Inversion, history matching, clustering and linear algebra

Andrew R. Conn
arconn@us.ibm.com

IBM T.J. Watson Research Center

January, 2016, 4-8, Tenth U.S.-Mexico Workshop on Optimization and its Applications,

Merida, Mexico

Automatic History Matching

IBM

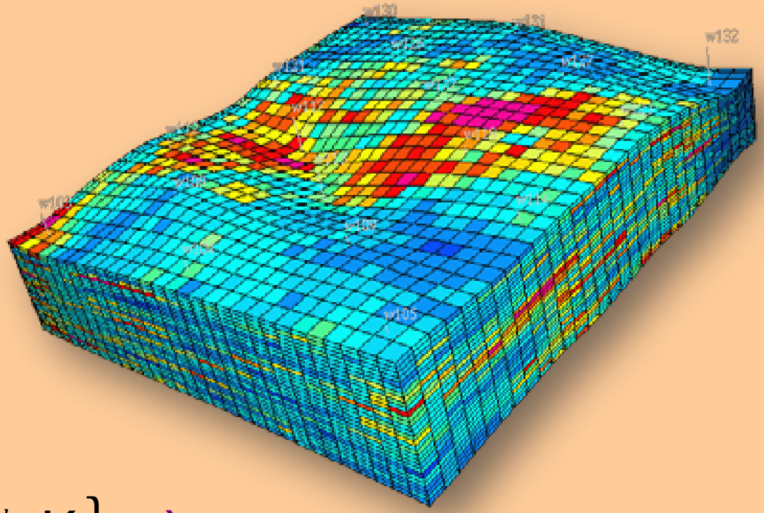
§ History matching

$$\hat{m} = \arg\min_{\hat{m}} \underbrace{\left\| F(m,y) - d(y) \right\|^2}_{\text{data misfit}} + \underbrace{R(m)}_{\text{regularization}}$$

$$\text{s.t.} \quad \underbrace{d(y) = F(m,y) + \epsilon + \mu}_{\text{constraints}}$$

§ Static model parameters

$$m_s \equiv \{k, \phi, \ldots\}$$



§ Dynamic model state

$$m_d \equiv \{p, S, \ldots\}$$

§ Observed data

$$d = \{p^w, S^w, q^{p,w}, \ldots\} + \epsilon$$

§ Simulation (observation)

$$F(m) = \{p, S, q^p, \ldots\}$$

Mathematician's point of view

- Typically an undetermined least-squares problem
- "Classical" fit based on local well data. Many good fits.
- The vast null space means the problem is intrinsically ill-posed
- Our purpose is to predict the future based upon (past) data.
- Very few of the fits will do this successfully
- Need to make the problem less underdetermined

Two obvious things one should do

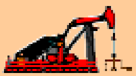
- Listen to and incorporate what the geologists can tell us
- Use more global data than just the well-logs

If relevant and possible

- Use smart linear algebra and updating

Integration of 4D seismic data into reservoir models

Use more global data than just the well-logs



	spatial resolution		temporal resolution	alignment
	areal	vertical		
production data	low	high	high	geological layers – simulation grid
seismic data	High	low	low	seismic trace grid

Compensate for the spatial sparsity of the production data via seismic information

Furthermore we can exploit existing adjoint functionality of modern simulators by transforming the seismic data to “equivalent” pseudo-wells

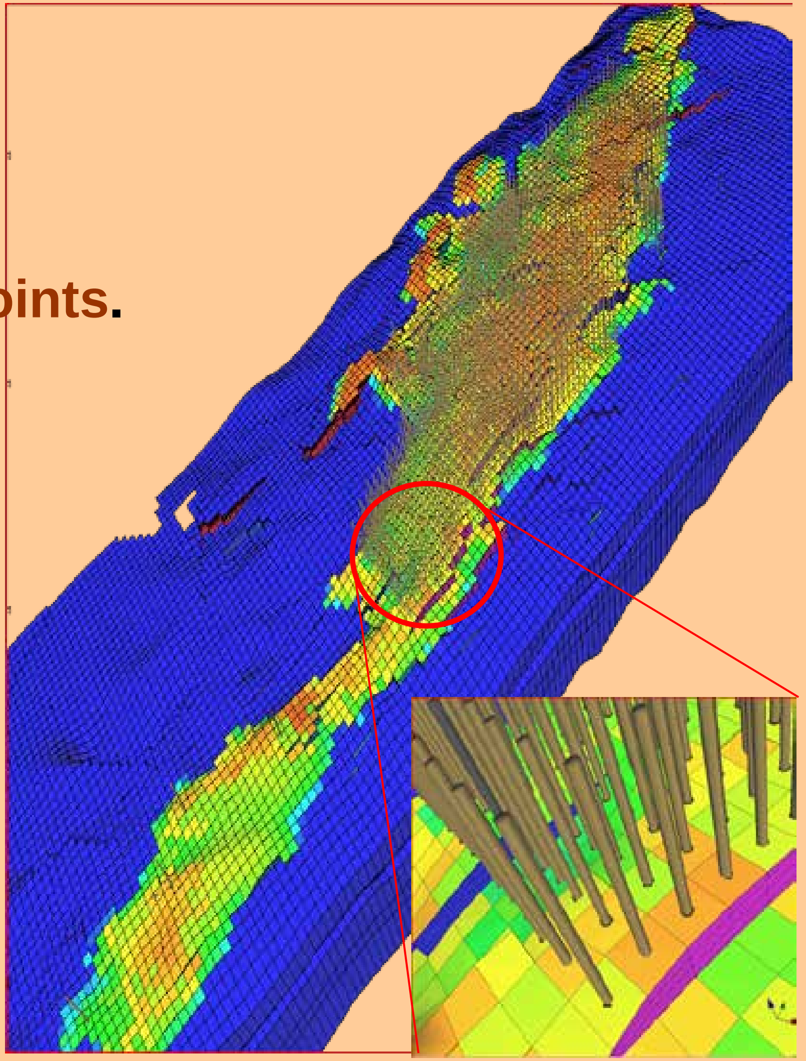
IMPORTANT FOR THE OPTIMIZATION

First Trick : change into something you can solve

Adjoint (Derivative) Based History Matching with Virtual Wells

Advanced industrial simulators offer adjoint /derivative computation capability for wells

