Australian Meteorological and Oceanographic Journal 59 (2010) 7-16

A Bayesian approach to state and parameter estimation in a Phytoplankton-Zooplankton model

Emlyn Jones[1], John Parslow[1] and Lawrence Murray[2]
[1]CSIRO Marine and Atmospheric Research, Hobart, Tasmania, Australia
[2]CSIRO Mathematical and Information Sciences, Floreat, Western Australia, Australia

Complex marine biogeochemical (BGC) models are now being used to inform management decisions at a variety of scales, from local coastal management issues through to the global effects of climate change. A majority of BGC models are still deterministic in nature with model tuning and calibration performed in a heuristic manner. This method does not allow for a quantitative estimate of model or parameter uncertainty. If these models are reformulated in a physical-statistical framework, using a stochastic process model, formal state and parameter estimation routines can be implemented, yielding quantitative estimates of model uncertainty. We have performed twin experiments using an idealised stochastic-dynamic non-linear phytoplankton-zooplankton model to trial two Markov Chain Monte Carlo (MCMC) Algorithms. The first uses a Particle Filter (PF) with a Metropolis-Hastings (MH) update step for state-estimation embedded within a MH MCMC for hyper-parameter estimation; we have named this approach MH-PF-MH. The second approach uses Gibbs sampling for state estimation and MH MCMC over hyper-parameters; referred to as MH-Gibbs. Both algorithms performed well in the twin-experiments, allowing both state and parameter estimation. The hybrid MH-Gibbs is more efficient than the MH-PF-MH algorithm, forming a reliable posterior sample with up to 99.9% fewer model trajectories. However, the MH-PF-MH algorithm is expected to be more flexible in its implementation.

Introduction

Within the last 15 years there have been significant advances in statistical techniques available for data assimilation (DA), largely driven by a combination of the availability of high performance computing and dense observational data-sets (Lorenc 2003; Evensen 2007). Marine Biogeochemical (BGC) modelling, unlike the disciplines of numerical weather prediction and operational oceanography, has been slow in adopting these new and evolving methods. This is not surprising given the lack of high quality data-sets with adequate temporal and spatial resolution. Recent advances in sensor platforms (e.g. coastal gliders, nutrient sensors, etc.) and remote sensing algorithms are providing new data streams that could be used in data assimilation routines for biogeochemical state and parameter estimation.

Corresponding author address: Emlyn Jones, CSIRO Marine and Atmospheric Research, GPO Box 1538, Hobart, TAS 7001, Australia.
Email: emlyn.jones@csiro.au

Unlike atmospheric and hydrodynamic models, which are based on sound physical laws such as the Navier-Stokes equations, biogeochemical models represent processes which are highly parametrised and often based on empirical studies (Miller 2004). It is becoming more apparent that the traditional form of the deterministic biogeochemical model using fixed parameters is an unrealistic representation of many key processes included in the current models. The natural diversity of phytoplankton and zooplankton assemblages is represented by a small number (often 2-3) of functional groups, and the variability of community composition within functional groups is ignored. This variability can be overcome by reformulating many of the deterministic parametrisations with stochastic parameters. Given the high prior uncertainty in many biogeochemical parameters, a combination of DA and parameter estimation routines must be invoked to adequately constrain the stochastic parameters.

There are three broad techniques that have been used in DA routines within the BGC modelling community; Variational/Adjoint techniques (Li et al. 2008), a broad range of Kalman filters (Natvik and Evensen 2001; Eknes and Evensen 2002; Bertino et al. 2003) and Markov Chain Monte Carlo (MCMC) algorithms (Dowd 2006; Dowd 2007; Arhonditsis et al. 2008). Of the three techniques, MCMC techniques are the most flexible and have the fewest assumptions; however the computational burden associated with these methods can be prohibitive. Harmon and Challenor (1997) proposed the use of the Metropolis-Hastings MCMC algorithm for state estimation in a simple Nutrient-Phytoplankton-Zooplankton-Detritus (NPZD) model (Miller 2004), which was further extended by Dowd (2006, 2007) using a stochastic growth parameter for phytoplankton. This approach also yielded probability distributions for the growth parameter at each assimilation step, however the hyper-parameters (mean and variance) of the growth parameter were constant. There have been few if any studies to our knowledge that include both hyper-parameters and state in the inference routine for stochastic biogeochemical models.

In this paper we focus on applying MCMC techniques to estimate the mean and variance (hyper-parameters) of a single stochastic parameter in an idealised biogeochemical model. Twin experiments are used to test the inference techniques, where all the statistical parameters are known, and each technique is evaluated for computational efficiency and ability to estimate the hyper-parameters.

Methods

Model description

The differential equations that describe the simple phytoplankton-zooplankton model are based on a modified version of the traditional Lotka-Volterra predator-prey model (Lotka 1920; Volterra 1926):

