# Simulation of Interannual Variability of Tropical Storm Frequency in an Ensemble of GCM Integrations

F. VITART

*Program in Atmospheric and Oceanic Science, Princeton University, Princeton, New Jersey*

J. L. ANDERSON AND W. F. STERN

*GFDL/NOAA, Princeton University, Princeton, New Jersey*

(Manuscript received 5 April 1996, in final form 12 July 1996)

ABSTRACT

The present study examines the simulation of the number of tropical storms produced in GCM integrations with a prescribed SST. A 9-member ensemble of 10-yr integrations (1979–88) of a T42 atmospheric model forced by observed SSTs has been produced; each ensemble member differs only in the initial atmospheric conditions. An objective procedure for tracking-model-generated tropical storms is applied to this ensemble during the last 9 yr of the integrations (1980–88). The seasonal and monthly variations of tropical storm numbers are compared with observations for each ocean basin.

Statistical tools such as the Chi-square test, the F test, and the *t* test are applied to the ensemble number of tropical storms, leading to the conclusion that the potential predictability is particularly strong over the western North Pacific and the eastern North Pacific, and to a lesser extent over the western North Atlantic. A set of tools including the joint probability distribution and the ranked probability score are used to evaluate the simulation skill of this ensemble simulation. The simulation skill over the western North Atlantic basin appears to be exceptionally high, particularly during years of strong potential predictability.

## 1. Introduction

Tropical storms, which are defined as tropical cyclonic systems with a maximum sustained wind larger than 17 m $s^{-1}$, are among the most devastating natural disasters. So the prediction of their frequency, their trajectories, and their intensities is a very important problem. Hurricane predictions over the western North Atlantic, the North Indian Ocean, and the western North Pacific are of special interest since these areas have a dense population and tropical storms over these basins have in the past caused great losses of human life. The vast majority of research on tropical storm prediction has been concerned with the evolution of individual storms over a period of a few days. Here, we are interested in a different problem—the prediction of the interannual variability of tropical storm frequency.

To date there have been three ways to try to predict tropical storm interannual variability. One technique consists of using statistical means (Gray et al. 1992, 1993, 1994; Hess et al. 1995; Nicholls 1992; Basher and Zheng 1995). Another method (Ryan et al. 1992; Watterson et al. 1995) uses a seasonal genesis parameter from a GCM. Gray (1979) identified various large-scale parameters that are believed to play an essential role in tropical storm cyclogenesis. From these parameters a seasonal cyclogenesis index is computed, and the number of tropical storms formed per region is deduced from this index. A third way to predict tropical storm frequency is to track individual tropical storms produced by a GCM. The present study examines the last technique.

Manabe et al. (1970) described for the first time disturbances similar to observed tropical storms in a low-resolution GCM. Bengtsson et al. (1982) reported a study of intense hurricane-type vortices in the European Centre for Medium-Range Weather Forecasts (ECMWF) operational model from 1 January 1980 to 31 December 1980. This model produced vortices with a structure close to that of observed tropical storms in the same regions and at the same time of year. They stressed the important role of sea surface temperature in the cyclogenesis. Krishnamurti (1988) looked at tropical storms generated at different resolutions: T42, T63, and T106. He evaluated how the track and the structure of a tropical storm are affected by increasing resolution. Broccoli and Manabe (1990) used R15 and R30 spectral models coupled with a simple ocean model to investigate the impact of atmospheric $CO_2$ increase on tropical

*Corresponding author address:* Frederic Vitart, NOAA/GFDL, Princeton University, P.O. Box 308, Princeton, NJ 08542.
E-mail: fvitart@phoenix.princeton.edu

storm frequency. They obtained a distribution of tropical storm numbers close to the observed, except over the South Atlantic, where the model generated tropical storms (no tropical storms are observed in this region), and over the eastern North Pacific, where the model greatly underestimated the number of tropical storms. Haarsma et al. (1993) also investigated the effect of doubling atmospheric $CO_2$ on tropical storm frequency using the 11-layer U.K. Meteorological Office atmospheric GCM coupled to a 50-m mixed layer ocean. They obtained a tropical storm distribution for observed $CO_2$ concentrations similar to the one obtained by Broccoli and Manabe (1990), except over the eastern North Pacific, where they obtained a tropical storm frequency much closer to the observations. Wu and Lau (1992) integrated an R15 model forced by observed SSTs. They found a significant correlation between eastern equatorial Pacific SST anomalies and tropical storm frequency over the western North Pacific, the western North Atlantic, and the western South Pacific, in agreement with climate studies. Bengtsson et al. (1995) utilized a 5-yr run of a T106-L19 GCM forced by climatological SSTs. They found that the tropical storms created by this model have a thermal and dynamic structure close to observed tropical storms, and that the frequency and distribution of these tropical storms resemble the observed climatology for all basins. They also found a considerable interannual variability without any interannual SST variations. Tsutsui et al. (1995) presented the tropical storm distribution obtained with a T42-L18 model. They noticed that the model tropical storms they obtained are shorter lived and have a more northward track than observed tropical storms.

The ability of GCMs to simulate tropical storms remains a controversial subject, and some of the papers cited above have been criticized (McBride 1984; Evans 1992; Lander 1993). Most of these critics focused on the fact that the tropical storms generated by GCMs are not entirely dynamically consistent with observed tropical storms.

All these studies on tropical storm frequency used a single GCM run. The present study uses, for the first time for this purpose, a 9-member ensemble of an atmospheric circulation model forced by observed SSTs, which allows us to measure how the tropical storm frequency in this model depends on the details of the SST forcing and to evaluate the skill of this GCM in simulating tropical storms by separating significant signal from noise. If model-generated storm frequency is significantly correlated with observed frequency, the exact details of the model storm dynamics are not directly relevant to the model's ability to simulate storm frequency. The goal of this study is not to prove that the tropical storms generated by the GCM result from the same cyclogenesis process or have exactly the same physical structure as an observed tropical storm, but to investigate whether their interannual variability is consistent with observations.

All the statistics concerning observations have been obtained from the National Hurricane Center (NHC, in Miami, Florida) and the Joint Typhoon Warning Center (JTWC, in Guam).

Section 2 presents a brief description of the design of this 9-member ensemble experiment. In section 3 the objective procedure for tracking tropical storms is described. In section 4 the mean structure of a simulated tropical storm is presented. The tropical storm seasonal and monthly distributions are discussed in section 5. Potential predictability of tropical storm frequency is presented in section 6. In section 7 some statistical tools are applied to evaluate the skill of this GCM in simulating tropical storm frequency.

## 2. Model description

The model used for this study is a T42 atmospheric model with 18 vertical levels. It was developed at the Geophysical Fluid Dynamics Laboratory and has been described by Gordon and Stern (1982). The major physical parameterizations include a ''bucket'' hydrology (Manabe 1969); orographic gravity wave drag (Stern and Pierrehumbert 1988); large-scale condensation and moist convective adjustment (Manabe et al. 1965) (both using a condensation criterion of 100%); shallow convection (based on Tiedtke 1988); cloud prediction [interactive with the radiation (Gordon 1992)]; radiative transfer (12-h averaged), which varies seasonally; stability-dependent vertical eddy fluxes of heat, momentum, and moisture throughout the surface layer, the planetary boundary layer and the free atmosphere (Sirutis and Miyakoda 1990); and $k\nabla^4$ horizontal diffusion. The surface temperatures over land and over sea ice are determined by solving surface heat balance equations (Gordon and Stern 1982).

This atmospheric model is forced by observed SSTs. This observed SST dataset is the one used in the Atmospheric Model Intercomparison Project (Gates 1992) and represents monthly means. Daily values of SSTs are obtained using linear interpolation from the two nearest monthly values.

A 9-member ensemble was integrated for 10 yr from 1 January 1979 to 31 December 1988. The initial conditions for the ensembles were taken from analyses for 12 December 1978 through 21 January 1979, sampled every 5 days; each of these analyses was then used as an initial condition as if it were the analysis for 1 January 1979. The first year of the ensemble integrations has been discarded in an attempt to eliminate direct effects of the initial conditions.

## 3. Objective procedure for tracking model tropical storms

The main problem in tracking model tropical storms is eliminating extratropical disturbances without eliminating tropical disturbances. In the real world it is usu-

ally easy to make a distinction between tropical and extratropical storms using the size of the cyclone as one criterion—extratropical storms tend to be much larger than tropical storms. When using a relatively low-resolution GCM, this approach does not work, as tropical storms generated by the model tend to have a size comparable to that of extratropical storms. Model tropical storms do not have an eye in their center or the spiral shape of observed tropical storms. Bengtsson et al. (1982), Broccoli and Manabe (1990), and Lau (1991) used only velocity, vorticity, or surface pressure as criteria for defining model tropical storms. Their definition of tropical storms is more adapted to coarse resolution models. The presence of a warm core above 500 mb at the center of the cyclone was an additional criterion considered by Haarsma et al. (1993) and Bengtsson et al. (1995).

This T42 model has a horizontal resolution of about 2.8°, which represents about 300 km in the Tropics. To locate the position of the center of the storm with a higher precision, an algorithm similar to the one described by Murray and Simmonds (1991) is used. Bicubic splines interpolate the fields from the grid point values, then a conjugate gradient algorithm locates the position of a maximum or a minimum from a given point.

An objective procedure for tracking model tropical storms is developed using data for 1980 from the first member of the ensemble and 1989 global analyses from ECMWF (Trenberth 1992, with a resolution of 2.5° × 2.5°). The criteria defining this objective procedure are chosen and adjusted to give the most realistic tracks for these datasets. When applied to ECMWF global analyses, 85% of the events detected by this objective procedure correspond to observed tropical storms. Most of the events that do not correspond to observed tropical storms occur in the polar regions and may correspond to the polar lows described by Rasmussen (1985) and Emanuel (1988). These events occur in high latitudes, so they do not have an impact on the statistics of the number of tropical storms. In the model these events have been filtered by the objective procedure since they do not last more than 2 days with a highest sustained velocity larger than 17 m $s^{-1}$. The criteria defining the objective procedure are not modified when applied to all other members and years of the ensemble, as our main goal is to investigate the tropical storm number interannual variability.

