Joint state–parameter estimation for nonlinear stochastic energy balance models

Fei Lu\textsuperscript{1} Nils Weitzel\textsuperscript{2} Adam Monahan\textsuperscript{3}

\textsuperscript{1}Department of Mathematics, Johns Hopkins
\textsuperscript{2}Meteorological Institute, University of Bonn, Germany
\textsuperscript{3}School of Earth and Ocean Science, University of Victoria, Canada

SIAM DS 2019, May 22
1. An SPDE from paleoclimate reconstruction
   - Stochastic energy balance model
   - State space model representation

2. Bayesian joint state-parameter estimation
   - Sampling the posterior: Particle MCMC
   - Ill-posedness: regularized posterior

3. Numerical results
   - Parameter estimation
   - State estimation
Paleoclimate: reconstruct past climate temperature from proxy data

- the temperature: a spatio-temporal process
  - physically laws: energy balance → SPDEs
  - discretized: a high-D process with spatial correlation

Sparse and noisy data

- Proxy data: historical data, tree rings, ice cores, fossil pollen, ocean sediments, coral etc.

Plan: inference of SPDEs from sparse noisy data

- joint state-parameter estimation
The SPDEs: stochastic Energy Balance Models

Idealized atmospheric energy balance (Fanning&Weaver1996)

\[
\begin{align*}
\partial_t u &= Q_T + Q_{SW} + Q_{SH} + Q_{LH} + Q_{LW} - Q_{LPW} \\
&= \nabla \cdot (\nu \nabla u) + \theta_0 + \theta_1 u + \theta_4 u^4 + \mathcal{W}(t, x) \\
\end{align*}
\]

- \(\theta = (\theta_k)\): unknown parameters
  - prior: a range of physical values
  - \(g_\theta(u)\) has a \textbf{stable} fixed point

- \(\mathcal{W}(t, x)\): Gaussian noise,
  - white-in-time Matern-in-space

Data: sparse noisy observations
State space model formulation

SEBM: \[ \frac{\partial}{\partial t} u = \nabla \cdot (\nu \nabla u) + \sum_{k=0,1,4} \theta_k u^k + W(t, x) \]

Observation data: \[ y_{t_i} = H(u(t_i, x)) + V_i \]

Discretization (simplification):
- finite elements in space
- semi-backward Euler in time

State space model

SEBM: \[ U_n = g(\theta, U_{n-1}) + W_n \]

Observation data: \[ y_n = HU_n + V_n \]
Joint parameter-state estimation

SEBM: \[ U_n = g(\theta, U_{n-1}) + W_n \]
Observation data: \[ y_n = HU_n + V_n \]

**Goal:** Given \( y_{1:N} \), we would like to jointly estimate \((\theta, U_{1:N})\)
- Gaussian prior for \( \theta \)
- 12 spatial nodes, 100 time steps

![Solution at time step n =10](image1)

![Trajectories of all 12 nodes](image2)
Bayesian joint state-parameter estimation

Bayesian approach:

\[ p(\theta, u_{1:N}|y_{1:N}) \propto p(\theta)p(u_{1:N}|\theta)p(y_{1:N}|u_{1:N}) \]

- Posterior: quantifies the uncertainties

Approximate the posterior by sampling

- high dimensional (> \(10^3\)),
- non-Gaussian, mixed types of variables \(\theta, u_{1:N}\)
- Gibbs Monte Carlo: \(U_{1:N}|\theta\) and \(\theta|U\) iteration
  - \(U_{1:N}|\theta\) needs highD proposal density \(\rightarrow\) Sequential MC
  - combine SMC with Gibbs (MCMC) \(\rightarrow\)

Particle MCMC methods based on conditional SMC
Sampling: particle MCMC

Particle MCMC (Andrieu&Doucet&Holenstein10)