$$\frac{dP}{dt} = \alpha P - cPZ \qquad \ldots 1$$

$$\frac{dZ}{dt} = ecPZ - m_l Z - m_q Z^2 \qquad \ldots 2$$

Here P and Z are the state variables of Phytoplankton (prey) and Zooplankton (predator) respectively. In more realistic models, P and Z would share a common currency of nitrogen (or carbon) in kilograms of nitrogen (carbon) per cubic meter, kg N m^{-3}; in this simple example the absolute units are non-consequential. The parameter α is the phytoplankton growth rate, c is the zooplankton clearance rate, e is the zooplankton growth efficiency, and m_l and m_q are the linear and quadratic zooplankton mortality terms. These equations would be neutrally stable except for the quadratic mortality term, $m_q Z^2$. In the positive quadrant the deterministic model has a globally stable equilibrium. To convert the deterministic model into a stochastic model, the phytoplankton growth parameter α is sampled from a normal distribution with a mean, μ_α, and a standard deviation, σ_α, $N(\mu_\alpha,\sigma_\alpha)$, and updated at discrete time periods, in this case one day. The parameters μ_α and σ_α characterise the distribution of the stochastic system parameters, and are referred to here as 'hyper-parameters'. The other parameters (c, e, m_l and m_q) are assigned known constant values. We denote the state vector containing P and Z as $\mathbf{X}$, where $\mathbf{X}(t)$ is the state at time t, similarly, the vector of observations as $\mathbf{Y}$, and the hyper-parameter vector as θ. The stochastic dynamic model described by Eqns 1 and 2 prescribes $[\mathbf{X}(t+1)|\mathbf{X}(t),\theta]$ and an observational model is defined as $[\mathbf{Y}(t)|\mathbf{X}(t)]$.

In our simple twin experiment, we have used the stochastic equations to produce a single realisation of $\mathbf{X}(t)$, $0<t<T$, for known hyper-parameters, and $\mathbf{Y}(t)$ is a synthetic data-set generated by:

$$\mathbf{Y}(t) = \mathbf{X}(t).\exp(\zeta), \qquad \ldots 3$$

where $\zeta \sim$ IID $N(0,\sigma_{obs})$. The log-normal error model was adopted as errors in the estimates of plankton density are typically better represented by log-normal multiplicative error than by additive normal error (Campbell 1995). Another benefit of the log-normal multiplicative error model are that all synthetic observations would be non-negative with a constant coefficient of variation. The observation errors are assumed to be independent over time, reflecting either analytical error or (more likely) uncorrelated small-scale variation in concentrations.

The synthetic data-set was created using the following parameters from the quasi-linearised model used in the Gibbs sampling:

$\mu_\alpha = 0.25$

$\sigma_\alpha = 0.1$

$\sigma_{obs} = 0.2$

The value for μ_α has been chosen such that the quasi-linearised form of the equations remain stable and would be considered low but realistic in either low light or nutrient limited conditions, and for this reason we have chosen an accompanying low value for σ_α. The observation noise, $\sigma_{obs} = 0.2$ is a realistic measurement error level for observations of phytoplankton biomass.

The combined data assimilation and parameter estimation inference routine is designed to obtain a large random sample from the posterior distribution $[\mathbf{X}(1)\ldots\mathbf{X}(T)\ \theta|\mathbf{Y}(1)\ldots\mathbf{Y}(T)]$. Two different MCMC algorithms are trialled to recover the hyper-parameters from the synthetic data. The following algorithms are used in the twin experiments:

- Particle Filter (PF) with Metropolis-Hastings (MH) update step for state-estimation embedded within a MH MCMC for hyper-parameter estimation; we have named this approach MH-PF-MH.
- Gibbs sampling for state estimation and MH MCMC over hyper-parameters; referred to as MH-Gibbs.

The differential equations are solved using a Runge-Kutta 4/5 solver in the particle filter, while the Gibbs sampling uses a quasi-linearised form of the equations solved using a first order Euler approximation.

The synthetic observations of Phytoplankton are assimilated into the model at daily intervals. Neither of the approaches attempted to assimilate Zooplankton observations, so these experiments also test our ability to recover information about an unobserved state variable. This is highly relevant, because in situ and remote sensors are readily available for phytoplankton biomass, but not zooplankton biomass.

Particle filter with Metropolis-Hastings update

The stochastic PZ model prescribes [X(t+1)|X(t)], while the observation model gives the likelihood, [Y(t)|X(t)]. A forwards particle filter with a Metropolis-Hastings measurement update is implemented in state space to estimate [X(t)|Y(l)... Y(t)](Harmon and Challenor 1997; Doucet et al. 2001; Dowd 2006; Dowd 2007). The M-H MCMC algorithm is less affected by the problems of sample impoverishment associated with other particle filtering algorithms such as sequential importance sampling and sequential importance resampling. This algorithm favours particles with a high likelihood, yet as it only evaluates a ratio between the last accepted particle and the present proposal, a spread of particles is typically maintained. A pseudo algorithm for the Particle Filter with Metropolis-Hastings (PF-MH) using N particles is implemented in the following way:

Algorithm 1: PF-MH Algorithm for state estimation

The observations with multiplicative log-normal noise are used as the prior for X at t = 0.

- At t = 0, draw the initial values of X(0) from a prior which is parametrised as:
  $X(0) \sim Y(0) \,.\, \exp(\zeta)$
  The MH-PF is then run from t = 1…T
- For t = 1:T
  - Draw N random samples in a resampling with replacement fashion from [X(t-1)|Y(1)…Y(t-1)] and propagate them forward using the stochastic PZ model, this gives [X*(t)|X(t-1)]
  - Initialise the chain X_i=1(t) with Y(t)
  - For i = 2…N

    $$LR = \frac{[Y(t)|X_i^*(t)]}{[Y(t)|X_{i-1}(t)]}$$

    Draw $\psi \sim U(0,1)$

    if min(1,LR)> ψ
      $X_i(t)=X_i^*(t)$
    else
      $X_i(t)=X_{i-1}(t)$
    end
  - end
- end

