

Changes of Probability Associated with El Niño

PRASHANT D. SARDESHMUKH, GILBERT P. COMPO, AND CÉCILE PENLAND

NOAA-CIRES Climate Diagnostics Center, University of Colorado, Boulder, Colorado

(Manuscript received 2 September 1999, in final form 15 February 2000)

ABSTRACT

Away from the tropical Pacific Ocean, an ENSO event is associated with relatively minor changes of the probability distributions of atmospheric variables. It is nonetheless important to estimate the changes accurately for each ENSO event, because even small changes of means and variances can imply large changes of the likelihood of extreme values. The mean signals are not strictly symmetric with respect to El Niño and La Niña. They also depend upon the unique aspects of the SST anomaly patterns for each event. As for changes of variance and higher moments, little is known at present. This is a concern especially for precipitation, whose distribution is strongly skewed in areas of mean tropospheric descent.

These issues are examined here in observations and GCM simulations of the northern winter (January–March, JFM). For the observational analysis, the 42-yr (1958–99) reanalysis data generated at NCEP are stratified into neutral, El Niño, and La Niña winters. The GCM analysis is based on NCEP atmospheric GCM runs made with prescribed seasonally evolving SSTs for neutral, warm, and cold ENSO conditions. A large number (180) of seasonal integrations, differing only in initial atmospheric states, are made each for observed climatological mean JFM SSTs, the SSTs for an observed warm event (JFM 1987), and the SSTs for an observed cold event (JFM 1989). With such a large ensemble, the changes of probability even in regions not usually associated with strong ENSO signals are ascertained.

The results suggest a substantial asymmetry in the remote response to El Niño and La Niña, not only in the mean but also the variability. In general the remote seasonal mean geopotential height response in the El Niño experiment is stronger, but also *more* variable, than in the La Niña experiment. One implication of this result is that seasonal extratropical anomalies may not necessarily be more predictable during El Niño than La Niña. The stronger seasonal extratropical variability during El Niño is suggested to arise partly in response to stronger variability of rainfall over the central equatorial Pacific Ocean. The changes of extratropical variability in these experiments are large enough to affect substantially the risks of extreme seasonal anomalies in many regions. These and other results confirm that the remote impacts of individual tropical ENSO events can deviate substantially from historical composite El Niño and La Niña signals. They also highlight the necessity of generating much larger GCM ensembles than has traditionally been done to estimate reliably the changes to the full probability distribution, and especially the altered risks of extreme anomalies, during those events.

1. Introduction

The coupling of the atmosphere and ocean in the tropical Pacific basin gives rise to the quasi-regular interannual El Niño–Southern Oscillation (ENSO) whose effects are felt around the globe. Several elements of the phenomenon, such as the sea surface temperature (SST) anomalies in the eastern equatorial Pacific, can now be predicted two or more seasons in advance. This raises the hope that at least some aspects of interannual variability may also be predictable in other regions at similar lead times. This hope is being pursued in the climate modeling community to improve seasonal predictions in both the Tropics and extratropics. Unfortu-

nately, the ENSO signal is small in the extratropics (e.g., Horel and Wallace 1981; Lau 1985; Barnston 1993; Livezey and Mo 1987; Palmer and Anderson 1994; Barnett 1995; Stern and Miyakoda 1995; and many others). Even so, it is perhaps the strongest *predictable* interannual signal there, and the question is how best to use it in seasonal predictions.

This paper was originally motivated by the fact that even a small ENSO signal-to-noise ratio of 0.5, while not large enough to affect appreciably the expected seasonal mean of an extratropical variable x , can still greatly affect the probability of extreme values of x . This follows from elementary considerations of the normal probability density function (PDF) in Fig. 1. Figure 1a illustrates a scenario in which ENSO results in, say, a shift of 0.5 (in units of the standard deviation) of the PDF of x to the right. Even though the shift is “small,” the likelihood (or risk) of obtaining an extreme positive value of x , defined here as greater than $+1$, increases

Corresponding author address: Dr. Prashant D. Sardeshmukh, NOAA-CIRES Climate Diagnostics Center, Mail Code R/CDC, 325 Broadway, Boulder CO 80303-3328.
E-mail: pds@cdc.noaa.gov

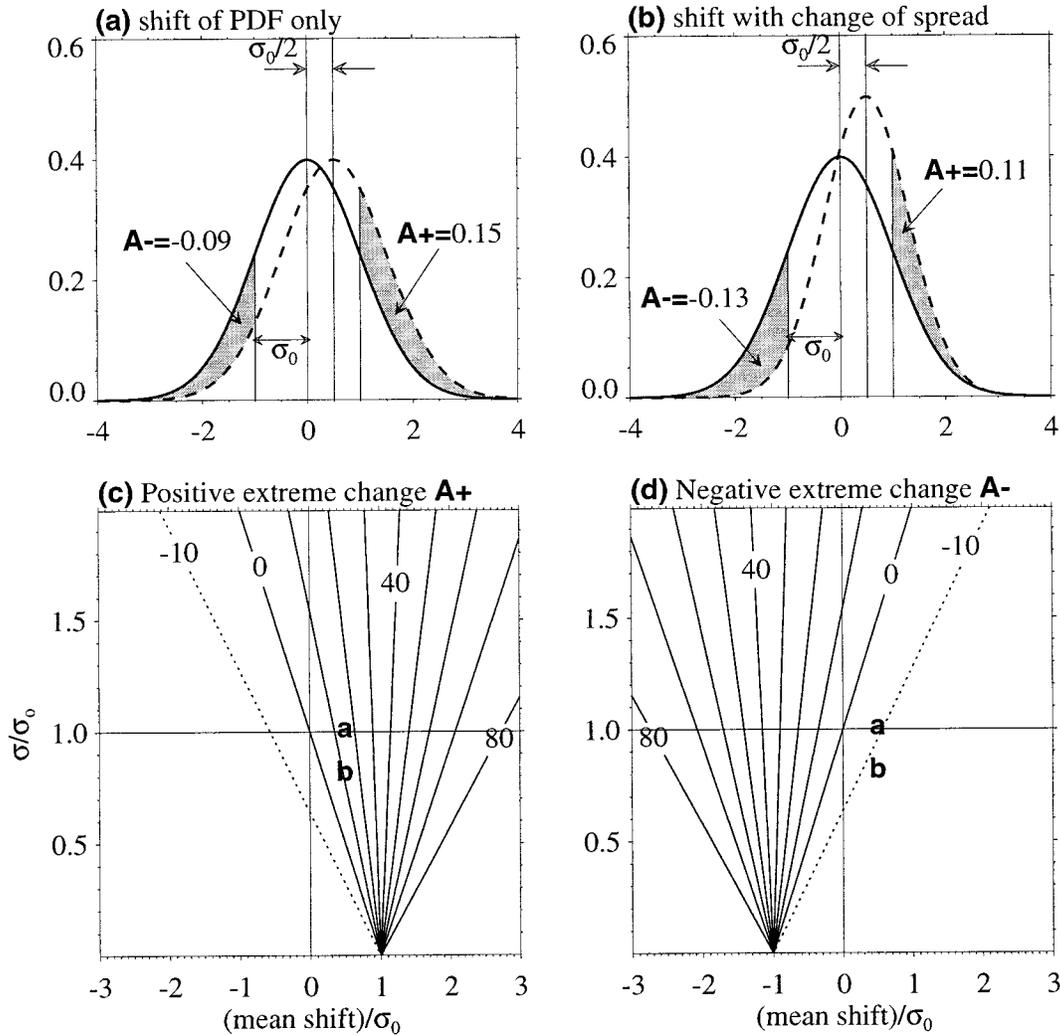


FIG. 1. (a) Normal probability density function (PDF) $N(0, 1)$ (solid curve) and $N(0.5, 1)$ (dashed curve) with means 0 and 0.5, respectively, and standard deviations 1.0. Shaded areas indicate difference in probability of extreme values greater than +1 (=A+) and difference in probability of extreme values less than -1 (=A-). (b) Same as (a) but for PDFs $N(0, 1)$ and $N(0.5, 0.8^2)$. (c) Here A+ is expressed as a percentage for normal distributions $N(\mu_0, \sigma_0^2)$ and $N(\mu, \sigma^2)$. The symbols "a" and "b" on the plot refer to the cases illustrated in the upper two panels. (d) As in (c) but for A-.

from 16% to 31%. Similarly, the likelihood of obtaining an extreme negative value of less than -1 decreases from 16% to 7%. Thus without ENSO the risks of extreme positive and negative anomalies are the same, but with ENSO the risk of an extreme positive anomaly becomes $31/7 = 4.4$ times the risk of an extreme negative anomaly.

When even small PDF shifts imply large changes of the likelihood of extremes, determining those shifts accurately becomes important. For this reason it is also important to determine how ENSO affects the extratropical noise, represented in Fig. 1a by the spread of the PDF. Figure 1b depicts another scenario in which the ENSO signal (i.e., the shift) is the same as in Fig. 1a, but the noise is reduced from 1 to 0.8. The risk of an extreme positive anomaly is now 27%, while that of

an extreme negative anomaly is 3%, giving an altered positive/negative risk ratio of $27/3 = 9$ that is more than double the ratio of 4.4 for the scenario of Fig. 1a. The remaining two panels of Fig. 1 show the changes of risk of extreme positive (Fig. 1c) and extreme negative (Fig. 1d) anomalies for other combinations of shift and spread. A wide range of possibilities is depicted in Figs. 1c and 1d. Note that for small shifts any reduction of spread implies a *reduced* risk of both positive and negative extremes. Clearly, knowledge of not just the mean shift but the change of the entire distribution is necessary if the altered risks are to be estimated reliably.

One might think of approaching this problem empirically, using observations of the past 50+ winters. For example, one might fit a linear regression model to the observations and use it to predict the shift of the PDF

during an individual ENSO event. The noise might be estimated as the standard error of the regression fit. Unfortunately, the observational record is not long enough for the regression parameters to be estimated with sufficient accuracy for our purpose. Given the small ENSO signal-to-noise ratios in the extratropics and relatively few (50+) seasonal samples, it is not hard to see how relatively small uncertainties in the estimated parameters would translate into unacceptably large uncertainties in the predicted risks of extremes. Such an empirical approach would also assume that (i) the extratropical response to ENSO is linear, (ii) ENSO does not affect the extratropical noise, and (iii) all the relevant PDFs are Gaussian. This scenario is essentially that of Fig. 1a, but with the added simplification that the shifts are linear with respect to ENSO forcing: El Niño events cause a shift one way, La Niña the other, and strong events cause proportionately larger shifts than weak ones. It is important to ascertain the extent to which such assumptions are valid, since they have a large bearing on the problem at hand.

The obvious alternative is to approach the problem with a dynamical numerical model. It is reasonable to expect general circulation models (GCMs), with their comprehensive nonlinear dynamics and also with many more degrees of freedom than can be treated adequately in the limited observational record, to perform much better than empirical models. Nonetheless, two factors work against the fulfillment of such expectations. First, in a chaotic atmosphere with a deterministic predictability limit of about two weeks, the problem of predicting seasonal averages even using deterministic GCMs becomes a probabilistic one. As numerous modeling groups have found, an ensemble of seasonal mean atmospheric GCM forecasts made with the same prescribed anomalous SST boundary forcing but with slightly different initial conditions has a spread that is comparable to the climatological standard deviation of observed seasonal means. This chaos is associated with unpredictable nonlinear interactions in the atmosphere. Only the statistical properties of the ensemble are meaningful in such a situation, not individual ensemble members. The best one can do is to issue the ensemble mean as an "expected value" forecast, and interpret the ensemble spread as indicative of the reliability of that forecast. This presentation format is similar to that of the empirical models of "expected value" and "standard error." To the extent that the nonlinear interactions are unpredictable (or parameterizable on seasonal scales as linear terms plus unpredictable noise), a GCM's nonlinear advantage over an empirical model is then lost. For it to have an advantage, a GCM must nonlinearly affect the ensemble mean and spread of atmospheric variables in response to anomalous tropical SST forcing. The issue is not yet settled (e.g., Peng et al. 2000) and provides another motivation for this paper.

Second, although a GCM can treat many more degrees of freedom than is feasible in an empirical frame-

work, that advantage may be more apparent than real. A large part of interannual tropical Pacific SST variability can be described with about 20 degrees of freedom (see, e.g., Penland and Sardeshmukh 1995), and possibly an even smaller subset is important in tropical-extratropical interactions. There is evidence that the extratropical atmosphere is sensitive only to large-scale features of the SST anomaly fields and not to their details (see, e.g., Barsugli et al. 1997). This raises the possibility that even if the tropical SST anomaly patterns differ considerably between ENSO events, their impact on the extratropical circulation may not. In such a situation a relatively small number of empirical regression parameters may suffice, and may be adequately estimated from the available observations. If true, this too would undercut a GCM's advantage.

