Sequential Monte Carlo Methods for State and and Parameter Estimation (with application to ocean biogeochemistry)

Michael Dowd
Dept of Mathematics and Statistics
Dalhousie University
Halifax, N.S., Canada

NSERC CRSNG

Canadian Foundation for Climate and Atmospheric Sciences (CFCAS)
Fondation canadienne pour les sciences du climat et de l'atmosphère (FCSCA)

NATIONAL PROGRAM ON COMPLEX DATA STRUCTURES

BIRS Data Assimilation Workshop, *February 3-8, 2008*

Outline

1. Motivation: observations and dynamic models for ocean biogeochemistry

2. The state space model for nonlinear and nonGaussian systems: filtering and smoothing

3. Sequential Monte Carlo approaches: resampling/bootstrap and MCMC

4. Static parameter estimation for stochastic dynamics: likelihood and state augmentation

Statistical Estimation for Nonlinear NonGaussian Dynamic Systems

Approaches …

• Statistical emulators/ computer experiments for studying large scale dynamical model and perhaps DA.

• Functional data analysis applied to estimating differential equations

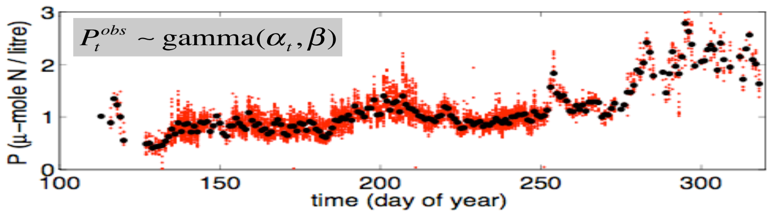
• Hierarchical Bayes and Markov Chain Monte Carlo

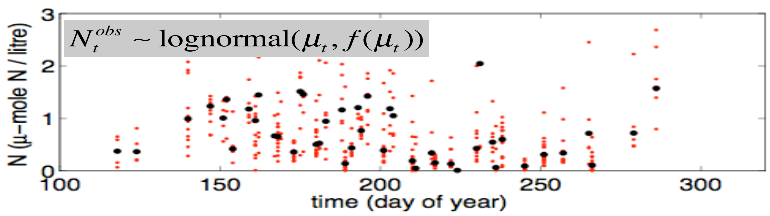
• Sequential Monte Carlo approaches*

Data

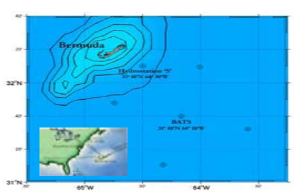
Ocean Biogeochemical Time Series

- Ocean Observatory at Lunenburg, Canada

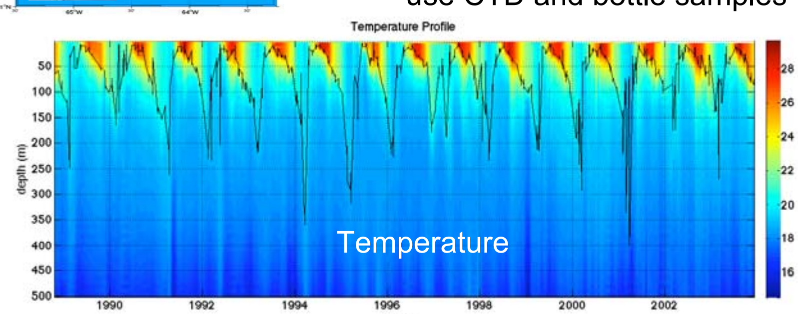


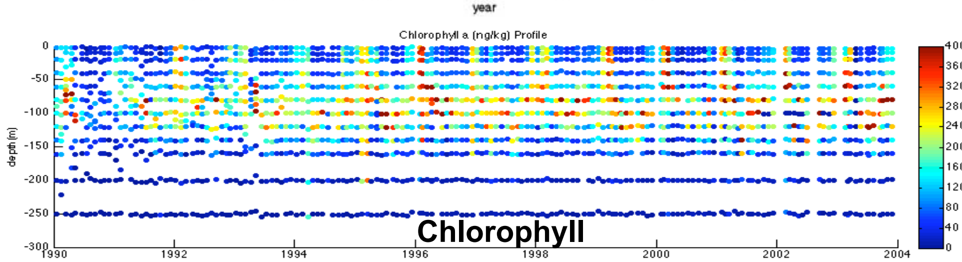


Long Term Ocean Time - Space Series



Bermuda Atlantic Time Series:
- 15 years, monthly cruises
- measure depth profiles of biogeochemical variables
- use CTD and bottle samples