The difficulty arises when we want to start estimating the hyper-parameters μ_α and σ_α. There have been few if any studies to our knowledge that have specifically dealt with the problem of estimating the hyper-parameters of a stochastic parameter in a non-linear time evolving model. Herein lies a fundamental difference between our study and that of Dowd (2006; 2007). The marginal likelihoods of the observations given the hyper-parameters can be approximated by Eqn 4 (for a derivation see Appendix A), if the number of particles is large enough (in this instance more than 200 particles were required to achieve an estimate of the marginal likelihood independent of ensemble size):

$$[Y(T)...Y(l)\,|\theta] \approx \prod_{i=1}^{T} \left\langle [Y(t)|X^*(t)] \right\rangle \quad ...4$$

The <> brackets denote the mean value of the likelihood function for all trial particles (X*(t)) at time t before any rejection has been implemented. An MH routine is then implemented in hyper-parameter space whereby a bi-variate Gaussian proposal density **Q** is used to propose the next value in the Markov Chain. The following pseudo code is used:

Algorithm 2: MH–PF-MH Algorithm for hyper-parameter estimation

- $\theta_{j=1}$ is initialised by drawing μ_α and σ_α from a prior, parametrised by an IID bi-variate normal distribution:

  $$\theta = \begin{pmatrix} \mu_\alpha \sim N(1,1) \\ \sigma_\alpha \sim N(1,1) \end{pmatrix}$$

  We assume there is no information regarding the initial values of $\theta_{j=1}$ and have therefore assigned an arbitrary, broad prior.

- The PF-MH (Algorithm 1) is then run to evaluate [Y(T)…Y(1)|θ_j].

  The following step is iterated NE times (where NE is the number of ensembles to run, in this case NE = 10000) to build a Markov chain of hyper-parameters.

- for j = 2:N
  - A new proposal (θ*) is generated using the proposal density, **Q**:

    $$\theta^*_j=\theta_{j-1}+Q$$

    $$\text{where } Q = \begin{pmatrix} Q_\mu \sim N(0,0.02) \\ Q_\sigma \sim N(0,0.01) \end{pmatrix}$$

  - The PF-MH (Algorithm 1) is then used to evaluate:

    $$[Y(T)...Y(l)\,|\theta] \approx \prod_{i=1}^{T} \left\langle [Y(t)|X^*(t)] \right\rangle$$

  - A M-H step is then used to either accept or reject θ*

    $$ELH = \frac{[Y(T)...Y(1)|\theta_j^*]}{[Y(T)...Y(1)|\theta_{j-1}]}$$

    Draw $\psi \sim U(0,1)$

    if min(1,ELH)> ψ
      Accept θ*, $\theta_j = \theta_j^*$
    else
      Reject θ*, $\theta_j = \theta_{j-1}^*$
    end
- end

MH-Gibbs sampling

The standard Gibbs sampler (Gelman 2003; Gamerman and Lopes 2006) generates a Markov Chain to produce a random sample from the joint posterior distribution $[X(1)...X(T)\ \theta \mid Y(1)...Y(T)]$ by sampling repeatedly, for $t = 1...T$, from the conditional distribution $[X(t) \mid X(1)...X(t\text{-}1)\ X(t+1)...X(T)\ \theta\ Y(1)...Y(T)]$, and then from $[\theta \mid X(1)...X(T)\ Y(1)...Y(T)]$. The chain is initialised at k=1 with observations and a value of θ drawn randomly from the hyper-parameter prior distribution. At the kth iteration, we need to obtain a random sample of $X_k(t)$ from the conditional distribution:

$$[X_k(t) \mid X_k(1)\ldots X_k(t\text{-}1)\ X_{k\text{-}1}(t+1)\ldots X_{k\text{-}1}(T)\ Y(1)\ldots Y(T)\ \theta_{k\text{-}1}] \quad \ldots 5$$

Due to the Markovian nature of the process model, the conditional distribution (Eqn 5) is proportional to:

$$[X_k(t)|X_k(t\text{-}1)\ \theta_{k\text{-}1}].[X_{k\text{-}1}(t+1)|X_k(t)\theta_{k\text{-}1}].[Y(t)|X_k(t)] \quad \ldots 6$$

$X_k(1)$ and $X_k(T)$ are special cases. $X_k(1)$ is sampled randomly from the conditional distribution proportional to:

$$[X_{k\text{-}1}(1)|X_k(2)\theta_{k\text{-}1}].[Y(1)|X_k(1)] \quad \ldots 7$$

And $X_k(T)$ is drawn from the conditional distribution proportional to:

$$[X_k(T)|X_k(T\text{-}1)\ \theta_{k\text{-}1}].[Y(T)|X_k(T)] \quad \ldots 8$$

Attempts to sample randomly from the conditional distribution (Eqn 6) by using the model to generate a proposal for $X_k^*(t)$ from $[X_k(t)|X_k(t\text{-}1)\theta_{k\text{-}1}]$, and then rejection sampling against $[X_{k\text{-}1}(t+1)|X_k(t)\ \theta_{k\text{-}1}].[Y(t)|X_k(t)]$ failed, due to acceptance probabilities less than 1 in 10000. The quasi-linearisation described in Appendix B allows us to compute an analytical formula for the joint conditional (Eqn 6), and draw directly from that joint conditional, thereby avoiding entirely the acceptance issue.