The apparently large advantages of a GCM over simple empirical linear regressions (or other linear statistical methods) in predicting ENSO-related global signals, and especially the altered risks of extreme anomalies, may thus be limited even in principle. In practice, they are reduced further by errors in the representation of tropical diabatic heating and the associated divergent circulation that continue to plague GCMs (Gates et al. 1999). There is already evidence suggesting that seasonal GCM forecasts are not unambiguously superior to empirical forecasts based on the simplest historical ENSO composites (Fig. 2 of Kumar et al. 1996; see also Peng et al. 2000).

With these considerations in mind, we will first estimate here, using past observations, the altered PDFs of global 500-mb height and precipitation for an individual El Niño winter [January–March (JFM) 1987] and an individual La Niña winter (JFM 1989) given the historical El Niño and La Niña composite anomaly fields. The composites will be derived with respect to observed SST anomalies in the Niño-3.4 area of the eastern equatorial Pacific. All anomalies will be defined as departures from the means of "neutral" ENSO winters. This will give us 10 El Niño, 10 neutral, and 10 La Niña winters in the 42-yr record from which to construct the composites. In the context of Fig. 1, our empirically predicted shifts for 1987 and 1989 will consist of adjusting the amplitudes of these historical composite fields to be consistent with those of the Niño-3.4 SST anomalies in 1987 and 1989, respectively. Note that in view of the limited samples, we will only be able to estimate shifts of the mean, but not the changes of spread, by this approach.

We will then ask to what extent an atmospheric GCM (AGCM) would predict substantially different PDFs for these two winters. Using the NCEP atmospheric GCM for this purpose, we will estimate the PDFs for JFM 87 and JFM 89 from large ensembles (with 180 members each) of seasonal integrations made with prescribed seasonally evolving observed global SSTs for these two winters. We will estimate the climatological PDF as-

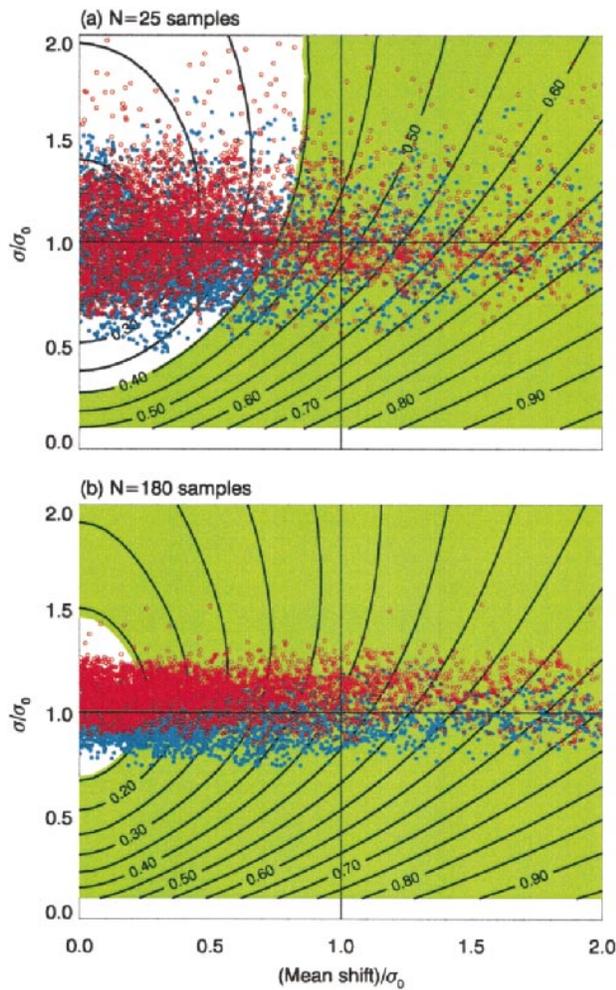


FIG. 2. Contours indicate the expected value of D_{ks} between sampled normal distributions $N(\mu_0, \sigma_0^2)$ and $N(\mu, \sigma^2)$, based on 50 000 Monte Carlo trials of (a) 25 and (b) 180 samples. Green shading indicates values significant at the 5% level, that is, the sample $N(\mu, \sigma^2)$ distribution can be distinguished from the sample $N(\mu_0, \sigma_0^2)$ distribution. Scattered points represent the magnitudes of the mean and standard deviation of JFM-mean 500-mb heights in the El Niño (red) and La Niña (blue) GCM ensembles relative to those in the neutral ensemble, (a) using only the first 25 members of the 180 available members in the GCM ensembles, and (b) using all 180 members.

sociated with the neutral winters through a parallel set of 180 integrations made with climatological JFM SSTs.

The distinguishing feature of these calculations is the very large number of AGCM ensemble members generated for SST forcing corresponding to an observed warm and an observed cold ENSO event. As will be shown below, one needs large ensembles to generate reliable estimates of the ENSO-induced changes of spread. Also, the aim here is to determine not just that there are statistically significant remote ENSO signals, but to what extent those signals might differ from event to event and also between El Niño and La Niña events. Large ensembles are necessary for a reliable estimation

of such formally second-order effects. With the computational constraints of generating large GCM ensembles in mind, we will interpret the difference between our single-event El Niño and *sign-reversed* La Niña signals as roughly indicative of the range of signal variation to expect between events, in addition to an asymmetry of the response to El Niño and La Niña.

Our focus in this paper will thus be on clarifying (i) the extent to which a GCM's predicted response (i.e., the PDF shifts) can differ for individual El Niño and La Niña events and also from empirically predicted shifts, (ii) whether there are appreciable changes of spread associated with ENSO, and if so, (iii) whether those changes differ in character for El Niño and La Niña. The answers to these questions have implications not only for the predictability of extreme risks but also of the seasonal averages themselves during ENSO events.

The paper is organized as follows. Because the relevant PDFs are in general not Gaussian, we discuss assessing the differences between them in terms of a nonparametric Kolmogorov–Smirnov measure in section 2. The important issue of sampling, that is, the number of ensemble members required to claim statistically significant differences between two PDFs, is also considered there. Section 3 describes the observational and GCM datasets used in the study. Section 4 compares the PDF shifts for JFM 87 and JFM 89 predicted by the empirical and dynamical methods and presents, perhaps for the first time in the literature, the changes of spread of seasonal means predicted by the dynamical method. Section 5 addresses special problems associated with assessing changes of precipitation PDFs, which can be substantially non-Gaussian. Section 6 discusses some potentially important asymmetries in the PDF changes for El Niño and La Niña. Section 7 speculates on the link between the changes of spread of the 500-mb height PDFs and those of the tropical precipitation PDFs. The implications of these changes of spread for the risks of extreme seasonal anomalies and the predictability of seasonal averages are discussed in sections 8 and 9, and concluding remarks are made in section 10.

2. Assessing differences between PDFs

The Student's t and Fisher's F tests are standard tests for assessing shifts of the mean and changes of variance, respectively, and we will use them where appropriate. For convenience we will consider an equal number of seasonal members N in our neutral, El Niño, and La Niña ensembles, with $N = 10$ for observations and $N = 180$ for the GCM. When looking at the results, it will be useful to keep the following rules of thumb in mind for the 5% significance level. (Other significance levels will have different constant factors in the expressions below, but the same form.) To establish the significance of a shift \bar{x} of the mean (in units of the sample standard deviation s_0 of the original PDF), the number of mem-

TABLE 1. Northern Hemisphere winters (Dec–Apr) classified as El Niño, La Niña, and neutral winters in this paper. Year in the table refers to that of the January of the event listed. The number below the year is the intensity of the SST index in the El Niño-3.4 area defined in the text.

El Niño:	1983, 1.70,	1998, 1.68,	1992, 1.30,	1958, 1.07,	1973, 0.89,	1966, 0.87,	1969, 0.86,	1995, 0.79,	1987, 0.76,	1993, 0.46
La Niña:	1974, −1.64,	1989, −1.39,	1976, −1.39,	1971, −1.34,	1999, −1.24,	1985, −0.79,	1975, −0.74,	1968, −0.64,	1965, −0.53,	1967, −0.51
Neutral:	1959, 0.29,	1978, 0.29,	1964, 0.27,	1980, 0.26,	1990, 0.15,	1982, 0.09,	1979, −0.07,	1961, −0.08,	1997, −0.14,	1960, −0.16

bers drawn from each distribution should exceed approximately

$$N > \sim 8/\bar{x}^2 \quad (1)$$

according to the t test. On the other hand, to establish that the sample standard deviation s is significantly different from 1 (again in units of s_0), it is easy to show that the number of members should exceed approximately

$$N > \sim 3/(s - 1)^2 \quad (2)$$

according to the F test. This approximation is excellent for $s \sim 1$. To get a feel for these numbers, consider again the $\bar{x} = 0.5$, $\sigma = 0.8$ scenario of Fig. 1. Equation (1) then states that N should exceed 32 to establish the shift of the mean, whereas Eq. (2) states that N should exceed 75 to establish the change of spread.

We will also assess our PDF changes in terms of the nonparametric two-sample Kolmogorov–Smirnov measure D_{ks} (see appendix A), defined as the *magnitude of the maximum difference* between the corresponding cumulative distribution functions (CDFs). Note that D_{ks} lies between 0 and 1, with larger values indicating progressively more different PDFs. The contours in Fig. 2 show the expected value of D_{ks} estimated from $N = 25$ (upper panel) and $N = 180$ (lower panel) members drawn from two Gaussian PDFs $N(0, 1)$ and $N(\bar{x}, \sigma^2)$ for a range of values of \bar{x} and σ . The shading indicates values significant at the 5% level, that is, higher than a value that could be obtained by chance for two sets of N members drawn from the same distribution. With 25 members only values greater than 0.4 are significant, whereas with 180 members, values as small as 0.15 are significant. The scattered dots highlight the importance of the sampling issue in assessing the significance of ENSO-induced PDF changes in the atmosphere. The dots in the lower panel represent the magnitude of the 180-member ensemble mean anomalies and standard deviations of our GCM-predicted 500-mb heights for JFM 1987 and 1989 relative to those of the climatological JFM 500-mb heights at (nearly) all model grid points. (The excluded points are in the Tropics where the shifts are greater than 2.0.) The dots in the upper panel were derived using only the first 25 of the 180 members. The scatter in the upper panel is much larger, and a much larger fraction (51%) of points lies in the statistically insignificant white area than in the lower panel (19%).

The dots in Fig. 2 illustrate the well-known fact that

away from the Tropics the PDF changes associated with ENSO are minor. They also highlight the requirement of large samples to establish the statistical significance of those PDF changes. The $N \sim 25$ (or fewer) members available in the observational record are simply not enough in this regard. The $N \sim 25$ (or fewer) members generated in many GCM studies for *individual* ENSO events are also not enough.

3. Observational and GCM datasets

For brevity we will focus on 500-mb height and precipitation fields in this paper. The observational height fields were obtained from the 42-yr (1958–99) NCEP–NCAR reanalysis dataset (Kalnay et al. 1996). An ENSO SST index, derived from the GISST-2.3b SST dataset (Parker et al. 1995), was defined as the area-averaged SST anomaly in the Niño-3.4 area 5.5°N–5.5°S, 170°–120°W. The anomalies were constructed by removing a least-squares fit to the first three harmonics of the annual cycle and smoothing with a 5-month running mean filter (Trenberth 1997). The SST index was then ranked, and winters (December–April) with the highest 10, lowest 10, and middle 10 ranks were classified as El Niño, La Niña, and neutral winters, respectively (see Table 1). Composites were constructed as averages over the 10 cases, and finally the composite anomalies for El Niño and La Niña were determined as departures from the neutral composite. The composite SST indices for El Niño and La Niña determined in this manner were +1.04° and −1.02°C, respectively. Note that since the standard errors for these were $\pm 0.30^\circ$ and $\pm 0.31^\circ$ C, the composite SST index *amplitudes* were not statistically different from each other at the 5% level according to the t test.

The SST indices for JFM 1987 and 1989 were +0.76 and −1.39. Thus, according to this measure, the 1987 El Niño event was “moderate” and the 1989 La Niña event was “strong.” Our empirical estimates of the 500-mb height ENSO signals in these two winters were then obtained by multiplying the El Niño and La Niña composite anomaly fields by the appropriate scale factors: 0.73 (=0.76/1.04) for JFM 1987 and 1.36 (=1.39/1.02) for JFM 1989.

