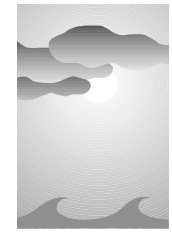
How Much Skill Was There in Forecasting the Very Strong 1997–98 El Niño?



Christopher W. Landsea* and John A. Knaff+

ABSTRACT

The very strong 1997–98 El Niño was the first major event in which numerous forecasting groups participated in its real-time prediction. A previously developed simple statistical tool—the El Niño–Southern Oscillation Climatology and Persistence (ENSO–CLIPER) model—is utilized as a baseline for determination of skill in forecasting this event. Twelve statistical and dynamical models were available in real time for evaluation. Some of the models were able to outperform ENSO–CLIPER in predicting either the onset or the decay of the 1997–98 El Niño, but none were successful at both for a medium-range two season (6–8 months) lead time. There were no models, including ENSO–CLIPER, able to anticipate even one-half of the actual amplitude of the El Niño's peak at medium-range (6–11 months) lead. In addition, none of the models showed skill (i.e., lower root-mean-square error than ENSO–CLIPER) at the zero season (0–2 months) through the two season (6–8 months) lead times. No dynamical model and only two of the statistical models [the canonical correlation analysis (CCA) and the constructed analog (ANALOG)] outperformed ENSO–CLIPER by more than 5% of the root-mean-square error at the three season (9–11 months) and four season (12–14 months) lead time. El Niño impacts were correctly anticipated by national meteorological centers one half-year in advance, because of the tendency for El Niño events to persist into and peak during the boreal winter. Despite this, the zero to two season (0–8 month) forecasts of the El Niño event itself were no better than ENSO–CLIPER and were, in that sense, not skillful—a conclusion that remains unclear to the general meteorological and oceanographic communities.

1. Introduction

The very strong 1997–98 El Niño caused dramatic worldwide changes to weather patterns, such as drought in Indonesia, extreme rains in Peru and Ecuador, and a shutdown of the Atlantic hurricane season (Bell and Halpert 1998). While past El Niño events have also produced similar though not as large effects, this was the first El Niño in which national meteorological centers made accurate forecasts of its impacts several months in advance (Barnston et al. 1999a,b). One issue that remains unresolved is how much skill did the various El Niño–Southern Oscillation (ENSO) forecast methodologies have for the event itself at various lead times. This issue was originally addressed in Barnston et al. (1999a), which analyzed the eight three-monthly forecast times from June 1996 to March 1998 for the operationally available statistical and dynamical models. They found that some of the models performed worse than a persistence control forecast, while most of the models performed quite well relative to persistence for the duration of the 1997–98 El Niño event. Both statistical and dynamical models were found in each categorization of performance. However, three key issues require a revisitation of this topic. The first is that an explicit analysis of the models' performance is needed for the onset and decay of the event. Second, stratification of the various forecasts would allow for analysis of how skill changes with increased lead time. Finally, the error analysis should be put into the context of evaluating the available schemes with respect to a common "no-skill" threshold that is more challenging than simple persistence alone: the El Niño–

*NOAA/AOML/Hurricane Research Division, Miami, Florida.
+NOAA/Cooperative Institute for Research in the Atmosphere, Fort Collins, Colorado.
Corresponding author address: Dr. Chris Landsea, NOAA/AOML/Hurricane Research Division, 4301 Rickenbacker Causeway, Miami, FL 33149.
E-mail: landsea@aoml.noaa.gov
In final form 28 January 2000.

Southern Oscillation Climatology[1] and Persistence (ENSO–CLIPER)[2] model (Knaff and Landsea 1997).

In evaluating ENSO prediction models, two aspects should be considered: do the forecasts present useful information and do the forecasts have skill? The first does not automatically imply the second holds true and vice versa. The first aspect of "usefulness" is whether the predictions can differentiate between the phases (El Niño, La Niña, and neutral) and, when an El Niño or La Niña is present, to know the approximate magnitude of it. Barnston et al. (1999a) delineate ENSO phases into the five categories based upon approximately 1.0°C differences for the Niño-3.4 SST anomaly region: strong/very strong El Niño, weak/moderate El Niño, neutral ENSO, weak/moderate La Niña, and strong/very strong La Niña. (The same divisions could be done with approximately 1.35°C for the Niño-3 SST anomaly region.) We apply this 1.0°C (1.35°C for Niño-3) threshold as a criterion below which determines when a prediction is "useful" from a root-mean-square error (rmse) analysis (Spiegel 1988). This terminology will be utilized for evaluating onset and decay, as well as the whole 1997–98 El Niño event.

Second, to provide a baseline of skill in seasonal ENSO forecasting, we utilize ENSO–CLIPER (Knaff and Landsea 1997) as a simple, statistical model that takes best advantage of the climatology of past ENSO events, persistence of the initial conditions, and their current multimonth trend. The output from ENSO–CLIPER replaces the use of persistence of anomalies as a skill threshold, although it is recognized that other simple statistical models could be employed for this test. "Skill" is then redefined as *the ability to improve upon ENSO–CLIPER:* a more difficult task. ENSO–CLIPER was developed as a multiple least squares regression from a total pool of 14 possible predictors that were screened for the best predictors based upon 1950–94 developmental data. A range of zero to four predictors were chosen in determining the regression models, developed separately for each initial calendar month. Most (72%) of the 480 (12 months × 5 predictands × 8 leads) ENSO–CLIPER regression equations developed contain only zero, one, or two predictors. The predictands to be forecast include Niño-1+2 (0°–10°S, 80°–90°W), Niño-3 (5°N–5°S, 90°–150°W), Niño-3.4 (5°N–5°S, 120°–170°W), and Niño-4 (5°N–5°S, 150°W–160°E) sea surface temperature (SST) anomaly indices for the equatorial eastern and central Pacific and the Southern Oscillation Index (SOI) (Fig. 1) at lead times ranging from zero seasons (0–2 months) through seven seasons (21–23 months). The previous one-, three-, and five-month observed values and corresponding time trends of the various Niño SST indices and the SOI also serve as possible predictors. The pool contains 14 potential predictors: six from the predictand itself and two each from the other four ENSO indices (Niño SST regions and SOI). Prospective predictors were retained only if they correlated in the regression test at a significance level beyond 95% using a Student's t-test and increased the total variance explained by at least 2.5%. To reduce the chance of statistical overfitting, only one of the three time averages for predictors (one, three, or five months) and only one of the three time trends are allowed to be chosen from each of the various Niño SST indices and the SOI. Though hindcast ability is strongly seasonally dependent, substantial improvement is achieved over simple persistence wherein largest improvements occur for the two to seven season (6–23 months) lead times.

