South Eastern Australian
Climate initiative

Final report for Project 3.1.5

3.1.5 An assessment of the skill of dynamical seasonal prediction models for South Eastern Australia

Principal Investigator: Ian Smith

CSIRO Marine and Atmospheric Research, Ph: 03 9239 4539, Fax: 03 9239 4444, Email: ian.smith@csiro.au, PBM 1, ASPENDALE, VIC, 3195.

Completed: 28 March 2007

Abstract
We have assessed the performance of 7 state-of-the-art coupled seasonal climate prediction models in terms of their ability to predict rainfall for 2 regions within south-eastern Australia. The assessment is based on a comparison of hindcast and observed data for the Mallee region and the upper Murrumbidgee catchment region. The target periods and start dates include May to October from May 1 and August to October from August 1. Up to 44 years of data were assessed, with each model generating an ensemble of 9 hindcasts for each time period. The assessment reveals that the models are characterised by only modest skill and that most of this is due to being able to capture dry and wet episodes associated with either El Nino or La Nina events. Typically the models were successful at hindcasting above or below median rainfall between only 50% and 60% of the time. There was no evidence of a single superior model and there was also no evidence that the construction of a multi-model ensemble adds any significant skill.

Introduction
The basis of most current seasonal forecast schemes is the fact that El Nino Southern Oscillation (ENSO) events represent the largest source of interannual climate variability beyond the seasonal cycle. ENSO events can be predicted to some extent since they evolve over the course of several months and there are a number of indices which can be used to predict the likelihood of occurrence. This method of seasonal prediction is described as statistical (as opposed to dynamical) and tends to only provide information about key ENSO indices such as sea surface temperatures or the Southern Oscillation Index (SOI). In general, dynamical based prediction schemes take an initial state of the atmosphere and ocean, and predict the evolution of both well into the future. This is analogous to weather prediction except that the aim is to predict seasonal averages of quantities such as rainfall and temperature rather than specific events. The genesis of ENSO events appears to lie in a complex interaction between Pacific surface wind stresses, sub-surface heat content and sea surface temperatures during the early part of the year. One potential advantage of dynamical schemes over statistical schemes lies in the fact that they can be initialised with this type of information and should, in theory, be more capable of correctly predicting the evolution of ENSO events.

Another advantage of dynamical prediction schemes is that they tend to be global and predict a range of climate variables such as temperature and rainfall. Reliable long-term rainfall predictions would, obviously, be of enormous benefit to a wide range of industries – particularly in Australia where the year to year variability of rainfall is relatively high compared to other continents. However, predictive skill varies considerably with the variable being predicted, the geographic location, the time of year and lead time. While researchers who develop prediction models mainly focus on maximizing their level of skill, possibly more important is the need to convey any predictive information in a manner that provides end-users with the best opportunity to benefit from any skill (Hartmann et al, 2002, various authors, 2005).

A study by Smith (2005) focussed on the skill of a suite of dynamical prediction models at predicting seasonal rainfall and inflows for the Burrinjuck Dam located within the upper Murrumbidgee catchment (UMC) region of south east Australia (SEA). The models exhibited only moderate skill, mainly associated with extremes associated with ENSO events. Furthermore, the value of this moderate level of skill was somewhat questionable. Here we extend the analysis of rainfall predictions from the same suite of models to include another region within SEA where the major activity is growing crops. The aim is to determine if there is any skill at predicting rainfall from early in the season with sufficient time to modify on-farm management decisions early in the year. In particular, we focus on the benefits that may be associated with adopting a multi-model approach to constructing a seasonal forecast. These have been identified in several studies including Palmer et al. (2004), Hagedorn et al. (2005) and Doblas-Reyes et al. (2005).

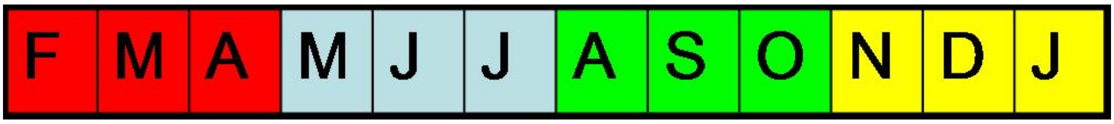
Data and methods

DEMETER is the acronym of the EU-funded project entitled "Development of a European Multimodel Ensemble system for seasonal to inTERannual prediction". The objective of the project is to develop a well-validated European coupled multi-model ensemble forecast system for reliable seasonal to interannual prediction. This obviously involves models capable of simulating ENSO events. A fundamental aspect is to establish the practical utility of such a system, particularly to the agriculture and health sectors (see http://www.ecmwf.int/research/demeter/).

