



The University of Tulsa
Petroleum Reservoir Exploitation Projects

ES-MDA, Data Assimilation and Uncertainty Quantification with MCMC

Al Reynolds, Xin Li and Javad Rafiee

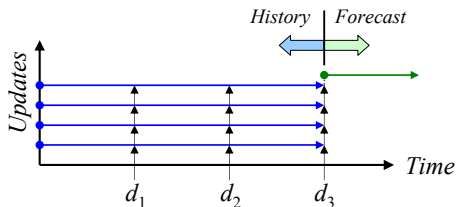
Assisted History Matching
IPAM, March 23, 2017

- Ensemble Smoother with Multiple Data Assimilation (ES-MDA).
- Discrepancy principle and choice of inflation factors in ES-MDA
- Pre-RML with MCMC for uncertainty quantification (UQ)
- Concluding comments

Ensemble Smoother With Multiple Data Assimilation

- ES-MDA was proposed by Emerick and Reynolds in 2011. Motivated by analogy of ES/EnKF with a single iteration of Gauss-Newton iteration with a full step (Reynolds et al., 2005, 2006) with the same average sensitivity matrix used for the update of all ensemble members and the need to provide damping at early iterations.
- Parameter estimation/simulation problem. Avoids statistical inconsistencies between updated models parameters and states that can occur with EnKF.
- Truly black box and completely parallelizable.

- Parameter estimation method.



- Update equation:

$$m_j^{a,i} = m_j^{f,i} + \tilde{C}_{MD}^{f,i} \left(\tilde{C}_{DD}^{f,i} + \alpha_i C_D \right)^{-1} \left(d_{uc,j}^i - d_j^{f,i} \right).$$

ES-MDA Algorithm

- 1 Choose the number of data assimilations, N_a , and the coefficients, α_i for $i = 1, \dots, N_a$.
- 2 Generate initial ensemble $\{m_j^{f,1}\}_{j=1}^{N_e}$
- 3 For $i = 1, \dots, N_a$:
 - (a) Run the ensemble from time zero,
 - (b) For each ensemble member, perturb the observation vector with the inflated measurement error covariance matrix, i.e., $d_{uc,j}^i \sim \mathcal{N}(d_{obs}, \alpha_i C_D)$.
 - (c) Use the update equation to update the ensemble.

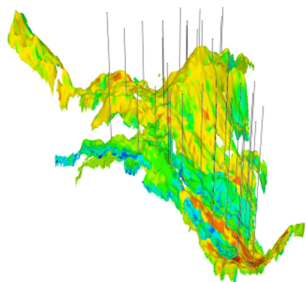
$$m_j^{a,i} = m_j^{f,i} + \Delta M^{f,i} (\Delta D^{f,i})^T \left[\Delta D^{f,i} (\Delta D^{f,i})^T + \alpha_i C_D \right]^{-1} \left(d_{uc,j}^i - d_j^{f,i} \right)$$
$$m_j^{f,i+1} = m_j^{a,i}$$

- Comment: Requires $\sum_{k=1}^{N_a} \frac{1}{\alpha_k} = 1$.

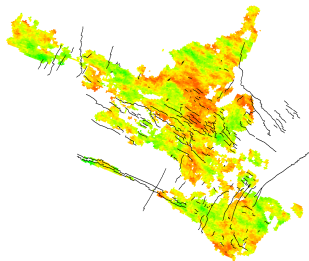
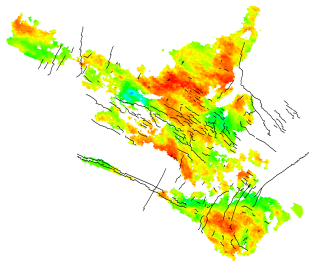
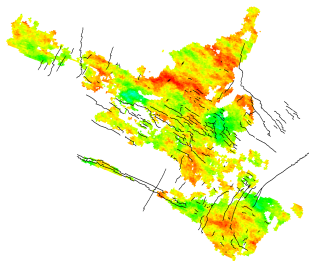
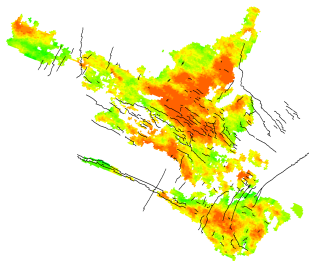
ASSISTED HISTORY MATCHING - GOALS

- 1 Honor Observations.
- 2 Maintain geological realism (or need to resolve geo-modeling issues).
- 3 Have predictive power not just in terms of matching future field production but in terms of fluid distributions.
- 4 Give a least some reasonable estimate of uncertainties for reservoir development and to manage risk - ideally characterize the posterior pdf.
- 5 Be compatible with company/institution simulation tools.
- 6 Be conceptually understandable to the user and management given a reasonable amount of training.
- 7 Be computationally feasible.