[1]Study of the mean physical state of the atmosphere together with its statistical variations in both space and time as reflected in the weather behavior over a period of many years (WMO 1992). In addition to the presentation of climatic data, it includes the analysis of the causes of differences of climate, and the application of climatic data to the solution of specific design or operational problems (Huschke 1959).

[2]The use of "CLIPER" types of models as the standard for comparison has a long history in other forecasting arenas, as for example, in tropical cyclone track and intensity forecasting (Neumann 1972; Jarvinen and Neumann 1979). Such optimal combinations of climatology, persistence, and trend have proven to be invaluable tools in validating new tropical cyclone prediction schemes, in both real time and hindcasts (DeMaria et al. 1990; DeMaria 1996).

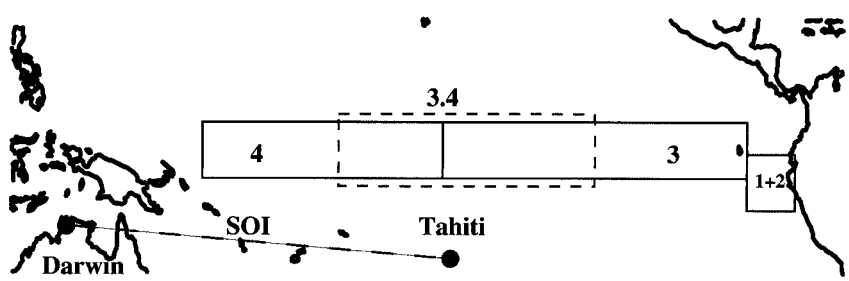


FIG. 1. Location of the four SST indices and of the SOI stations utilized as predictors and predictands in ENSO–CLIPER.

We interpret ENSO–CLIPER as capturing the climatological aspects of the whole ENSO complex in depicting both mean conditions and propagation of those features in time. In essence this model, given initial conditions of ENSO (SST anomalies in Niño regions 1+2, 3, 3.4, and 4, and the SOI) and the recent past valid at a particular time, will fit using regression techniques the best evolution from those initial conditions. The method has been frozen following its development (over 40 years of dependent data) and yields the mean climatological evolutions for that period. The ENSO–CLIPER model thus offers a baseline no-skill forecast of ENSO variability and a useful (as defined earlier) ENSO prediction as well.[3]

2. Results

Operational forecasts from the various ENSO prediction models were provided through digitizing of results shown in the *Experimental Long-Lead Forecast Bulletin* (ELLFB; Barnston 1995, 1996, 1997a; Kirtman 1998) and the *Climate Diagnostics Bulletin* (Kousky 1995, 1996, 1997, 1998) with confirmation of a few cases' values from A. G. Barnston (1999, personal communication). [Note that while ENSO–CLIPER has been run in real time from late 1996 onward, June 1997 marked its first appearance in the ELLFB, just before the publication of Knaff and Landsea (1997).] Only models that provided at least a two season lead forecast valid for the entire duration of the El Niño event were considered.[4] The zero and one season forecasts (0–5 months) will be termed "short-range" predictions. The two and three season forecasts (6–11 months) will be called "medium-range" predictions. "Long-range" predictions are those for the remaining four to seven season leads (12–23 months). All forecast SST anomalies were adjusted to the standard 1950–79 base period climatology, resulting in small (0°–0.2°C) alterations in some of the model predictions. An analysis and comparison of the model forecasts for seven nonoverlapping seasons was performed from just before the onset of the El Niño in early 1997 to just after the decay of the event in mid-1998.

Models for forecasting the ENSO phenomena can be broadly broken into two categories: statistical and dynamical. The statistical models range from simple linear regression schemes to more sophisticated techniques such as neural networks, time series analysis, multivariate multiple regression, and ensemble methods. Dynamical ENSO models range from simplified linear shallow water equations for both the ocean and atmosphere, to the intermediate models with two active layers representing the ocean, to the hybrid coupled models that have a comprehensive ocean system coupled to a statistical atmosphere, to the comprehensive coupled models with multilevel representations of both the ocean and atmosphere. However, even the comprehensive coupled models still require statistical corrections to account for systematic biases in the model output, likely related to the extremely difficult task of ideally specifying the transfer of heat, moisture, and momentum from ocean to atmosphere and vice versa.

The models evaluated for prediction of Niño-3.4 SST anomalies were the following: ENSO–CLIPER (Knaff and Landsea 1997), the constructed analog statistical model (ANALOG; Van den Dool 1994), the consolidation (ensemble) statistical method (CONSOL; Unger et al. 1996), the National Center for Environmental Prediction comprehensive dynamical model (NCEP; Ji et al. 1996), the neural network statistical model (NEURAL; Tangang et al. 1997), the Scripps/Max Planck Institute hybrid dynamical model (SCR/MPI; Barnett et al. 1993), the canonical correlation analysis statistical model (CCA; Barnston and Ropelewski 1992), and the University of Oxford intermediate dynamical model (OXFORD;[5] Balmaseda et al. 1994). The models evaluated for predictions of Niño-3 SST anomalies were ENSO–CLIPER, the Bureau of Meteorology Research Centre simplified dynamical model (BMRC; Kleeman et al. 1995), the Center for Ocean–Land–Atmosphere Studies comprehensive dynamical model (COLA; Kirtman et al.

[3]Details about the ENSO–CLIPER model, including its anticipated forecast performance and its predictor selection rules, can be found in Knaff and Landsea (1997). The program to run ENSO–CLIPER and independent forecasts from ENSO–CLIPER since 1 January 1993 are available online at http://www.aoml.noaa.gov/hrd/Landsea/landsea_bio.html.

[4]Here we follow the nomenclature of Barnston and Ropelewski (1992), wherein zero lead indicates predictions for the next immediately upcoming month (their Fig. 5). For example, a forecast issued on 1 February for February through April conditions is termed a zero lead seasonal forecast. A 1 February forecast for May through July is a one season lead forecast, and so forth.

[5]These forecasts are actually for the region "Eq. 2" (5°N–5°S, 130°–170°W), which is quite similar to Niño-3.4

1997), the Lamont-Doherty simplified dynamical model (LDEO; Zebiak and Cane 1987), the singular spectrum analysis/maximum entropy method statistical model (SSA/MEM; Keppenne and Ghil 1992), and the linear inverse statistical model (LIM; Penland and Magorian 1993). An important caveat is that some of these models were not specifically designed for *only* forecasting ENSO. However, the purpose of this assessment is to analyze the performance of the models in predicting ENSO, as measured by the Niño-3 or Niño-3.4 SST anomalies. Forecasts based upon the traditional persistence (PERSIS) of anomalies from the initial season were also utilized for comparison purposes.