The DEMETER set of results refers to hindcasts from 7 different European coupled models made for different periods up to the end of year 2001. The longest sets of results are associated with models which begin in 1958 and the shortest set is with the CERFACS model which begins in 1980 (Table 1). For each year, the DEMETER hindcasts were initialised on four specific dates (February1, May 1, August 1 and November 1) and then run forward in time for 6 months (Figure 1).Each set of results comprise an ensemble of 9 members initialised slightly differently on the same date. For further details see Palmer et al. (2004).

Table 1. DEMETER model and hindcast periods.

UKMO (UK)	1959-2001
MPI (Germany)	1969-2001
METEO (France)	1958-2001
LODYC (France)	1974-2001
CERF (France)	1980 -2001
INGV (Italy)	1973-2001
ECMWF	1958-2001



1st 1st 1st 1st
*…………………………………………….>
*---.>
*………………………………………..>
*……………………
……………………>

Figure 1. Start dates and length of hindcast runs.

Hindcast data were extracted for the Mallee region (-37o to -34o S, 141o to 144o E) and the UMC region (-36o to -34o S, 148o to 150o E) see Figure 2. These two regions were selected because they correspond to crop growing and water resource management activities respectively. Furthermore, because ENSO-related rainfall anomalies are large-scale (e.g. see Smith, 2004), any assessment of skill within these regions is believed to be representative of the larger region.

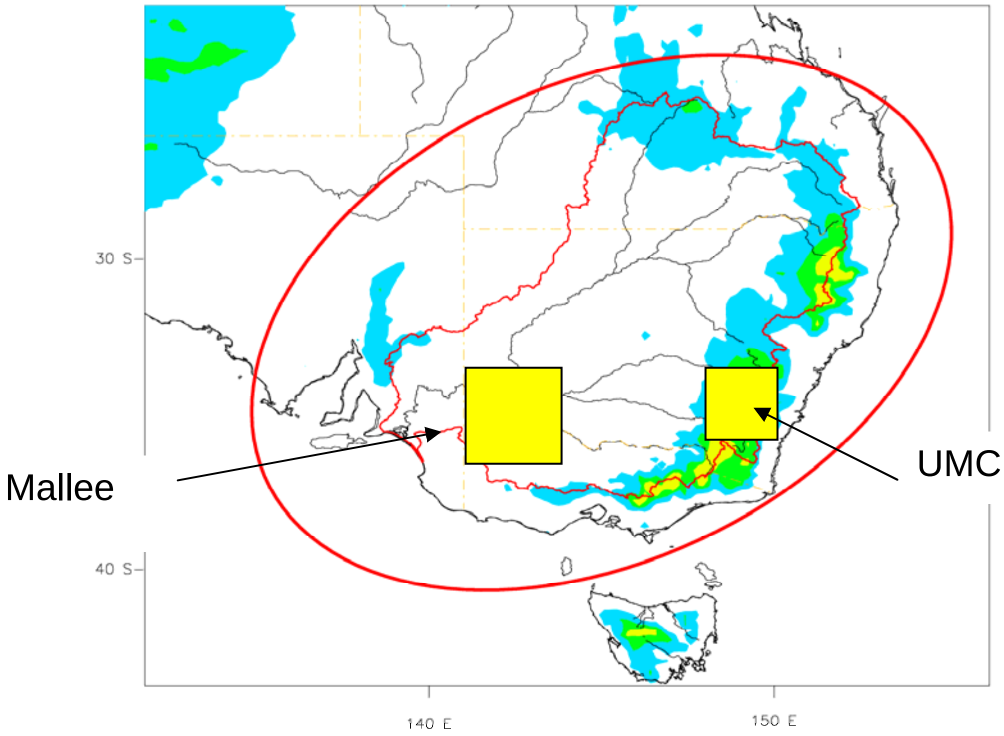


Figure 2. Location of regions within the SEA broader region where model predictions have been assessed.

For both regions, we assess the rainfall predictions for the May to October season, from May 1 start dates, and also the August to October rainfall predictions from August 1 start dates. These represent medium and longer-term seasonal predictions and are selected since these times of the year correspond to management opportunities for both water resource managers and crop growers. Secondly, these are the times of the year when any ENSO-related climate signals are at their strongest. The observed rainfall totals for each of the sub regions were extracted from the Bureau of Meteorology Interactive Australian Rainfall and Surface Temperature portal (see: http://www.bom.gov.au/cgi-bin/silo/cli_var/area_timeseries.pl).

