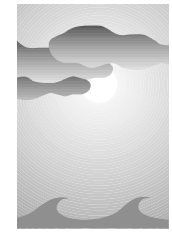
# The IRI Seasonal Climate Prediction System and the 1997/98 El Niño Event



Simon J. Mason,* Lisa Goddard,* Nicholas E. Graham,*
Elena Yulaeva,* Liqiang Sun,* and Philip A. Arkin+

## ABSTRACT

The International Research Institute for Climate Prediction (IRI) was formed in late 1996 with the aim of fostering the improvement, production, and use of global forecasts of seasonal to interannual climate variability for the explicit benefit of society. The development of the 1997/98 El Niño provided an ideal impetus to the IRI Experimental Forecast Division (IRI EFD) to generate seasonal climate forecasts on an operational basis. In the production of these forecasts an extensive suite of forecasting tools has been developed, and these are described in this paper. An argument is made for the need for a multimodel ensemble approach and for extensive validation of each model's ability to simulate interannual climate variability accurately. The need for global sea surface temperature forecasts is demonstrated. Forecasts of precipitation and air temperature are presented in the form of "net assessments," following the format adopted by the regional consensus forums. During the 1997/98 El Niño, the skill of the net assessments was greater than chance, except over Europe, and in most cases was an improvement over a forecast of persistence of the latest month's climate anomaly.

## 1. Introduction: The International Research Institute for Climate Prediction

Until fairly recently, no method was available for reliably predicting the variability of the global climate system on seasonal to interannual timescales. However, beginning in the mid-1960s, observational and diagnostic studies of the ocean and atmosphere began to make it clear that certain behaviors of the coupled system might indeed be predictable, including in particular the El Niño–Southern Oscillation (ENSO) phenomenon (see, among many others, Bjerknes 1966, 1969, 1972; Davis 1976; Horel and Wallace 1981; Wallace and Gutzler 1981; Rasmusson and Carpenter 1982; Shukla and Wallace 1983; Philander 1983; Latif et al. 1998; Neelin et al. 1998; Stockdale et al. 1998a). In 1985, the World Climate Research Programme (WCRP) initiated the Tropical Oceans Global Atmosphere (TOGA) program (WCRP 1985). Under TOGA, increased attention was devoted to the development of physical/mathematical models of the ocean and atmosphere in the tropical Pacific Ocean, as well as to the establishment of observing systems to provide the data such models required. As TOGA progressed, the Tropical Atmosphere Ocean Array was designed and implemented (Hayes et al. 1991; McPhaden et al. 1998), and models began to show evidence of the capability to make useful real-time predictions (Cane and Zebiak 1985; Cane et al. 1986).

By 1990, research results inspired an international group of scientists to prepare a plan for the International Research Institute for Climate Prediction (IRI). This plan outlined the scientific background justifying the creation of the institute as well as the benefits to be derived from its implementation. Following the

*International Research Institute for Climate Prediction, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California.
+International Research Institute for Climate Prediction, Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York.
*Corresponding author address:* Dr. Simon J. Mason, International Research Institute for Climate Prediction, Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0235.
E-mail: simon@lacosta.ucsd.edu
In final form 19 May 1999.

International Forum on Forecasting El Niño: Launching an International Research Institute (6–8 November 1995, Washington, D.C.), the National Oceanic and Atmospheric Administration (NOAA) committed itself to support an interim IRI, and to facilitate the transition to a multinational institute. The IRI was established in late 1996 at Lamont-Doherty Earth Observatory and Scripps Institution of Oceanography (SIO) (Carson 1998).

The mission of the IRI is to foster the improvement, production, and use of global forecasts of seasonal to interannual climate variability for the explicit benefit of society. To accomplish this mission, the IRI, in coordination and collaboration with the international climate research and applications communities, develops and implements

- improved mathematical models of the physical climate system;
- better techniques for forecasting seasonal to interannual variations in the physical climate system;
- techniques for monitoring such variations, and for disseminating climate monitoring and forecasting information products to potential users;
- more advanced applications of forecasts and other climate information products; and
- methods of training potential users of IRI products so as to ensure the existence of a growing cadre of enthusiastic proponents ready to apply such information to the benefit of their societies.

The IRI consists of four divisions and a training program. These include the Modeling Research Division (MRD), Experimental Forecasting Division (EFD), Climate Monitoring and Dissemination Division (CMD), and Applications Research Division (ARD). The Modeling Research Division aims to provide a continually evolving suite of tools representing the state of the art for seasonal to interannual climate prediction and forecast applications. In the near term, the MRD will focus its efforts on the development and implementation of better models of the coupled ocean–land–atmosphere system, improving methods of ocean data assimilation, and developing methods for using ensembles of realizations from individual models as well as from multiple models, in collaboration with the Experimental Forecast Division. The Experimental Forecast Division, currently located at SIO, is the operational forecasting arm of the IRI. Its mission is to produce regularly the most useful and accurate possible forecasts of seasonal to interannual climate variability at global and regional scales. The division took an active role in providing seasonal climate forecasts throughout the 1997/98 El Niño episode. The Climate Monitoring and Dissemination Division aims to provide the formal distribution of IRI products, emphasizing service to the applications community. The CMD maintains the datasets required to permit IRI scientists to monitor current climatic conditions and conduct their research, and distributes IRI forecasts and other products, both directly to users in the Climate Information Digest and via the IRI Web site. The climate information digest contains a synopsis of current climatic conditions, a description of the most recent IRI forecasts, and a summary of the impacts of both current and expected events on several applications sectors. The aim of the Applications Research Division is to maximize the utility and accessibility of climate forecast products for societies around the world, building on the efforts initiated by NOAA (Buizer et al. 1999). It strives to facilitate the use of IRI forecast guidance products to improve planning and decision making in climate-sensitive sectors, and to demonstrate how climate information can enhance sustainable economic growth and reduce vulnerability to climate-related hazards including extreme events and disease outbreaks. The training program is an outreach effort whose ultimate purpose is to increase the global awareness of the need for IRI products. It seeks to create a cadre of international experts, from a variety of economic and social sectors, each of whom understands and is able to make efficient application of the kinds of climate products that the IRI generates.

The principal purpose of this paper is to describe the developing IRI climate forecast system and its performance during the intense El Niño of 1997/98. This introduction is followed by an explanation of the historical and intellectual basis for the IRI EFD forecast process. The modeling methodology and the statistical tools used are described in sections 3 and 4. Section 5 discusses the validation of the forecast process, while section 6 describes the future plans for the operation.

## 2. Forecasting at the IRI Experimental Forecasting Division

The IRI operational forecast system is based on the premise that the atmospheric response to sea surface temperature (SST) variability provides the potential to produce forecasts of seasonal climate anomalies for

many areas of the world (Palmer and Anderson 1994; Shukla 1998). Recent advances in understanding and modeling of the global ocean–atmosphere system have permitted the production of such forecasts on a regular basis. A key development has been the ability demonstrated in the mid-1980s to forecast SSTs in the equatorial Pacific Ocean with lead times of as much as a year (Zebiak and Cane 1987), although lead times of about six months are now considered more realistic (Barnston et al. 1994, 1999a). The El Niño phenomenon constitutes the strongest signal in the interannual variability of global SST and exhibits major effects on climate variability in many parts of the world (Ropelewski and Halpert 1987, 1989; Halpert and Ropelewski 1992; Trenberth et al. 1998). However, its effects in some areas are far less robust, and the climate in these parts of the world may instead be affected by SST variability in ocean basins other than the Pacific. Seasonal forecasts of global atmospheric anomalies therefore depend upon an ability to forecast SSTs in areas beyond the equatorial Pacific.

The development of the 1997/98 El Niño provided an ideal opportunity to generate seasonal climate forecasts on an operational basis. A successful forecast of an El Niño had first been provided in 1986 (Cane et al. 1986). The 1991/92 event was again forecast successfully, notwithstanding some false alarms that were issued in 1990 (Mo 1993; Ropelewski et al. 1993). Forecasts of the 1997/98 event were unique, however, in that for the first time forecasts of the event, and of seasonal atmospheric climate anomalies, were made widely available to the general public (Barnston et al. 1999b; Buizer et al. 1999). Forecasts of a strong El Niño were available to the public as early as June 1997 (Barnston et al. 1999a). The IRI EFD took an active role in forecasting this El Niño and global climate anomalies during the 1997/98 season. In this paper, some of the tools and methods that contributed to the climate forecasts issued by the IRI EFD during the height of the 1997/98 El Niño are described.

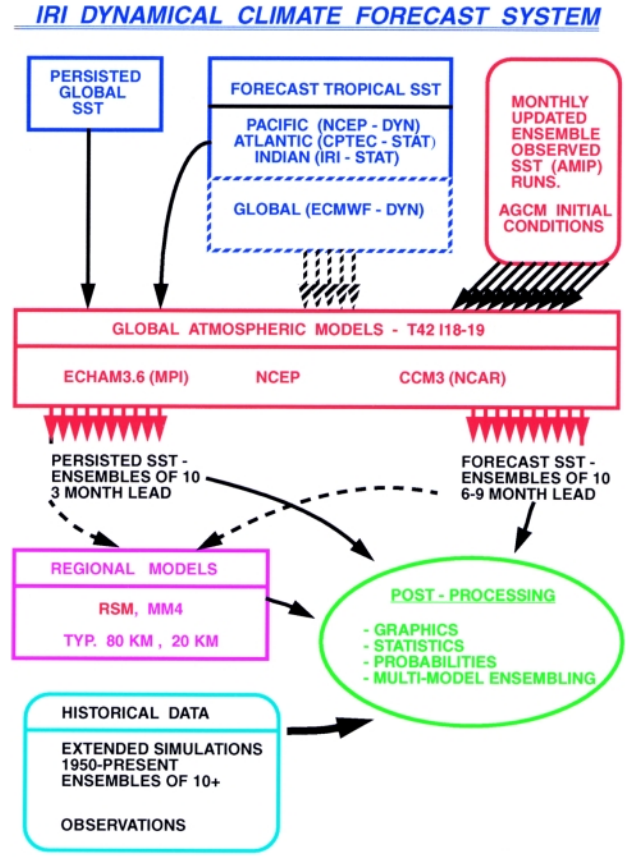



FIG. 1. Schematic diagram illustrating the structure of the IRI two-tiered numerical forecast system during the 1997/98 El Niño.

## 3. Two-tiered dynamical modeling

The IRI EFD uses a two-tiered approach to dynamical seasonal climate prediction (Fig. 1): SST forecasts are produced first, which then serve as the lower boundary condition forcing for an atmospheric general circulation model (AGCM) (Bengtsson et al. 1993; Palmer and Anderson 1994). The procedure for constructing a global SST field for the upcoming season(s) is not obvious, however, since the models used to produce forecasts of El Niño/La Niña generally are confined to the tropical Pacific Ocean. For many parts of the world, even those places where climate variability apparently is related strongly to El Niño/La Niña SSTs, it is not possible to simulate correctly the climate variability with SST anomalies defined only in the tropical Pacific. For example, Indian Ocean SST anomalies are critical for simulating the proper climate signal over eastern Africa (Goddard and Graham 1999). Similarly, the tropical Atlantic Ocean has an important modulating effect on El Niño's impact over South America (Moura and Shukla 1981; Wallace et al. 1998) and western Africa (Lamb and Peppler 1991; Rowell et al. 1995). Hence predicted SSTs must be prescribed over all tropical ocean basins.

### *a. Global sea surface temperature forecasts*

The simplest SST prediction involves the persistence of the current observed SST anomalies on top of the observed seasonal cycle. Retrospective 3-month forecasts indicate that in many areas and most seasons

this approach to SST prediction is almost as skillful as if one had known the true SST values for the forecast period. Thus persisted SST anomaly forecasts are useful for short-term climate forecasting, particularly when there is no significant ENSO variability present and the evolution of SSTs may be difficult to determine. For longer lead times, and when an ENSO extreme of either sign is developing or decaying, such as in 1997 and 1998, it is preferable to use a prediction of evolving SST anomalies.

Operational forecasts of SST anomalies for the tropical Pacific were obtained from the National Centers for Environmental Prediction (NCEP) coupled model (Ji et al. 1998). The model Pacific, and hence the SST forecasts, are constrained to cover the area from 30°N to 25°S and from 120°E to 70°W. The ocean model is initialized by assimilating observed surface and subsurface data. The ocean model is initialized by assimilating observed surface and subsurface data. The ocean model is then coupled to the atmospheric model, and the coupled model integration extends 6 months into the future. The coupled model uses stress, heat, and salinity flux anomalies obtained from the coupled interaction between ocean and atmosphere models added to observed climatological fluxes to provide the forcing to the OGCM and AGCM. The resulting SST fields are statistically corrected offline.