Figure 2 presents the two season lead forecasts and the observed SST anomalies, referred to as verifications, for all 12 schemes as well as ENSO–CLIPER and persistence. Except for the observations, the forecast values in the figure are not a time series as the individual points came from separate runs from the models. It is apparent that none of the models performed extremely well for this medium-range (6–8 months) forecast. Some of the models did well for the onset of the El Niño event (e.g., NCEP and CCA), some were able to capture the decay of the event (e.g., ENSO–CLIPER, SCR/MPI, and LIM), but the peak SST anomalies that occurred in late 1997 were dramatically underestimated (by one-half the amplitude or greater) by all models at this lead. Additionally, none of the models captured *both* the onset and the decay of the El Niño event with success at this two season lead.

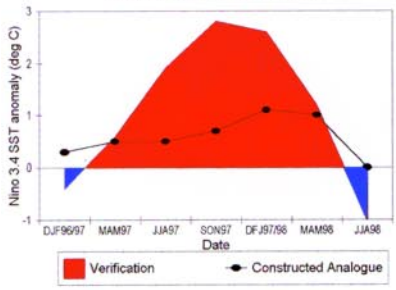
To better quantify how the models performed for the life cycle of the El Niño event, an evaluation of forecast skill was conducted for the times of the El Niño's onset and decay, defined similarly to that of Trenberth (1997). It is not expected that the ENSO prediction models could (or should) perform better during the onset/decay phases relative to the whole ENSO lifecycle. However, these phases of El Niño are of keen climatological and societal interest because that is when the ENSO teleconnections typically begin and end. The "onset" stage is defined as the first three-month period when the Niño-3.4 SST anomalies exceed +0.4°C, which was March–May 1997. For Niño-3, the onset occurred during the same months. The "decay" stage is defined as the first three-month period that Niño 3.4 SST anomalies averaged lower than +0.4°C once again, which was April–June 1998. For Niño-3, the decay began slightly later during May–July 1998.

The two three-month forecasts from each scheme that bracket the onset and decay phase are evaluated for rmse in Table 1. For onset, nearly all prediction schemes proved to be useful for the short- and medium-range forecasts. The exceptions were LDEO (not useful at either short or medium ranges) and SCR/MPI (not useful at the medium range). At the long-range forecasts, only ANALOG, BMRC, CCA, ENSO–CLIPER, and LIM provided useful predictions, though several (NCEP, NEURAL, OXFORD, SCR/MPI, and SSA/MEM) are not run this far into the future. However, when one places the additional constraint of having to outperform ENSO–CLIPER, the number of models that were useful and showed skill for the El Niño onset are reduced: short range—BMRC, CCA, COLA, LIM, and SSA/MEM; medium range—ANALOG, BMRC, COLA, CONSOL, LIM, NCEP, and SSA/MEM; and long range—BMRC. Note that it is quite possible to be useful and show skill at a medium-range forecast, but for these not to hold true at the short range (e.g., ANALOG, CONSOL and NCEP). For the medium-range, two season (6–8 months) forecasts depicted in Fig. 2, NCEP had by far the most useful and skillful predictions of the El Niño onset improving upon ENSO–CLIPER by 31% in terms of rmse.

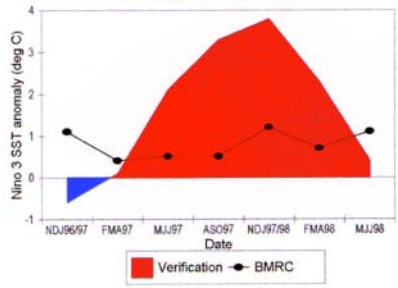
For the decay of the El Niño in the boreal spring of 1998, all of the ENSO schemes provided useful short- and medium-range predictions except for CONSOL (not useful at either the short or medium range), LDEO (not useful at the medium range), NCEP (not useful at the medium range), and OXFORD (not useful at either the short or medium range). For the long range, only ANALOG, COLA, and ENSO–CLIPER provided useful forecasts; though again several models (NCEP, NEURAL, OXFORD, SCR/MPI, and SSA/MEM) are not run out this far into the future. When the analysis also includes improving upon ENSO–CLIPER's forecast of the decay, only a few models are both useful and show skill: short range—ANALOG, CCA; medium range—ANALOG, COLA, and LIM; and long range—none. For the medium-range, two season (6–8 months) predictions depicted in Fig. 2, LIM had the most skillful prediction of the

FIG. 2 (*facing page*). Two-season (6–8 month) lead forecasts of Niño-3.4 or Niño-3 SST anomalies from the following prediction schemes: ANALOG, BMRC, CCA, COLA, CONSOL, ENSO–CLIPER, LDEO, LIM, NCEP, NEURAL, OXFORD, PERSIS, SCR/MPI, and SSA/MEM.

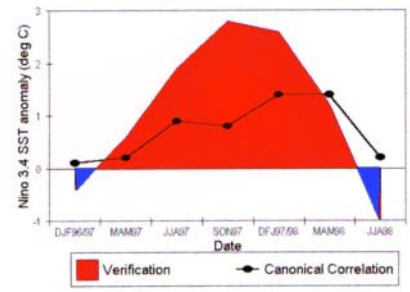
Constructed Analogue Forecasts
Two season lead for Nino 3.4



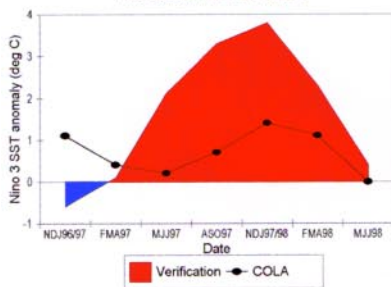
BMRC Forecasts
Two season lead for Nino 3



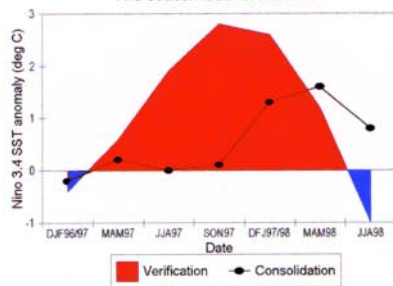
Canonical Correlation Forecasts
Two season lead for Nino 3.4



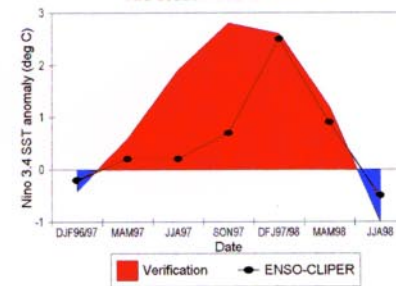
COLA Forecasts
Two season lead for Nino 3



