Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

An Uncertainty Quantification Framework for Land Models

Maoyi Huang[1], Zhangshuan Hou[1], Jaideep Ray[2], Laura Swiler[2], L. Ruby Leung[1]

[1]Pacific Northwest National Laboratory
[2]Sandia National Laboratories

Background and motivation

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Most land models in Earth System Models include numerous sub-models, each representing key processes with mathematical equations and model parameters.

- Quantifying parametric uncertainties and optimizing the parameter values may improve model skill in capturing observed behaviors.

- The land models are highly computationally expensive. It is crucial to take advantage of advances in applied mathematics (e.g., efficient sampling and surrogate model construction) and high performance computing (e.g., big data analytics and parallel algorithms).

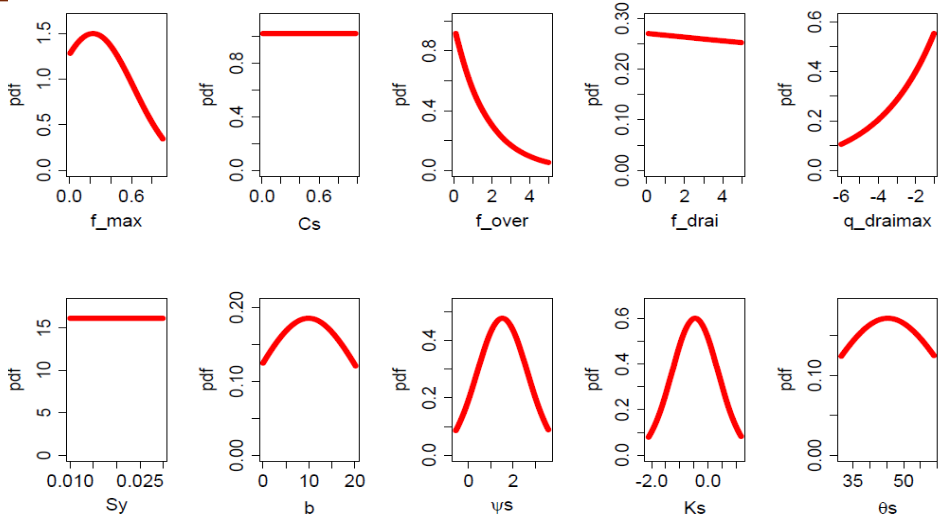
An Uncertainty Quantification Framework for CLM4SP hydrologic parameters

Parameterization
- parameters of interest

Closed-form Prior probability density functions (pdfs)

Quasi Monte Carlo sampling

Realizations of parameter sets



CLM forward modeling

Output responses:
- Latent heat fluxes (LH)
- Sensible heat fluxes (SH)
- Total runoff

Calculate selected metrics for
- Global sensitivity analysis;
- Parameter screening;
- Assessing parameter transferability

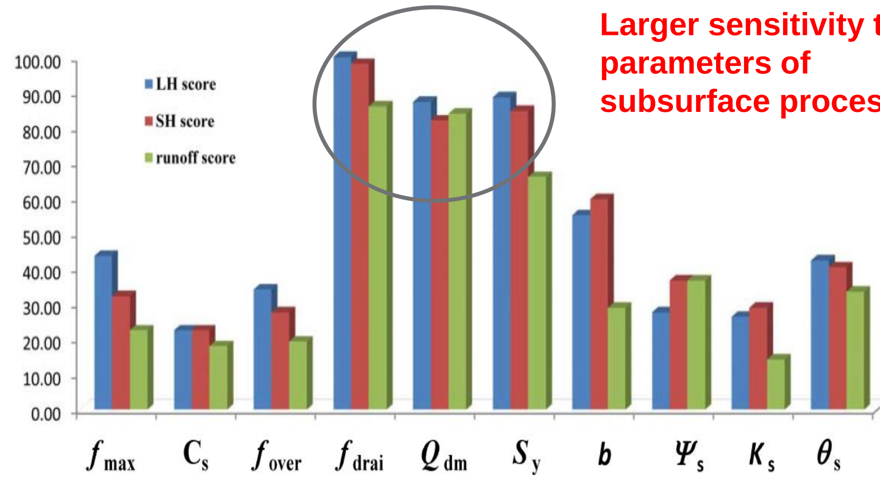
- Surrogate construction
- Bayesian inversion using surrogates/real model