These forecasts of SST anomalies for the tropical Pacific Ocean were supplemented by statistical forecasts of SSTs for the Indian and Atlantic Oceans. The Indian Ocean usually warms during El Niño and cools during La Niña episodes (Pan and Oort 1983; Cadet 1985; Meehl 1993; Hastenrath et al. 1993; Latif et al. 1994; Latif and Barnett 1995; Nagai et al. 1995; Nicholson 1997; Landman and Mason 1999), with the tropical Pacific SST anomalies leading by approximately three months (Nicholson 1997; Goddard and Graham 1999). A canonical correlation analysis (CCA) model has been developed on the basis of this lagged association, relating Indian Ocean SSTs to SSTs in the tropical Pacific Ocean. This method of predicting Indian Ocean SST anomalies is intended only as an interim solution until such time as high quality global SST predictions become available from coupled ocean–atmosphere models. Similarly, the tropical Atlantic Ocean evolution currently is forecast using a CCA model that was developed at Centro de Previsão de Tempo e Estudos Climáticos (CPTEC)/Instituto Nacional de Pesquisas Espaciais (INPE) in Brazil (Pezzi et al. 1998). In the mid- and high latitudes of all three oceans, observed sea surface temperature anomalies were slowly damped with an *e*-folding time of 90 days.

*b. Global atmospheric forecasts*

Because of the requirement to provide global forecasts, a multimodel ensemble approach has been adopted so that the weaknesses of one AGCM can be offset by the strengths of another. In the initial selection of AGCMs, a wide range of models was considered with the aim of identifying a manageable subset that could be used in the production of the operational forecasts. The criterion used was the models' ability to simulate observed climate variability; this skill can be tested in a number of ways. One method used to assess the skill of the ensemble mean variability involves a comparison of the primary modes of observed and simulated variability. The first two empirical orthogonal functions (EOFs) from the observed data are calculated for various continental-scale regions, and the model simulations are projected onto those. A correlation matrix is then formed between the actual (observed) and projected (model) EOF amplitudes to determine how much of the dominant observed variability is captured by the model. Various performance scores can be calculated to determine the percentage variability that can be extracted from the model fields and has helped determine which AGCMs would be chosen for operational work.

After testing a wide range of AGCMs, the IRI EFD selected three to be used operationally: ECHAM3 (Max Planck Institute), MRF9 (NCEP), and CCM3 (National Center for Atmospheric Research). While these three models perform approximately equally well on average, differences in model performance for specific variables in different regions and seasons are evident. Since the IRI EFD produces forecasts for the international community, it was essential that a multimodel ensemble approach be taken. All three models were run at T42 resolution (approximately 2.8° of latitude and longitude) with vertical resolutions of 19, 18, and 18 levels, respectively, and all were forced with both persisted and forecast SSTs. Predictions using persisted SSTs were confined to 3 months because of the rapid loss of skill at longer leads, but predictions out to 6 months were generated using forecast SST anomalies. A comparison of scenarios based on persisted versus evolving SST anomalies provided information on the sensitivity of the climate system to the evolving boundary layer forcing and thus facilitated the interpretation of AGCM predictions.

Each month an ensemble of at least 10 runs was computed for each SST scenario and each AGCM, thus providing a total ensemble size of more than 60. Debate still exists over the optimal number of ensemble members, but a compromise must be found between an acceptable number and the amount of computer time required to generate the ensemble. In most cases an ensemble of 10 appeared to be sufficient to sample the probability distribution of the seasonal climate variability with a given AGCM and SST scenario. Because initialization shocks may result from using AGCM-based reanalysis products as initial conditions in alternative AGCMs, the individual ensemble members were initialized from continually updated simulations using observed sea surface temperatures, and thus they did not include any information about the actual initial conditions of the observed atmosphere.

## 4. Statistical tools

The raw output of numerical models, whether of the atmosphere alone or of the coupled ocean–atmosphere system, is not enough to permit a forecast to be made. The great sensitivity of the behavior of the atmosphere to small differences in initial conditions makes it necessary to use ensembles of many model executions, but it adds a complex step in the need to interpret the ensemble output. Statistical analyses of the model outputs therefore are required. In addition, at the present state of such models, statistical predictions from historical data can provide useful ancillary information.

*a. Ensemble means*

Ensemble methods typically are used to indicate the range of possible climate outcomes for a given SST boundary forcing (Milton 1990; Murphy 1990; Mureau et al. 1993; Tracton and Kalnay 1993; Déqué et al. 1994; Harrison 1995; Anderson 1996), but methods of presenting and interpreting the forecast information contained within the ensemble are not obvious. In addition, because of model errors and biases, the ensemble distribution may not accurately represent the true distribution of possible outcomes. The IRI EFD has adopted and developed a range of methods for displaying AGCM ensemble output.

The simplest form of presentation of ensemble information is the ensemble mean (Fig. 2a). Maps of the ensemble mean climate predictions together with information based on the historical performance of the model are produced. A mask is applied to the model-predicted anomaly map such that the anomalies are shaded only in regions where the model skill, as measured by the correlation between the ensemble mean anomaly and the observed anomaly, is statistically significant (Fig. 2b). For the temperature predictions, the anomaly fields are plotted; for the precipitation, the anomalies are expressed as an absolute quantity (mm day$^{-1}$) (Fig. 3) and as a relative quantity (% normal, where 100% normal implies no anomaly) (Fig. 4).

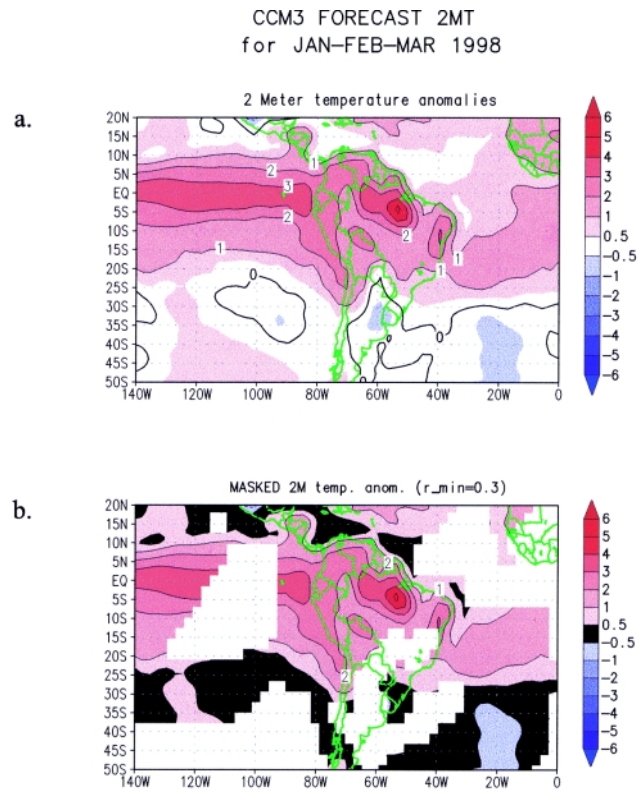



FIG. 2. (a) Ensemble mean prediction of January–March 1998 air temperature anomalies (°C) relative to a 1961–90 climatology. The prediction was produced in December 1997 and consists of 10 ensemble members from the CCM3 model forced with forecast tropical Pacific and Indian Ocean SSTs, with persisted anomalies elsewhere. In (b) the temperature anomalies are masked where the correlation between simulated and observed variability does not exceed the 90% confidence level for this season over the years 1950–94.

The correlation mask indicates where the model simulates the observed climate variability reasonably well. Because the correlation coefficient is an imperfect measure of model performance (Potts et al. 1996), the correlation mask may hide the output in some places where there is some model skill. In other places,

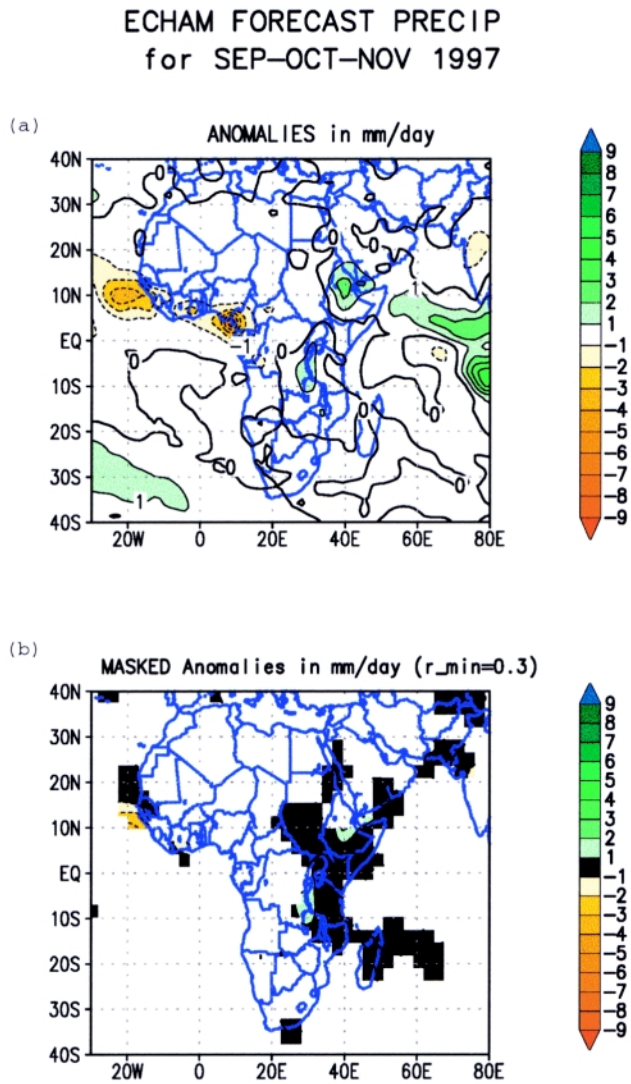



FIG. 3. (a) Ensemble mean prediction of September–November 1997 anomalous precipitation rates (mm day$^{-1}$) relative to a 1961–90 climatology. The prediction was produced in September 1997 and consists of 10 ensemble members from the ECHAM3 model forced with forecast SST anomalies for the tropical Pacific and Indian Oceans, and persisted anomalies elsewhere. In (b) the anomalous precipitation rates are masked where the correlation between simulated and observed variability does not exceed the 90% confidence level for this season over the years 1950–94.

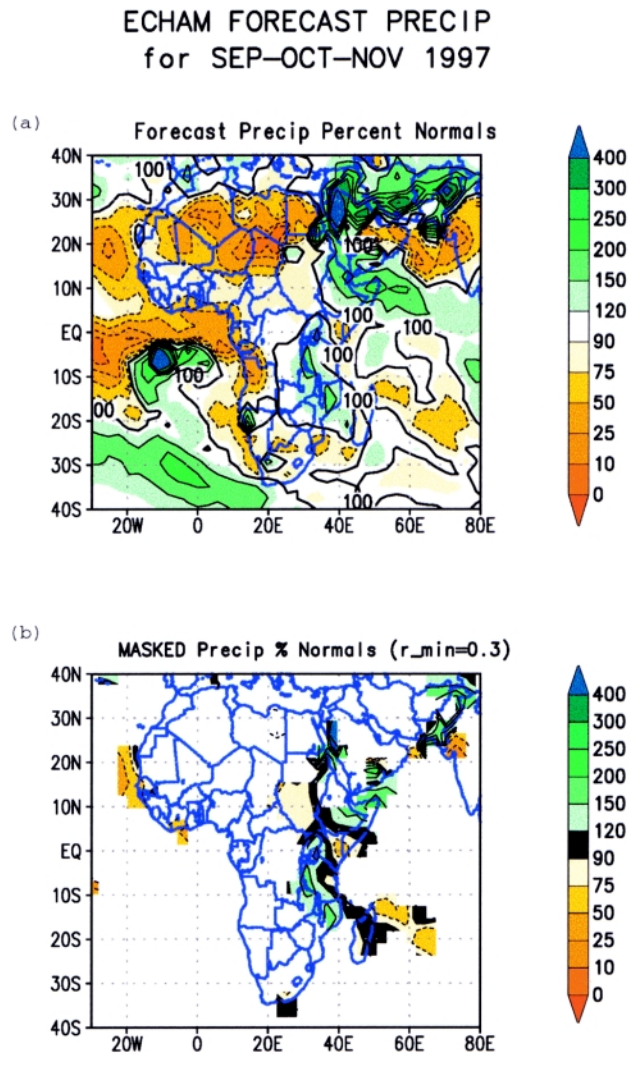



FIG. 4. (a) Ensemble mean prediction of September–November 1997 anomalous precipitation as a percentage of average seasonal rainfall relative to a 1961–90 climatology. The prediction was produced in September 1997 and consists of 10 ensemble members from the ECHAM3 model forced with forecast SST anomalies for the tropical Pacific and Indian Oceans, and persisted anomalies elsewhere. In (b) the anomalous precipitation is masked where the correlation between simulated and observed variability does not exceed the 90% confidence level for this season over the years 1950–94.

