Year 1 Technical Report on South-East Australia Climate Initiative (SEACI) Project 1.5.5: Hierarchical frameworks for physical-statistical climate models.

Edward Campbell

Technical Report 07/18

6 February 2007

Table of Contents:

Executive Summary

Project Description: Physical-statistical models using Bayesian hierarchical methods are a recent innovation. They combine both physical and statistical elements in one model. Building on these ideas we will develop a hierarchical framework linking suitable physical processes, such as the El Niño-Southern Oscillation (ENSO), to rainfall and temperature in the region.

Project Methodology:

- Determine climate variables of interest to stakeholders, along with potential predictors and key physical processes;
- Use Bayesian hierarchical modelling to develop a framework to link a physical process model with a statistical model for the climate observations of interest. This will allow us ultimately to produce probabilistic forecasts for climate variables of interest that integrate physical knowledge with observations. Given the small size of this project, we will be limited to using conceptual models of climate physics. The so-called physical-statistical model developed will be implemented using contemporary methods of Bayesian computation, based on particle filters.
- Compare a variety of such approaches and benchmark available algorithms for computational efficiency and ease of use.
- Select the most promising approach and document.

Milestones for Year 1:

1. Complete & document relevant literature with modelling options.
2. Develop proto-type physical-statistical model.
3. Benchmark performance of most promising model-fitting algorithm.

Outlook for Year 2:

Year 2 will be focused on software implementation and investigation of potential forecast skill.

Summary of Progress:

The objective of this project is to develop and test a prototype physical-statistical model linking climate variables and physical drivers, to be implemented in project 3.2.7. Physical-statistical models using Bayesian hierarchical methods are a recent innovation. Hierarchical modelling is based on the fact that the joint probability distribution of a collection of random variables can be decomposed into a series of simpler conditional probability models. It combines both physical and statistical elements, and expert knowledge or judgement in one model (Berliner, 2003).

A key choice to be made has been which physical processes 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. We have extended this model to incorporate seasonal forcing and measurement 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.

To set up project 3.2.7 we have also explored and benchmarked 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.

1 Introduction

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 plan to 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 framework adopted is probabilistic in nature, providing a natural means to integrate and quantify all sources of uncertainty. 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 very small with some risk attached as this is a method development activity. Resources are therefore focused on proof-of-concept, 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 a well established literature on physically-based models for the phenomenon. Indeed, there is a significant literature on relatively simple models that capture the essential features of ENSO. We note in particular the literature on delayed oscillators (Suarez and Schopf, 1988) and recharge oscillators (Burgers *et al.*, 2005), the models for which can be written as quite simple equations. Despite this simplicity of representation these models are capable of considerable complexity when realised.

We consider two modes of forecasting:

1. Calibrating a forecasting model and then applying it as new data arise, which we term off-line processing;
2. Running the model in on-line mode with streaming data, much as in a data assimilation problem.

Conventional Bayesian model-fitting methodologies tend to focus on Markov chain Monte Carlo (Robert and Casella, 2004). 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. The algorithms developed in this field are known by a number of names, but perhaps most often as particle filters. Particle filters use a weighted sample (the so-called "particles") from the state space of interest, which are resampled at each step of the filter to ensure only the fittest survive. Particle filters generalise readily to nonlinear models and/or non-Gaussian error structures, so we explore these methods in particular.

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 identify the following key types of uncertainty:

1. Prior uncertainty captured using expert knowledge regarding physical parameter values and boundary conditions;
2. Representational uncertainty, recognising that no physical model is entirely correct;

3. Measurement error, recognising that we cannot take perfect observations of any physical system. In many geophysical applications this source of uncertainty can be substantial.

We account for each of these uncertainty sources in the model we develop.

The remainder of this report is structured as follows. In the next section we describe the stakeholder consultation we have undertaken and the scope of the project. In section 3 we review the literature on models for ENSO that capture its essential features. In section 4 we develop a physical-statistical model, and scope its essential components. In section 5 we identify and benchmark algorithms for fitting this model. The final section provides a discussion of findings and some conclusions.

2 Stakeholder Consultation

Based on consultations with CSIRO Climate, the primary focus in order of priority will be:

- Rainfall
- Temperature
    - Maximum
    - Minimum
    - Range

Given the small size of this project (0.8 FTE total effort over 2 years) we will focus attention on processes influenced by the El Niño-Southern Oscillation (ENSO). The objective of this project is to develop a physical-statistical model for seasonal forecasting purposes to be implemented in project 3.2.7. Our effort is focused on developing a suitable demonstrator of the proposed approach.

3 Models for ENSO

Two well known models for the fundamental ENSO mechanism are the delayed oscillator of Suarez and Schopf (1988) and the recharge oscillator of Jin (1996, 1997). Both approaches yield relatively simple differential equation models that we can use to assist in forecasting. A key paper exploring a delayed oscillator mechanism from a physical model perspective is Battisti and Hirst (1989).

Suarez and Schopf (1988) proposed a nonlinear delayed oscillator model for ENSO, which we refer to here as the S&S model. This model sought to explain the periodic behaviour of ENSO exhibited in simple circulation models (e.g., Cane and Zebiak, 1985). The S&S model conceives of ENSO arising as linear growth of some initial disturbance, which is tempered by local nonlinear effects, such as advective processes in the ocean and moist processes. These local nonlinear effects are approximated using a cubic term, so on a suitable scale the model may be written as:

$$\dot{f}(t) = f - f^3, \tag{1}$$

where f represents the magnitude of the growing disturbance. The S&S model now invokes the argument that the growing disturbance generates (westward propagating) Rossby waves, which reflect as (eastward propagating) Kelvin waves. These Kelvin waves act as a delayed damping effect, so the model may be extended to:

$$\dot{f}(t) = f - f^3 - \alpha f(t-\delta). \tag{2}$$

Here δ is a non-dimensional delay term and α measures the influence of the returning wave relative to damping effects.

