Predictive Skill of Statistical and Dynamical Climate Models in SST Forecasts during the 1997–98 El Niño Episode and the 1998 La Niña Onset

Anthony G. Barnston,* Michael H. Glantz,+ and Yuxiang He*

ABSTRACT

Critical reviews of forecasts of ENSO conditions, based on a set of 15 dynamical and statistical models, are given for the 1997–98 El Niño event and the initial stages of the 1998–99 La Niña. While many of the models forecasted some degree of warming one to two seasons prior to the onset of the El Niño in boreal spring of 1997, none predicted its strength until the event was already becoming very strong in late spring. Neither the dynamical nor the statistical models, as groups, performed significantly better than the other during this episode. The best performing statistical models and dynamical models forecast SST anomalies of about +1°C (vs 2.5°–3° observed) in the Niño 3.4 region prior to any observed positive anomalies. The most comprehensive dynamical models performed better than the simple dynamical models. Once the El Niño had developed in mid-1997, a larger set of models was able to forecast its peak in late 1997 and dissipation and reversal to cold conditions in late spring/early summer 1998. Overall, however, skill for these recent two years does not appear greater than that found over an earlier (1982–93) period. In both cases, median model correlation skill averaged over lead times of one to three seasons is near or just above 0.6.

Because ENSO extremes usually develop in boreal spring or early summer and persist through the following winter, forecasting impact tendencies in extratropical North America for winter (when impacts are most pronounced) at 5 months of lead time is not difficult, requiring only good observations of the summer ENSO state and knowledge of the winter teleconnections. Because of the strength of the 1997–98 El Niño and the consequent skill of 5-month lead forecasts of U.S. winter 1997–98 impacts, the success of these forecasts was noticed to an unprecedented extent by the general public. However, forecasting impacts in austral winter that occur simultaneously with the initial appearance of an ENSO extreme (e.g., in Chile, Uruguay, Kiribati, Ecuador, and Peru) require forecasting the boreal spring/summer onset of ENSO events themselves at several months of lead time. This latter task is formidable, as evidenced by the fact that formal announcements of an El Niño did not occur until May, leaving little time for users in the above regions to prepare.

Verbal summaries of ENSO forecasts issued to users worldwide during the 1997–98 El Niño event contained ambiguities. To address the needs for forecasts to be expressed verbally for nontechnical users and also to be precise enough for meaningful utility and verification, a simple numerically based verbal classification system for describing ENSO-related forecasts is presented.

1. Introduction

In the last decade and a half, our improved ability to forecast El Niño–Southern Oscillation (ENSO) warm and cold events (El Niño and La Niña, respectively) at longer lead times can be attributed to several factors. Foremost among these are our increasingly improved data observing and analysis/assimilation systems, higher computer speed and storage capacity, and increased understanding of the tropical oceanic

*Climate Prediction Center, NCEP/NWS/NOAA, Camp Springs, Maryland.
+National Center for Atmospheric Research, Boulder, Colorado. NCAR is sponsored by the National Science Foundation.
Corresponding author address: Dr. Anthony G. Barnston, Climate Prediction Center, W/NMC 51, NOAA/NWS/NCEP, World Weather Building, Rm. 604, 5200 Auth Rd., Camp Springs, MD 20746-4304.
E-mail: wd51ab@ncep.noaa.gov
In final form 20 November 1998.

and atmospheric physics underlying the evolution of the ENSO phenomenon. This last achievement has taken the form of new insights into tropical air–sea interactions and their improved representation in coupled dynamical models. Societal interest in using ENSO forecasts has sharply increased as a result of improved understanding about the ENSO phenomenon and its influence on global climate.

Barnston et al. (1994) compared the performances of several dynamical and statistical models in the prediction of tropical Pacific sea surface temperatures (SSTs) over the 1982–93 period. They described the state of the art in ENSO prediction as having progressed to a moderate level, with model correlations with actual observations at about 0.6 for 6-month lead forecasts of 3-month mean conditions (i.e., 7.5 months between the time of the forecast and the center of the predicted period). At that time, dynamical and statistical models showed comparable skills. Forecasts at the 0.6 skill level are useful but far from excellent. That study focused as much as possible on real-time forecasting in which sample forecasts were made without knowledge of the correct answer.

The real-time aspect was stressed because that is the condition under which the forecasts are of service to customers. The performance of statistical forecasts was included because they are much simpler and less expensive to develop. The statistical models serve as a baseline reference against which the skill of the more complex dynamical models can be compared.

This paper serves as an addendum to that 1994 evaluation and also as a case study in climate prediction. During 1997–98, an El Niño occurred that was approximately the same strength as that of the very strong 1982–83 El Niño. Prior to and during 1997–98, real-time forecasts of the ENSO state (using SST in the central Pacific as the indicator) had been made every 3 months and were published in the *Experimental Long-Lead Forecast Bulletin* (henceforth referred to as *ELLF Bulletin*; Barnston 1996a–c, 1997a–d, 1998) issued by the Climate Prediction Center (CPC) of the National Oceanic and Atmospheric Administration (NOAA).[1] Forecasts were made that did not appear in the *ELLF Bulletin*, but this paper examines those that did over much of the period of interest. Forecasts produced from mid-1996 through the beginning of 1998 are examined; these cover the entire El Niño episode and the beginning of the La Niña episode that began in mid-1998.

Evaluating models by their performance on a single episode is, of course, unfair at worst and unrepresentative at best. There is no guarantee that models that performed well (or poorly) in 1996–98 would perform at a similar level for the next strong warm or cold ENSO episode. Even the 12-yr period used in the Barnston et al. (1994) study was considered too short to provide robust conclusions. The main purpose of this study is to get an idea of the general skill of the models in a collective sense and to revisit the question of whether the numerical models have begun to outperform the statistical models. This is addressed in section 4, following information about the data, methods, and models in sections 2 and 3. A second objective of the study, treated in section 5, is to assess the more qualitative, societal aspects of the set of forecasts of the 1997–98 ENSO event. A summary and discussion about the above and related topics are in section 6.

2. Data and analysis method

This section provides information about the data, briefly describes the dynamical and statistical models whose forecasts are examined, and then explains how the evaluations were performed.

a. Data

The SST data used for all the models discussed here come from combinations of the Comprehensive Ocean Atmosphere Data Set (COADS; Slutz et al. 1985) and the National Centers for Environmental Prediction (NCEP) (Reynolds 1988; Reynolds and Smith 1994), the latter being used by all models during and after 1980. Some of the models use an empirical orthogonal functions (EOF) reconstruction of the COADS SST for years prior to 1982 (Smith et al. 1996). Forecasts were made and verified for area-average SST over discrete regions, as specified below. Even observed area-average SST anomalies may differ slightly for the same time and region from one model to another, both because of slight differences among the SST datasets and because the periods over which climatological means are based are different. These interorganization data calibration differences are inevitable given the geographical and political diversity of the forecast producers. For verification, NCEP's optimum interpola-

[1]Beginning in March 1998, the production responsibilities for the *ELLF Bulletin* were transferred from CPC/NOAA to the Center for Ocean–Land–Atmosphere Studies.

tion SST data of Reynolds and Smith (1994) are used, as are NCEP's definitions of the commonly used SST indices such as Niño 3 and Niño 3.4. The forecast data were digitized from maps or graphs that appeared in the *ELLF Bulletin.* While desirable, it is impractical to verify the forecasts of each model using the slightly varying SST datasets and "homemade" SST indices of each institution. The anomalies used in verification are based on the normal defined from the 1961–96 period. The miscalibrations introduced by these decisions, while as high as 0.25°C in some instances, are minor compared with the amplitude of the SST fluctuations in the eastern half of the tropical Pacific during 1997–98.

b. Analysis method

Our interest is in forecasts made from the period immediately prior to the onset of the 1997–98 El Niño through the El Niño episode itself, as well as its ending in late northern spring of 1998 and the onset and initial development of the cold episode in summer and fall of 1998. Because some of the models do not produce forecasts out to a year in advance, we include forecasts made for one-, two-, and three-season leads. For example, suppose a forecast is produced early in December (either using observed data through November or through an earlier time due to delays caused by data processing requirements). Then the target periods examined for that forecast would be February–March–April (3.5-month lead), May–June–July (6.5-month lead), and August–September–October (9.5-month lead). If forecasts are made for individual months rather than for 3-month periods, then the forecasts for March, June, and September would be used in place of the above 3-month periods. While model performance for leads longer than three seasons are of interest, the value of forecasts from one to three seasons of lead time is sufficiently high practically and economically.

The specific forecast target periods used here, and their lead times and start times, are shown in Table 1. The numbers in the cells of Table 1 indicate the start time, as defined below the table. The start times are approximate, as the various models are not all run using the same observed data cutoff times, nor are they run during the same week. Forecasts for most of the target periods are made approximately 3.5 months, 6.5 months, and 9.5 months prior to the center of the middle month. For example, forecasts for February–March–April 1997 are made in November or early December 1996 (3.5-month lead), August or early September 1996 (6.5-month lead), and May or early June 1996 (9.5-month lead). The three lead times can be thought of as roughly one, two, and three seasons of lead time. For forecasts made in February or early March 1998, only one- to two-season lead forecasts are made, making August–September–October 1998 the final season forecasted. Because of the short period used, performances for the three lead times above are pooled together in most of the skill analyses. Thus, 23 forecasts are used for each model (eight start times, seven of which produce forecasts at three lead times, plus two forecasts for the final start time). Most of the forecasts used here appeared in the CPC's *ELLF Bulletin.* There are a few exceptions, such as models whose appearance in the *ELLF Bulletin* began after the beginning of the study period (e.g., the Colorado State University or CLIPER regression model first appeared in the June 1997 *ELLF Bulletin*), or when there was a major model change during the period [e.g., the switch from the Scripps–Max Planck Institute (MPI) hybrid coupled model version 1 to version 3 beginning with the September 1996 *ELLF Bulletin,* or the replacement of the Lamont LDEO-2 by LDEO-3 beginning with

TABLE 1. Forecasted 3-month periods (ASO, for example, refers to the August through October period) and their lead times, as used in this study. There is a total of 23 forecasts. The numbers in the cells indicate start times, as defined below the table. Note that there is up to a month of variation in start time among the models; thus, the indicated lead times are also approximate.

Lead time	ASO 1996	NDJ 1996	FMA 1997	MJJ 1997	ASO 1997	NDJ 1997	FMA 1998	MJJ 1998	ASO 1998
3.5 month	1	2	3	4	5	6	7	8	
6.5 month		1	2	3	4	5	6	7	8
9.5 month			1	2	3	4	5	6	7

Start times:
1) May or early Jun 1996
2) Aug or early Sep 1996
3) Nov or early Dec 1996
4) Feb or early Mar 1997
5) May or early Jun 1997
6) Aug or early Sep 1997
7) Nov or early Dec 1997
8) Feb or early Mar 1998

the March 1998 *ELLF Bulletin* (Chen et al. 1998)]. In the case of the CLIPER, a retroactive real-time (Barnston et al. 1994) version of the four sets of missing forecasts was obtained. Similarly, a set of retroactive real-time forecasts of the Scripps/MPI model version 3 was obtained for the June 1996 start time. In the case of the LDEO2, there is a discontinuity between the December 1997 and March 1998 start times when LDEO3 began. While this transition occurred too late for the LDEO3 performance to be presented as though it had been available in real time, its performance from a hindcasting perspective will be briefly mentioned.

Model skill scores are expressed in two ways: root-mean-square error (rmse) and temporal correlation. Other verification measures such as the spatial anomaly correlation would be equally useful, but are not used here because some of the models only predict SST averaged over index regions. Rmse scores are usually given to reflect the errors in physical units. However, because the target areas vary among the models, rmse is computed on standardized anomalies. This enables the models to be scored after adjusting, for example, for the greater SST variance in Niño 3 than in Niño 3.4. Similarly, standardization by individual season is done to account for the seasonal variation in interannual variability, so that models are penalized for error in the context of the interannual distribution specific to the season, as opposed to actual error. The temporal correlation is computed with respect to long-term statistics rather than the statistics of the short period considered here (see the appendix of Barnston et al. 1994). Thus, the means for the short period are not removed in the computation; rather, the long-term means, relative to which the anomalies are expressed, are removed. In calculating the correlation this way, models that forecast the broad features of the SST anomaly (e.g., positive anomalies for the final half of 1997 and early portion of 1998) are rewarded even if the shorter-term fluctuations within the brief study period are not in phase with those observed.

3. Fifteen forecast models

The 15 models whose real-time forecasts are examined here include 8 dynamical models and 7 statistical models, as shown in Tables 2 and 3, respectively. A brief description of each model follows. These descriptions are intended for readers with a special interest in some of the model features considered important to the groups in charge of the respective models. The descriptions are not uniform across the models; features mentioned for one or more models may not be mentioned for others. For readers mainly interested in a skill summary or seeking more general information, the descriptions that follow may be bypassed without significant loss of continuity.

a. Dynamical models

The Lamont simple coupled model (LDEO1; Cane et al. 1986; Zebiak and Cane 1987) was the first dynamical model to routinely predict the ENSO-related SST fluctuations in the eastern tropical Pacific Ocean. It covers the tropical Pacific region and predicts monthly anomalies using simplified linear shallow water dynamics for both the ocean and atmosphere. However, it uses more complicated nonlinear forms for atmospheric heating and ocean mixed layer thermodynamics. The model is initialized using wind stress anomalies derived from The Florida State University (FSU) analyses (Goldenberg and O'Brien 1981). Originally constructed to simulate ENSO rather than predict it (Zebiak 1984; Cane et al. 1986), the model has been used for retrospective forecasts of the tropical Pacific SST from 1970. The model has remained unchanged since late 1985 and has continuously produced independent, real-time forecasts. Its skill, which is highest in the Niño 3 region (5°N–5°S, 90°–150°W), is thought to be due partly to successful reproduction of the heat storage mechanism in the subsurface western and central equatorial Pacific Ocean (Wyrtki 1985), attributable to ocean wave dynamics. Model output is adjusted for known systematic errors. Forecasts are made for the SST anomaly field across the tropical Pacific basin for lead times out to a year and beyond.