- **Idea: Use virtual wells that mimic the**
- **(interpreted) saturation measurement**
- **of seismic information. So we have adjoints.**



Impact on simulated fluid flows can be marginalized by:

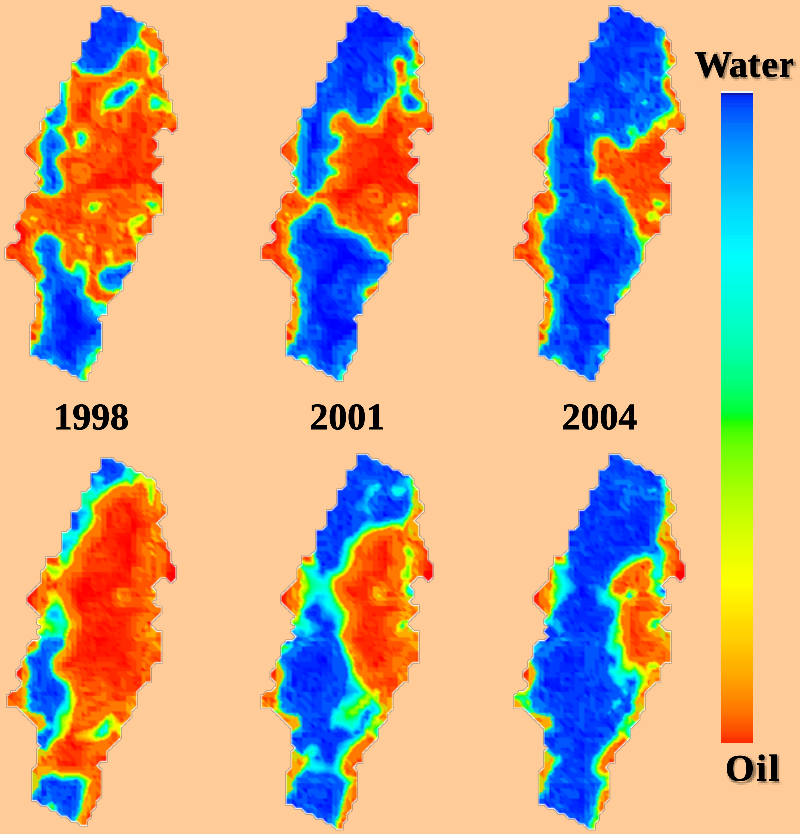
- **Volumetric sample of insignificant size – does not interfere with fluids flow simulation**
- **Using a very short time-step when a saturation 'measurement' is conducted**
- **Shutting in virtual-wells when no measurement is taken**

FIELD RESULTS

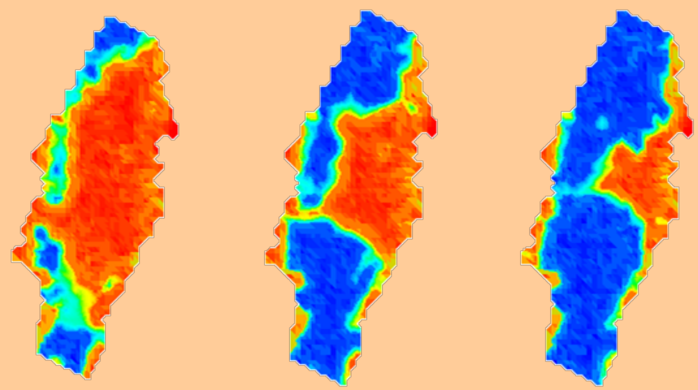
COMBINED PRODUCTION DATA & 4D SEISMIC

- Combined History Matching of production and 4D seismic leads to significant improvement in model performance (x10 improved match)
- Highly efficient workflow (hours replacing months)
- Understanding of boundaries and reservoir connectivity