Dynamic Models

A Physical - Biogeochemical Model

$$\frac{dP}{dt} = g\left\{N;k_N\right\}\gamma P - \lambda_P P - g\left\{P;k_P\right\}IZ + n_P$$

$$\frac{dZ}{dt} = \varepsilon g\left\{P;k_P\right\}IZ - \lambda_Z Z + n_Z$$

$$\frac{dN}{dt} = \phi D + \beta g\left\{P;k_P\right\}IZ - g\left\{N;k_N\right\}\gamma P + \nu\lambda_Z Z + n_N$$

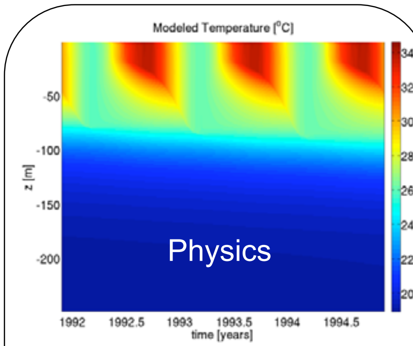
$$\frac{dD}{dt} = -\phi D + \lambda_P P + (1-\varepsilon-\beta) g\left\{P;k_P\right\}IZ + (1-\nu)\lambda_Z Z + n_D$$

$$\frac{\partial X}{\partial t} - \frac{\partial}{\partial z}\left(K_t'\frac{\partial X}{\partial z}\right) = SMS(X)$$

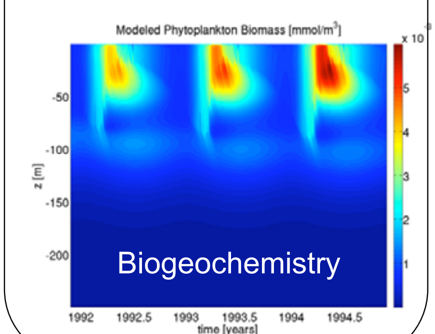
$$\frac{\partial u}{\partial t} - \frac{\partial}{\partial z}\left(K_t\frac{\partial u}{\partial z}\right) - fv = F_u \qquad \frac{\partial S}{\partial t} - \frac{\partial}{\partial z}\left(K_t'\frac{\partial S}{\partial z}\right) = F_S$$

$$\frac{\partial v}{\partial t} - \frac{\partial}{\partial z}\left(K_t\frac{\partial v}{\partial z}\right) + fu = F_v \qquad \frac{\partial T}{\partial t} - \frac{\partial}{\partial z}\left(K_t'\frac{\partial T}{\partial z}\right) = F_T$$

where turbulence sub-model computes K_t and K_t' from u,v,T and S



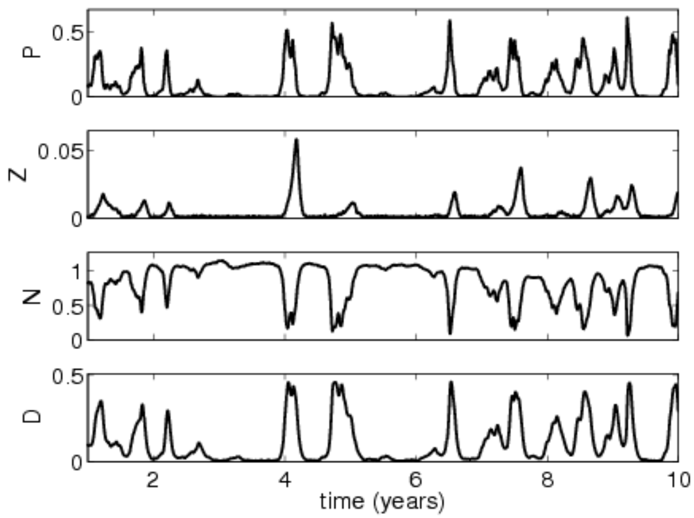
Sample Output



Incorporating Stochasticity

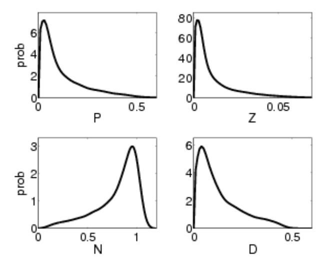
(stochastic photosynthetic parameter + system noise)