Similar procedures were applied to the observational precipitation fields in the 20-yr (1979–98) Xie–Arkin dataset (Xie and Arkin 1997). The National Centers for Environmental Prediction (NCEP) reanalysis precipi-

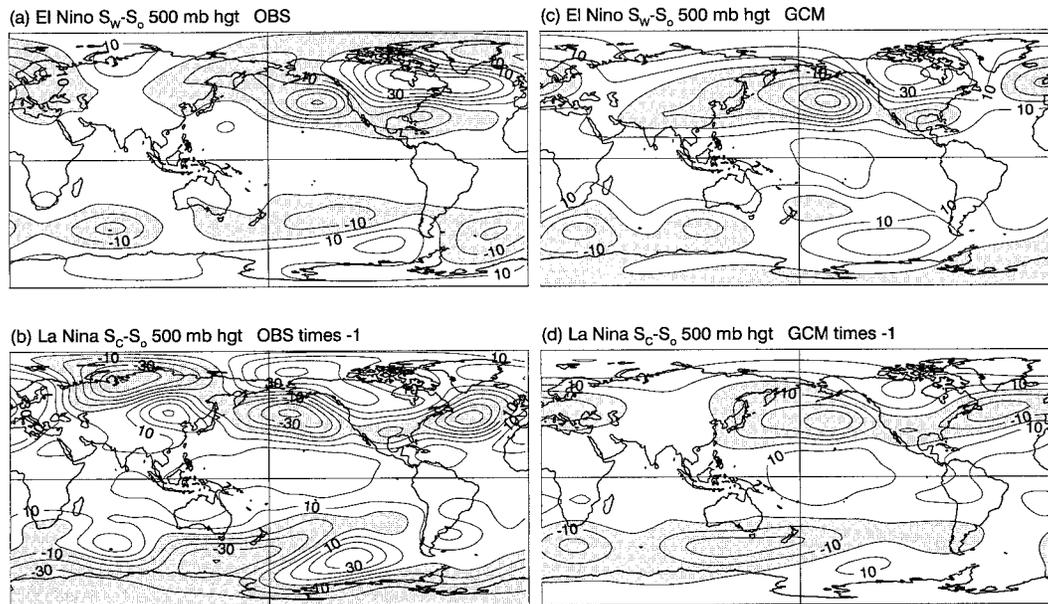


FIG. 3. The ENSO signal in JFM-mean 500-mb heights during the (top) 1987 El Niño and (bottom) 1989 La Niña estimated from the (left) historical observations and the (right) NCEP GCM ensembles. The observational fields were obtained by first forming 10-event composite fields for El Niño and La Niña in the 1958–99 NCEP reanalysis dataset, and then adjusting their amplitudes to account for the relatively weak 1987 El Niño and relatively strong 1989 La Niña events, respectively. The AGCM fields represent the average of 180 seasonal integrations with prescribed global SST fields for JFM 1987 and JFM 1989. Contour interval in all panels is 10 m and negative values are shaded. Note that the La Niña fields are shown with their signs reversed to enable a quick grasp of their similarity to the El Niño fields. See text for further details.

tation data were not used in this study because of their relatively low interannual variability over most of the globe compared to the Xie–Arkin estimates (see Trenberth and Guillemot 1996). This shorter 20-yr dataset was also stratified into El Niño, La Niña, and neutral sets based on the SST index, but now with only five members each.

The NCEP atmospheric GCM used was identical to that employed by Kumar et al. (1996) and Chen and van den Dool (1997). The model has a spatial discretization of T40 in the horizontal (about 3° lat \times 3° long) and 18 sigma (normalized pressure) levels in the vertical. It differs from earlier versions of the NCEP model (Kanamitsu et al. 1991) chiefly in a closer prescribed link between SST and the onset of deep convection in the convection scheme, and in its restriction of the radiative effects of convective clouds to deep clouds in the radiation scheme (see Kumar et al. 1996). Seasonal GCM integrations were made with observed global climatological JFM SSTs and observed global SSTs for JFM 87 and JFM 89. Ensembles of 180 integrations, differing only in initial atmospheric states, were made for each of these SST boundary conditions. The initial atmospheric conditions were drawn from a separate 50-yr seasonal cycle run made with climatological SSTs as boundary forcing.

Finally, ensemble means and spreads were determined for each of the three 180-member ensembles.

Note that the term “spread” here refers to the standard deviation of the 180 simulated seasonal (JFM) averages. For all model fields, the JFM 1987 (1989) ENSO signal was defined as the JFM 1987 (1989) ensemble mean minus the climatological JFM ensemble mean.

4. Results for 500-mb height

For brevity, we will refer to JFM 1987 and JFM 1989 as simply El Niño and La Niña where there is no possibility of confusion. We will also display the mean La Niña anomaly fields with their sign reversed to enable a quicker grasp of their similarity or dissimilarity to the El Niño fields.

Figure 3 shows the ENSO signal in seasonal mean 500-mb height in JFM 1987 (top) and JFM 1989 (bottom) as estimated from the historical record (left) and the GCM ensembles (right). The contour interval is 10 m in all panels, and negative values are shaded. Note again that the observational patterns are identical to the historical composite El Niño and La Niña patterns, but their amplitudes have been multiplied by 0.73 and 1.36 to account for the moderate 1987 and strong 1989 events, respectively. The empirically predicted signal patterns resemble each other only broadly: in the Tropics, in the Pacific–North American (PNA) sector, and over the south Pacific Ocean. In most other regions there are differences not only of magnitude but also of sign.

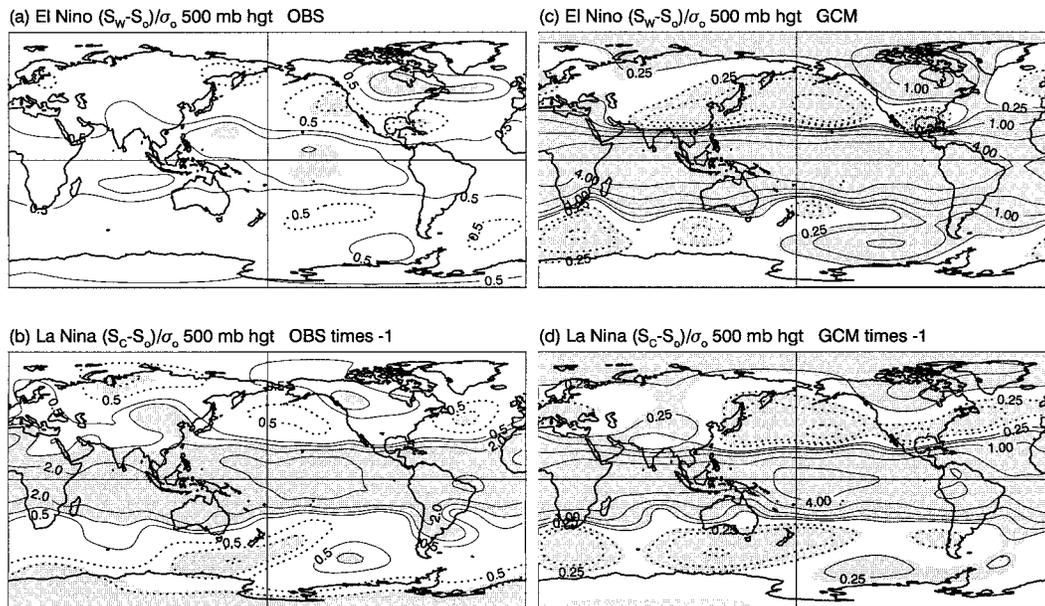


FIG. 4. As in Fig. 3 but with the local signal normalized by the local standard deviation of winter-mean 500-mb heights in the neutral ensembles. The standard deviations used in the observational and GCM plots were estimated from 10 “neutral” years in the observational dataset and 180 members in the GCM’s neutral ensemble. Contouring begins at ± 0.5 (± 0.25) standardized units in the observational (GCM) plots and the contour value doubles at each interval. Negative values are dashed. Shading indicates values significantly different from zero at the 5% level using a two-sided Student’s *t*-test.

Given the small samples, it is difficult to judge whether these differences reflect (i) sampling error, (ii) different linear responses to the slightly different composite SSTs anomaly patterns for El Niño and La Niña, or (iii) truly nonlinear responses. The GCM’s predicted signal patterns for JFM 1987 and JFM 1989, on the other hand, are much more similar to one another. Inadequate sampling is now much less of a concern, so the differences are almost certainly due to different SST forcing for JFM 1987 and JFM 1989 and/or nonlinearity.

Figure 4 is in a similar format to Fig. 3 except that it shows the mean anomalies in units of the local climatological spread, that is, the local standard deviation of the JFM-mean 500-mb heights in the neutral winter ensemble. These represent precisely the standardized ENSO-induced PDF shifts of Fig. 1. The contour value doubles at each interval starting at ± 0.5 in the observational and ± 0.25 in the GCM panels. Negative contours are dashed. The shading indicates values significant at the 5% level according to a local two-sided *t* test. For the observations, the statistically significant signals outside the Tropics are confined mainly to the Pacific–American sector. On the other hand, every plotted contour is significant in the GCM panels.

Although the signal patterns for El Niño and La Niña are generally similar to one another, some notable differences exist. In the GCM, the El Niño response is stronger in both the Tropics and the Pacific–American sector, despite the weaker tropical SST forcing. This is true of only the PNA region in the observations. On the

other hand, the GCM shows a weaker El Niño signal south of Australia that is consistent with observations. There is little else to compare between the GCM and observations because of the limited observational samples. Note that the shading in the El Niño and sign-reversed La Niña panels only indicates values that are statistically different from the “neutral” values, and not from each other. In areas such as the North Atlantic where observations suggest a strong asymmetric signal of the same sign in El Niño and La Niña, the significance of that asymmetry is therefore questionable. It is interesting that the GCM also suggests an asymmetric signal in the North Atlantic, but one that is displaced farther west toward Newfoundland. Similar remarks may be made about the observational and GCM signals over China. These asymmetries between El Niño and La Niña have not been noted previously in the literature, and we will return to them in section 6. An asymmetry that has been emphasized previously (e.g., Hoerling et al. 1997) is the slight eastward shift of the North Pacific low in El Niño relative to (the sign reversed) La Niña. The significance of this observed North Pacific asymmetry, however, can again be questioned for the reason given above. Note that it is much less apparent in Fig. 3, and is also absent in our GCM simulations.

Figure 5 shows the spread of the seasonal mean 500-mb heights in the El Niño (top) and La Niña (bottom) GCM ensembles relative to the spread of the neutral ensemble. The contour interval is 0.1, and the 1 contour is suppressed. Ratios greater (less) than 1 are indicated

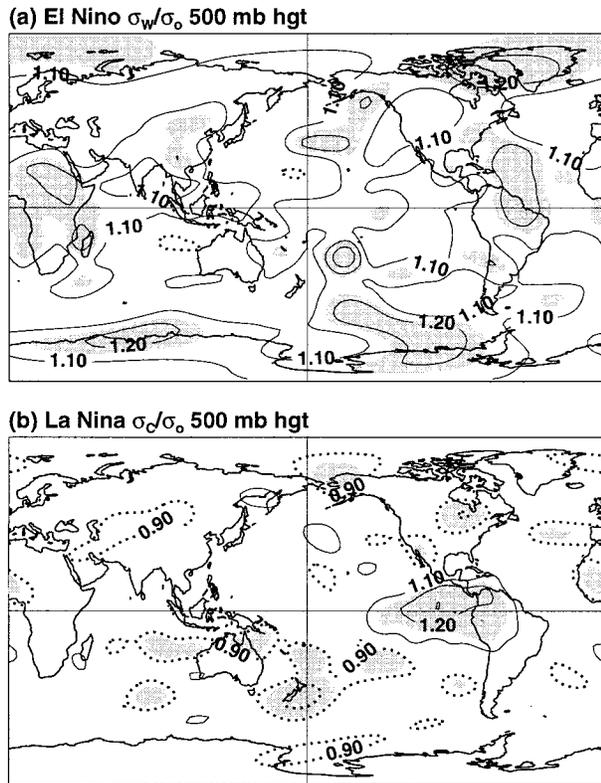


FIG. 5. (a) Ratio of the standard deviation of winter-mean 500-mb heights in the 180-member El Niño GCM ensemble to the standard deviation in the 180-member neutral ensemble. (b) Ratio of the standard deviation of the La Niña and neutral GCM ensembles. Ratios less than 1.0 are dashed. Shading indicates values significantly different from 1.0 at the 5% level using a two-sided F test.

by solid (dashed) contours. The major features to note are (i) over much of the globe, El Niño is associated with an *increase* of the spread of seasonal mean 500-mb heights, whereas La Niña is associated with a decrease, (ii) the patterns of the change of spread are different for El Niño and La Niña, and (iii) the patterns show considerable symmetry about the equator on planetary scales.