The hyper-parameters, θ_k are drawn from the conditional distribution:

$$[\theta_k|X_k(1)\ldots X_k(T)\ Y(1)\ldots Y(T)] \quad \ldots 9$$

which simplifies to:

$$[\theta_k|X_k(1)\ldots X_k(T)] \sim [\theta]^{prior}.\Pi_t[X_k(t+1)|X_k(t)\ \theta_k] \quad \ldots 10$$

If one chooses appropriate (conjugate) priors for θ, it is possible to construct an analytic expression for (Eqn 10) using the quasi-linear approximation, and sample directly for θ_k. That would constitute a straight Gibbs sampler for $X(1..T)$ and θ. In the examples here, we have employed a more flexible hybrid approach, in which θ is updated from the conditional distribution (Eqn 10) using a MH step. Specifically, at step k, candidates θ_k^* are drawn from the prior for $[\theta]^{prior}$, and are accepted or rejected depending on whether a random variate drawn from U(0,1) is less or greater than the likelihood ratio:

$$\Pi_t[X_k(t+1)|X_k(t)\ \theta_k^*]\ /\ \Pi_t[X_k(t+1)|X_k(t)\ \theta_{k\text{-}1}] \quad \ldots 11$$

Note that because $X(1)...X(T)$ are given, the quasi-linearisation allows us to compute this ratio quickly and efficiently.

Results

The effect of the stochastic variable, α, causes a change in the model behaviour, from neutrally stable to a more quasi-cyclic nature (Miller 2004). In this study α is treated as a fixed variable over a discrete time period of one day and is then resampled from the distribution $N(\mu_\alpha,\sigma_\alpha)$.

To assess the robustness of both algorithms in state and parameter estimation, a synthetic observational data-set was generated using both a first order Euler approximation to the equations and a Runge-Kutta 4/5 ordinary differential equation (ODE) solver. Neither of the algorithms were sensitive to the method used to solve the ODEs to generate the synthetic observations. We therefore use the synthetic observations generated by the first order Euler approximation in all further twin experiments. True values for P and Z are saved along with the synthetic observations generated using Eqn 3.

Both methods adequately handle the state estimation with the MH-Gibbs method generating a smoother median trajectory (Fig. 1). The confidence intervals using both approaches broaden when the Phytoplankton concentration peaks and narrow at the local minima in the troughs, which is attributed to the choice of noise model. A result of not assimilating Zooplankton observations can be seen in observations falling outside the 95 per cent credibility intervals. Interestingly, most of the 'true' values fall within the credibility intervals.

To visually inspect the ability of both algorithms to recover the true values of P and Z, a scatter plot of the true values and median values of P and Z was generated (Fig. 2). The root mean square error (RMSE) between the true values and the median values for P and Z highlighted the superior performance of the MH-Gibbs algorithm. Furthermore, the results of a regression analysis yielded correlation coefficients of 0.85 and 0.94 for the Metropolis-Hastings and MH-Gibbs respectively. For the purposes of state estimation, the MH-Gibbs algorithm substantially outperforms the Metropolis-Hastings.

The predicted median values for μ_α were 0.250 and 0.245 for the MH-PF-MH and MH-Gibbs algorithms respectively (Fig. 3). When compared with the true value of 0.25, both schemes performed remarkably well. The coefficient of variation (CV) for μ_α using the MH-PF-MH was 0.0443 while the CV using the MH-Gibbs approach was 0.0432. This appears as a slightly broader peak in the MH-PF-MH histogram.

Both the MH-PF-MH and MH-Gibbs slightly under predicted σ_α with values of 0.093 and 0.096 respectively, when compared with the true value of 0.1 (Fig. 4). Furthermore, the CV for σ_α using MH-PF-MH was 0.1553, while the MH-Gibbs approach did slightly better with a CV of 0.1379.

Fig. 1 The state estimates for the median concentrations (solid lines) of Phytoplankton (blue) and Zooplankton (red) using (a) MH-PF-MH and (b) MH-Gibbs sampling. Assimilated observations (+), true values (.) and the 95% credibility intervals for the state estimate (dashed lines) are also included.