systematic model biases may result in poor correlations with observed climate variability, but simple corrections to the model output may provide some useful information. One approach to capturing this information is to compile "ensemble mean contingency tables," thus comparing the historical performance of the model to the observations according to tercile classifications rather than by some form of squared error statistic, such as the correlation coefficient. The climate anomalies are classified such that, in the case of precipitation (temperature), the wettest (warmest) third of the values is defined as "above normal," the middle third is "near normal," and the driest (coldest) third is "below normal." The contingency table is built by counting how often the observed anomaly was above, near, or below normal given the predicted category. A table is built for each grid point for the season under consideration, and the information is displayed as

four maps (Fig. 5). The first shows the spatial probabilities associated with an above-normal prediction (Fig. 5a), the second shows the probabilities of near normal (Fig. 5b), the third shows the probabilities of below normal (Fig. 5c), and the fourth is a rebuilt forecast (Fig. 5d). The rebuilt forecast highlights a category that has been assigned a probability of at least 50%. For those points for which two neighboring categories are approximately equally likely, and the third has been assigned a probability of less than 30%, then the rebuilt forecast indicates the improbability of the "unlikely category." Following this logic, the rebuilt forecast consists of five categories: "dry," "not wet," "near normal," "not dry," and "wet," in the case of precipitation.

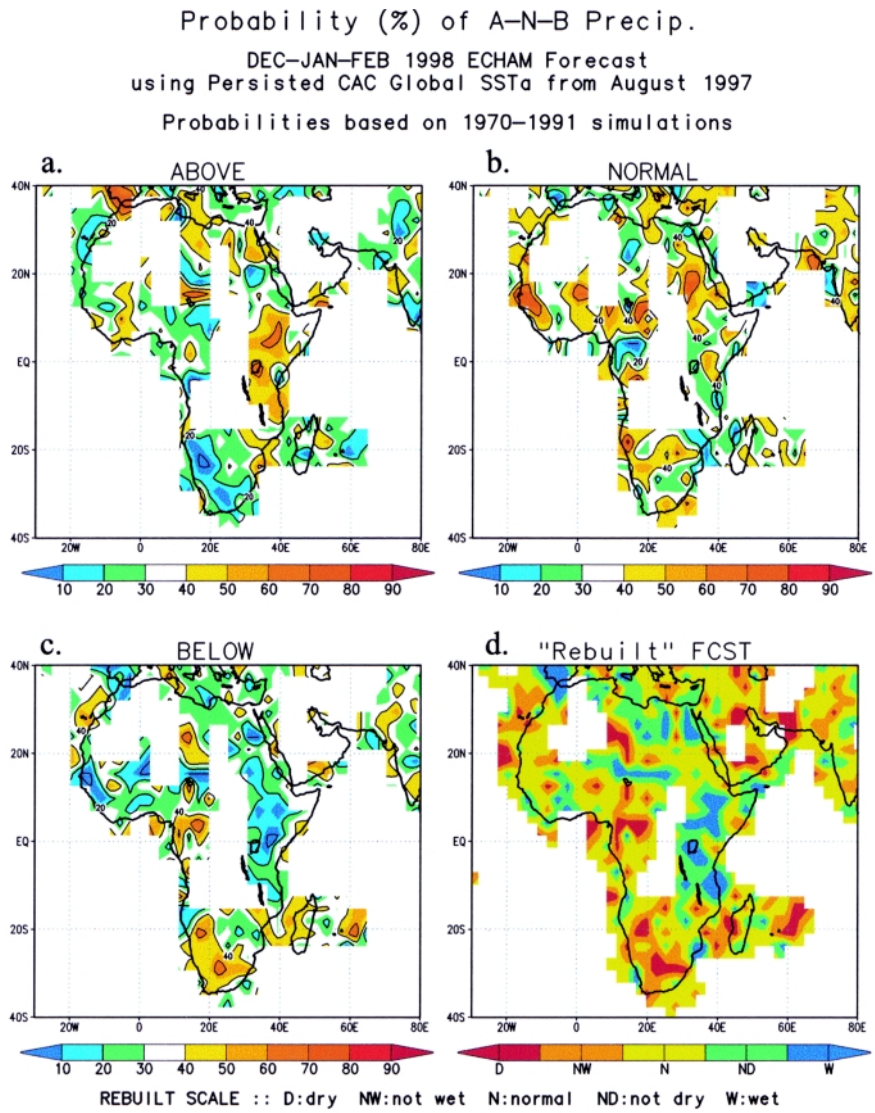



FIG. 5. Probability that the precipitation will fall into the (a) above-normal, (b) near-normal, or (c) below-normal tercile, given the current ensemble mean forecast shown in Figs. 2 and 3 and the previous behavior of the AGCM's ensemble mean climate response; (d) a "rebuilt" forecast based on the category shown in (a)–(c) that has the greatest probability. Full details of how the rebuilt forecast is made are provided in the text.

*b. Estimation of forecast probabilities*

In addition to displaying the ensemble mean predictions and forecast probabilities estimated from the ensemble mean, several methods for extracting and displaying the information contained in the distribution of the ensemble members are used. Although it can be demonstrated that the ensemble mean provides a more skillful forecast than an individual ensemble member, there can be additional important information in the ensemble spread. For a number of predefined regions, the distribution of the ensemble members, the ensemble mean, and the observational anomaly are plotted for each year (Fig. 6a). The plot indicates the year-to-year variability in the spread of the ensemble members, how that spread may be related to the accuracy of the prediction, and whether the model has performed better for positive or negative anomalies. In addition, the role of extreme ensemble members and the appearance of bimodal distributions can be investigated from the plot. The historical distribution of the ensemble members is then used to construct a climatological probability distribution function/curve for the region, against which the distribution of the current ensemble of predictions from the same model can be compared (Fig. 6b). A climate signal in the regional prediction should appear as a discernable shift of the forecast distribution relative to the climatological distribution.

Another simple approach for calculating and representing forecast probabilities is to calculate the percentages of ensemble members with positive or negative anomalies, or that fall within the upper, middle, and lower terciles. Maps showing the percent-

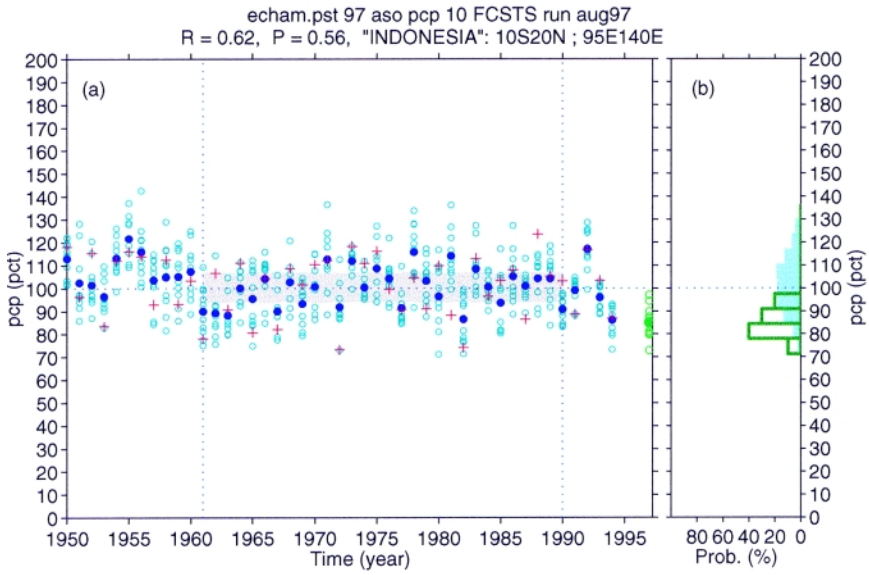



FIG. 6. (a) Historical performance of the ECHAM3 model for the August–October season compared to observations averaged over the "Indonesia" region (10°S–20°N, 95°–140°E). The open blue circles show the model anomaly for individual ensemble members (expressed as a percentage of long-term mean), solid blue circles show the ensemble mean, and red crosses indicate the observed anomaly. The green circles are for the current forecast, and the solid green circle represents the ensemble mean. The gray-shaded area indicates the range of the near-normal tercile based on the climatological period 1961–90. The numbers at the top of the graph indicate the correlation between the ensemble mean simulation and the observed anomalies ($R$) and the tercile hit score ($P$). (b) Distribution of forecast members for August–October 1997 (open green bars) relative to the climatological distribution (solid blue bars).

ages of ensemble members simulating temperature and rainfall anomalies in each tercile are produced by the IRI EFD for each AGCM and sea surface temperature scenario. Similar maps are produced indicating the percentage of ensemble members in the upper and lower 15th percentiles. An example is provided in Fig. 7, showing the ECHAM3 prediction for Africa for September–November 1997 from August 1997, using persisted SST anomalies. The model indicated high probabilities of extremely wet conditions over much of eastern Africa. In some areas in the region over 500% of the long-term average rainfall for September–November was received.

While it can be demonstrated that even when the probabilities are unreliable, such probabilistic forecast information is of potentially greater value to decision makers than deterministic forecasts (Thompson 1962; Murphy 1977; Krzysztofowicz 1983), clearly it is preferable for the forecast probabilities to correspond with the observed relative frequency of the forecast event as closely as possible (Murphy and Winkler 1987). Some adjustment for reliability is therefore made to the ensemble percentages based on the capture rates of the ensembles. The adjustments are usually small and do not fully correct for model biases to the same extent as the contingency table–based approach described above.

### c. *Statistical inflation of ensemble size*

Given that model ensembles are used to estimate the probability distribution of possible climate outcomes, the model's response to boundary forcing should sample the observed distribution as accurately and with as much resolution as possible. Larger ensembles provide greater resolution in defining the shape of the probability distribution (Buizza et al. 1998), but considerable computer resources are required to produce large ensemble sizes. A simple, computationally efficient, nonparametric statistical approach for inflating ensemble sizes has been developed (Graham et al. 1999, manuscript submitted to *Mon. Wea. Rev.*). This method, known as ensemble likelihood values from inferred statistics, is based upon the assumption that, apart from the influence of the SST boundary conditions, monthly values within any individual ensemble member are independent. The assumption, if valid, permits seasonal values to be calculated by combining monthly values from different ensemble members. This method of inflating the ensemble size appears to provide a notable improvement in forecast skill in areas where there is a high degree of ensemble scatter and positive model skill.

### d. *ENSO-related climate probabilities*

The studies of Ropelewski and Halpert (1987, 1989) and Halpert and Ropelewski (1992) provide a guide to the expected climatic impacts of ENSO extremes (i.e., El Niño and La Niña). Simply knowing that an ENSO episode is under way appears to be sufficient to make a forecast that would improve upon one predicting that conditions would be the same as the average for that time of year over the past decade or two (such a forecast

is referred to as climatology). The IRI forecast process uses historical manifestations of ENSO extremes as a guide to check and improve upon the model results.

The historical impacts of El Niño and La Niña events on rainfall anomalies can be illustrated effectively by calculating the observed percentage of times that seasonal rainfall has been in the upper, middle, and lower climatological terciles during ENSO extremes. Three-month mean NIÑO3 indices were calculated using the Kaplan et al. (1998) sea surface temperature data, and the warmest 10 episodes since 1950 were identified for each season. The number of times that the observed precipitation and temperature anomalies during these warm extremes were in each tercile were calculated for each grid point and were expressed as relative frequencies (Fig. 8). These values give some indication of the likelihood of observing a climate anomaly in each of the categories during strong El Niño or La Niña years. The significance of the difference between the calculated percentages and those expected by chance can be calculated using the hypergeometric theorem, or by permutation methods.