- **Observed data:**
 - 20 producers: oil rate, water rate, GOR, bottom-hole pressure.
 - 10 water injection: bottom-hole pressure.
- **Initial ensemble:**
 - 200 models.
 - Porosity and permeability (> 125,000 active gridblocks).
 - Anisotropic ratio k_v/k_h .
 - Rock compressibility.
 - End point of water relative permeability curve.
- Data assimilation with **ES-MDA (4×)** with **localization**.

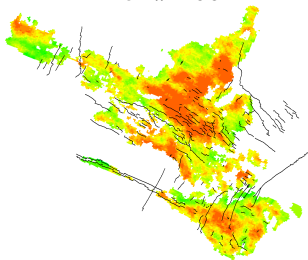


Field Case 2: Model Plausibility – Permeability

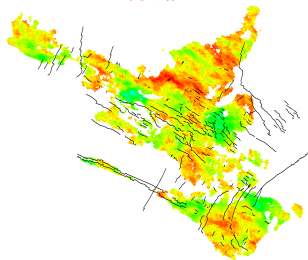


Field Case 2: Model Plausibility – Permeability

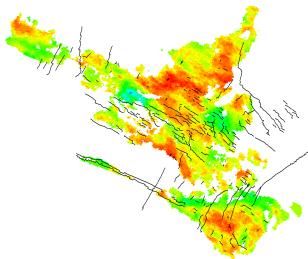
Prior # 200



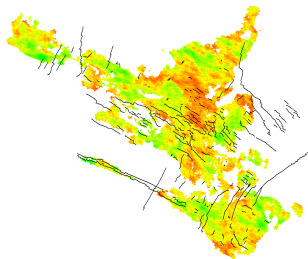
Post # 1



Post # 200

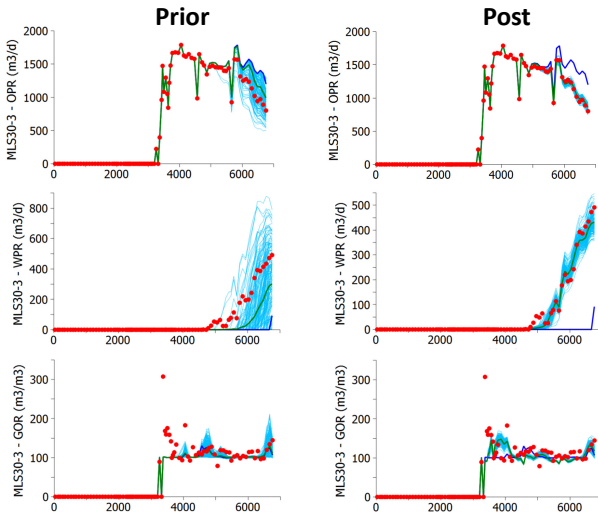


Prior # 1



Field Case 2: Well Data

Well # 30



Dimensionless Sensitivity

- The dimensionless sensitivities control the change in model parameters that occurs when assimilating data (Zhang et al., 2003; Tavakoli and Reynolds, 2010). The standard dimensionless sensitivity is defined as

$$\widehat{G}_D^i \equiv C_D^{-1/2} G(\bar{m}^i) C_M^{1/2}, \quad (1)$$

where $G(m)$ is the sensitivity matrix for $d^f(m)$ where

$$\widehat{g}_{i,j} = \frac{\partial d_i^f(m)}{\partial m_j}. \quad (2)$$

- Dimensionless sensitivity matrix components are

$$g_{i,j} = \frac{\sigma_{m,j}}{\sigma_{d,i}} \frac{\partial d_i^f}{\partial m_j}. \quad (3)$$

- The direct analogue of the standard dimensionless sensitivity matrix in ensemble based methods is given by

$$G_D^i \equiv C_D^{-1/2} \Delta D^{f,i} \approx C_D^{-1/2} G(\bar{m}^i) \Delta M^i. \quad (4)$$

Recall the ES-MDA update equation

$$m_j^{a,i} = m_j^{f,i} + \Delta M^{f,i} (\Delta D^{f,i})^T \left[\Delta D^{f,i} (\Delta D^{f,i})^T + \alpha_i C_D \right]^{-1} \left(d_{uc,j}^i - d_j^{f,i} \right)$$

Using the definition of the dimensionless sensitivity

($G_D^i \equiv C_D^{-1/2} \Delta D^i$), we can write ES-MDA update equation as

$$m_j^{a,i} = m_j^{f,i} + \Delta M^{f,i} (G_D^i)^T \left[G_D^i (G_D^i)^T + \alpha_i I_{N_d} \right]^{-1} C_D^{-1/2} \left(d_{uc,j}^i - d_j^{f,i} \right),$$

for $j = 1, \dots, N_e$.