Note however that the S&S model does not incorporate periodic forcing, as was done by Minobe and Jin (2004):

$$\dot{f}(t) = f - f^3 - \alpha f(t-\delta) + A_f \cos\left(2\pi t / T_f\right). \tag{3}$$

We refer to this as the M&J model, which was shown by the authors to demonstrate threshold behaviour arising from positive feedback in the model. Their results imply that the feedback mechanisms present in this model can have significant implications for the nature of medium to long term variability of ENSO, and so processes driven by ENSO.

The most recent contribution to the literature on recharge oscillators for ENSO is Burgers *et al.* (2005), who note that eastern Pacific sea surface temperature (SST), denoted T_E, and mean equatorial thermocline depth (h) are key variables related to ENSO. They generalise the recharge oscillator of Jin(1997), which is driven by two prognostic and two diagnostic equations:

$$\begin{aligned} \dot{h}_W(t) &= -r\left(h_W + \alpha\tau\right) \\ \dot{T}_E(t) &= -\varepsilon_1\left(T_E - \gamma_h h_E\right) \\ \tau &= bT_E \\ h_E &= h_W + \tau. \end{aligned} \tag{4}$$

Here h_W denotes the western pacific thermocline depth anomaly; h_E denotes eastern Pacific thermocline depth anomaly; τ denotes the central Pacific zonal wind stress anomaly; T_E denotes the eastern Pacific SST anomaly.

The first equation gives the collective response of the western Pacific to wind stress changes through Kelvin waves, Rossby waves and western boundary reflection. The last equation states that the thermocline tilt reacts essentially instantaneously to wind stress. We do not reproduce here the full argument made by Burgers *et al.* (2005), which leads to a simple linear system of differential equations:

$$\begin{aligned} \dot{X}(t) &= AX(t), \\ \text{where } X &= \left(T_E,\, h\right)'. \end{aligned} \tag{5}$$

The coefficients of the matrix A are easily obtained in terms of the parameters of the model. Burgers *et al.* (2005) also considered the fit of this model to available data, and its capacity to forecast ENSO. The data they used for their study are described in Table 1.

Based on the results of their empirical study, Burgers *et al.* (2005) propose setting

$$A = \begin{pmatrix} -2\gamma & \omega_0 \\ -\omega_0 & 0 \end{pmatrix}. \tag{6}$$

These model equations (6) are identical to a classical damped oscillator, with momentum T_E and position h. The parameter estimates in model (6) are found to vary significantly with the seasonal cycle. We will return to this point later in the report.

Table 1 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

3.1 Which ENSO Model to Use?

We have a number of choices open to us:

1. Use the S&S model fitted separately to different periods of the annual cycle;
2. Use the M&J model to incorporate periodic forcing, the frequency of which could be estimated or be assumed to be annual;
3. The Burgers model (6) could be used, fitted to different parts of the annual cycle or periodic forcing could be introduced.

Given the relatively small amount of effort available in this project, we will focus on the Burgers model for winter season rainfall.

A typical realisation of this model is shown in Figure 1 below, and we see that the system has a stable limit point at (0, 0); the parameter values for this example are drawn from Burgers *et al.* (2005). In practice of course this pattern would be disturbed by interactions with other physical processes, and in our forecast scheme we will essentially be assimilating current data and using this model to maintain a forecast of the climate outputs of interest.

4 Developing a Physical-Statistical Model

An exciting recent development is the use of Bayesian hierarchical methods to develop hybrid physical-statistical models, which provide for a sophisticated combination of physical and statistical modelling. The idea driving these methods is that there are many sources of information available to aid understanding of physical systems. We may make use of observations of various kinds, as well as models of various sub-systems. The Bayesian hierarchical approach allows us to integrate these sources of information, including the uncertainty in each component. For a general introduction see Wikle (2003). Some examples of applying this thinking to physical processes may be found in Berliner et al. (2003), Berliner (2003) and Berliner et al. (2000).

Suppose that we are studying a physical process P, which may be a collection of sub-processes, with physical parameters η. In observing the process P we generate data D

and so statistical parameters θ, such as a measurement error variance. We assume that all of these elements are subject to uncertainty, and seek to develop a model for the joint probability distribution denoted $[D, P, \eta, \theta]$. We may apply Bayes' theorem (Bernardo and Smith, 1994 pp 2) to factorise this joint probability model as

$$[D, P, \eta, \theta] = \left[D \middle| P, \eta, \theta\right]\left[P \middle| \eta, \theta\right][\eta, \theta]. \quad (7)$$