4D seismic



Prediction from initial geological model

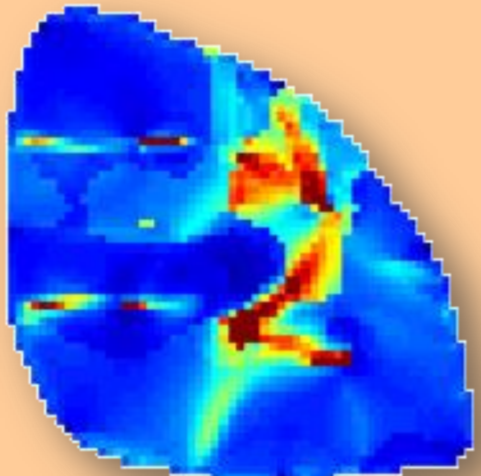


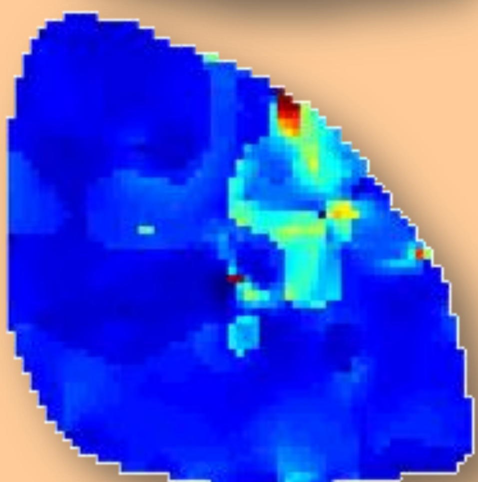
Combined production - 4D History Match

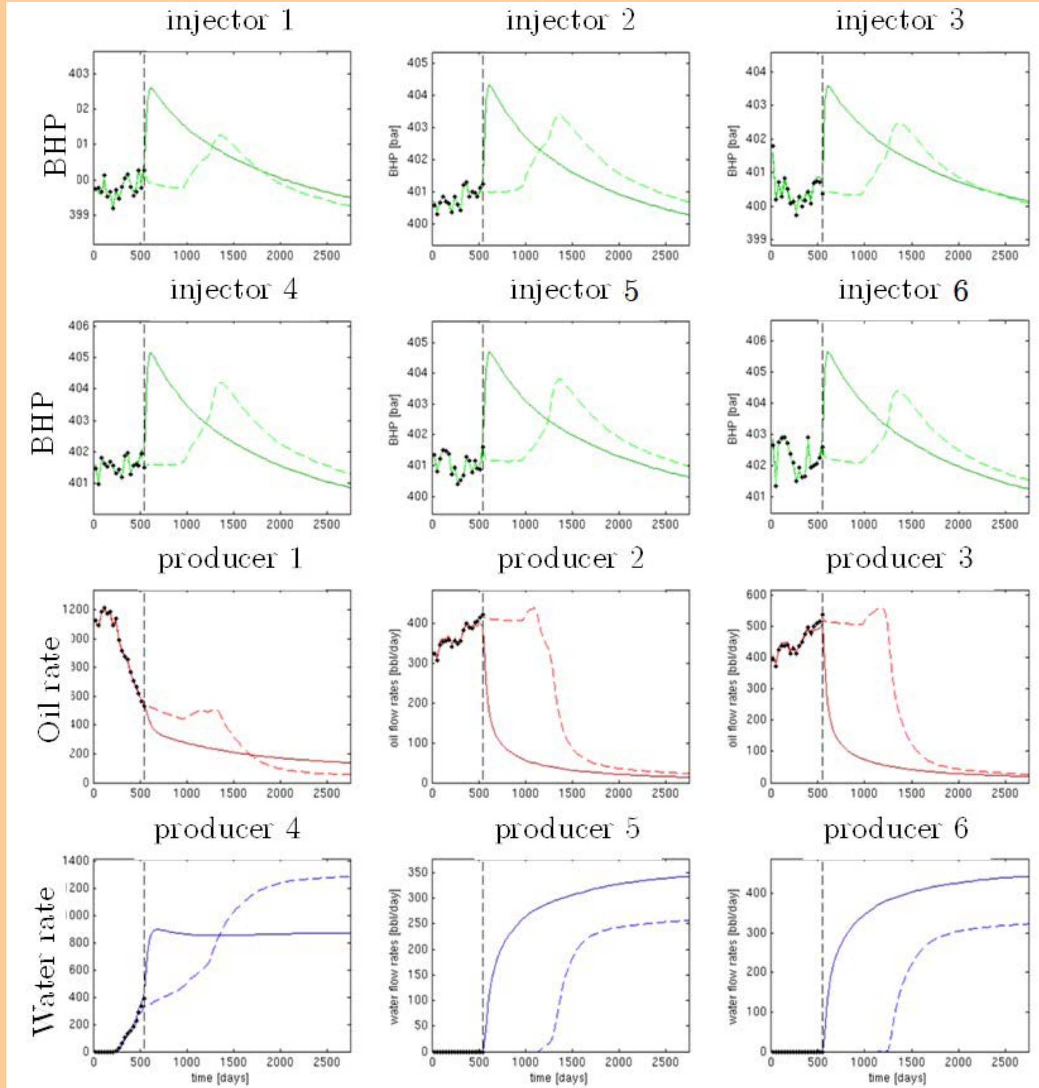
The Practical effect of an underdetermined problem

DECISION RELEVANT PRIOR SAMPLING – THE FUTURE

- … with very different predictions
  and predicted Net Present Value

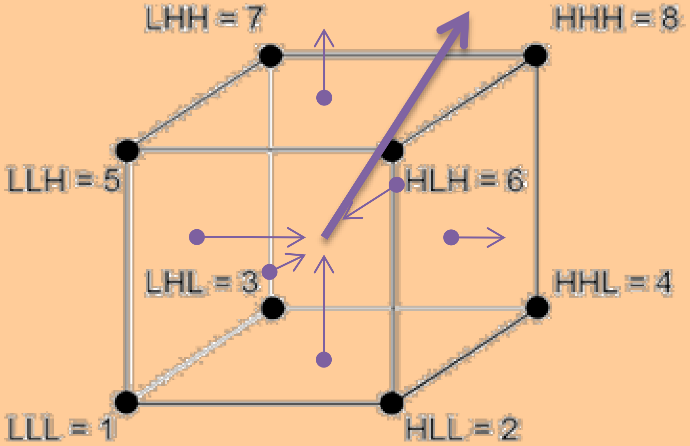






REDUCED ORDER MASS FLUX REPRESENTATION

§ For each realization mass flux vector fields is computed $\dot{F}_i(x, y, z; t)$



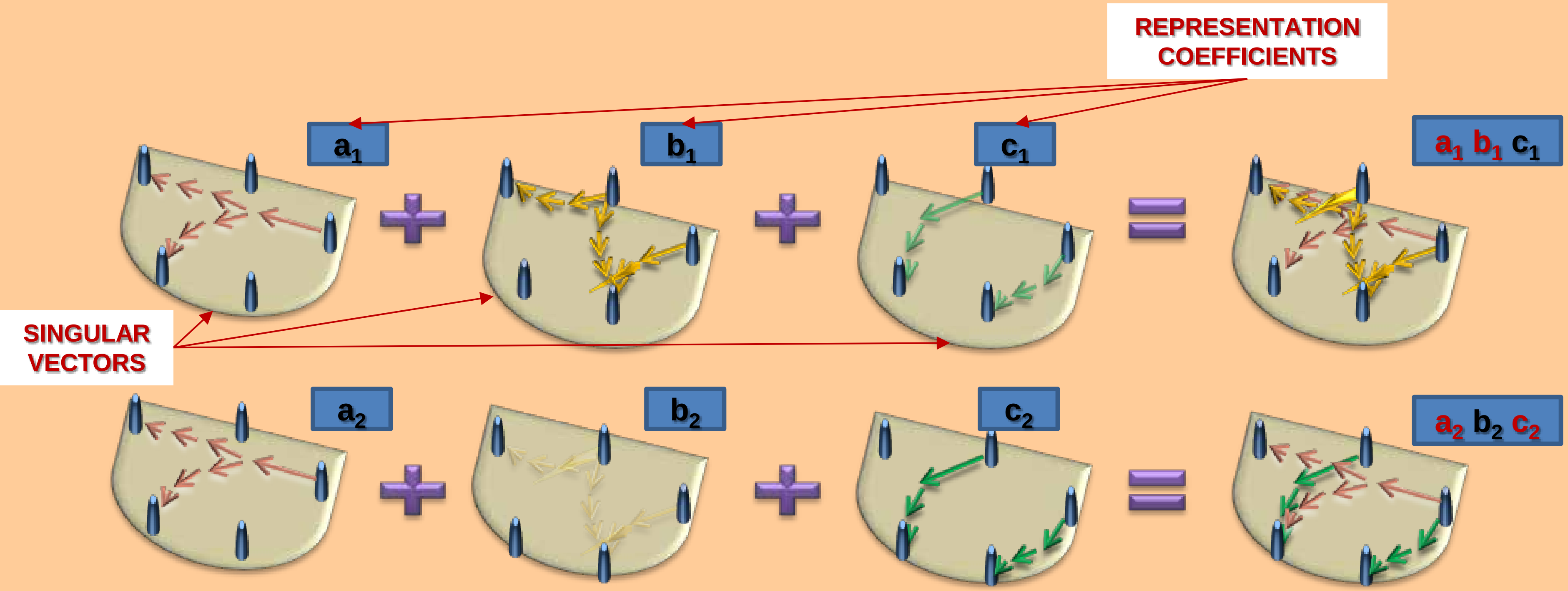
- **Fluxes capture chief characteristics of dynamics, yet, 4D vector fields are of a large dimension (3 x grid cells x time steps)**
- **Clustering in such large dimensional space is intractable**
- **Instead, reduced order representation of each flux is considered**

