

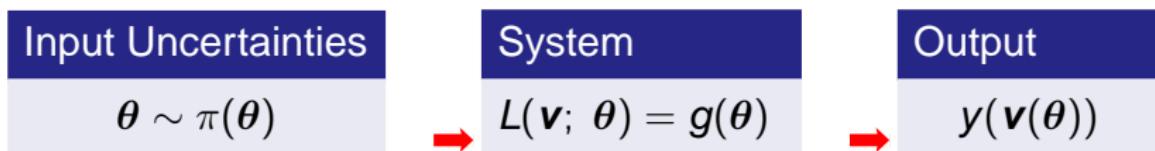
# Accurate Uncertainty Quantification Using Inaccurate Models

P.S. Koutsourelakis

Cornell University  
pk285@cornell.edu

California Institute of Technology  
January 14 2010

# Uncertainty Quantification



- Find *output* statistics, e.g:
  - $Pr[y(\theta) \geq y_0]$
  - $E[y(\theta)]$
  - $E[h(y(\theta))]$
  - ...

# Uncertainty Quantification

Some important characteristics:

- Uncertainties  $\theta \in \mathbb{R}^d$ ,  $d \gg 1$
- A *deterministic* solver for the governing equations of the system is available (e.g. Finite Element code)
- quite frequently, we are interested in *rare* events (i.e.  $Pr[y(\theta) \geq y_0] \ll 1$ ) or expectations of multimodal functions  $E[h(y(\theta))]$  of the output.

# Uncertainty Quantification

## What is available:

- The only general, (asymptotically) exact method is Monte Carlo
- A naive implementation can be extremely or prohibitively expensive
- Consider  $p_0 = \Pr[y(\theta) \geq y_0] \ll 1$  - then one needs to call *deterministic solver*  $\approx 10/p_0$  times.
- If  $p_0 = 10^{-3}$ , we need  $\approx 10^4$  calls to the deterministic solver.

# Uncertainty Quantification

## What is available:

- Advanced simulation techniques like **Sequential Monte Carlo (SMC)**
- Instead of sampling from  $\pi(\theta)$  we operate on a sequence of distributions that facilitate identification of *important* regions and accelerate convergence.
- Number of samples needed  $\approx 100 \log^2 p_0$ , so if  $p_0 = Pr[y(\theta) \geq y_0] = 10^{-3}$ , we need  $\approx 10^3$  calls to the deterministic solver (e.g. Au & Beck 2001, Johansen et al. 2006 )<sup>1</sup>.

---

<sup>1</sup>efficiency can be even greater for smaller  $p_0$

# Uncertainty Quantification

## Some notes:

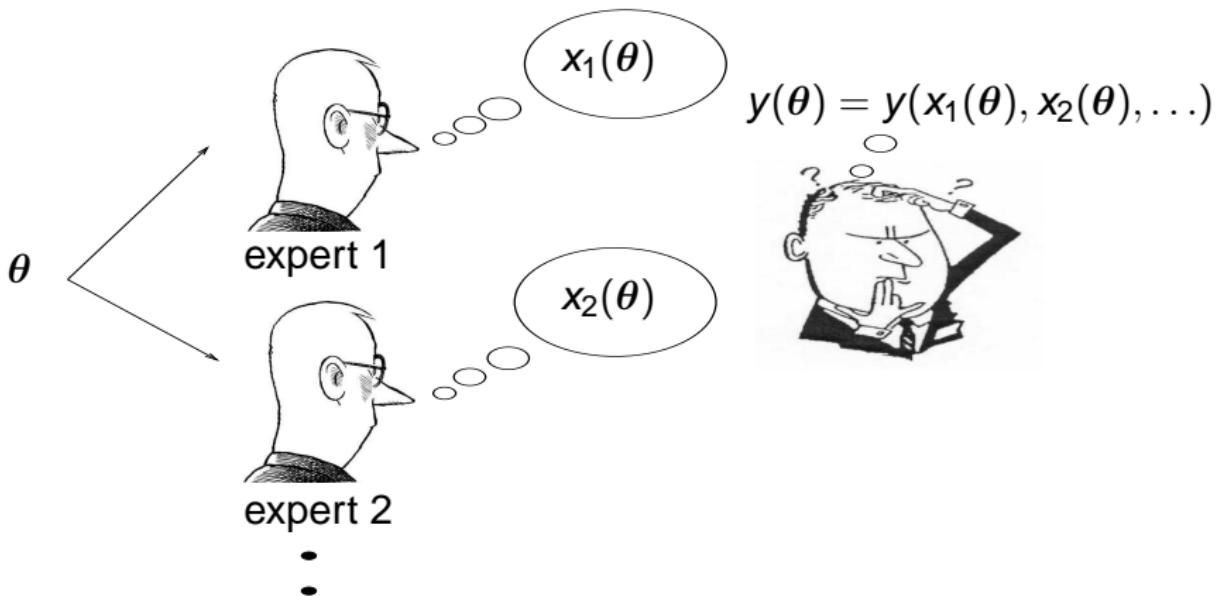
- The fact that Monte Carlo methods treat the deterministic solver as **black box** makes them more general, but at the same time doesn't exploit the analyst's knowledge/physical understanding/experience (if available)
- For example we know that, using a coarser grid in a Finite Element solution will not be accurate but nevertheless give *some indication* of the exact output.
- Similarly one can think of several **approximate solvers**, e.g:
  - ODE integrator with larger time step  $\Delta t$
  - Nonlinear solver using fewer Newton-Raphson iterations
  - ..
- Can we utilize, less-expensive and less-accurate predictions provided by *approximate solver(s)* in order to estimate consistently statistics of the *exact solver*?

# Uncertainty Quantification

## Some notes:

- In some cases, the *accurate model* might not even be available in the form of a computational model, but rather be an *experiment*.
- Example:
  - Consider a system (e.g. a structure) that is subjected it to a random excitation (e.g. earthquake) and we want to evaluate its response statistics (e.g. probability of failure)
  - We do not fully trust any of the computational models available.
  - We can carry out a few experiments under specified loading time histories, but obviously we want to restrict the number of experiments we perform.
  - *Design of Experiment*: Can we use the *approximate* computational models to reduce the number of the experiments?

# Methodology



# Methodology

- Problem: Estimate  $p_0 = \Pr[y(\theta) \geq y_0]$
- Assumptions:
  - 1 accurate and expensive model  $y(\theta)$
  - 2 approximate and inexpensive model(s)  $x(\theta)$
- The two extremes:

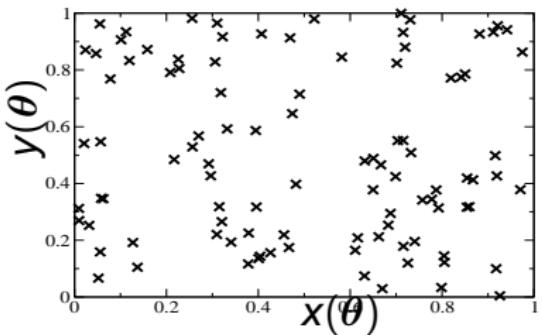


Figure: independent

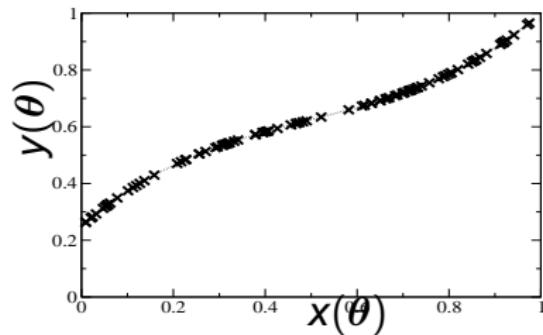


Figure: one-to-one

# Methodology

- Problem: Estimate  $p_0 = \Pr[y(\theta) \geq y_0] = \int \mathbf{1}_{y(\theta) \geq y_0} \pi(\theta) d\theta$
- But:<sup>2</sup>

$$p_0 = \int \Pr[y \geq y_0 | x] p(x) dx$$

where:

