Permeability Estimation in Heterogeneous Formations with Derivative-Free Optimization Based On Observed Data

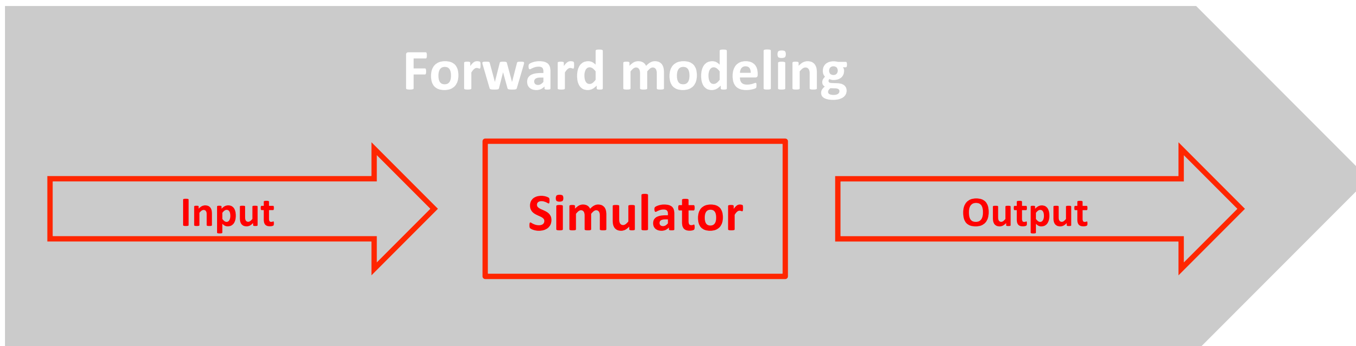
Yan Zhang[1,2] and Nick Sahinidis[1,2]
[1]National Energy Technology Laboratory, Pittsburgh, PA
[2]Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA
CAPD Annual Meeting, March 2013

DISCLAIMER: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

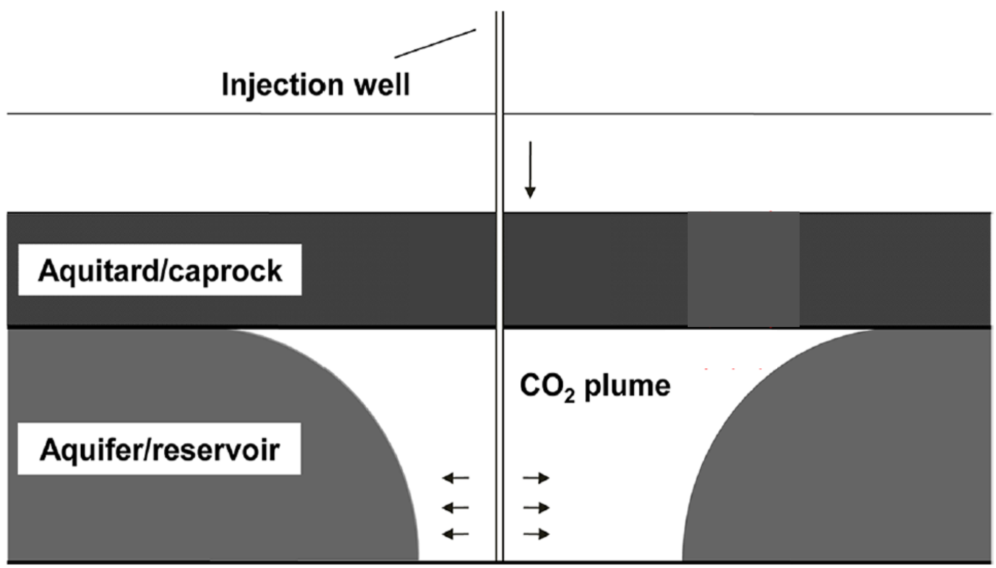
Outline

- Three key components of parameter estimation
- Work flow of parameter estimation
- Introduction to derivative-free optimization (DFO)
- Two case studies
- Conclusions

Three key components: System model

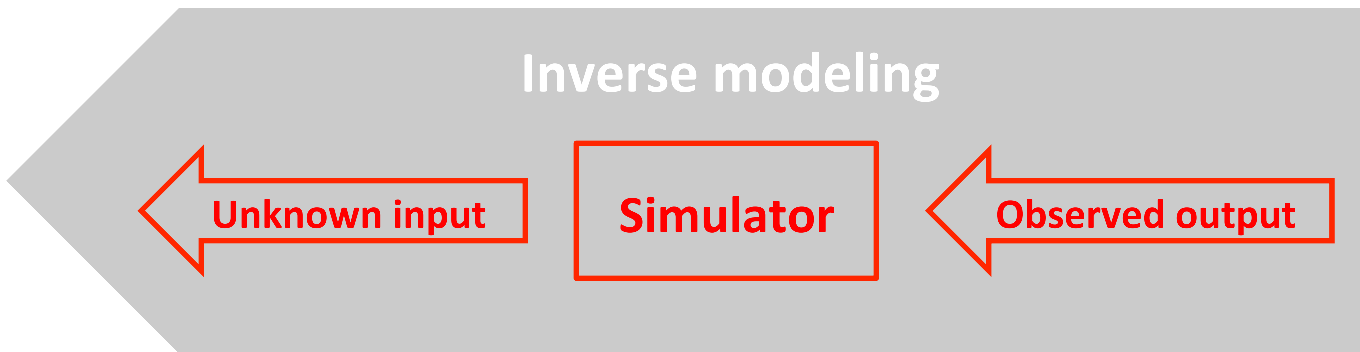


Porosity, permeability, etc

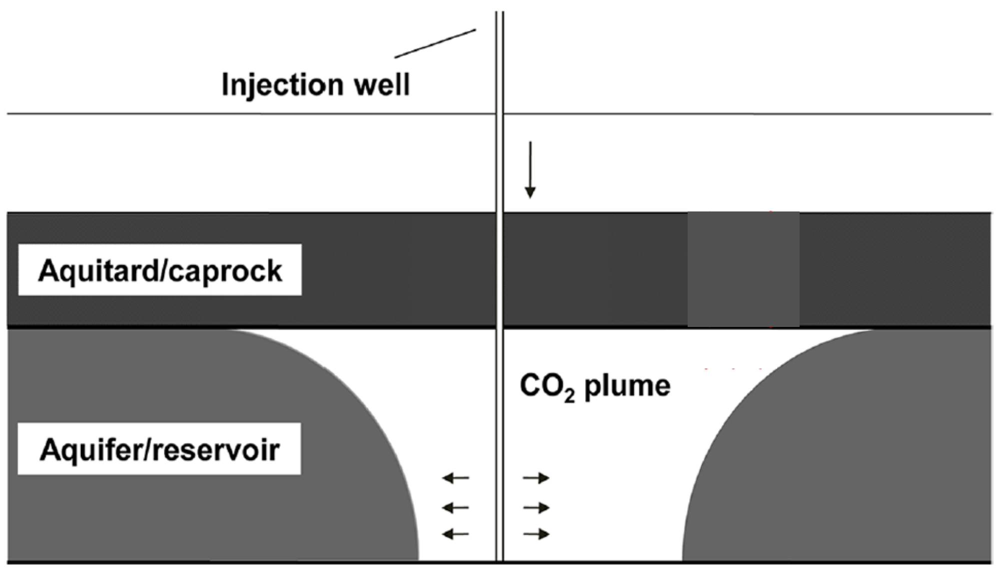


Pressure, gas saturation, etc

Three key components: System model



Unknown porosity, permeability, etc



Observed pressure, gas saturation, etc

Three key components: Fitness measure

- **Least squared error:**
  - **Mismatch between simulation and observation:** $\sum_i w_i \left(\frac{\boxed{y_i^{\mathrm{sim}}(p)} - y_i^{\mathrm{obs}}}{\sigma_i} \right)^2$

