Properties of iterative ensemble smoothers and strategies for conditioning on production data

Geir Evensen

IPAM: May 22-26, 2017

History matching problem

Given:

- Prior parameters $\boldsymbol{x}$ described by $f(\boldsymbol{x})$.
- A model $\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x})$ described by $f(\boldsymbol{y}|\boldsymbol{x}) \propto \delta(\boldsymbol{y} - \boldsymbol{g}(\boldsymbol{x}))$.
  - Perfect model.
- Measurements $\boldsymbol{d}$ of $\boldsymbol{y}$ described by $f(\boldsymbol{d}|\boldsymbol{y})$.

The joint conditional pdf becomes

$$f(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{d}) \propto f(\boldsymbol{x}, \boldsymbol{y}) f(\boldsymbol{d}|\boldsymbol{y}) \propto f(\boldsymbol{x}) f(\boldsymbol{y}|\boldsymbol{x}) f(\boldsymbol{d}|\boldsymbol{y})$$

Posterior marginal pdf for $\boldsymbol{x}$

$$
\begin{aligned}
f(\boldsymbol{x}|\boldsymbol{d}) &\propto \int f(\boldsymbol{x}) f(\boldsymbol{y}|\boldsymbol{x}) f(\boldsymbol{d}|\boldsymbol{y}) d\boldsymbol{y} \\
&= \int f(\boldsymbol{x}) \boldsymbol{\delta}(\boldsymbol{y} - \boldsymbol{g}(\boldsymbol{x})) f(\boldsymbol{d}|\boldsymbol{y}) d\boldsymbol{y} \\
&= f(\boldsymbol{x}) f(\boldsymbol{d}|\boldsymbol{g}(\boldsymbol{x}))
\end{aligned}
$$

Assume Gaussian prior and measurement errors

$$
f(\boldsymbol{x}|\boldsymbol{d}) = \exp -\frac{1}{2} \left\{ (\boldsymbol{x} - \boldsymbol{x}^{\mathrm{f}})^{\mathrm{T}} \left(\boldsymbol{C}^{\mathrm{f}}_{xx} \right)^{-1} (\boldsymbol{x} - \boldsymbol{x}^{\mathrm{f}}) + (\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{d})^{\mathrm{T}} \boldsymbol{C}_{dd}^{-1} (\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{d}) \right\}
$$

History-matching problem

Sampling the posterior

$$f(\boldsymbol{x}|\boldsymbol{d}) = \exp -\frac{1}{2}\left\{(\boldsymbol{x}-\boldsymbol{x}^{\mathrm{f}})^{\mathrm{T}}\left(\boldsymbol{C}^{\mathrm{f}}_{xx}\right)^{-1}(\boldsymbol{x}-\boldsymbol{x}^{\mathrm{f}}) + (\boldsymbol{g}(\boldsymbol{x})-\boldsymbol{d})^{\mathrm{T}}\boldsymbol{C}^{-1}_{dd}(\boldsymbol{g}(\boldsymbol{x})-\boldsymbol{d})\right\}$$

or minimizing the cost function

$$J(\boldsymbol{x}) = (\boldsymbol{x}-\boldsymbol{x}^{\mathrm{f}})^{\mathrm{T}}\left(\boldsymbol{C}^{\mathrm{f}}_{xx}\right)^{-1}(\boldsymbol{x}-\boldsymbol{x}^{\mathrm{f}}) + (\boldsymbol{g}(\boldsymbol{x})-\boldsymbol{d})^{\mathrm{T}}\boldsymbol{C}^{-1}_{dd}(\boldsymbol{g}(\boldsymbol{x})-\boldsymbol{d}).$$

Solution methods

- Gradient methods
  - Best to use adjoints.
  - Local minima and error estimates?
- MCMC or genetic algorithms.
  - Dimension of solution space is the number of parameters.
  - Possible to fit only $\mathcal{O}(10)$ parameters to the measurements.
- Ensemble methods (EnKF, ES, ES–MDA, IES, PF).
  - Ensemble sub-space solution.
  - Handle large number of measurements and parameters.

