South Eastern Australian
Climate initiative

Final report for Project 1.5.5

1.5.5 Hierarchical frameworks for physical-statistical climate models

Principal Investigator: Edward Campbell,

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

Updated: October 2007

Abstract
This project is focused on developing physical-statistical methods for seasonal climate forecasting. Our intended output is a probability distribution for the forecast quantity of interest, so we make use of a contemporary statistical technology known as Bayesian hierarchical modelling (Berliner, 2003). The essence of these methods is that, where possible, we use physically-based models for the processes concerned blended with empirical models where this is not possible. The outputs are probability distributions which we can use to summarise quantities of interest, particularly physical model parameters and forecasts of climate outputs (rainfall, temperature etc).

This project is a small method development activity, with implementation to follow in Project 3.2.7. Given this, in consultation with CSIRO Climate we decided to focus this project on forecasting climate processes strongly influenced by the El Niño-Southern Oscillation (ENSO). The principal reason for this is that ENSO is quite well understood, with an established literature on simplified physical models. We note in particular the literature on delayed oscillators (Suarez and Schopf, 1988) and recharge oscillators (Burgers *et al.*, 2005).

We consider two modes of forecasting: on-line (akin to data assimilation) and off-line (repeated use of a fixed forecast rule). Conventional Bayesian model-fitting methodologies tend to focus on Markov chain Monte Carlo (Smith and Roberts, 1993). This is an iterative technique, so is quite suitable for off-line problems. However, this may not be the case for streaming data and a methodology known as sequential Monte Carlo (Doucet *et al*., 2001) has been developed for this situation. A common approach is to use a weighted sample (so-called “particles”) from the state space of interest, which are resampled at each step of the filter to ensure only the fittest survive. These particle filters generalise readily to nonlinear models and/or non-Gaussian error structures.

A key concern in model development is proper accounting for uncertainty, and integration of uncertainty into any forecasts that we derive. This is a strength of the Bayesian approach since it is based on probabilistic representation of knowledge acquisition. We may integrate prior uncertainty (parameter values and boundary conditions), representational uncertainty (recognising that no physical model is entirely correct) and measurement error in particular.

Summary of Results
The project objective was:

- Develop and test a prototype physical-statistical model linking climate variables and physical drivers

The results obtained are discussed below against each of the project’s three milestones:

- *Review & document relevant literature with modelling options.*

A key choice to be made has been which physical processes and climate outputs to focus on. Through consultation with CSIRO Climate we decided to focus on ENSO as the key driving process, and selected the recharge oscillator model of Burgers *et al*. (2005) for the physical component. This represents the most contemporary work on simplified models for ENSO, but we have also considered the delayed oscillator model of Suarez and Schopf (1988) in particular. This latter model incorporates local nonlinear effects, which is an extension we may consider if time permits in project 3.2.7. In terms of climate outputs, we will focus on rainfall and temperature (maximum, minimum, range) in order of priority.

In practical terms, the results obtained under this milestone provide a suitable physical model to build a forecasting system around.

- *Develop prototype physical-statistical model.*

A key task under this milestone was to develop a suitable model framework to integrate the physical model component with the empirical components. The empirical components are needed to capture measurement error and to incorporate climate forcing that is not captured via a physical model. This was accomplished, and is described in the next section in more detail.

We have also extended the Burgers *et al.* model to incorporate seasonal forcing and representational error, and successfully tested our implementation for computational efficiency. We have compiled all the data needed to fit this model to observed data. The physical-statistical model also incorporates the possibility of other driving processes, which are captured via empirical relationships.

In practical terms, the results obtained under this objective allowed us to develop a physical-statistical probability model with an extended version of the Burgers *et al*. (2005) model.

- *Benchmark performance of most promising model-fitting algorithm.*

To set up Project 3.2.7 we have also explored and benchmarked[1] appropriate algorithms to implement the model. We have examined two scenarios: off-line data processing and on-line, streaming data. The latter case provides a mechanism to dynamically update the model as new data became available. In both cases we find that methods based on Markov chain Monte Carlo simulation are viable. For streaming data a further choice is the use of particle filters, which may be thought of as a general approach to data assimilation. Particle filters will likely be the only viable option if it is necessary to incorporate nonlinear/non-Gaussian features into the model.
In practical terms, the results obtained under this objective identified leading candidates for algorithms to implement the physical-statistical model that has been developed.