- $p(x(\theta))$  is the density of the *approximate solver*
- $\Pr[y \geq y_0 | x]$  is the conditional probability that  $y(\theta) \geq y_0$  when  $x(\theta) = x$ .
- We need to learn the conditional  $p(y | x)$
- Note that in the two extreme (degenerate) cases:
  - 1 “independent”:  $p(y | x) = p(y)$
  - 2 “one-to-one”:  $p(y | x) = \delta(y - h(x))$

---

<sup>2</sup>e.g. Kennedy & O'Hagan 2000, Au 2006

# Methodology

Two important points:

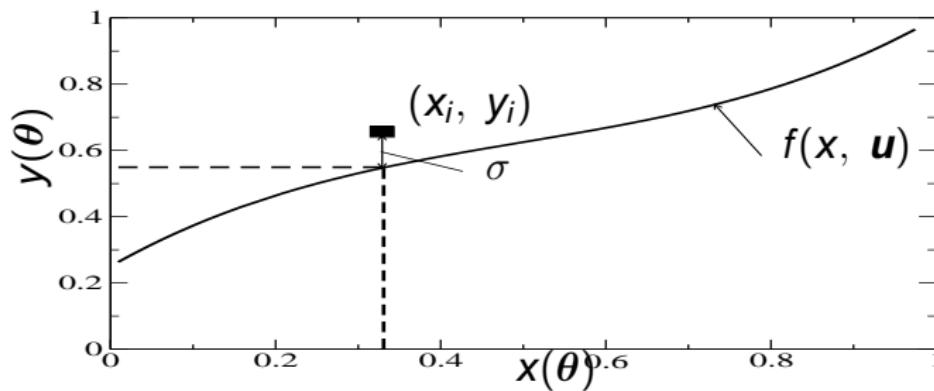
- The idea is that learning  $p(y | x)$  will be less expensive/more efficient than calculating  $p(y)$  directly.
- The cost is defined as the number of calls to the *exact solver* to evaluate  $y(\theta)$

# Learning from data

- Given a dataset  $D_n = \{x_i, y_i\}_{i=1}^n$ , we postulate:

$$y_i = f(x_i; \boldsymbol{u}) + \sigma Z_i, \quad Z_i \sim N(0, 1) \text{(i.i.d)}$$

where  $\boldsymbol{u}$  are *model parameters*



# Learning from data

- Nonparametric model

$$y_i = f(x_i; \boldsymbol{u}) + \sigma Z_i, \quad Z_i \sim \mathcal{N}(0, 1) \text{ (i.i.d.)}$$

where:

$$f(x_i; \boldsymbol{u}) = a_0 + \sum_{j=1}^k a_j K(x_i; x_j, \tau_j)$$

and:

$$K(x; x_j, \tau_j) = \exp\{-\tau_j \|x - x_j\|^2\}$$

- Unknowns:  $\boldsymbol{u} = \left\{ k, \{a_j\}_{j=0}^k, \{x_j\}_{j=1}^k, \{\tau_j\}_{j=1}^k \right\}$
- The number of kernels  $k$  (and as a result the number of parameters) is **not fixed** but inferred from the data.

# Nonparametric Bayesian Model

- Model:

$$f(x_i; \boldsymbol{u}) = a_0 + \sum_{j=1}^k a_j \exp\{-\tau_j \|x - x_j\|^2\}$$

- Unknowns  $\boldsymbol{u} = \left\{ k, \{a_j\}_{j=0}^k, \{x_j\}_{j=1}^k, \{\tau_j\}_{j=1}^k \right\}$
- (Hierarchical) Priors:
  - $p(k | \lambda) \propto \begin{cases} \frac{\lambda^k}{k!} & k \leq k_{max} \\ 0 & \text{otherwise} \end{cases}, p(\lambda) \propto s e^{-\lambda} s$
  - $p(\{a_j\}_{j=0}^k | \sigma_a^2) = \mathcal{N}(0, \sigma_a^2 \mathbf{I}), p(\sigma_a^2) = \text{InvGamma}(\gamma, \delta)$
  - $p(\{\tau_j\}_{j=1}^k) = \prod_{j=1}^k \frac{1}{\tau_j}$
  - $p(\{x_j\}_{j=1}^k) = \prod_{j=1}^k U(x_{min}, x_{max})$

# Nonparametric Bayesian Model

- Also:

$$y_i = f(x_i; \boldsymbol{u}) + \sigma Z_i$$

with prior  $p(\sigma^2) = \text{InvGamma}(\alpha, \beta)$

- Likelihood:

$$p(\underbrace{\{x_i, y_i\}_{j=1}^n}_{\text{data } D_n} \mid \boldsymbol{u}, \sigma^2) = \mathcal{N}(\mathbf{f}(\mathbf{x}; \boldsymbol{u}), \sigma^2 \mathbf{I})$$

- Posterior:

$$p(\boldsymbol{u}, \sigma^2 \mid \text{data } D_n) \propto \underbrace{p(\text{data } D_n \mid \boldsymbol{u}, \sigma^2)}_{\text{likelihood}} \underbrace{p(\boldsymbol{u}) p(\sigma^2)}_{\text{prior}}$$

# Methodology

- Problem: Estimate  $p_0 = \Pr[y(\theta) \geq y_0]$ , given data  $D_n = \{x_i, y_i\}_{i=1}^n$ , where:

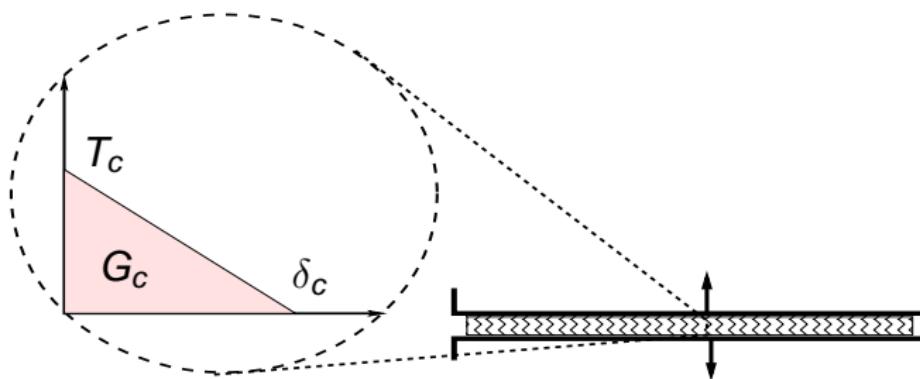
$$p_0 = \int \Pr[y \geq y_0 | x] p(x) dx$$

- Based on the nonparametric model that relates the two solvers  $x$  and  $y$ , and the data  $D_n$ :

$$\begin{aligned} \Pr[y \geq y_0 | x, D_n] &= \int p(y \geq y_0, (\boldsymbol{u}, \sigma^2) | x, D_n) d\boldsymbol{u} d\sigma^2 \\ &= \underbrace{\int p(y \geq y_0 | (\boldsymbol{u}, \sigma^2), x)}_{likelihood} \underbrace{p(\boldsymbol{u}, \sigma^2 | D_n)}_{posterior} d\boldsymbol{u} d\sigma^2 \\ &= \int \Phi\left(\frac{f(x; \boldsymbol{u}) - y_0}{\sigma}\right) p(\boldsymbol{u}, \sigma^2 | D_n) d\boldsymbol{u} d\sigma^2 \end{aligned}$$

# Numerical Example

- Consider a *cohesive interface*:

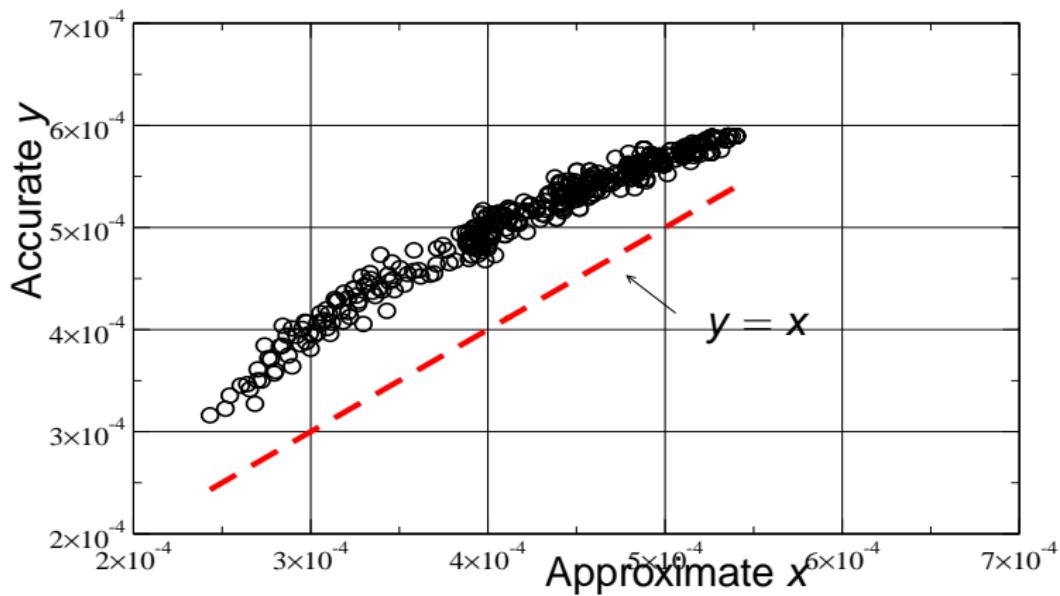


- Uncertainties  $\theta$  relate to the cohesive strength  $T_c$  and fracture energy  $G_c$  across the interface.

# Numerical Example

- Output  $y$ : Fracture energy for uniform separation of  $0.5 \times 10^{-3}$ .
- **Accurate model**: 1,000 cohesive elements with load increments of  $0.5 \times 10^{-6}$
- **Approximate Model**: 10 cohesive elements with load increments  $0.5 \times 10^{-5}$ .  $T_c$  is set to the minimum and  $G_c$  is averaged.
- Approximate model is  $\approx 1,069$  times faster than the accurate.

