A coupled method for initializing El Niño Southern Oscillation forecasts using sea surface temperature

By NOEL KEENLYSIDE[1]*, MOJIB LATIF[1], MICHAEL BOTZET[2], JOHANN JUNGCLAUS[2] and UWE SCHULZWEIDA[2], [1]*Leíbniz-Institut für Meereswissenschaften, Düsternbrooker Weg 20, D-24105 Kiel, Germany;* [2]*Max-Planck-Institut für Meteorologie, Hamburg, Germany*

ABSTRACT

A simple method for initializing coupled general circulation models (CGCMs) using only sea surface temperature (SST) data is comprehensively tested in an extended set of ensemble hindcasts with the Max-Planck-Institute (MPI) climate model, MPI-OM/ECHAM5. In the scheme, initial conditions for both atmosphere and ocean are generated by running the coupled model with SST nudged strongly to observations. Air–sea interaction provides the mechanism through which SST influences the subsurface. Comparison with observations indicates that the scheme is performing well in the tropical Pacific.

Results from a 500-yr control run show that the model's El Niño Southern Oscillation (ENSO) variability is quite realistic, in terms of strength, structure and period. The hindcasts performed were six months long, initiated four times per year, consisted of nine ensemble members, and covered the period 1969–2001. The ensemble was generated by only varying atmospheric initial conditions, which were sampled from the initialization run to capture intraseasonal variability. At six-month lead, the model is able to capture all the major ENSO extremes of the period. However, because of poor sampling of ocean initial conditions and model deficiencies, the ensemble-mean anomaly correlation skill for Niño3 SST is only 0.6 at six-month lead. None the less, the results presented here demonstrate the potential of such a simple scheme, and provide a simple method by which SST information may be better used in more complex initialization schemes.

1. Introduction

Our understanding of the El Niño Southern Oscillation phenomenon (ENSO) and our ability to predict it have improved significantly over the last decade (Latif et al., 1998) due to better observations (McPhaden et al., 1998) and to modelling studies (Delecluse et al., 1998). However, our ability to predict ENSO is still far from perfect (Barnston et al., 1999; Landsea and Knaff, 2000). Advances in ENSO prediction must come from both model improvements and better constraining initial conditions. The latter is particularly important in the ocean, because the memory for ENSO resides there (Neelin et al., 1998). Indeed, the importance of ocean subsurface data in making ENSO predictions has been demonstrated in a number of studies (e.g. Kleeman et al., 1995; Ji and Leetmaa, 1997; Rosati et al., 1997).

The focus of this paper is to explore the use of sea surface temperature (SST) data in initializing ENSO forecasts with coupled general circulation models (CGCMs). Because the relationship between ocean heat content and SST is highly non-linear, its inclusion into ocean data assimilation schemes has been very limited, and in general its influence has been restricted to the model's SST. However, SST is a critical quantity for air–sea interaction, and it is available from satellites with a high degree of accuracy, with high spatial and temporal resolution, and in near real time. These data are also relatively cheap in comparison to ocean-based observing systems. Thus, its use in initializing scheme needs to be further studied.

The initialization method implemented here is simply to run the coupled model with the SST strongly nudged to observations. Through air–sea interaction, the SST is able to initialize the ocean heat content by forcing the atmospheric winds. Such methods have been used successfully for initialization ENSO predictions with intermediate complexity models (ICMs; e.g. Zhang et al., 2003). The skill of these ICMs is competitive with the best CGCM results, and thus is a good demonstration of the high amount of information contained in SST data and the potential of such a simple scheme.

To our knowledge, there have only been two attempts to use such an initialization method in a CGCM. The first was by Rosati et al. (1996), but they had only limited success.

*Corresponding author.
e-mail: nkeenlyside@ifm-geomar.de

The second was by Oberhuber et al. (1998) for a retrospective forecast of the 1997–1998 El Niño event. The forecast evolution was highly realistic. In this paper, this coupled scheme is tested in a state-of-the-art model, MPI-OM/ECHAM5, by performing an extended set of ensemble hindcasts for the period 1969–2001. In contrast to Oberhuber et al. (1998), this model is not flux-corrected, and we insert SST and not SST anomalies (SSTa).

The choice to insert SST and not SSTa was made for several reasons. First, the model's representation of ENSO and the associated teleconnections are likely to be better represented when the model's mean state is close to the observed. Secondly, the assimilation of SSTa is not completely consistent, because there are differences between the spatial structure and amplitude of modelled and observed SSTa. Thirdly, it is our opinion that much can be gained in terms of model improvements by testing the model in such a way that model deficiencies are identified, as has been the case in this work. Finally, we assumed, somewhat naively, that the model's skill would not be strongly affected by systematic errors.

Another recent effort to better include SST data in initializing ENSO forecasts has been proposed by Tang et al. (2004). In their scheme, a statistically derived relationship between SST and subsurface temperature is used in combination with an ocean data assimilation system. They demonstrate that their scheme produces good results in ENSO hindcast experiments with a hybrid coupled model.

The initialization scheme presented here has several advantages: it is simple, it is dynamically consistent (in particular, in terms of equatorial ocean currents; e.g. Burgers et al., 2002), it is a consistent method of initializing both atmospheric and ocean components of a CGCM, and it makes use of good- and high-quality data. There are also several disadvantages: no use of subsurface data is made, atmospheric variability unrelated to SST variability is not captured, there is a heavy reliance on models (although they are clearly far from perfect), and because the scheme is simple it is also not optimal.

An outline of the paper is as follows. In Section 2 we describe the model, initialization scheme, ensemble member generation method and experiments. In Section 3, we present the results from a 500-yr climate control run of the model, with a focus on tropical Pacific variability. Results of the SST-based initialization scheme and of the ensemble member generation method are given in Sections 4 and 5, respectively. Hindcast results are presented in Section 6. In the final section, we give a discussion and conclusions.

The hindcast simulations in this study were performed as part of European Union (EU) project DEMETER (Palmer et al., 2004). In this paper only results relevant to the assessment of the initialization scheme are presented. An independent and comprehensive analysis of the hindcast performance of the model (and all other DEMETER models) is available on the DEMETER web page (http://www.ecmwf.int/research/demeter/).

2. Model description and experimental design

The Max-Planck-Institut (MPI) coupled model consists of the ECHAM5 atmospheric GCM and the MPI-OM ocean GCM, coupled with the OASIS coupler (Valcke et al., 2003). No flux correction is applied in the model. The coupled model was used, for instance, by Latif et al. (2004), to study the multidecadal variability in the North Atlantic. ECHAM5 is the latest version of the ECHAM spectral model; here, it is run at T42 resolution with 19 vertical levels. Model dynamics and physics are described in detail by Roeckner et al. (2003). MPI-OM is the latest version of the HOPE model. It is fully described by Marsland et al. (2003). The model equations are discretized on an orthogonal curvilinear C-grid. The grid used here has its poles over Greenland and Antarctica (Fig. 1). Equatorial horizontal grid refinement in the tropics gives an average longitude–latitude grid spacing of $2.8° \times 0.5°$ between $10°$S and $10°$N. There are 23 unevenly spaced levels in the vertical, with 10 over the upper 300 m. A sea-ice model, whose dynamics are based on Hibler (1979), is included in the oceanic component.

A simple analysis scheme was implemented into the model to generate the initial conditions for the ENSO hindcasts. The scheme is essentially that described by Oberhuber et al. (1998), except that here the full SSTs, as opposed to anomalies, were ingested into the model. Specifically, the coupled model is run with SSTs strongly nudged to observations: between $30°$S and $30°$N the damping constant equals 0.25 d (3.8×10^3 Wm^{-1} K^{-1}); poleward of these latitudes the damping constant decreases linearly to zero at $60°$S and $60°$N. The National Center for Environmental Prediction (NCEP) reanalysis (Kalnay et al., 1996) skin temperature (obtained from the Climate Diagnostics Center; see http://www.cdc.noaa.gov) is used for the observed SST. Post November 1981, these data correspond to an optimal interpolation analysis of SST linearly interpolated to daily values. Prior to this, the data are an EOF-based reconstruction of observed SSTs.

The basis for this scheme is that in the tropical Pacific (and throughout large regions of the tropics) on ENSO time-scales air–sea interaction is responsible for a strong relationship among anomalous SST, convection, surface wind stress and ocean heat content (e.g. McPhaden et al., 1998). Thus, the observed SSTs,
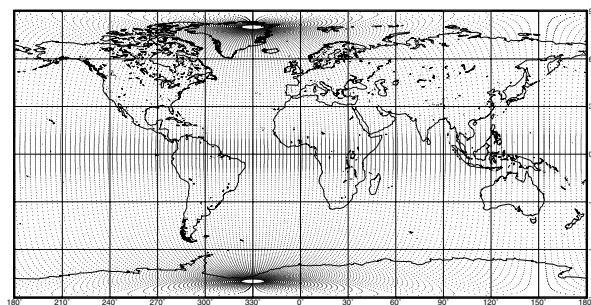

Fig 1. Ocean model grid.

through forcing the atmospheric winds, are able to (dynamically) force ocean heat content. Sensitivity to the nudging strength was not investigated, because the aim was to force the ocean as close to the observations as possible. However, this tight constraint may have some undesired side effects, as discussed below.

The hindcasts were performed in ensemble mode, nine members per hindcast. Ensemble forecasting recognizes that the system considered is non-linear and, as such, it aims to provide a probability distribution of possible outcomes based on known uncertainties. These uncertainties not only include uncertainties in initial conditions but also, for example, the influence of stochastic processes and of uncertainties in model formulations. The importance of the latter has been demonstrated by the EU project DEMETER (Palmer et al., 2004). Here, only uncertainties in initial conditions are considered.

To provide a realistic measure of uncertainties in initial conditions, initial conditions must project on to the maximal growing modes or singular vectors of the system. However, the calculation of singular vectors of a CGCM is computationally expensive, because it involves the calculation of the adjoint of the coupled system. An alternative method for calculating climatically relevant singular vectors has been proposed by Kleeman et al. (2003). Here, instead, a more pragmatic approach is taken that attempts to sample intraseasonal variability on the 30–60 d time-scale. Observational and modelling studies indicate that intraseasonal variability may have a significant impact on the evolution and may even trigger El Niño events (Latif et al., 1991; McPhaden, 1999; Lengaigne et al., 2004).

The initial conditions for the nine ensemble members only differ in the atmospheric component; ocean initial conditions are unchanged. The atmospheric initial conditions are taken from the coupled initialization run with a 5-d separation, starting 20 d before the hindcast start date and finishing 20 d after. Because only atmospheric conditions are varied, an imbalance between ocean and atmosphere components exists. The shock of this imbalance has not been investigated explicitly, but appears to be of minor importance, because no obvious systematic difference exists between the trajectories of perturbed and unperturbed hindcasts.