Ensemble methods

Sampling the priors

$$\boldsymbol{f}(\boldsymbol{x}) = N(\boldsymbol{x}^{\mathrm{f}}, \boldsymbol{C}^{\mathrm{f}}_{xx}) \rightarrow \{\boldsymbol{x}^{\mathrm{f}}_j\}$$
$$\boldsymbol{f}(\boldsymbol{d}|\boldsymbol{g}(\boldsymbol{x})) = N(\boldsymbol{d}, \boldsymbol{C}_{dd}) \rightarrow \{\boldsymbol{d}_j\}$$

and solving for posterior samples $\boldsymbol{x}_j$ from minimizing

$$J(\boldsymbol{x}_j) = (\boldsymbol{x}_j - \boldsymbol{x}^{\mathrm{f}}_j)^{\mathrm{T}} \left(\boldsymbol{C}^{\mathrm{f}}_{xx}\right)^{-1} (\boldsymbol{x}_j - \boldsymbol{x}^{\mathrm{f}}_j) + (\boldsymbol{g}(\boldsymbol{x}_j) - \boldsymbol{d}_j)^{\mathrm{T}} \boldsymbol{C}^{-1}_{dd} (\boldsymbol{g}(\boldsymbol{x}_j) - \boldsymbol{d}_j)$$

- Minimization for each realization j as in IES.
- Approximate direct solution as in ES and ES–MDA.

Redundancy in data vs. dependent data

Dependent data:

$$f(\boldsymbol{d}|\boldsymbol{g}(\boldsymbol{x})) \neq \prod_{i=1}^{m} f(d_i|\boldsymbol{g}(\boldsymbol{x}))$$

i.e., measurement errors are correlated.

Redundant information:

$$\boldsymbol{H}\boldsymbol{C}_{xx}^{\mathrm{f}}\boldsymbol{H}^{\mathrm{T}} \quad \text{is of low rank.}$$

Measurements contain some of the same information.

Scalar update with one measurement

Update equations:

$$x^{\mathrm{a}} = x^{\mathrm{f}} + C^{\mathrm{f}}_{xx}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}C^{\mathrm{f}}_{xx}\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{C}_{dd})^{-1}(\boldsymbol{d} - \boldsymbol{H}x^{\mathrm{f}})$$

$$C^{\mathrm{a}}_{xx} = \left(1 - C^{\mathrm{f}}_{xx}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}C^{\mathrm{f}}_{xx}\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{C}_{dd})^{-1}\boldsymbol{H}\right)C^{\mathrm{f}}_{xx}$$

Prior:

$$x^{\mathrm{f}} = 0, \qquad C^{\mathrm{f}}_{xx} = 1$$

Measurement:

$$d_1 = 1, \qquad C_{dd} = 1$$

Solution:

$$x^{\mathrm{a}} = 0.5, \qquad C^{\mathrm{a}}_{xx} = 0.5$$

Scalar update with two measurements:

Uncorrelated errors

Prior:

$$x^{\mathrm{f}} = 0, \qquad C^{\mathrm{f}}_{xx} = 1$$

Measurements:

$$d_1 = 1, \quad d_2 = 1, \qquad \boldsymbol{H} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \boldsymbol{C}_{dd} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Inversion of

$$\boldsymbol{H} C^{\mathrm{f}}_{xx} \boldsymbol{H}^{\mathrm{T}} + \boldsymbol{C}_{dd} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Solution:

$$x^{\mathrm{a}} = \frac{2}{3} \approx 0.67, \qquad C^{\mathrm{a}}_{xx} = \frac{1}{3} \approx 0.33$$

Sequential processing of measurements

IRIS

Prior (posterior from above):

$$x^{\mathrm{f}} = 0.5, \qquad C^{\mathrm{f}}_{xx} = 0.5$$

Second measurement:

$$d_2 = 1, \qquad C_{dd} = 1$$

Solution:

$$x^{\mathrm{a}} = \frac{2}{3} \approx 0.67, \qquad C^{\mathrm{a}}_{xx} = \frac{1}{3} \approx 0.33$$

Scalar update with two measurements:

Correlated errors

Prior:

$$x^{\mathrm{f}} = 0, \qquad C^{\mathrm{f}}_{xx} = 1$$

Measurements:

$$d_1 = 1, \quad d_2 = 1, \qquad \boldsymbol{H} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \boldsymbol{C}_{dd} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

Inversion of

$$\boldsymbol{H} C^{\mathrm{f}}_{xx} \boldsymbol{H}^{\mathrm{T}} + \boldsymbol{C}_{dd} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

Solution:

$$x^{\mathrm{a}} = \frac{4}{7} \approx 0.57, \qquad C^{\mathrm{a}}_{xx} = \frac{3}{7} \approx 0.43$$

Learning

Redundant information in measurements leads to low-rank $\boldsymbol{H}\boldsymbol{C}_{xx}^{\mathrm{f}}\boldsymbol{H}^{\mathrm{T}}$.

- Take care when inverting $\boldsymbol{H}\boldsymbol{C}_{xx}^{\mathrm{f}}\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{C}_{dd}$

Independent measurements can be processed recursively.