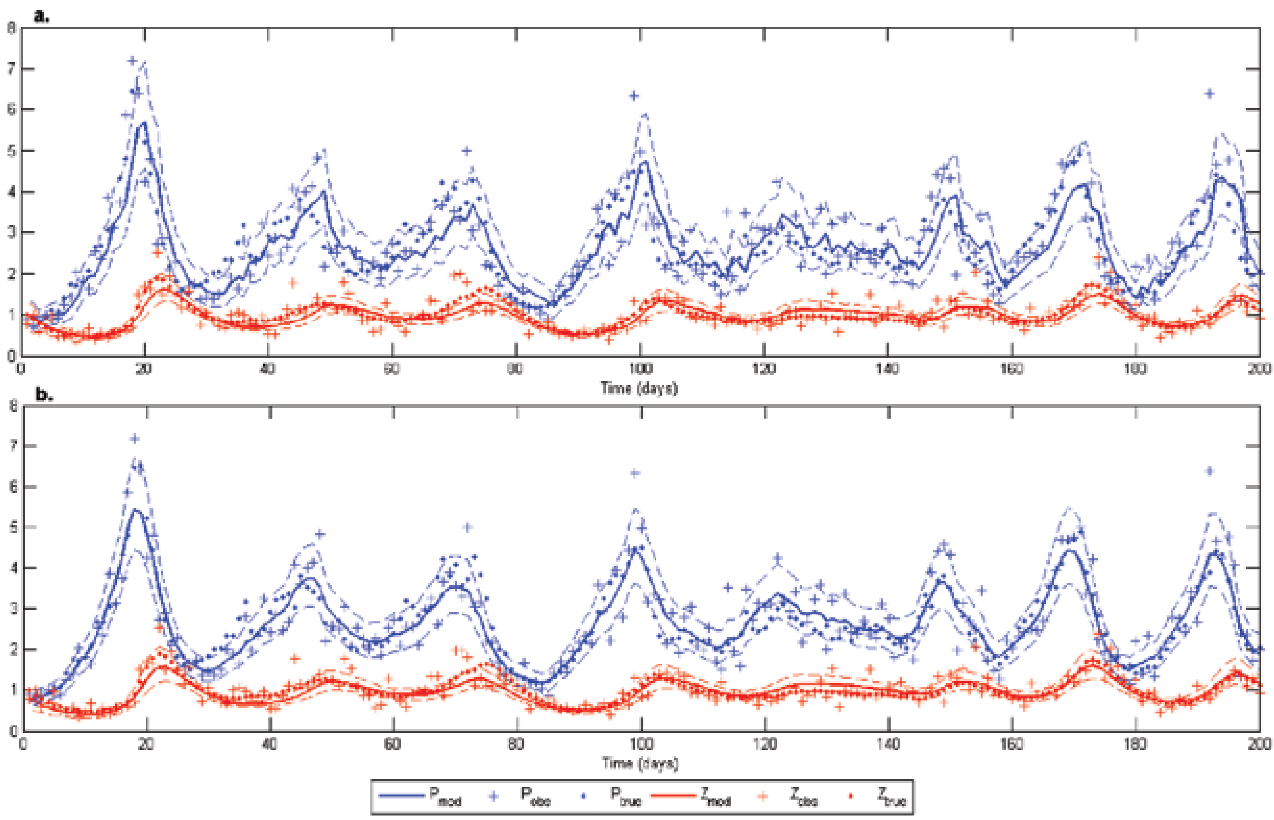


Fig. 2 Scatter plots of modelled values and true values for (a) phytoplankton and (b) zooplankton. Root mean square errors (RMSE) for each of the algorithms are shown in the lower right of each panel.

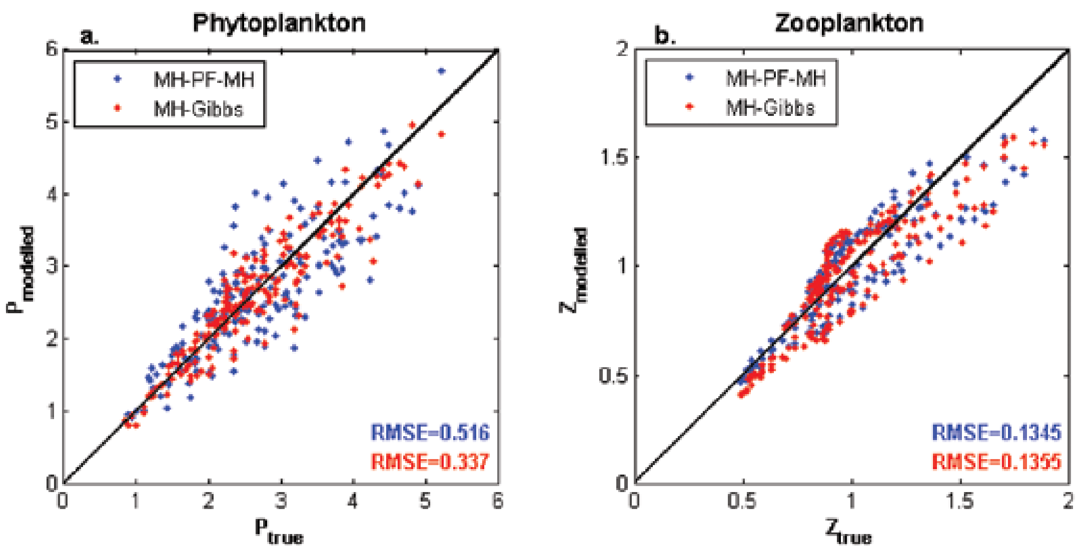


Fig. 3 Estimates for μ_α using MH-PF-MH (left) and MH-Gibbs sampling (right).

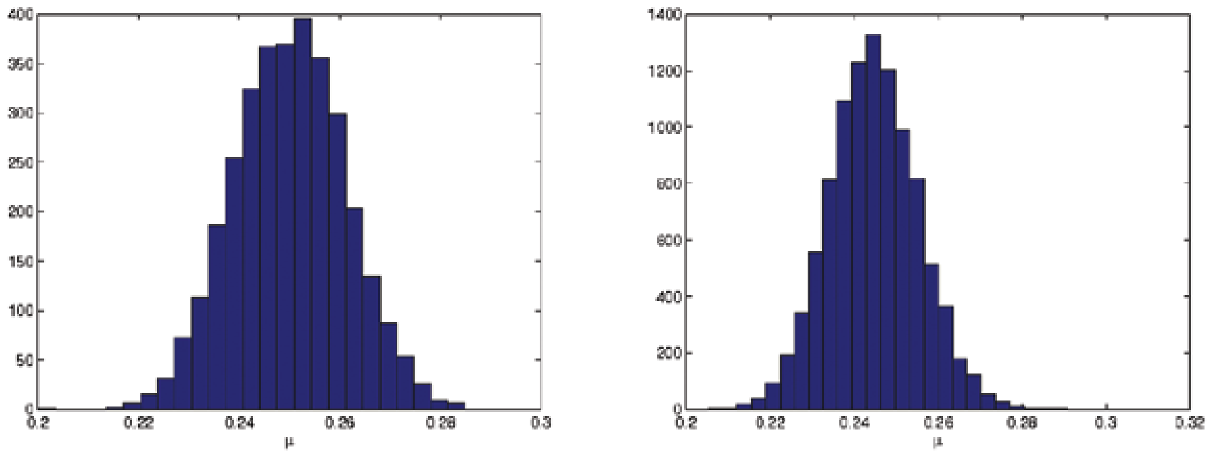
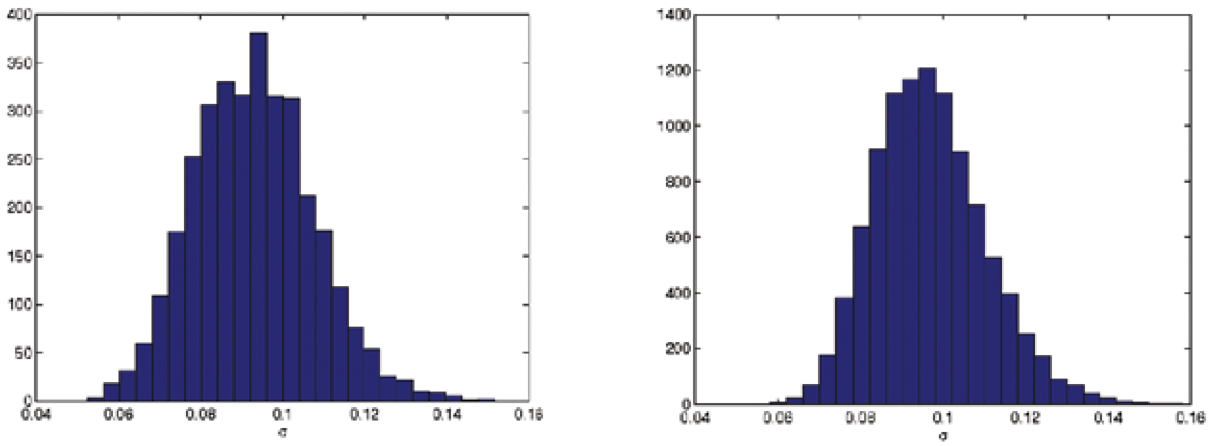


