Assessing The Skill And Value Of Seasonal Climate Predictions

I.N.Smith

CSIRO Marine and Atmospheric Research, Aspendale, VIC 3195, ian.smith@csiro.au

Keywords*: Seasonal predictions, skill, El Nino/La Nina, rainfall, inflows, Burrinjuck Dam*

EXTENDED ABSTRACT

The main driver for the large research effort devoted to developing and improving seasonal climate prediction models is the fact that El Nino Southern Oscillation (ENSO) events (quasi-periodic fluctuations in Indo-Pacific Ocean sea surface temperatures and mean sea level pressure) represent, on a global scale, the greatest source of interannual climate variability and are, to some extent, predictable. Australia is considerably impacted by these fluctuations and although is served by several operational prediction schemes, the associated degree of skill is, at best, only moderate.

There now exist quite a number of dynamically-based seasonal prediction models which are global in extent and there is considerable interest in developing methods for maximizing and quantifying their skill and utility to potential end-users. It is also possible to assess their performance by accessing hindcast (retrospective prediction) data. One of these models was developed by CSIRO and is based on the CSIRO Mk3 global coupled climate model. Results from seven other models which comprise the DEMETER ("Development of a European Multimodel Ensemble system for seasonal to inTERannual prediction") project were also assessed.

This paper focuses on an assessment of the skill of the models at predicting rainfall for a catchment region of south-east Australia. In each case, rainfall hindcasts are compared with observed rainfall totals and also compared with observed inflows into one of the major reservoirs, the Burrinjuck dam.

The major findings are:

- It is not possible to distinguish between the performance of the different models due to different sample sizes and periods for which hindcasts are available.
- Overall, the models exhibit an ability to capture, to some extent, variations in seasonal rainfall associated with ENSO events and this is evident in the fact that they exhibit skill in the extreme categories but not in the average category.
- The average success rate, while greater than that expected by chance or the strategy which assumes climatology, is not high and is expressed in the slight shifts in the probabilities for below average and above average tercile categories. As a rough guide, the model-based predictions provide an advantage over climatology 1 year in every 10.
- Taking into account the fact that rainfall and inflow predictions can be somewhat redundant when dealing with water storages, this may overestimate the potential utility to end –users.
- Finally, it has to be recognised that the economic value of predictions, no matter how skilful, can be diminished according to the costs/benefits associated with decisions made by the end-user. Assessing value is a more task which needs to be done on a case-by-case basis.

1. 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 initialized 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 high level of skill does not automatically translate into economic value. In fact, the overall value of seasonal predictions to the community can be difficult to assess because of the multitude of factors which affect decisions made by climate affected industries (grazing, cropping, water resource management etc.).

Here we assess the skill of a suite of dynamical seasonal prediction models, with regard to how well they can predict rainfall fluctuations over a region of Australia. Specifically, we assess the potential for better predicting inflows into a dam within the Murrumbidgee water catchment region since these affect total storage at the end of winter/spring which, in turn, is a large determinant of water releases made over the dry summer season. Rainfall in this region is known to be affected by ENSO events (c.f. Smith, 2004) which, if predictable, could potentially provide much earlier, and therefore more useful, estimates of likely water allocations to downstream irrigators.

2. Burrinjuck dam – rainfall and inflows

The Burrinjuck dam was the first major dam built for irrigation in New South Wales and is situated in the upper catchment of the Murrumbidgee River (see Figure 1). It was one of the first dams in NSW to have environmental flow releases based on inflows. The holding capacity is 1,026 gigalitres, (almost half the volume of water in Sydney Harbour), the surface area is 5,500 hectares (more than 8,000 football fields), and its catchment area is 13,953 square kilometres (larger than the catchment area of the whole of the Snowy Mountains).