# Numerical Example



# Numerical Example

## ■ Approximate model

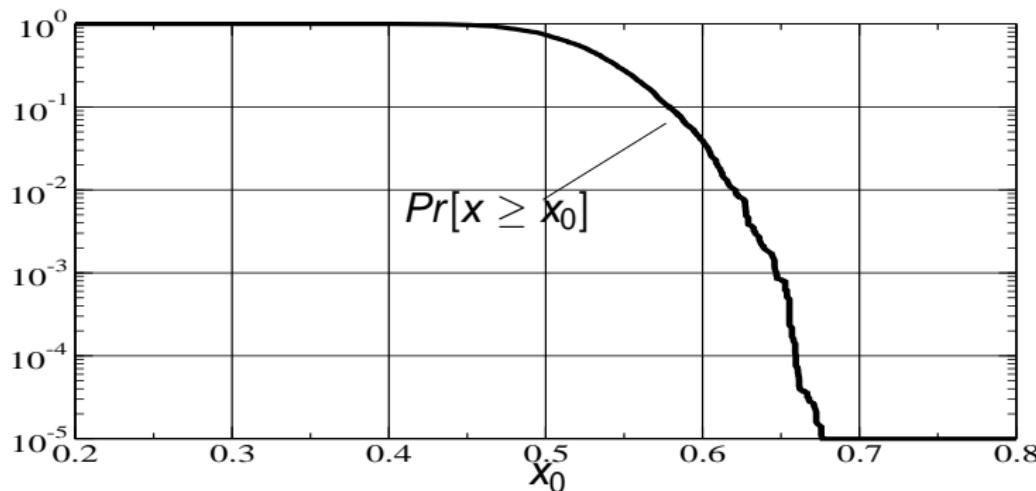


Figure: Computational Cost  $\approx 5$  calls of the accurate solver

# Numerical Example

- Model:  $y_i = f(x_i; \boldsymbol{u}) + \sigma Z_i$  - Data Points: 10

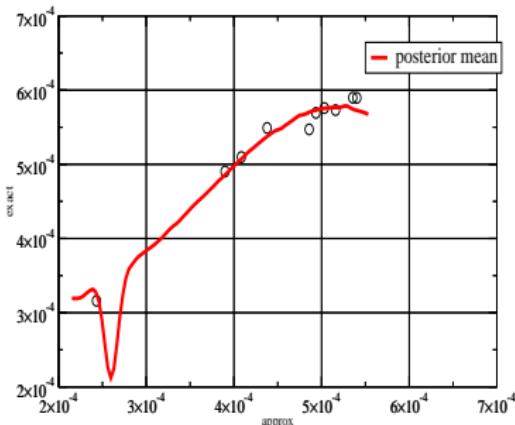


Figure: posterior of  $f(x; \boldsymbol{u})$

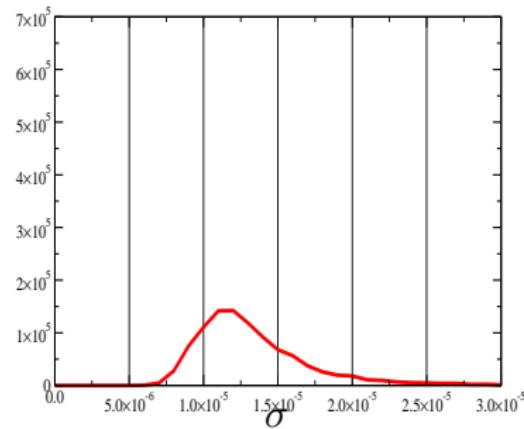


Figure: posterior of  $\sigma$

# Numerical Example

- Model:  $y_i = f(x_i; \boldsymbol{u}) + \sigma Z_i$  - Data Points: 50

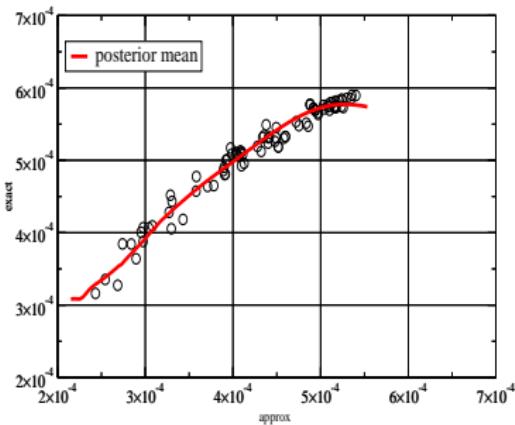


Figure: posterior of  $f(x; \boldsymbol{u})$

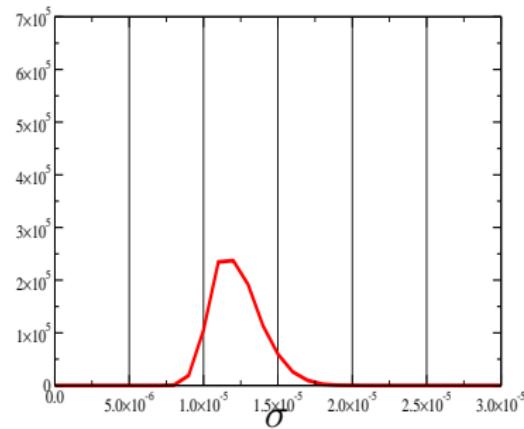


Figure: posterior of  $\sigma$

# Numerical Example

- Model:  $y_i = f(x_i; \boldsymbol{u}) + \sigma Z_i$  - Data Points: 150

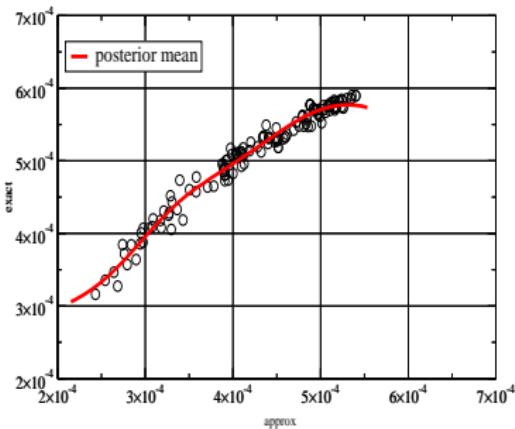


Figure: posterior of  $f(x; \boldsymbol{u})$

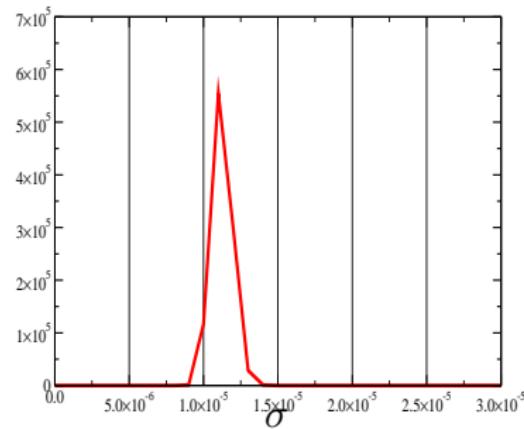


Figure: posterior of  $\sigma$

# Numerical Example

- Model:  $y_i = f(x_i; \boldsymbol{u}) + \sigma Z_i$  - Data Points: **10**

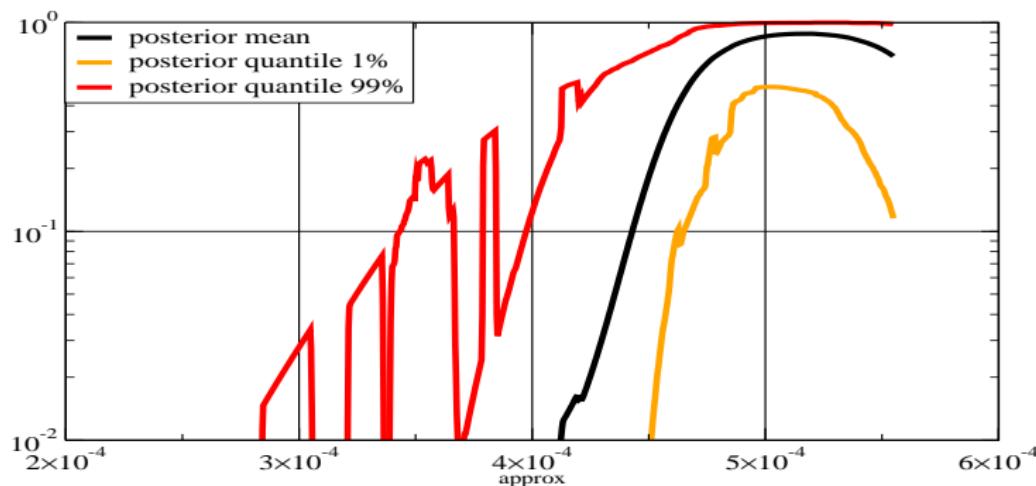


Figure: posterior  $Pr[y \geq 5.601 \times 10^{-4} | x, \text{ data } D_{10}]$

# Numerical Example

- Model:  $y_i = f(x_i; \boldsymbol{u}) + \sigma Z_i$  - Data Points: 50

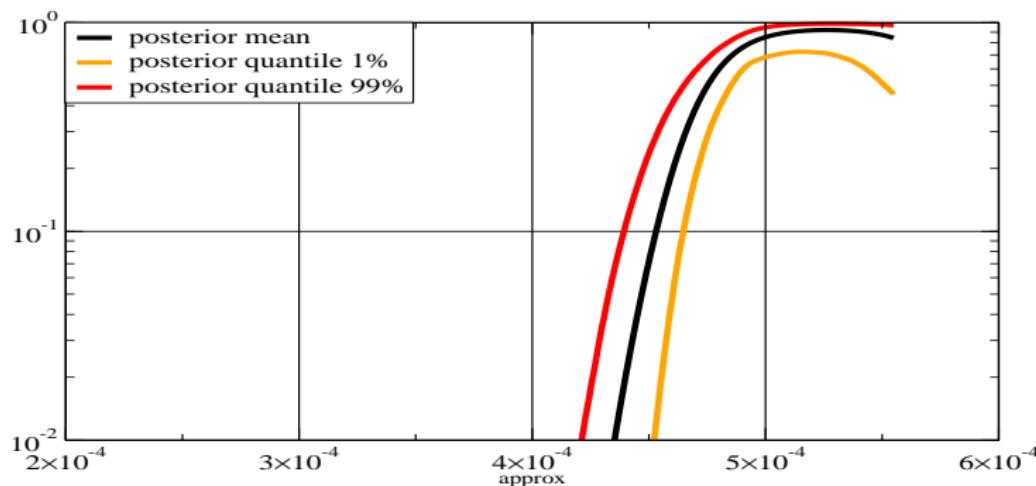


Figure: posterior  $\Pr[y \geq 5.601 \times 10^{-4} | x, \text{ data } D_{50}]$

# Numerical Example

- Model:  $y_i = f(x_i; \boldsymbol{u}) + \sigma Z_i$  - Data Points: **150**

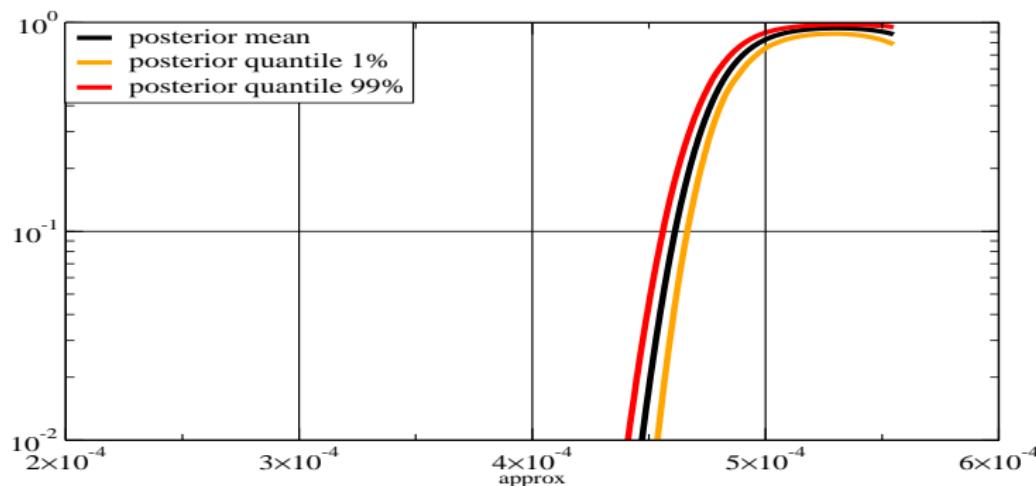


Figure: posterior  $\Pr[y \geq 5.601 \times 10^{-4} | x, \text{ data } D_{150}]$

# Numerical Example

- $p_0 = \Pr[y \geq 5.601 \times 10^{-4}] = 1.00 \pm 0.23 \times 10^{-3}$  - with  
1,500 calls to accurate model

Number of samples	Posterior mean	Posterior quantile 1%	Posterior quantile 99%
10	$6.36 \times 10^{-3}$	$5.91 \times 10^{-4}$	$7.17 \times 10^{-2}$
50	$1.75 \times 10^{-3}$	$7.39 \times 10^{-4}$	$3.55 \times 10^{-3}$
150	$1.01 \times 10^{-3}$	$7.07 \times 10^{-4}$	$1.42 \times 10^{-3}$

# Numerical Example

- Because we learn  $p(y | x)$  we can calculate statistics of any event related to  $y$

