

# **Application of Bayesian Hierarchical Modelling to a Delayed Action Oscillator Model for the El Niño-Southern Oscillation**

Technical Report 2004/139

Edward P. Campbell  
CSIRO Mathematical & Information Sciences  
Centre for Environment and Life Sciences  
Private Bag 5, PO Wembley, WA 6913.  
Tel: (08) 9333 6203  
[eddy.campbell@csiro.au](mailto:eddy.campbell@csiro.au)

September 2004

## Abstract

In this report we continue previously reported ideas on physical-statistical models using Bayesian hierarchical approaches. We develop a physical-statistical model using a simple model for the El Niño-Southern Oscillation to illustrate the ideas. The model-fitting performance is assessed and we briefly examine the issues involved in forecasting. The results are very promising, suggesting that more complex physical models be looked at next.

Keywords: Bayes' theorem; El Niño-Southern Oscillation; hierarchical methods; physical-statistical models; importance sampling-resampling; uncertainty.

## 1 Introduction

In a previous report (Campbell, 2004) we provided an introduction to physical-statistical modelling using Bayesian hierarchical methods. The objective of this report is to develop a climate research-related application that is suitably complex to be realistic, but is sufficiently simple to be transparent.

In the next section we describe a nonlinear conceptual model for the well known El Niño-Southern Oscillation (ENSO) phenomenon. We provide a brief overview of the motivation for this model and describe our numerical approach to using the model. In section 3 we develop a hierarchical model to integrate the physical model, incorporating parameter uncertainty, with a statistical measurement error model for the observations. An algorithm to fit the model using importance sampling-resampling is described, and in section 4 we explore a number of case studies to examine the performance of the algorithm. In section 5 we briefly examine forecasting using the hierarchical model, a strength of the Bayesian approach. In section 6 we provide some discussion and conclusions. An Appendix is included deriving the resampling probabilities required by the model fitting algorithm.

## 2 The Delayed Action Oscillator

Suarez and Schopf (1988) developed a simple nonlinear conceptual model for ENSO. ENSO may be thought of as a periodic disturbance in the ocean-atmosphere circulation. The model incorporates a negative, delayed feedback mechanism to explain the long time scale of ENSO (typically 2-4 years). There is a considerable amount of work in the literature on linear mechanisms for ocean-atmosphere interactions to model the growth rate of ENSO, with nonlinear damping effects limiting its growth.

Suarez and Schopf (1988) assume the nonlinear form

$$dT/dt = kT - bT^3,$$

where  $T$  represents the amplitude of the growing disturbance,  $k$  its growth rate and  $bT^3$  all nonlinear effects acting on it. Scaling time by  $k^{-1}$  and  $T$  by  $(k/b)^{1/2}$  we find

$$dT/dt = T - T^3. {}^1$$

Wind anomalies driven by sea surface temperature perturbations drive westward propagating (Rossby) waves. On reaching the western boundary they reflect into eastward propagating (Kelvin) waves. Upon returning to the central/eastern part of the Pacific basin they will tend to have a damping effect. This suggests the inclusion of a delayed negative term:

$$dT/dt = T - T^3 - \mathbf{a}T(t - \mathbf{d}), \quad (1)$$

where  $\mathbf{d}$  is a non-dimensional delay (wave transit time) and  $0 < \mathbf{a} < 1$  measures the influence of the returning signal relative to that of the local feedback.

The model equation (1) is easy to solve numerically using step-by-step integration, given a suitable initial condition  $T_0$  and values for  $\mathbf{a}$  and  $\mathbf{d}$ . If we appeal to the definition of a differential equation, we may think of (1) as being approximated by

$$\frac{T(t + \mathbf{d}t) - T(t)}{\mathbf{d}t} \approx T - T^3 - \mathbf{a}T(t - \mathbf{d}),$$

for a small time increment  $\mathbf{d}t > 0$ , leading to:

$$T(t + \mathbf{d}t) \approx T(t) + \mathbf{d}t [T - T^3 - \mathbf{a}T(t - \mathbf{d})]. \quad (2)$$

Thus a numerical solution for  $T$  can be found by repeated application of equation (2), starting from  $T_0$  and a suitably small choice for the time step  $\mathbf{d}t$ . This is the approach used to produce the results described in this report.

### 3 Bayesian Hierarchical Model

We assume that the amplitude  $T$  is measured with error, yielding observations  $Y$  that are normally distributed with mean  $T$  and known variance  $t^2$  (a simplifying assumption, but not necessary). Prior information about the physical parameters is summarised via a probability distribution  $[\mathbf{d}, \mathbf{a}]$ , and we will focus on learning about these parameters using available data and the physical model (1). Using the results of Campbell (2004), we find that the distribution for the physical parameters given the observed data can be written as

$$[\mathbf{d}, \mathbf{a} | Y] \propto [Y | T, t^2] [T | \mathbf{d}, \mathbf{a}] [\mathbf{d}, \mathbf{a}]. \quad (3)$$

The term  $[Y | T, t^2]$  is the measurement error model for the observations  $Y$  described above. The probability model for the physical process  $T$  is the second term,  $[T | \mathbf{d}, \mathbf{a}]$ , which arises due to uncertainty in the physical parameters  $\mathbf{d}$  and  $\mathbf{a}$ . In the absence of more information it seems reasonable to assume that the physical parameters are *a priori* independent, so that  $[\mathbf{d}, \mathbf{a}] = [\mathbf{d}][\mathbf{a}]$ , which does not imply posterior independence.

---

<sup>1</sup> Technically we should replace  $T$  by  $T'$ , say, because of the change of variable, but neglect this here.

### 3.1 Fitting the Model

A number of possibilities were discussed by Campbell (2004), and we focus here on the basic importance sampling-resampling method. Recall that this algorithm is applied as follows:

1. Generate a sample of size  $N$  from the prior parameters model  $[\mathbf{d}, \mathbf{a}]$ , yielding  $\{\mathbf{d}^i, \mathbf{a}^i : i = 1, \dots, N\}$ .
2. Using this sample, generate an ensemble of physical process realisations, yielding  $\{T^i, \mathbf{d}^i, \mathbf{a}^i\}$ . This is accomplished by numerical solution of the model equation using (2) for each  $\{\mathbf{d}^i, \mathbf{a}^i\}$ .
3. Resample the ensembles using the acceptance probability  $\frac{[Y|T^l, \mathbf{t}^2]}{\sum_j [Y|T^j, \mathbf{t}^2]}$ ,  $l = 1, \dots, N$ , to provide a sample of the desired size<sup>2</sup>, drawn approximately from  $[\mathbf{d}, \mathbf{a}|Y]$ .

More details can be found in Bernardo and Smith (1994) pp350-352, with a geophysical application provided by Kivman (2003)

## 4 Results

### 4.1 Unknown $\mathbf{a}$ and $\mathbf{d}$

Given that there is a close relationship between  $\mathbf{d}$  and the periodicity of  $T$  it is likely that this parameter will be relatively easy to pin down from prior information. We simulate data for  $(\mathbf{d}, \mathbf{a}) = (6, 0.75)$  using equation(2), then add normally distributed measurement error with mean 0 and variance  $\mathbf{t}^2 = 1.0$ . The resulting time series is shown in Figure 1, with 100 physical process ensembles, assuming a starting value of  $T_0 = 0.4$ . These ensembles were generated by first simulating 100 realisations  $\{\mathbf{d}^i, \mathbf{a}^i : i = 1, \dots, 100\}$  from the prior distributions defined by  $\mathbf{d} \sim U\{4, 5, 6, 7, 8\}$  (uniform on the integers 4, 5, 6, 7, 8) and  $\mathbf{a} \sim N(0.75, 0.05^2)$ . For each pair  $\{\mathbf{d}^i, \mathbf{a}^i\}$  a physical process realisation was generated using equation (2). Note that we could also incorporate initial condition uncertainty by giving  $T_0$  a prior distribution, but for the purposes of this report chose to fix  $T_0 = 0.4$ .

It is clear that many ensembles are far removed from the simulated data, which is to be expected given that  $\mathbf{d}$  has been assumed unknown; we would also expect increasing divergence with time, and this is evident.