The tropical storm detection algorithm first locates the position of intense vortices with a warm core for each day as follows.

- A local maximum of vorticity larger than $3.5 \times 10^{-5}$ $s^{-1}$ at 850 mb is located.
- The closest local minimum sea level pressure is defined as the center of the storm.
- The closest local maximum of average temperature between 500 and 200 mb is located and is defined as the center of the warm core. The distance between the center of the warm core and the center of the storm must not exceed 2° latitude. From the center of the warm core the temperature must decrease by at least 0.5°C in all directions within a distance of 8° latitude.
- The closest local maximum thickness between 1000 and 200 mb is located. The distance between this local maximum and the center of the storm must not exceed 2° latitude. From this local maximum, the thickness must decrease by at least 50 m in all directions within a distance of 8° latitude.

Although the criteria on maximum temperature anomaly and maximum thickness between 1000 and 200 mb seem to be redundant since both of them characterize the presence of a warm core, the use of both of these criteria was a useful way to eliminate several extratropical storms that satisfied one of these criteria. All of these criteria must be satisfied to define a storm. After storms are located for each day, an objective procedure is applied to find storm trajectories as follows.

1) For a given storm, we examine whether there are storms that appear on the following day at a distance of less than 800 km.
2) If there is no such storm, then the trajectory is considered to have stopped. Otherwise, in 95% of the cases there is only one appropriate storm in the following day, which is then considered to belong to the same trajectory as the initial storm, and the first step is repeated. For the remaining 5% of cases, which occur mainly over the western North Pacific, there is more than one storm within 800 km, and a first preference is given to the storms located in the northwestern quadrant in the Northern Hemisphere or southwestern quadrant in the Southern Hemisphere, relative to the initial storm. Next, the closest storm is chosen as belonging to the same trajectory as the initial storm. The first step is then repeated.
3) To be considered as a model tropical storm trajectory, a trajectory must last at least 2 days and have a maximum wind velocity within an 8° circle centered on the middle of the storm, which must be larger than 17 m $s^{-1}$ during at least 2 days (not necessarily consecutive).

Cases satisfying all these criteria will henceforth be referred to as tropical storms.

## 4. Mean structure of a simulated tropical storm

Previous studies, for example, Wu and Lau (1992) and Bengtsson et al. (1995), have displayed many similarities between the structure of a tropical storm generated by a GCM and an observed tropical storm. Wu and Lau (1992) presented a composite pattern of a tropical storm (Fig. 3 in Wu and Lau 1992) generated by an R15 model over the western North Pacific basin. It appeared that although their model had a tendency to

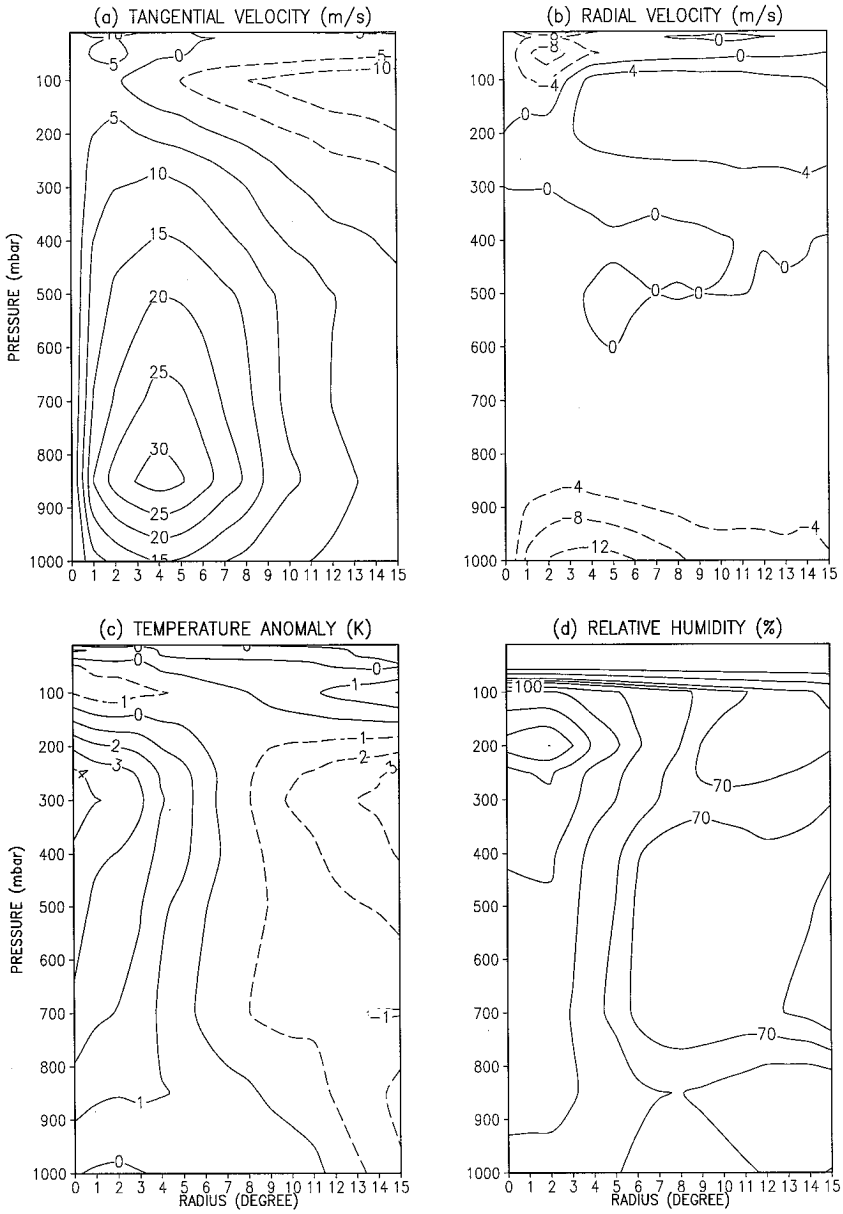


FIG. 1. Two-dimensional cross section of (a) tangential wind (m $s^{-1}$), (b) radial wind (m $s^{-1}$), (c) temperature anomaly (K), and (d) relative humidity (%) of a tropical storm generated over the western North Pacific by one element of the ensemble on 30 August 1980.

underestimate the storm intensity, the 200-, 950-, and 1000-mb wind circulations generated by the R15 GCM had many qualitative similarities to their observed counterparts: convergence, high moisture content, strong upward motion, and heavy precipitation at the lower levels of the tropical storm, and anticyclonic vorticity and divergence at 200 mb. Bengtsson et al. (1995) presented a 2D cross section of a hurricane generated by T-42 and T-106 models. Their study showed that the tropical storm generated by the T-106 model agrees very well with a Pacific composite observed typhoon (Frank 1977). With a T-42 resolution, the structure of the simulated storm still looks very realistic, although it has a larger horizontal extent.

The realism of the structure of a model tropical storm depends on the resolution of the model, as Bengtsson et al. (1995) emphasized, but it may also depend strongly on the model parameterization. For this reason, the mean structure of a simulated tropical storm may be very different from one GCM to another. Figure 1 presents a 2D cross section of mean tangential velocity, mean radial velocity, mean temperature anomaly, and mean relative humidity from a tropical storm that has been generated by one member of our ensemble. The

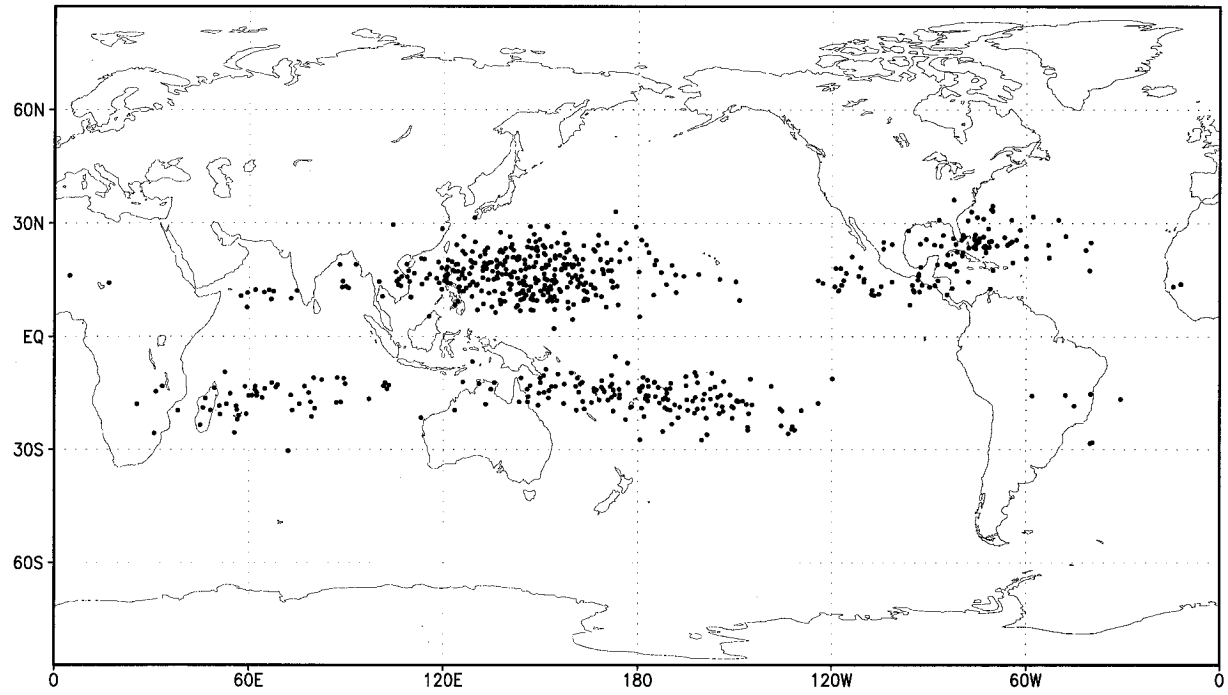


FIG. 2. First position of all the tropical storms generated by 1 member of the ensemble during the 9 yr of the simulation. Each point represents the first appearance of at least one tropical storm.

tropical storm that has been chosen for this study occurred over the western North Pacific, and Fig. 1 represents its state on 30 August 1980.

