ENVIRONMETRICS
Environmetrics 2006; **17**: 435–455
Published online 19 December 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/env.780

A sequential Monte Carlo approach for marine ecological prediction

Michael Dowd*,†

Department of Mathematics and Statistics, Dalhousie University, Halifax, Nova Scotia, Canada B3H 3J5

SUMMARY

This study considers the problem of marine ecological prediction in the context of online estimation and forecasting. Process oriented dynamic ecosystem models are combined with marine observations. The nonlinear, nonGaussian state space model provides the statistical framework. The associated filtering (nowcasting) and prediction (forecasting) problems are addressed via sequential Monte Carlo methods, in this instance a sequential importance resampler combined with Metropolis-Hastings MCMC. The specific focus is on a prototypical marine ecosystem model comprised of four interacting populations (phytoplankton, zooplankton, nutrients and detritus; PZND) whose co-evolution is described by system of coupled nonlinear differential equations. Stochastic environmental variation is introduced through a stochastic growth parameter, as well as through dynamical noise in the state evolution equations. The dynamic behaviour of this stochastic ecosystem model is complex: it regularly transitions through a Hopf bifurcation and exhibits episodic blooms of variable magnitude and duration. The model is applied to a case with weak seasonality, that is the oceanic mixed layer in the eastern equatorial Pacific. A partially observed state is considered comprised of a five year satellite (SeaWiFS) derived time series of ocean phytoplankton concentration at 12°N 95°W. Filtering estimates for the ecosystem state and a dynamic parameter were obtained using the sequential Monte Carlo approach. These showed predictor-corrector behaviour at observation times, including abrupt shifts in the median level after forecasts over measurement void. A corresponding variance (also skewness and kurtosis) growth and subsequent collapse was also seen. Forecasting experiments indicate some negative bias, and suggest there is predictive skill for forecasts out to 10–15 days.

KEY WORDS: particle filters; complex dynamics; state space models; marine ecosystem model; nonlinear time series; data assimilation

1. INTRODUCTION

Prediction for the lower trophic levels of the marine ecosystem has received increased attention in recent years. This is due to improvements in mathematical models describing the underlying biological processes, as well as their linkage with ocean circulation models (Fennell and Neumann, 2004). Perhaps, most importantly, during the last decade there has been a revolution in automated

*Correspondence to: Michael Dowd, Department of Mathematics and Statistics, Dalhousie University, Halifax, Nova Scotia, Canada B3H 3J5.
†E-mail: mdowd@mathstat.dal.ca

Contract/grant sponsor: NSERC Discovery Grant.

Received 28 February 2005
Accepted 21 October 2005

observation of the microscopic plants (phytoplankton) that are the foundational elements of the marine ecosystem using ocean optical properties (Cullen *et al.*, 1997). Satellite remote sensing in the visible and near-infrared bands from ocean colour sensors, such as SeaWiFS and MODIS, now routinely provide daily global ocean estimates for the plant pigments (chlorophyll) contained in the phytoplankton. In addition, *in situ* measurements from ocean observing systems, comprised of extensive mooring arrays measuring oceanographic and meteorological variables, are both underway (Gould, 2003) and planned (Tunnicliffe *et al.*, 2003). The development of analysis methods which effectively synthesize the information in mechanistic models and observations is a key element for the advancement of marine ecology. In this study, a statistical-dynamical approach to the prediction and forecasting of the lower trophic levels of the marine ecosystem is presented.

The major elements to be considered in ecological prediction are dynamic process models and oceanographic measurements. Marine ecosystem, or biogeochemical, models are posed as nonlinear systems of coupled ordinary differential equations, but solved numerically as time dependent difference equations. They typically support complex dynamical behaviour including multiple equilibrium states, bifurcation points and chaotic dynamics (Edwards, 2001). Further complexity arises when stochastic environmental forcing is introduced (Bailey *et al.*, 2004). These models are currently being coupled to fluid dynamic models of ocean circulation, yielding very high-dimensional problems (Hofmann and Lascara, 1998). Available *in situ* observations of marine ecological state variables, such as plankton, are generally sparse, noisy and have multiplicative nonGaussian errors (Campbell, 1995; Dowd *et al.*, 2003). For emerging ocean prediction systems, observations are routinely available for only a subset of the prognostic ecosystem variables which lend themselves to automated analysis, most notably phytoplankton chlorophyll through satellite remote sensing of ocean colour. Observations of other ecosystem components must rely on dedicated sampling or measurement programs. However, it is encouraging that some longer and more comprehensive time series are becoming available at selected global study sites (Siegel *et al.*, 2001).

The problem of combining time-dependent marine ecosystem models with available measurements (referred to as data assimilation in oceanography) has been approached from a variety of perspectives. For retrospective analyses or model calibration, a common approach is to estimate unknown or uncertain parameters of a model using observations of the state (Lawson *et al.*, 1995; Vallino, 2000). These may be viewed as problems in nonlinear regression (Thompson *et al.*, 2000), but are more typically treated as inverse problems and solved using optimisation techniques which take advantage of the time-dependent structure. Evans and Stark (2002) point out, however, that such inverse approaches have a strong parallel in statistical estimation. To further generalise these hindcasting approaches to include model errors, stochastic forcing and nonGaussian observations, Markov chain Monte Carlo (MCMC) methods have also been used, both for ecological parameter estimation (Harmon and Challenor, 1997) as well as for joint state and parameter estimation (Dowd and Meyer, 2003). Unfortunately, while these MCMC methods are general and flexible (Carlin *et al.*, 1992; Meyer and Christensen, 2000), they are also computationally costly and not ideally suited for time-dependent state estimation and prediction.

For real-time nowcasts (filtering) and short-term forecasts (prediction) using dynamic models, the state space model along with modern sequential Monte Carlo procedures appear promising (Lui and Chen, 1998). These treat nonlinear time-dependent dynamics and nonGaussian observations by providing for discrete representations of the time-evolving probability density function (pdf) of the state, which may be augmented with dynamic parameters (Kitagawa, 1998). Overviews of these sampling-based solutions can be found in Doucet *et al.* (2001) and Arulampalam *et al.* (2002). Recent studies have argued the potential of these methods for data assimilation in oceanography

(Bertino *et al.*, 2003). In marine ecology, for example, experiments have been carried out with the ensemble Kalman filter (Allen *et al.*, 2003; Natvik and Evensen, 2003) and sequential importance resampling (Losa *et al.*, 2003). In this study, we investigate a relatively sophisticated sequential Monte Carlo method (a sequential importance resampler feeding a Metropolis-Hastings MCMC) for use in marine ecological forecasting. The major emphasis is on time-dependent online state and dynamic parameter estimation for complex nonlinear dynamics for the case of a partially observed state (note that estimating static parameters is not considered here). A stochastic ecosystem model is used that supports episodic plankton blooms, which are driven by high frequency transitions of a stochastic dynamic parameter across a bifurcation point. Satellite time series observations for phytoplankton in the eastern equatorial Pacific from SeaWiFS are to estimate the ecosystem state, and to assess short-term forecast skill.

This paper is structured as follows. Section 2 introduces the ecological dynamics, the satellite time series observations, and the state space framework. It also outlines the sequential Monte Carlo method used to address the associated filtering and prediction problems. In Section 3, the application of the stochastic dynamic ecological model and results from the filtering and prediction experiments are given. A discussion and conclusion follows in Section 4.

2. METHODS

2.1. Ecological dynamics

The prototypical marine ecosytem model describes the co-evolution of the following state variables: phytoplankton P (microscopic plants), zooplankton Z (microscopic animals), nutrients N (dissolved inorganics), and detritus D (particulate organic non-living matter). These ecological state variables are generally considered as the foundational elements of the marine ecosystem (Fasham, 1993).