Fig. 4 Estimates for σ_α using MH-PF-MH (left) and MH-Gibbs sampling (right).



The differences in the marginal distributions from Figs 3 and 4 are obvious when the non-normalised 2D kernel density estimates (KDE) are examined (Fig. 5). The 2D KDE using the MH-PF-MH appears to be fairly symmetrical in shape with the exception of the peak. In this region there is a definite non-concentric peak located quite near the true values of μ_α and σ_α. The MH-PF-MH gave a broader posterior than the MH-Gibbs sampling algorithm.

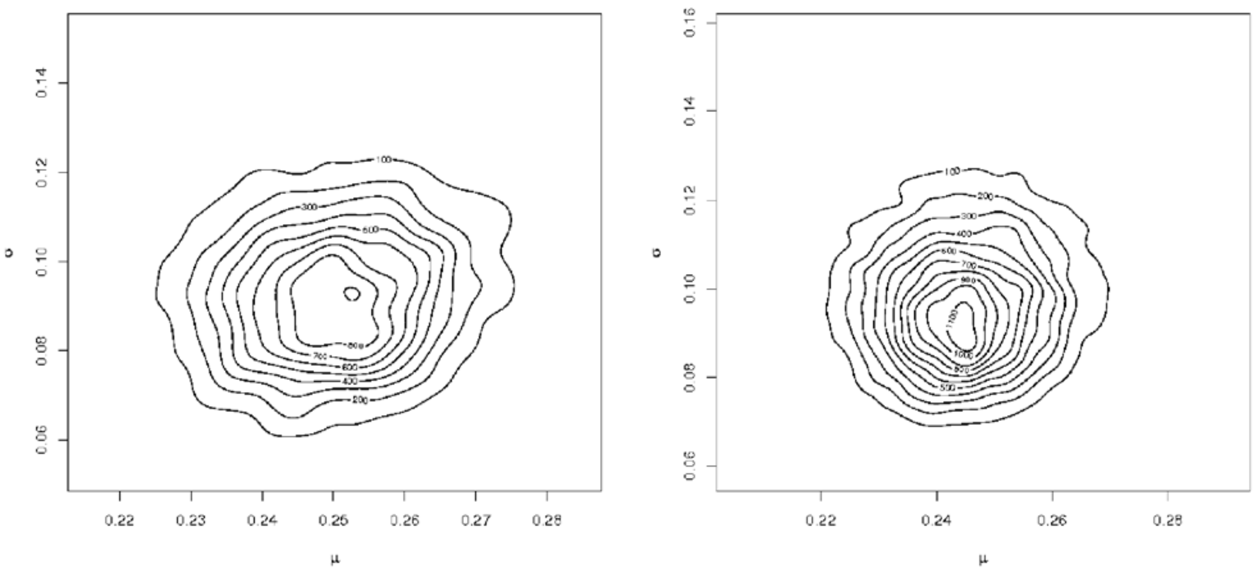
Discussion

The CVs for σ_α were higher than those for μ_α by a factor of three to four. It therefore appears that the data are not as informative about σ_α as they are for μ_α. This occurs regardless of the methods used and is likely to be related to the level of measurement error in the observation model.

The results reported in this study use the quasi-linearised model to generate the synthetic observations. A second experiment (not reported) was conducted whereby the synthetic observations were generated using a Runge-Kutta 4/5 solver. It is encouraging that neither of the methods are sensitive to the method used to generate the synthetic observations and furthermore the use of a generating model which differs slightly from the estimation model doesn't appear to bias the parameter estimates. This is a positive result given that real world processes don't necessarily follow a well defined process model.

The choice for the initial values of θ in the MH-PF-MH algorithm influences the burn-in time for the chain. Where the initial values of θ differ significantly from the true values, a longer burn-in period is required. In this study approximately 50 – 80 iterations of Algorithm 2 were required for the hyper-parameter chain to reach convergence. The proposal

Fig. 5 Estimates for non-normalised 2D kernel density using MH-PF-MH (left) and MH-Gibbs sampling (right).



density Q was tuned to maximise the mixing of the θ chain, while achieving an acceptance rate of approximately 20 per cent, which is suggested for a bi-variate Gaussian distribution (Gamerman and Lopes 2006).