Defining measures of similarity

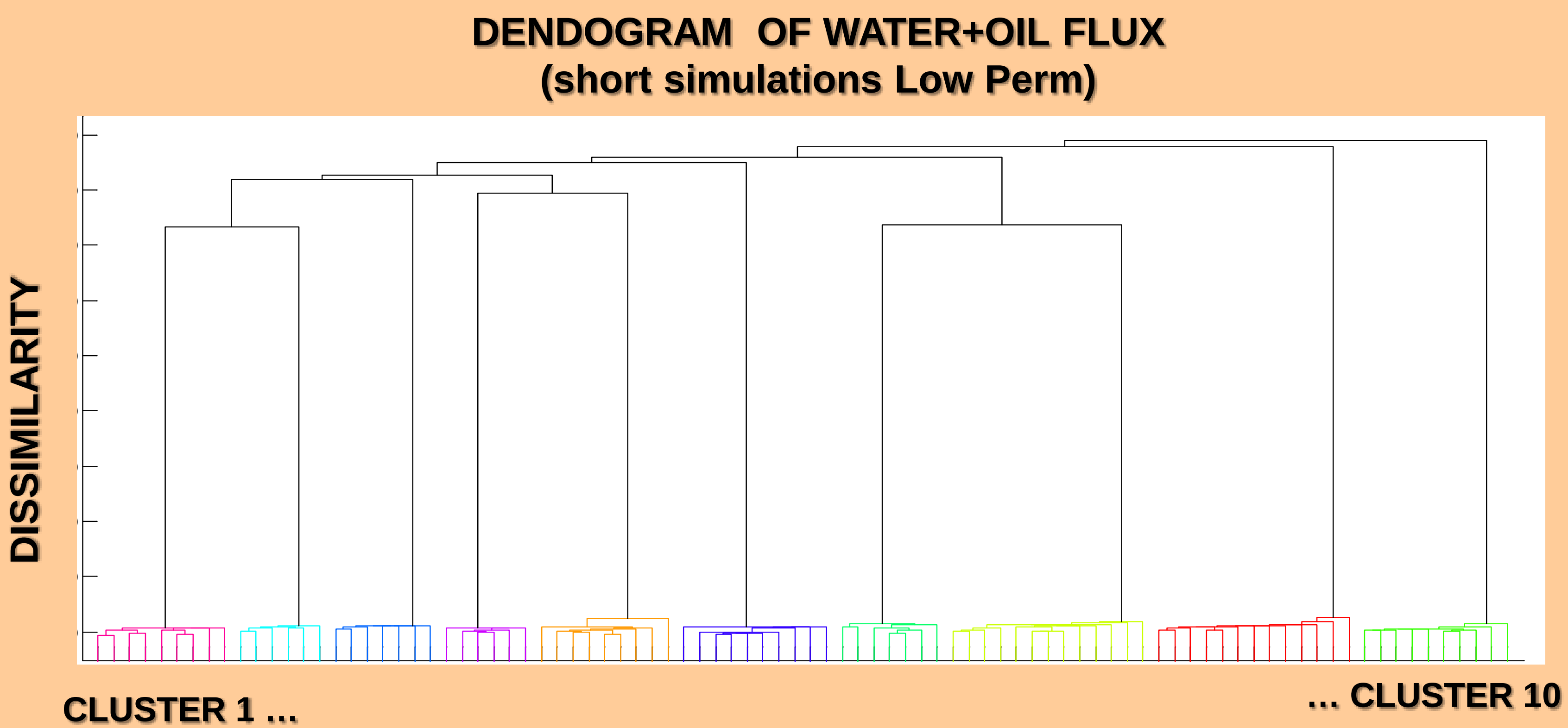
Mass Flux Representation in Reduced Space

- **Singular value decomposition of vector fields from all realizations enables reduced order representation**

$$USV^{T} = \left[\vec{F}_1(x,y,z;t), \vec{F}_2(x,y,z;t), \ldots, \vec{F}_n(x,y,z;t)\right]$$



