Two-stage MCMC for groundwater problems

L. Josset[1], V. Demyanov[2], A. H. Elsheikh[2], I. Lunati[1]
[1]ISTE, University of Lausanne, Switzerland, [2]IPE, Heriot-Watt University, Edinburgh, UK

1. Problem statement

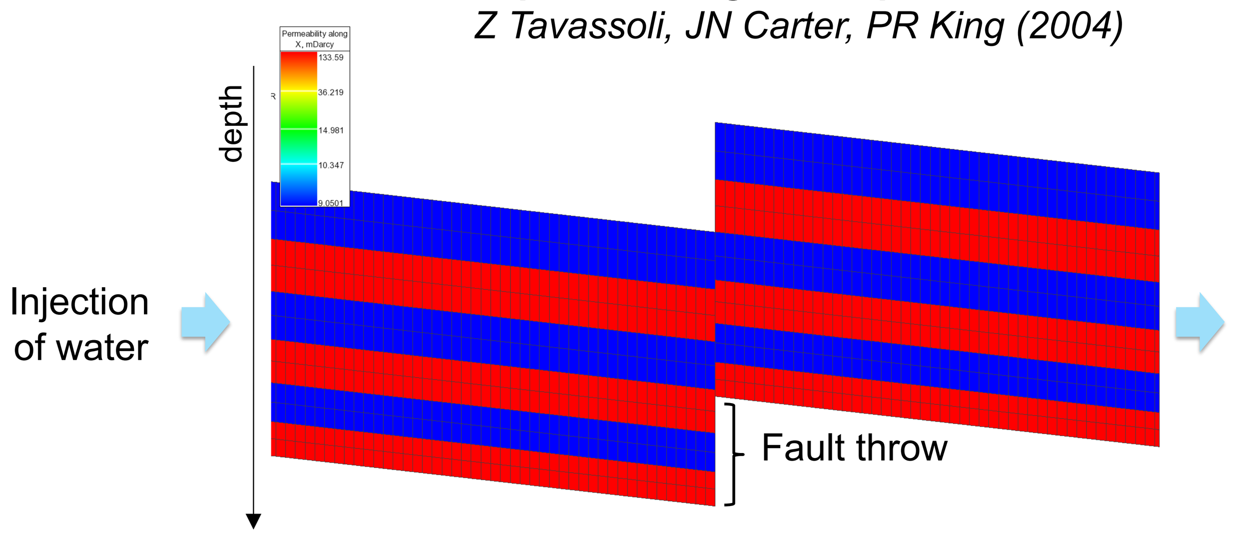
Challenges in groundwater problem

Question of interest:
What is the concentration of pollutant?

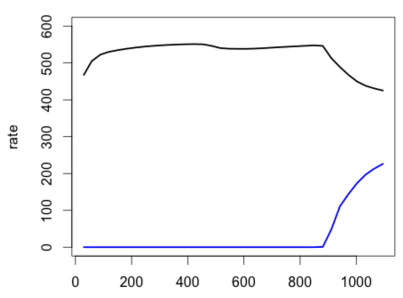
Problems:
- Underground properties are unknown
  - stochastic approaches are required
- Complex physical processes
  - e.g. two-phase flow simulations
  - computational cost becomes prohibitive

Synthetic problem

Imperial College Fault problem
Z Tavassoli, JN Carter, PR King (2004)



3 parameters:
- Fault throw = ?
- Conductivity K_{high} = ?
- Conductivity K_{low} = ?



Observed data:
Oil production rate -
Water production rate -

Goal:
Sample the parameters given the observed data

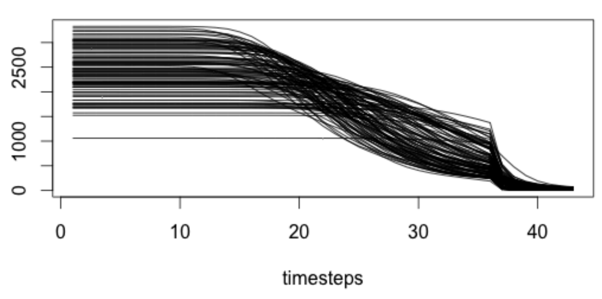
$$p(\theta|y) \propto \mathcal{L}(\theta;y)p(\theta)$$