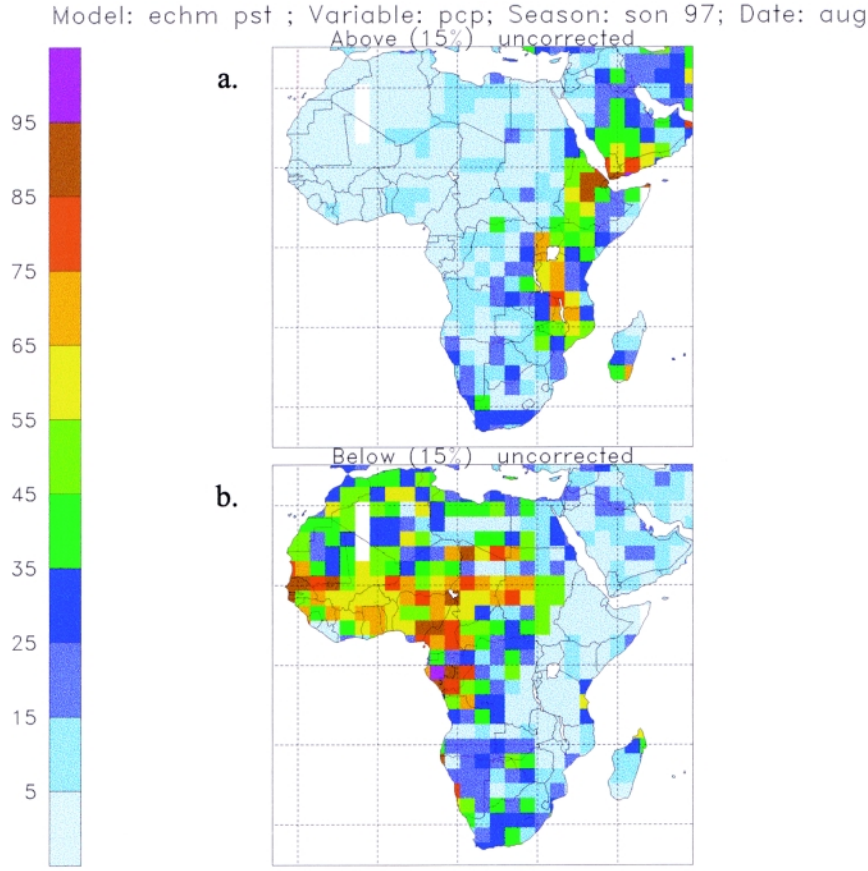



FIG. 7. Percentages of ECHAM3 ensemble member predictions for September–November 1997 in the extreme 15th percentiles. The predictions were produced in August 1997 using persisted sea surface temperature anomalies.

*e. Analog years*

The value of using ENSO conditions in general as a guide to likely climate anomalies is limited. ENSO episodes vary in strength and configuration, and some consideration must be given to these variations. Given the extreme nature of the 1997/98 sea surface temperature anomalies in the equatorial Pacific and tropical Indian Oceans, the atmospheric response to the global boundary layer almost inevitably would be dominated by the Indo–Pacific forcing. Therefore some similarities to the 1982/83 El Niño, which was of an approximate equal strength (peak NIÑO3.4 and monthly Southern Oscillation index anomalies actually exceed those in 1997/98), could be expected. Maps of 3-month climate anomalies during 1982/83 were produced, to highlight areas with marked climate anomalies, which would possibly suggest strong climate responses to major ENSO episodes. The AGCM predictions for 1997/98 were compared with those for 1982/83 to help indicate systematic model biases. As an example, a systematic northeastward shift of a band of anomalously wet conditions over central southeastern South America was found to occur (Fig. 9), and so forecasts were adjusted accordingly.

*f. Process*

The various model predictions that are generated each month by the IRI EFD are not released to the general public. Instead the model predictions and statistical inputs are combined in a subjective manner to produce a "net assessment," an example of which is illustrated in Fig. 10. Inputs from external sources such as the European Centre for Medium-Range Weather Forecasts (ECMWF), national weather services (e.g., Barnston et al. 1999b), and the various regional consensus climate outlooks are considered when available.

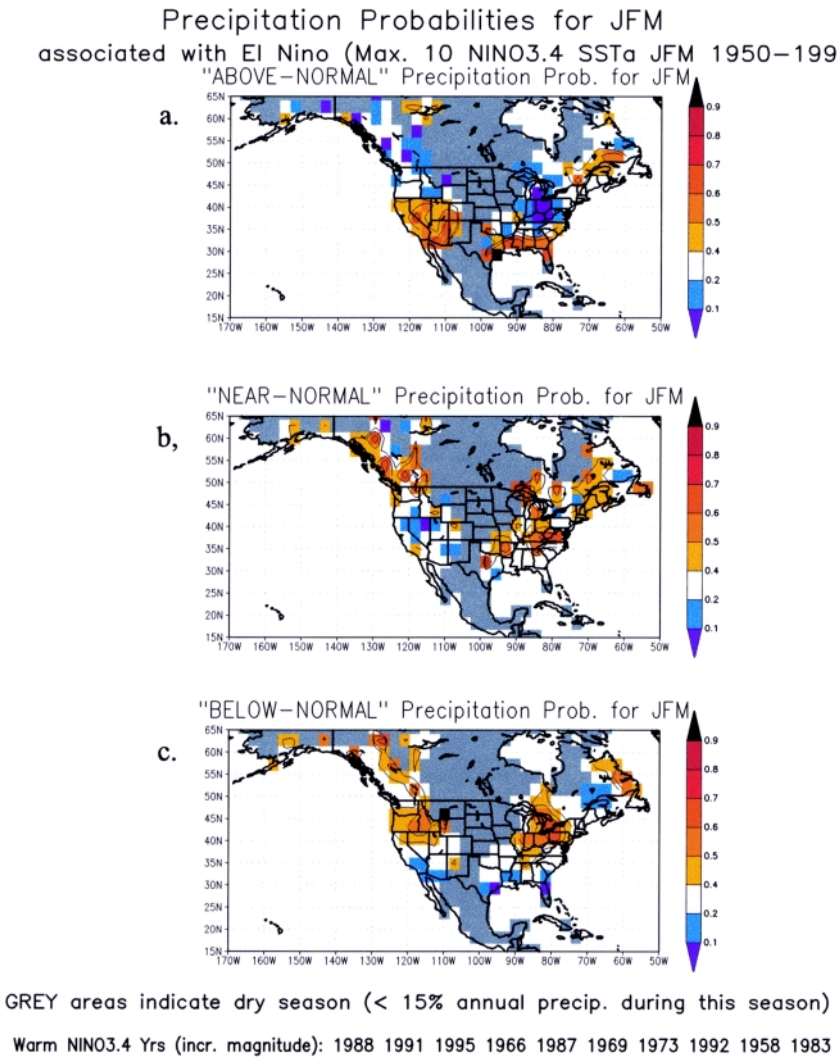



FIG. 8. Frequency of occurrence of (a) above-normal, (b) near-normal, and (c) below-normal January–March precipitation during the 10 strongest El Niño episodes since 1950 (as measured by the average SST anomaly within NIÑO3.4: 5°S–5°N, 120°–170°W).

The respective value of all available information is assessed using available skill measurements, and a best-estimate forecast is provided in the net assessments. The net assessment forecasts are expressed in terms of probabilities of the respective season's rainfall (temperature) being in the wettest (warmest) third, in the driest (coldest) third of years, and in the third centered upon the climatological median. Indications of inflated risk of precipitation anomalies being in the extreme 15th percentiles have been provided since April 1998. The April–June 1998 first risk of extremes map is indicated in Fig. 11a. The extreme forecasts, including those for April–June 1998, have consistently shown a strong tendency to underforecast, such that warnings were not provided for many extreme events that did occur. Nevertheless, the hit rate of the forecasts has been high (Fig. 11b).

The production of the net assessments follows a procedure that is somewhat similar to that used in the regional consensus forecast forums. Output from these regional forums constitutes an important input to the IRI net assessments. Where available, regions are delimited using boundaries that have been defined by regional experts using various statistical methods, including EOFs and cluster analysis. The results are largely unpublished, but in many cases, most notably in Africa, the regions have been defined at preforum workshops. In areas such as southern Africa, the same predefined regions are used at each consensus meeting but may be combined in varying ways into larger areas where forecasts for the current season are identical. Where regional expertise is unavailable, the regions used in the net assessments are defined at the IRI on the basis of model skill and forecast signal, as well as an objective analysis of interannual variability based on known SST-related climate patterns. The forecast regions are defined by subjectively delimiting areas where there is some skill in at least one of the models, and where there is an indication of a spatially coherent signal in the current model predictions.

Once a region has been subjectively delimited, or provided by regional expertise, estimates of the forecast probabilities are made. Probabilities provided by the regional forecast forums are revised if the net assessment is for a slightly different period or if new predictions have become available since the regional consensus was reached. Where no regional guidance is provided, the probabilities are estimated from the following considerations: the level of agreement between different dynamical and statistical model pre-

dictions, the strength of the signal in the various predictions, the respective skill levels of the models, and subjective confidence in the different predictions, based on confidence in the various SST forecasts used to force the AGCMs, for example.

Clearly there is considerable scope for automating and reducing the subjective component involved in defining the forecast regions and estimating the forecast probabilities. Work is in progress to combine the various model predictions objectively, using information about historical performance, and to provide model output for predefined regions rather than model grid points, possibly including some form of statistical correction. In the meantime, operational requirements necessitate the large subjective element involved in the production of the net assessments. The operational production of the forecasts therefore is an evolving process as new tools are gradually developed; the process did not remain constant through the 1997/98 episode. Sea surface temperature scenarios used to force the AGCMs have developed from tropical Pacific-only forecasts in the early stages of the El Niño, to global tropical sea surface temperatures that are used currently. Postprocessing tools have been developed on an ongoing basis, and a growing list of externally produced forecasts and comments has been considered in the production of the net assessments.

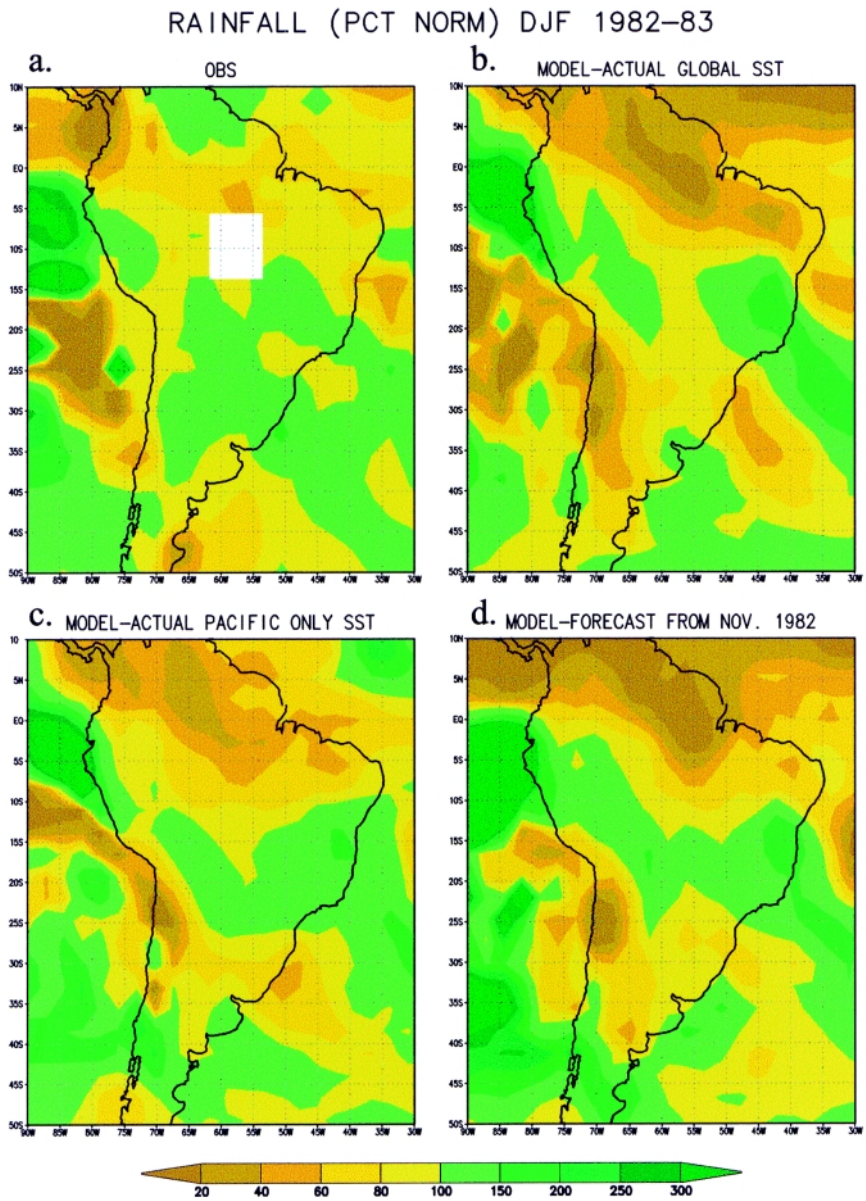



FIG. 9. Comparison of (a) observed December–February precipitation anomalies during the 1982/83 El Niño with simulated anomalies by the ECHAM3 model forced with (b) observed SSTs, (c) observed Pacific SSTs and climatological SSTs elsewhere, and (d) forecast tropical Pacific SSTs and climatological SSTs elsewhere.

