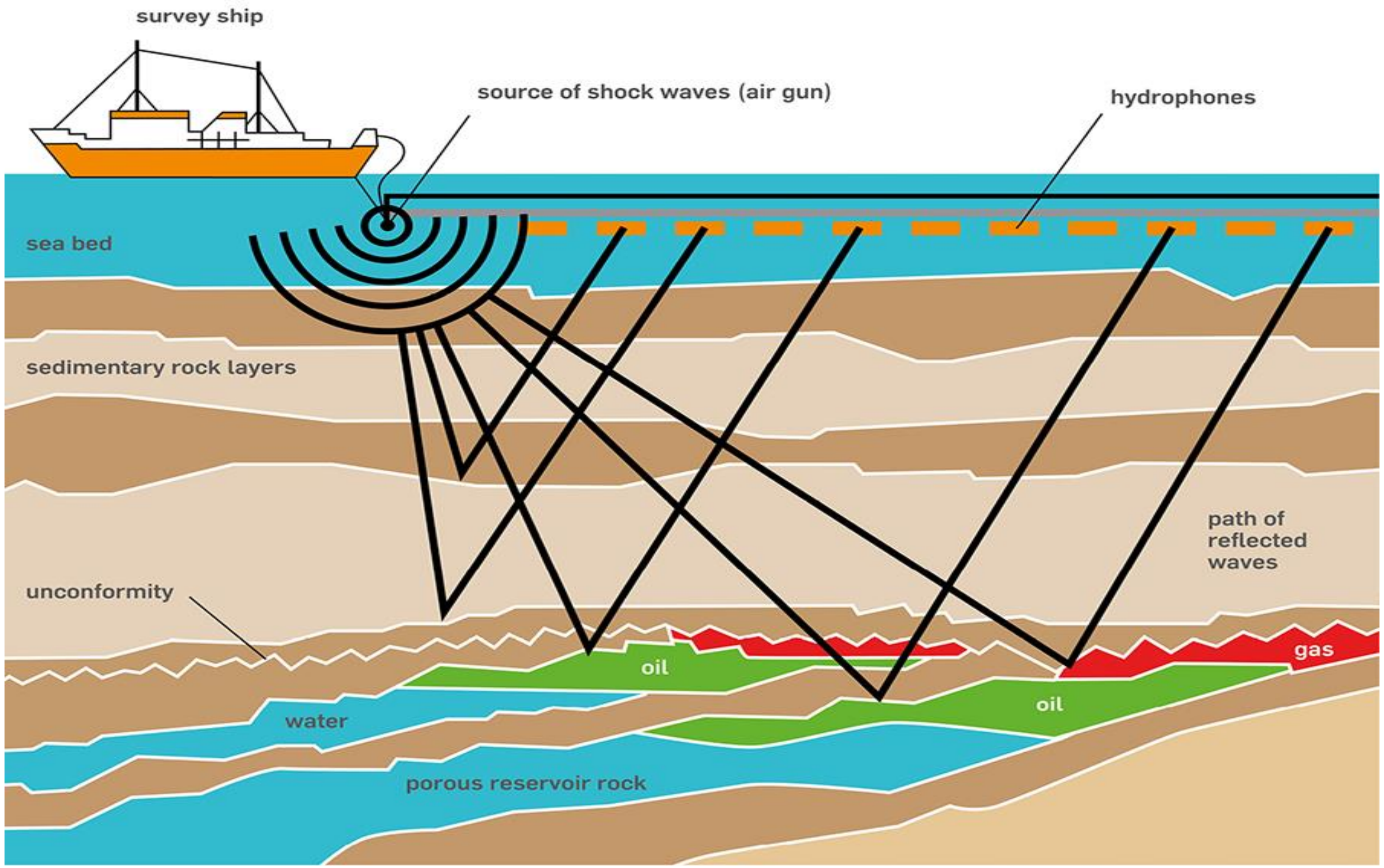
An ensemble-based kernel learning framework to handle data assimilation problems with imperfect forward simulators

Xiaodong Luo, NORCE

Outline

- Background and motivation
- An ensemble-based kernel algorithm for supervised learning
- From supervised learning to data assimilation with model errors
- Synthetical examples and a real field application
- Discussion and conclusion