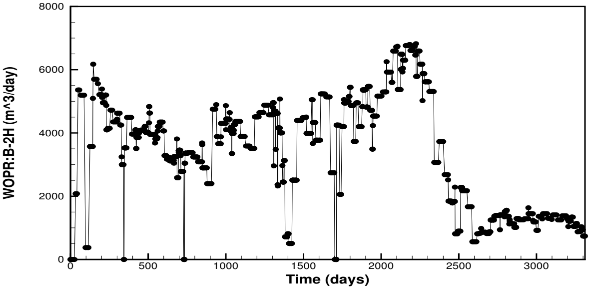
- Underlying assumption in EnKF.

Neglecting correlations in measurement errors leads to:

- Too strong update of estimate.
- Too strong reduction of posterior variance.

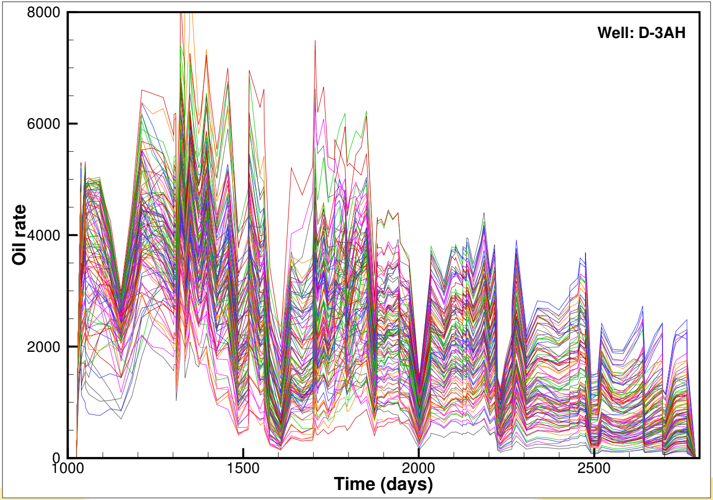
Historical rate data

Time series of oil, water and gas rates from each well.

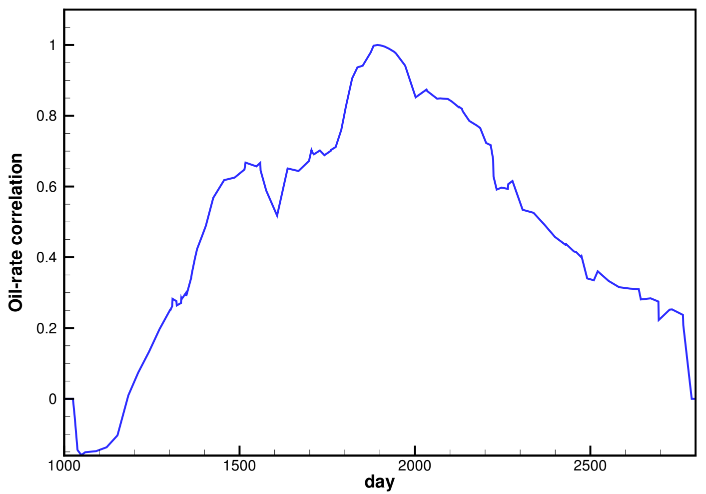


- Redundant information in measurements?
- Correlated measurement errors?

Redundant information in rate data?



Redundant information in rate data?



Learning

IRIS

Information from rate data may be highly dependent in time.

- Most of the variability in the data is a result of external forcing.
- We need to handle poor conditioning of $\boldsymbol{H}\boldsymbol{C}^{\mathrm{f}}_{xx}\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{C}_{dd}$.

Correlated errors in rate data?

- Rates are often derived from a rate-allocation table.
- Table constructed from infrequent separator tests.
- Leads to strong time correlations of measurement errors.

Thus, we can not assume a diagonal C_{dd}!

Observations from using EnKF

- Implicitely assumes independent measurement errors.

$$f(\boldsymbol{d}|\boldsymbol{y}) = \prod_i f(\boldsymbol{d}_i|\boldsymbol{y})$$

- EnKF has lead to “ensemble collapse.”
- Attributed to ensemble inbreeding or negligence of model errors.
- Resolved by introducing localization and inflation.

Observations from using ES/IES/ES–MDA

- Also uses diagonal C_{dd}.
- Also observed to lead to "ensemble collapse."
- Still attributable to inbreeding in ensemble?

Neglecting measurement error correlations or using poor inversion methods may lead to ensemble collapse.