The efficiency of the parameter estimation scheme is an important consideration when evaluating the relative performance of the two techniques. Typical runtime ratios for a single trajectory of a BGC model are around 1:300-600 as a single thread on a quad-core work station. The MH-Gibbs approach presented here is far more efficient than the MH-PF-MH for two reasons. Firstly the MH-PF-MH requires an ensemble size of 200 particles or more, whereas the MH-Gibbs method effectively integrates only one trajectory for each accepted hyper-parameter set. Moreover, the MH-Gibbs method is able to evaluate the marginal likelihood for new hyper-parameter proposals without computing a new trajectory. The MH-PF-MH approach needs to integrate the particle filter forward for each new hyper-parameter proposal to compute the marginal likelihood, and 80 per cent of proposals are rejected. The MH-Gibbs algorithm requires up to 99.9 per cent fewer trajectories and yields more accurate results than the MH-PF-MH algorithm, therefore, if efficiency and accuracy are the sole criteria on which these algorithms are judged, the MH-Gibbs approach is the clear winner.

However, the efficiency of the MH-Gibbs approach relies on using the quasi-linearisation to allow an explicit analytical representation of, and direct sampling from, the marginal joint distribution (Eqn 6). A version of the Gibbs model, which attempted to use rejection sampling to sample from the joint distribution (Eqn 6), failed due to extremely low acceptance ratios. Extension of the MH-Gibbs approach to more complex non-linear models, with more parameters, will therefore depend on developing a suitable linear approximation. The MH-PF-MH scheme can in principle be easily extended to multivariate assimilation in more complex non-linear models with numerous stochastic parameters. We plan to apply these hybrid methods to NPZD models (Fasham et al. 1990). While the simple quasi-linear approximation proved adequate in this example, it may not prove adequate in NPZD models, which can prove quite stiff (Hairer et al. 1987), and require higher order adaptive integrators.

Gibbs approaches are well-known to have limited ability to explore state space in the presence of highly correlated variables, and to be susceptible to being trapped in local modes of the posterior. In the current example, there appears to be only one mode. However, in the presence of multi-modal distributions, block Gibbs approaches may be more robust (Gamerman and Lopes 2006).

The advantage of the MH-PF-MH approach over other filtering approaches (e.g. Kalman filtering) is that any distribution can be represented and multimodal distributions could be approximated. However, in this case where the joint distribution of μ_α and σ_α appears to be approximately a bi-variate Gaussian, other approaches such as the Ensemble Kalman Filter (Evensen 2007) or variational methods (Lorenc 2003) could adequately represent this distribution. These approaches have not been trialled in this study and are an area that will be explored in future research as we begin working in higher dimensional systems.

The effect of altering the assimilation frequency was not reported in this study and remains an area of interest. Preliminary results suggest that MH-PF-MH ensemble size will need to be increased if the data are assimilated into the model at intervals greater than three days. Results from the MH-Gibbs algorithms suggest that if the assimilation interval is greater than three days, the Gibbs propagation using the quasi-linear approximation becomes unstable. Again, the adoption of block Gibbs approaches may help.

Conclusion

In this study we have presented two candidate MCMC techniques that could be used for both parameter estimation and data assimilation in marine biogeochemical models. In the twin experiments presented, both models performed well in both state and parameter estimation. The MH-Gibbs approach yielded slightly better constrained parameters estimates than the MH-PF-MH, however the present implementation depends on a quasi-linear approximation to the process model. The results indicate that observations are less informative in constraining the standard deviation of the stochastic parameters, but result in well constrained means for the stochastic parameters. It is likely that both schemes will require significant computational power as high dimensional problems are encountered.

Acknowledgments

The authors gratefully acknowledge the support and funding from the CSIRO Computational and Simulation Sciences (CSS) Transformational Capability Platform. We also wish to thank Nugzar Margvelashvili and two anonymous reviewers for their useful comments and suggestions relating to this manuscript.

Appendix A

Derivation of the hyper-parameters' conditional probability.

The M-H MCMC algorithm in hyper-parameter space requires the marginal likelihood:

$$[Y(T)\ldots Y(1)|\theta] \qquad \ldots A1$$

to be evaluated during the accept/reject step. Due to the implementation of a particle filter, it is possible to calculate A1 sequentially as part of the filter. We can write:

$$[Y(t+1)\ Y(t)\ \ldots\ Y(1)\ |\ \theta] = [Y(t+1)\ |\ Y(t)\ldots Y(1)\ \theta].[Y(t)\ |\ Y(t-1)\ldots Y(1)\ \theta].\ \ldots\ [Y(1)\ |\theta] \qquad \ldots A2$$

Assuming that observations and parameters are conditionally independent given the state, we introduce our process model:

$$[Y(t+1)\ X(t+1)\ |\ Y(t)\ \ldots\ Y(1)\ \theta] = [Y(t+1)\ |\ X(t+1)]\ .\ [X(t+1)\ |\ Y(t)\ \ldots\ Y(1)\ \theta] \qquad \ldots A3$$

Integrating over X(t+1), A3 yields:

$$[Y(t+1)\ |\ Y(t)\ \ldots\ Y(1)\ \theta] = \int [Y(t+1)\ |\ X(t+1)]\ .\ [X(t+1)\ |\ Y(t)\ \ldots\ Y(1)\ \theta]\ dX(t+1) \qquad \ldots A4$$