## 5. Forecast validation

Because of systematic errors in the models used, it is essential that estimates of confidence in the model predictions be made. Such estimates provide the forecaster with indications of strengths and weaknesses in the various models, and assist in the subjective combination of the model outputs and statistical products into the net assessment forecasts. Many of the statistical methods used to validate the models are used to verify the forecasts produced operationally. Forecast verification is an essential process for ensuring credibility for users, and for monitoring forecast reliability. Here the observational datasets used, the processes through which skill estimates are derived, and limited efforts to determine the quality of forecast performance are described.

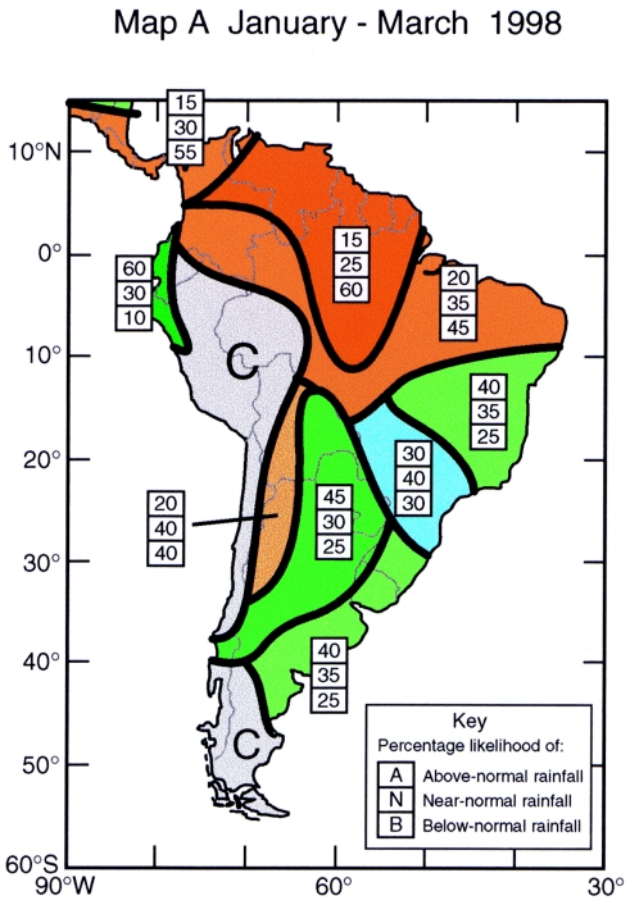



FIG. 10. Net assessment for January–March 1998 precipitation produced in January 1998.

*a. Observational datasets*

Production of meaningful climate forecasts requires extensive critical evaluation of the AGCM output based on comparisons with observational data. Good agreement between model-simulated anomalies and observed anomalies over a sufficiently long period is necessary before a forecast can be made with any confidence. Therefore, to test the skill of model hindcasts, the availability of high quality observational datasets is imperative.

The Jones et al. (1997) air temperature dataset, which contains SST data over the oceans, has been used. The dataset extends from 1854 to the present, but only data for the period 1950–94 are used because of the preliminary nature of the data since 1994. The spatial coverage is nearly complete; the only region lacking consistent data is confined to central Brazil. This dataset displays a larger and more realistic interannual variance in seasonal temperature anomalies in the midlatitudes than others that were examined.

A merged precipitation dataset was formed with the intention of creating a high quality dataset with the maximum global coverage, for the longest possible record. Precipitation data from Hulme (1994) serve as the foundation. This dataset covers the period 1900–95, although preliminary data are available for more recent years. The station data from which this gridded dataset has been constructed are an extension of the original Climate Research Unit (CRU)/U.S. Department of Energy data described in Eischeid et al. (1991). Additional work by Hulme and the CRU at University of East Anglia extended the station time series and increased the station network to include over 9000 stations. These station data have undergone extensive quality control. However, there are no data over the oceans and several regions over land are not covered by the Hulme dataset. These holes have been filled by using other precipitation datasets. The Global Historical Climatology Network (GHCN) precipitation data cover a long period and are based on nearly identical station data as the Hulme dataset, but they have not been as rigorously controlled. During the years 1950–94, if the Hulme data are missing, and the GHCN are not, then the GHCN value is used. After 1979, satellite measurements of rainfall are available. The Climate Prediction Center Merged Analysis Precipitation (CMAP) dataset (Xie and Arkin 1996, 1997, 1998) uses several estimates of precipitation as measured by satellite over land and ocean, as well as gauge data over land. Beginning in 1979, CMAP data are used over the ocean and over land where neither Hulme nor GHCN data exist.

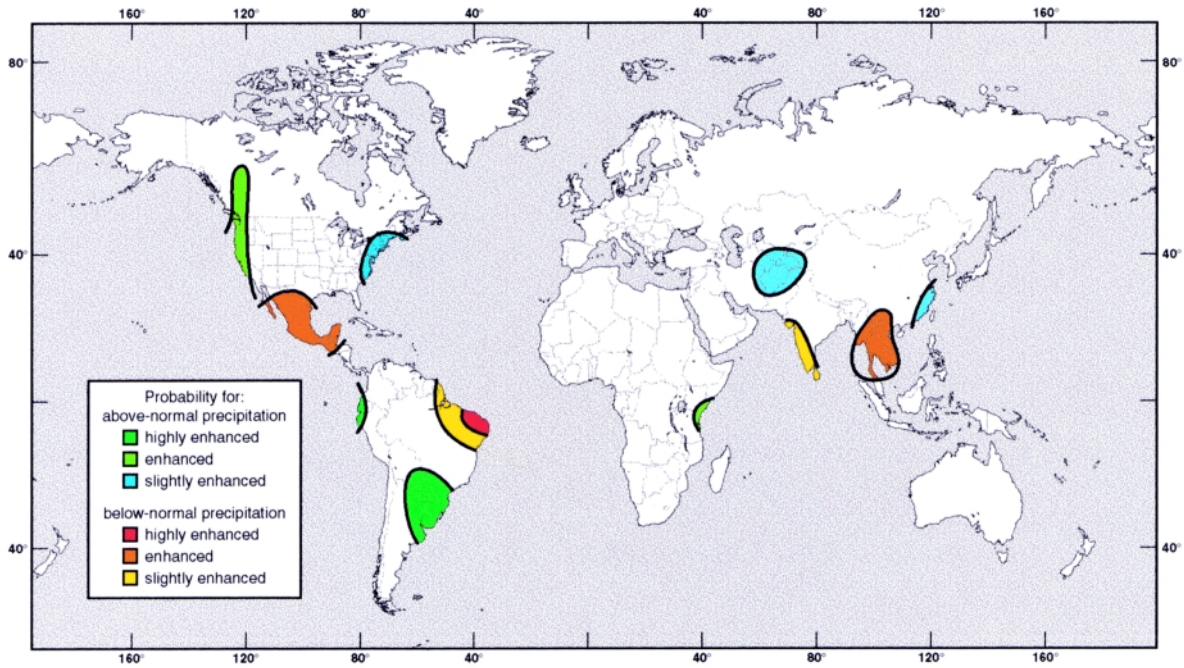
Recently use has been made of the reanalysis data from NCEP (Kalnay et al. 1996) and from ECMWF (Gibson et al. 1997). These data represent a hybrid between observations and models, since the observations are fed through models to produce dynamically consistent atmospheric fields. The reanalysis data potentially can indicate where discrepancies between models and observations originate. To the extent that reanalysis data do represent nature, they are useful for diagnostic studies, particularly for fields such as moisture fluxes that are not observed directly.

FIG. 11. (a) Net assessment for April–June 1998 indicating areas with enhanced risks of precipitation anomalies being within the wettest and driest 15% of climatological occurrences. The forecast was produced in April 1998. (b) Observed precipitation anomalies for April–June 1998 where the observed anomaly was in the wettest and driest 15% of climatological occurrences.

# Global Precipitation Anomalies

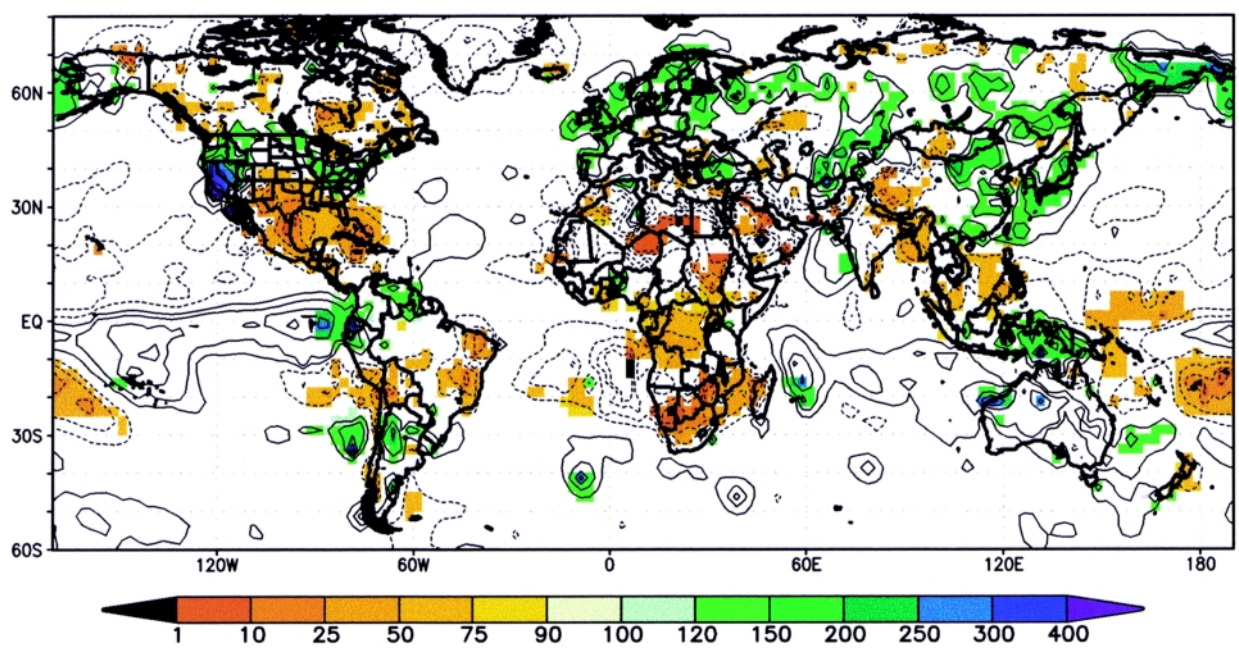
## April - June 1998

a. Forecast (produced April 1998)




b. Observations (% of average)

Shaded for "EXTREME Wet" (above 85%–ile) and "EXTREME Dry" (below 15%–ile)
[CAMS_OPI data, courtesty of NCEP/CPC]



*b. Long runs and retrospective forecasting*

It is essential to have long enough records to define reliable estimates of the skill of AGCM simulations. At the same time, it is necessary to have sufficient ensemble members for each model in order to reduce the climatic noise in the model simulations (of course it is not possible to remove the climatic noise from the observations) and to obtain reliable estimates of ensemble variability. In order to have confidence in the validation statistics, and to determine how stable the patterns of variability are, simulations of at least 25 years are required; longer runs are preferable. The limit of how far back in the past the simulations are useful is dependent on the quality of the SST data used to force the model and also on the quality of the observational climate data against which the model simulations are compared.

To get consistently long runs with sufficient estimates of ensemble variability, simulations with all three AGCMs covering the period from 1950 to the present for at least 10 ensemble members have been completed. The ECHAM3 simulations were forced with the Reynolds reconstructed SST data (Reynolds and Smith 1994) for the period 1950–96 and Reynolds optimally interpolated (OI) SST data (Reynolds 1988) beginning in 1997. NCEP provided a 13-member ensemble for the years 1950–94. All 13 NCEP runs were forced with Reynolds reconstructed SSTs. Recently, runs using CCM3 were completed yielding a set of 10 simulations also covering 1950–94; five of those runs were forced with Reynolds reconstructed SSTs throughout the whole period, and five were forced with Reynolds reconstructed SSTs through 1980 and OI data beginning in 1981.