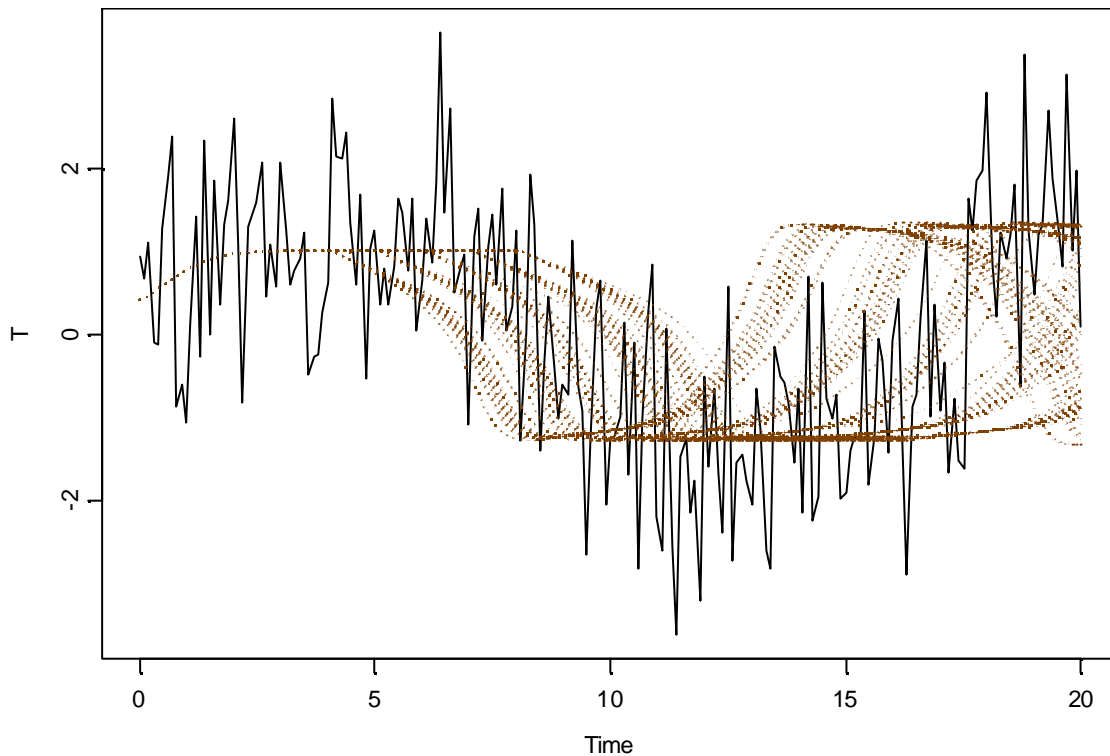
We look next at the goodness of fit of the ensembles as functions of the parameters. An intuitive way to do this is via the sum of squared differences between the simulated data

---

<sup>2</sup> In all the examples to be shown we generate a sample of size 1000.

and the ensembles. These are plotted in Figure 2 and Figure 3 for  $\mathbf{d}$  and  $\mathbf{a}$  respectively. It is clear that the best fit is provided by the true value  $\mathbf{d} = 6$ , whilst a number of regimes are evident for  $\mathbf{a}$  corresponding to different values of  $\mathbf{d}$ . The best fitting regime corresponds to  $\mathbf{d} = 6$ , and there is clearly some uncertainty as to the value of  $\mathbf{a}$ .

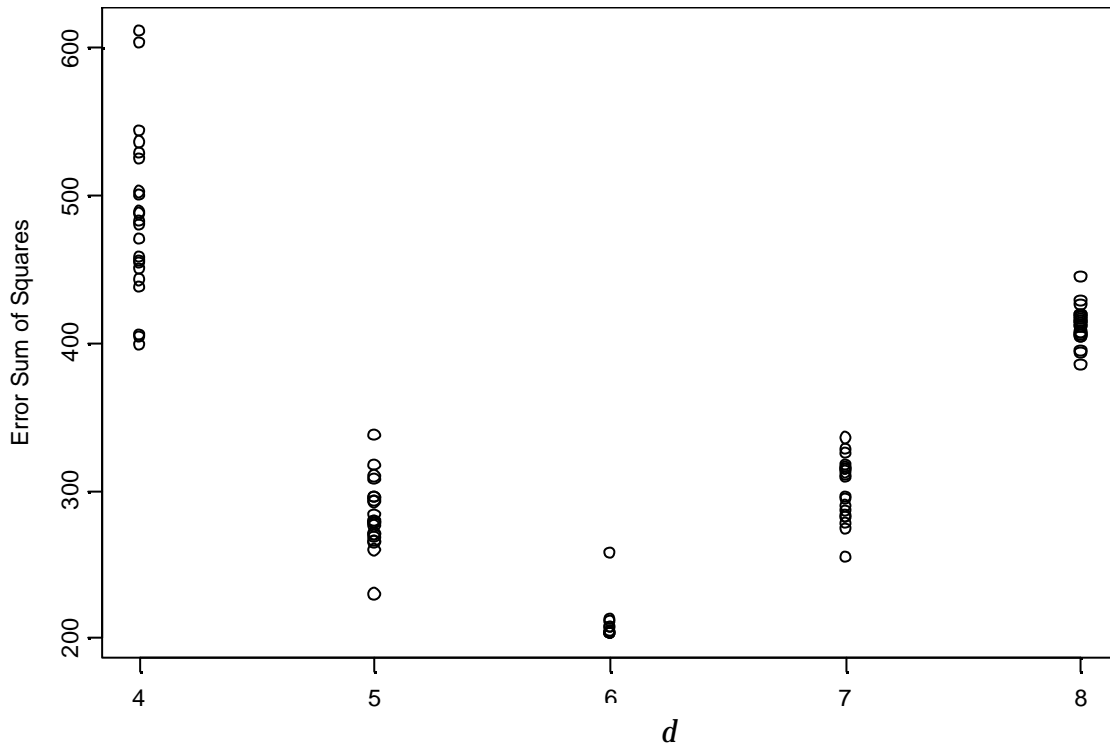
A histogram of resampling probabilities (calculated as per the Appendix) is shown in Figure 4, and it is clear that most of the ensembles essentially have probability 0 of being resampled, and that a range of ensembles have been resampled with similar probability.



**Figure 1** 100 physical process ensembles, assuming both  $\mathbf{d}$  and  $\mathbf{a}$  to be unknown. The simulated time series is the solid line; the physical process ensembles are shown as dotted lines.

The resampled ensembles (1000 in all) are summarised in Exhibit 1, and in practice we would conduct a detailed examination of these to examine their contribution to the posterior sample. We see that the credibility interval for  $\mathbf{a}$  is favouring values greater than the true value of 0.75.

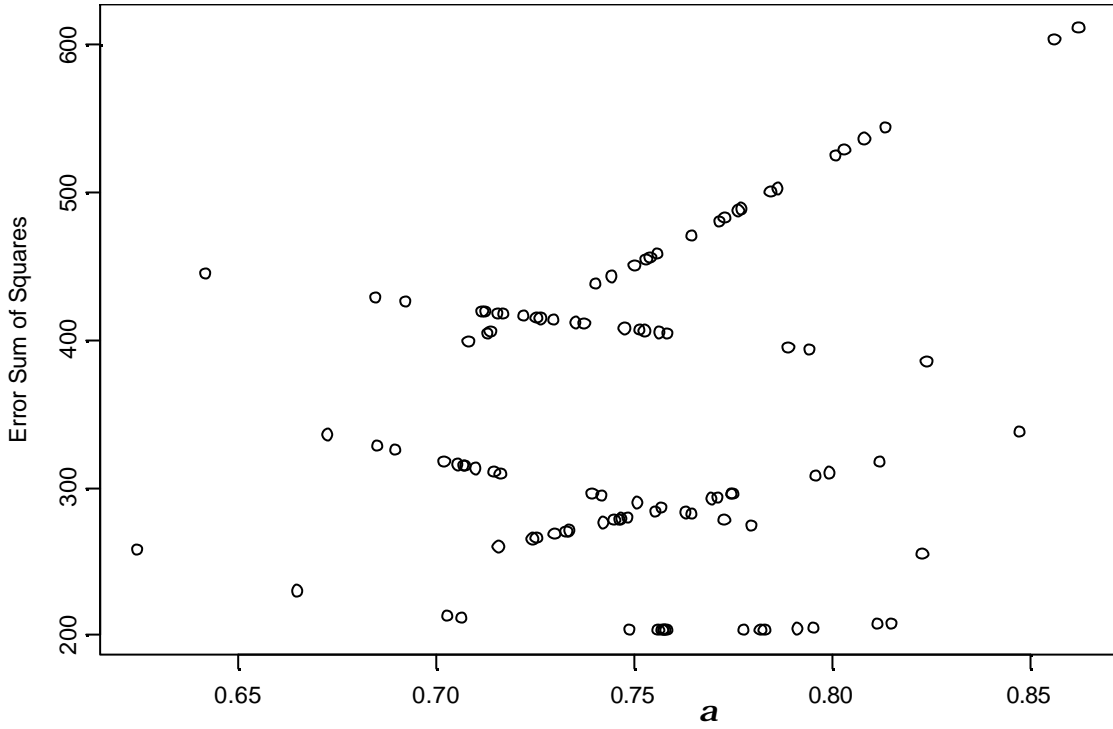
A comparison using boxplots<sup>3</sup> of the prior and posterior samples for  $\mathbf{a}$  is shown in Figure 5, and the greater information content in the posterior distribution compared to the prior distribution is evident in the much reduced variability.