The mathematical form of the PZND model is given by the following system of coupled nonlinear ordinary differential equations,

$$\frac{\mathrm{d}P}{\mathrm{d}t} = f\{N; k_N\}\gamma P - \lambda_p P - f\{P; k_P\}IZ \tag{1}$$

$$\frac{\mathrm{d}Z}{\mathrm{d}t} = \epsilon f\{P; k_P\}IZ - \lambda_z Z \tag{2}$$

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \phi D + \beta f\{P; k_P\}IZ - f\{N; k_N\}\gamma P + \nu\lambda_z Z \tag{3}$$

$$\frac{\mathrm{d}D}{\mathrm{d}t} = -\phi D + \lambda_p P + (1 - \epsilon - \beta)f\{P; k_P\}IZ + (1 - \nu)\lambda_z Z \tag{4}$$

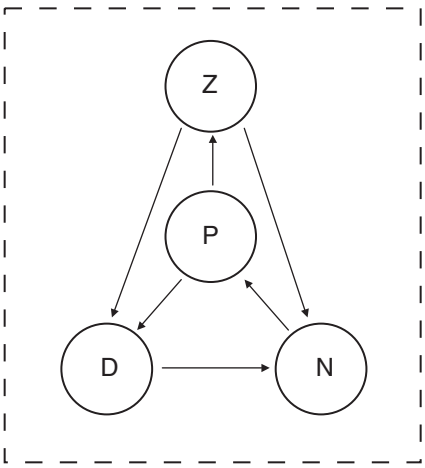

Figure 1. Conceptual diagram of the ecosystem box model. Prognostic ecosystem state variables are phytoplankton (P), zooplankton (Z), nutrients (N) and detritus (D). Arrows represent the direction of mass, or nutrient, fluxes between these populations

A conceptual diagram of the model is given in Figure 1 and describes the linkages amongst these ecosystem state variables. The state variables and parameters used in equations (1)–(4) are summarised in Table 1. The nondimensional modulation functions f are of the form

$$f\{X; k_X\} = \frac{X}{k_X + X} \tag{5}$$

Table 1. Definition of terms in the ecosystem model

Quantity	Units	Value	Definition
(i) State variables			
P	$\mathrm{gC/m^3}$	Eqn. (1)	Phytoplankton concentration
Z	$\mathrm{gC/m^3}$	Eqn. (2)	Zooplankton concentration
N	$\mathrm{gC/m^3}$	Eqn. (3)	Nutrients concentration
D	$\mathrm{gC/m^3}$	Eqn. (4)	Detritus concentration
(ii) Parameters			
k_N	$\mathrm{gC/m^3}$	0.2	Rate constant for N uptake by P
γ	$\mathrm{d^{-1}}$	0.14	P growth rate (photosynthesis)
λ_p	$\mathrm{d^{-1}}$	0.1	Loss/mortality term for P
I	$\mathrm{d^{-1}}$	0.6	Ingestion rate of P by Z
k_P	$\mathrm{gC/m^3}$	0.1	Rate constant for Z ingestion of P
ϵ	—	0.3	Assim fraction for Z ingested ration
β	—	0.4	Excreted fraction for Z ingested ration
λ_z	$\mathrm{d^{-1}}$	0.1	Loss/mortality term for Z
ν	—	0.5	Fraction of Z loss to N
ϕ	$\mathrm{d^{-1}}$	0.1	Remineralisation rate of D to N

For each quantity the following information is given: units, its numerical value (or its source), and a brief definition. Here gC denotes grams carbon, m is metres, and d is days.

where X in the above is one of P or N, and k_X is the rate constant. State variables and model parameters are nonnegative. The right hand side of equations (1)–(4) describes the cycling of matter between the ecosystem components. Each term represent a mass flux associated with a particular ecological process, the sign corresponding to the direction of mass flow as indicated by arrows in Figure 1. State variables are measured in units of a most limiting nutrient (usually nitrogen), but are expressed here in units of carbon concentration equivalents (gC/m^3) (Steele and Henderson, 1981). Since the sum of the right hand side of equations (1)–(4) is zero, the system conserves its total mass, which is specified by initial conditions. In practice, the system is solved numerically as difference equations.

Three major ecological processes are described by this model:

1. *Photosynthetic production*: $f\{N; k_N\}\gamma P$ represents the conversion of inorganic N to organic P by photosynthesis.
2. *Predator-prey interactions*: $f\{P; k_P\}IZ$ describes the flux of P to herbivorous Z through grazing.
3. *Biogeochemistry*: ϕD is the flux of organic matter D to inorganic N by marine bacteria (remineralisation).

Mortality (loss) processes are described by λ_p and λ_z, for P and Z respectively. The ingested ration of P by Z is partitioned amongst assimilated fraction for tissue growth ϵ, metabolic excretion β, and fecal production $1 - \beta - \epsilon$. The parameter ν partitions Z losses between N and D. This ecological model is based on a simplified version of the one in Dowd (2005) and details can be found there. PZND models are the basis for most current descriptions of marine ecology and ocean biogeochemistry (Fennell and Neumann, 2004). The emphasis here is on variability in photosynthetic growth processes, due to its central role in the ecological dynamics.

The ecosystem model was applied to the eastern equatorial Pacific. This region is characterised by weak seasonality but strong synoptic (meteorological) variability, as well as interannual changes as a result of climatic processes such as El Niño (Talley *et al.*, 2005). The ocean biogeochemistry in this region, as defined by (1)–(4), takes place in the oceanic-mixed layer, that is the upper $\sim 50\,\mathrm{m}$ (Edwards *et al.*, 2004). Stochastic environmental and biological variability was included through the following processes: (i) photosynthetic production is described by the daily averaged growth rate parameter γ. This was considered here as a stochastic dynamic parameter driven by changes in surface and underwater light field (clouds, water properties), as well as the interaction of phytoplankton photosynthetic physiology with light and mixed layer physics (Denman and Gargett, 1983; Cullen and Lewis, 1988); (ii) dynamical noise occurs due from turbulent mixing and associated mass fluxes; these were included as an additive error term for each of (1)–(4) (Bailey *et al.*, 2004; Monahan and Denman, 2004). Note that the remaining static parameters are assumed fixed and known, with their values taken from field and lab studies (cf. Dowd, 2005).

2.2. *Observations*

Satellite ocean colour imagery were used to provided time series of P concentration. These data were obtained from NASA's SeaWiFS satellite which provides calibrated quantitative data on global ocean optical properties (Hooker and McClain, 2000). The concentration of plant pigments (chlorophyll) in the upper ocean can be derived from these ocean colour data (O'Reilly *et al.*, 1998). For this study, SeaWiFS imagettes were extracted for a 30 km box centred at 12°N 95°W in the eastern tropical Pacific for the period from late 1997 to mid 2002. Since a large portion of these SeaWiFS chlorophyll were obscured by clouds or otherwise contaminated, all valid pixels within the imagettes were

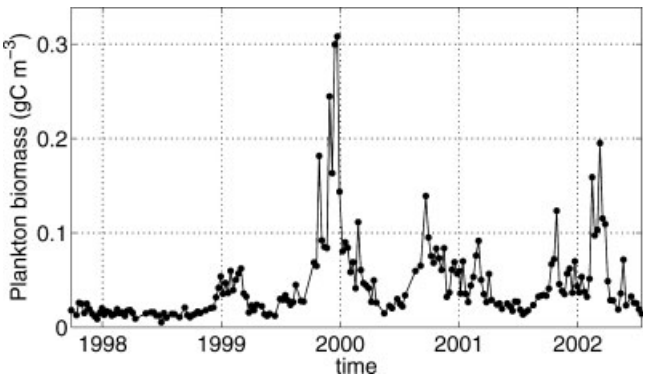

Figure 2. Satellite derived time series of phytoplankton biomass concentration from SeaWiFS in the eastern equatorial Pacific (12^{c}*irc*N 95°W). Dates here indicate the beginning of calendar year

spatially and temporally averaged to yield a representative point measures of chlorophyll on an approximately weekly basis. Chlorophyll was converted to carbon concentrations using a carbon to cholorphyll ratio of 100 (Cloern *et al.*, 1995). The terms 'phytoplankton' and 'chlorophyll' are used interchangeably hereafter.

The resultant phytoplankton time series is shown in Figure 2. There are 190 data points distributed over the approximately 5-year period. The shortest sampling interval is approximately weekly, but at times valid observations are absent for up to a month or more. The series exhibits plankton blooms which are quasi-periodic and irregular in their magnitude, shape and duration. These may to be linked to climatic processes.[1] It is well established that bio-optical cholophyll measurements follow a lognormal distribution (Campbell, 1995). Examination of the extensive comparisons of *in situ* and SeaWiFS derived chlorophyll estimates compiled by O'Reilly *et al.* (1998) and McClain *et al.* (2000) lead us to choose a standard deviation $\sigma = 0.3$ for the lognormal distribution which characterises observation uncertainty.

