NATIONAL RESEARCH
FLAGSHIPS

Water for a Healthy Country

An assessment of the role and potential utility of dynamical seasonal forecasts for Australia

I. N. Smith, S. G. Wilson and P. McIntosh

CSIRO Marine and Atmospheric Research

June 2005

CSIRO

An assessment of the role and potential utility of dynamical seasonal forecasts for Australia

I. N. Smith, S. G. Wilson and P. McIntosh

June 2005

CSIRO Marine and Atmospheric Research
Private Bag No. 1,
Aspendale, Victoria, Australia 3195
T: (03) 9239 4400
F: (03) 9239 4444

Head office:
GPO Box 1538 Hobart, Tasmania, Australia 7001

Smith, I. N. (Ian Noble), 1953- .
An assessment of the role and potential utility of
dynamical seasonal forecasts for Australia.

ISBN 1 921061 86 3.

1. Dynamic climatology.
2. Climatic changes - Australia.
3. Australia - Climate - Forecasting.
   I. CSIRO. Marine and Atmospheric Research.
   II. Title.

551.6594

Water for a Healthy Country is one of six National Research Flagships established by CSIRO in 2003 as part of the National Research Flagship Initiative. Flagships are partnerships of leading Australian scientists, research institutions, commercial companies and selected international partners. Their scale, long time-frames and clear focus on delivery and adoption of research outputs are designed to maximise their impact in key areas of economic and community need. Flagships address six major national challenges; health, energy, light metals, oceans, food and water.

The Water for a Healthy Country Flagship is a research partnership between CSIRO, state and Australian governments, private and public industry and other research providers. The Flagship aims to achieve a tenfold increase in the economic, social and environmental benefits from water by 2025.

For more information about Water for a Healthy Country Flagship visit or the National Research Flagship Initiative at www.csiro.au.

Contents

Acknowledgments

Much of the work described in this report was initiated during a 3-day workshop held at CSIRO Atmospheric Research, March 23-25, 2004. The participants included the authors, Deborah Abbs, Redha Beddek, Bob Cechet, Stephen Charles, Roger Jones, Yun Li and Brian Ryan. The authors acknowledge the software support provided by Harvey Davies and assistance with graphics provided by Leanne Webb. Additional funding for this research has been provided under the Indian Ocean Climate Initiative (IOCI). The authors also acknowledge the role of Barrie Hunt (who initiated and led the development of the coupled model) and Hal Gordon (who actually performed the bulk of the development). Mark Collier kindly provided comments on a version of this report and Julie Siedses kindly assisted with the formatting.

Executive summary

The potential utility of dynamical seasonal climate forecasts is addressed by analyzing the hindcast results of the CSIRO COCA2 model spanning the period 1980 to 2003. The analysis includes an estimate of the skill of the model at predicting large scale climate variability that arises due to El Nino Southern Oscillation (ENSO) events. It also includes as assessment of the utility of predicted variables for several key Australian regions.

It is found that

- The hindcast results for the NINO34 sea surface temperature index are more skilful than persistence during the first half of the year but are similar (although still skillful) to persistence once the predictability barrier has been passed. The level of skill at a 6 month lead-time also appears to be equal to, or greater than, that achieved by other models.

- Southern Oscillation Index (SOI) forecasts are most skillful for spring and summer and can be achieved with lead times up to 6 months.

- It is possible to provide, with longer lead times than currently provided, estimates of shifts in probabilities of above/below median rainfall for Australia using either the NINO34 or SOI forecasts.

- Forecasts of target variables over the key regions indicate relatively low potential skill in the southwest of the continent but higher potential skill in the north-east of the continent where ENSO has its strongest impacts.

- The most useful forecasts appear to be those started on July 1 start dates.

- There appear to be potentially useful statistical links between regional target variables and forecasts of climate indices such as regional rainfall, SOI or region mean sea level pressure (MSLP).

- This assessment may underestimate potential skill since it does not deal with ensemble results. A more thorough assessment can be performed by accessing the large hindcast data base created by the European Centre for Medium-Range Weather Forecasts (ECMWF) under the DEMETER project.

- This assessment has mainly focused on the use of linear correlations but there are a number of alternative measures which appear to be more relevant to end-users (c.f. Hartmannn et al.,2002) These can be employed in future assessments.

1 Introduction

In this study we consider the skill, and potential for application, of a dynamical seasonal forecast model by analyzing the results of retrospective forecasts, or hindcasts, made for the period 1980 to 2003. This first section of the report provides some background and details of the model, while Section 2 deals with a general assessment of the skill of the model at predicting ENSO events and considers methods for making use of this skill in applications-type studies. Section 3 deals explicitly with comparing predicted and observed climatic variables at three key locations: Southwest Western Australia (SWWA), the Murray-Darling Catchment basin (MDB) and the Burdekin region of north Queensland

1.1 Background

Australia is currently served by two operational forecast schemes, both of which are statistical schemes. The first is the 3-month seasonal outlook service provided by the National Centre of the Australian Bureau of Meteorology (see: http://www.bom.gov.au/climate/ahead). This service commenced in 1989, has evolved over time, and now provides an estimate of shifts in probabilities of regions receiving above (or below) median rainfall (and temperatures) for the next 3-month outlook period. These probabilities are based on the behavior of 2 predictors comprising the leading modes of Pacific and Indian Ocean sea surface temperature (SST) variability (Drosdowsky and Chambers, 2001). A feature of this scheme is that it represents the (independent) influences of 2 oceans on Australian rainfall. The second scheme provides similar information for 3-month rainfall probabilities and is provided by the Queensland Department of Primary Industry (QDPI) (see: www.longpaddock.qld.gov.au/SeasonalClimateOutlook/RainfallProbability/index.html). In this case the probabilities are based on the behavior of the monthly SOI (Stone et al., 1996) and mainly reflect a Pacific-only influence. These schemes apply for 3-month periods and they only provide a lead time (i.e., “advance warning”) of 2 to 3 weeks before the forecast period.

It is not often realized that the skill of any seasonal forecast system cannot be high due to the relatively large amount of inherently unpredictable noise in the climate system. This means that the predictive information of any scheme can only be manifested as moderately large correlation coefficients or as only moderate shifts in the probabilities of above/below median rainfall. Despite these limitations, the moderate predictability that does exist presents an opportunity to provide information to end-users. This information can have an economic impact where an industry or enterprise is affected by seasonal climate variability.

Two major issues confront those who develop and issue seasonal forecasts. The first is to try and maximize the level of skill associated with their forecast schemes and the second is to convey the forecast information in a manner which provides end-users with the best opportunity to benefit from this skill. This can involve convincing end-users that, in some situations, seasonal predictions can assist decision making (c.f. Hartmann et al., 2002) or converting the model outputs into climate variables at relevant time and space scales (c.f. Wood and Maurer, 2002).

Methods for improving the performance of seasonal forecast models is a large topic and the reader is referred to Palmer et al. (2004) and references therein for some details. However, an important aspect of forecast skill is the use of ensembles, particularly as it has been demonstrated that multi-model ensembles produce more reliable probability forecasts than do single-model ensembles (Palmer et al., 2004). Within this report we deal with an assessment of a single member ensemble from a single model. As such, the results are likely to represent an underestimate of the potential skill.

With regard to utility, Hartmann et al., (2002) (HEA) for example, have evaluated seasonal climate outlooks issued by the U.S. National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Centre (CPC) from a user perspective. These outlooks are very similar in content to those issued by the Australian National Climate Centre and the problems associated with interpretation and misinterpretation of these by end-users appears universal. HEA also make the important point that the nature (e.g. categorical, probabilistic etc.), timing (e.g. lead time, time of year etc.) and location of seasonal forecasts required by water resource managers can be diverse and indicate that successful uptake of the information requires close dialog between the providers and users. HEA focused on what they regarded as useful measures of forecast skill including the probability of detection (POD), false alarm rate (FAR), the Brier score, the ranked probability score and conditional distribution diagrams. They noted that winter precipitation outlooks made during autumn and winter were better than climatology forecasts but that this skill was more relevant to water managers of the lower Colorado River basin compared to the upper basin. They also found that there was a difference in the utility of the forecasts for ranchers operating within the southwest United States- in particular those with grasslands that peaked in winter appeared to have a competitive advantage over those whose grassland peak in summer. Finally, they noted that only wildland fire managers in those regions with an important winter fire risk (such as southern California) could benefit from any forecast skill.

It is apparent that most published assessments of skill make comparisons between different systems difficult because:

- There appears to be no accepted methods for describing skill although basic measures such as correlation coefficients and RMS errors are often quoted
- Different systems have been assessed over different time periods, some of which are easier to predict than others
- The number of years involved in each assessment can differ from study to study
- The effective sample size can vary even if the number of years is the same
- The sample size associated with many hindcasts is often relatively small and the significance associated with any measure of skill is not always assessed
- Skill is known to strongly dependent on both lead time and time of year. Quite often the skill measures from hindcasts started at different times of the year are combined into a single measure which can be misleading

"Real", as opposed to "raw" skill refers to the improvement in skill over and above that which could be achieved using a simple forecast strategy such as persistence. This is an important issue but is sometimes neglected in some studies.

Depending on the enterprise and the risk exposure, forecast skill does not necessarily translate into economic value. Richardson (2000) notes that (with regard to weather forecasts) "..the relationship between skill and value is complex". For example, it can be shown that unless forecast skill can exceed some critical threshold value, a hypothetical wheat grower can be better off in the long-term by following a fixed "no-skill" strategy of applying fertilizer at the beginning of each growing season – regardless of the forecast. While this can lead to a negative return when the season turns out poor, it is more than compensated by the relatively large gains when the season turns out to be good or even average. These ideas of critical skill thresholds may not often be appreciated by seasonal forecasters assume that all skill can potentially translate into economic value. This issue is beyond the scope of the present study but will be addressed in the near future.

It is clearly outside the scope of this report to address all the above issues. Consequently, it focuses on the potential for improving skill utilizing dynamical (as opposed to statistical) seasonal forecast methods. One of the main advantages of these methods is that they offer the opportunity to increase the lead time over the current statistical schemes. In the case of water resources, The Water for a Healthy Country Flagship program is focused on maximizing the benefits of water utilization across the continent. This report deals with the potential for providing useful information to water resource managers in the south-west of Western Australia, and in the Murray Darling Catchment Basin. It also focuses on crop growth in the Burdekin region of north Queensland. In this region skillful rainfall forecasts could affect on-farm managements decisions aimed at maximizing yield and optimizing the use of available soil moisture.

