A Scalable Multi-chain Markov Chain Monte Carlo Method for Inverting Subsurface Hydraulic and Geological Properties

Jie Bao, Huiying Ren, Zhangshuan Hou, Maoyi Huang
Pacific Northwest National Laboratory

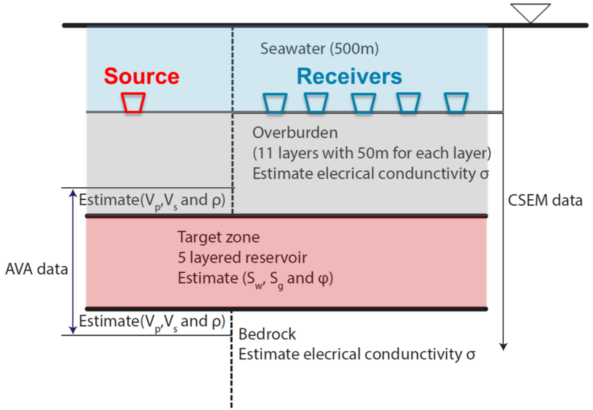
Jaideep Ray, Laura Swiler
Sandia National Laboratories

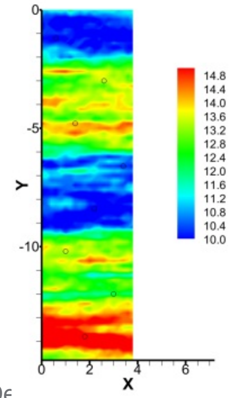
December 14, 2015

SAND2015-10841C

Outlines

- Scalable Multi-chain Markov Chain Monte Carlo Method
- Case 1: Reservoir porosity and saturation through invert marine seismic amplitude versus angle (AVA) and controlled-source electro-magnetic (CSEM) data





- Case 2: Soil moisture variations through ground penetrating radar (GPR) travel time data

Scalable Multi-chain Markov Chain Monte Carlo Method
Bayesian Formulation

- Generate posterior distributions on model parameters, given
  - Experimental data
  - A prior distribution on model parameters
  - A presumed probabilistic relationship between experimental data and model output that can be defined by a likelihood function

$$\pi(\theta \mid d) \propto \pi(\theta) L(d \mid \theta)$$

Posterior parameter distribution

Model parameters

Observed Data

Prior parameter distribution

Likelihood function which Incorporates the model

Scalable Multi-chain Markov Chain Monte Carlo Method
Bayesian Formulation

- Experimental data = Model output + error

$$d_i = G(\boldsymbol{\theta}_i) + \varepsilon_i$$

- If we assume error terms are independent, zero mean Gaussian random variables with variance σ^2, the likelihood is:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(d_i - G(\boldsymbol{\theta}_i))^2}{2\sigma^2} \right]$$

- Markov Chain Monte Carlo (MCMC)

  Generating a sampling density that is approximately equal to the posterior.

Markov Chain Monte Carlo

- MCMC generates samples that approximate the posterior distribution

- MCMC requires a "proposal density" which is used for generating θ_{i+1} in the sequence, conditional on θ_i.

- Metropolis-Hastings is a commonly used algorithm
  - Sample a candidate Y from the proposal density function $q(Y|\theta_i)$
  - Calculate the acceptance ratio $\alpha(\theta_i, Y) = \min\left[1, \frac{\pi(Y)q(Y \mid \theta_i)}{\pi(\theta_i)q(\theta_i \mid Y)}\right]$
  - If $\alpha(\theta_i, Y) > U$, set $\theta_{i+1} = Y$, else set $\theta_{i+1} = \theta_i$.
  - Increment i.

Markov Chain Monte Carlo

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- MCMC requires more than 10,000 evaluations of forward simulation model
- We want to avoid surrogates

COMPUTATIONALLY VERY EXPENSIVE

- Parallel MCMC

MCMC is inherently sequential

SaChES: Scalable Adaptive Chain-Ensemble Sampling

Scalable Multi-chain Markov Chain Monte Carlo Method
SaChES: Scalable Adaptive Chain-Ensemble Sampling

- Hybrid method that incorporates:
  - DREAM (DiffeRential Evolution Adaptive Metropolis) to utilize multiple chains to obtain high-quality proposal densities
  - DRAM (Delayed Rejection Adaptive Metropolis) to obtain posterior distributions efficiently
  - Parallel chains to accelerate computations

More details about the method is available on the poster

Bayesian calibration of the Community Land Model using a multi-chain Markov chain Monte Carlo method