Solutions

1. Proper modeling of C_{dd}.
2. Error-subspace inversion (*Evensen*, 2009, Chap. 14).
3. Alternatively, strongly subsample the rate data.

Norne example

No practical difference in using rates vs accumulated production.

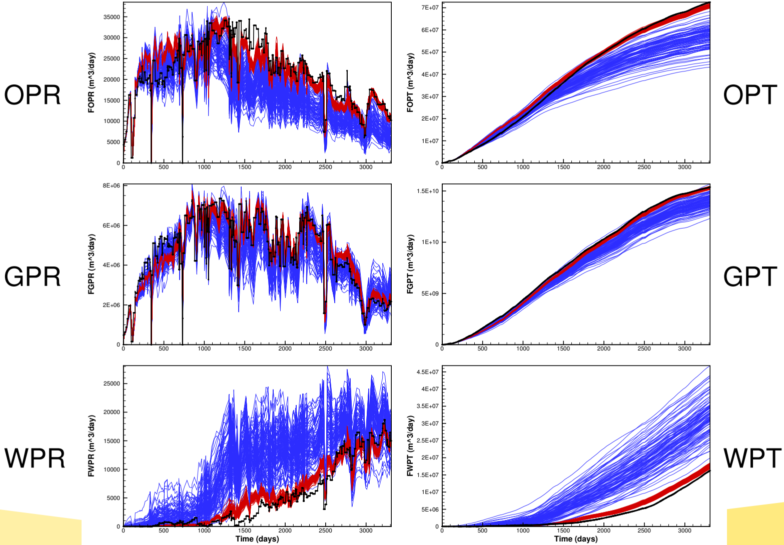
Conditioning only on total accumulated production per well.

- Observations: Final WOPT, WGPT, and WWPT.
- 22 wells: i.e. 66 measurements.
- Should be ok to use diagonal C_{dd}.
- Numerically stable with 100 realizations.
- Results from ES, IES and ES–MDA.

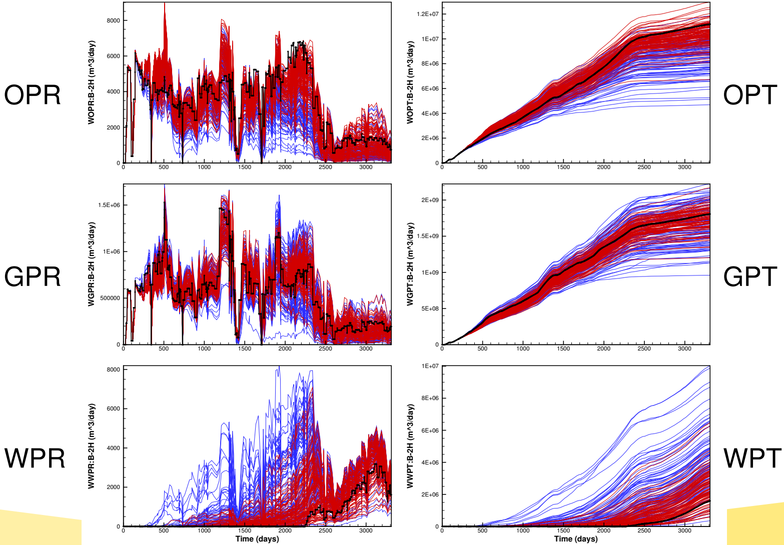
Parametrization

Parameter	Dimension	Distribution
PORO	44927	NORMAL
PERMX	44927	LOG-NORMAL
MULTFLT	53	LOG-UNIFORM $[0.00001 : 1.0]$
MULTRGT	3	LOG-UNIFORM $[0.00001 : 1.0]$
MULTZ	6	LOG-UNIFORM $[0.00001 : 1.0]$
RELPERM KRW	4	UNIFORM $[0.8 : 1.5]$
RELPERM KRG	4	UNIFORM $[0.8 : 1.0]$
OWC:Garn (C,D)	1	NORMAL N(2682.0 m, 5 m)
OWC:Garn (G)	1	NORMAL N(2585.5 m, 5 m)
OWC:Garn (E)	1	NORMAL N(2638.0 m, 5 m)
OWC:Ile, Tilje (G)	1	NORMAL N(2400.0 m, 5 m)
OWC:Ile, Tilje (C,D,E)	1	NORMAL N(2693.3 m, 5 m)