1.2 ENSO forecast systems

The basis of most current seasonal prediction schemes is the fact that El Nino Southern Oscillation (ENSO) events represent the largest source of interannual climate variability beyond the seasonal cycle. There are numerous schemes providing predictions of ENSO events at long lead times (i.e. 3 to 12 months ahead in time). Many of these schemes are experimental and there has been some debate about whether the relatively complex dynamical schemes (usually coupled ocean-atmosphere models) provide more useful skill than do relatively simple statistical schemes. Evaluations of many of these schemes over the past two decades have provided indications of absolute skill and the potential for improvement over time but, until recently, it has been unclear what the limit to predictability is and how much further improvement is possible. Furthermore, while evaluations of skill have tended to focus on the ability of schemes to correctly predict sea surface temperature (SST) indices (such as NINO3 or NINO34) there have been few studies dealing with methods for converting seasonal predictions into useful information for the benefit of potential end-users.

Barnston et al. (1994) evaluated the performance (at predicting SST indices) of 5 prediction schemes over the period 1982-93 and noted that both statistical and dynamical schemes achieved a moderate level of skill at 6-month lead times. In a follow-up study, Barnston et al. (1999) evaluated 15 schemes over the period June 1996 to March 1998 with similar results. A study by Landsea and Knaff (2000) analyzed the performance of 12 statistical and dynamical schemes in predicting both the onset and decay of the 1997-98 ENSO event and found that none performed better than a statistical tool (Knaff and Landsea, 1997) based on combining climatology and persistence. Landsea and Knaff (2000) maintained that this apparent lack of skill also occurred over the 1993-1996 period.

The findings of Barnston et al. (1994,1999) and Landsea and Knaff (2000) were potentially crucial since they implied that the limit of predictability had probably been approached and that this limit corresponded to, at best, only moderate skill. They also implied that there was little gain in investing in relatively complex dynamical schemes since the case studies indicated that the gain in skill over simple statistical schemes was probably not worthwhile. However, the Landsea and Knaff (2000) conclusions were based on only one assessment technique and, only one El Nino event (1997-1998). A more thorough assessment of seasonal prediction skill would ideally consider a large number of events and also a variety of assessment techniques. This has been hampered by the relatively small effective sample size associated with most dynamical schemes which make use of observed sea surface temperatures, and/or surface wind stresses and/or sub-surface information. Consequently, retrospective predictions (or hindcasts) have tended to span just the past 2 to 3 decades, when observations were of sufficient quality. This has meant that actual number of independent ENSO events available for study has only been about 10. More recently, Chen et al. (2004) analyzed the predictions of an intermediate coupled ocean-atmosphere model for the period 1857 to 2003, the only inputs being reconstructed SSTs. This study encompassed 24 El Nino events and 23 La Nina events, the largest (to date) sample that has been studied. The analysis demonstrated predictive skill at lead times out to 24 months and concluded that "El Nino has been more predictable than previously envisaged". This suggests that the findings of Barnston et al. et al. (1994,1999) and Landsea and Knaff (2000) are perhaps too pessimistic and that further study is warranted.

1.3 The COCA2 dynamical forecast system

The COCA2 dynamical prediction system employs the CSIRO Mk3 global coupled GCM (Gordon et al., 2002). The Mk3 model is an evolution of the CSIRO Mk 2 coupled model which has been used in a numerous coupled model studies (see for example Hirst et al., 1996; Gordon and O'Farrell, 1997; Hirst, 1999; and Hirst et al., 2000). The model is global and simulates both the evolution of SSTs and associated climate variables including rainfall, mean sea level pressure (MSLP), winds etc.

The new Mk3 model now runs without the use of flux adjustments and has strong ENSO variability (Cai et al., 2003). The horizontal resolution of the Mk 3 atmospheric model is spectral T63 (1.875 degrees latitude by 1.875 degrees longitude per grid box) with 18 vertical levels on a hybrid sigma-pressure vertical coordinate. The atmospheric model includes a comprehensive cloud microphysical parameterization (Rotstayn et al. 2000), and the convection parameterization is based on that used in the Hadley Centre model (Gregory and Rowntree, 1990). This convection parameterization has been linked to the cloud microphysics scheme via detrainment of liquid and frozen water at the cloud top. Atmospheric moisture advection (vapor, liquid, and frozen) is carried out by the semi-Lagrangian method (McGregor, 1993). There is a land surface scheme (six layers of moisture and temperature) with a vegetation canopy (Kowalczyk et al., 1994). Atmospheric concentrations of greenhouse gases (expressed as an equivalent CO2 concentration) and stratospheric ozone, and the direct effect of sulfate aerosols, are varied each year according to the observed values.

The Mk 3 ocean model is based upon the Modular Ocean Model version 2.2 (MOM2.2) of the Geophysical Fluid Dynamics Laboratory (GFDL) model. The ocean component has horizontal resolution matching that of the atmospheric model's grid in the east–west direction, and twice that in the north–south direction. Thus the grid boxes are 0.937 degrees latitude by 1.875 degrees longitude and there are two ocean grid boxes per atmospheric grid box in the meridional direction thereby allowing the atmospheric model and ocean model to have identical land–sea masks. There are 31 levels in the vertical, with the spacing of the levels gradually increasing with depth, from 10 m at the surface to 400 m in the deep ocean. The ocean model includes a Gent–McWilliams parameterization for mixing of tracers

(Griffies et al., 1998; Griffies, 1998), and improved vertical mixing that achieves a sharp equatorial thermocline in the tropical Pacific (Wilson, 2000, 2002).

The atmosphere and ocean models are directly coupled with a synchronous time step of 15 minutes. To permit initialization of the COCA2 model for seasonal prediction, some changes have been made to the standard coupling of the Mk 3 model. Whilst salinity and heat fluxes remain directly coupled as in the standard Mk 3 model, surface wind stress and SST are anomaly coupled at each time step by a process called coupled-surface nudging (CSN). anomaly coupled by a process called coupled-surface nudging (CSN)[1].

Part of a CSN initialization sequence is shown in Figure 1. In this figure, the longitudes plotted are for the equatorial Pacific Ocean, and only years 1992 to 2003 have been shown. Wind stress anomalies generated by the Atmospheric General Circulation Model (AGCM) are blended with observed wind stress anomalies before coupling of the blended wind stress field to the Mk3 Ocean General Circulation Model (OGCM). In addition, sea surface temperature (SST) anomalies generated by the OGCM are blended with observed SST anomalies before coupling of the blended SST field to the AGCM. This CSN initialization is carried out over the period from 1978 to current-day. In this way the COCA2 model is pre-conditioned (nudged) towards the pattern of behaviour seen in reality. In Figure 1, the good match of observed FSU wind stress anomalies with the Mk 3 atmospheric model's wind stress anomalies, and the good match of observed Reynolds SST anomalies with the Mk 3 model's SST anomalies, indicates that this is being achieved.

The CSN sequence does not attempt to force or modify wind stress produced within the atmospheric model, or to modify temperature produced within the ocean model. Instead, a blended field is formed from ocean or atmospheric model surface data and observed surface data, and the blended field is passed by the coupling to the other model and used as its surface forcing. This initialization by CSN avoids a number of potential problems:

- Coupling shock is avoided because the model is fully coupled during the lengthy CSN initialization from 1978 to the present. Coupling shock can occur with systems that assimilate observational data into an uncoupled ocean model, and then, at the transition from initialization mode to prediction mode, begin coupling the ocean model to the atmospheric model. Because there is no opportunity for the component models to adjust to each other before the transition to prediction mode, the component models do so abruptly at the beginning of the prediction, and this abrupt adjustment, called "coupling shock" can reduce predictability (Chen et al., 1995).

- Release shock is avoided because the component models are forced only from the surface. With CSN, observational information only penetrates below the ocean surface in accord with the ocean model's own physics so that no dynamical

[1] Use of the term "nudging" to describe techniques for initializing dynamical seasonal forecast models can be confusing because two different techniques have been referred to in the literature as "nudging". Chen et al. (1995) used the present COCA2 model's *blending and coupling* technique and called it nudging. Rosati et al. (1997) relaxed SSTs towards SST observations *within* the ocean model component of their coupled model, and also called it nudging, as do other modelers in the nested modeling community. To distinguish between these two types of nudging, we refer to the present blending and coupling technique as coupled-surface-nudging

imbalances are generated. In contrast, with sub-surface ocean data assimilation, observational information is injected into the ocean model at depth. If this information is not dynamically consistent, or, if the model and observed climatologies differ (as they do in most models), the ocean model will not be able to accept this sub-surface information, and will only appear to do so for as long as the ocean model is constrained by the data assimilation system. At the transition from initialization mode to prediction mode when the sub-surface data assimilation is switched off, an abrupt dynamical adjustment, called "release shock" can occur. Release shock typically generates an ocean Kelvin wave that propagates along the equator and can be confused with ENSO variability. Sophisticated Ensemble Kalman Filter techniques are currently been developed to address the dynamical imbalance part of this release shock problem.

The observed wind stresses employed in the CSN process are: for the equatorial Pacific Ocean, the Florida State University (FSU) objective wind stresses (Bourassa et al., 2001), and for the Indian Ocean, the FSU subjective wind stresses (Legler et al., 1989). The FSU stresses are only available only within latitudes 30 degrees North to 30 degrees South in the tropical Indian and Pacific oceans, they are embedded within the global Hellerman and Rosenstein (1983) wind stress climatology. To avoid discontinuities in the latitudinal transition between the FSU and Hellerman and Rosenstein stresses, data are blended over three latitudinal grid points. The seasonal climatologies used for calculating inter-annually varying wind stress anomalies are: for the coupled model, a climatology calculated from a 20 year freely coupled run, and for the observations, in the Pacific Ocean, the mean seasonal cycle of 1978-2000 Florida State University (FSU) objective wind stresses, and in the Indian Ocean, the mean seasonal cycle of 1970-1992 Florida State University (FSU) subjective wind stresses. The observed SSTs employed in the CSN process are the NOAA OI.v2 monthly SSTs (Reynolds et al., 2002) and the seasonal climatology used for calculating inter-annually varying SST anomalies is the mean 1978-2000 NOAA OI.v2 SSTs (ibid).

Finally, in the CSN process:
- The blending of the atmospheric model and observed wind stress anomalies at each spatial location is uniformly 50% model and 50% observation at all latitudes.
- The blending of the ocean model and observed SST anomalies at each spatial location is a 50% model and 50% observation within latitudes 5ºN/S, rising linearly with increasing latitude to 75% observations and 25% model at 15degrees North and 15 degrees South and is this blend is maintained poleward of these latitudes.

Finally, methods for improving the performance of seasonal prediction models is a large topic and the reader is referred to Palmer et al. (2004) and references therein for some details. It is also worth noting that an important aspect of prediction skill is the use of ensembles, particularly as it has been demonstrated that multi-model ensembles produce more reliable probability predictions than do single-model ensembles (Palmer et al., 2004). Here we deal with an assessment of a single member ensemble from the COCA2 model. As such, the results are likely to represent an underestimate of the potential skill.