Simulation/black-box
No algebraic form for gradient

- **Maximum Likelihood**
  - **Likelihood of observation being present under parameter *p*:**

$$f\left(y^{\mathrm{obs}} | p\right) = \prod_i f_i\left(y_i^{\mathrm{obs}} | p\right) \propto \exp\left\{ -\frac{1}{2} \sum_i \left(\frac{y_i^{\mathrm{sim}}(p) - y_i^{\mathrm{obs}}}{\sigma_i} \right)^2 \right\}$$

- **Bayesian inference – posterior distribution**
  - **Prior probability distribution for *p*:** $f(p)$
  - **Posterior probability distribution for *p*:** $f\left(p | y^{\mathrm{obs}}\right) = f\left(y^{\mathrm{obs}} \middle| p\right) f(p)$

Three key components: Optimization algorithm

Derivative-based or algebraic solvers

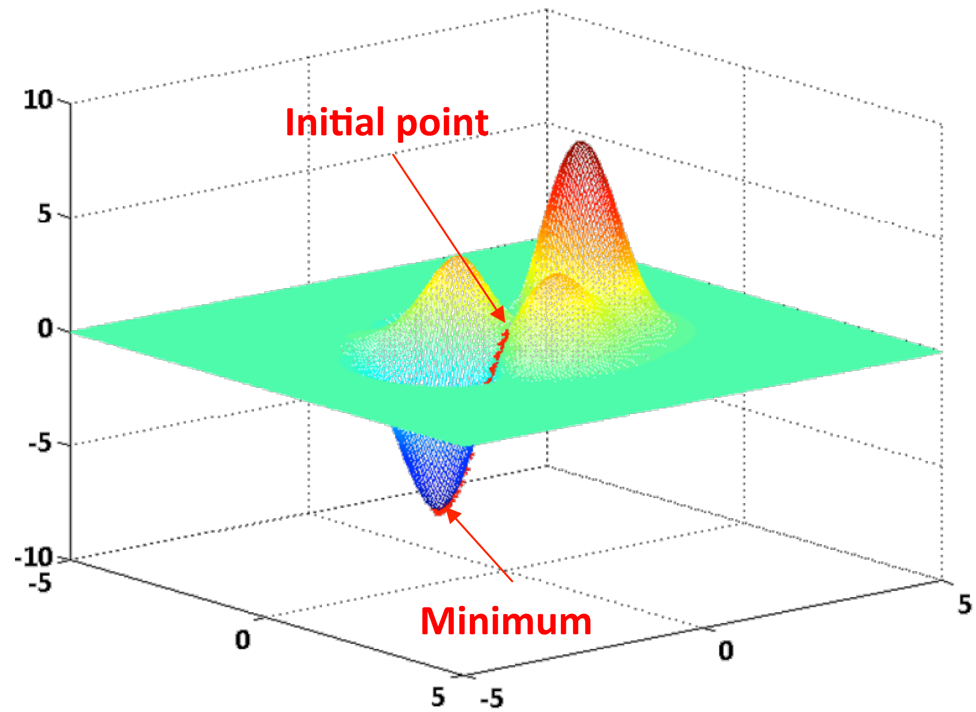
(BARON, CPLEX, IPOPT, CONOPT, etc.)

Pros:

Fast, efficient solutions with provably optimal solutions

Cons:

Need an algebraic model (access to derivative information)



Estimate gradient using central differencing

Three key components: Optimization algorithm

Derivative-free or black-box solvers

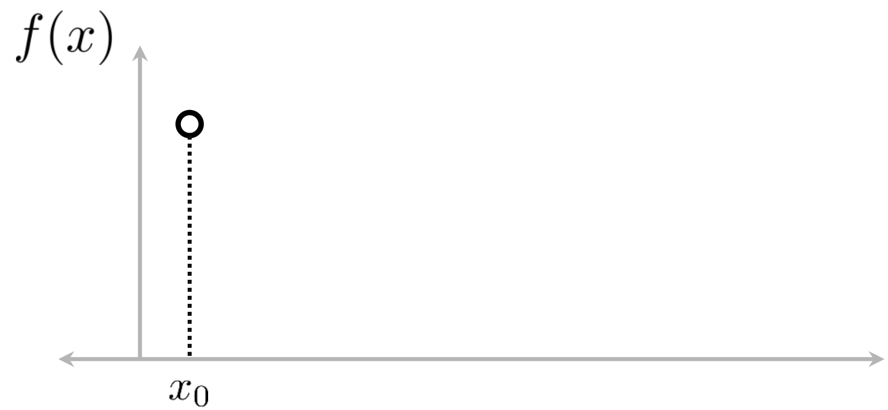
(Genetic algorithms, Matlab's fminsearch(), etc.)

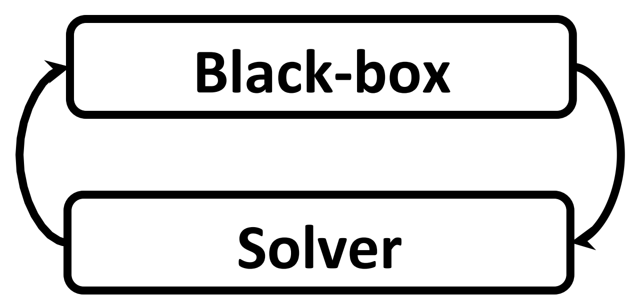
Pros:

Can solve problems where derivative information is unavailable, unreliable, or prohibitively expensive

Cons:

Slow and little to no guarantees that you have an optimal solution in the end





Three key components: Optimization algorithm

Derivative-free or black-box solvers

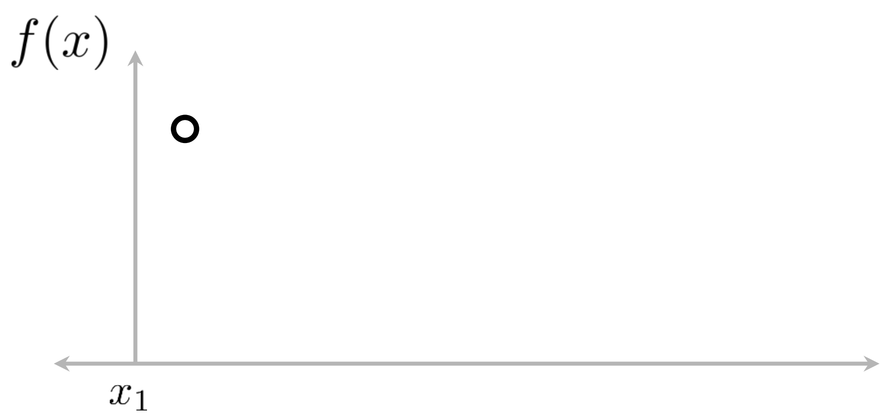
(`Genetic algorithms, Matlab's fminsearch(), etc.`)

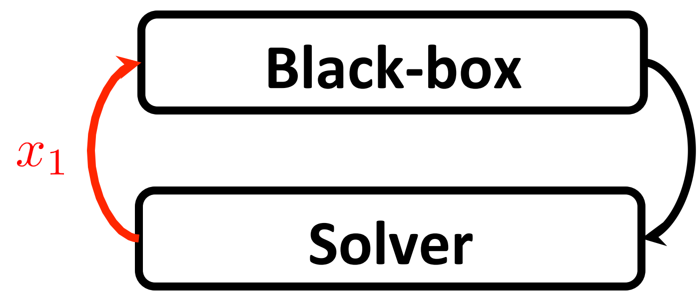
Pros:

Can solve problems where derivative information is unavailable, unreliable, or prohibitively expensive

Cons:

Slow and little to no guarantees that you have an optimal solution in the end





Three key components: Optimization algorithm

Derivative-free or black-box solvers

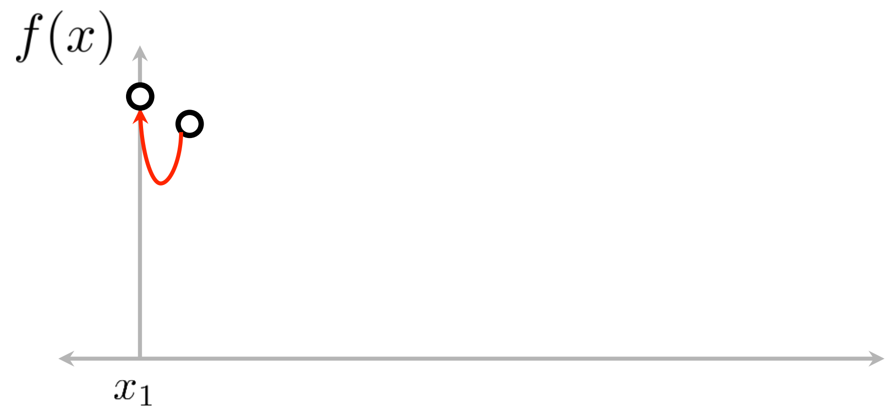
`(Genetic algorithms, Matlab's fminsearch(), etc.)`

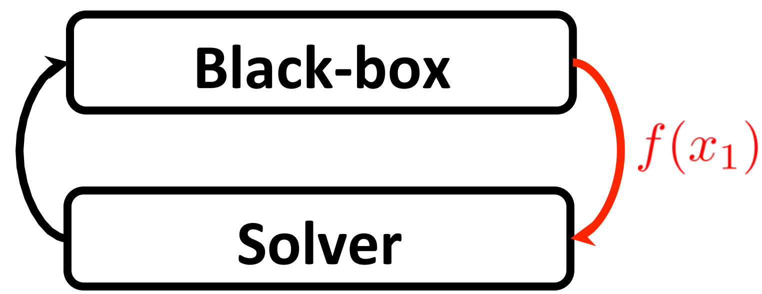
Pros:

Can solve problems where derivative information is unavailable, unreliable, or prohibitively expensive

Cons:

Slow and little to no guarantees that you have an optimal solution in the end





Three key components: Optimization algorithm

Derivative-free or black-box solvers

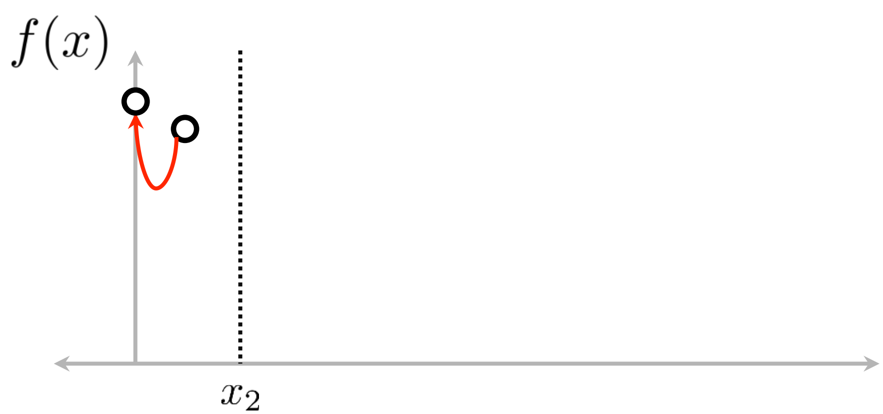
`(Genetic algorithms, Matlab's fminsearch(), etc.)`

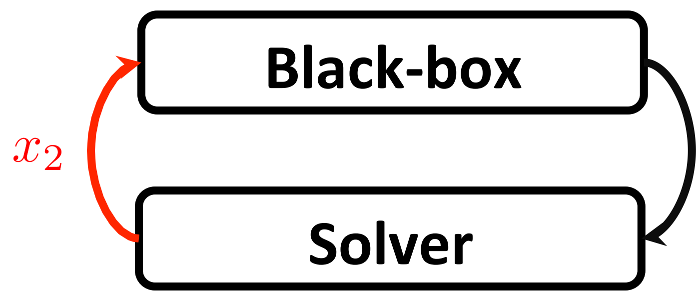
Pros:

Can solve problems where derivative information is unavailable, unreliable, or prohibitively expensive

Cons:

Slow and little to no guarantees that you have an optimal solution in the end





Three key components: Optimization algorithm

Derivative-free or black-box solvers

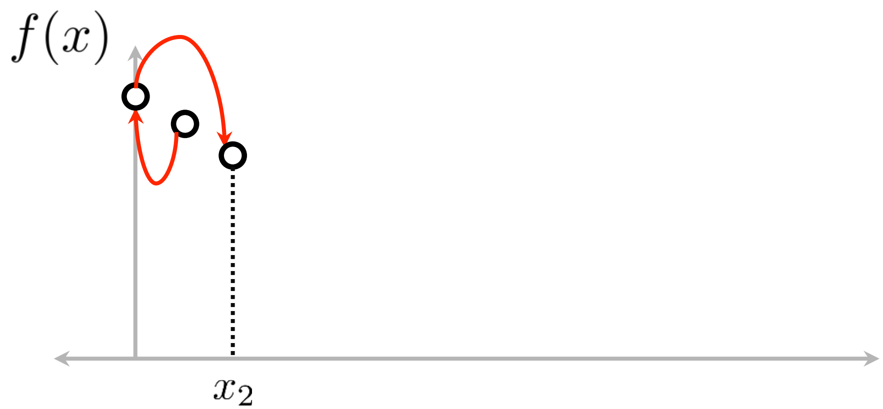
(Genetic algorithms, Matlab's fminsearch(), etc.)