A figure such as Fig. 5 has not appeared previously in the literature. It shows evidence of significant systematic changes of spread induced by ENSO, which can locally be as high as 25% for both El Niño and La Niña. The changes of spread for both El Niño and La Niña are also “field significant” in the sense of Livezey and Chen (1983). This was established by performing a bootstrap (i.e., a resampling with replacement) procedure on two sets of 360-member ensembles containing the 180 warm and 180 neutral, and the 180 cold and 180 neutral, members to generate 1000 synthetic warm/neutral and 1000 cold/neutral variance ratio maps, respectively. The percentage of the area in each of these synthetic maps where the 5% significance level was exceeded *locally* was then calculated. This percentage was greater than 6.9% and 8.2%, respectively, in fewer

than 5% (i.e., 50) of the 1000 maps. The percentages of area in Figs. 5a and 5b exceeding the local 5% level were considerably higher, 25.4% and 12%. We will discuss these changes of spread further in section 7.

5. Results for precipitation

Figure 6 shows, in a similar format to Fig. 3, the ENSO signal in seasonal precipitation in JFM 1987 (top) and JFM 1989 (bottom), as estimated from the historical record (left) and the GCM (right). As in Fig. 3, the results for JFM 1989 are shown with the signs reversed. The contour value doubles at each interval starting at $\pm 0.5 \text{ mm day}^{-1}$ in the observational and $\pm 0.25 \text{ mm day}^{-1}$ in the GCM panels. Values less than the first negative contour value are shaded. We repeat that the observational patterns are identical to the historical composite patterns (but now based on only five samples each), whose overall amplitudes are adjusted to reflect the moderate 1987 El Niño and strong 1989 La Niña events.

In terms of anomaly magnitudes, the largest precipitation signals are confined to the Tropics and the subtropics. The observational fields are considerably noisier due to the small number of samples, and many of their details are not significant. They are included here mainly to provide reassurance that the NCEP AGCM’s precipitation response to ENSO is not grossly unrealistic. A large degree of symmetry is evident in the GCM’s response to El Niño and La Niña. However, one can also discern many regions with an asymmetric response. Indeed it is striking how the precipitation response is asymmetric almost everywhere *except* in regions identified with strong ENSO signals in observations using El Niño minus La Niña composite differences, such as the western and central (but not eastern) tropical Pacific, northeast Brazil, and the southeast United States (see also Ropelewski and Halpert 1987, 1996).

To establish the statistical significance of these results, the non-Gaussian character of precipitation variability needs to be taken into account. Precipitation PDFs, unlike 500-mb height PDFs, can be highly non-Gaussian, especially in areas of mean tropospheric descent. Figure 7 demonstrates this for *monthly* precipitation in the winter months of January, February, and March in observations (top panel) and the GCM (middle panel). The observational panel is based on 30 monthly fields in the 10 neutral winters in the NCEP–National Center for Atmospheric Research (NCAR) dataset, and the GCM panel on 540 monthly fields in the 180-member neutral winter ensemble. The thick solid contour in both maps separates areas of mean descent from areas of ascent at 500 mb, and the thin contour shows areas of strong descent. The shading indicates areas in which the precipitation PDFs are significantly different from Gaussian PDFs with the same mean and variance according to the Lilliefors test (see appendix A). These areas correspond well with the areas of descent, with

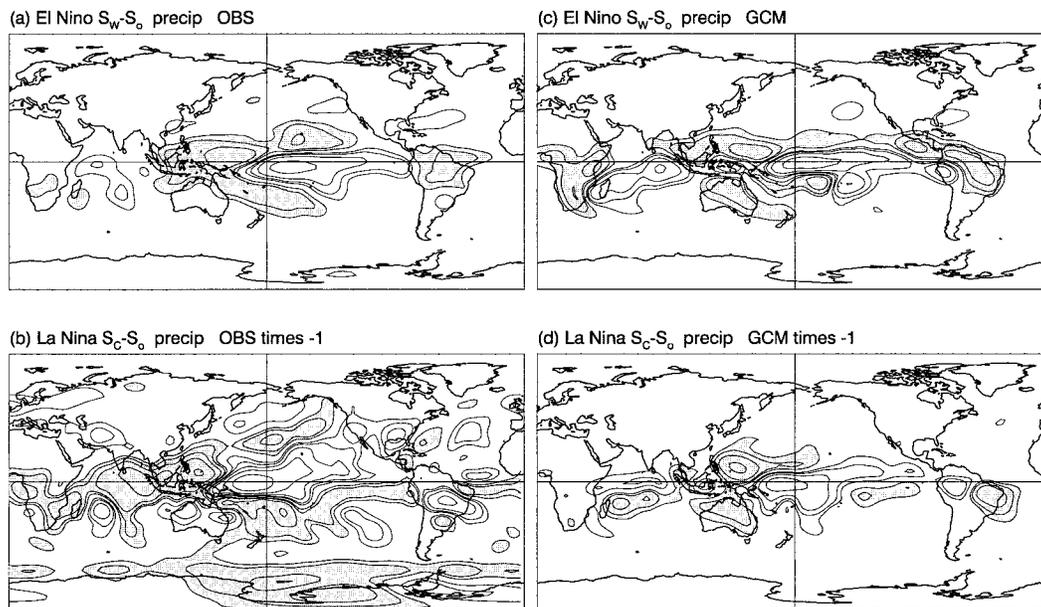


FIG. 6. As in Fig. 3 but for precipitation. The observational fields were obtained by first forming five-event composite fields for El Niño and La Niña in the 1979–98 Xie–Arkin precipitation dataset, and then adjusting their amplitudes to account for the relatively weak 1987 El Niño and relatively strong 1989 La Niña events, respectively. Contour value doubles at each interval in all panels starting from ± 0.5 (± 0.25) mm day^{-1} in the observations (GCM). Negative values are shaded. The zero contour is suppressed.

the strongest departures from normality generally occurring in the areas of the strongest descent. In terms of the D_{ks} statistic, these departures can be greater than 0.2. Note again, however, that these values are for the PDFs of *monthly* and not *seasonal* mean precipitation. One might expect the latter to be more Gaussian by virtue of time averaging and the Central Limit Theorem. The bottom panel of Fig. 7 shows the departure of the PDFs of seasonal (JFM) mean precipitation from Gaussianity in the neutral GCM ensemble in an identical format to that of the middle panel. The D_{ks} values are indeed smaller than in the middle panel, but remain highly significant in the areas of descent.

To view our precipitation results in the format of Fig. 4, and also to demonstrate their significance in terms of the familiar t statistic, we therefore transformed the GCM's seasonal precipitation values to standardized Gaussian deviates using a probability-integral transformation described in appendix B. Figure 8 shows the GCM's seasonal ensemble-mean precipitation response so standardized, in the format of Fig. 4. The contours are ± 0.25 , ± 0.5 , ± 1 , ± 2 , ± 4 , \dots , etc. and negative values are indicated by dashed contours. Areas in which the precipitation response is significant at the 5% level using a two-sided t test are shaded. The figure shows that this GCM can evidently produce substantially different precipitation response patterns for the different SST forcings associated with individual El Niño and La Niña events. For example, El Niño (La Niña) is generally associated with wet (dry) winters over California and the southeast United States. Figure 8 is consistent

with this general tendency, but with differences of detail that would be important when making seasonal precipitation forecasts in these regions.

Figure 8 also reproduces the well-known tendency of the central equatorial Pacific rainfall anomaly extremum during El Niño to occur east of the corresponding extremum during La Niña. This is also evident in Fig. 6, and has been speculated by Hoerling et al. (1997) to be responsible for the observed eastward shift of the geopotential low in the subtropical northern Pacific in Fig. 4. As remarked earlier, the GCM produces much less of this eastward geopotential shift, despite producing a strong eastward rainfall shift along the equator.

Our large ensemble sizes enable us to be confident that the asymmetries of the precipitation response evident in Fig. 8 are not due to sampling error. The strong asymmetry over the Indian Ocean is interesting. There is a hint of it in the observational panels of Fig. 6, but more importantly, it is also evident in several 100-yr station records of precipitation in the region (not shown).

The GCM also predicts significantly different spreads (both locally and in the “field significance” sense) of the seasonal mean precipitation in the El Niño and La Niña ensembles. We will postpone presentation of these results until section 7.

6. ENSO-induced PDF changes

An ENSO event affects the mean as well as the standard deviation of the PDFs of atmospheric variables

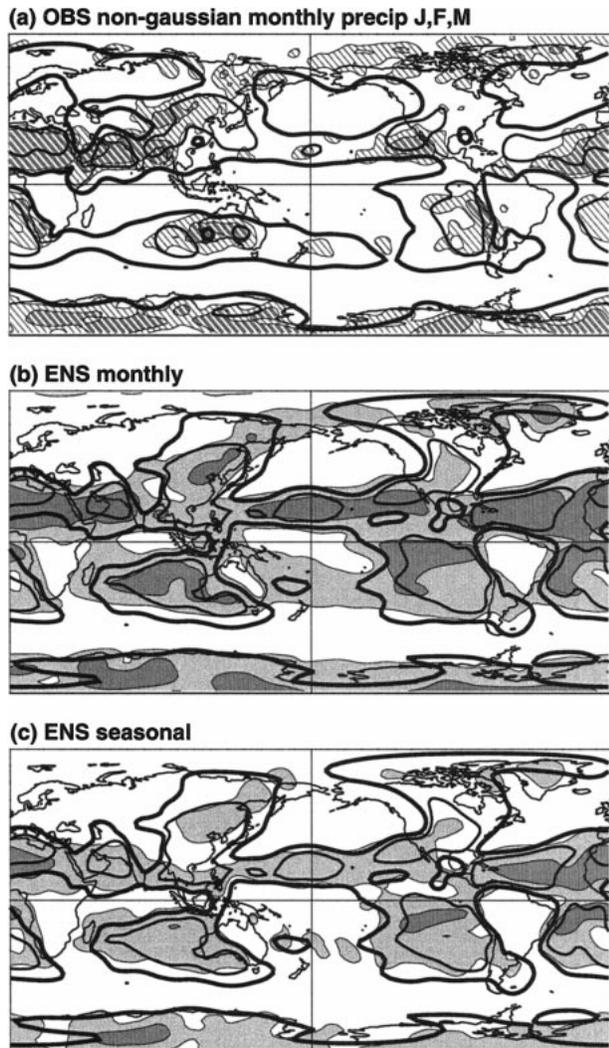


FIG. 7. Shading highlights regions in which the probability distribution of monthly precipitation is significantly non-Gaussian in (a) the 30 months of the 10 neutral winters in the NCEP reanalyses for 1958–99, (b) the 540 months in the 180-member neutral winter GCM ensemble, and (c) the 180 winter (JFM) means in the neutral GCM ensemble. The (a) stippled and (b, c) shaded contours indicate the Kolmogorov–Smirnov statistic D_{ks} between the precipitation distribution and a Gaussian distribution with the same mean and variance. The stippled contours in (a) are 0.175 and 0.25, which are well above the 5% significance level with 30 samples ($=0.161$) according to the Lilliefors test (Wilks 1995), and in (b) and (c) are 0.075 and 0.15, which are well above the 5% significance level with 540 samples ($=0.035$), and with 180 samples (0.060). In all panels, the thick contour separates areas of JFM-mean 500-mb ascent from areas of descent, and the thin contour shows areas of descent with vertical velocity equal to 0.025 Pa s^{-1} in the observational and neutral GCM ensembles.

worldwide. As discussed earlier, the D_{ks} statistic is useful for summarizing these and other changes to the PDF. In an important sense, D_{ks} measures the total impact of ENSO on the probability distribution associated with an atmospheric variable.