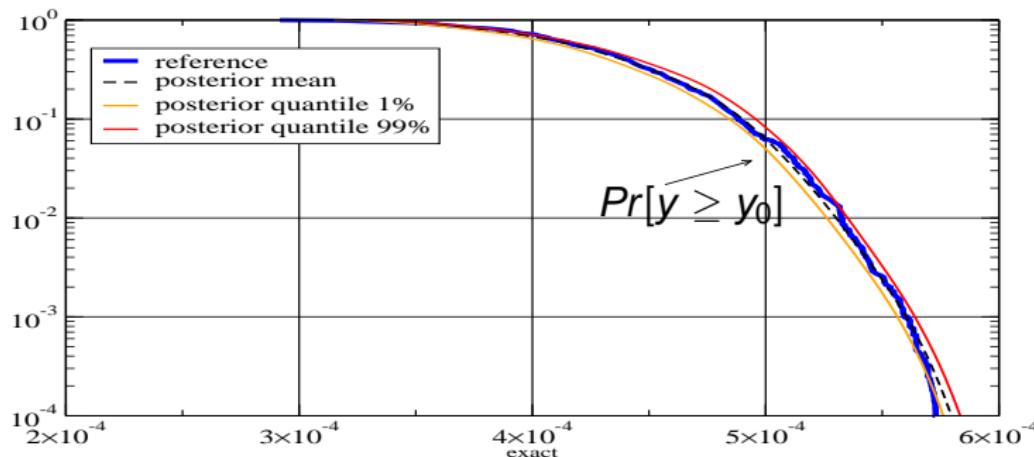


Figure: posterior for  $\Pr[y \geq y_0 | \text{data } D_{150}]$

# Numerical Example

- Consider a random heterogeneous medium occupied by two-phases of *elastic-perfectly plastic* materials with yield

$$\text{stresses } \frac{\sigma_y^{(\text{inclusion})}}{\sigma_y^{(\text{matrix})}} = 10$$

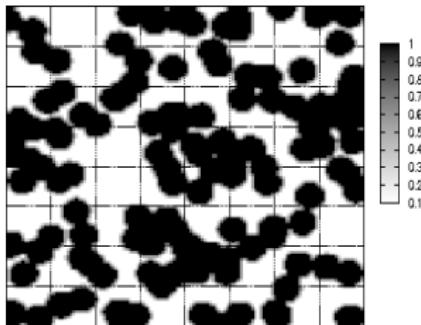


Figure: low strength

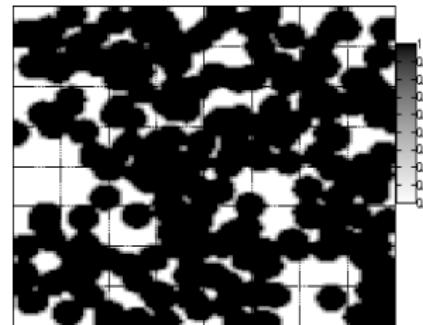


Figure: high strength

- Uncertainties  $\theta$  relate to the random geometry i.e. number and locations of disks (following a Boolean model)

# Numerical Example

- **Accurate model:**  $128 \times 128$  FE mesh (i.e. 11 elements per inclusion disk diameter) with  $\approx 33,000$  dof.
- **Approximate Model:**  $8 \times 8$  FE mesh with 135 dof, using as the yield stress within each element the *log-average*.

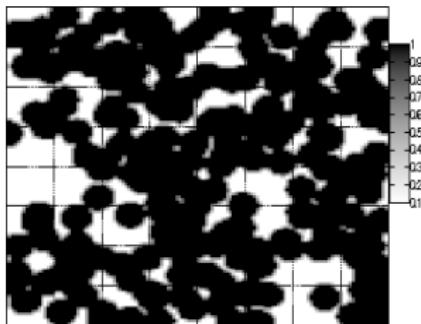


Figure: average comp. time  
700 sec

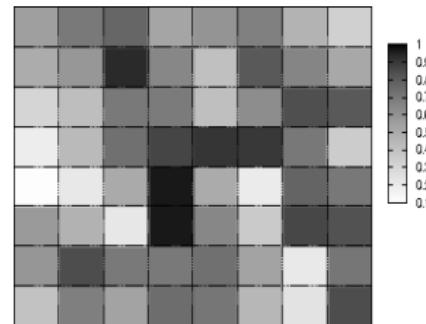
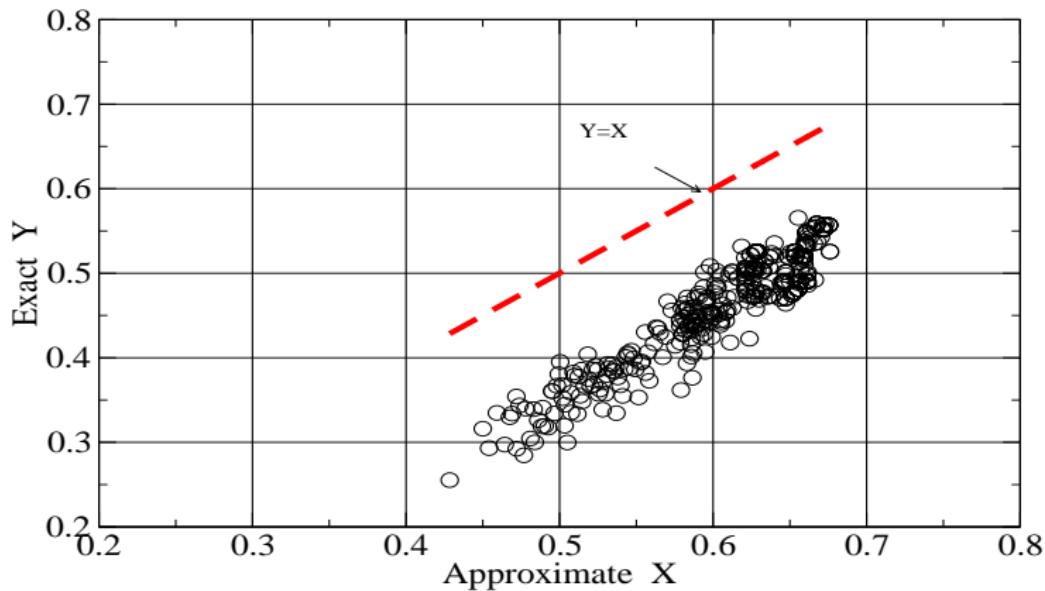


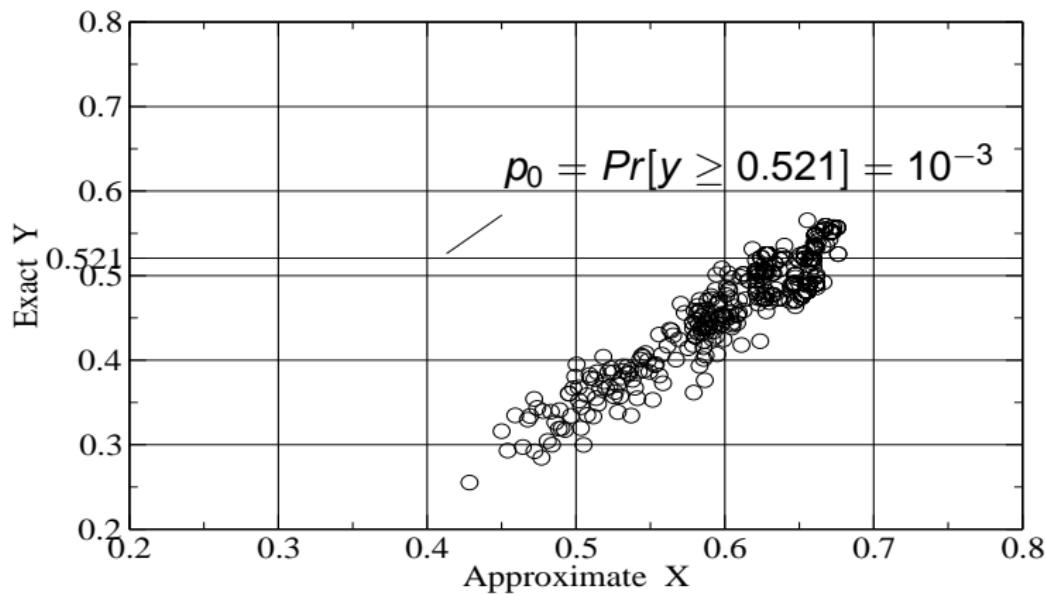
Figure: average comp. time  
0.15 sec

- Approximate model is  $\approx 5,000$  times faster than the accurate.