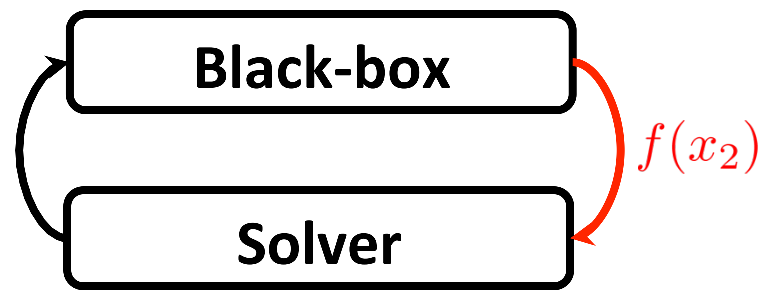
Pros:

Can solve problems where derivative information is unavailable, unreliable, or prohibitively expensive

Cons:

Slow and little to no guarantees that you have an optimal solution in the end





Three key components: Optimization algorithm

Derivative-free or black-box solvers

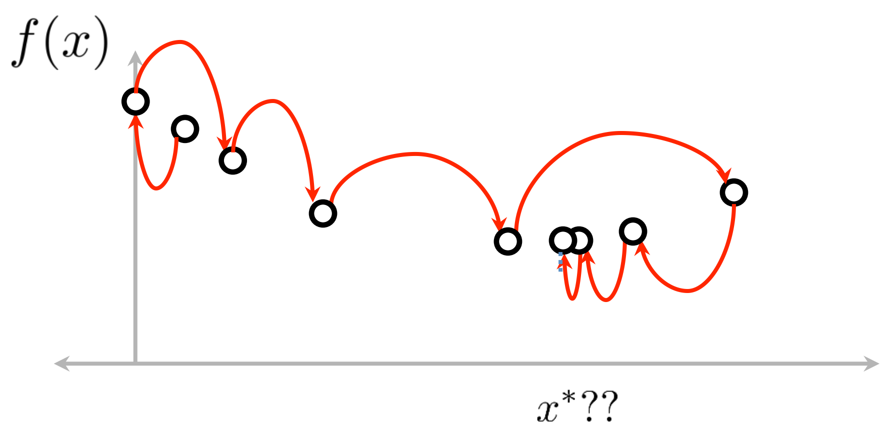
`(Genetic algorithms, Matlab's fminsearch(), etc.)`

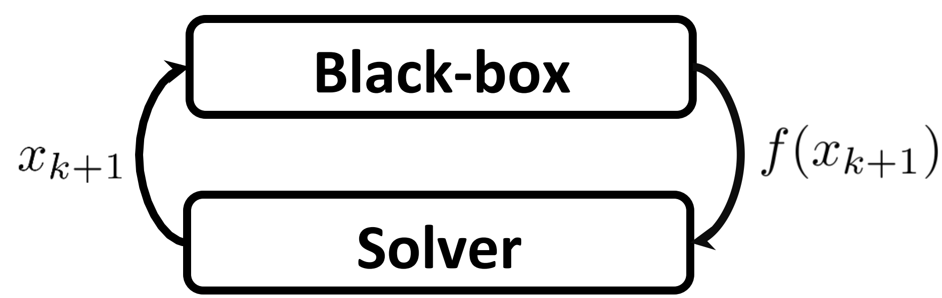
Pros:

Can solve problems where derivative information is unavailable, unreliable, or prohibitively expensive

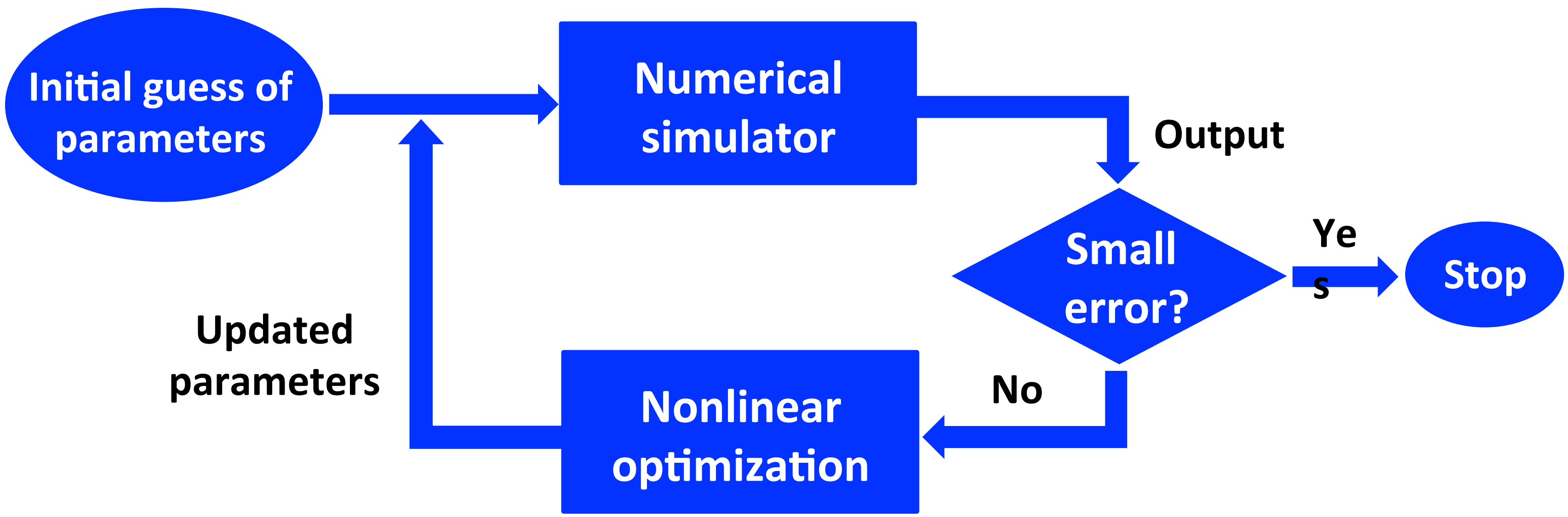
Cons:

Slow and little to no guarantees that you have an optimal solution in the end

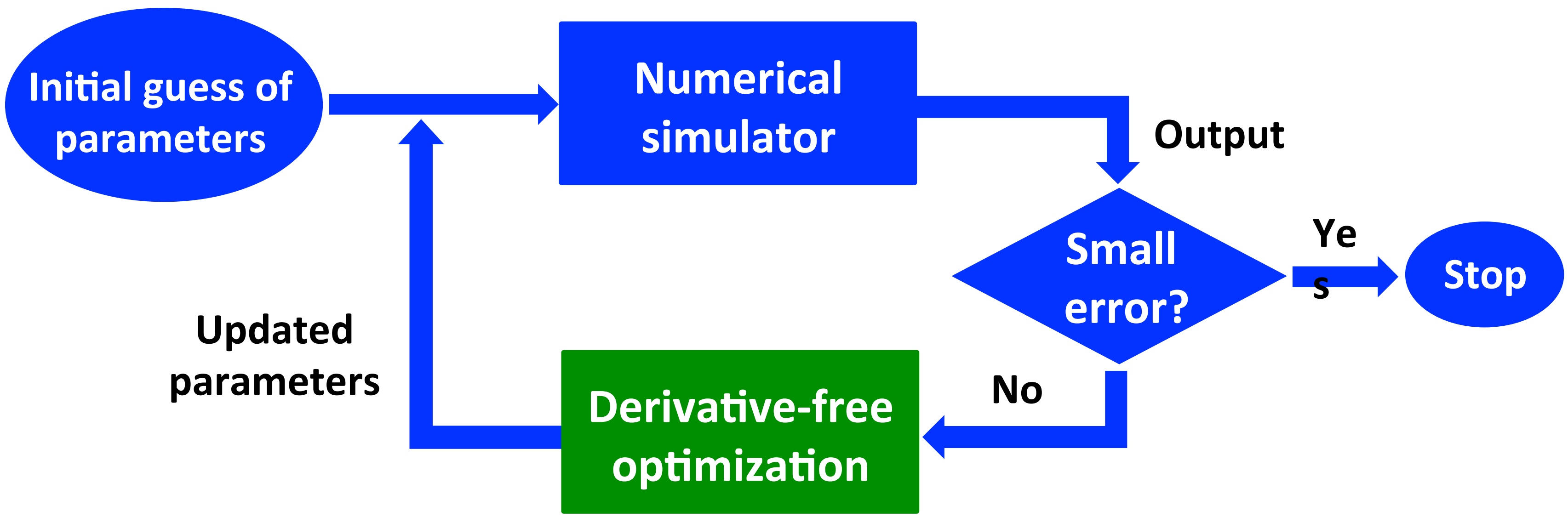




Procedure for estimating parameters

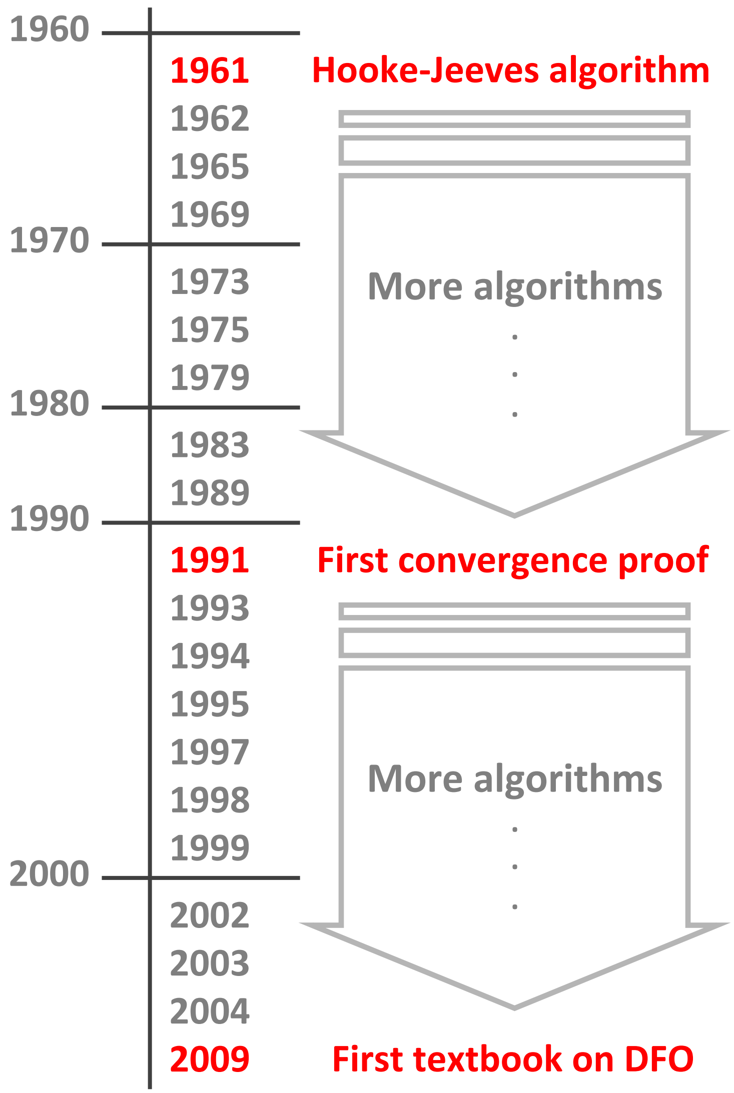


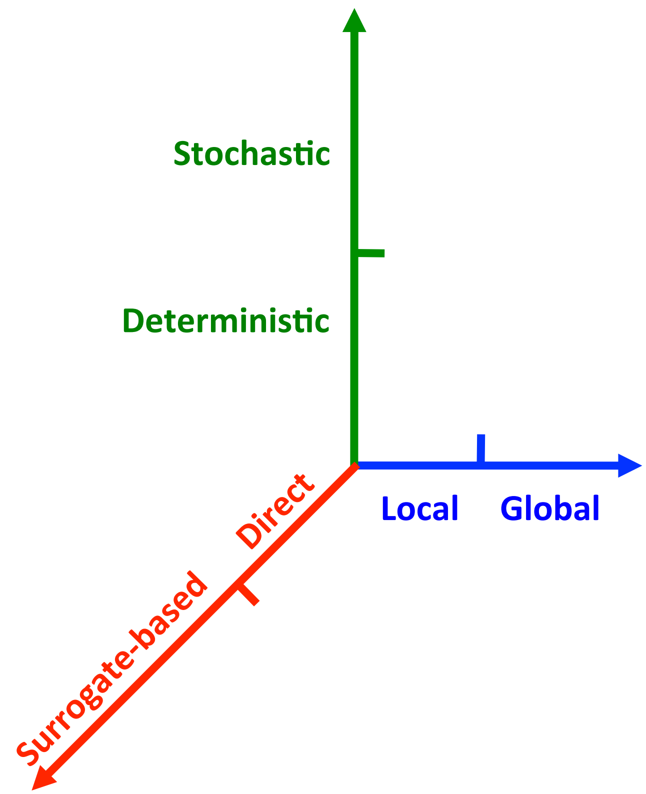
Procedure for estimating parameters



- **Performance of DFO solvers**
- **No systematic comparison of DFO solvers for parameter estimation**