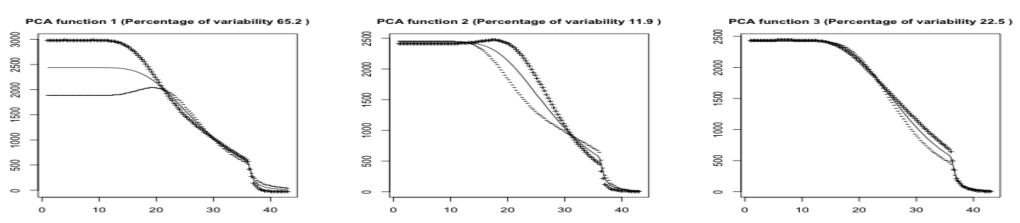
Usual approach MonteCarlo Markov Chain
But forward model is very expensive
→ 2-stage MCMC

3.a Error Model

Reduction of the dimensionality using Functional PCA

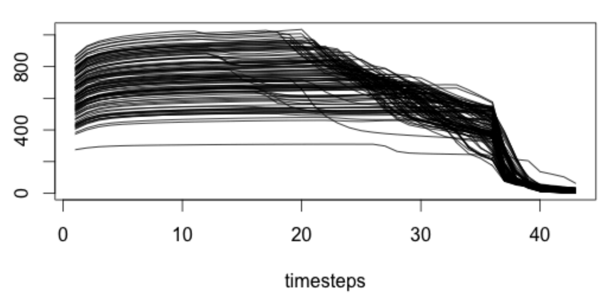
Single-phase breakthrough curves

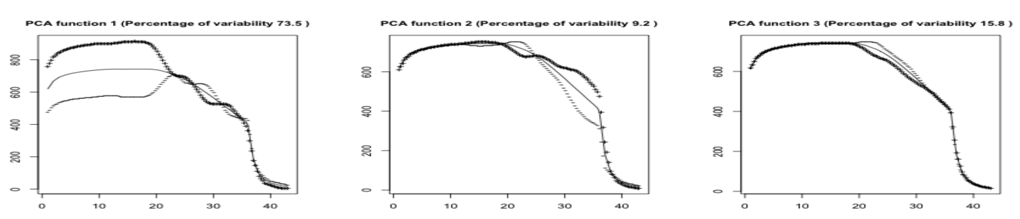




Variability explained with the 3 first PCs: 99.6%

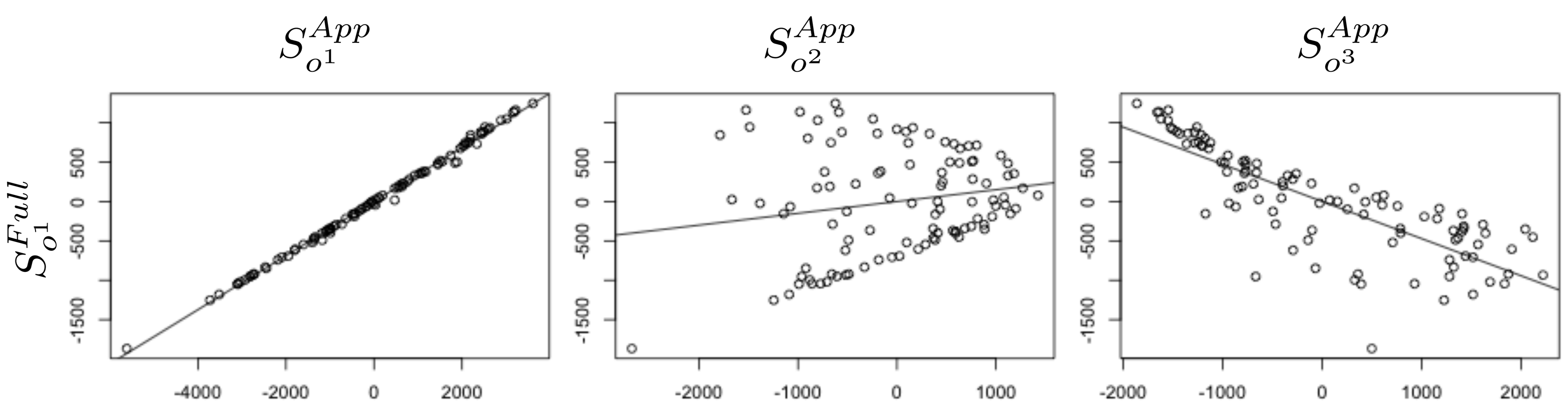