The simulated tropical storm structure presented in Fig. 1 seems to be consistent with observations by Frank (1977), although the simulated storm is too big compared with observations—the radius of maximum winds is about 400 km in the simulation versus around 25 km in observations (Weatherford and Gray 1988). The tangential velocity has its maximum in the lower troposphere, and an anticyclonic circulation is visible in the upper troposphere; the radial wind is convergent in the lower troposphere and divergent above 300 mb. The negative radial wind near the center of the storm and above 100 mb may be due to a slight tilt of this tropical storm in the upper troposphere; the temperature anomaly (defined as the temperature minus the mean temperature within a 15° circle centered at the middle of the tropical storm) presents a warm core located at about 300 mb and a weak cold core at 100 mb, which is consistent with Fig. 3 in Frank (1977). The relative humidity is very high particularly near the center of the storm where it reaches 100%.

## 5. Model tropical storm distribution

Although the primary goal of this paper is to investigate the interannual variability of model tropical storms, the study of their climatological distribution may give some information about the behavior of the model in each ocean basin.

The objective procedure for tracking model tropical storms is applied to the 9-member ensemble for the 1980–88 period. Figure 2 represents the first position of all the tropical storms generated by 1 member of the ensemble during the 9 yr of the simulation. This ensemble of initial positions changes little from one member of the ensemble to another.

All of the tropical storms generated by the model appear at the same latitudes as in the real world. The majority form in the same regions and take place at the same time of year, but in the model some tropical disturbances are created in the following places:

- in the South Atlantic (about 1 $yr^{-1}$),
- over land (about 2 $yr^{-1}$), and
- too far eastward in the South Pacific.

The creation of tropical storms over the South Atlantic has already been noticed in numerous studies (Broccoli and Manabe 1990; Wu and Lau 1992; Haarsma et al. 1993; Tsutsui et al. 1995). It is possible that the occurrence of some land tropical storms in GCMs is due to an insufficiently accurate surface temperature prediction (Tuleya 1994). It is likely too that some of the land storms identified are monsoon depressions, which do occur in the real world (McBride 1987).

Figure 3 represents all of the tropical storm trajectories generated by one member of the ensemble during 3 yr. These trajectories appear to be more poleward and shorter than observed tropical storm trajectories (Gray 1979). The fact that they are more poleward may explain

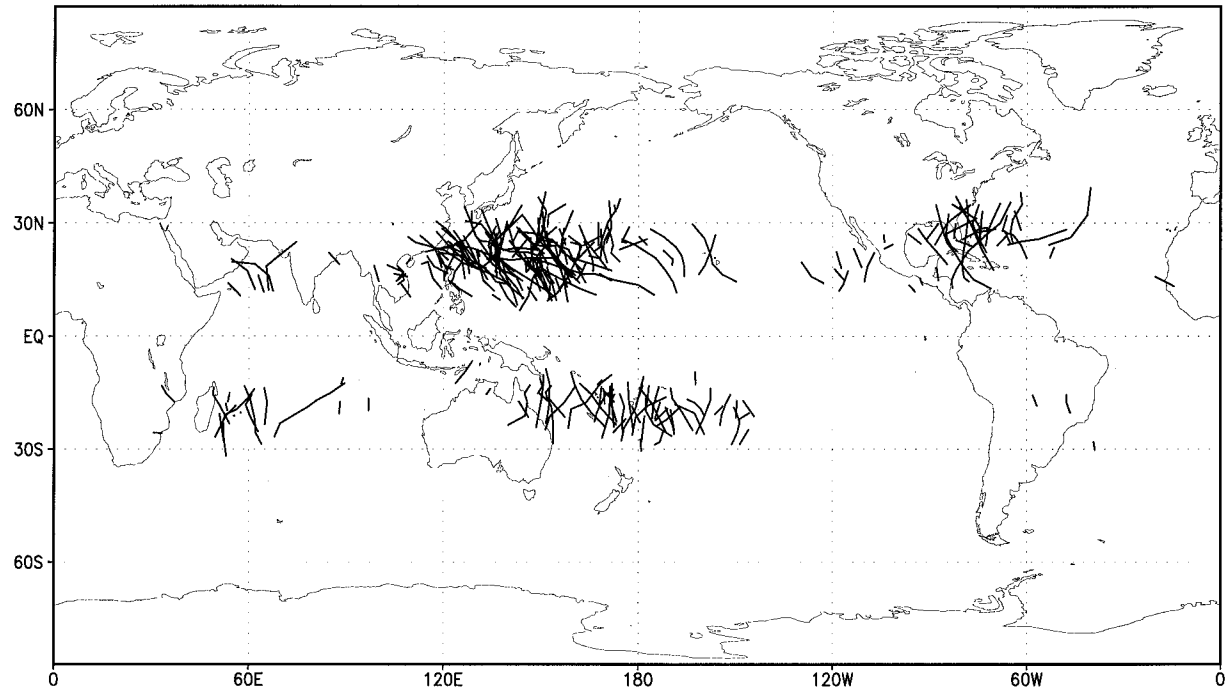


FIG. 3. Trajectories of all the tropical storms generated by one member of the ensemble during 3 yr (1980, 1981, and 1982).

why the model tropical storms have shorter lives than those observed, since they move faster to colder water.

As presented in Fig. 4, seven basins where tropical storms occur are defined: the western North Atlantic, the eastern North Pacific, the western North Pacific, the north Indian Ocean, the south Indian Ocean, the Australian Basin, and the South Pacific. The term ''total'' refers to the tropical storms that occur over these seven

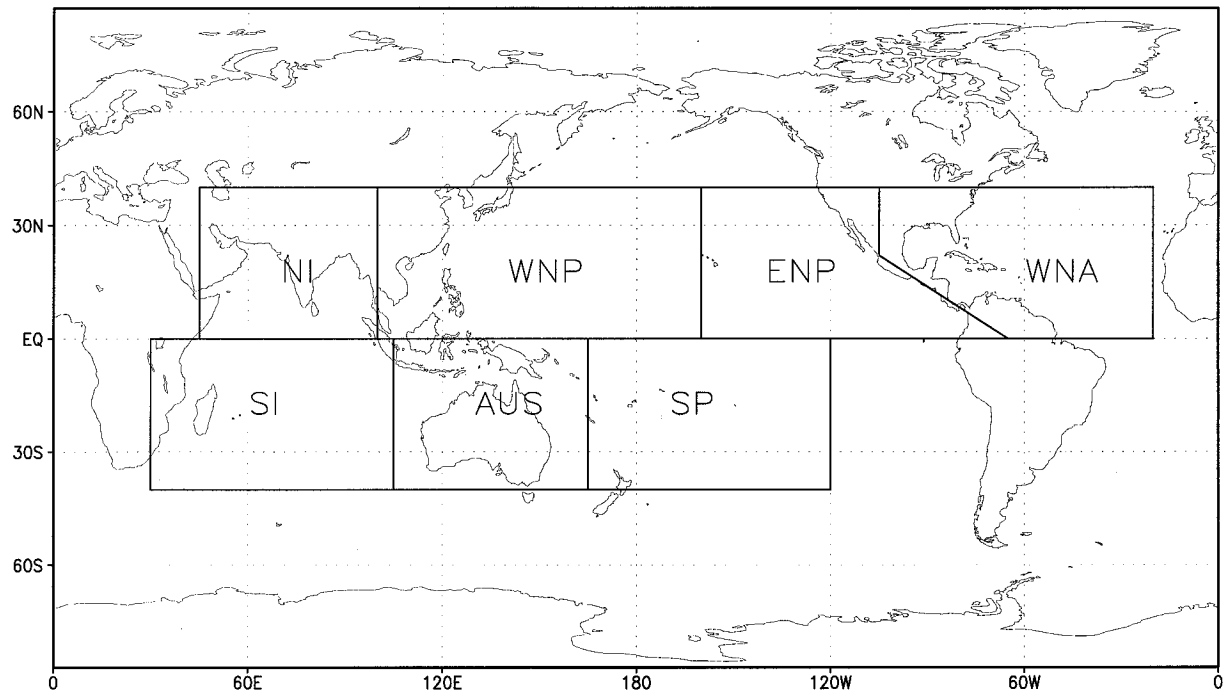


FIG. 4. Specification of the ocean basins. WNA = western North Atlantic, ENP = eastern North Pacific, WNP = western North Pacific, NI = north Indian Ocean, SI = south Indian Ocean, AUS = Australian Basin, and SP = South Pacific.

TABLE 1. Average tropical storm number per season.

| Basin | WNA | ENP | WNP | NI | SI | AUS | SP | Total |
|---|---|---|---|---|---|---|---|---|
| Model | 8.7 | 3 | 33.8 | 2.4 | 6.5 | 4.5 | 11.5 | 70.4 |
| Observations | 8.8 | 17.4 | 27.3 | 4.3 | 12 | 10.3 | 5.5 | 86 |

basins (it does not include tropical storms that occur over land or over the South Atlantic). These basins are similar to those presented by Gray (1979), but in this study the eastern North Pacific and the western North Pacific extend to the Hawaiian Islands and the South Pacific extends to 120°W. These differences are due to the fact that Gray (1979) was concerned with tropical storm cyclogenesis, while this study focuses on mature tropical cyclones.

