The BMRC Coupled General Circulation Model ENSO Forecast System

GUOMIN WANG, RICHARD KLEEMAN, NEVILLE SMITH, AND FAINA TSEITKIN
Bureau of Meteorology Research Centre, Melbourne, Victoria, Australia

(Manuscript received 15 May 2000, in final form 24 August 2001)

ABSTRACT

An El Niño–Southern Oscillation (ENSO) prediction system with a coupled general circulation model and an ocean data assimilation scheme has been developed at the Australian Bureau of Meteorology Research Centre (BMRC). The coupled model consists of an R21L9 version of the BMRC climate model and a global version of the Geophysical Fluid Dynamics Laboratory modular ocean general circulation model with resolution focused in the tropical region and 25 vertical levels. A univariate statistical interpolation method, with 10-day data ingestion windows, is used to assimilate ocean temperature data and initialize the coupled model. The coupling procedure does not use any flux corrections. Hindcasts have been carried out for the period 1981–95 for each season (60 in all), for up to a lead time of 12 months. This paper will describe these initial experiments and show that the skill of sea surface temperature (SST) hindcasts in the tropical Pacific is comparable to other published coupled models. The skill of the model is strongest in the central Pacific. SST skill tends to be lower during the earlier 1990s than during 1980s in the eastern Pacific but not in the central Pacific. Since the ENSO SST anomaly in the central Pacific is the most important forcing of regional and global climate anomalies, the high SST prediction skill and its insensitivity over the hindcast period in this region in this model give grounds for optimism in the use of coupled general circulation models.

1. Introduction

The El Niño–Southern Oscillation (ENSO) is known to be the strongest climate variation on seasonal to interannual timescales. The ENSO phenomenon, as a coupled oscillation of the tropical Pacific ocean and atmosphere, has widespread and systematic influence on the ocean–atmosphere system (Trenberth et al. 1998). The embedded quasiperiodicity in this phenomenon suggests that ENSO may be predictable with a lead time of more than a year. Indeed the 10 years of the Tropical Ocean Global Atmosphere (TOGA) program introduced a hierarchy of ENSO prediction schemes, which outperform the persistence forecast in predicting gross indices of ENSO on lead times of several months to a year (Latif et al. 1998).

The ENSO forecast schemes include both statistical techniques and dynamical models. With regard to dynamical forecasting of ENSO, there are typically two prototypes of model systems: one is the intermediate model, which is designed to simulate only the anomaly part of the ocean–atmosphere system while treating the mean state of the system as prescribed. The other is the comprehensive coupled ocean–atmosphere general circulation model (CGCM), which is intended to simulate the full behavior of the system, that is, both the mean cycle and the anomaly deviation from it. The motivation for developing CGCMs is that, with increasing physical understanding of ENSO and the representation of all relevant processes in the coupled model, comprehensive dynamical models have more potential for increased ENSO forecast skill than purely statistical models and simplified dynamical models.

The first successful ENSO prediction using a CGCM was made by Latif et al. (1993). This CGCM, consisting of a tropical ocean general circulation model (OGCM) and a global low-resolution atmospheric general circulation model (AGCM), was initialized through separate spinup integrations of the ocean and atmospheric models in which only observed wind stress anomalies were needed, and, then, integrated forward in coupled mode without using flux correction. Another CGCM ENSO forecasting system was developed by Kirtman et al. (1997). This model consists of a high-resolution Pacific basin ocean model and a global AGCM. Ocean initial conditions were derived through an iterative wind stress initialization procedure. An anomaly coupling strategy was employed, with the surface zonal wind stress anomalies determined empirically from the AGCM 850-hPa zonal wind anomalies.

The above coupled models did not use any subsurface ocean measurements in initializing predictions. Results from both intermediate models (Kleeman et al. 1995) and general circulation models (Rosati et al. 1997; Ji and Leetmaa 1997; Ji et al. 1998) have demonstrated
that subsurface ocean data have a positive impact on the prediction skill. Forecasts for the period 1982–88 from the Geophysical Fluid Dynamics Laboratory (GFDL) CGCM, which consists of a high-resolution global ocean model coupled to an AGCM, indicate that data assimilation is crucial in obtaining skillful forecasts (Rosati et al. 1997). Several versions of the National Centers for Environmental Prediction (NCEP) CGCM, which consist of a high-resolution Pacific basin OGCM coupled to a modified version of an operational weather forecasting model, treating the surface heat flux in a combination of full and anomaly coupling, have been developed over the past few years (Ji and Leetmaa 1997; Ji et al. 1998; among others). All NCEP models use ocean initial conditions derived from a full ocean data assimilation system. Evaluation of ENSO prediction results, from the latest version of the model (CMP12), show that the improved coupled model, along with more accurate ocean initial conditions, gives higher predictive skill.