The second version of the Lamont simple coupled model (LDEO2; Chen et al. 1995) is basically the same as LDEO1, except that it has a modified initialization scheme that combines the observed wind field with the model's wind field. This produces more skillful forecasts at the beginning of the forecast run, making possible higher skills at longer leads as well. It should be noted that while LDEO2 did well in retrospective forecasts of Niño 3 during 1971–96, it failed to predict the 1997–98 El Niño. Because this paper deals mainly with that event, poor performance scores can be anticipated for LDEO2. Recent study at Lamont has led to insight into this failure, and a third version of the model (LDEO3; Chen et al. 1998) has been developed, which additionally includes sea level data in the initialization process, producing skillful forecasts for

1997–98 as well as the earlier years in hindcast mode. Unfortunately, LDEO3 became operational too recently (early 1998) to be included in this real-time forecast study. LDEO2's unexpected poor behavior during 1997–98 serves as a reminder of what may occasionally occur with any model.

The Australian Bureau of Meteorology Research Center (BMRC) low-order coupled model is based on the physics of ENSO predictability. A simple ocean model with thermodynamical equations governing the SST is coupled to a simple atmospheric model that performs well when forced by ENSO-related SST anomalies (Kleeman 1991). The coupled model is somewhat similar to LDEO1 (Cane and Zebiak 1987) but differs in aspects of the coupling, atmospheric convection and heating, and ocean thermodynamics. Hindcast skills were tested over the 1972–86 period using FSU winds, and optimal skill occurred when the ocean model SST was determined purely by equatorial thermocline perturbations. Initialization was improved by using a space–time variational (adjoint) technique to assimilate subsurface thermal data, as well as the usual wind data, into the ocean model (Kleeman et al. 1995). The BMRC model is one of the only models (along with the NCEP coupled model) to use subsurface data in its initialization. The BMRC model predicts the Niño 3 region out to over a year in advance.

The original University of Oxford intermediate coupled ocean–atmosphere model has been used to predict SST anomalies in the tropical Pacific in the At-

TABLE 2. Basic information about the dynamical models.

Dynamical model	References	Type of dynamical model	Predictand
LDEO1 simple coupled, Lamont-Doherty Earth Observatory, Palisades, NY	Cane et al. (1987) Zebiak and Cane (1987)	Simplified	Gridded field; Niño 3
LDEO2 simple coupled, Lamont-Doherty Earth Observatory, Palisades, NY	Chen et al. (1995)	Simplified	Gridded field; Niño 3
BMRC low-order coupled, Bureau of Meteorology Research, Melbourne, Australia	Kleeman (1991) Kleeman (1993) Kleeman et al. (1995)	Simplified	Niño 3
University of Oxford intermediate coupled, Oxford, United Kingdom	Balmaseda et al. (1994)	Simplified	Eq 2*: (5°N–5°S, 130°–170°S)
University of Oxford intermediate coupled, Oxford, United Kingdom	Balmaseda et al. (1994)	Simplified; same as above, except forecast defined as change from "forecast" at initial condition time	Eq 2*: (5°N–5°S, 130°–170°S)
Scripps–MPI hybrid coupled, University of California, San Diego La Jolla, CA	Barnett et al. (1993)	Comprehensive ocean, statistical atmosphere	Gridded field; Niño 3.4
COLA coupled, Center for Ocean–Land–Atmosphere Studies, Calverton, MD	Kirtman et al. (1997)	Comprehensive ocean and atmosphere	Gridded field; Niño 3.4
NCEP (NOAA) coupled, CPC/NCEP/NOAA Camp Springs, MD	Ji et al. (1994a) Ji et al. (1994b) Ji et al. (1996)	Comprehensive ocean and atmosphere	Gridded field; Niño 3.4

*Other discrete regions also predicted (this designation not given if gridded field predicted).

TABLE 3. Basic information about the statistical models.

Statistical model	References	Type of statistical model	Predictand
CSU climatology-persistence (CLIPER), CSU, Fort Collins, CO	Knaff and Landsea (1997)	"Leaps and bounds" multiple regression (all possible subsets of zero to four predictors)	Niño 3.4*
CPC constructed analog, CPC/NCEP/NOAA Camp Springs, MD	Van den Dool (1994)	Analog, multiple regression, empirical orthogonal functions	Niño 3.4
CPC CCA, CPC/NCEP/NOAA Camp Springs, MD	Barnston and Ropelewski (1992)	Multivariate multiple regression, empirical orthogonal functions	Niño 3.4*
CDC linear inverse modeling, CDC, University of Colorado, Boulder, CO	Penland and Magorian (1993)	Autocorrelation, time series analysis, empirical orthogonal functions	Gridded field; Niño 3.4
UCLA SSM/MEM, UCLA, Los Angeles, CA	Keppenne and Ghil (1992)	Autocorrelation, time series analysis, empirical orthogonal functions, max entropy method	Niño 3
UBC neural net with "clearning," University of British Columbia, Vancouver, BC, Canada	Tangang et al. (1997)	Neurological network (nonlinear and recursive)	Niño 3.4
CPC consolidated forecast, three- or four-way multiple regression, CPC/NCEP/NOAA, Camp Springs, MD	Cane et al. (1987); Ji et al. (1996); Van den Dool (1994); Barnston and Ropelewski (1992); (regression by D. Unger of CPC, personal communication)	Multiple regression [uses forecasts of 1) Lamont-LDEO2,** 2) NCEP coupled model, 3) CPC constructed analog, and 4) CPC CCA]	Niño 3.4

*Other discrete regions also predicted (this designation not given if gridded field predicted).
**Discontinued during study period, leaving three models.

mospheric, Oceanic and Planetary Physics Department at the University of Oxford. The model consists of a tropical Pacific ocean model with two active layers coupled to a statistical model that relates SST anomalies, heat content anomalies, and surface wind stress anomalies (Balmaseda et al. 1994). The ocean model is first forced by observed wind stress (based on FSU winds) during 1961–91. The output of this run is used to build the statistical atmospheric model, which assumes that the wind stress anomalies are a linear function of the first six principal components of the model SST and heat content anomalies, with seasonal variation. The model predicts SST in the Niño 3 region and in the region bounded by 5°N–5°S, 130°–170°W (called Eq 2). Forecasts are made out to 9 months in advance.

The second version University of Oxford intermediate coupled model is identical to the original version, except that its forecasts are redefined in terms of the initial observations and the tendency of the SST relative to the zero-lead SST simulation:

$$F(t) = \mathrm{Obs}(t_0) + [F_{\mathrm{MDL}}(t) - F_{\mathrm{MDL}}(t_0)],$$

where t_0 is the initial time of the forecast run. Thus, the forecasts of the original model are adjusted for the difference between $F_{\mathrm{MDL}}(t_0)$ and $\mathrm{Obs}(t_0)$, yielding forecasts that are free of initial error.

The Scripps–MPI hybrid coupled model, or HCM, of the tropical ocean–atmosphere system (Barnett et al. 1993) represents a major increment in the complexity of the ocean model relative to the models described above. The ocean model, created at MPI (Latif 1987), is a fully nonlinear GCM bounded by 30°N–30°S and by Asia and South America. It has 13 vertical levels, 10 being within the top 300 m. The seasonal cycle is governed by a Newtonian heat flux and observed wind stress from FSU. The vertical mixing scheme is dependent on the vertical stability and shear (Pacanowski and Philander 1981). The atmospheric model is a statistical canonical correlation analysis (CCA)-like model, deriving the wind stress forcing for the ocean GCM using the GCM's SST. Systematic errors in the SST fields produced by the ocean GCM are accounted for in the coupling. The model is initialized with wind stress fields derived from observed SST data. The 1965–86 period is used for model development and 1986 onward for independent forecasting. The model's skill is highest in the central equatorial Pacific (140°W–180°) and for boreal winter. An improved version of the model, called HCM-3 (Pierce 1996) uses the HOPE2 ocean model from MPI (Wolff and Maier-Reimer 1992). While similar to the original HCM, the numerical scheme has reduced numerical diffusion, especially vertically, resulting in better thermocline representation across the tropical Pacific. Systematic errors, now much smaller, are still corrected. Winds of da Silva et al. (1994) are now used, subject to some presmoothing. The HCM forecasts the tropical Pacific SST field out to over a year in advance.

The Center for Ocean–Land–Atmosphere Studies (COLA) comprehensive coupled model contains a sophisticated dynamical ocean as well as atmospheric model to make long-lead forecasts of the tropical Pacific SST anomaly field (Kirtman et al. 1997). The COLA atmospheric GCM (Kinter et al. 1988) includes a state-of-the-art land surface model (Xue et al. 1991) and physical parameterizations of radiation, convection, and turbulence. It is a global spectral model with a horizontal resolution of T30 and 18 unevenly spaced vertical sigma levels. The oceanic model is a Pacific basin version of the Geophysical Fluid Dynamics Laboratory (GFDL) ocean model (Pacanowski et al. 1993) with 20 vertical levels, 16 being in the upper 400 m. Zonal resolution is 1.5° long by 0.5° lat between 20°N and 20°S; further detail is provided in Huang and Schneider (1995). Skill has been improved by reducing zonal wind stress errors in the atmospheric model by using zonal winds at the top of the boundary layer to redefine the zonal wind stress at the surface (Huang and Shukla 1997). Zonal wind stress simulations have been further improved (Kirtman and Schneider 1996), leading to better initial conditions for coupled forecasts (Kirtman et al. 1997). The COLA coupled model forecasts the tropical Pacific SST anomaly field out to over a year in advance.

The NCEP comprehensive coupled model was developed for long-lead climate forecasts (Ji et al. 1994a,b; Ji et al. 1996). The NCEP Medium Range Forecast (MRF) atmospheric model is used with a dynamic Pacific basin ocean model developed at GFDL by Bryan (1969) and Cox (1984) and subsequently improved by Philander et al. (1987). It covers the domain 45°S–55°N, 120°E–70°W. Zonal resolution is 1.5°; meridional resolution is 0.33° within 10° of the equator, decreasing gradually to 1° between 10° and 20° away from the equator. There are 28 vertical levels, with higher vertical resolution in the upper ocean. The MRF atmospheric GCM has low resolution (18 vertical levels and about 3° horizontal resolution), and the convective parameterizations are tuned for more realistic tropical air–sea interaction and convection. Exchange of surface momentum and heat fluxes and SST at the air–sea interface occur at 5-day intervals. The ocean thermal field is initialized with an ocean data assimilation system (Ji et al. 1995). This is one of the only models (along with the BMRC model) to use subsurface sea data in its initialization. The coupled system has undergone incremental improvements over the last six years: refinement of the flux climatology, insertion of a model output statistics correction for wind stress, and anomaly coupling for the net heat flux forcing (Ji et al. 1996). In 1996 the data assimilation system was improved, as was the ocean model mixing and the anomalous evaporation–precipitation (e–p) flux forcing in the coupling. The model forecasts the tropical Pacific SST anomaly field out to 9 months advance.

b. Statistical models

The Colorado State University (CSU)/Atlantic Oceanographic and Meteorological Laboratory (AOML) CLIPER regression model is a simple statistical model using persistence and the trends in recent observed SST conditions, in a multiple regression framework (Knaff and Landsea 1997), intended to provide a baseline of skill for more complex prediction models that is harder to beat than simple persistence. The CLIPER model selects an optimal combination of a subset of zero to four out of 14 candidate predictors including several timescales of persistence, a month-to-month trend of initial conditions, and climatology. The 1950–94 period is used to build the models, which are trained separately to predict the SST in each Niño region for each calendar month at different lead times. Skill for all possible subsets of the given number of predictors is tested; if no model is suitable, a climatology forecast is issued. Predictors include 1-, 3-, or 5-month averages of initial predictor anomalies as well as their recent trends; predictors are the predictand SSTs themselves at earlier times. Limits on predictor selection were imposed to reduce overfitting. Skills are degraded from training sample results to estimate independent forecast skill following Davis (1979) and Shapiro (1984). Final skill estimates are found to be at levels comparable to those of more sophisticated statistical and dynamical models. CLIPER forecasts are given for the tropical Pacific SST anomaly in the Niño regions out to over a year in advance.

The CPC of NCEP constructed analog model combines the concepts of analogs and multiple regression to form SST forecasts for the Niño 3.4 region (Van den Dool 1994). Because close natural analogs are rarely obtainable, construction of a synthetic analog from selected pieces of natural ones is helpful. The construction is a linear combination of observed anomaly patterns in the predictor fields such that the combination is as close as desired to the base. The analog selection criterion, or predictor, is the set of five leading EOFs of the global SST field at four consecutive 3-month periods prior to the forecast time. Data from 1955 to the present are used to develop the model. For a given base time, a linear combination is made of the first five EOFs of global SST from all 40+ yr (excluding the base year), so as to match the SST pattern of the base time. This is done as a multiple regression, using each year's SST state as a predictor to which a weight is assigned. These weights are then applied to the subsequently occurring Niño 3.4 SST in the predictand period of these years, forming a forecast for the base year that can be extended from the initial time to as far into the future as desired. While analog year weighting is determined with a least square error criterion, the forecasts are not damped to satisfy such a criterion. If nature were linear, and observations error free, the constructed analog process would give an exact integration in time. Forecasts are made for the Niño 3.4 SST anomaly out to over a year in advance.

