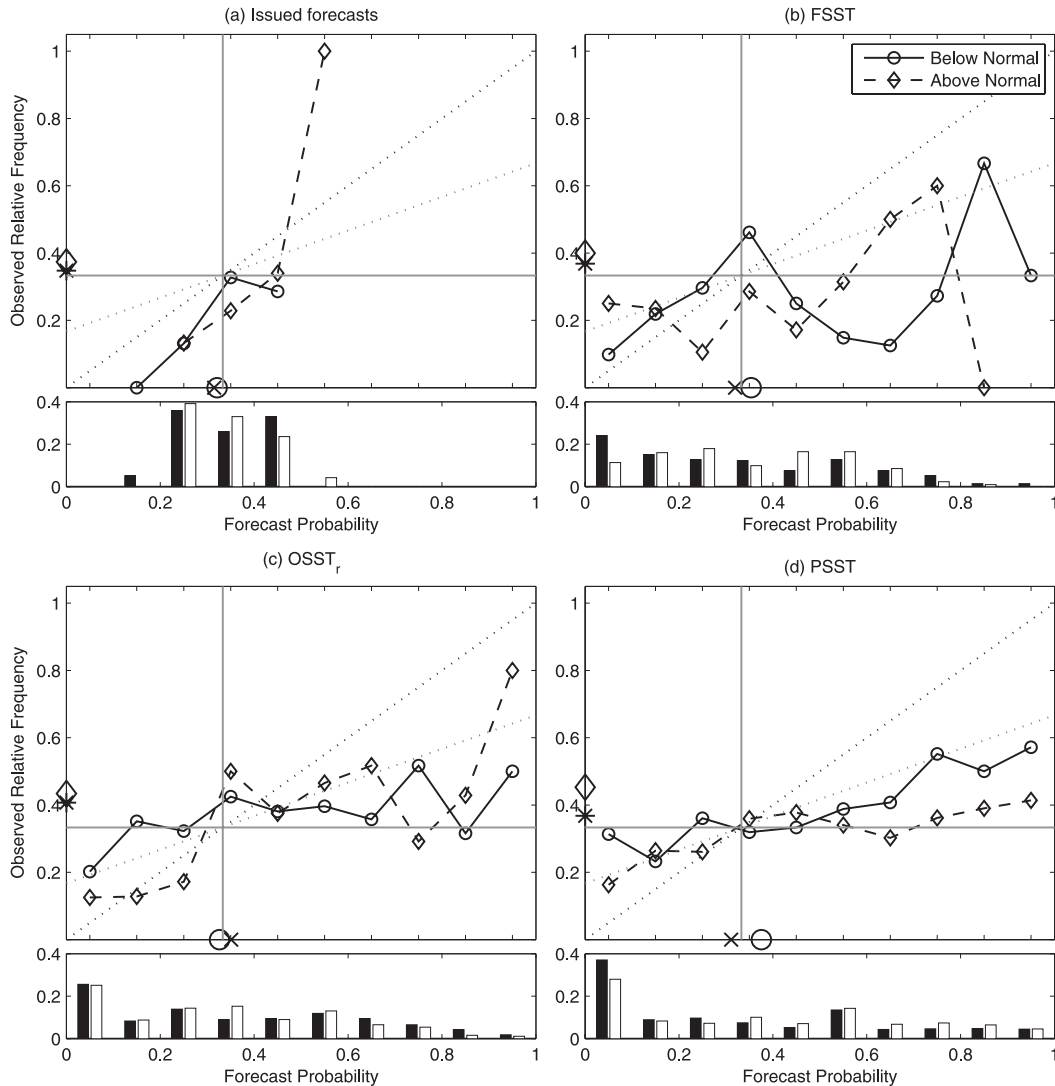


FIG. 5. Reliability diagrams for above (circles) and below (diamonds) normal categories: (a) issued forecasts, (b) FSST, (c) OSST<sub>r</sub>, and (d) HSS<sub>T<sub>p</sub></sub>. The histograms [below normal (black bars); above normal (white bars)] below each plot show the percentage frequency with which each category of probability was forecast. The circle (criss-cross) indicates the overall mean of the forecast probabilities for the above (below) normal categories, and the diamond (asterisk) indicates the overall mean but for observed relative frequencies. The vertical and horizontal lines indicate the climatologically expected forecast probability and observed relative frequency, respectively. The ideal reliability is shown by the dotted 45° diagonal line. The dotted line with the shallower slope is the slope above which positive skill would be realized in RPSS, and below which (for positive slope) RPSS would be negative but correlation skill for corresponding deterministic forecasts would usually be positive, suggesting informational value.

in mid-June, with observed data through May) during the period of issued TC forecasts, with the corresponding observed seasonal SST. A moderate El Niño (EN) occurred during 2002–03, with weak ENs in 2004–05 and late 2006. A weak, brief La Niña (LN) condition was observed in late 2005 and early 2006, and a stronger LN developed during mid-2007. The average of the observed Niño-3.4 SST anomalies over the approximately 5-yr period is 0.45, compared with an average 2-month

lead forecast anomaly of 0.37, indicating a small forecast bias. The uncentered correlation coefficient for the period is in the range of 0.70–0.79 for forecasts for the Northern Hemisphere peak seasons, and in the range of 0.80–0.89 for forecasts for the Southern Hemisphere peak season, suggesting somewhat skillful forecasts of tropical Pacific SST fluctuations for these peak TC seasons.