For each year, we calculate individual model ensembles (based on the average of the 9 predictions that each model generates for the specified target seasons). We also average these individual ensemble values (a total of 7) to obtain a single multi-model ensemble value for each target season each year.

Results for the Mallee region

Figure 3 provides an indication of the skill of the models as it compares the individual ensemble mean values for May to October rainfall from each model with the observed values, for every year when data is available. It also shows the multi-model ensemble mean for each year. Figure 4 shows the same results associated with the August to October predictions.

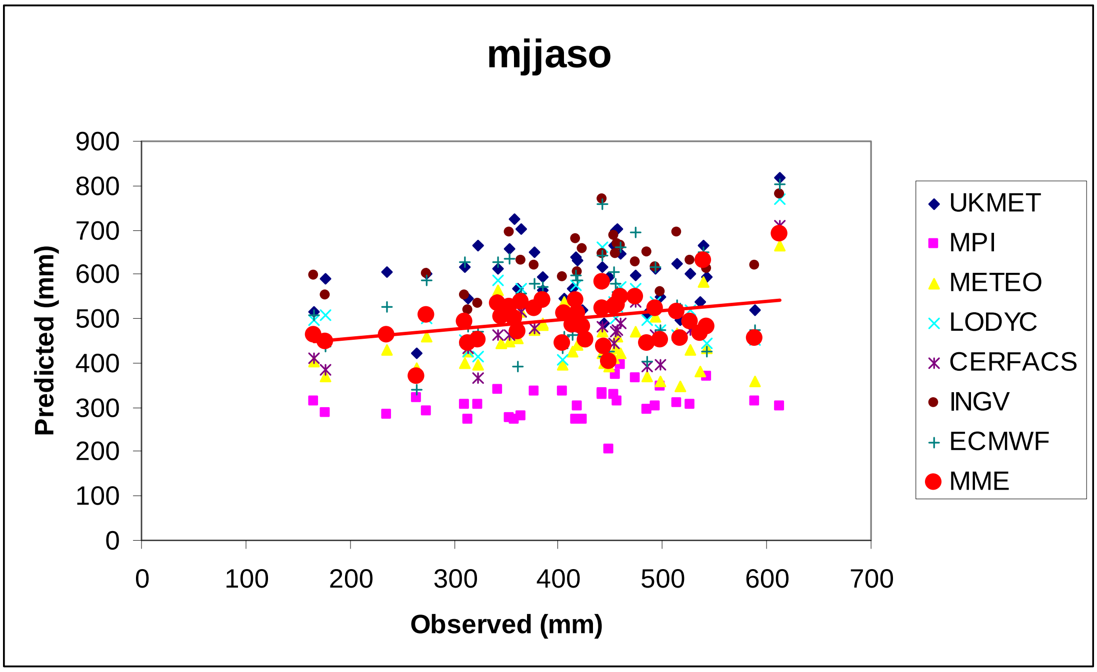


Figure 3. Predicted versus observed May to October rainfall for the Mallee region. The different symbols refer to the ensemble means for each model for each year. The heavy red symbols correspond to the multi-model ensemble means (MME) which are accompanied by a line of best fit.