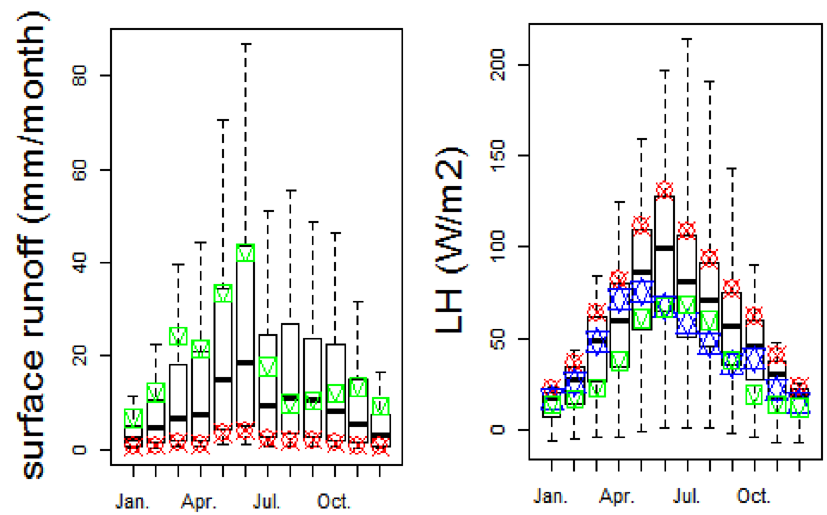
Sensitivity of Simulated Surface Fluxes and Runoff to Hydrologic Parameters

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by **Battelle** *Since 1965*

- CLM4-SP simulated water/energy fluxes show the largest sensitivity to subsurface runoff generation parameters.
- Simulations using default parameters (red) are significantly different from observations at ARM SGP (blue) and a co-located MOPEX site (green).
- With the observations falling within the range of parameter uncertainties, it is feasible to use model inversion to improve water/energy simulations.