This paper presents results of the Australian Bureau of Meteorology Research Centre (BMRC) efforts to develop an ENSO forecast system based on an ocean analysis/assimilation component and a CGCM component. We will show that our coupled model exhibits relatively small systematic biases in SST in the tropical Pacific even though flux corrections have not been adopted. We will also show that the use of ocean data assimilation in deriving oceanic initial conditions can significantly improve ENSO prediction skill, consistent with results described above. Section 2 describes the component models used in the system. The BMRC ocean thermal data analysis and assimilation scheme is presented in section 3. Sections 4 and 5 show the hindcast forecast design and results emphasizing the coupled model’s performance in predicting sea surface and subsurface temperature anomalies in the tropical Pacific. Some sensitivities of the forecast skill of the coupled model are given in section 6 and are followed by conclusions in section 7.

2. The component models

a. Atmospheric model

The global AGCM used in the coupled system is an R21L9 (roughly 5.6° latitude × 3.2° longitude in the horizontal and nine levels in the vertical) version of the BMRC climate model. This model has been used in various climate and climate change studies. The simulation of present-day climate by a closely related version of the AGCM with observed sea surface temperature (SST) is given by Frederiksen et al. (1999). The present version of the model uses boundary layer parameterizations and vertical diffusion based on the stability-dependent bulk formulations of Louis et al. (1981). A linear sixth-order horizontal diffusion is applied on levels with pressure >75 hPa, while a linear second-order form is applied in the upper part of the atmosphere with pressure <75 hPa. Gravity wave drag is determined using the formulation of Palmer et al. (1986). Shortwave and longwave radiation schemes follow the methods, respectively, of Lacis and Hansen (1974) and Schwarzkopf and Fels (1991) with slight modifications. Pene-trative convection is treated by the mass flux scheme of Tiedtke (1989), but without inclusion of momentum effects. Shallow convection is parameterized in terms of the model’s vertical diffusion scheme following Tiedtke (1988). Cloud fractions are calculated diagnostically in three distinct layers. A simple sea ice model is included in the AGCM (Colman and McAvaney 1992). Further model details are described by Hart et al. (1988, 1990) and McAvaney and Hess (1996) and references therein.

b. Ocean model

The OGCM component of the coupled model system is a global version of the GFDL Modular Ocean Model (Pacanowski et al. 1991; Power et al. 1995). The resolution near the equator and near the surface has been enhanced to allow better representation of the seasonal variation of the mixed layer and the equatorial wave guide, which is thought to be crucial to the dynamics of ENSO. The grid spacing is 2° in the zonal direction, while meridional spacing varies from 0.5° within 7° of the equator, smoothly changing to a maximum of 5.8° near the North Pole. There are 25 vertical levels, with 12 concentrated in the top 185 m of the ocean. The bathymetry is represented by a smoothed approximation to the high-resolution dataset of Gates and Nelson (1975) with a maximum depth of 5000 m. In order to save computational cost, only Australia (combined with New Guinea), New Zealand, and Antarctica are separated from the remaining land points. No-slip boundary conditions are used along all sidewalls. The vertical diffusion is determined via a mixing scheme based on the scheme described by Chen et al. (1994). The scheme consists of two parts: a turbulent kinetic energy equation for the surface mixed layer, and a gradient Richardson number scheme for the ocean interior. Further details on the ocean model are described by Power et al. (1995). A very similar OGCM has been used in studying dynamics and thermodynamics of the Indian Ocean (Schiller et al. 1998).

c. Coupling strategy

The coupling strategy is described in Power et al. (1998). The AGCM has a time step of 1350 s (64 time steps per day), while the OGCM time step is 1800 s (48 time steps per day). A full day is used as the time step of the ice model. The coupled model proceeds as a series of 1-day coupling intervals where the atmospheric model runs separately for 1 day with fixed SSTs and sea ice extents, accumulating relevant time-mean surface fluxes. These are then used to drive the corresponding 1-
day integration of the ocean and sea ice models, following which the updated SSTs and sea ice extents are passed back to the atmosphere model for the next coupling interval.

The quantities needed to drive the ocean are wind stress (applied at both open ocean and sea ice points), freshwater flux (i.e., precipitation minus evaporation), and downward energy flux, composing net shortwave and longwave radiation, and turbulent fluxes of sensible and latent heat. All components of the energy flux are assumed to be absorbed exclusively in the top layer of the ocean with the exception of shortwave radiation, a portion (45%) of which penetrates to deeper layers according to the scheme of Paulson and Simpson (1977). In addition, surface friction velocity (proportional to the cube of the surface wind speed) is provided for use in the mixing scheme. No flux corrections are made to these interfacial forcing fields.

A preliminary coupled simulation using the CGCM (with slightly different atmospheric physics) exhibits significant interannual variability and a number of the essential ingredients of the delayed action oscillator in the tropical Pacific (Power et al. 1998).