The CPC of NCEP CCA model has been used to forecast the area-averaged SST anomaly in the Niño 3.4 region since 1990 (Barnston and Ropelewski 1992). CCA linearly models relationships between evolving patterns in the predictor fields and the SST predictand field (Barnett and Preisendorfer 1987; Graham et al. 1987a,b; Barnston 1994). The predictor fields include global sea level pressure, the tropical Pacific thermocline depth, and the tropical Pacific SST itself. These predictor fields are taken from four consecutive 3-month periods immediately prior to the forecast time. The mean SST in a set of eight large regions spanning the tropical Pacific and Indian Oceans is the predictand, to be linearly related to the evolving patterns in the four prior consecutive predictor fields. These pattern relationships are defined over the training period and then applied to the future year being forecast, based on the time–space patterns in the predictors immediately preceding the time of the current forecast. Forecasts are inflated to equalize their interannual variance to that of the observations. The CCA forecasts of Niño 3.4 (and seven other regions) are made out to five seasons in advance.

The Climate Diagnostics Center (CDC)/University of Colorado's linear inverse modeling (Penland and Magorian 1993) is a linear multivariate technique used to predict the field of Indo–Pacific SST anomalies using the past history of that field. A prediction at a given lead time is made by applying a statistically obtained Green function for that lead time to an observed initial condition consisting of SST anomalies in the region 30°N–30°S, 30°E–70°W. While similar to principal oscillation pattern analysis (Xu and von Storch 1990), a relatively large number of oscillation patterns (modes) is used rather than just one or two, to describe the anomaly field in considerable detail. Here, 20 EOFs of the SST are included. Although the linear model parameters are obtained statistically, the technique also hinges on the existence of specific dynamically preferred scenarios in the ocean–atmosphere system (Penland 1989; Penland and Sardeshmukh 1995). The field of SST anomaly over the Indo–Pacific basin is made out to a lead time of four seasons.

The University of California, Los Angeles (UCLA) singular spectrum analysis/maximum entropy method (SSA/MEM), developed in the Department of Atmospheric Sciences and Institute of Geophysics and Planetary Physics, uses SSA and MEM to predict the SST anomaly in the Niño 3 region. SSA is an EOF analysis applied to a matrix made of a single vector time series repeated at incrementally increasing lag times (Vautard and Ghil 1989; Keppenne and Ghil 1992). The resulting EOFs and their time components represent oscillations of preferred frequency. Significant EOFs are then used to reconstruct a filtered time series containing frequencies of a priori interest, namely, 2-yr [quasi-biennial; Rasmusson et al. (1990)] and 4-yr components of ENSO. Extrapolating the filtered time series into the future (i.e., forecasting) is then done using the autoregressive MEM (Penland et al. 1991). The SSA/MEM model is based on nearly 50 yr of historical data. Multichannel SSA (MSSA) has also been used (Keppenne and Ghil 1993; Jiang et al. 1995), using the time series of selected EOFs of the whole tropical Pacific basin as input to an SSA/MEM procedure. MSSA provides SST information other than that of the predictand, potentially benefitting the forecasts. The (M)SSA/MEM forecasts for Niño 3 SST are made out to a year in advance.

The University of British Columbia (UBC) neurological network model attempts to capture the nonlinear aspects of the ENSO system. While most dynamical models contain nonlinearity, the neurological network is the only nonlinear statistical model in this study. Neural networks usually model a set of empirical data (e.g., predictor–predictand pairs), adjusting a set of network weights to minimize forecast errors. Often the nonlinear functions are not examined directly by the modeler; a set of neurons between the input and a "hidden layer," and a set between the hidden layer and the output, define the relationships. In the version used at UBC (Tangang et al. 1997; Hsieh and Tang 1998), the predictors are the first four EOFs of the FSU wind stress data and the Niño 3 SST anomaly for two consecutive prior 3-month periods. The predictands are the same five variables at the later time. There are several sigmoidal neurons in the hidden layer, generating 69 weights (see references). Clearly, a multitude of nonlinear functions can result. The UBC's neural net systems have evolved through schemes such as "clearning" (where the data themselves are changed), "optimal brain damage" (where unproductive neurons are deleted), and "bagging" (using ensembles of neural net models). Since late 1997 the sea level pressure field has replaced the FSU wind stress across the tropical Pacific (Tangang et al. 1998a; Tangang et al. 1998b). Neural net forecasts for Niño 3.4 SST are made out to a year in advance.

The CPC of NCEP consolidated forecast is a four-way linear multiple regression, using as predictors the forecasts of two dynamical models (the Lamont LDEO2 and the NCEP coupled model) and two statistical models (CCA and constructed analog). The purpose is to combine several forecasts having differing strengths into a consensus forecast with higher overall skill than that of its constituents. Beginning with the September 1997 start time, when it became clear that the Lamont LDEO2 model was missing the strong El Niño episode, that model was eliminated, leaving a three-way multiple regression. Because the NCEP coupled model forecast data begins in 1983 (rather than much earlier for the other models), the regression coefficients are developed using only the recent period and pooling running 3-month periods to bolster the small sample size. Because the NCEP coupled model forecasts go out to only 9 months in advance, the consolidated model forecasts for longer lead times are made from the other models, allowing forecasts for Niño 3.4 SST out to 13 months in advance.

4. Model performance results

Table 4 shows the real-time forecast skill of the 15 models in predicting tropical Pacific SST anomalies over the 1997–98 period, shown as rmse and temporal correlation scores. A large variation in model skill is evident both for the dynamical models (listed first) and the statistical models. While part of these differences may be due to variations in model quality, much of it is caused by sampling considerations (related to differing atmospheric initial conditions in the dynamical models), given the shortness of the study period. This implies that the ranks of model performance would be expected to shift considerably if a different 2-yr period had been sampled. Because the most difficult season for which to make a forecast is boreal summer when the forecast is made before spring, rmse skill scores for forecasts for June and September of 1997 (when the strong El Niño initially developed) and for June and September of 1998 (after the El Niño quickly dissipated and a La Niña emerged) are shown individually in the last two columns of Table 4. The rmse scores for the first of those periods are considerably higher than those for the entire period, due to our underestimation of the anomalies. The skills of climatology and persistence are shown at the bottom of the table, and it is noted that most of the models outperformed these simple controls.

Figure 1 shows graphs of the forecasts of each of the 15 models compared with the observations. Each graph is shown for the region forecast by the particular model, and in the units chosen for the models—either actual SST anomaly (°C) or standardized anomaly. When the model forecasts are not given for a discrete region but rather are issued for the entire field of tropical Pacific SST anomaly, an area-averaged

TABLE 4. Synopsis of performance of forecast models in predicting tropical Pacific SST at eight 3-monthly forecast times from June 1996 to March 1998. Forecasts made at lead times of 3.5, 6.5, and 9.5 months are used for all except the March 1998 forecasts, for which 9.5-month lead forecasts are excluded, resulting in a total of 23 forecasts per model. Rmse is computed on standardized anomalies. Correlation is computed with respect to long-term statistics (i.e., means for this short study period are not removed in the computation).

Model	Rmse	Correlation	Rmse for Jun–Sep 1997 from Dec, Mar	Rmse for Jun–Sep 1998 from Dec, Mar
Lamont LDEO-1 coupled	2.61	–0.38	3.63	0.76
Lamont LDEO-2 coupled*	2.82*	–0.48*	3.94*	1.11*
BMRC coupled	1.81	0.70	2.35	1.22
Oxford-1 coupled	1.81	0.26	2.80	1.14
Oxford-2 coupled	1.88	0.50	1.94	3.28
Scripps–MPI hybrid coupled	1.39	0.70	2.26	1.58
COLA coupled	1.18	0.81	2.24	0.39
NCEP coupled	1.15	0.83	1.26	2.37
Mean for the eight dynamical models	**1.83 (1.69)****	**0.37 (0.49)****	**2.43 (2.22)****	**1.29 (1.27)****
CLIPER regression	1.09	0.84	1.81	1.33
Constructed analog	1.18	0.83	1.96	1.04
Canonical correlation analysis	1.05	0.84	1.70	1.43
SSA/MEM	2.11	0.38	2.83	1.90
Linear inverse model	1.42	0.66	2.92	0.48
Neural network	1.64	0.55	3.23	0.79
Three- or four-way consolidation	1.65***	0.51***	2.42***	2.16***
Mean for the seven statistical models	**1.45**	**0.66**	**2.31**	**0.99**
Mean for all 15 models	**1.65 (1.57)****	**0.50 (0.57)****	**2.38 (2.27)****	**1.15 (1.13)****
Control: climatology	1.79	undefined	2.55	0.78
Control: persistence—Niño 3	2.33	0.41	3.31	3.35
Control: persistence—Niño 3.4	2.04	0.40	2.76	3.13

*Scores for the new LDEO3 model run in hindcast mode, from left to right, would have been 1.57, 0.81, 2.71, and 1.06.
**Score computed excluding the LDEO2 model.
***Forecasts from Sep 1997 onward (8 of 23 forecasts) excluded the LDEO2 model. If the LDEO2 model had been excluded throughout, the scores, from left to right, would have been 1.19, 0.80, 1.38, and 1.71.

value for the Niño 3.4 region (or whatever region the model was designed to forecast best) was derived from the anomaly field. Because of the model-to-model variation in units and forecasted region, the observations (shown by the solid line in Fig. 1) vary somewhat among graphs. The dashed line indicates the model forecasts for each of eight start times, all but the last of which produces three seasonal forecasts (centered at 3.5, 6.5, and 9.5 months following the start time). Thus, there are two dashed straight-line segments associated with most start times, and the start time is always 3.5 months earlier than the central month of the first time being forecast. A dotted line connects the observation at the approximate start time to the first forecasted time. Recall that there is model-to-model variation in the actual start time; some models were initialized two to three weeks earlier and did not "know" the start-time observation from which the dotted line begins. Additionally, some models' initialization may simulate an SST condition at the start time that differs from the latest observations. It is understood that models that can have significant errors in their simulations at the initial time generally tend to have lower skills at both short and longer leads (e.g., Chen et al. 1995). Some evidence of this is found in the performance of the two Oxford models discussed here—one with and one without an a posteriori linear adjustment for initialization errors. The skill difference, while not huge, is quite noticeable (Table 4, Fig. 1). Among the dynamical models discussed here, only the initializations of the NCEP and BMRC coupled models include an assimilation of subsurface ocean observations at their start times. As shown in Table 4 and Fig. 1, the NCEP model was one of the best performers. The BMRC model performed fairly well, but not better than some of the models without assimilation (e.g., COLA coupled, Scripps–MPI hybrid coupled). From these few cases it appears that initialization accuracy is somewhat important to overall model skill. It is unfortunate that the skill of dynamical models that accurately represent the physics of the climate system may be degraded due to the initialization accuracy problem.

Inspection of the graphs shows that in early 1997 most of the models predicted some degree of warming of the tropical Pacific SST for the forthcoming season, and thereby demonstrated skill above that of climatology or persistence forecasts. However, a prevalent and serious error in this set of forecasts was the low magnitude in the predicted warming: No model came close to predicting the size of the 1997–98 El Niño as of March 1997, the time at which the SST warmed to its climatological value from its previously cooler state and was about to warm very rapidly. Figure 1 shows that the two models that came closest to the correct magnitude were the dynamical NCEP coupled model and the statistical CSU CLIPER regression model, with the COLA and NCEP CCA and constructed analog models nearly as good. For the December 1996 start time, the NCEP coupled model forecasted mild warming in the Niño 3.4 region between March and June 1997. By March 1997 it forecasted warming to nearly 1.8°C by the end of the year. In June, when the warming had largely materialized, the model predicted substantial further warming and predicted the maximum anomaly accurately. It was slow, however, to weaken and end the event in the first half of 1998. The NCEP coupled model tends to simulate El Niño episodes more readily than La Niñas and may not show as much skill in forecasts of the 1998–99 La Niña period as it did for the 1997–98 El Niño. The CSU/AOML CLIPER regression, developed mainly as a more competitive benchmark than persistence for comparing skills of more complicated models, performed comparably to the NCEP coupled model, handling the onset of the warm event less skillfully than the NCEP model but performing better than NCEP in the dissipation phase. In both phases of the event it lagged the observations by 1 to slightly over 2 months, with the greatest lags occurring near the summer 1997 target period. For the March 1997 start time, CLIPER correctly increased the SST anomaly sequentially over the three target periods; these increases were substantial (bringing the SST anomaly to 1.7°C by year's end) but still not sufficient to forecast what occurred.