Larger sensitivity to parameters of subsurface processes



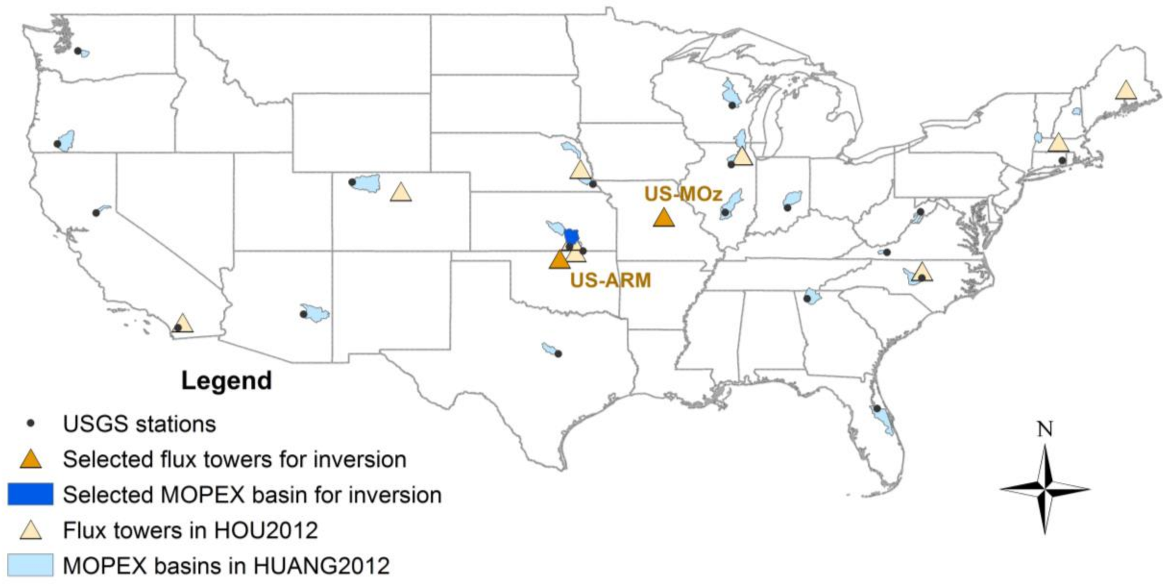


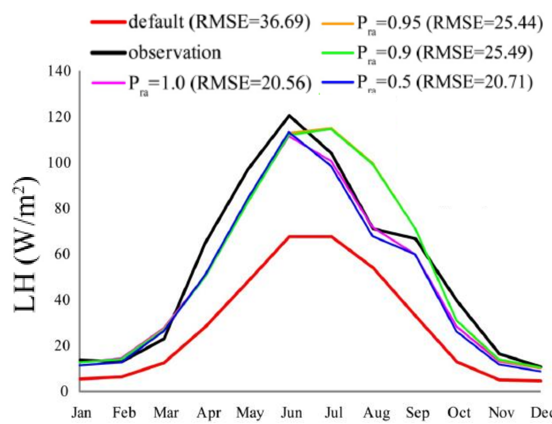
Hou et al. 2012, JGR; Huang et al., 2013, JHM;

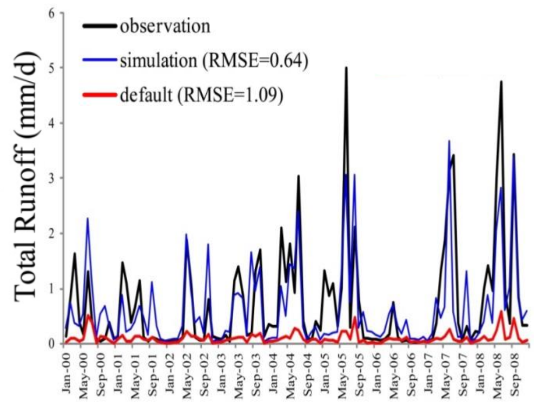
Inverse Modeling of Hydrologic Parameters using Surface Flux and Runoff Observations

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by **Battelle** *Since 1965*

- A Markov Chain Monte Carlo (MCMC) – Bayesian inversion algorithm was implemented to CLM4;
- We evaluated the effects of surface flux and streamflow observations on the inversion results and compare their consistency and reliability using both monthly and daily observations;
- Our results suggest that parameter inversion of CLM4SP is possible, at least at the site level;







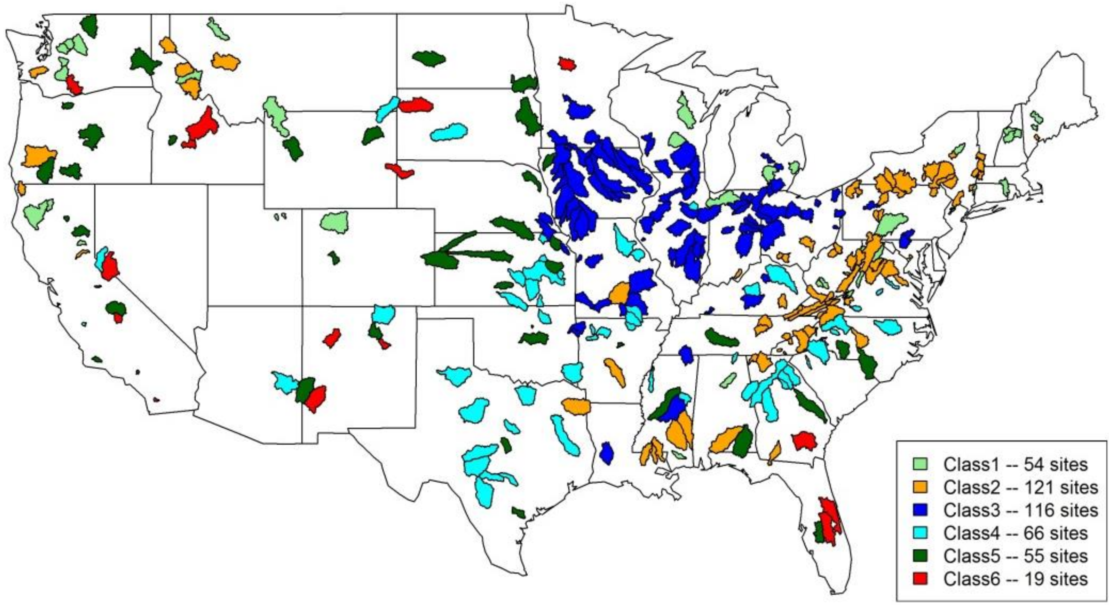
Sun et al., 2013, HESS

MCMC-Bayesian calibrated parameters can significantly improve CLM simulation of energy fluxes and runoff