3. The ocean analysis and assimilation system

a. The assimilation method

The ocean data assimilation method is based on the BMRC ocean analysis system (Smith et al. 1991; Smith 1995a,b; Smith and Meyers 1996). The scheme is a univariate implementation of the statistical interpolation method described by Lorenc (1986). The data are analyzed in 10-day bins ignoring the temporal distribution of the data within that bin and using the ocean model fields at the start of the 10-day period $T_{\text{model}}$ as an initial estimate (first guess). The merging of the model field and data is done according to

$$\bar{T} = T_{\text{model}} + \sum w_m (T_{\text{obs}} - T_{\text{model}}),$$

where the sum is over all observations $m$, $T_{\text{obs}}$ is the observed value, $w_m$ are the optimum interpolation weights, and $T_{\text{anal}}$ is the analyzed value. The $w_m$ depend upon the distribution of data about the target model grid point and upon various statistical assumptions used in the interpolation (forecast and observation error). This method is applied separately for each model level and only for temperature. No further adjustments are made to the model state, which does lead to some initialization shock subsequent to the assimilation step.

The statistical parameterizations are similar to those adopted in the above-mentioned papers. The model errors are assumed to be anisotropic in the equatorial region with scales of 1500 km east–west and 250 km north–south. Outside ±10° latitudes the scales are 500 km and isotropic. The noise-to-signal ratio is kept at around 1. The observation errors are assumed to be isotropic with spatial and temporal scales of 150 km and 5 days, respectively, and amplitude 1°C.

The top boundary condition of the OGCM during the assimilation integration are specified from the observed wind stress derived from the Florida State University (FSU) pseudostress (Goldenberg and O’Brien 1981) and observed SST (Reynolds and Smith 1994).

b. The data

The temperature data are a consolidated set from various real-time and delayed-mode sources and from various special projects and data archives.

Real time: Includes all subsurface data received at the Australian Bureau of Meteorology via the Global Telecommunication System (GTS) since around 1987.

Near-real time: Through the agencies of the U.S. National Oceanographic Data Center, the Global Temperature Salinity Profile (nee Pilot) Project, the Canadian Marine Environmental Data Service, and the Tropical Atmosphere-Ocean array (TAO), we have compiled further data that have been subject to some quality control. In cases of duplication we give this stream precedence over the real-time stream.

Data assembly centres: The World Ocean Circulation Experiment established a system of data assembly centres for upper ocean thermal data. As these data have been made available they have been added to the database and given precedence over the real-time and near-real-time streams. In the case of TAO, their quality controlled dataset is used.

Levitus’s World Ocean Atlas 1992: The data described in Levitus and Boyer (1994) have also been added to the database. These constitute the principal source of data for the 1980s. It was assumed they had been subject to the same levels of quality control as the scientific data assembly centres.

c. Quality control

Quality control is handled through the objective procedures described in Smith (1991). As noted in that paper and in Smith (1995a,b), the efficacy of quality control is strongly dependent on the quality of information provided and, in particular, on the accuracy of the forecast. When the “model” is climatology, we are faced with a poor estimate in extreme conditions (e.g., El Niño) and this does lead to bad decisions with respect to rejections. A dynamical model alleviates this problem to some extent since the model is better able to capture (predict) extreme events given accurate wind forcing. However this advantage can be offset by the intrusion of systematic errors into the data assimilation process in which case biased estimates may lead to erroneous decisions.

Our experience with the present system and within
the operational environment described in Smith (1995b) suggests the methodology is effective but imperfect. The European Centre for Medium-Range Weather Forecasts (ECMWF) experience with this scheme (Stockdale et al. 1998) has been similar.

4. Hindcast experiment design

a. Initial conditions

The initial conditions for the atmosphere are provided by driving the AGCM component model with observed surface forcing up to the start time of the forecast. The forcing consists of SSTs from Reynolds and Smith (1994) and sea ice extents inferred from the SST analysis. The ocean initial conditions are provided by the BMRC ocean data assimilation system as described in section 3.

In this paper results from 60 hindcasts of 1-yr integrations are presented. These experiments start from the last day of the selected months, February, May, August, and November, for the years 1981 through 1995. In the subsequent sections, forecast lead time is defined as the time lag between the initial month and the target month. For example an August forecast with lead time of 6 months is the August monthly mean from a coupled model run that is initiated on the last day of the previous February.

b. Postprocessing

As both the AGCM and OGCM have deficiencies in their respective stand-alone mode simulations and no flux corrections are used in the coupling procedure, a systematic drift from reality is expected to occur during the course of the coupled model’s forecasts. Hence the forecasts from the coupled model are bias corrected before being validated against observation. Stockdale (1997) notes that removal of the coupled model bias should make the corrected forecasts a better estimate for the expected variability. However it is likely that only the linear part of the bias can be estimated with confidence in this manner (Stockdale 1997). The bias is estimated from the mean drift of the forecasts, as a function of the initial months and the forecast lead times. The model drift is defined as the difference of model climatology and observational climatology. The model SST climatology for a given start month, for example, is obtained by averaging all predictions from the start month of all 15 years and over all lead times as follows:

$$\text{SST}_{\text{climatology}}^k(\tau) = \frac{1}{N} \sum_{j=1}^{N} \text{SST}_j^k(\tau),$$

where the summation goes from 1981 to 1995 so that $N = 15$ and $\text{SST}_j^k(\tau)$ is the $\tau$ lead time monthly mean forecasts initialized in month $k$ of year $j$. The model climatologies for other variables are calculated in the same manner. We use the same 15 years of hindcast period to define the observational climatologies.

