Calibration of Oil Reservoir Simulation Codes

Maurizio Filippone

Department of Statistical Science
University College London
maurizio@stats.ucl.ac.uk

February 8th, 2011

What is the meaning of calibration?

Given a model of a physical system, calibration means:

- frequentist: estimation of model parameters
- Bayesian: inference of model parameters

Motivating application

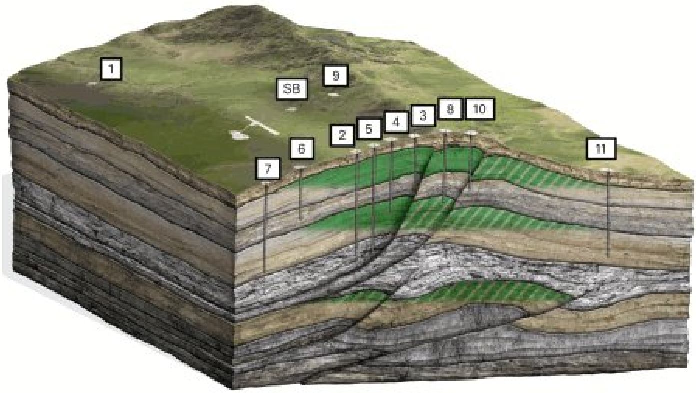
What is the problem?

- maximizing economic recovery of hydrocarbons from subsurface reservoirs

What's available?

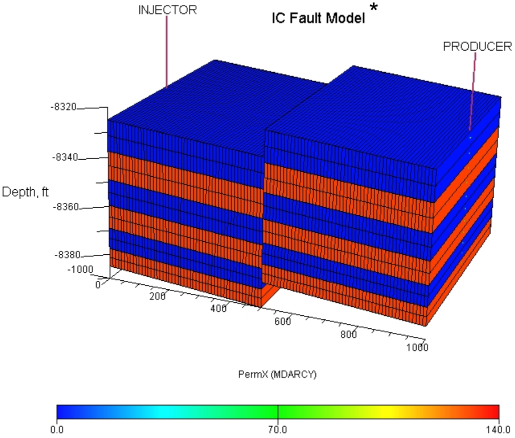
- models of the reservoirs
- geophysical data gathered on site

Example - Umiat oil field



Oil reservoir simulator

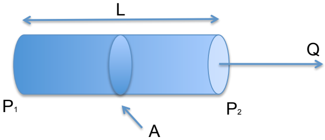
Eclipse oil reservoir simulation software by Schlumberger



Oil reservoir simulator

Reservoir simulator principles:

- conservation of energy and mass
- isothermal fluid phase behavior
- Darcy approximation (fluid flow through porous media):

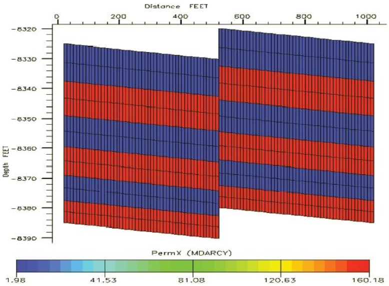


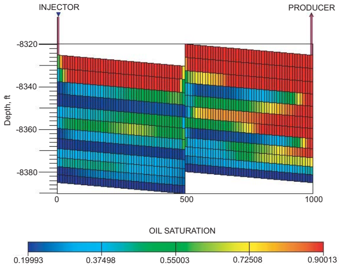
$$Q = -\frac{kA}{\mu}\frac{(P_2 - P_1)}{L}$$

- k permeability, μ viscosity
- Q discharge rate (m^3/s)

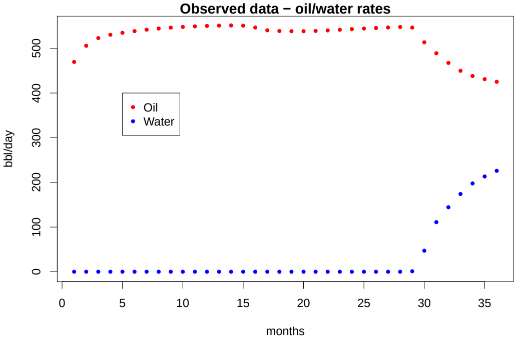
The problem - Oil reservoir model

- 2D 100 $\times$ 12 grid blocks (Tavassoli et al. 2004)
- Three parameters:
    - poor quality sand permeability (k_{low})
    - good quality sand permeability (k_{high})
    - discontinuity (throw)