A favorable forecast for ACE in the western North Pacific took place in 2002. Figure 2d shows that the

TABLE 10. Comparison of skill ( $\times 10^2$ ) between real-time probability forecasts for TC activity based directly on the AGCM (FSST), and those issued by IRI forecasters following subjective modification. Skills for forecasts for all five ocean basins for both NTC and ACE over the approximately 6-yr period are aggregated using RPSS, likelihood, and scaled ROC area for forecasts when above normal (AN) or below normal (BN) TC activity was observed. The likelihood score is computed as the  $n$ th root of the product of the probabilities given to the tercile category that was indeed observed, spanning temporally through all  $n$  forecasts, and then linearly scaled as zero for perpetual climatology forecasts and unity for 100% probability always given to the observed category. AGCM probabilities of zero for the correct category are set to 0.01 to avert degeneracy in the likelihood score. The ROC score is scaled as  $2 \times (\text{area} - 0.5)$  for increased comparability to the other skill measures.

No. of forecasts		48	48	48	50
Lead		3	2	1	0
Skill score	Type				
RPSS	FSST	-24	-12	-22	-37
RPSS	Issued	9	9	8	6
Likelihood	FSST	2	-1	-1	-5
Likelihood	Issued	4	5	4	4
ROC (AN)	FSST	37	27	28	10
ROC (AN)	Issued	50	42	59	47
ROC (BN)	FSST	15	14	21	15
ROC (BN)	Issued	24	25	21	15

observation was in the above-normal category, and that the AGCM forecasts were not far from this number for the four lead times. For ACE in the western North Pacific, the ENSO condition is key, with EN (LN) associated with higher (lower) ACE. Between April and June of 2002 it became clear that an EN was developing, although the SST predictions contained a weaker EN than was observed (Fig. 7). Nonetheless, the SST predictions contained ENSO-related anomaly patterns of sufficient amplitude to force an above normal ACE prediction that verified positively. The favorable AGCM forecasts are shown probabilistically in Fig. 3d, with a positive RPSS verification shown in Fig. 4d.

An unfavorable forecast outcome occurred for ASO 2004, when the North Atlantic ACE was observed to be  $2.41 \times 10^6 \text{ kt}^2$ , the highest on record after 1970 for this season, but the AGCM forecasts from all five lead times were for between  $0.5$  and  $1.0 \times 10^6 \text{ kt}^2$ , only in the near-normal category. A weak EN developed just prior to the peak season, which, while somewhat underpredicted, was present in the SST forecasts. But despite weak EN conditions during the 2004 peak season, NTC and especially ACE were well above normal (Figs. 2e and 2f). A feature of the EN that likely weakened its inhibiting effect on Atlantic TC development was its manifestation mainly in the central part of the tropical Pacific, rather than in the Niño-3 region that appears more

critical. Coupling of the warmed SSTs to the overlying atmosphere was also modest in ASO. However, aspects of the SST that were not well predicted were those that mattered more critically in this case: the North and tropical Atlantic SSTs (Goldenberg et al. 2001; Vimont and Kossin 2007), including the main development region (Goldenberg and Shapiro 1996). These regions developed markedly stronger positive anomalies than had been observed in April and May or forecast for the forthcoming peak season months, and are believed to have been a major cause of the high 2004 Atlantic TC activity level.

Both examples described above highlight the importance of the quality of the SST forecast for the peak TC season in the relevant tropical and subtropical ocean regions. ENSO-related Pacific SST is known to have some predictability, but there is room for improvement in capturing it, and the seasonal prediction of SST in the tropical Atlantic is a yet more serious challenge.

#### 4. Comparison with simple statistical predictions

One reasonably might ask whether the skill levels of the AGCM simulations and predictions are obtainable using statistical models derived purely from the historically observed TC data and the immediately preceding environmental data such as sea level pressure or SST conditions. How much does the dynamical approach to TC prediction offer that is not obtainable using empirical approaches? Here, we explore this for deterministic skill, using observed predictors in multiple regressions. To minimize the artificial skill associated with “fishing” for accidentally skillful predictors, four restrictions are imposed: a maximum of two predictors is used for each basin, the same predictors are used for NTC as for ACE for each basin, all predictors are SSTs averaged over rectangular index regions,<sup>9</sup> and all predictors must have a plausible physical relevance to the TC activity. “Leave out one” cross validation is applied to assess the expected real-time predictive skill of the statistical models. We use mainly SST because of the well-documented influence of SST anomaly patterns, including in particular the state and direction of the evolution of ENSO, on the inter-annual variability of TC activity in most ocean basins. Statistical predictions are made at a lead time of 1 month (e.g., June SST predicting the Atlantic peak season of ASO). A similar “prediction” is made for the simulation of TC activity using predictors simultaneous with the center month of the peak TC season.<sup>10</sup> The simulation

<sup>9</sup> The only exception to this is Australia, in which sea level pressure at Darwin is used.

<sup>10</sup> The second month is used for both 3- and 4-month peak seasons.