5. Results

In this section we first focus on the SST field and look at its bias and forecast skill as a measure of performance of the coupled model, followed by an examination of subsurface temperature forecast skill. Then we show comparisons of model ENSO composites with observations to examine the model’s ability to capture the major ENSO response characteristics.

a. SST bias

The model hindcast and observed SST Niño-3 index (SST averaged over the region $5^\circ S$–$5^\circ N$ and $90^\circ W$–$150^\circ W$) climatologies and the corresponding bias for each start month are shown in Fig. 1. The overall coupled model SST Niño-3 means show a very weak seasonal cycle (Fig. 1a). The bias is dominantly positive and peaks around September and October with amplitude about 1.5°C, similar to the observed amplitude of interannual SST variability (Fig. 1b). The bias is close to the negative of the annual cycle for forecasts starting in February and May. Those starting in August and November follow the seasonal cycle reasonably well for the first 6–8 months, but then the model remains in a warm state when reality cools. This means that the eastern Pacific cold phase is not well simulated. Based on these results we opted to remove the bias as an a posteriori correction to the forecast, following Stockdale (1997).

Shown in Fig. 2 is the spatial pattern of systematic error of SST in the tropical Pacific at lead times of 3, 6, 9, and 12 months for the February start case. The largest bias occurs in the eastern Pacific and during the cold phase of the annual cycle (Figs. 2b and 2c, corresponding to forecast months of August and November,
respectively), while only a minor bias exists during the warm phase (see Figs. 2d and 2a, for forecast months of February and May, respectively). In the subtropics a general cooling in the North Pacific and warming in the South Pacific are seen to develop, which seems to be an inherent bias of our coupled model (Power et al. 1998). However these might be expected to have only a small impact on the ENSO forecasts, which are the major focus of this paper. In the following analyses the systematic errors have been removed from the original hindcasts.

b. SST forecast skill

The 1-yr evolution of the SST anomalies in the Niño-3 region for all hindcasts is shown in Fig. 3. Most predictions track the observed anomaly well for the first half year of hindcasts, although there are some notable exceptions. The forecasts initialized before the peak phase of the 1982/83 ENSO event produce colder than observed anomalies, while forecasts started just after the peak phase produce too warm anomalies. Several forecasts, such as the ones initialized in November 1986 and May 1988, show very large deviations from the observation even at short lead times. Forecasts initialized in the late 1980s and early 1990s tend to be biased toward warmer conditions and show larger case to case variability. During the 1992–95 period, the forecast SST anomalies are often of the wrong sign, although the actual amplitude is relatively small.

Two commonly used metrics of forecast skill are the anomaly correlation coefficient (ACC) and the root-

FIG. 2. Mean SST bias in the tropical Pacific for all Feb start forecasts at lead times of (a) 3, (b) 6, (c) 9, and (d) 12 months. Contour interval is 1°C and areas with bias exceeding 1°C are shaded.

FIG. 3. Time evolution of the Niño-3 SST anomalies of all hindcast cases (black lines, thick for Feb start forecasts) and the observation (gray line). Year ticks indicate Jan of each year.
mean-square (rms) error between observed and predicted anomalies. Such metrics for the Niño-3 SST anomaly are shown in Fig. 4 based on the same 60 cases used in Fig. 3. For comparison, skills for persistence forecast are also shown. It is seen that for all lead times our coupled model forecast is more skillful than persistence in terms of both ACC and rms error metrics. Using a correlation of 0.6 as a minimum value for useful forecast skill, Fig. 4 indicates that the forecast is skillful for lead times up to about 8 months.

As indicated in Fig. 1 the modeled seasonal cycle is too weak. What does the modeled interannual variability look like? We produce in Fig. 4c the observed and the model-estimated Niño-3 SST standard deviations as a measure of interannual variability. After the first few months during which the coupled model seems to undergo a fast adjustment from a likely initial imbalance, the estimated model interannual variability has been maintained at a level of about 70% of the observed magnitude. Interestingly the level of interannual variability seems related to the level of forecast skill for the Niño-3 SST as shown in section 6b.

It is difficult to quantitatively compare the skill scores from this model to those from other forecast systems because of the large diversity in the models’ formulation and treatment of coupling, as well as in data initialization methods. However a general comparison can be made against the recent results presented by Kirtman et al. (2000) in which an assessment of the current status of Niño-3 SST anomaly forecast skill from six dynamical and two statistical systems is conducted in a uniform format. Here we only compare our model results with that of the Center for Ocean–Land–Atmosphere Studies (COLA) model and NCEP model, both of which are comprehensive coupled GCMs. To be consistent with Kirtman et al. (2000) the same common 8-yr subset (1982-84, 1986-89, and 1991) is used. Our results are based on 32 cases initialized 3 months apart while COLA and NCEP results are on 96 cases initialized 1 month apart. The ACC skills from COLA, NCEP, and the present models are, respectively, 0.70, 0.79, and 0.77 at lead times of 6 months, and 0.66, 0.69, and 0.68 at lead times of 9 months. This indicates that our model performance is comparable to that of the two other CGCMs in terms of predicting a gross SST anomaly index.