In this paper, we present results from a number of different experiments. The tropical variability of the model is presented as simulated in a 500-yr control run (Section 3). For this simulation, the ocean initial conditions were taken from an extended climatologically forced ocean-only simulation (Jungclaus et al., private communication). Coupled initialization runs were performed over two periods: January 1967 to December 1986 and January 1984 to December 2001. The data from these runs were used to initialize hindcasts over the period 1969–2001. No data were used from the first two years of the two initialization runs to allow for model spin up. The only reason the initialization was split into two periods was that the early set of hindcasts was not initially planned. The hindcasts were started four times per year (1 January, 1 May, 1 August and 1 November), were six months long, and consisted of nine ensemble members.

3. Coupled ENSO variability

The coupled model has been run with no flux correction for 500 yr. After an initial drift, of no longer than 50 yr, the model's climate remains stable. For the global mean 10-m ocean temperature, the initial drift is around 0.5°C, and thereafter the mean remains close to 18.1°C. The global (Atlantic) mean meridional overturning streamfunction has a maximum of 18 Sv (20 Sv) at 20°N (30°N). At 20°N, the northward poleward heat transport is 1.8 PW in the global average, well within the range of observed estimates, and 0.85 PW averaged over the Atlantic, somewhat weaker than observed estimates. The model also has a realistic simulation of sea-ice variability and extent. These and other details of the model's global climate simulation are further described by Jungclaus et al. (private communication).

Common to many coupled models (Latif et al., 2001; Davey et al., 2002), the model exhibits significant mean biases (Fig. 2). Of most significance to ENSO simulation, the model's equatorial cold tongue in the Pacific is too pronounced and extends too far westward, with SST in the region more than 4°C colder than observed. This SST cold bias is almost entirely established with in the first few years of the integration, and after 30 yr the bias remains unchanged. Extensive coupled and uncoupled sensitivity experiments have been performed to determine the nature of and to reduce the tropical SST cold bias. However, these tests have remained largely unsuccessful, with the cold bias a robust feature of the model. Associated with the SST bias, the zonal winds across the tropics are too strong, and a double intertropical convergence zone (ITCZ) is simulated in the Pacific. The model's annual cycle in the tropics (not shown) is also not well simulated: the SST in the eastern equatorial Pacific has a
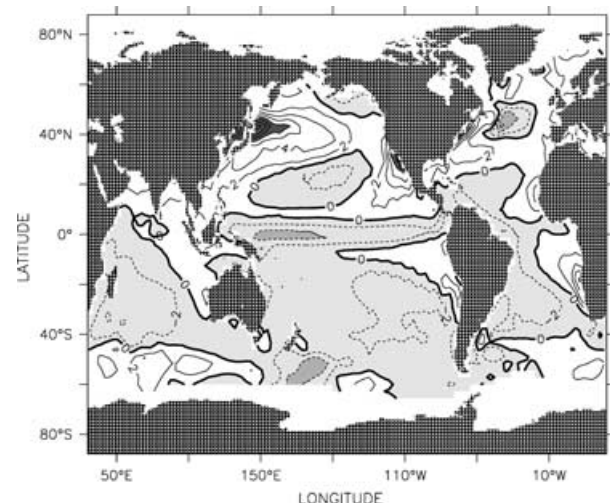

Fig 2. Mean model SST bias (modelled minus observed). Observed SST are from Reynolds and Smith (1994). Mean model SST is calculated from years 31 to 400 of the control run. The contour interval is 2°C. Negative values are shaded lightly, and positive values greater than 8°C are shaded the darkest.
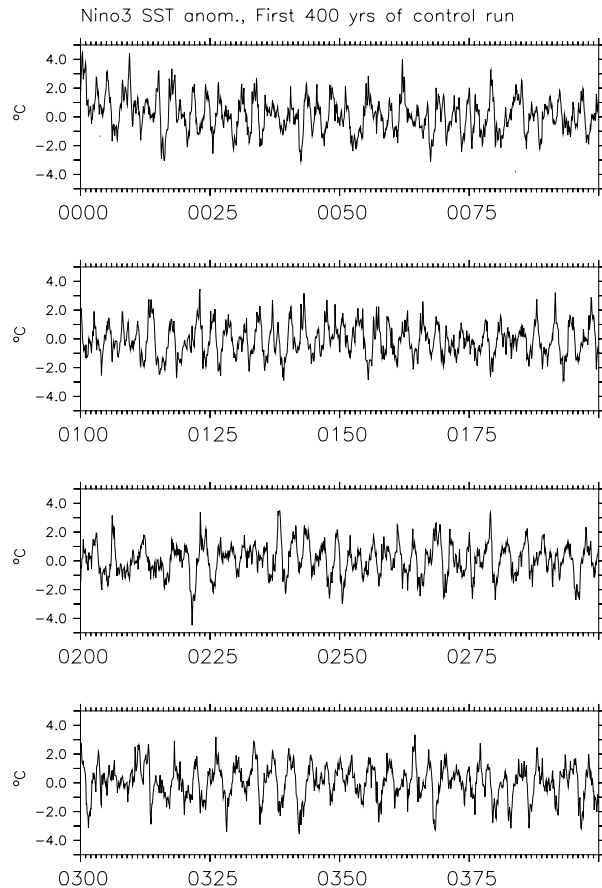
Niño3 SST anom., First 400 yrs of control run

Fig 3. Niño3 averaged SSTa from the control run.

Spectrum of Niño3 SST anomalies
Detrended HADSST (1870-1998) and model (30-400) data

Fig 4. Modelled and observed Niño3 SSTa power spectra. Modelled spectra calculated from years 31 to 400 of the control run. Observations (1870–1998) are the HADISST data set (Rayner et al., 2003) from the Hadley Centre, UK Meteorological Office. Data are detrended, and spectra are calculated using a Bartlett window of 8 yr.

semi-annual cycle with peaks (troughs) in June (May) and November (September), in contrast to the observed annual cycle. Related to the poorly simulated annual cycle, variability of eastern Pacific SSTa is also semi-annual with variability peaking in boreal winter and summer. This contrasts with the observed phase locking to the annual cycle, with strongest variability occurring in boreal winter.

Despite the deficiencies in the model's climatology, the model realistically simulates ENSO variability. The strength of tropical Pacific SST variability compares well to observations, as can be seen from the Niño3 (150°W–90°W, 5°S–5°N averaged) time series (Fig. 3). The standard deviation of the modelled Niño3 SSTa (year 51–500) is 1.1°C, whereas the observed is between 0.8°C and 1°C, depending on the period considered. However, the model simulates stronger than observed La Niña events: the simulated minimum Niño3 SSTa is close to −4°C, whereas in observations it is not more than −2°C. The spectrum of the model's Niño3 SST variability has a clear peak at 4 yr, comparing well with the observed spectrum (Fig. 4).

The structure of the ENSO-related variability can be seen from the correlation of Niño3 SSTa with SST and atmospheric fields (Fig. 5). The pattern of SST variability (Fig. 5a) is in good
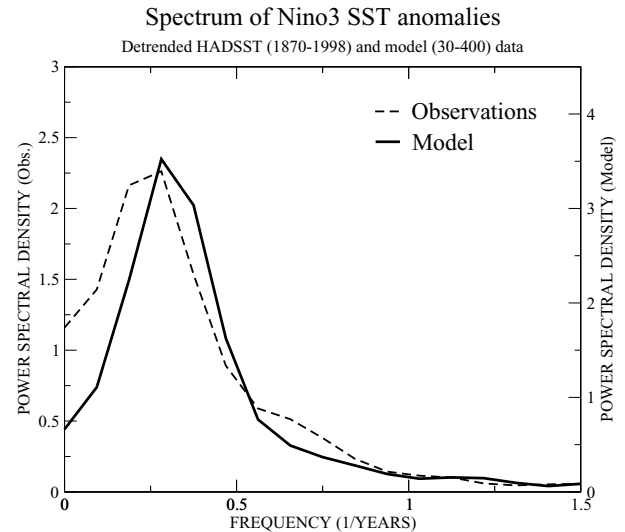
agreement with observations (Slutz et al., 1985), both in structure and strength. Notable discrepancies are that SST variability in the Pacific is too equatorially confined and extends across the warm pool. The latter is consistent with the extension of the model's equatorial SST cold bias into the warm pool. The variability in the Indian and Atlantic oceans is quite realistic, except for the positive correlations (above 0.25) in equatorial Atlantic region; a weak negative correlation is seen in observations (Slutz et al., 1985).

The SLP variability (Fig. 5c) with a dipole structure across the tropical basins is quite realistic, although the correlation in the eastern Pacific is stronger than observed (Slutz et al., 1985). Similar to the SST variability, SLP variability is somewhat equatorially confined in the Pacific. Zonal and meridional wind variabilities (Figs. 5d and e) show an in-phase weakening of the Walker circulation and a strengthening the Hadley circulation, in general agreement with observations, although correlations are again stronger than observed and obvious biases exist (e.g. the zonal wind response is too far east; Slutz et al., 1985). The precipitation pattern (Fig. 5f) is somewhat unrealistic, with negative precipitation anomalies in the central eastern Pacific and a double-banded precipitation structure across the entire Pacific. The latter is associated with the model's double ITCZ.

Hovmoeller plots of equatorial zonal wind stress, heat content and SST show that the relation among these variables is roughly consistent with observations: wind stress anomalies in the western and central Pacific force heat content anomalies, which propagate eastward and influence SST variability in the eastern Pacific (Fig. 6). However, wind stress anomalies propagate too rapidly and too far eastward, and are weaker than
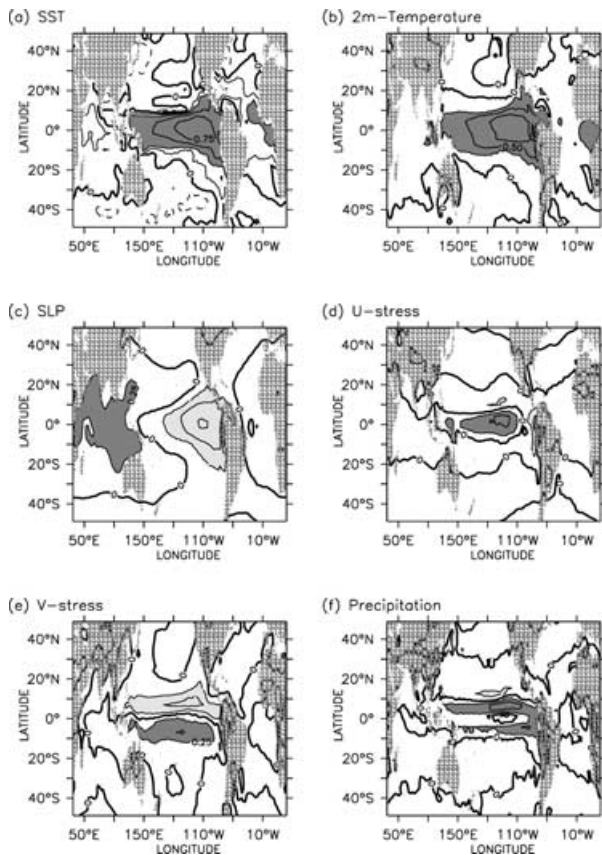
Fig 5. Correlation between modelled Niño3 SSTa and anomalies of (a) SST, (b) surface air temperature, (c) sea level pressure, (d) zonal and (e) meridional wind stress, and (f) precipitation. Model data are from years 31 to 400 of the control run. The contour interval is 0.25, and dark (light) shading indicates values larger (less) than 0.25 (−0.25). There is an extra solid (dashed) contour for 0.125 (−0.125) for Fig. 5a.

observed. Cross-correlation analysis between Niño3 SSTa and equatorial heat content (not shown) indicates the buildup of equatorial heat content anomalies in the west and their subsequent eastward propagation prior to an ENSO extreme is realistic, although slightly weaker and faster than observed. The strength of SST variability in the east is as observed, but there is a too pronounced westward propagating component in the SSTa, which extend unrealistically across the warm pool.