The upper panel of Fig. 9 shows, for seasonal mean

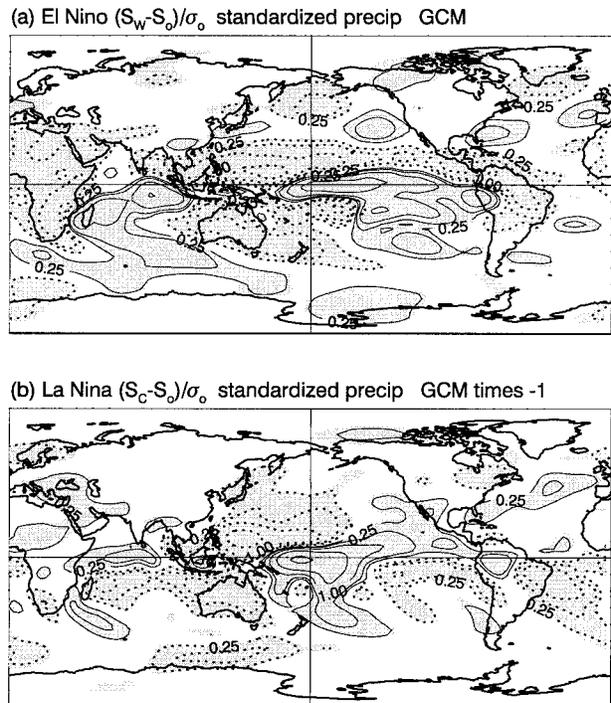
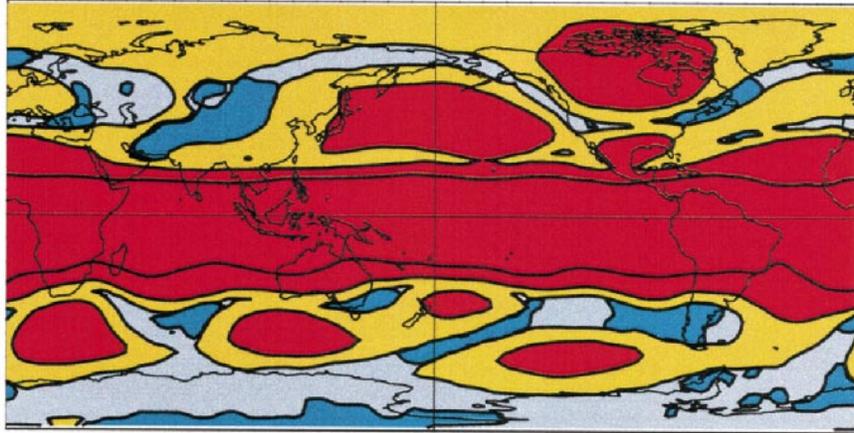


FIG. 8. As in Fig. 4 but for the GCM's transformed ensemble-mean precipitation anomalies normalized by the standard deviation of the transformed winter-mean precipitation in the 180-member neutral ensemble. Contours double with each interval beginning at ± 0.25 . Shading indicates values significantly different from zero at the 5% level using a two-sided Student's *t*-test.

500-mb heights, the D_{ks} of the El Niño and La Niña distributions derived from the El Niño and La Niña GCM ensembles, respectively. The lower panel shows the same quantity for seasonal mean precipitation. The plotting conventions are identical in the two panels. The contours drawn are 0.15, 0.4, and 0.9, with orange and red colors indicating regions in which D_{ks} is significant at the 5% level with 180 members ($D_{ks} \geq 0.144$). In other words, these are regions in which the PDFs for El Niño and La Niña can be distinguished from one another. In the light blue regions, the PDFs cannot be distinguished even with 180 members. In the dark blue areas, the PDFs are still indistinguishable from one another; however, the El Niño and/or the La Niña distribution can be distinguished from the neutral distribution. The responses in these dark blue areas are therefore both highly significant and highly asymmetric with respect to the neutral distribution.

Thus, in all regions except the light blue regions, there is a statistically significant response to ENSO in these GCM experiments, where the term “response” is to be interpreted more broadly as any change of the PDF rather than a shift of the mean alone. The red regions in Fig. 9 are already familiar as regions with significant ENSO signals in 500-mb height and precipitation, and are entirely consistent with the emphasis on them in

(a) 500 mb height



(b) Precipitation

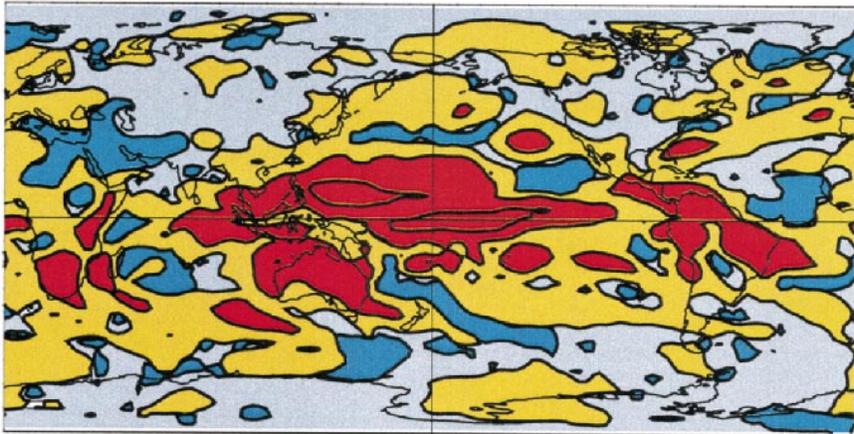


FIG. 9. Kolmogorov–Smirnov measure D_{ks} of the difference between the probability distributions of El Niño and La Niña GCM ensembles for (a) JFM-mean 500-mb heights and (b) JFM-mean precipitation. Contours drawn are 0.15, 0.40, and 0.90. The orange and red regions indicate regions in which D_{ks} is significant at the 5% confidence level ($D_{ks} \geq 0.144$). The blue regions indicate where the El Niño and La Niña distributions are not significantly different from each other at the 5% level. The dark blue regions indicate where the El Niño and/or the La Niña distribution is significantly different from the neutral distribution, but where the El Niño and La Niña distributions are still indistinguishable from each other. All ensembles have 180 members.

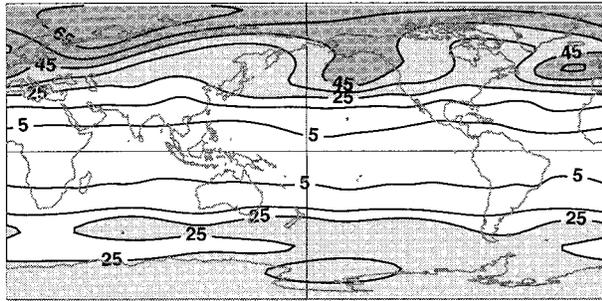
previous ENSO studies. The dark blue regions are perhaps the most interesting. It is possible that they represent regions with a highly nonlinear response to tropical SST forcing, but our experimental design precludes us from addressing this cleanly. They can alternatively, and unambiguously, be interpreted as areas in which there can be a substantial impact of ENSO during *some* events, which may be obscured by examining El Niño minus La Niña differences. The 500-mb responses over central Asia and China, the northeast United States, southeast Australia, and southern South America are noteworthy in this regard. The dark blue regions are equally extensive in the precipitation panel of Fig 9.

The Nordeste region of Brazil, as well as the eastern Mediterranean and the Middle East, emerge as sensitive areas in this respect.

7. Another look at the changes of spread

A reliable estimation of ENSO-induced changes of variability is one of the original contributions of this paper, made possible by our generation of large ensembles. To get a feel for these changes, Fig. 10 shows the standard deviation of seasonal mean 500-mb heights and precipitation in the 180-member neutral ensemble. The contour interval is 10 m for the heights, and values

(a) CLIM σ_o 500 mb hgt



(b) CLIM σ_o precip

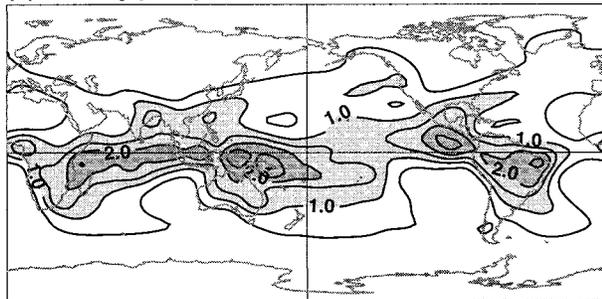


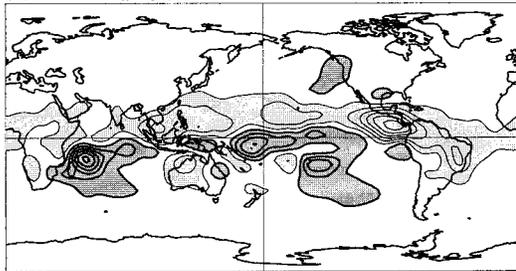
FIG. 10. (a) Standard deviation of JFM-mean 500-mb heights in the neutral GCM ensemble. Contours are at 10-m intervals starting at 5 m. (b) As in (a) but for JFM-mean precipitation. Contour interval is 0.5 mm day⁻¹.

greater than 25 (45) m are indicated by light (dark) shading. The contour interval for precipitation is 0.5 mm day⁻¹, with values greater than 1.0 (2.0) mm day⁻¹ indicated by light (dark) shading. The figure shows generally increasing variability of the 500-mb heights as one moves poleward from the Tropics, with regional maxima over the northeast Pacific and Atlantic Oceans. The precipitation variability is largely confined to the Tropics, with a tendency for the maxima to be collocated with the maxima of the mean precipitation itself (not shown).

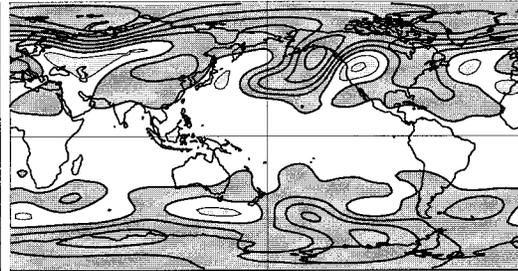
Figure 11 shows how the values in Fig. 10 are affected by El Niño and La Niña forcing in our GCM experiments. It shows the change of the standard deviation of seasonal mean 500-mb heights (right) and precipitation (left) for El Niño (top) and La Niña (bottom). (Note that unlike in Figs. 3, 4, 6, and 8, the La Niña fields *do not* have their signs reversed in any of the variance-change plots.) The contours in the height plots are $\pm 1, \pm 3, \pm 5, \dots$ m; those in the precipitation plots are $\pm 0.1, \pm 0.3, \pm 0.5, \dots$ mm day⁻¹. In all the panels, positive (negative) values are indicated by dark (light) shading.

The height panels in Fig. 11 convey the information of Fig. 5 (and also Fig. 2) in a different manner. The general increase of extratropical 500-mb height variance during El Niño, and decrease during La Niña, are now clearly evident. So is the general symmetry of these changes with respect to the Northern and Southern Hemispheres. The changes of the height spread are typically around 5 m in the extratropics, although locally they can be as high as 9 m in the PNA region.

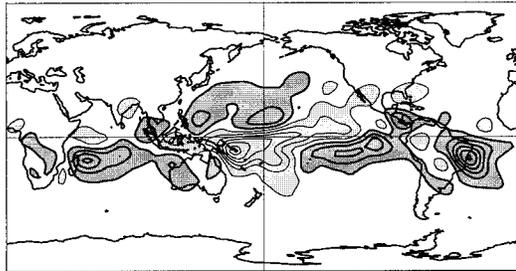
(a) El Niño $\sigma_w - \sigma_o$ precip



(c) El Niño $\sigma_w - \sigma_o$ 500 mb hgt



(b) La Niña $\sigma_c - \sigma_o$ precip



(d) La Niña $\sigma_c - \sigma_o$ 500 mb hgt

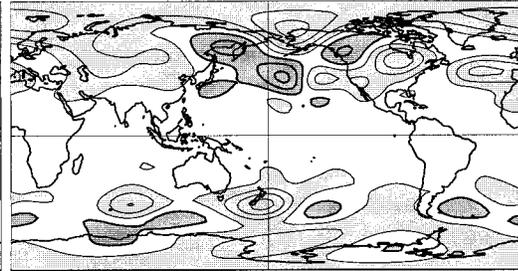


FIG. 11. The difference in the standard deviations of the (left) JFM-mean precipitation and (right) 500-mb height El Niño and (top) neutral GCM ensembles and (bottom) La Niña and neutral GCM ensembles. Contours in the precipitation plots are drawn at intervals of 0.2 mm day⁻¹ starting at 0.1 mm day⁻¹. Those in the 500-mb height plots are drawn at 2-m intervals starting at 1 m. Positive values are indicated by dark and negative values by light shading.