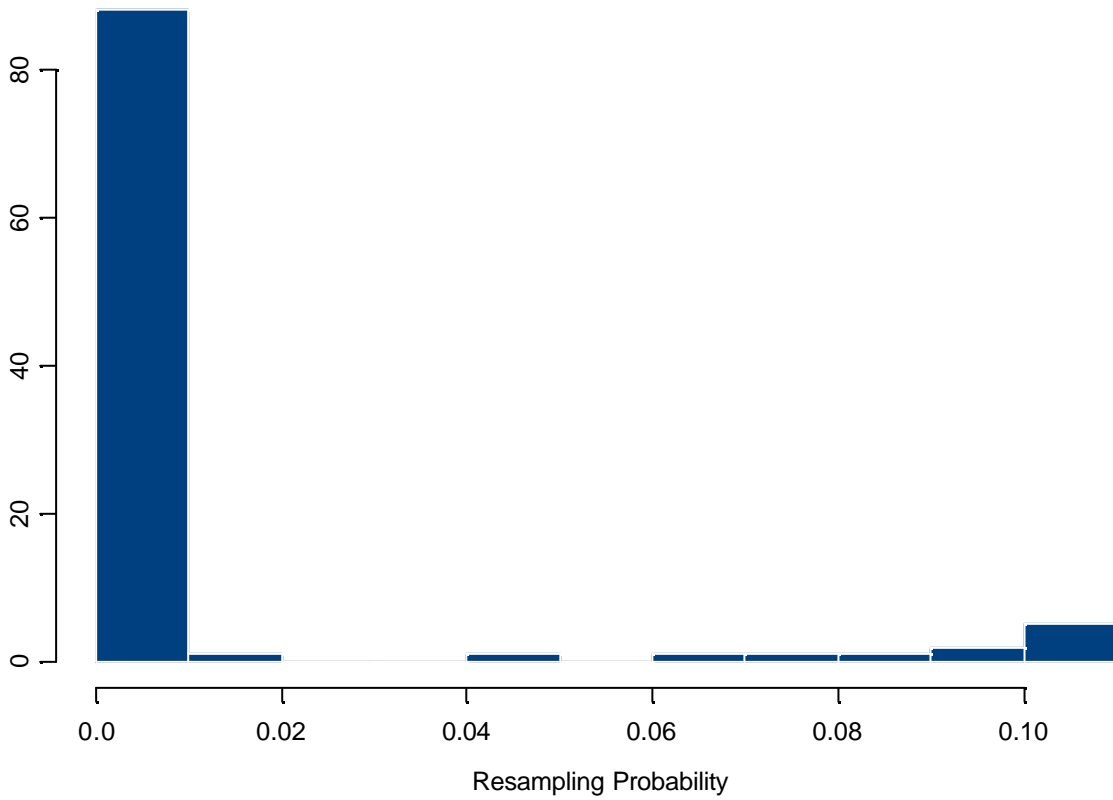
**Figure 2** Error sum of squares as a function of  $d$ .

---

<sup>3</sup> The central box ranges from the lower to the upper quartile, whilst the so-called ‘whiskers’ stretch essentially to the maximum and minimum points, unless they are deemed to be outliers, which are identified as circles. The central dot in the box is the median, or half-way point.



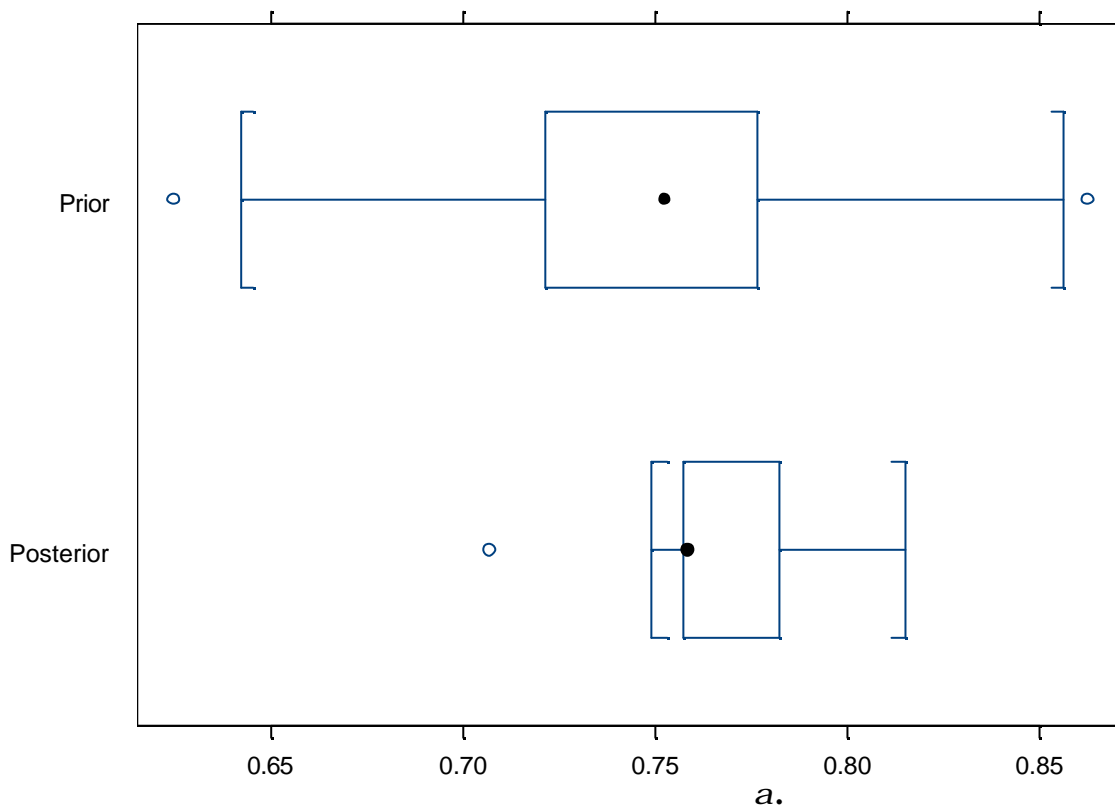
**Figure 3** Error sum of squares as a function of  $a$ .