O-D ODE based PZND model



- *Frequent transitioning across bifurcation*
- *aperiodic/episodic*
- *dynamical dependencies maintained*
- *note (high freq) forcing versus (low freq) response*

Describe ensemble properties as a distribution →



The estimation problem
for system state and parameters …

The State Space Model

dynamics equation	measurement equation
$x_t = f(x_{t-1}, \theta, n_t)$	$y_t = h(x_t, \phi, \nu_t)$
or	*or*
$x_t \sim p(x_t \mid x_{t-1}, \theta)$	$y_t \sim p(y_t \mid x_t, \phi)$

- Given $Y_\tau = (y_1', \ldots, y_\tau')'$

$\rightarrow$ want to jointly estimate the state x_t *and* static parameters θ and ϕ

*General Case (Nonlinear Stochastic Dynamics): Filtering and Smoothing for State Estimation**

Filtering:

$$p(x_t \mid Y_t, \theta, \phi) \propto p(y_t \mid x_t, \phi) \int p(x_t \mid x_{t-1}, \theta)\, p(x_{t-1} \mid Y_{t-1}, \theta, \phi)\, dx_{t-1}$$

for $t=1, \ldots, T$, given $p(x_0)$

Smoothing:

$$p(x_{1:T} \mid Y_T, \theta, \phi) = p(x_0) \prod_{t=1}^{T} p(x_t \mid x_{t-1}, \theta) \prod_{t=1}^{T} p(y_t \mid x_t, \phi)$$

$\rightarrow$ nonlinear, non-Gaussian case can be treated with sampling based solutions (via sequential MC methods)

* treat parameter estimation later on …

Sequential Monte Carlo Approaches

1. *Stochastic dynamic prediction*: numerical integration of stochastic dynamic system (generate forecast ensemble)

$$x_{t|t-1}^{(i)} = f(x_{t-1|t-1}^{(i)}, n_t^{(i)}, \theta), \quad i = 1,....,n$$

$$\rightarrow \quad \left\{x_{t|t-1}^{(i)}\right\} \sim p(x_t \mid Y_{t-1}, \theta, \phi)$$

Ensemble must cover the part of state space with non-negligible values of the predictive density

2. *Bayesian blending* of measurements and numerical model predictions (e.g. resampling, MCMC). $\rightarrow$

Sequential Bayesian Monte Carlo

$$\left\{x_{t|t-1}^{(i)}\right\} \sim p(x_t \mid Y_{t-1}, \theta, \phi) \longrightarrow \left\{x_{t|t}^{(i)}\right\} \sim p(x_t \mid Y_t, \theta, \phi)$$

(a) SIR - compute: $w_t^{(i)} = w_{t-1}^{(i)} \; p(y_t \mid x_{t|t-1}^{(i)}), \quad i = 1,....,n$

- weighted resample of $\left\{x_{t|t-1}^{(i)}, w_t^{(i)}\right\} \rightarrow \left\{x_{t|t}^{(i)}\right\} \sim p(x_t \mid Y_t)$

or

(b) Sequential Metropolis Hastings MCMC $\longrightarrow$

(c) Resample - Move (SIR/MCMC)

(d) Ensemble/unscented Kalman filter (approximate)

+ …

Sequential Metropolis-Hastings

Basic Idea: Given $\left\{x_{t-1|t-1}^{(i)}\right\} \sim p(x_{t-1} \mid Y_{t-1}, \theta, \phi), \quad i = 1, \ldots, n$

loop over k

1. Generate candidate from predictive density:

$$x_{t|t-1}^{*} \sim p(x_t \mid Y_{t-1}, \theta, \phi) \qquad \text{via: } x_{t|t-1}^{*} = f(x_{t-1|t-1}^{*}, n_t^{*}, \theta)$$

