Bayesian data assimilation for flows in heterogeneous media

Yalchin Efendiev

Texas A&M University

Collaborators: A. Datta-Gupta, R. Jana, B. Mallick, B. Mohanty, A. Mondal, J. Wei, J. Xie
Subsurface characterization is to identify subsurface properties taking into account various data sources and their precisions and predict future reservoir performance.
Soil Moisture / Brightness Temperature Measurement Platforms and Scales

Space-borne

Air-borne

Ground-based

In situ

by B. Mohanty
Bayesian Approach: An Illustration

\[ P(I | O) \propto P(O | I)P(I) \]

- \( I \) - permeability
- \( O \) - data
Problem Statement and Challenges

• Good priors are needed

• Static Models Must be Updated with Dynamic Response from the Reservoir

• Integration of Dynamic Data is Difficult and Time Consuming
  – Inverse Problem and Ill-posed
  – Expensive Forward Simulations
Prior modeling

- Goal: to keep the parameter space dimension as small as possible and data relevant.
- SVD-based parameterization for fields described by two-point correlation (expensive?)
- For non-Gaussian fields (channelized permeability fields), level set approaches combined with SVD-based techniques.
- The data are largely affected by the distinct geologic facies with sharp contrasts in properties across facies boundaries (Koltermann and S. M. Gorelick WRR 1996,....)
Karhunen-Loeve Expansion

- **Decompose covariance function**
  \[ C(x, y)\phi_i(y) = \lambda_i\phi_i(x), \text{ where } \phi_i(x) \text{ are eigenvectors and } \lambda_i \text{ are eigenvalues.} \]

- **Expand permeability field in terms of eigenvectors**
  \[ Y(x) = \sum_{i=1}^{\infty} \theta_i \sqrt{\lambda_i} \phi_i(x), \text{ where } \lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n \]

- **Truncate expansion using dominant eigenvalues**
  \[ Y(x) = \sum_{i=0}^{N} \theta_i \sqrt{\lambda_i} \phi_i(x), \text{ k(x)=exp(Y(x))} \]
  \[ \text{where } \theta_i \sim \text{N}(0, \sigma_i^2) \]

Permeability field can be represented by a small number of random parameters while the underlying two-point geostatistical structure is still maintained. It can be conditioned at well locations.
Multiple Gaussian fields

- In soil moisture applications, the soil content at each pixel is described by the percentage of sand, silt, and clay.
- Their percentages are used to calculate soil properties
- In this case, SVD-based approaches can be applied to correlated Gaussian fields.
Channelized fields. Level-set deformations

• Clustering (Aharon et al., 2005, …)

\[ \frac{\partial \varphi(t,x)}{\partial t} + \nu(x)|\nabla \varphi(t,x)| = 0 \text{ OR } \frac{\partial \varphi(t,x)}{\partial t} + \nu(x) \cdot \nabla \varphi(t,x) = 0 \]

\[ \varphi(t,x) = 0 \] represents the facies boundaries

• Parameterization of facies is now that of the velocity field.

• Smooth interfaces – smooth velocity fields
Prior modeling. Level-set deformations. Cont’d.

- Additive properties of the flow map needs to be preserved.
- We typically use “fixed streamlines” and random velocity magnitudes

\[ \frac{\partial \varphi(t, x)}{\partial t} + v(x) \cdot \nabla \varphi(t, x) = 0 \text{ OR } \frac{\partial \varphi(t, x)}{\partial t} + v(x) \cdot \nabla \varphi(t, x) = 0 \]

\[ \varphi(t, x) = 0 \text{ represents the facies boundaries} \]
Workflow – Level Set

Signed Distance Function

Updated Signed Distance

Velocity Field
Level Set Method

- The level set method was developed in the 1980s by Osher and Sethian\(^{(1)}\)

- Numerical algorithm for dynamic implicit surfaces and their evolution

The implicit surface is defined as function \(\varphi(t, x)\)

\[
\Gamma = \{x \in \Omega | \varphi(t, x) = 0\}
\]

\[
\Omega^+ = \{x \in \Omega \text{ for } \varphi(t, x) > 0\}
\]

\[
\Omega^- = \{x \in \Omega \text{ for } \varphi(t, x) < 0\}
\]

\[
\Omega = \Omega^+ \cup \Omega^- \text{ and } \Omega^+ \cap \Omega^- = 0
\]
Implicit Surface Function

- Application in facies modeling (Moreno et al., 2008)
  - Zero level set represents facies boundaries
- Additionally, $2^n$ facies can be characterized by $n$ level set functions.
- By solving level set equation, facies boundaries – level set – will evolve into a different shape with time
- Level set equation: Motion in the normal direction
  $$\frac{\partial \phi(t, x)}{\partial t} + \nu(x) |\nabla \phi(t, x)| = 0$$
Continuous Facies Boundaries