The model's comparitively good simulation of ENSO variability indicates that the model realistically captures air–sea interactions in the tropical Pacific. As is the case in other coupled models, a realistic simulation of the mean climate and annual cycle of the tropical Pacific does not appear to be a prerequisite for a good ENSO simulation (Latif et al., 2001).

4. Assimilation scheme

In this section, results from the analysis period 1984–2001 are compared with observations. The earlier period is not considered,

because subsurface observations are scarce. The mean zonal and meridional winds in the tropical Pacific simulated by the analysis (not shown) were compared with Florida State University (FSU; Stricherz et al., 1997) and Hellerman and Rosenstein (1983) observations and to the NCEP reanalysis (Kalnay et al., 1996), and in terms of structure and strength were within the uncertainties of these wind products. Comparison of simulated and observed surface zonal wind stress anomalies averaged over the Niño4 region, a region of central importance to ENSO, shows that the atmospheric model realistically captures the observed interannual variability (Fig. 7), with a correlation around 0.8. The model does not capture the observed higher frequency variability. This is consistent with variability associated with weather and the Madden–Julian Oscillation not being strongly related to SSTa.

The simulated mean thermocline (20° isotherm) depth, a proxy for ocean heat content, in the equatorial Pacific is in good agreement with observations (not shown). Simulated interannual variability (Fig. 8) agrees well with tropical atmosphere ocean (TAO)/TRITON observations (McPhaden et al., 1998). The correlation is close to 0.9 in the west and 0.8 in the east (Fig. 9a). The root-mean-square (RMS) error is around 10 m in the west and 18 m in the east, which is smaller than the observed standard deviation of 20° isotherm depth variability for the same period: around 16 m at 180°E, 0°N and 26 m at 125°W, 0°N (Fig. 9c). However, the strength of the simulated variability is 10–20% stronger than observed. This is particularly evident during the 1997–1998 El Niño event, with both maximum positive and negative anomalies 20–40 m larger than observed (Fig. 8).

For comparison, results are shown from an ocean model run forced by daily fluxes of momentum, fresh water and heat (as calculated with standard bulk formulas) from the NCEP reanalysis (Kalnay et al., 1996) from 1948 to 2000, using an identical set-up of the ocean model as in the coupled and hindcast simulations. Such a run could have provided an alternative and more conventional initialization method for the hindcasts performed here. Furthermore, it might be expected that these initial conditions would be more accurate than those generated by our method, because the atmospheric forcing contains the observed intraseasonal variability. However, the simulation of 20° isotherm depth anomalies along the equator is not noticeably better than those from the initialization run (Figs. 9a and c). In fact, the RMS error in the western and central Pacific is significantly higher, and also in the earlier part of the year. On the equator, 20° isotherm depth variability of the NCEP forced run is up to 40% weaker than observed in the western central Pacific, while in the east it is 10–20% too strong (Fig. 9a). This result suggests that the too strong variability simulated in the initialization run is partly atmospheric in origin. Analysis of the simulated atmospheric winds shows that atmospheric variability is overly strong in the eastern central Pacific, implying the atmospheric sensitivity to SSTa is overestimated by the model.
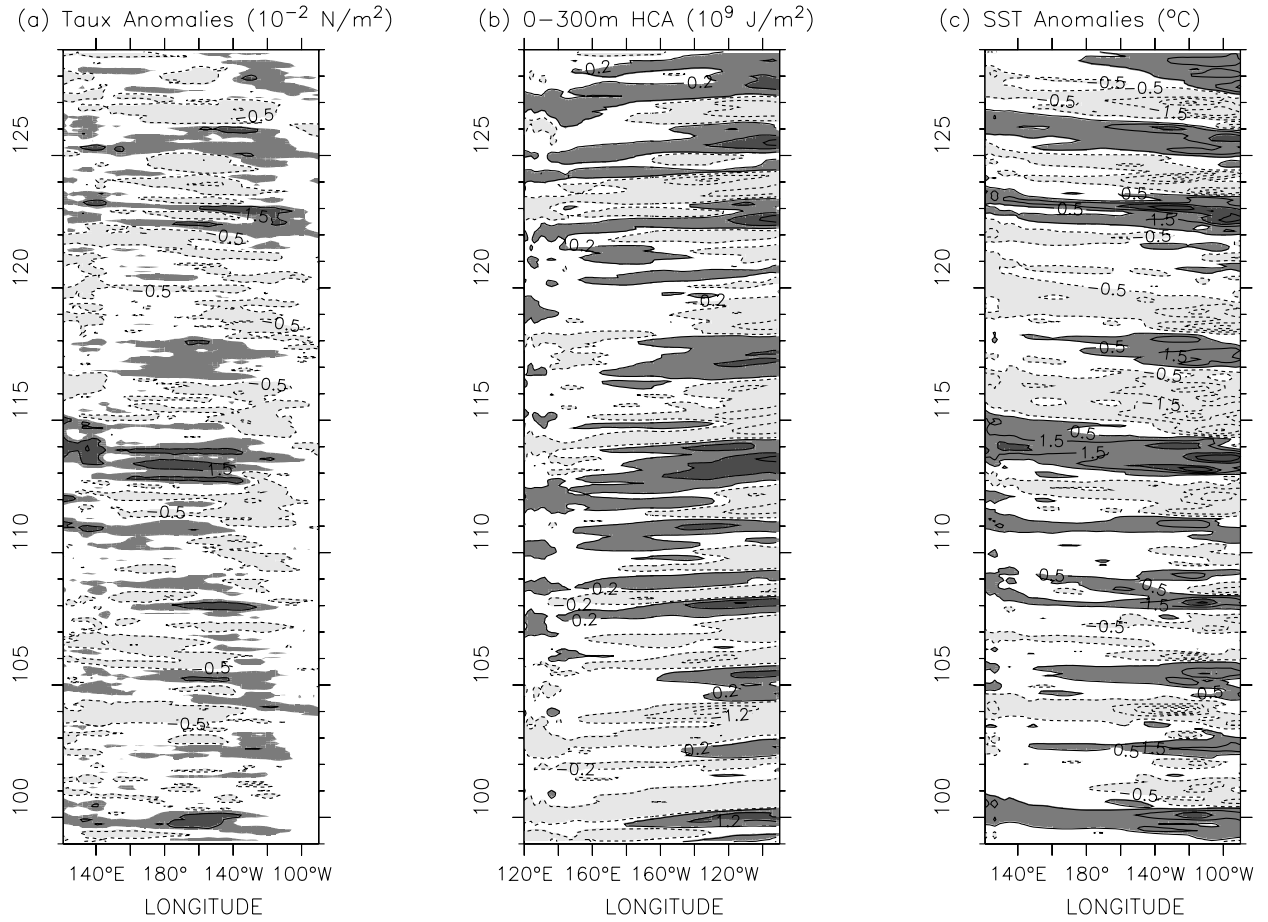
(a) Taux Anomalies (10^{-2} N/m^2)   (b) 0–300m HCA (10^9 J/m^2)   (c) SST Anomalies (°C)

LONGITUDE   LONGITUDE   LONGITUDE

Fig 6. Hovmoeller diagrams of 2°S–2°N averaged anomalies of (a) zonal wind stress, (b) 0–300 m heat content and (c) SST. A three-month running mean has been applied. The contour intervals for (a), (b) and (c) are 2, 1 and 1, respectively.

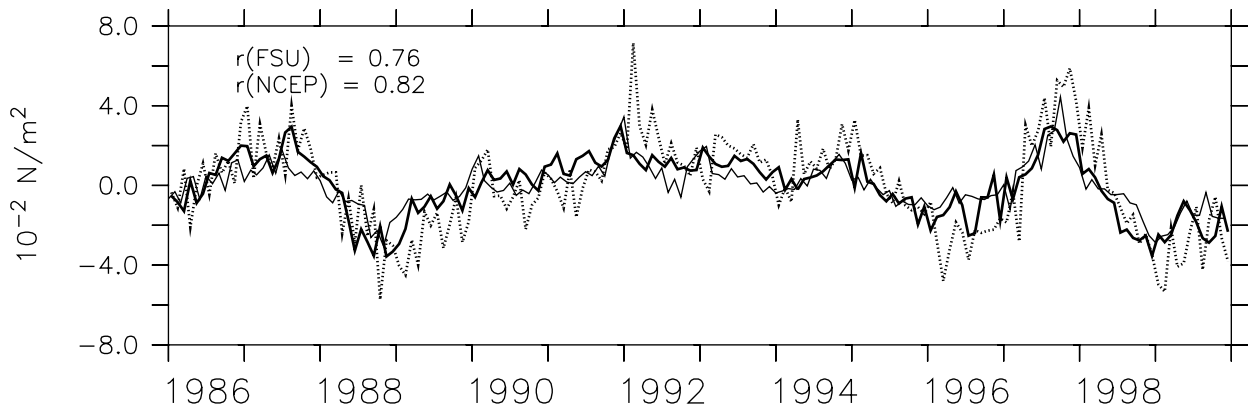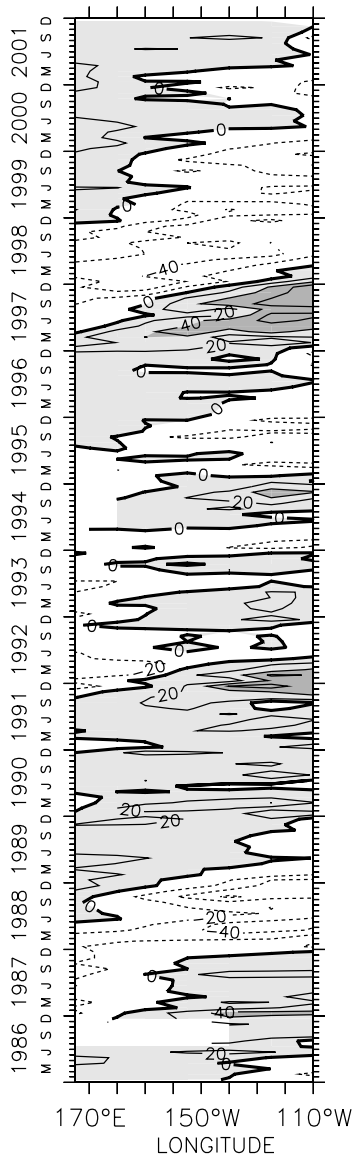## Nino4 average zonal wind stress anomalies