2. Evaluate acceptance probability

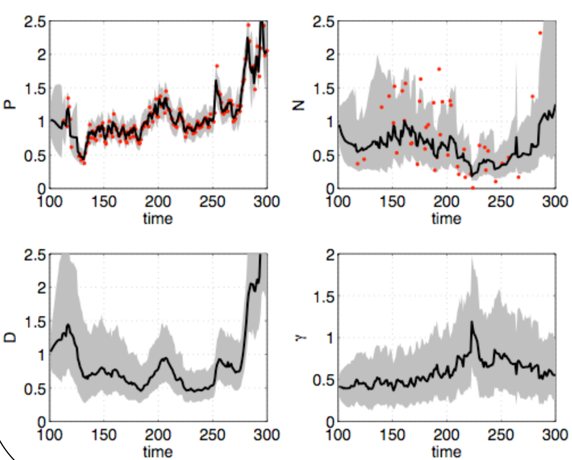
$$\alpha = \min\left(1, \frac{p(y_t \mid x_{t|t-1}^{*}, \theta, \phi)}{p(y_t \mid x_{t|t}^{(k)}, \theta, \phi)}\right), \text{ choose } x_{t|t-1}^{*} \text{ or } x_{t|t}^{(k)}$$

$\longrightarrow$ *sample from target:* $\left\{x_{t|t}^{(i)}\right\} \sim p(x_t \mid Y_t, \theta, \phi)$

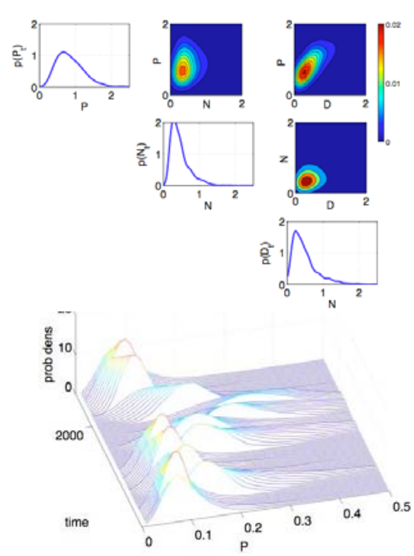
- ***Flexible and configurable, e.g adaptive ensemble***
- ***EFFICIENT PROPOSALS ARE KEY, e.g prior, or from EnKF?***