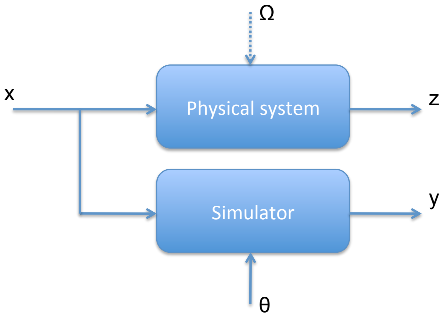


The problem - Oil reservoir model



- running time of the 2D oil reservoir model in Tavassoli et al. (2004): couple of seconds
- more complex and realistic 3D models - hours/days to run

Calibration of simulators



- We aim at inferring model parameters to quantify uncertainty in predictions

Relevance of the problem

- Calibration of simulators has application in very many application fields, e.g.:
  - high energy physics (Higdon et al. 2005)
  - geophysics (Cui et al. 2009)
  - astrophysics (Kaufman et al. 2010)
  - industrial processes (Forrester 2010)
  - ecology (Schneider et al. 2006)
  - climatology (Guillas et al. 2004, Wilkinson 2001)
  - systems biology (Wilkinson 2010)
- quantifying uncertainty is of paramount importance for balancing risks/costs of decisions

Importance of quantifying uncertainty

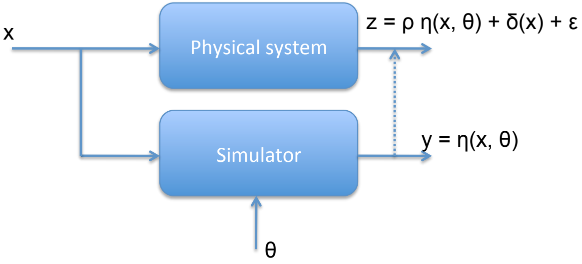
- this is what we might want
  - inferring model parameters
  - obtaining predictive distributions (balance cost of decisions)
  - Other desirables:
    - including prior information
    - approaching sequential estimation
    - doing model selection

Importance of quantifying uncertainty

- this is what we might want
  - inferring model parameters
  - obtaining predictive distributions (balance cost of decisions)
  - Other desirables:
    - including prior information
    - approaching sequential estimation
    - doing model selection
- Bayesian framework seems to be appropriate

Bayesian calibration of simulators

- Kennedy at al. (2001):

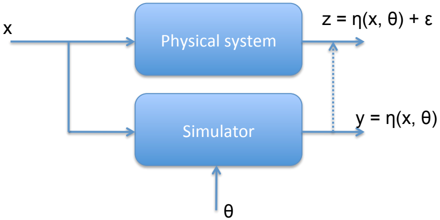


- $\eta(\mathbf{x}, \boldsymbol{\theta})$ and $\delta(\mathbf{x})$ independent and modeled using GPs
- $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ and i.i.d.

Interpretability of calibration results

- Can we really interpret the inferred simulator parameters as the “real” ones?
- if the model differs from the physical system, then the interpretation of θ is questionable
- Model misspecification and consequences in Bayesian inference - see e.g., White (1982) and Müller (2009)
- Loeppky et al. (2006) studied the problem of model misspecification in calibration problems

Simplified calibration model - Likelihood



- we have direct access to the log-likelihood:

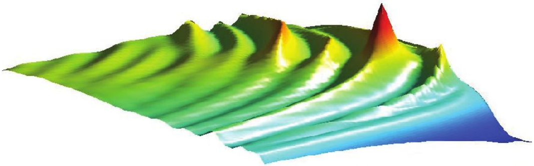
$$\log[p(\mathrm{Data}|\boldsymbol{\theta})] = \sum_i \log\left[\mathcal{N}(z_i|\eta(\mathbf{x}_i, \boldsymbol{\theta}_i), \sigma_\varepsilon^2)\right]$$

- if we place priors $\pi(\boldsymbol{\theta})$ we can infer $\boldsymbol{\theta}$ - simulation using Markov chain Monte Carlo (MCMC)

$$p(\boldsymbol{\theta}|\mathrm{Data}) \propto p(\mathrm{Data}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$$

Posterior inference - Multimodalities

- Given that the simulator is usually modeling a complex physical system, the likelihood is often multimodal