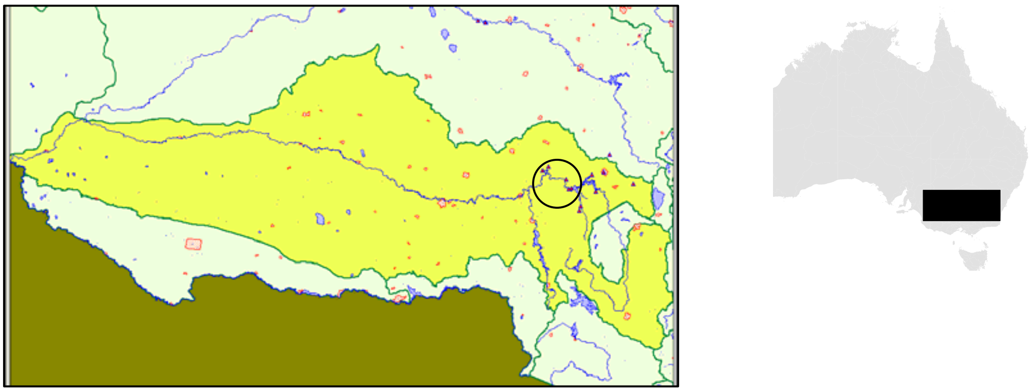

Figure 1. Map of the Murrumbidgee catchment region with the location of the Burrinjuck reservoir indicated.

Figure 2 indicates how total storage in the Burrinjuck Dam has varied over two recent years. In general, storages are replenished during the winter/spring period and diminish over the summer part of the year when water is released for both downstream irrigation purposes and environmental flows. In 2001 there was much more water available for release than in subsequent years which were persistently dry. Irrigators, who depend on releases over the summer are provided with estimates of likely releases and hence water allocations well before summer. While this allows time for appropriate management decisions to be made, it would be more advantageous to extend this lead time by actually predicting likely winter/spring inflows. Therefore, although the water resource management strategy is fixed, predictive information about inflows combined with information about current storage levels could improve forward estimates of summer releases which affect the irrigators.

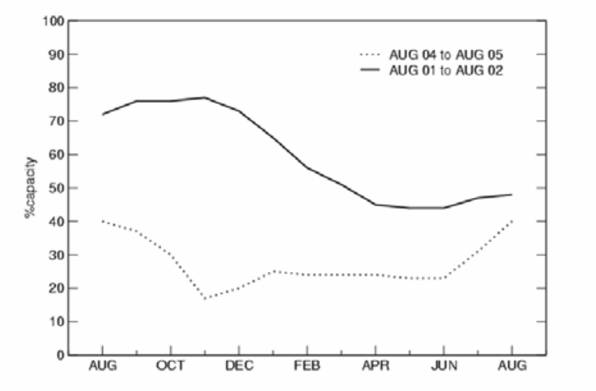

Figure 2. Burrinjuck Dam total storage as a percentage of capacity for two recent 12-month periods, August 2001 to August 2002 and August 2004 to August 2005**.**

Figures 3 provides an indication of the relationship between observed total rainfall over the catchment region for the 6-month period May to October and observed total inflows over the same period, for each year 1967 to 2003. As expected, inflows are related, to some extent, to rainfall with the correlation between both time series being +0.70. Another way of expressing this relationship is to note how often inflows are above average (i.e. fall within the top tercile of all value) when rainfall does the same, and how often they are below average (bottom tercile) when rainfall does the same. If we describe these tercile categories as “extremes”, then there were 15 extreme rainfall years (out of a total of 24) which coincided with extreme inflows. This represent a success rate of 63% compared to the expected success rate of only 33% due to chance (or random guesswork). It should be noted that this represents a relatively crude comparison since; (a) it assumes that: the gridded, observed rainfall data set used to calculate total rainfall is representative of what actually fell within the catchment region; (b) neither total inflows nor total rainfall are normally distributed quantities (for example, the total inflow in 1974 is about 4 standard deviations above the mean); (c) inflows are not simply dependent on in-season total rainfall, but can also depend on pre-season soil moisture content and therefore pre-season rainfall. Despite these issues, it is apparent that a skillful prediction of total rainfall can provide information about expected total inflows.

