Aspects of sequential and simultaneous assimilation

May 29, 2013

Motivation

- There exist several algorithms for conditioning a model to data, s.a. EnKF, RML, ES, EnRML, MDA,...
- All methods generate approximate samples from the same distribution
  - Methods sample correctly for linear problems
  - Methods give different results for non-linear problems
- Differences between methods are defined by some key characteristics
- Focus on: Sequential vs. simultaneous assimilation of data for updating static parameters

Motivation

- Formal Bayesian expression
  - Seq. data assimilation = sim. data assimilation

Motivation

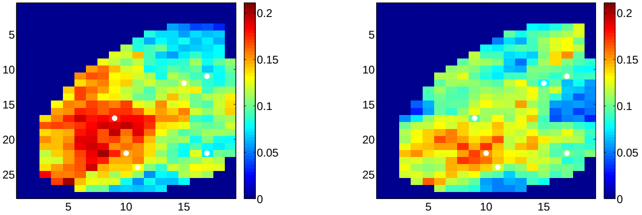
- Formal Bayesian expression
  - Seq. data assimilation = sim. data assimilation
- Approximate methods
  - Linear forward models: seq. data assimilation = sim. data assimilation

Motivation

- Formal Bayesian expression
  - Seq. data assimilation = sim. data assimilation
- Approximate methods
  - Linear forward models: seq. data assimilation = sim. data assimilation
  - Non-linear forward models: seq. data assimilation $\neq$ sim. data assimilation

Motivation

- Formal Bayesian expression
  - Seq. data assimilation = sim. data assimilation
- Approximate methods
  - Linear forward models: seq. data assimilation = sim. data assimilation
  - Non-linear forward models: seq. data assimilation $\neq$ sim. data assimilation



Seq. scheme

Sim. scheme

Analytical strategy

Goal: Understand the importance of seq. and sim. assimilation when combining data with different degrees of non-linearity

Analytical strategy

Goal: Understand the importance of seq. and sim. assimilation when combining data with different degrees of non-linearity

- Note: Analytical result exist for seq. vs sim. RML with linear data

Analytical strategy

Goal: Understand the importance of seq. and sim. assimilation when combining data with different degrees of non-linearity

- Note: Analytical result exist for seq. vs sim. RML with linear data
- Strategy:

Analytical strategy

Goal: Understand the importance of seq. and sim. assimilation when combining data with different degrees of non-linearity

- Note: Analytical result exist for seq. vs sim. RML with linear data
- Strategy:
  - Define comparable variants of seq./sim. RML and EnKF/ES

Analytical strategy

Goal: Understand the importance of seq. and sim. assimilation when combining data with different degrees of non-linearity

- Note: Analytical result exist for seq. vs sim. RML with linear data
- Strategy:
  - Define comparable variants of seq./sim. RML and EnKF/ES
  - Analyze differences between the methods

Analytical strategy

Goal: Understand the importance of seq. and sim. assimilation when combining data with different degrees of non-linearity

- Note: Analytical result exist for seq. vs sim. RML with linear data
- Strategy:
  - Define comparable variants of seq./sim. RML and EnKF/ES
  - Analyze differences between the methods
  - Extend linear RML result to new RML variants for combination of linear and non-linear data

Analytical strategy

Goal: Understand the importance of seq. and sim. assimilation when combining data with different degrees of non-linearity

- Note: Analytical result exist for seq. vs sim. RML with linear data
- Strategy:
  - Define comparable variants of seq./sim. RML and EnKF/ES
  - Analyze differences between the methods
  - Extend linear RML result to new RML variants for combination of linear and non-linear data
  - Extend linear RML result for variants of EnKF/ES

Analytical strategy

Goal: Understand the importance of seq. and sim. assimilation when combining data with different degrees of non-linearity

- Note: Analytical result exist for seq. vs sim. RML with linear data
- Strategy:
  - Define comparable variants of seq./sim. RML and EnKF/ES
  - Analyze differences between the methods
  - Extend linear RML result to new RML variants for combination of linear and non-linear data
  - Extend linear RML result for variants of EnKF/ES