Summary of Methods
The Burgers *et al*. (2005) model is two-dimensional in terms of sea surface temperature anomaly and thermocline depth, and a typical realisation is shown in Figure 1 below. We see that following an initial shock the system settles down to a stable limit point.

[1] We developed benchmark algorithms via a review of the literature on computation in this field according two criteria: (i) Effectiveness of automatic implementation and (ii) Anticipated run time.

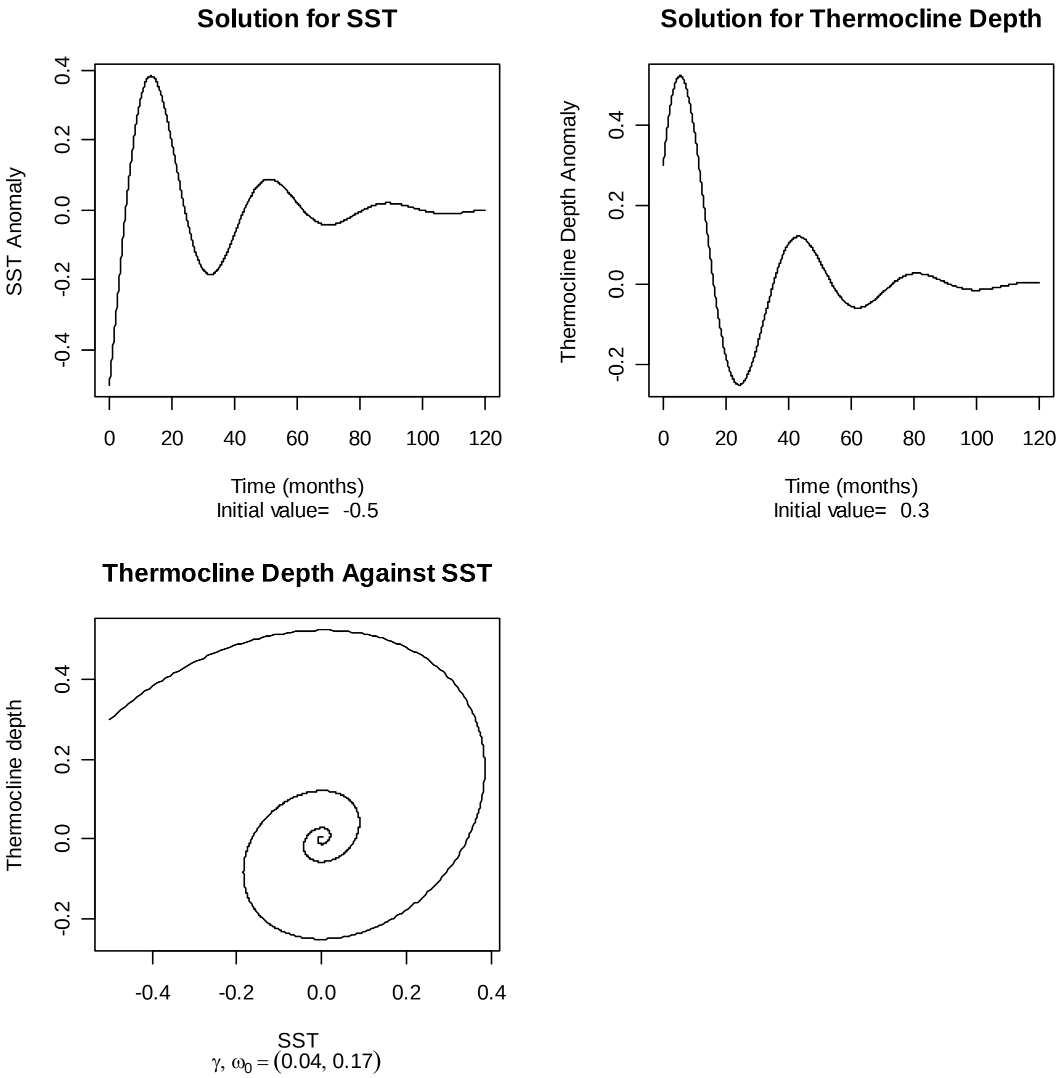

Figure 1 Realisation of the Burgers *et al.* model with starting value (-0.5, 0.3).