Seismic survey for hydrocarbon reservoir monitoring and management

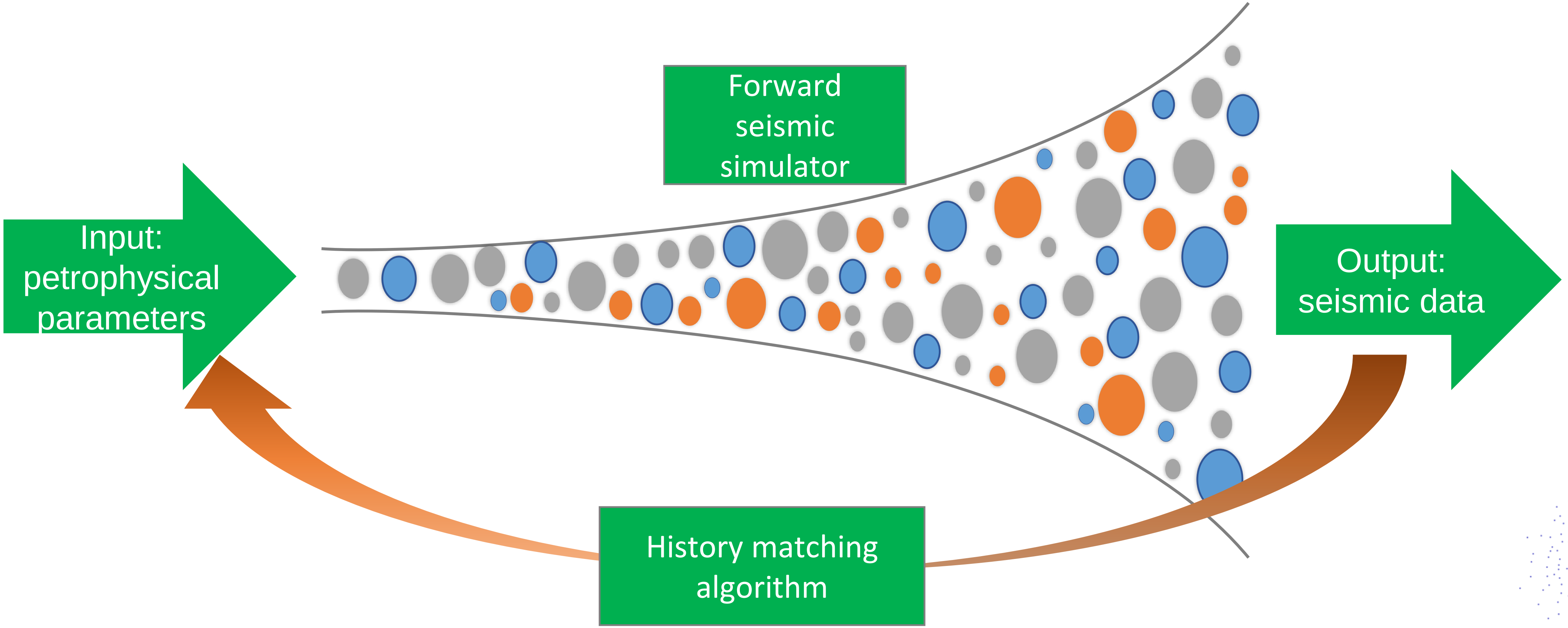


Source: https://oilnow.gy/

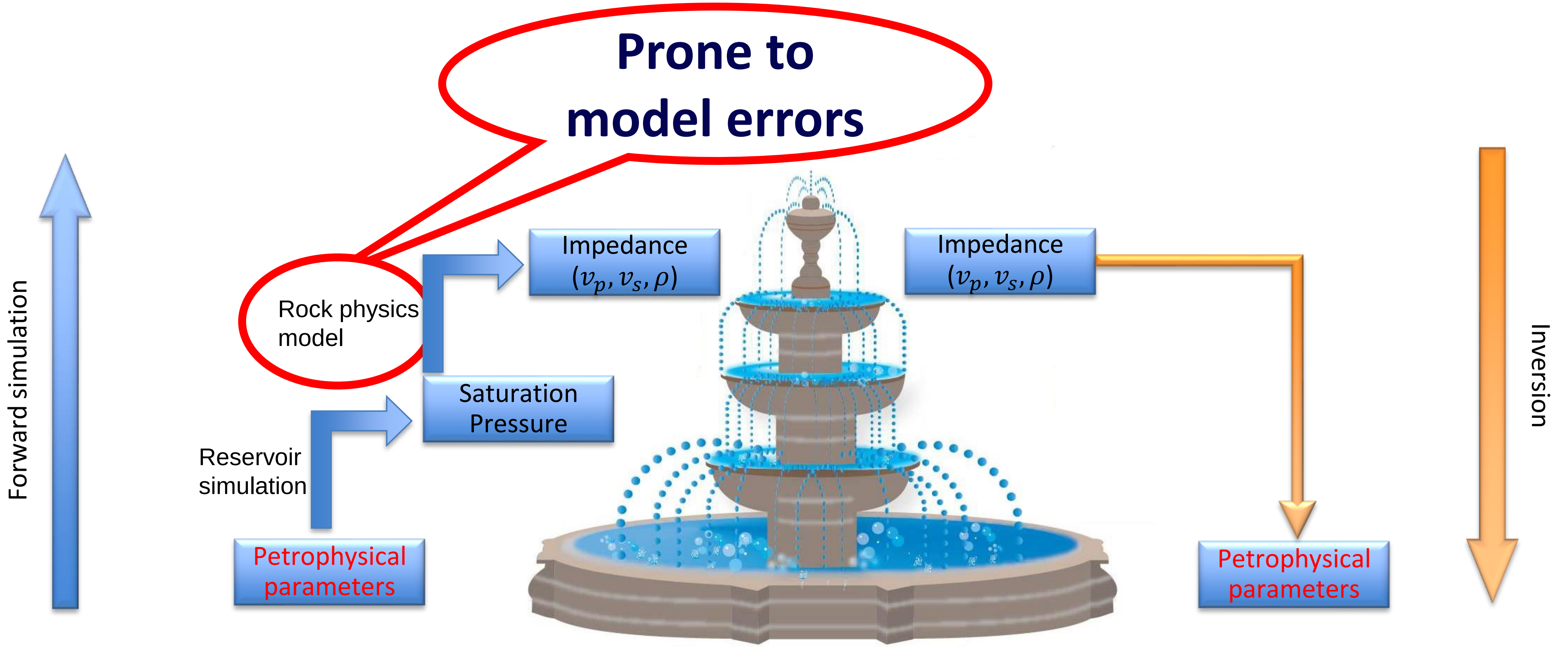
More advanced techniques available e.g., **Ocean Bottom Cable (OBC)** or even **Permanent Reservoir Monitoring (PRM)** system

Seismic history matching (SHM)

SHM involves using (3D or 4D) seismic data to estimate properties of reservoir formations



Forward seismic simulation and inversion



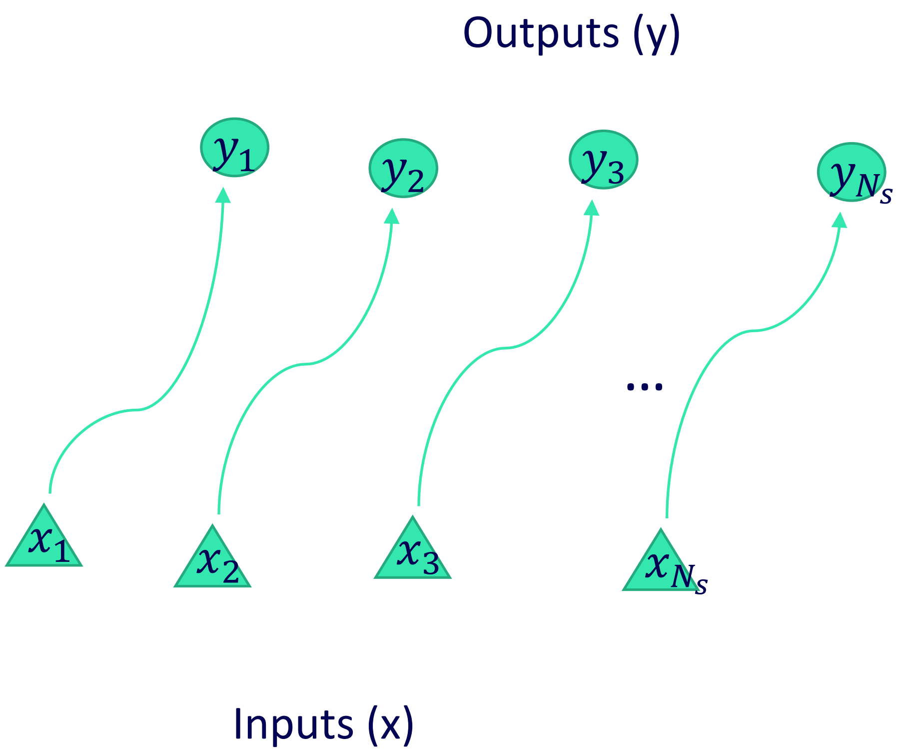
Motivation

Develop a workflow to account for model errors in **rock physics models** (RPM)

Outline

- Background and motivation
- An ensemble-based kernel algorithm for supervised learning
- From supervised learning to data assimilation with model errors
- Synthetical examples and a real field application
- Discussion and conclusion

Supervised learning (1/3)