Characteristics & Algorithms

- Define variants of EnKF and the RML method
- Remove impact of other characteristics than seq./sim. by ensuring
  1. Updates based on ensemble
  2. Perform one complete run
  3. Focus on static parameters
- Choose versions of RML and EnKF honoring 1-3

Characteristics & Algorithms

- EnKF honors
  - Point 1: Updates based on ensemble
  - Point 2: Perform one complete run

Characteristics & Algorithms

- EnKF honors
  - Point 1: Updates based on ensemble
  - Point 2: Perform one complete run
- EnKF does not honor
  - Point 3: Focus on static parameters

Characteristics & Algorithms

- EnKF honors
  - Point 1: Updates based on ensemble
  - Point 2: Perform one complete run
- EnKF does not honor
  - Point 3: Focus on static parameters
- Solution
  - Restart from initial time after each assimilation

Characteristics & Algorithms

- EnKF honors
  - Point 1: Updates based on ensemble
  - Point 2: Perform one complete run
- EnKF does not honor
  - Point 3: Focus on static parameters
- Solution
  - Restart from initial time after each assimilation
  - EnKF $\rightarrow$ Half-iterative EnKS (Hi-EnKS)

Characteristics & Algorithms

- EnKF honors
  - Point 1: Updates based on ensemble
  - Point 2: Perform one complete run
- EnKF does not honor
  - Point 3: Focus on static parameters
- Solution
  - Restart from initial time after each assimilation
  - EnKF $\rightarrow$ Half-iterative EnKS (Hi-EnKS)
- If data are assimilated simultaneously: Hi-EnKS $\rightarrow$ ES

Characteristics & Algorithms

- EnKF honors
  - Point 1: Updates based on ensemble
  - Point 2: Perform one complete run
- EnKF does not honor
  - Point 3: Focus on static parameters
- Solution
  - Restart from initial time after each assimilation
  - EnKF $\rightarrow$ Half-iterative EnKS (Hi-EnKS)
- If data are assimilated simultaneously: Hi-EnKS $\rightarrow$ ES
- Hi-EnKS: sequential scheme honoring 1-3
- ES: simultaneous scheme honoring 1-3

Characteristics & Algorithms

- RML honors
  - Point 3: Focus on static parameters

Characteristics & Algorithms

- RML honors
  - Point 3: Focus on static parameters
- RML does not honor
  - Point 1: Updates based on ensemble

Characteristics & Algorithms

- RML honors
  - Point 3: Focus on static parameters
- RML does not honor
  - Point 1: Updates based on ensemble
- Solution
  - EnRML updates using an ensemble approximation to gradient: RML $\rightarrow$ EnRML

Characteristics & Algorithms

- RML honors
  - Point 3: Focus on static parameters
- RML does not honor
  - Point 1: Updates based on ensemble
  - Point 2: Perform one complete run
- Solution
  - EnRML updates using an ensemble approximation to gradient: RML $\rightarrow$ EnRML

Characteristics & Algorithms

- RML honors
  - Point 3: Focus on static parameters
- RML does not honor
  - Point 1: Updates based on ensemble
  - Point 2: Perform one complete run
- Solution
  - EnRML updates using an ensemble approximation to gradient: RML $\rightarrow$ EnRML
  - Minimize utilizing one full Gauss-Newton step: EnRML $\rightarrow$GN-EnRML

Characteristics & Algorithms

- RML honors
  - Point 3: Focus on static parameters
- RML does not honor
  - Point 1: Updates based on ensemble
  - Point 2: Perform one complete run
- Solution
  - EnRML updates using an ensemble approximation to gradient: RML $\rightarrow$ EnRML
  - Minimize utilizing one full Gauss-Newton step: EnRML $\rightarrow$GN-EnRML
- Sim. GN-EnRML: Simultaneous scheme honoring 1-3
- Seq. GN-EnRML: Sequential scheme honoring 1-3