Least Squares Problem

Similarly, one can update the mean of m directly as

$$\bar{m}^{a,i} = \bar{m}^{f,i} + \Delta M^{f,i} (G_D^i)^T \left[G_D^i (G_D^i)^T + \alpha_i I_{N_d} \right]^{-1} C_D^{-1/2} \left(d_{\text{obs}} - d^f(\bar{m}^{f,i}) \right).$$

It is easy to show that $\bar{m}^{a,i}$ is the solution of the regularized least squares problem given by

$$x^{a,i} = \arg \min_x \left\{ \frac{1}{2} \|G_D^i x - y\|^2 + \frac{\alpha_i}{2} \|x\|^2 \right\},$$

where

$$x = (\Delta M^{f,i})^+ (m - \bar{m}^{f,i}), \quad (5)$$

$$y = C_D^{-1/2} \left(d_{\text{obs}} - d^f(\bar{m}^{f,i}) \right), \quad (6)$$

where $(\Delta M^{f,i})^+$ is the pseudo-inverse of $\Delta M^{f,i}$.

Assume that

$$\|y\| = \|C_D^{-1/2} (d_{\text{obs}} - \bar{d}^{f,i})\| > \eta, \quad (7)$$

where η is the noise level

$$(\eta^2 = \|C_D^{-1/2} (d_{\text{obs}} - d^f(m_{\text{true}}))\|^2 \approx N_d).$$

Based on the **discrepancy principle** the minimum regularization parameter, α_i , should be selected such that

$$\eta = \|G_D^i x^{\alpha,i} - y\|. \quad (8)$$

Discrepancy Principle

From Eqs. 7 and 8 we can write

$$\|C_D^{-1/2} (d_{\text{obs}} - \bar{d}^{f,i})\| > \eta = \alpha_i \left\| [G_D^i (G_D^i)^T + \alpha_i I_{N_d}]^{-1} C_D^{-1/2} (d_{\text{obs}} - \bar{d}^{f,i}) \right\|. \quad (9)$$

Therefore, for some $\rho \in (0, 1)$

$$\rho \|C_D^{-1/2} (d_{\text{obs}} - \bar{d}^{f,i})\| = \alpha_i \left\| [G_D^i (G_D^i)^T + \alpha_i I_{N_d}]^{-1} C_D^{-1/2} (d_{\text{obs}} - \bar{d}^{f,i}) \right\|. \quad (10)$$

Hanke (1997) proposed the following condition for the regularization parameter in Levenberg-Marquardt (LM) algorithm

$$\rho^2 \left\| C_D^{-1/2} (d_{\text{obs}} - \bar{d}^{f,i}) \right\|^2 \leq \alpha_i^2 \left\| [G_D^i (G_D^i)^T + \alpha_i I_{N_d}]^{-1} C_D^{-1/2} (d_{\text{obs}} - \bar{d}^{f,i}) \right\|^2. \quad (11)$$

Iglesias (2015) used Eq. 11 for choosing inflation factors in his version of ES-MDA (IR-ES).

Le et al. (2015) used a much stricter condition based on Eq. 11 for choosing inflation factors in ES-MDA-RLM.

An Analytical Procedure for Calculation of Inflation Factors

Recall that

$$\rho^2 \left\| C_D^{-1/2} (d_{\text{obs}} - \bar{d}^{f,i}) \right\|^2 \leq \alpha_i^2 \left\| [G_D^i (G_D^i)^T + \alpha_i I_{N_d}]^{-1} C_D^{-1/2} (d_{\text{obs}} - \bar{d}^{f,i}) \right\|^2. \quad (11)$$

Using the definitions of $y = C_D^{-1/2} (d_{\text{obs}} - \bar{d}^{f,i})$ and $C \equiv G_D^i (G_D^i)^T + \alpha_i I_{N_d}$,

$$\rho^2 \leq \alpha_i^2 \frac{\|C^{-1}y\|^2}{\|y\|^2}.$$

$$\rho^2 \leq \alpha_i^2 \max_j \gamma_j^2 = \alpha_i^2 \max_j \frac{1}{(\lambda_j^2 + \alpha_i)^2}$$

where γ_j 's are the eigenvalues of C^{-1} and λ_j 's are the singular values of G_D^i .

Instead of enforcing

$$\rho^2 \leq \alpha_i^2 \max_j \frac{1}{(\lambda_j^2 + \alpha_i)^2},$$

we use

$$\rho^2 \leq \alpha_i^2 \frac{1}{(\bar{\lambda}^2 + \alpha_i)^2},$$

or

$$\alpha_i = \frac{\rho}{1 - \rho} \bar{\lambda}^2$$

where $\bar{\lambda}$ is the average singular value of the dimensionless sensitivity matrix, G_D^i , and is given by

$$\bar{\lambda} = \frac{1}{N} \sum_{j=1}^N \lambda_j.$$

We use $\rho = 0.5$, so $\alpha_i = \bar{\lambda}^2$.