A CSN initialization sequence is shown in Figure 1. In this figure, the longitudes plotted are for the equatorial Pacific Ocean, and only years 1992 to 2003 have been shown. Wind stress anomalies generated by the atmospheric model (indicated in the figure as AGCM) are blended with observed wind stress anomalies before coupling of the blended wind stress field to the Mark3 ocean model (indicated in the figure as OGCM). In addition, sea surface temperature (SST) anomalies generated by the ocean model are blended with observed SST anomalies before coupling of the blended SST field to the Mark 3 atmospheric model. This CSN process is carried out over the period from 1978 to current-day. In this way the COCA2 model is pre-conditioned (nudged) towards the pattern of behaviour seen in reality. In Figure 1, the good match of observed FSU wind stress anomalies with the Mark 3 atmospheric model's wind stress anomalies, and the good match of observed Reynolds SST anomalies with the Mark 3 model's SST anomalies, indicates that this is being achieved.

The CSN sequence does not attempt to force or modify wind stress produced *within* the atmospheric model, or to modify temperature produced *within* the ocean model. Instead, a blended field is formed from ocean or atmospheric model surface data and observed surface data, and the blended field is passed by the coupling to the other model and used as its surface forcing. This initialization by CSN avoids a number of potential problems:

- Coupling shock is avoided because the model is fully coupled during the lengthy initialization procedure. Coupling shock can occur with systems that assimilate observational data into an *uncoupled* ocean model, and then, at the transition from initialization mode to prediction mode, begin coupling the ocean model to the atmospheric model. Because there is no opportunity for the component models to adjust to each other *before* the transition to prediction mode, the component models do so abruptly at the beginning of the prediction, and this abrupt adjustment, called "coupling shock" can reduce predictability (Chen et al., 1995).
- Release shock is avoided because the component models are forced only from the surface. With CSN, observational information only penetrates below the ocean surface in accord with the ocean model's own physics so that no dynamical imbalances are generated. In contrast, with sub-surface ocean data assimilation, observational information is injected into the ocean model at depth. If this information is not dynamically consistent, or, if the model and observed climatologies differ, the ocean model will not be able to accept this sub-surface information, and will only appear to do so for as long as the ocean model is constrained by the data assimilation system. At the transition from initialization mode to prediction mode when the sub-surface data assimilation is switched off, an abrupt dynamical adjustment, called "release shock" can occur. Release shock typically generates an ocean Kelvin wave which propagates along the equator and which can be confused with ENSO variability. Sophisticated Ensemble Kalman Filter techniques are currently been developed to address the dynamical imbalance part of this release shock problem.

The observed wind stresses employed in the CSN process are: for the equatorial Pacific Ocean, the Florida State University (FSU) objective wind stresses (Bourassa et al., 2001), and for the Indian Ocean, the FSU subjective wind stresses (Legler et al., 1989). The FSU stresses are obtained with a constant drag coefficient, $C_D = 1.5 \times 10^{-3}$, and as these stresses are available only within 30ºN-30ºS in the tropical Indian and Pacific oceans, they are embedded within the global Hellerman and Rosenstein (1983) wind stress climatology. To avoid discontinuities in the latitudinal transition between the FSU and Hellerman and Rosenstein stresses, data are blended over three latitudinal grid points. The climatologies used for calculating inter-annually varying wind stress anomalies are: for the coupled model, a climatology calculated from a 20 year freely coupled run, and for the observations, in the Pacific Ocean, the mean seasonal cycle of 1978-2000 Florida State University (FSU) objective wind stresses, and in the Indian Ocean, the mean seasonal cycle of 1970-1992

Florida State University (FSU) subjective wind stresses. The observed SSTs employed in the CSN process are the NOAA OI.v2 monthly SSTs (Reynolds et al., 2002) and the climatology used for calculating inter-annually varying SST anomalies is the mean 1978-2000 NOAA OI.v2 SSTs.

Finally, in the CSN process:

- The blending of the atmospheric model and observed wind stress anomalies at each spatial location is uniformly 50% model and 50% observation at all latitudes.
- The blending of the ocean model and observed SST anomalies at each spatial location is a 50% model and 50% observation within latitudes 5ºN/S, rising linearly with increasing latitude to 75% observations and 25% model at 15ºN/S, and is this blend is maintained poleward of 15ºN/S.

2 General assessment of skill

2.1 Introduction

This assessment deals mainly with the results of model hindcasts over the period 1980 to 2003. 1980 represents a point in time when both observed SSTs and observed winds are regarded as more reliable than previous years. For each year, the model was initialized at four specific dates -January 1, April 1, July 1 and October 1 - and each time it was then run forward for 12 months. This resulted in 24 separate 12-month sets of results for January to December, April to March, July to June and October to September - in all, a total of 96. We refer here to a 6-month lead time as being associated with a result for the month of June from a January 1 start date, or the result for August from an April 1 start etc. Ideally, a set of hindcasts would comprise the results based on initializing the model at the start of every calendar month. Furthermore, it would also comprise an ensemble of results based on starting the hindcasts from slightly different initial conditions. However, the computational load associated with such a large number of coupled model runs becomes relatively large. The sample set analysed here, comprising just the four start dates and a single realization in each case, is assumed to be sufficient to provide a reliable estimate of skill.

2.2 Spatial correlations

We firstly focus on the SST hindcasts made with a 6-month lead time. Figures 2(a to d) show the correlations between the predicted and observed SST anomalies for the target months of June (from January 1), September (from April 1), December (from July 1) and March (Year+1, from October 1). Given that each correlation value (r) is based on a sample size of 24 values that we assume to be independent, only values exceeding 0.3 in magnitude are displayed. Values exceeding 0.33 are nominally significant at the 90% level while values exceeding 0.5 are nominally significant at the 99% level.

Figure 2a indicates that the results for June SSTs (from January 1) appear to be skilful (r>+0.5) in north-west Indian Ocean, much of the equatorial Pacific Ocean and two other regions in the Pacific at higher latitudes. The maximum value for r (~+0.9) occurs near 5^{o}S, 120^{o}W. Figure 2b indicates that the results for September SSTs (from April 1) are, overall, less skilful. This is most noticeable in the Indian Ocean where only a few small regions have values which exceed +0.5. This lack of large-scale coherency suggest these are not statistically significant. The results in the Pacific Ocean are broadly similar to those shown in Figure 2a. The maximum value is about +0.8. Figure 2c shows that the results for December SSTs (from July1) in the equatorial Pacific Ocean are more skilful than those made earlier in the year. The pattern of correlations is much more coherent and the region where values exceed +0.8 is greater than previously. Finally, Figure 2d indicates that the results for March SSTs (from October 1) appear to be, overall, the most skilful of the four start dates. There is skill in the Indian Ocean, though less than if Figure 2a, and a large-scale pattern of relatively high values throughout the equatorial Pacific.

These results are consistent with the existence of the so-called “predictability barrier” during the austral autumn months. In effect, it is easier to predict equatorial Pacific SSTs at long lead times after this period than before, i.e., once the climate system becomes locked into either a “warm” or a “cold” event at the end of autumn, it tends to follow a more predictable path.

The situation with regard to SSTs in the Indian Ocean is different. Here it appears that SSTs during autumn and early winter are more predictable than at other times of the year.

2.3 NINO34 SST index

Commonly used ENSO indices are the average of sea surface temperature anomalies in key regions of the equatorial Pacific Ocean. These regions have been selected because they effectively characterize the state of El Niño or La Niña events. The NINO3 index (which refers to the box region 5^oS to 5^oN, 150^oW to 90^oW) is now regarded as less appropriate than the NINO3.4 index (which refers to the box region 5^oS to 5^oN, 170^oW to 120^oW that is located more centrally in the equatorial Pacific Ocean, Barnston et al. ,1997).

Figures 3(a to d) compare the predicted and observed NINO34 anomalies for the four start dates. Each graph shows the set of 24 individual results as discrete 12-month time series (black line segments) while the observations are shown as a continuous (red) line. Each SST anomaly (for a particular month, from a particular start date) is calculated as the difference between the model SST value and the average of all the model SST values for the same month and start date. This calculation effectively removes both the effect of any bias in the model climatology plus the effect of any drift (no matter how small) in the model results. The first value in each discrete time series represents a 1-month lead time prediction whilst the last represents a 12-month lead time.

Figure 3 shows that, while the predicted NINO34 anomalies appear to capture the magnitude of the warm events, they tend to overestimate the strength of cool events. For example Figure 3a, January 1, 1983 represents the peak of one of the largest warm events ever observed. The initially warm anomalies from this start date show a rapid decay over the remainder of the year before switching to relatively cool anomalies (i.e. magnitude in excess of -2^oC) by the end of the year. While the observed anomalies did decay rapidly, they did not exceed $-1\ ^oC$. Figure 3b shows that the results from April 1 1983 also indicate cooling, but not as severe as those from January 1. However, they also suffer from an overestimate of the cooling which took place later that year. Similar problems can be seen with the predictions for 1988 and 1994. Figure 3c shows that, even with a later start date (July 1), the results eventually overestimate the cooling at the end of 1983 and into 1984, and also at the end of 1998 and into 1999. Otherwise, these results appear to capture the magnitude of the warm events. Finally, Figure 3d also shows that the results from October 1 start dates tend to overestimate the strength of cool events during 1983-84, 1998-89 and 1992-93, and also the strength of warm events during 1997-98, 2000-01 and 2001-02.

The NINO3.4 forecast skill at various lead times is summarized in Figures 4(a to d) which show the correlation between the model and observed NINO34 values as a function of lead time for each start date. These correlations are compared with forecasts based on simple persistence, i.e., taking the observed NINO3.4 anomaly from the month before the start of the forecasts and maintaining (persisting) that value over the 12-months of the forecast. The 95% significance level for a sample size of 24 is estimated to be 0.4 and the model values exceed this except for the October forecasts at 9-month lead time.

Figure 4a shows that the NINO34 results from January 1 start dates are relatively skilful (i.e. correlation values exceed +0.7) for the first 5 months. Thereafter values drop to between +0.4 and +0.5 by July with little change by the end of the year. These values can be compared with those for persistence which are relatively skilful for the first 4 months but drop below +0.5 by May and below zero by July. The results from April 1 start dates (Figure 4b) are skilful only for the first 2 months. Thereafter, values drop below +0.7 but remain above +0.5 out to June of the following year. This indicates that, despite the predictability barrier, model forecasts initialized at the start of April still maintains some skill in forecasting ENSO evolution over the remainder of the year. More importantly, the results clearly demonstrate the additional skill that the model can provide over and above persistence. Figure 4c shows that, once beyond the predictability barrier, it is less difficult to predict the evolution of the index over the rest of the year since it exhibits a large amount of persistence. Only when the predictability barrier in April of the following year is encountered do the correlations deteriorate. Finally, Figure 4d shows that both the results from October 1 start dates and persistence are skilful for about 6 months, at which lead time the next April predictability barrier is encountered. At 9 month lead time, values fall below +0.4 but this is the only case where this occurs.