Assessing Clustering Results

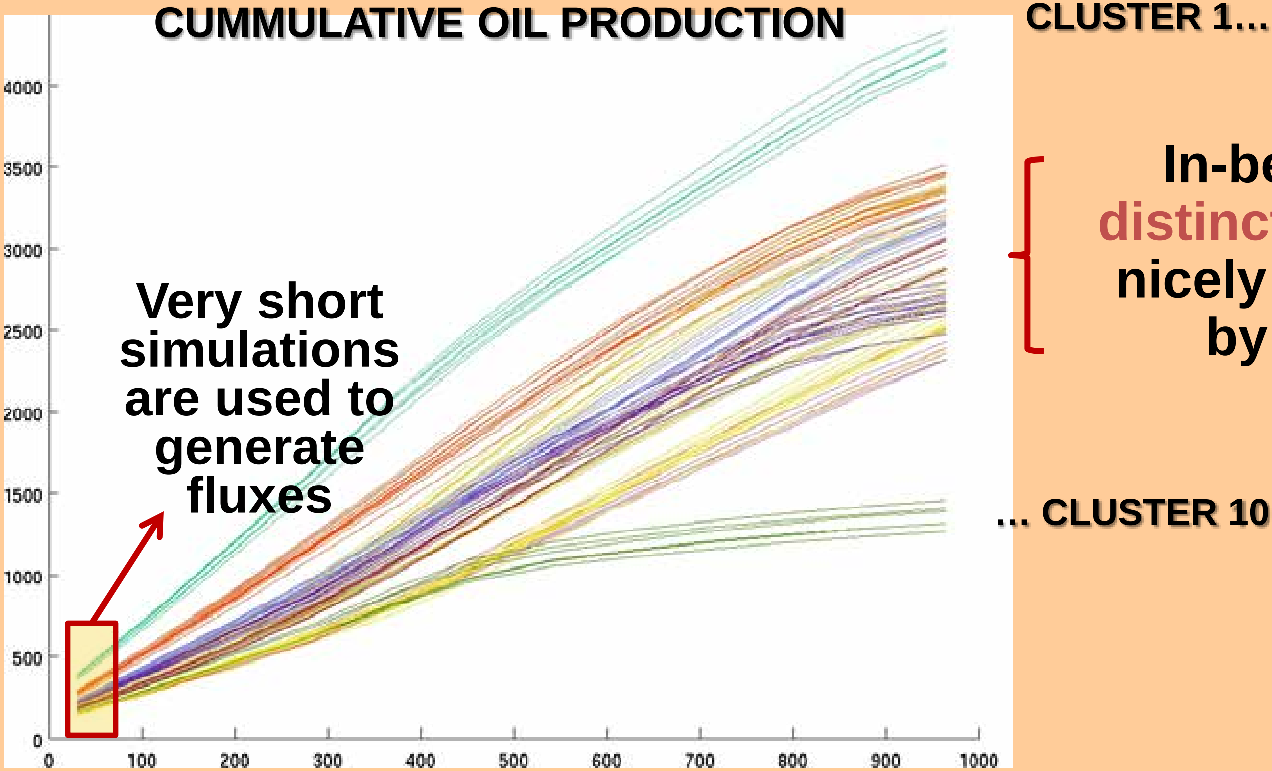


- Flux clustering pick up complete spectrum of training rock models

- Big question! do these clusters provide different production scenarios ?

Assessing Clustering Results

Dendograms based on OIL+WAT fluxes (Low Perm)



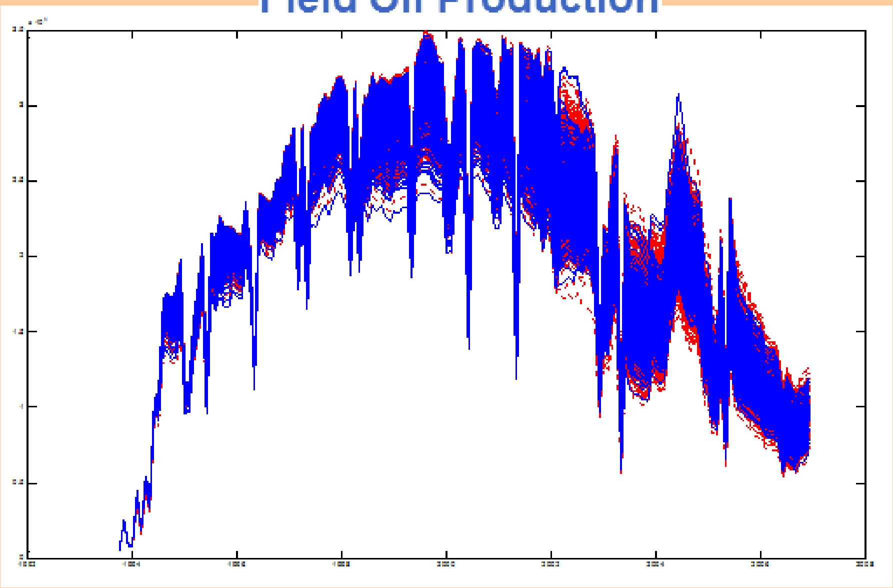
In-between **flow-distinct scenarios** are nicely discriminated by clustering

- A very **narrow window** of time (**well modulations are key**) is used and still we're able to pick up **long-term trends** in production data
- **Representatives** can now be extracted

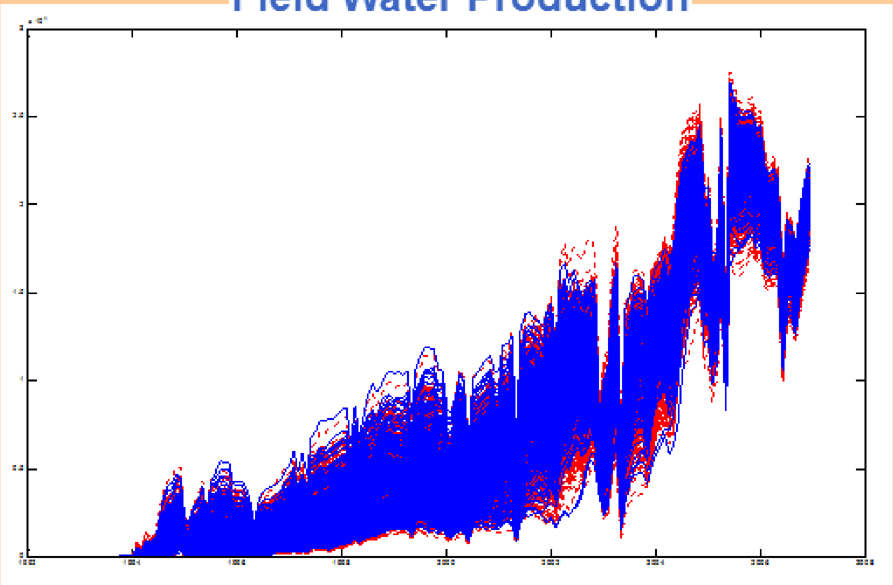
FLOW RELEVANT MODELS SAMPLING NORWEGIAN OFF-SHORE FIELD – SAMPLES PRESERVING UNCERTAINTY