2.3. State space framework

The nonlinear and nonGaussian state space model provides the basis for combining ecological dynamics with measurements. The vector $\boldsymbol{x}_t$ represents the state of the ecological system at time t. It includes the abundances of P, Z, N and D, as well as the current value of the time-varying stochastic parameter γ (Kitagawa, 1998). The model and measurement equations are:

$$\boldsymbol{x}_t = \Phi(\boldsymbol{x}_{t-1}, v_t) \tag{6}$$

$$\boldsymbol{y}_t = h_t(\boldsymbol{x}_t, n_t) \tag{7}$$

The state evolution equation (6) is comprised of a Markovian transition linking $\boldsymbol{x}_t$ with $\boldsymbol{x}_{t-1}$. The discrete version of the nonlinear dynamics of equations (1)–(4) are embodied in the time-invariant

[1]The beginning of the record in late 1997 to mid 1998 was a strong El Niño period. This rapidly transitioned to a La Niña in mid-1998 to early 2000 after which the equatorial Pacific remained in a neutral phase until the end of the analysis period in mid-2002.

operator Φ. System noise v_t includes the stochastic forcing for γ and the dynamical noise processes. The measurement equation (7) links the observations $\boldsymbol{y}_t$ with the state $\boldsymbol{x}_t$ incorporating the measurement error n_t. As defined for this study the operator h_t is straightforward: only P is observed directly, but not at all model time steps. Both v_t and n_t and assumed uncorrelated in time, and with one another.

Define the cumulative observation set as $Y_\tau = (y_1', \ldots, y_\tau')'$ and consider a particular time t. Determination of the pdf $p(\boldsymbol{x}_t \mid Y_\tau)$ provides a solution for (6)–(7). On this basis, three categories of problem are defined: (i) $t < \tau$ or smoothing; (ii) $t = \tau$ or filtering; and (iii) $t > \tau$ or prediction. In this study, the focus is on filtering and prediction.

Suppose we are given $p(\boldsymbol{x}_{t-1} \mid Y_{t-1})$, or the joint pdf of $\boldsymbol{x}_{t-1}$ given all information (observations) up to and including time $t-1$. For online estimation, the target distribution of interest is $p(\boldsymbol{x}_t \mid Y_t)$. The single stage transition involves two steps:

1. *Prediction* (or forecast):

$$p(\boldsymbol{x}_t \mid Y_{t-1}) = \int p(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})\, p(\boldsymbol{x}_{t-1} \mid Y_{t-1}) \mathrm{d}\boldsymbol{x}_{t-1} \tag{8}$$

2. *Measurement* (or update):

$$p(\boldsymbol{x}_t \mid Y_t) \propto p(\boldsymbol{y}_t \mid \boldsymbol{x}_t)\, p(\boldsymbol{x}_t \mid Y_{t-1}) \tag{9}$$

Prediction relies on knowledge of the state at the previous time $p(\boldsymbol{x}_{t-1} \mid Y_{t-1})$ which is multipled by a transition density $p(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})$ identified with the dynamical prediction (6). Marginalisation over $\boldsymbol{x}_{t-1}$ yields the predictive density $p(\boldsymbol{x}_t \mid Y_{t-1})$. At the measurement step, the newly available observation, $\boldsymbol{y}_t$, is used to update the predictive density according to Bayes formula to yield $p(\boldsymbol{x}_t \mid Y_t)$. This uses the likelihood $p(\boldsymbol{y}_t \mid \boldsymbol{x}_t)$, and treats the predictive density $p(\boldsymbol{x}_t \mid Y_{t-1})$ as the prior.

Only for a very restricted class of problems can these quantities be computed analytically, most notably the Kalman filter for the linear, Gaussian version of Equations (6) and (7). There are many approximate solutions for the filtering problem for nonlinear and/or nonGaussian systems. Examples include the extended Kalman filter (Jazwinski, 1970), the ensemble Kalman filter (Evensen, 2003), the Gaussian sum filter (Alspach and Sorenson, 1972), the dynamic generalised linear model (West *et al.*, 1985) and the nonGaussian filter (Kitagawa, 1987). MCMC approaches can also be used to treat the general smoothing problem (Carlin *et al.*, 1992). Sequential Monte Carlo methods have, however, become established as efficient and effective sampling-based solutions to the general filtering problem (e.g. Gordon *et al.*, 1993; Kitagawa, 1996; Lui and Chen, 1998; Doucet, 2001; Arulampalam *et al.*, 2002).

In this study, sequential Monte Carlo methods were used to solve the filtering problem, or to carry out the single stage transition from $p(\boldsymbol{x}_{t-1} \mid Y_{t-1})$ to $p(\boldsymbol{x}_t \mid Y_t)$ involving prediction (8) and measurement (9). They rely on discrete representations of the relevant pdfs. At time $t-1$ we are given

$$\left\{ \boldsymbol{x}_{t-1\,|\,t-1}^{(i)}, w_{t-1}^{(i)} \right\} \sim p(\boldsymbol{x}_{t-1} \mid Y_{t-1})$$

where the above notation says that $\boldsymbol{x}_{t-1\,|\,t-1}^{(i)}$, $i = 1, \ldots, m$, are m weighted samples from $p(\boldsymbol{x}_{t-1} \mid Y_{t-1})$, where the associated weights are $w_{t-1}^{(i)}$. Prediction (8) and measurement (9) then occur as

1. *Prediction*: The model (6) updates particle positions: $\boldsymbol{x}_{t\,|\,t-1}^{(i)} = \Phi(\boldsymbol{x}_{t-1\,|\,t-1}^{(i)}, v_t^{(i)})$ where $v_t^{(i)}$ is an independent realisation of the system noise. This yields

$$\left\{\boldsymbol{x}_{t\,|\,t-1}^{(i)}, w_{t-1}^{(i)}\right\} \sim p(\boldsymbol{x}_t \,|\, Y_{t-1})$$

2. *Measurement*: Weights are then updated as $w_t^{(i)} = w_{t-1}^{(i)}\, p(\boldsymbol{y}_t \,|\, \boldsymbol{x}_{t\,|\,t-1}^{(i)})$, and re-normalised so that

$$\left\{\boldsymbol{x}_{t\,|\,t-1}^{(i)}, w_t^{(i)}\right\} = \left\{\boldsymbol{x}_{t\,|\,t}^{(i)}, w_t^{(i)}\right\} \sim p(\boldsymbol{x}_t \,|\, Y_t)$$

Note that this particular weight update defines the importance density as the prediction density $p(\boldsymbol{x}_t \,|\, Y_{t-1})$, but other choices are possible (Liu and Chen, 1998).

The above procedure represents sequential importance sampling, and has the well known problem of sample degeneracy wherein after a few time steps only a few particles have nonnegligible weights (Kong *et al.*, 1994). To alleviate this sample degeneracy, a weighted resampling (with replacement) of $\{\boldsymbol{x}_{t\,|\,t}^{(i)}, w_t^{(i)}\}$ is generally carried out (Gordon *et al.*, 1993). This provides for an (unweighted) sample

$$\left\{\boldsymbol{x}_{t\,|\,t}^{(i)}\right\} \sim p(\boldsymbol{x}_t \,|\, Y_t)$$

where all the $w_t^{(i)} = 1/m$ after resampling. The issue with this sequential importance resampling (SIR) procedure is sample impoverishment, wherein the particle ensemble may have a large number of repeated copies due to resampling.

In this study, it was found that when SIR as carried out according to the above procedure, the particle ensemble became impoverised due to the abrupt corrections and variance collapse at measurement times (see Subsection 3.2). Motivated by Gilks and Berzuini (2001), the particle ensemble was transformed by feeding the results from the SIR procedure into a Metropolis-Hastings MCMC algorithm. The end result was that an improved sample $\{\boldsymbol{x}_{t\,|\,t}^{(i)}\}$ of the posterior $p(\boldsymbol{x}_t \,|\, Y_t)$ was obtained. The details and rationale for this procedure for the state space model are given in Appendix A.

3. RESULTS

3.1. Stochastic dynamic model

The ecosystem model (1)–(4) was discretised using a Euler method. A unit time increment in what follows corresponds to one day. An ecological refuge was included wherein each of the state variables are restricted to not drop below 1 percent of their equilibrium values (defined below). This both enforces nonnegativity and provides for a seed populations, which in practice results from spatial patchiness.