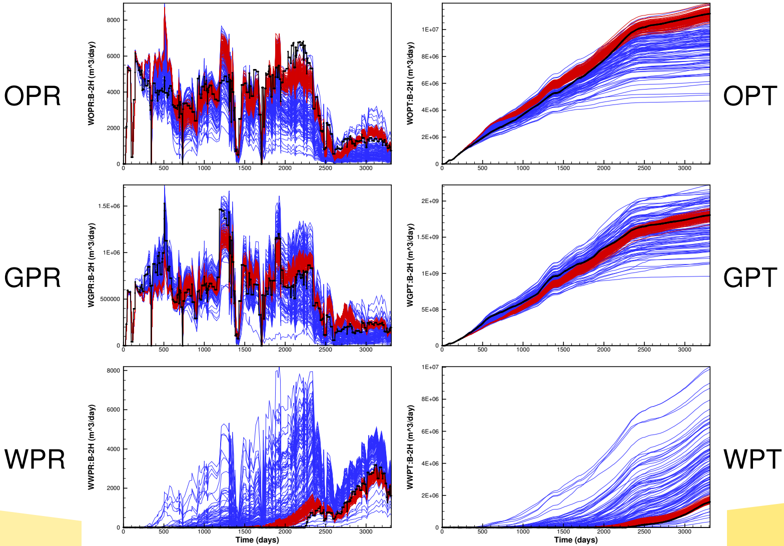
Total field production: ES–MDA 8



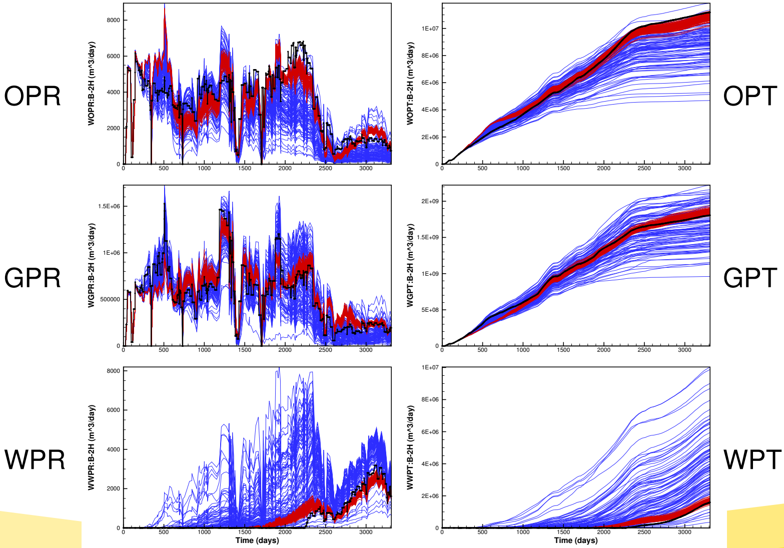
Well B-2H: ES


OPR
GPR
WPR
OPT
GPT
WPT
WOPR:B-2H (m^3/day)
WGPR:B-2H (m^3/day)
WWPR:B-2H (m^3/day)
WOPT:B-2H (m^3/day)
WGPT:B-2H (m^3/day)
WWPT:B-2H (m^3/day)
Time (days)
Time (days)

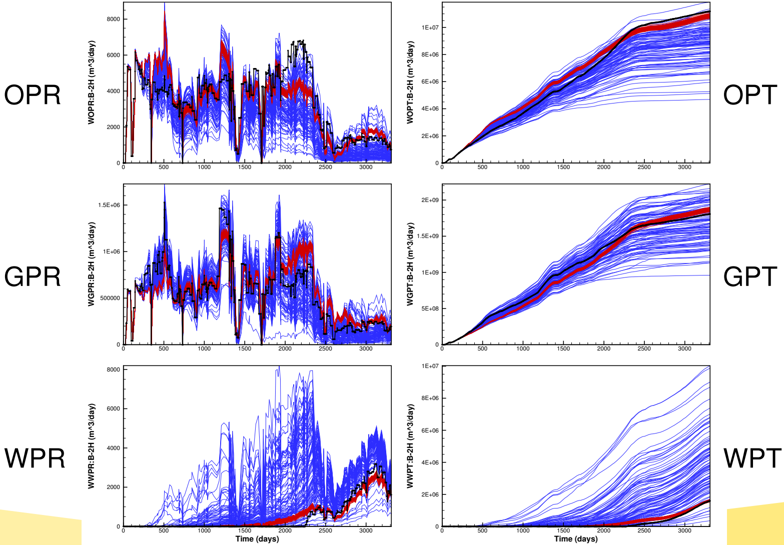
Well B-2H: ES–MDA 8



Well B-2H: ES–MDA 16



Well B-2H: IES


OPR
OPT
GPR
GPT
WPR
WPT
WOPR:B-2H (m^3/day)
WOPT:B-2H (m^3/day)
WGPR:B-2H (m^3/day)
WGPT:B-2H (m^3/day)
WWPR:B-2H (m^3/day)
WWPT:B-2H (m^3/day)
Time (days)
Time (days)

Learning

Conditioning only on total accumulated production per well:

- Both IES and ES–MDA lead to acceptable match.
  - Forcing ECLIPSE with RESV gives correct variability in time.
  - Conditioning leads to correct total production of oil, gas, and water.
- Resolves issue with ensemble collapse.
- Var(ES) > Var(ES–MDA 8) > Var(ES–MDA 16) > Var(IES).