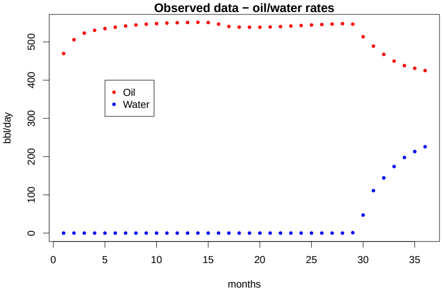


Oil reservoir data

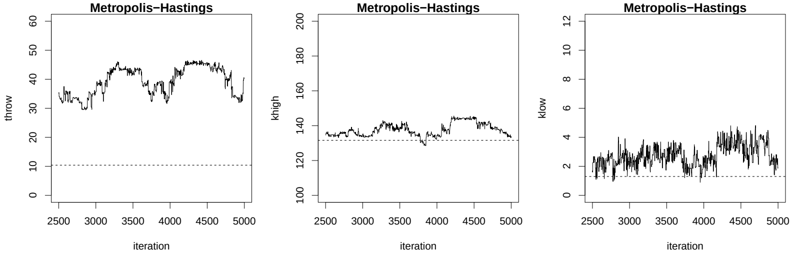
- As a working example let us consider the 2D oil reservoir model by Tavassoli et al. (2004)
- "Real" data are generated from the simulator:

$$z_i = \eta(\mathbf{x}_i, \tilde{\theta}) + \varepsilon_i$$

where $\tilde{\theta}$ represents a set of "true" parameters



Inference using Metropolis-Hastings



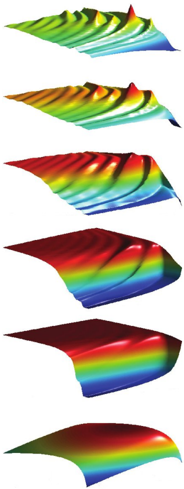
- Poor exploration of the parameter space
- Proper exploration of the parameter space via population based Markov chain Monte Carlo (Pop-MCMC)

Inference using Pop-MCMC

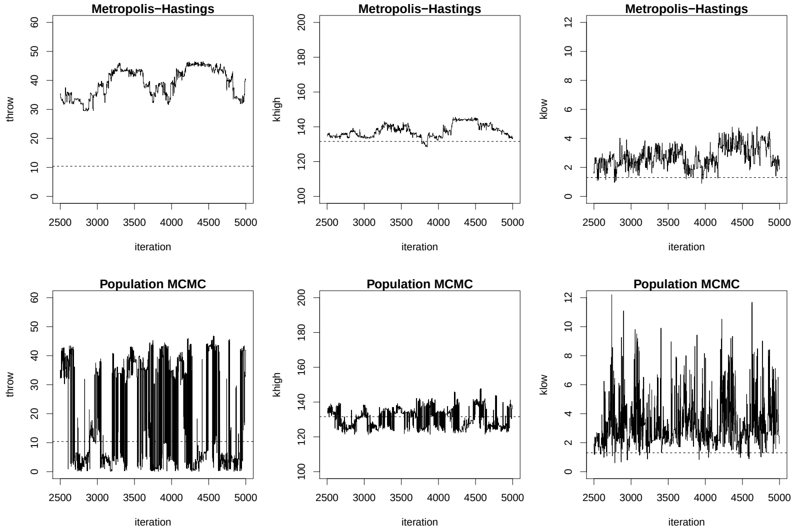
- bridge from the prior to the posterior via tempering

$$\text{tempered posterior} \propto p(\mathrm{Data}|\boldsymbol{\theta})^t \pi(\boldsymbol{\theta})$$

  with $t \in [0, 1]$
- one sampler for each tempered posterior
- samplers sampling independently and exchanging samples so that invariance of $p(\boldsymbol{\theta}|\mathrm{Data})$ is preserved

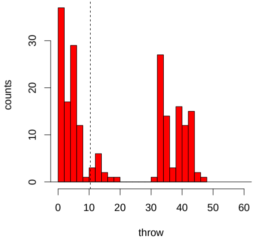


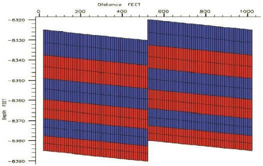
Inference using Pop-MCMC

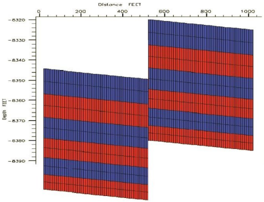


Inference using Pop-MCMC - Results

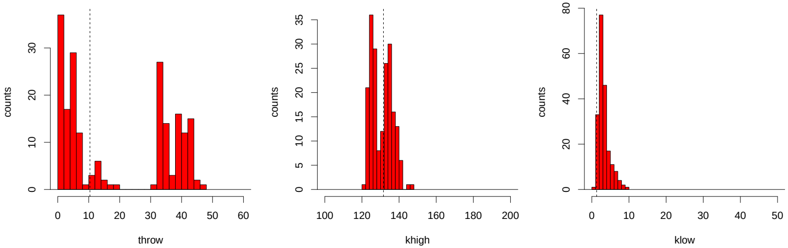
Interpretation of the multimodal posterior - throw parameter