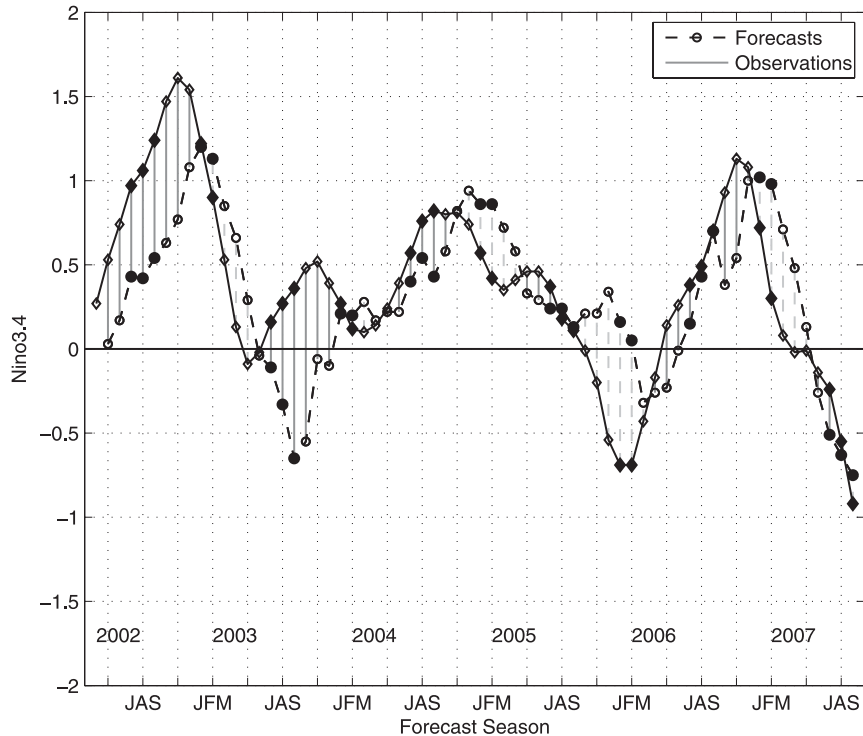


FIG. 6. Two-month lead IRI forecasts (circles) and observations (diamonds) of Niño-3.4 SST index anomalies. Seasons for which TCs are forecast are shown with filled symbols. The differences between the forecasts and observations are shown by solid (dotted) gray vertical lines when the observation was warmer (colder) than the forecast.

predictors are usually the same as those used for the 1-month lead prediction.

Selection of the predictor SST indices is based on previous studies and on examination of the geographical distribution of the interannual correlation between SST and the TC variables using 1970–2005 data. For example, Fig. 7 shows the correlation field for SST in June versus Atlantic NTC during the ASO peak season, indicating the well-known inverse relationship with warm ENSO, and positive association with SSTs in the North Atlantic, associated with the Atlantic meridional

mode (Vimont and Kossin 2007) and the Atlantic multidecadal oscillation (Goldenberg et al. 2001). When September SST is used, simultaneous with the Atlantic TC activity, these same two key regions remain important, but even stronger correlations appear for SST in the main development region (Goldenberg and Shapiro 1996).

Table 11 identifies the predictors used for each ocean basin for 1-month lead forecasts and for simultaneous simulations. In the cases of the Atlantic and western North Pacific forecasts, the first predictor contains both

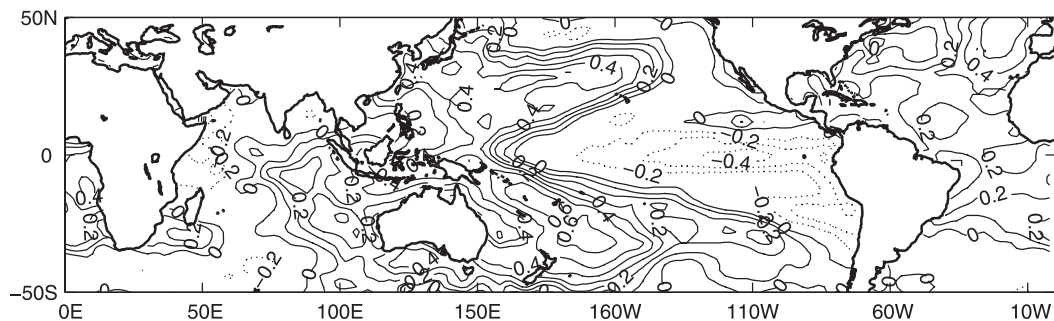


FIG. 7. Correlations between June SST and ASO Atlantic NTC, 1970–2006. Contour interval is 0.1. Zero and positive (negative) contours represented by solid (dotted) lines. The  $-0.1$  contour is not shown.

TABLE 11. Predictors for statistical tropical cyclone 1-month lead forecasts and simulations. The month of the two SST predictors is indicated in parentheses. MDR is the main development region of the Atlantic (10°–20°N, 82°–20°W). AMO is the Atlantic multidecadal oscillation region (here, 40°–50°N, 75°W–0°). Darwin is located in northern Australia (12.4°S, 130.9°E).