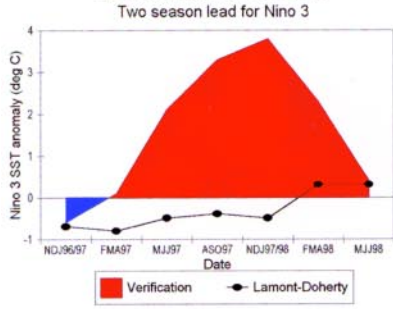
Consolidation Forecasts
Two season lead for Nino 3.4



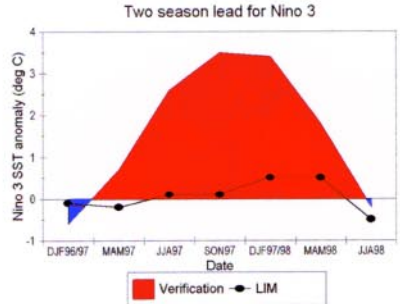
ENSO-CLIPER Forecasts
Two season lead for Nino 3.4



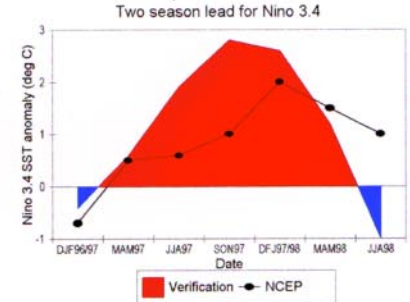
Lamont-Doherty Forecasts
Two season lead for Nino 3



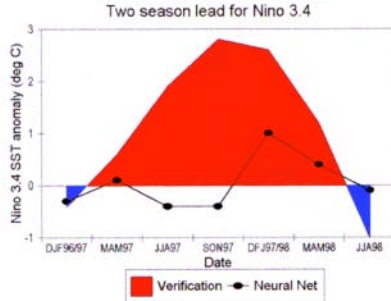
Linear Inverse Forecasts
Two season lead for Nino 3



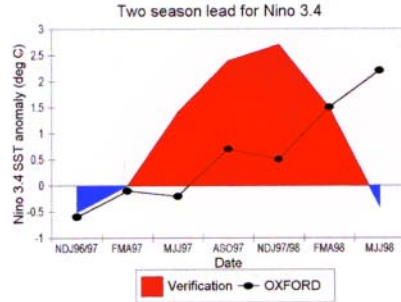
NCEP Coupled Model Forecasts
Two season lead for Nino 3.4



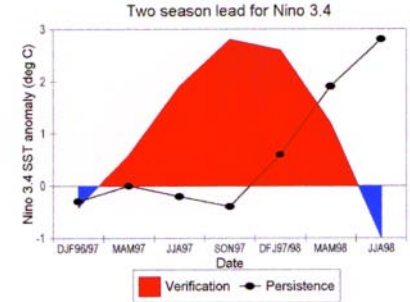
Neural Network Forecasts
Two season lead for Nino 3.4



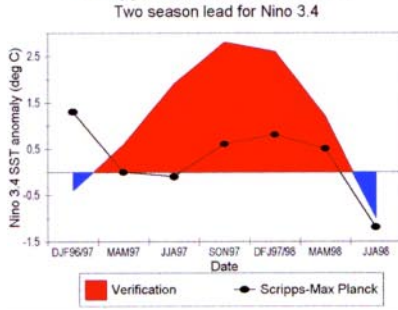
Oxford Coupled Forecasts
Two season lead for Nino 3.4



Persistence Forecasts
Two season lead for Nino 3.4



Scripps-Max Planck Forecasts
Two season lead for Nino 3.4



Singular Spectrum & MEM Forecasts
Two season lead for Nino 3

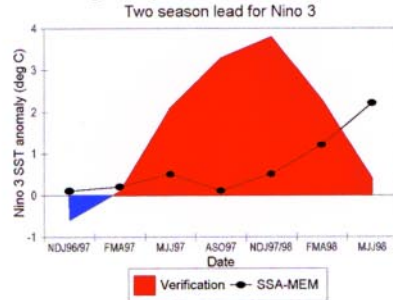


TABLE 1. Root-mean-square error (rmse) statistics for the onset and decay of the 1997–98 El Niño event by the various ENSO forecasting methodologies in °C. "DJF" indicates that the forecast periods were Dec–Feb, Mar–May, Jun–Aug, Sep–Nov. "NDJ" indicates that the forecast periods were Nov–Jan, Feb–Apr, May–Jul, Aug–Oct. Values in boldface indicate that the model in question outperformed ENSO–CLIPER (i.e., had smaller rmse).

Scheme	Lead 0	Lead 1	Lead 2	Lead 3	Lead 4	Lead 5	Lead 6	Lead 7
ONSET: Niño-3.4								
Verification dates: FMA and MJJ 1997								
ENSO–CLIPER–NDJ	0.57	0.85	0.85	0.64	0.74	0.86	0.92	0.92
PERSIS–NDJ	1.05	1.36	1.21	1.21	1.55	1.68	1.41	1.05
OXFORD–NDJ	0.92	0.86	0.85	1.03				
Verification dates: DJF 1996–97 and MAM 1997								
ENSO–CLIPER–DJF	0.45	0.41	0.32	0.55	0.57	0.61	0.57	0.60
PERSIS–DJF	0.72	0.63	0.43	0.70	1.01	0.96	0.65	1.08
ANALOG–DJF	0.67	0.45	0.50	**0.50**	0.57	0.65		
CCA–DJF	0.45	**0.35**	0.45	0.86	0.78			
CONSOL–DJF	—	0.45	0.32	**0.16**				
NCEP–DJF	0.50	0.54	**0.22**					
NEURAL–DJF	1.20	0.93	0.36					
SCR/MPI–DJF	—	0.67	1.14					
ONSET: Niño-3								
Verification dates: FMA and MJJ 1997								
ENSO–CLIPER-NDJ	1.17	1.28	1.27	1.27	1.35	1.48	1.42	1.42
PERSIS–NDJ	1.50	1.94	1.80	1.80	1.88	2.15	2.07	1.70
BMRC–NDJ	**0.41**	**1.06**	**1.15**	1.37	**1.12**	**1.20**	**0.74**	
COLA–NDJ	**0.79**	**1.26**	1.39	**1.08**				
LDEO	1.77	1.97	1.95	1.95	2.06			
SSA/MEM–NDJ	**1.06**	1.44	**1.13**	1.34				
Verification dates: DJF 1996–97 and MAM 1997								
ENSO–CLIPER–DJF	0.74	0.51	0.71	0.75	0.44	0.82	0.71	0.55
PERSIS–DJF	0.94	0.79	**0.60**	**0.71**	0.85	1.03	**0.50**	1.08
LIM–DJF	0.76	0.57	0.73	**0.71**	0.60			