- We have a set of inputs $X \equiv \{x_i\}_{i=1}^{N_s}$ with N_s samples; and a corresponding set of **noisy** outputs $Y \equiv \{y_i\}_{i=1}^{N_s}$
- We want to learn a function h so that $h(x_i)$ match y_i to a good extent, for $i = 1, 2, \ldots, N_s$

Supervised learning (2/3)

To this end, we solve a functional optimization (known as ***empirical risk minimization***, ***ERM***) problem to find the optimal h^*

$$h^* = \underset{h}{\operatorname{argmin}} \frac{1}{N_s} \sum_i \left(y_i - h(x_i)\right)^2 + \gamma\, R(||h||)$$

- γ: regularization parameter
- R: regularization functional to avoid overfitting, e.g., $R(x) = x^2$
- $||h||$: functional norm in a certain function space

Supervised learning (3/3)

To solve the ***ERM*** problem, in practice, one strategy is to adopt a parametric model that can be used to approximate a functional

Then the ***ERM*** problem is converted to a parameter estimation problem, i.e.,

$$h^* = \underset{h}{\operatorname{argmin}} \frac{1}{N_s} \sum_i \left(y_i - h(x_i)\right)^2 + \gamma\, R(||h||)$$

$$\Downarrow \; h(\theta; x_i) \approx h(x_i)$$

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \frac{1}{N_s} \sum_i \left(y_i - h(\theta; x_i)\right)^2 + \gamma\, R(\theta);$$

Examples of parametric model for functional approximation: $h(\theta; x_i) \approx h(x_i)$

generalized linear models

support vector machines (SVM)

(shallow or deep) neural networks

Ensemble-based supervised learning (1/2)

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \frac{1}{N_s} \sum_i \left(y_i - h(\theta; x_i)\right)^2 + \gamma\, R(\theta)$$

vectorize

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \left(Y - H(\theta; X)\right)^2 + \gamma\, R(\theta)$$

Similar to a Variational Data Assimilation (Var) problem

Naturally, in light of the developments of ensemble based data assimilation methods, instead of estimating a single set θ of parameters, we can estimate an ensemble

$$\Theta \equiv \{\theta_j\}_{j=1}^{N_e}$$

of such parameters

Ensemble-based supervised learning (2/2)

NORCE

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \left(Y - H(\theta; X)\right)^2 + \gamma\, R(\theta)$$

ensemblize

$$\Theta^* = \underset{\Theta = \{\theta_j\}_{j=1}^{N_e}}{\operatorname{argmin}} \frac{1}{N_e} \left\{ \sum_j \left(Y - H(\theta_j; X)\right)^2 + \gamma\, R(\theta_j) \right\}$$

We will obtain all the benefits in using ensemble based methods:

- Adjoint free
- Uncertainty quantification
- Fast implementation

Iterative ensemble smoothers, e.g., Luo et al. 2015*, can be used to solve the ensemble-based (supervised) learning problem

*Luo, X., Stordal, A. S., Lorentzen, R. J., & Naevdal, G. (2015). Iterative Ensemble Smoother as an Approximate Solution to a Regularized Minimum-Average-Cost Problem: Theory and Applications. *SPE Journal*, *20*, 962-982.

Kernel method for functional approximation

$$h(x;\theta) = \sum_k c_k\, K(\left\|x - x_k^{cp}\right\|; \beta_k)$$

$$\theta = \left[c_1, c_2, \dots c_{N_{sp}}; \beta_1, \beta_2, \dots \beta_{N_{sp}}\right]^{\mathrm{T}}$$

for a set of "center points" x_k^{cp} $(k = 1, 2, \dots, N_{sp})$, where

- c_k and β_k are parameters associated with the k-th center point
- K is a certain kernel function. Here we use Gaussian kernel

$$K\left(\left\|x - x_k^{cp}\right\|; \beta_k\right) = e^{-\beta_k^2 (x - x_k^{cp})^2}$$

Outline

- Background and motivation
- An ensemble-based kernel algorithm for supervised learning
- From supervised learning to data assimilation with model errors
- Synthetical examples and a real field application
- Discussion and conclusion

Problem statement (1/3)

- Problem in consideration:

$$\boldsymbol{y}^{o} = \boldsymbol{f}(\boldsymbol{x}^{tr}) + \boldsymbol{\epsilon}$$

where

- $\boldsymbol{y}^{o}$: observed output (observation)
- $\boldsymbol{x}^{tr}$: underlying true model variables that generate $\boldsymbol{y}^{o}$ through the true forward simulator $\boldsymbol{f}$
- $\boldsymbol{f}$: true (but unknown) forward simulator
- $\boldsymbol{\epsilon}$: observation noise. $\boldsymbol{\epsilon} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{C_d})$

Problem statement (2/3)

- In history matching (data assimilation), we may use the following forward simulation system

$$\boldsymbol{y}^{sim} = \boldsymbol{g}(\boldsymbol{x})$$

where

- $\boldsymbol{y}^{sim}$: simulated observation
- $\boldsymbol{x}$: model variables to be estimated
- $\boldsymbol{g}$: imperfect forward simulator

Problem statement (3/3)

$$\boldsymbol{y^o} = \boldsymbol{g}(\boldsymbol{x}) + [\boldsymbol{y^o} - \boldsymbol{g}(\boldsymbol{x})]$$
$$\approx \boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{r}(\boldsymbol{x}, \boldsymbol{\theta})$$

Kernel methods (or other machine learning models) can be used to reparametrize/approximate the residual term*

$$\boldsymbol{r}(\boldsymbol{x}, \boldsymbol{\theta}) \equiv \boldsymbol{r}(\boldsymbol{x}, \boldsymbol{\theta}; \boldsymbol{y^o}, \boldsymbol{y^o_{cp}}, \boldsymbol{x_{cp}})$$

so instead of trying to find an optimal functional form for $\boldsymbol{r}$, we optimize/estimate a set $\boldsymbol{\theta}$ of parameters (as well as $\boldsymbol{x}$) instead.

X. Luo, 2019. *Ensemble-based kernel learning for a class of data assimilation problems with imperfect forward simulators.* Available from arXiv:1901.10758

Ensembled-based data assimilation with kernel approximation to the residual term

$$\Theta^* = \underset{\Theta=\{[\boldsymbol{x}_j;\boldsymbol{\theta}_j]\}_{j=1}^{Ne}}{\operatorname{argmin}} \sum_j \left(\boldsymbol{y}^{\boldsymbol{o}} - \boldsymbol{g}(\boldsymbol{x}_j) - \boldsymbol{r}(\boldsymbol{x}_j, \boldsymbol{\theta}_j)\right)^T C_d^{-1} \left(\boldsymbol{y}^{\boldsymbol{o}} - \boldsymbol{g}(\boldsymbol{x}_j) - \boldsymbol{r}(\boldsymbol{x}_j, \boldsymbol{\theta}_j)\right) + \gamma\, R\left([\boldsymbol{x}_j; \boldsymbol{\theta}_j]\right)$$