In summary, the hindcast results for the NINO34 index are more skilful than persistence during the first half of the year but are similar (although still skillful) to persistence once the predictability barrier has been passed. The level of skill at the 6 month lead-time also appears to be equal to, or greater than that evident in the results reported by Barnston et al.(1994) for 5 different models over the 1982-1993 period. However, because the COCA2 results are based on a much larger sample size they are more statistically significant. Stockdale (1997) analyzed the performance of two early versions of the ECMWF seasonal forecast model for the period 1981 to 1990. In one case (Stockdale, 1997, Fig.8) there was a marked difference in skill for hindcasts started in January compared to those started in July. Although different samples are involved, the COCA2 results exhibit similar behavior but higher levels of skill. Taking all start dates into consideration, COCA2 also appears to exhibit equal if not higher skill, and definitely more statistically significant levels of skill than 5 models (including the 2 ECMWF models) (Stockdale, 1997, Fig.5).

2.4 Southern Oscillation Index

The Southern Oscillation Index (SOI).has traditionally been used to track Australian climate variability, is used as the basis of some ENSO forecast systems, and is incorporated in to some end user application models, e.g., for pasture and crop growth. Also, fluctuations in the observed SOI index have traditionally been linked to fluctuations in Australian rainfall totals over much of the continent for much of the year. So, while the model demonstrates some skill at predicting the NINO34 index, another important test is to ask how well it can simulate the SOI.

The model values for the SOI are calculated in a similar fashion to the NINO34 index. For each start date, there are 24x12-month long time series of mean sea level pressure values for grid points corresponding to Tahiti and Darwin. The SOI for any particular month within these time series is calculated according to the difference between the Tahiti and Darwin pressures, relative to the long-term average difference over the 24 years. The anomalous difference is then scaled by the standard deviation of the differences and multiplied by 10 to make it comparable with the official SOI values.

Figures 5 (a to d) show, for each of the four start dates, the model SOI values as a sequence of 24x12-month time series, compared with the official values shown as a continuous time series. As with Figure 3, each model time series comprises values corresponding to a 1-month lead time through to values corresponding to a 12-month lead time. The first feature worth noting is that there is far more volatility associated with the monthly SOI values than the NINO34 values yet, despite this, there is noteworthy agreement between the model values and observations in each of the figures.

A clearer indication of the SOI skill at various lead times is summarized in Figures 6 (a to d) which, as was done for the NINO34 index in Figures 4 (a to d), show the correlation between the model and observed values based on the results from each of the four start dates. Again, the 95% significance level for a sample size of 24 is estimated to be 0.4, and this is shown as a red dotted line on each graph. For January 1 start dates, Figure 6a indicates these are skilful up to April, after which the correlation values drop to near zero by May. This again is consistent with the predictability barrier and, although the correlations peak near +0.5 for the month of July, this single value is not meaningful. For the results from April 1 start dates, Figure 6b indicates little skill. There is a moderate value (+0.5) for the first month (April), a deterioration over months 2 (May) and 3 (June), then a peak near +0.5 for months 4 (July), 8 (November) and 12 (March the following year). These are not regarded as particularly meaningful since there is large month-to-month variability in the SOI combined with a medium sample size (24). Figure 6c indicates a similar unclear pattern of variation in the correlations for the results from July 1 start dates. However, Figure 6d indicates that, for the results from October1 start dates, there is relatively high skill over the first few months and, on average, minor skill out to March of the following year. In summary, these SOI results, like those for NINO3.4 in Figure 4 (a to d), are consistent with the existence of a predictability barrier during the autumn period.

The pattern of skill as a function of season and lead time differs from that associated with the NINO34 results and appears to be affected by the relatively large variability associated with the SOI in combination with the limited sample size. The large variability can be diminished if we consider the results for 3-month running averages for the SOI. Figure 7 summarizes the skill associated with these average results and displays the skill curves for each of the four start dates as a function of calendar month. Month 1 corresponds to January, month 13 to January of the following year etc. This display provides a better indication than Figure 5 of when, and for how long, SOI predictive skill occurs. For the January start there appears to be some skill at predicting the SOI until the predictability barrier in April, but little skill thereafter. There is little skill associated with the April 1 start dates, but the results from the July 1 and October 1 start dates appear to be skillful for much of the remainder of the year and into the early part of the subsequent year, until the next April predictability barrier in month 16. Figure 7 not only suggests that the most skillful SOI predictions are those for spring (July start) and summer (October start), but also that this skill can be achieved with lead times up to 6 months.

2.5 Conditional Probabilities

Here we look at skill from the perspective of risk analysis. Rather than focusing on correlation as a measure of skill, we consider how well the model is capable of predicting the observed tercile categories: Below Average (BA) or cool event, Average (A) or neutral, and Above Average (AA) or warm event, for the NINO34 index over the 24-year period 1980 to 2003. This provides a more meaningful measure for many applications since it addresses the direct question of: “How often is the model correct compared to how often is it incorrect?”. Tables 1a to 1d summarize the performance of the model for the four start dates. Each table analyzes four lead times (3, 6, 9 and 12 months), e.g., for the April start dates in Table 1b, a lead time of 3 months = Predicted June, a lead time of 6 months = Predicted September, and so on. There are 9 cells for each lead time corresponding to the 3 predicted x 3 observed tercile matrix of possible outcomes. It should be noted that the 24 years analyzed here represents a relatively small sample size for this type of analysis since it only yields 8 values in each of the three categories. Consequently, this introduces a relatively large amount of uncertainty into the derived probabilities. Within each set of 9 cells, the bolded values along the top-left to bottom-right diagonals indicate how often the model predicted the same category as observed (i.e. successful hits) while the off-diagonal values indicate how often the model incorrectly predicted the observed category. The row below each 9 cell matrix lists the total number of *correct* categorical predictions expressed as a percentage score (or percentage correct “hits” (H %)). For this 3x3 equi-probable contingency table, the expected number of correct “hits” due to chance is 33% and this provides a benchmark score for assessing skill.

Table 1. Predicted and observed tercile categories for the NINO34 index (1980-2003):

(a) January 1 start dates
(b) April 1 start dates
(c) July1 start dates
(d) October 1 start dates

(a)		Predicted March			Predicted June			Predicted September			Predicted December		
		BA	A	AA	BA	A	AA	BA	A	AA	BA	A	AA
Observed	BA	**7**	1	0	**6**	2	0	**5**	2	1	**4**	3	1
	A	1	**5**	2	2	**2**	4	2	**3**	3	3	**2**	2
	AA	0	2	**6**	0	4	**4**	1	3	**4**	1	3	**5**
H%		75%			50%			50%			46%		

(b)		Predicted June			Predicted September			Predicted December			Predicted March		
		BA	A	AA	BA	A	AA	BA	A	AA	BA	A	AA
Observed	BA	5	2	1	4	4	0	5	3	0	5	2	1
	A	3	3	2	2	2	4	2	3	2	2	3	3
	AA	0	3	5	2	2	4	1	2	6	1	3	4
H%		46%			42%			58%			50%		

(c)		Predicted September			Predicted December			Predicted March			Predicted June		
		BA	A	AA	BA	A	AA	BA	A	AA	BA	A	AA
Observed	BA	5	3	0	6	2	0	5	3	0	5	2	1
	A	3	2	3	1	4	2	3	2	3	3	1	4
	AA	0	3	5	1	2	6	0	3	5	0	5	3
H%		50%			67%			50%			38%		

(d)		Predicted December			Predicted March			Predicted June			Predicted September		
		BA	A	AA	BA	A	AA	BA	A	AA	BA	A	AA
Observed	BA	6	2	0	5	2	1	5	2	1	4	2	2
	A	2	4	1	2	5	1	1	4	3	3	1	4
	AA	0	2	7	1	1	6	2	2	4	1	5	2
H%		71%			67%			54%			29%		

The results for the January 1 start dates show that H is a maximum (75%) at a 3-month lead time but decreases to 50% by June and September, and then 46% by December. While this indicates that at least one half of the predictions are incorrect from June onwards, the skill scores are still greater than could be expected due to chance. In the case of the April 1 starts, the skill score for the 3-month lead time is the lowest of all 4 start dates and corresponds to the difficulty presented by the predictability barrier. The score for the 6-month lead time (42%) is lower than the scores for the 9- and 12-month lead times (58% and 50%). This is unexpected and is likely to be an artifact of the relatively small sample size. The 3-month lead time score for the July 1 start dates (50%) is higher than for April 1, as expected, once the predictability barrier is passed. While the 6-month and 9-month lead times are unexpectedly high (67% and 50%) the score for the 12-month lead time (38%) is consistent with the fact that June of the following year lies on the other side of the predictability barrier. Finally, the scores for the October 1 predictions are a high 71% at 3 month lead time and then fall with increasing lead time, as expected. At 6- month lead times the score is still high (67%), decreases to 54% after 9 months and, by 12 months, the predictions have no effective skill, since the score (29%) is less than what could be expected due to chance.

Focusing on just the 6-month lead time scores, poorest skill (42%) occurs with the April 1 predictions, followed by the January 1 predictions (50%) , while the most successful predictions (67% and 67%) are those from July1 and October 1. Thus, despite the relatively small sample size, this method for assessing skill yields results that are consistent with the correlations presented in Figure 4. Importantly, this method also provides an indication of the uncertainty associated with any prediction of the NINO34 index that is not conveyed by correlation values. Thus, the April 1 predictions have relatively low skill, but are still better than what could be achieved by chance. The July 1 and October 1 predictions are more skilful, but are far from being correct all of the time. For example, the 6-month lead predictions from July 1 of a warm (i.e. above average) event in December could be expected:

- To be correct for 6 out 8 occurrences, i.e., to be correct 75% of the time,
- To incorrectly predict a neutral (i.e. average) event instead of the correct warm event for 2 out 8 occurrences, i.e., 25% of the time and
- To almost never incorrectly predict a cool (below average) event instead of the correct warm event.

In Table 2 we convert the tercile categories from Table 1 into conditional probabilities for just the predicted cool and warm tercile events at 6-month lead times. Random skill corresponds to each column containing percentages close to 33.3% (i.e. there is no discrimination). Perfect skill corresponds to 100% in the correct category and 0% in the other two categories. The most skilful predictions are for cool events in June (from January 1) and in December (from July 1) and for warm events in December (from July 1) and March (from October 1). The least successful predictions are for cool events in March (from October 1). Note that predictions of cool events in September (from April 1) are also poor since they were followed by warm events on 25% of occasions.

Table 2. Conditional probabilities associated with NINO34 predictions at 6-month lead time.