Results from a deterministic simulation (i.e. time integration of (1)–(4), or equivalently (6) with $v_t = 0$) are shown in Figure 3. Initial conditions are taken as the equilibrium points, defined as the values of the state variables where the time derivatives of (1)–(4) vanish ($P^* = 0.125, Z^* = 7.08 \times 10^{-3}, N^* = 0.764, D^* = 0.136$). Results are reported as phase plane plots wherein pairwise combinations of the ecosystem state variables are plotted against one another. The trajectories each

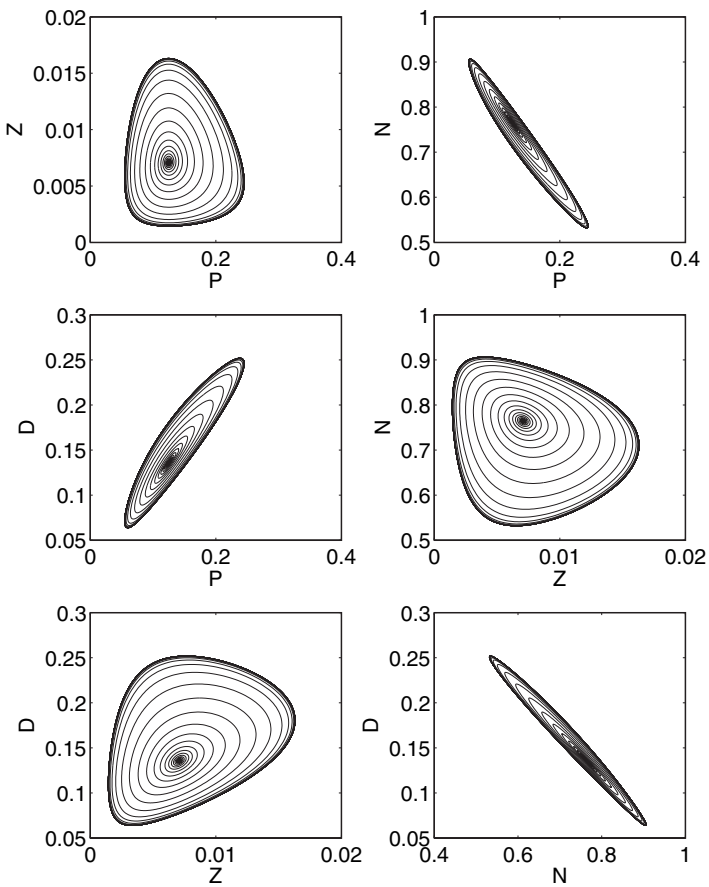

Figure 3. Phase plots for pairwise combinations of the ecosystem state variables. In all cases, the system is started at the equilibrium point (dot) and spirals outwards as time progresses to achieve a periodic steady state or limit cycle

start in the vicinity of the equilibrium point and spiral outwards until a limit cycle is achieved. The variable pairs (P, N), (D, P) and (D, N) are tightly coupled and nearly in phase; variable pairs involving Z are in near quadrature. The nonelliptical shapes results from model nonlinearities. Once the periodic steady state is reached, the deterministic simulation has regular and repeating cycles with a period slightly less than 1 year.

The growth rate parameter γ varies due to changes in environmental conditions and physiological adapatation. Figure 4 shows how varying γ in the deterministic case produces a Hopf bifurcation near $\gamma = 0.13$. As γ increases the position of equilibrium points in state space change, and the character of these point alters from a stable attractor at low γ, to a periodic point as γ increases. The case where $\gamma < 0.12$ may be viewed as one where the biology is turned off, that is almost all the mass of the system resides in inorganic N. Increasing γ beyond this activates the biology and initiates the cycling of nutrients amongst the ecosystem components.

Stochastic environmental variations were specified through both a dynamic parameter and through the state evolution equations according to the following:

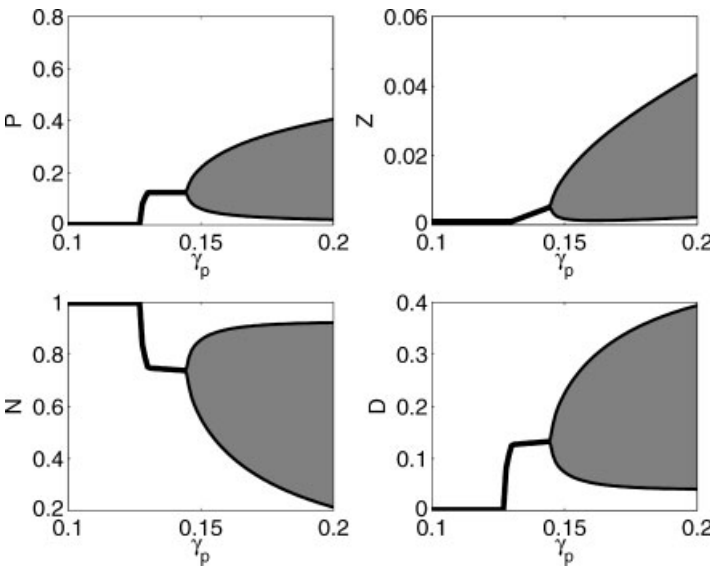

Figure 4. Hopf bifurcation for ecosystem state variables resulting from varying the parameter γ. For γ below bifurcation point, the solid line represents the equilibrium value for the ecosystem state variables (stable attractor). For γ above this value, it represents upper and lower limits of a periodic orbit, with the gray shaded area indicating the range

1. *Growth rate* γ: The mean value, $\bar{\gamma}$, is 0.14 and its daily variability, $\Delta\gamma$, follows an AR(1) process, that is

$$\Delta\gamma_t = \alpha\Delta\gamma_{t-1} + \varepsilon_t$$

where $\alpha = 0.9$ and $\varepsilon_t \sim \text{NID}(0, 0.01)$. The daily varying growth rate is reconstructed as $\gamma_t = \bar{\gamma} + 3\Delta\gamma_t$. These values are reasonable for a depth averaged mixed layer γ computed with representative light levels. The decorrelation time matches that of the meteorological band. With γ_t as defined above, the ecosystem will regularly transition across the bifurcation point.

2. *Dynamical noise*: Additive zero-mean normally distributed i.i.d. noise was added to each of the discretised (1)–(4). Its standard deviation is $0.01 \times X^*$, where X^* is one of P^*, Z^*, N^* and D^*. These values are the same order of magnitude as those of Bailey *et al.* (2004).

These complete the specification of the state Equation (6). The remaining parameters are assumed fixed and known. Note that the stochastic system is constructed to conserve total mass in an ensemble sense such that its expected value remains constant.[2]

Figure 5 shows a realisation of the stochastic simulation. The effect of adding stochastic variability is dramatic when compared to the regular cycles of the deterministic system. The realisation exhibits irregular and aperiodic oscillations punctuated by periods of little biological activity. These are due

[2]During simulation, the dynamical noise acts as source and sink terms which changes the total mass. Careful treatment of the ecological refuge is required to prevent model drift.

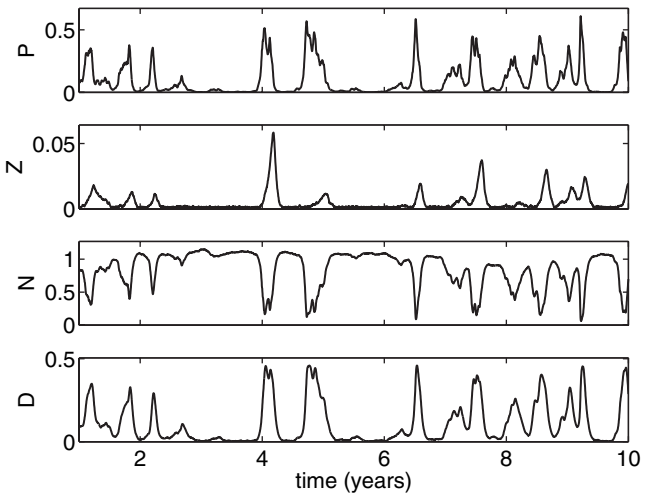

Figure 5. Sample realisation generated from the stochastic ecosystem model. Shown is time evolution for all ecosystem state variables based on a 9-year model integration

largely to γ_t which acts on the system at relatively short periods (weeks). However, the emergent response is that of episodic plankton blooms with a much longer period.