Nature tends to be far more complex, so we have incorporated periodic forcing and representation error into the model; a typical realisation is shown in Figure 2 below. We see now that there is no simple limit point. This physical model has been embedded in a physical-statistical model, which is depicted as a graphical model (Lauritzen and Spiegelhalter, 1988) in Figure 3.
This depicts the ENSO process (*P*) having parameters (Ψ_P) and boundary conditions (B_P); observations on ENSO are shown as *Y*, and statistical parameters arising such as measurement error as θ_Y. We then have links to the climate outputs of interest (C_1, …, C_m), which may be further influenced by a set of climate predictors (A_i) in each case. These are empirical relationships, so giving rise to a set of statistical parameters (θ_i) in each case. Using this graph it is possible to write down a probability model for all the components to use for forecasting, and a detailed description of the methods are described by Campbell (2007-attached).

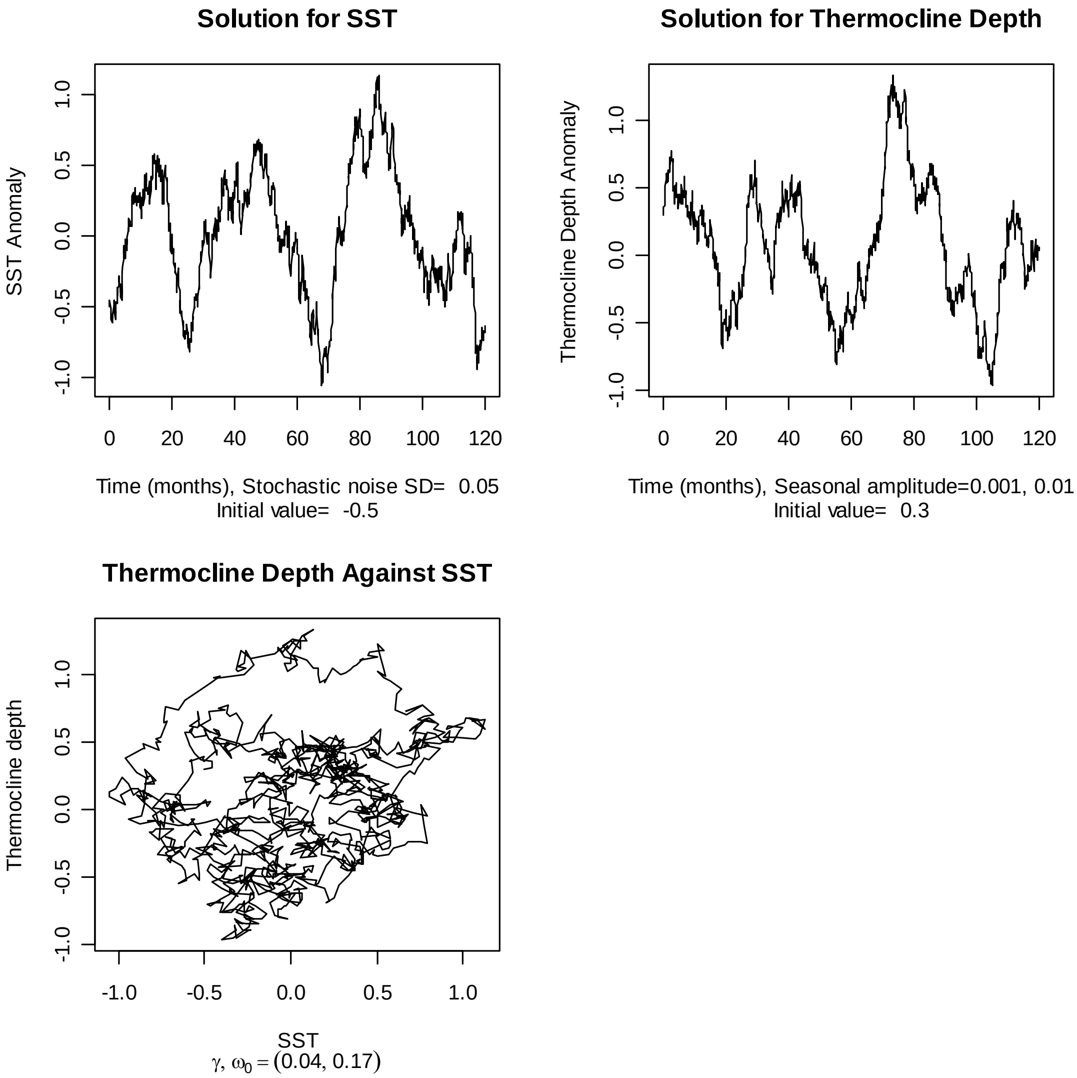

Figure 2 The Burgers *et al.* model with periodic forcing and stochastic representation error.

The final method required is for model fitting. Effective model-fitting, or calibration as it is sometimes known, is crucial to developing a skilful forecast model. In the Bayesian framework this amounts to characterising the probability distributions of uncertain quantities of interest, such as predictive distributions for forecast quantities. Expressions for these can be written down but rarely evaluated analytically, so numerical methods are required.

Simulation methods are the dominant approach to solving such problems, but we first need to consider more deeply how a forecast scheme might be applied in practice. We distinguish two cases:

1. The model is calibrated and validated, then applied essentially as-is as new data arrive. This approach is termed *off-line* processing.
2. The model is updated as new data arrive, so this approach is termed *on-line* processing. In this context the data are said to be *streaming*.

The most widely-used approach to Bayesian computation is an iterative method known as Markov chain Monte Carlo (MCMC). This can be very computationally intensive by the standards of conventional statistical methods. An alternative approach, which is commonly used with streaming data, is known as *particle filtering*. The fundamental idea of the particle filter is that whilst a particular probability distribution may be difficult to sample from, it is typically easy to evaluate values that are proportional to the probability density. We may then draw a sample of so-called particles from essentially any distribution, then use values of the density function we are interested in to resample these particles. After resampling the particles follow the distribution of interest. Details of these approaches are provided by Campbell (2005, 2007).