Some questions

- Why does it help to iterate?
- What are we iterating?
- What is the difference between IES and ES–MDA?
- What does IES and ES–MDA converge to?

ES–MDA recursion

Bayes'

$$f(x|d) \propto f(x) \prod_{i=1}^{N} f(d|g(x_i))^{\frac{1}{\alpha_i}}$$

Recursive Bayes'

$$f(x_1|d) \propto f(x) f(d|g(x_1))^{\frac{1}{\alpha_1}}$$

$$f(x_2|d) \propto f(x_1|d) f(d|g(x_2))^{\frac{1}{\alpha_2}}$$

$$\vdots$$

$$f(x_N|d) \propto f(x_{N-1}|d) f(d|g(x_N))^{\frac{1}{\alpha_N}}$$

ES–MDA

ES–MDA solves for each realization j

$$J(x_{j,i+1}) = (x_{j,i+1} - x_{j,i}) \left((C^{\mathrm{e}}_{xx})_i\right)^{-1} (x_{j,i+1} - x_{j,i}) \\ + (g(x_{j,i+1}) - d - \sqrt{\alpha_i}\epsilon) \left(\alpha_i C^{\mathrm{e}}_{dd}\right)^{-1} (g(x_{j,i+1}) - d - \sqrt{\alpha_i}\epsilon),$$

ES–MDA equations

$$x_{j,i+1} = x_{j,i} + (C^{\mathrm{e}}_{xy})_i \left((C^{\mathrm{e}}_{yy})_i + \alpha_i C^{\mathrm{e}}_{dd}\right)^{-1} (d + \sqrt{\alpha_i}\epsilon_j - g(x_{j,i}))$$

$$y_{j,i+1} = g(x_{j,i+1})$$

Ensemble smoothers

Linear model:

1. ES, ES–MDA and IES exactly sample the posterior.

Nonlinear model:

1. ES computes one step based on linearization around x^{f}.
2. ES–MDA applies a sequence of local linearizations.
3. IES approximately minimizes the cost function for each realization.

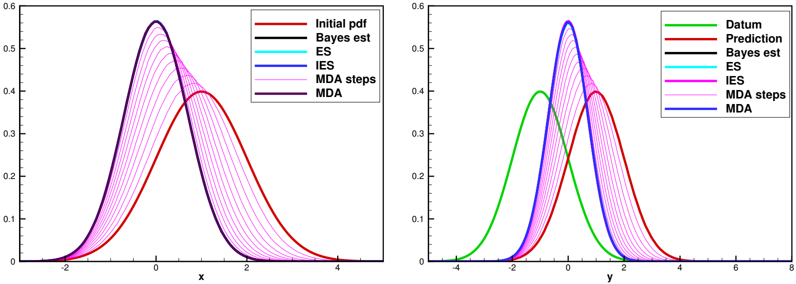
Scalar example

Model

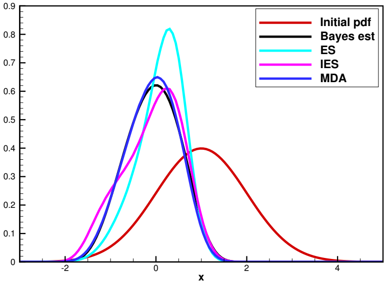
$$y = x(1 + \beta x^2)$$

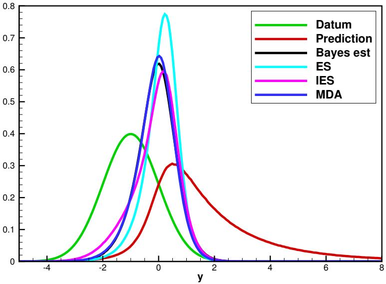
- Linear case: $\beta = 0$.
- Nonlinear case: $\beta = 0.2$.
- Prior ensemble for x: $N(1, 1)$.
- Likelihood for measurement of y: $N(-1, 1)$.