The spatial distribution of the SST anomaly correlation coefficient at a lead time of 6 months for the coupled model forecasts and persistence is shown in Fig. 5. The coupled model has the largest correlation skill over the central to eastern Pacific and within 10° of the equator (Fig. 5a). Unlike many intermediate models, such as Kleeman et al. (1995) and Chen et al. (1995), the forecast skill in the far eastern Pacific is relatively poor. This region in our model has the largest SST bias (see Fig. 2). It is an area where both oceanic processes (such as horizontal advection, mixed layer entrainment, upwelling) and atmospheric processes (low-level stratus cloud formation) contribute to the SST variability. Unfortunately most general circulation models deal poorly with some of these processes, even in stand-alone mode with prescribed forcing (Schiller et al. 2000; Saji and Goswami 1997; Haskins et al. 1995). The skill of persistence, shown in Fig. 5b, is significantly lower than that achieved by the coupled model. The skillful area just around and east of the date line is a sensitive area having significant regional and global impacts on climate and weather (Hoerling and Kumar 1997; Nicholls 1989; Barsugli et al. 1999). We will further stress this point in section 6a.

c. Heat content forecast skill

The existence of independent subsurface thermal data analyses from the BMRC ocean analysis system (Smith 1995b; see section 3 of this paper) allows similar sub-

---

**Fig. 4.** Niño-3 SST anomaly (a) correlation coefficient, (b) rms error, and (c) standard deviation (STD). The solid and dashed lines, respectively, indicate results for the coupled model forecast and persistence in (a) and (b) and for the coupled model forecast and observation in (c).
surface temperature skill estimates as done for SST above. We validate the forecast anomaly of heat content, defined as the vertically integrated temperature of the upper 400 m, against the BMRC analysis. Figure 6 shows the heat content ACC for both forecast and persistence at lead time of 6 months. Interestingly the maximum skill for the heat content is mainly in the western Pacific and off the equator, a pattern reminiscent of Rossby wave incidence and Kelvin wave reflection on the western boundary, as implied by delayed action oscillator theory.

The ocean temperature forecast skill tends to be sustained longer in the subsurface than at the surface. Figure 7 shows a depth–longitude section of temperature forecast and persistence ACCs along the equator at lead times of 6 and 12 months. The skillful region at 6-month lead time is near the top layer in the central Pacific and extends into the western Pacific at a depth of about 150 m. At 12-month lead time, the surface levels have virtually no skill but in the western Pacific at a depth of 100–200 m there is still a significant area with ACC greater than 0.5. The ACC skill from persistence is less than 0.5 almost everywhere at the given lead times.

The subsurface skill in the subsurface of the western Pacific does not show up at the surface because of the well-known weak correlation between the subsurface and the surface temperature in this area. Such a weak correlation mainly results from the fact that the warm pool SST is strongly influenced by mixed layer processes driven by surface fluxes associated with atmospheric convective variability on a range of shorter timescales (Fasullo and Webster 1999). But the existence of a strong signal underneath the surface of the warm pool, with comparable forecast skill as the SST in the central Pacific, provides a promising perspective for enhancing SST skill in this region.

d. ENSO composite

The above analysis demonstrates that there is useful SST forecast skill in the central tropical Pacific in our coupled model. Observational studies have shown that SST anomalies in the Tropics can have a significant influence on global climate. It is interesting to examine the ability of the CGCM to capture regional and global responses associated with tropical Pacific SST anomalies. Due to the presence of systematic errors in the model and unpredictable internal atmospheric noise, the common skill measurements, such as the ones used in the previous section, are poor indicators of model skill. Instead, an ENSO composite analysis of various fields from both the forecasts and observations, which enhances the signal-to-noise ratio, will be presented here,
in an attempt to understand some of the attributes of the coupled model in response to ENSO.

There are three warm (1982/83, 1986/87, and 1991/92) and two cold (1984/85 and 1988/89) ENSO events during the hindcast period. We form composites for the boreal winter season (December, January, and February) from August start forecasts. The composites are thus the two-season lead forecasts initialized a few months prior to the mature phase of ENSO.

Figure 8 shows the SST anomaly composites for
warm and cold events. It indicates that the amplitude of the warm SST anomaly produced by the coupled model is about half of, but with spatial extent similar to, the observations. In contrast, the simulated cold event composite is close to the observations in its strength. Although we have only considered a small number of cases in our composites, they show that the model distribution of equatorial Pacific SST anomalies has a different skewness to that observed; that is, the model does not capture the full strength of the warm events. This is a feature commonly seen in the majority of coupled ocean–atmosphere ENSO forecast models [e.g., see Landsea and Knaff (2000) for forecasts of the 1997/98 El Niño]. In the subtropical western Pacific the anomalies are of opposite sign in the observations and almost nonexistent in the prediction. The composite for the subsurface thermal anomaly is given in Fig. 9. The overall pattern of anomalous thermal structure and reversal between warm and cold event are well reproduced by the model, but the strength is still too weak, especially for the warm composite. In addition the subsurface anomaly does not extend as deep as the analyses.