Figure 6 shows the kernel smoothed long-run equilibrium pdfs associated with each of the ecosystem state variables. The biotic components P, Z and D are strongly right skewed, reflecting the fact that the system is most often in a state of relatively inactive biology with episodic blooms

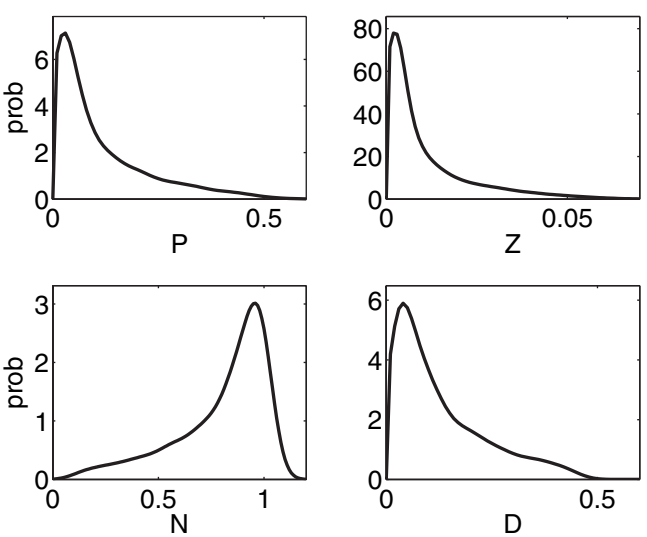

Figure 6. Estimated long run, or equilibrium, kernel smoothed equilibrium probability density functions for the ecosystem state variables

occuring occasionally. Similar reasoning explains why the N pdf is left skewed. The timescale for the system to reach a statistical steady state was determined using a scaled Kullback-Liebler divergence metric. This quantified the approach of the state to its equilibrium pdf, starting from a known initial state, which was taken here to be the equilibrium values (P^*, Z^*, N^*, D^*). The Kullback-Leibler divergence falls off exponentially with a timescale for system adjustment of around 30 days for all ecosystem state variables.

3.2. Filtering and prediction

In order to combine the satellite phytoplankton measurements with the ecological dynamics, the filtering problem was treated with the sequential Monte Carlo method of section 2.3. A particle ensemble with 10^4 members was used as the basis for representing the ecosystem state. The state evolution equations (6) were implemented as in section 3.1. The measurement Equation (7) considers only phytoplankton observations and takes the form $P_t^{\text{obs}} = P_t \exp(n_t)$ where $n_t \sim \text{NID}(0, 0.3)$ (see section 2.2). The analysis period coincides with the observation period. Initial conditions for the ecosystem state took the form of a diffuse prior, with the result that the pdfs were effectively reset upon encountering the first observation (the initial conditions have little influence on the remainder of the analysis). The SIR procedure used the resampler of Kitagawa (1996), and the MCMC step was based on the Metropolis-Hasting algorithm outlined in Appendix 1.

Filtering results for the observable state variable P are given in Figure 7. These show the time-varying estimates for P reported in terms of the median and approximate 95 percent confidence intervals. Figure 7a shows the full analysis period, while Figure 7b zooms in on the central time period (mid-1999 to mid-2000) corresponding to the largest plankton bloom. The median of P follows closely the observations. Abrupt jumps occur when a new observation is encountered after a smooth model forecast over an observation void. A related feature is also evident in the confidence intervals: the width of the confidence interval increases exponentially during the forecast phase, and then abruptly decreases as additional measurement information is incorporated into the filter estimate of the state.

The filter also reconstructs estimates for unobserved state variables (Z, N, D) and the dynamic parameter (γ) using the dynamical linkages. These are given in Figure 8. The estimated Z shows episodic blooms whose peaks lag slightly those in the corresponding P series, with magnitudes which depend on the recent time history of P. The width of the confidence interval scales with this magnitude. The N values generally decline over the analysis period, with a pronounced dip during the 2000 bloom (this reflects the general upward trend of P over the course of the analysis period and is not due to nonconservative mass). The D series is a near mirror image of the N series. All the unobserved state variables exhibit similar growth in the width of the 95 percent confidence interval over the observation voids between P measurements. Abrupt changes are less evident in the unobservables since the dynamical linkages lead to low pass filtering properties. The recovered time series for the parameter γ shows how the observations and dynamics together estimate the high-frequency changes in this dynamic parameter. The autocorrelation of the estimated γ series is consistent with the postulated AR(1) process, but there is some evidence of nonnormality (heavy tails).

Figure 9 shows the time evolution of the kernel smoothed pdf for P during a 100-day time period centered around the year 2000 plankton bloom. Its most obvious feature is the abrupt shifting of the mean level of the pdf, as well as the corresponding decrease in variance, which occurs as measurements are encountered after forecasting over an observation void. Asymmetry in the pdf is evident at low P. Variance increase are seen with greater P abundance, and are consistent with the lognormal observation errors.

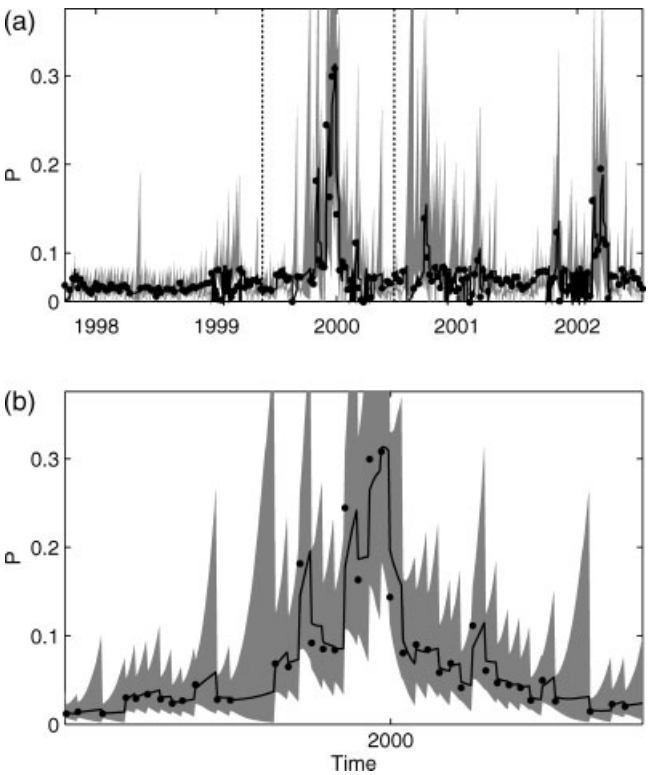

Figure 7. Filtering results for the observed state variable P. Shown are the filter estimate of the median (black line), the observations (dots), the approximate 95 per cent confidence intervals (gray shaded area). a: Results for the full analysis period from mid-1997 to mid-2002. b: Results for the central time period as designated in panel a by the vertical-dashed lines

The moments for the filter estimated P from mid-1999 to mid-2000 are given in Figure 10. The mean level shows similar patterns to the median. The variance scales with the mean level and shows exponential growth between observation times due to prediction uncertainty. For low P abundance, skewness and kurtosis increase between observation times with maximal values corresponding to the longest forecasts. An interesting feature is that skewness, and to a lesser extent kurtosis, shows a pronounced dip at the bloom peak implying that far from edge effects (nonnegativity constraints) the P distribution is closer to normal.

Forecast errors are defined as the difference between a P observation and the median value of the associated forecast ensemble at that time. The forecast period is the time difference between the last observation used to create the forecast ensemble, and the (future) observation used to compute the forecast error. (Note then that, a forecast period of zero yields the analysis errors at the measurement time.) Forecast errors were compiled for all possible forecasts for periods up to 50 days. These were binned on a roughly weekly basis to reflect the sampling interval (and allow a sufficient number of realisations for summary statistics to be compiled). Figure 11 reports some properties of these forecast errors. Figure 11a indicates that the root mean square (RMS) forecast error increases from the analysis RMS error of 0.01, until it asymptotes after 30 days to around 0.05. This overall error includes

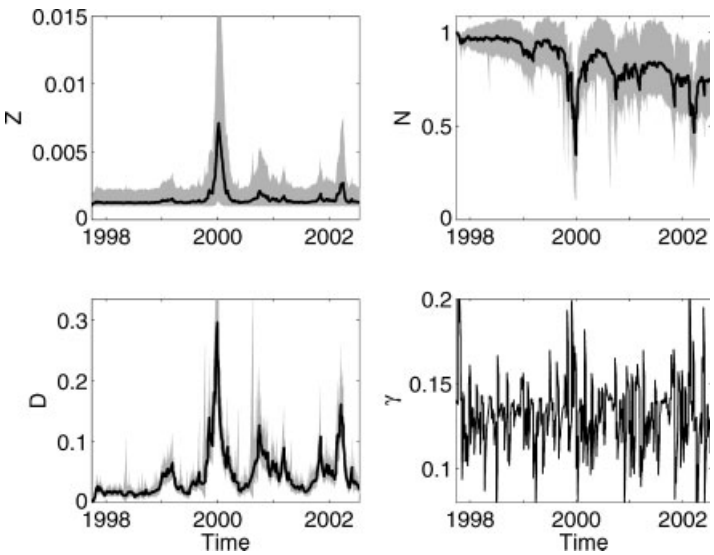

Figure 8. Filtering results for the unobserved state variables Z, N and D, as well as the dynamic parameter γ. The median (black line) is given for all state variables. For Z, N and D the approximate 95 per cent confidence intervals are also shown (gray-shaded area)

components due to both bias and variance. Figure 11b shows that the forecast errors have a negative bias which increases linearly with the forecast period. This under-forecasting is pronounced for low values of P since, relative to the SeaWiFS observations, the model system under-predicts P when the biology is turned off (see Figure 5). The forecast errors also exhibit significant variability. Figure 11c shows the forecast error standard deviation (the variability about their respective weekly means).