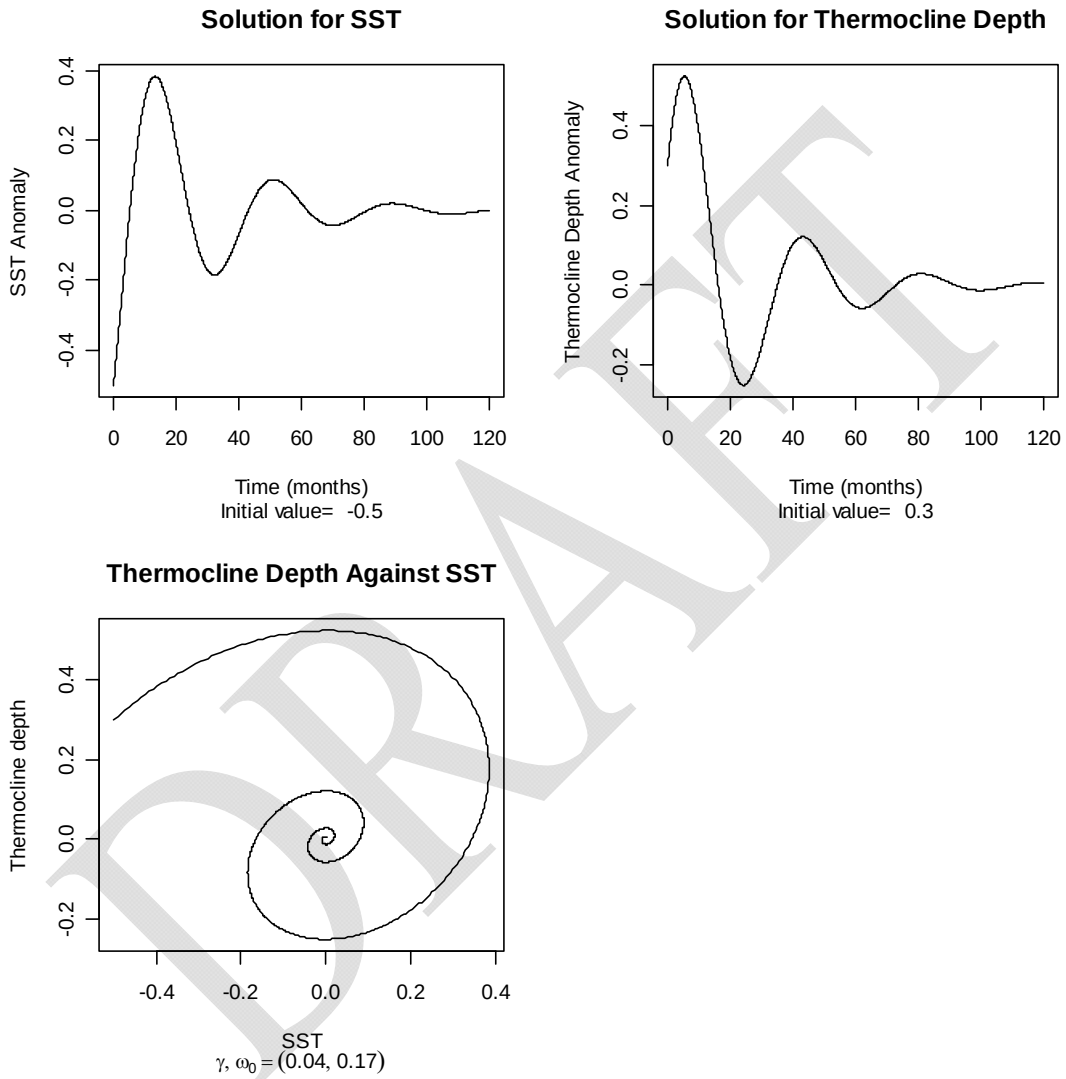

Figure 1 Realisation of the Burgers *et al.* model with starting value (-0.5, 0.3).

We may now make some modelling assumptions. In the first term, conditional on P and θ there is no further information in the physical parameters η about the data D. Similarly for the second term, given η there is no further information in the statistical parameters θ on the physical process P. We may therefore simplify (7) to

$$[D, P, \eta, \theta] = \left[D \middle| P, \theta\right]\left[P \middle| \eta\right][\eta, \theta]. \quad (8)$$

We see that the joint probability model is the product of a model for the data, a process model and a prior parameters model. The prior parameters model captures

available information on the parameters before the data are collected. For more details see Berliner (2003). A key point to note about this so-called physical-statistical model is the interconnection between the data and process models. The physical and statistical components are coupled by conditioning the data model on the physical process *P*.

It can be shown that the distribution of the process and parameters conditional on the data, the so-called posterior distribution, is such that

$$[\eta, P, \theta | D] \propto [D | P, \theta][P | \eta][\eta, \theta]. \quad (9)$$

In this way we can learn about the physical parameters through observation.

Algorithms for fitting physical-statistical models represent an active area of research. Campbell (2005) uses the importance sampling Monte Carlo approach of Berliner et al. (2003). This requires us to generate a relatively small ensemble from the prior parameters model, and pass each member of the ensemble through the physical process model. This physical process ensemble is then resampled so that a much larger sample drawn approximately from the posterior distribution (9) is obtained. This is done by assigning probabilities to each member of the ensemble, calculated using the observed data, and then sampling from them with replacement. Ensembles close to the observed data will be assigned a relatively high probability.

In this project we are seeking to develop a physical-statistical model for climate variables that are influenced by ENSO. A suitable graphical model (Lauritzen and Spiegelhalter, 1988) is shown in Figure 2 below, with quantities defined in Table 2. A graphical model is a way of picturing a probability model. In a graphical model of this type the probability distribution of each node is independent, conditional on its so-called parent nodes- the nodes with arrows leading to it. We see that the physical process is dependent upon a set of process parameters and boundary conditions. A number of quantities are then dependent upon the physical process.