Basin	Season	1-month lead forecasts	Simultaneous simulations
ENP	JJAS	1) 20°–30°N, 110°–130°W (Apr) (north subtropical version of Niño-3) 2) 20°–30°N, 180°–140°E (Apr) (north part of outer ENSO horseshoe)	1) Same as forecasts, but for Jul 2) Same as forecasts, but for Jul
WNP	JASO	1) Niño-3.4 (May) + Niño-3.4 change from Feb–Mar to May 2) 20°–30°N, 110°–150°W (May) (northeast subtropical Pacific)	1) Niño-3.4 (Aug) 2) Same as forecasts, but for Aug
ATL	ASO	1) Niño-3 (Jun) + Niño-3 change from Mar–Apr to Jun 2) MDR (weighted 1/3) + AMO (weighted 2/3)	1) Niño-3 (Sep) 2) MDR (Sep)
AUS	JFM	1) Niño-3.4 to Niño-4 (Nov) (5°N–5°S, 120°W–160°E) 2) Darwin SLP (SON)	1) Same as forecasts, but for Feb 2) Same as forecasts, but for JFM
SP	DJFM	1) Niño-3 to Niño-3.4 (Oct) (5°N–5°S, 90°–170°W) 2) 25°–35°S, 170°W–170°E (Oct) (south part of outer ENSO horseshoe)	1) Same as forecasts, but for Jan 2) Same as forecasts, but for Jan

a recent SST level and a recent time derivative for the same region, to capture the ENSO status and direction of evolution. Many of the statistical predictors are ENSO related. The Niño-3 SST in the east-central tropical Pacific is found to be more relevant to Atlantic TC activity (Gray et al. 1993) than the location more central to ENSO itself [i.e., Niño-3.4; Barnston et al. (1997)]. Niño-3.4 is used for western North Pacific TC activity, with the second predictor being SST in the subtropical northeastern Pacific associated with the North Pacific atmospheric circulation pattern that is found to be linked with the TC activity (Barnston and Livezey 1987; Chan et al. 1998). For northeast Pacific TC activity, the SST regions highlight an ENSO-related east–west dipole at northern subtropical latitudes, while for the South Pacific and Australia the regional TC predictions, also ENSO governed, are tailored to their Southern Hemisphere locations.

Table 12 indicates the strengths of the relationship between each of the predictors and the predictand, the

predictors’ correlations with one another, and the resulting multiple correlation coefficients both within the model development sample and upon using cross validation. The latter skill is considered to be a less biased skill estimate with which the AGCM-based skill (shown in the subsequent column in Table 12) can be compared. Results for this comparison are mixed. The dynamical forecasts and simulations are slightly more skillful for South Pacific NTCs, as well as in the eastern North Pacific basin in most cases. The statistical model produces higher skill levels in most cases in the Atlantic and western North Pacific for forecasts and for simulations in some cases. Statistical tests indicate that none of the dynamical–statistical skill differences are significant for the 36-case sample. Considering this, and the alternation of skill rank between the approaches over the basins, there is no clear suggestion that one approach is generally superior to the other. That the dynamical approach tended to yield higher skill levels in the South Pacific, and results that were no lower than the

TABLE 12. Diagnostics for the two-predictor statistical TC forecasts and simulations, 1970–2005. The pr1 (pr2) columns show the correlations ( $\times 10^2$ ) between the first (second) predictor and the observed TC variable (NTC or ACE) in two-predictor multiple regression predictions at 1-month lead time, per ocean basin. Sim1 (sim2) columns are similar but for simulations in which the SST forcing is prescribed as that observed during the peak TC season. Correlation between the two predictors is indicated in the “1vs2” column. The next two columns show the full sample and the one-year-out cross-validated multiple regression coefficients, the latter to be regarded as the skill estimate for real-time forecasts for comparison with dynamical (AGCM based) skill levels, shown in the subsequent column. Dynamical predictive skill comes from the  $HSST_p$  at 1-month lead, and simulation skill from  $OSST_p$ . Statistically significant skills are shown in boldface.

		Forecasts						Simulations					
		pr1	pr2	1vs2	R	Rcv	Dyn	sim1	sim2	1vs2	R	Rcv	Dyn
ENP	NTC	29	–30	–33	<b>37</b>	14	32	32	–28	–18	<b>39</b>	20	<b>37</b>
ENP	ACE	<b>40</b>	–33	–33	<b>45</b>	31	20	<b>47</b>	–27	–18	<b>50</b>	<b>35</b>	<b>45</b>
WNP	NTC	19	<b>41</b>	<b>47</b>	<b>41</b>	23	11	23	14	<b>54</b>	23	–6	28
WNP	ACE	<b>60</b>	<b>53</b>	<b>47</b>	<b>66</b>	<b>58</b>	6	<b>67</b>	<b>56</b>	<b>54</b>	<b>71</b>	<b>66</b>	<b>33</b>
ATL	NTC	–38	<b>49</b>	–19	<b>57</b>	<b>46</b>	<b>33</b>	–45	<b>58</b>	–2	<b>72</b>	<b>66</b>	<b>55</b>
ATL	ACE	–27	<b>60</b>	–19	<b>62</b>	<b>55</b>	<b>43</b>	–34	<b>47</b>	–2	<b>57</b>	<b>50</b>	<b>60</b>
AUS	NTC	–40	–30	<b>83</b>	<b>41</b>	23	26	–44	–53	<b>81</b>	<b>53</b>	<b>42</b>	<b>40</b>
SP	NTC	<b>38</b>	–28	–55	<b>39</b>	20	<b>42</b>	<b>42</b>	–37	–48	<b>46</b>	30	<b>42</b>

statistical method in the Australian region, could be related to the comparatively lower quality of the SST predictor data south of the equator, as well as NTC data in the Southern Hemisphere, particularly in the 1970s. It is possible that less accurate SST data would degrade the statistical forecasts more than the AGCM forecasts forced by the SST because the SST indices used in the statistical forecasts represent relatively smaller regions than the aggregate of the SST regions influencing the behavior of the AGCM. The larger areas of SST influencing the model may allow the opportunity for opposing error impacts, leading to smaller net impacts.