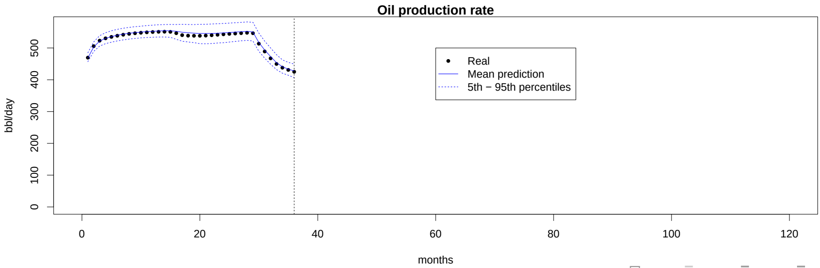




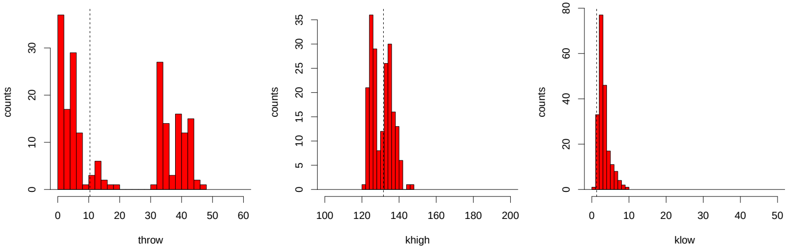


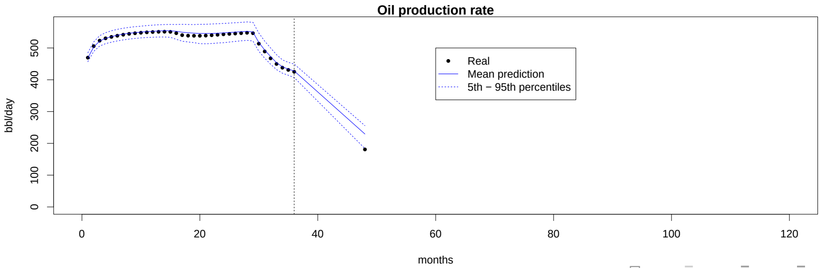
Inference using Pop-MCMC - Results



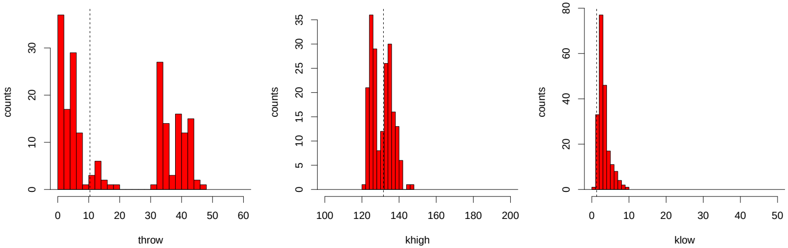


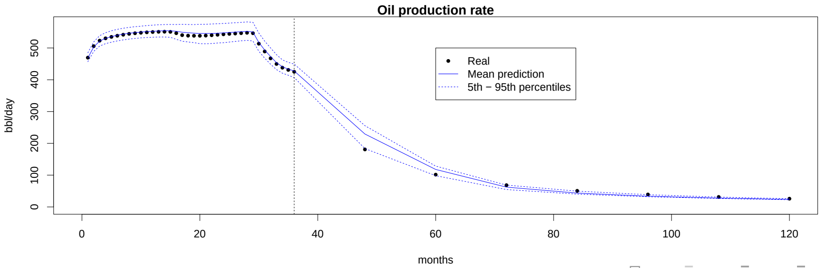
Inference using Pop-MCMC - Results





Inference using Pop-MCMC - Results



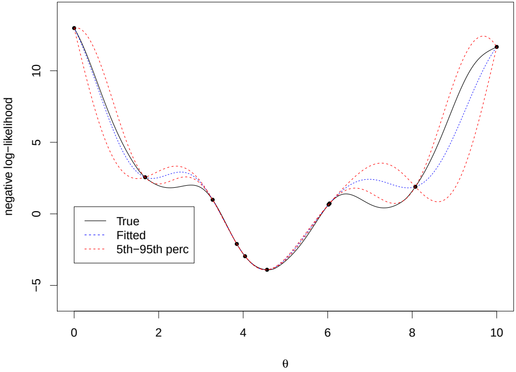


What if the running time of the simulator is prohibitive?