- Specify the number of data assimilation steps (N_a).
- We assume that the inflation factors form a monotonically decreasing geometric sequence

$$\alpha_{i+1} = \beta^i \alpha_1,$$

- Determine

$$\alpha_1 = \bar{\lambda}^2 = \left(\frac{1}{N} \sum_{j=1}^N \lambda_j \right)^2.$$

- Recall that ES-MDA requires that

$$1 = \sum_{i=1}^{N_a} \frac{1}{\alpha_i} = \sum_{i=1}^{N_a} \frac{1}{\beta^{i-1} \alpha_1}$$

- Solve

$$\frac{1 - (1/\beta)^{N_a-1}}{1 - (1/\beta)} = \alpha_1,$$

for β .

- We call our proposed practical method ES-MDA-GEO.

Comments on “Convergence” of ES-MDA

- Classifying ES-MDA as an iterative ES may be augmentable; stops when $\sum_{k=1}^{N_a} \frac{1}{\alpha_k} = 1$.
- Criterion based on ensuring methods samples correctly in the linear Gaussian case as ensemble size goes to infinity.
- Analogue of Hanke’s suggestion for RLM, should terminate ES-MDA when

$$\frac{1}{N_d} \|C_D^{-1/2} (d_{\text{obs}} - \bar{d}^{f,i})\|^2 < \tau^2$$

where $\tau > 1/\rho = 2$.

- This means, terminate when the normalized objective function is less than 4.

- The performance of ES-MDA-GEO is compared to IR-ES, ES-MDA-RLM, ES-MDA-EQL, and M-IR-ES for two synthetic examples.
- To investigate the performance of the methods, we define the following measures:

$$\text{RMSE} = \frac{1}{N_e} \sum_{j=1}^{N_e} \left(\frac{1}{N_m} \sum_{k=1}^{N_m} (m_{\text{true},k} - m_{j,k})^2 \right)^{1/2}, \quad (12)$$

$$\bar{\sigma} = \frac{1}{N_m} \sum_{k=1}^{N_m} \sigma_k, \quad (13)$$

$$O_{Nd} = \frac{1}{N_e N_d} \sum_{j=1}^{N_e} (d_j^f - d_{\text{obs}})^T C_D^{-1} (d_j^f - d_{\text{obs}}). \quad (14)$$

Example 1: 2D Waterflooding

Two-dimensional waterflooding problem:

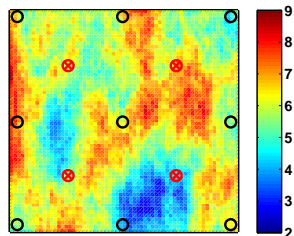
- $64 \times 64 \times 1$ grid.
- 9 production wells (BHP control).
- 4 injection wells (BHP control).

Observed data:

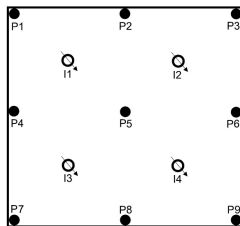
- Oil and water production rates and water injection rates.
- Standard deviations of measurement error: 3% of true data.
- Data from the first 36 months are history-matched and data for next 20 are used for prediction.

Model parameters:

- The gridblock log-permeabilities are considered as the model parameters.



True permeability field



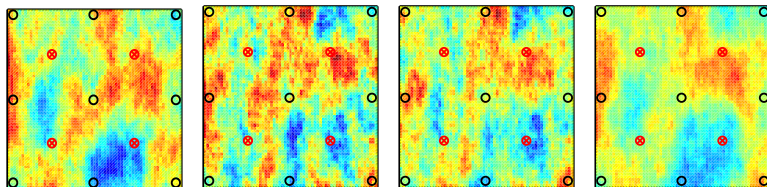
Well locations

Example 1: Results

- An ensemble of 300 realizations is generated from the prior distribution.
- First inflation factor from DP is 1049.4; N_a of 4 and 6, respectively, give β equal to 0.102 and 0.264.
- Comment IR-ES with $\rho = 0.8$ did not converge after 200 iterations.

	Prior	ES-MDA-RLM	IR-ES	M-IR-ES	ES-MDA-EQL		ES-MDA-GEO	
		$\rho=0.5$	$\rho=0.5$	$\rho=0.5$	$N_a=4$	$N_a=6$	$N_a=4$	$N_a=6$
RMSE	2.23	0.613	0.902	0.680	1.45	1.09	0.586	0.633
$\bar{\sigma}$	0.995	0.334	0.517	0.363	0.258	0.255	0.380	0.362
O_{Nd}	16121	1.06	8.14	6.90	8.45	1.344	25.2	5.78
Iter	-	21	9	4	4	6	4	6