- Combines Sequential MC with MCMC:
  - SMC: seq. importance sampling $\rightarrow$ highD proposal density
  - conditional SMC: keep a reference trajectory in SMC
  - MCMC transition by conditional SMC
    $\rightarrow$ target distr invariant even w/ a few particles

- Particle Gibbs with Ancestor Sampling (Lindsten&Jordan&Schon14)
  - Update the ancestor of the reference trajectory
  - Improving mixing of the chain
Ill-posed inverse problem

For the Gaussian prior $p(\theta)$,
unphysical samples of posterior: systems blowing up
Ill-posed inverse problem

For the Gaussian prior $p(\theta)$, unphysical samples of posterior: systems blowing up

Parameter estimation is ill-posed:

Singular Fisher infomation matrix

$\rightarrow$ large oscillation in sample $\theta$ from Gibbs $\theta | \hat{U}_{1:N}$

Std of errors of MLE from noisy observations
Regularized posterior

Recall the regularization in variational approach

Variational: \[ (\hat{\theta}, \hat{u}_{1:N}) = \arg \min_{(\theta, u_{1:n})} C_{\lambda,y_{1:N}}(\theta, u_{1:N}) \]

Bayesian: \[ p_{\lambda}(\theta, u_{1:N}|y_{1:N}) \propto p(\theta)^{\lambda} p(y_{1:N}|u_{1:N})p(u_{1:N}|\theta) \]

\[
C_{\lambda,y_{1:N}}(\theta, u_{1:N}) = \lambda \log p(\theta) + \log[p(y_{1:N}|u_{1:N})p(u_{1:N}|\theta)]
\]

\[
= \lambda \left( \log p(\theta) + \frac{1}{\lambda} \log[p(y_{1:N}|u_{1:N})p(u_{1:N}|\theta)] \right)
\]

- \( \lambda = 1 \): Standard posterior \( \overset{N \to \infty}{\sim} \) likelihood\(^1\)
- \( \lambda = N \): regularized posterior

\[ p_{\lambda}(\theta, u_{1:N}|y_{1:N}) \propto p(\theta) \left[ p(y_{1:N}|u_{1:N})p(u_{1:N}|\theta) \right]^{1/N} \]

\(^1\)Bernstein-von Mises theorem
posterior close to prior;

Errors in 100 simulations

<table>
<thead>
<tr>
<th></th>
<th>$\theta_0$</th>
<th>$\theta_1$</th>
<th>$\theta_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posterior mean</td>
<td>$-0.44 \pm 0.58$</td>
<td>$0.09 \pm 0.42$</td>
<td>$0.11 \pm 0.20$</td>
</tr>
<tr>
<td>MAP</td>
<td>$-0.32 \pm 0.61$</td>
<td>$0.02 \pm 0.42$</td>
<td>$0.03 \pm 0.21$</td>
</tr>
</tbody>
</table>
State estimation

Observed node: noise filtered

Unobserved node: large spread
State estimation

Observed node: noise filtered

Unobserved node: large spread

Sample trajectories of node 1. Relative error of Mean = 0.007

Sample trajectories of node 2. Relative error of Mean = 0.008

Time steps

State

Probability
Observing more or less nodes:

When more modes are observed:

- State estimation gets more accurate

- Parameter estimation does not improve much:
  The posterior keeps close to prior due to the need of regularization
Bayesian approach to jointly estimate parameter-state

- a stochastic energy balance model
- sparse and noisy data
- Ill-posed parameter estimation problem
  (The parameters are correlated on a lowD manifold)

Introduced a regularized posterior:

- Enabling state estimation
- Large uncertainty in parameter estimation due to ill-posedness
Open questions

1. Re-parametrization/ nonparametric to avoid ill-posedness?

2. How many nodes need to be observed (for large mesh)?
   (theory of determining modes)
Open questions

1. Re-parametrization/ nonparametric to avoid ill-posedness?

2. How many nodes need to be observed (for large mesh)? (theory of determining modes)

Thank you!