**Figure 4** Histogram of resampling probabilities.

Frequency Table of Ensembles Selected													
1	2	3	8	18	41	48	74	75	81	85	88	93	94
9	91	95	11	119	98	91	2	114	93	41	83	65	88
Summary statistics for $d$													
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.								
6	6	6	6	6	6								
Standard Error= 0													
95% Credibility Interval= [6, 6]													
Summary statistics for $a$													
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.								
0.7065830	0.7573346	0.7583640	0.7686252	0.7822383	0.8151645								
Standard Error= 0.0158													
95% Credibility Interval= [0.749, 0.796]													

**Exhibit 1** Summary of resampled ensembles and posterior statistics.



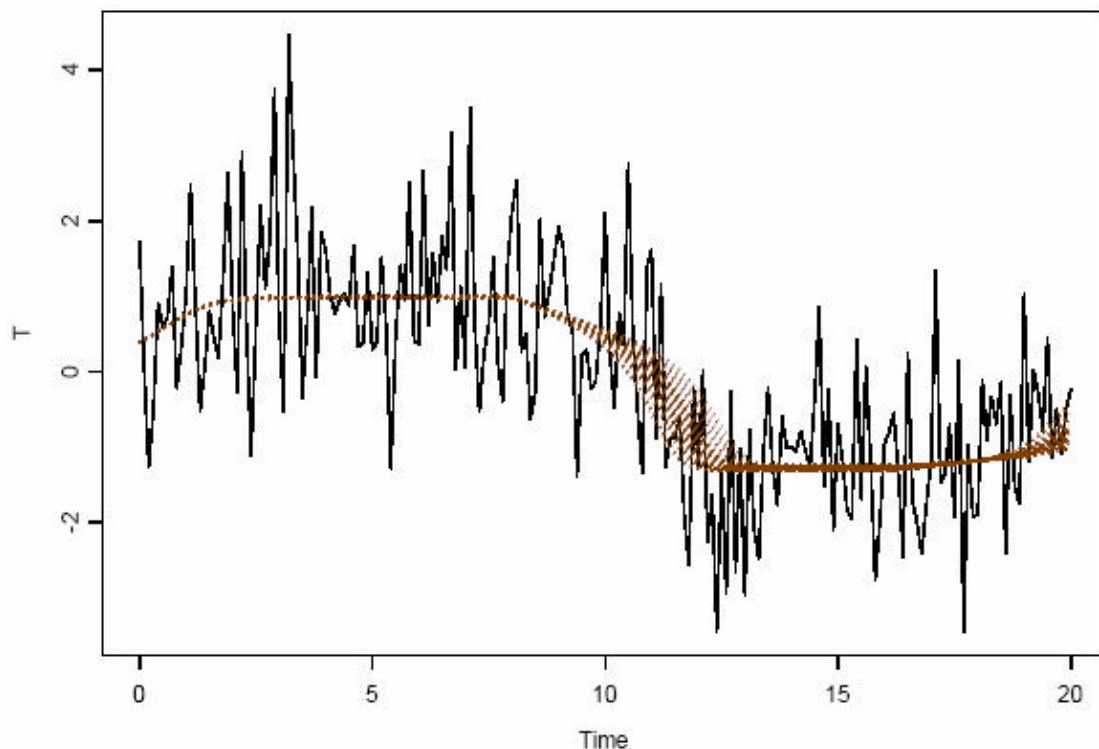
**Figure 5** Boxplot of prior and posterior samples for  $a$ .

## 4.2 Known $d = 8$ And Unknown $a$

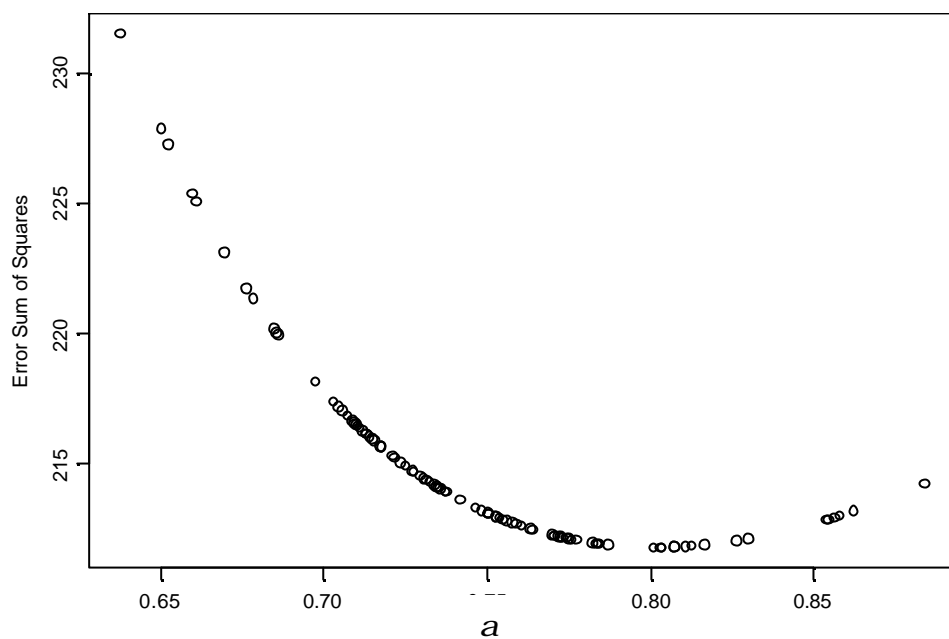
The simulated data and physical process ensembles are shown in Figure 6. It is immediately evident, although not surprising, that the physical process ensembles provide a much better fit to the data with known  $d$ .

The error sum of squares as a function of  $a$  is shown in Figure 7, which indicates an optimal value of about 0.8 in this case, with some degree of uncertainty. We see from the histogram of resampling probabilities in Figure 8 that many ensembles have contributed to the posterior distribution.

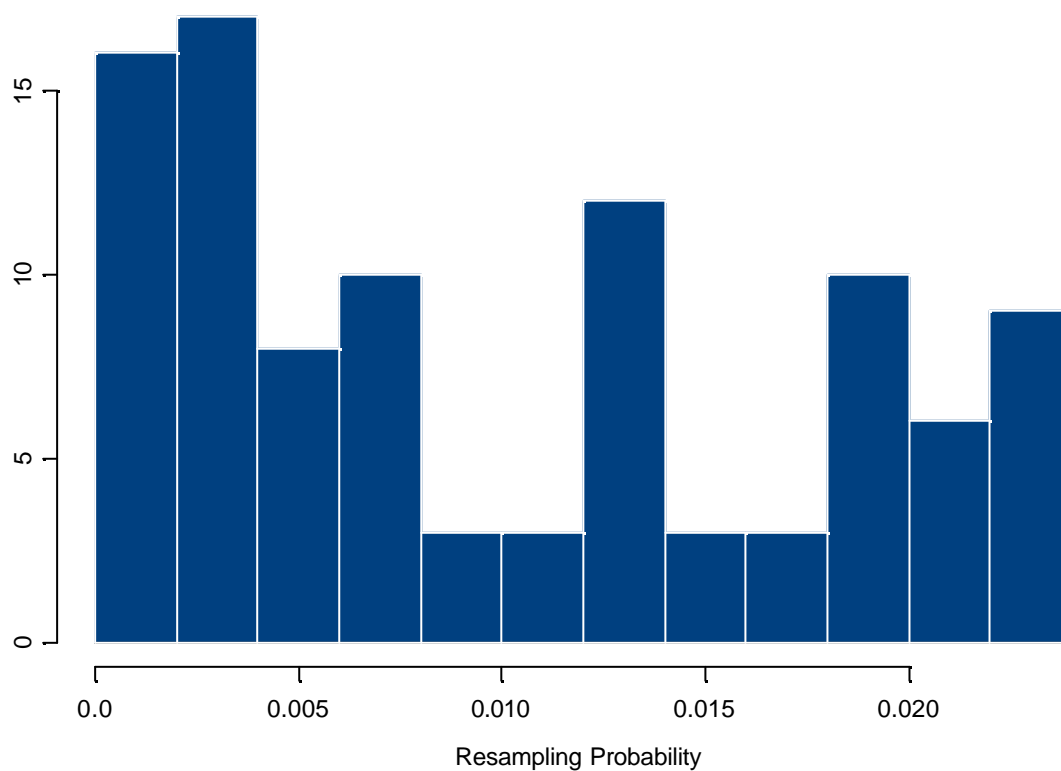
Posterior summary statistics are given in Exhibit 2, which confirms our assessment with a posterior mean of 0.78, with a 95% credibility interval ranging from 0.712 to 0.858. Boxplots of the prior and posterior samples are shown in Figure 9, and again it is evident that the information content of the posterior is greater than the prior. There is a shift in location, and some reduction in variability.