Filter State Estimates

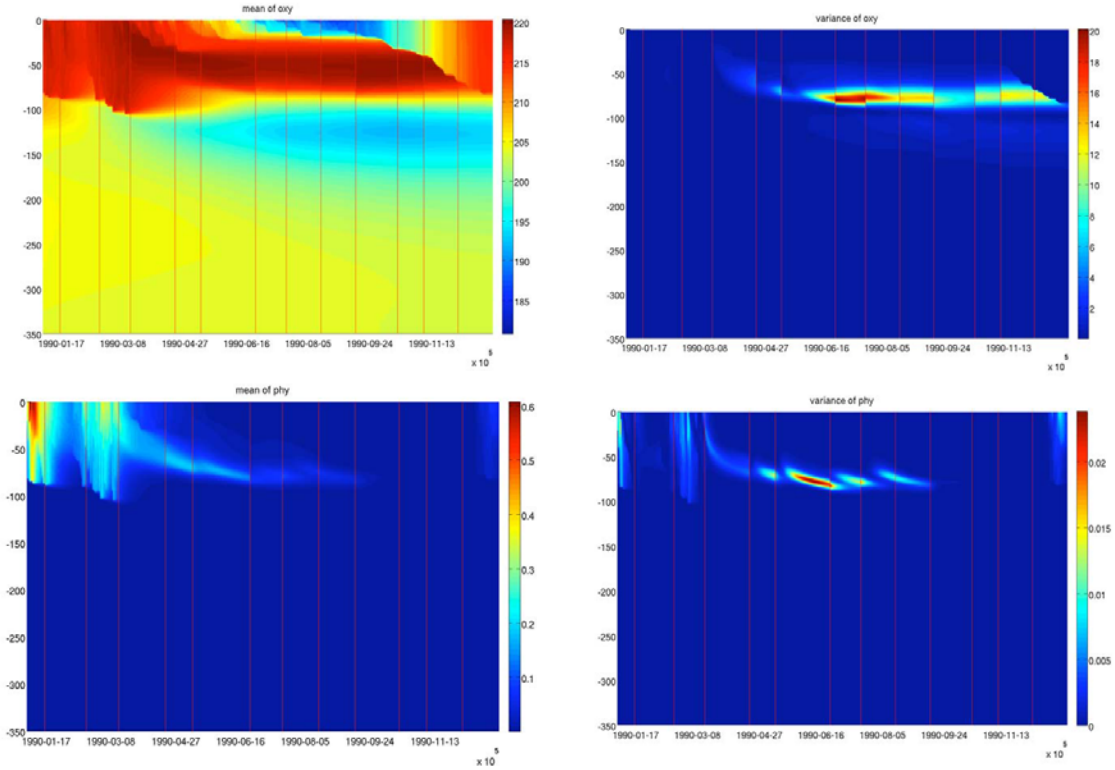
Time series of median and percentiles



Distributions



Example SIR Results from Physical Biogeochemical Model



Comparison of SMC Methods : Convergence of Distributions

$$\text{<K-L divergence>} = \left\langle \int p(x_t \mid y_{1:t}) \log \frac{p(x_t \mid y_{1:t})}{\tilde{p}(x_t \mid y_{1:t})} dx_t \right\rangle$$

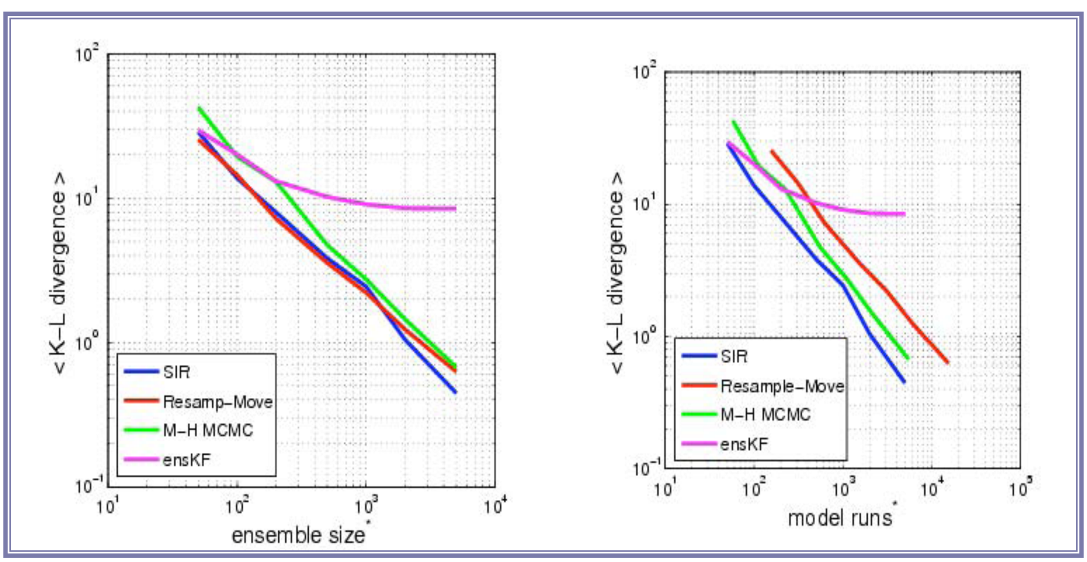
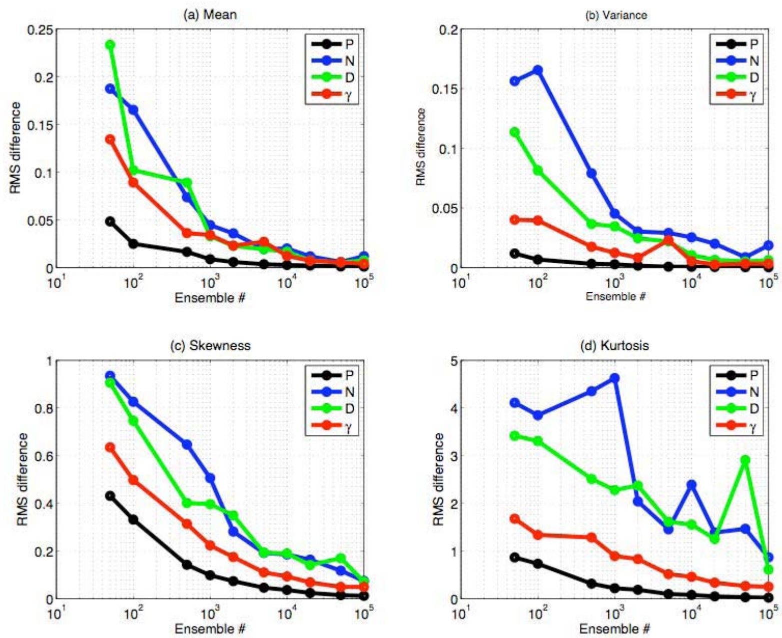