To the left-hand side of the figure we have the observations of the physical process, inducing a set of parameters to represent features such as measurement errors. To the bottom right we have observations of climate outputs that are though to be dependent on the physical process. In each case, as for the observations on the physical process, a set of parameters of a statistical nature is induced. We assume in this case that no physical model is available to map *P* onto the climate outputs, so these models will be empirical in nature, including possibly a set of additional empirical predictors in each case. In this sense the approach being developed is novel when compared to other physical-statistical models in the literature.

A feature of graphical models is that we can write down a probability model for the uncertain quantities in the graph by working top-down through to the so-called terminal nodes that have no subsequent node in the graph:

$$[P | \Psi_P, B_P] \times [Y | P, \theta_Y] \times \prod_{i=1}^{m} [C_i | P, \theta_i, A_i]. \quad (10)$$

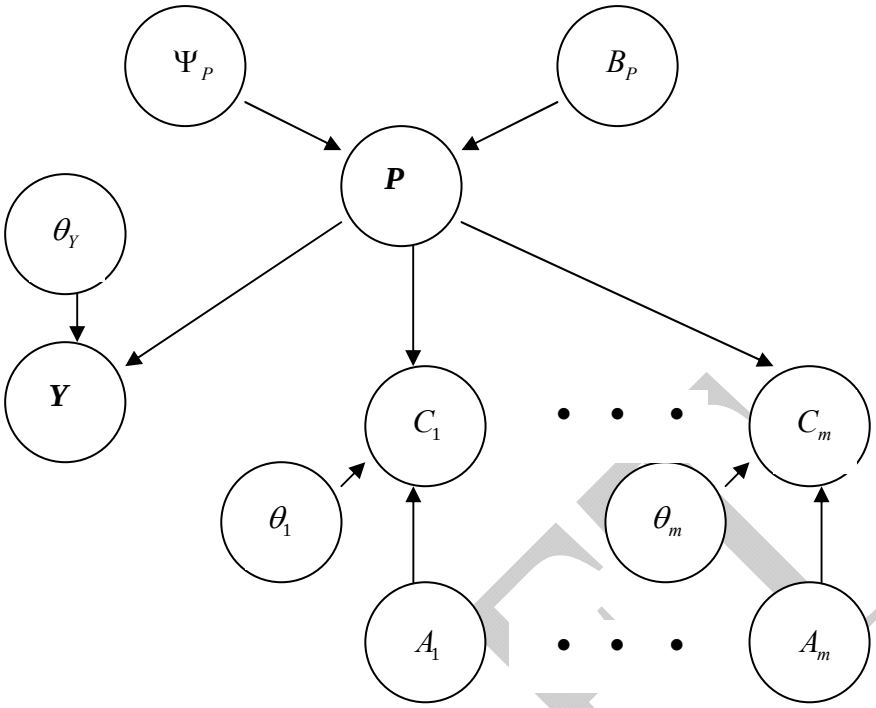

Figure 2 Graphical model displaying inter-connections between the ENSO process (*P*), observations on ENSO (*Y*) and various climate outputs ($C_1, \ldots, C_m$).

4.1 Mathematical Formulation

To express the graphical model of Figure 2 in more mathematical terms some additional notation is helpful:

$$\Theta = \{\theta_Y, \theta_1, \ldots, \theta_m\}$$
$$\mathbf{C} = \{C_1, \ldots, C_m\}$$
$$\mathbf{D} = \{Y, \mathbf{C}\}$$
$$\mathbf{A} = \{A_1, \ldots, A_m\}.$$

We may now apply Bayes' theorem to develop an expression for the joint probability of the random quantities in the graphical model shown in Figure 1:

$$[\mathbf{D}, \mathbf{A}, \Theta, P, \Psi_P, B_P] = [\mathbf{D}|\Theta, P, \mathbf{A}][P|\Psi_P, B_P][\Theta, \Psi_P],$$

after applying standard arguments to simplify the resulting conditional distributions (Berliner, 2003). The key difference to a conventional case is the data model:

$$[\mathbf{D}|\Theta, P, \mathbf{A}] = [Y, \mathbf{C}|\Theta, P, \mathbf{A}]$$
$$= \left\{\prod_{i=1}^{m}[C_i|\theta_i, P, A_i]\right\}[Y|\theta_Y, P],$$

assuming conditional independence of Y and **C,** which matches equation (10). In a more conventional case the empirical predictors $\{A_i\}$ would not be present, but as we are unlikely to explain all variation via ENSO alone it seems wise to retain these in our model framework.

The implication of equation (10) is that we need to characterise the relationship between the the climate outputs $\{C_i\}$ and ENSO (P), as well as the empirical predictors $\{A_i\}$. This will require some experimentation with the full data base, and will be taken up in project 3.2.7.

Table 2 List of variables used in Figure 2.

Uncertain Quantity	Description
Ψ_P	Physical process parameters
B_P	Boundary conditions
$\boldsymbol{P}$	Physical process model
$\boldsymbol{Y}$	Observations of the physical process
θ_Y	Statistical parameters induced by observing the physical process
C_i	Observations of the i^{th} climate output of interest
A_i	Empirical predictors additional to P
θ_i	Statistical parameters induced by observing linking in empirical predictors $\{A_i\}$