In the PF-MH method described by Algorithm 1, we start with a random sample from [X(t) | Y(t) … Y(1) θ] and use the stochastic model to generate a random sample X*(t+1) from [X(t+1 | Y(t) … Y(1) θ]. Note that X*(t+1) has been taken from the prior before any rejection step has taken place. Therefore, if the sample size of X*(t+1) is large enough, we can approximate the integral in Eqn A4 by the sample mean:

$$[Y(t+1)\ |\ Y(t)\ \ldots\ Y(1)\ \theta] = \langle[Y(t+1)\ |\ X^*(t+1)]\rangle \qquad \ldots A5$$

where the <> brackets denote the mean. Therefore, the marginal likelihood (A1) for the entire time series can be approximated by:

$$[Y(T)\ \ldots\ Y(1)\ |\ \grave{e}\ ] \approx \prod_{i=1}^{T} \left\langle [Y(t)|X^*(t)] \right\rangle \qquad \ldots A6$$

Appendix B

MH-Gibbs Sampling Implementation Procedure

MH-Gibbs sampling requires the evaluation of [X(t+1) | X(t)] in the calculation of the transition probability, r, from X(t) to X(t+1). There is no simple way to sample from this distribution using the traditional forward integration of the ODEs. An alternative method is to use a first order Euler approximation to the ODEs to explicitly calculate [X(t+1) | X(t)]. This approximation is outlined below with a time step (Δt =1). However, these equations can become stiff, and a quasi-linearised form is used. It is possible to divide each of the ODEs by P and Z respectively to obtain:

$$\frac{d \ln P}{dt} = \alpha - cZ \qquad \ldots B1$$

$$\frac{d \ln Z}{dt} = ecP - m_i - m_q Z \qquad \ldots B2$$

Now applying the Euler approximation we obtain the following quasi-linearised equations:

$$P(t+1) = P(t).\exp(\alpha - cZ) \qquad \ldots B3$$

$$Z(t+1) = Z(t).\exp(ecP - m_i - m_q Z) \qquad \ldots B4$$

where $\alpha \sim N(\mu_\alpha, \sigma_\alpha)$, $\alpha = \mu_\alpha + \varepsilon$ where $\varepsilon \sim N(0, \sigma_\alpha)$. It is then possible to compute the value of ε given μ_α associated with the transition from (P(t),Z(t)) to (P(t+1),Z(t+1)):

$$\varepsilon = \ln\left(\frac{P(t+1)}{P(t)}\right) - \mu_\alpha + cZ(t) \qquad \ldots B5$$

References

Arhonditsis, G. B., Papantou, D., Zhang, W. T., Perhar, G., Massos, E. and Shi, M. L. 2008. Bayesian calibration of mechanistic aquatic biogeochemical models and benefits for environmental management. *J. Marine Sys., 73(1-2)*, 8-30.

Bertino, L., Evensen, G. and Wackernagel, H. 2003. Sequential data assimilation techniques in oceanography. *Int. Stat. Rev., 71(2)*, 223-41.

Campbell, J.W. 1995. The lognormal-distribution as a model for biooptical variability in the sea. *J. Geophys. Res.-Oceans 100(C7)*, 13237-54.

Doucet, A., De Freitas, N. and Gordon, N. 2001. *Sequential Monte Carlo methods in practice*. New York, London, Springer.

Dowd, M. 2006. A sequential Monte Carlo approach for marine ecological prediction. *Environmetrics, 17(5)*, 435-55.

Dowd, M. 2007. Bayesian statistical data assimilation for ecosystem models using Markov Chain Monte Carlo. *J. Marine Sys., 68(3-4)*, 439-56.

Eknes, M. and Evensen, G. 2002. An Ensemble Kalman filter with a 1-D marine ecosystem model. *J. Marine Sys., 36(1-2)*, 75-100.

Evensen, G. 2007. *Data assimilation: the Ensemble Kalman filter*. Berlin, Springer.

Fasham, M.J.R., Ducklow, H.W. and McKelvie, S.M. 1990. A nitrogen-based model of plankton dynamics in the oceanic mixed layer. *J. Marine Res., 48(3)*, 591-639.

Gamerman, D. and Lopes, H.F. 2006. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. Boca Raton, Chapman and Hall/CRC Press.

Gelman, A. 2003. *Bayesian data analysis*. Boca Raton, Fla., Chapman and Hall/CRC Press.

Hairer, E., Norsett, S.P. and Wanner, G. 1987. *Solving ordinary differential equations*. Berlin, New York, Springer-Verlag.

Harmon, R. and Challenor, P. 1997. A Markov chain Monte Carlo method for estimation and assimilation into models. *Eco. Modelling, 101(1),* 41-59.

Li, Z.J., Chao, Y., McWilliams, J.C. and Ide, K. 2008. A Three-Dimensional Variational Data Assimilation Scheme for the Regional Ocean Modeling System. *J. Atmos. Ocean. Tech., 25(11*), 2074-90.

Lorenc, A.C. 2003. Modelling of error covariances by 4D-Var data assimilation. *Q. Jl R. Met. Soc., 129(595)*, 3167-82.

Lotka, A.J. 1920. Undamped oscillations derived from the law of mass action. *J. Amer. Chem. Soc., 42*, 1595-9.

Miller, C.B. 2004. *Biological oceanography*. Malden, MA, Blackwell Pub.

Natvik, L.J. and Evensen, G. 2001. Assimilation of ocean colour data into a biochemical model of the North Atlantic - Part 1. Data assimilation experiments. *33rd International Liege Colloquium on Ocean Dynamics,* Liege, Belgium, Elsevier Science Bv.

Volterra, V. 1926. Fluctuations in the abundance of a species considered mathematically. *Nature, 118*, 558-60.