IBM

Field Oil Production



Field Water Production



3,125 Realizations

300 Realizations – from the clusters

- Size of initial ensemble can readily be reduced by orders of magnitude
- Each representative can be regarded as a sample from a density function
- This density function can further be used for History Matching, model maturation
- Practicality is not compromised as no full simulations are performed

AUGMENTED SINGULAR VALUE DECOMPOSITION (SVD)

- Let A be an augmentation of the matrices the matrices $A_1 \in \mathbb{R}^{m_1 \times n}, A_2 \in \mathbb{R}^{m_2 \times n}$

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \in \mathbb{R}^{(m_1+m_2) \times n}$$

- Let the SVD decomposition of these matrices be given by:

$$A_1 = U_1 S_1 V_1^*$$

$$A_2 = U_2 S_2 V_2^*$$

with $U_i \in \mathbb{R}^{m_i \times n}, S_i \in \mathbb{R}^{n \times n}, V_i \in \mathbb{R}^{n \times n}$

- We seek the decomposition $A = USV^*$ of the augmented matrix A

AUGMENTED SVD - FORMULATION

- By definition

$$\underbrace{A^* A}_{K} V = \left(VS^* U^*\right)\left(USV^*\right)V = VS^2$$

- Starting with

$$\underbrace{A^* A}_{K} V = \begin{bmatrix} A_1^* & A_2^* \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} V = \begin{bmatrix} \underbrace{A_1^* A_1}_{M_1} + \underbrace{A_2^* A_2}_{M_2} \end{bmatrix} V = VS^2$$

- Then solve the (relatively small) eigen-problem $KV = VS^2$

- U is then be given by $U = AVS^{-1} = \left(USV^*\right)VS^{-1}$

In our context we can ignore the reduction in stability

AUGMENTED SVD - FORMULATION

- Note that here, we can save some computation by utilizing the small n×n (number of columns) product

$$A_1^* A_1 = M_1$$

from the previous run, and therefore, we retain the product

$$A_2^* A_2 = M_2$$

for future use

- This process can be repeated further giving $M_1, M_2, \ldots, M_k$

$$\left[M_1 + M_2 + \ldots + M_k \right] V = V S^2$$

MULTI-LEVEL DISTRIBUTED REDUCED SPANNING SET

- Let us assume that a set of $n < m$ model realizations $A \in \mathbb{R}^{m \times n}$

- Further assume their effective rank k is relatively small $k \ll n$

$$\left\| A - U^{(k)} S^{(k)} V^{(k)T} \right\|_2 \le \delta_k$$

- Partition A into s subsets for which we can effectively compute their SVD

$$A = \left[A_1, A_2, \ldots, A_s \right]$$

- SVD of each can be computed in parallel

$$U_1 S_1 V_1^T = A_1, \quad U_2 S_2 V_2^T = A_2, \quad \ldots, \quad U_s S_s V_s^T = A_s$$

- Given singular values, we select the top singular entries

$$\sum_i k_i = \kappa_s \ge \kappa$$

MULTI-LEVEL DISTRIBUTED REDUCED SPANNING SET

- Re-orthogonalize the union of the selected SVs

$$\left[U_1^{(k_1)} S_1^{(k_1)} V_1^{(k_1)T}, U_2^{(k_2)} S_2^{(k_2)} V_2^{(k_2)T}, \ldots, U_s^{(k_s)} S_s^{(k_s)} V_s^{(k_s)T} \right]$$

- 2nd truncation can be performed now

- The output would be $k_T \leq k_s$ ordered spanning vectors

- If needed, randomly mix the remaining vectors for further distributed processing

- The process is repeated until a sufficiently small set is obtained

- Finding a spanning set is a key problem for a broad range of numerical algorithms but for large scale matrices it is computationally intensive [of the order of min(mn^2, m^2n) for an m x n matrix] or even unattainable

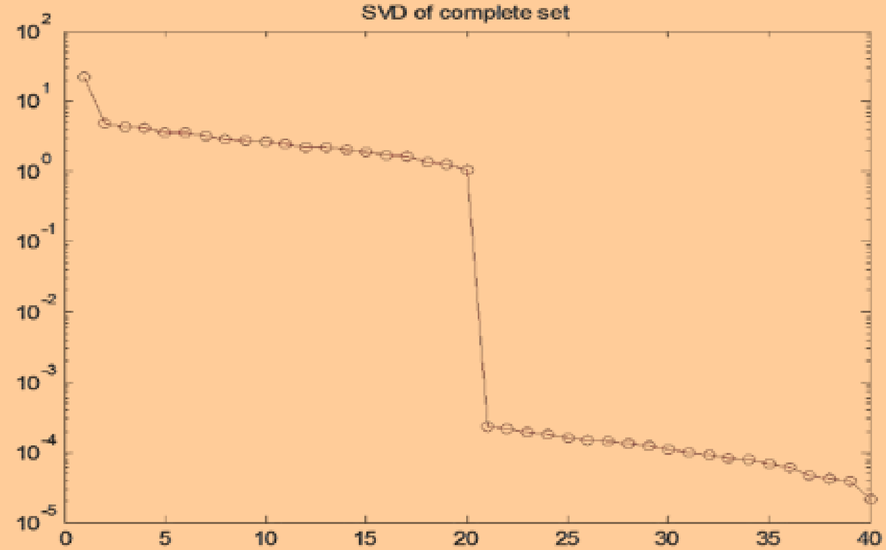
MULTI-LEVEL DISTRIBUTED SPANNING SET – TEST CASE