- **Signed Distance Function (SDF)**

\[
\varphi(t, \mathbf{x}) = \begin{cases} 
\text{distance}(\mathbf{x}, \Gamma) & \forall \mathbf{x} \in \Omega^+ \\
-\text{distance}(\mathbf{x}, \Gamma) & \forall \mathbf{x} \in \Omega^- 
\end{cases}
\]

- **The Re-initialization Equation**

- Convert an implicit surface into a signed distance function

\[
\frac{\partial \varphi(t, \mathbf{x})}{\partial t} + \text{sgn}(\varphi(0, \mathbf{x}))(\| \nabla \varphi(t, \mathbf{x}) \|- 1) = 0
\]

where, \( \text{sgn}(\varphi(0, \mathbf{x})) = \begin{cases} 
-1 & \mathbf{x} \in \Omega^- \\
1 & \mathbf{x} \in \Omega^+ 
\end{cases} \)

SDF has the nice property that \( |\nabla \varphi(t, \mathbf{x})| = 1 \)
Prior modeling. Level-set deformations, cont’d

- The dimension reduction in the velocity space can be achieved by restricting it to a region.
- Given velocity samples (e.g., realizations of Gaussian fields or elements of the cluster), we generate (using SVD) a best basis with a given weight function \( m(x) \)

\[
C(x, y) = E(m(x)m(y)v(x, \omega)v(y, \omega))
\]

- We assume variable dimensional parameter space.
Permeability within facies

- Grid-connectivity based parameterization or covariance based parameterization
- Karhunen-Loeve parameterization.
Likelihood setup

• **Forward model is nonlinearly coupled flow and transport equations**

\[
E_k^p (p) = q, \quad E_k^p (p) = -\nabla \cdot \sum_{j=1}^{n_p} \lambda_{ij} k(x)(\nabla p - \rho_j g \nabla x_3)
\]

\[
E_k^S (S) = q
\]

• **Data:** well data (water-cut, oil-cut, GOR),
time-lapse seismic data (uses spatial \( S(x) \));
soil moisture data,…

• **Likelihood**

\[
L(p, S) = \exp \left( -(D^w(p, S) - D^w_{obs})C_{D^w}^{-1}(D^w(p, S) - D^w_{obs}) - (D^I(p, S) - D^I_{obs})C_{D^I}^{-1}(D^I(p, S) - D^I_{obs}) \right).
\]
Likelihood regularity

• How does the likelihood depend on the input parameters (or how regular is the posterior measure w.r.t. the prior)?

• Infinite dimensional vs. finite dimensional (continuous fields, absolute continuity,…, e.g., A. Stuart, 2010)

• Finite dimensional permeability statistics and infinite dimensional PDE analysis

• Exact and approximate posterior measure

\[
\pi(\theta_M, \tau) \propto L(\theta_M, \tau) \pi_0(\tau) \prod_{i=1}^{s} \pi_0(\theta_{i1}, \ldots, \theta_{iM_i}) \pi_0(\theta_{i, M_i+1}, \ldots, \theta_{iN_i}),
\]

\[
\pi(\theta_N, \tau) \propto L(\theta_N, \tau) \pi_0(\tau) \prod_{i=1}^{s} \pi_0(\theta_{i1}, \ldots, \theta_{iM_i}) \pi_0(\theta_{i, M_i+1}, \ldots, \theta_{iN_i})
\]

• Estimate

\[
| \int f(\theta_N, \tau) \pi(\theta_N, \tau) d(\theta_N, \tau) - \int f(\theta_N, \tau) \pi(\theta_M, \tau) d(\theta_N, \tau) | \leq \text{related to input}
\]

• We show that constants are independent of the dimension.
Posterior

• Prior for Reservoir Model

\[ P(k, F) \]

• Likelihood Function

\[
p(d_{obs} \mid k) = c_2 \exp\left\{-\frac{1}{2} \left[ (g(k) - d_{obs})^T C_D^{-1} (g(k) - d_{obs}) \right]\right\}
\]

• Posterior PDF for Reservoir Model

\[
p(k, F \mid d_{obs}) = c \exp\left\{-\frac{1}{2} \left[ (g(k) - d_{obs})^T C_D^{-1} (g(k) - d_{obs}) \right]\right\} P(k \mid F) P(F)
\]
Sampling the posterior

- **Posterior Distribution**
  - Very high-dimensional
  - Non-Gaussian and Multimodal
  - Unknown Normalizing constant

- **Requires large number of simulation runs**

- **Very expensive for high resolution models**
MCMC: Metropolis-Hastings Algorithm

\[
\alpha = \min \left[ 1, \frac{\pi(k^*) q(k^* | k^*)}{\pi(k^i) q(k^* | k^i)} \right]
\]