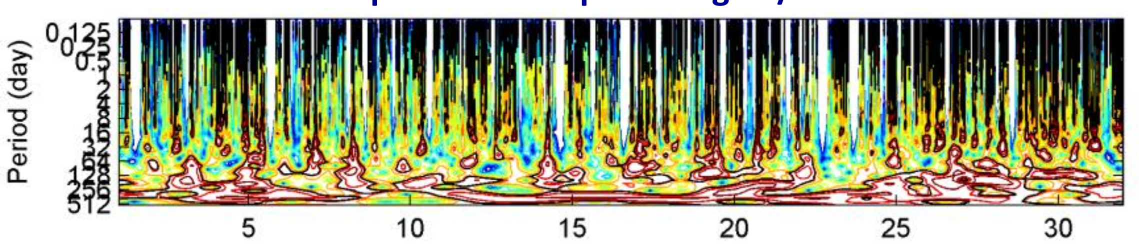
Classification of Hydrological Parameter Sensitivity and Evaluation of Parameter Transferability across 431 US Basins

- Use parameter sensitivity patterns/ attributes, together with climate and soil conditions to classify the basins. The classification yields six classes with unique sensitivity of streamflow simulations to variations in hydrological parameters.
- By grouping a large number basins into a reasonably small number of classes with similar sensitivity behaviors, the same optimization strategy can be used within each class. Model optimization effort can be further reduced given the parameter similarity and transferability.

Sensitivity-based classification of the 431 MOPEX Basins basins



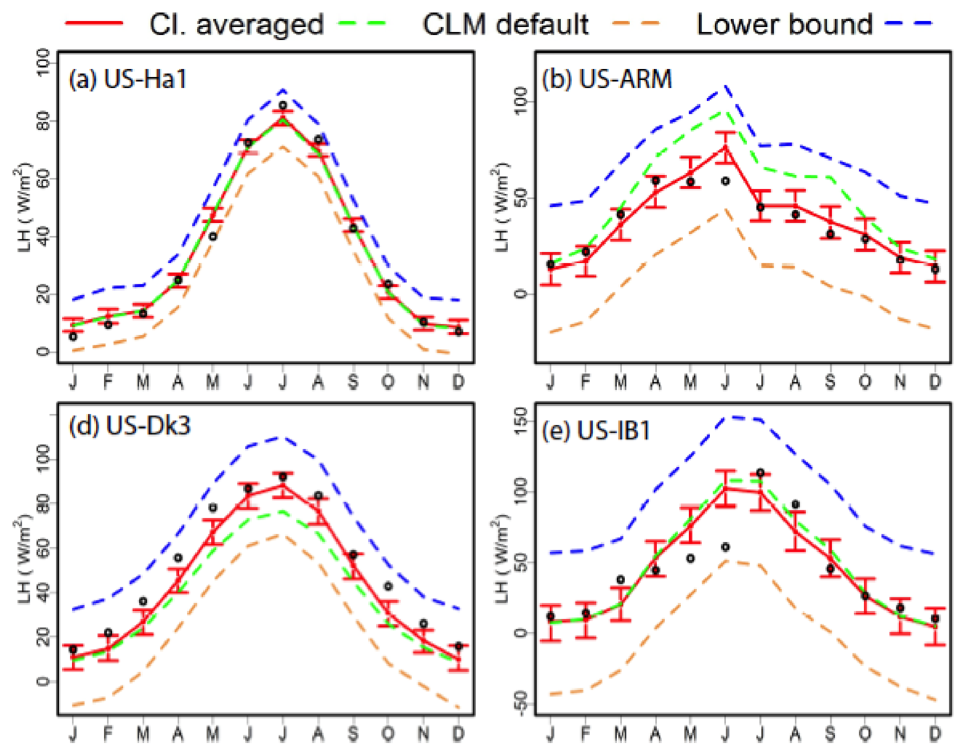
Wavelet decomposition to separate signal/noise for calibration



Ren et al., J Hydrology, 2016

Surrogate-based MCMC-Bayesian Inversion : Case Studies at Flux Tower Sites

- Assessed the feasibility of applying a Bayesian calibration technique in combination with surrogates to estimate CLM4SP parameters;
- Simulated LH from CLM using the calibrated parameters are generally improved at all sites;
- The calibration method also results in credibility bounds around the simulated mean fluxes which bracket the measured data;



- The computational cost is significantly reduced when surrogates are used. On the other hand, a surrogate-based calibration procedure is intrinsically subject to errors as a result of approximating a complex model using simplified functions.