The COLA coupled model and the NCEP CCA and NCEP constructed analog statistical models performed somewhat less well than the NCEP coupled and CSU CLIPER models during the El Niño portion of the period, but about as well as them for the entire period (Table 4); slightly down from these are the dynamical BMRC and Scripps–MPI hybrid coupled models. All five of these models underpredicted the warming during 1997 and were slow to "catch up" as the large magnitude of the event became evident in the start time observations. The Scripps–MPI coupled and NCEP constructed analog models began forecasting the event's dissipation too early, while the NCEP CCA was repeatedly slow to end the event. The COLA model (Fig. 1) ended the event more accurately. That model had forecast a warming for late 1996, but when

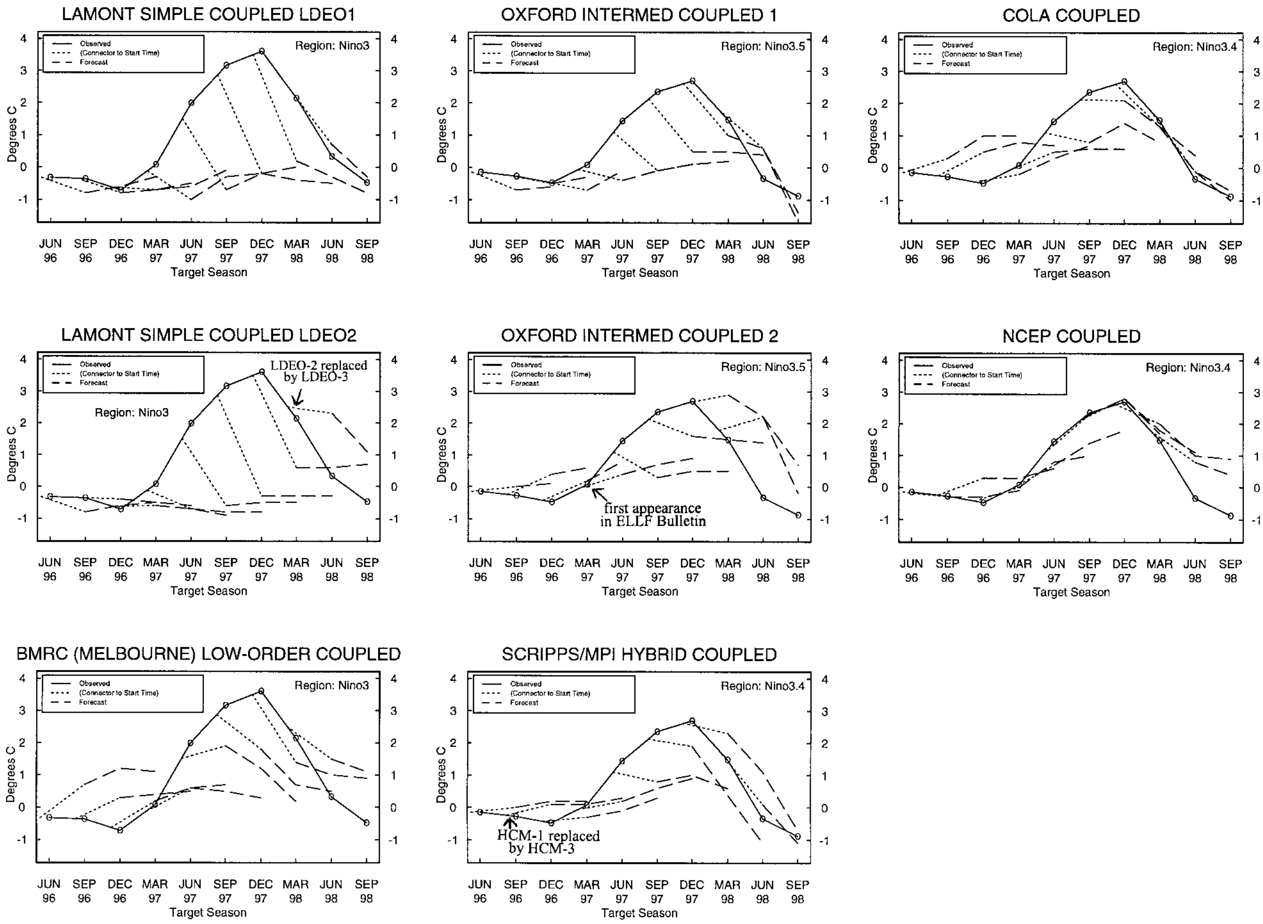

FIG. 1. Plots of forecasts (dashed lines) and observations (solid lines) of SST over the Jun 1996–Sep 1998 period for 15 models projecting the state of ENSO at lead times of 3.5, 6.5, and 9.5 months. The predicted regions and SST units vary by model as shown in the upper-right corner and by the vertical axis label, respectively. Dotted lines connect the observation centered at the approximate start time to the earliest (3.5-month lead) forecast. This is done for readers' convenience only; the beginning of the dotted line does not represent the model's actual initialization nor its forecast at very short lead times. The 3-month periods shown on the *(continued)*

that did not occur it weakened (but did not discontinue) its warming forecast into 1997. The CDC inverse model had good timing and direction but was too weak. The NCEP three- or four-way consolidated statistical model did not attain high skill in real time, but performed well when rerun retrospectively as a three-way consolidation throughout the period using only the NCEP coupled, CCA, and constructed analog models.

As seen in Fig. 1, some of the models failed to forecast, or even acknowledge, significant warming after the El Niño was already well under way. In the dynamical models, this may reflect an inadequate assimilation of the initial conditions. Negative tropical Pacific wind stress anomalies (relaxation of the trades in spring 1997) should imply a presence of El Niño conditions. This problem is apparent in the two LDEO models and in the original version of the Oxford coupled model. In the statistical models, sluggishness in acknowledging current observed conditions is often related to a "watering down" of the high-frequency events of the episode (e.g., a sudden onset) due to long averaging periods. For example, the NCEP constructed analog and CCA use the most recent 3-month mean as the most current predictor data and thus may not have an accurate snapshot of what is occurring in the present when conditions change quickly. Additionally, because statistical models usually train using historical data, they often average many observed past outcomes into one solution (e.g., an "ENSO mode") and may not be able to use details that distinguish the current situation from the generic ENSO event. As discussed in Chen et al. (1999), for example, errors in the National Aeronautics and Space Administration scatterometer wind data in the southeastern tropical Pacific had a marked impact on the LDEO3 forecasts even though the heart of the trades was accurately represented. Another

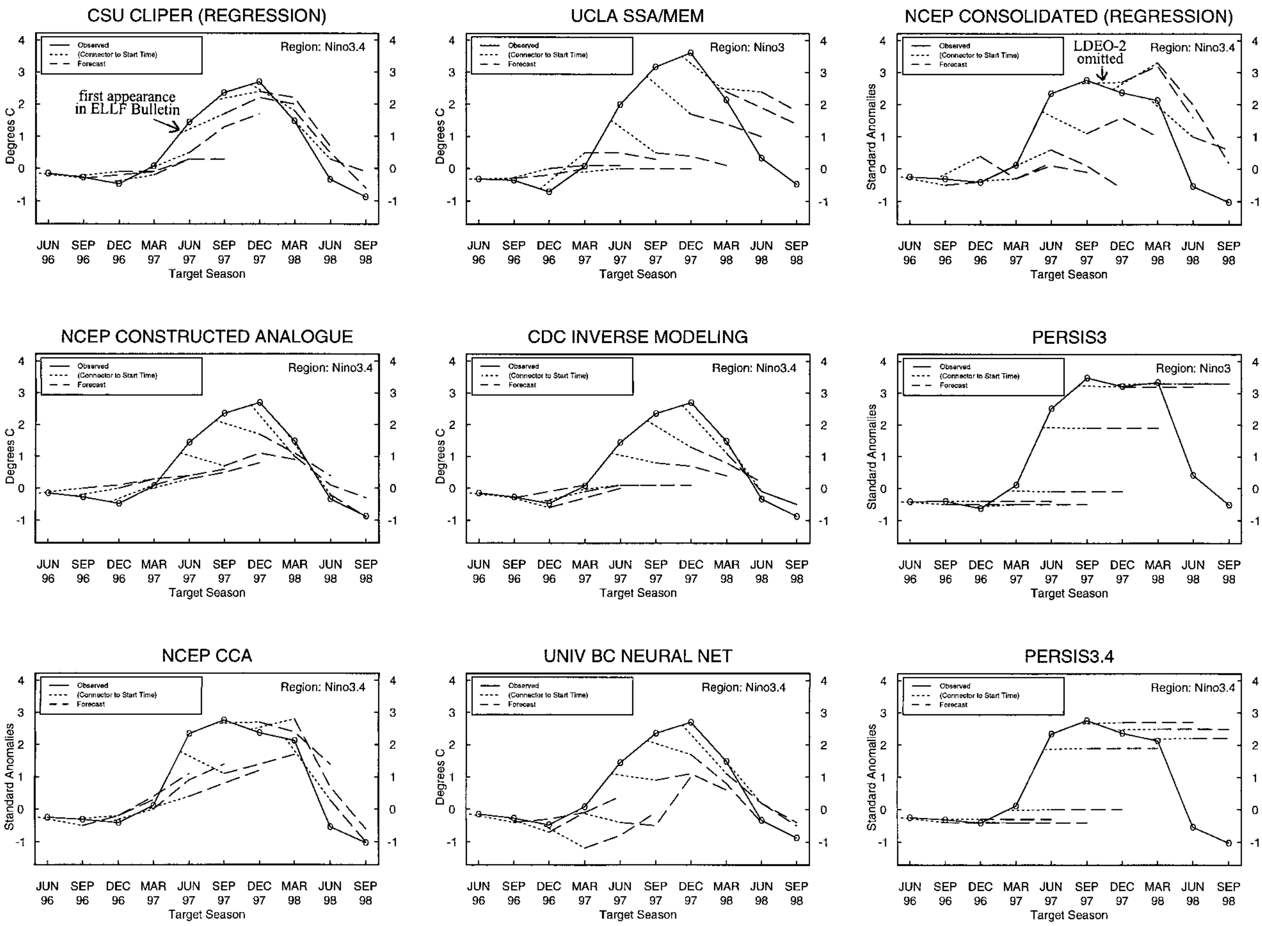

Fig. 1 *(continued)*. abscissa are denoted by their middle month. The Sep 1998 data are used to represent the Aug–Oct 1998 observations, to avoid a 1-month publication delay. The first eight panels show the performances of the dynamical models (indicated above each panel, and in same order as shown in Tables 2, 3, and 4), the next seven the statistical models, and the last two the persistence controls. Persistence forecasts use the mean SST of the period centered about 3 weeks prior to a regular 3-month seasonal observation, making them correspond to the approximate start times of the forecasts.

common limitation in many statistical models is linearity, in which El Niño and La Niña are modeled symmetrically both in terms of the SST and the remote teleconnections. This characteristic is beneficial in reducing sampling variability, but would only serve this purpose if the climate system is in fact quasi-linear.

While judgment on the relative skills of individual models is not meaningful over a short period such as this, there may be some value in comparing the mean skill of the dynamical models to that of the statistical models. The mean skills for the two model types shown in Table 4 suggest that the statistical models slightly outperformed the dynamical models over this period, even when the Lamont LDEO-2 model (whose 1997 failure prompted its replacement by LDEO-3) is omitted. However, the intermodel variability is much greater than the mean differences, and with only one to two independent time realizations the mean differences are statistically insignificant. When only the most skillful dynamical models are compared to the highest scoring statistical models, no meaningful skill difference is evident. While more statistical models ranked in the top four performers than dynamical models, it is important to realize that the three best performing statistical models (CLIPER, CCA, constructed analog) are broadly doing the same task—using some form of linear regression (or analogy, which often but not always leads to a similar result) to relate predictor and predictand data, differing largely in the details of the predictor data choice and the procedure. While the two comprehensive models here (NCEP and COLA) may be said to be broadly doing the same thing as well—using many of the same equations of oceanic and atmospheric motion—their forecasts during 1997–98 had somewhat greater differences than those of the three best statistical models. Part of the reason for this is that the NCEP model is initialized with subsurface ocean information and SST, while the COLA model is initialized without subsurface data.

The conclusion of this skill evaluation for 1997–98 is that today's dynamical and statistical models have a useful, but only moderate, level of skill in forecasting the ENSO state, and that neither type of model shows clear superiority over the other. This is the same conclusion that was reached in the model skill evaluation several years ago (Barnston et al. 1994) using five models, but a more meaningful period of record of ENSO fluctuations (12 yr, although not all 12 contained real-time forecasts). Performance during this recent period was no better than that of the former 12-yr period. Because the computational requirements for statistical models are relatively small and much of their potential may have already been realized, it stands to reason that dynamical models have more upside potential than statistical models. If computer capability were astronomically increased and dynamical modeling ideas could be evaluated quickly and exhaustively, some of them would likely result in skills higher than those found here. As increases in computer power come about, there is reason to hope that our long-term efforts will bring additional scientific knowledge and, hopefully, higher forecast skill. On the other hand, we do not know whether the approximate equality of skill presently found between empirical and dynamical models represents a unique "moment" along the evolutionary path of the dynamical models, or if we are ignoring the possibility that dynamical models will never be able to materially outperform empirical models.