Jaideep Ray, Laura Swiler, Maoyi Huang, Zhangshuan Hou

Thursday, 17 December, 13:40 – 18:00

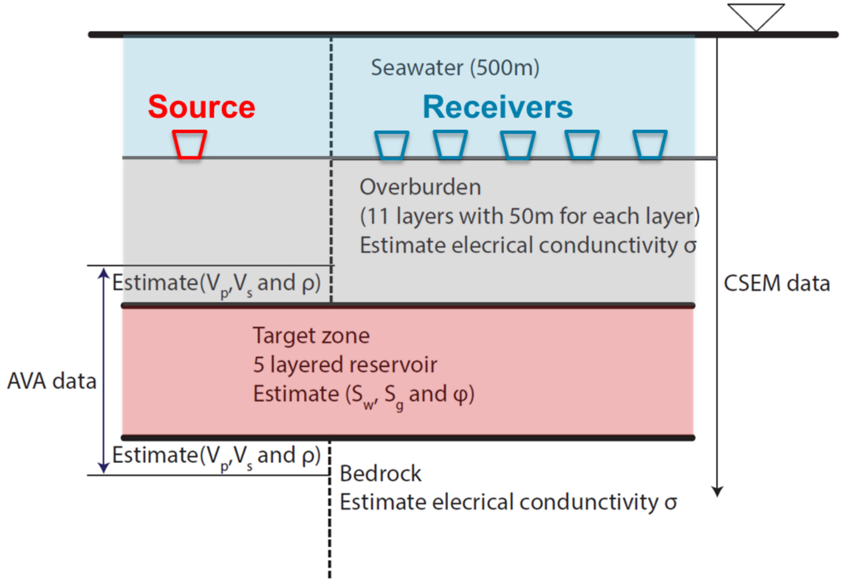
Moscone South – Poster Hall

Case 1: Gas Saturation Estimation

- Inversion domain

Seismic amplitude versus angle(AVA)

Controlled-source electro-magnetic(CSEM)



5-layered reservoirs from the upper to bottom with water saturations:

0.95, 0.05, 0.6, 0.9 and 0.1

and the porosity:

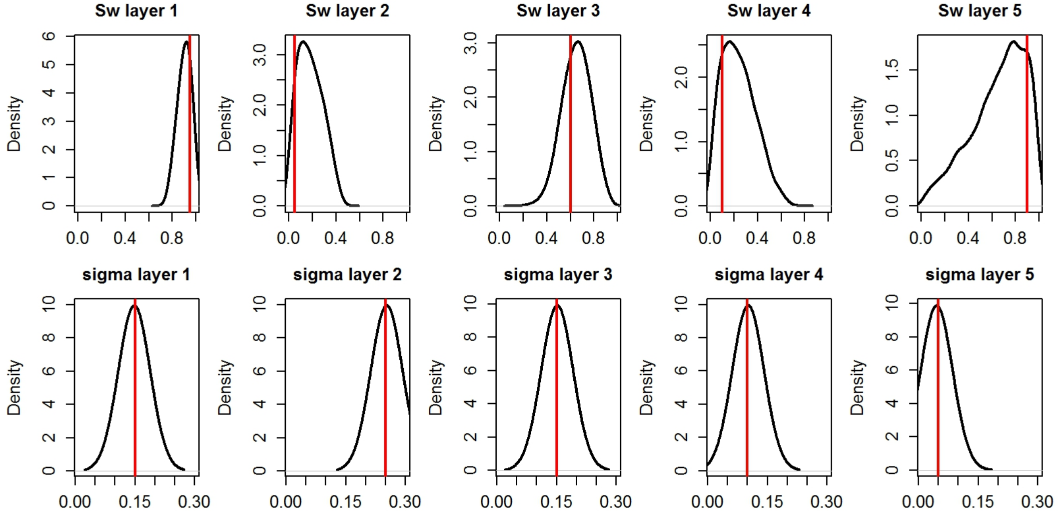
0.15, 0.25, 0.15, 0.1 and 0.05

The source and receivers were both located 50m above the seafloor. 21 receivers were away from electrodes from 500m to 5000m.

Case 1: Gas Saturation Estimation

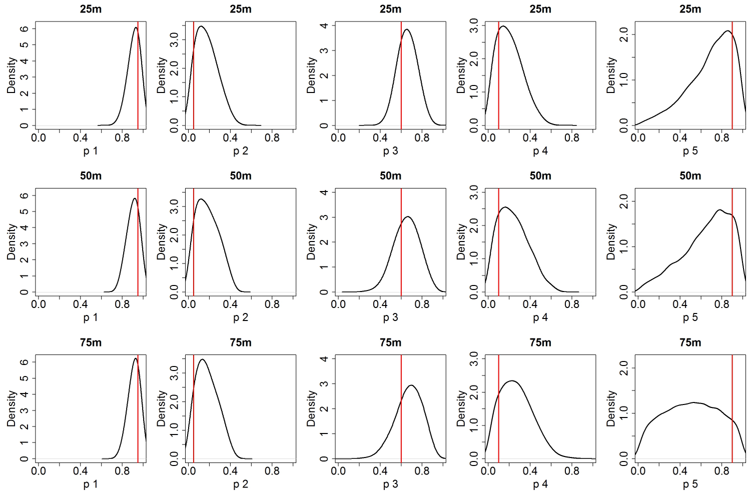
- Seismic AVA data (80 time steps) was used to estimate porosity and narrow bounds were obtained for each layer, then estimate water saturation.
- The reservoir thickness is 50m
- CSEM data were obtained from 2Hz channel