Example 1: The posterior mean of the log-permeability

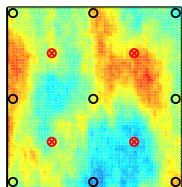


(a) True

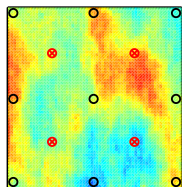
(b) ES-MDA-EQLx4

(c) ES-MDA-EQLx6

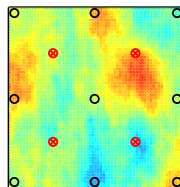
(d) ES-MDA-GEOx4



(e) ES-MDA-GEOx6

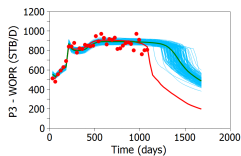


(f) ES-MDA-RLM 0.5

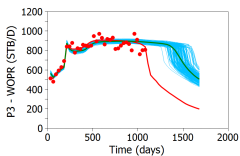


(g) IR-ES 0.5

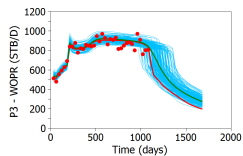
Example 1: Data Match - P3 Oil Rate



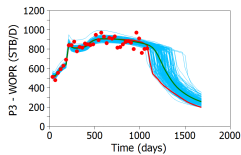
(a) ES-MDA-EQLx4



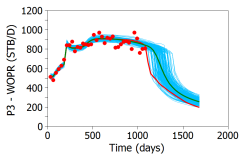
(b) ES-MDA-EQLx6



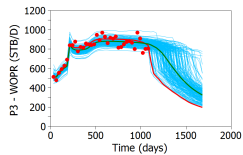
(c) ES-MDA-GEOx4



(d) ES-MDA-GEOx6

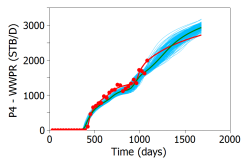


(e) ES-MDA-RLM 0.5

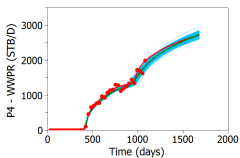


(f) IR-ES 0.5

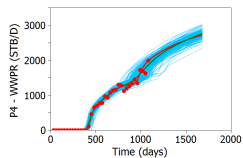
Example 1: Data Match - P4 Water Rate



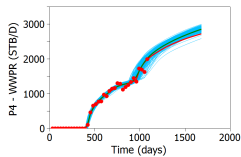
(a) ES-MDA-EQLx4



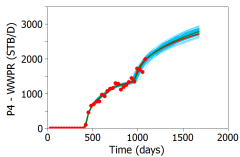
(b) ES-MDA-EQLx6



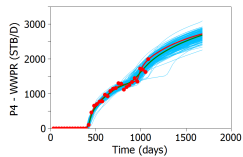
(c) ES-MDA-GEOx4



(d) ES-MDA-GEOx6



(e) ES-MDA-RLM 0.5



(f) IR-ES 0.5

Posterior PDF for Gaussian Prior and Gaussian Measurement/Modeling Errors

- Assume the prior model is multivariate Gaussian,
 $N(m_{\text{prior}}, C_M)$,

$$f(m) = \frac{1}{(2\pi)^{N_d/2} \det C_M} \exp\left(-\frac{1}{2}(m - m_{\text{prior}})^T C_M^{-1}(m - m_{\text{prior}})\right)$$

$$f(m|d_{\text{obs}}) = a \exp(-O(m))$$

$$O(m) = \frac{1}{2}(g(m) - d_{\text{obs}})^T C_D^{-1}(g(m) - d_{\text{obs}}) + \frac{1}{2}(m - m_{\text{prior}})^T C_M^{-1}(m - m_{\text{prior}})$$

- Metropolis-Hastings:
 - 1 From current state m_n , propose a state y according to $q(y|m_n)$.
 - 2 Accept state y with probability $\alpha(y|m_n) = \min(1, \frac{\pi(y)q(m_n|y)}{\pi(m_n)q(y|m_n)})$.
 - 3 If y is accepted, set $m_{n+1} = y$, otherwise, $m_{n+1} = m_n$.
- If $q(y|m_n) = N(m_n, \sigma^2 C_M)$, we refer to this strategy as random walk MCMC.
- σ is called the scaling factor, it controls the performance of the chain.
- Adaptive - use states to adapt mean, covariance and even σ to get acceptance rates close to 0.234. (Hario et al. (2001), Roberts and Rosenthal (2007), Andrieu and Thoms (2008), Holden et al. (2009),...