**Figure 6** Simulated data and physical process ensembles with  $d$  known and  $a$  unknown.



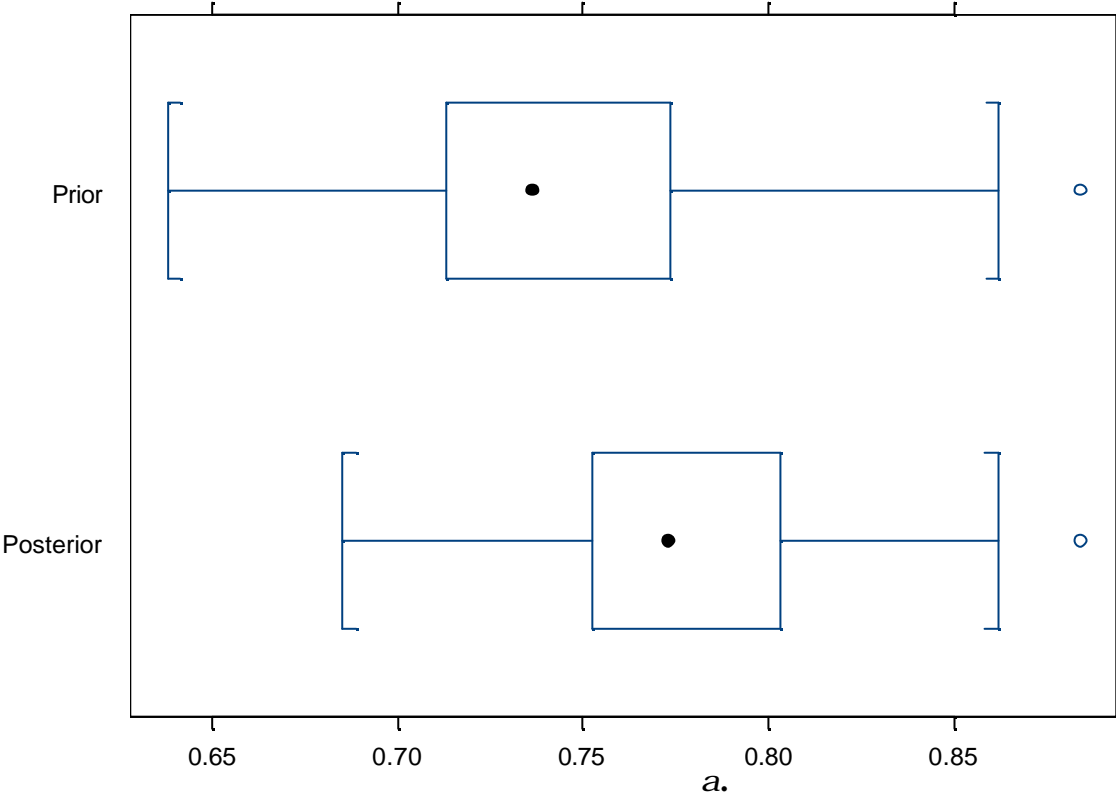
**Figure 7** Error sum of squares as a function of  $a$ .



**Figure 8** Histogram of resampling probabilities.

Summary statistics for ALPHA						
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
0.6849161	0.7527148	0.7730340	0.7782655	0.8034836	0.8840849	
Standard Error= 0.0380						
95% Credibility Interval= [0.712, 0.858]						

**Exhibit 2** Posterior summary statistics for *a*.



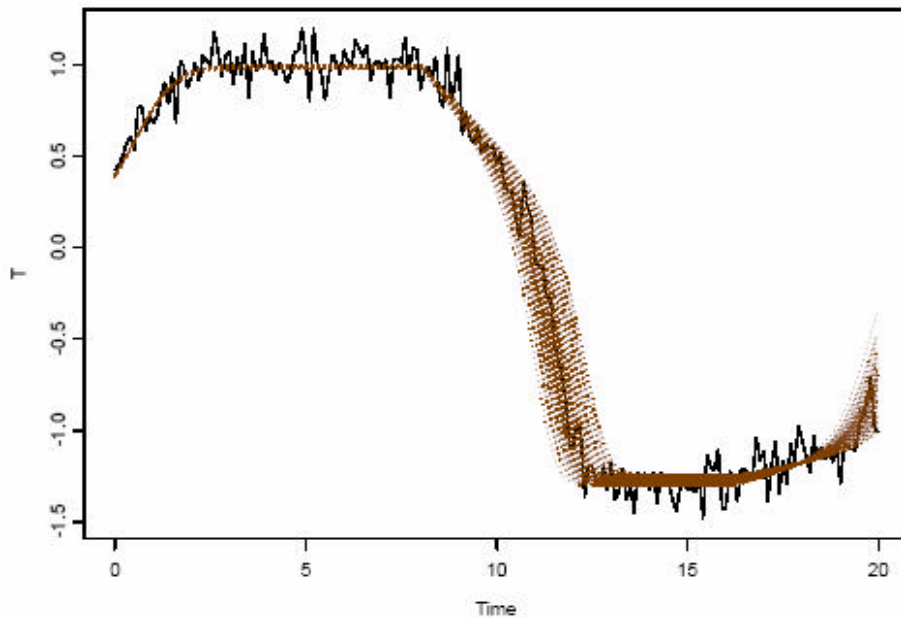
**Figure 9** Boxplots of prior and posterior samples for *a*.

### 4.3 Reduced Measurement Error

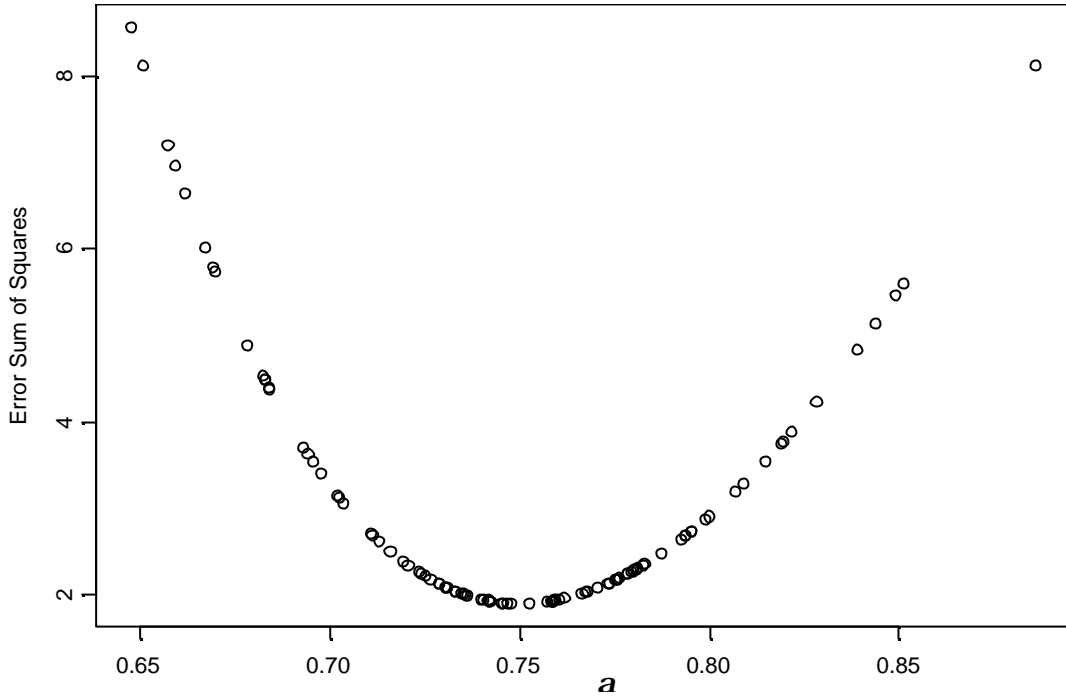
In this case we assume a significant reduction in measurement error to  $t^2 = 0.1^2$ , anticipating much more precise inference for  $\mathbf{a}$ . The simulated data and ensembles are shown in Figure 10.

The error sum of squares is plotted in Figure 11, and it is clear that the true value of  $\mathbf{a}$  is identified with some precision. The histogram of resampling probabilities in Figure 12 confirms that many ensembles have essentially been rejected, having very small resampling probabilities. Thus only samples close to  $\mathbf{a}=0.75$  have been resampled. Boxplots of the prior and posterior samples are shown in Figure 13, confirming this.