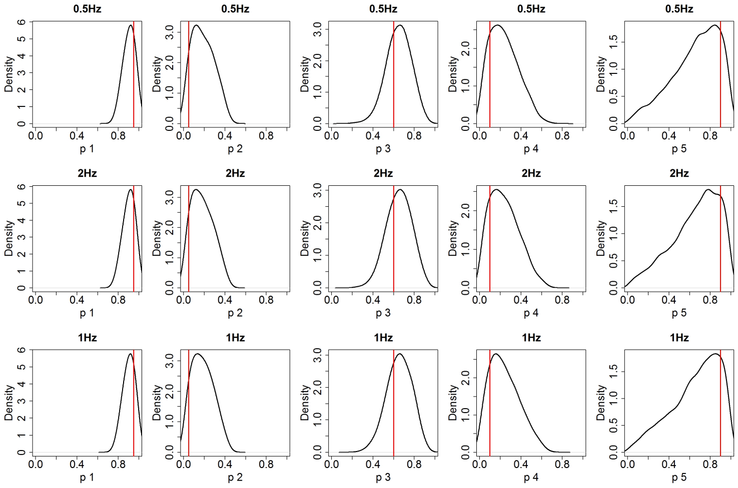
Posterior distribution of the parameters



Case 1: Gas Saturation Estimation

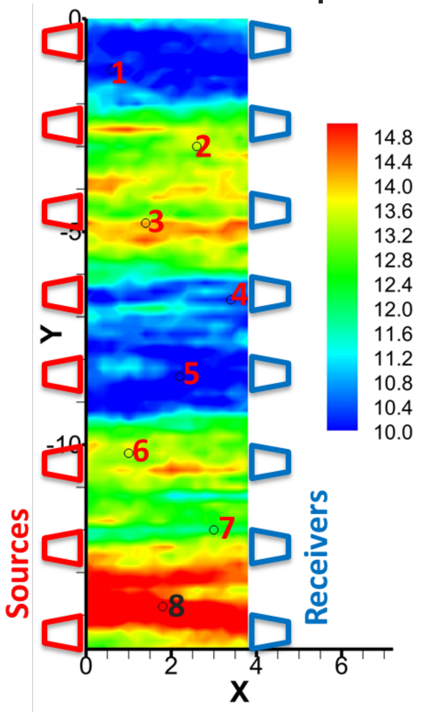
- The effect of thickness for each reservoir layer
  - Each layer thickness: 25m, 50m and 75m
- The effect of CSEM data frequency
  - CSEM data frequency: 0.5Hz, 1Hz and 2Hz





Case 2: Soil Moisture Variations

- "True" dielectric permittivity field



- Synthetic test case

  3.8 X 15 m; 20 X 75 points

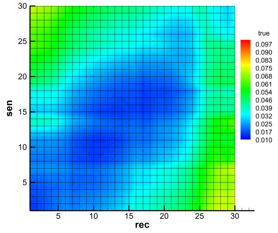
  8 pilot points, and range (correlation length)

  Generate a random dielectric field in SGSIM

- Ground penetrating radar (GPR) travel time simulation

  Velocity: $v = v_l / \sqrt{\varepsilon}$

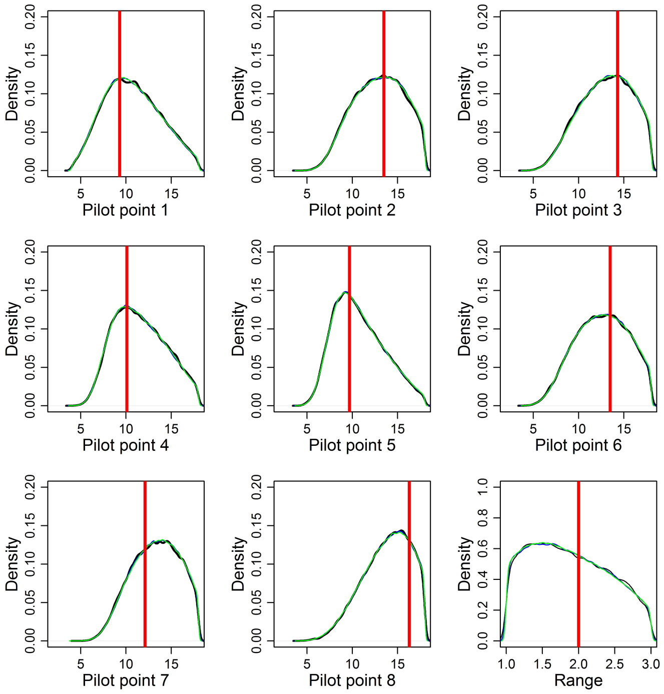
  Calculate the radar signal travel time between each source and receiver