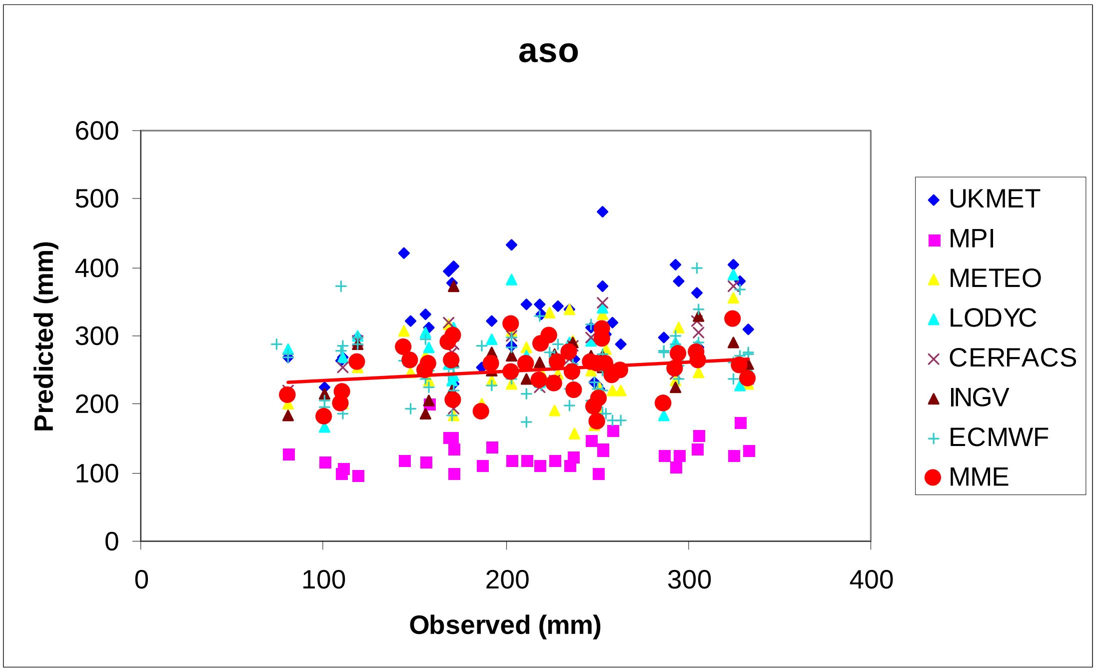


Figure 4. As for Figure 3 except August to October rainfall.

Table 1 quantifies the skill levels (for both target periods) in terms of (a) how often each model was able to predict below average (BA) rainfall, average (AV) rainfall and above average (AA) rainfall categories (a 3-category prediction), (b) how often each model was able to predict above/below median rainfall (a 2-category prediction) and (c) the percentage variation of the observations explained by the predictions.

Table 1. Skill of DEMETER model ensembles at predicting Mallee region rainfall (all available years).

Model	% correct in tercile categories				% correct for above/below median	% variation explained
	BA	AV	AA			
	May to October from May 1					
UKMET	33	36	33	`	52	3%
MPI	37	34	44		57	7%
METEO	36	39	35		46	3%
LODYC	40	39	31		58	7%
CERF	43	35	37		55	32%
INGV	49	32	43		63	27%
ECMWF	35	30	39		55	5%
Average	39	35	37		55	
Multi-model ensemble	40	27	33		50	27%
	August to October from August 1					
UMMET	44	35	37	`	56	4%
MPI	41	38	38		53	6%
METEO	40	27	44		54	4%
LODYC	35	29	35		46	4%
CERF	40	32	25		47	19%
INGV	48	22	32		53	13%
ECMWF	44	30	47		57	2%
Average 42	30	37	52			
Multi-model ensemble	40	29	40		50	5%

For the full period, May to October, it can be seen that there is some indication of skill since the average 3-category scores for BA and AA rainfall are slightly greater than expected by chance (33%) while the average category-2 score (55%) is also above that expected by chance (50%). However, there is not much significant difference in the individual scores for the different models. The variance explained appears to indicate that the CERF and INGV models perform best, but it is also the case that these model results represent relatively small sample sizes (22 and 29 years respectively). The multi-model ensemble appears at first to be relatively skilful since the percentage variance explained is relatively high (27.1%). However, in terms of the 3- and 2-catgory scores it is not much better than chance.

A similar result is evident when analysing the August to October results. In this case we might expect a much higher level of skill since the models are only predicting 3 months ahead in time and, by August1, the models should have been initialised with a much more definite El Nino or La Nina temperatures compared to May 1. In fact, there is very little significant difference between the skill scores for this later period. The multi model ensemble mean results are also unremarkable.

Results for the Upper Murrumbidgee Catchment region

The same assessment of hindcasts for the UMC region was also performed using Bureau of Meteorology Interactive Australian Rainfall and Surface Temperature portal (see: http://www.bom.gov.au/cgi-bin/silo/cli_var/area_timeseries.pl) for the observed rainfall totals.

These data are very similar to the values interpolated by Smith (2005) from the original monthly gridded data.

The results are summarized in Table 2. In this case the average results are very similar to the Mallee region. For both target periods, the average 2-category score is 56% and 54% respectively with most models exhibiting skill at predicting below average conditions (average score 45% and 40% respectively). Again, there is no apparent skill at predicting average conditions (34% and 34% respectively). It is also apparent that there is not a great deal of difference between the performances of the different models. The multi-model ensemble mean does appear to exhibit superior skill at predicting below average conditions (47% and 47% respectively), but less so for the 2-category scores (56% and 55% respectively). In each case, it is outscored by at least one other model.