r(FSU) = 0.76
r(NCEP) = 0.82

Fig 7. Niño4 averaged wind stress anomalies from the analysis run (thick line), FSU (dotted line) and NCEP reanalysis (thin line). FSU wind stresses are calculated using a drag coefficient of 0.0013 and an air density of 1.2 kg m^{-3}.
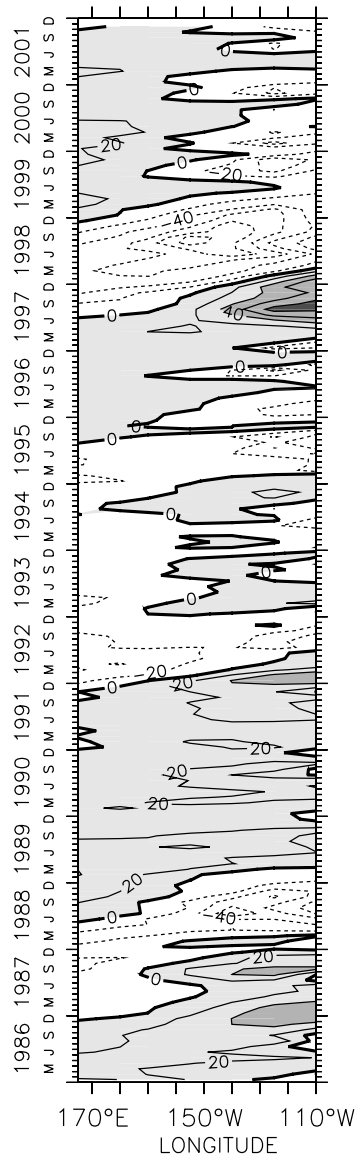
(a) TOGA/TAO

(b) Model

Fig 8. The 20° isotherm depth anomalies from the (a) analysis and (b) observations (TAO) along the equator. Anomalies are relative to the climatology of the period 1986–2001. TAO data have been filled with one point linear interpolation in longitude. Units are meters. The contour interval is 20 m, with positive values shaded.

A certain fraction of the error in the initial conditions may be attributed directly to the inability of the coupled initialization scheme to capture intraseasonal variability. To assess this contribution, the standard deviation of the low pass filtered observed and simulated 20° isotherm depth variability was calculated. The data were low pass filtered by applying a four-month running mean and then subtracting the result from the monthly mean values. The standard deviation of the observations is around 8 m in the east and 5 m in the west – slightly stronger than the McPhaden (2003) estimate, because of the slightly different filtering applied here – whereas the simulated (initialization and NCEP runs) is around 9 m in the east and 3 m in the west (Fig. 9c). If intraseasonal variability were the only source of error, then the

RMS error would be equal to $\sqrt{2}$ times the standard deviation of the intraseasonal variability, and so would be between 11 and 14 m in the east and between 4 and 7 m in the west. This is consistent with the RMS difference between the initialization run and a second run (Fig. 9c), identical to the initialization run, but numerically perturbed (details below). Thus, the schemes inability to capture intraseasonal variability accounts for somewhere between 40–80% of the total error. None the less, uncertainties due to intraseasonal variability alone cannot account for all the RMS error. Thus, clearly other processes (errors in surface forcing and ocean model physics) also contribute strongly to analysis errors. This is supported by the fact that the RMS errors of the initialization run and the NCEP
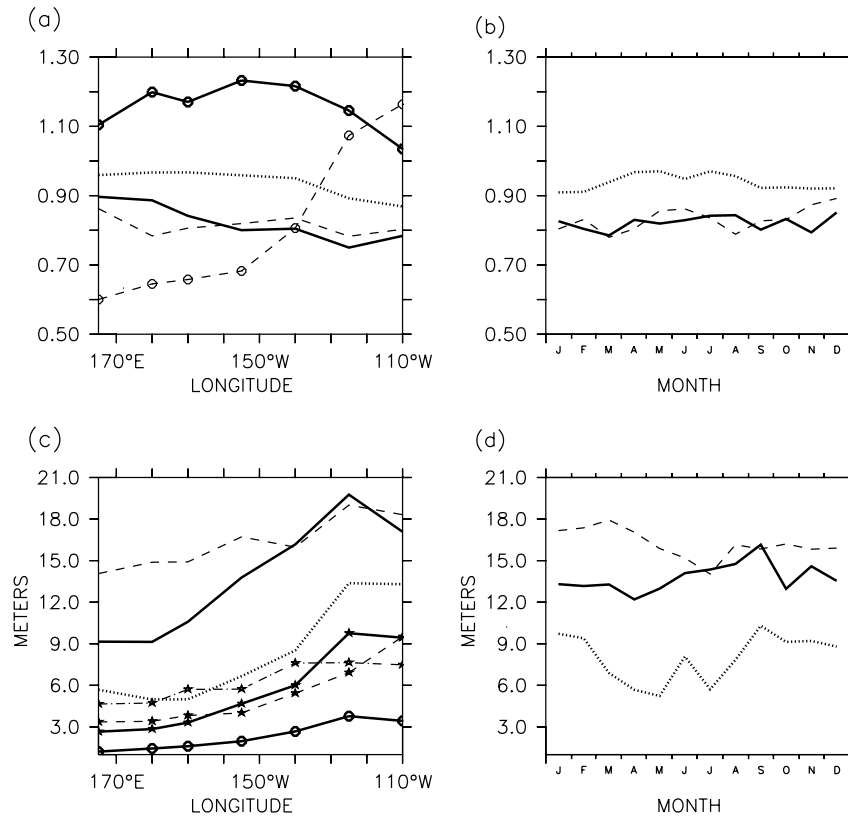
Fig 9. Statistics for 20° isotherm anomalies along the equator for the period 1986–2001 (except for the NCEP run, which is only until 2000). (a) Correlation between TAO observations and the initialization run (thick solid line), between TAO observations and the NCEP run (dashed line), and between the two perturbed initialization runs (dotted line). Also shown are the ratio of standard deviations (model to observed) between TAO observations and the initialization run (thick solid line with open circles), and between TAO observations and the NCEP run (dashed line with open circles). (b) Zonally averaged (165°E–110°W) correlations as in (a) except as a function of calendar month. (c) RMS errors between TAO observations and the initialization run (thick solid line) and between TAO observations and the NCEP run (dashed line). RMS difference between the two perturbed initialization runs (dotted line). Standard deviation of intraseasonal variability for observations (dash-dotted line with stars), initialization run (thick solid line with stars) and NCEP run (dotted line with stars). Mean ensemble spread in first hindcast month (thick solid line with open circles). (d) Zonally averaged RMS errors and difference as in (c) except as a function of calendar month.

forced run are comparable (Fig. 9c), even though the latter forcing should contain the observed atmospheric intraseasonal variability.

A second coupled initialization run was performed for the period 1984–2001, to both quantify the impact of intraseasonal atmospheric variability on ocean initial conditions and to analyse the nudging term in the SST equation (used to constrain the SST to observation) and its impact on the hindcasts. The latter is discussed in Section 6. The run apart from a recompilation of the same model with heat budget terms output should have been identical to the initialization run discussed here, and indeed the results were very similar. The differences between these runs result from atmospheric variability not forced by SST variability. Thus, the comparison of these runs provides an estimate of the upper limit of skill of the coupled initialization scheme. The correlation of equatorial 20° isotherm depth variability between the two schemes is close to 1 in the west and 0.9 in the east (Fig. 9a). The RMS difference is around 6 m in the west and 13 m

in the east (Fig. 9b), which as discussed above is consistent with the magnitude of intraseasonal variability. Clearly, unsampled intraseasonal variability is responsible for significant errors in ocean initial conditions.

Seasonally, the correlations (RMS difference) between these two runs are highest (lowest) in boreal spring and summer (Figs. 9b and d). This may seem counterintuitive, both because one may have expected that the atmosphere is only sensitive to larger SSTa, which tend to peak in boreal winter, and because of the well-known spring predictability barrier. Seasonally, the correlation skill and RMS error of the initialization run remain quite constant (Figs. 9b and d), indicating that the skill of the scheme is largely independent of the strength of SSTa. The seasonal skill of the NCEP forced run is quite similar to the initialization run, whereas the RMS error is poorer (Figs. 9b and d). Thus, there is no indication that the skill of our scheme is biased to any particular phase of the ENSO cycle, or worse than that of a more standard initialization method.
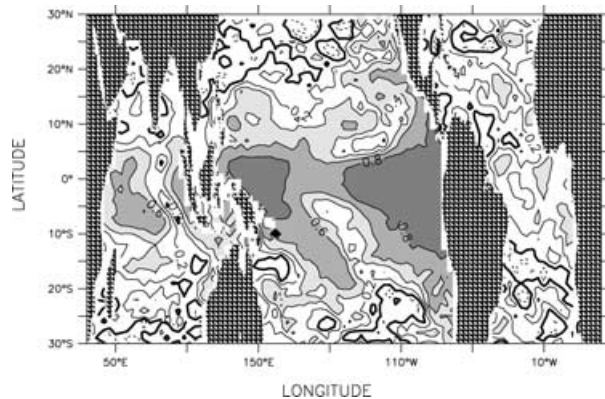
Fig 10. Correlation between modelled and observed sea level anomalies. The observations are from *TOPEX/Poseidon* obtained from the International Research Institute for Climate Prediction (IRI) data library (http://ingrid.ldgo.columbia.edu/); the analysis was for the period 10/1992 to 12/2001. The contour interval is 0.2.

A spatial assessment of simulated heat content variability is given by the comparison of simulated and observed (*TOPEX/Poseidon*) sea level height anomalies (Fig. 10). Correlations are high throughout the equatorial Pacific. In particular, in the eastern and western Pacific correlations exceed 0.8. The weaker correlations about the dateline mark the nodal point in sea level variability in the Pacific. High correlations are also found in the equatorial Indian Ocean. In the tropical Atlantic, the correlations are weak. The differences between the Pacific and the Atlantic are consistent with sea level variability in the Atlantic on these time-scales being significantly weaker. Outside the tropics, correlations are weak as expected: in the extratropics on seasonal interannual time-scales atmospheric anomalies generally drive SSTa (e.g. Peña et al., 2003). In comparison, the NCEP forced simulation achieves somewhat better skill (not shown): in the tropical Pacific, the pattern remains similar, but the regions of skill above 0.8 now extend right along the equator, and further along the South Pacific convergence zone. In the Atlantic and Indian oceans, the regions of skill above 0.6 are more extensive.