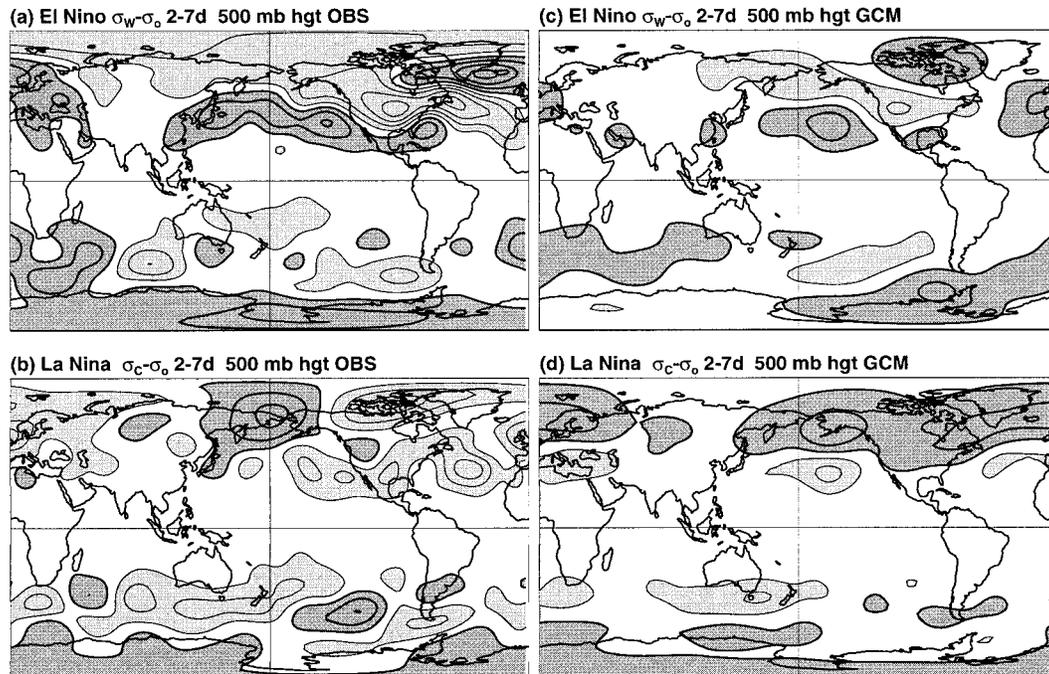


FIG. 12. The difference in the standard deviations of bandpass-filtered 500-mb heights between (left) ENSO and neutral winters in the NCEP reanalyses (left) and (right) the GCM's ENSO and neutral winter ensembles. Results are shown for El Niño minus neutral in the upper and La Niña minus neutral in the lower panels. Contours in all panels are drawn at 2-m intervals starting at 1 m. Positive values are indicated by dark and negative values by light shading, and the zero contour is suppressed. The bandpass filter passes periods between 2 and 7 days.

The precipitation panels of Fig. 11 tell a different story. There is some tendency for the patterns to be of opposite sign for El Niño and La Niña, although again primarily in those regions identified with large ENSO signals through El Niño minus La Niña composite differences, such as the central and eastern equatorial Pacific, northeast Brazil, and the southern United States. Elsewhere, the patterns are highly asymmetric. The asymmetry over the Indian Ocean is again striking.

It is tempting to understand the altered spread of the 500-mb heights in terms of the altered spread of precipitation, but the precipitation panels in Fig. 11 do not suggest an obvious link. One may speculate that to the extent that the mean rainfall anomalies in Fig. 6 in the central equatorial Pacific are important in determining the mean extratropical response, the greater (lesser) variability of the precipitation in this region in Fig. 11 during El Niña (La Niña) may also partly account for the greater (lesser) variability of the extratropical response. The typically 10%–20% changes of the precipitation spread are consistent with the 10%–20% changes of the geopotential height spread in this regard. The central equatorial Pacific has also been identified by Barsugli et al. (1997) as a sensitive forcing region for generating a large global response. A detailed dynamical diagnosis is needed, and will be reported in a future publication.

Figure 11 will be unfamiliar to many readers, espe-

cially since an observational counterpart does not yet exist. Still, the result that El Niño is associated with *increased* extratropical seasonal height variability may come as somewhat of a surprise. Such a result must ultimately be reproduced by other GCMs and for other ENSO events before one can gain confidence in it. Note that the impact of ENSO on *synoptic* variability is very different from its impact on *seasonal* variability. Figure 12 shows ENSO-induced changes of the standard deviation of 500-mb heights on 2–7-day timescales. Results are shown for El Niño (top) and La Niña (bottom) for the GCM (right) and the NCEP–NCAR reanalyses (left). As with the other comparisons of GCM and observational fields in this paper, the comparison in Fig. 12 is not clean in that the observational results are for 10 sets of El Niño and La Niña events whereas the GCM results are for one set. Nevertheless, the observed southward shift of the storm tracks over the eastern Pacific and United States during El Niño, and the northward shift during La Niña, are reasonably well captured by the GCM for this particular set of events.

8. Altered risks of extreme seasonal anomalies

As discussed in section 1, the original motivation for this paper was to explore how to maximize the utility of seasonal forecasts even when the signal-to-noise ratio is small, by focusing on the altered risks of extreme

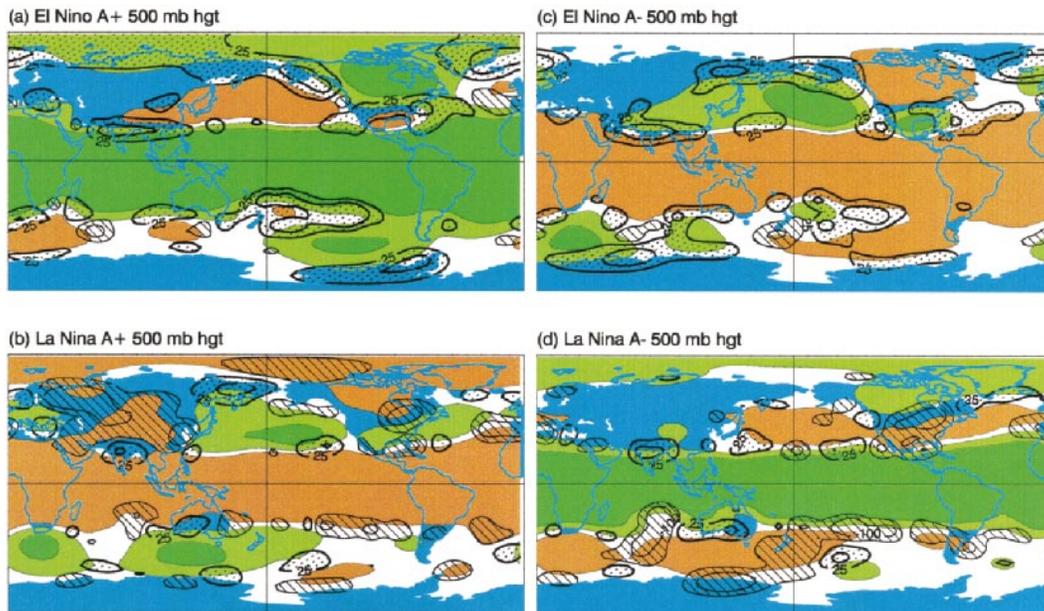


FIG. 13. The change in the probability of extreme values of JFM-mean 500-mb height estimated from the AGCM for (top) El Niño and (bottom) La Niña ensembles. Colored shading depicts the changes of probability of (left panels) extreme positive values $A+$ and (right panels) extreme negative values $A-$ defined in Fig. 1. Green (orange) shading indicates an increase (decrease) of the probability of extreme values. The shading contours are 5% and 25%. (Note that the negative changes cannot be greater than 16%; see Fig. 1). The stippled (striped) regions indicate where the ENSO-induced changes of variability make relatively large positive (negative) contributions to the values of $A+$ and $A-$. The contours drawn are 25% and 100%. A 100% contour indicates that the ENSO-induced change of variability makes just as large a contribution to the value of A as the ENSO-induced shift of the mean.

seasonal anomalies. Having established statistically significant ENSO-induced changes of variability as well as shifts of the mean through our large GCM ensembles, we are now in a position to address this issue.

Figure 13 summarizes the results for 500-mb heights in the framework of Fig. 1. The colored shading in all four panels indicates the change in the likelihood of extreme positive ($A+$) and extreme negative ($A-$) JFM-mean anomalies for La Niña and El Niño SST forcing as predicted by this particular GCM for this particular set of ENSO events. Positive values are colored green and negative orange; for clarity only the 5% and 25% shading levels are shown. In addition to all the well-known aspects of the 500-mb height response to ENSO, the asymmetries of the response discussed in the previous sections are also evident in Fig. 13. But perhaps the most interesting features are the stippled and striped regions. These show where the ENSO-induced changes of noise (i.e., the seasonal variability changes in Fig. 11) make a substantial positive (stippled) and negative (striped) contribution to the predicted values of $A+$ and $A-$. In other words, in these regions a scenario such as that of Fig. 1b is relevant; in other regions, that of Fig. 1a is sufficient. Figure 13 suggests that in many regions around the globe, but especially in the PNA and North Atlantic Oscillation regions of interest to the climate research community, the changes of seasonal variance can be as effective as the mean ENSO signal in

affecting the risks of extreme seasonal anomalies in ENSO winters. Figure 14 is a similar figure for seasonal precipitation, in an identical format, and further supports this conclusion.

9. Implications for predictability

Finally, our results also have implications for the predictability of seasonal averages during El Niño and La Niña events. This issue is conveniently addressed in the context of Fig. 15, as explained below.

Let us say that the multivariate distribution $P(\langle \mathbf{x} \rangle, \mathbf{C}_0)$ represents the altered PDF of seasonal averages associated with an ENSO event, where $\langle \mathbf{x} \rangle$ is the population-mean anomaly state vector and \mathbf{C}_0 is the covariance matrix of the variations \mathbf{x}' of the anomaly state vector around $\langle \mathbf{x} \rangle$. Note that P need not be multinormal. We define a perfect GCM as one that correctly reproduces P with an infinite-member ensemble. The PDF of the ensemble mean of finite n -member ensembles is therefore $P(\langle \mathbf{x} \rangle, n^{-1}\mathbf{C}_0)$. A vector \mathbf{y} from this distribution is issued as the seasonal average forecast. The real atmosphere picks a vector \mathbf{x} from $P(\langle \mathbf{x} \rangle, \mathbf{C}_0)$ as its seasonal average. The average anomaly correlation of the observed and predicted vectors can then be shown to be

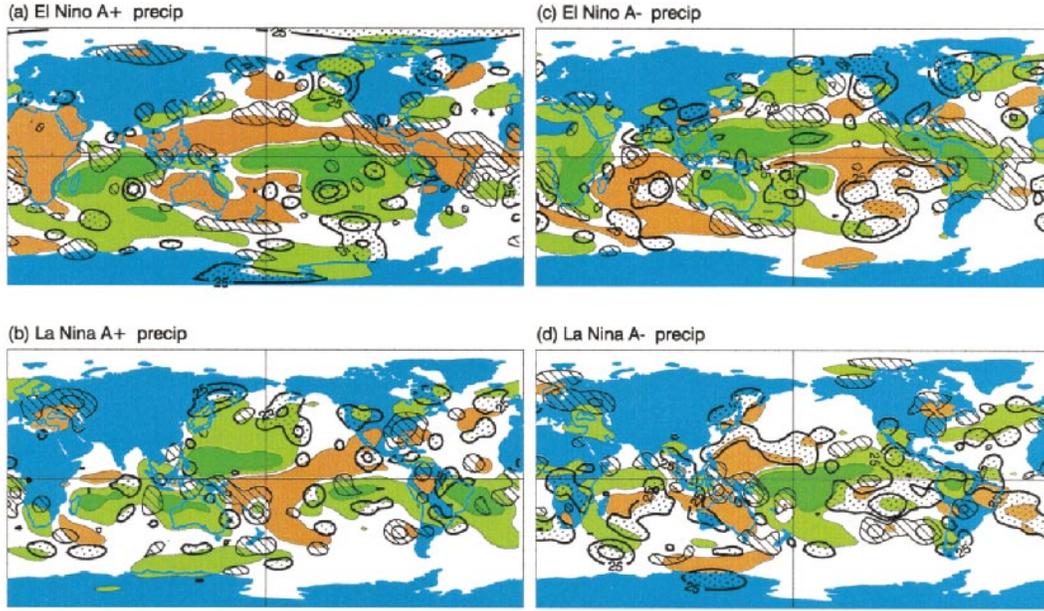


FIG. 14. Same as Fig. 13 but for precipitation.