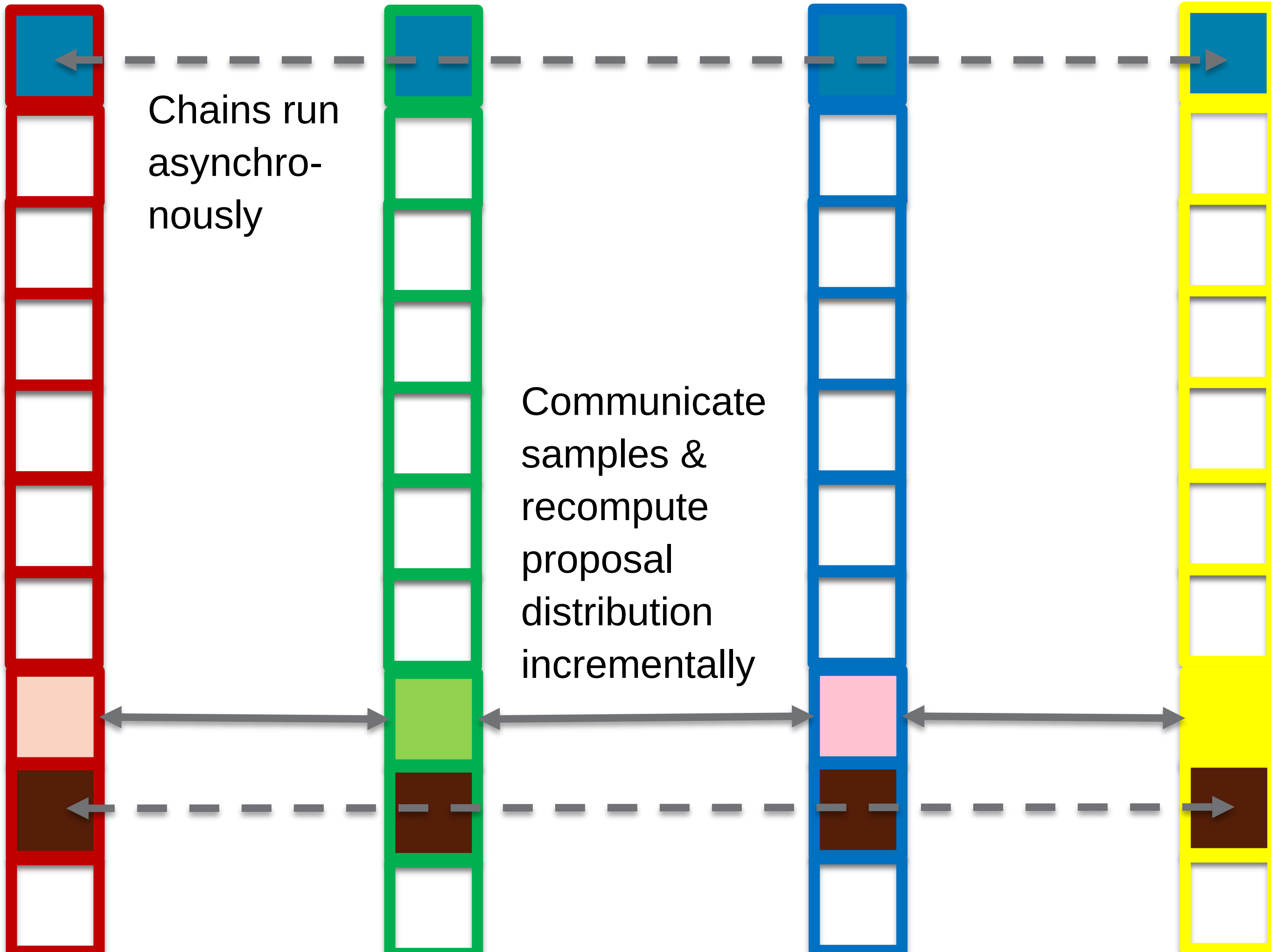
Ray et al., SIAM-JUQ, 2015
Huang et al., JGR, in revision

Scalable adaptive chain ensemble sampling (SACHES):
a parallel MCMC method for calibrating computationally expensive models

