Joint state–parameter estimation for nonlinear stochastic energy balance models

Fei Lu [1] Nils Weitzel[2] Adam Monahan[3]

[1]Department of Mathematics, Johns Hopkins

[2]Meteorological Institute, University of Bonn, Germany

[3]School of Earth and Ocean Science, University of Victoria, Canada

SIAM DS 2019, May 22

Outline

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

An SPDE from paleoclimate reconstruction

Paleoclimate: reconstruct past climate temperature from proxy data

- the temperature: a spatio-temporal process
  - physically laws: energy balance $\to$ 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

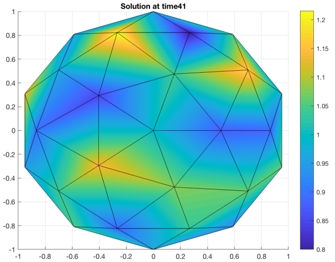
Idealized atmospheric energy balance (Fanning&Weaver1996)

$$\partial_t u = \underbrace{Q_T}_{\text{transport}} + \underbrace{Q_{SW}}_{\substack{\text{absorbed}\\\text{shortwave}}} + \underbrace{Q_{SH}}_{\substack{\text{sensible}\\\text{heat}}} + \underbrace{Q_{LH}}_{\substack{\text{latent}\\\text{heat}}} + \underbrace{Q_{LW}}_{\substack{\text{longwave}\\\text{surf.}\to\text{atmos.}}} - \underbrace{Q_{LPW}}_{\substack{\text{longwave}\\\text{into space}}}$$

$$= \nabla \cdot (\nu \nabla u) + \underbrace{\theta_0 + \theta_1 u + \theta_4 u^4}_{g_\theta(u)} + W(t,x)$$

- $\theta = (\theta_k)$: unknown parameters
  - prior: a range of physical values
  - $g_\theta(u)$ has a **stable** fixed point
- $W(t,x)$: Gaussian noise,
  - white-in-time Matern-in-space

Data: sparse noisy observations



State space model formulation

$$\begin{aligned} \text{SEBM:} \quad \partial_t u &= \nabla \cdot (\nu \nabla u) + \sum_{k=0,1,4} \theta_k u^k + W(t,x) \\ \text{Observation data:} \quad y_{t_i} &= H(u(t_i,x)) + V_i \end{aligned}$$

Discretization (simplification):

- finite elements in space
- semi-backward Euler in time

State space model

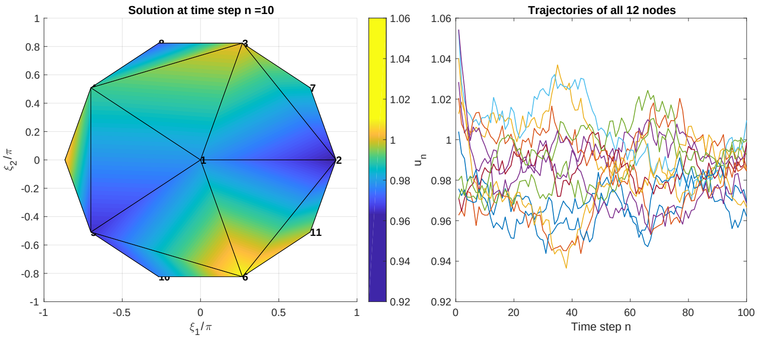
$$\begin{aligned} \text{SEBM:} \quad U_n &= g(\theta, U_{n-1}) + W_n \\ \text{Observation data:} \quad y_n &= HU_n + V_n \end{aligned}$$

Joint parameter-state estimation

$$\begin{aligned} \text{SEBM:} \quad U_n &= g(\theta, U_{n-1}) + W_n \\ \text{Observation data:} \quad y_n &= HU_n + V_n \end{aligned}$$

Goal: Given $y_{1:N}$, we would like to jointly estimate $(\theta, U_{1:N})$

- Gaussian prior for θ
- 12 spatial nodes, 100 time steps



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

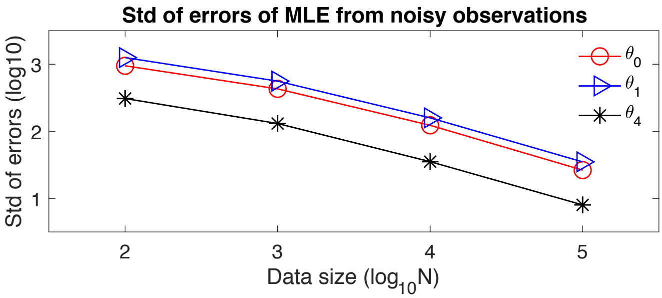
For the Gaussian prior $p(\theta)$,
unphysical samples of posterior: systems blowing up

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 θ from Gibbs $\theta|\widehat{U}_{1:N}$



Regularized posterior

Recall the regularization in variational approach

Variational: $$(\widehat{\theta}, \widehat{u}_{1:N}) = \underset{(\theta, u_{1:n})}{\arg\min}\ C_{\lambda, y_{1:N}}(\theta, u_{1:N})$$