The two critical variables of the atmospheric response to the tropical SST anomaly are the zonal wind stress and precipitation. Their composites are shown in Figs. 10 and 11, respectively. Along the equator the predictions contain many features of ENSO: a reduced easterly zonal wind stress and enhanced precipitation accompanying the eastward shift of the intertropical convergence zone (ITCZ) during warm periods in the central Pacific, and the reverse during cold phases, although the amplitude and extent during the warm events is less than observed, consistent with the deficiencies in the SST anomalies discussed above. However the predictions lack amplitude away from the equator particularly in the Northern Hemisphere where the strongest response is expected to occur during the winter season (see panels a and c in Figs. 10 and 11).

Global composites for mean sea level pressure, 500-hPa height, and 200-hPa zonal wind are presented in Figs. 12–14, respectively. Over the tropical Pacific and Indian Oceans, a seesaw-type pattern in surface pressure is evident in the forecast and in the verification, but the signals diminish quickly toward the Indian Ocean in the forecast (Fig. 12). In the upper troposphere the upper branch of the Walker circulation weakens during warm periods (and strengthens during cold phases) in both the forecast and verification, but the magnitude in the forecast is only half of that in the verification (Fig. 14). This is consistent with deficiencies in the precipitation and zonal wind stress responses (Figs. 10 and 11), indicating too weak convection in the model atmosphere. The anomalous upper-level zonal wind associated with tropical convection forcing is seen to propagate away from the equator into midlatitudes of both hemispheres but the responses are weak poleward of about 30°. In the Pacific–North American regions the responses shown in the verification are strong, while in the forecast the responses are much weaker. The centers of major maxima and minima are also located differently (Figs. 12 and 13). Given the relatively weak tropical forcing in our coupled model, particularly for warm events as seen in the precipitation composites (Fig. 11), it is not surprising that the overall extratropical response patterns are weak and shifted somewhat relative to their observational counterparts, as they are largely dependent on both strength and location of the tropical forc-
The composite results indicate that our coupled model is capable of reproducing many basic features of ENSO in the tropical Pacific but is unable to capture desirable responses in midlatitudes both in strength and pattern. In addition to weak SST and convection anomalies in the coupled model, the weak midlatitude response to the tropical forcing is likely to be an inherent feature of our AGCM. A similar experiment using an earlier version of the BMRC AGCM forced with observed SST exhibits a weak response as well (McAvaney and Colman 1993). Thus improvements in atmospheric response to ENSO forcing in the AGCM are needed so as to take full advantage of the global CGCM framework. Furthermore coupled model multimember ensemble runs are also needed to enhance the signal-to-noise ratios in the extratropical regions as was shown by Stockdale et al. (1998).

6. Skill sensitivities

a. Seasonal and decadal variations

For lead times of 0–8 months the SST anomaly forecast skill from our model does not show clear seasonal dependence. After 8 months the skill decreases more rapidly for forecasts initialized at August and November (figures not shown). This implies that the “spring prediction barrier” may not be a significant factor in our model. In the NCEP coupled model (Ji et al. 1996) the seasonal skill dependence compared to the previous versions tended to become less apparent with continuous improvement in various parts of the coupled model system. The introduction of a new initialization scheme in the Zebiak and Cane model almost eliminated the spring barrier from their model (Chen et al. 1995). These results might suggest that the spring barrier to ENSO prediction may not be an intrinsic feature of the real climate system. However the present model does have significant bias and this might be masking the presence of such effects.

On the other hand, our coupled model appears to show a reduction in skill of SST anomaly forecast in the eastern Pacific from the 1980s to the early 1990s. Figure 15a compares Niño-3 SST anomaly correlation skill calculated for cases initialized during the earlier hindcast period (1981–88) with that from the later period (1989–95). In the eastern Pacific both the forecast skill and the persistence skill in the earlier period (dark lines in Fig. 15a) are better than that based on all cases (see Fig. 4). However the skill for the later period (light lines in Fig. 15a) is much worse. Such skill variation is also observed in hindcast results from other dynamical models (e.g., Chen et al. 1995; Ji et al. 1996). There is no consensus
Fig. 11. As in Fig. 8 but for precipitation anomaly composites. Contours are drawn from \( \pm 1 \) mm day\(^{-1}\) with contour interval of 1 mm day\(^{-1}\). Positive and negative values are denoted by heavy and light shading, respectively.

Fig. 12. As in Fig. 8 but for the global mean sea level pressure anomaly composites. Contours are drawn from \( \pm 0.5 \) hPa with contour interval of 1.5 hPa. Positive and negative values are denoted by heavy and light shading, respectively.
on the origins of such skill variation over decadal timescales. It might be attributed to the decadal variation of autocorrelation of tropical Pacific SST anomalies (Balmaseda et al. 1995) or to the presence of an SST decadal mode (Latif et al. 1997). Interestingly, we notice that when the anomaly correlation skill for the Niño-4 region is examined, as shown in Fig. 15b, there is little skill variation between the earlier and the later periods. Figure 5 shows that the skill in the Niño-4 region is concentrated in the eastern two-thirds of that region, with the peak skill in the central Pacific.