Two-phase breakthrough curves





Variability explained with the 3 first PCs: 97.4%

Understanding the relationship between the two models



Complete model:

$$S_{w^i}^{Full} = f(S_{o^1}^{App}, S_{o^2}^{App}, S_{o^3}^{App}, S_{w^1}^{App}, S_{w^2}^{App}, S_{w^3}^{App})$$

Regression R^2

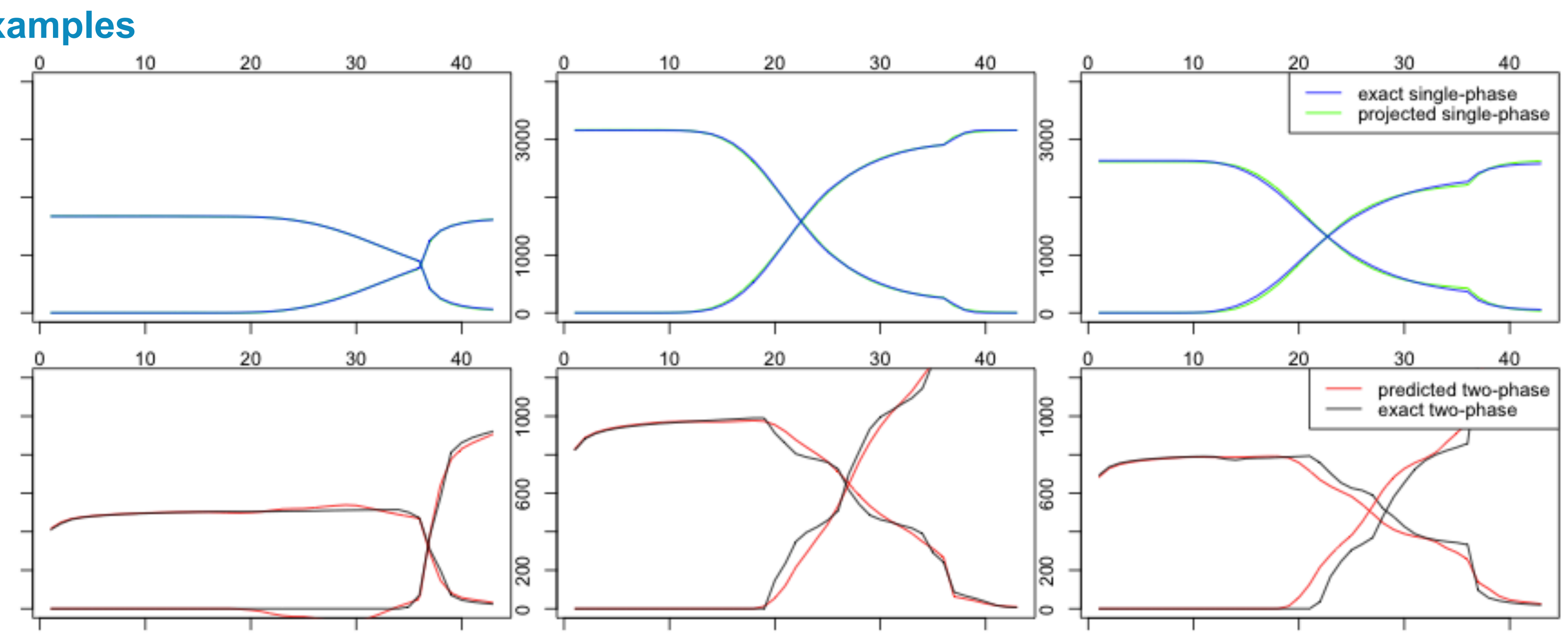
S^1_{oil} : 0.99	S^1_{water} : 0.99
S^2_{oil} : 0.93	S^2_{water} : 0.99
S^3_{oil} : 0.95	S^3_{water} : 0.99