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- **Problems with MCMC**
  - *Sampling cost:* Many samples needed; each sample leads to 1 model evaluation
  - *Poor proposals:* If proposal distribution is sub-optimal, most proposals will be rejected
  - *Bad start:* What's a good place to start
- **Solutions:**
  - *Sampling cost:* Distribute sampling over *m* chains
  - *Poor proposals:* adaptive Metropolis-Hasting sampling
    - Periodically, use samples collected to compute a multivariate Gaussian approximation to f(: | :)
    - Inflate its variance and use it as a proposal
    - Only works if you have some samples to work with
  - *Bad start:* Have *m* chains start from an over-dispersed set of $\mathbf{p}_0$

SACHES: Addressing sampling cost



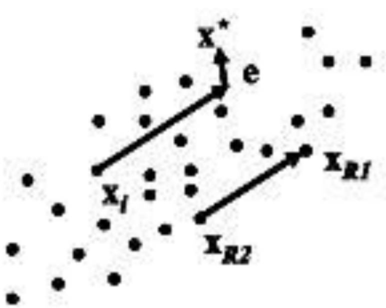
Generation i

Each generation consists of

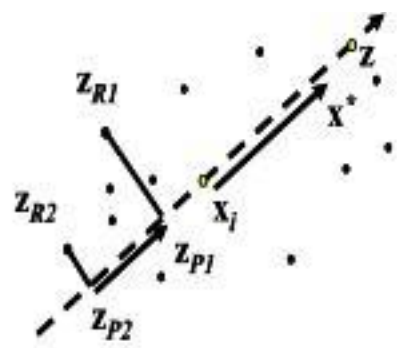
1. proposal generation
2. model run
3. accept/reject of proposal

SACHES: Addressing bad starts

- When there aren't enough samples, how to make a good proposal distribution?
  - Use genetic algorithm (Differential Evolution) to collect a few good samples
  - Use parallel and snooker updates to construct proposals