Sampling Multimodal Distributions

- Minimize objective function with different initial guesses with adjoint gradients and quasi-Newton trust region method to find local minimum of OF to obtain modes of the posterior.
- Cluster models, then build a Gaussian mixture model (GMM); each Gaussian has the minimum (mode) as its mean and associated inverse Hessian as covariance.
- GMM as proposal distribution in MCMC:

$$q(m | m_i) = \sum_{k=1}^{N_m} \frac{1}{N_m} G_k(m | m_k^*, C_k).$$

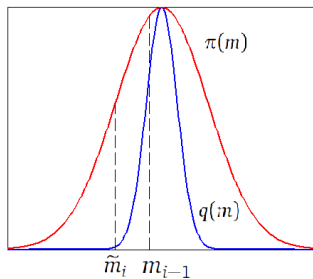
- The acceptance probability for \tilde{m}_{i+1} is

$$\alpha = \min\left(1, \frac{\pi(\tilde{m}_{i+1})q(m_i)}{q(\tilde{m}_{i+1})\pi(m_i)}\right).$$

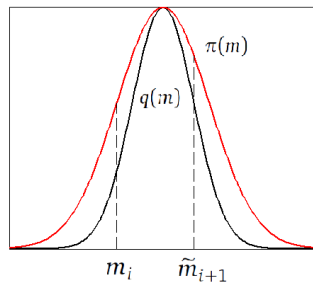
- Scale Gaussians at beginning by replacing C_k by $\gamma_k C_k$ and apply covariance matrix adaptation.

Covariance Matrix Adaptation

- The acceptance probability for \tilde{m}_i is $\alpha = \min(1, \frac{\pi(\tilde{m}_i)q(m_{i-1})}{q(\tilde{m}_i)\pi(m_{i-1})})$.
- The acceptance probability for \tilde{m}_{i+1} is $\alpha = \min(1, \frac{\pi(\tilde{m}_{i+1})q(m_i)}{q(\tilde{m}_{i+1})\pi(m_i)})$.



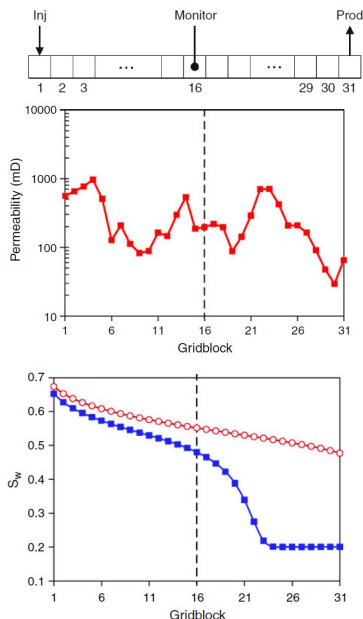
(a) Before Adaptation



(b) After Adaptation

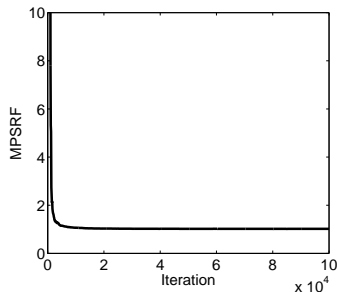
Example One: Case Description

- 1D; $31 \times 1 \times 1$ reservoir.
- Initial reservoir pressure is 3500 psi.
- One injector and one producer at each end of the reservoir.
- Injector BHP=4000 psi, Producer BHP=3000 psi
- One monitor well located at the center of reservoir.
- Pressure data at the monitor well are assimilated every 30 days until 360 days.



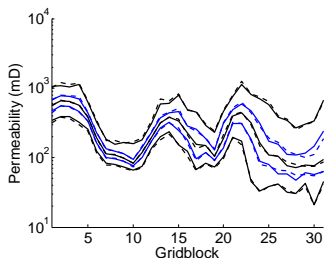
Two-stage MCMC Convergence

- 200 different initial guesses are used.
- We use the models whose normalized objective function value below 1.5 to generate 25 modes by k-medoids clustering.
- Figure below shows the convergence rate of five parallel chains, based on MPSRF (Brooks and Gelman, 1998).
- Samples from 10,000 to 15,000 of each chain are mixed and form the posterior distribution.

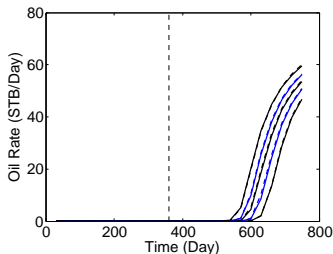


Check Two-stage MCMC Convergence

- Figure below shows the comparison of posterior distribution using all states with indices from 10 thousand to 15 thousand and using all states with indices from 60 thousand to 200 thousand.
- The acceptance rate for this case is 15%.



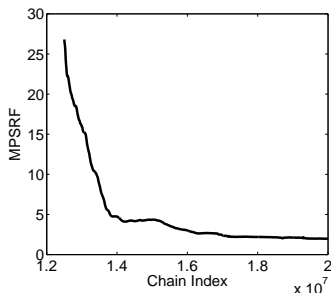
(a) Permeability Distribution



(b) Water Rate Distribution

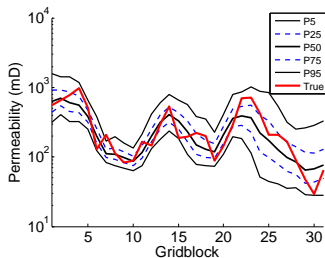
Random Walk Convergence

- The results using random walk are used as the base case for comparison with the two-stage MCMC method.
- For random walk, we initialize five parallel chains, the length of each chain is 23 million.
- Figure below shows the convergence rate of five parallel chains, based on MPSRF (Brooks and Gelman, 1998).
- The last 500 thousand samples from each chain are mixed to approximate the posterior distribution.