Analytical strategy

Goal: Understand the importance of seq. and sim. assimilation when combining data with different degrees of non-linearity

- Note: Analytical result exist for seq. vs sim. RML with linear data
- Strategy:
  - Define comparable variants of seq./sim. RML and EnKF/ES
  - Analyze differences between the methods
  - Extend linear RML result to new RML variants for combination of linear and non-linear data
  - Extend linear RML result for variants of EnKF/ES

Comparing Hi-EnKS & GN-EnRML

- GN-EnRML update:

$$m_j^a = m_j^f + C_m \tilde{G}^T \left[\tilde{G} C_m \tilde{G}^T + C_d \right]^{-1} \left(d_j - g\left(m_j^f \right) \right)$$

- Hi-EnKS parameter update

$$m_j^a = m_j^f + \tilde{C}_{mg} \left[\tilde{C}_{gg} + C_d \right]^{-1} \left(d_j - g\left(m_j^f \right) \right)$$

Comparing Hi-EnKS & GN-EnRML

- GN-EnRML update:

$$m_j^a = m_j^f + C_m \tilde{G}^T \left[\tilde{G} C_m \tilde{G}^T + C_d \right]^{-1} \left(d_j - g\left(m_j^f \right) \right)$$

- Hi-EnKS parameter update

$$m_j^a = m_j^f + \tilde{C}_{mg} \left[\tilde{C}_{gg} + C_d \right]^{-1} \left(d_j - g\left(m_j^f \right) \right)$$

- The two methods are equal if
  - $C_m \tilde{G}^T = \tilde{C}_{mg}$
  - $\tilde{G} C_m \tilde{G}^T = \tilde{C}_{gg}$

Comparing Hi-EnKS & GN-EnRML

$$\tilde{C}_m \tilde{G}^T = \tilde{C}_{mg}$$

- Ensemble gradient given by the pseudo inverse

$$\tilde{G} = \Delta d \Delta m^\dagger$$

- Ensemble covariance

$$\tilde{C}_m = \frac{1}{N_e - 1} \Delta m \Delta m^T$$

Comparing Hi-EnKS & GN-EnRML

$$\boxed{\tilde{C}_m \tilde{G}^T = \tilde{C}_{mg}}$$

- Ensemble gradient given by the pseudo inverse

$$\tilde{G} = \Delta d \Delta m^{\dagger}$$

- Ensemble covariance

$$\tilde{C}_m = \frac{1}{N_e - 1} \Delta m \Delta m^T$$

- Rewriting

$$\tilde{C}_m \tilde{G}^T = \frac{1}{N_e - 1} \Delta m \Delta m^T \left(\Delta m^{\dagger}\right)^T \Delta d^T$$

$$\Rightarrow \tilde{C}_m \tilde{G}^T = \frac{1}{N_e - 1} \Delta m \Delta d^T = \tilde{C}_{mg}$$

Comparing Hi-EnKS & GN-EnRML

$$\boxed{\tilde{G}\tilde{C}_m\tilde{G}^T = \tilde{C}_{gg}}$$

- Inserting for $\tilde{G}$ and $\tilde{C}_m$

$$\tilde{G}\tilde{C}_m\tilde{G}^T = \frac{1}{N_e - 1}\Delta d V_p V_p^T \Delta d^T$$

Comparing Hi-EnKS & GN-EnRML

$$\boxed{\tilde{G}\tilde{C}_m\tilde{G}^T = \tilde{C}_{gg}}$$

- Inserting for $\tilde{G}$ and $\tilde{C}_m$
$$\tilde{G}\tilde{C}_m\tilde{G}^T = \frac{1}{N_e - 1}\Delta d V_p V_p^T \Delta d^T$$
- $N_e \leq N_m \Longrightarrow V_p V_p^T = I \Longrightarrow \tilde{G}\tilde{C}_m\tilde{G}^T = \frac{1}{N_e-1}\Delta d \Delta d^T = \tilde{C}_{gg}$