5. Optimizing the format of forecast summaries for users: A user's perspective

Together with the predictions of SSTs, their presentation is equally important to the user community. In this section, the way that the forecasts of the 1997–98 SST conditions were expressed in verbal summaries is examined. The *ELLF Bulletin*, produced quarterly, is posted on the Internet and is also sent to approximately 1500 subscribers each quarter. While some of the readers are scientists and academicians, others are users who lack technical expertise but have critical need for the information. Many of these lay users rely mainly on the brief paragraphs on the first page, which serve as an executive summary of the highlights of the El Niño predictions derived from the 15 models. To appreciate how the different predictions of the 1997–98 El Niño event evolved over time using these abbreviated statements of the ENSO forecasts, the Executive Summaries of December 1996 to September 1997 were examined. These brief qualitative statements summarize the longer, more detailed and quantitative treatments appearing in the body of the *ELLF Bulletin* for each of the modeling groups. (The graphs shown in Fig. 1 are subsets of these quantitative submissions.) We assume the Executive Summary accurately portrays the forecasts detailed in the longer papers: no modeling group had challenged any of the summaries in the study period.[2] The nature of the statements in these summaries is examined next to evaluate their utility to those who rely on them.

Table 5 contains the statements that appeared in the Executive Summary of *ELLF Bulletins* issued from December 1996 (several months prior to the onset of the strong El Niño of 1997–98) through September 1997 (when the El Niño was near its peak). Using these statements, we seek to determine how well the models predicted the appearance and evolution of the 1997 El Niño. The statements essentially summarize the graphs shown in Fig. 1, enabling the reader to track over a year's time the projections of each model.

Examination of the statements in Table 5 reveals variation in how forecasts are expressed. These differences can affect how users come to view or understand El Niño forecasts, and they raise the question of whether condensed, verbally expressed forecasts contain sufficient information to be helpful in their decision making processes. Three areas of concern to potential El Niño forecast users are 1) the way time is expressed, 2) the adjectives (or other modifiers) used to describe the predicted changes in tropical Pacific SST, and 3) the set of phrases used in a specific forecast.

a. *Expressions of time*

Table 6 provides some examples of the references in the forecasts to time for the dynamical models (part a) and the statistical models (part b). The parentheses at the beginning bracket the range of times referred to in the set of all forecasts for the given quarterly issue of the *ELLF Bulletin*, followed by all of the time references made in the forecasts for that quarter. There is considerable variety among the forecasts' references to time. This is partly related to the fact that statistical

[2]Responding to a written inquiry from the editor, some of the modelers felt the summaries were, indeed, representative.

models tend to forecast out to longer lead times than dynamical models, due in part to the known error saturation that can develop in long dynamical model integrations. The precision indicated in the time references is highly variable: sometimes high precision is implied (e.g., "by mid-January"), while in other cases precision is too low for effective use in decision making (e.g., "by summer" or "later in the year"). In some cases terms are not defined. For example, what does "early-", "mid-", or "late-spring" mean? What is meant by "through (a season)"? These terms are subject to differing interpretations by users.

b. Descriptors

The following are some examples taken from the *ELLF Bulletin*'s summaries to describe the changes in the various aspects of ENSO: somewhat decreasing SOI (Southern Oscillation index), positive anomalies emerging, El Niño period peaking, below-normal SSTs dissipating, switching to warm, SOI decreasing from high-normal to normal, slowly increasing to normal, coolish to normal, warming slightly, cool but moderating SSTs, largely dissipating, somewhat warm, etc. While the author of a forecast or the *ELLF Bulletin* editor may know the meaning of such statements, they are subject to differing interpretations by users. Scientists and researchers themselves may differ on the meaning of "neutral" with regard to the ENSO state. For instance, some believe that there are basically only two states: El Niño and La Niña. Others argue that only El Niño and normal are meaningful (La Niña being within the normal classification). Many others recognize all three categories, but differ in where the limits of the categories should be placed. Hence, the meanings of normal and abnormal differ among the forecast producers. Furthermore, meanings depend on the use of the SST information by the users. The modifying terms used in verbal summaries of the ENSO forecasts would be more informative and useful if a set of definitions were provided, thereby tightening the range of possible user interpretations. A proposal for a set of terms and their associated numerical cutoffs for SST is offered below.

c. Phrasing

The aggregate of the time references and the descriptors discussed above make up the phrasing of a verbal forecast. How would such a forecast be verified? Using verbal phrases alone makes an evaluation difficult. For example, consider the December 1996 experimental forecast of COLA: "forecasts cool Niño 3 SST in boreal winter 1996–97, warming to somewhat above normal by October 1997, lasting through winter 1997–98." This forecast can be divided into three parts: 1) Forecasts cool Niño 3 SST in the boreal winter of 1996–97 (in fact, cool Niño 3 SST was observed at this time); 2) warming to somewhat above normal by October 1997 (it warmed rapidly from March through September 1997 and was confirmed as a strong El Niño event by June; while the tendency to warming was stated, the rate, magnitude, and timing were underestimated); 3) lasting through the winter of 1997–98 (this statement of duration was correct). Verification of the forecast as a whole can be done only roughly using the verbal summary, due to the lack of precision in the phrases. Does "somewhat above normal" SST include the possibility of a full-fledged El Niño? The phrases of this COLA forecast can be diagrammed, as illustrated in Table 7, and the accuracy of each phrase assessed. Suppose that one point is credited for each phrase that turned out to be correct, zero is given for items that were not clearly correct or incorrect, and a point is subtracted for each incorrect phrase. If the phrase pertaining to a cool 1996–97 winter is credited a full point and the phrase referring to somewhat above normal SST from October 1997 to winter 1997–98 is not credited because of the magnitude underestimation, then the forecast would appear to be fairly good (a score of 1, with a possible range of –2 to 2). However, the two diagrammed phrases may not be weighted equally in terms of the length of time encompassed, nor in the potential impacts of the anomaly prediction.

A more precise version of a forecast summary would need to be quantitatively based. Graphs such as those of Fig. 1 require too much space to fit in a one-page summary and also would not be reader friendly for nontechnical users. One solution is to develop a simple numerical coding system that would convey the forecasts and their timing more precisely than verbal phrases, and that would also have defined verbal equivalents. An example of such a format is shown in Tables 8 and 9, using only selected subsets of 6 of the 8 models out of 15 that forecast the actual SST (i.e., not standardized) in the Niño 3.4 region. In both tables the forecasts and the observations are coded in one digit (with minus sign if negative) according to the legend at the bottom of Table 8. An asymmetry in magnitude is noted between El Niño and La Niña in the legend, since the former can assume greater strength. In Table 8, the information is blocked by forecast start time (denoted by "s"), and the forecasts for the different models are shown within

TABLE 5. ENSO-related forecast statements pertaining to each of 13 models, as they appeared in the Executive Summary ("summary of forecasts") of the *ELLF Bulletins* issued between Dec 1996 and Sep 1997.

	NCEP coupled	BMRC coupled	Scripps–MPI hybrid coupled	Oxford coupled, 1 and 2
Dec 96	Calls for warming through the neutral range in winter 1996–97 becoming somewhat warm by Jul 1997	Predicts slightly cool conditions for DJF 1996–97, becoming slightly warm by JAS 1997	Predicts mildly cool conditions for winter 1996–97, moderate warming for winter 1997–98	1) Calls for moderately negative SST anomalies boreal winter 1996–97, warming to normal by JAS 1997; 2) no forecast in this issue
Mar 97	Considerable warming through fall 1997	Predicts slightly warm conditions by JAS 1997, slowly weakening to normal by summer 1998	Predicts warming through winter 1997–98 into spring 1998	1) Calls for warming to normal by id- to late 1997; 2) forecasts greater warming, to moderate positive anomalies by winter 1997–98
Jun 97	Calls for strong El Niño conditions through winter 1997–98	Predicts moderate to strong El Niño conditions this summer and fall, largely dissipating by late winter 1997–98	Predicts warming through winter 1997–98 peaking slightly before winter	1) Calls for neutral to slight warming by late 1997; 2) forecasts moderate El Niño conditions for fall and winter 1997–98
Sep 97	Calls for strong El Niño conditions through winter 1997–98, moderating by summer 1998	Predicts moderate to strong El Niño conditions, peaking near Nov 1997, returning toward normal by June 1998, and below normal by Feb 1999	Predicts strong warmth through winter 1997–98, dropping to normal by May 1998, cold conditions by Oct 1998	1) Calls for slight warming by late 1997; 2) forecasts strong El Niño conditions for the rest of 1997 and early 1998
	COLA coupled	**CDC inverse modeling**	**UCLA SSA/MEM**	**UBC neural network**
Dec 96	Forecasts cool Niño 3 SST boreal winter 1996–97, warming to somewhat above normal by Oct 1997, lasting through winter 1997–98	Predicts mildly below normal east-central Pacific SST for winter/spring 1996–97, followed by near normal conditions for summer/fall 1997	Predicts near to slightly above normal Niño 3 SST to fall 1997, and SOI moving toward normal from above	Predicts a slightly cool Niño 3 for winter 1996–97, recovering in spring 1997
Mar 97	Forecasts warming to moderately above normal by summer and especially fall and winter	Predicts cool east-central Pacific SST weakening to neutral by late spring 1997, becoming warmish by summer/fall/winter 1997–98	Predicts Niño 3 SST slowly becoming slightly warm by winter 1997–98, SOI moving toward normal from above	Predicts some weakening of the cool Niño 3 SST toward neutral during 1997

TABLE 5. *Continued.*

	COLA coupled	CDC inverse modeling	UCLA SSA/MEM	UBC neural network
Jun 97	Forecasts a moderate to strong El Niño episode, peaking near winter 1997–98	Predicts mature El Niño conditions through winter 1997–98	Predicts somewhat warm Niño 3 SST peaking in fall 1997 and declining during winter 1997–98; SOI moving to below normal	Predicts continuing moderate El Niño conditions through winter 1997–98
Sep 97	Forecasts El Niño conditions until Jan 1998, rapidly cooling through neutral by Jun 1998 and to strongly cold by winter 1998–99	Predicts moderately strong El Niño conditions in boreal fall 1997, weakening to only slight positive anomalies by summer 1998	Predicts moderately warm Niño 3 SST peaking in fall/winter 1997–98, declining during 1998; SOI remaining below normal through mid-1998	Forecast did not appear in this issue, but was retrieved in retroactive real-time mode for the evaluation in section 4
	NCEP/CPC CCA	**NCEP/CPC contruct. analog**	**NCEP/CPC consolidated**	**CSU CLIPER regression**
Dec 96	Predicts slightly below normal Niño 3.4 (120°–170°W) SST JFM 1997 becoming warm from boreal summer 1997 through following winter	Predicts Niño 3.4 SST rising through normal by late winter 1996–97, becoming moderately warm by winter 1997–98	Calls for coolish conditions winter 1996–97, rising to normal by spring, remaining normal until late 1997 when warming is predicted	Forecasts did not yet appear in these two issues, but were retrieved in retroactive real-time mode for the evaluation in section 4
Mar 97	Predicts Niño 3.4 (120°–170°W) SST warming through neutral by spring 1997 to warm levels from late summer 1997 to early 1998	Predicts Niño 3.4 SST becoming slightly above normal by summer, reaching moderate warmth by winter 1997–98 then weakening	Calls for warmish conditions by winter 1997–98, cooling to somewhat below normal by winter 1997–98, returning to warm in spring 1998	
Jun 97	Predicts moderate El Niño conditions for Niño 3.4 (120°–170°W) for the next 12 months, peaking in spring 1998	Predicts El Niño conditions for Niño 3.4 through spring 1998, peaking in fall/winter 1997–98	Calls for moderate El Niño conditions through spring 1998, peaking in late fall 1997 and again in late spring 1998	Predicts fairly strong El Niño conditions, peaking in winter 1997–98
Sep 97	Predicts strong El Niño conditions for Niño 3.4 (120°–170°W) through mid-spring 1998, peaking near Jan 1998 in late summer 1998	Predicts quite strong El Niño conditions for Niño 3.4 through spring 1998, peaking in winter 1997–98, then rapidly declining	Calls for strong El Niño conditions through late spring 1998, peaking in early spring, then rapidly declining for the rest of 1997 and early 1998	Predicts strong El Niño conditions peaking in winter 1997–98 and returning to normal or slightly below by fall 1998

TABLE 6. Quarterly forecast issued in these months referred to the following time periods.

Dynamical models	
Dec 1996:	(Summer 1997–winter 1997–98): summer 1997, winter 1997–98, Jul 1997, DJF 1996–97, JAS 1997, boreal winter 1996–97, JAS 1997, Oct 1997
Mar 1997:	(Spring 1997–98): summer 1997, fall 1997, winter 1997–98, summer 1998, mid- to late 1997, spring 1998
Jun 1997:	(Aug 1997–spring 1998): fall 1997, winter 1997–98, 1998, spring 1998, late 1997
Sep 1997:	(Fall 1997–98): Jul 1998, winter 1997–98, summer 1998, Nov 1998, Jun 1998, Feb 1999, late 1997, early 1998, winter 1998–99, May 1998, Jun 1998, Oct 1998, fall 1997, early 1998
Statistical models	
Dec 1996:	(Winter 1996–winter 1997/98): winter–spring 1996–97, summer 1997 and fall 1997, Feb 1997, fall 1997, winter 1996/97, spring 1997, JFM 1997, boreal summer 1997, through winter 1997/98, late winter 1997, winter 1997/98, late 1997
Mar 1997:	(Spring 1997–spring 1998): late spring 1997, by summer, fall, and winter 1997–98, May 1997, winter 1997/98, during 1997, spring 1997, late summer 1997, to early 1998, spring 1998
Jun 1997:	(Aug 1997–spring 1998): through winter 1997/98, to Aug 1997, fall 1997, winter 1997/98, late spring 1998, fall–winter 1997–98, late fall 1997, late spring 1998
Sep 1997:	(Fall 1997–during 1998): fall 1997, summer 1998, fall–winter 1997–98, during 1998, through mid-1998, through midspring 1998, near Jan 1998, late summer 1998, through spring 1998, winter 1997/98, late spring 1998, early spring 1998, fall 1998

each block. In Table 9, the information is blocked by individual model, and the different forecast start times are shown within each model block. Further space could be conserved by eliminating the negative signs and using a different font to signify negative anomalies. In these tables the models' forecasts are expressed with moderate precision, both in magnitude and timing, in a compact format that users could understand with little difficulty. Because the range of possible anomalies varies with season and tropical Pacific region, the categorical definitions would need adjustment. To accomplish such an adjustment, the legend relating the one-digit codes to their verbal ENSO magnitude labels should be applied to standardized SST anomalies rather than physical (°C) anomalies. Use of standardized values would establish equivalency across seasons and regions but might present problems for users who prefer physical units. While such details would need to be resolved, it is clear that the tables provide far more detailed information than would be possible using verbal phrases that occupy the same space. With this user-friendly forecast information, users may more easily learn the implications of a given one-digit ENSO

TABLE 7. Diagram of the summary of the COLA coupled model forecast made in December 1996, appearing in the *ELLF Bulletin's* Summary of Forecasts. The forecast summary was "cool Niño 3 SST boreal winter 96–97, warming to somewhat above normal by Oct 97, lasting through winter 97–98." The entries in parentheses are implied, while those not in parentheses are stated outright.