Over each of these basins the numbers of tropical storms generated by the model per month and per season are counted for each element of the ensemble. The season over the Northern Hemisphere begins in January and ends in December; over the Southern Hemisphere it begins in July and ends in June. There are nine tropical storm seasons over the Northern Hemisphere and eight over the Southern Hemisphere. In this section we present only results obtained with ensemble means.

### *a. Climatological means*

The average duration of the model tropical storms is 4 days, with an average maximum surface wind of 31.4 m $s^{-1}$. Based on observations, real tropical storms have an average duration of 6 days, with an average maximum surface velocity of 37.6 m $s^{-1}$, so the model simulates shorter and less intense tropical storms than those observed.

Table 1 presents the average number of model and observed tropical storms per season. Over the western North Atlantic basin the model and the observations give a similar average number of tropical storms per season. Over the western North Pacific the model simulates a number of tropical storms close to observations, although there is a slight overestimation. This overestimation may be related to the unrealistically strong moisture convergence center created by this T42 model over the western tropical Pacific (Stern and Miyakoda 1995). Over the northern Indian Ocean both the model and observations indicate a low number of tropical storms per season. The poor skill of this T42 model in simulating the Indian monsoon (Sperber and Palmer 1996) may explain the slight underestimation of tropical storm number over the North Indian Ocean. Over all the other basins the simulation and the observations do not agree as well. In particular, over the eastern North Pacific the model greatly underestimates the average number of tropical storms per season. This may be due to the fact that this basin is the only one where tropical storms form close to high mountain regions (Sierra Madre Occidental). The model we use for this study is a spectral model, which has a tendency to generate ripples of sea surface elevation in areas close to mountain regions (Navarra et al. 1994).

In the Southern Hemisphere the tropical storm distribution per basin is much further from observations than in the Northern Hemisphere. In the Southern Hemisphere the model has a tendency to create tropical storms further eastward than they are in the real world. That may explain why the model greatly overestimates the number of tropical storms over the South Pacific and severely underestimates it over the two other basins in the Southern Hemisphere.

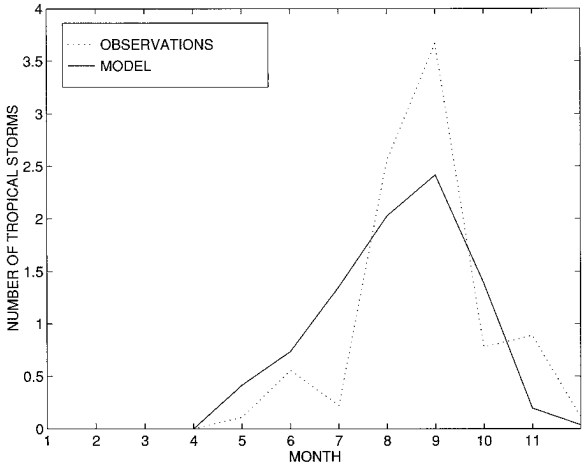


FIG. 5. Ensemble mean number of tropical storms and observed number of tropical storms as a function of month over the western North Atlantic.

### *b. Monthly variability*

Figures 5 and 6 display the average number of model and observed tropical storms for each month of the year over the western North Atlantic and the western North Pacific.

Over the western North Atlantic the observed tropical storm season ranges from May to December, with a maximum number in September. The model reproduces this, but it has a tendency to create more tropical storms in the beginning of the season and in October, and fewer in September, November, and December. Over the eastern North Pacific (not shown) tropical storms occur at the same period as over the western North Atlantic, but their maximum number is observed in July. The model underestimates their number and produces a maximum frequency in August instead of July. Over the western North Pacific (Fig. 6) both observations and the model

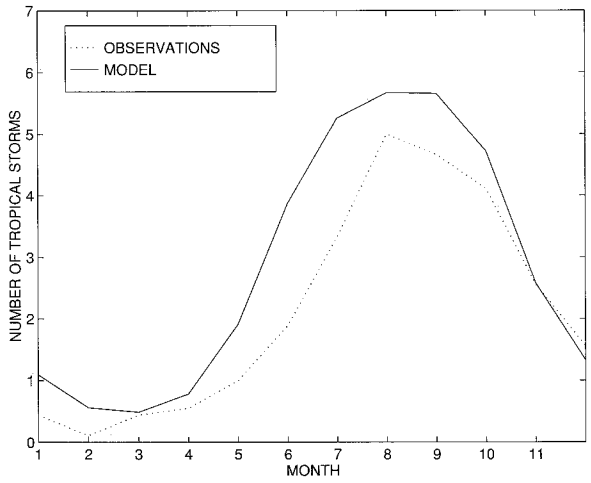


FIG. 6. As in Fig. 5 but for western North Pacific.

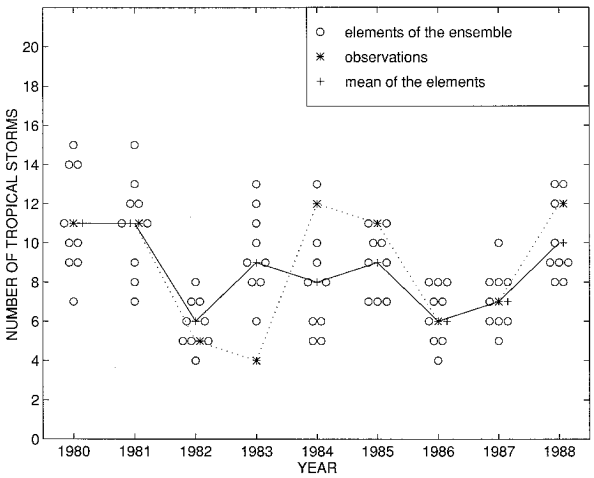


FIG. 7. Ensemble simulation of number of tropical storms and observed number of tropical storms over the western North Atlantic. The solid line represents observed tropical storm numbers. The dotted line represents the mean tropical storm numbers of the elements of the ensemble. Each circle represents the tropical storm number of one element of the ensemble.

display a tropical storm season ranging from January to December, with a maximum in August–September. Over the north Indian Ocean (not shown) it has been observed that tropical storms occur primarily during two short periods of the year: May–June and October–November. Their frequency is much higher during the second period. The model is successful in simulating the occurrence of the first tropical storm period, with a maximum number in June, but fails to simulate the tropical storms occurring in October–November. Over the Southern Hemisphere (not shown) the monthly number of observed tropical storms is close to the number of model tropical storms, except during the period of the maximum in January–February, when the model simulates fewer tropical storms than are observed.

### *c. Interannual variability*

Gray (1979) noted that a property of observed tropical storms is that their total number displays very little interannual variability. The annual tropical storm total number from 1980 to 1988 had a standard deviation of 7.4 according to observations. The model simulates a standard deviation of 5.25, which is small compared to the average total number of model tropical storms, which is 70.

Over the western North Atlantic the observations (Fig. 7) display a small number of tropical storms in 1982, 1983, 1986, and 1987. Figure 7 shows that the model is able to simulate a low mean number of tropical storms during the years 1982, 1986, and 1987, and the mean number of tropical storms generated by the model has an interannual variability close to observations, except in 1983 and 1984. Over the western North Pacific (Fig. 8) the model simulates tropical storm number interannual variability similar to the observed, but with more amplitude. Over the eastern North Pacific (Fig. 9) the model simulates realistic interannual variability in tropical storm number, although the climatological number of tropical storms is much less in the model than in observations. Over the south Indian Ocean (not shown) and the Australian Basin (not shown) the observations display a reduced number of tropical storms in 1983 and 1987. The model simulates these decreases over the Australian Basin, but gives the opposite interannual variability over the south Indian Ocean. Over the South Pacific (Fig. 10) the observations display an increase in the number of tropical storms in 1983 and 1987, when the model simulates a decrease. This may be related to the fact that tropical storms are created further eastward than in the real world over the Southern Hemisphere.

Several previous studies pointed out that the interannual variability of tropical storm frequency over some ocean basins was a consequence of El Niño events; Gray

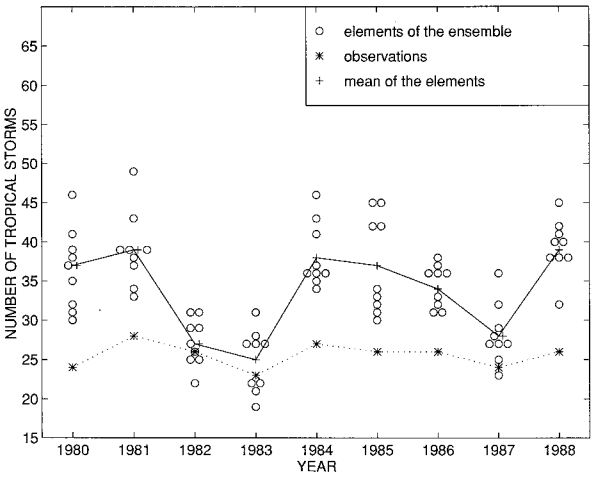


FIG. 8. As in Fig. 7 but for the western North Pacific.

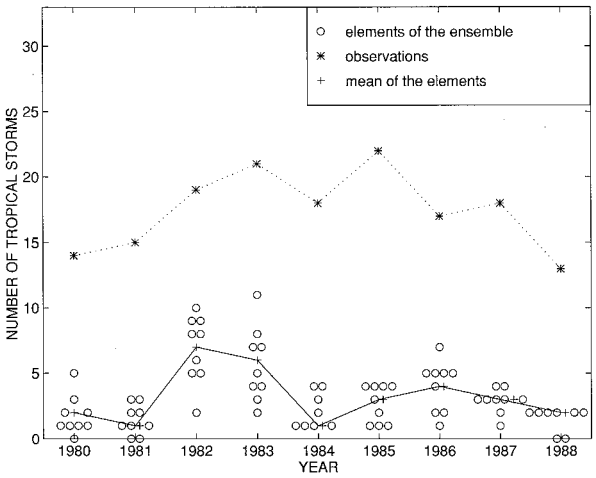


FIG. 9. As in Fig. 7 but for the eastern North Pacific.