- Simulators could be so complex to require hours or even days to run
- Fully Bayesian treatment is not viable
- We aim at using the available computational resources to find parameters and an estimate of the uncertainty (not Bayesian)

Need for emulators

- Emulators as a proxy for the expensive likelihood
- Start from a set of design points (latin hypercubes)
- Emulator using Gaussian Process



Need for emulators

- Marginals of the emulator are Gaussian:

$$y(\boldsymbol{\theta}) \simeq \mathcal{N}(\mu(\boldsymbol{\theta}), s^2(\boldsymbol{\theta}))$$

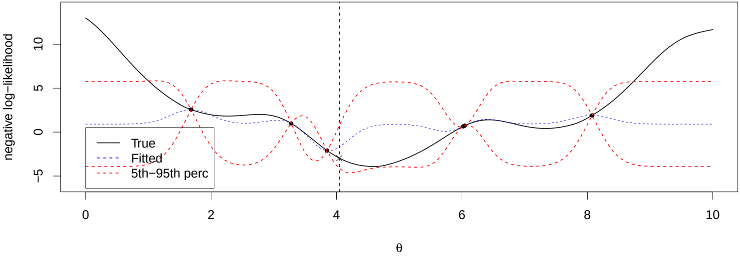
- Incremental design for minimization optimizing a utility (improvement) function (Jones et al. 1998)

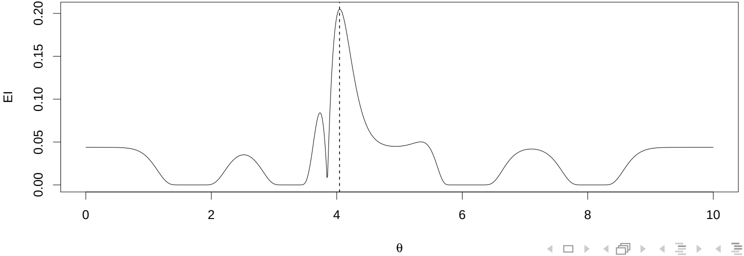
$$I(\boldsymbol{\theta}) = \max(f_{\text{min}} - y(\boldsymbol{\theta}), 0)$$

expected value:

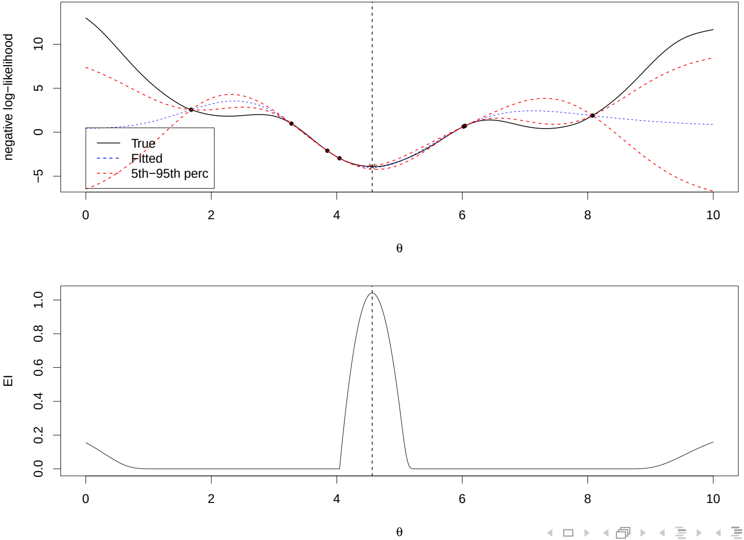
$$\mathrm{E}[I(\boldsymbol{\theta})] = (f_{\text{min}} - \mu(\boldsymbol{\theta}))\Phi\left(\frac{f_{\text{min}} - \mu(\boldsymbol{\theta})}{s(\boldsymbol{\theta})}\right) + s(\boldsymbol{\theta})\mathcal{N}\left(\frac{f_{\text{min}} - \mu(\boldsymbol{\theta})}{s(\boldsymbol{\theta})}\right)$$