TABLE 1. *Continued.*

Scheme	Lead 0	Lead 1	Lead 2	Lead 3	Lead 4	Lead 5	Lead 6	Lead 7
DECAY: Niño-3.4								
Verification dates: FMA and MJJ 1998								
ENSO–CLIPER–NDJ	0.65	0.50	0.36	1.17	0.81	0.86	0.92	0.98
PERSIS–NDJ	1.59	2.28	1.98	1.66	1.44	1.27	1.20	1.49
OXFORD–NDJ	—	2.09	1.84	1.46				
Verification dates: MAM and JJA 1998								
ENSO–CLIPER–DJF	0.57	0.76	0.41	0.79	0.96	1.14	1.10	1.02
PERSIS–DJF	1.78	2.79	2.73	2.09	1.60	**1.08**	**1.02**	1.27
ANALOG–DJF	**0.16**	**0.35**	0.72	0.87	1.03	**1.08**		
CCA–DJF	**0.45**	**0.67**	0.86	1.06	1.42			
CONSOL–DJF	—	1.10	1.30	1.10	1.14			
NCEP–DJF	0.65	1.06	1.43					
NEURAL–DJF	0.79	0.79	0.81					
SCR/MPI–DJF	—	—	0.51					
DECAY: Niño-3								
Verification dates: FMA and MJJ 1998								
ENSO–CLIPER–NDJ	0.75	0.76	0.92	1.63	1.53	1.56	1.36	1.36
PERSIS–NDJ	1.71	2.51	2.06	1.97	2.06	2.04	1.99	1.99
BMRC–NDJ	1.10	1.24	1.23	**1.42**	**1.51**	**1.51**	1.65	
COLA–NDJ	0.98	1.01	**0.89**	**0.92**	**1.41**			
LDEO–NDJ	0.86	1.14	1.42	1.93	**1.35**			
SSA/MEM–NDJ	1.17	1.64	1.49	**1.60**				
Verification dates: MAM and JJA 1998								
ENSO–CLIPER	0.57	0.71	1.20	1.13	1.21	1.12	1.22	1.22
PERSIS	1.81	2.82	2.68	2.13	1.81	1.51	1.35	1.49
LIM	0.92	1.12	**0.94**	**1.08**	1.64			

El Niño decay improving upon ENSO–CLIPER by 22% in terms of rmse.

To best assess the event-long performance (onset, peak, and decay) of the individual forecasting schemes, the rmse was calculated for each lead time in Table 2.[6] Note that the rmse errors for PERSIS, which has been the traditional standard for determining skill, are quite large and are only exceeded by LDEO (at a zero and one-season lead) and LIM (at a zero season lead). Additionally, the two-season lead forecasts of LEDO had errors equal to those of PERSIS. However, to state—as has been done traditionally—that these two-season lead forecasts from LDEO in Fig. 2 are on the threshold of having "skill" based on persistence as a control highlights the need for a more stringent standard. For the duration of the El Niño event, the models have only limited ability to show usefulness. At the short range, ANALOG, BMRC, CCA, COLA, ENSO–CLIPER, NCEP, NEURAL, and SSA/MEM provide useful predictions. None of the models give useful forecasts at the medium- and long-range lead times.

Figure 3 helps to summarize the details found in Table 2. These diagrams provide a direct comparison of the rmse of the various models versus ENSO–CLIPER. Persistence is not plotted because of the extremely poor predictions utilizing this methodology, which would dominate the *y* axis in the figure. It is readily apparent that none of the ENSO models, statistical and dynamical, were able to provide skillful forecasts at the short-range lead times. Only CCA, ANALOG, and COLA were able to outperform ENSO–CLIPER at the medium and long ranges: 23% and 12% lower rmse for CCA, 6% and