Prediction of the two-phase from the single-phase response

Workflow

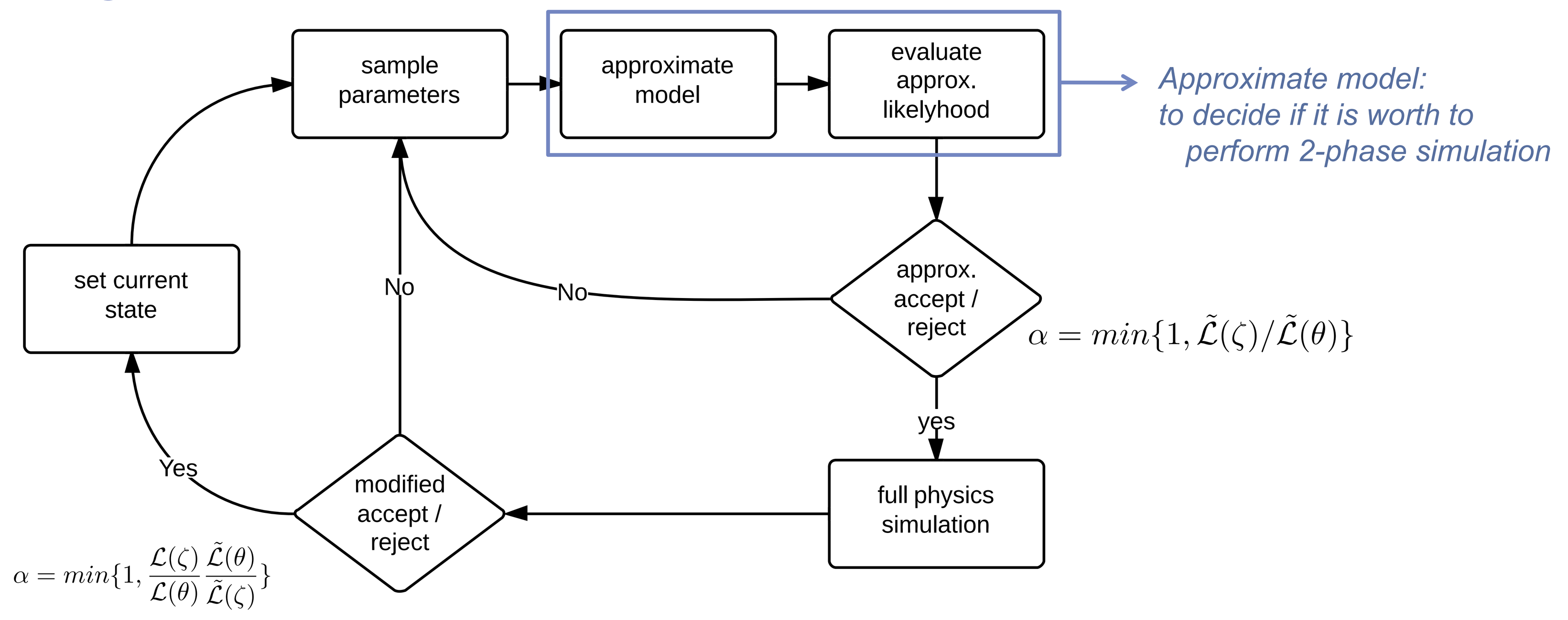
Single-phase simulation → Projection on rotated FPCs → Prediction of the scores → Reconstruction of the curve