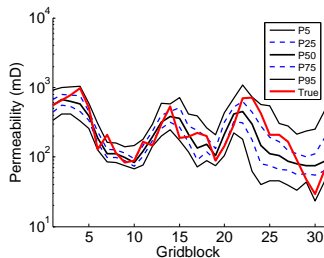


Permeability Distribution

- The left figure represents permeability distribution using random walk.
- The right figure represents permeability distribution using two-stage MCMC.



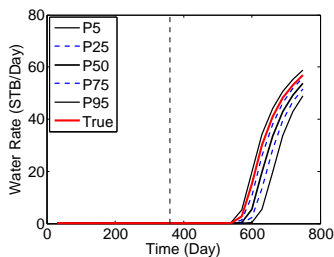
(a) Random walk



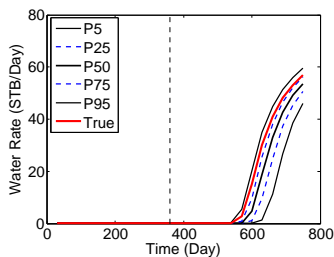
(b) Two-stage MCMC

Water Rate Prediction

- The left figure represents water production rate prediction using random walk.
- The right figure represents water production rate prediction using two-stage MCMC.



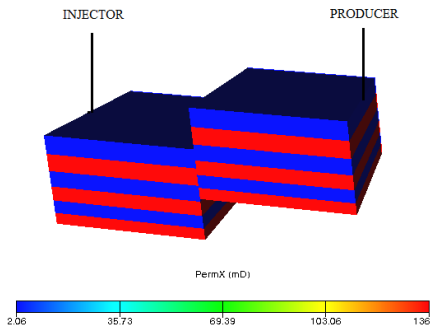
(a) Random walk



(b) Two-stage MCMC

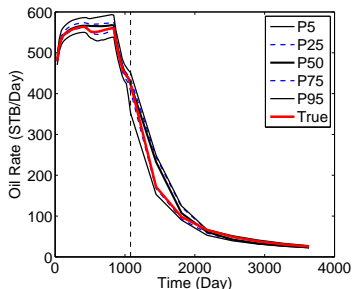
Example Two: IC-Fault Model

- 2D; 100×12 gridblocks.
- Injector is on the left side, producer is on the right side.
- Injector BHP=4700 psi, Producer BHP=4300 psi
- The reservoir contains six alternating layers of high and low quality sands with high and low permeability.
- Water injection rate, oil production rate and water production rate are assimilated every 30 days until 3 years.
- Three parameters: permeability of good and poor sand, fault throw.
- Compare with Population MCMC

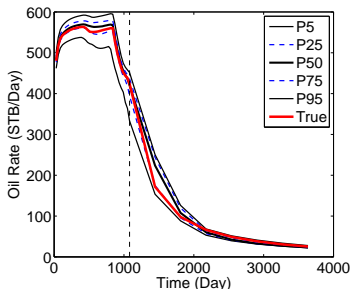


Results using Algorithm I

- The left figure represents oil production rate using two-stage MCMC (states from 1200 to 3000).
- The right figure represents oil production rate using population MCMC (states from 1000 to 2000).



(a) Two-stage MCMC



(b) Population MCMC

Algorithm II: Approximate Two-stage MCMC Method

- Starting from different initial guesses, we use an optimization algorithm to locate multiple modes of the PDF.
- Selecting representative modes using clustering algorithm.
- Generate a number of samples from each mode and form a set S of all samples.
- Sample m_l randomly from S , and define the acceptance probability by $\alpha = \min(1, \frac{\pi(m_l)q(m_{l-1})}{\pi(m_{l-1})q(m_l)})$. Here,

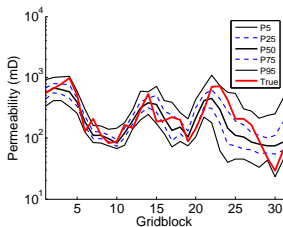
$$q(m) = \sum_{k=1}^{N_m} \frac{1}{N_m} N(m_k^*, C_k).$$

- Adapt the covariance matrix.

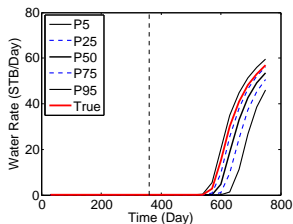
The advantage of this method is that the computational cost is relatively low, and we only need evaluate the objective function a few thousand times.

Results using Algorithm II

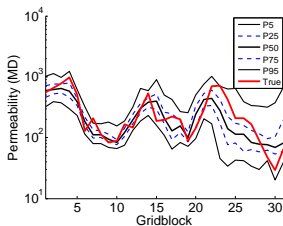
- Generate 200 samples from each of the 10 modes.