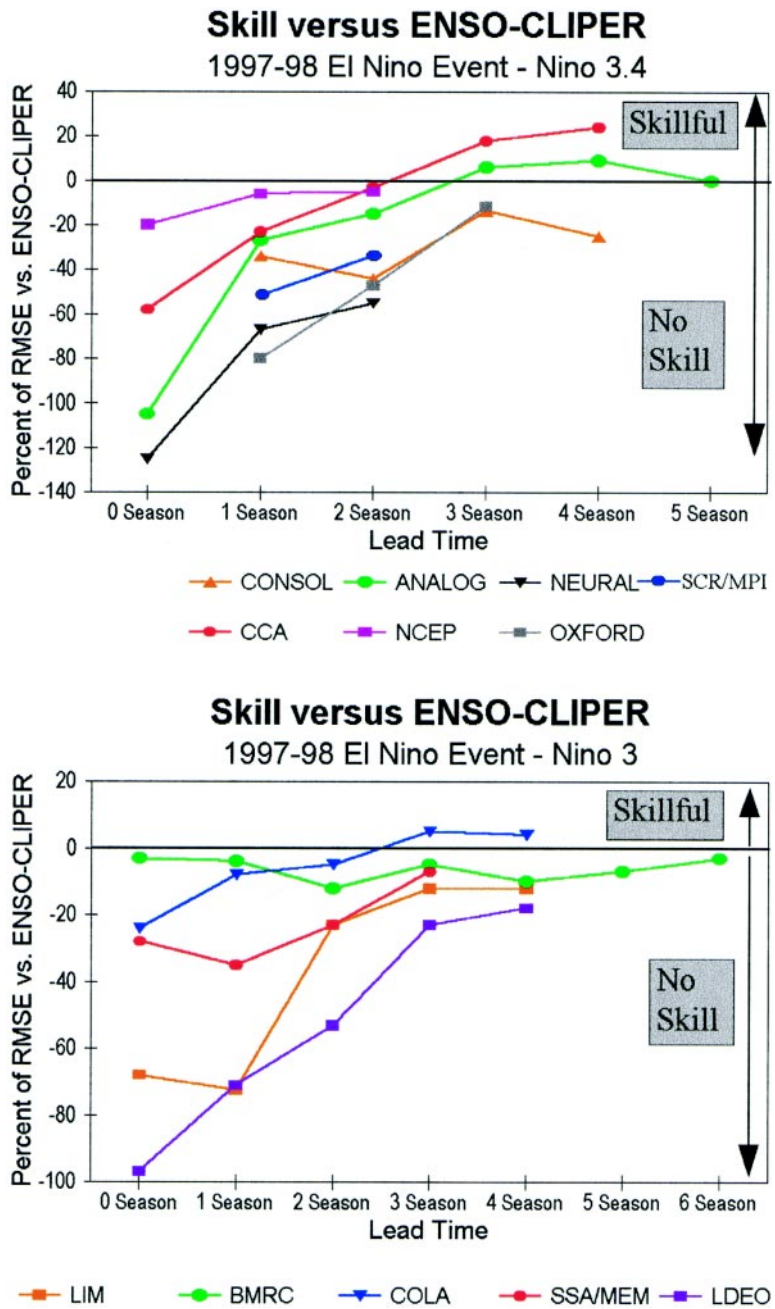


FIG. 3. Skill of various ENSO forecasting methodologies for the 1997–98 El Niño event. Error values for lead times ranging from zero to six seasons in advance are expressed as a percentage of the rmse relative to ENSO–CLIPER. Negative values indicate a forecast performance worse than ENSO–CLIPER and thus no skill. Positive values are improvements over ENSO–CLIPER and thus are "skillful" predictions. Error values are shown for forecasts of (top) Niño-3.4 and (bottom) Niño-3 SST anomalies.

[6]The 1 May 1996 and 1 August 1996 NEURAL forecasts were for the Niño-3 region. The remaining NEURAL forecasts were issued for the Niño-3.4 region. These NEURAL RMSE values should be compared with the following homogeneous ENSO–CLIPER statistics: zero season lead 0.40, one season lead 0.79, two season lead 1.08. The SCR/MPI forecast proved difficult to score because of a lack in standardization of predictions presented in the *Experimental Long-Lead Forecast Bulletin.* Only six one-season lead and seven two-season lead forecasts could be verified. The corresponding verification statistics from ENSO–CLIPER are the ones appropriate for comparison: one season lead 0.81 and two season lead 1.06.

TABLE 2. Root-mean-square error (rmse) statistics for various ENSO forecasting methodologies in °C over the entire 1997–98 El Niño event. "DJF" indicates that the forecast periods were Dec–Feb, Mar–May, Jun–Aug, Sep–Nov. "NDJ" indicates that the forecast periods were Nov–Jan, Feb–Apr, May–Jul, Aug–Oct. All verifications are performed for seven three-month (seasonal) forecasts from early 1997 through mid-1998, except where noted. Values in boldface indicate that the model in question outperformed ENSO–CLIPER (i.e., had smaller rmse).

Scheme	Lead 0	Lead 1	Lead 2	Lead 3	Lead 4	Lead 5	Lead 6	Lead 7
Niño 3.4								
ENSO–CLIPER–NDJ	0.48	0.84	1.03	1.30	1.24	1.14	1.46	1.51
PERSIS–NDJ	1.08	1.75	1.96	1.93	1.88	1.96	1.85	2.12
OXFORD–NDJ	—	1.51	1.51	1.46				
ENSO–CLIPER–DJF	0.40	0.77	1.06	1.38	1.48	1.46	1.32	1.16
PERSIS–DJF	1.19	1.96	2.06	2.12	1.98	1.98	2.06	2.17
ANALOG–DJF	0.82	0.98	1.22	**1.30**	**1.35**	1.46		
CCA–DJF	0.63	0.95	1.09	**1.13**	**1.12**			
CONSOL–DJF	—	1.03	1.53	1.56	1.85			
NCEP–DJF	0.48	0.82	1.11					
NEURAL–DJF	0.90	1.32	1.67					
SCR/MPI–DJF	—	1.22	1.43					
Niño 3								
ENSO–CLIPER–NDF	0.87	1.30	1.64	1.96	1.96	1.98	1.93	2.09
PERSIS–NDJ	1.32	2.17	2.51	2.59	2.59	2.65	2.73	2.78
BMRC–NDJ	0.90	1.35	1.83	2.06	2.12	2.12	1.98	
COLA–NDJ	1.08	1.40	1.72	**1.85**	**1.88**			
LDEO–NDJ	1.72	2.22	2.51	2.41	2.30			
SSA/MEM–NDJ	1.11	1.75	2.01	2.09				
ENSO–CLIPER–DJF	0.85	1.11	1.61	1.93	1.93	1.96	1.93	1.90
PERSIS–DJF	1.35	2.06	2.51	2.59	2.38	2.33	2.30	2.54
LIM–DJF	1.43	1.90	1.98	2.17	2.17			

9% for ANALOG, and 5% and 4% for COLA at the three- and four-season lead, respectively. The small number of forecasts (seven per lead time) do not make for very meaningful significance testing at this time. It is to be noted that, in general, models improve relative to ENSO–CLIPER as lead time is increased. Thus it is quite possible that models currently only run out of a two-season lead (such as NCEP and SCR/MPI)

might have shown skill if they were integrated further out in time. Finally, *none* of the models are both useful and skillful for the duration of the 1997–98 El Niño event at *any* lead time.

Linear correlation coefficients were also calculated for all models at all forecast periods versus observed anomalies (Table 3). Results were quite similar qualitatively to those of the rmse analysis.

TABLE 3. Linear correlation coefficients (r) for various ENSO forecasting methodologies versus observed anomalies over the entire 1997–98 El Niño event. "DJF" indicates that the forecast periods were Dec–Feb, Mar–May, Jun–Aug, Sep–Nov. "NDJ" indicates that the forecast periods were Nov–Jan, Feb–Apr, May–Jul, Aug–Oct. All verifications are performed for seven three-month seasonal forecasts from early 1997 through mid–1998. Values in boldface indicate that the model in question outperformed ENSO–CLIPER (i.e., had a higher r).