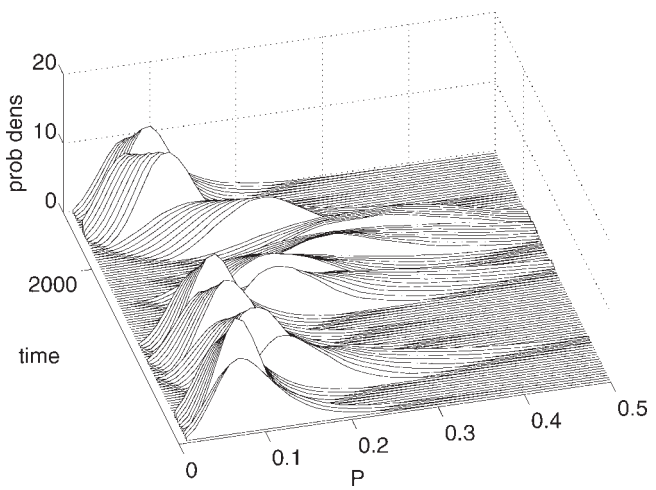

Figure 9. Time evolution of the pro pdf for P as estimated by the filter. The period covered is 100 days. The beginning of the 2000 calendar year is indicated

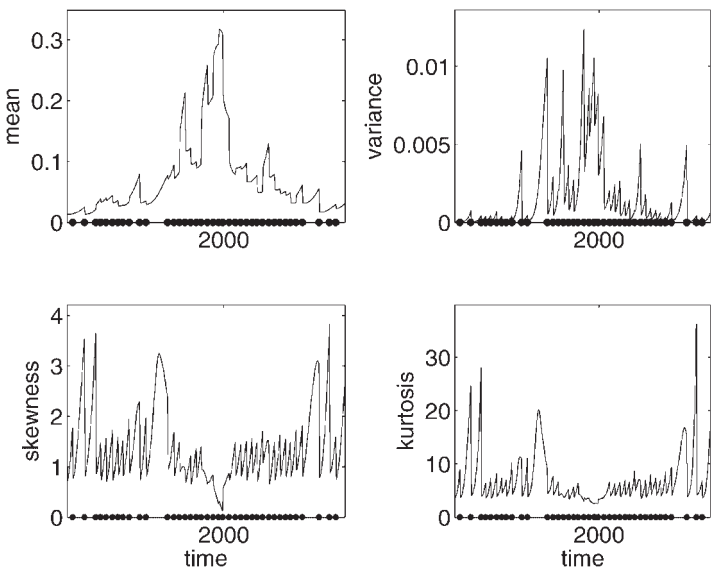

Figure 10. Time evolution of the moments for estimated phytoplankton *P*. The time period here is from mid-1999 to mid-2000 as indicated in Figure 7. Observation times are shown as dots along the time axis

These resemble the overall RMS error, but have slightly lower values. On the basis of these forecast diagnostics, it is concluded that the system may have useful predictive skill out to 10–15 days. However, the estimated uncertainty of model forecasts grows rapidly when unconstrained by observations.

4. DISCUSSION AND CONCLUSIONS

This study has considered the problem of marine ecological prediction using dynamic population models together with observations. The general nonlinear, nonGaussian state space model provided a unifying statistical framework for considering the joint estimation of the time varying ecological state and parameters. Like Meyer and Christensen (2000), the emphasis of this study is on the use of sampling-based Monte Carlo methods for estimation in a complex nonlinear system. Other studies in marine ecological data assimilation are concerned primarily with hindcasting the system state and parameter estimation (e.g. Lawson *et al.*, 1995; Harmon and Challenor, 1997; Dowd and Meyer, 2003; Allen *et al.*, 2003). The focus here, however, is on sequential methods for online state, and dynamic parameter, estimation (filtering and prediction) in the context of a dynamically complex stochastic marine ecosystem model.

The simplified PZND ecosystem model used in this study provides a mathematical description of the time-evolution of the lower trophic levels of the marine ecosystem (Fasham, 1993; Fennel and Neumann, 2004). Ecological dynamics are described as coupled nonlinear ODEs, which are discretised and numerically solved as difference equations. The dynamical behaviour supported by the PZND model is very complex (Edwards, 2001). Stochastic ecological dynamics were applied here for the oceanic mixed layer for the case of weak seasonality, such as found in equatorial regions. Quasi-regular time series measurements of one component of the ecosystem state, *P*, were obtained

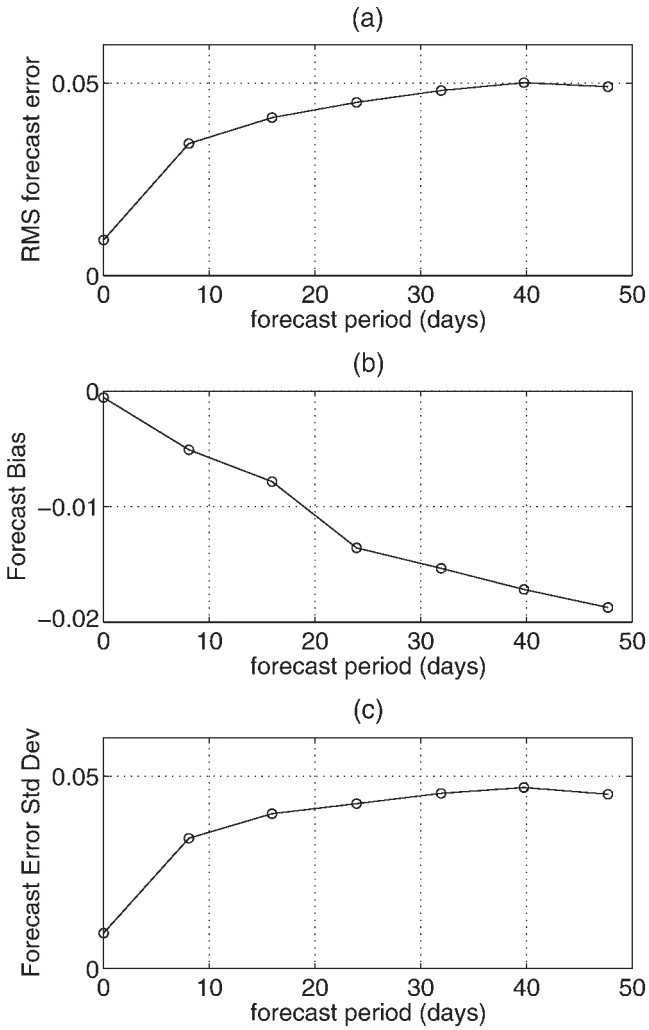

Figure 11. Forecast errors statistics for phytoplankton, *P*, for forecast periods up to 50 days. a: Overall RMS forecast error. b: Bias of the forecast error. c: Standard deviation of the forecast error

from the SeaWiFS satellite from the eastern tropical Pacific. This configuration mimics a prototype ocean forecasting system wherein only one of the ecosystem components is routinely observable, and information on the other components must be diagnosed via dynamical relations.