- This optimization problem can still be solved through an iterative ensemble smoother
- We need to jointly estimate/update $\boldsymbol{x}_j$ and $\boldsymbol{\theta}_j$
- In implementation, it just means that we augment $\boldsymbol{x}_j$ and $\boldsymbol{\theta}_j$ into model variable vectors that will be updated

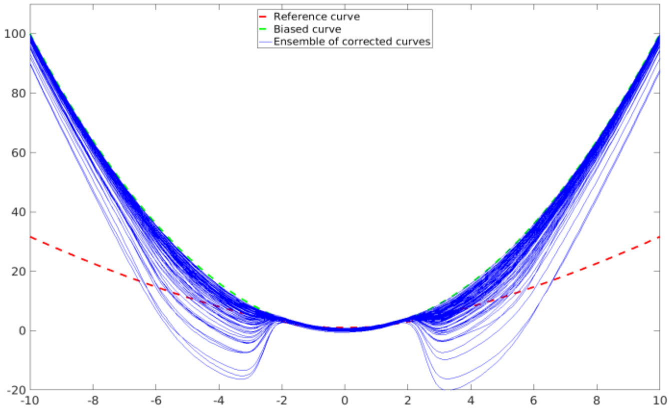
Outline

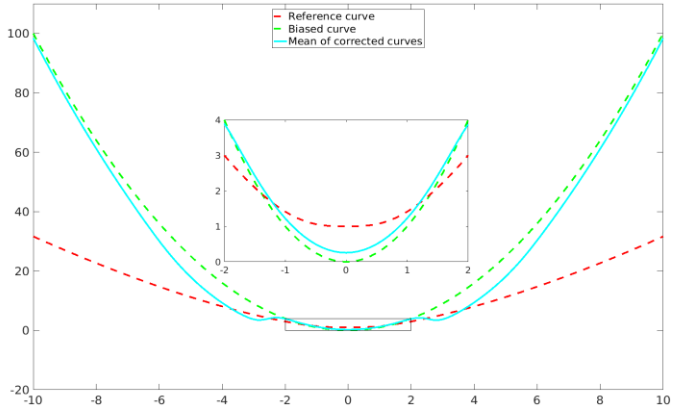
- Background and motivation
- An ensemble-based kernel algorithm for supervised learning
- From supervised learning to data assimilation with model errors
- Synthetical examples and a real field application
- Discussion and conclusion

Synthetic example 1: supervised learning

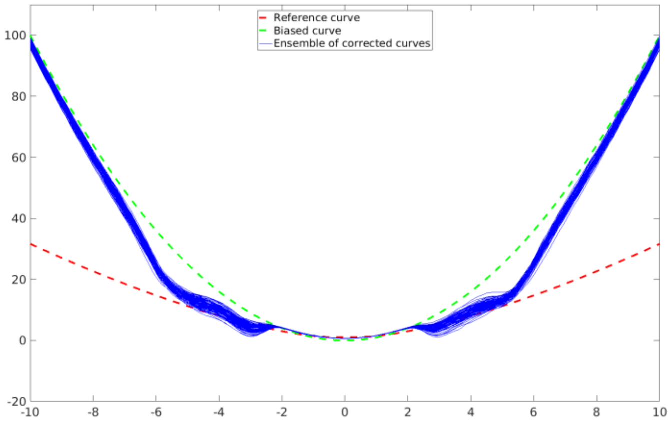
Blue: Ensemble of predicted functions
Red (dashed): reference function
Green (dashed): biased function

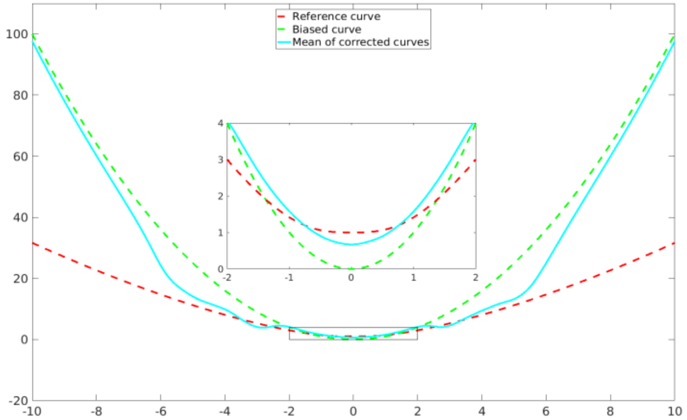
Cyan (solid): Ensemble mean
Red (dashed): reference function
Green (dashed): biased function