Parallel



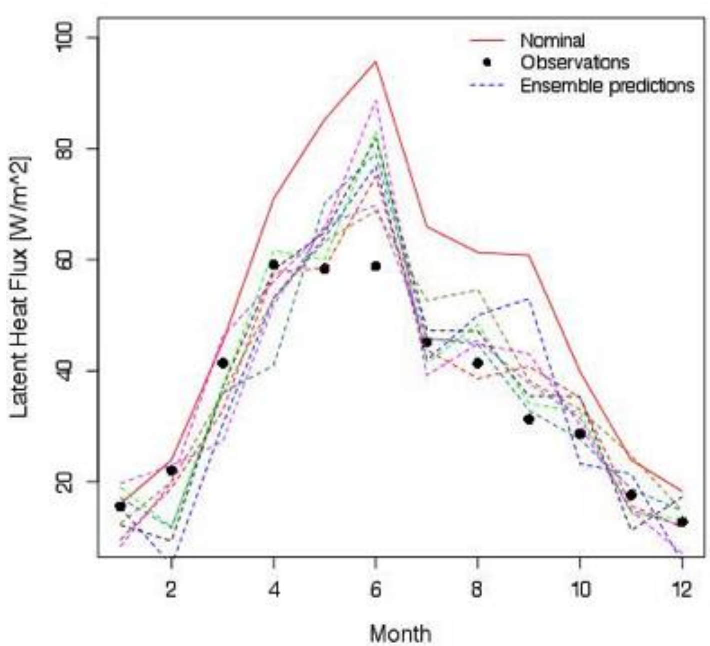
Snooker

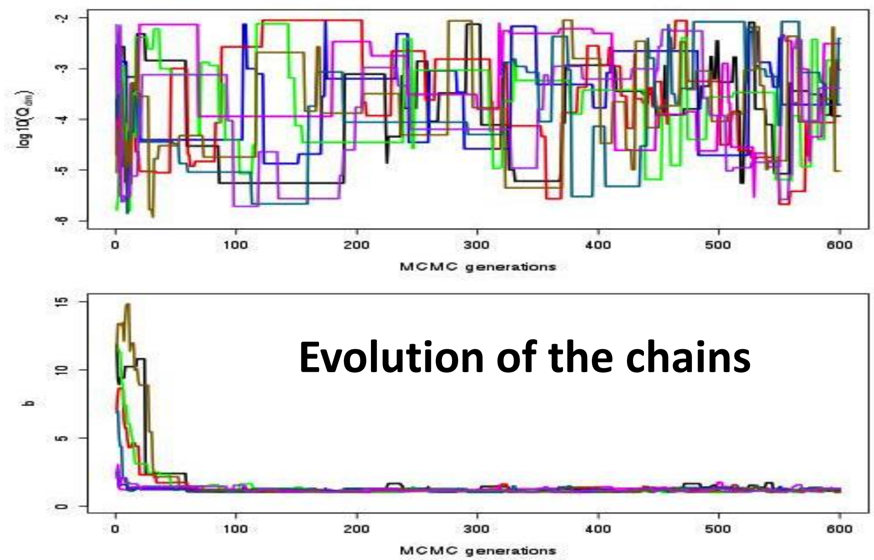
  - Switch to adaptive Metropolis-Hastings when we have a few good samples

SACHES: CLM calibration with real LH observations

- Calibrate: F_{drai}, $\log(Q_{dm})$, b
- Use observations from ARM/SGS site for 2003
  - Observations are latent heat fluxes
  - Averaged to their monthly value

Predictions using posteriors

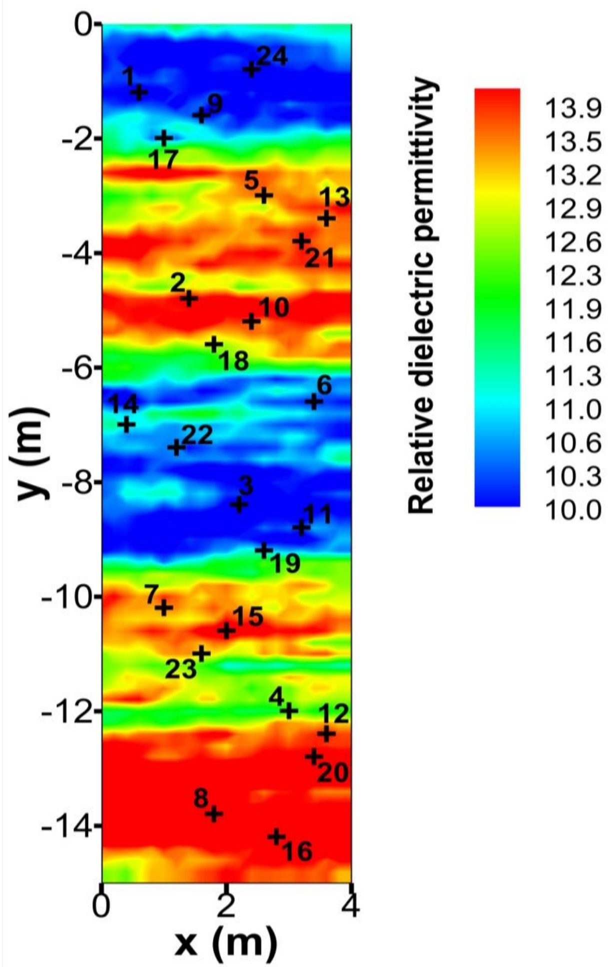




- The likelihood is flat near the minimum error point, hard to converge:
  - The chain for b has converged
  - The other chains are still wandering
  - Far from convergence @ 600 generations
- Even so, simulated LH is improved based on the posterior parameters

Ray et al., in preparation

SACHES: inverting soil moisture field using data from cross-hole ground penetrating radar



Bao et al., mathematical Geoscience, submitted;

- Monitoring soil moisture variations using tomographic ground penetrating radar (GPR) travel time data
- Tomographic GPR is a borehole-based geophysical imaging technique.
- It involves transmitting an electromagnetic (EM) pulse from a source in one borehole and recording the arrival of EM energy at a receiver position in a separate borehole.
- Inversion of the first arrival times of the EM energy is used to estimate the velocity and the dielectric permittivity (ϵ) distribution between the boreholes.
- Use of pilot points to model the dielectric permittivity field
- Challenges exist in the inversion of GPR tomographic data for handling non-uniqueness and high-dimensionality of unknowns.
- Explain pilot point random field model.
- Forward problem is linear y = travel time, x=permittivity field
- Reconstruction: estimating a random field, show random field realizations

Acknowledgement:

Department of Energy Office of Science

- BER: Earth System Modeling Program
- ASCR: Applied Mathematics Program