Accepted \( k^* \) with probability \( \alpha \)

This algorithm generates Markov chain with the steady state distribution \( \pi(k) \)

\[
\pi(k) = C \exp \left[ -\frac{1}{2} \left( g(k) - d_{obs} \right) C_D^{-1} \left( g(k) - d_{obs} \right) \right] P(k, F)
\]

Requires flow simulation run

There is a burn-in time after which the accepted proposals are collected
Key Features of MCMC

• **Samples from the posterior rigorously**

• **Problems:**
  - Low acceptance rates for proposals
  - Computationally expensive (requires reservoir simulation run for each proposed model)
Uncertainty Quantification
Multistage preconditioned MCMC

• **Key features of the multistage MCMC**
  – Rigorous sampling using MCMC; high acceptance rate; reduces computation time (Christen and Fox 2005, Efendiev et al. 2006)

• **Approach**
  – Use of approximate likelihood to filter obvious rejections
  – Minimizes the number of simulation runs
Approximate models

- Solutions on a coarse grid.
- Upscaling/multiscale methods. The main idea is to solve flow and transport equations on a coarse grid.
- Ensemble level multiscale methods (Chen and Durlofsky 2004, Aarnes and Efendiev, 2008). Coarse-scale parameters are computed using only a few realizations (sparse grids…) and interpolated for any other one.
Multistage MCMC

• **Posterior**
  \[ \pi(k) \propto \exp(-L(k))\pi_0(k) \]

- **Approximate posterior**
  \[ \hat{\pi}(k) \propto \exp(-\hat{L}(k))\pi_0(k) \]

- **Error models**
  
  \[ L(k) \text{ vs. } \hat{L}(k) \text{ cross-plot for a number of initial runs and model the error} \]

• **First stage**
  
  – Make a proposal (with “coarse” Langevin \( q(x|y) \)) and accept/reject based on approximate model response and accept with probability

  \[ \alpha = \min(1, \frac{\hat{\pi}(k^*)q(k_n | k^*)}{\hat{\pi}(k_n)q(k^* | k_n)}) \]

  If accepted, proceed to the 2\text{nd} stage; otherwise start with a new proposal
Multistage MCMC cont’

- **Second stage**
  - Run simulator \( g(k^*) \)
  - Calculate likelihood \( \pi(k^*) \) (up to a constant)
  - Check acceptance with modified probability \( \alpha \)

\[
\alpha = \min \left( 1, \frac{\hat{\pi}(k_n)\pi(k^*)}{\hat{\pi}(k^*)\pi(k_n)} \right)
\]

- Reversible jump MCMC with variable parameter dimension (Mondal et al., 2009)
Workflow

1. Initialize Parameters ($\theta$)
2. Perturb $\theta$ to Generate Proposal
3. Calculate Approx Likelihood
4. Accepted?
   - Yes: Promote to Proposal
   - No: Calculate Exact Likelihood
5. Calculate Exact Likelihood
6. Accepted?
   - Yes: Collect Sample
   - No: Perturb $\theta$ to Generate Proposal
7. Enough samples?
   - Yes: Collect Sample
   - No: Repeat from Step 2
A Synthetic Example

(a) Reference Model

(b) Initial Model
History Matching Water-Cut
Multistage MCMC: Synthetic Case

Acceptance rate

RMS reduction

60%
20%
700
2000
Collected Samples

Selected samples
2-D Synthetic Channelized Example

- Grid size: 50*50
- Waterflood example
- Two phase flow (oil & water)
- 3 injectors: pressure control
- 4 producers: oil rate control
- History matching:
  - Oil rate
  - Water-cut
  - Producer BHP
  - Water injection rate
MCMC Result

Data Misfit vs Step
3-D Synthetic Example

- Grid size: 50*50*6
- Waterflood example
- Two phase flow (oil & water)
- 3 injectors: pressure control
- 3 producers: oil rate control
- History matching
  - Oil rate
  - Water-cut
  - Producer BHP
  - Water injection rate
MCMC Result

![MCMC Result](image-url)
# Permeability Comparison

<table>
<thead>
<tr>
<th>Sand/Shale</th>
<th>4 UPDATED MODELS</th>
<th>TRUE</th>
<th>INITIAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6737</td>
<td>0.6985</td>
<td>0.7150</td>
<td>0.6635</td>
</tr>
</tbody>
</table>

Layer 2

Layer 4

Layer 6
Dynamic Data Match
- Observe - Initial - Update

Oil Rate

Prod 1

Prod 2

Prod 3
Conclusions

• Prior modeling
• Likelihood setup and regularity
• Efficient posterior sampling (MCMC, EnKF,...)
• Applications