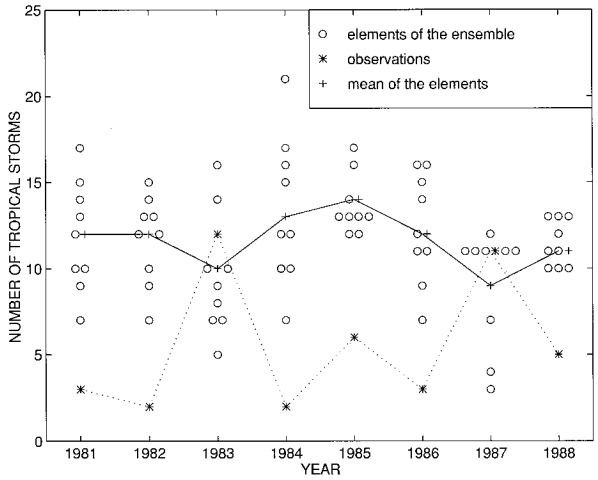


FIG. 10. As in Fig. 7 but for the South Pacific.

(1984) noticed that the western North Atlantic hurricane number was reduced during El Niño events, and Cayan and Webb (1992) conclude that this is also true for the eastern North Pacific. Pan (1981) and Chan (1985) identified a strong connection between northwest Pacific tropical cyclone frequency and El Niño events. Wu and Lau (1992) found, in agreement with Pan (1981) and Chan (1985), that there is a strong correlation between eastern tropical Pacific sea surface temperature anomalies and the number of tropical storms generated by their R15 model in the western North Pacific.

One simple way to evaluate the ability of the model to simulate realistic interannual variations of tropical storm frequency is to examine the linear correlation between the observations and the mean of the 9-member ensemble. The results for each basin are displayed in Table 2.

The correlation is highest over the western North Pacific and is greater than 0.5 over the eastern North Pacific, western North Atlantic, and Australian Basin. This suggests that the model simulates realistic interannual variations in the number of tropical storms over these basins with some skill. The correlation is highly negative over the north Indian Ocean, the south Indian Ocean, and the South Pacific, which indicates that the model has an unrealistic behavior over these three basins.

Table 3 presents the linear correlations between the interannual variability of tropical storm numbers over the western North Atlantic simulated by the model for each member of the ensemble and the observed. It appears from this table that only 2 members of the ensemble (members 4 and 9) have a linear correlation with observations greater than the linear correlation between the mean of the ensemble and observed number of tropical storms. This result, which is similar over all basins, illustrates how useful ensemble runs are in comparison with a single run. The high correlations obtained over the western North Atlantic, eastern North Pacific, and western North Pacific are mainly due to the use of large ensembles. This result is in agreement with previous studies (Leith 1974; Murphy 1988), who showed that an ensemble run by filtering the noise generated by the model performs better than most of the individual runs that are part of the ensemble.

## 6. Potential predictability of tropical storm number

The term ''potential predictability'' has been used with different meanings in numerous previous studies (Madden 1976; Madden 1981; Shukla 1981; Hayashi 1986; Chervin 1986). In the present study, the ensemble simulation is said to have potential predictability if the ensemble distribution of tropical storm numbers for a given year can be distinguished from the model ''climatological'' distribution of this quantity in a statistically significant way. In other words, this 9-member ensemble displays potential predictability if the simulation it provides for a given year is significantly different from the simulation produced by randomly selecting 9 elements from the model climatological distribution. The model climatological distribution is defined in this study as being the 9-yr ensemble

TABLE 2. Linear correlation for each ocean basin between observations and the mean of the 9-member ensemble.

| | Total | WNA | ENP | WNP | NI | SI | AUS | SP |
|---|---|---|---|---|---|---|---|---|
| Correlation | 0.41 | 0.56 | 0.545 | 0.66 | −0.63 | −0.66 | 0.57 | −0.775 |
| Significance | 0.69 | 0.89 | 0.88 | 0.945 | 0.93 | 0.93 | 0.87 | 0.977 |

TABLE 3. Linear correlation for the western North Atlantic between observations and each member of the ensemble.

| Member | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Correlation | 0.43 | 0.14 | 0.44 | 0.84 | −0.29 | 0.51 | 0.55 | 0.11 | 0.61 |
| Significance | 0.77 | 0.28 | 0.76 | 0.99 | 0.56 | 0.84 | 0.87 | 0.23 | 0.92 |

distribution for the Northern Hemisphere basins and the 8-yr ensemble distribution for the Southern Hemisphere basins. This climatological distribution has 81 members in the Northern Hemisphere and 72 members in the Southern Hemisphere. The main difficulty consists in finding the best statistical tool to evaluate this potential predictability for discrete quantities like tropical storm number. Different classical tools that measure the statistical difference between the yearly and model climatological distribution of tropical storm number are used and compared over the western North Atlantic. The results are presented in Table 4.

### *a. Measure of the spread*

One way to evaluate the potential predictability for a given year is to use a measure of the ensemble spread (Stern and Miyakoda 1995). In this study the dispersion or scatter of the members of the ensemble is referred to as the ensemble spread. The variance for a given year divided by the variance for the 9-yr ensemble represents a good quantitative measure of the spread. Traditionally, when this quantity is significantly smaller than 1, the model is considered to have skill in making predictions and so to have potential predictability. The F test is used to evaluate the significance of the spread for each year; a large significance of the F test indicates that the variance for that year is significantly smaller than the climatological variance.

According to Table 4, over the western North Atlantic the significance of the F test is larger than 90% in 1982, 1985, 1986, and 1987, which indicates strong potential predictability during these 4 yr.

As stressed by Anderson and Stern (1996), the principal defect of this method is that it makes implicit assumptions about the normality of the underlying probability distributions and the nature of the differences between ensemble distributions for different years.

### *b. Student's t-test for significantly different means*

Another way to evaluate how different two distributions are is to look at their means. For this purpose we use the unequal variance *t* test (Press 1986). According to Table 4, the ensemble distributions during the years 1980, 1981, 1982, 1986, 1987, and 1988 display significant differences in the mean from the climatological mean, which implies potential predictability during these years.

### *c. Kolmogorov–Smirnov and Kuiper tests*

The Kolmogorov–Smirnov test (KS test) (Press 1986; Knuth 1981) is probably the most widely used measure of the difference between two distributions. It is generally used for unbinned distributions. The Kuiper (KP test) statistic (Press 1986; Kuiper 1980) is a variant of the KS test.

From Table 4 it appears that 5 yr have a significantly different distribution than the climatology: 1980, 1981, 1982, 1986, and 1987. The KS test does not detect the significant difference of variance in 1985. The KS test is known to be good at finding changes in the median value, but it is not always so good at finding differences of spreads. The Kuiper test is a good tool for finding differences of spreads. From Table 4 the Kuiper test indicates significant difference only for 3 yr: 1982, 1986, and 1987. It seems to fail at detecting a difference in the mean between two distributions since the significance of this test for the years 1980 and 1981 is low.

### *d. Chi-square test*

The 9 members of the ensemble and the 9 yr (1980–88) define 81 events (72 events for the Southern Hemisphere basins). These 81 events are distributed over eight bins, which are defined such that this distribution is as uniform as possible. For example, over the western North Atlantic, the number of tropical storms ranges from 4 to 15. This range is divided into eight bins such that there are about 10 events in each of these bins. For each individual year, there are 9 events (the 9 members of the ensemble), which are distributed over these eight bins. If this distribution is indistinguishable from the 9-yr climatological distribution, there is no potential predictability. The Chi-square test is used to evaluate

TABLE 4. Significance of several statistical tests for potential predictability over the western Atlantic basin.

| | 1980 | 1981 | 1982 | 1983 | 1984 | 1985 | 1986 | 1987 | 1988 |
|---|---|---|---|---|---|---|---|---|---|
| F test | 0.4 | 0.5 | 0.985 | 0.71 | 0.46 | 0.92 | 0.956 | 0.952 | 0.79 |
| *T* test | 0.965 | 0.969 | 0.999 | 0.72 | 0.62 | 0.42 | 0.99 | 0.97 | 0.928 |
| KS test | 0.91 | 0.93 | 0.998 | 0.359 | 0.59 | 0.29 | 0.997 | 0.975 | 0.80 |
| KP test | 0.68 | 0.72 | 0.988 | 0.29 | 0.18 | 0.72 | 0.98 | 0.90 | 0.72 |
| Chi-square | 0.94 | 0.992 | 0.996 | 0.26 | 0.36 | 0.95 | 0.92 | 0.91 | 0.76 |

TABLE 5. Significance of the Chi-square test for each ocean basin and for each year.

| | 1980 | 1981 | 1982 | 1983 | 1984 | 1985 | 1986 | 1987 | 1988 |
|---|---|---|---|---|---|---|---|---|---|
| WNA | 0.94 | 0.992 | 0.996 | 0.26 | 0.36 | 0.95 | 0.92 | 0.91 | 0.76 |
| ENP | 0.72 | 0.95 | 1.0 | 0.98 | 0.70 | 0.98 | 0.999 | 0.999 | 0.999 |
| WNP | 0.14 | 0.992 | 0.996 | 0.999 | 0.915 | 0.85 | 0.987 | 0.991 | 0.999 |
| NI | 1.0 | 0.84 | 0.39 | 0.24 | 0.77 | 0.88 | 0.07 | 0.93 | 0.21 |
| SI | — | 0.34 | 0.44 | 0.89 | 0.83 | 0.91 | 0.90 | 0.98 | 0.37 |
| AUS | — | 0.98 | 0.99 | 0.87 | 0.28 | 0.04 | 0.56 | 0.26 | 0.99 |
| SP | — | 0.02 | 0.16 | 0.75 | 0.57 | 0.97 | 0.91 | 0.99 | 0.99 |
| Total | — | 0.17 | 0.23 | 0.87 | 0.29 | 0.60 | 0.97 | 0.84 | 0.994 |

whether the distribution for a given year is significantly different from the distribution for all 9 yr. The choice of eight bins is a compromise between using more bins to resolve additional details of the distribution and using fewer bins to increase the significance of the results. The results are not qualitatively sensitive to changes in the number of bins.