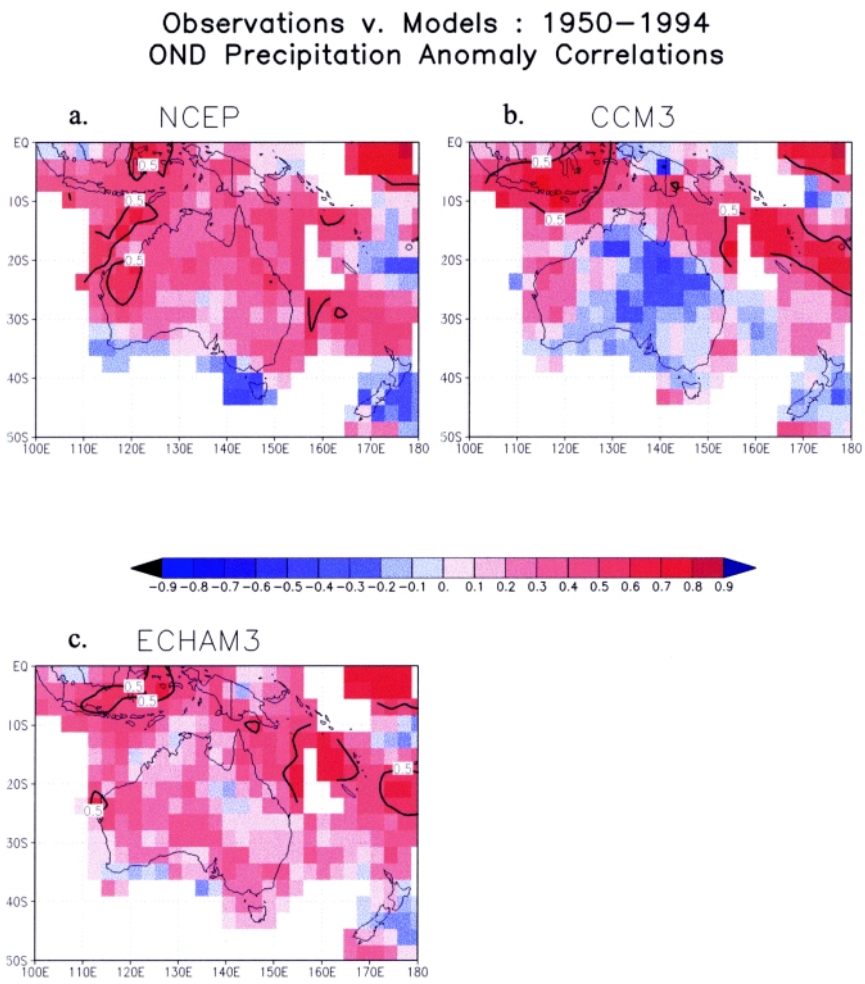



FIG. 12. Correlations between observed and ensemble mean October–December precipitation for the (a) NCEP, (b) CCM3, and (c) ECHAM3 models. The correlations were calculated using data for 1950–94.

It is unfortunate that the long run simulations of the different atmospheric models are not forced with the same SST fields. Given the uncertainty in the observational data, though, the differences in SST fields for the historical simulations are assumed small in comparison to differences in model formulation. However, a small but noticeable improvement in model skill is evident in the years following the switch to the optimally interpolated data in the CCM3 model. How much this difference in skill is attributable to the quality of SST data, how much to the quality of the observations against which the models are compared, and how much to a potential increase in predictability during that period is, as yet, undetermined.

Because the long run simulations are forced with observed SSTs, they provide estimates of the confidence that can be placed in the model output given perfect SST forecasts. These Atmospheric Model Intercomparison Project (AMIP) style runs give an estimate of potential predictability that could be achieved

given perfect SST forecasts. Clearly the persistent and forecast SST fields used in the operational forecasts are imperfect, and so the true confidence may be overestimated. In order to estimate the loss of predictability incurred when using imperfect SST forecasts, retrospective forecasts using persisted SST anomalies have been produced using the ECHAM3 model. These retrospective forecasts have been completed for the period 1970 to date for the December–February, March–May, June–August, and September–November seasons and are validated in the same manner as the long runs, as discussed below.

*c. Model validation*

The long run simulations of the AGCMs used by the IRI EFD, and the retrospective forecast skill of the ECHAM3 model, have been validated extensively. Correlations of ensemble mean simulated precipitation and air temperature with the respective observations have been calculated on a grid-by-grid basis using the 45 yr of long AMIP-style and 23 yr of retrospective runs (Fig. 12). Correlations for selected area averages are calculated in addition (Fig. 6). All calculations are performed using 3-month average precipitation rates and temperatures. In general, correlations for temperature are greater than for precipitation, and they are higher for both variables in the Tropics than they are in the midlatitudes. However, the models do show significant skill over many regions in the midlatitudes for at least part of the year. Comparing the correlation maps of different models (Fig. 12) indicates which model(s) are most likely to produce a reliable forecast over a particular area, for a particular variable in a particular season.

Similarly, relative operating characteristic (ROC) scores (Swets 1973; Mason 1982; Mason and Graham 1999) for precipitation and temperature have been calculated for all 3-month seasons. For all three AGCMs the scores (defined as the area beneath the ROC curve as plotted on linear axes) have been calculated using the ensemble mean, a set of 10 ensemble members, and statistically inflated sets of ensemble members (Graham et al. 1999, manuscript submitted to *Mon. Wea. Rev.*). Events have been defined for the lower, upper, and middle terciles, and the lower and upper 15th percentiles. As with the correlations, the ROC scores have been calculated on a grid-by-grid basis and for selected area averages. For the retrospective runs, ROC scores have been calculated using the five ensemble members available. In general, the ROC scores mirror the areas and seasons of high predictability evident from the correlation maps. However, the ROC curves provide additional information indicating the models' biases toward accurate simulations of wet or dry conditions. Consistently, the curves indicate poor predictability of near-normal conditions for both precipitation and temperature in all areas and all seasons. In some cases, the models appear to be able to simulate accurately above- but not below-normal conditions, or vice versa. As an example, both the ECHAM3 and NCEP models have higher skill in simulating wet, rather than dry, conditions of the September–November eastern African rains (Mason and Graham 1999). In some cases, the ROC curve is able to indicate predictability not evident from the ensemble mean correlations. For example, there is some indication that the ECHAM3 model is able to simulate accurately dry conditions of the March–May eastern African rains (Fig. 13), which are generally considered difficult to predict (Mutai et al. 1998).

*d. Sensitivity experiments*

Sensitivity experiments have been conducted to determine the qualitative effects of SST variability in

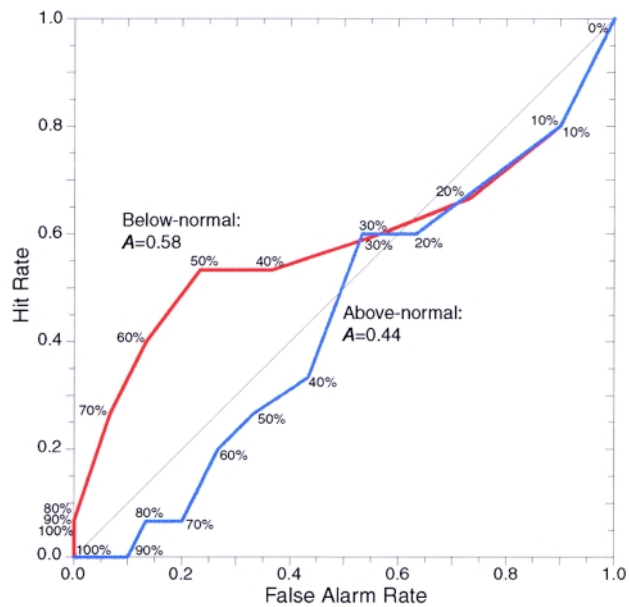



FIG. 13. ROC curves for March–May area-averaged rainfall for eastern Africa (10°N–10°S, 30°–50°E) from 1950 to 1994. The hit and false alarm rates were calculated using rainfall simulated by the ECHAM3–T42 general circulation model forced with observed sea surface temperatures and using 10 ensemble members. Results are shown for the simulation of rainfall in the upper (blue), middle (green), and lower (red) terciles. Rates are indicated using different minimum percentages of ensemble members simulating rainfall in the respective tercile to issue a warning, as indicated by the values along the curves. The areas, A, beneath the curves are indicated also.

various tropical oceans. In these experiments, the SSTs in a particular ocean basin were varied, and in the other ocean areas seasonally varying climatological SST values were prescribed. To date, two of these experiments have been completed, one for the tropical Pacific (20°S–20°N, 120°E–75°W) and one for the Indian Ocean (north of 40°S and west of 120°E). These experiments are referred to as Pacific Ocean Global Atmosphere and Indian Ocean Global Atmosphere. It was found that the Indian Ocean SST anomalies are critical for simulating the proper climate signal over southern and eastern Africa (Goddard and Graham 1999). Rainfall variability in eastern Africa shows a strong association with the occurrence of ENSO. However, the simulation indicates that the teleconnection between ENSO and rainfall in this region of Africa occurs through the Indian Ocean. Thus, to predict correctly seasonal climate anomalies over eastern Africa, it is necessary to predict Indian Ocean temperatures in addition to predicting tropical Pacific SSTs.

*e. Forecast value*

Initial estimates of forecast value have been made based on the $2 \times 2$ contingency table estimates of forecast quality that form the basis of the ROC curves described above. The contingency table can be used to calculate a set of hit and false alarm rates for a set of forecasts of predefined meteorological or climatological events, such as the seasonal rainfall total being within the driest 10% of historical values. The same contingency table is used in the cost–loss model of forecast value estimation (Murphy 1994), by defining the cost of mitigation against an expected event, the mitigated loss if the event occurs, and the loss incurred if no warning is provided. Forecast value can then be calculated given information on event frequency, forecast quality, and on the costs and losses. From a set of interviews with forecast users in the southern African region, cost–loss tables were completed for events of interest to the users. Using the ECHAM3 retrospective forecasts for December–February rainfall, averaged over the area approximately between 23° and 36°S, and between 15° and 32°E, savings of the order of at least U.S. $10 billion to $10 trillion per year were indicated.

*f. Forecast performance*

No quantitative measure of forecast performance has been constructed as yet. One obvious difficulty in producing such a metric is that the format of the IRI forecasts is probabilistic, as is generally the case with climate forecasts. Thus any single forecast can hardly be said to have any errors, since no zero probability terciles are forecast and so any observed value can be said to have been allowed for. Nevertheless, the methods used to validate a set of forecasts made through time can be applied to estimate the performance of a specific forecast over an area. Heidke skill scores (Wilks 1995) for the precipitation net assessments produced by the IRI for the 12-month period October 1997–September 1998 are summarized in Tables 1 and 2. The Heidke skill score treats the forecasts deterministically and so gives a rather crude estimate of their skill but it does provide a useful and relatively simple estimate of skill. The net assessments are being verified more comprehensively, and there are plans to report on the results in detail elsewhere.

In general, the skill of the net assessments for precipitation during the lifetime of the 1997/98 El Niño was highest for the Southern Hemisphere, at the shorter lead time. The net assessment forecast skill for South America was consistently high, while skill for Europe and Asia was low. In most cases the forecasts have been more skillful than chance, and they were a small improvement over forecasts of persistence of the last month's climate anomaly and over the use of pure ENSO-related climate statistics. It is expected that the quality of the net assessment forecasts will improve as the tools used in their production are developed, and the experience of the forecasters increases.

## 6. Future plans

The IRI EFD operational climate forecast system has proven itself already to be a valuable contribution to a global community that needs to find better means of protecting itself from variations in weather and climate. In order to improve upon this performance, several means of enhancing the numerical model forecasts have been identified, and a number of these are being implemented. The main areas of development include an improvement in spatial resolution through downscaling and by higher-resolution AGCMs, and the implementation of a fully coupled ocean–atmosphere model in the IRI EFD. In the shorter term, an agreement has been reached with ECMWF to validate and experiment with the use of the SST forecasts from their fully coupled model (Stockdale et al. 1998b). Some initial progress in these areas is detailed here.