Figure: Convergence to “exact” solution for different SMC methods

Convergence of Moments (M-H MCMC)



Parameter Estimation via Likelihood

The likelihood arising from the state space model* is

$$L(\theta \mid Y_T) = p(Y_T \mid \theta) = \prod_{t=1}^{T} p(y_t \mid Y_{t-1}, \theta) = \prod_{t=1}^{T} \int p(y_t \mid x_t, \theta) p(x_t \mid Y_{t-1}, \theta) dx_t$$

From sequential MC filter we can compute predictive density

$$\{x_{t|t-1}^{(i)}\} \sim p(x_t \mid Y_{t-1}, \theta)$$

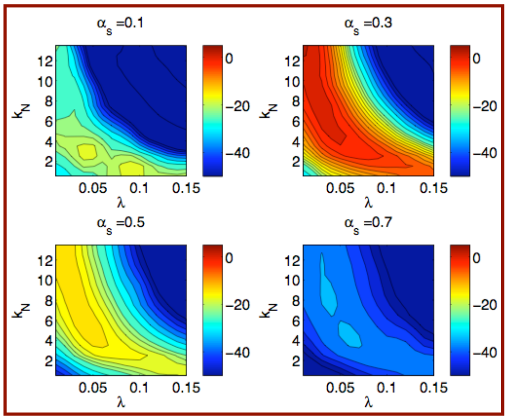
and so compute the likelihood as

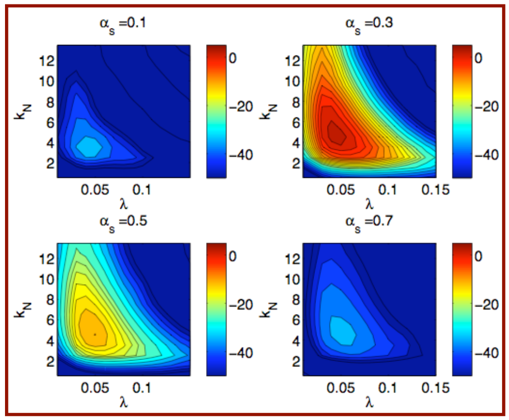
$$L(\theta \mid Y_T) \approx \frac{1}{n} \prod_{t=1}^{T} \left(\sum_{i=1}^{n} p(y_t \mid x_{t|t-1}^{(i)}, \theta \right)$$

*assume ϕ is given, and suppress the explicit dependence)

Distributions for Parameters

Likelihood | *Posterior: using prior info*





↑ Parameter identifiability issues

↑ priors 'focus' the likelihood

(sample based) likelihood surface is 'rough'
→ challenge for optimizers (stochastic gradients)

Parameter Estimation via State Augmentation

Idea: *Append state* to include parameters, $\tilde{x}_t = \begin{pmatrix} x_t \\ \theta_t \end{pmatrix}$