Environmental and biological stochasticity are important drivers for population dynamics in general (Marion *et al.*, 2000; Bjørnstad and Grenfell, 2001), and the PZND ecosystem in particular (Bailey *et al.*, 2004). Dynamical noise was appended to the PZND state evolution equations to represent mixing or model error (cf. Monahan and Denman, 2003). The primary source of stochasticity was introduced through the parameter γ, the daily growth rate of phytoplankton, and the other parameters are considered fixed. This stochastic dynamic parameter varies due to changes in the underwater light field and physiological adaptations by phytoplankton (Denman and Gargett,

1983; Cullen and Lewis, 1988). Variations in γ followed an AR(1) model with a decorrelation time-scale of 7–10 days (the meteorological band). Analysis of the deterministic system indicated that a Hopf bifurcation occurred with respect to γ. The stochastic process for γ thus ensured that the system frequently transitioned across this bifurcation point. A rich set of dynamical behaviour emerged which supported episodic plankton blooms with variable timing, duration and magnitude, and consistent with plankton observations from the eastern tropical Pacific.

A major goal of this study was to investigate the use of modern sequential Monte Carlo methods as a basis for ocean ecosystem nowcasting and forecasting. These sampling-based solutions for nonlinear, nonGaussian systems allow determination of the time evolution of the pdf of the ecosystem state, or any desired summary quantity. Filtered state estimates showed the expected predictor/corrector behaviour, that is a smooth forecast with exponentially increasing variance, followed by an abrupt correction and variance collapse once an observation is encountered. At each measurement step, sequential importance resampling (Gordon *et al.*, 1993; Kitagawa, 1996) followed by Metropolis-Hastings MCMC step (Appendix A) provided samples (or particles) from the posterior distribution of the ecosystem state vector and for the dynamic growth rate parameter. Such a hybrid procedure avoids sample impoverishment (Gilks and Berzuini, 2001). The estimation of static parameters has not been considered in this study; these quantities were assumed known based on literature values (cf. Dowd, 2005). Inclusion of MCMC procedures into the particle filter may provide a means for their estimation (Lee and Chia, 2002).

From an implementation perspective, sequential Monte Carlo algorithms are straightforward to apply, but computationally intensive. This has motivated simplified algorithms such as the ensemble Kalman filter (Evensen, 2003) which has been used in marine ecology (Allen *et al.*, 2003; Natvik and Evensen, 2003). This method uses the prediction step of (8), but replaces the measurement step with the Kalman filter updating equations. In contrast, this study gave particular attention to the measurement step through use of the hybrid SIR/MCMC procedure. (Little emphasis was given to computational considerations since they were quite reasonable for the problem under consideration). However, it is also clear that for high-dimensional problems, such as when ecosystem models coupled to ocean circulation models, computational considerations and effective approximations are important (Bertino *et al.*, 2003).

To assess predictive skill, forecasts for the observable state variable P were examined. Prediction errors are central diagnostics for the linear, Gaussian state space model (e.g. Harvey, 1989), and are also readily computed for the nonlinear and nonGaussian case. As an outer bound on forecast skill, note that the adjustment time-scale for the system to approach the equilibrium pdf, as measured here via the Kullback-Liebler divergence, was of order 30 days. (Interestingly, this matches the results from stability analysis of the deterministic system). Forecast errors showed a negative bias, and an increase in variance until an asymptote is reached at about 30 days. The rate of change in the RMS forecast error suggest there may be useful forecast skill out to 10–15 days.

In summary, sequential Monte Carlo methods are becoming established as a general approach for the analysis of dynamic systems (e.g. Doucet, 2001). Realistic applications requires one to give careful consideration to the interaction of the dynamics and the estimation problem. Results here suggest that Monte Carlo filtering appears promising for addressing the prediction problem for nonlinear, partially observable ecosystem models. They may also facilitate Bayesian model identification and selection (Carlin and Chib, 1995), which is an important for determining ecological model structure and complexity. It is hoped that this study will motivate further application of these general and flexible state space approaches to realistic problems in marine ecology.

ACKNOWLEDGEMENTS

The author thank Dr. Bruce Smith, Dr. Keith Thompson and Dr. Andy Edwards of Dalhousie University, as well Dr. Alain Vezina and Dr. Svetlana Losa of the Bedford Institute of Oceanography, for useful discussion. Chi Ding and Moritz Lehmann of Dalhousie Oceanography assisted with the SeaWiFS data extraction. The SeaWiFS Project and SeaDAS Development Group at NASA GSFC are gratefully acknowledged. This work was supported by an NSERC Discovery Grant to Michael Dowd.

APPENDIX

A. Metropolis-Hastings algorithm for state space filtering

In Section 2.3, it was indicated that Metropolis Hastings (M-H) algorithm is appended to the sequential importance resampling (SIR) procedure at the measurement (update) step. This improves the properties of the particle ensemble as a discrete representation of $p(\boldsymbol{x}_t \mid Y_t)$ by eliminating particle repeats due to resampling (sample impoverishment). M-H algorithms proceed by drawing a candidate $\boldsymbol{x}_t^*$ from a trial distribution, followed by computation of an acceptance probability (e.g. Chib and Greenberg, 1995). Below, this procedure is outlined for the filtering problem of this study.

For state space models it is difficult to evaluate the posterior directly. Consider the following

$$\begin{aligned} p(\boldsymbol{x}_t \mid Y_t) &\propto p(\boldsymbol{y}_t \mid \boldsymbol{x}_t) \cdot p(\boldsymbol{x}_t \mid Y_{t-1}) \\ &\propto p(\boldsymbol{y}_t \mid \boldsymbol{x}_t) \cdot \int p(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})\, p(\boldsymbol{x}_{t-1} \mid Y_{t-1}) \mathrm{d}x_{t-1} \end{aligned}$$

If $\boldsymbol{x}_t^*$ can be drawn from $p(\boldsymbol{x}_t \mid Y_{t-1}) = \int p(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}) p(\boldsymbol{x}_{t-1} \mid Y_{t-1}) \mathrm{d}\boldsymbol{x}_{t-1}$ then the M-H acceptance probability is

$$\alpha = \min\left(1, \frac{p(\boldsymbol{y}_t \mid \boldsymbol{x}_t^*)}{p(\boldsymbol{y}_t \mid \boldsymbol{x}_{t\mid t}^{r,(i)})}\right)$$

which requires only a straightforward evaluation of a likelihood ratio. The particle $\boldsymbol{x}_{t\mid t}^{r,(i)}$ in the denominator is the ith resampled particle after SIR (but before the M-H-based adjustment of its position).

The above procedure requires draws (candidate) samples from $p(\boldsymbol{x}_t \mid Y_{t-1})$. Note that the particle ensemble associated with the previous filter estimate is available as

$$\left\{\boldsymbol{x}_{t-1\mid t-1}^{m,(i)}\right\} \sim p(\boldsymbol{x}_{t-1} \mid Y_{t-1}), \quad i = 1, \ldots, M$$

where $\boldsymbol{x}_{t-1\,|\,t-1}^{m,(i)}$ is the ith particle after both the SIR and the M-H steps. Since each of these are equally likely, one is randomly chosen one as precursor candidate $\boldsymbol{x}_{t-1\,|\,t-1}^{*}$. To generate the required random sample from $p(\boldsymbol{x}_t \mid Y_{t-1})$, $\boldsymbol{x}_{t-1\,|\,t-1}^{*}$ is then passed through the state evolution Equation (6), that is

$$\boldsymbol{x}_{t\,|\,t-1}^{*} = \Phi\left(\boldsymbol{x}_{t-1\,|\,t-1}^{(i)}, v_t\right)$$

with v_t being an independent realisation of the system noise. This moves the particle and provides for the required candidate drawn from the predictive distribution $p(\boldsymbol{x}_t \mid Y_{t-1})$. It is designated $\boldsymbol{x}_t^{*} = \boldsymbol{x}_{t\,|\,t-1}^{*}$.

These $\boldsymbol{x}_t^{*}$ from the above are used for the evaluation of the acceptance probability α above. Other than the above procedure for generating candidates, the M-H algorithm proceeds in the standard fashion. The end result is a set of particles which are the desired draws from the posterior

$$\left\{\boldsymbol{x}_{t\,|\,t}^{(i)}\right\} = \left\{\boldsymbol{x}_{t\,|\,t}^{m,(i)}\right\} \sim p(\boldsymbol{x}_t \mid Y_t)$$

The procedure outlined above is a specific application of the resample-move algorithm of Gilks and Berzuini (2001). It is tailored specifically for state space models and the filtering problem. Note that in the analysis of this paper the SIR step could be omitted altogether, but then Markov chain would then require a burn-in period since it would not be able to start with draws from the posterior (which are provided by SIR).

REFERENCES

Allen JI, Eknes M, Evensen G. 2003. An ensemble Kalman filter with a complex marine ecosystem model: hindcasting phytoplankton in the Cretan Sea. *Annales Geophysicae* **21**: 399–411.