Some notable features of this methodological comparison are (i) the statistical models used here were restricted to be fairly simple, and may not be near optimum; (ii) despite the use of cross validation, some “fishing” may still have occurred in selecting the predictor SST indices, and there may be some artificial skill; and (iii) the one-year-out cross validation design has a negative skill bias in truly low predictability situations (Barnston and van den Dool 1993). Such caveats of opposing implications suggest that the skill comparisons should be considered as rough estimates, intended to detect obvious skill differences—and such differences are not revealed here. One might expect that much of the skill of a near-perfect dynamical model would be realizable by a sophisticated (e.g., containing nonlinearities) statistical model if accurate observed data were available, since the observations should occur because of, and be consistent with, the dynamics of the ocean–atmosphere system with noise added. Seasonal climate has been shown to be statistically modeled fairly well using only linear relationships (Peng et al. 2000). However, linearity may compromise statistical skill in forecasting seasonal phenomena such as TC activity, with its highly nonlinear hydrodynamics in individual storms that may not reduce to linear behavior even upon aggregating over a TC peak season.

## 5. Conclusions

The IRI has been issuing experimental TC activity forecasts for several ocean basins since early 2003. The forecasts are based on TC-like features detected and tracked in a climate model at low horizontal resolution. The model is forced at its lower boundary by SSTs that are predicted first, using several other dynamical and statistical models. The skill of the model’s TC predictions using historical observed SSTs are discussed as references against which skill levels using several types of predicted SSTs (including persisted SST anomalies) are compared. The skill of the raw model output is also compared with that of subjective probabilistic forecasts

actually developed since mid-2001, where the subjective forecasts attempt to correct the “overconfident” probabilistic forecasts from the AGCM. The skill levels of the AGCM-based forecasts are also compared with those from simple statistical forecasts based on observed SSTs preceding the period being forecast.

Results show that low-resolution uncoupled climate models deliver statistically significant, but fairly modest, skill in predicting the interannual variability of TC activity. The levels of correlation skill are comparable to the levels obtained with simple empirical forecast models—here, models employing two-predictor multiple regression using preceding area-average SST anomalies and their recent time derivative. In ocean basins where observed SST predictor data are of questionable quality, statistical prediction is less effective. Despite that this same SST is used as the boundary forcing for the climate model, the dynamical predictions tend to slightly outperform the statistical predictions in this circumstance.

In a two-tiered dynamical prediction system such as that used in this study, the effect of imperfect SST prediction is noticeable in the skill levels of TC activity compared with skill levels when the model is forced with historically observed SSTs.

Similar to climate forecasts made by AGCMs, the probabilistic reliability of the AGCM’s forecasts for TC activity forecasts is not favorable in that the model ensemble forecasts usually deviate too strongly from the climatological distribution, due to too narrow an ensemble spread or to too large a shift in the ensemble mean from climatology. This overconfidence of the AGCM forecasts is partly due to their being based on specific representations of the physics, through parameterizations, and their own hindcast performances are not taken into account in forming ensemble forecasts. Upon subjective human intervention, the forecasts are made more probabilistically conservative and reliability is improved, leading to higher probabilistic verification scores than for the uncalibrated AGCM forecasts.

The potential skill levels (i.e., discriminating information, but needing calibration) seen in the approximately 6-yr-period real-time AGCM TC predictions are tentative and nonrobust due to the small sample size. However, these skill levels are not inconsistent with those of longer-period AGCM-based hindcasts forced by SST anomalies persisted from the previous month, as seen in comparing the correlation skills for leads of 0 and 1 of FSST and HSST<sub>p</sub> in Tables 4 and 5. Thus, assuming a calibration step, real-time skill levels are expected to be fairly modest but contain useful informational value (e.g., most short-lead HSST<sub>p</sub> correlations are 0.2–0.5) with the highest skill potential appearing in the Atlantic, eastern Pacific, and South Pacific, and

somewhat lower levels of skill potential found in the western Pacific and Australia regions. Skill levels are generally seen to decrease slowly with increasing lead time, such that forecasts issued several months prior to the peak season onset are also expected to have some informational value.

We plan to examine the skill of other models in hopes of being able to add more information, and hopefully skill, to our seasonal TC forecasts. The problem of overconfidence in AGCMs is relieved to some extent by the use of multimodel ensembles (Kharin and Zwiers 2002; Vitart 2006; Tippett et al. 2007): adding additional models should help restrain the probabilistic amplitude exhibited by a single model. The merging of TC forecasts made by AGCMs and by statistical methods may also prove beneficial. Another possibility that could be explored in the future is to combine dynamical forecasts using the direct method (tracking of model storms) and the indirect method (using only the large-scale fields of the model).

Issues not examined here are the role of AGCM spatial resolution in governing predictive skill, and the impact of using a fully coupled dynamical system rather than a two-tiered system as is employed here. Although the prospects for the future improvement of dynamical TC prediction are uncertain, it appears likely that additional improvements in dynamical systems will make possible better TC predictions. As is the case for dynamical approaches to ENSO and near-surface climate prediction, future improvements will depend on a better understanding of the underlying physics, more direct physical representation through higher spatial resolution, and substantial increases in computer capacity. Hence, improved TC prediction should be a natural by-product of the improved prediction of ENSO, global tropical SST, and climate across various spatial scales.