Overview of Derivative-free optimization





Local DFO algorithms and software

Direct	• Nelder-Mead simplex algorithm	FMINSEARCH
	• Generalized pattern search and generating search set	DAKOTA PATTERN HOPSPACK SID-PSM NOMAD
Surrogate-based	• Implicit filtering	IMFIL
	• Trust-region methods	DFO BOBYQA NEWUOA

Global DFO algorithms and software

Direct

- Deterministic
    - Lipschitzian-based partitioning
    - Multilevel coordinate search

  `DAKOTA SOLIS-WETS`
  `DAKOTA DIRECT`
  `TOMLAB GLBSOLVE`
  `TOMLAB GLCSOLVE`
  `MCS`

- Stochastic
    - Hit-and-run
    - Simulated annealing
    - Genetic algorithms
    - Particle swarm

  `ASA`
  `CMA-ES`
  `DAKOTA`
  `EA`
  `GLOBAL`
  `PSWARM`

Surrogate-based

- Response surface methods

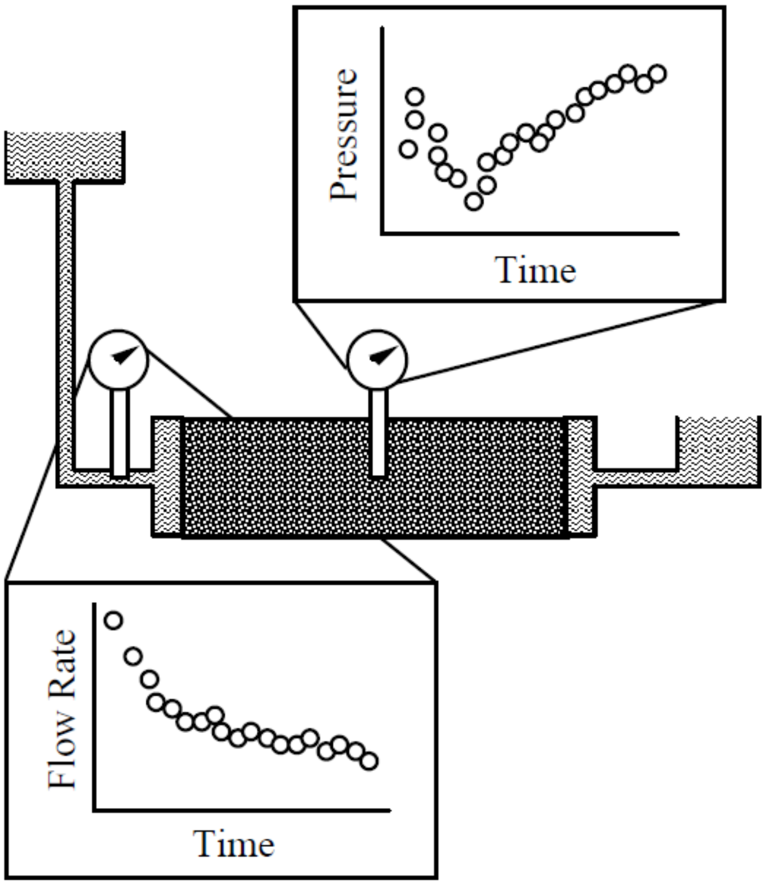
  `TOMLAB EGO`
  `TOMLAB RBF`

- Branch-and-fit

  `SNOBFIT`
  `TOMLAB LGO`

- Surrogate management framework

Case study I: Three parameters to be estimated



Water is injected at constant pressure into a 1D column filled with partially saturated sand

Known:

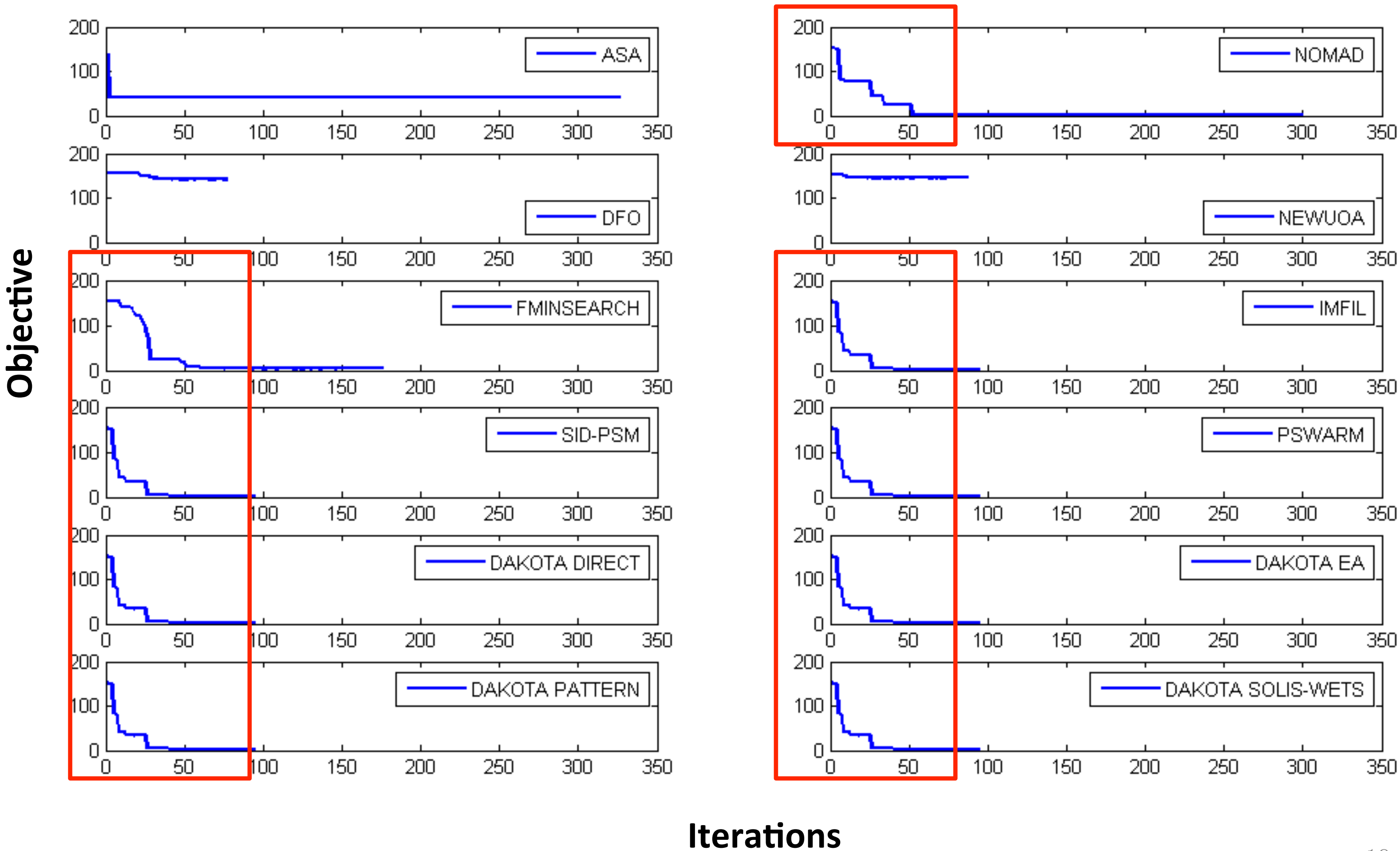
- **Pressure profile**
- **Flow rate profile**