Alspach DL, Sorenson HW. 1972. Nonlinear Bayesian estimation using Gaussian sum approximations. *IEEE Transactions on Automatic Control* **AC-17**: 439–448.

Arulampalam MS, Maskell S, Gordon N, Clapp T. 2002. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing* **50**(2): 174–188.

Bailey B, Doney SC, Lima ID. 2004. Quantifying the effects of dynamical noise on the predictability of a simple ecosystem model. *Environmetrics* **15**: 337–355.

Bertino L, Evensen G, Wackernagel H. 2003. Sequential data assimilation techniques in oceanography. *International Statistical Review* **71**(2): 223–241.

Bjørnstad ON, Grenfell BT. 2001. Noisy clockwork: time series analysis of population fluctuations in animals. *Science* **293**: 638–643.

Campbell JW. 1995. The lognormal distribution as a model for bio-optical variability in the sea. *Journal of Geophysical Research* **100**: 13237–13254.

Carlin BP, Chib S. 1995. Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society* **B 57**(3): 473–484.

Carlin BP, Polson NG, Stoffer DS. 1992. A Monte Carlo approach to nonnormal and nonlinear state space modelling. *Journal of the American Statistical Association* **87**: 493–500.

Chib S, Greenberg E. 1995. Understanding the Metropolis-Hastings Algorithm. *The American Statistician* **49**(4): 327–335.

Cloern JE, Grenz C, Vidergar-Lucas L. 1995. An empirical model of the plankton chlorophyll: carbon ratio—the conversion factor between productivity and growth rate. *Limnology and Oceanography* **40**(7): 1313–1321.

Cullen JJ, Lewis MR. 1988. The kinetics of algal photoadaptation in the context of vertical mixing. *Journal of Plankton Research* **10**: 1039–1063.

Cullen JJ, Ciotti AM, Davis RF, Lewis MR. 1997. Optical detection and assessment of algal blooms. *Limnololgy and Oceanography* **42**: 1223–1239.

Denman KL, Gargett AE. 1983. Time and space scales of vertical mixing and advection of phytoplankton in the upper ocean. *Limnology and Oceanography* **28**(5): 801–815.

Doucet A, de Freitas N, Gordon N. 2001. *Sequential Monte Carlo Methods in Practice*. Springer: New York; 581.

Dowd M. 2005. A biophysical coastal ecosystem model for assessing environmental effects of marine bivalve aquaculture. *Ecological Modelling* **183**(2–3): 323–346.

Dowd M, Meyer R. 2003. A Bayesian approach to the ecosystem inverse problem. *Ecological Modelling* **169**: 39–55.
Dowd M, Martin JL, LeGresley MM, Hanke A, Page FH. 2003. Interannual variability in a plankton time series. *Environmetrics* **14**(1): 73–86.
Edwards AM. 2001. Adding detritus to a nutrient-phytoplankton-zooplankton model: a dynamical systems approach. *Journal of Plankton Research* **23**: 389–413.
Edwards AM, Platt T, Sathyendranath S. 2004. The high nutrient, low chlorophyll regime of the ocean: limits on biomass and nitrate before and after iron enrichment. *Ecological Modelling* **171**: 103–125.
Evans SN, Stark PB. 2002. Inverse problems as statistics. *Inverse Problems* **18**: R55–R97.
Evensen G. 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics* **53**: 343–367.
Fasham MJR. 1993. Modelling the marine biota. In *The Global Carbon Cycle*, Heimann M (ed.). Spinger-Verlag: Berlin; 457–504.
Fennell W, Neumann T. 2004. *Introduction to the Modelling of Marine Ecosystems*. Elsevier: Amsterdam; 330.
Gilks WR, Berzuini C. 2001. Following a moving target—Monte Carlo inference for dynamic Bayesian models. *Journal of the Royal Statistical Society B* **63**(1): 127–146.
Gordon NJ, Salmond DJ, Smith AFM. 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F* **140**(2): 107–113.
Gould WJ. 2003. WOCE and TOGA the foundations of the Global Ocean Observing System. *Oceanography* **16**(4): 24–30.
Harmon R, Challenor P. 1997. A Markov chain Monte Carlo method for estimation and assimilation into models. *Ecological Modelling* **101**: 41–59.
Harvey AE. 1989. Forecasting, structural time series models and the Kalman filter. Cambridge University Press: Cambridge; 554.
Hofmann EE, Lascara CM. 1998. Overview of interdisciplinary modeling for marine ecosystems. In *The Sea: Ideas and Observations on Progress in the Study of the Seas. The Global Coastal Ocean: Processes and Methods*, Vol. 10, Brink KH, Robinson AR (eds). Wiley: New York.
Hooker SB, McClain CR. 2000. The calibration and validation of SeaWiFS data. *Progress in Oceanography* **3–4**: 427–465.
Jazwinski AH. 1970. *Stochastic Processes and Filtering Theory.* Academic Press: New York; 376.
Kalman RE. 1960. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, Transactions AMSE, Series D* **82**: 35–45.
Kitagawa G. 1998. A self-organizing state space model. *Journal of the American Statistical Association* **93**(443): 1203–1215.
Kitagawa G. 1996. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics* **5**: 1–25.
Kitagawa G. 1987. Non-Gaussian state-space modeling of nonstationary time series. *Journal of the American Statistical Association* **82**: 1032–1041.
Kong A, Liu JS, Wong WH. 1994. Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association* **89**(425): 278–288.
Lawson LM, Sptiz YH, Hofmann EE, Long RL. 1995. A data assimilation technique applied to a predator-prey model. *Bulletin of Mathematical Biology* **57**: 593–617.
Lee DS, Chia NKK. 2002. A particle algorithm for sequential Bayesian parameter estimation and model selection. *IEEE Transactions on Signal Processing* **50**(2): 326–336.
Losa SN, Kivman GA, Schroter J, Wenzel M. 2003. Sequential weak constraint parameter estimation in an ecosystem model. *Journal of Marine Systems* **43**: 31–49.
Lui JS, Chen R. 1998. Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association* **93**(443): 1032–1044.
Marion G, Renshaw E, Gibson G. 2000. Stochastic modelling of environmental variation for biological populations. *Theoretical Population Biology* **57**: 197–217.
McClain CR, Barnes RA, Eplee RE, *et al*. 2000. SeaWiFS Postlaunch Calibration and Validation Analyses, Part 2. NASA Technical Memorandum 2000–206892, Vol. 10. Hooker SB, Firestone ER (eds). NASA Goddard Space Flight Center, 57.
Meyer R, Christensen NL. 2000. Bayesian reconstruction of chaotic dynamical systems. *Physical Review E* **62**: 3535–3542.
Monahan AH, Denman KL. 2004. Impacts of atmospheric variability on a coupled upper-ocean/ecosystem model of the subarctic Northeast Pacific. *Global Biogeochemical Cycles* **18**: GB2010.
Natvik LJ, Evensen G. 2003. Assimilation of ocean color data into a biochemical model of the North Atlantic. Part 2. Statistical Analysis. *Journal of Marine Systems* **40–41**: 155–169.
O'Reilly JE, Maritorena S, Mitchell BG, *et al*. 1998. Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research* **103**: 24937–24953.
Seigel DA, Karl DM, Michaels AF. 2001. Interpretations of biogeochemical processes for the US JGOFS Bermuda and Hawaii time series sites. *Deep Sea Research II* **48**: 1403–1404.
Steele JH, Henderson EW. 1981. A simple plankton model. *American Naturalist* **117**: 676–691.

Talley LD, Fryer G, Lumpkin R. 2005. Physical Oceanography of the Topical Pacific. In *Geography of the Pacific Islands*, pp. 19–32, Rapaport M (ed.).

Thompson KR, Dowd M, Lu Y, Smith B. 2000. Oceanographic data assimilation and regression analysis. *Environmetrics* **11**: 183–196.

Tunnicliffe V, Dewey R, Smith D. 2003. Research plans for a cabled seafloor observatory in western Canada. *Oceanography* **16**(4): 53–59.

West M, Harrison PJ, Migon HS. 1985. Dynamic generalized linear models and Bayesian forecasting (with discussion). *Journal of the American Statistical Association* **80**: 73–97.

Vallino JJ. 2000. Improving marine ecosystem models: use of data assimilation and mesocosm experiments. *Journal of Marine Research* **58**: 117–164.