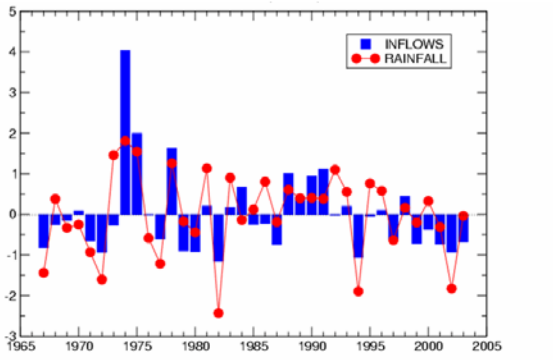

Figure 3. The relationship between inflows to Burrinjuck Dam and catchment area average May to October rainfall 1967 to 2001. The values shown are normalized anomalies.

Figure 4 shows the same comparison, except for the late season period August to October. In this case the correlation between the two time series is higher (+0.77) and the number of matched extreme events is 16 (or 67%). Despite the sample comprising only 37 years, these numbers also suggest that there may be useful information contained with skilful predictions of late season rainfall. In both cases, the occurrence of average (middle tercile) total rainfall does not appear to be useful since average inflows only occur on about 33% of these occasions (i.e. the same fraction that could be expected by chance).

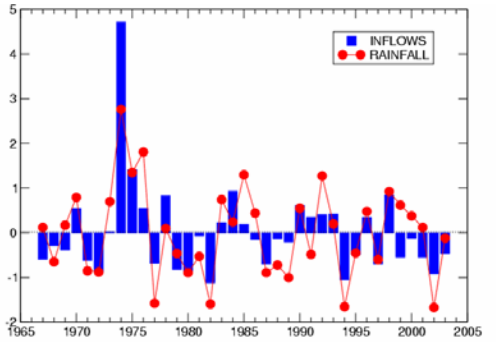

Figure 4. As for Figure 2 except for August to October.

3. Models

Here we asses the performance of rainfall hindcasts (i.e. retrospective predictions) from a suite of 8 dynamical models. The models are fundamentally similar insofar as they are global atmospheric and ocean models, they differ in terms of their horizontal and vertical resolution, parameterizations and methods of initialization for prediction purposes.

One of these models was developed by CSIRO and is referred to by the acronym COCA2 and is based on the CSIRO Mk3 global coupled climate model. 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., 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. For each year, the COCA2 model was initialized at four specific dates (January 1, April 1, July 1 and October 1) and each time 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. Ideally an ensemble of results would be generated for analysis but this has not been possible since the resolution of the model (331,776 atmospheric grid points and 1,142,784 oceanic grid points) and the time step employed (15 minutes) places a serious demand on computing resources.

DEMETER is the acronym of the EU-funded project entitled "Development of a European Multimodel Ensemble system for seasonal to inTERannual prediction". (DEMETER was the goddess of fertility in ancient Greece and in classical Rome became Ceres). 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. For each year, the DEMETER hindcasts were initialized on four specific dates (February1, May 1, August 1 and November 1) and each time run forward for 6 months. Unlike the COCA2 results, the DEMETER results comprise an ensemble of 9 members for each hindcast. For further details see Palmer et al. (2004).

4. Assessment of rainfall hindcasts

Given that observed rainfall and observed inflows are related, we firstly compare predicted and observed rainfall at model grid points which approximately cover the catchment region of the Burrinjuck dam. Again, the two periods analysed are May to October and August to October. Note that the number of years available for comparison vary from model to model. Secondly, we also compare predicted rainfall with observed inflows. In this case inflow data is only available from 1967 onwards and so the number of years available for comparison is less than that for rainfall.