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

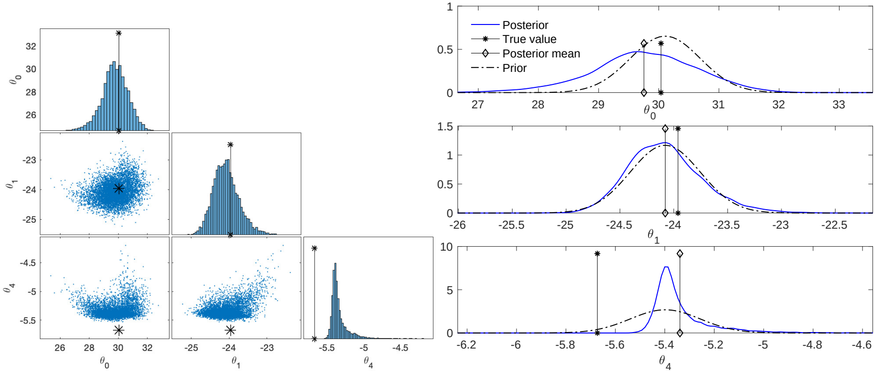
$$\begin{aligned} C_{\lambda, y_{1:N}}(\theta, u_{1:N}) &= \underbrace{\lambda \log p(\theta)}_{\text{regularization}} + \underbrace{\log[p(y_{1:N}|u_{1:N})p(u_{1:N}|\theta)]}_{\text{likelihood}} \\ &= \lambda\left(\log p(\theta) + \frac{1}{\lambda}\log[p(y_{1:N}|u_{1:N})p(u_{1:N}|\theta)]\right) \end{aligned}$$

- $\lambda = 1$: Standard posterior $\xrightarrow{N\to\infty} \sim$ 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

Parameter estimation

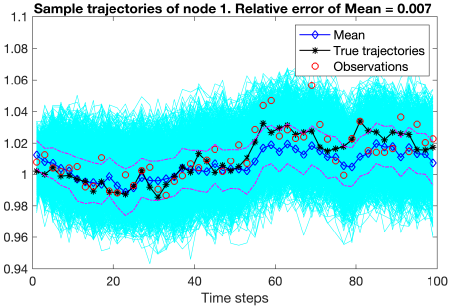


- posterior close to prior;
- Errors in 100 simulations

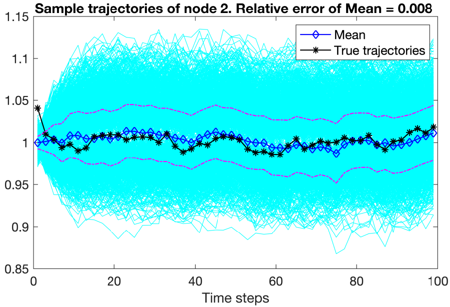
	θ_0	θ_1	θ_4
Posterior mean	-0.44 $\pm$ 0.58	0.09 $\pm$ 0.42	0.11 $\pm$ 0.20
MAP	-0.32 $\pm$ 0.61	0.02 $\pm$ 0.42	0.03 $\pm$ 0.21

State estimation

Observed node: noise filtered

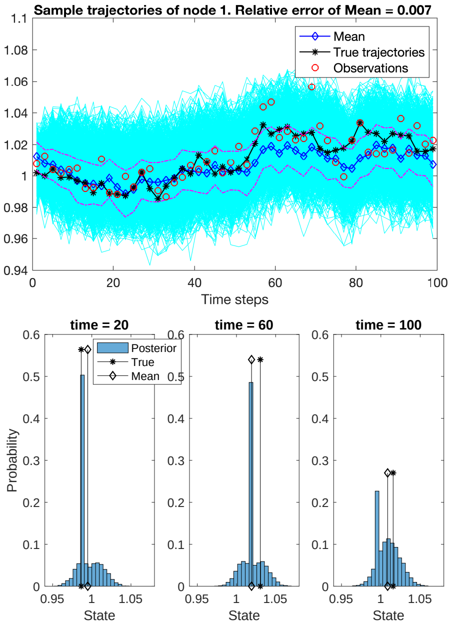


Unobserved node: large spread

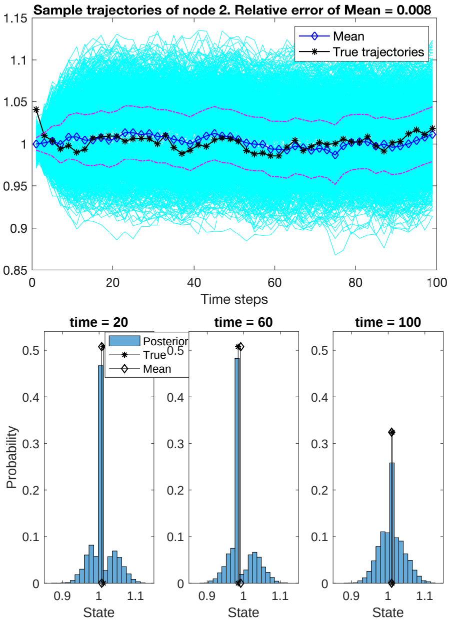


State estimation

Observed node: noise filtered



Unobserved node: large spread



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

Summary and discussion

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!