Table 2. As for Table 1 except for UMC region rainfall

Model	% correct in tercile categories BA	AV	AA	% correct for above/below median	% variation explained
	May to October from May 1				
UMMET	45	42	43	56	8%
MPI	40	22	34	54	2%
METEO	44	37	41	57	16%
LODYC	43	37	41	56	15%
CERF	46	36	38	55	21%
INGV	56	35	38	61	20%
ECMWF	38	30	44	55	12%
Average	45	34	40	56	
Multi-model ensemble	47	47	43	59	15%
	August to October from August 1				
UMMET	41	39	47	54	14%
MPI	37	32	43	56	5%
METEO	43	28	44	54	8%
LODYC	41	42	35	52	4%
CERF	33	36	41	55	19%
INGV	50	30	39	61	12%
ECMWF	37	33	36	45	2%
Average	40	34	41	54	
Multi-model ensemble	47	33	29	55	10%

Conclusions

We have assessed the skill of various European coupled models at predicting seasonal rainfall for two sub regions (the Mallee region and the upper Murrumbidgee catchment region) within the southeastern Australia study area. These sub regions cover activities where climate information can potentially assist decision making – specifically, crop management and water resource management. Furthermore, given that rainfall anomalies and the influence of ENSO events on rainfall is relatively large-scale, the results from these sub regions are believed to provide a good indication of model performance over the larger region.

The target predictions assessed are for May to October rainfall from May 1, and August to October rainfall from August 1. These predictions represent time frames when the predictions can add value and also correspond to the time of the year when ENSO-related climate anomalies are strongest.

The assessment is based on available hindcast predictions made for the years 1958 to 2001. There were no comparable data available from the POAMA model that could be included in this assessment. These will be assessed under SEACI Projects 3.1.3 and 3.2.2

For both regions and both target periods, there is evidence of some skill at predicting rainfall anomalies. However, the level of skill is relatively low, with most of it apparently arising from the ability of the models to partly capture extremes associated with either El Nino or La Nina events. There is no evidence of significant skill at predicting average conditions, but there is evidence that the models do slightly better at predicting below average conditions rather than above average conditions.

There is no significant difference between the skill of the various models. i.e. there is no indication of a consistently superior or inferior performance by any single model.

Despite claims in the literature of the benefits of constructing multi-model ensembles, in the case of south-eastern Australia winter and spring rainfall we find no evidence that these provide extra skill. In fact, the 2-category and 3-category scores for the multi-model ensembles tend to be more inferior than not, when compared with individual model scores.

Even if there was a slight improvement in skill, the fact that the baseline levels are so low indicates that the expense of generating operational multi-model ensembles is not a practical way forward in the Australian context.

Acknowledgements
This study was made possible by funding provided to CSIRO under the South-East Australia Climate Initiative and by funding provided under the Water for a Healthy Country Flagship Program.

References
Doblas-Reyes, F.J, R. Hagedorn and T.N. Palmer (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting – II. Calibration and combination. Tellus, 57A, 3, 234-252.

Hagedorn, R., F.J. Doblas-Reyes and T.N. Palmer (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting – I. Basic concept. Tellus, 57A, 2,19-233.

Hartmann H.C., T. C. Pagano, S. Sorooshian, and R. Bales (2002) Confidence builders; Evaluating seasonal climate forecasts from user perspectives. Bulletin of the American Meteorological Society, 83 (5), 683–698.

Palmer, T.N., A. Alessandri, U. Andersen, P. Cantelaube, M. Davey, P. Délécluse, M. Déqué, E. Díez, F. J. Doblas-Reyes, H. Feddersen, R. Graham, S. Gualdi, J.-F. Guérémy, R. Hagedorn, M. Hoshen, N. Keenlyside, M. Latif, A. Lazar, E. Maisonnave, V. Marletto, A. P. Morse, B. Orfila, P. Rogel, J.-M. Terres, M. C. Thomson (2004) Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER), Bulletin of the American Meteorological Society, 85(6), 853–872.

Smith, I.N. (2004) Trends in Australian rainfall - are they unusual? Australian Meteorological Magazine, 53(3), 163-173.

Smith, I.N. (2005) Assessing the skill and value of seasonal climate predictions. In Zerger, A. and Argent, R.M. (eds) MODSIM 2005 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, December 2005, pp. 170-176. ISBN: 0-9758400-2-9. http://www.mssanz.org.au/modsim05/papers/smith_1.pdf

Various authors, (2005) Special issue of Tellus A, 57(3).

Project Milestone Reporting Table

Milestone description	Performance indicators	Completion date	Budget for Milestone ($)	Progress (1- 3 dot points)	Recommended changes to work plan (1- 3 dot points)
1.Extract observed and model hindcast data for selected regions and target periods	All time series available for analysis	30/04/2006	35k	Completed	Nil
2. Evaluate performance of each model. Develop a multi-model approach.	Skill scores and significance values tabulated. Multi-model approach developed and results compared with individual models.	31/08/2006	35k	Completed Completed	Nil Nil
3.Writing up	Technical report describing findings completed. (End-of-project report).	31/12/2006	35k	Completed	