*Acknowledgments.* This work was supported by a grant/cooperative agreement from the National Oceanic and Atmospheric Administration (Grant NA05OAR431 1004). The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its subagencies.

## APPENDIX

### Acronyms and Their Definitions

ACE	accumulated cyclone energy
AGCM	atmospheric general circulation model
AMIP	Atmospheric Model Intercomparison Project (AGCM is forced using observed SST)
CCA	canonical correlation analysis

CFS	Climate Forecast System (global coupled model of NOAA/NCEP)
DJFM	December–January–February–March (and similarly for other multimonth periods)
ECHAM4.5	ECMWF–Hamburg, Germany, AGCM, version 4.5
ECMWF	European Centre for Medium-Range Weather Forecasts
EN	El Niño
ENSO	El Niño–Southern Oscillation
EUROSIP	European Seasonal to Interannual Prediction superensemble
FSST	forecasted SST used for real-time AGCM forecasts, 2001 onward
FSST <sub>e</sub>	FSST using evolving (predicted) SST anomalies
FSST <sub>p</sub>	using persisted SST anomalies observed from previous month
FSST	hindcasted SST, covering a long past history, using persisted anomalies observed from the previous month
HSST <sub>p</sub>	International Research Institute for Climate and Society
IRI	Lamont-Doherty Earth Observatory (a campus of Columbia University)
LDEO	La Niña
LN	mean squared error skill score
MSESS	National Centers for Environmental Prediction
NCEP	National Oceanic and Atmospheric Administration
NOAA	number of tropical cyclones
NTC	observed SST used for a long past history of AGCM hindcasts, starting in 1950
OSST	OSST for a relatively more recent period, starting in 1970
OSST <sub>r</sub>	relative operating characteristics
ROC	ranked probability skill score
RPSS	sea surface temperature
SST	tropical cyclone
TC	

## REFERENCES

- Aldrich, J., 1997: R. A. Fisher and the making of maximum likelihood 1912–1922. *Stat. Sci.*, **12**, 162–176.
- Anderson, J. L., 1996: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate*, **9**, 1518–1530.
- Barnston, A. G., 1992: Correspondence among the correlation, RMSE, and Heidke forecast verification measures—Refinement of the Heidke score. *Wea. Forecasting*, **7**, 699–709.
- , and R. E. Livezey, 1987: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Mon. Wea. Rev.*, **115**, 1083–1126.