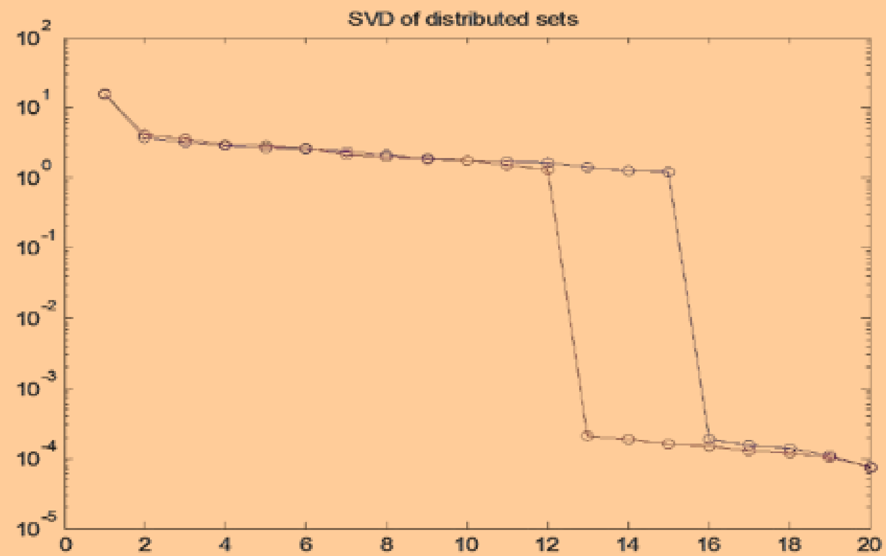
§ A set is constructed of 50x20 random vectors

§ Variability of additional 20 entries is simulated via noisy linear combination of the 50x20 set

§ 50x40 set was split into two 50x20 sets

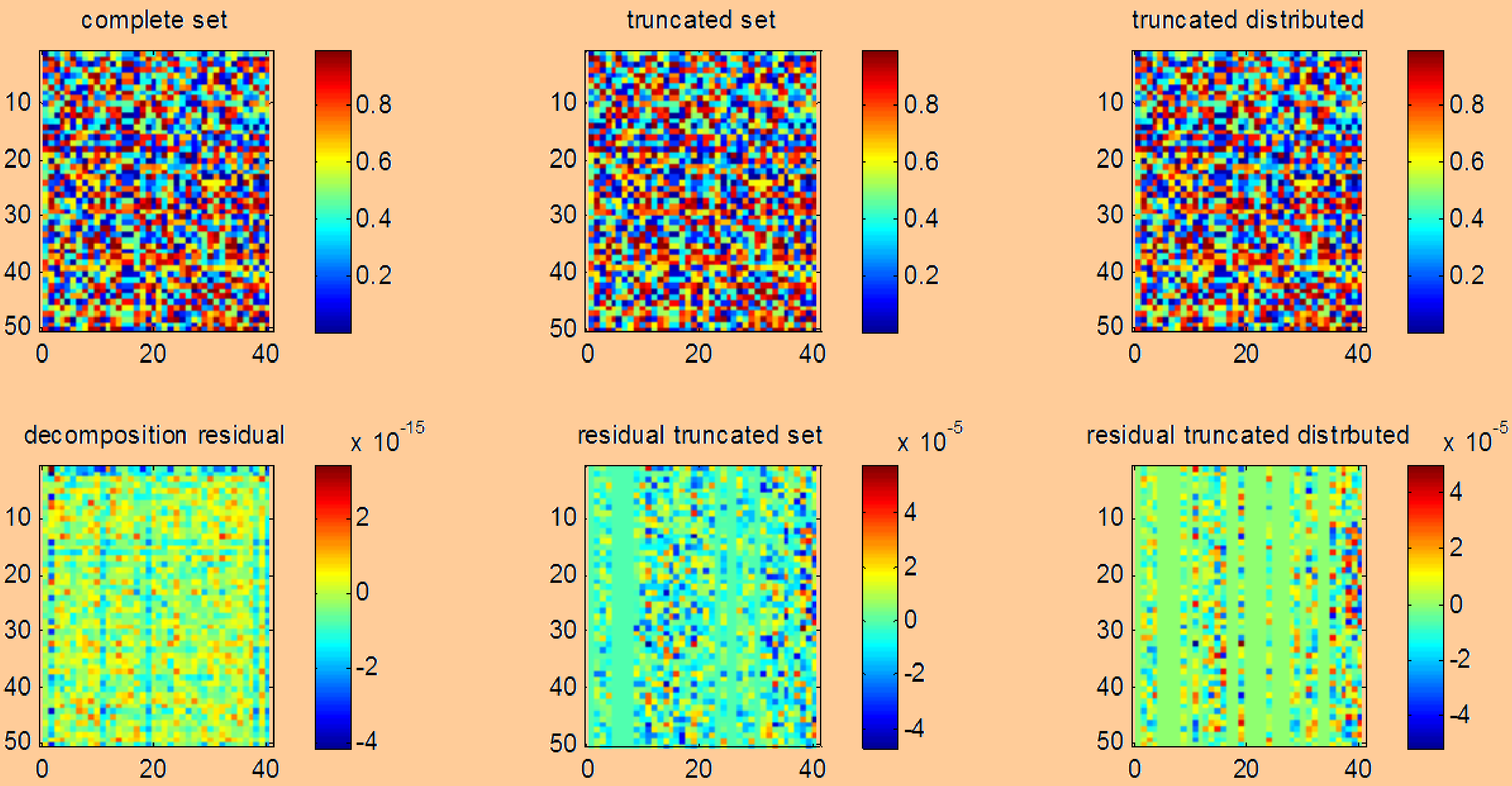
§ More than 10 SVDs were retain from each set





MULTI-LEVEL DISTRIBUTED SPANNING SET TRUNCATION ERROR

§ Following independent SVD and composition



Four Fundamental Unproved Theorems:

Asymptotics are rarely seen in practise but the best methods in theory are the best in practise.

A sensible person normally gives up on determining global optima. (So a sensible person doesn't try to solve MINLPs ???????)

It is always better to obtain and use derivatives if you can.

Simulated Annealing, Genetic Algorithms etc are usually for the ignorant or the desperate.

Four Fundamental Unproved Theorems:

Asymptotics are rarely seen in practise but the best methods in theory are the best in practise.

A sensible person normally gives up on determining global optima.
(So a sensible person doesn't try to solve MINLPs ???????)

It is always better to obtain and use derivatives if you can.

Simulated Annealing, Genetic Algorithms etc are usually for the ignorant or the desperate.

Four Fundamental Unproved Theorems:

Asymptotics are rarely seen in practise but the best methods in theory are the best in practise.