Specify $p(\theta_0)$ and allow parameter to evolve as $\theta_t = \theta_{t-1}$

Choose θ_T as estimate for parameter

For practical implementation with finite sample, we must

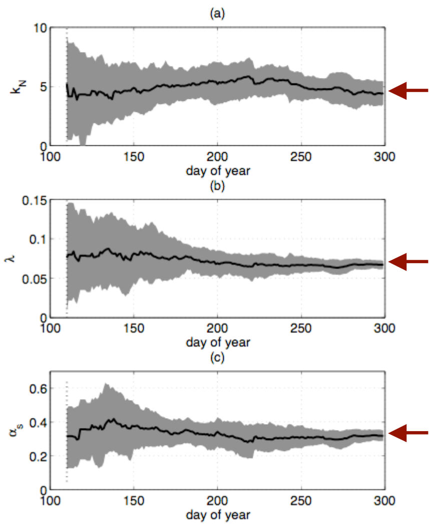
(1) Specify *initial ensemble* $\{\theta_0^{(i)}\} \sim p(\theta)$ (including dependence structure between the parameters)

(2) At each *t*, introduce *smoothed bootstrap* of $\{\theta_t^{(i)}\}$ (with dispersion correction) to generate diversity in parameter ensemble, while maintaining distributional properties.

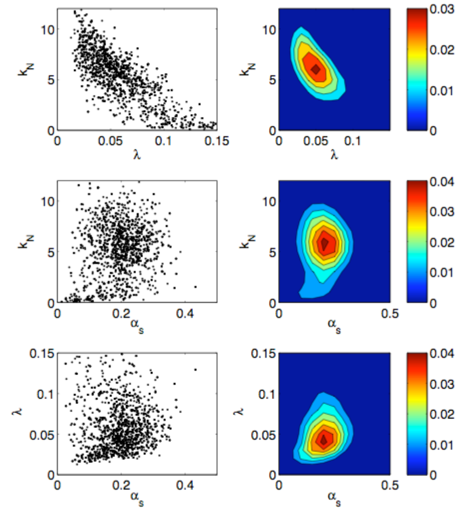
Easy and seems to work in practice, little theoretical guidance on convergence.
Does not seem to work well with EnKF

State Augmentation Example

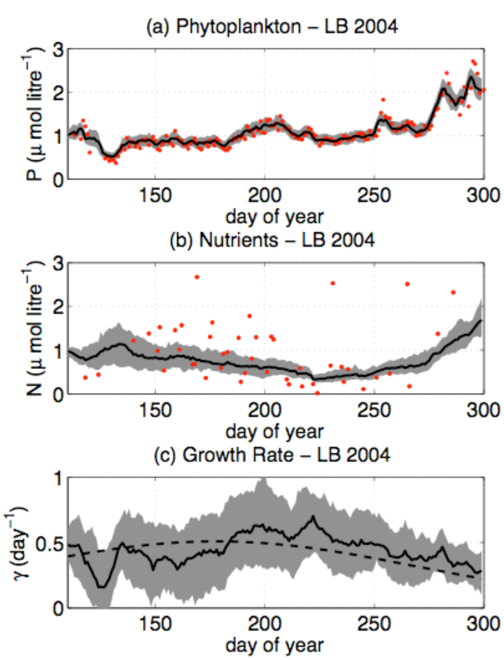
Trace plot of one realization



Parameter values from 2000 realizations



State Reconstruction by Smoother



- Fixed interval smoother using optimal parameters.
- Uses forward M-H MCMC filter
- Smoother realizations provided by backwards sweep smoother algorithm of Godsill et al (2004)

Remarks and Outstanding Issues on Fully Bayesian DA via Sequential Monte Carlo

1. Sequential MC approaches allow for state and parameter estimation in nonlinear nonGaussian dynamic systems. Wide variety available (bootstrap or MCMC) and easy to implement, but computationally ….

2. Static parameter estimation in SDEs outstanding statistical issue. Likelihood (via predictive density). State augmentation (via filter density). EM algorithm (via smoother density)

3. Effective stochastic simulation (integration) and specification of model errors a key feature.

4. Adaptation for (large dimension) dynamical systems! $\rightarrow$ need small ensembles (100-1000) to represent large dimensional state space. *Efficient proposal distribution is paramount*, e.g use information flow via dynamics.

5. Methods for computationally efficient smoothing also needed.

6. Information based metrics for assessing improvements and comparing approaches