Scheme	Lead 0	Lead 1	Lead 2	Lead 3	Lead 4	Lead 5	Lead 6	Lead 7
Niño 3.4								
ENSO–CLIPER–NDJ	0.92	0.75	0.69	0.17	0.90	0.91	0.51	0.15
PERSIS–NDJ	0.59	0.00	–0.27	–0.23	0.24	–0.14	–0.80	–0.81
OXFORD–NDJ	—	0.28	0.04	**0.20**				
ENSO–CLIPER–DJF	0.96	0.84	0.75	0.56	0.69	0.52	0.85	0.78
PERSIS–DJF	0.57	–0.13	–0.48	–0.45	–0.09	0.00	–0.69	–0.75
ANALOG–DJF	0.96	**0.87**	**0.79**	**0.74**	0.51	0.11		
CCA–DJF	0.91	0.74	0.75	**0.77**	0.58			
CONSOL–DJF	—	0.44	0.13	–0.26	–0.12			
NCEP–DJF	0.94	0.81	0.54					
NEURAL–DJF	0.84	0.64	0.33					
SCR/MPI–DJF	—	0.62	0.39					
Niño 3								
ENSO–CLIPER–NDJ	0.90	0.81	0.81	0.63	0.73	0.69	0.70	0.47
PERSIS–NDJ	0.67	0.18	–0.14	–0.23	0.24	0.21	–0.69	–0.87
BMRC–NDJ	0.87	0.74	–0.08	–0.72	–0.52	–0.42	–0.16	
COLA–NDJ	**0.98**	0.80	0.29	–0.09	–0.29			
LDEO–NDJ	0.62	0.37	0.14	**0.70**	0.11			
SSA/MEM–NDJ	0.69	0.14	–0.13	0.00				
ENSO–CLIPER–DJF	0.92	0.89	0.71	0.68	0.95	0.51	0.74	0.80
PERSIS–DJF	**0.96**	0.30	–0.08	–0.26	0.00	0.50	–0.45	–0.83
LIM–DJF	0.92	0.87	0.71	–0.29	–0.06			

3. Discussion

The analyses here conclude that the best performing model for forecasting the entirety of the very strong 1997–98 El Niño at the short-range lead was the statistical ENSO–CLIPER model (Knaff and Landsea 1997), while the best at the medium-range lead was the canonical correlation analysis statistical model (Barnston and Ropelewski 1992). Thus the use of more complex, physically realistic dynamical models does not automatically provide more reliable forecasts. Increased complexity can increase by orders of magnitude the sources for error, which can cause degradation in skill. Despite the lack of skill in forecasting ENSO itself up to 8 months in advance, national meteorological centers were able to correctly anticipate the effects of the 1997–98 El Niño because of the tendency for El Niño events to persist into and peak during the boreal winter (Barnston et al. 1999a). Indeed, the U.S. Climate Prediction Center's most skillful tools (measured by the Heidke skill score) for predicting the U.S. seasonal precipitation anomalies were the statistical ENSO composites and the statistical optimal climate normals, rather than the NCEP coupled model (Barnston et al. 1999b). (For seasonal temperature anomalies in the United States, the two statistical and one dynamical tools were about equal in skill.) No dynamical models were needed to anticipate a wet and stormy winter for the southern tier of the United States and a warm winter for the northern tier of states.

Within this paper we have utilized ENSO–CLIPER as the baseline methodology against which other prediction schemes can be judged for skill, depending on whether they outperformed ENSO–CLIPER ("skillful") or not ("no skill"). One may, however, alternatively interpret ENSO–CLIPER to be more than a strict combination of climatology and persistence, since it does allow for phase propagation of ENSO within the Niño regions. As we argue in the introduction, this should not invalidate ENSO–CLIPER as a no-skill test since it is providing the climatological evolution of past ENSO events and is simpler than many statistical and all numerical models. Even if one does not agree with this reasoning, two points are still clearly evident: 1) a distinct need exists for a standard against which skill is to be measured in predicting the ENSO phenomena. Use of the simple persistence of anomalies is much too easy a benchmark to exceed. If not ENSO–CLIPER, some other comparative test is essential for ENSO forecasting. [See Qin and Van den Dool (1996) for another creative no-skill benchmark test.] Also, 2) the multiple regression–based ENSO–CLIPER outperformed all of the more sophisticated models, both other statistical schemes as well as numerical techniques, for zero- to two-season lead (0–8 months). Thus these more complex models may not be doing much more than carrying out a pattern recognition and extrapolation of their own. National meteorological centers may wish to consider carefully their resource priorities (both personnel and computers) when the current best tools are the relatively cheap statistical systems, relative to the expensive (developmentally and computationally) dynamical models that have not yet produced skillful and useful operational ENSO forecasts at any forecast lead times.

The results herein may be surprising given the general perception that seasonal El Niño forecasts from dynamical models have been quite successful and may even be considered a solved problem (e.g., Kerr 1998; Stern and Easterling 1999; Porter 1999). Kerr's (1998) report "Models win big in forecasting El Niño" in *Science*, in particular, generated widespread publicity for the success in forecasting the 1997–98 El Niño onset by the comprehensive dynamical models. His report was based upon Barnston's (1997b) unrefereed and incomplete (since only the onset was considered) analysis. No followup mention in *Science* was forthcoming when the Barnston et al. (1999a) paper was finally published showing that the comprehensive dynamical models did not "win big" after all. [It is worth mentioning that the results from Barnston et al. (1999a) do indeed agree quite well in general with what is shown here, though the interpretation is very different.]

Also disturbing is that others are using the supposed success in dynamical El Niño forecasting to support other agendas. As an example, an overview paper by Ledley et al. (1999) to support the American Geophysical Union's "Position Statement on Climate Change and Greenhouse Gases" said the following:

> Confidence in [comprehensive coupled] models [for anthropogenic global warming scenarios] is also gained from their emerging predictive capability. An example of this capability is the development of a hierarchy of models to study the El Niño–Southern Oscillation (ENSO) phenomena. . . . These models can predict the lower frequency responses of the climate system, such as anomalies in monthly and season averages of the sea surface temperatures in the tropical Pacific.

On the contrary, with the results of this study, one could even have *less* confidence in anthropogenic global warming studies because of the lack of skill in predicting El Niño (or, alternatively, the inability of dynamical models to outperform relatively simple statistical schemes). The bottom line is that the successes in ENSO forecasting have been overstated (sometimes drastically) and misapplied in other arenas.

A followup study will assess forecast skill of the strong 1998–2000 La Niña event, which immediately followed the El Niño. This examination of ENSO forecast skill only analyzed seven forecasts per lead time, so the findings here are rough indications of the relative skills of the various models and approaches. It may be that with consideration of the most recent, complete ENSO warm and cold cycle that truly skillful predictions from models are available. But the current answer to the question posed in this article's title is that there was essentially no skill in forecasting the very strong 1997–98 El Niño at lead times ranging from 0 to 8 months when using the performance of ENSO–CLIPER as the no-skill baseline. Moreover, the lack of skill at the short- to medium-range lead times continues to confirm what was also observed in independent tests of real-time ENSO prediction models for the period 1993–96 (Knaff and Landsea 1997).

4. Conclusions

In this study, we have utilized the simple ENSO–CLIPER statistical model as a new baseline standard for determination of "skill" in predicting the very strong 1997–98 El Niño primarily through analysis of the rmse.

