South Eastern Australian Climate initiative

Final report for Project #3.2.7

Implementing Physical-Statistical Climate Models

Principal Investigator: Edward P. Campbell

CSIRO Mathematical & Information Sciences
Private Bag 5, PO Wembley, WA 6913
e-mail: eddy.campbell@csiro.au
tel: (08) 9333 6203

Co-Authors: Mark J. Palmer

CSIRO Mathematical & Information Sciences
Private Bag 5, PO Wembley, WA 6913
e-mail: mark.palmer@csiro.au
tel: (08) 9333 6293

Completed: 17 January 2008

GRDC Region: South-eastern corner of the Murray-Darling Basin

Abstract

The objective of this project is to implement a prototype physical-statistical model, developed in project 1.5.5, to proof-of-concept stage linking climate output variables, such as rainfall and temperature, via a physically-based process model for ENSO and empirical predictors.

We have successfully implemented the approach, developing a Monte Carlo-based algorithm for processing streaming data. This yields an estimate of the form of the required predictive distribution, which is summarised using a so-called adaptive Metropolis algorithm, now often used for Bayesian statistics. This is very computationally demanding, but there is considerable scope through parallel processing to speed up the execution of the algorithm. The exciting feature of this approach is that we can integrate physical and empirical models to extract the best of both using this method.

In our case study we found that monthly temperature is a regular periodic series, so we focused on monthly rainfall as a more difficult application. ENSO alone did not have very strong predictive ability, so we retained time-coincident empirical predictors as proxies for an additional system process describing regional pressure characteristics. These predictors are essentially mean sea level pressure and a range of geo-potential height variables. Further work to develop a process model capable of representing these variables, coupled with the ENSO model, is likely to be beneficial.

Significant research highlights, breakthroughs and snapshots

1. As far as we can see in the literature this is the first time that a fully integrated physical-empirical seasonal forecasting model has been developed. This puts in place a new set of techniques for expanding forecast lead times based on this prototype.
2. Forecasts are probabilistic in nature, providing a risk-informed forecast product.
3. Encouraging results have been found in a case study, and there are clear paths to improvement.
4. The approach has been designed to be modular, so as system knowledge grows it may easily be incorporated.
5. The technical aspects of the algorithm have been implemented to require minimal user intervention; being adaptive in nature it learns from the data.

Statement of results, their interpretation, and practical significance against each objective

> **Objective 1:** Implement and assess a prototype physical-statistical model for forecasting rainfall and temperature using ENSO and other physical drivers.

The results obtained are discussed below against each of the project's three milestones. A detailed description is provided by Campbell and Palmer (2008).

1. *Complete assessment for non-streaming data.*

It was considered unlikely that simple multiple linear regression models would be effective in this type of modelling, so a more flexible approach using Generalized Additive Models (GAMs- Wood, 2006) was implemented. We found temperature to be a regular periodic series, so we focused our attention on monthly rainfall as a more difficult task. The key lag for ENSO variables was at 5 months; Nino3 (mid-Pacific sea surface temperature anomaly) had a particularly strong relationship with monthly rainfall along with interactions involving thermocline depth.

2. *Complete assessment for streaming data and document results.*

We first had to implement the model from project 1.5.5, which developed a number of implementation options. So far as we can see the method implementation in project 3.2.7 (described briefly in the next section) has not been published elsewhere, and provides a direct means to integrate an empirical forecast model with a physically-based process model. The evidence is that the forecast skill for ENSO alone is quite limited and perhaps stronger in ENSO years.

3. *Assess skill improvement from additional physical drivers.*

It seems clear that a physical process describing features of regional pressure characteristics is required. We used the pressure variables derived by Charles (2007), but could not find skill in lagged versions of these variables, only when time-coincident. Thus a dynamical forecast capability is required, coupled with the ENSO process. We retained the time-coincident variables as a proxy for this physical process, which indicated good forecasts may be possible. In addition we have also informally investigated the use of an Indian Ocean Dipole process to assist forecast skill. Preliminary indications suggest this process may be influential in at least some years.

Summary of methods and modifications (with reasons)

Method and Algorithm: The implementation is summarised by equation (3) of Campbell and Palmer (2008), which is closely related to the one-step forecast equation used in data assimilation. The last term is the one-step forecast distribution for the system, which drives the seasonal forecast. The integration with the empirical model in this prototype is achieved by the middle term, the empirical model linking predictors to the climate output. The empirical terms are held fixed at the current time step, and the forecast distribution for the system is used to provide a trajectory for the climate output. The first term is an observation model for the process variables. Campbell and Palmer (2008) go on to modify an established particle filter algorithm (Kitagawa and Sato, 2001 pp191) to calculate the forecast distribution.

A path to improvement here is to add detail to the middle term, particularly the development of a 1-step prediction model for the empirical predictors. The Bayesian approach makes it quite straightforward to incorporate this additional detail. Empirical predictors could also be replaced by incorporating further physical variables as process understanding grows.