4.2 Posterior Distributions

We are primarily interested in learning about the physical process parameters and statistical parameters linking the climate outputs to the physical process. We do this by calculating probability distributions for these parameters given the observed data, and these are known as posterior distributions. Applying Bayes' theorem we see that

$$\begin{aligned}\left[\Psi, \Theta \middle| \mathbf{D}, P\right] &\propto \left[\mathbf{D}, P \middle| \Psi, \Theta\right]\left[\Psi, \Theta\right] \\ &= \left[\mathbf{D} \middle| P, \Theta\right]\left[P \middle| \Psi\right]\left[\Psi, \Theta\right] \\ &= \left\{\prod_{i=1}^{m}\left[C_i \middle| \theta_i, P\right]\right\}\left[Y \middle| \theta_Y, P\right]\left[P \middle| \Psi\right]\left[\Theta, \Psi\right].\end{aligned} \tag{11}$$

4.3 Predictive Distributions

By forecasting we mean projecting the model into the future, and capturing the uncertainty in the forecast. This is very easy to do in a Bayesian framework, particularly when implemented via simulation methods. For each ensemble in the

posterior sample we simply continue to evaluate the model beyond the current observed maximum time point. These so-called predictive samples can be used to summarise the forecast.

It is important to note that the forecast is being driven in part by a physically-inspired model for ENSO. Campbell (2005) found that forecast lead times in a physical-statistical model can be longer than those obtained via a purely empirical approach. This result was found in the context of a turning point in the physical process, a phenomenon that is difficult to forecast by purely empirical means.

5 Model-Fitting Algorithms

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. Expressions for these can be written down, as above, but it is rarely possible to evaluate them analytically; numerical methods are therefore typically 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 set of algorithms proceeds by generating a large sample from the posterior distribution of interest via a Markov random walk through the corresponding parameter or state space of interest. 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.

We provide some more details on these approaches below.

5.1 Bayesian Computation

5.1.1 Markov Chain Monte Carlo (MCMC) Methods

At the heart of MCMC is the *Metropolis-Hastings algorithm*. To define the algorithm we have to define the structure of the Markov chain. The Metropolis-Hastings algorithm constructs a probability function $P(\Theta,\Theta')$ describing the state transition from $\Theta^t = \Theta$ to $\Theta^{t+1} = \Theta'$ at iteration t as follows. First, generate a *candidate value* Θ' for Θ^{t+1} from a probability distribution $q(\Theta,\Theta')$, which for now is essentially

arbitrary. Accept this value with probability $\alpha(\Theta,\Theta')$ and move to $\Theta^{t+1} = \Theta'$. If this move is rejected, with probability $1-\alpha(\Theta,\Theta')$, set $\Theta^{t+1} = \Theta$ and so remain at the current state. This forms a Markov chain with transition probabilities

$$P(\Theta,\Theta') = q(\Theta,\Theta')\alpha(\Theta,\Theta')$$

$$\begin{aligned} P(\Theta,\Theta) &= \int_{\Theta''} q(\Theta,\Theta'')\big(1-\alpha(\Theta,\Theta'')\big)d\Theta'' \\ &= 1-\int_{\Theta''} q(\Theta,\Theta'')\alpha\big(\Theta,\Theta''\big)d\Theta'' \end{aligned}$$

If we write the posterior distribution of interest given data $\mathbf{x}$ $\pi(\Theta) = f(\Theta|\mathbf{x})$ for convenience and define

$$\alpha(\Theta,\Theta') = \begin{cases} \min\left\{\dfrac{\pi(\Theta')q(\Theta',\Theta)}{\pi(\Theta)q(\Theta,\Theta')},1\right\} & \text{if } \pi(\Theta)q(\Theta,\Theta') > 0 \\ 1 & \text{if } \pi(\Theta)q(\Theta,\Theta') = 0 \end{cases}$$

then

$$\pi(\Theta)\,p(\Theta,\Theta') = \pi(\Theta')\,p(\Theta',\Theta) \qquad (12)$$

Equation (12) defines the condition of *reversibility*, a sufficient condition for $\pi(\Theta)$ to be the limiting distribution of the chain provided $q(\Theta,\Theta')$ is chosen to be irreducible (i.e., all states can be sampled from any given state) and aperiodic (ie, the number of steps until the chain returns to its current state is not a multiple of an integer greater than 1) (Smith and Roberts, 1993).

A number of variants of this general algorithm are available (Tierney, 1994). Amongst the most important are *single-site* updating algorithms. Rather than generating candidate values for Θ, these algorithms generate candidate values for each element of Θ in turn. Thus for p-dimensional Θ, a candidate value θ_i^* from a univariate density $q_i(\Theta,\theta_i^*)$ is proposed and the move from $\Theta = (\theta_1,\ldots,\theta_{i-1},\theta_i,\ldots,\theta_p)'$ to $\Theta^* = (\theta_1,\ldots,\theta_{i-1},\theta_i^*,\ldots,\theta_p)'$ is accepted with probability

$$\alpha_i(\Theta,\Theta^*) = \min\left\{\frac{\pi(\Theta^*)q_i(\Theta^*,\theta_i)}{\pi(\Theta)q_i(\Theta,\theta_i^*)},1\right\}.$$

Chan and Geyer (1994) show that the single-site updating algorithm also converges to $\pi(\Theta)$ if the univariate candidate generation densities are irreducible and aperiodic. This technique also extends to updating blocks of parameters at a time using multivariate updates. This is important in cases where some of the parameters are highly correlated.

Generally the most efficient algorithms use the current state as the mean for the next proposed state, allowing the algorithm to locate regions of high probability and explore them. In the terminology of Tierney (1994) these are random walk chains. An alternative is to propose candidate states from the same distribution at each step, regardless of the current state. A chain constructed in this way is known as an independence chain. If carefully chosen this can work well, but random walk chains tend to be more robust and efficient, provided the spread of the proposal densities is chosen appropriately.