Descriptive terms	Anomaly prediction for Niño 3 SST	Forecast of when anomaly would begin	Forecast of when anomaly would end
	Cool	Winter 1996/97	(Before October 1997)
Warming to	Somewhat above normal	October 1997	(During, or later than,winter 1997/98)

intensity code with respect to impacts affecting them in their location.

6. Summary and discussion

Critical reviews of forecasts of ENSO conditions, based on dynamical and statistical model output, are necessary for an improved understanding of the current level of success of model forecasts. In this instance we have used the 1997–98 El Niño event, one of the strongest in a century (rivaling the event of 1982–83), as a focal point for the assessment of forecast skill and forecast usefulness. Skill during the onset phase of the La Niña of 1998–99 is also examined. It becomes clear on review of the various forecasts that were issued quarterly in the *ELLF Bulletin* that most of the forecasts did not identify a tendency toward an El Niño onset until the first quarter (i.e., March) of 1997, when rapid warming was just beginning to occur. Similarly, they did not acknowledge the appearance of an exceptionally strong event until the June 1997 forecast, when the event was becoming very strong. On the positive side, many of the models had been forecasting some degree of warming in December 1996, several months before any warming began. While none of them forecast a very strong warming, 2 or 3 out of 15 predicted warmth on the order of 1°C and subsequently predicted further strong warming once the warming became evident in March 1997. Neither the dynamical nor the statistical models, as groups, tended to perform significantly better than the other: certain models from either group were identified as the best performers. This review is not exhaustive, as some of the newest, potentially skillful dynamical models were in experimental stages and did not yet appear in the *ELLF Bulletin.* Examples are the LDEO3 (Chen et al. 1998), the hybrid coupled model used at UCLA/Jet Propulsion Laboratory (Syu et al. 1995), the model used at GFDL (Gordon and Stern 1982), and the European Centre for Medium-Range Weather Forecasts coupled model (Stockdale et al. 1998).

The large spread among the SST predictions, even for a strong event like the 1997–98 El Niño, has led to a high intermodel disparity of performance scores. This variation is expected, given the brevity of the study period, including only one to two independent realizations of the climate state over the 1–2-yr period. The small sample factor manifests itself through the intermodel differences in initial conditions (for the dynamical models), reflecting the inherent limit of predictability related to the chaotic nature of the observed system over which the consistent effects of the large-scale physics must rise. A second source of intermodel differences is that of model imperfections. For the dynamical models this encompasses flawed representation of oceanic and atmospheric physics, such as faulty parameterizations, possibly related to overtuning on the developmental data. For the statistical models, imperfections include an inappropriate choice of predictors, an unrealistic model design, or an overfitted system (e.g., using too many predictors) that scores far more poorly on independent forecasts than on the developmental data.

It cannot be overstated that a 2-yr period is insufficient to make meaningful comparisons among mod-

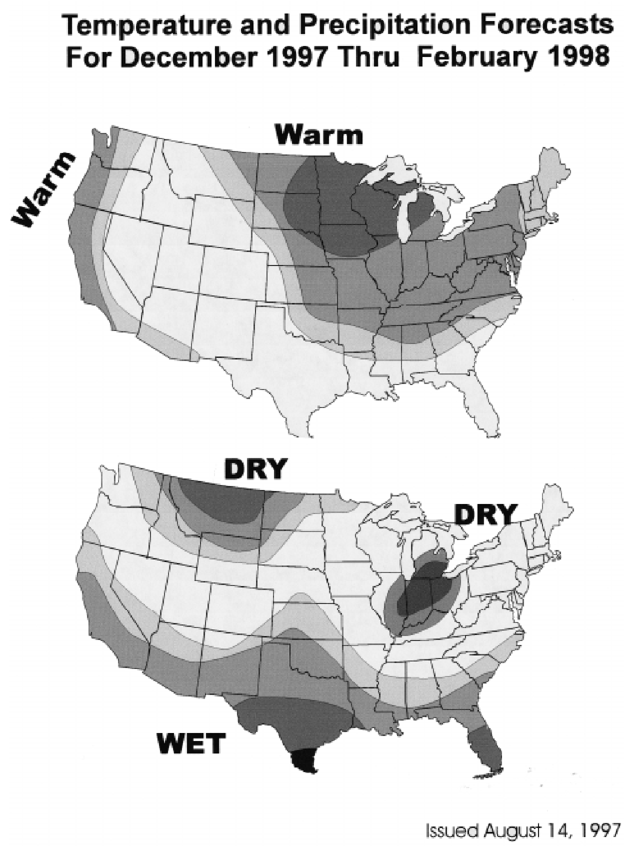

FIG. 2. Temperature (top) and precipitation (bottom) probability anomaly forecasts for Dec–Feb 1997–98, issued by the NCEP/CPC in mid-August 1997. The lightest shading indicates probability anomalies from 0 to 0.05, the next darkest shading from 0.05 to 0.10, next from 0.10 to 0.15, etc. Anomaly predictions for temperature included only positive anomalies, while anomalous dryness was forecast for the northern Rockies and Ohio Valley and enhanced precipitation was forecast for the southern tier. Climatological probabilities are forecast in the white areas, implying no forecastable tilts of the odds of any of the three tercile categories.

TABLE 8. Summarization of model forecasts by one-digit forecasting coding, against observations, blocked by start time for several forecast models. Numerical symbol definitions are shown at the bottom. A selected subset of six of the eight models forecasting the Niño 3.4 region in physical units (°C) is shown. Start time indicated by "s."

1996							1997												1998									Model
summer				fall								summer				fall								summer			fall	
Jun			Sep			Dec			Mar			Jun			Sep			Dec			Mar			Jun			Sep	
−1	−1	−1	−1	−1	−2	−2	−1	−0	0	1	2	4	5	6	7	7	8	8	7	6	4	2	0	−1	−2	−3	−3	**Observed**
s			0	0	0	0	0	1	1																			NCEP construc analog
s			−1	−1	−1	−1	−0	−0	0																			Linear inverse mdl
s			−1	−1	−1	−1	−1	−0	−0																			CSU CLIPER (regresn)
s			1	2	3	3	3	3	3																			COLA coupled mdl
s			0	0	0	1	1	1	1																			Scripps–MPI HCM
s			−1	−1	−1	−1	−1	−0	−0																			NCEP coupled mdl
−1	−1	−1	−1	−1	−2	−2	−1	−0	0	1	2	4	5	6	7	7	8	8	7	6	4	2	0	−1	−2	−3	−3	**Observed**
						s			0	0	1	1	1	2	2													NCEP construc analog
						s			0	0	0	0	0	0	1													Linear inverse mdl
						s			−1	−0	0	1	1	1	1													CSU CLIPER (regresn)
						s			−1	−0	1	1	2	2	2													COLA coupled mdl
						s			−1	−1	−0	0	1	1	1													Scripps–MPI HCM
						s			0	1	2	2	2	3	3													NCEP coupled mdl
−1	−1	−1	−1	−1	−2	−2	−1	−0	0	1	2	4	5	6	7	7	8	8	7	6	4	2	0	−1	−2	−3	−3	**Observed**
												s			2	2	3	3	3	3	3							NCEP construc analog
												s			3	3	2	2	2	2	1							Linear inverse mdl
												s			5	6	6	7	7	6	6							CSU CLIPER (regresn)
												s			2	3	4	4	4	3	2							COLA coupled mdl
												s			2	3	3	3	3	2	2							Scripps–MPI HCM
												s			7	7	8	8	7	6	5							NCEP coupled mdl
−1	−1	−1	−1	−1	−2	−2	−1	−0	0	1	2	4	5	6	7	7	8	8	7	6	4	2	0	−1	−2	−3	−3	**Observed**
																		s			3	2	1	0	−0	−1	−1	NCEP construc analog
																		s			3	2	1	−1	−1	−2	−2	Linear inverse mdl
																		s			5	3	2	1	0	−1	−2	CSU CLIPER (regresn)
																		s			4	3	2	−0	−1	−2	−3	COLA coupled mdl
																		s			7	5	4	3	1	−1	−2	Scripps–MPI HCM
																		s			6	5	4	3	3	3	3	NCEP coupled mdl
−1	−1	−1	−1	−1	−2	−2	−1	−0	0	1	2	4	5	6	7	7	8	8	7	6	4	2	0	−1	−2	−3	−3	**Observed**

Table 8. *Continued.*

Legend: One-digit magnitude code definitions for Niño 3.4 SST (°C), with verbal equivalents. (Ideally, standardized SST anomalies would be used.)

9		3.00	and	upward			extraordinary El Niño	−0	between	0.00	and	−0.14	center is	−0.068	neutral, below zero
8	between	2.50	and	3.00	center is	2.75	outstanding El Niño	−1	between	−0.14	and	−0.40	center is	−0.27	slightly cold
7	between	2.17	and	2.50	center is	2.33	very strong El Niño	−2	between	−0.40	and	−0.67	center is	−0.53	somewhat cold
6	between	1.83	and	2.17	center is	2.00	strong El Niño	−3	between	−0.67	and	−0.93	center is	−0.80	moderately cold
5	between	1.50	and	1.83	center is	1.67	moderately strong El Niño	−4	between	−0.93	and	−1.20	center is	−1.07	standard La Niña
4	between	1.17	and	1.50	center is	1.33	standard El Niño	−5	between	−1.20	and	−1.47	center is	−1.33	moderately strong La Niña
3	between	0.83	and	1.17	center is	1.00	moderately warm	−6	between	−1.47	and	−1.73	center is	−1.60	strong La Niña
2	between	0.50	and	0.83	center is	0.67	somewhat warm	−7	between	−1.73	and	−2.00	center is	−1.87	very strong La Niña
1	between	0.17	and	0.50	center is	0.33	slightly warm	−8	between	−2.00	and	−2.40	center is	−2.20	outstanding La Niña
0	between	0.00	and	0.17	center is	0.085	neutral, above zero	−9		−2.40	and downward				extraordinary La Niña

els and is at best marginal for comparisons between model types. The results presented here should be regarded as only a glimpse of the 1990s state of the art in climate prediction. Moreover, model evaluations during high-amplitude observed events do not necessarily provide more revealing performance information than evaluations over periods with average fluctuations; we do not know if strong events are inherently more predictable. On the basis of this glimpse of model skill, it is suggested that our ability to forecast ENSO conditions is moderate at lead times of one to three seasons, with median correlation skill in the neighborhood of 0.6. This skill level is clearly usable and helpful, but not excellent. For this particular period, certain statistical models and the relatively comprehensive dynamical models appear to have done particularly well, with correlation scores near 0.8. This study did not attempt to assess model skill at lead times of a year or greater. While skill would clearly become more marginal at these leads, whether it would be usable is an open question.

It is important to recognize that forecast success for the ENSO phenomenon does not necessarily imply a comparable success for the impacts over remotely teleconnected regions such as North America. Because ENSO extremes often emerge in boreal spring or early summer (between March and June) and because ENSO behavior during this time of year is especially difficult to predict, forecasting the onset of ENSO events is still something at which we are not very skillful even at 4–6 months of lead time. Knowledge of the subsurface sea temperature data has resulted in some skill increase, as subsurface temperature anomalies in the western equatorial Pacific have some tendency to become SST anomalies in the key ENSO SST regions the following year (Smith et al. 1995). In 197, many of the models produced their initial strong El Niño forecasts at about the same time that TOGA–TAO (tropical atmosphere–ocean) observations indicated westerly wind bursts in the western and central tropical Pacific, and rapid increases in east-central Pacific SST were observed. Thus, the forecasting of a strong El Niño in 1997 did not appear in the media until June 1997, following press releases of various modeling groups. On a climate timescale this does not represent a major forecast success, and regions that experience immediate impacts (e.g., Kiribati, or the coasts of Ecuador or Peru) had little time to prepare.

However, because the ENSO condition during boreal summer tends to persist through the remainder of the calendar year and into the first 1–3 months of

TABLE 9. Summarization of model forecasts by one-digit forecast coding, against observations, blocked by individual model for consecutive start times. A selected subset of six of the eight models forecasting the Niño 3.4 region in physical units (°C) is shown.