6-month lead time		Predicted June		Predicted September		Predicted December		Predicted March	
		Cool	Warm	Cool	Warm	Cool	Warm	Cool	Warm
Observed	Cool Neutral Warm	**75%** 25% 0%	0 50% **50%**	**50%** 25% 25%	0% 50% **50%**	**74%** 13% 13%	0% 25% **75%**	**42%** 25% 13%	13% 13% **74%**

A similar tercile category assessment of the SOI results has been performed but with a major difference being that the 3-month (or seasonal) average values have been analyzed instead of the individual monthly values in Table 1 and 2. Tables 3a and 3b summarize the conditional probabilities associated with just the negative and positive tercile events at 3- and 6-month lead times. Note that because of the seasonal averaging a 3-month lead time for any season refers to the mid-month of that season. For example, the February to April season represent a 3-month lead time for January 1 start dates. Table 3a shows the 3-month lead time hindcasts of both negative SOI (i.e., El Nino or warm event) and positive SOI (i.e., La Nina or cool event) conditions. These forecast were, with one exception, successful between 63% and 75% of the time. At the 6-month lead time, Table 3b shows that hindcasts of both February to April SOI from October 1, and May-July SOI from January 1 were least skillful with only 50% of either negative of positive conditions successful. There appears to be slight improvement when predicting August to October conditions from April 1, and the most skillful results were for November to January SOI from July1 with 75% of either negative or positive conditions successful. This pattern is again consistent with the existence of the predictability barrier.

Table 3 Conditional probabilities associated with (seasonal average) SOI hindcasts:
(a)

3-month lead time		Predicted Feb-Apr		Predicted May-Jul		Predicted Aug-Oct		Predicted Nov-Jan	
		Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive
Observed	Negative Neutral Positive	**75%** 25% 8%	0% 25% **75%**	**50%** 25% 25%	12% 12% **75%**	**63%** 25% 12%	0% 37% **63%**	**63%** 25% 12%	12% 12% **75%**

(b)

6-month lead time		Predicted May-Jul		Predicted Aug-Oct		Predicted Nov-Jan		Predicted Feb-Apr	
		Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive
Observed	Negative Neutral Positive	**50%** 25% 25%	0% 50% **50%**	**50%** 25% 25%	12% 25% **63%**	**75%** 12% 12%	0% 25% **75%**	**50%** 27% 12%	25% 25% **50%**

2.6 Using index forecasts to estimate rainfall.

Having established that the model's predicted NINO34 and SOI values exhibit some skill, we now consider their potential use to predict seasonal rainfall. In the observations, below median NINO34 values and above median SOI values (La Nina-like conditions) are associated with above median rainfall and vice-versa (El Nino-like conditions). For this analysis we use the gridded (0.25 degree by 0.25 degree) Australian monthly rainfall data as provided by the Bureau of Meteorology.

We concentrate on the late spring (OND) season since, according to Figure 7, this is a time of year when the hindcasts appear to be more skillful than at other times of the year. Secondly, the observed correlation between rainfall and ENSO (as measured by either the NINO34 or SOI indices) is strongest during the spring. We firstly calculate the seasonal average values for the hindcasted NINO34 and SOI indices and then rank them into either above median or below median categories according to the 24 values from 1980 to 2003. Above/below median NINO34 values correspond to EL Nino-like/La-Nina-like conditions whereas below/above median SOI values correspond to EL Nino-like/La-Nina-like conditions. The same procedure is followed with the average rainfall for the same 3-month season at each grid point. An estimate of skill is the frequency with which the observed seasonal average rainfall is above median when the hindcasts indicate La Nina-like conditions. Because of symmetry, this frequency is also the frequency with which the rainfall falls below median when the hindcasts indicate El Nino-like conditions. The baseline frequency is 50% since this is the expected frequency, over the long term, associated with random predictions. A more detailed assessment of skill is probably not justified because the sample size of 24 is relatively small. This means that the derived frequencies are subject to relatively large uncertainty (of the order of 10%). Given this uncertainty, a conservative threshold for estimating potential skill is 60%. However, any measure of skill needs to be compared with that which could potentially be achieved by other means, including the use of statistically based schemes that are currently in operational use. One simple alternative scheme is persistence, i.e., to assume that the NINO3.4, or SOI, for OND will be the same as that for the preceding season JAS.

Figure 8a shows the level of skill associated with hindcasts of the NINO34 index for OND made from start dates on October 1, a lead time of 1.5 months. The highest values appear in the far north-east and the south-east of Australia. Moderate values are evident over much of the eastern half of the continent and over the far western part of the continent. Figure 8b shows the level of skill associated with SOI hindcasts from start dates on October 1. The pattern is similar to Figure 8a but is characterized by slightly higher values throughout the centre of the continent and slightly lower values in the east and west. This is compared in Figure 8c with the persisted pattern obtained by assuming the mean SOI for July to September applies to the OND season. Again, the skill associated with persistence is very similar, again indicating no significant increase in skill over persistence at this short lead time.

Figure 7 suggests that the SOI during spring (i.e. mid September) is predictable from as far back as July 1, a lead time of 4.5 months. Figures 9 (a to c) compare the skill associated with hindcast values for both NINO34 and the SOI from July1 start dates and persisted values for the SOI index. In this case the persisted SOI value is the mean value for April to June. Figure 9a indicates that relatively large areas of skill (values exceeding 70%) are associated with the NINO34 predictions. In the case of the long-lead SOI predictions, a comparison of Figures 9b and 9c suggest no gain in skill over persistence. Thus it appears the greatest source of additional skill may be that associated with long-lead predictions of the NINO34 index from start dates after the predictability barrier in autumn.

2.7 Latitude of the sub-tropical ridge

Here we use the results from the COCA2 CSN initialization, i.e., “nudged” run to estimate the ability of the model to capture variability in the latitude of the sub-tropical ridge in the Australian region. This is essentially the latitude of the maximum in mean sea-level pressure (MSLP) between about 17°S and 55°S. It is thought to represent the latitude at which high-pressure systems traverse from west to east, and therefore is related to meteorological conditions at the surface. The CSN initialization run is based on continually using observations of SSTs and winds to “nudge” the model towards the observed state. Because it is nudged by observations it represents the optimum simulation that can be achieved by the model since it does not involve any predictions. In effect, if the model does not exhibit any skill when run in this configuration, then it will not exhibit any skill when run in forecast mode. In effect, we are testing the role of blended model and observed SSTs in determining the ridge position.

Pittock (1973) has calculated the latitude of the sub-tropical ridge (L) at longitude 150°E from monthly mean station data along the east coast of Australia. McIntosh et al. (2004a) calculated L values from NCEP reanalysis data. Here we calculate L values from the COCA2 CSN initialization run for comparison with the observational values. We also calculate L values for all southern hemisphere longitudes. At each longitude, the latitude of the maximum pressure is determined between 17.5°S and 55°S. At those longitudes where a maximum does not occur, L is determined by interpolation from neighboring (in a longitudinal sense) values of L. The interpolation used is piecewise cubic hermite polynomial interpolation. This method produces a curve with a continuous first derivative, and does not introduce spurious extrema. We then apply a triangular moving average to the L values as a function of longitude, with half-width of five grid points (each NCEP grid point is 2.5° of longitude, each COCA2 grid point is 1.875°).

Figure 10 shows the annual L (average of monthly L) at 150°E from NCEP and COCA2. Although there is not a very good year-to-year correlation (r=0.3), there is good general agreement, including the poleward shift of L in the mid-to late-1990’s. The mean seasonal cycle of L is shown in Figure 11 and indicates that COCA2 has a larger southward excursion of L in summer (DJF) by around 2°.

Finally, Figure 12 shows the longitudinal dependence of the yearly mean L for both NCEP and COCA2. The largest differences are in the mid-Pacific (~200^{o} E) and in the south Atlantic (~340^{o} E) where the COCA2 values indicate the ridge is simulated to be too far south by about 5°. Differences near the west coast of South America (~300^{o} E) can probably be ignored because of the difficulty in extrapolating to MSLP values in the presence of significant topography.

In summary, these results demonstrate that the climatology of the COCA2 model, insofar as the location of the MSLP maxima in the southern hemisphere is concerned, compares well with the observations. They also indicate that year-to-year fluctuations in SSTs appear to affect the location of the sub-tropical ridge over eastern Australia on an annual basis. This is not unexpected since we know that warm and cool events in the equatorial Pacific Ocean are strongly linked to MSLP fluctuations over Australia (as evidenced by fluctuations in the Southern Oscillation Index (SOI)). While the link is not strong, Figure 10 suggests that skilful predictions of SSTs may manifest themselves as moderately skilful predictions of L-values.

3 Applications

This section deals explicitly with comparing predicted and observed climatic variables at three key locations: Southwest Western Australia (SWWA), the Murray-Darling catchment basin (MDB) and the Burdekin region of north Queensland (see Figure 13). The averaged observed climate variables used in these comparisons are generated from the NCEP reanalysis data set (Kalnay et al. , 1996) using the data extraction tools provided by the NOAA-CIRES Climate Diagnostics Center web site at http://www.cdc.noaa.gov/

3.1 South West Western Australia

SWWA is a winter-rainfall dominated regime but, unlike eastern Australia, seasonal rainfall is still largely unpredictable there. Here we analyze some of the hindcasted fields from the COCA2 April 1 hindcast runs in order to detect any predictability of subsequent peak winter (June July August) rainfall.

Figure 14 compares the observed winter rainfall totals over the SWWA region (see IOCIP, 2002) with those hindcasted by COCA2 over a box area covering most of this region (116 to 118^{o}E, 35 to 32^{o}S). Also shown are the NCEP rainfall totals for the same box area. The hindcasted values underestimate the observations by about a factor of 2 while the NCEP values underestimate the observations by about a factor of three. This is partly the result of averaging over slightly different regions but is mainly due to the relatively coarse resolution of the COCA2 and NCEP models. It has been shown that the use of finer horizontal resolution provides a better representation of topography which contributes to higher rainfall amounts. Despite this bias, and the (partly related) reduced variability, the NCEP values are highly correlated with observations (r =+0.88) over the 24-year period (see Table 4). It should be remembered that the NCEP values are not hindcasts but are continuously forced model generated values driven by observed SSTs and daily observations of pressure, winds, moisture etc. So the NCEP values are proxy observations rather than predictions. In contrast, the COCA2 hindcasts are predictions at a lead time of 3.5 months. The correlation between the hindcast and observed values is only +0.19 which indicates that, for this region at this time of year, there is essentially no predictability associated with the model rainfall values.

Previous studies have documented the strong inverse relationship between observed Perth mean sea level pressure (MSLP) and SWWA winter rainfall (Allan and Haylock, 1993; Smith et al. 2000; IOCIP 2002). In addition, statistical downscaling of large scale atmospheric fields also identified the north-south MSLP gradient and dew-point temperature depression at 850hPa (RH850) as important factors which affect point-scale daily rainfall (Charles et. al, 2004). Charles et al. (2004) found there was some predictability associated with the probability of rainfall occurrence but little associated with rainfall amount. An alternative assessment of potential predictability is to focus on the skill of COCA2 in simulating mean winter MSLP over SWWA from the April 1 start dates. Figure 15 compares the observed and hindcasted JJA MSLP time series over the period 1980 to 2003. Apart from a bias of about +4 hPa, the hindcasts show very little skill since the correlation coefficient between the two series is only -0.19. This suggests that there would be relatively little skill associated with statistical downscaling of the COCA2 MSLP fields over SWWA. Table 4 shows that hindcast JJA MSLP values are only weakly correlated (r=+0.20) with the observed rainfall.