In summary, the analysis in this section indicates that the skill of our coupled scheme in capturing atmospheric and oceanic initial conditions in key ENSO regions is reasonably good, and in particular no worse than those of an NCEP forced ocean model run. Furthermore, there is no indication of a seasonal skill dependence of our scheme. However, unconstrained intraseasonal atmospheric variability results in significant RMS errors in ocean initial conditions.

5. Ensemble member generation

The purpose of ensemble forecasting is to provide a forecast of the probability density function of possible outcomes, given uncertainties in initial conditions. As such, the success of the ensemble generation method can only be assessed in the forecasts themselves. However, how well any particular ensemble generation scheme performs is model-dependent. In this section, a first assessment of the scheme used here is made by comparing ensemble spread against estimates of observed uncertainties. Here, ensemble spread refers to the standard deviation of the ensemble, and the mean ensemble spread is the average of ensemble spread over the period 1986–2001. In this section, we present the ensemble spread at the initial time and averaged over the first hindcast month.

Simulated 10-m zonal wind speed variability over the Niño4 region is in good agreement with estimates of the observed variability (see Fig. 7 for zonal wind stress). The standard deviation of Niño4 averaged 10-m zonal winds from the NCEP reanalysis (Kalnay et al., 1996) is 1.2 m s^{-1}, and that from the model is 1.4 m s^{-1}. The correlation and RMS difference between the two are 0.85 and 0.8 m s^{-1}, respectively. The mean ensemble member spread for the Niño4 averaged 10-m zonal winds at the initial start time is 1.0 m s^{-1}. Thus, this scheme is able to generate variability of appropriate magnitude (i.e. compared to the RMS error) in this key region.

As a second test, the mean ensemble member spread of 20° isotherm depth along the equator is compared to the RMS error of simulated variability and to observed intraseasonal variability from the TAO array (Fig. 9c). As ocean conditions are not perturbed at the initial time, the ensemble member spread is an average over the first hindcast month. Assuming that subsurface variability in the first month is mostly due to the perturbed initial wind field, rather than to coherent air–sea interaction, this provides an estimate of the uncertainties generated in the ocean due to the uncertainties in surface forcing. This estimate of uncertainty is significantly smaller than the magnitude of observed intraseasonal variability and the estimated RMS errors (Fig. 9c). The mean ensemble spread is around 30% of the observed intraseasonal variability, and between a quarter and a seventh the size of the RMS error. Thus, the simple scheme here is not able to adequately represent the uncertainties in the analysis. Nor is it able to fully represent the uncertainties due to intraseasonal variability that it was designed to do.

With hindsight, a more effective method to sample initial conditions would have been better to perform an ensemble of coupled initialization runs, each slightly perturbed to provide a different realization of atmospheric variability independent of the SST.

6. Hindcast results

Results from six-month-long hindcast experiments conducted for the period 1969–2001 are now presented. The assessment is restricted to SST variability only and is in terms of traditional deterministic scores (correlation and RMS error), because the primary aim of this study is to assess the model's initialization scheme. An independent assessment of the model (and of the other DEMETER models) in terms of both deterministic and

probabilistic measures is available on the DEMETER web page (http://www.ecmwf.int/research/demeter/).

The hindcasts are bias-corrected in two stages. First, to correct for a small drift in the hindcasts, which occurs over the initialization periods and is related to a slow adjustment process in the ocean analysis, the results are separately detrended over the two analysis periods. Although the constraining of model SST to observations effectively provides a flux correction to the coupled system, the subsurface temperatures drift slowly. Secondly, the mean hindcast drift as a function of lead time and calendar month is also removed, as is common practice with CGCM seasonal forecasts (Stockdale, 1997). The mean drift, which largely resembles Fig. 2, is discussed below and can be seen on the DEMETER web page.

At six-month lead, the model is able to capture all the stronger events of the period (Fig. 11). The anomaly correlation to observations is 0.6 (Fig. 12; Table 1). In particular, the model captures the 1972–1974 and 1987–1988 El Niño–La Niña events well. However, the strength and duration of the 1982–1983 event is underestimated, and the strength of the 1997 El Niño event is initially underestimated (consistent with contemporary forecasts of this event; Barnston et al., 1999), and then later overestimated. The model also performs poorly for the weaker events of the 1990s, which is a common problem among ENSO models (e.g. Kleeman et al., 1996).

A global picture of the skill at six-month lead (Fig. 13) shows the model has skill over large parts of the tropics. The correlations are, not surprisingly, higher in the central and eastern Pacific; however, RMS errors there are larger than the standard deviation of observed SST variability. Skill is seen in the north tropical Atlantic Ocean and south-western tropical Indian Ocean. Seasonal
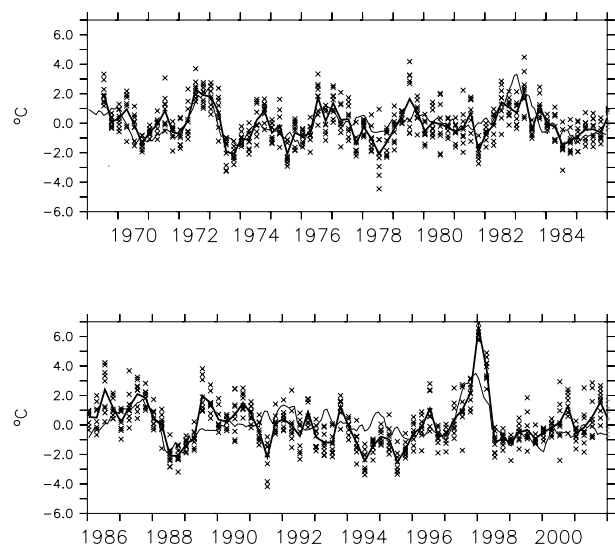

Fig 11. Niño3 SSTa for the ensemble mean hindcasts at six-month lead time (thick line) and observed SST (thin line) (Reynolds and Smith, 1994). Individual ensemble members are indicated by crosses.

to interannual variability in both these regions is strongly related to ENSO variability (Enfield and Mayer, 1997; Xie et al., 2002). The western tropical Atlantic and regions of the extratropical Pacific also show significant skill. These regions have perhaps a greater potential for skill increase with model improvements than the equatorial Pacific itself. In particular, higher model resolution may play an important role in capturing teleconnection patterns (Merkel and Latif, 2002), and ENSO behaviour itself (Guilyardi et al., 2004) better.

The ensemble mean hindcast skill for Niño3 averaged SSTa drops off linearly with lead time to a value of 0.6 at six months (Fig. 12a; Table 1). After the first month, the model performs better than persistence. At all lead times, the ensemble mean is substantially more skilful than the individual ensemble members, indicating that uncertainties in the initial conditions are being sampled to a certain degree.

For the first five months, the RMS error for Niño3 averaged SSTa grows rapidly and is larger than that of persistence (Fig. 12c; Table 1). The model error then levels off, so that at six months it is $1.1°C$, just dropping below persistence. The standard deviation of the observed SSTa for the same period is $1°C$. Thus, after five months the hindcast error is greater than the observed variability. The large RMS error is a signature of the model's over-reactive behaviour, as seen in the six-month lead Niño3 time series (Fig. 11). For example, the behaviour in both 1989 and 1998 shows severely overestimated SSTa. Possible reasons for this behaviour are discussed below.

The ratio of the standard deviation of the Niño3 SSTa model against observations shows that hindcast SST variability at all lead times is strongly overestimated (Table 1). The largest discrepancy occurs at four-month lead, with the model's variability 1.6 times larger than observed, and 1.5 times larger than that of the coupled control simulation ($1.1°C$). The mean ensemble spread, as measured by the standard deviation of the ensemble, increases rapidly with lead time, reaching a value of $0.8°C$ at six-month lead. Thus, the individual ensemble members have significantly higher variability than the ensemble mean, and correspondingly lower correlation skill and higher RMS error. The ensemble spread rapid growth is consistent with both a stochastic (atmospheric transient) mechanism for error growth within the CGCM and the fact that the model has significantly higher SST variability than observed when the ocean mean state is close to observed (see discussion below).

Although the model is able to capture all the major ENSO events of the period, the overall skill in forecasting SSTa is disappointing, especially when compared to more conventional initialization methods (i.e. with oceanic and atmospheric initial conditions generated separately by uncoupled data assimilation processes). Reasons for the model's poor skill can be grouped into three categories: (1) errors in initial conditions, including poor sampling of their uncertainties; (2) errors in model physics, such as biases or incorrectly represented modes of variability; (3) errors due to inconsistencies between model and
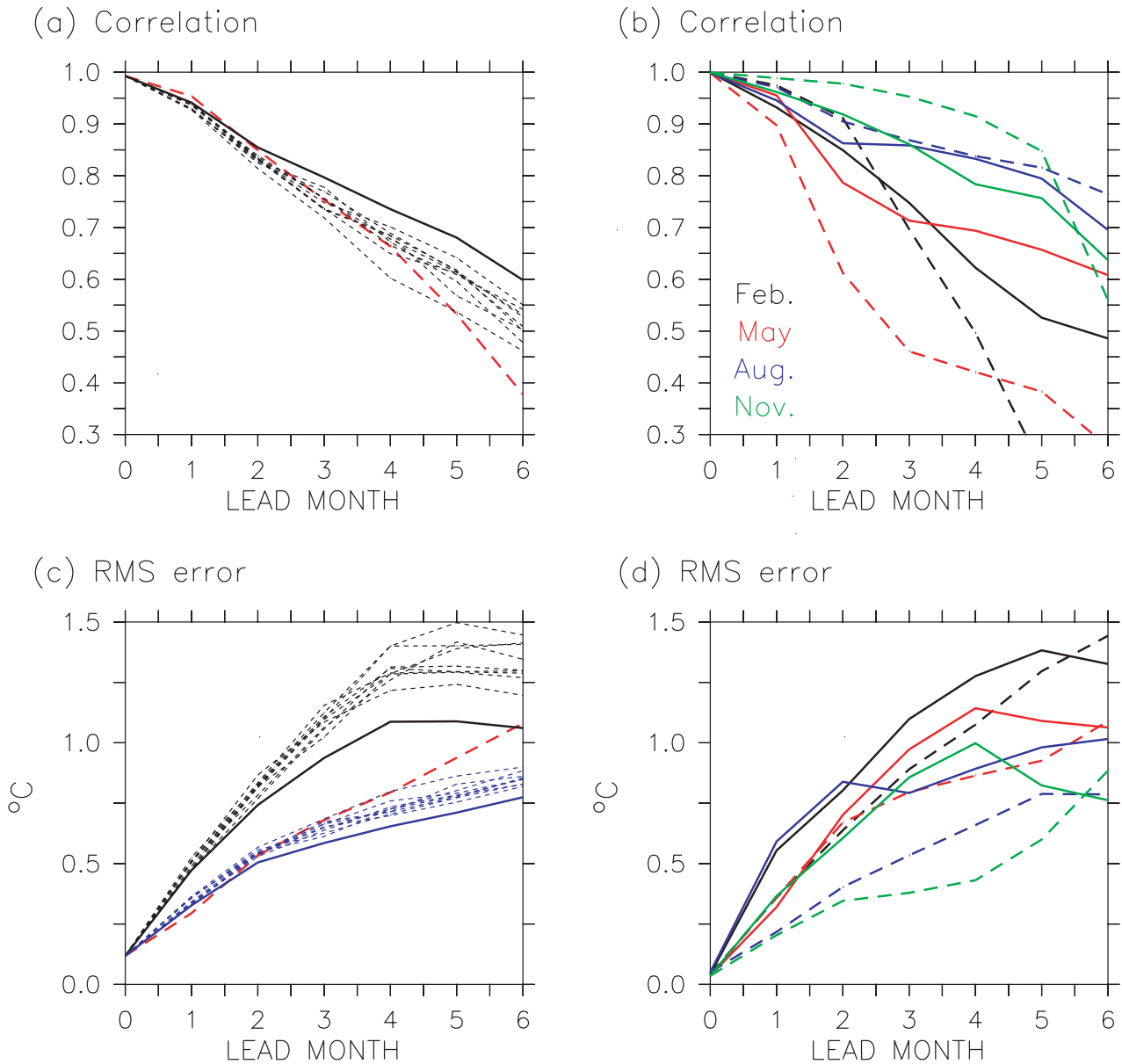
Fig 12. (a) Correlation skill for Niño3 averaged SSTa as a function of lead time for the ensemble mean (black solid line), individual ensemble members (black dotted line), and for persistence (red dashed line). (b) As in (a) except as for different start months as indicated by the colours. Solid (dashed) lines indicate the ensemble-mean (persistence) skill. (c) As in (a) except for RMS error. In addition, the RMS error after correcting for the models too strong variability is shown (blue). (d) As in (b) but for RMS error. Values are calculated for the period 1969–2001.