Software: We have implemented the algorithm in the R statistical computing environment (http://cran.au.r-project.org/) to facilitate uptake. The software is designed to be modular to enable alternative process models or updates to be incorporated.

Case Study: We collaborated with Dr Steven Charles who led SEACI projects 1.3.2 and 1.3.4 and used the same station list being employed for statistical downscaling studies. We then sought the advice of the Bureau of Meteorology through Bertrand Timbal (SEACI Theme Leader) on the quality of the stations, which left the 3 stations listed in Table 1 below.

In general the fits obtained for the non-streaming case were moderate-to-mediocre, and as a summary the values of the R^2- adjusted statistic are shown in Table 2 below. Arguably the best fit was obtained for station 75012, where *a priori* we might have expected this for station 88060 as the most North-easterly station. For station 75012 we see a strong main effect due to the North-South mean sea level pressure (MSLP-time coincident) gradient as an empirical predictor. There appears to be no skill in these variables however when lagged for forecasting purposes. The key system variables are a main effect due to Nino3 and interactions involving thermocline depth (east and west Pacific) and thermocline depth (west Pacific) and Nino3 respectively. All of these terms are highly significant at a lag of 5 months. The highly significant main effect due to Nino3 is common to all three models, but the significance of the interaction terms varies with location. We note that Indo-Pacific thermocline depth has been used to forecast streamflow in Australia by Ruiz *et al.* (2006, 2007).

Table 1 Meteorological observing stations used in the case study

Station Number	Station Name	Latitude (degrees)	Longitude (degrees)	Elevation (m)
70028	Yass (Derringullen)	-34.7419	148.8895	595
75012	Wakool (Calimo)	-35.4217	144.5983	84
88060	Wallaby Creek Weir	-37.4489	145.2144	520

Table 2 Values of R^2- adjusted from the empirical (non-streaming data) prediction models.

Station	R^2- adjusted
70028	0.217
75012	0.405
88060	0.348

We have run the forecast model for each from April to October in each of the years 1981 to 2006, and estimated the forecast and analysis distributions. The figures described below show the forecast mean (solid profile) and 95% credibility limits (dashed lines). The full set of figures is provided by Campbell and Palmer (2008).

Overview of the Analysis Results: In almost all cases the analysis tracks the observed state variables very closely, although there are some cases of ‘drift’. Examples include 1981 (Nino3) and 1984 (Nino3); we used published parameter estimates for the system model, and this drift problem could be improved by estimating the parameters specifically for the case study data.

Overview of the Prediction Results: As it was clear that ENSO alone explains a relatively small part of the variation in the rainfall data we limit our summary to a description of the results obtained rather than detailed quantitative summaries.

First, the results are very computationally intensive to obtain- the set shown here required 268 hours of CPU time. It is likely that the summaries retain a significant noise component due to Monte Carlo error, so the credibility intervals are indicative only. High confidence can be attached to the mean profiles found. There is considerable room for speed improvement to remedy this, and the results suggest that this would be worthwhile if a suitably predictive physical process model is found. Our exploratory work suggests that a process model incorporating mean sea level pressure fields and geo-potential height as state variables would be attractive. We have also undertaken preliminary work on the Indian Ocean Dipole series, and there are some promising indications, supported by the work of Ruiz *et al.* (2006, 2007). We break our discussion down by ENSO events in the case study period, using the years listed on the Bureau of Meteorology's web site.

El Niño Years

1982: The rainfall patterns are followed quite closely, especially well for station 75012 (). The analysis tracks the observed state variables very closely.

1987: The mean profiles are quite accurate in all cases, especially for station 70028; the analysis tracks well with a little bit of drift within the credibility interval.

1991: The forecasts are very good for station 88060 (Figure 2) and very effective for the other two stations, with the exception of June. The analysis tracks well, although there is a persistent bias present, especially for the Nino3 analysis.

1993: The forecast mean profiles are accurate in response shape with a few exceptions, but are biased in the case of station 88060 in particular. The analysis drifts for Nino3 in particular.

1994: The analysis tracks well; perhaps the most striking feature is spectacular over-prediction for August to October at stations 70028 (Figure 3) and 88060.

1997: Stations 75012 and 88060 are forecast quite well; the analysis tracks well.

2002: Overall there is consistent over-prediction, especially for station 75012 although the pattern of the mean profile is largely correct. The analysis tracks well.

La Niña Years

1988: Good reproductions are found for each profile, with a notable exception for June at station 75012 where a large over-prediction is made. The highest rainfall was observed at station 88060 and this was well forecast, including a sustained relatively high rainfall around this month.

1998: The forecasts are accurate in all cases, except April at station 88060. The analysis tracks well. []

Non-ENSO years

Performance here seems to be very similar, which could be driven by the time-coincident pressure variables that were left in the model to be proxies for a pressure system process. There are examples of poor analyses (1981 and 1984) and biases (various).