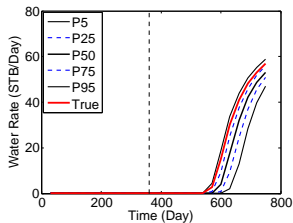
(a) Algorithm I



(b) Algorithm I



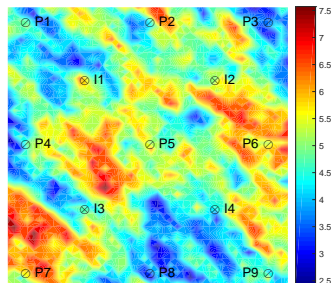
(c) Algorithm II



(d) Algorithm II

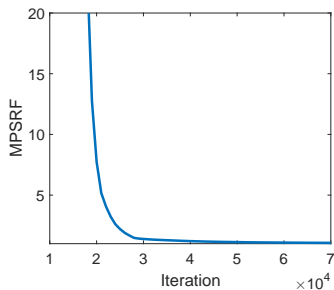
Example Three: 2D Synthetic Case

- 2D; 44×44 gridblocks.
- Model parameters: gridblock log permeabilities.
- 9 producers and 4 injectors.
- Initial reservoir pressure is 3,000 psi.
- All producers (2800 psi) and injectors (4500 psi) are in bottomhole pressure control.
- The true model is built using an exponential covariance function with major correlation length 2100 ft, minor correlation length 1100 ft.
- Gaussian random noise with zero mean and standard deviation equal to 5% (minimum of 2 bbl) of the true data are added.
- Water injection rate, oil production rate and water production rate are assimilated every 30 days until 3 years.
- The prediction period is 20 months.



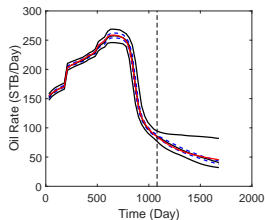
Two-stage MCMC Convergence

- 250 different initial guesses are used.
- We use the models which give a normalized objective function value below 1.5 to generate 35 modes by k-medoids clustering.
- Figure below shows the convergence rate of five parallel chains, based on MPSRF (Brooks and Gelman, 1998).
- Samples from 30,000 to 40,000 of each chain are mixed and form the posterior distribution. (acceptance rate is 9%)

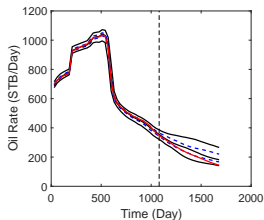


Comparison using Two-Stage MCMC Algorithms

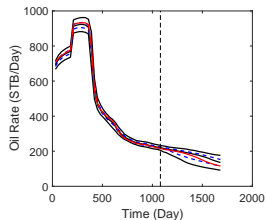
- 200 samples from each of the 35 modes.
- The approximate two-stage MCMC method takes 9000 simulation runs.



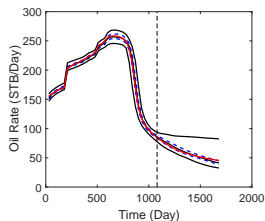
(a) Producer 2



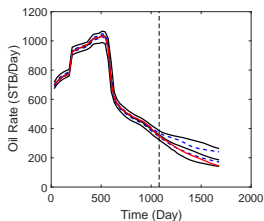
(b) Producer 5



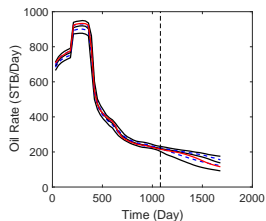
(c) Producer 8



(d) Producer 2



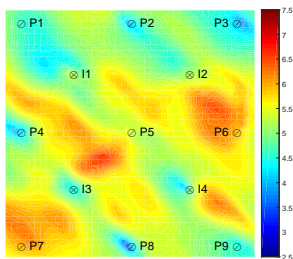
(e) Producer 5



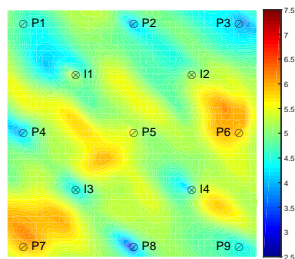
(f) Producer 8

Comparison using Two-stage MCMC Algorithms

- Mean of the log-permeability posterior distribution.



(a) Two-stage MCMC



(b) Approximate Two-stage MCMC

- ES-MDA-Geo with $N_a = 4$ is a good first try.
- Work to make ES-MDA perform better with object-based modeling is underway.
- PreRML-GMM-MCMC with covariance matrix adaptation may be plausible (feasible?) for UQ. At the least, it can be used to provide benchmark problems for testing other UQ methodology.
- Perhaps useful for cases with reduced parameter sets on the order of 25 or less.
- If adjoint not available, use a localized iterative ES to estimate modes - work underway.