In practice it is found that candidate generation densities having variances about the same or somewhat larger than the marginal posteriors work well. This is in the sense of exploring the posterior density in an efficient manner, one measure of which is the acceptance rate of the algorithm. Clearly if proposed states are only rarely accepted the algorithm is very inefficient. Conversely, too high an acceptance rate suggests that only a small region of the posterior density is being explored. A balance needs to be struck between these extremes, and an acceptance rate of between 30% and 70% seems to work well. Some theoretical and practical justification for this rule of thumb has been found (e.g., Roberts and Poulson, 1994; Roberts and Smith, 1994; Weir, 1997).

Perhaps the most well known MCMC method is Gibbs sampling. Gibbs sampling requires that each of the full conditional distributions $\left[\theta_j \middle| \Theta_{(j)}, \mathbf{X}\right]$ be specified, where the notation (j) denotes that the j^{th} parameter is omitted. The $[\cdot]$ notation is used to denote distribution or density as required.

Gibbs sampling proceeds as follows:

1. Select a starting value for the parameter vector, $\Theta_0 = \left(\theta_1, \ldots, \theta_p\right)_0'$.
2. Sample a value of θ_1 from the distribution $\left[\theta_1 | \Theta_{(1)}\right]$.
3. Sample a value of θ_2 from the distribution $\left[\theta_2 | \Theta_{(2)}\right]$.
4. …
5. Sample a value of θ_p from the distribution $\left[\theta_p | \Theta_{(p)}\right]$, which completes a sample for the parameter vector Θ.
6. If enough sample have been completed then stop. Otherwise, return to step 1 replacing Θ_0 by Θ found at step 5.

We see that Gibbs sampling is a single-site updating Metropolis-Hastings algorithm in which the candidate generation densities are the full conditionals, and the acceptance probability is set to 1 because the full conditionals attains a technical condition known as reversibility (see equation (12)) without the need for a rejection step.

For off-line processing in project 3.2.7 we would expect to use mixtures of Gibbs sampling and Metropolis-Hastings steps, depending on the nature of the empirical models we develop. Ideally we will use Gibbs sampling as much as possible as it tends to be faster overall, unless efficient one-step multivariate proposals are available.

5.1.2 Particle Filters

As noted above, MCMC methods are iterative in nature so normally best suited to off-line processing applications. In situations where processing is on-line with streaming data iterative methods tend not be practical. For computing posterior distributions in such situations a collection of techniques under the general heading *Sequential Monte Carlo* methods have been developed (Doucet *et al*., 2001).

The physical process model we have chosen is relatively simple, so if we applied a normally distributed error model then in principle we could make use of the Kalman filter to update posterior distributions of interest. However, it is well known that the Kalman filter is not extendable to nonlinear/non-Gaussian models. We therefore choose to focus on methods known by various names, which we refer to as *particle filters*. These methods use weighted realisations from the state space of interest (a future observation, say) to summarise the relevant probability distribution (predictive distribution).

To illustrate these methods, first consider the directed graph component for the ENSO process alone taken from Figure 2, but now with a time component (t), shown in Figure 3 below.

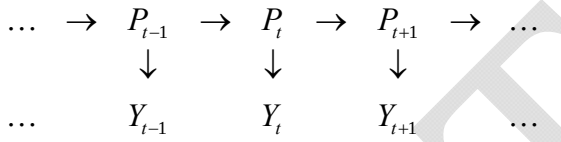

Figure 3 Directed graph component for the ENSO process alone

Of primary interest is predicting the ENSO state P_t given observations $y_{1:t-1}$, and it is straightforward to show that

$$p\left(P_t \middle| y_{1:t-1}\right) = \int p\left(P_t \middle| P_{t-1}\right) p\left(P_{t-1} \middle| y_{1:t-1}\right) dP_{t-1}. \tag{13}$$

The first component inside the integral is the discretised physical process model, whilst the second component represents an update step. In practice it is not straightforward to evaluate this integral, so Monte Carlo methods are used because their rate of convergence is independent of the dimension of the integrand. For any deterministic method the rate of convergence decreases as the dimension of the integrand increases.

The most common approach is based on importance sampling (Smith and Gelfand, 1992). This technique works as follows. Suppose we wish to estimate the integral

$$E_p\left[h(x)\right] = \int h(x)\, p(x) dx,$$

for some function $h()$ and probability density function $p()$ via Monte Carlo. We first generate a random sample $x_1, \ldots, x_N$ from the probability function $p()$, then clearly

$$E_p\left[h(x)\right] \approx N^{-1} \sum_j h\left(x_j\right).$$

This is the standard Monte Carlo estimator, but suppose that $p()$ is hard to simulate from but some (ideally similar) probability function is very easy to simulate from. By definition:

$$\begin{aligned} E_p\left[h(x)\right] &= \int \frac{h(x)\, p(x)}{\pi(x)} \pi(x) dx \\ &= \int h(x)\, w(x)\, \pi(x) dx \\ &= E_\pi\left[h(x)\, w(x)\right]. \end{aligned}$$

Thus, if we generate a random sample $x_1, \ldots, x_N$ from the probability function $\pi()$ then

$$E_p\left[h(x)\right] \approx N^{-1}\sum_j h(x_j)w(x_j),$$

where $w(x_j) = p(x_j)/\pi(x_j)$. In the language of sequential Monte Carlo, the sample members $\{x_j\}$ are known as particles, with weights $\{w_j\}$.