A sensible person normally gives up on determining global optima. (So a sensible person doesn't try to solve MINLPs ???????)

It is always better to obtain and use derivatives if you can.

Simulated Annealing, Genetic Algorithms etc are usually for the ignorant or the desperate.

Four Fundamental Unproved Theorems:

Asymptotics are rarely seen in practise but the best methods in theory are the best in practise.

A sensible person normally gives up on determining global optima.
(So a sensible person doesn't try to solve MINLPs ???????)

It is always better to obtain and use derivatives if you can.

Simulated Annealing, Genetic Algorithms etc are usually for the ignorant or the desperate.

Four Fundamental Unproved Theorems:

Asymptotics are rarely seen in practise but the best methods in theory are the best in practise.

A sensible person normally gives up on determining global optima.
(So a sensible person doesn't try to solve MINLPs ???????)

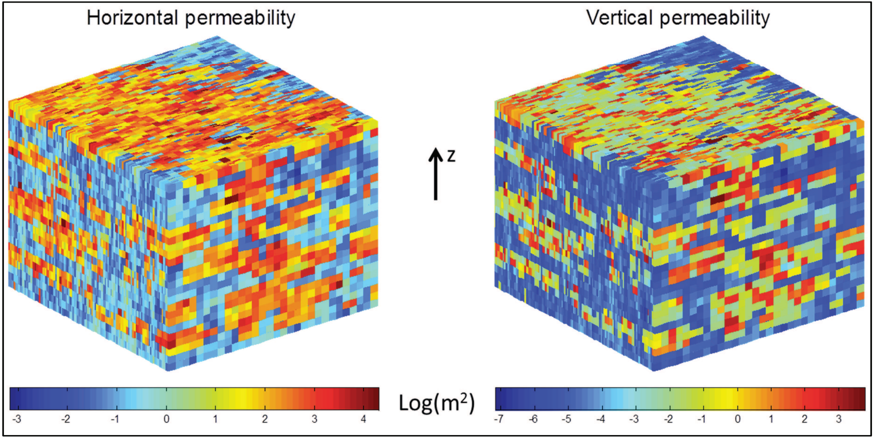
It is always better to obtain and use derivatives if you can.

Simulated Annealing, Genetic Algorithms etc are usually for the ignorant or the desperate.

Numerical Results: History Matching

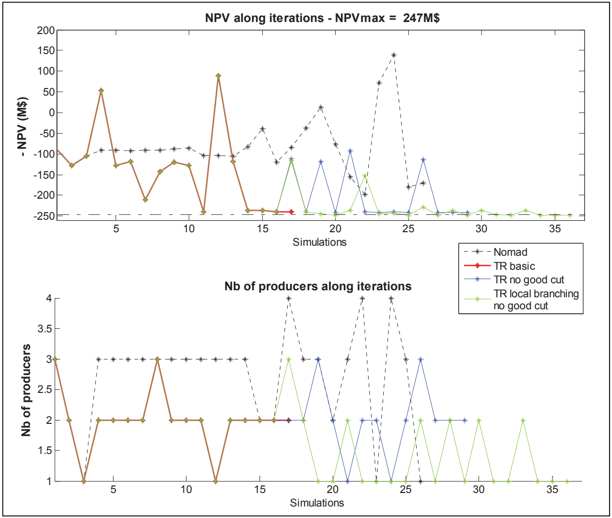
50 layers of 2′ with 60 × 220 cells 20′ × 10′
Up-scaled to 30 × 110 × 25 cells of 80′ × 40′ × 4′
10 yrs production: 1 injector well, 1 — 4 producers.

Optimize the number of wells and their locations to maximize the NPV of the field.



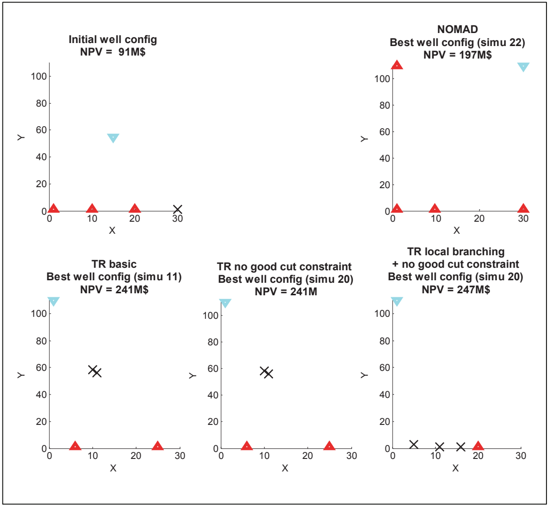
Numerical Results (continued)

Number of variables being set is 14 continuous and 4 binary variables



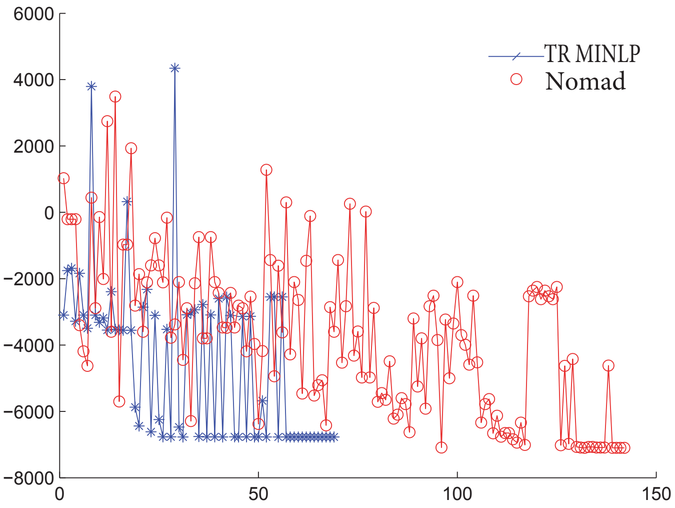
Numerical Results Compare NOMAD solutions & ours

Run with 3 different tunings. The initial configuration is displayed at top left.



Numerical Results (continued)

Number of variables being set is 4 continuous and 8 binary variables



Numerical Results (continued)

Number of variables being set is 4 continuous and 8 binary variables