- , and H. M. van den Dool, 1993: A degeneracy in cross-validated skill in regression-based forecasts. *J. Climate*, **6**, 963–977.
- , M. Chelliah, and S. B. Goldenberg, 1997: Documentation of a highly ENSO-related SST region in the equatorial Pacific. *Atmos.–Ocean*, **35**, 367–383.
- , S. J. Mason, L. Goddard, D. G. DeWitt, and S. Zebiak, 2003: Multimodel ensembling in seasonal climate forecasting at IRI. *Bull. Amer. Meteor. Soc.*, **84**, 1783–1796.
- , A. Kumar, L. Goddard, and M. P. Hoerling, 2005: Improving seasonal prediction practices through attribution of climate variability. *Bull. Amer. Meteor. Soc.*, **86**, 59–72.
- Bell, G. D., and Coauthors, 2000: Climate assessment for 1999. *Bull. Amer. Meteor. Soc.*, **81** (6), S1–S50.
- Bengtsson, L., 2001: Hurricane threats. *Science*, **293**, 440–441.
- , H. Böttger, and M. Kanamitsu, 1982: Simulation of hurricane-type vortices in a general circulation model. *Tellus*, **34**, 440–457.
- , U. Schlese, E. Roeckner, M. Latif, T. P. Barnett, and N. E. Graham, 1993: A two-tiered approach to long-range climate forecasting. *Science*, **261**, 1026–1029.
- , M. Botzet, and M. Esch, 1995: Hurricane-type vortices in a general circulation model. *Tellus*, **47A**, 175–196.
- , —, and —, 1996: Will greenhouse gas-induced warming over the next 50 years lead to higher frequency and greater intensity of hurricanes? *Tellus*, **48A**, 57–73.
- , K. I. Hodges, and M. Esch, 2007a: Tropical cyclones in a T159 resolution global climate model: Comparison with observations and re-analysis. *Tellus*, **59A**, 396–416.
- , —, —, N. Keenlyside, L. Kornbluh, J.-J. Luo, and T. Yamagata, 2007b: How may tropical cyclones change in a warmer climate? *Tellus*, **59A**, 539–561.
- Broccoli, A. J., and S. Manabe, 1990: Can existing climate models be used to study anthropogenic changes in tropical cyclone climate? *Geophys. Res. Lett.*, **17**, 1917–1920.
- Buckley, M. J., L. M. Leslie, and M. S. Speer, 2003: The impact of observational technology on climate database quality: Tropical cyclones in the Tasman Sea. *J. Climate*, **16**, 2640–2645.
- Camargo, S. J., and S. E. Zebiak, 2002: Improving the detection and tracking of tropical storms in atmospheric general circulation models. *Wea. Forecasting*, **17**, 1152–1162.
- , and A. H. Sobel, 2004: Formation of tropical storms in an atmospheric general circulation model. *Tellus*, **56A**, 56–67.
- , and A. G. Barnston, 2008: Description and skill evaluation of experimental dynamical seasonal forecasts of tropical cyclone activity at IRI. International Research Institute for Climate and Society Tech. Rep. 08-02, Columbia University, Palisades, NY, 35 pp. [Available online at <http://portal.iri.columbia.edu>.]
- , —, and S. E. Zebiak, 2005: A statistical assessment of tropical cyclones in atmospheric general circulation models. *Tellus*, **57A**, 589–604.
- , —, P. J. Klotzbach, and C. W. Landsea, 2007a: Seasonal tropical cyclone forecasts. *WMO Bull.*, **56**, 297–309.
- , H. Li, and L. Sun, 2007b: Feasibility study for downscaling seasonal tropical cyclone activity using the NCEP Regional Spectral Model. *Int. J. Climatol.*, **27**, 311–325, doi:10.1002/joc.1400.
- , A. H. Sobel, A. G. Barnston, and K. A. Emanuel, 2007c: Tropical cyclone genesis potential index in climate models. *Tellus*, **59A**, 428–443.
- Chan, J. C. L., J. E. Shi, and C. M. Lam, 1998: Seasonal forecasting of tropical cyclone activity over the western North Pacific and the South China Sea. *Wea. Forecasting*, **13**, 997–1004.
- , —, and K. S. Liu, 2001: Improvements in the seasonal forecasting of tropical cyclone activity over the western North Pacific. *Wea. Forecasting*, **16**, 491–498.
- Chen, D., M. A. Cane, A. Kaplan, S. E. Zebiak, and D. Huang, 2004: Predictability of El Niño over the past 148 years. *Nature*, **428**, 733–735.
- Chu, J.-H., C. R. Sampson, A. S. Levine, and E. Fukada, cited 2002: The Joint Typhoon Warning tropical cyclone best-tracks, 1945–2000. Naval Research Laboratory, NRL/MR/7540-02-16. [Available online [http://metocph.nmci.navy.mil/jtwc/best\\_tracks/TC\\_bt\\_report.html](http://metocph.nmci.navy.mil/jtwc/best_tracks/TC_bt_report.html).]
- Epstein, E. S., 1969: A scoring system for probability forecasts of ranked categories. *J. Appl. Meteor.*, **8**, 985–987.
- Goddard, L., A. G. Barnston, and S. J. Mason, 2003: Evaluation of the IRI's "net assessment" seasonal climate forecasts: 1997–2001. *Bull. Amer. Meteor. Soc.*, **84**, 1761–1781.
- Goldenberg, S. B., and L. J. Shapiro, 1996: Physical mechanisms for the association of El Niño and West African rainfall with Atlantic major hurricane activity. *J. Climate*, **9**, 1169–1187.
- , C. W. Landsea, A. M. Mestas-Núñez, and W. M. Gray, 2001: The recent increase in Atlantic hurricane activity: Causes and implications. *Science*, **293**, 474–479.
- Gray, W. M., C. W. Landsea, P. W. Mielke Jr., and K. J. Berry, 1993: Predicting Atlantic basin seasonal tropical cyclone activity by 1 August. *Wea. Forecasting*, **8**, 73–86.
- Harper, B. A., S. A. Stroud, M. McCormack, and S. West, 2008: A review of historical tropical cyclone intensity in northwestern Australia and implications for climate change and trend analysis. *Aust. Meteor. Mag.*, **57**, 121–141.
- Hsu, W.-R., and A. H. Murphy, 1986: The attributes diagram: A geometrical framework for assessing the quality of probability forecasts. *Int. J. Forecast.*, **2**, 285–293.
- Ji, M., D. W. Behringer, and A. Leetmaa, 1998: An improved coupled model for ENSO prediction and implications for ocean initialization. Part II: The coupled model. *Mon. Wea. Rev.*, **126**, 1022–1034.
- Kharin, V. V., and F. W. Zwiers, 2002: Climate predictions with multimodel ensembles. *J. Climate*, **15**, 793–799.
- Klotzbach, P. J., 2007a: Recent developments in statistical prediction of seasonal Atlantic basin tropical cyclone activity. *Tellus*, **59A**, 511–518.
- Knutson, T. R., J. J. Sirutis, S. T. Garner, and R. E. T. I. M. Held, 2007: Simulation of the recent multidecadal increase of Atlantic hurricane activity using an 18-km-grid regional model. *Bull. Amer. Meteor. Soc.*, **88**, 1549–1565.
- Landman, W. A., A. Seth, and S. J. Camargo, 2005: The effect of regional climate model domain choice on the simulation of tropical cyclone-like vortices in the southwestern Indian Ocean. *J. Climate*, **18**, 1263–1274.
- Landsea, C. W., B. A. Harper, K. Horau, and J. A. Knaff, 2006: Can we detect trends in extreme tropical cyclones? *Science*, **313**, 452–454.
- Manabe, S., J. L. Holloway, and H. M. Stone, 1970: Tropical circulation in a time-integration of a global model of the atmosphere. *J. Atmos. Sci.*, **27**, 580–613.
- Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291–303.
- Mason, S. J., 2004: On using climatology as a reference strategy in the Brier and ranked probability skill scores. *Mon. Wea. Rev.*, **132**, 1891–1895.
- , L. Goddard, N. E. Graham, E. Yulaeva, L. Q. Sun, and P. A. Arkin, 1999: The IRI seasonal climate prediction system and the 1997/98 El Niño event. *Bull. Amer. Meteor. Soc.*, **80**, 1853–1873.