It turns out to be quite straightforward to adapt this method to sequential estimation problems (Doucet *et al*, 2001, pp 9-10), such as defined by equation (13). Unfortunately it can be shown that naïve application of sequential importance sampling will lead to skewed weights as the time step grows larger, quite quickly leading to just one particle having non-zero weight. A solution to this problem is provided by the so-called *bootstrap filter* (Gordon *et al.*, 1993) which resamples the particles using the importance weights. This ensures that only the 'fittest' particles are allowed to survive, ideally with equal weights.

It is possible to improve the performance of particle filters in many respects, but it seems that the bootstrap filter is well suited to our application. Our proposed approach to project 3.2.7 is therefore as follows:

1. Use the boostrap filter to maintain a recursive estimate of the prediction density (13);
2. Build an off-line model for $p\left[C_i(t)\middle|P_t, A_{1:t}\right]$;
3. Using the particles from step 1 and the off-line model for 2, generate ensemble for predictive distribution for C_i.

5.2 The Physical Process Model

It is clear from the above discussion on fitting physical-statistical models that we have to calculate realisations from the process model, which requires solution of the model equations. The model of Burgers *et al*. (2005) can be written as $\dot{X}(t) = \mathbf{A}X(t)$, so the fundamental solution is given by $X(t) = \exp(\mathbf{A}t).\mathbf{k}$, where **k** is an arbitrary vector determined from the initial conditions. A general method of solution will be to note that the matrix **A** can be eigen-decomposed as $\mathbf{A} = E^{-1}JE$, where J is in Jordan canonical form and the columns of E are formed from the eigenvectors of **A**. It is then straightforward to calculate

$$\exp(At) = E^{-1}\exp(Jt)E.$$

However, it will be much faster to solve the equations by forward integration, noting that:

$$\dot{X}(t).\delta t \approx X(t+\delta t) - X(t)$$

for small $\delta t > 0$. The forward integration is therefore driven by

$$X(t+\delta t) \approx (I + \delta t A)X(t), \qquad (12)$$

commencing from a suitable initial condition $X(t_0)$. This essentially discretises the process model.

5.2.1 Physical Model Extensions

A couple of extensions of the Burgers *et* al. model are considered: periodic forcing and representational error. By representational error we mean uncertainty in the physical model itself, recognising that we are not completely certain of the dynamics of ENSO. Equation (5) may therefore be generalised to:

$$\dot{X}(t) = AX(t) + P(t) + S(t), \tag{13}$$

where

$$P(t) = \begin{pmatrix} a_1 \cos(2\pi t/12) \\ a_2 \cos(2\pi t/12) \end{pmatrix}, \qquad S(t) \sim N(\mathbf{0}, \Sigma_p).$$

In principle the period of the forcing could be estimated also (Campbell, 2005), but we neglect this for the present. An illustration of the behaviour that can be displayed by this system is shown in Figure 4 below. We see that the behaviour is much more complex than the base model (Figure 1); the periodic forcing is evident as is the stochastic error, so no simple limiting behaviour is observed in the phase diagram showing thermocline depth against SST (anomaly).

6 Discussion and Conclusions

Application of physical-statistical methods using Bayesian hierarchical modelling are growing in the literature, and have been used in many geophysical applications to date. Notable among these are ENSO forecasting by Berliner *et al*. (2000). Current forecasts using this approach are maintained at http://www.stat.ohio-state.edu/~sses/collab_enso.php, and the probabilistic nature of Bayesian forecasts is apparent. The probability triangle plot at http://www.stat.ohio-state.edu/~sses/collab_enso_regime_forecast.php is particularly effective tool in the context of this approach which incorporates regime-dependence. For each regime a posterior forecast probability is given, with a mean sea surface temperature plot for each regime.

In this method-focused project we have developed a framework using relatively simple physical models for ENSO to develop a forecasting model, allowing for off-line processing and streaming data. We have identified suitable algorithms in each case, which will be implemented in project 3.2.7. The forecasts obtained will be in the form of probability distributions, which can be used to summarise key information for users of the forecasts.

The resources available to this project are somewhat limited, but given the methodological process that has been made it seems that aiming for a working prototype in project 3.2.7 is not unreasonable.