Examples



2. Method

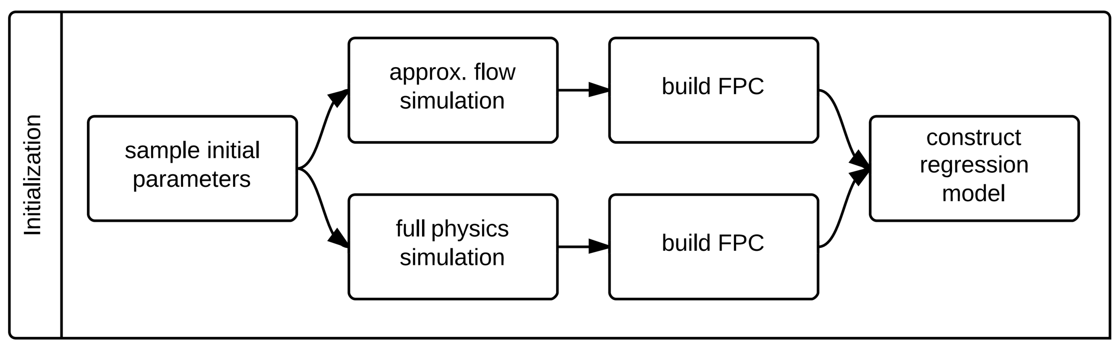
Two-stage MCMC



Approximate model

	Single-phase flow simulation	+	Functional PCA regression
Motivation	Provides information on the connectivity → Cheap in terms of computation Pressure problem is solved only once		Can we recover the missing physics ? → Learn from a training set

Construction of the regression model

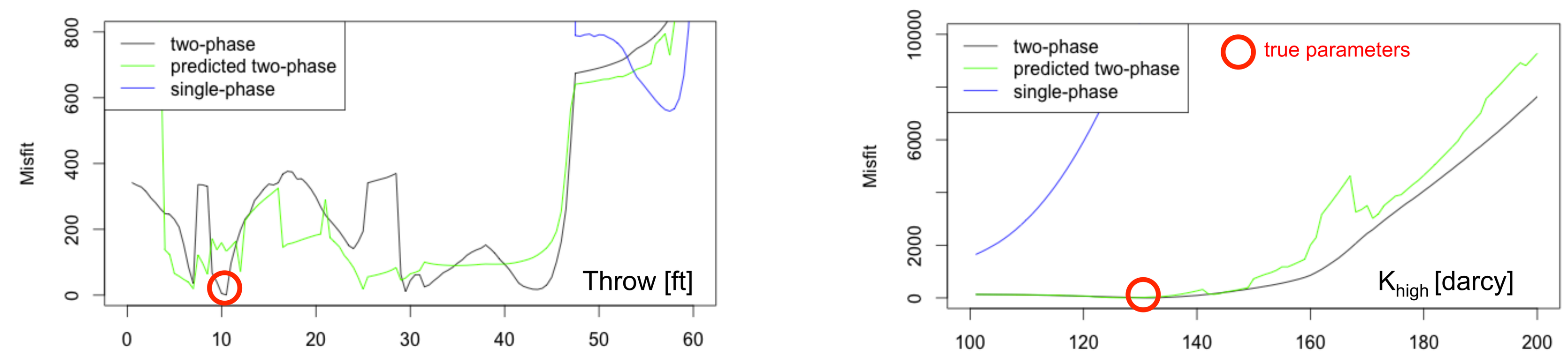


3.b Two-stage MCMC

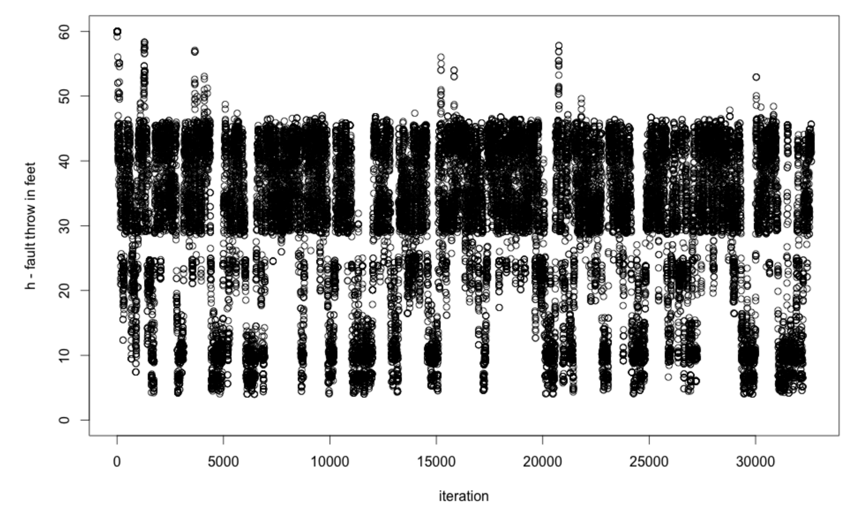
Misfit definition: l_2 distance with the reference