- Matsuura, T., M. Yumoto, and S. Iizuka, 2003: A mechanism of interdecadal variability of tropical cyclone activity over the western North Pacific. *Climate Dyn.*, **21**, 105–117.
- Owens, B. F., and C. W. Landsea, 2003: Assessing the skill of operational Atlantic seasonal tropical cyclone forecasts. *Wea. Forecasting*, **18**, 45–54.
- Peng, P. T., A. Kumar, A. G. Barnston, and L. Goddard, 2000: Simulation skills of the SST-forced global climate variability of the NCEP-MRF9 and the Scripps-MPI ECHAM3 models. *J. Climate*, **13**, 3657–3679.
- Rajagopalan, B., U. Lall, and S. E. Zebiak, 2002: Categorical climate forecasts through regularization and optimal combination of multiple GCM ensembles. *Mon. Wea. Rev.*, **130**, 1792–1811.
- Repelli, C. A., and P. Nobre, 2004: Statistical prediction of sea-surface temperature over the tropical Atlantic. *Int. J. Climatol.*, **24**, 45–55.
- Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate*, **15**, 1609–1625.
- Robertson, A. W., U. Lall, S. E. Zebiak, and L. Goddard, 2004: Optimal combination of multiple atmospheric GCM ensembles for seasonal prediction. *Mon. Wea. Rev.*, **132**, 2732–2744.
- Roeckner, E., and Coauthors, 1996: The atmospheric general circulation model ECHAM-4: Model description and simulation of present-day climate. Max Planck Institute for Meteorology Tech. Rep. 218, Hamburg, Germany, 90 pp.
- Royer, J.-F., F. Chauvin, B. Timbal, P. Araspin, and D. Grimal, 1998: A GCM study of the impact of greenhouse gas increase on the frequency of occurrence of tropical cyclones. *Climatic Change*, **38**, 307–343.
- Ryan, B. F., I. G. Watterson, and J. L. Evans, 1992: Tropical cyclone frequencies inferred from Gray's yearly genesis parameter: Validation of GCM tropical climate. *Geophys. Res. Lett.*, **19**, 1831–1834.
- Saha, S., and Coauthors, 2006: The NCEP Climate Forecast System. *J. Climate*, **19**, 3483–3517.
- Saunders, M. A., and A. S. Lea, 2004: Seasonal prediction of hurricane activity reaching the coast of the United States. *Nature*, **434**, 1005–1008, doi:10.1038/nature03454.
- Thorncroft, C., and I. Pytharoulis, 2001: A dynamical approach to seasonal prediction of Atlantic tropical cyclone activity. *Wea. Forecasting*, **16**, 725–734.
- Tippett, M. K., A. G. Barnston, and A. W. Robertson, 2007: Estimation of seasonal precipitation tercile-based categorical probabilities from ensembles. *J. Climate*, **20**, 2210–2228.
- Trewin, B., 2008: An enhanced tropical cyclone data set for the Australian region. Preprints, *20th Conf. on Climate Variability and Change/Tropical Meteorology Special Symposium*, New Orleans, LA, Amer. Meteor. Soc., JP3.1. [Available online at <http://ams.confex.com/ams/pdfpapers/128054.pdf>.]
- van den Dool, H. M., 1994: Searching for analogues, how long must we wait? *Tellus*, **46A**, 314–324.
- , 2007: *Empirical Methods in Short-Term Climate Prediction*. Oxford University Press, 215 pp.
- Vimont, J. P., 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, doi:10.1256/qj.05.65.
- , and J. L. Anderson, 2001: Sensitivity of Atlantic tropical storm frequency to ENSO and interdecadal variability of SSTs in an ensemble of AGCM integrations. *J. Climate*, **14**, 533–545.
- , and T. N. Stockdale, 2001: Seasonal forecasting of tropical storms using coupled GCM integrations. *Mon. Wea. Rev.*, **129**, 2521–2537.
- , J. L. Anderson, and W. F. Stern, 1997: Simulation of interannual variability of tropical storm frequency in an ensemble of GCM integrations. *J. Climate*, **10**, 745–760.
- , —, and —, 1999: Impact of large-scale circulation on tropical storm frequency, intensity and location, simulated by an ensemble of GCM integrations. *J. Climate*, **12**, 3237–3254.
- , D. Anderson, and T. Stockdale, 2003: Seasonal forecasting of tropical cyclone landfall over Mozambique. *J. Climate*, **16**, 3932–3945.
- , 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.
- Walsh, K. J. E., and B. F. Ryan, 2000: Tropical cyclone intensity increase near Australia as a result of climate change. *J. Climate*, **13**, 3029–3036.
- , K. C. Nguyen, and J. L. McGregor, 2004: Fine-resolution regional climate model simulations of the impact of climate change on tropical cyclones near Australia. *Climate Dyn.*, **22**, 47–56.
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. International Geophysics Series, Vol. 59, Academic Press, 627 pp.
- Wu, G., and N. C. Lau, 1992: A GCM simulation of the relationship between tropical storm formation and ENSO. *Mon. Wea. Rev.*, **120**, 958–977.