In an attempt to downscale the information provided by the AGCMs at T42 resolution, the NCEP

TABLE 1. Heidke skill scores for the 0-month lead precipitation net assessments. The scores in italic are for a forecast of persistence of the latest month's climate anomaly (middle column) and for a forecast based on the frequency of occurrence of above-normal, near-normal, and below-normal precipitation during the 10 strongest El Niño episodes since 1950 (as measured by the average SST anomaly within NIÑO3: 5°S–5°N, 90°–150°W) (third column).

| | OND 1997 | | | JFM 1998 | | | AMJ 1998 | | | Average | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Africa** | 14.4 | *–9.1* | *2.2* | 13.9 | *20.7* | *1.7* | 25.2 | *22.7* | *–6.0* | 17.8 | *11.4* | *–0.7* |
| **Asia** | 5.2 | *3.8* | *–8.7* | –11.9 | *–34.3* | *3.3* | 11.3 | *19.8* | *5.7* | 1.5 | *–16.8* | *0.1* |
| **Australasia** | 39.0 | *45.3* | *1.8* | 36.3 | *50.0* | *37.5* | 11.1 | *6.5* | *–57.1* | 28.8 | *33.9* | *–5.9* |
| **Europe** | 16.5 | *–35.9* | *–17.6* | –59.3 | *–5.1* | *28.1* | –7.6 | *–24.5* | *16.8* | –16.8 | *–21.8* | *9.1* |
| **North America** | 12.2 | *27.3* | *–9.4* | 12.1 | *14.1* | *–8.0* | –4.9 | *29.3* | *–6.3* | 6.5 | *23.6* | *–7.9* |
| **South America** | 31.3 | *9.9* | *8.9* | 35.9 | *25.7* | *22.8* | 18.5 | *–5.2* | *12.2* | 28.6 | *10.1* | *14.6* |
| **Globe** | 14.5 | *8.9* | *–4.5* | 6.3 | *8.2* | *6.4* | 11.1 | *5.3* | *2.6* | 10.6 | *7.5* | *1.5* |

nested regional spectral model (RSM) has been used to simulate September–January rainfall variability over eastern Africa for the period 1970/71 to 1996/97 (Sun et al. 1999, manuscript submitted to *J. Geophys. Res.*). The regional model has been run at a resolution of 80 km with a very high resolution 20-km nest over Kenya. There are 19 vertical layers. The nested system has realistic vegetation and detailed topography. The outputs from the ECHAM3 atmospheric climate model provide the large-scale circulation forcing for the nested system. From two numerical integrations, an ensemble mean has been calculated and validated using an observational network of over 300 rainfall stations throughout Kenya. The nested system captures both

TABLE 2. Heidke skill scores for the 3-month lead precipitation net assessments. The scores in italic are for a forecast of persistence of the latest month's climate anomaly (middle column) and for a forecast based on the frequency of occurrence of above-normal, near-normal, and below-normal precipitation during the 10 strongest El Niño episodes since 1950 (as measured by the average SST anomaly within NIÑO3: 5°S–5°N, 90°–150°W) (third column).

| | OND 1997 | JFM 1998 | | | AMJ 1998 | | | Average | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Africa** | N/A | 12.0 | *23.3* | *1.7* | 4.1 | *2.2* | *–6.0* | 8.1 | *12.8* | *–2.1* |
| **Asia** | N/A | –7.1 | *–27.9* | *3.3* | 12.8 | *–8.7* | *5.7* | 2.9 | *–18.3* | *4.5* |
| **Australasia** | N/A | 43.1 | *52.8* | *37.5* | –15.4 | *–20.9* | *–57.1* | 13.9 | *16.0* | *–9.8* |
| **Europe** | N/A | –22.2 | *–6.2* | *28.1* | –28.2 | *–20.5* | *16.8* | –25.2 | *–13.4* | *22.5* |
| **North America** | N/A | 9.4 | *11.9* | *–8.0* | 1.1 | *23.0* | *–6.3* | 5.3 | *17.5* | *–7.1* |
| **South America** | N/A | 34.7 | *24.4* | *22.8* | 9.6 | *–6.2* | *12.2* | 22.2 | *9.1* | *17.5* |
| **Globe** | N/A | 9.7 | *10.4* | *6.4* | 3.3 | *1.6* | *2.6* | 6.5 | *6.0* | *4.5* |

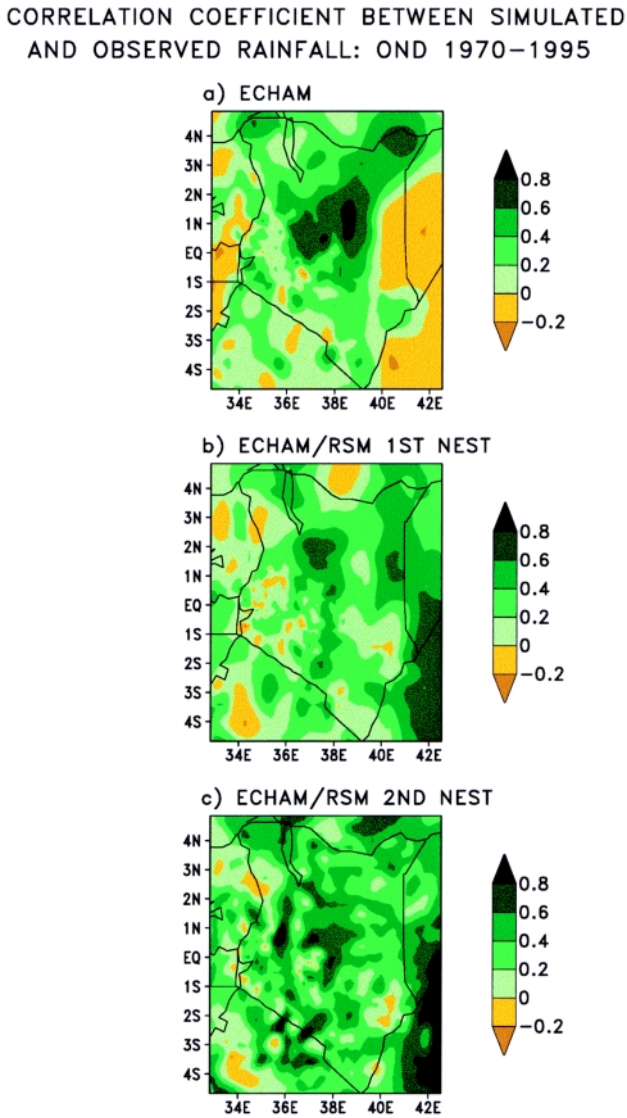



FIG. 14. Correlation between simulated and observed October–December rainfall over Kenya from 1970 to 1996 for (a) the ECHAM3–T42 model, (b) the RSM model at 80-km resolution, and (c) the RSM model at 20-km resolution.

the large-scale characteristics of the circulation and the terrain-induced local features such as local precipitation maxima. The model reproduced the observed interannual variability of the Kenyan short rains and of the timing of the onset of rains in most of the years. Gridpoint correlations between simulated precipitation at the 20-km resolution and the observations exceed 0.4 over much of Kenya (the critical correlation at the 95% level of confidence is 0.38) and locally exceed 0.8 over the 27-yr period in a few places (Fig. 14).

An alternative method for improving the spatial detail of the numerical predictions is to run the AGCMs at a higher resolution (Buizza et al. 1998). The ECHAM3 model has been configured at T106 resolution (approximately 1° horizontal resolution) and forced with the seasonal cycle of SSTs. These runs show improvement in the simulated distribution of annual mean precipitation of the underlying topography, which impacts the simulated climate both locally and remotely (through orographic forcing of atmospheric long waves). An experimental prediction was produced with this high-resolution AGCM during the 1997/98 El Niño, but it was not included in the net assessments because the skill of the model with this configuration has not yet been determined. The model currently is being forced with observed SSTs dating back to 1950 in order to produce an ensemble of AMIP-style simulations similar to those available at T42 resolution.

In addition to improving forecast skill by downscaling methods, the IRI EFD is investigating means of improving the global SST forecasts. An agreement has been reached with ECMWF to use their global SST forecasts as boundary conditions for the AGCMs used by the IRI, as an alternative to the blend of dynamical, statistical, and persistence forecasts that are used currently. Retrospective forecasts have been produced by ECMWF since 1991 (Stockdale et al. 1998b). Although this is too short a period to validate the model robustly, some estimates of the skill of the SST forecasts will be calculated before incorporating these fields into the operational AGCM forecasts. At a later date, the IRI EFD in collaboration with the IRI MRD will use a global, fully coupled ocean–atmosphere model, permitting the eventual execution of a one-tiered forecasting system.

Enhancements in the forecast format are planned, and close collaboration with the IRI CMD and ARD will be maintained to define the most useful directions for improvements and new products. Enhancements to the skill of the net assessments should be attainable through the availability of output from statistical post-processing of the model simulations, which are being tested and implemented. By fostering collaboration with other research groups and national meteorological services, additional gains in skill can be achieved through an extension of the inclusion of regional expertise in the production of the net assessments.

## 7. Concluding remarks

The International Research Institute for Climate Prediction (IRI) was formed in late 1996 with the aim

of fostering the improvement, production, and use of global forecasts of seasonal to interannual climate variability for the explicit benefit of society. The development of an El Niño in early 1997, and its rapid development into a major episode, provided an ideal opportunity to generate and distribute seasonal climate forecasts on an operational basis. The IRI EFD has taken an active role in forecasting this El Niño and global climate anomalies during the 1997/98 season and has continued to produce forecast information during the 1998/99 La Niña.

The IRI EFD has developed a two-tiered forecasting system, in which forecasts of global tropical SSTs are generated first and then used as boundary forcing for a suite of atmospheric models. The need for forecasts of global tropical SSTs has been demonstrated from experience of forecasting climate variability over Africa, where the use of persisted SST anomalies can result in incorrect climate signals in some areas. In producing global climate forecasts, the IRI EFD has made use of a multimodel ensemble approach since no one model is able to provide accurate simulations of climate variability in all regions and all seasons. All the models used have been validated extensively to identify model strengths and weaknesses, and to ensure that it is possible to make reliable estimates of the confidence that can be placed in the model predictions. In addition to the various numerical and statistical models that are run at the IRI, an effort is made to access as much forecast information as possible from centers around the world, in the production of the net assessments.

Forecasts of precipitation and air temperature are presented in the form of "net assessments," following the format adopted by the regional consensus forums, and are distributed via the World Wide Web (http://iri.ucsd.edu/forecasts/net_asmt) and with the assistance of the IRI Climate Monitoring and Dissemination Division (IRI CMD). During the 1997/98 El Niño, the skill of the net assessments for precipitation was greater than chance, except over Europe, and in most cases was an improvement over a forecast of persistence of the latest month's climate anomaly and over ENSO-related climate statistics. Net assessments indicating enhanced risk of extreme precipitation anomalies have been highly successful.

*Acknowledgments.* This paper was funded in part by a grant/cooperative agreement from the National Oceanic and Atmospheric Administration (NOAA). The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its subagencies. The computer assistance of M. Tyree, M. Olivera, J. del Corral, and N. Sheridan is gratefully acknowledged.

## References

Anderson, J. L., 1996: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate,* **9,** 1518–1530.

Barnston, A. G., and Coauthors, 1994: Long-lead seasonal forecasts—Where do we stand? *Bull. Amer. Meteor. Soc.,* **75,** 2097–2114.

——, Y. He, and M. H. Glantz, 1999a: Predictive skill of statistical and dynamical climate models in forecasts of SST during the 1997–98 El Niño episode and the 1998 La Niña onset. *J. Climate,* **12,** 217–244.

——, and Coauthors, 1999b: NCEP forecasts for the El Niño of 1997–98 and its U.S. impacts. *Bull. Amer. Meteor. Soc.,* **80,** 1829–1852.

Bengtsson, L., U. Schlese, E. Roeckner, M. Latif, T. P. Barnett, and N. E. Graham, 1993: A two-tiered approach to long-range climate forecasting. *Science,* **261,** 1027–1029.

Bjerknes, J., 1966: A possible response of the atmospheric Hadley circulation to equatorial anomalies of ocean temperature. *Tellus,* **18,** 820–829.

——, 1969: Atmospheric teleconnections from the equatorial Pacific. *Mon. Wea. Rev.,* **97,** 163–172.

——, 1972: Large-scale atmospheric response to the 1964–65 Pacific equatorial warming. *J. Phys. Oceanogr.,* **2,** 212–217.

Buizer, J. L., J. Foster, and D. Lund, 1999: Global impacts and regional actions: Preparing for the 1997–98 El Niño. *Bull. Amer. Meteor. Soc.,* in press.