Initial ensemble

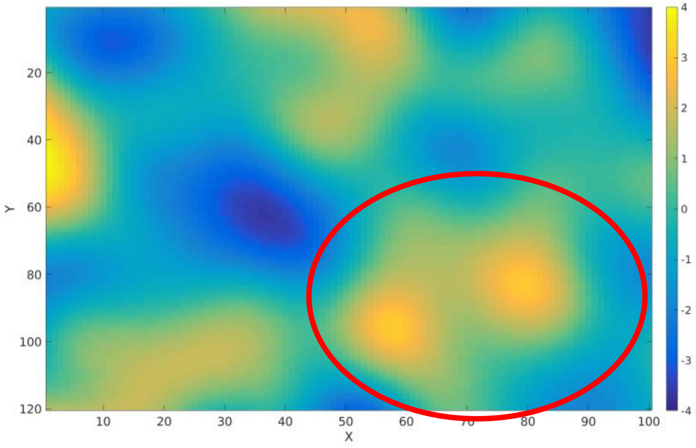




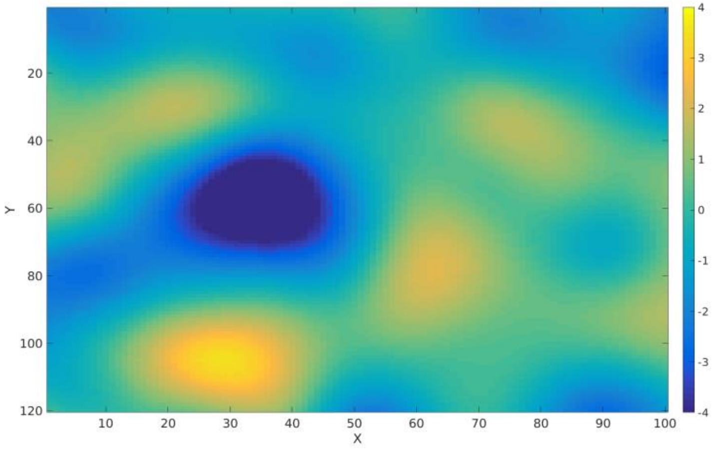
Final ensemble

Synthetic example 2: data assimilation

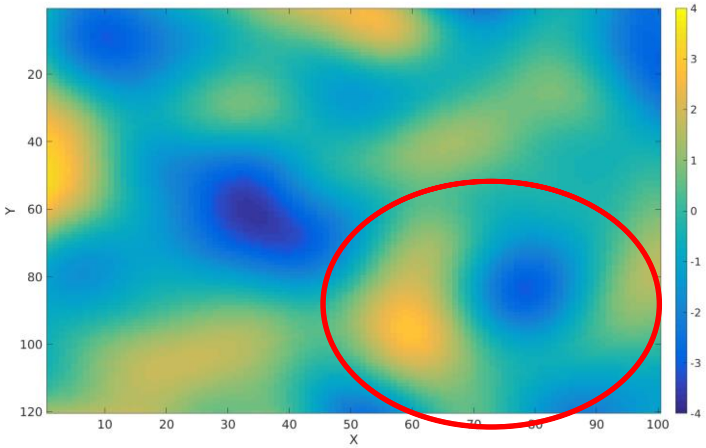
Truth



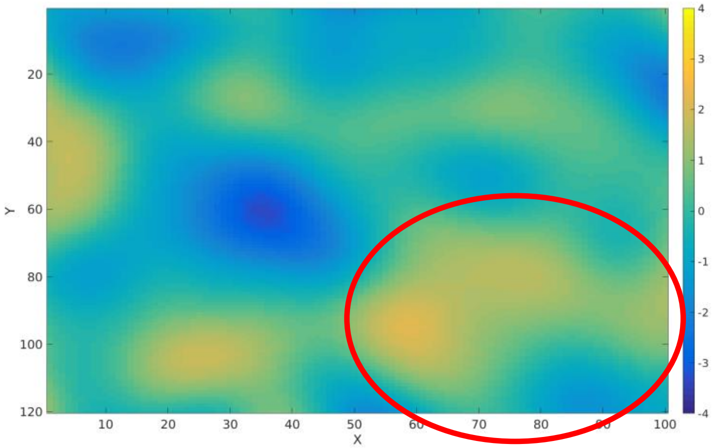
Mean of initial ensemble



Mean of final ensemble
(no model error correction)

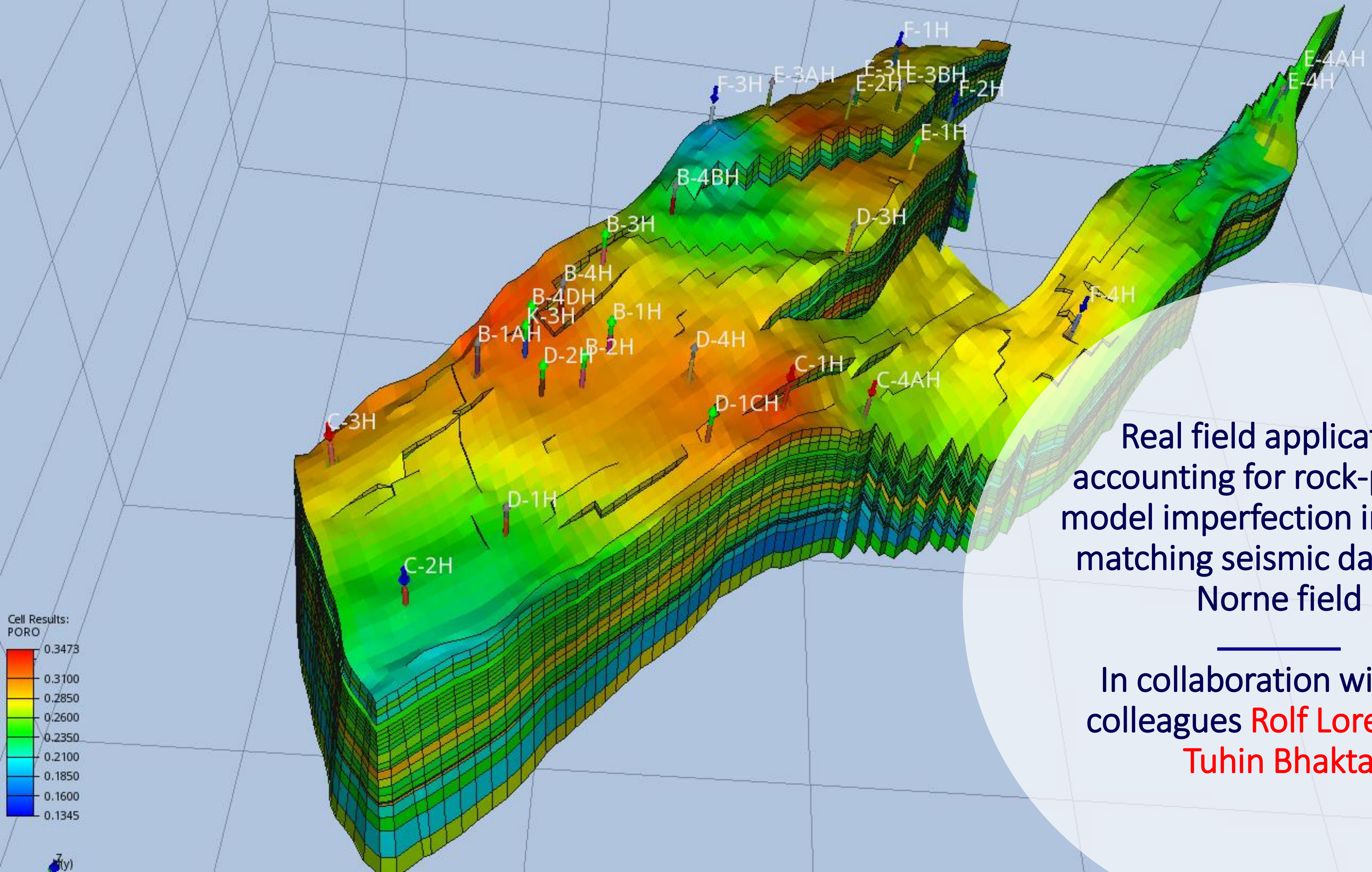


Mean of final ensemble
(with model error correction)



More information and results of both synthetical examples (supervised learning and data assimilation) can be found in the preprint

X. Luo, 2019. *Ensemble-based kernel learning for a class of data assimilation problems with imperfect forward simulators.* Available from arXiv:1901.10758



Real field application: accounting for rock-physics-model imperfection in history matching seismic data from Norne field