Unknown:

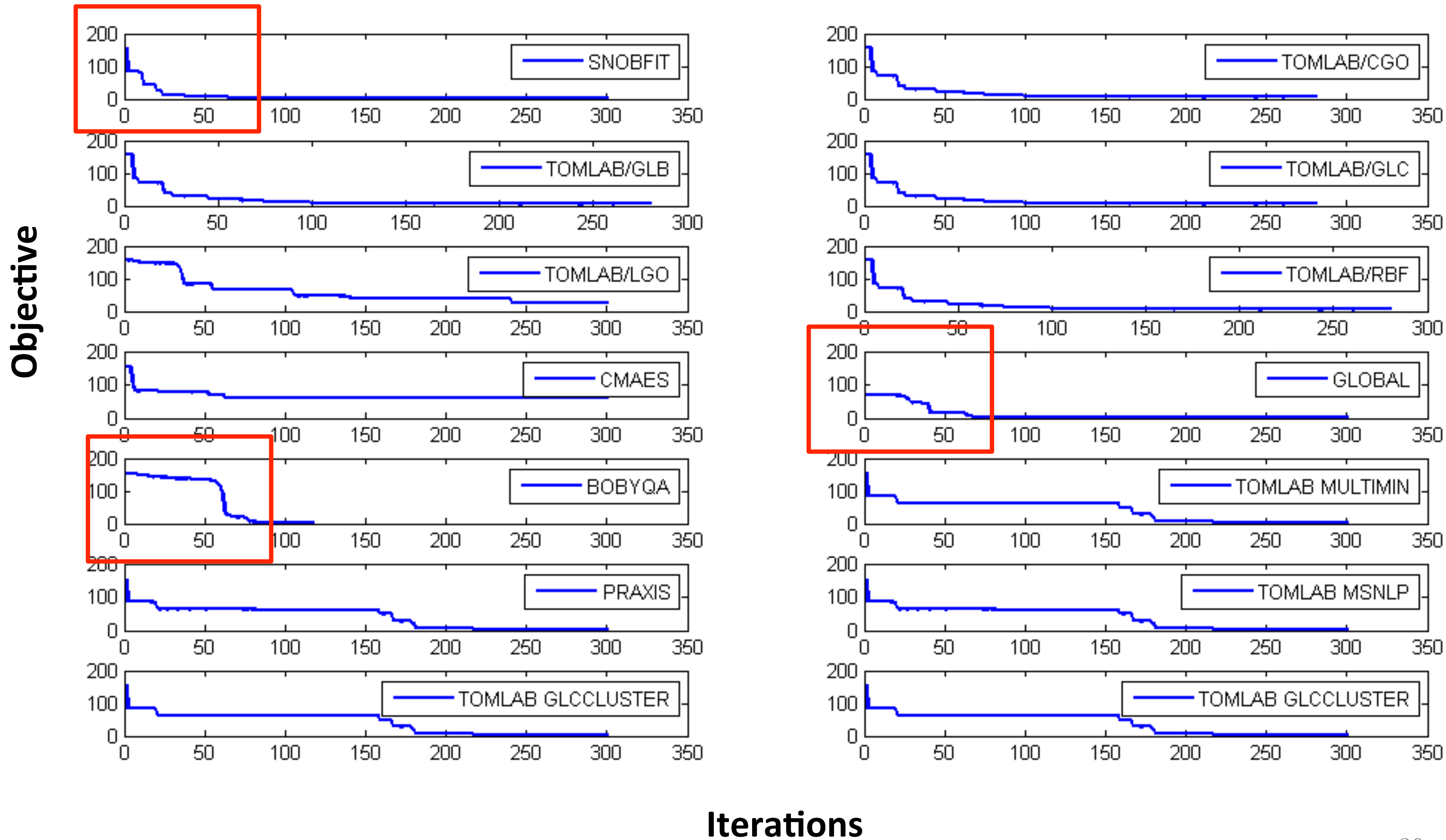
- **Porosity**
- **Permeability**
- **Initial gas saturation**