Years 2005 & 2006

These years were not used for model-fitting. In both cases the analysis tracks well and the mean forecast profiles are quite accurate.

Summary of links to other projects

The links for this project are focused on downscaling activities within SEACI, principally projects 1.4.2 and 2.1.3.

- The core methodology for this project was developed in project 1.5.5, and a technical report describing the work completed is about to be reviewed.
- Project 1.3.3 (Atmospheric Predictor Selection for Statistical Downscaling)

- Project 2.1.2 (Extraction and Assessment of Statistical Downscaling Predictors)

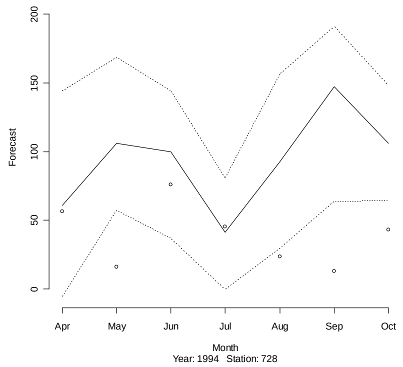

Figure 2 Forecast results for 1991, station 88060.

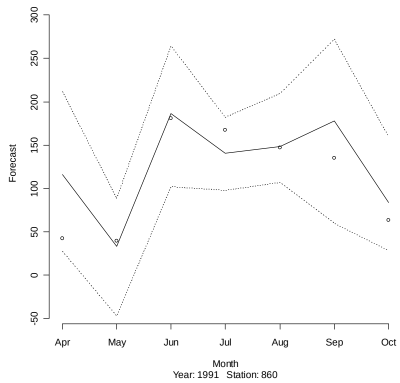

Figure 1 Forecast results from 1994, station 70028.

Publications arising from this project

Technical report: Campbell and Palmer (2008)
Manuscript in preparation: Campbell, E. P. and Palmer, M. J. (2008). Physical-Statistical Seasonal Forecasting. *To be submitted to Journal of Climate*. This manuscript will be submitted for approval to the SEACI working group when it enters the internal review process.

Acknowledgement: This project was funded by the South East Australian Climate Initiative. We are grateful to Steven Charles and Bertrand Timbal for their assistance.

Recommendations for changes to work plan from your original table

An amendment was notified in April 2007; no further changes were required.

References

Berliner, L. M. (2003). Physical-statistical modeling in geophysics. *Journal of Geophysical Research- Atmospheres*, 108 (D24), doi:10.1029/2002JD002865.

Campbell, E. P. and Palmer, M. J. (2008). Year 2 Technical Report on South-East Australia Climate Initiative (SEACI) Project 3.2.7: Implementing physical-statistical climate models.. CSIRO Mathematical and Information Sciences, Technical Report 08/08, 8 January 2008.

Charles, S. (2007) SEACI End of project report: 1.3.4: Calibration of Statistical Downscaling Models.

Doucet, A., de Freitas, N. and Gordon, N. (Eds.) (2001). *Sequential Monte Carlo Methods in Practice*, Springer, New York.

Kitagawa, G. and Sato, S. Monte Carlo Smoothing and Self-Organising State-Space Model. Chapter 9 In: Doucet *et al.* (2001).

Ruiz, J. E., I. Cordery and A. Sharma (2006). “Impact of mid-Pacific Ocean thermocline on the prediction of Australian rainfall.” *J. Hydrol.*, 317, 1-2, 104-122.

Ruiz, J. E., I. Cordery and A. Sharma (2007). “Forecasting streamflows in Australia using the tropical Indo-Pacific thermocline as predictor.” *J. Hydrol.*, 341, 156-164.

Wood, S.N. (2006). *Generalized Additive Models: An Introduction with R.* Chapman and Hall/CRC, Texts in Statistical Science Series.

Project Milestone Reporting Table

Milestone description[1] (one line) (up to 33% of project activity)	Performance indicators[2] (1- 3 dot points)	Completion date[3] xx/xx/xxxx	Budget[4] for Milestone ($)	Progress[5] (1- 3 dot points)	Recommended changes to workplan[6] (1- 3 dot points)
1. Complete assessment for non-streaming data.	- Full database compiled. - Software assessed against full database. - Implement algorithm for non-streaming data. - Commence skill testing. - Identify additional non-ENSO drivers.	30/4/2007	-	- Data base compiled - Delay to staff availability has prevented progress on subsequent activities	- Move completion date of Milestone 1 to 30/7/2007 - Move completion date of Milestone 2 to 30/9/2007 - Milestone 3 is unaffected
2. Complete assessment for streaming data and document results.	- Implement algorithm for streaming data on full data base. - Complete skill testing.	30/6/2007	25k	Completed according to revised milestone.	
3. Assess skill improvement from additional physical drivers.	- Include additional non-ENSO drivers. - Assess skill improvement. - Final report complete by milestone date	31/12/2007	25k	Completed on time. Technical report to complement final project report is in review (draft supplied with final project report).	