Table 1. Niño3 SST statistics for the ensemble mean forecast

Lead time (months)	0	1	2	3	4	5	6
Correlation	0.99	0.94	0.85	0.80	0.74	0.68	0.60
RMS error	0.12	0.47	0.74	0.94	1.10	1.10	1.10
Std. dev. model/obs.	1.00	1.29	1.42	1.55	1.63	1.53	1.35
Ens. member spread	–	0.15	0.34	0.55	0.72	0.78	0.80

ocean initial conditions, otherwise known as coupling shock. To gain some insight into the relative size of the various contributions to the model's poor skill, results are presented from the analysis of the SST budget of the initialization run, the analysis of a flux-corrected coupled simulation and the analysis of a pair of ensemble hindcast with two different ocean initial conditions. In addition, the initial conditions and hindcasts from another of the DEMETER models (here referred to as DEMETER-X) with significantly better skill than ours were also analysed.

In prelude to the analyses it is instructive to revisit the concept of bias correction. A hindcast may be considered as consisting of three components: a signal, which we wish to forecast, a systematic error and a non-systematic error (i.e. hindcast = signal + systematic error + error). It is common in seasonal forecasting to apply a seasonal mean bias correction, as was applied here. This is calculated by averaging over a large ensemble of hindcasts (started from the same calendar month) so that the signal and error average to zero. This definition assumes that the systematic error is independent of the signal or initial state. When this does not hold completely, then a mean bias correction may not be effective. In situations where the systematic error and signal are of comparable size, then separating systematic error, signal and non-systematic error is likely non-trivial, especially when they interact non-linearly.

Now, let us reconsider the quality of the ocean initial conditions. The analysis of ocean initial conditions indicated that, although uncertainties due to intraseasonal variability were not well sampled, there was little reason to believe the scheme itself was at fault, because the quality of the initial conditions was equivalent to that of a forced ocean run (Section 4). The analysis of the ocean initial conditions from DEMETER-X showed that the initial conditions from both schemes were also roughly equivalent. The sea level variability in the central Pacific and in off-equatorial regions of the western Pacific was better correlated with observations (by ≈ 0.1) than those of our scheme, but otherwise skill levels were similar; in the case of $20°$ isotherm depth variability along the equator, this model's skill was poorer in the western Pacific (by about 0.1) than ours, while in the east they were roughly equal. So from the analysis of ocean initial conditions, our initialization scheme seems not to be the main problem.

However, subtle differences in the initial conditions not captured by correlation and RMS measures may exist. In particular, as discussed earlier, the scheme, because it relies on coupled interactions, may be only able to forecast ENSO extremes that are already well developed, and so suffers from a strong spring predictability barrier. The skill of the model as a function of start month (Figs. 12b and d) is very seasonally dependent, and a spring predictability barrier is clearly present in the model, with skill dropping significantly across boreal spring. However, in comparison to DEMETER-X, the drop in skill across spring is not particularly strong. The difference between the two models is primarily that DEMETER-X is able to better maintain skill in other months. This is particularly obvious for hindcasts beginning in August and November, where our model has trouble to maintain skill above persistence. The system should in principle be highly predictable from these months. This model deficiency may be partly related to the model's poor simulation of the annual cycle of SST and SST variability, and to the mean drift, which, associated with the poorly simulated annual cycle, is strongest for August and November start dates (see below).

The model's too strong variability alone does not affect the linear correlation skill. The skill is degraded by a non-uniform increase in variability across the spectrum. Visual inspection of the Niño3 time series at six months shows that the model hindcasts have strong variability at frequencies higher than the ENSO time-scale (Fig. 11). This high frequency variability may be partly introduced by intraseasonal atmospheric variability. Thus, another reason for the model's poor skill may be no sampling of ocean initial conditions. Analysis of the initialization run (Section 4) indicates that intraseasonal variability can lead to RMS errors in $20°$ isotherm depth of up to 12 m in the eastern Pacific. From the regression relationship derived from the flux-corrected coupled simulation described below, this would be equivalent to an SST error of $1.5°C$. It would be surprising if anomalies of such strength did not impact the hindcast trajectory. To test this idea, a separate nine-member ensemble hindcast, starting in November 1987, using initial conditions from the earlier initialization period, was performed. The evolution of this ensemble and that started from the second initialization period were quite different, and by six months Niño3 SSTa differed by more than $2°C$ (Fig. 14). Because the most significant difference between these two ensembles was variations in ocean initial conditions, it can be concluded that their poor sampling contributes strongly to our model's poor skill.

The mean systematic error of the hindcasts or mean hindcast drift was analysed in detail; plots are not presented here, but are available from the DEMETER web page (http://www.ecmwf.int/research/demeter/). In the tropical Pacific, the mean hindcast drift is similar in structure to the mean coupled model bias (Section 3; Fig. 2). It is characterized by a too strong equatorial cold tongue, too strong trade winds, and by a double ITCZ. The bias patterns over the tropics are similar for all start months. In contrast to the coupled model's mean bias, which is most pronounced in the western equatorial Pacific, the hindcast drift is strongest in the eastern Pacific. The mean SST drift varies strongly with season. It is strongest for hindcasts beginning in August and November, and weakest for hindcasts starting in May. For August and November starts, the SST bias is already well developed in the first month, and by six months exceeds $6°C$ in the equatorial cold tongue. The seasonal variations in drift reflect errors in the model's annual cycle (see Section 3).

Coupled model biases and mean hindcast drift are long-standing problems (e.g. Davey et al., 2002). They result from the amplification, by coupled feedbacks, of errors in the individual model components. In our hindcasts, the drift evolution is similar for all start months. The drift begins with an initial cooling of the ocean surface in the eastern Pacific that is driven by a mismatch between ocean and atmospheric heat flux budgets: the ocean needs more heat than the atmosphere can provide. This initial cooling leads to a strengthening of the easterly trade winds and a weakening of the meridional Hadley circulation. The stronger trade winds result in stronger equatorial upwelling,
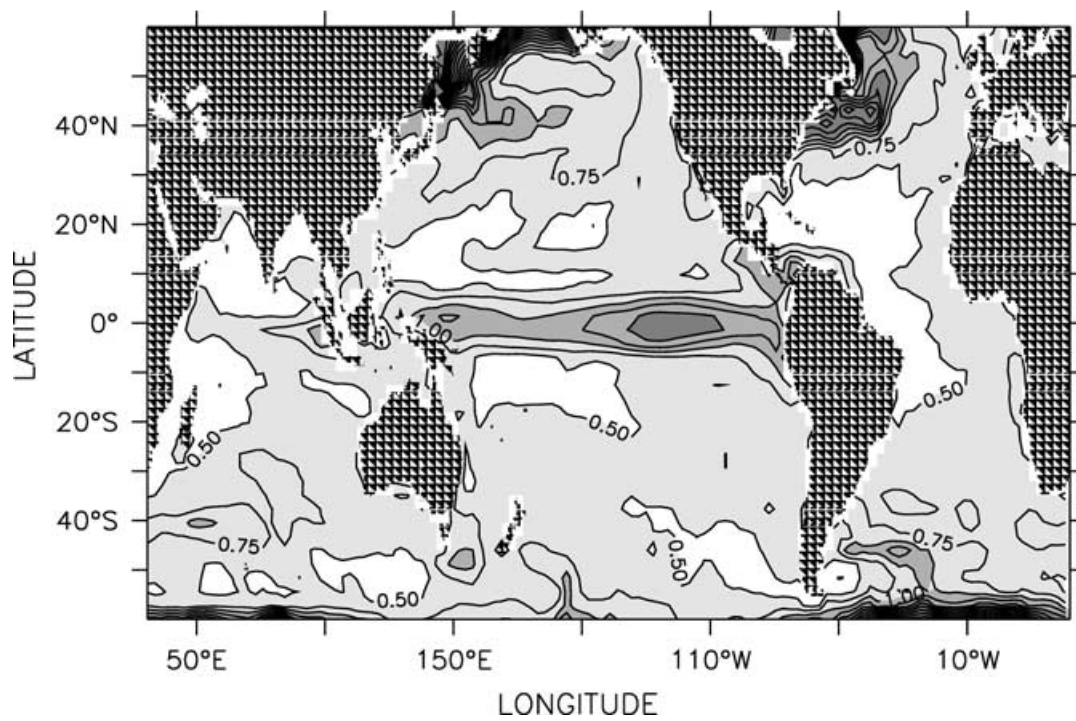
(a) Correlation
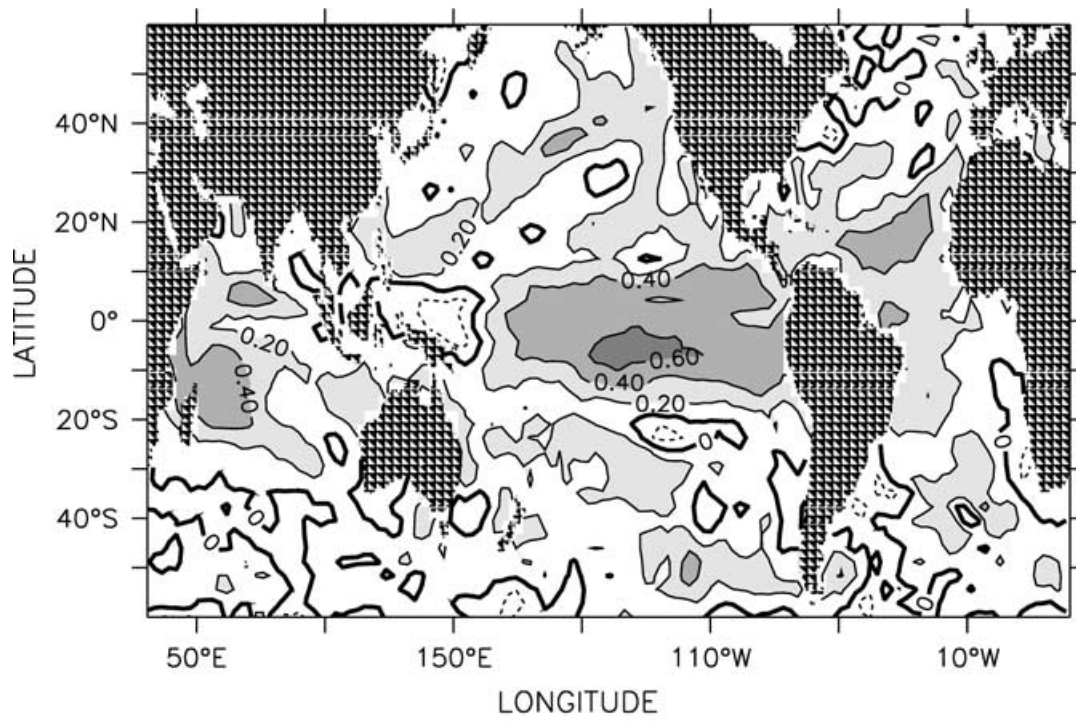
(b) RMS error (°C)