As has been demonstrated by many studies the skill variation is closely related to the strength of signal and noise in these indices. Figure 15c gives the observed
Fig. 15. (a) Niño-3 SST anomaly correlations of the forecast (solid lines) and the persistence (dashed lines) for separate periods of years 1981–88 (dark lines) and years 1989–95 (light lines). (b) Same as in (a) but for Niño-4 anomaly correlations. (c) The observed Niño-3 (dark lines) and Niño-4 (light lines) SST standard deviations (°C) for separate periods of years 1981–88 (solid lines) and years 1989–95 (dashed lines) displayed as in (a) and (b).

Niño-3 and Niño-4 SST standard deviations calculated and displayed in the same manner as in Figs. 15a and 15b. The Niño-3 region experiences a dramatic change with the standard deviation in the earlier period being more than twice that in the later period (dark solid and dotted lines in Fig. 15c), whereas the Niño-4 region undergoes less dramatic change (light solid and dotted lines in Fig. 15c). This result is consistent with the findings of Kirtman and Schopf (1998). Extensive prediction experiments using a simple coupled model show that during decades with high (low) amplitude of the Niño-3 SST variance the forecast skill is relatively high (low) and the delayed action oscillator (external noise forcing) is the dominant mechanism controlling the interannual SST variability (Kirtman and Schopf 1998). Notice that in the present model the eastern Pacific is also the region where there is the largest systematic error (see Fig. 2). It may not be the decadal variability that causes such skill variation but a sensitivity to the systematic errors of the model.

The relatively high skill in our model for the 1980s and 1990s in the central equatorial Pacific has important implications for predictability in terms of tropical forcing of extratropical circulation anomalies (although our present AGCM only weakly captures tropical/extratropical interactions, as discussed in section 5d). The strong forcing of the atmosphere on seasonal and interannual timescales is closely related to latent heat release from tropical large-scale deep convection (Horel and Wallace 1981). In both warm and cold ENSO phases the convection anomalies, which are a function of both SST anomaly and total SST, mainly reside in the central Pacific (Hoerling et al. 1997; or see Figs. 11a and 11c in this paper). The convection anomalies in this region can create significant impact in various regions of the world (Livezey et al. 1997; Trenberth et al. 1998; among others). It is also well established that there is a strong relationship between SST in the central Pacific and wintertime rainfall in eastern Australia (Nicholls 1989).

b. Sensitivity to ocean initial conditions

We have demonstrated in preceding sections that our coupled model has skill in predicting SST anomalies in the central tropical Pacific, comparable with that of other coupled models. It is interesting to examine how such skill is sensitive to changes to the physics of the model and to changes in the initial condition used to produce the forecasts. These aspects of sensitivity can tell us how well ENSO-related processes are represented and how much the skill might be enhanced by improving physics and initial conditions. These issues are the subject of future research and here only a brief discussion on skill sensitivity to ocean data assimilation will be given.

We compare results from “control” coupled forecasts (CN), which are initialized by an ocean-only run that utilizes ocean subsurface data and SST, and which have been described in previous sections, and forecasts initialized from an ocean-only run without assimilation of subsurface thermal data (NA). These two ocean-only runs also differ in the strength of the relaxation of the model SST to observed SST (relaxation timescale was 3 days for CN experiments and 20 days for NA experiments). Figure 16 summarizes the hindcast skill measured by the anomaly correlation of the SST Niño-3 index in the two sets of experiments. It is clear from Fig. 16 that the SST Niño-3 skill in the NA forecasts is significantly worse than that of the CN forecasts. The use of different relaxation timescales in generating ocean initial conditions for the CN and NA runs should only have a small impact on forecast skill because the wind stress (which is the same in both cases) is the main forcing that affects the subsurface thermocline where most of the skill resides (Fig. 7) and both timescales
are still relatively short compared to interannual time-scales. Hence, the increased skill in the CN run is due to the assimilation of subsurface temperature observations. This conclusion is consistent with many other studies, such as Rosati et al. (1997) and Ji and Leetmaa (1997). It is worth pointing out that the interannual variability of the Niño-3 SST anomaly index for NA forecasts is weaker than that of the CN forecasts at all lead times. For example at 9-month lead time the variability for NA is about 50% of observed while for CN it is about 70% (not shown). The ocean assimilation will have two impacts in the initial states: first, to correct the interannual temperature anomalies, and second to correct the model’s temperature mean state. Each of these can potentially contribute to the higher forecast skill and the stronger interannual variability as seen in the CN run (also see Fig. 17). From these experiments alone it is not possible to separate the relative contribution of each of these improvements due to the data assimilation. This is a topic for future research.

A issue of concern is that the skill of the forecasts initialized without subsurface data is only comparable to persistence for the first 8 months. This indicates that either the ocean model is unable to produce the correct initial ocean state for a given wind forcing or there are errors in the wind forcing. Certainly the model is not perfect and observations are incomplete. By direct insertion of in situ observational oceanographic data in the assimilation stage, at least part of these problems are overcome at initial time and the forecast skill is greatly improved.