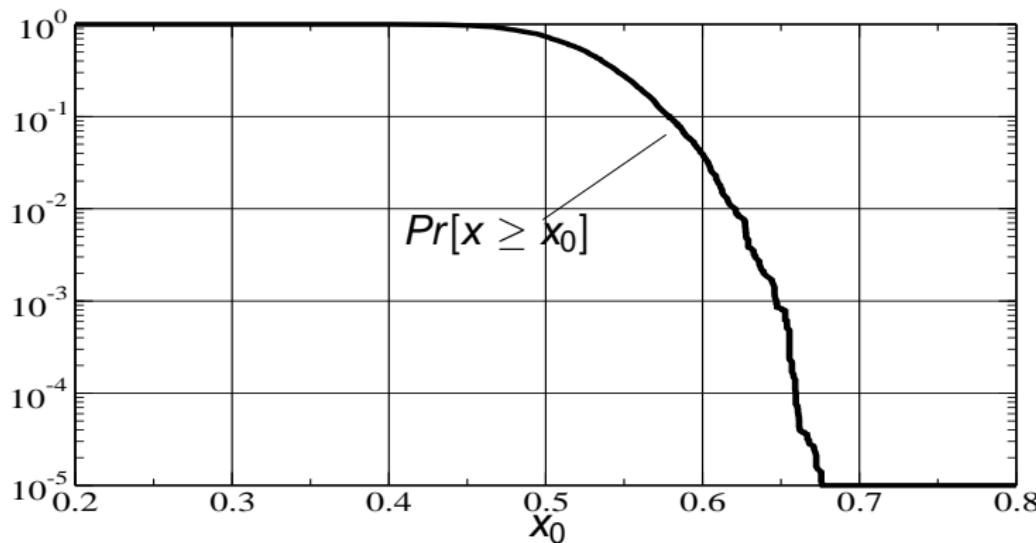
# Numerical Example



# Numerical Example



# Numerical Example



**Figure:** Computational Cost  $\approx 1$  call of the accurate solver

# Numerical Example

- Model:  $y_i = f(x_i; \boldsymbol{u}) + \sigma Z_i$  - Data Points: 10

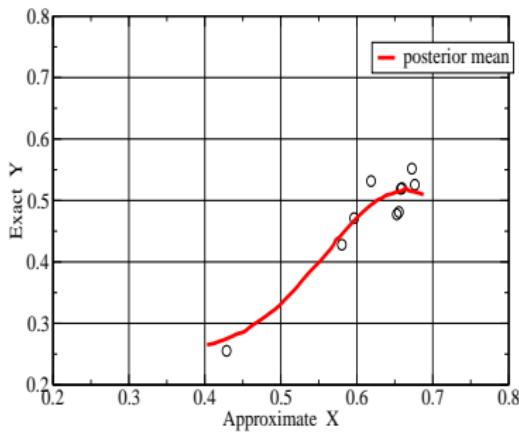


Figure: posterior of  $f(x; \boldsymbol{u})$

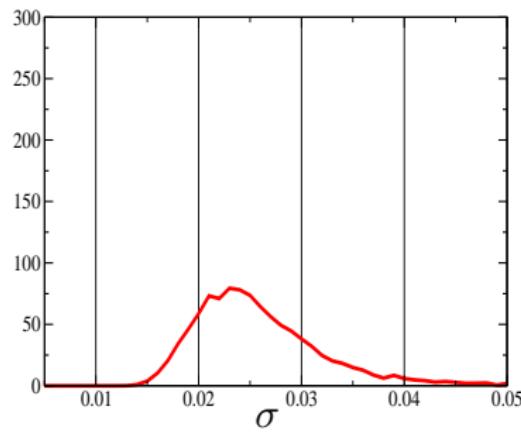


Figure: posterior of  $\sigma$

# Numerical Example

- Model:  $y_i = f(x_i; \boldsymbol{u}) + \sigma Z_i$  - Data Points: 20

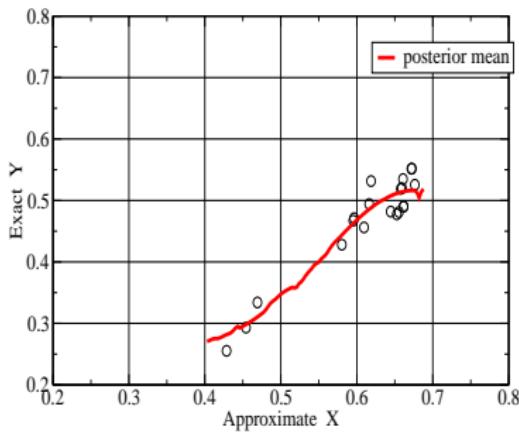


Figure: posterior of  $f(x; \boldsymbol{u})$

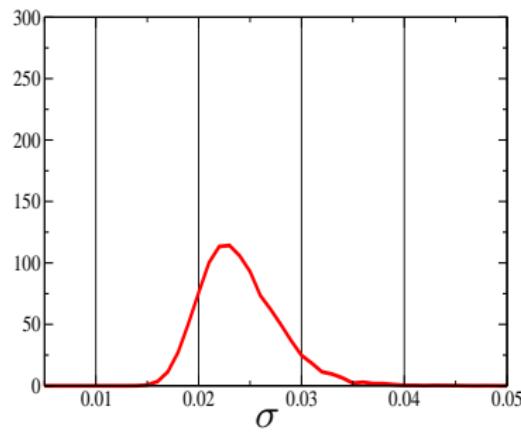


Figure: posterior of  $\sigma$

# Numerical Example

- Model:  $y_i = f(x_i; \boldsymbol{u}) + \sigma Z_i$  - Data Points: 30

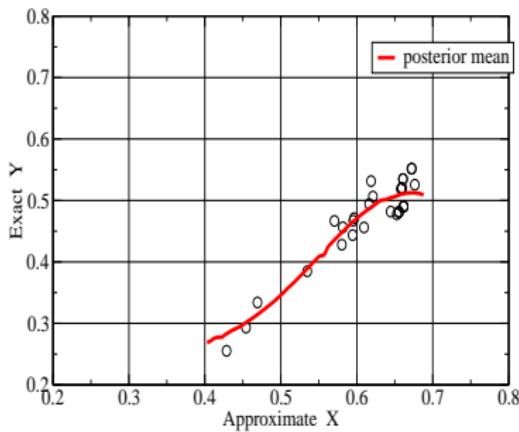


Figure: posterior of  $f(x; \boldsymbol{u})$

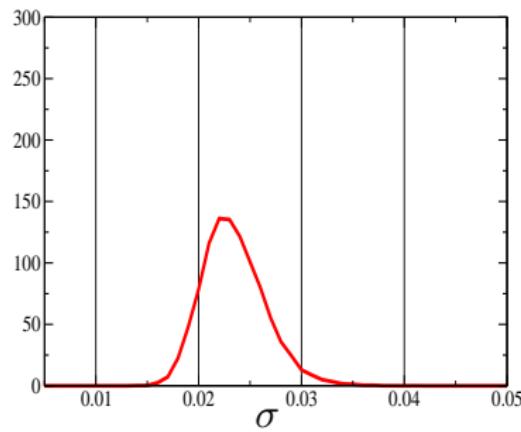


Figure: posterior of  $\sigma$

# Numerical Example

- Model:  $y_i = f(x_i; \boldsymbol{u}) + \sigma Z_i$  - Data Points: 50

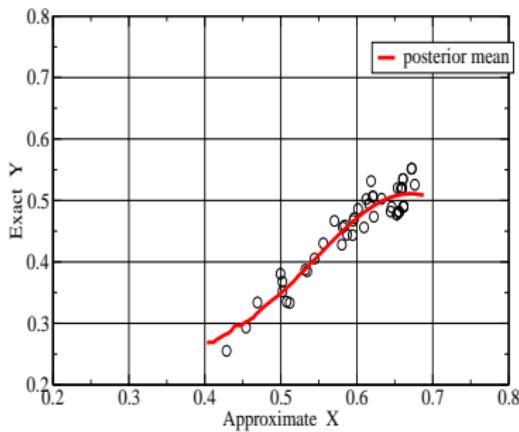


Figure: posterior of  $f(x; \boldsymbol{u})$

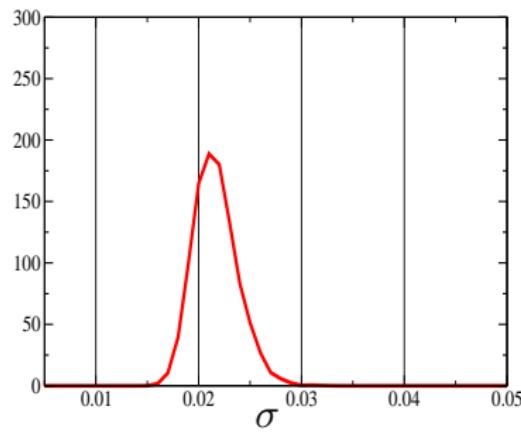


Figure: posterior of  $\sigma$

# Numerical Example

- Model:  $y_i = f(x_i; \boldsymbol{u}) + \sigma Z_i$  - Data Points: 100

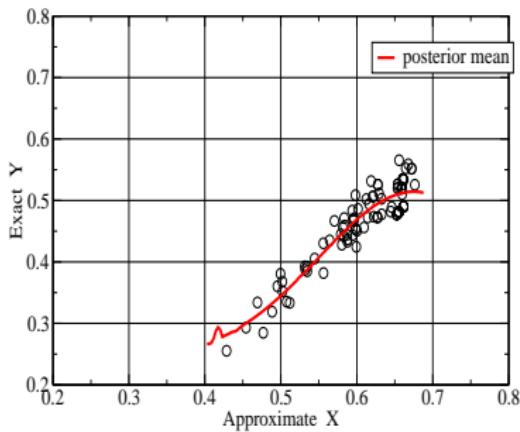


Figure: posterior of  $f(x; \boldsymbol{u})$

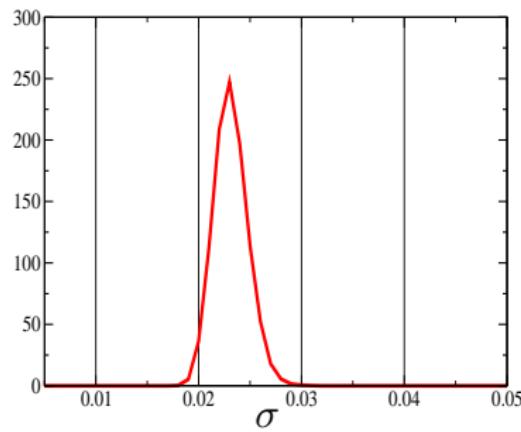


Figure: posterior of  $\sigma$