The results are tabulated in Table 1 which lists each model, the hindcast years used in the analysis, and then the percentage of successes in each of three tercile categories (BA=below average, A=average, AA=above average). The columns refer to the different periods and the two rows of results for each model represent the comparisons with observed rainfall and observed inflows. Note that in each case the hindcasts and observations are ranked over the period for which hindcasts are available. This means that there are equal (or near to equal) numbers in each of the three categories. Therefore, the “no-skill” percentage value is 33% and values close to this indicate little or no skill. Values in excess of this value indicate potential skill, but may not be significant due to the relatively small sample size involved. For example, the COCA2 sample size is only 24 for rainfall corresponding to 8 years in each of the three categories. The success rates in each category can therefore be regarded as uncertain by about 13 % (i.e. 1 out of 8) and so cannot be regarded as certain as the METEO model values (total sample size = 44, uncertainty in each category ~7%).

Table 1. The skill of the different models as measured by the percentage matching of hindcast rainfall categories with observed rainfall categories and observed inflow categories. The two hindcast periods are May to October and August to October. Values less than 40% in smaller font, values greater or equal to 50% are bolded.

	May to October:			August to October:		
		Rainfall Inflows			Rainfall Inflows	
	Below Average	Average	Above Average	Below Average	Average	Above Average
UKMO						
1959-2001	**51**	38	42	46	41	45
1967-2001	47	34	25	**53**	34	44
MPI						
1969-2001	45	29	38	45	30	43
1969-2001	40	29	36	47	34	42
METEO						
1958-2001	45	32	41	41	30	47
1967-2001	46	33	47	**53**	35	46
LODYC						
1974-2001	49	39	35	38	41	37
1974-2001	44	36	40	47	41	36
CERFACS						
1958-2001	47	39	37	40	49	49
1967-2001	48	38	43	40	38	35
INGV						
1973-2001	**53**	36	44	49	37	**53**
1973-2001	**53**	33	44	50	27	34
ECMWF						
1958-2001	49	33	44	37	30	43
1967-2001	40	34	45	47	31	48
COCA2						
1980-2003	**51**	38	42	**63**	38	38
1980-2001	38	25	**50**	**63**	38	**50**

The acronyms for the models refer to the DEMETER modelling partners: CERFACS (European Centre for Research and Advanced Training in Scientific Computation, France), ECMWF (European Centre for Medium-Range Weather Forecasts, International Organization), INGV (Istituto Nazionale de Geofisica e Vulcanologia, Italy), LODYC (Laboratoire d'Océanographie Dynamique et de Climatologie, France), METEO (Centre National de Recherches Météorologiques, Météo-France, France), UKMO (The Met Office, UK) and MPI (Max-Planck Institut für Meteorologie, Germany). COCA2 refers to the CSIRO model.

The fact that different sample sizes and different time periods are involved makes it very difficult to discriminate between the performances of the different models. However, it is clear that there exists a consistent pattern across all the model results corresponding to the presence of potential skill in the extreme categories but none in the average category (when the success rate is not very different to the 33% expected by chance). This is understandable if the models are capable of predicting the extreme phases of ENSO (i.e. El Nino and La Nina events) when the effect on rainfall is strongest, compared to other times when the signal is weak and rainfall tends to be controlled by other factors.

If there exists some potential skill at predicting the 6-month seasonal rainfall, it could be expected that there should be greater skill at predicting the 3-month seasonal rainfall – simply because the lead time is less and it could be expected that skill would be higher during the later part of the year compared to autumn. However, the results do not indicate any such improvement. Furthermore, given the existence of potential skill at predicting catchment rainfall, it could be expected that some of this might be apparent, albeit degraded, in predictions of inflows using Figures 2 and 3 as a guide. It can be seen that while the skill often deteriorates, in some cases it improves. There is no obvious reason why this should be the case other than the effect of uncertainty due to the different sample sizes.

Given the degree of consistency in the results shown in Table 1, an approximate measure of overall performance of the models can be obtained by simply averaging the values, as shown in Table 2. In summary:

- potential skill is apparent in predicting rainfall extremes, possibly with more skill associated with below average conditions than above average conditions. This may suggest the models are better at predicting El Nino events (which are associated with below average Australian rainfall) compared to La Nina events, but has not been investigated here
- there is no evidence of skill at predicting average rainfall conditions
- the skill at predicting 3-month late-season rainfall is not greater than the skill at predicting the 6-month rainfall earlier in the year
- skill at predicting catchment rainfall translates into potential skill at predicting inflows, although the values are not large

Table 2. Arithmetic average of the results of Table 1. Values less than 40% in smaller font, values greater or equal to 50% are bolded.