Comparing Hi-EnKS & GN-EnRML

$$\tilde{G}\tilde{C}_m\tilde{G}^T = \tilde{C}_{gg}$$

- Inserting for $\tilde{G}$ and $\tilde{C}_m$

$$\tilde{G}\tilde{C}_m\tilde{G}^T = \frac{1}{N_e - 1}\Delta d V_p V_p^T \Delta d^T$$

- $N_e \leq N_m \implies V_p V_p^T = I \implies \tilde{G}\tilde{C}_m\tilde{G}^T = \frac{1}{N_e-1}\Delta d \Delta d^T = \tilde{C}_{gg}$

$$N_e \leq N_m \wedge \tilde{C}_m = \frac{1}{N_e-1}\Delta m \Delta m^T$$

Hi-EnKS = GN-EnRML

Comparing Hi-EnKS & GN-EnRML

$$\boxed{\tilde{G}\tilde{C}_m\tilde{G}^T = \tilde{C}_{gg}}$$

- Inserting for $\tilde{G}$ and $\tilde{C}_m$

$$\tilde{G}\tilde{C}_m\tilde{G}^T = \frac{1}{N_e - 1}\Delta d V_p V_p^T \Delta d^T$$

- $N_e \leq N_m \Longrightarrow V_p V_p^T = I \Longrightarrow \tilde{G}\tilde{C}_m\tilde{G}^T = \frac{1}{N_e - 1}\Delta d \Delta d^T = \tilde{C}_{gg}$

$$N_e \leq N_m \wedge \tilde{C}_m = \frac{1}{N_e - 1}\Delta m \Delta m^T$$

Hi-EnKS = GN-EnRML

- Same result when comparing ES & sim. GN-EnRML

Comparing Hi-EnKS & GN-EnRML

- For $N_e > N_m$ the difference between GN-EnRML and Hi-EnKS determined by

$$\tilde{C}_{gg} = \frac{1}{N_e - 1}\Delta d_t \Delta d_t \qquad \tilde{G}\tilde{C}_m\tilde{G}^T = \frac{1}{N_e - 1}\Delta d_p \Delta d_p^T$$

- Δd_t: True predicted data pertubation
- Δd_p: Predicted linearized data pertubation
  $\Delta d_p = \tilde{G}\Delta m = \Delta d_t V_p V_p^T$

Comparing Hi-EnKS & GN-EnRML

- For $N_e > N_m$ the difference between GN-EnRML and Hi-EnKS determined by

$$\tilde{C}_{gg} = \frac{1}{N_e - 1}\Delta d_t \Delta d_t \qquad \tilde{G}\tilde{C}_m\tilde{G}^T = \frac{1}{N_e - 1}\Delta d_p \Delta d_p^T$$

- Δd_t: True predicted data pertubation
- Δd_p: Predicted linearized data pertubation
  $\Delta d_p = \tilde{G}\Delta m = \Delta d_t V_p V_p^T$
- Difference depends on the non-linearity of the data:

$$\Delta d_t - \Delta d_p = \Delta e \left(I - V_p V_p^T \right)$$

- $\Delta e = e_j - \overline{e}$
- e_j: Truncation error

Analytical strategy

Goal: Understand the importance of seq. and sim. assimilation when combining data with different degrees of non-linearity

- Note: Analytical result exist for seq. vs sim. RML with linear data
- Strategy:
  - Define comparable variants of seq./sim. RML and EnKF/ES
  - Analyze differences between the methods
  - Extend linear RML result to new RML variants for combination of linear and non-linear data
  - Extend linear RML result for variants of EnKF/ES

Sequential & Simultaneous assimilation (GN-EnRML)

- Utilizing seq./sim. GN-EnRML extend linear RML result for combination of linear and non-linear data
- Choice of covariance update for seq. GN-EnRML