1996							1997												1998									Model
summer				fall								summer				fall								summer			fall	
Jun			Sep			Dec			Mar			Jun			Sep			Dec			Mar			Jun			Sep	
−1	**−1**	**−1**	**−1**	**−1**	**−2**	**−2**	**−1**	**−0**	**0**	**1**	**2**	**4**	**5**	**6**	**7**	**7**	**8**	**8**	**7**	**6**	**4**	**2**	**0**	**−1**	**−2**	**−3**	**−3**	**Observed**
s			0	0	0	0	0	1	1																			
			s			0	0	1	1	1	1	1																
						s			0	0	1	1	1	2	2													
									s			1	1	1	1	1	2	2										NCEP constructed analog
												s			2	2	3	3	3	3	3							
															s			5	4	4	3	2	2	1				
																		s			3	2	1	0	0	−1	−1	
−1	**−1**	**−1**	**−1**	**−1**	**−2**	**−2**	**−1**	**−0**	**0**	**1**	**2**	**4**	**5**	**6**	**7**	**7**	**8**	**8**	**7**	**6**	**4**	**2**	**0**	**−1**	**−2**	**−3**	**−3**	**Observed**
s			−1	−1	−1	−1	−1	0	0																			
			s			−2	−1	0	0	0	1	1																
						s			−4	−3	−2	−2	−1	0	0													
									s			−1	−1	−1	−1	1	2	3										UBC neural net
												s			3	3	3	3	3	3	2							
															s			5	4	3	2	1	0	−1				
																		s			3	2	1	1	0	−1	−1	
−1	**−1**	**−1**	**−1**	**−1**	**−2**	**−2**	**−1**	**−0**	**0**	**1**	**2**	**4**	**5**	**6**	**7**	**7**	**8**	**8**	**7**	**6**	**4**	**2**	**0**	**−1**	**−2**	**−3**	**−3**	**Observed**
s			−1	−1	−1	−1	−1	0	0																			
			s			0	0	0	0	0	1	1																
						s			−1	0	0	1	1	1	1													
									s			1	2	4	4	4	5	5										CSU CLIPER (regression)
												s			5	6	6	7	7	6	6							
															s			7	7	7	7	5	3	2				
																		s			5	3	2	1	0	−1	−2	
−1	**−1**	**−1**	**−1**	**−1**	**−2**	**−2**	**−1**	**−0**	**0**	**1**	**2**	**4**	**5**	**6**	**7**	**7**	**8**	**8**	**7**	**6**	**4**	**2**	**0**	**−1**	**−2**	**−3**	**−3**	**Observed**

Table 9. *Continued.*

1996							1997												1998									Model
summer				**fall**								**summer**				**fall**								**summer**			**fall**	
Jun			**Sep**			**Dec**			**Mar**			**Jun**			**Sep**			**Dec**			**Mar**			**Jun**			**Sep**	
–1	**–1**	**–1**	**–1**	**–1**	**–2**	**–2**	**–1**	**–0**	**0**	**1**	**2**	**4**	**5**	**6**	**7**	**7**	**8**	**8**	**7**	**6**	**4**	**2**	**0**	**–1**	**–2**	**–3**	**–3**	**Observed**
s			0	0	0	0	1	2	2																			
			s			1	1	2	2	2	3	3																
						s			1	1	2	2	2	3	3													University of Oxford
									s			1	2	2	2	3	3	3										intermed coupl mdl 2
												s			1	1	2	2	2	2	2							
															s			5	5	5	5	5	4	4				
																		s			8	8	7	7	5	3	2	
–1	**–1**	**–1**	**–1**	**–1**	**–2**	**–2**	**–1**	**–0**	**0**	**1**	**2**	**4**	**5**	**6**	**7**	**7**	**8**	**8**	**7**	**6**	**4**	**2**	**0**	**–1**	**–2**	**–3**	**–3**	**Observed**
s			0	0	0	1	1	1	1																			
			s			0	0	0	0	0	1	1																
						s			–1	–1	0	0	1	1	1													Scripps–MPI
									s			1	1	2	2	2	3	3										hybrid coupled model
												s			2	3	3	3	3	2	2							
															s			6	5	3	2	0	–2	–3				
																		s			7	5	4	3	1	–1	–2	
–1	**–1**	**–1**	**–1**	**–1**	**–2**	**–2**	**–1**	**–0**	**0**	**1**	**2**	**4**	**5**	**6**	**7**	**7**	**8**	**8**	**7**	**6**	**4**	**2**	**0**	**–1**	**–2**	**–3**	**–3**	**Observed**
s			–1	–1	–1	–1	–1	0	0																			
			s			1	1	1	1	1	2	2																
						s			0	1	2	2	2	3	3													
									s			2	3	3	4	4	5	5										NCEP coupled model
												s			7	7	8	8	7	6	5							
															s			8	7	6	5	4	4	3				
																		s			6	5	4	3	3	3	3	
–1	**–1**	**–1**	**–1**	**–1**	**–2**	**–2**	**–1**	**–0**	**0**	**1**	**2**	**4**	**5**	**6**	**7**	**7**	**8**	**8**	**7**	**6**	**4**	**2**	**0**	**–1**	**–2**	**–3**	**–3**	**Observed**

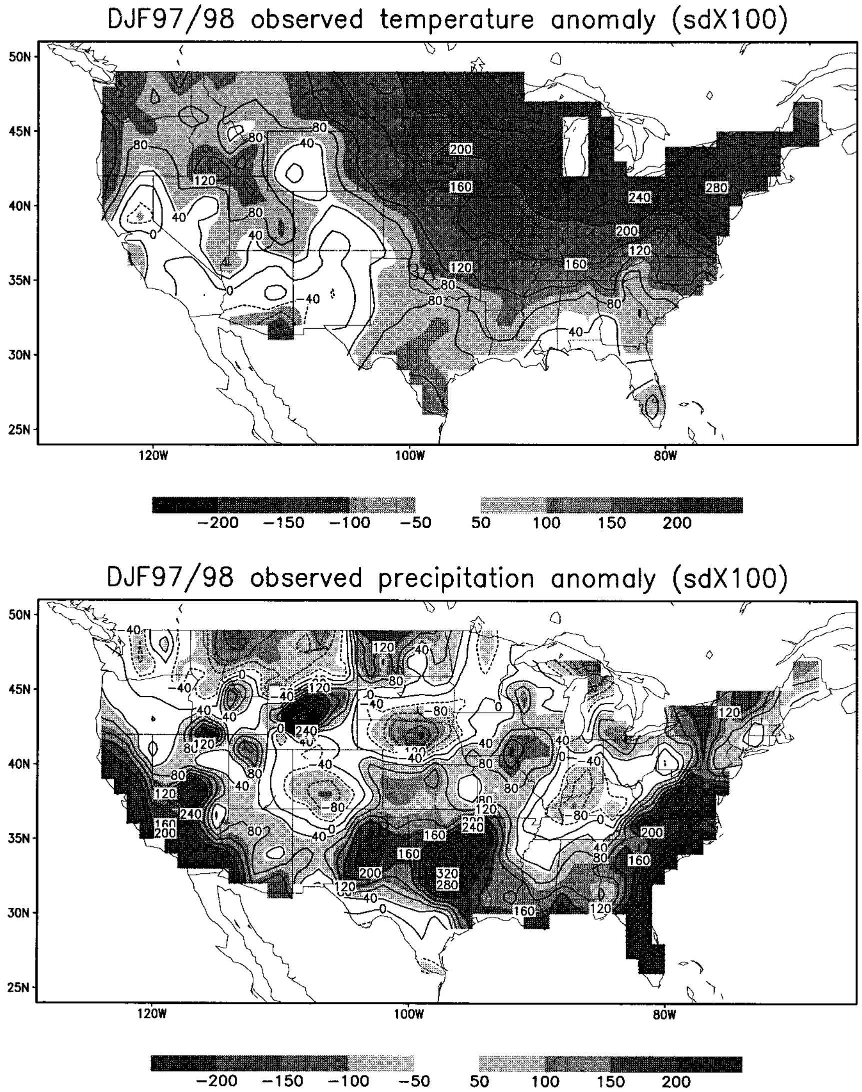

FIG. 3. The observations for Dec–Feb 1997–98, whose forecasts are shown in Fig. 2. Observations are expressed in local standard deviation units (× 100). Contour interval is 40. Negative values are dashed.

the following year, and because North American impacts are strongest during the cold half of the year, successful forecasts of likely winter impacts in North America were possible 5–6 months in advance. This was a result of knowledge of El Niño's simultaneous teleconnections (Horel and Wallace 1981; van Loon and Madden 1981; Ropelewski and Halpert 1986, 1996; Barnett et al. 1994; Graham and Barnett 1995). These winter impacts included heavy rains in southern and central California, mild temperatures in the northern United States, and enhanced rainfall in the Gulf states. A quiet Atlantic fall hurricane season also could have been correctly predicted[3] (Gray et al. 1993). In essence, fairly good 5-month lead forecasts of boreal winter impacts in the Pacific–North American region are possible by merely observing the ENSO condition in late boreal summer, without any proactive ENSO forecasting requirement other than its persistence between August and January. Figure 2 shows the temperature and precipitation probability anomaly forecasts for the mainland United States for December through February 1997–98, issued by the CPC in mid-August 1997. The corresponding observations for that 3-month period are shown in Fig. 3, expressed in local standard deviation units. While the forecasts are far from perfect, they capture many of the gross features of the observed anomaly patterns and are considered successful by climate forecasting standards. On a hit versus miss scoring system using three climatologically equally likely categories (below, near, and above normal), these forecasts achieved Heidke scores of 47 for temperature and 52 for precipitation. In the Heidke scoring system, zero is assigned when the number of correctly forecast locations equals that expected by chance (i.e., 1 out of 3 in this case), and 100 is assigned when all locations are correctly forecast. When only the areas receiving nonzero probability anomalies (the nonblank areas in Fig. 2) are included in the verification, these scores become 86 and 80, respectively. Because oceanographers and climatologists in various groups across the United States and abroad are aware of the North American teleconnections associated with El Niño, this level of fore-

[3] An average Atlantic hurricane season was predicted by Dr. W. Gray of Colorado State University because some of the predictors other than the ENSO state had indicated enhanced activity. Earlier in 1997, when only a mild to moderate warming was expected by late boreal summer, Dr. Gray had forecast somewhat above average hurricane activity.

casting success could have been attained using informal forecasts by any of these groups. Nonetheless, the CPC/NCEP, charged with the responsibility of national public safety, issued its forecasts "loudly" and with ample lead time, allowing users in the mainland and the Pacific Islands to take needed disaster mitigation measures well in advance of the coming climate impacts. A secondary effect of the strong forecasts was an unprecedented awareness of the El Niño by the media and the general public.

The probability anomaly format used in CPC's forecast maps describes the spatial variation of the anomaly of the most favored category in detail. A disadvantage is a lack of information about the remaining two categories, unless at least one additional map is produced. As approximations, CPC uses simple rules in which the missing probabilities are determined from the probability of the favored class. An alternative forecast format adopted by the International Research Institute overcomes this deficiency by specifying the probabilities of all three categories within demarcated regions of the globe. While this provides complete probability information, it suffers from a lack of spatial resolution with large regions sharing the same probabilities. With the advent of the Internet and its high information storage capacity, ways to use either of the above forecast formats in an expanded form are possible. The ultimate in information detail might be provided by a system in which a user clicks the cursor on any map location of interest and obtains a probability density curve on a continuum, with the option of calculating probability anomalies between any two limits. Such a system is currently being developed at CPC for locations within the United States.

Ambiguities were identified in the verbal summaries of the ENSO forecasts in the *ELLF Bulletin.* These occurred specifically in words selected to describe the magnitude, time of onset, and duration of the 1997 El Niño event. Because some of the forecast users lack technical expertise, there is a need for the forecasts to be expressed simply and briefly, but also precisely enough for meaningful verification. It is the forecast community's responsibility to address these concerns. As an initial step toward establishing standards for verbal descriptions of quantitative SST forecasts, a simple system for categorizing the magnitudes of ENSO-related SST forecasts has been presented, with an accompanying verbal classification.

Acknowledgments. The comments of the three official reviewers were valuable and constructive. Earlier on, informal reviews by Arun Kumar, Wayne Higgins, and Huug Van den Dool enriched both the text and the final choice of analyses. Robert Churchill supplied Fig. 2 and Huug Van den Dool provided Fig. 3.

References

Balmaseda, M. A., D. L. T. Anderson, and M. K. Davey, 1994: ENSO prediction using a dynamical ocean model coupled to statistical atmospheres. *Tellus,* **46A,** 497–511.

Barnett, T. P., and R. W. Preisendorfer, 1987: Origins and levels of monthly and seasonal forecast skill for North American surface air temperatures determined by canonical correlation analysis. *Mon. Wea. Rev.,* **115,** 1825–1850.

——, M. Latif, N. Graham, M. Flugel, S. Pazan, and W. White, 1993: ENSO and ENSO-related predictability. Part I: Prediction of equatorial Pacific sea surface tempertures with a hybrid coupled ocean–atmosphere model. *J. Climate,* **6,** 1545–1566.

——, and Coauthors, 1994: Forecasting global ENSO-related climate anomalies. *Tellus,* **46A**, 381–397.

Barnston, A. G., 1994: Linear statistical short-term climate predictive skill in the Northen Hemisphere. *J. Climate,* **7,** 1513–1564.