Table 4. Correlation between various indices and SWWA observed winter (Jun-Aug) rainfall.

Index	**(1980-2003)**	**(1948-2003)**
NCEP rainfall	+0.88	+0.63
NCEP MSLP	-0.80	-0.82
NCEP RH850	+0.21	+0.40
SOI (obs)	+0.43	+0.34
COCA2 rainfall	+0.19	
COCA2 MSLP	+0.20	
COCA2 SOI	-0.13	

While there is little evidence of predictability associated with the April runs, it could be expected that hindcasts beginning on July 1 might be more skilful, if only because they begin after the ENSO predictability barrier. Table 5 shows the results from the same analyses of the April hindcasts, the only difference is that the target season is now the July to October (or late winter) season. The relationship the NCEP-derived and observed rainfall values is again relatively high (r=+0.75), but the COCA2 rainfall hindcasts again correlate poorly (r=+0.18). There is some improvement in the correlations between MSLP and rainfall (r =-0.14 compared to +0.20) and between the model SOI and rainfall (r=+0.35 compared to -0.13). Figure 16 compares the result of using the hindcast SOI values to derive rainfall anomalies (via linear regression) and suggests that this index may provide the most useful predictive information.

Table 5. Correlation between various indices and SWWA observed late winter (Jul-Oct) rainfall.

Index	**(1980-2003)**	**(1948-2003)**
NCEP rainfall	+0.75	+0.59
NCEP MSLP	-0.80	-0.78
NCEP RH850	+0.16	+0.41
SOI (obs)	+0.39	+0.56
COCA2 rainfall	+0.18	
COCA2 MSLP	-0.14	
COCA2 SOI	+0.35	

Weak predictability of SWWA winter rainfall is partly due to the fact that ENSO events arise in the Pacific Ocean region and have their greatest impacts over the eastern part of the Australian continent. Even so, these impacts tend to be strongest in the tropics, rather than the mid-latitudes, and during the spring (SON), rather than the winter season. While it has been hoped that SSTs outside of the equatorial Pacific Ocean may be potential sources of additional predictability, observational studies to date (e.g. Smith et al., 2000) have yet to identify any statistically significant links. This suggests that SSTs at mid- to high- latitudes may not be well observed or that there are no such links to be discovered. Climate model studies based on observed SSTs all seem to indicate the latter. Experiments with both the CSIRO Mk3 atmospheric model and the high resolution conformal cubic model (CCAM) forced by observed SSTs all fail to reproduce both decadal and interannual variability of SWWA winter rainfall. On the other hand, recent experiments with the CCAM model using both observed SSTs and observed winds as forcing, do reproduce the major fluctuations in rainfall (IOCIP, 2004). The implication from these studies is that rainfall is primarily driven by fluctuations in the large-scale circulation as measured by the winds and pressure patterns, and not by the SSTs. The question as to what drives these fluctuations remains unanswered at present, but is likely to involve processes at the high latitudes such as fluctuations in the extent of Antarctic sea ice. Whatever the processes, observational products are still of questionable quality in these data-sparse regions. This will limit our ability to identify any useful links to rainfall and, by definition, make improved seasonal predictions difficult. At this stage, the SOI index offers the only source of (albeit) weak predictability of SWWA winter rainfall.

3.2 Murray Darling Basin

In this analysis we focus on the links between both rainfall over the Murray Darling Basin catchment regions (defined here as the box-region 140 to 150^{o}E, 33^{o}S to 37^{o}S, see Figure 13) and measured inflows into the Burrinjuck dam. The mean rainfall and mean inflow are calculated for May to October and cover the period 1967 to 2003 for which inflow data are available.

Table 6 summarizes the simultaneous correlations between Burrinjuck inflows and key climatic indices. While inflows are related to some extent to rainfall over the larger catchment region (r =+0.67), they would, in fact, be more closely related to rainfall over just the upper catchment of the Murrumbidgee River. The link to large scale variables is indicated by the moderate (but significant) correlations between NCEP 850 hPa relative humidity (r=+0.50), mean sea level pressure (r=-0.49) and SOI (r=+0.51). As is expected, ENSO events have more of an impact in this region than they do in SWWA. However, the analysis of the COCA2 hindcasts from April suggests that neither the hindcasted rainfall, MSLP or SOI for the May to October period provides useful information about Burrinjuck inflows.

May to October rainfall over the MDB region (Table 7), can also be seen to be related to the large-scale atmospheric variables including what appears to be a link to northerly winds at 850 hPa, (V850). However, none of the COCA2 April hindcast variables provide any useful information. This lack of predictability is primarily due to the fact that the April 1 hindcasts begin within the predictability barrier and, as indicated by the assessments in Section 1 above, have limited skill.

Table 6. Correlation between various indices and Burrinjuck May to October inflows.

Index	1967 to 2003	1980 to 2003
Observed rainfall	+0.67	
NCEP rainfall	+0.43	
NCEP RH850	+0.50	
NCEP MSLP	-0.49	
NCEP U850	0.00	
NCEP V850	+0.03	
SOI May-Oct (Obs)	+0.51	+0.53
COCA2 rain May to Oct		+.24
COCA2 MSLP May-Oct		-0.21
COCA2 SOI May-Oct		+.01

Table 7. Correlation between various indices and Murray Darling Basin catchment rainfall May to October.

Index	1967 to 2003	1980 to 2003
NCEP rainfall	+0.74	
NCEP RH850	+0.65	
NCEP MSLP	-0.54	
NCEP U850	+.11	
NCEP V850	-0.41	
SOI May-Oct (Obs)	+0.63	+0.53
COCA2 rainfall		-0.03
COCA2 MSLP May-Oct		+0.07
COCA2 SOI May-Oct		-0.18

The July 1 hindcasts, made as they are after the end of the predictability barrier, are more skilful. This is indicated by Table 8 which shows that the July to October inflows are weakly linked to hindcasts of July to October rainfall over the MDB region (r=+0.37) and SOI (r=+0.36). Surprisingly, there is less predictability associated with July to October catchment rainfall (Table 9) with only the hindcasted SOI possibly providing some useful information (r=+0.32). Figure 17 compares the observed July to October Burrinjuck inflows with those derived from a linear regression of the observations and the hindcast SOI values for the same period. The comparison of the observed and derived values therefore indicates how well a simple statistical forecast scheme would perform based on the model-generated SOI values. Figure 18 does the same for July to October MDB catchment rainfall.

Table 8. Correlation between various indices and Burrinjuck July to October inflows.

	1980 to 2003
COCA2 rain Jul-Oct	+0.37
COCA2 MSLP Jul-Oct	+0.33
COCA2 SOI Jul-Oct	+0.36

Table 9. Correlation between various indices and Murray Darling Basin catchment rainfall July to October.

	1980 to 2003
COCA2 rain Jul-Oct	+0.23
COCA2 MSLP Jul-Oct	+0.07
COCA2 SOI Jul-Oct	+0.32

In summary, the predicted SOI values from July onwards appear to provide some level of predictability for both the inflows and catchment rainfall over the ensuing period July to October. This level is not high but is consistent with what is known about the strength of the relationship between the SOI and rainfall for this part of the continent at this time of year. McBride and Nicholls (1983), for example, show that the simultaneous correlation is about +0.5, consistent with the values reported in Table 7. It is to be expected that the correlation between the *predicted* SOI and observed rainfall will be less than this.

3.3 Burdekin region

3.3.1 Burdekin rainfall

In this analysis we focus on rainfall over the Burdekin region defined here as the box-region (146.5 to 147^{o}E, 19.5 to 21^{o}S). Firstly, we consider the predictability of July to September and October to December seasonal rainfall from the July 1 hindcasts.

Unlike the SWWA and MDB regions, Table 10 indicates that for the Burdekin region COCA2 hindcast rainfall does appear to contain some moderate skill (r=+0.35). This is expected because the Burdekin region is more affected by ENSO events. However, this skill appears to decrease rapidly since there is no skill (r=+0.01) associated with October to December rainfall. There is less evidence of a link between model SOI and rainfall (r=+0.23). However, there is a much stronger link with MSLP (r=-0.57) which is still apparent for the later October to December period (r=-0.53). For the Burdekin, this suggests a useful role for MSLP-based dynamical or statistical downscaling.

Table 10. Correlation between COCA2 hindcast indices and Burdekin rainfall.

	July to September 1980 to 2003	**October to December 1980 to 2003**
Index		
COCA2 rainfall	+0.35	+0.01
COCA2 MSLP	-0.57	-0.53
COCA2 SOI	+0.23	-0.07

Figure 18 compares the observed July to September rainfall with that derived from a linear regression fit to observed rainfall using the forecast MSLP as a predictor.. The derived values capture the relative wet years 1981, 1984, 1988 and 1998, and the relatively dry years 1982 to 1983, 1987, 1990 to and 1992, and 1997. These results indicate MSLP predictions can provide a useful basis for predicting July to September rainfall for Burdekin region. This potential is greater than indicated by the results for the other regions and seasons considered. It is consistent with the fact that ENSO has its strongest impacts over Australia in the east and north-east of the continent during the spring season.

3.3.2 Plant growth predicted from daily average values of climate indices

Here we test whether the COCA2 model hindcasts can be used to make predictions of plant growth in the Burdekin region. In a first case we investigate the use of daily averages of hindcast variables from the model within a plant growth model. In a second case we investigate whether SST hindcasts can be used to predict plant growth given links between observed SST patterns and growth (McIntosh et al., 2004b)

Growth index days (GID) are the number of days in a month that conditions are suitable for plant growth. Hence monthly values lie between 0 and the number of days in the month. For north-east Queensland, a day is considered suitable for growth if a moisture index is greater than 0.2. The moisture index varies from 0 to 1, and is the fractional soil water content relative to the maximum available water at field capacity; see McDonald (1994) for more details.

Daily averages over the Burdekin region have been extracted from both the COCA2 CSN initialization (nudged) run and the July 1 start date hindcasts. For comparison, a daily average time series has been constructed using data from nine stations in the Burdekin region. GID were calculated from both the model and observational daily data. The GID were calculated using a model called GRIM (McDonald, 1994), with parameters tuned for north-east Queensland conditions. Values were calculated each day from daily data provided by the Bureau of Meteorology (SILO), and then binned up into months. The GID calculation began on 1 January 1957, the beginning of SILO data. The starting value for soil moisture is assumed to be 60mm but is effectively “forgotten” after a few months. The starting value for soil moisture for the model calculation is taken to be the soil moisture from the SILO calculation on 1 January 1980 (36.7mm).

Figure 19 compares the SILO-derived and the model-derived (nudged run) monthly GID from 1980-2003. The SILO-derived data vary from 0 to 31 and vary on yearly and longer timescales. The model-derived GID is always equal to the number of days in the month, except for the final value which is 24. Clearly there is a problem causing the model GID to be too high.