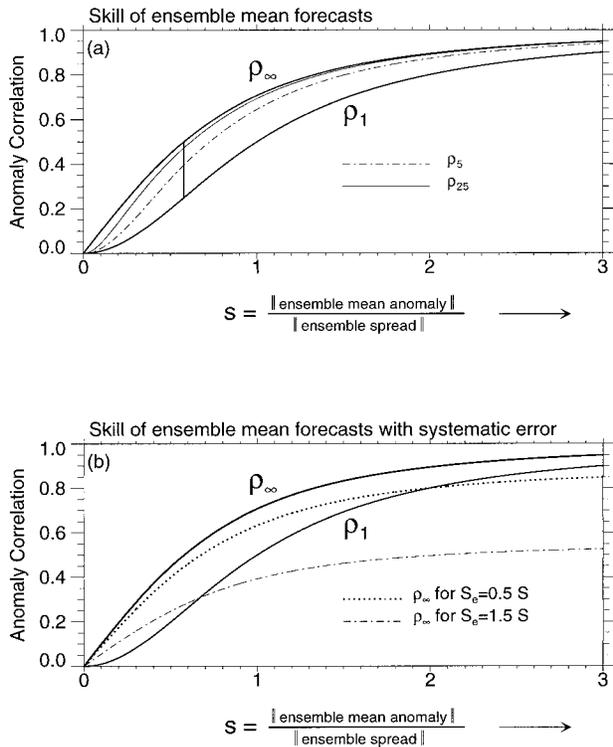


FIG. 15. (a) Theoretical predictability curves for ensemble sizes of 1, 5, 25, and infinity as a function of the signal-to-noise ratio s based on Eq. (4). The solid vertical bar indicates the maximum distance ($=0.25$) between the ρ_∞ and ρ_1 curves obtained at $s = 1/\sqrt{3}$. (b) The ρ_∞ and ρ_1 curves are copied from the upper panel. The dashed curves now show the modified predictability ρ_∞ [Eq. (6)] with a nonzero systematic error s_e whose magnitude is 50% and 150% that of s . See text for a full explanation.

$$\begin{aligned} \rho_n &= \langle \mathbf{x} \cdot \mathbf{y} \rangle / (\langle \mathbf{x} \cdot \mathbf{x} \rangle \langle \mathbf{y} \cdot \mathbf{y} \rangle)^{1/2} \\ &= \langle \mathbf{x} \rangle \cdot \langle \mathbf{x}' \rangle \\ &\quad \div [(\langle \mathbf{x} \rangle \cdot \langle \mathbf{x} \rangle + \langle \mathbf{x}' \cdot \mathbf{x}' \rangle) (\langle \mathbf{x} \rangle \cdot \langle \mathbf{x} \rangle + n^{-1} \langle \mathbf{x}' \cdot \mathbf{x}' \rangle)]^{1/2}. \end{aligned} \tag{3}$$

In this expression the dot product signifies a general scalar product of the form $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{W} \mathbf{y}$, where \mathbf{W} is any suitable positive-definite weight matrix. Note that $\langle \mathbf{x}' \cdot \mathbf{x}' \rangle = \text{Tr}[\mathbf{W}^{1/2} \mathbf{C}_0 \mathbf{W}^{1/2}]$. The weight matrix \mathbf{W} can be chosen to emphasize one variable at one grid point, a linear combination of variables over a region, or be set equal to identity and examine forecast skill over the atmosphere as a whole.

The average anomaly correlation can be expressed more concisely as

$$\rho_n = s^2 / [(s^2 + 1)(s^2 + n^{-1})]^{1/2}, \tag{4}$$

where $s = (\langle \mathbf{x} \rangle \cdot \langle \mathbf{x} \rangle / \langle \mathbf{x}' \cdot \mathbf{x}' \rangle)^{1/2}$ is a measure of the population mean anomaly relative to the population spread, that is, the “true” signal-to-noise ratio. Figure 15a shows how ρ_n varies as a function of s and ensemble size n . Note that $\rho_1 = \rho_\infty^2$, a fact also noted by Rowell (1998). The difference between the ρ_1 and ρ_∞ curves shows the gain to be made in forecast skill by using forecast ensembles instead of a single forecast. The maximum gain in anomaly correlation of 0.25 is achieved for $s = 1/\sqrt{3}$. Note that most of this gain is achieved with 25 members.

The ρ_∞ curve represents a hard predictability limit with a perfect model, and shows that to produce “useful” forecasts with anomaly correlations greater than 0.6, s needs to be greater than 0.75. To produce “excellent” forecasts with anomaly correlations greater

than 0.9, s needs to be greater than 2. The question is what the appropriate values of s are during individual ENSO events for different atmospheric variables in different parts of the globe. Figures 4 and 8 are useful in this regard for seasonal mean 500-mb height and precipitation. Note that these only represent *estimates* of s , both because of finite samples and more subtly, because the signals in those panels were divided by the standard deviation of the original rather than the altered distributions. With these caveats, we estimate s to be greater than 2 for geopotential heights only in the Tropics and for precipitation only in the central equatorial Pacific. The weaker requirement of $s > 0.75$ for useful forecasts is met additionally for precipitation over Australia and northeast Brazil, and for geopotential heights in the PNA region. These statements are entirely consistent with the emphasis on these regions in numerous ENSO studies. The predictability of seasonal precipitation over the United States, however, emerges as less promising according to these measures.

In this paper we have stressed the need of large ensembles and the fact that s can be different for individual El Niño and La Niña events due to differences of both signal and noise. *All of these factors are important for locating oneself on the s axis* in Fig. 15. For example, one might expect seasonal averages to be less predictable during La Niña than El Niño since the signal is generally weaker. However, this ignores our result in Fig. 11 that the noise is also weaker in some regions during La Niña, so values of s need not in fact be smaller. Similarly, 25 ensemble members may suffice for estimating the population *mean* in most cases. However, many more than 25 members are needed to estimate the population *spread* within 20%, whose inaccurate estimates can lead to both false predictability estimates and a reduced ability to forecast the forecast skill.

Note that Fig. 15a was generated under the “perfect model” assumption. If an imperfect model predicts a distribution $P(\langle \mathbf{x} + \mathbf{x}_e \rangle, \mathbf{C}_{0m})$ instead of $P(\langle \mathbf{x} \rangle, \mathbf{C}_0)$, where $\langle \mathbf{x}_e \rangle$ represents the model error in predicting the population mean, and \mathbf{C}_{0m} the model’s covariability around $\langle \mathbf{x} + \mathbf{x}_e \rangle$, then Eq. (3) becomes

$$\rho_n = \frac{\langle \mathbf{x} \rangle \cdot \langle \mathbf{x} + \mathbf{x}_e \rangle / [\langle \langle \mathbf{x} \rangle \cdot \langle \mathbf{x} \rangle \rangle + \langle \langle \mathbf{x}' \cdot \mathbf{x}' \rangle \rangle]}{\times (\langle \langle \mathbf{x} \rangle \cdot \langle \mathbf{x} + \mathbf{x}_e \rangle \rangle + n^{-1} \langle \langle \mathbf{x}' \cdot \mathbf{x}' \rangle \rangle_m)^{1/2}} \quad (5)$$

This expression can also be simplified if we assume that $\mathbf{x} \cdot \mathbf{x}_e = 0$ and $\mathbf{C}_{0m} = \mathbf{C}_0$. We then have

$$\rho_n = s^2 / [(s^2 + 1)(s^2 + s_e^2 + n^{-1})]^{1/2}, \quad (6)$$

where $s_e = [\langle \langle \mathbf{x}_e \rangle \cdot \langle \mathbf{x}_e \rangle \rangle / \langle \langle \mathbf{x}' \cdot \mathbf{x}' \rangle \rangle]^{1/2}$. Figure 15b shows how the potential predictability ρ_∞ is degraded if $s_e = 0.5s$ and $s_e = 1.5s$. For the latter, the skill even using infinite member-ensembles is much less than the skill of a perfect model using a single member.

10. Concluding remarks

Our GCM results suggest that the global atmospheric response to individual El Niño and La Niña events can differ substantially from historical composite signals. They also indicate substantial asymmetry in the response to El Niño and La Niña. This asymmetry is due to both multilinearity (i.e., the fact that the tropical SST anomaly patterns for El Niño and La Niña are not mirror images of each other) and nonlinearity (i.e., some asymmetry would be expected even if the SST patterns were mirror images). The asymmetry of the mean response is particularly marked for precipitation over the central equatorial Pacific and western Indian Oceans, and for 500-mb heights over the northern Atlantic Ocean, eastern Asia, and south of Australia.

In general the seasonal response to El Niño is stronger, but also *more* variable, than the response to La Niña in our GCM experiments. We have speculated that the greater seasonal variability of the 500-mb heights may be partly due to the greater variability of seasonal rainfall in the central equatorial Pacific during El Niño. If true, this would provide an interesting link between “tropical chaos” and “extratropical chaos.” Another reason for the altered extratropical variability could be an adjustment to the altered mean state. Changes of synoptic variability are probably controlled by this mechanism. According to the linear storm-track theory of Whitaker and Sardeshmukh (1998), the approximately opposite mean anomalies forced by El Niño and La Niña could induce the approximately opposite changes in the extratropical storm tracks seen in Fig. 12. Further, those altered storm tracks could induce opposite effects on the lower-frequency variability through opposite eddy/mean-flow feedbacks. All of these possibilities need to be investigated in a careful dynamical diagnosis of the problem.

That both the population mean and spread of seasonal averages can differ for individual ENSO episodes has implications for the predictability of seasonal averages during such episodes. Because of the relatively small signal-to-noise ratio s , the predictability of seasonal averages is modest outside the Tropics. According to Fig. 15, the average seasonal anomaly correlation skill ρ_∞ of ensemble-mean seasonal forecasts will be less than 0.6 for variables with $s < 0.75$ even for a perfect model utilizing infinite-member ensembles. However, because ρ_∞ increases rapidly with s for small s , an accurate determination of s becomes relatively more important in such situations, both for improving the skill of the forecast and for assessing its reliability. In this connection it is interesting that our GCM results do not always imply lower extratropical predictability during La Niña. Although the La Niña signal is weaker, the noise can also be weaker, so s , and therefore ρ_∞ , can remain relatively unchanged. Note that ρ_∞ refers strictly to the *average* forecast skill of infinite-member ensemble means over an infinite number of forecast cases in a

perfect model setting. This should be borne in mind when considering the results of say Chen and van den Dool (1997), who report greater average hindcast skill during El Niño than La Niña based on 13-member mean hindcasts for seven El Niño and seven La Niña cases.

A GCM should in principle be able to capture the detailed changes of s from event to event. Most empirical models, because of their assumption of linearity (or at most low-dimensional multilinearity; see Peng et al. 2000) imposed by the limited observational record, do not have the flexibility to represent such details. Their erroneous prediction of signal, plus their inability to predict changes of noise, amount in effect to a systematic error s_e in Fig. 15b. In several extratropical regions in Fig. 4, the differences between the empirical and GCM signals are of the order of the signal itself, that is, $s_e \sim s$. If these differences are interpreted as reflecting errors of the empirical model, then a GCM would indeed have an advantage in predicting seasonal averages. However, that advantage would be slight for $s < 0.75$, and become evident only after averaging over many cases, because the spread of forecast skill in Fig. 15 is large for small s (not shown). It is, of course, equally possible that the differences between the GCM and empirical signals in Fig. 4 reflect GCM error. Such an interpretation should certainly motivate GCM improvements with the goal of approaching the ρ_∞ curve in Fig. 15a.

The gap between the ρ_∞ and ρ_1 curves in Fig. 15 is a measure of the gain in forecast skill to be realized by running GCM ensembles. Although there is always something to be gained by running large forecast ensembles rather than a single forecast, the usefulness of that gain can be moot. For $s > 3$, a single forecast is already “excellent,” whereas for $s < 0.75$ the gain, though relatively large, is still not large enough to render the ensemble-mean forecast “useful.” In this sense, ensembles are unnecessary for large s , and for small s they remain useless, for predicting *seasonal averages*. For intermediate values of s , however, say between 0.75 and 3, the gain of anomaly correlation skill of up to 0.25 can be both substantial and useful. Figure 15b nonetheless provides a sobering reminder that even this gain can be nullified by a model’s systematic error.

In conclusion, the ρ_∞ curve in Fig. 15 strongly constrains the predictability of seasonal averages during ENSO events, and because of this even a perfect GCM’s advantage over an empirical model may well turn out to be modest. The results of this paper show, however, that a GCM’s apparently minor advantages in predicting variations of the ENSO signal from event to event, and certainly the changes of noise, would translate into enormous advantages in predicting the variations of extreme risks. Accurate estimation of these altered risks would benefit not only from GCM improvements, but also from running large ensembles to determine accurately the apparently minor but nonetheless important changes of noise.