The results as presented in Table 4 indicate significant potential predictability in 1980, 1981, 1982, 1985, 1986, and 1987 over the western North Atlantic. These 6 yr include the years with different means and the years with different variances than the climatology. From all the statistical tests we used, this one indicates the largest number of potentially predictable years. The results obtained with this test are close to the one obtained with the KS test, except for 1985, for which the KS test failed to detect a difference between the 1985 distribution and the climatological distribution. That is probably because, as mentioned above, this test is not good at finding differences of spreads. This significant difference is detected only by the Chi-square test and by the measure of the spread. The Chi-square test also succeeds in detecting potential predictability in 1980 and 1981, while the KP test fails in detecting a difference of means with the climatological distribution. Therefore, the Chi-square test appears to be the most satisfactory test for potential predictability for discrete values. This conclusion is similar to one obtained by Best (1994). Therefore, we used this test to measure the potential predictability over all the basins. Table 5 displays the significance of this test over the eight basins we have defined.

The results displayed in Table 5 show that there is no specific year with strong potential predictability over all the basins. The western North Atlantic, western North Pacific, and eastern North Pacific display more years with strong potential predictability than the other basins. Over these three basins, the significance of the F test is larger than 90% during a majority of years. It can be concluded that the 9-member ensemble has a particularly strong potential predictability over the western North Atlantic, western North Pacific, and eastern North Pacific. On the other hand, the potential predictability seems to be particularly poor over the two Indian Oceans basins.

## 7. Simulation verification

The potential predictability tests presented in the previous section evaluate the ability of the simulation to give a prediction that is distinct from the model climatology, but they do not indicate whether this prediction is in agreement with observations. The goal of simulation verifications is to evaluate whether the model simulations and the observations are related.

The correlation between the ensemble mean and the observations gives an incomplete evaluation of skill in this ensemble simulation. The choice of the mean as the quantity to compare to observations is restrictive, and this test does not take into account the spread of the ensemble simulation for a given year.

Murphy and Winkler (1987) proposed a method for forecast verification. Denoting the forecast by $f$ and the observation by $x$, they suggest that the behavior of the probability distributions $P(x \mid f)$, $P(f \mid x)$, $P(f)$, and $P(x)$ gives us valuable information for forecast verification, where $P(x \mid f)$ represents the conditional probability distribution for a given forecast, $P(f \mid x)$ the conditional probability distribution for a given observation, $P(f)$ the marginal probability distribution of forecasts, and $P(x)$ the marginal probability distribution of observations. These four quantities are not completely independent since

$$P(x \mid f)P(f) = P(f \mid x)P(x) = P(x, f), \qquad (1)$$

where $P(x, f)$ represents the joint probability distribution. Over all the ocean basins defined previously, these four quantities are examined, with $f$ denoting a simulation instead of a forecast and applying to anomalies in the number of tropical storms instead of the total number of tropical storms. The tropical storm number anomaly is defined as the difference between the number of tropical storms and the climatological number of tropical storms. For tropical storms generated by the model the climatological number of tropical storms is the 9-yr ensemble mean of tropical storm numbers; for observed tropical storms the climatological number of tropical storms is the average number of observed tropical storms that occurred during the period 1980–88. The values of tropical storm number anomalies, which are real numbers, are rounded to the closest integer.

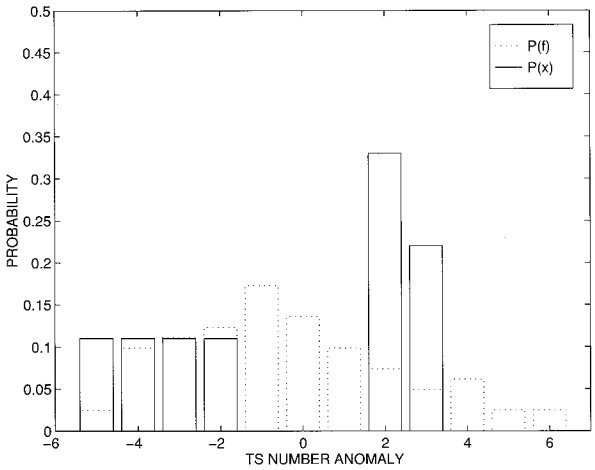


FIG. 11. Probability distributions of observed tropical storm number anomaly (base rate) for the period 1980–88 (solid line) and probability distribution of simulated tropical storm number anomaly (refinement) for the 9-yr ensemble (dashed line) over the western North Atlantic.

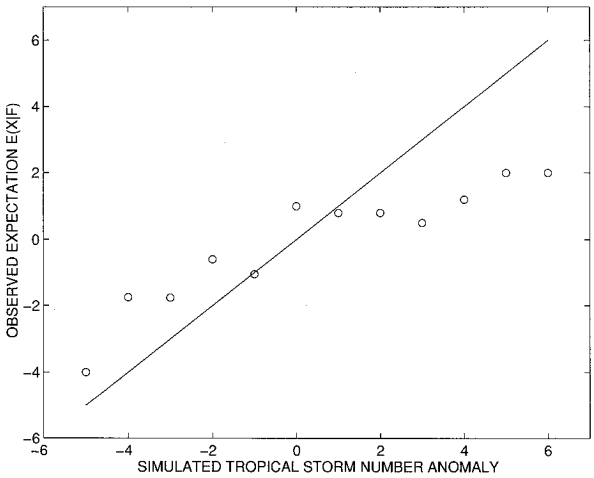


FIG. 12. The circles represent the expected value of observed tropical storm number anomaly for a given simulated tropical storm number anomaly over the western North Atlantic. The solid line represents the ideal case $E(x \mid f) = f$, where these two anomalies are equal.

### *a. Base rate P(x)*

The base rate is completely independent of the simulation since it is based only on observations. This probability distribution indicates how $x$ should be predicted in the absence of a model. Figure 11 displays the probability distribution of observed tropical storm number anomaly over the western North Atlantic (base rate) from 1980 to 1988. During these 9 yr there have been only 6 different anomaly values. The base rate indicates that over the western North Atlantic the probability of having a climatological number of tropical storms is small and that there is a tendency for each year to have either a large or a small number of tropical storms. The probability of having an anomaly of $+2$ or $+3$ is particularly high during this period. As for the other ocean basins (not shown), the base rate is particularly uniform over the Australian Basin, the South Pacific, the north Indian Ocean, and eastern North Pacific. Over the western North Pacific the observed anomalies have the smallest range during the 1980–88 period, and there is a probability larger than 0.5 of having 27 or 28 tropical storms per season. This result suggests that for the period 1980–88, the utility of a model tropical storm frequency simulation over the western North Pacific is low, since this frequency is almost constant.

### *b. Refinement P(f)*

The marginal distribution $P(f)$ tells us how sharp or refined the model is—if the model simulation always gives the climatological number of tropical storms, then it is considered to be nonrefined. This quantity is related to potential predictability since the model is said to be perfectly refined if it allows us to make a distinction between years of large anomalies and years of low anomalies. Refinement is completely independent of observations, but its comparison with the base rate gives some information about the behavior of the model and how simulations compare with observations.

Figure 11 displays the observed probability distribution (refinement) $P(f)$ over the western North Atlantic basin. The refinement $P(f)$ has been calculated from the 81 seasonal tropical storm numbers simulated by the 9-member ensemble over the western North Atlantic from 1980 to 1988. The shapes of $P(f)$ and $P(x)$ are different. The simulation probability distribution indicates that the most probable events over the western North Atlantic are the climatology and an anomaly of $-1$. There are indications of refinement over all basins, except over the north Indian Ocean (not shown).

### *c. Calibration P(x | f)*

The conditional distribution $P(x \mid f)$ indicates how often different observations have occurred for a particular simulated anomaly. This quantity relates to the calibration or reliability of the simulation. If the model is perfect, then $P(x \mid f)$ is equal to 1 for $x = f$ and equal to zero everywhere else. In other words, the model is perfectly calibrated if $E(x \mid f) = f$, where $E(x \mid f)$ is the expected value of $x$ given the simulated anomaly $f$, $E(x \mid f) = \Sigma_x P(x \mid f)x$.

The 9-member ensemble simulates 81 seasonal tropical storm number anomalies from 1980 to 1988. These 81 anomalies take on only 12 different values over the western North Atlantic and 31 over the western North Pacific. Figures 12 and 13 represent the expected value of observations for each of the different values defined by the 81 seasonal tropical storm number anomalies simulated by the GCM over the western North Atlantic and the western North Pacific, respectively. The diag-

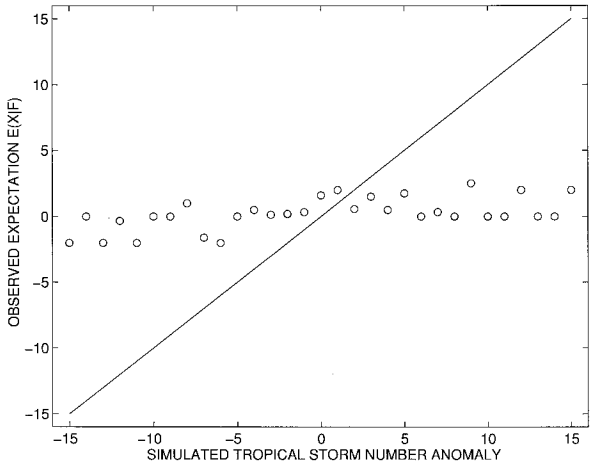


FIG. 13. The circles represent the expected value of observed tropical storm number anomaly for a given simulated tropical storm number anomaly over the western North Pacific. The solid line represents the ideal case $E(x \mid f) = f$, where these two anomalies are equal.