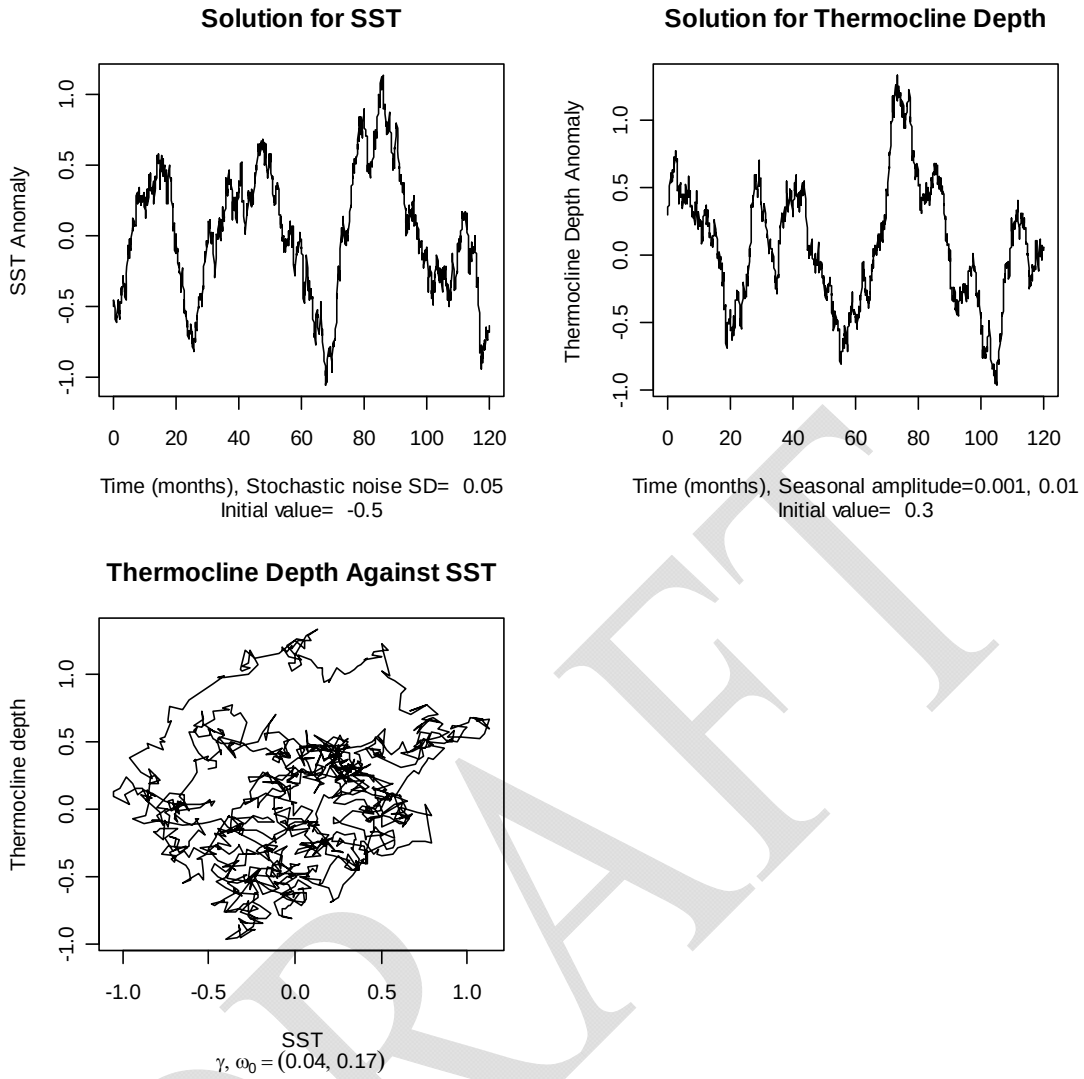

Figure 4 The Burgers *et al.* model with periodic forcing and stochastic representation error.

Acknowledgements: This project was funded by the South East Australian Climate Initiative. The project leader is grateful for the input of Dr Bryson Bates (Director, CSIRO Climate) in helping to frame the scope of this work and the review of Mark Palmer.

References

Battisti, D. S. and Hirst, A. C. (1989). Interannual variability in a tropical atmosphere-ocean model: Influence of the basic state, ocean geometry and nonlinearity. *Journal of the Atmospheric Sciences,* **46,** 1687-1712.

Berliner, L. M., Wikle, C. K. and Cressie, N. (2000). Long-lead prediction of Pacific SST via Bayesian dynamic modeling. *Journal of Climate,* **13,** 3953-3968.

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

Berliner, L. M., Milliff, R. F. and Wikle, C. K. (2003). Bayesian hierarchical modeling of air-sea interaction. *Journal of Geophysical Research,* **108,** 1-1:1-18.

Bernardo, J. M. and Smith, A. F. M. (1994) *Bayesian Theory,* John Wiley & Sons, Chichester.

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.

Cane, M. A. and Zebiak, S. E. (1985). A theory for El Niño and the Southern Oscillation. *Science*, 228, 1085-1087.

Chan, K. S. and Geyer, C. J. (1994). Markov-Chains for Exploring Posterior Distributions - Discussion. *Annals of Statistics*, 22, 1747-1758.

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

Gordon, N. J., Salmond, D. J. and Smith, A. F. M. (1993). Novel approach to nonlinear/non-gaussian Bayesian state estimation, *IEE Proceedings-F*, 140(2), 107-113.

Jin, F.-F. (1996). Tropical ocean-atmosphere interaction, the Pacific Cold Tongue, and the El Niño/Southern Oscillation. *Science,* **274,** 76-78.

Jin, F.-F. (1997). An equatorial ocean recharge paradigm for ENSO. Part I: Conceptual model. *Journal of the Atmospheric Sciences,* **54,** 811-829.

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).

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer texts in statistics, 2nd Edition, New York.

Roberts, G. O. and Polson, N. G. (1994). On the Geometric-Convergence of the Gibbs Sampler. *J. Roy. Statist. Soc.* **B**, 56, 377-384.

Roberts, G. O. and Smith, A. F. M. (1994). Simple Conditions for the Convergence of the Gibbs Sampler and Metropolis-Hastings Algorithms. *Stochastic Processes and Their Applications*, 49, 207-216.

Smith, A. F. M. and Gelfand, A. E. (1992). Bayesian statistics without tears: a sampling-resampling perspective. *Am. Statist.*, 46(4), 84-88.

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.

Tierney, L. (1994). Markov chains for exploring posterior distributions, *The Annals of Statist., 22*(4), 1701-1762 (including discussion).

Weir, I. S. (1997). Fully Bayesian reconstructions from single-photon emission computed tomography data. *J. Am. Statist. Assoc.*, 92, 49-60.

Wikle, C. K. (2003). Hierarchical models in environmental science. *International Statistical Review,* **71,** 181-199.