Example of incremental design

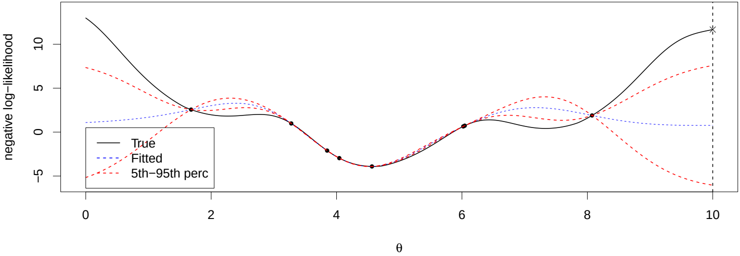


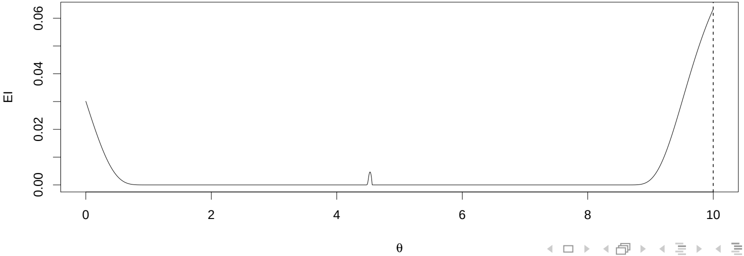


Example of incremental design

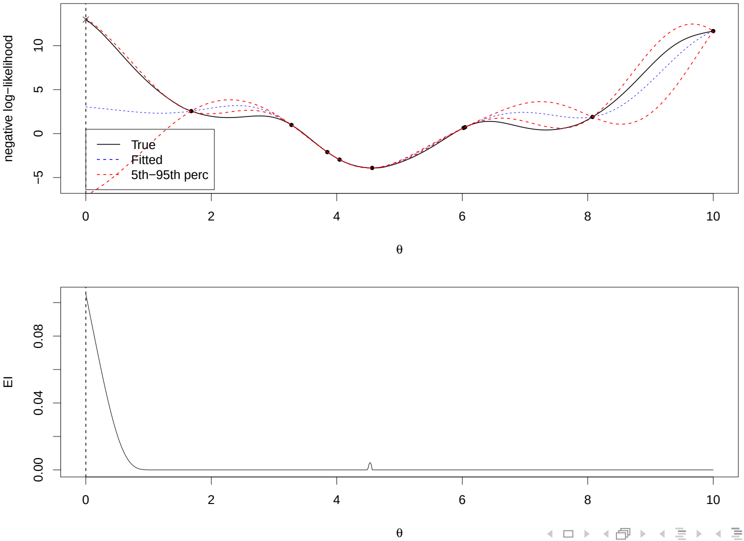


Example of incremental design

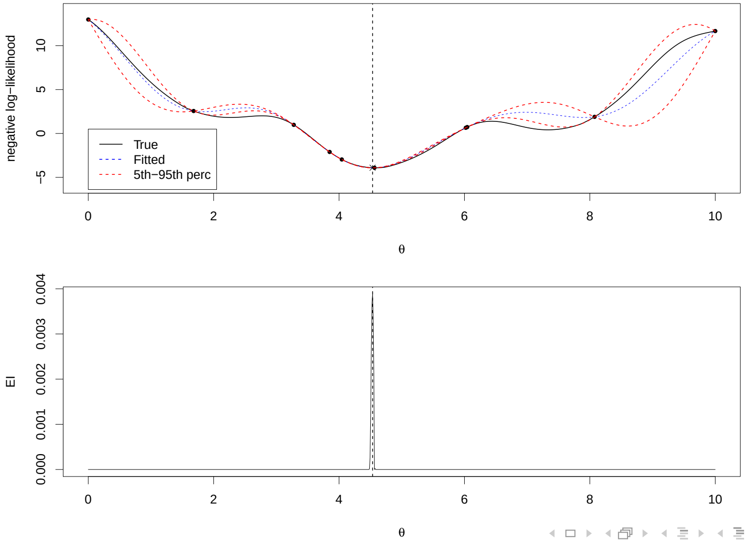




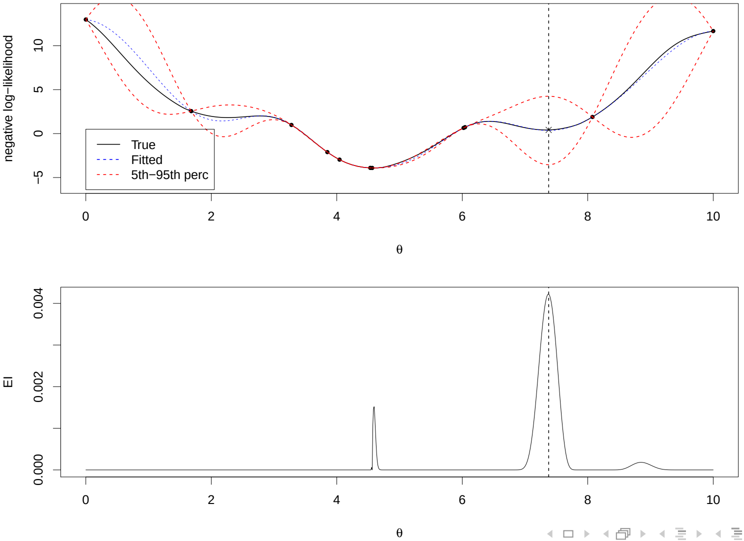
Example of incremental design



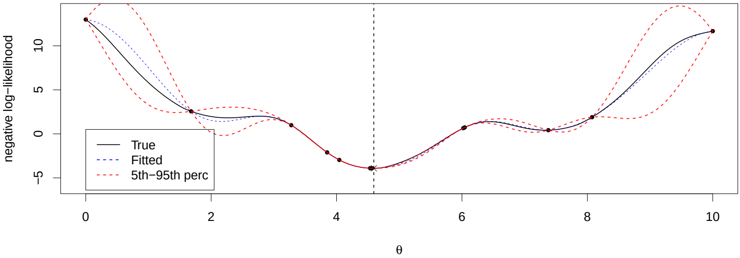
Example of incremental design

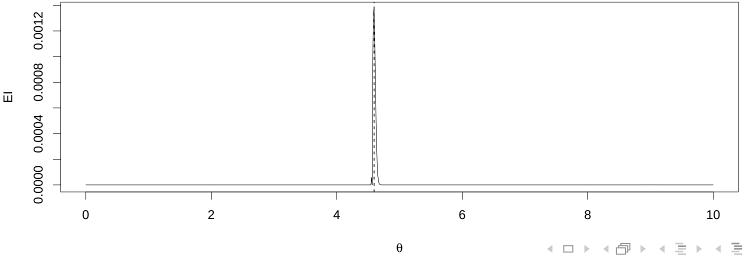


Example of incremental design

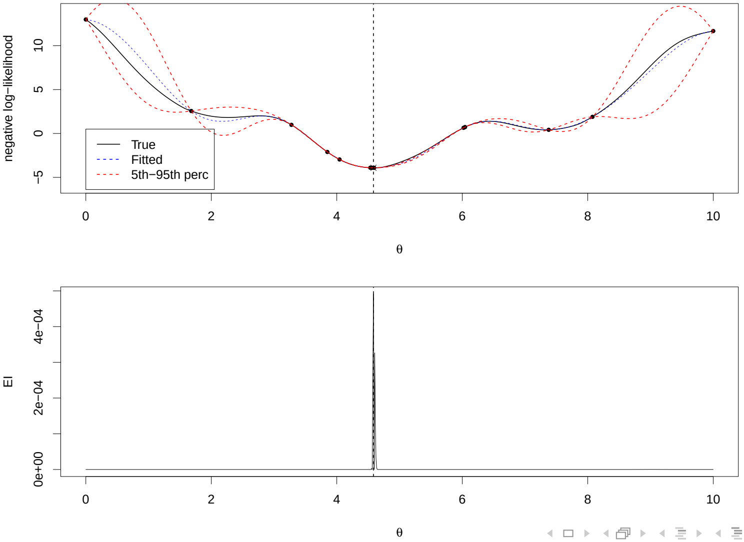


Example of incremental design

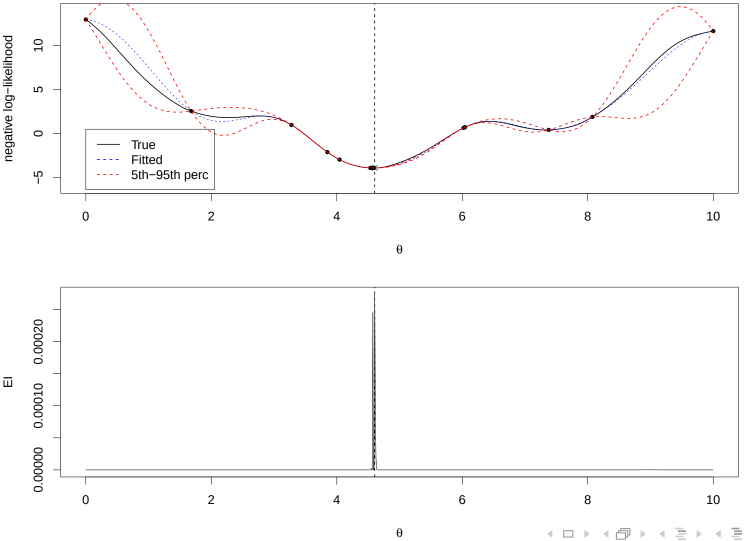




Example of incremental design



Example of incremental design

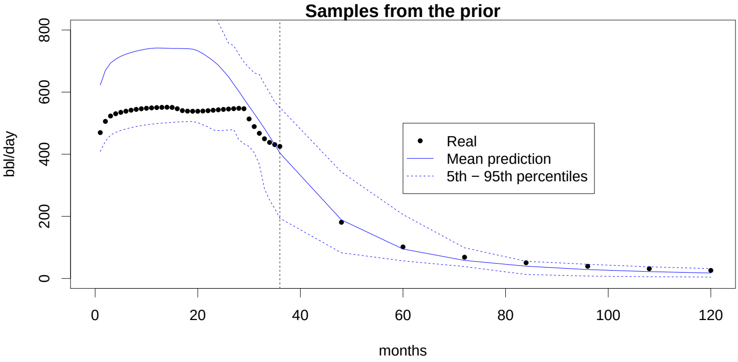


Incremental design - Oil reservoir - Results

- incremental design on the oil data
- we assumed that computational resources allowed only 100 runs of the simulator