1. $C_m^a = C_m^f - C_m^f G^T \left[G C_m^f G^T + C_d \right]^{-1} G C_m^f$
2. $\tilde{C}_m^a = \frac{1}{N_e - 1} \Delta m \Delta m^T$

Sequential & Simultaneous assimilation (GN-EnRML)

- Utilizing seq./sim. GN-EnRML extend linear RML result for combination of linear and non-linear data
- Choice of covariance update for seq. GN-EnRML

1. $C_m^a = C_m^f - C_m^f G^T \left[G C_m^f G^T + C_d \right]^{-1} G C_m^f$
2. $\tilde{C}_m^a = \frac{1}{N_e - 1} \Delta m \Delta m^T$

- We choose 1 for seq. GN-EnRML

Sequential & Simultaneous assimilation (GN-EnRML)

- Compare sim. assimilation with seq. assimilation of two data groups

Sequential & Simultaneous assimilation (GN-EnRML)

- Compare sim. assimilation with seq. assimilation of two data groups

Case 1:

$$\begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} g(m_j) \\ Gm_j \end{bmatrix}$$

After assimilation:

$$m_j^{seq} = m_j^{sim}$$

Sequential & Simultaneous assimilation (GN-EnRML)

- Compare sim. assimilation with seq. assimilation of two data groups

Case 1:

$$\begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} g(m_j) \\ Gm_j \end{bmatrix}$$

After assimilation:

$$m_j^{seq} = m_j^{sim}$$

Case 2:

$$\begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} Gm_j \\ g(m_j) \end{bmatrix}$$

After assimilation:

$$m_j^{seq} \neq m_j^{sim}$$

Sequential & Simultaneous assimilation (GN-EnRML)

- Compare sim. assimilation with seq. assimilation of two data groups

Case 1:

$$\begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} g(m_j) \\ Gm_j \end{bmatrix}$$

After assimilation:

$$m_j^{seq} = m_j^{sim}$$

Case 2:

$$\begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} Gm_j \\ g(m_j) \end{bmatrix}$$

After assimilation:

$$m_j^{seq} \neq m_j^{sim}$$

- GN-EnRML: seq. $=$ sim. for non-linear $\rightarrow$ linear
- GN-EnRML: seq. $\neq$ sim. for linear $\rightarrow$ non-linear

Analytical strategy

Goal: Understand the importance of seq. and sim. assimilation when combining data with different degrees of non-linearity

- Note: Analytical result exist for seq. vs sim. RML with linear data
- Strategy:
  - Define comparable variants of seq./sim. RML and EnKF/ES
  - Analyze differences between the methods
  - Extend linear RML result to new RML variants for combination of linear and non-linear data
  - Extend linear RML result for variants of EnKF/ES

Analytical results

- Remember:

$$N_e \leq N_m \wedge \tilde{C}_m = \frac{1}{N_e - 1} \Delta m \Delta m^T$$

Hi-EnKS = GN-EnRML

Analytical results

- Remember:

$$N_e \leq N_m \wedge \tilde{C}_m = \frac{1}{N_e - 1} \Delta m \Delta m^T$$

$$\text{Hi-EnKS} = \text{GN-EnRML}$$

- Seq. GN-EnRML updated covariance as:
$C_m^a = C_m^f - C_m^f G^T \left[G C_m^f G^T + C_d \right]^{-1} G C_m^f$

Analytical results

- Remember:

$$N_e \leq N_m \wedge \tilde{C}_m = \frac{1}{N_e - 1} \Delta m \Delta m^T$$

Hi-EnKS = GN-EnRML

- Seq. GN-EnRML updated covariance as:
  $C_m^a = C_m^f - C_m^f G^T \left[G C_m^f G^T + C_d \right]^{-1} G C_m^f$
- Need: $N_e \to \infty$ for $C_m^a = \tilde{C}_m$
  - GN-EnRML $\neq$ Hi-EnKS