Acknowledgments. The encouragement of the CDC Director R. Dole to pursue this problem is much appreciated. Assistance with running the NCEP AGCM by our CDC colleagues, especially J. Barsugli, as well as members of NCEP’s Coupled Modeling Branch is gratefully acknowledged. The GISST data were kindly provided by D. Rowell and the Hadley Centre of the U.K. Meteorological Office. The NCEP–NCAR reanalysis data were obtained from the CDC archives (<http://www.cdc.noaa.gov>). This research was partly supported by a grant from NOAA’s Office of Global Programs.

APPENDIX A

The Kolmogorov–Smirnov Distance between Two Probability Distributions

Our interest in this paper is in ENSO-induced changes to the entire PDF and not only in the changes of the mean and variance. The t and F tests are not entirely appropriate in this context. They are strictly applicable only to Gaussian distributions, and further, the t test is strictly applicable only when there is no change of variance. One way to assess the difference between two PDFs is in terms of the nonparametric two-sample Kolmogorov–Smirnov measure D_{KS} (see also Anderson and Stern 1996). Figure A1 illustrates the concept for two Gaussian PDFs $N(0, 1)$ and $N(0.5, 1.2^2)$. The D_{KS} mea-

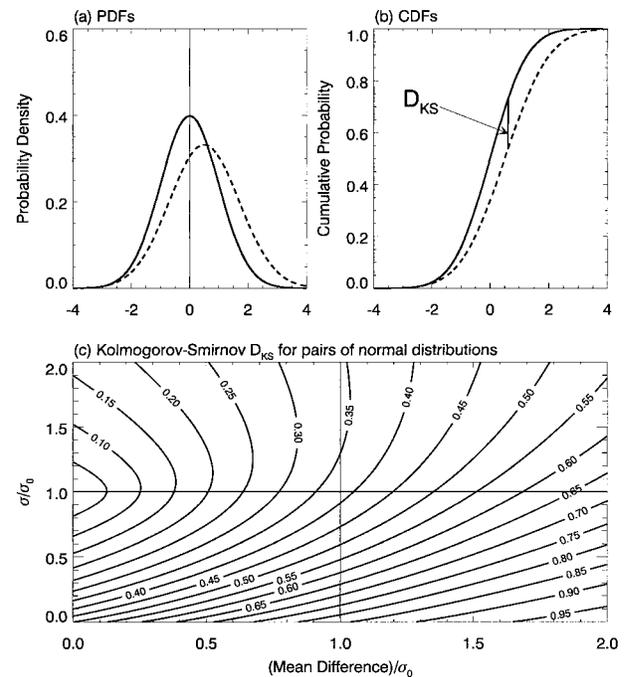


FIG. A1. (a) Normal PDFs $N(0, 1)$ (solid curve) and $N(0.5, 1.2^2)$ (dashed curve). (b) CDFs corresponding to the PDFs in (a). The vertical segment indicates the maximum distance between the two curves and is referred to as the Kolmogorov–Smirnov statistic D_{KS} . (c) Values of D_{KS} between normal distributions $N(\mu_0, \sigma_0^2)$ and $N(\mu, \sigma^2)$, where the mean shift is $\mu - \mu_0$.

sure of their difference is defined as the *magnitude of the maximum difference* between their corresponding cumulative distribution functions (CDFs), which in this case is 0.19. Since all CDFs rise from 0 to 1, the value of D_{ks} also always lies between 0 and 1. To get a feel for what gives rise to high values of D_{ks} , the lower panel of Fig. A1 gives the D_{ks} between $N(0, 1)$ and $N(\bar{x}, \sigma^2)$ for a range of values of \bar{x} and σ . Note that the plot is symmetric about the $\bar{x} = 0$ axis, so only the right half-plane is shown. It is clear that changes of spread ($\sigma \neq 1$) can be as effective in increasing D_{ks} as shifts of the mean ($\bar{x} \neq 0$).

As in the t and F tests, one again needs to address the question of whether the number of members drawn from each distribution are sufficient for a sample value of D_{ks} obtained to be statistically significant, that is, higher than a value that could be obtained by chance for two sets of N samples drawn from the same distribution. This issue is taken up in Fig. 2.

The Kolmogorov–Smirnov distance can also be used to test a sample distribution’s goodness-of-fit to a *specified* theoretical distribution such as a Gaussian. When the parameters of the specified distribution are estimated from the same sample data instead of being known a priori, the standard KS significance levels are not applicable and modifications must be made to them (Crutcher 1975). This modified goodness-of-fit KS test, called a Lilliefors test (Wilks 1995), was developed by Lilliefors (1967) and apparently independently by Stephens (1970).

APPENDIX B

The Probability-Integral Transformation to Standard Gaussian Deviates

Suppose x is a random variable with PDF $f(x)$ and CDF $F(x)$ such that $dF = f(x) dx$. Let z be a standard Gaussian variable with zero mean and unit variance, with PDF $g(z)$ and CDF $G(z)$ such that $dG = g(z) dz$. We transform x to z by equating probability such that

$$dG(z) = dF(x).$$

Integrating the left and right hand sides from $-\infty$ to z , and $-\infty$ to x , respectively, gives

$$G(z) = F(x) + C,$$

where C is a constant of integration which must be zero since $G(\infty) = F(\infty) = 1$. The transformation linking z to x is therefore

$$z = G^{-1}F(x) \quad (\text{B1})$$

This equation states that the cumulative probability distribution function of x can be inverted to find the quantiles of the $N(0, 1)$ distribution with the same cumulative probability. Note that $F(x)$ can be any cumulative distribution function and can even be empirical, given N values $F(x_i) = i/(N + 1)$.

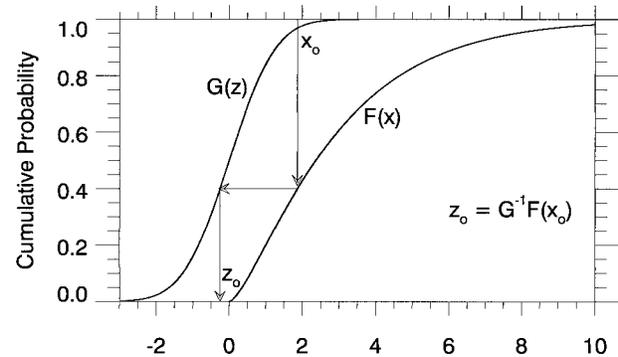


FIG. B1. A graphical illustration of the probability-integral transform for transforming a random variable x with cumulative distribution function $F(x)$ to a standard Gaussian deviate z . In the example shown, x_0 is mapped into z_0 .

The above idea is of course the same as used in transforming a random deviate with uniform density in the interval $[0, 1]$ to a deviate in any interval $[a, b]$ with any specified distribution (e.g., Kendall and Stuart 1977). It is routinely employed, for example, in random number generators. It does not, however, appear to have been used in meteorology to transform precipitation to Gaussian deviates. Figure B1 illustrates the transformation. Note that precipitation takes on only positive values [i.e., $F(x)$ is zero for negative x]. A value x_0 is mapped into z_0 by first dropping a vertical from x_0 to $F(x_0)$, moving horizontally across to $G(z_0)$, and finally dropping vertically to z_0 . All values of x between 0 and $+\infty$ can thus be mapped into values of z between $-\infty$ and $+\infty$.

To transform the precipitation values at each grid point in our GCM runs, $F(x)$ was estimated by fitting a Gamma distribution (Wilks 1995) to the 180 seasonal values in the neutral ensemble. This $F(x)$ was then used in Eq. (B1) to transform the 180 values in the warm and cold ensembles. In this way, the values in the neutral ensemble were guaranteed to map into $N(0, 1)$, and those in the warm and cold distributions to have standardized means and variances relative to $N(0, 1)$.

REFERENCES

- Anderson, J. L., and W. F. Stern, 1996: Evaluating the potential predictive utility of ensemble forecasts. *J. Climate*, **9**, 260–269.
- Barnett, T. P., 1995: Monte Carlo climate forecasting. *J. Climate*, **8**, 1005–1022.
- Barnston, A. G., 1993: Statistical short-term climate predictability in the Northern Hemisphere. *J. Climate*, **6**, 1513–1564.
- Barsugli, J. J., P. D. Sardeshmukh, and S. Zhang, 1997: Identifying the most sensitive areas of tropical SST forcing for midlatitude seasonal prediction. Preprints, *Seventh Conf. on Climate Variations*, Long Beach, CA, Amer. Meteor. Soc., 166–167.
- Chen, W. Y., and H. M. van den Dool, 1997: Atmospheric predictability of seasonal, annual, and decadal climate means and the role of the ENSO cycle: A model study. *J. Climate*, **10**, 1235–1254.
- Crutcher, H. L., 1975: A note on the possible misuse of the Kolmogorov–Smirnov test. *J. Appl. Meteor.*, **14**, 1600–1603.

- Gates, W. L., and Coauthors, 1999: An overview of the results of the Atmospheric Model Intercomparison Project (AMIP). *Bull. Amer. Meteor. Soc.*, **80**, 29–55.
- Hoerling, M. P., A. Kumar, and M. Zhong, 1997: El Niño, La Niña, and the nonlinearity of their teleconnections. *J. Climate*, **10**, 1769–1786.
- Horel, J. D., and J. M. Wallace, 1981: Planetary-scale atmospheric phenomenon associated with the Southern Oscillation. *Mon. Wea. Rev.*, **109**, 2080–2092.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Kanamitsu, M., and Coauthors 1991: Recent changes implemented into the global forecast system at NMC. *Wea. Forecasting*, **6**, 425–435.
- Kendall, M., and A. Stuart, 1977: *The Advanced Theory of Statistics*. 3d ed. Vol. 1. Macmillan, 283 pp.
- Kumar, A., M. P. Hoerling, M. Ji, A. Leetmaa, and P. D. Sardeshmukh, 1996: Assessing a GCM's suitability for making seasonal predictions. *J. Climate*, **9**, 115–129.
- Lau, N.-C., 1985: Modeling the seasonal dependence of the atmospheric response to observed El Niños in 1962–76. *Mon. Wea. Rev.*, **113**, 1970–1996.
- Lilliefors, H. W., 1967: On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *Amer. Stat. Assoc. J.*, **62**, 399–402.
- Livezey, R. E., and W. Y. Chen, 1983: Statistical field significance and its determination by Monte Carlo techniques. *Mon. Wea. Rev.*, **111**, 46–59.
- , and K. C. Mo, 1987: Tropical–extratropical teleconnections during Northern Hemisphere winter. Part II: Relationships between monthly mean Northern Hemisphere circulation patterns and proxies for tropical convection. *Mon. Wea. Rev.*, **115**, 3115–3132.
- Palmer, T. N., and D. L. T. Anderson, 1994: The prospects for seasonal forecasting—A review paper. *Quart. J. Roy. Meteor. Soc.*, **120**, 755–793.
- Parker, D. E., C. K. Folland, A. Bevan, M. N. Ward, M. Jackson, and K. Maskell, 1995: Marine surface data for analyses of climate fluctuations on interannual to century time scales. *Natural Climate Variability on Decade to Century Time Scales*, D. G. Martinson et al., Eds., National Academy Press, 241–250.
- Peng, P., 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.
- Penland, C., and P. D. Sardeshmukh, 1995: The optimal growth of tropical sea surface temperature anomalies. *J. Climate*, **8**, 1999–2024.
- Ropelewski, C. F., and M. S. Halpert, 1987: Global and regional scale precipitation patterns associated with the Southern Oscillation. *Mon. Wea. Rev.*, **115**, 1606–1626.
- , and —, 1996: Quantifying Southern Oscillation–precipitation relationships. *J. Climate*, **9**, 1043–1059.
- Rowell, D. P., 1998: Assessing potential seasonal predictability with an ensemble of multidecadal GCM simulations. *J. Climate*, **11**, 109–120.
- Stephens, M. A., 1970: Use of the Kolmogorov–Smirnov, Cramer–Von Mises and related statistics without extensive tables. *J. Roy. Statist. Soc.*, **B32**, 115–122.
- Stern, W., and K. Miyakoda, 1995: The feasibility of seasonal forecasts inferred from multiple GCM integrations. *J. Climate*, **8**, 1071–1085.
- Trenberth, K. E., 1997: The definition of El Niño. *Bull. Amer. Meteor. Soc.*, **78**, 2771–2777.
- , and C. J. Guillemot, 1996: Evaluation of the atmospheric moisture and hydrological cycle in the NCEP reanalysis. NCAR Tech. Rep. NCAR/TN-430 + STR, 308 pp.
- Whitaker, J. S., and P. D. Sardeshmukh, 1998: A linear theory of extratropical synoptic eddy statistics. *J. Atmos. Sci.*, **55**, 237–258.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.
- Xie, P., and P. A. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78**, 2539–2558.