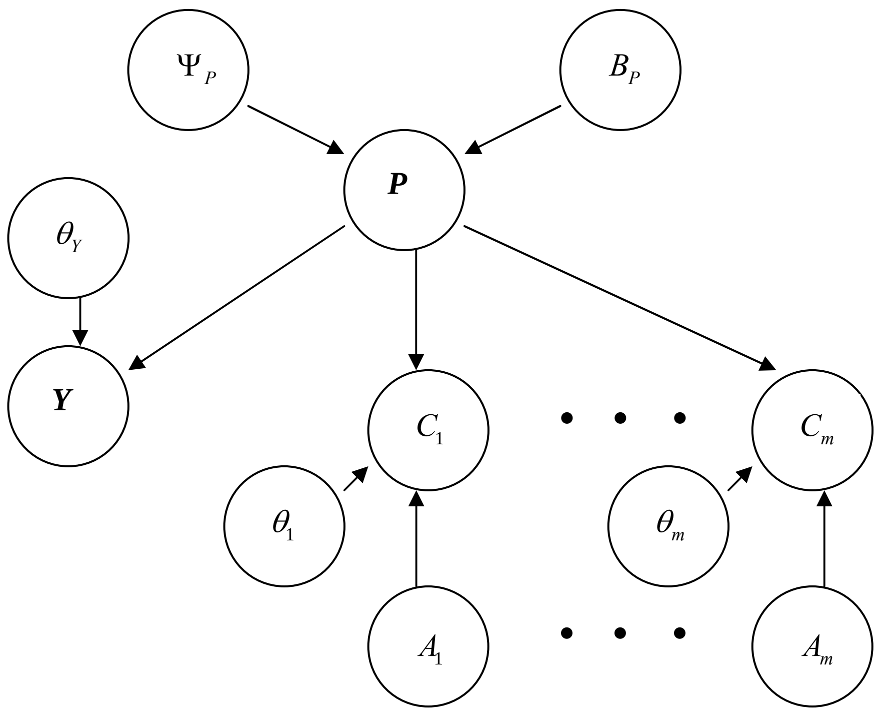

Figure 3 Graphical model displaying inter-connections between the ENSO process (*P*), observations on ENSO (*Y*) and various climate outputs ($C_1, \ldots, C_m$).

Conclusions

Project 1.5.5 had been focused on method development, with key achievements including a framework for the physical-statistical forecast model, algorithms to implement it and an extension of the physical model to be used. This model will next be used to develop and test forecasts of rainfall and temperature in the southeast Murray-Darling Basin. The objective will be to provide a proof-of-concept application via Project 3.2.7, to be concluded by December 30, 2007.

Acknowledgements: This project was partly funded by the South East Australian Climate Initiative, with additional support provided by CSIRO for statistical method development. The project leader is grateful for the input of Dr Bryson Bates in helping to frame the scope of this work.

References

Full project technical report:

Campbell, E. P. (2007). Year 1 Technical Report on South-East Australia Climate Initiative (SEACI) Project 1.5.5: Hierarchical frameworks for physical-statistical climate models. CSIRO Mathematical and Information Sciences, Technical Report 07/18, 6 February 2007.

Scientific Literature

Berliner, L. M. (2003). Physical-statistical modelling in geophysics. *Journal of Geophysical Research- Atmospheres,* 108 (D24).

Burgers, G., Jin, F. F. and van Oldenborgh, G. (2005). The simplest ENSO recharge oscillator. *Geophys. Res. Lett.,* 32**,** doi:10.1029/2005GL022951.

Campbell, E. P. (2005). Statistical modelling in nonlinear systems. *J. Clim.,* **18,** 3099-3110.

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

Lauritzen, S. L. and Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *J. Roy. Statist. Soc.* **B**, 50, 157-224 (incl. discussion).

Smith, A. F. M. and G. O. Roberts (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods, *J. Roy. Statist. Soc., B, 55*(1), 3-23.

Suarez, M. J. and Schopf, P. S. (1988). A delayed action oscillator for ENSO. *Journal of the Atmospheric Sciences,* 45**,** 3283-3287.

Project Milestone Reporting Table

To be completed prior to commencing the project				Completed at each Milestone date	
Milestone description[1] (brief) (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 & document relevant literature with modelling options.	- Document climate variables of interest. - Document relevant physical processes and candidate conceptual models. - Select a conceptual model to progress.	30/05/2006	15k	• Initial consultation suggested a focus on ENSO. • Review of literature on conceptual models for ENSO is complete. • Have selected a conceptual model to progress.	Milestone Complete
2. Develop proto-type physical-statistical model.	- Establish and document hierarchical model.	30/09/2006	15k	• A model framework has been developed. • A draft technical report is in progress is about to be submitted for peer review.	Milestone Complete
3. Benchmark performance of most promising model-fitting algorithm.	- Select test data. - Identify candidate algorithms. - Benchmark most promising algorithm. - Document as a report.	31/12/2006	20k	• We have selected algorithms appropriate for streaming and non-streaming data and benchmarked them for computational efficiency. • Draft technical report is to be submitted for peer review imminently.	Milestone Complete.

Project Outputs: Methodology technical report and software to test model-fitting algorithms.

Appendix: Data Compiled for this Project

Data sources for the model fitting and ENSO forecasting study of Burgers *et al.* (2005).

Model Variable	Data Set
T_E	Observed NCEP Niño3 index
τ	Average of the FSU objective pseudo wind stress
Thermocline depth	BMRC data set of the 20° isotherm depth: h_W - Average over 130°E-170°E h_E - Average over 150°W-90°W h- Average over 130°E-80°W