Analytical results

- Remember:

$$N_e \leq N_m \wedge \tilde{C}_m = \frac{1}{N_e - 1} \Delta m \Delta m^T$$

Hi-EnKS = GN-EnRML

- Seq. GN-EnRML updated covariance as:
  $C_m^a = C_m^f - C_m^f G^T \left[G C_m^f G^T + C_d \right]^{-1} G C_m^f$
- Need: $N_e \rightarrow \infty$ for $C_m^a = \tilde{C}_m$
  - GN-EnRML $\neq$ Hi-EnKS
- Difference between methods depend on non-linearity
  - Perform numerical studies for Hi-EnKS with weakly non-linear data

Numerical studies

- Difference between Hi-EnKS and GN-EnRML varies with Δe
- Numerical study to investigate:
  - Optimal assimilation strategy
  - GN-EnRML result valid for Hi-EnKS

Numerical studies

- Difference between Hi-EnKS and GN-EnRML varies with Δe
- Numerical study to investigate:
  - Optimal assimilation strategy
  - GN-EnRML result valid for Hi-EnKS
- Numerical experiments
  - Univariate:
    - Simple forward model
    - One linear data group
    - One non-linear data group
  - Multivariate
    - 1D Reservoir
    - One weakly non-linear data group
    - One data group with stronger non-linearity

Numerical studies

- Difference between Hi-EnKS and GN-EnRML varies with Δe
- Numerical study to investigate:
  - Optimal assimilation strategy
  - GN-EnRML result valid for Hi-EnKS
- Numerical experiments
  - Univariate:
    - Simple forward model
    - One linear data group
    - One non-linear data group
  - Multivariate
    - 1D Reservoir
    - One weakly non-linear data group
    - One data group with stronger non-linearity
- Assess quality of Hi-EnKS/ES by Kullback-Leibler Divergence (KLD) to McMC samples
  - Nearest Neighbor kernel density estimator

Univariate example

- Simple forward model

$$d_i = m^{r_i}$$

- Assimilate $d_1 \rightarrow d_2$ and $d_2 \rightarrow d_1$ with Hi-EnKS, and assimilate simultaneously with ES

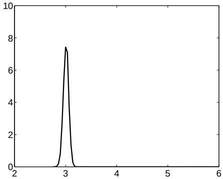
m_{ref}	3
Prior mean	8
Prior Var	1

r_1	1
r_2	2
$\sigma^2_{d_{1/2}}$	0.1

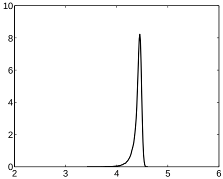
Ensemble size	1×10^5
McMC iterations	1×10^5
McMC acceptance rate	0.2267

Table: Numerical details

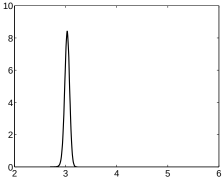
Univariate example



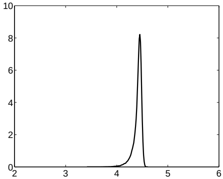
McMC



ES (1)

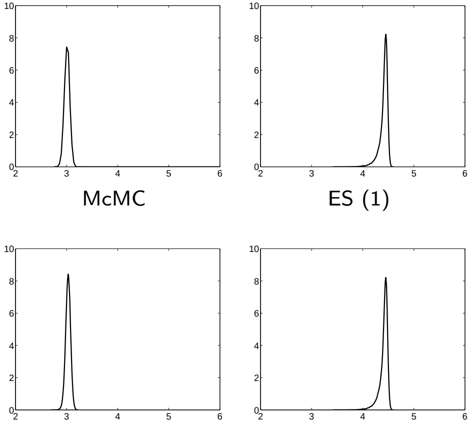


Hi-EnKS $d_1 \rightarrow d_2$ (2)



Hi-EnKS $d_2 \rightarrow d_1$ (3)

Univariate example



McMC | ES (1)

Hi-EnKS $d_1 \to d_2$ (2) | Hi-EnKS $d_2 \to d_1$ (3)