# Numerical Example

- Model:  $y_i = f(x_i; \boldsymbol{u}) + \sigma Z_i$  - Data Points: **10**

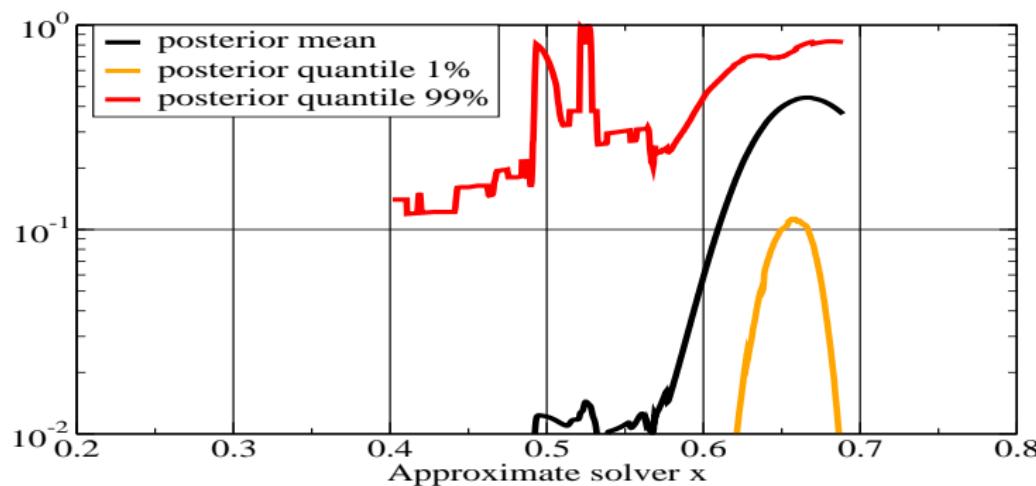


Figure: posterior  $\Pr[y \geq 0.521 | x, \text{ data } D_{10}]$

# Numerical Example

- Model:  $y_i = f(x_i; \boldsymbol{u}) + \sigma Z_i$  - Data Points: **20**

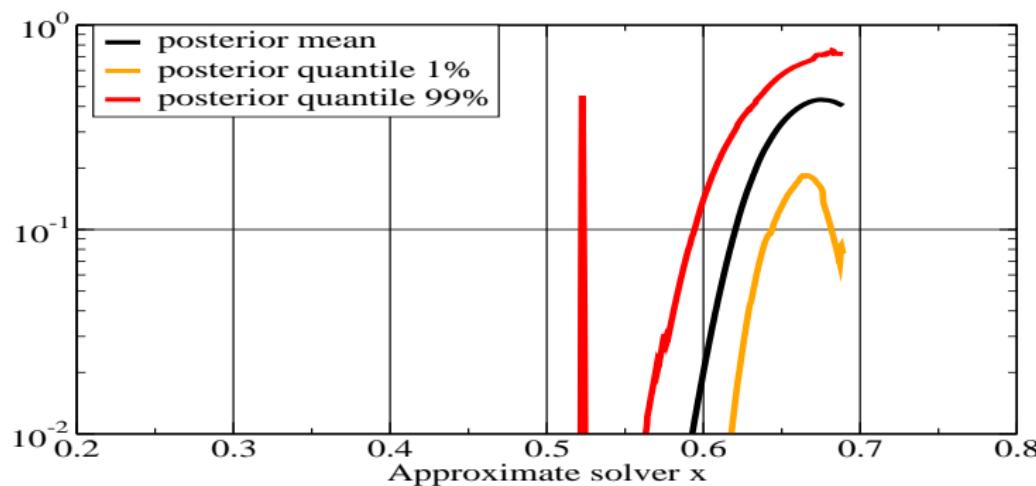


Figure: posterior  $\Pr[y \geq 0.521 | x, \text{ data } D_{20}]$

# Numerical Example

- Model:  $y_i = f(x_i; \boldsymbol{u}) + \sigma Z_i$  - Data Points: 30

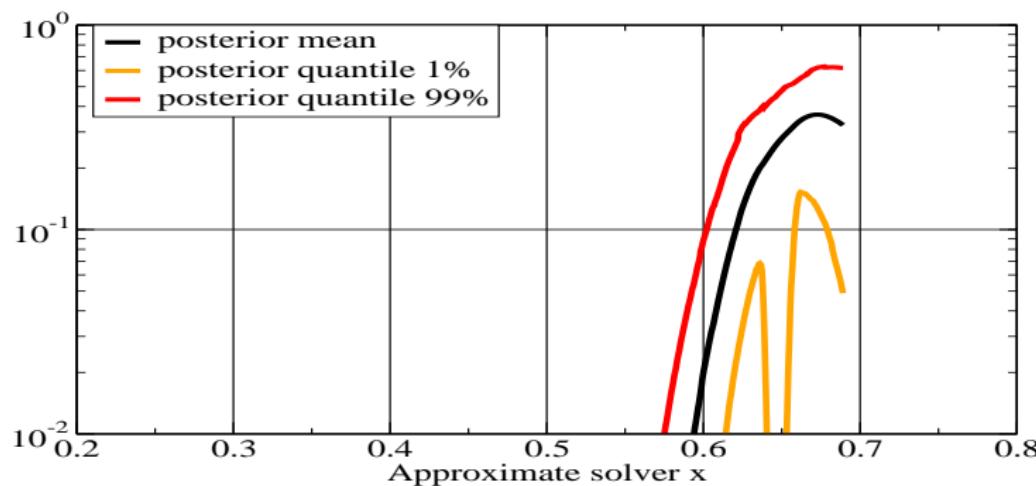


Figure: posterior  $Pr[y \geq 0.521 | x, \text{ data } D_{30}]$

# Numerical Example

- Model:  $y_i = f(x_i; \boldsymbol{u}) + \sigma Z_i$  - Data Points: 50

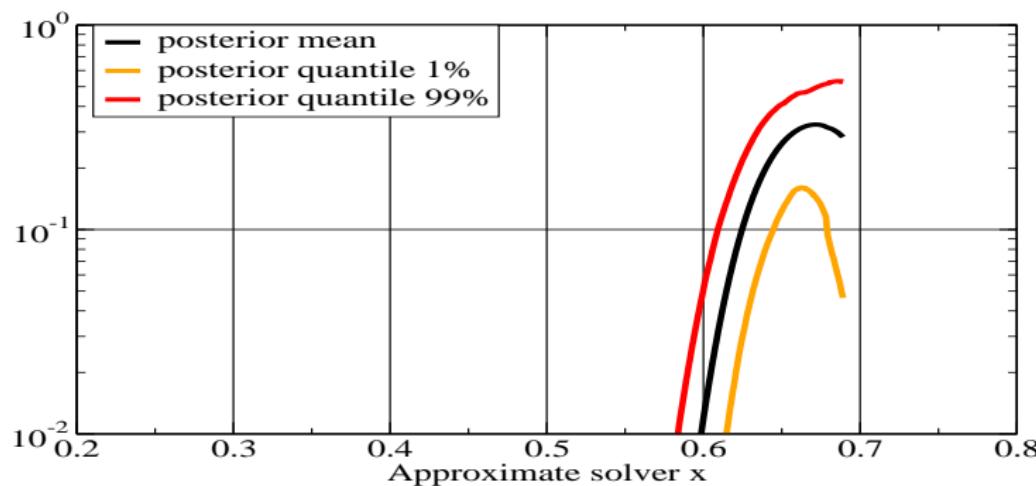


Figure: posterior  $\Pr[y \geq 0.521 | x, \text{ data } D_{50}]$

# Numerical Example

- Model:  $y_i = f(x_i; \boldsymbol{u}) + \sigma Z_i$  - Data Points: **100**

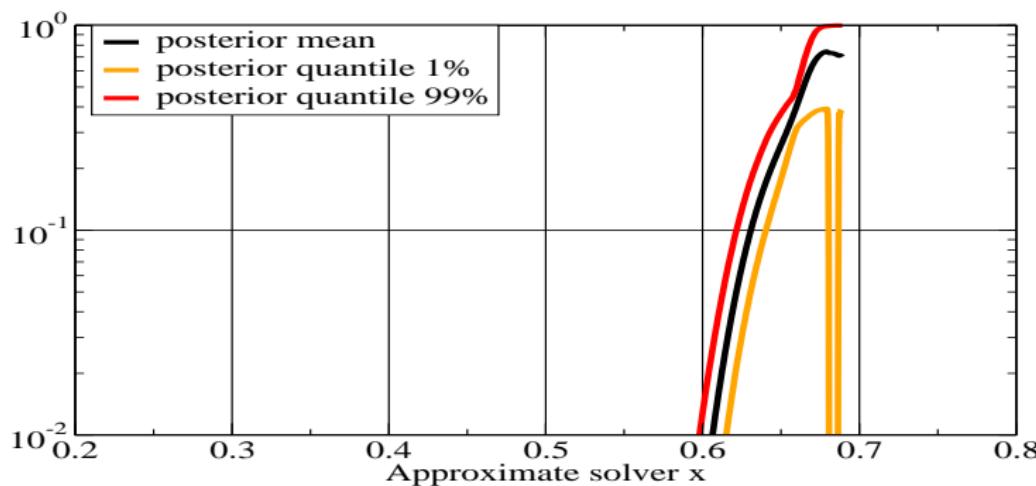


Figure: posterior  $\Pr[y \geq 0.521 | x, \text{ data } D_{100}]$

# Numerical Example

- $p_0 = \Pr[y \geq 0.521] = 1.00 \pm 0.23 \times 10^{-3}$  - with 1,500 calls to accurate model

Number of samples	Posterior mean	Posterior quantile 1%	Posterior quantile 99%
10	$1.47 \times 10^{-2}$	$2.33 \times 10^{-4}$	$3.80 \times 10^{-1}$
20	$6.24 \times 10^{-3}$	$3.56 \times 10^{-3}$	$1.90 \times 10^{-2}$
30	$2.64 \times 10^{-3}$	$3.50 \times 10^{-4}$	$8.55 \times 10^{-3}$
50	$2.64 \times 10^{-3}$	$4.25 \times 10^{-4}$	$5.23 \times 10^{-3}$