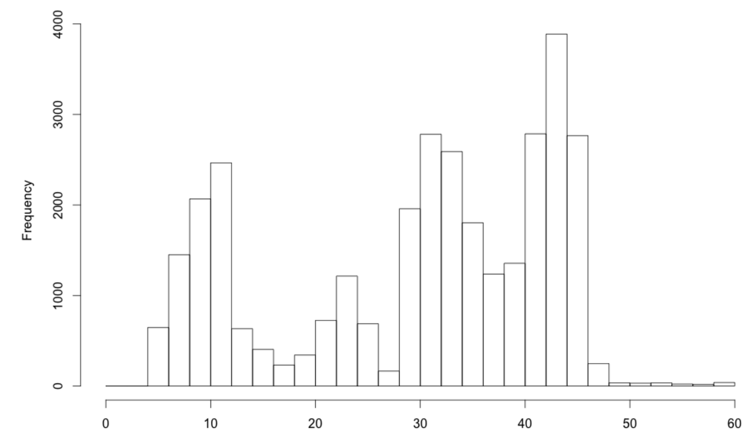
$$M = \frac{1}{36}\sum_{t=1}^{36}\frac{(C_{ref}^{oil}(t) - C^{oil}(t))^2}{2\sigma^2} + \frac{1}{7}\sum_{t=30}^{36}\frac{(C_{ref}^{water}(t) - C^{water}(t))^2}{2\sigma^2}$$

1D response surfaces for the 3 models



1D chain: run of 32'000 iterations





32'600 iterations
26'556 approximate accepted
→ 19'026 accepted after exact simulation
→ 7'530 rejected

		Approximate accepted	Approximate rejected	
Exact	accepted	19'897	1'470	21'367
Exact	rejected	6'659	4'593	11'252
		26'556	6'063	

4. Conclusion

Key ideas

2-stage MCMC using an approximate model
- 2-stage = evaluation of complete model when useful
- Approximate model = 1-phase + FPCA regression

Single-phase flow simulations:
- Connectivity is what varies between realisations
- Provides information on the advection part of the physics
- Cheap: pressure is solved only once

Regression model on FPCA scores:
- Response surfaces do not match if only single-phase
- Missing physics has to be taken in account

First results
- 1D chains of ~32'000 iterations
- 3 main modes of the exact posterior distribution are observed

Next steps
- Additional tests
- Investigate the MCMC set up
- Prediction
- Improve the approximate model in an iterative set-up

Acknowledgements

This project is supported by the Swiss National Science Foundation as a part of the ENSEMBLE project (Sinergia Grant No. CRSI22-132249/1)

References

- J. Andres Christen and Colin Fox. Markov chain Monte Carlo using an approximation. J. Comput. Graph. Statist., 14 (4):pp. 795–810, (2005)
- Y. Efendiev, A. Datta-Gupta, V. Ginting, X. Ma, and B. Mallick. An efficient two-stage Markov chain Monte Carlo method for dynamic data integration. WRR (2005)
- M. Febrero-Bande and M. Oviedo de la Fuente, "Statistical computing in functional data analysis: the R package fda. usc. Journal of Statistical Software", 51(4) (2012) 1-28.
- J. Ramsay, G. Hooker and S. Graves, "Functional data analysis with R and MATLAB", Springer (2009)
- Z. Tavassoli, J.N. Carter and P.R. King, "Errors in history matching", SPE 86883 (2004)