	Below Average	Average	Above Average
May to October rainfall	49	36	40
May to October inflows	45	33	41
August to October rainfall	45	37	45
August to October inflows	**50**	35	42

To put these overall results into perspective, consider that over 100 years there will be close to 67 extreme (wet or dry) years. Chance, or guesswork would successfully predict only 22 of these (with 44 failures) yet the model-based predictions appear capable of successfully predicting about 30 (with 36 failures). Alternatively, if average inflows were assumed each and every year ("climatology"), these would be correct about 33 times (with 67 failures) compared to the model-based predictions which, taking all years into account, would be successful about 43 times. i.e. the model-based predictions would be successful 4 years out of 10 compared to climatology, which would only be successful 3 years out of 10.

While only moderate, this apparent skill may, in fact, be far less useful for a number of reasons. Firstly, inflows only determine the total storage of any dam while it is less than 100% full. If this is achieved early in the year, then it is likely to remain at 100% independent of the winter/spring season inflows. Thus, for those occasions when the dam is close to capacity, seasonal predictions of rainfall/inflows are of negligible value. The reverse situation is also relevant here. When storages are extremely low at the start of the season, then a prediction of average inflows may have very little value if these are insufficient to boost total storages. In other words, depending on the rate of recharge, a near-empty dam may never reach total capacity in one season. In this situation, predictions of inflows can be of limited value since the end-of-season total storage may already be limited.

For those occasions when a seasonal prediction can be valuable, it is important to know the costs and benefits associated with a correct prediction compared to an incorrect prediction. The conservative strategy of assuming average inflows/rainfall, will be correct 33% of the time and, when incorrect, will only be incorrect by one category (e.g. predicted average and below average occurred). Any other scheme which attempts to predict extreme values can suffer from two-category errors (e.g. predicted above average and below average occurred). If management decisions are based on these predictions then, over time, the gains/profits made when the predictions are correct could be swamped by the losses incurred on the relatively few occasions they suffer from two-category errors. It is quite possible that a predictive scheme can be objectively quantified as more skilful according to a number of measures, but may actually be less valuable than a simple strategy based on climatology. Quantifying this is difficult since it depends very much on how the end user reacts to predictive information and the associated costs and benefits.

Finally, it also needs to be recognised that streamflow or reservoir inflow predictions can be better obtained via a calibrated hydrological model and downscaled rainfall predictions. The method described here is relatively simple but sufficient, it is believed, to broadly quantify potential skill.

5. Summary and Conclusions

A suite of state-of-the-art seasonal dynamical models have been assessed with respect to how well they can predict rainfall (and inflows) for a catchment region within south-east Australia. The major findings are:

- It is not possible to distinguish between the performance of the different models due to different sample sizes and periods for which hindcasts are available.
- Overall, the models exhibit an ability to capture, to some extent, variations in seasonal rainfall associated with ENSO events and this is evident in the fact that they exhibit skill in the extreme categories but not in the average category.
- The average success rate, while greater than that expected by chance or the strategy which assumes climatology, is not high and is expressed in the slight shifts in the probabilities for below average and above average tercile categories. As a rough guide, the model-based predictions provide an advantage over climatology 1 year in every 10.
- Taking into account the fact that rainfall and inflow predictions can be somewhat redundant when dealing with water storages, this may overestimate the potential utility to end –users.
- Finally, it has to be recognised that the economic value of predictions, no matter how skilful, can be diminished according to the costs/benefits associated with decisions made by the end-user. Assessing value is a more task which needs to be done on a case-by-case basis.

6. Acknowledgements

This study has been funded under the CSIRO Water for A Healthy Country Flagship Program.

7. References

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.

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. *Journal of Climate*, 13, 139–163.

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.

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