The dominant factor determining GID is rainfall. Figure 20 compares the Burdekin region monthly rainfall totals from both SILO and COCA2. Although the model captures the seasonal cycle, it overestimates the rainfall amounts. The mean rainfall is observed to be 49 mm/month, whereas the mean COCA2 rainfall is 72 mm/month. However, the observed median rainfall is 25 mm/month, which is much greater than that for COCA2 - 9 mm/month. A histogram of monthly rainfall (not shown) confirms that COCA2 produces far more extreme rainfall events, both high and low, than observed.

The next factor influencing GID is evaporation. Figure 21 compares monthly average evaporation from both SILO and COCA2. There is considerably less evaporation occurring in COCA2 than observed, perhaps by a factor of 3. In addition, the timing of the maximum evaporation from COCA2 appears to occur one to two months later than observed.

Given that COCA2 overestimates rainfall and underestimates evaporation compared to the SILO data, it is not surprising that COCA2 GID is much higher. In effect, the model implies that there is always enough soil moisture to drive plant growth.

It also implies that runoff must be overestimated. This quantity is not directly observed but can be estimated using a hydrological model such as GRIM. We compare “observed” values (based on SILO data) with those calculated directly by COCA2. In addition we can also derive model runoff values using GRIM and the model values for rainfall and evaporation. Figure 22 compares the three time series and confirms that the COCA2 results (either direct or as derived) lead to severe overestimates. The mean “observed” runoff is only 12 mm/month, compared to a directly calculated value of 21 mm/month, and a derived value of 61 mm/month. This result also indicates that there is a considerable difference between the GRIM and the COCA2 soil models.

Finally, another important factor in plant growth is mean temperature - defined here as the average of the daily maximum and minimum temperatures. Figure 23 compares observed and model values and indicates that, while the seasonal cycle is well simulated, the model overestimates temperatures by about 2^{o}. Examination of the individual time series (not shown) indicates that, while daily maxima are well simulated, the model overestimates minima by about 4°C.

In summary, while there are many climatological features of the Burdekin region that can be reasonably well simulated by the model, the important variable rainfall is severely overestimated. In terms of applying the data to plant growth models and similar, it is clear that the raw model values are not appropriate and that the anomaly values should be used, as when doing correlation analysis. This would remove the effect of the bias in the mean rainfall and more clearly demonstrate if the model can reproduce important interannual signals. This remains to be analyzed.

3.3.2 Plant growth predicted from SST patterns

As demonstrated previously, model hindcasts of raw rainfall seem to carry less information than model hindcasts of related variables such as MSLP, SSTs or SOI values. Here we investigate the feasibility of using seasonal hindcasts of SSTs, together with some knowledge of the relationship between SST and rainfall or plant growth (GID). Statistical prediction models, such as used in the Oceans to Farms project (McIntosh et al. 2004b), use a similar technique. In the case of statistical models, the relationship between SST and GID must be a lagged relationship in order that a true prediction is possible. When using SST from a dynamical model, there is the additional possibility of using a simultaneous relationship. We compare the skill associated with the predicted SSTs to that associated with the pure statistical method.

The partial least squares (PLS) method is used to make a statistical connection between SST and Burdekin GID and hence make a prediction. The PLS method is described in McIntosh et al. (2004b). The calculations are done in a software package called SSTman, which was written for the Oceans to Farms project described in McIntosh et al. (2004b), and subsequently updated for more general use in data visualization and correlation.

GID values were calculated using SILO meteorological data at 9 sites using the method described in McDonald (1994). The 9 values were averaged to produce a regional average, and then averaged over the 9 month period July-March. The model SST data were taken from hindcasts started on April 1 and July 1. Because of known biases, the data were converted into anomalies. Availability of model data and observed SST data sets at the time of this analysis has restricted the calculations to the 21-year period 1982-2002. In all cases the skill values are for cross-validated skill. This means that when predicting GID values for a particular year, data from that year is excluded from the calculation. In particular, with the PLS method, this means that the SST/GID correlation patterns are calculated with the hindcast year removed.

3.3.2.1 SST forecasts from July 1 starts

If a prediction of July-March GID is to be made from observed SST, then the latest month that can be used is June. The correlation between SST at a grid point and GID (r) is treated as a weighting factor. The PLS method then selects those grid points with the highest correlations as potential predictors. The subsequent SST values at these grid points are then used to generate a set of GID predictions. A single GID prediction is calculated as the weighted average of this set. Therefore, those grid points with the highest SST/GID correlations contribute proportionally more to the prediction than other grid points.

Figure 26. shows the correlation pattern between observed June SSTs and subsequent July-March GID. This indicates that warm SSTs adjacent to the continent, and an ENSO pattern of cool SSTs over much of the equatorial Pacific Ocean, are associated with high GID. This is consistent with a positive link between Australian rainfall and ENSO cool (La Nina) events. As a first test, we calculate the equivalent correlation pattern using the COCA2 CSN initialization (nudged) SST data for June (the last month before the July predictions). We expect that, because the SST anomalies outside the tropics are nudged up to the month prior to the prediction start date, they should closely resemble the observed SST anomalies and, produce a similar correlation map as shown in Figure. 24. Figure. 25 shows that they are indeed very similar, although there are also a few differences.

Next we look at the cross-validated hindcast skill using SST from various months (June through to March) to predict GID. Note that the only true hindcast is that based on June data. Figure.26 compares the resultant skill using observations with that based on the predicted SSTs. The first thing to note is that although the June SST correlation maps are very similar, the resultant COCA2 skill (r=+0.51), is considerably lower than the skill obtained using the observed data (r=+0.66).

The second thing to note is that the model skill never exceeds +0.60 for any month. The implication is here is that the optimum prediction for July-March GID is that obtained using observed June SST rather than using model predictions of SST. Another point to note is that even a perfect model would not give a much better result, as indicated by the fact that the skill based on observed SSTs after June does not change very much.

Figure.27 shows the correlation map between observed SSTs and GID , not only for June, but also for each month thereafter through until March the following year. The equivalent set of maps based on the COCA2 July 1 predicted SSTs is shown in Figure.28. While these correlation patterns are qualitatively fairly similar, there are also some differences. Probably the major difference of relevance to Australia is the westward extension of the negative correlations associated with the predicted SSTs along the equator in the west Pacific and the fact that the area of negative correlation along the whole equatorial Pacific is narrower than that associated with the observed SSTs.

3.3.2.2 SST forecasts from April 1 starts

In this case, we analyse the results of model predictions starting on April 1 start dates in order to see whether the earlier start date provides useful predictive information at a longer lead time. Figure. 29 shows the cross-validated skill at predicting the July-March GID using SST in various months. Here the model and observed skills in the month prior to starting a prediction (March) are very similar, which indicates that the model SST anomalies in March are more similar to the observed than they are for June. Beyond March however, when the prediction has started, the model quickly loses skill compared to the observations. This is to some extent expected, given that the earlier start lies within the predictability barrier and any skill (associated with predicting ENSO events) will be relatively poor. While a prediction of GID based on the model result in April is slightly better than a prediction based on observed SST in March, this is not regarded as significant. In fact, there is more information to be gained by waiting until the end of the month when the observed SST for April can be seen to be relatively skilful (r=+0.60), even if less than that for the end of June (r=+0.66).

The evolution of the SST correlation patterns are shown for both the observed and COCA2 predcitions in Figures. 30 and 31 respectively. While the March patterns are similar, as might be expected from the similar skill figures, there is a fairly rapid divergence in similarity over the next 4 months. In particular, the model correlation on the equator near the dateline decreases in magnitude, while the observations do not exhibit this feature. Limiting the spatial extent of the SST used to the region where the model performs well (120°E to 300°E, 10°S to 10°N) does not improve the model skill. The same is true for predictions starting in July.

The major conclusions from this analysis are that, as far as July to March Dalrymple GID is concerned:

- There is little skill associated with SST predictions made from April 1
- There is more skill associated with SST predictions made from July 1 but this is less than that associated with observed SSTs for June.
- There is some discrepancy between the model SST anomalies in June used as initialization, and the observed anomalies. It is possible that the model performance could be improved by addressing the reason for these differences, but any such improvement is unlikely to lead to significantly greater skill than that based on using the observed June SST as a predictor.

4 Summary of findings

An assessment has been performed of the performance of a coupled ocean-atmosphere seasonal forecast model (COCA2) which has been developed to predict El Nino-Southern Oscillation (ENSO) events. The assessment deals with the results from a series of hindcast runs with the model from four start dates per year, for the 24 year period 1980 to 2003. The model is firstly assessed with regard to its ability to predict two key indices- the NINO34 sea surface temperature index and the Southern Oscillation Index (SOI).

- It is found that, in terms of the NINO34 and SOI indices, the model exhibits levels of skill that appear, at worst, equal to other forecast systems discussed in the literature and are most likely higher in most cases.

- It is found that skill at predicting these indices is strongly dependent on time of year with the austral autumn period (the so-called ENSO predictability barrier) acting as a strong inhibitor. The greatest skill is associated with hindcasts initialized after this period and appears to be significant for relatively long lead times. The level of skill associated with the NINO34 hindcasts, appears at least as high, if not higher, than that reported for other seasonal forecast systems although making these comparisons is made difficult by the different sample sizes and periods that have been analyzed.

- The utility of skillful forecasts of these indices is further analysed by considering their implications for Australian rainfall. It is found that, for the season when the indices are most skillfully predicted (OND) the long-lead forecasts for the NINO34 index imply shifts in rainfall that appear to be more skillful than could be achieved by simple persistence.

- In a related study, effectively involving the use of prescribed sea surface temperatures, it is found that there is no real skill associated with predicting the latitude of the sub-tropical ridge.

An assessment is also performed of the potential value of model output variables to three regions: (1) southwest Western Australia, (SWWA), (2) the Murray Darling Catchment Basin (MDB) (including the Burrinjuck Dam) and (3) the Burdekin region of north Queensland. In this assessment, seasonal mean values of model output variables were compared with observed mean values for each of the 24 hindcast years. The degree of correspondence between both time series is expressed by the correlation coefficient with statistically significant values (at the 5% level) being those which exceed about 0.39 in magnitude. Variables and predictors which exhibit correlations near or exceeding this value are as follows:

- For SWWA rainfall (July to October) the best predictor is the SOI predicted from July 1 start dates (correlation r= +0.35)

- For Burrinjuck July to October inflows the best predictor is the rainfall predicted from July 1 (correlation r=+0.37)

- For MDB catchment rainfall (July to October) the best predictor is the SOI predicted from July 1 (correlation r=+0.32)

- For Burdekin rainfall (July to September) the best predictor is mean sea level pressure (MSLP) predicted from July 1 (correlation r=-0.57)

- For Burdekin rainfall (October to December) the best predictor is MSLP predicted from July 1 (correlation r=-0.53)

This particular assessment indicates relatively low skill in the southwest of the continent but increasing towards the north-east of the continent where ENSO has its strongest impacts on Australian climate variability. It is also consistent with the fact that the ENSO signal is strongest and also most predictable during the latter half of the year.