——, Ed., 1996a: *Experimental Long-Lead Forecast Bulletin.* Climate Prediction Center, National Weather Service, U.S. Department of Commerce, **5** (2). [Available from CPC, World Weather Building, 5200 Auth Rd., Room 604, Camp Springs, MD 20746-4304.]

——, Ed., 1996b: *Experimental Long-Lead Forecast Bulletin.* Climate Prediction Center, National Weather Service, U.S. Department of Commerce, **5** (3). [Available from CPC, World Weather Building, 5200 Auth Rd., Room 604, Camp Springs, MD 20746-4304.]

——, Ed., 1996c: *Experimental Long-Lead Forecast Bulletin.* Climate Prediction Center, National Weather Service, U.S. Department of Commerce, **5** (4). [Available from CPC, World Weather Building, 5200 Auth Rd., Room 604, Camp Springs, MD 20746-4304.]

——, Ed., 1997a: *Experimental Long-Lead Forecast Bulletin.* Climate Prediction Center, National Weather Service, U.S. Department of Commerce, **6** (1). [Available from CPC, World Weather Building, 5200 Auth Rd., Room 604, Camp Springs, MD 20746-4304.]

——, Ed., 1997b: *Experimental Long-Lead Forecast Bulletin.* Climate Prediction Center, National Weather Service, U.S. Department of Commerce, **6** (2). [Available from CPC, World Weather Building, 5200 Auth Rd., Room 604, Camp Springs, MD 20746-4304.]

——, Ed., 1997c: *Experimental Long-Lead Forecast Bulletin.* Climate Prediction Center, National Weather Service, U.S. Department of Commerce, **6** (3). [Available from CPC, World Weather Building, 5200 Auth Rd., Room 604, Camp Springs, MD 20746-4304.]

——, Ed., 1997d: *Experimental Long-Lead Forecast Bulletin.* Climate Prediction Center, National Weather Service, U.S. Department of Commerce, **6** (4). [Available from CPC, World Weather Building, 5200 Auth Rd., Room 604, Camp Springs, MD 20746-4304.]

——, Ed., 1998: *Experimental Long-Lead Forecast Bulletin.* Climate Prediction Center, National Weather Service, U.S. Department of Commerce, **7** (1). [Available from CPC, World Weather Building, 5200 Auth Rd., Room 604, Camp Springs, MD 20746-4304.]

——, and C. F. Ropelewski, 1992: Prediction of ENSO episodes using canonical correlation analysis. *J. Climate,* **5,** 1316–1345.

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

Bryan, K., 1969: A numerical method for the study of the World Ocean. *J. Comput. Phys.,* **4,** 347–376.

Cane, M., and S. E. Zebiak, 1987: Prediction of El Niño events using a physical model. *Atmospheric and Oceanic Variability,* H. Cattle, Ed., Royal Meteorological Society Press, 153–182.

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

Chen, D., S. E. Zebiak, A. J. Busalacchi, and M. A. Cane, 1995: An improved procedure for El Niño forecasting: Implications for predictability. *Science,* **269,** 1699–1702.

——, M. A. Cane, S. E. Zebiak, and A. Kaplan, 1998: The impact of sea level data assimilation on the Lamont model prediction of the 1997/98 El Niño. *Geophys. Res. Lett.,* **25,** 2837–2840.

——, ——, and ——, 1999: The impact of NSCAT winds on predicting the 1997/98 El Niño: A case study with the Lamont model. *J. Geophys. Res.,* in press.

Cox, M. D., 1984: A primitive, 3-dimensional model of the ocean. GFDL Ocean Group Tech. Rep. 1, Geophysical Fluid Dynamics Laboratory, 143 pp.

da Silva, A. M., C. C. Young, and S. Levitus, 1994: *Atlas of Surface Marine Data 1994, Vol. 1: Algorithms and Procedures.* NOAA Atlas NESDIS 6, 83 pp. [Available from User Services Group, National Oceanographic Data Center, NOAA/NESDIS E/OC1, SSMC3, 4th Floor, 1315 East-West Highway, Silver Spring, MD 20910-3282.]

Davis, R. E., 1979: A search for short range climate productivity. *Dyn. Atmos. Oceans,* **3,** 485–497.

Goldenberg, S. D., and J. J. O'Brien, 1981: Time and space variability of tropical Pacific wind stress. *Mon. Wea. Rev.,* **109,** 1190–1207.

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

Graham, N. E., and T. P. Barnett, 1995: ENSO and ENSO-related predictability. Part II: Northern Hemisphere 700-mb height predictions based on a hybrid coupled ENSO model. *J. Climate,* **8,** 544–549.

——, J. Michaelsen, and T. Barnett, 1987a: An investigation of the El Niño–Southern Oscillation cycle with statistical models. 1. Predictor field characteristics. *J. Geophys. Res.,* **92,** 14 251–14 270.

——, ——, and ——, 1987b: An investigation of the El Niño–Southern Oscillation cycle with statistical models. 2. Model results. *J. Geophys. Res.,* **92,** 14 271–14 289.

Gray, W. M., C. W. Landsea, P. Mielke, and K. Berry, 1993: Predicting Atlantic basin seasonal tropical cyclone activity by 1 August. *Wea. Forecasting,* **8,** 73–86.

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

Hsieh, W. W., and B. Tang, 1998: Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bull. Amer. Meteor. Soc.,* **79,** 1855–1870.

Huang, B., and E. K. Schneider, 1995: The response of an ocean general circulation model to surface wind stress produced by an atmospheric general circulation model. *Mon. Wea. Rev.,* **123,** 3059–3085.

——, and J. Shukla, 1997: An examination of AGCM simulated surface stress and low level winds over the tropical Pacific Ocean. *Mon. Wea. Rev.,* **125,** 985–998.

Ji, M., A. Kumar, and A. Leetmaa, 1994a: A multiseason climate forecast system at the National Meteorological Center. *Bull. Amer. Meteor. Soc.,* **75,** 569–577.

——, ——, and ——, 1994b: An experimental coupled forecast system at the National Meteorological Center: Some early results. *Tellus,* **46A,** 398–418.

——, A. Leetmaa, and J. Derber, 1995: An ocean analysis system for seasonal to intraannual climate studies. *Mon. Wea. Rev.,* **123,** 460–481.

——, ——, and V. E. Kousky, 1996: Coupled model predictions of ENSO during the 1980s and 1990s at the National Centers for Environmental Prediction. *J. Climate,* **9,** 3105–3120.

Jiang, N., D. Neelin, and M. Ghil, 1995: Quasi-quadrennial and quasi-biennial variability in the equatorial Pacific. *Climate Dyn.,* **12,** 101–112.

Keppenne, C. L., and M. Ghil, 1992: Adaptive filtering and prediction of the Southern Oscillation index. *J. Geophys. Res.,* **97,** 20 449–20 454.

——, and ——, 1993: Adaptive filtering and prediction of noisy multivariate signals: An application to subannual variability in atmospheric angular momentum. *Int. J. Bif. Chaos,* **3,** 625–634.

Kinter, J. L., III, J. Shukla, L. Marx, and E. K. Schneider, 1988: A simulation of winter and summer circulations with the NMC global spectral model. *J. Atmos. Sci.,* **45,** 2486–2522.

Kirtman, B. P., and E. K. Schneider, 1996: Model based estimates of equatorial Pacific wind stress. *J. Climate,* **9,** 1077–1091.

——, J. Shukla, B. Huang, Z. Zhu, and E. K. Schneider, 1997: Multiseasonal predictions with a coupled tropical ocean global atmosphere system. *Mon. Wea. Rev.,* **125**, 789–808.

Kleeman, R., 1991: A simple model of the atmospheric response to ENSO sea surface temperature anomalies. *J. Atmos. Sci.,* **48,** 3–18.

——, A. M. Moore, and N. R. Smith, 1995: Assimilation of subsurface thermal data into an intermediate tropical coupled ocean–atmosphere model. *Mon. Wea. Rev.,* **123,** 3103–3113.

Knaff, J. A., and C. W. Landsea, 1997: An El Niño–Southern Oscillation climatology and persistence (CLIPER) forecasting scheme. *Wea. Forecasting,* **12,** 633–652.

Latif, M., 1987: Tropical ocean circulation experiments. *J. Phys. Oceanogr.,* **17,** 246–263.

Pacanowski, R. C., and S. G. H. Philander, 1981: Parameterization of vertical mixing in numerical models of tropical oceans. *J. Phys. Oceanogr.,* **11,** 1443–1451.

——, K. Dixon, and A. Rosati, 1993: The GFDL modular ocean model users guide, version 1.0. GFDL Ocean Group Tech. Rep. 2.

Penland, C., 1989: Random forcing and forecasting using principal oscillation pattern analysis. *Mon. Wea. Rev.,* **117,** 2165–2185.

——, and T. Magorian, 1993: Prediction of Niño 3 sea surface temperatures using linear inverse modeling. *J. Climate,* **6,** 1067–1076.

——, and P. D. Sardeshmukh, 1995: The optimal growth of tropical sea surface temperature anomalies. *J. Climate,* **8,** 1999–2024.

——, M. Ghil, and K. M. Weickmann, 1991: Adaptive filtering and maximum entropy spectra, with application to changes in atmospheric angular momentum. *J. Geophys. Res.,* **96,** 22 659–22 671.

Philander, S. G. H., W. J. Hurlin, and A. D. Seigel, 1987: A model of the seasonal cycle in the tropical Pacific Ocean. *J. Phys. Oceanogr.,* **17,** 1986–2002.

Pierce, D. W., 1996: The Hybrid Copled Model, Version 3: Technical Notes.Scripps Instiution of Oceanography Ref. Series No. 96-27, 66 pp. [Available from Climate Research Division, Mail Stop 0224, Scripps Institution of Oceanography, La Jolla, CA 92093-0224.]

Rasmusson, E. M., X. Wang, and C. F. Ropelewski, 1990: The biennial component of ENSO variability. *J. Mar. Sys.,* **1,** 71–96.

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 analyses using optimum interpolation. *J. Climate,* **7,** 929–948.

Ropelewski, C. F., and M. S. Halpert, 1986: North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO). *Mon. Wea. Rev.,* **114,** 2352–2362.

——, and ——, 1996: Quantifying Southern Oscillation–precipitation relationships. *J. Climate,* **9,** 1043–1059.

Shapiro, L. J., 1984: Sampling errors in statistical models of tropical cyclone motion: A comparison of predictor screening and EOF techniques. *Mon. Wea. Rev.,* **112,** 1378–1388.

Slutz, R., S. J. Lubler, J. D. Hiscox, S. D. Woodruff, R. J. Jenne, D. H. Joseph, P. M. Steurer, and J. D. Elius, 1985: Comprehensive Ocean Atmosphere Data Set. National Oceanic and Atmospheric Administration, Boulder, CO, 268 pp. [Available from Climate Research Program, ERL, R/E/AR6, 325 Broadway, Boulder, CO 80303.]

Smith, T. M., A. G. Barnston, M. Ji, and M. Chelliah, 1995: The impact of Pacific Ocean subsurface data on operational prediction of tropical Pacific SST at the NCEP. *Wea. Forecasting,* **10,** 708–714.

——, R. W. Reynolds, R. E. Livezey, and D. C. Stokes, 1996: Reconstruction of historical sea surface temperatures using empirical orthogonal functions. *J. Climate,* **9,** 1403–1420.

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

Syu, H.-H., J. D. Neelin, and D. Gutzler, 1995: Seasonal and interannual variability in a hybrid coupled GCM. *J. Climate,* **8,** 2121–2143.

Tangang, F. T., W. W. Hsieh, and B. Tang, 1997: Forecasting the equatorial Pacific sea surface temperatures by neural network models. *Climate Dyn.,* **13,** 135–147.

——, ——, and ——, 1998a: Forecasting regional sea surface temperatures of the tropical Pacific by neural network models, with wind stress and sea level pressure as predictors. *J. Geophys. Res.,* **103,** 7511–7522.

——, B. Tang, A. H. Monahan, and W. W. Hsieh, 1998b: Forecasting ENSO events: A neural network–extended EOF approach. *J. Climate,* **11,** 29–41.

Van den Dool, H. M., 1994: Searching for analogues, how long must we wait? *Tellus,* **46A**, 314–324.

van Loon, H., and R. A. Madden, 1981: The Southern Oscillation. Part I: Global associations with pressure and temperature in northern winter. *Mon. Wea. Rev.,* **109,** 1150–1162.

Vautard, R., and M. Ghil, 1989: Singular spectrum analysis in nonlinear dynamics with applications to paleoclimatic time series. *Physica D,* **35,** 395–424.

Wolff, J.-O., and E. Maier-Reimer, 1992: HOPE, the Hamburg ocean primitive equation model. Max-Planck-Institut für Meteorologie, 81 pp. [Available from Max-Planck-Institut für Meteorologie, Hamburg D-20146, Germany.]

Wyrtki, K., 1985: Water displacements in the Pacific and the genesis of El Niño cycles. *J. Geophys. Res.,* **90** (C4), 7129–7132.

Xu, J.-S., and H. von Storch, 1990: Predicting the state of the Southern Oscillation using principal oscillation pattern analysis. *J. Climate,* **3,** 1316–1329.

Xue, Y., P. J. Sellers, J. L. Kinter III, and J. Shukla, 1991: A simple biosphere model for global climate studies. *J. Climate,* **4,** 345–364.

Zebiak, S. E., 1984: Tropical atmosphere–ocean interaction and the El Niño/Southern Oscillation phenomenon. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, 261 pp.

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