Buizza, R., T. Petroliagis, T. Palmer, J. Barkmeijer, M. Hamrud, A. Hollingsworth, A. Simmons, and N. Wedi, 1998: Impact of model resolution and ensemble size on the performance of an Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.,* **124,** 1935–1960.

Cadet, D. L., 1985: The Southern Oscillation over the Indian Ocean. *J. Climatol.,* **5,** 189–212.

Cane, M. A., and S. E. Zebiak, 1985: A theory for El Niño and the Southern Oscillation. *Science,* **228,** 1085–1087.

——, ——, and S. C. Dolan, 1986: Experimental forecasts of El Niño. *Nature,* **321,** 827–832.

Carson, D. J., 1998: Seasonal forecasting. *Quart. J. Roy. Meteor. Soc.,* **124,** 1–26.

Davis, R. E., 1976: Predictability of sea surface temperature and sea level pressure anomalies over the North Pacific Ocean. *J. Phys. Oceanogr.,* **6,** 249–266.

Déqué, M., J. F. Royer, and R. Stroe, 1994: Formulation of Gaussian probability forecasts based on model extended-range integrations. *Tellus,* **46A,** 52–65.

Eischeid, J. K., H. G. Diaz, R. S. Bradley, and P. D. Jones, 1991: A comprehensive precipitation data set for global land areas. Tech. Rep. TR051, U.S. Dept. of Energy, Carbon Dioxide Research Division, 81 pp. [Available from Climate Diagnostics Center, 325 Broadway, Boulder, CO 80303.]

Gibson, J. K., P. Kallberg, A. Nomura, A. Hernandez, and E. Serrano, 1997: ERA Description. *ECMWF Re-Analysis Project Report Series,* Vol. 1. [Available from ECMWF, Shinfield Park Reading RG2 9AX, United Kingdom.]

Goddard, L., and N. E. Graham, 1999: The importance of the Indian Ocean for simulating rainfall anomalies over eastern and southern Africa. *J. Geophys. Res.,* in press.

Halpert, M. S., and C. F. Ropelewski, 1992: Surface temperature patterns associated with the Southern Oscillation. *J. Climate,* **5,** 577–593.

Harrison, M. S. J., 1995: Long-range seasonal forecasting since 1980: Empirical and numerical prediction out to one month for the United Kingdom. *Weather,* **50,** 440–448.

Hastenrath, S., A. Nicklis, and L. Greischar, 1993: Atmospheric-hydrospheric mechanisms of climate anomalies in the western equatorial Indian Ocean. *J. Geophys. Res.,* **98,** 20 219–20 235.

Hayes, S. P., L. J. Mangum, J. Picaut, A. Sumi, and K. Takeuchi, 1991: TOGA-TOA: A moored array for real-time measurements in the tropical Pacific Ocean. *Bull. Amer. Meteor. Soc.,* **72,** 339–347.

Horel, J. D., and J. M. Wallace, 1981: Planetary-scale atmospheric phenomena associated with the Southern Oscillation. *Mon. Wea. Rev.,* **109,** 813–829.

Hulme, M. H., 1994: Validation of large-scale precipitation fields in General Circulation Models. *Global Precipitation and Climate Change,* M. Desbois and F. Desalmund, Eds., NATO ASI Series, Springer-Verlag, 387–406.

Ji, M., A. Leetmaa, and V. E. Kousky, 1998: An improved coupled model for ENSO prediction and implications for ocean initialization. Part II: The coupled model. *Mon. Wea. Rev.,* **126,** 1022–1034.

Jones, P. D., T. J. Osborn, and K. R. Briffa, 1997: Estimating sampling errors in large-scale temperature averages. *J. Climate,* **10,** 2548–2568.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

Kaplan, A., M. A. Cane, Y. Kushnir, A. Clement, M. Blumenthal, and R. Rajagopalan, 1998: Analyses of global sea surface temperature 1856–1991. *J. Geophys. Res.,* **103,** 18 567–18 589.

Krzysztofowicz, R., 1983: Why should a forecaster and a decision maker use Bayes theorem. *Water Resour. Res.,* **19,** 327–336.

Lamb, P. J., and R. A. Peppler, 1991: West Africa. *Teleconnections Linking Worldwide Climate Anomalies: Scientific Basis and Societal Impact,* M. H. Glantz, R. W. Katz, and N. Nicholls, Eds., Cambridge University Press, 121–189.

Landman, W. A., and S. J. Mason, 1999: Change in the association between Indian Ocean sea-surface temperatures and summer rainfall over South Africa and Namibia. *Int. J. Climatol.,* in press.

Latif, M., and T. P. Barnett, 1995: Interactions of the tropical oceans. *J. Climate,* **8,** 952–964.

——, A. Sterl, M. Assenbaum, M. M. Junge, and E. Maier-Reimer, 1994: Climate variability in a coupled GCM. Part II: The Indian Ocean and monsoon. *J. Climate,* **7,** 1449–1462.

——, and Coauthors, 1998: A review of the predictability and prediction of ENSO. *J. Geophys. Res.,* **103,** 14 375–14 393.

Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.,* **30,** 291–303.

Mason, S. J., and N. E. Graham, 1999: Conditional probabilities, relative operating characteristics, and relative operating levels. *Wea. Forecasting,* in press.

McPhaden, M. J. and Coauthors: 1998: The Tropical Ocean–Global Atmosphere observing system: A decade of progress. *J. Geophys. Res.,* **103,** 14 169–14 240.

Meehl, G. A., 1993: A coupled air–sea biennial mechanism in the tropical Indian and Pacific regions: Role of the oceans. *J. Climate,* **6,** 31–41.

Milton, S. F., 1990: Practical extended-range forecasting using dynamical models. *Meteor. Mag.,* **119,** 221–233.

Mo, K. C., 1993: The global climate of September–November 1990: ENSO-like warming in the western Pacific and strong ozone depletion over Antarctica. *J. Climate,* **6,** 1375–1391.

Moura, A. D., and J. Shukla, 1981: On the dynamics of droughts in northeast Brazil: Observations, theory and numerical experiments with a general circulation model. *J. Atmos. Sci.,* **38,** 2653–2675.

Mureau, R., F. Molteni, and T. N. Palmer, 1993: Ensemble prediction using dynamically conditioned perturbations. *Quart. J. Roy. Meteor. Soc.,* **119,** 299–323.

Murphy, A. H., 1977: The value of climatological, categorical and probabilistic forecasts. *Mon. Wea. Rev.,* **105,** 803–816.

——, 1994: Assessing the economic value of forecasts: An overview of methods, results and issues. *Met. Apps.,* **1,** 69–73.

——, and R. L. Winkler, 1987: A general framework for forecast verification. *Mon. Wea. Rev.,* **115,** 1330–1338.

Murphy, J. M., 1990: Assessment of the practical utility of extended range ensemble forecasts. *Quart. J. Roy. Meteor. Soc.,* **116,** 89–125.

Mutai, C. C., M. N. Ward, and A. W. Colman, 1998: Towards the prediction of the East Africa short rains based on sea-surface temperature–atmosphere coupling. *Int. J. Climatol.,* **18,** 975–997.

Nagai, T., Y. Kitamura, M. Endoh, and T. Tokioka, 1995: Coupled atmosphere–ocean model simulations of El Niño/Southern Oscillation with and without an active Indian Ocean. *J. Climate,* **8,** 3–14.

Neelin, J. D., D. S. Battisti, A. C. Hirst, F. F. Jin, Y. Wakata, T. Yamagata, and S. E. Zebiak, 1998: ENSO theory. *J. Geophys. Res.,* **103,** 14 261–14 290.

Nicholson, S. E., 1997: An analysis of the ENSO signal in the tropical Atlantic and western Indian Oceans. *Int. J. Climatol.,* **17,** 345–375.

Palmer, T. N., and D. L. T. Anderson, 1994: The prospects for seasonal forecasting. *Quart. J. Roy. Meteor. Soc.,* **120,** 755–793.

Pan, Y.-H., and A. H. Oort, 1983: Global climate variations connected with sea surface temperature anomalies in the eastern equatorial Pacific Ocean for the 1958–1973 period. *Mon. Wea. Rev.,* **111,** 1244–1258.

Pezzi, L. P., C. R. Repelli, P. Nobre, I. F. A. Cavalcanti, and G. Sampaio, 1998: Forecasts of tropical Atlantic SST using a statistical ocean model at CPTEC/INPE - Brazil. *Exp. Long-Lead Forecast Bull.* Vol. 7, No.1, 28–31.

Philander, S. G. H., 1983: El Niño Southern Oscillation phenomenon. *Nature,* **302,** 295–301.

Potts, J. M., C. K. Folland, I. T. Jolliffe, and D. Sexton, 1996: Revised "LEPS" scores for assessing climate model simulations and long-range forecasts. *J. Climate,* **9,** 34–53.

Rasmusson, E. M., and T. H. Carpenter, 1982: Variations in tropical sea surface temperature and surface wind fields associated with the Southern Oscillation/El Niño. *Mon. Wea. Rev.,* **110,** 354–384.

Reynolds, R. W., 1988: A real-time global sea surface temperature analysis. *J. Climate,* **1,** 75–86.

——, and T. M. Smith, 1994: Improved global sea surface temperature analysis using optimal interpolation. *J. Climate,* **7,** 929–948.

Ropelewski, C. F., and M. S. Halpert, 1987: Precipitation patterns associated with El Niño/Southern Oscillation. *Mon. Wea. Rev.,* **115,** 1606–1626.

——, and ——, 1989: Precipitation patterns associated with the high index phase of the Southern Oscillation. *J. Climate,* **2,** 268–284.

——, P. J. Lamb, and D. H. Portis, 1993: The global climate for June–August 1990: Drought returns to sub-Saharan Africa and warm Southern Oscillation episode conditions develop in the central Pacific. *J. Climate,* **6,** 2188–2212.

Rowell, D. P., C. K. Folland, K. Maskell, and M. N. Ward, 1995: Variability of summer rainfall over tropical North Africa (1906–92): Observations and modelling. *Quart. J. Roy. Meteor. Soc.,* **121,** 669–704.

Shukla, J., 1998: Predictability in the midst of chaos: A scientific basis for climate forecasting. *Science,* **282,** 728–731.

——, and J. M. Wallace, 1983: Numerical simulation of the atmospheric response to equatorial Pacific sea surface temperature anomalies. *J. Atmos. Sci.,* **40,** 1613–1630.

Stockdale, T. N., A. J. Busalacchi, D. E. Harrison, and R. Seager, 1998a: Ocean modeling for ENSO. *J. Geophys. Res.,* **103,** 14 325–14 355.

——, D. L. T. Anderson, J. O. S. Alves, and M. A. Balmaseda, 1998b: Global seasonal rainfall forecasts using a coupled ocean-atmosphere model. *Nature,* **392,** 370–373.

Swets, J. A., 1973: The relative operating characteristic in psychology. *Science,* **182,** 990–1000.

Thompson, J. C., 1962: Economic gains from scientific advances and operational improvements in meteorological prediction. *J. Appl. Meteor.,* **1,** 13–17.

Tracton, M. S., and E. Kalnay, 1993: Operational ensemble prediction at the National Meteorological Center: Practical aspects. *Wea. Forecasting,* **8,** 379–398.

Trenberth, K. E., G. W. Branstator, D. Karoly, A. Kumar, N. G. Lau, and C. F. Ropelewski, 1998: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. *J. Geophys. Res.,* **103,** 14 291–14 324.

Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height fields during the Northern Hemisphere winter. *Mon. Wea. Rev.,* **109,** 784–812.

——, E. M. Rasmusson, T. P. Mitchell, V. E. Kousky, E. S. Sarachik, and H. von Storch, 1998: On the structure and evolution of ENSO-related climate variability in the tropical Pacific: Lessons from TOGA. *J. Geophys. Res.,* **103,** 14 241–14 259.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 467 pp.

WCRP, 1985: CLIVAR science plan: A study of climate variability and predictability. WCRP-89, WMO/TD No. 690, ICSU, WMO, UNESCO, 157 pp. [Available from World Meteorological Organization, Case Postale 2300, CH-1211 Geneva 2, Switzerland.]

Xie, P., and P. A. Arkin, 1996: Analyses of global monthly precipitation using gauge observations, satellite estimates and numerical model predictions. *J. Climate,* **9,** 840–858.

——, and ——, 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.

——, and ——, 1998: Global monthly precipitation from satellite-observed outgoing longwave radiation. *J. Climate,* **11,** 137–164.

Zebiak, S. E., and M. A. Cane, 1987: A model El Niño–Southern Oscillation. *Mon. Wea. Rev.,* **115,** 2262–2278.