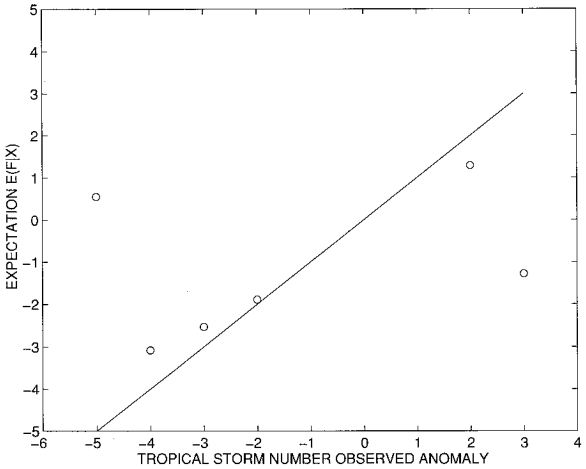


FIG. 14. The circles represent the expected value of simulated tropical storm number anomaly for a given observed tropical storm number anomaly over the western North Atlantic. The solid line represents the ideal case $E(f \mid x) = x$, where these two anomalies are equal.

onal line represents the perfect case, $E(x \mid f) = f$. Over the western North Atlantic $E(x \mid f)$ does not lie far from the perfect case; $E(x \mid f)$ is increasing with $f$ and lies further from $f$ for large anomaly predictions than for small anomaly predictions. Over the western North Atlantic the model can be considered to have some calibration skill.

Figure 13 indicates that the calibration over the western North Pacific is poor since $E(x \mid f)$ varies little with $f$ and lies far from the perfect case.

Over the eastern North Pacific (not shown), and to a lesser extent over the Australian Basin (not shown), the calibration is much higher than over all other basins. This may indicate that although the model climatology over the eastern North Pacific is far from the observed climatology, the model is able to simulate realistic tropical storm number anomalies over this basin.

Over the South Indian Ocean (not shown) and the South Pacific (not shown), $E(x \mid f)$ is a decreasing function of $f$, which is consistent with the negative correlation we obtained between the mean ensemble simulation and observations.

### d. Likelihood $P(f \mid x)$

The quantity $P(f \mid x)$ is the likelihood associated with the simulated anomaly $f$. It indicates how well the simulation discriminates among the various values of $x$. The simulation system is perfectly discriminatory if $P(f \mid x)$ is equal to zero for all values of $x$ except one.

The 9 observed tropical storm number anomalies from 1980 to 1988 took on only 6 different values over the western North Atlantic. Figure 14 represents the simulation expectation as a function of observation for each different observed tropical storm number anomaly over the western North Atlantic, and the solid line represents the ideal case $E(f \mid x) = x$; $E(f \mid x)$ is almost equal to $x$ for $x = -4, -3, -2$, and 2. For small amplitude anomalies the expectation is correct, but for extreme anomalies the expectation is wrong: $E(f \mid x = -5) = 0.55$ and $E(f \mid x = 3) = -1.28$. The eastern North Pacific (not shown) is the only other basin where $E(f \mid x)$ is close to $x$ for most values of $x$. Over the other basins the expectation $E(f \mid x)$ lies far from the ideal case, particularly over the north Indian (not shown) and the south Pacific Oceans (not shown).

### e. Ranked probability score

The ranked probability score (RPS) score described by Epstein (1969) is defined for each year and for each basin by the formula

$$S_j = \frac{3}{2} - \frac{1}{2(K-1)} \sum_{i=1}^{K-1} \left[ \left( \sum_{n=1}^{i} P_n \right)^2 + \left( \sum_{n=i+1}^{K} P_n \right)^2 \right] - \frac{1}{K-1} \sum_{i=1}^{K} |i-j| P_i. \quad (2)$$

Here, $K$ represents the total number of classes defined from the different values of simulated anomalies, $j$ represents the class in which the observed tropical storm number anomaly ranges, and $P_i$ represents the probability of a simulation for the $i$th class.

The RPS score ranges from 0 to 1, with 1 corresponding to perfect simulations ($P_j$ is equal to 1). When the probability distribution $P$ takes nonzero values far from the class where the observation occurs, then the last term of (2) becomes important and reduces the RPS score.

The RPS scores are presented in Table 6 for each season, and Table 7 displays the yearly average of this score for each ocean basin. The comparison of RPS scores from one basin to those from another basin is

TABLE 6. RPS score obtained over each ocean basin for each tropical storm season.

| | 1980 | 1981 | 1982 | 1983 | 1984 | 1985 | 1986 | 1987 | 1988 |
|---|---|---|---|---|---|---|---|---|---|
| WNA | 0.92 | 0.94 | 0.95 | 0.60 | 0.71 | 0.89 | 0.95 | 0.96 | 0.88 |
| ENP | 0.91 | 0.97 | 0.89 | 0.91 | 0.79 | 0.61 | 0.92 | 0.94 | 0.81 |
| WNP | 0.95 | 0.94 | 0.79 | 0.83 | 0.96 | 0.93 | 0.97 | 0.90 | 0.86 |
| NI | 0.55 | 0.80 | 0.94 | 0.92 | 0.92 | 0.82 | 0.88 | 0.49 | 0.91 |
| SI | — | 0.92 | 0.92 | 0.65 | 0.75 | 0.93 | 0.77 | 0.94 | 0.85 |
| AUS | — | 0.88 | 0.92 | 0.80 | 0.80 | 0.66 | 0.58 | 0.96 | 0.57 |
| SP | — | 0.74 | 0.81 | 0.81 | 0.71 | 0.89 | 0.83 | 0.70 | 0.85 |

difficult, since the number of classes and events are not the same. On the other hand, this score gives valuable indications about the comparison from one year to another for a given basin. Over the western North Atlantic basin, this score indicates the lowest simulation skill in 1983 and 1984 and the highest in 1981, 1982, 1986, and 1987. There is no special year where the simulation skill is particularly high or particularly low for all the basins, as was the case for potential predictability.

When producing an ensemble *forecast*, one knows the potential predictability a priori. A high correlation between potential predictability and forecast skill would be very useful. In other words, when the model is able to predict something, this prediction can be trusted. The relationship between potential predictability and simulation (forecast) skill may be measured by a linear correlation between the significance of the Chi-square test (Table 5) performed in section 6 and the RPS score (Table 6) for each basin. The results are displayed in Table 8. Over the western North Atlantic there is a very significant correlation between potential predictability and simulation skill. Over all other basins there is no significant correlation between these two quantities. This result plus the fact that potential predictability and simulation skill are very high for some years over the western North Atlantic suggests that when the model is able to predict something (as in 1982, 1986, or 1987) over this basin, then the result is very close to observations.

## 8. Conclusions

A 9-member ensemble of 10-yr integrations with a T42 atmospheric model forced by observed SSTs has been produced. Seven basins are defined, and the number of model tropical storms per basin and per year is evaluated by applying an objective procedure for tracking tropical storms. This objective procedure uses the presence of a warm core as a criterion of selection for model tropical storms.

The tropical storms simulated by this 9-member ensemble have a physical structure that is qualitatively similar to observed tropical storms with convergence at low levels, divergence above 300 mb, a warm core located at about 300 mb and a weak cold core above 100 mb, and very strong relative humidity near the center of the storm.

Results show that this T42 model simulates a tropical storm number climatology close to the observed over the Northern Hemisphere basins, except in the eastern North Pacific, where the model simulates considerably fewer tropical storms than are observed. Over the Southern Hemisphere basins the model creates tropical storms that are located too far eastward. For most basins, the most active month of the season is the same in the model as in the real world.

The potential predictability and the simulation skill depend on the basin. The best results with this model are obtained over the western North Atlantic, the eastern North Pacific, and the western North Pacific. There, the potential predictability is very high for most of the years. Over the western North Atlantic and western North Pacific, the climatological number of model tropical storms is very close to observed. The western North Atlantic basin is the only basin where years of very high potential predictability correspond to years of very high simulation skill, which is a useful property when producing true ensemble forecasts.

The years of largest potential predictability or simulation skill are not necessarily El Niño years, and we cannot immediately conclude that El Niño events play a dominant role in increasing potential predictability or the skill of this simulation over a specific basin.

This study indicates that the tropical storm number interannual variability simulated by the model is consistent with observations over the western North Atlantic, the eastern North Pacific, and the western North Pacific. This indicates that the large-scale mechanisms that regulate the model tropical storm number are consistent with observations in these three basins. Another

TABLE 7. Mean RPS score obtained over each ocean basin.

| WNA | ENP | WNP | NI | SI | AUS | SP |
|---|---|---|---|---|---|---|
| 0.87 | 0.87 | 0.90 | 0.80 | 0.84 | 0.77 | 0.79 |

TABLE 8. Linear correlation between potential predictability (Chi-square test significance) and predictive skill (RPS score).

| | WNA | ENP | WNP | NI | SI | AUS | SP |
|---|---|---|---|---|---|---|---|
| Correlation | 0.97 | 0.01 | −0.35 | −0.68 | −0.33 | 0.06 | 0.32 |
| Significance | 0.99 | 0.01 | 0.64 | 0.95 | 0.57 | 0.11 | 0.57 |

important point that can be deduced from this study concerns the role of SSTs in regulating tropical storm number interannual variability. The only common point between the 9 members of the ensemble is the observed SSTs. Therefore, the very strong potential predictability obtained over some of the basins and during some of the years proves that the SSTs play a major role in model tropical storm interannual variability.

Future plans include examining why the potential predictability and the simulation skill are not as high in Southern Hemisphere basins as over the western North Atlantic, the eastern North Pacific, and the western North Pacific. These ensemble simulations provide an excellent opportunity to investigate the influence of large-scale circulation on model tropical storm cyclogenesis. The cause of the large ensemble spread that occurs in certain years over certain basins (e.g., in 1980 over the western North Atlantic) will be studied, along with the mechanism by which the SSTs affect the interannual variability of model tropical storms.