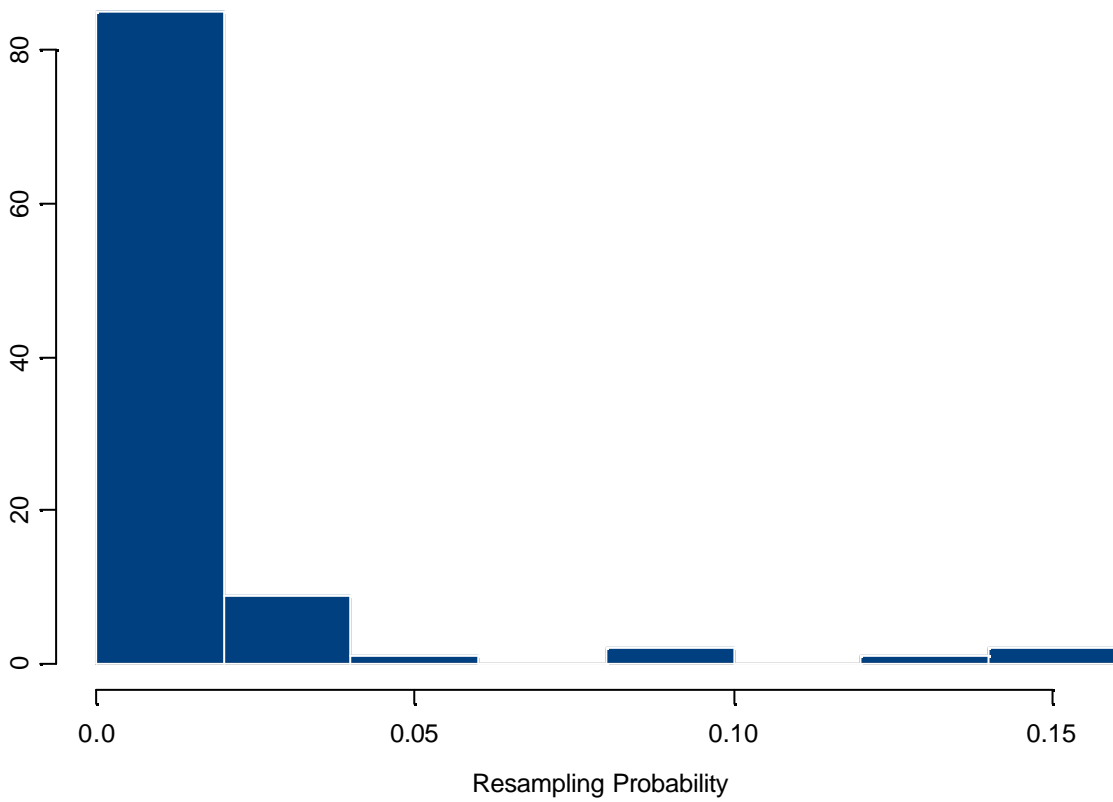
The method has clearly performed as we would expect.



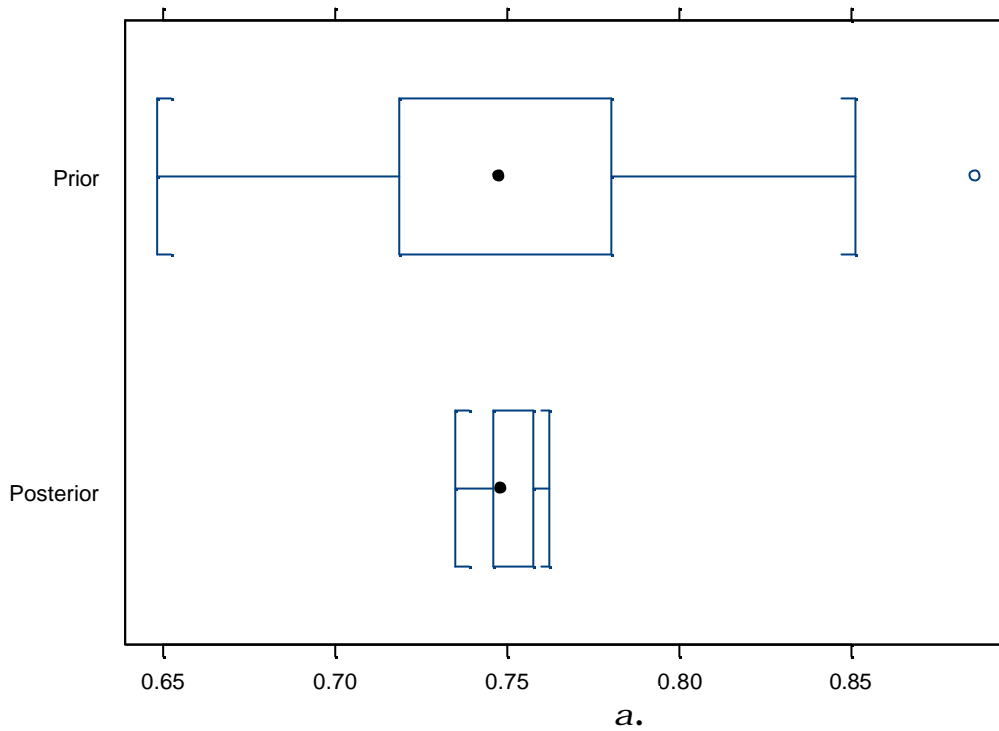
**Figure 10** Simulated data and ensembles with reduced measurement error.



**Figure 11** Error sum of squares as a function of  $a$ .



**Figure 12** Histogram of resampling probabilities.



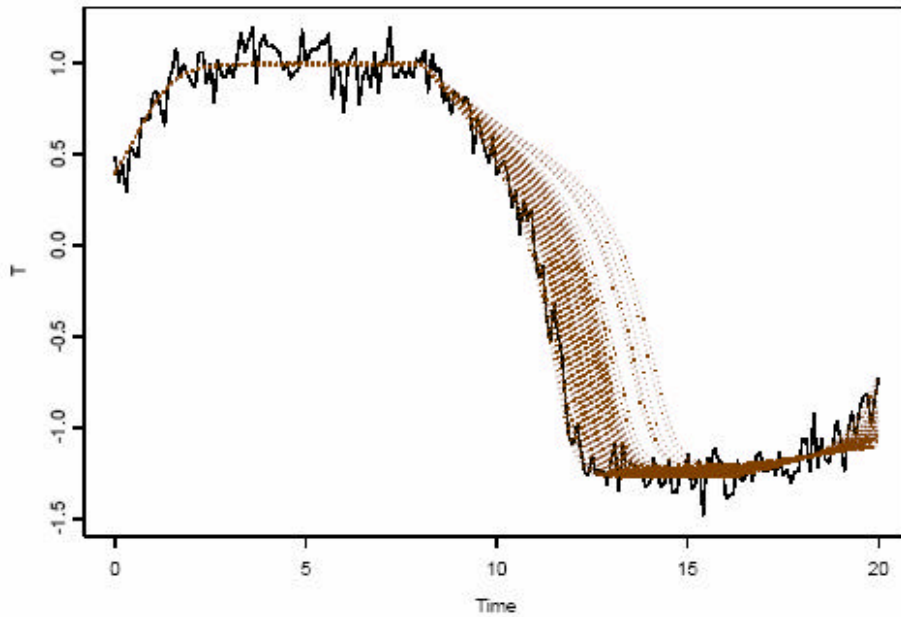
**Figure 13** Boxplots of prior and posterior samples for **a**.