Another example of how model forecast variables can be used is to use daily values (of rainfall, temperature solar radiation etc.) to drive a crop growth model. A preliminary assessment of this approach for the Burdekin region indicated that the raw model values can be (and are most likely to be) biased to an extent which severely limits this type of application. While model biases can be overcome, this still needs to be assessed further.

Finally, there are some studies which indicate a link between SST patterns (not necessarily those associated with ENSO) and indices of pasture growth. The potential application here is to forecast the SST patterns and thereby generate some information about expected yields. Preliminary assessments of this approach suggest that any skill is less than or equal to that associated with the SST patterns observed at the time of the forecast.

The results presented here indicate that, unless the inhibitions posed by the ENSO predictability barrier can be overcome, it is probably more efficient to concentrate future assessments on the performance of hindcasts (or forecasts) made after this time of the year.

However, this assessment only provides an indication of the limits and potential for ENSO forecast models. It is constrained by the number of years analyzed (24), the fact that only four start dates per year were considered and only one 12-month hindcast was performed for each start date. The skill estimates presented here, referring as they do to a single-model single member ensemble, most likely represent an underestimate of true skill. This issue will be addressed in future work by analyzing the multi-model data sets that have been made available by the European Centre for Medium Range Weather Forecasts (ECMWF) under the DEMETER project (Palmer et al., 2004). At the same time, the time scales for model development and model experimentation are such that future work will also allow for the inclusion of any results generated from an updated version of the COCA2 model, as well as any results generated by the new Bureau of Meterology POAMA model (Oscar Alves, private communication).

The DEMETER data comprise hindcast results, for 6 months ahead in time, from each of 4 start dates per year, from 7 global coupled ocean-atmosphere models, spanning up to 43 years in some cases. Each model has been used to generate an ensemble of 9 members. It can be appreciated that DEMETER data represent an enormous resource and provide the opportunity for a thorough assessment of potential forecast skill for Australia. Work is underway to begin extracting from this database, key climatic fields for the key Australian regions. It is planned to assess these data in the manner described in this report but also using a range of measures such as described by Hartmann et al. (2000).

5 References

Allan, R.J. and Haylock, M.R. (1993). Circulation features associated with the winter rainfall decrease in southwestern Australia. J. Climate, 6, 1356-1367.

Barnston, A.G., M.Chelliah and S.B.Goldenburg (1997) Documentation of a highly ENSO-related SST region in the equatorial Pacific. Atmosphere-Ocean, 35(3), 367-383.

Barnston A.G. and Coauthors (1994) Long-lead seasonal forecasts – where do we stand? Bull. Amer. Meteor. Soc., 75, 2097-2114.

Barnston A.G., M.H.Glantz and Y.He (1999) Predictive skill of statistical and dynamical climate models in SST forecasts during the 1997-98 El Niño episode and the 1988 La Niña onset. Bull. Amer. Meteor. Soc., 80, 217-243.

Bourassa, M. A., S. R. Smith, and J. J. O'Brien, 2001: A new FSU winds and flux climatology. 11th Conference on Interations of the Sea and Atmosphere, San Diego, CA, Amer. Meteor. Soc., 912.

Cai, W., M. A. Collier, H. B. Gordon and L. J. Waterman (2003). Strong ENSO Variability and a Super-ENSO Pair in the CSIRO Mark 3 Coupled Climate Model. Mon. Wea. Rev., 131, 1189-1210.

Charles, P.C., Bates, B.C, Smith, I.N. and J.P.Hughes (2004) Statistical downscaling of observed and modelled atmospheric fields. Hydrological Processes, 18 (8), 1373-1394.

Chen, D., S. Zebiak, A. J. Busalacchi, and M. A. Cane (1995). An improved procedure for El Niño forecasting. Science, 269, 1699–1702.

Chen, D., M.A. Crane, A. Kaplan, S.E. Zebiak and D. Huang (2004) Predictability of El Nino over the past 148 years. Nature, 248, 733-736.

Drosdowsky, W. and L.E. Chambers (2001) Near global sea surface temperature anomalies as predictors of Australian seasonal rainfall, J.Clim., 14, 1677-1687.

Gregory, D., and P. R. Rowntree (1990) A mass flux convection scheme with representation of cloud ensemble characteristics and stability dependent closure. *Mon. Wea. Rev.,* 118**,** 1483–1506.

Griffies, S. M. (1998) The Gent–McWilliams skew flux. *J. Phys. Oceanogr.,* **28,** 831–841.

Griffies, S. M, A. Gnanadesikan, R. C. Pacanowski, V. D. Larichev, J. K. Dukowicz, and R. D. Smith, (1998) Isoneutral diffusion in a *z*-coordinate ocean model. *J. Phys. Oceanogr.,* **28,** 805–830.

Gordon, H. B., and S. P. O'Farrell (1997) Transient climate change in the CSIRO coupled model with dynamic sea ice. Mon. Wea. Rev., 125, 875–907.

Gordon, H. B., L. D. Rotstayn, J. L. McGregor, M. R. Dix, E. A. Kowalczyk, S. P. O'Farrell, L. J. Waterman, A. C. Hirst, S. G. Wilson, I. G. Watterson and T. I. Elliot (2002). The CSIRO Mk3 Climate System Model. CSIRO Division of Atmospheric Research Tech. Paper 60, 130 pages.

Hartmann, H.C., T.C. Pagano, S. Sorooshian and R. Bales (2002) Confidence builders; Evaluating seasonal climate forescasts from user perspectives. Bull. Amer. Met. Soc., 83(5), 683-698.

Hirst, A., H. Gordon, and S. O Farrell (1996) Global warming in a coupled climate model including oceanic eddy-induced advection. Geophys. Res. Lett, 23, 3361–3364.

Hirst, A., S. O Farrell, and H. Gordon (2000) Comparison of a Coupled Ocean Atmosphere Model with and without Oceanic Eddy-Induced Advection. Part I: Ocean Spinup and Control Integrations. J. Climate, 13, 139–163.

Hirst, A. C. (1999) The Southern Ocean response to global warming in the CSIRO coupled ocean-atmosphere model. Environmental Modelling and Software, 14, 227–241.

Indian Ocean Climate Initiative Panel (IOCIP), (2002). Climate variability and change in south west Western Australia. September, 2002.

Indian Ocean Climate Initiative Panel (IOCIP), (2004). Report of Stage 2, Phase 2 activity. (to be published).

Knaff, J. A. and C. W. Landsea (1997) An El Niño -Southern Oscillation Climatology and Persistence (CLIPER) forecasting system. Wea. Forecasting, 12, 633-652.

Landsea C.W. and J.A. Knaff (2000) How much skill was there in forecasting the very strong 1997-98 El Niño ? Bull. Amer. Meteor. Soc., 81(9), 2107–2120.

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

Kowalczyk, E. A., J. R. Garratt, and P. B. Krummel (1994) Implementation of a soil canopy scheme into the CSIRO GCM Regional aspects of the model response. Technical Paper 32, CSIRO Atmospheric Research.

Legler, D., I. Navon, and J. O'Brien (1989). Objective Analysis of pseudostress over the Indian Ocean using a direct-minimization approach. Mon. Wea. Rev., 117, 709–720.

Levitus, S. (1982) Climatological Atlas of the World Ocean, NOAA Professional Paper 13, National Oceanic and Atmospheric Administration, Rockville, Maryland.

McDonald, C. K. (1994) Calculating climatic indices affecting plant growth. CSIRO Division of Tropical Crops and Pastures, Tropical Agronomy Technical Memorandum, No. 83, 35pp.

McIntosh, P.C., M.J. Pook and G.A. Meyers (2004a) Movement of the Sub-Tropical Ridge in the Southern Hemisphere (manuscript in preparation).

McIntosh, P.C., A.J. Ash, and M. Stafford Smith (2004b) From Oceans to Farms: Using Sea-Surface Temperatures in Agricultural Management. (Submitted to *Journal of Climate)*

McGregor, J. L. (1993) Economical determination of departure points for semi-Lagrangian models. *Mon. Wea. Rev.,* **121,** 221–230.

McGregor, J., H. Gordon, I. Watterson, M. Dix, and L.D. Rotstayn (1993) The CSIRO 9-level Atmospheric General Circulation Model, Technical Paper 26, CSIRO Atmospheric Research.

Pacanowski, R. C. (1995) MOM2 Documentation User's Guide and Reference Manual. Technical Report 3, Geophysical Fluid Dynamics Laboratory, Box 308, Princeton, NJ 08542.

Peters, H., M. C. Gregg, and J. M. Toole (1988) On the Parameterization of Equatorial Turbulence. Geophys. Res. (Oc.), 93, 1199–1218.

Pittock, B., (1973) Global meridional interactions in stratosphere and troposphere.*Quart. J. R. Met. Soc.*, **99,** 424-437.

Reynolds, R.W., N.A. Rayner, T.M. Smith, D.C. Stokes, and W. Wang (2002) An Improved In Situ and Satellite SST Analysis for Climate, J. Climate, 15, 1609-1625.

Rosati, A., K. Miyakoda, and R. Gudel (1997) The Impact of Ocean Initial Conditions on ENSO Forecasting with a Coupled Model. Mon. Wea. Rev., 125, 754–772.

Rotstayn, L. D., B. F. Ryan, and J. J. Katzfey (2000) A scheme for calculation of the liquid fraction in mixed-phase stratiform clouds in large-scale models. *Mon. Wea. Rev.,* 128**,** 1070–1088.

McBride, J.L. and Nicholls, N., 1983. Seasonal relationships between Australian rainfall and the Southern Oscillation. Mon. Weath. Rev., 111, 1998-2004.

Simonot, J.-Y., and H.L. Treut (1986) A climatological field of mean optical properties of the world ocean. Jour. Geophys. Res.-Oceans, 91, 6642-6646.

Smith I.N., P. McIntosh, T.J. Ansell, C.J.C. Reason and K.McInnes (2000) South-west Western Australian winter rainfall and its association with Indian Ocean climate variability, *International Journal of Climatology*, 20, 1913-1930

Stone, R.C., G.L. Hammer, T. Marcussen, (1996) Prediction of global rainfall probabilities using phases of the Southern Oscillation Index. Nature, 384, 252-255.

Wilson, S.G. (2000) How ocean vertical mixing and accumulation of warm surface water influence the "sharpness" of the equatorial thermocline. J.Climate, 13, 3638-3656.

Wilson, S. G. (2002). Evaluation of various vertical mixing parameterizations in a tropical Pacific Ocean GCM. Ocean Modelling, 4, 291-311.

Wood, A. W. and E.P. Maurer (2002) Long-range experimental hydrologic forecasting for the eastern United States. J. Geophys. Res., 107(D20), 4429,doi:10.1029/2001JD000659.