Linear problem

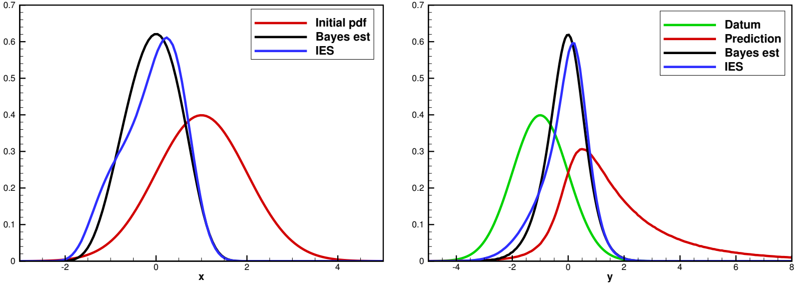


Nonlinear problem



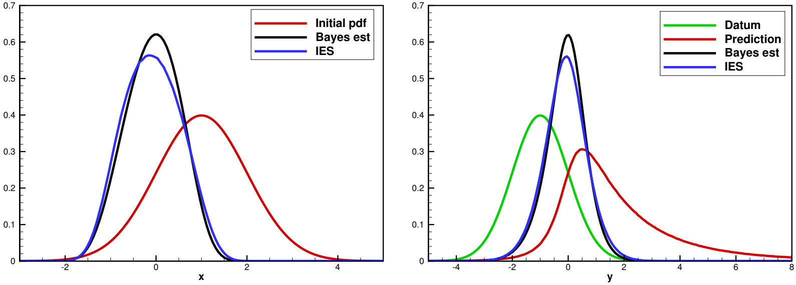


IES: Analytic $\nabla J(x)$



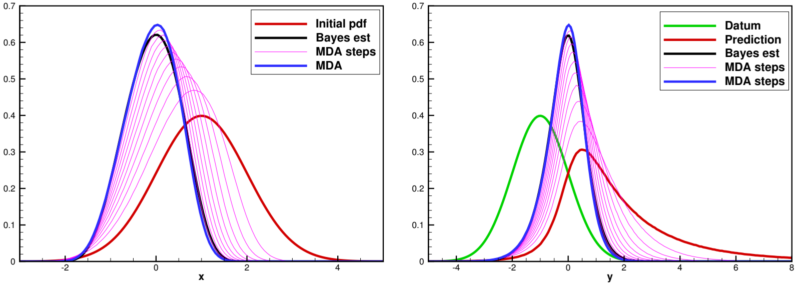
$$x_{j,i+1} = x_{j,i} - \gamma \frac{C_{dd}(x_{j,i} - x_j^{\mathrm{f}}) + g'(x_{j,i})C_{xx}(g(x_{j,i}) - d_j)}{g'(x_{j,i})C_{xx}g'(x_{j,i}) + C_{dd}}$$

IES: Ensemble $\nabla J(x)$

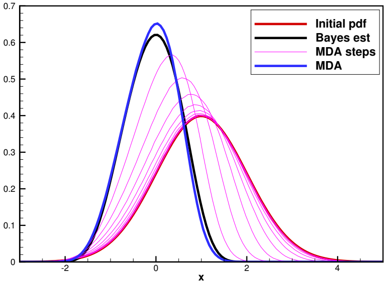


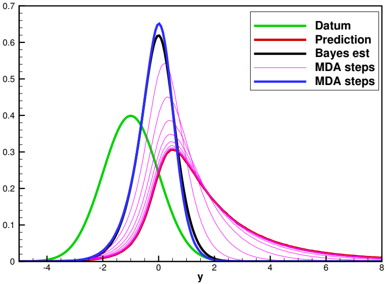
$$x_{j,i+1} = x_{j,i} - \gamma \frac{C_{dd}(x_{j,i} - x_j^{\mathrm{f}}) + P_{xy}^{\mathrm{e},i} \left(P_{xx}^{\mathrm{e},i}\right)^{-1} C_{xx}^{\mathrm{e}}(g(x_{j,i}) - d_j)}{P_{yy}^{\mathrm{e},i} + C_{dd}}$$