This GCM ensemble, which used observed SSTs, was a simulation experiment, not a forecast experiment. To test the ability of the model to forecast seasonal tropical storm numbers over the western North Atlantic basin, future plans include integrations of this T42 model coupled to a realistic ocean model. Plans also include studies of tropical storm tracks and strength in the existing ensemble of simulations.

*Acknowledgments.* The authors would like to thank R. Tuleya, Dr. N. Lau, and A. Broccoli for their stimulating discussions and help, and two anonymous reviewers whose comments proved invaluable in improving the presentation of the material.

## REFERENCES

Anderson, L. J., and W. F. Stern, 1996: Evaluating the potential predictive utility of ensemble forecasts in a perfect model setting. *J. Climate,* **9,** 260–269.

Basher, R. E., and X. Zheng, 1995: Tropical cyclones in the southwest Pacific: Spatial patterns and relationships to Southern Oscillation and sea surface temperature. *J. Climate,* **8,** 1249–1260.

Bengtsson, L., H. Böttger, and M. Kanamitsu, 1982: Simulation of hurricane-type vortives in a general circulation model. *Tellus,* **34,** 440–457.

——, M. Botzet, and M. Esh, 1995: Hurricane-type vortices in a general circulation model. *Tellus,* **47A,** 175–196.

Best, D. J., 1994: Nonparametric comparison of two histograms. *Biometrics,* **50,** 538–541.

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.

Cayan, D. R., and R. H. Webb, 1992: El-Niño–Southern Oscillation and streamflow in the western United States. *El Niño, Historical and Paleoclimatic Aspects of the Southern Oscillation,* H. F. Diaz and V. Markgraf, Eds., Cambridge University Press, 29–68.

Chan, J. C. L., 1985: Tropical cyclone activity in the northwest Pacific in relation to the El-Niño/Southern Oscillation phenomenon. *Mon. Wea. Rev.,* **113,** 599–606.

Chervin, R. M., 1986: Interannual variability and seasonal climate predictability. *J. Atmos. Sci.,* **43,** 233–251.

Emanuel, K. A., 1988: Polar lows as arctic hurricanes. *Tellus,* **41A,** 1–17.

Epstein, E. S., 1969: A scoring system for probability forecasts of ranked categories. *J. Appl. Meteor.,* **8,** 985–993.

Evans, J. L., 1992: Comment on ''Can existing climate models be used to study anthropogenic changes in tropical cyclone climate?'' *Geophys. Res. Lett.,* **19,** 1523–1524.

Frank, W. M., 1977: The structure and energetics of the tropical cyclones. I: Storm structure. *Mon. Wea. Rev.,* **105,** 1119–1135.

Gates, W., 1992: AMIP: The Atmospheric Model Intercomparison Project. *Bull. Amer. Meteor. Soc.,* **73,** 1962–1970.

Gordon, C. T., 1992: Comparison of 30-day integrations with and without cloud–radiation interaction. *Mon. Wea. Rev.,* **120,** 1244–1277.

——, and W. F. Stern, 1982: A description of the GFDL global spectral model. *Mon. Wea. Rev.,* **110,** 625–644.

Gray, W. M., 1979: Hurricanes: Their formation, structure and likely role in the tropical circulation. *Meteorology over the Tropical Oceans,* D. B. Shaw, Ed., Roy. Meteor. Soc., 155–218.

——, 1984: Atlantic seasonal hurricane frequency. Part I: El Niño and 30 mb quasi-biennial oscillation influences. *Mon. Wea. Rev.,* **112,** 1649–1668.

——, C. W. Landsea, P. W. Mielke Jr., and K. J. Berry, 1992: Predicting Atlantic basin seasonal hurricane activity 6–11 months in advance. *Wea. Forecasting,* **7,** 440–455.

——, ——, ——, and ——, 1993: Predicting Atlantic basin seasonal tropical cyclones activity by 1 August. *Wea. Forecasting,* **8,** 73–86.

——, ——, ——, and ——, 1994: Predicting Atlantic basin seasonal tropical cyclones activity by 1 June. *Wea. Forecasting,* **9,** 103–115.

Haarsma, R. J., J. F. B. Mitchell, and C. A. Senior, 1993: Tropical disturbances in a GCM. *Climate Dyn.,* **8,** 247–257.

Hayashi, Y., 1986: Statistical interpretations of ensemble-time mean predictability. *J. Meteor. Soc. Japan,* **64,** 167–181.

Hess, J. C., J. B. Elsner, and N. E. LaSeur, 1995: Improving seasonal hurricane predictions for the Atlantic basin. *Wea. Forecasting,* **10,** 425–432.

Knuth, D. E., 1981: *Seminumerical Algorithms.* Vol. 2, *The Art of Computer Programming,* Addison-Wesley, 688 pp.

Krishnamurti, 1988: Some recent results on numerical weather prediction over the tropics. *Aust. Meteor. Mag.,* **36,** 141–170.

Kuiper, N. H., 1960: Texts concerning random points on a circle. *Proc. K. Ned. Akad. Wet. Ser. A,* **63,** 38–47.

Lander, M. A., 1993: Comments on ''A GCM simulation of the relationship between tropical storm formation and ENSO.'' *Mon. Wea. Rev.,* **121,** 2137–2146.

Lau, K.-H., 1991: An observational study of tropical summertime synoptic scale disturbances. Ph.D. dissertation, Princeton University, 243 pp.

Leith, C. E., 1974: Theoritical skill of Monte Carlo forecasts. *Mon. Wea. Rev.,* **102,** 409–418.

Madden, R. A., 1976: Estimates of the natural variability of time-averaged sea-level pressure. *Mon. Wea. Rev.,* **104,** 942–952.

——, 1981: A quantitative approach to long-range prediction. *J. Geophys. Res.,* **86,** 9817–9825.

Manabe, S., 1969: Climate and the ocean circulation. I. The atmospheric circulation and the hydrology of the earth's surface. *Mon. Wea. Rev.,* **97,** 739–774.

——, J. Smagorinsky, and R. F. Strickler, 1965: Simulated climatology of a general circulation model with a hydrologic cycle. *Mon. Wea. Rev.,* **93,** 769–798.

——, 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.

McBride, J. L., 1984: Comments on ''Simulation of hurricane-type vortices in a general circulation model.'' *Tellus,* **36A,** 92–93.

——, 1987: The Australian summer monsoon. *Reviews of Monsoon Meteorology,* C. P. Chang and T. N. Krishnamurti, Eds., Oxford University Press, 203–231.

Murphy, A. H., and R. L. Winkler, 1987: A general framework for forecast verification. *Mon. Wea. Rev.*, **115,** 1330–1338.
Murphy, J. M., 1988: The impact of ensemble forecasts on predictability. *Quart. J. Roy. Meteor. Soc.*, **114,** 463–493.
Murray, R. J., and I. Simmonds, 1991: A numerical scheme for tracking cyclone centres from digital data. Part I: Development and operation of the scheme. *Aust. Meteor. Mag.*, **39,** 155–166.
Navarra, A., W. F. Stern, and K. Miyakoda, 1994: Reduction of the Gibbs oscillation in spectral model simulations. *J. Climate,* **7,** 1169–1183.
Nicholls, N., 1992: Recent performance of a method for forecasting Australian seasonal tropical cyclone activity. *Aust. Meteor. Mag.,* **40,** 105–110.
Pan, Y., 1981: The effect of the thermal state of eastern equatorial Pacific on the frequency of typhoons over western Pacific (in Chinese). *Acta Meteor. Sin.,* **40,** 24–32.
Press, W. R., 1986: *Numerical Recipes: The Art of Scientific Computing.* Cambridge University Press, 102 pp.
Rasmussen, E., 1985: A case study of a polar low development over the Barent Sea. *Tellus,* **37A,** 407–418.
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.
Shukla, J., 1981: Dynamical predictability of monthly means. *J. Atmos. Sci.,* **38,** 2547–2572.
Sirutis, J., and K. Miyakoda, 1990: Subgrid scale physics in 1-month forecasts. Part I: Experiment with four parameterization packages. *Mon. Wea. Rev.,* **118,** 1043–1064.
Sperber, K. R., and T. N. Palmer, 1996: Interannual tropical rainfall variability in general circulation model simulations associated with the Atmospheric Model Intercomparison Project. *J. Climate*, **9,** 2727–2750.
Stern, W., and R. Pierrehumbert, 1988: The impact of orographic gravity wave drag parameterization on extended range predictions with a GCM. Preprints, *Eighth Conf. on Numerical Weather Prediction,* Baltimore, MD, Amer. Meteor. Soc., 745–750.
——, and K. Miyakoda, 1995: The feasibility of seasonal forecasts inferred from multiple GCM simulations. *J. Climate,* **8,** 1071–1085.
Tiedtke, M., 1988: Parameterization of cumulus convection in large-scale models. *Physically-Based Modeling and Simulation of Climate and Climate Change,* M. Schlesinger. Ed., D. Reidel, 375–431.
Trenberth, K. E., 1992: Global analyses from ECMWF and atlas of 1000 to 10 mb circulation statistics. NCAR Tech. Note 373+STR, 191 pp.
Tsutsui, J.-I., A. Kasahara, and H. Hirakuchi, 1995: Identification of tropical cyclones in long-term simulations with NCAR Community Climate Model (CCM2). Preprints, *21st Conf. on Hurricanes,* Miami, FL, Amer. Meteor. Soc., 375–377.
Tuleya, R., 1994: Tropical storm development and decay: Sensitivity to surface boundary conditions. *Mon. Wea. Rev.,* **122,** 291–304.
Watterson, I. G., J. L Evans, and B. F. Ryan, 1995: Seasonal and interannual variability of tropical cyclogenesis: Diagnostics from large-scale fields. *J. Climate,* **8,** 3042–3066.
Weatherford, C., and W. M. Gray, 1988: Typhoon structure as revealed by aircraft reconnaissance. Part II: Structural variability. *Mon. Wea. Rev.*, **116,** 1044–1056.
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.