As the CN and NA forecasts differ only in their initial ocean conditions, it is worthwhile to examine how such differences affect the coupled system. Figure 17a gives an annual mean 20°C isothermal depth along the equator in the Pacific from the BMRC analysis and from the two runs. These values are simply the averages of all hindcasts over all lead times and thus are approximately representative of model annual means. They are systematically biased toward a reduced gradient; the upper ocean in the NA run is systematic warmer than that in the CN run with a thermocline depth difference of about 10 m in the east Pacific and 20 m in the west Pacific. The mean thermocline depth difference has a significant impact on the coupled system behavior as can be seen in Fig. 17b, which shows the local regression of zonal wind stress anomaly against normalized SST anomaly along the equator. The observed regression uses FSU pseudo stress and Reynolds SST datasets and shows a maximum wind stress regression of above 0.015 N m$^{-2}$ per positive unit of standard deviation of SST in the central Pacific. The wind stress/SST regression in the CN run is roughly the same shape as the observed but 30% weaker in the central Pacific. On the other hand the regression relation for the NA run seems much weaker than both the observed and the CN results; its maximum location is also biased westward. Detailed analysis of the relationship between the thermocline and wind stress/SST association in our model is beyond the scope of this paper. But we may infer that the larger warm error in the ocean’s subsurface introduced at the initial time in the NA run makes the SST less sensitive to changes in the thermocline. One of the consequences is the lack of strong coupled activity in the key area of
the tropical central Pacific, which in turn adversely affects SST evolution and skill.

7. Conclusions

This paper presents the Australian Bureau of Meteorology Research Centre coupled general circulation model ENSO forecast system and evaluates its performance. The system includes an ocean thermal data assimilation component (Smith et al. 1991; Smith 1995a,b) and a CGCM component combining an R21L9 version of the BMRC climate AGCM and a global version of the GFDL modular OGCM (Power et al. 1998). Ocean measurements are ingested through a window of 10 days in the assimilation system to provide ocean initial conditions for the hindcasts. Hindcasts are performed every season for the period 1981–95 (60 in all). The model is validated against observed surface and subsurface temperature datasets.

The model exhibits skill comparable to, and in some cases exceeding that of, other published models. The SST skill in the eastern Pacific (Niño-3) region is sustained out to lead times of around 8 months but higher skill is concentrated in the central Pacific (eastern two-thirds of the Niño-4) region from where potential seasonal forecasts can be exploited for many regions. The model seems to show a skill variation from the period 1981–88 to the period 1989–95 in the eastern Pacific but not in the central Pacific judging from a comparison between these two periods. Further analyses suggest that the skill variation may be related to the amplitude of the observed SST variance, which is consistent with the results of Kirtman and Schopf (1998).

This study is the first to provide a systematic evaluation of subsurface skill. The results suggest that the model draws significant benefit from the ingestion of subsurface data, consistent with the results of the current BMRC operational model (from Kleeman et al. 1995) and other CGCMs (Rosati et al. 1997; Ji and Leetmaa 1997), and that this information tends to be retained longer than at the surface.

The coupled model skill variation on longer timescales as demonstrated here and in many other papers is a challenge for ENSO prediction using coupled GCMs. The fact that Niño-3 SST anomaly skill from both coupled models and persistence in the 1980s is better than that in the early 1990s may reflect a possible climatic regime shift of the tropical Pacific ocean–atmosphere system. In addition to the weakening signal, the larger systematic errors in this region in this model could be another factor affecting model skill. This may explain why SST anomaly skill in the Niño-4 region shows little sensitivity over different periods of time, because both the signal reduction and the model errors are small here.

Another practical and outstanding question in ENSO forecasting is how to estimate the reliability of a forecast made, say, by a dynamical model like the present one. An ensemble forecasting technique is often used to address this issue (Kleeman and Moore 1999). Ideally the ensemble mean should reduce the uncertainty due to internal variability in the nonlinear ocean–atmosphere system and the ensemble spread should be an indication of reliability of the forecast. For the present model, ensemble forecasts produced by randomly perturbing initial conditions fail to give such useable a priori estimates of the uncertainty of a forecast (figure not shown). Kirtman et al. (2000) reached the same conclusion in a systematic evaluation of seven ENSO forecast schemes by looking at the same relationship of the ensemble spread versus the ensemble mean derived from forecasts initialized 1 month apart. The failure suggests that either the ensemble needs to be generated in a more sophisticated way (Moore and Kleeman 1998) or other reliability measures should be utilized (Kleeman and Moore 1999).

There are planned upgrades of the coupled model forecast system focusing on improving the model resolution and physics as well as the ocean data assimilation scheme. With better definition of the initial ocean states and better depiction of various atmospheric and oceanic physics in the new coupled model system we hope that better model performance will be achieved.

Acknowledgments. We wish to thank Scott Power, Robert Colman, Jim Fraser, and Bryant McAvaney for providing some detailed information of the coupled model. We would like to thank Oscar Alves and three anonymous reviewers for their constructive comments on the earlier versions of this manuscript. This work has been assisted by an LWRDRC grant since January 1999.

REFERENCES