---

In collaboration with my colleagues Rolf Lorentzen, Tuhin Bhakta

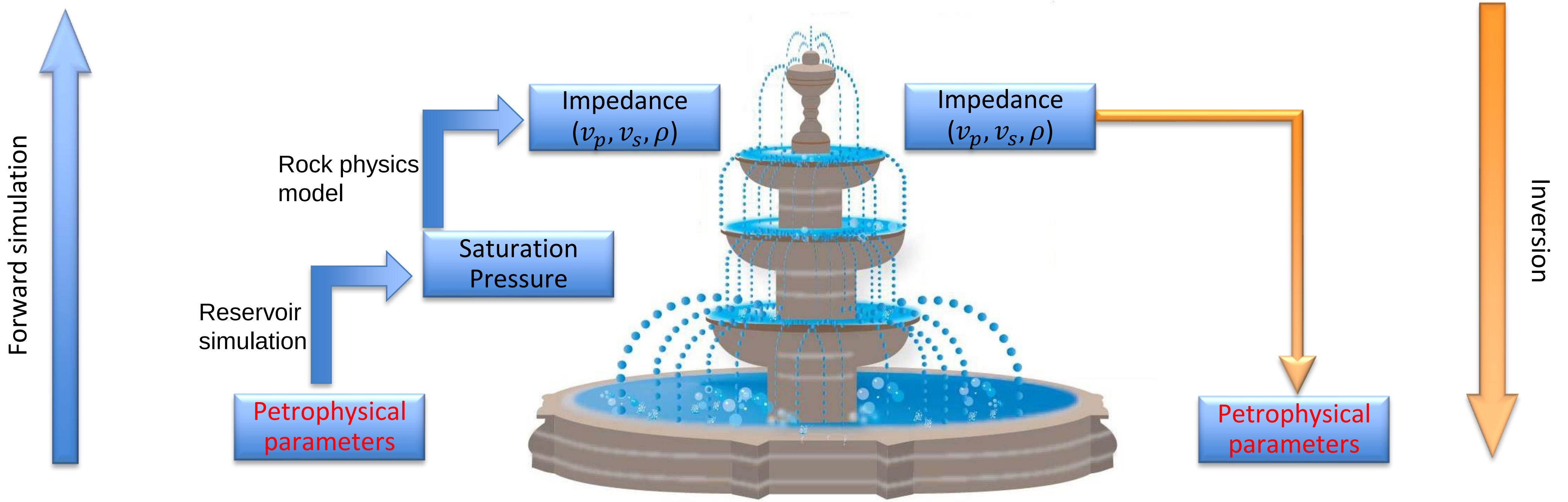
Experimental settings

Types of settings	Values/Info
Reservoir model size:	46 x 112 x 22
Seismic data (four surveys)	Acoustic impedance on each active gridblock Total number: 453,376; reduced to 24,232 through wavelet-based sparse representation*
Production data (1997 - 2006)	WOPRH, WGPRH, WWPRH Total number: 5,038
Model variables to estimate	PERM, PORO, NTG etc. Total number: 148,183
History matching algorithm	Iterative ES (Luo et al. 2015) + correlation-based adaptive localization (Luo et al. 2018, 2019)

*X Luo, T Bhakta, M Jakobsen, G Nævdal, 2017. An ensemble 4D-seismic history-matching framework with sparse representation based on wavelet multiresolution analysis. *SPE Journal*, 22, 985 - 1,010

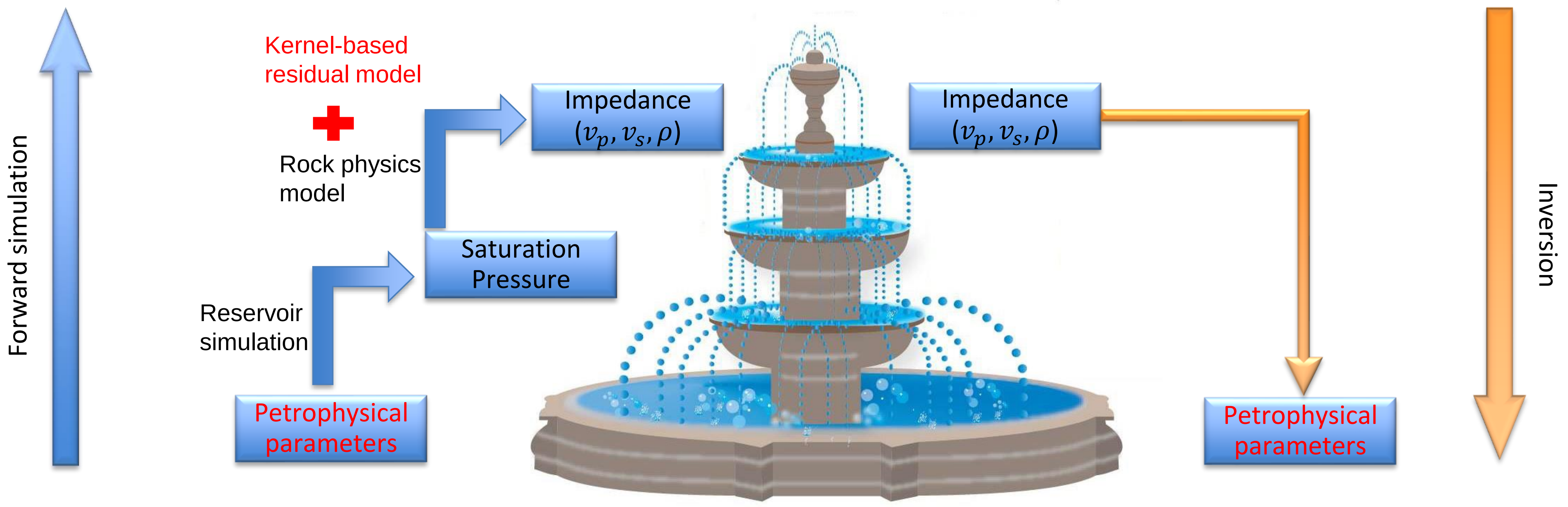
Experimental settings

The setting without model error correction (MEC)



Experimental settings

The setting with model error correction (MEC)