For the onset of the El Niño, the following models were both useful and provided skillful forecasts:

- short range (0–1 season lead): BMRC, CCA, COLA, LIM, SSA/MEM;
- medium range (2–3 season lead): ANALOG, BMRC, COLA, CONSOL, LIM, NCEP, SSA/MEM; and
- long range (4–7 season lead): BMRC.

For the decay of the El Niño, the following models were both useful and provided skillful forecasts:

- short range: ANALOG, CCA;
- medium range: COLA, LIM; and
- long range: (none).

For the overall depiction of the 1997–98 El Niño event from onset in spring 1997, to peak in winter 1997–98, to decay in spring 1998, the following models provided skillful forecasts:

- short range: (none);
- medium range: ANALOG, CCA, COLA; and
- long range: ANALOG, CCA, COLA.

However, since no models were able to provide useful predictions at the medium and long ranges, *there were no models that provided both useful and skillful forecasts for the entirety of the 1997–98 El Niño*. This is a conclusion that remains unclear to the general meteorological and oceanographic community.

Acknowledgments. We thank Tony Barnston and Dave Enfield for fruitful discussions on the topic of forecasting ENSO and what constitutes measures of skill. Stan Goldenberg, Dennis Mayer, Huug Van den Dool, and two anonymous reviewers provided very useful comments and suggestions on earlier versions of this paper.

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., 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 temperature with a hybrid coupled ocean–atmosphere model. *J. Climate*, **6,** 1545–1566.

Barnston, G. A., Ed., 1995: *Experimental Long-Lead Forecast Bulletin.* Vol. 4, Nos. 1–4.

——, Ed., 1996: *Experimental Long-Lead Forecast Bulletin.* Vol. 5, Nos. 1–4.

——, Ed., 1997a: *Experimental Long-Lead Forecast Bulletin.* Vol. 6, Nos. 1–4.

——, 1997b: ENSO forecasts for 1998: Experimental long lead predictions. *Proc. 22d Annual Climate Diagnostics and Prediction Workshop*, Berkeley, CA, NOAA, 6–9.

——, and C. F. Ropelewski, 1992: Prediction of ENSO episodes using canonical correlation analysis. *J. Climate*, **7,** 1316–1345.

——, M. H. Glantz, and Y. He, 1999a: Predictive skill of statistical and dynamical climate models in SST forecasts during the 1997–98 El Niño episode and the 1998 La Niña onset. *Bull. Amer. Meteor. Soc.*, **80,** 217–243.

——, A. Leetmaa, V. E. Kousky, R. E. Livezey, E. A. O'Lenic, H. Van den Dool, A. J. Wagner, and D. A. Unger, 1999b: NCEP forecasts of the El Niño of 1997–98 and its U.S. impacts. *Bull. Amer. Meteor. Soc.*, **80,** 1829–1852.

Bell, G. D., and M. S. Halpert, 1998: Climate assessment for 1997. *Bull. Amer. Meteor. Soc.*, **79,** S1–S50.

DeMaria, M., 1996: Hurricane forecasting, 1919–1995. *Historical Essays on Meteorology 1919–1995,* J. R. Fleming, Ed., Amer. Meteor. Soc., 263–305.

——, M. B. Lawrence, and J. T. Kroll, 1990: An error analysis of Atlantic tropical cyclone track guidance models. *Wea. Forecasting,* **5,** 47–61.

Huschke, R. E., 1959: *Glossary of Meteorology.* Amer. Meteor. Soc., 638 pp.

Jarvinen, B. R., and C. J. Neumann, 1979: Statistical forecasts of tropical cyclone intensity. NOAA Tech. Memo. NWS NHC-10, 22 pp. [Available from Environmental Science Information Center, Environmental Data Service, NOAA, U.S. Department of Commerce, 3300 Whitehaven St. NW, Washington, DC 20235.]

Ji, M., A. Leetmaa, and V. E. Kousky, 1996: Coupled model forecasts of ENSO during the 1980s and 1990s at the National Meteorological Center. *J. Climate,* **9,** 3105–3120.

Keppenne, C. L., and M. Ghil, 1992: Adaptive spectral analysis and prediction of the Southern Oscillation Index. *J. Geophys. Res.,* **97,** 20 449–20 554.

Kerr, R. A., 1998: Models win big in forecasting El Niño. *Science,* **280,** 522–523.

Kirtman, B., Ed., 1998: *Experimental Long-Lead Forecast Bulletin*. Vol. 7, Nos. 1–2. [Available online from http://grads.iges.org/ellfb/.]

——, 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., 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.

Kousky, V. E., Ed., 1995: *Climate Diagnostics Bulletin.* Vol. 95.

——, Ed., 1996: *Climate Diagnostics Bulletin.* Vol. 96.

——, Ed., 1997: *Climate Diagnostics Bulletin.* Vol. 97.

——, Ed., 1998: *Climate Diagnostics Bulletin.* Vol. 98.

Ledley, T. S., E. T. Sundquist, S. E. Schwartz, D. K. Hall, J. D. Fellows, and T. L. Killeen, 1999: Climate change and greenhouse gases. *Eos, Trans. Amer. Geophys. Union,* **80,** 454–458.

Neumann, C. J., 1972: An alternative to the HURRAN tropical cyclone model system. NOAA Tech. Memo. NWS SR-62, 22 pp. [Available from Environmental Science Information Center, Environmental Data Service, NOAA, U.S. Department of Commerce, 3300 Whitehaven St. NW, Washington, DC 20235.]

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

Porter, H. F., 1999: *Forecast: Disaster—The Future of El Niño.* Dell Publishing, 196 pp.

Qin, J., and H. M. Van den Dool, 1996: Simple extensions of an NWP model. *Mon. Wea. Rev.,* **124,** 277–287.

Spiegel, M. R., 1988: *Statistics-Schaums Outline Series.* 2d ed. McGraw-Hill, 504 pp.

Stern, P. C., and W. E. Easterling, Eds., 1999: *Making Climate Forecasts Matter*. Panel on the Human Dimensions of Seasonal-to-Interannual Climate Variability, National Research Council, National Academy Press, 175 pp.

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.

Trenberth, K. E., 1997: The definition of El Niño. *Bull. Amer. Meteor. Soc.,* **78,** 2771–2777.

Unger, D., A. Barnston, H. Van den Dool, and V. Kousky, 1996: Consolidated forecasts of tropical Pacific SST in Niño 3.4 using two dynamical models and two statistical models. *Experimental Long-Lead Forecast Bulletin.* Vol. 5, No. 1, 50–52.

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

WMO, 1992: *International Meteorological Vocabulary.* 2d ed. Secretariat of the World Meteorological Organization, 784 pp.

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