Frequency Table of Ensembles Selected																				
5	8	9	11	12	14	31	32	34	36	38	45	48	50	60	62	66	87	88	92	98
39	12	1	10	144	20	21	17	40	68	1	23	130	110	24	29	33	25	124	98	31
Summary statistics for ALPHA																				
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.															
0.7349020	0.7456840	0.7479229	0.7501030	0.7573002	0.7618743															
Standard Error= 0.00608																				
95% Credibility Interval= [0.741, 0.761]																				

**Exhibit 3** Posterior summary statistics for **a**.

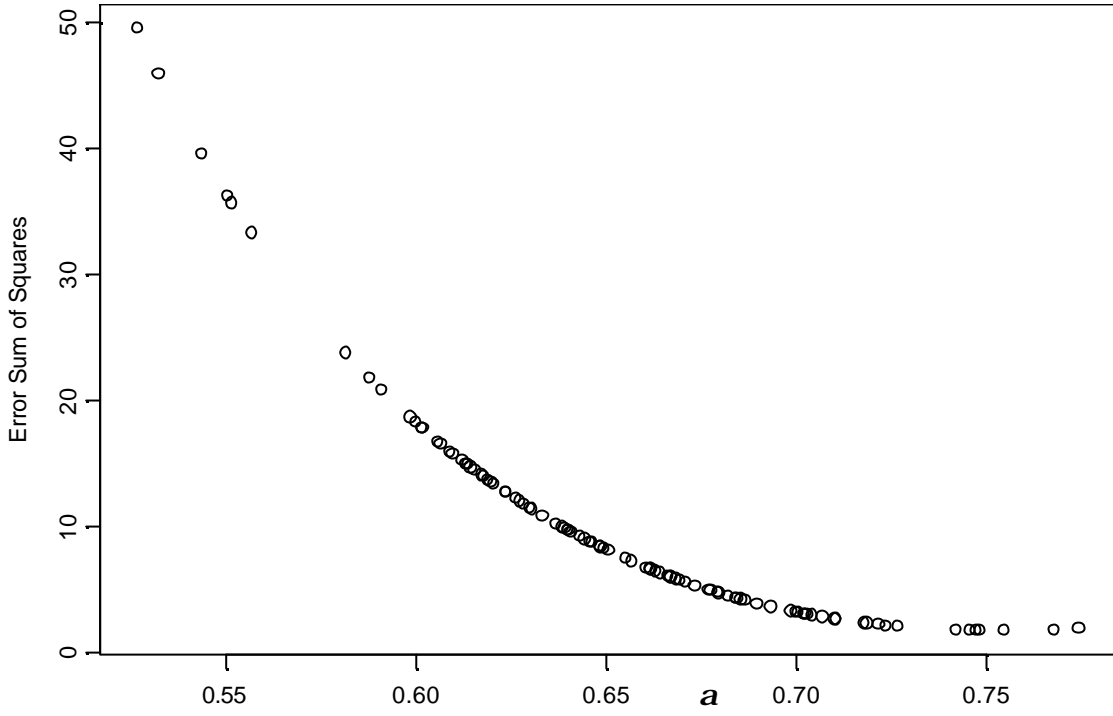
#### 4.4 Biased Prior

We test the method further now by assuming that the prior is substantially biased, so that *a priori*  $a \sim N(0.65, 0.05^2)$ . The simulated data and ensembles are shown in Figure 14, and not surprisingly most of the ensembles are biased.

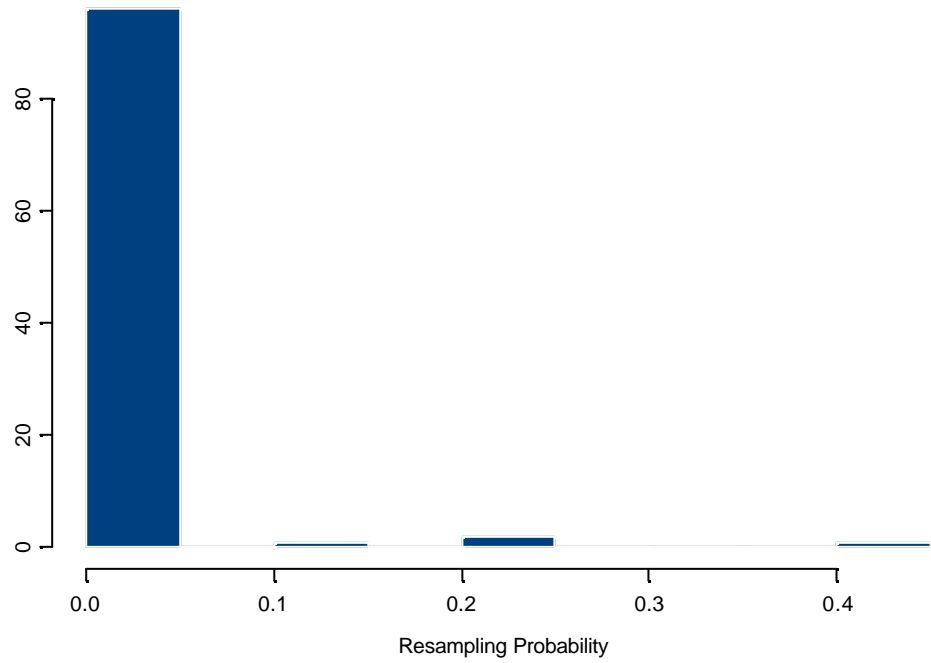


**Figure 14** Simulated data and ensembles with a biased prior for  $\mathbf{a}$ .

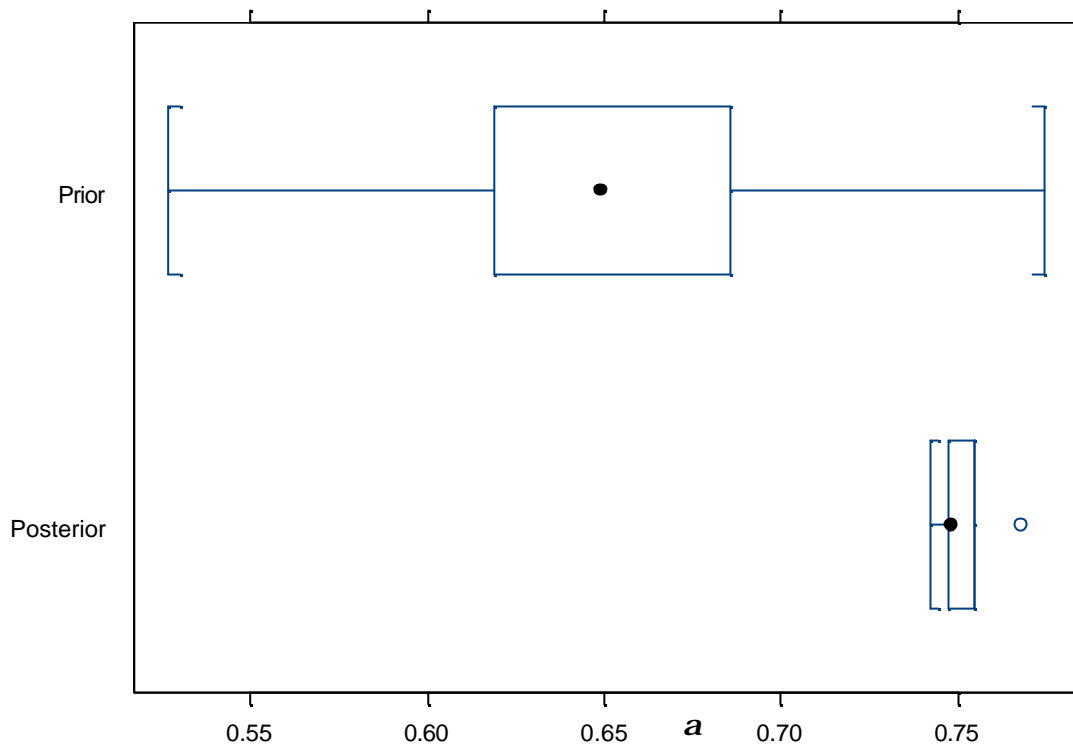
The error sum of squares is plotted in Figure 15, and it is clear that values of  $\alpha$  near 0.75 (the true value) are favoured. Because of the biased prior not many ensembles are consistent with this, as is evident from the resampling probabilities shown in Figure 16. It is clear from the boxplots of prior and posterior samples in Figure 17 that the posterior has been strongly influenced by the data, as we would hope. The biased prior has not had an unduly damaging impact on posterior inference for  $\mathbf{a}$ , and the posterior 95% credibility interval is [0.746, 0.755]. Only 5 ensembles were resampled, but this is an encouraging result.