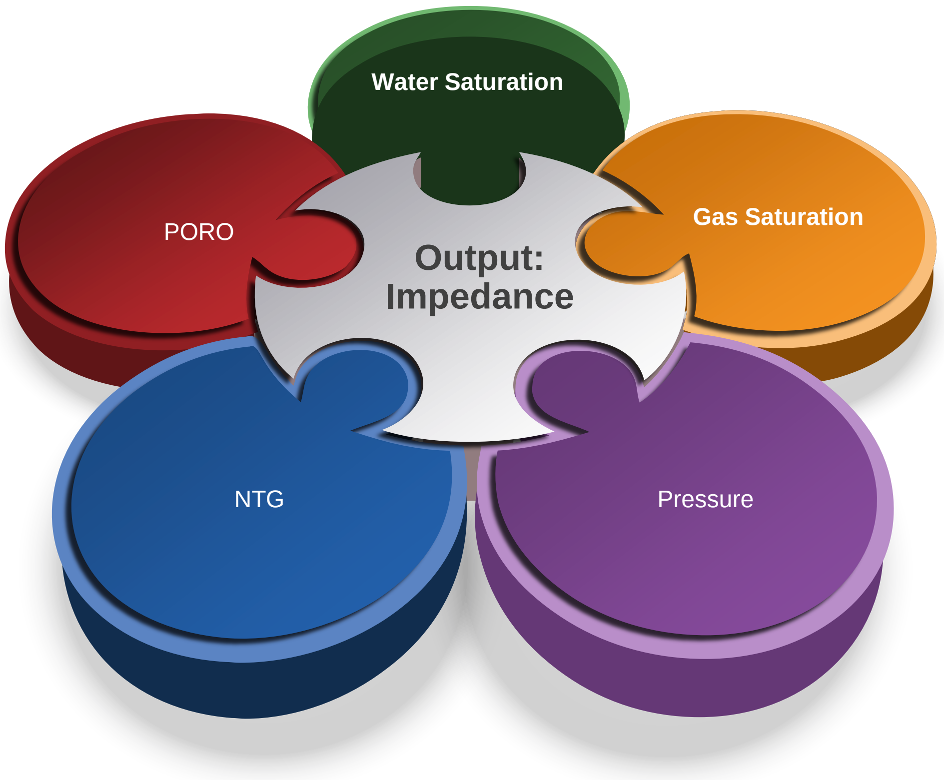


Experimental settings

Kernel-based residual model (inputs/output) at each active gridblock

Input 5: Water Saturation

Input 1: PORO



Input 4: Gas Saturation

Input 2: NTG

Input 3: Pressure

Total number of kernel parameters: 120,000 (with 20,000 center points)

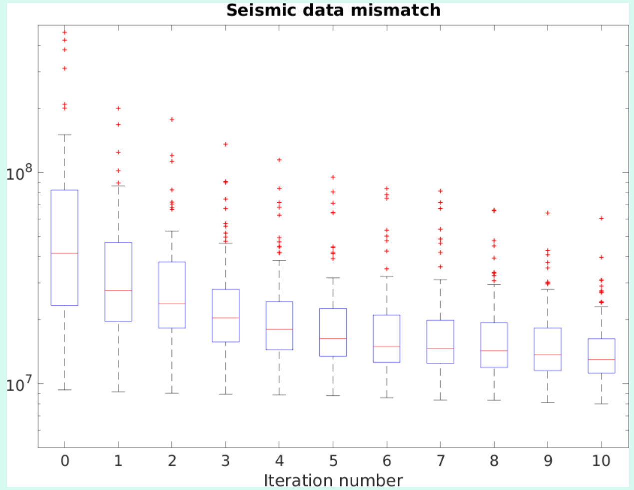
Experimental settings

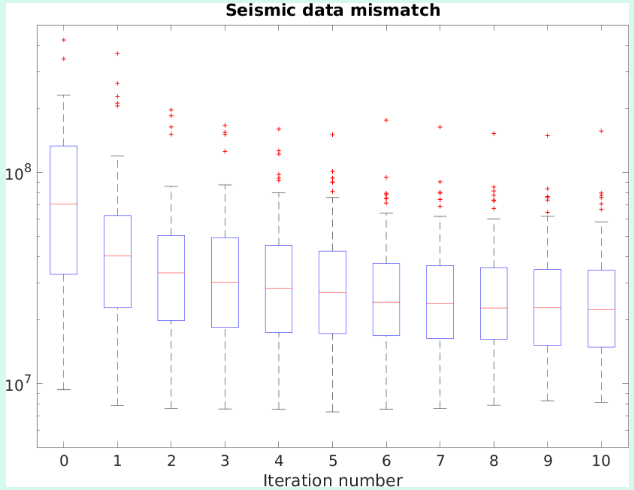
- Only seismic data are used history matching
- Production data are reserved for cross-validation

Experimental results: data mismatch

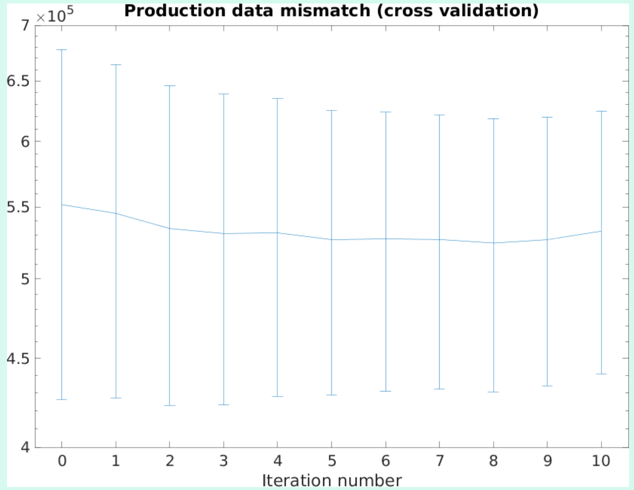
Results without MEC

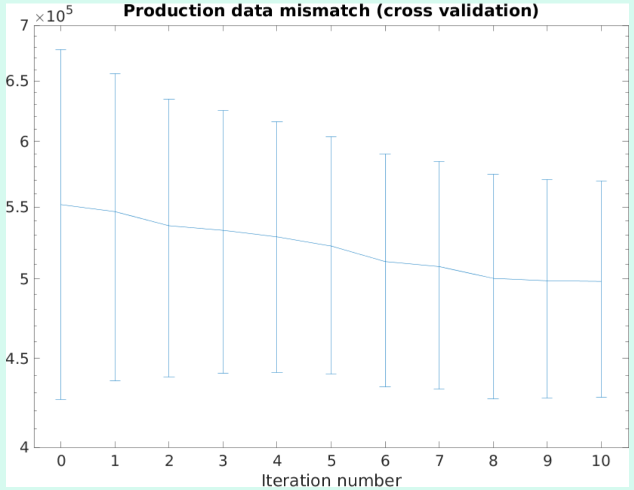
Results with MEC





Seismic data mismatch (history matching)