Fig 13. (a) Correlation and (b) RMS error (°C) for simulated SSTa at six-month lead. Contour interval is 0.2 (0.25) for correlation (RMS error), and values larger than 0.2 (0.5) are shaded.
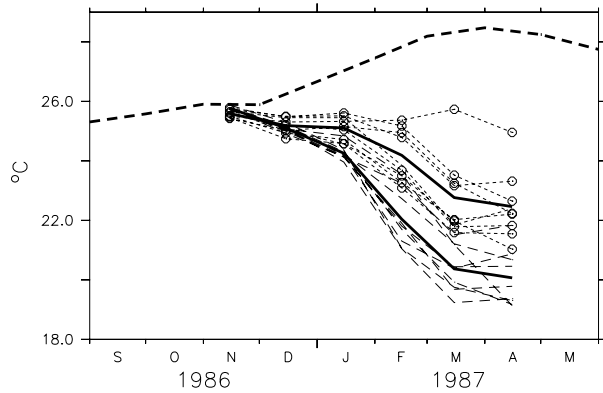
Fig 14. Niño3 SSTa for two ensembles of hindcasts started from November 1987 using ocean initial conditions from the earlier initialization period (dashed lines) and later initialization period (dotted lines with open circles). The ensemble means are shown as thick solid lines, and the observations as a thick dashed line.

Table 2. Correlation and regression values between the listed SST tendency terms and the rate of SST change. The standard deviations of the SST tendency are also given. Data are from the anomaly SST budget for the initialization period 1986–2001 averaged over the Niño3 region

	Correlation	Regression	σ
$\partial\,\mathrm{SST}/\partial t$	1.0	1.0	0.012
Nudging	0.2	1.3	0.074
$u(\partial\,\mathrm{SST}/\partial x)$	0.1	0.04	0.010
$v(\partial\,\mathrm{SST}/\partial y)$	0.1	0.24	0.027
$w(\partial\,\mathrm{SST}/\partial z)$	−0.001	−0.004	0.040
Mixing terms	−0.3	−0.60	0.023
Net surface heat flux	−0.1	−0.06	0.023
Nudging $+ w(\partial\,\mathrm{SST}/\partial z)$	0.4	1.3	0.044

and a shallowing and broadening of the eastern Pacific thermocline, which in turn leads to colder SST in the east. The feedback process continues until the ocean–atmosphere heat flux reaches equilibrium. Associated with the strengthening of the equatorial cold tongue and its extension into the western Pacific, a double ITCZ and a dry precipitation bias along the equator form. The colder SST in the eastern Pacific also allow stratus clouds to form over the equator. In summary, we find that although the initial cooling is not driven by ENSO physics, the feedbacks that amplify the initial error clearly are. Thus, it is clearly possible that signal and systematic errors will interact and degrade model skill. The situation is likely exasperated when the drift is as strong as the signal, as it is in our case.

Systematic error may not only consist of a mean drift. To investigate this possibility further, the nudging term from a second coupled initialization run for the period 1984–2001 (see Section 4) was also analysed. The nudging term by definition is the corrective heat flux required for a perfect hindcast. However, because the run discussed here is not numerically identical to the run used to initialize the hindcasts, this is not exactly true. The mean nudging term is largest in the eastern equatorial Pacific on the equator, where it has a strength of 0.6°C d^{-1}. The area coincides well with the maximum mean hindcast SST drift. In the Niño3 region, the nudging term is close to zero in boreal spring and maximum in boreal summer. This is consistent with the seasonality of the mean SST drift.

The nudging term has strong interannual variability. The analysis of the model's SSTa budget averaged over the Niño3 region shows that the nudging term is the dominant term (Table 2) driving SST variability. The next strongest term is vertical advection. Although it is weakly anticorrelated with the SST tendency term, it is strongly anticorrelated ($r = -0.87$) with the nudging term. In combination the two terms mainly determine the SST evolution (Table 2). The other advection terms are also driving terms,

but of lesser importance. The mixing terms and net surface heat flux are damping terms, as expected. Two important points can be inferred. First, because the nudging term is the strongest term in the SST budget, the action of setting it to zero when making a hindcast will produce a strong coupling shock. Further, given that the nudging term has strong interannual variability it seems that the coupling shock may be quite unpredictable. Secondly, because vertical advection is generally accepted to be a leading term in the eastern Pacific SST budget, the anticorrelation between it and the nudging term implies that part of the action of the nudging term is to counteract the growth of SSTa – in other words, to constrain the model's variability. This is an error that is likely more predictable.

Consistent with the idea that the nudging term acts to constrain SST variability, the nudging term is itself anticorrelated with SST variability ($r = -0.4$). The reason for the anticorrelation becomes obvious when a flux-corrected simulation of the coupled model is performed: the simulated SST variability of the model is now twice as strong as before. Analysis of the relevant ENSO feedbacks shows that the enhanced variability is primarily the result of a sharper and more realistic thermocline in the flux-corrected simulation. In particular, the regression relationship between Niño3 averaged anomalies of 20° isotherm depth and SST, with SST leading by two months, strengthens from 14 to 8 m °C^{-1}. The observed value is estimated to be around 14 m °C^{-1} (as calculated from the BMRC ocean analysis; Smith, 1995). The reasons for this may lie in the oceanic or atmospheric (through insufficient damping) components and are being investigated. Thus, in the initialization phase, when the model has a mean state close to observed the nudging term acts to constrain the SST variability, which would otherwise be too strong.

A relationship between the nudging term and the ensemble mean hindcast SST error should exist during the early stages of the hindcast, when linear assumptions still hold. Indeed, we find that, averaged over the first month of the hindcasts, the ensemble mean SST error is strongly anticorrelated with the nudging term:

the correlation equals −0.75 when computed for the Niño3 region. If the nudging term had been from the initialization run used to initialize the hindcasts (as opposed to a perturbed initialization run), then the strength of the correlation would likely have been higher. The relationship strongly weakens when averaged over the six months of the hindcasts. The anticorrelation is consistent with the nudging term acting as a forcing term. As may be anticipated, the ensemble hindcast SST error is strongly correlated with the ensemble mean SST: for the Niño3 region, correlations calculated for the whole hindcast period vary with lead month between 0.7 and 0.8. Based on the regression relationship calculated between hindcast SSTa and error, the hindcasts were corrected. The RMS error is significantly improved (Fig. 12c). However, because this correction is equal to a rescaling of the original signal, correlation skill is not affected.

In summary, the model achieves useful skill in predicting tropical Pacific SSTa up to six months in advance. However, the level of skill is lower than that of more conventional methods. Additional analyses of the hindcasts, initialization simulation and a flux-corrected run indicate two main sources of error. The first is that the model, when close to the observed mean state as it is at the start of the hindcasts, simulates too strong SST variability. This explains the difference between the strength of the Niño3 variability in the coupled control run and hindcasts. The second is inadequate sampling of oceanic intraseasonal variability. The initialization scheme does not constrain the atmosphere, and hence the ocean, on intraseasonal time-scales. This can lead to significant errors in subsurface initial condition and later in SST. In addition, the model's strong mean hindcast drift is driven by processes similar to the ENSO. It is thus likely that because the drift and the signal to be predicted are of similar magnitude, the two will interact and also lead to a degradation of skill. However, it is hard to assess the size of this contribution to the model's poor skill.

7. Discussion and conclusions

A simple scheme for initializing CGCMs for seasonal predictions is implemented into the MPI climate model. The model's simulation of the ENSO is relatively good, in terms of amplitude, structure and period. This is accomplished despite an overly pronounced cold tongue, which results in a strong equatorial cold bias, and a poor simulation of the annual cycle in the tropical Pacific.

The initialization scheme is simply to run the coupled model with SST strongly nudged to observations up and until the hindcast start time. Ocean–atmosphere coupling provides the mechanism for subsurface initialization. Comparisons with observations – the tropical ocean–global atmosphere (TOGA)/TAO programme and *TOPEX/Poseidon* – show that the scheme can skilfully generate subsurface conditions.

Six-month-long hindcasts are conducted with the model for the period 1969–2001. There are four hindcasts per year, once

per season, each consisting of nine ensemble members. The additional ensemble members are generated by a lead-lag initialization method in which atmospheric conditions are chosen to sample intraseasonal atmospheric variability. Ocean conditions are not varied, and thus subsurface uncertainties are not sampled. The initial spread in atmospheric initial conditions over the central Pacific is consistent with uncertainties in our atmospheric analysis. This initialization method is able to generate large ensemble spread, in contrast to that of other models (e.g. Vialard et al., 2003). However, this is no doubt partly due to the model's unrealistically strong variability.

Hindcast skill has been assessed only in terms of deterministic scores, and only using SST. At six-month lead, the model captures all the major ENSO extremes of the period. The linear correlation over the central and eastern Pacific is around 0.6, and RMS errors are of the order 1.5°C at six-month lead. However, in comparison to other GCMs, this level of skill is somewhat disappointing. For example, in the same region, the other DEMETER models have a level of skill around 0.8 at six-month lead time, and RMS error around 0.6–0.7°C.