S. Finsterle, iTOUGH2 sample problems, LBNL-40042, 2007

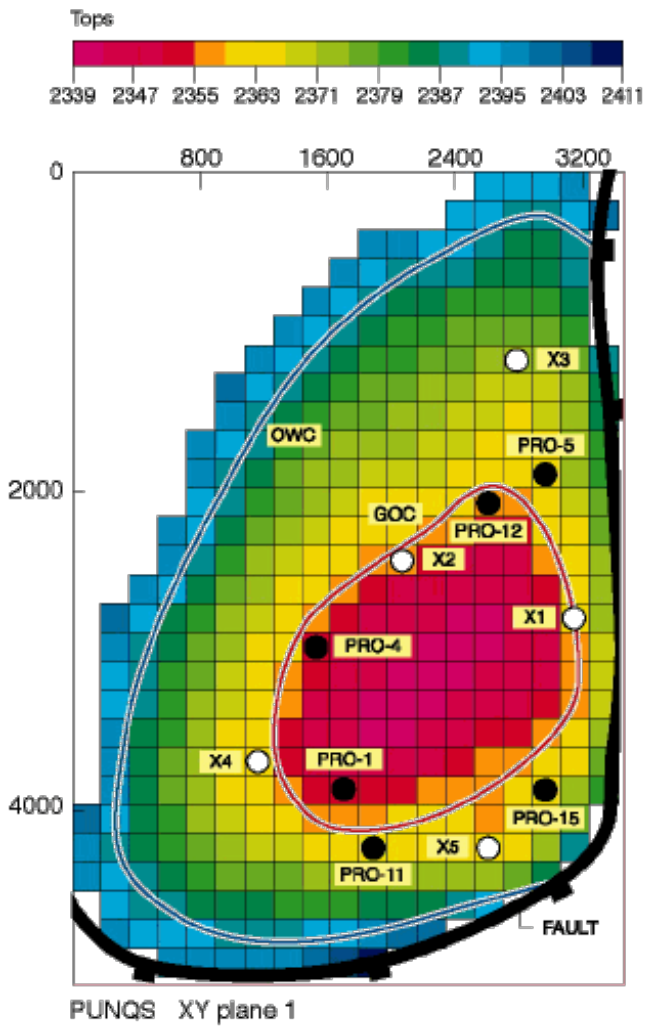
Performance of DFO solvers

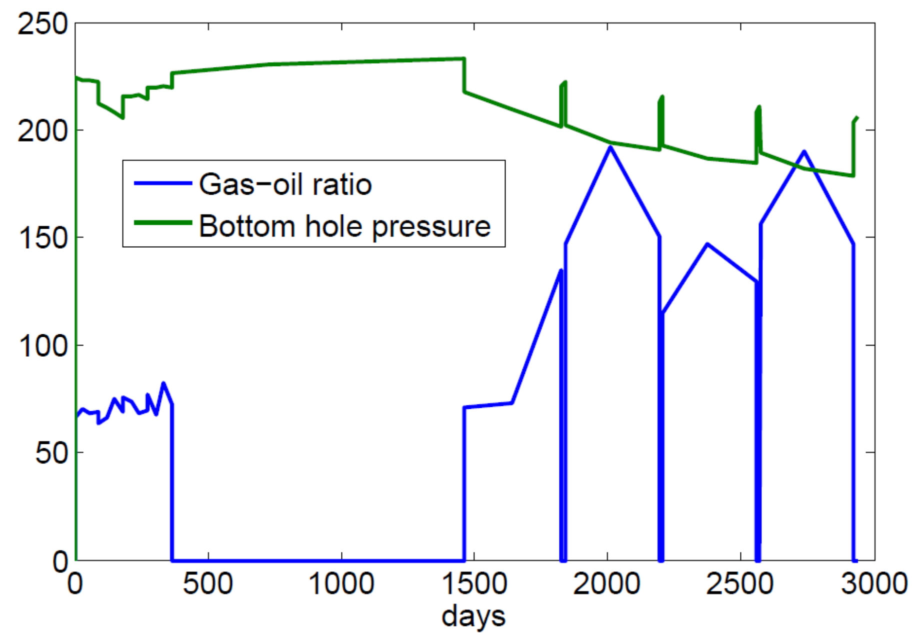


Performance of DFO solvers-cont'd



Case study II: 2660 parameters to be estimated



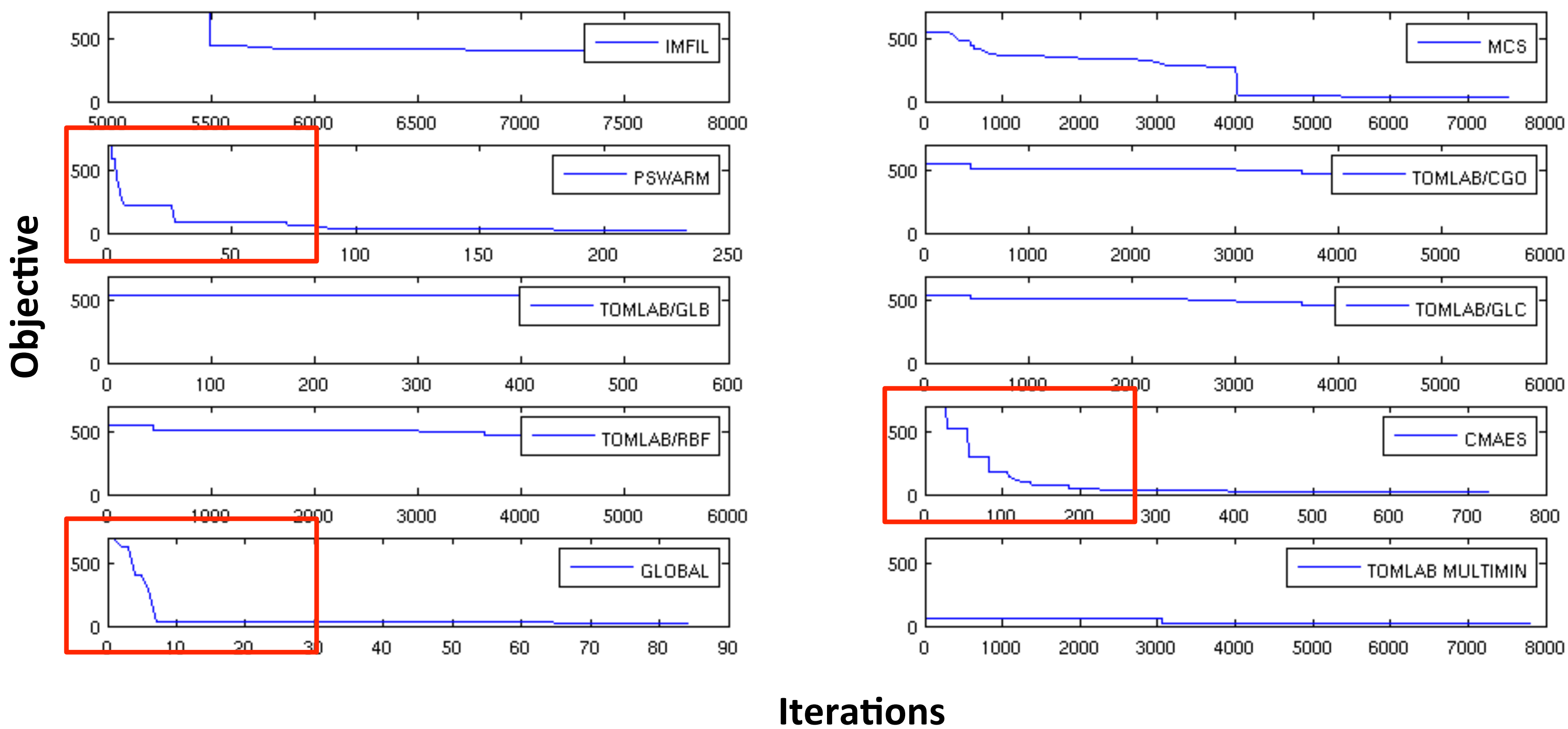


Production data for a well

- **Gridblocks: 19x28x5 = 2660**
- **2660 unknown porosities**
- **Permeability = function (porosity)**

http://www3.imperial.ac.uk/earthscienceandengineering/research/perm/punq-s3model

Performance of DFO solvers



Conclusions and future work

- **Conclusions**
  - **Systematic comparison of current DFO solvers in inverse modeling**
  - **DFO solver collection provides a robust package for inverse modeling**
    - **Small dimension problem:** `FMINSEARCH, SID-PSM, NOMAD, DAKOTA IMFIL, DFO, BOBYQA, NEWUOA`
    - **Large dimension problem:** `PSWARM, CMAES, GLOBAL`
  - **Performance depends on limits of computation resources such as CPU time and number of iterations**
- **Future work**
  - **Test with different measure of fitness**
  - **Apply DFO algorithms to permeability estimation of CO_2 sequestration in aquifer**