**Figure 15** Error sum of squares as a function of  $a$ .



**Figure 17** Histogram of resampling probabilities.

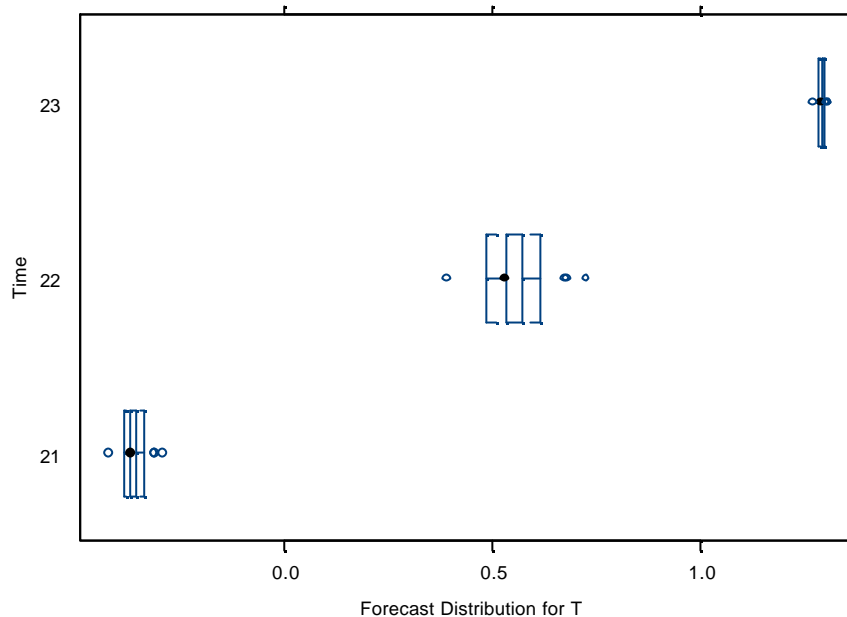


**Figure 17** Boxplots of prior and posterior samples for  $a$ .

## 5 Forecasting

By forecasting we mean projecting the model into the as yet unobserved 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.

We continue with the biased prior example examined previously, and derive probability distributions for the forecasts at times 21, 22 and 23 having observed the system up to time 20. The results are shown in Figure 18. We see that uncertainty increases from time 21 to 22, but due to the rapid upswing to time 23 there is little room for uncertainty. The key point to note here is that it's very easy to provide probability distributions for forecasts using this approach. If data subsequently become available for times 21 to 23, then an assessment of model skill could be undertaken. Alternatively some data could be withheld from model-fitting to assess predictive skill.



**Figure 18** Forecast distributions for biased prior example.

## 6 Discussion and Conclusions

We have applied the importance sampling-resampling method for obtaining approximate posterior samples from a simple statistical-physical model for the El Niño-Southern Oscillation. The method uses Bayesian hierarchical methods to integrate observations and physics. We have applied this physical-statistical model in both model-fitting and forecasting modes. The results are very encouraging, especially in the capacity of these methods to capture uncertainty in parameters and forecasts. Thus far however we have only used a very simple physical model. Despite this, the Bayesian approach is highly intuitive and simple to use once a reasonable algorithm has been developed.

The results were produced using S-Plus® software, without the use of DLLs for example to speed up the calculations required. Even so, especially in model fitting mode, the computations are reasonably fast. With more complex physical models it will be necessary to make use of a lower level language (FORTRAN or C) for speed; the algorithms adopted could also be made more efficient, particularly for forecasts.

The importance sampling algorithm appears to be very acceptable on current evidence. It is likely however to become inefficient in the absence of strong prior information. We tested this using a biased prior and the method did perform well. It may well be a different story if many parameters have priors far from values implied by the data. Campbell (2004) noted a number of possibilities from the literature for developing a more adaptive algorithm. Another promising possibility has been developed by Haario (2001); for an interesting hydrological application see Marshall et al. (2004).

In the case studies developed here the observations were statistically independent, given the physical model. In practice it is likely that a more general data model will be required. It is quite straightforward to build time-dependence into the modelling, but as this wasn't the purpose of this report it was not pursued here.

The key challenge now is to develop an application with a complex physical component, and our particular interest is in climate prediction. We note that if prediction of rainfall is important, then the physical model need not be for rainfall but for some key variables thought to influence rainfall. Simulated physical observations from a climate model could also be used, rather than a physical model *per se*. In this way the model can at least be constrained by some suitable physics.

**Acknowledgements** The work described here are being developed with the support of the Indian Ocean Climate Initiative (IOCI). IOCI is supported financially and in-kind by the WA State Government, by CSIRO and the Commonwealth Bureau of Meteorology. I am grateful to my colleagues Brent Henderson and Richard Morton for their comments on an earlier draft.

## References

- Bernardo, J. M. and Smith, A. F. M. (1994) *Bayesian Theory*, John Wiley & Sons, Chichester.
- Campbell, E. P. (2004). An introduction to physical-statistical modelling using Bayesian methods. CSIRO Mathematical and Information Sciences, Perth, Western Australia. Technical Report 2004/49, 18 pages.
- Haario, H., Saksman, E. and Tamminem, J. (2001). An adaptive Metropolis algorithm. *Bernoulli*, **7**, 223-242.
- Kivman, G. A. (2003). Sequential parameter estimation for stochastic systems. *Nonlinear Processes in Geophysics*, **10**, 253-259.
- Marshall, L., Nott, D. and Sharma, A. (2004). A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modelling. *Water Resources Research*, **40**.
- Suarez, M. J. and Schopf, P. S. (1988). A delayed action oscillator for ENSO. *Journal of the Atmospheric Sciences*, **45**, 3283-3287.

## Appendix: Calculation of Resampling Probabilities

It is possible to derive a very general expression for the resampling probabilities, which we may then apply to the specific problem of interest to us here. We suppose that a  $p$ -dimensional physical system  $\{X_t : t \in T\}$  is measured with a normally distributed error,

yielding an observed series  $\{\tilde{X}_t : t \in T\}$ . The structure of this error is assumed to have mean 0 and variance matrix  $\Sigma = \text{diag}(\mathbf{t}_t^2 : i=1, \dots, p)$ . Thus:

$$\begin{aligned} [\tilde{X} | X] &= \prod_i \prod_{l=1}^p (2p\mathbf{t}_t^2)^{-\frac{1}{2}} \exp\left[-(\tilde{X}_t^l - X_t^l)^2 / 2\mathbf{t}_t^2\right] \\ &= \prod_{l=1}^p (2p\mathbf{t}_t^2)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\mathbf{t}_t^2} \sum_t (\tilde{X}_t^l - X_t^l)^2\right], \end{aligned} \quad (4)$$

where  $n$  represents the number of sampling points. If we write the residual sum of squares over time for the  $i$ th variable as  $RSS_X^l$  then we may write this as

$$[\tilde{X} | X] = \prod_{l=1}^p (2p\mathbf{t}_t^2)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\mathbf{t}_t^2} RSS_X^l\right].$$

Given a collection of ensembles  $\{X_i : i=1, \dots, e\}$ , then the resampling probabilities are

$$\begin{aligned} q_i &= [\tilde{X} | X_i] / \sum_j [\tilde{X} | X_j] \\ &= \frac{\prod_{l=1}^p (2p\mathbf{t}_t^2)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\mathbf{t}_t^2} RSS_i^l\right]}{\sum_{j=1}^e \prod_{l=1}^p (2p\mathbf{t}_t^2)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\mathbf{t}_t^2} RSS_j^l\right]} \\ &\Rightarrow q_i^{-1} = 1 + \sum_{j \neq i} \exp\left[-\sum_{l=1}^p \frac{RSS_i^l - RSS_j^l}{2\mathbf{t}_t^2}\right]. \end{aligned} \quad (5)$$

□

In practice we could use this formula, or we could use equation (4) directly. It is likely however that the exponent in (4) will become very large and negative, in which case (5) would be more stable. It is also instructive to examine the form of (5), which shows a direct link between the goodness of fit of the ensembles and the probability of resampling.

We note that the observational model here is very straightforward, assuming mutual independence. In forming the likelihood defined by (4) it would be a relatively trivial matter to incorporate time dependence for example. We will not examine this issue here, but have noted it in the discussion.

First Draft: 22 June 2004

Second Draft: 28 June 2004

For internal review: 12 July 2004

Revisions: 6 September 2004

Final version: 14 September 2004