  - 50 simulations using latin hypercubes
  - 50 simulations using incremental design

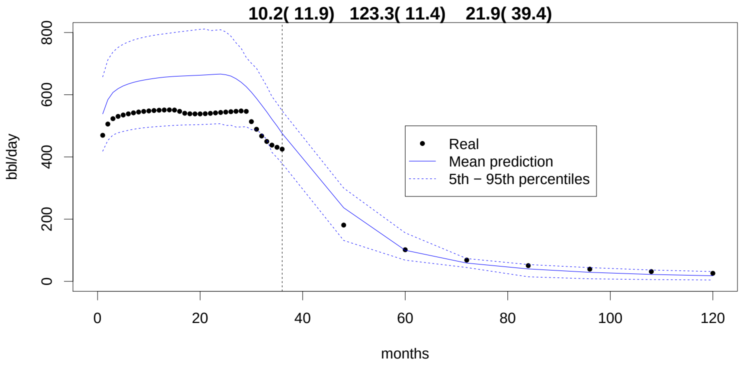
Incremental design - Oil reservoir - Results

true values: 10.4, 131.6, 1.3



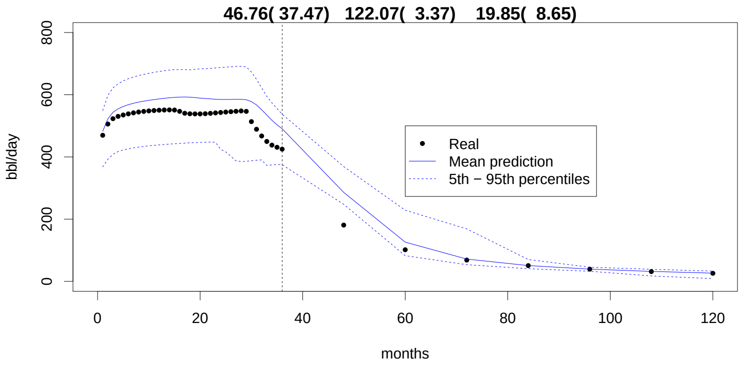
Incremental design - Oil reservoir - Results

true values: 10.4, 131.6, 1.3



Incremental design - Oil reservoir - Results

true values: 10.4, 131.6, 1.3



Conclusions and ongoing work

- This work is part of the ongoing research program of the Computational Statistics group at UCL on inference in complex systems
- Calibration of simulators is an important and promising research area
- It is a challenging problem even with simplifying assumptions:
  - inference over parameters of complex simulators
  - expensive to evaluate the likelihood

Conclusions and ongoing work

Ongoing research:

- Composite likelihoods
- Emulators for moderate/large sized data sets
- Consistency of estimators in incremental experiment design?
- Hybrid/Manifold Monte Carlo with emulator as potential field
- Design of covariance functions for emulators of differential equations based simulators
- Model selection

Acknowledgements

[1] L. Mohamed, B. Calderhead, M.Filippone, M. Christie, M. Girolami.
Population MCMC methods for history matching and uncertainty quantification.
Computational Geosciences. to appear

Collaborators:

- Mike Christie, Heriot Watt University, UK
- Derek Bingham, Simon Fraser University, Canada
- John Skilling, Maximum Entropy Data Consultants Ltd, UK

This research is funded by the Engineering and Physical Sciences Research Council (EPSRC), grant number EP/E052029/1.

EPSRC
Pioneering research and skills