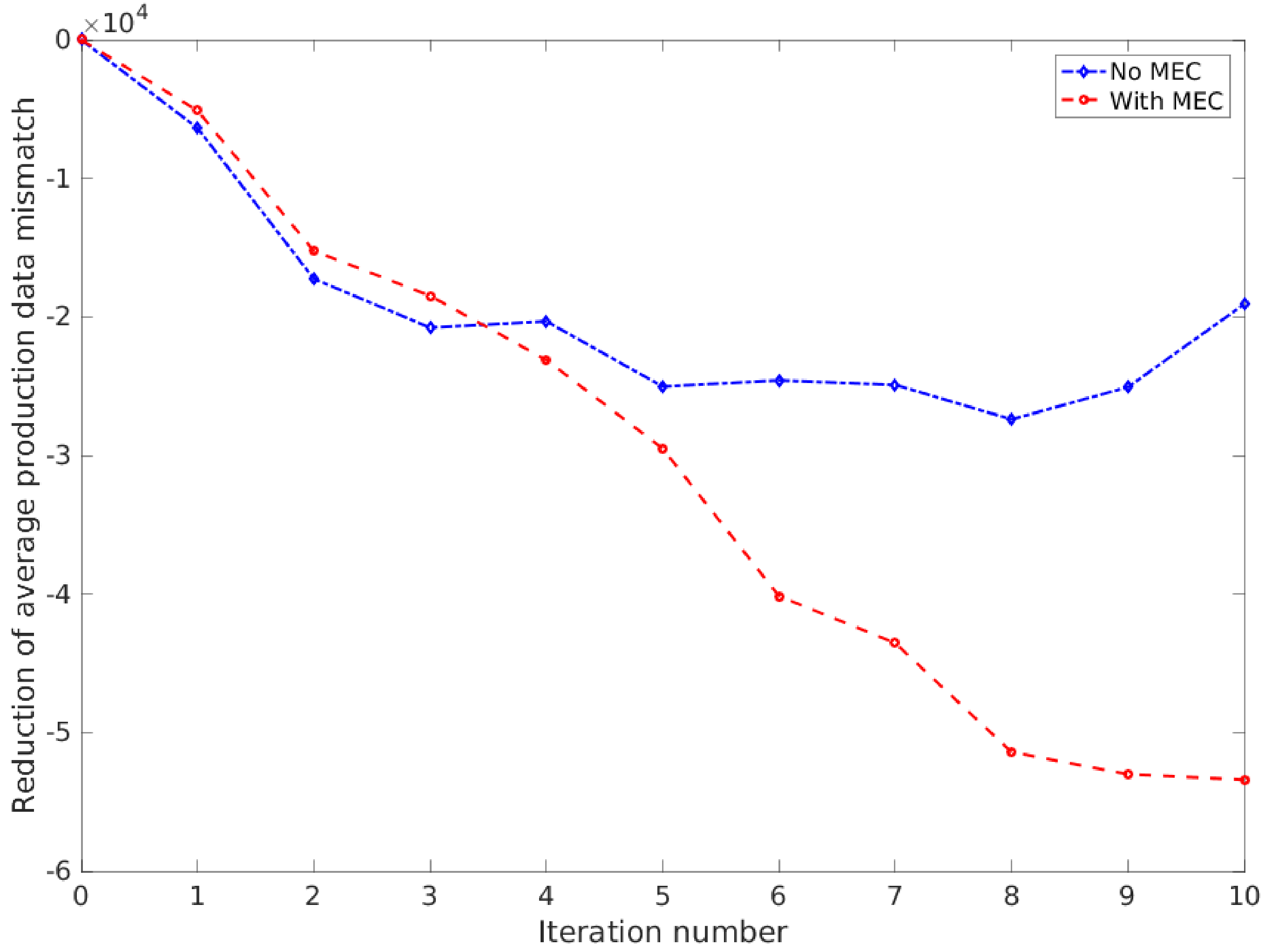




Production data mismatch (cross validation)

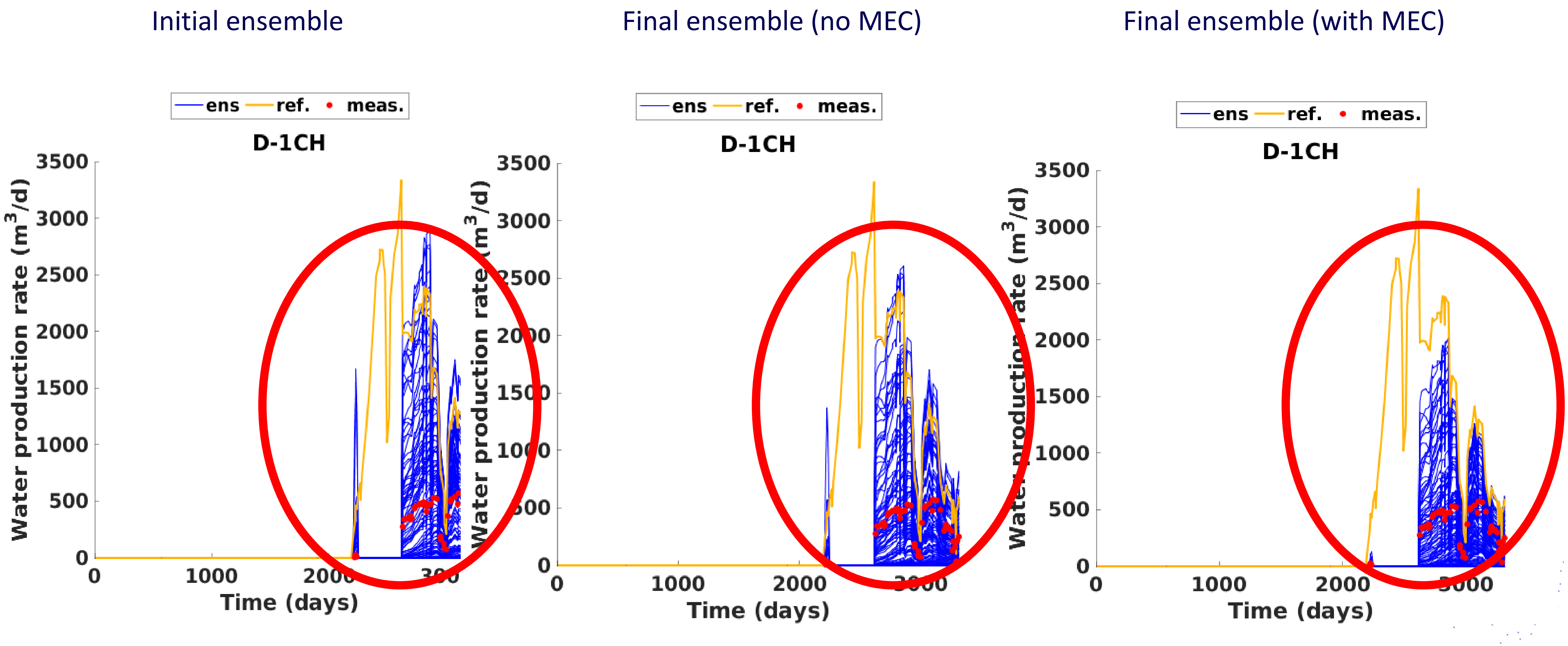
Experimental results: mismatch reduction

Reductions of average production data mismatch with respect to the initial ensemble



Experimental results: forecast

Predicted water production rates (WPR) at well D-1H



Outline

- Background and motivation
- An ensemble-based kernel algorithm for supervised learning
- From supervised learning to data assimilation with model errors
- Synthetical examples and a real field application
- Discussion and conclusion

Discussion and conclusion

- We show similarities between supervised learning and data assimilation; As such, it becomes natural for us to develop an ensemble-based framework for supervised learning problems
- With minor modifications, ensemble-based learning can also be extended to handle data assimilation problems in the presence of model errors
- The integrated data assimilation framework appears to be useful for improving DA performance in both synthetical and real-world problems presented here

Acknowledgement

XL acknowledges financial supports from the “DIGIRES” project (RCN no. 280473) funded by indutry partners: ***AkerBP ASA, DEA Norge AS, EquiNor ASA, Lundin Norway AS, Neptun Energy Norge AS, Petrobras and Vår Energi AS, as well as the Research Council of Norway***

Q&A