ES–MDA–10 : uniform α

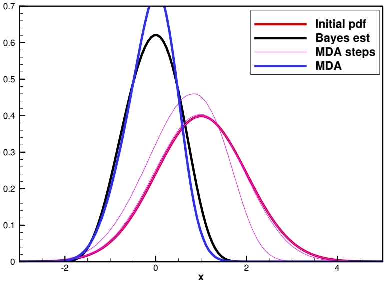


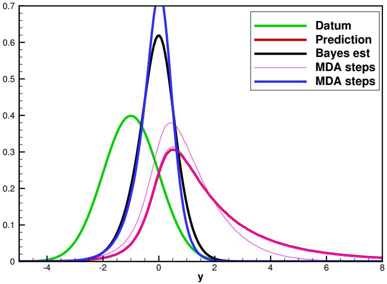
ES–MDA–10 : geometrical $\alpha(i) = \alpha(i-1)/2$



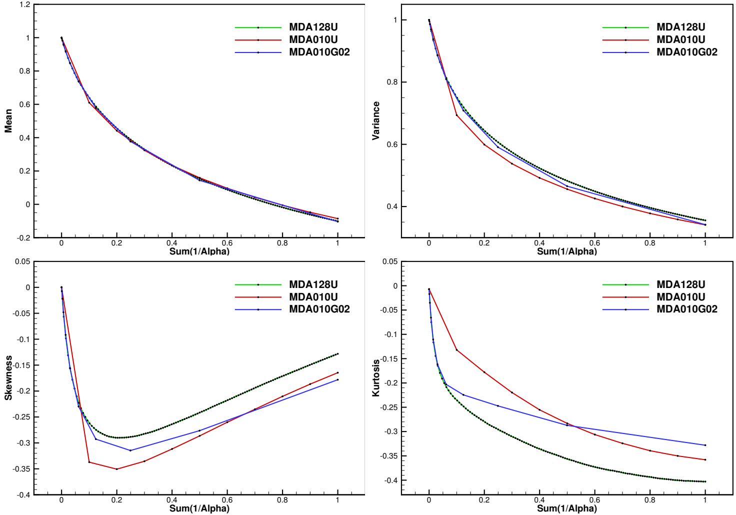


ES–MDA–10 : geometrical $\alpha(i) = \alpha(i-1)/10$

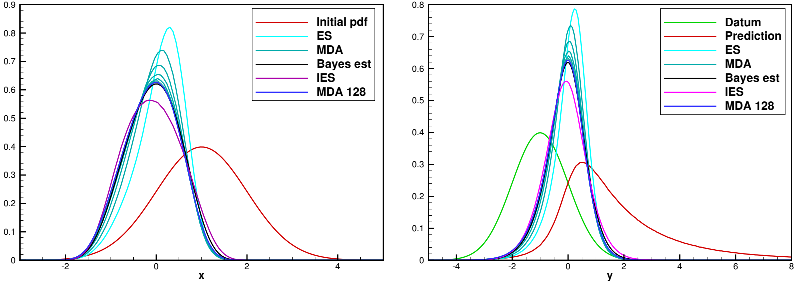




MDA-10 Statistics



ES–MDA convergence
(2, 4, 8, 16, 32, 64, 128 steps)



Acknowledgement

- Kjersti Solberg Eikrem for the Norne experiments.
- Joakim Hove/Statoil for developing ERT and making it available.
- Statoil for funding much of this work.

Chen, Y., and D. S. Oliver, Levenberg-Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification, *Computational Geosciences*, *17*, 689–703, 2013.

Emerick, A. A., and A. C. Reynolds, Ensemble smoother with multiple data assimilation, *Computers and Geosciences*, *55*, 3–15, 2013.

Evensen, G., *Data Assimilation: The Ensemble Kalman Filter*, 2nd ed., Springer, 2009.

Skjervheim, J.-A., G. Evensen, J. Hove, and J. Vabø, An ensemble smoother for assisted history matching, SPE 141929, 2011.

Stordal, A., and A. H. Elsheikh, Iterative ensemble smoothers in the annealed importance sampling framework, *Advances in Water Resources*, *86*, 231–239, 2015.

van Leeuwen, P. J., and G. Evensen, Data assimilation and inverse methods in terms of a probabilistic formulation, *Mon. Weather Rev.*, *124*, 2898–2913, 1996.

Takeaways

Rate data are dependent in time and contain redundant information.

- Need to properly model $\boldsymbol{C}_{dd}$.
- Need stable schemes for inverting $\boldsymbol{H}\boldsymbol{C}_{xx}\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{C}_{dd}$.
- Error-subspace inversion (*Evensen*, 2009, Chap. 14).

Rate data may be strongly subsampled.

- Sufficent to conditioning on total production per well?
- Time variability is introduced from RESV forcing.

Iterative smoothers like IES and ES–MDA should be used.

- Solves different problems but which method is the best?
- What does ES–MDA and IES converge to in the nonlinear case?
- What is the optimal number of ES–MDA steps and weights?