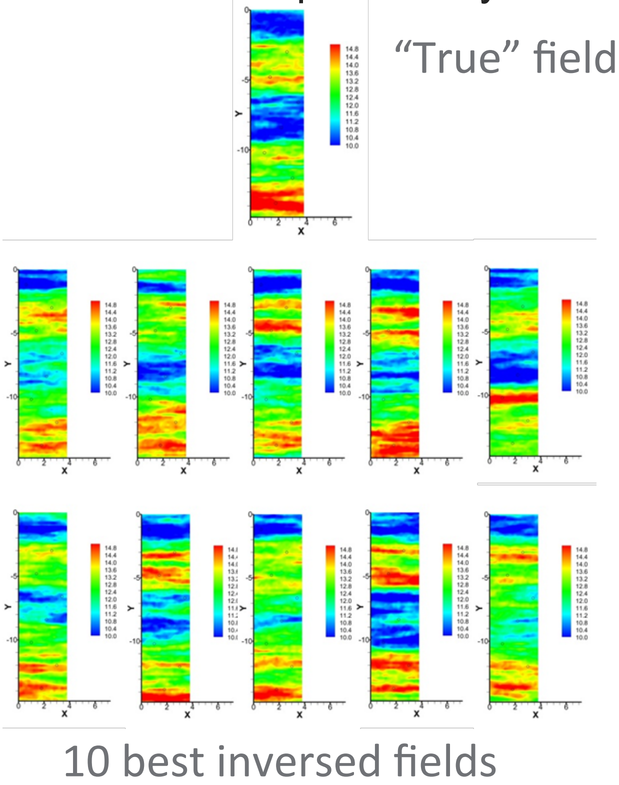
"Observations"
Travel time between 30 sources and 30 receivers

Case 2: Soil Moisture Variations

- Posterior distribution of the parameters



- Inversed dielectric permittivity field

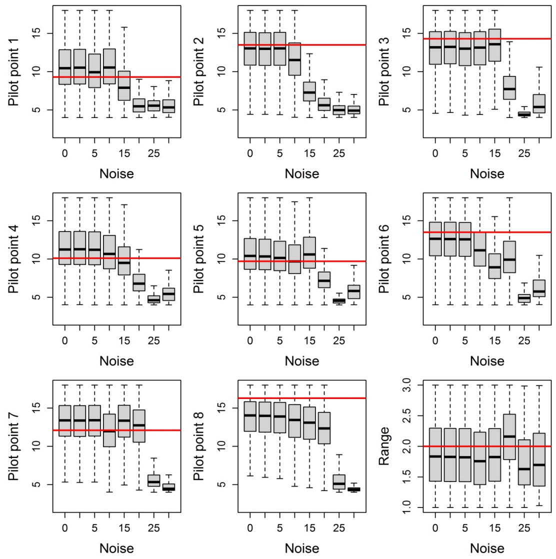


Case 2: Soil Moisture Variations

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

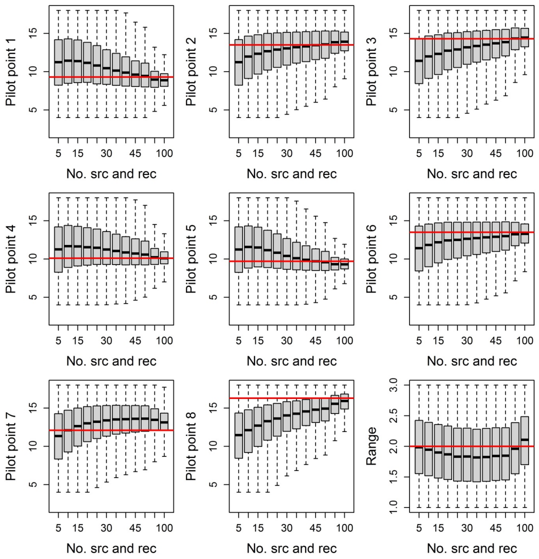
Noise on observation

Noise's standard deviation is defined as the percentage of the mean of the true observation



Number of sources and receivers

5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, and 100 sources and receivers



Questions?