There are three main reasons for poor model skill: (1) poor initial conditions, including sampling of their uncertainties; (2) poor model physics, such as biases and incorrectly simulated modes of variability; (3) coupling shock, i.e. the interaction between model and initial conditions, due to inconsistencies between them. Analysis of the hindcasts, initial conditions and an additional flux-corrected coupled simulation showed that the two main sources of error in our model were the models too strong variability and inadequate sampling of intraseasonal variability. The impact of the coupled model's strong biases also likely contributes to the model's poorer skill, particularly because the evolution of the ENSO and hindcast drift are driven by similar physics, and the drift and signal are of similar strength. However, it is hard to assess the real impact of hindcast drift on skill.

In the model, the relationship between eastern Pacific SST and thermocline variability is too strong when the mean state is realistic. The reasons for this are not clear and are being investigated. As the hindcasts are started from a realistic ocean state, they simulate too strong variability. Furthermore, the nudging term used to generate ocean initial conditions is large, because it not only pushes the model towards the correct trajectory, but also constrains the variability. Thus, the too strong variability results in a large coupling shock. A possible solution to these problems would have been to generate ocean initial conditions by inserting only SSTa and not the full SST as was done here.

The initialization scheme is not able to constrain intraseasonal atmospheric, and hence oceanic variability, because intraseasonal variability is not strongly related to SST variability. These intraseasonal uncertainties were demonstrated to be able to result in sizeable SST errors in the initial stages of the forecast, and to result in large forecast differences. With hindsight, it thus would have been sensible to have sampled uncertainties in ocean initial conditions. An easy way to do this would have

been to perform an ensemble of coupled initialization runs, generated for example by a slight perturbation of the atmosphere at the start of the initialization period. None the less, it remains an interesting result that intraseasonal variability can apparently have such a large impact on forecast skill.

In conclusion, we have tested a simple scheme for initializing CGCMs based only on SST data. It has been demonstrated that the scheme, although relatively simple, can produce realistic subsurface ocean initial conditions, which in turn can produce skilful ENSO hindcasts. However, the model's skill was comparatively poor, but causes of the poor skill were identified and methods to mitigate the causes were proposed. These will be investigated in future work. Finally, we hope that this work spurs further work in this area.

8. Acknowledgments

References

Barnston, A. G., Glantz, M. H. and He, Y. 1999. 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. Am. Meteorol. Soc.* **80**, 217–243.

Burgers, G., Balmaseda, M. A., Vossepoel, F. C., van Oldenborgh, G. J. and van Leeuwen, P. J. 2002. Balanced ocean-data assimilation near the equator. *J. Phys. Oceanogr.* **32**, 2509–2529.

Davey, M., Huddleston, M., Sperber, K. R., Braconnot, P., Bryan, F. and co-authors. 2002. STOIC: a study of coupled model climatology and variability in tropical ocean regions. *Climate Dyn.* **18**, 403–420.

Delecluse, P., Davey, M. K., Kitamura, Y., Philander, S. G. H., Suarez, M. and co-authors. 1998. Coupled general circulation modeling of the tropical Pacific. *J. Geophys. Res.* **103**, 14 357–14 373.

Enfield, D. B. and Mayer, D. A. 1997. Tropical Atlantic sea surface temperature variability and its relation to El Niño Southern Oscillation. *J. Geophys. Res.* **102**, 929–945.

Guilyardi, E., Gualdi, S., Slingo, J. M., Navarra, A., Delecluse, P. and co-authors. 2004. Representing El Niño in coupled ocean–atmosphere GCMs: the dominant role of the atmospheric component. *J. Climate* **17**, 4623–4629.

Hellerman, S. and Rosenstein, M. 1983. Normal monthly wind stress over the world ocean with error estimates. *J. Phys. Oceanogr.* **13**, 1093–1104.

Hibler, W. D. 1979. A dynamic thermodynamic sea-ice model. *J. Phys. Oceanogr.* **9**, 815–846.

Ji, M. and Leetmaa, A. 1997. Impact of data assimilation on ocean initialization and El Niño prediction. *Mon. Wea. Rev.* **125**, 742–753.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D. and co-authors. 1996. The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77**, 437–471.

Kleeman, R., Moore, A. M. and Smith, N. R. 1995. Assimilation of subsurface thermal data into a simple ocean model for the initialization of an intermediate tropical coupled ocean–atmosphere forecast model. *Mon. Wea. Rev.* **123**, 3103–3113.

Kleeman, R., Colman, R., Smith, N. R. and Power, S. B. 1996. A recent change in the mean state of the Pacific basin climate: observational evidence and atmospheric and oceanic responses. *J. Geophys. Res.* **101**, 20 483–20 499.

Kleeman, R., Tang, Y. and Moore, A. M. 2003. The calculation of climatically relevant singular vectors in the presence of weather noise as applied to the ENSO problem. *J. Atmos. Sci.* **60**, 2856–2868.

Landsea, C. W. and Knaff, J. A. 2000. How much skill was there in forecasting the very strong 1997–98 El Niño. *Bull. Am. Meteorol. Soc.* **81**, 2107–2119.

Latif, M., Biercamp, J. and von Storch, H. 1991. The response of a coupled ocean-atmosphere general circulation model to wind bursts. *J. Atmos. Sci.* **45**, 964–979.

Latif, M., Anderson, D., Barnett, T., Cane, M., Kleeman, R. and co-authors. 1998. A review of the predictability and prediction of ENSO. *J. Geophys. Res.* **103**, 14 375–14 393.

Latif, M., Sperber, K., Arblaster, J., Braconnot, P., Chen, D. and co-authors. 2001. ENSIP: the El Niño simulation intercomparison project. *Climate Dyn.* **18**, 255–276.

Latif, M., Roeckner, E., Botzet, M., Esch, M., Haak, H. and co-authors. 2004. Reconstructing, monitoring, and predicting multidecadal-scale changes in the North Atlantic thermohaline circulation with sea surface temperature. *J. Climate* **17**, 1605–1614.

Lengaigne, M., Guilyardi, E., Boulanger, J.-P., Menkes, C., Inness, P. and co-authors. 2004. Triggering of El Niño by westerly wind events in a coupled general circulation model. *Climate Dyn.* **23**, 601–620.

McPhaden, M. J. 1999. Genesis and evolution of the 1997–98 El Niño. *Science* **283**, 950–954.

McPhaden, M. J. 2003. Temperature balance on intraseasonal timescales in the equatorial Pacific Ocean. *J. Climate* **15**, 2632–2647.

McPhaden, M. J., Busalacchi, A. J., Cheney, R., Donguy, J. R., Gage, K. S. and co-authors. 1998. The tropical ocean–global atmosphere observing system: a decade of progress. *J. Geophys. Res.* **103**, 14 169–14 240.

Marsland, M., Haak, H., Jungclaus, J., Latif, M. and Röske, F. 2003. The Max-Planck-Institut global ocean/sea-ice model with orthogonal curvilinear coordinates. *Ocean Model.* **5**, 91–127.

Merkel, U. and Latif, M. 2002. A high-resolution AGCM study of the El Niño impact on the North Atlantic/European sector. *Geophys. Res. Lett.* **29**, 10.1029–10.1032.

Neelin, J. D., Battisti, D. S., Hirst, A. C., Jin, F. F., Wakata, Y. and co-authors. 1998. ENSO theory. *J. Geophys. Res.* **103**, 14 261–14 290.

Oberhuber, J. M., Roeckner, E., Christopher, M. Esch, M. and Latif, M. 1998. Predicting the 1997 El Niño event with a global climate model. *Geophys. Res. Lett.* **25**, 2273–2276.

Palmer, T., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M. and co-authors. 2004. Development of a European multi-model ensemble system for seasonal to interannual prediction (DEMETER). *Bull. Am. Meteorol. Soc.* **85** 853–872.

Peña, M., Kalnay, E. and Cai, M. 2003. Statistics of locally coupled ocean and atmosphere intraseasonal anomalies in reanalysis and AMIP data. *Nonlinear Proc. Geoph.* **10**, 245–251.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V. and co-authors. 2003. Global analyses of sea surface temperature, sea ice and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* **108**, doi:10.1029/2002JD002670.

Reynolds, R. W. and Smith, T. M. 1994. Improved global sea surface temperature analyses using optimum interpolation. *J. Climate* **7**, 929–948.

Roeckner, E., Bäuml, G., Bonaventura, L., Brokopf, R., Esch, M. and co-authors. 2003. The atmospheric general circulation model ECHAM5. Part I: model description. Report 349, Max-Planck-Institut für Meteorologie, Hamburg, Germany.

Rosati, A., Gudgel, R. and Miyakoda, K. 1996. Global ocean data assimilation system. In: *Modern Approaches to Data Assimilation in Ocean Modeling* (ed. P. Malanotte-Rizzoli), Elsevier, Amsterdam, 181–203.

Rosati, A., Miyakoda, K., and Gudgel, R. 1997. The impact of ocean initial conditions on ENSO forecasting with a coupled model. *Mon. Wea. Rev.* **125**, 754–772.

Slutz, R. J., Lubker, S. J., Hiscox, J. D., Woodruff, S. D., Jenne, R. L. and co-authors. 1985. Comprehensive Ocean–Atmosphere Data Set; Release 1, Climate Research Program, Boulder, CO.

Smith, N. R. 1995. An improved system for tropical ocean subsurface temperature analyses. *J. Atmos. Oceanic. Technol.* **12**, 850–870

Stockdale, T. N. 1997. Coupled ocean–atmosphere forecasts in the presence of climate drift. *Mon. Wea. Rev.* **125**, 809–818.

Stricherz, J. N., Legler, D. M. and O'Brien, J. J. 1997. TOGA Pseudo-stress Atlas 1985–1994, Volume II: Pacific Ocean, Florida State University.

Tang, Y., Kleeman, R. and Moore, A. 2004. SST assimilation experiments in a tropical Pacific ocean model. *J. Phys. Oceanogr.* **34** 623–642.

Valcke, S., Caubel, A., Declat, D., and Terray, L. 2003. OASIS3 ocean atmosphere sea-ice soil. Users guide. Prisim project report 2, CERFACS, Toulouse, France.

Vialard, J., Vitart, F., Balmaseda, M., Stockdale, T. N. and Anderson, D. L. T. 2003. An ensemble generation method for seasonal forecasting with an ocean–atmosphere coupled model. ECMWF Tech. Memo., No. 417 ECMWF, Reading, Berkshire, UK.

Xie, S. P., Annamalai, H., Schott, F. A. and McCreary, J. P. 2002. Structure and mechanisms of South Indian Ocean climate variability. *J. Climate* **15**, 864–878.

Zhang, R.-H., Zebiak, S. E., Kleeman, R., and Keenlyside, N. 2003. An improved intermediate coupled model for El Niño simulation and prediction. *Geophys. Res. Lett.* **30**, 2012 (doi:10.1029/2003GL018010).