$KLD_{(1)}$	11.96
$KLD_{(2)}$	0.079
$KLD_{(3)}$	11.96

Table: Univariate results

Multivariate example

- 1D reservoir consisting of 31 unknown parameters
- Two data groups
  - d_1 : 6 measurements of log(perm) field. Measurements made at wells marked Hard obs

$$d_1 = \log(perm)^{1.2}$$

  - d_2 : 6 pressure observations, made at a single well marked Pres obs

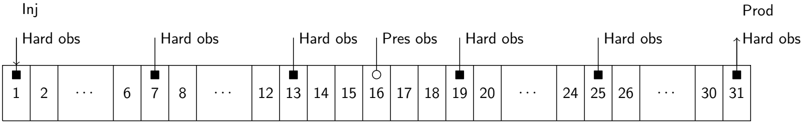


Figure: Grid blocks & well placement

Multivariate example

Ensemble size	5×10^4
McMC proposals	5×10^5
McMC acceptance rate	0.238

Table: Numerical details

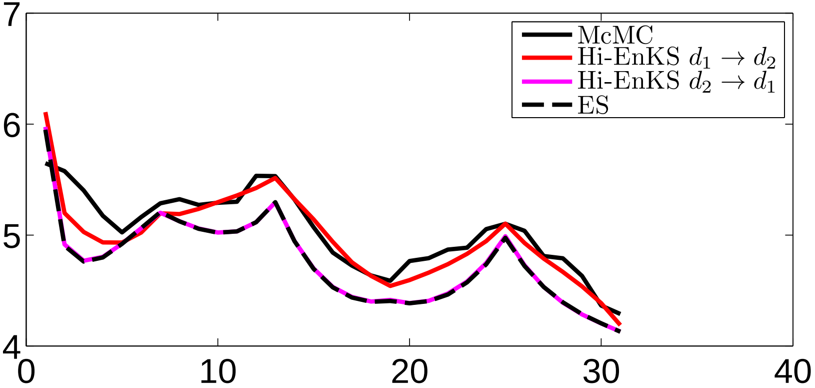


Figure: Mean values

Multivariate example

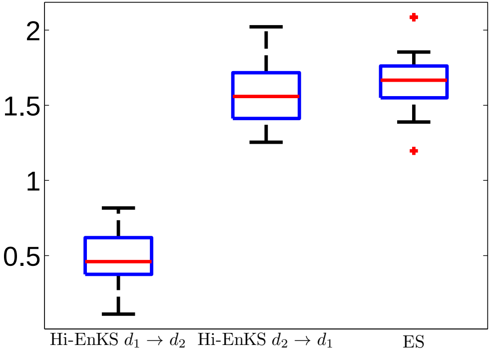


Figure: KLD

$\overline{\text{KLD}}_{d_1 \to d_2}$	0.48
$\overline{\text{KLD}}_{d_2 \to d_1}$	1.58
$\overline{\text{KLD}}_{ES}$	1.66

Table: Multivariate results

Conclusions

Analysis shows:

- GN-EnRML: seq. = sim. for non-linear data before linear data
- For $N_e \leq N_m \wedge \tilde{C}_m = \frac{1}{N_e - 1} \Delta m \Delta m^T$: GN-EnRML = Hi-EnKS
- For $N_e > N_m$: (GN-EnRML - Hi-EnKS) $\propto \Delta e$

Conclusions

Analysis shows:

- GN-EnRML: seq. = sim. for non-linear data before linear data
- For $N_e \leq N_m \wedge \tilde{C}_m = \frac{1}{N_e - 1} \Delta m \Delta m^T$: GN-EnRML = Hi-EnKS
- For $N_e > N_m$: (GN-EnRML - Hi-EnKS) $\propto \Delta e$

Numerical experiments show:

- Hi-EnKS: seq. = sim. non-linear data before linear data
- Data with weakest non-linearity first:
  - Best univariate and multivariate mean
  - Best univariate and multivariate KLD

Thank you