100	$1.06 \times 10^{-3}$	$4.75 \times 10^{-4}$	$2.14 \times 10^{-3}$

# Numerical Example

- Because we learn  $p(y | x)$  we can calculate statistics of any event related to  $y$
- For  $Pr[y \geq y_0]$ :

Probability (Threshold $y_0$ )	Posterior mean	Posterior quantile 1%	Posterior quantile 99%
$1.00 \pm 0.19 \times 10^{-2}$ (0.503)	$7.41 \times 10^{-2}$	$3.86 \times 10^{-3}$	$1.42 \times 10^{-2}$
$1.00 \pm 0.27 \times 10^{-4}$ (0.532)	$2.85 \times 10^{-4}$	$1.41 \times 10^{-4}$	$8.98 \times 10^{-4}$

- Computational Cost (for  $p_0 = 10^{-4}$ ): 2,000 vs 100 calls to exact solver

# Numerical Example

- Because we learn  $p(y | x)$  we can calculate statistics of any event related to  $y$

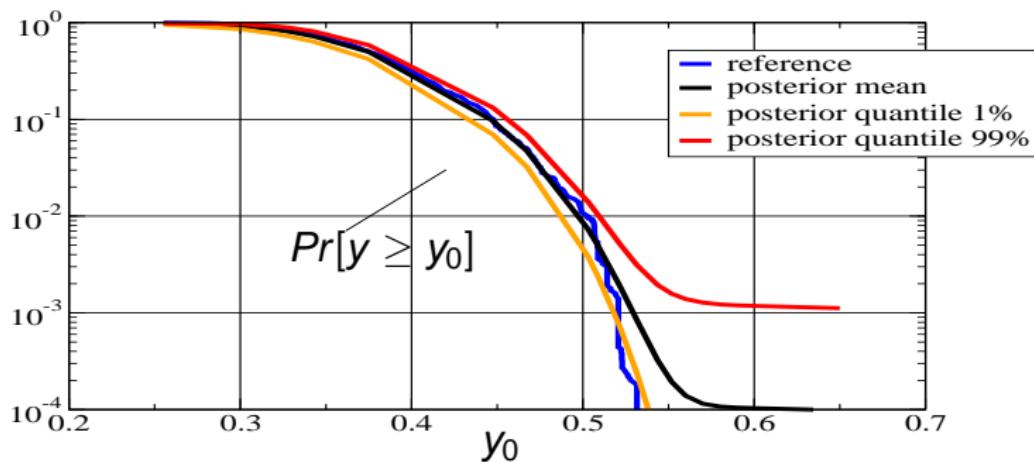


Figure: posterior for  $\Pr[y \geq y_0 | \text{data } D_{100}]$

## Numerical Example - Two experts

- **Accurate model:**  $128 \times 128$  FE mesh (i.e. 11 elements per inclusion disk diameter) with  $\approx 33,000$  dof.
- **Approximate Model 1:**  $8 \times 8$  FE mesh with 135 dof, using as the yield stress within each element the *log-average*.
- **Approximate Model 2:**  $8 \times 8$  FE mesh with 135 dof, using as the yield stress within each element the *average*. This leads to larger assigned yield stress compared to model 1 and therefore predictions  $x_2(\theta) > x_1(\theta)$
- Approximate model(s) are  $\approx 5,000$  times faster than the accurate.

# Numerical Example - Two experts

- Model:  $y_i = f(\mathbf{x}_i; \mathbf{u}) + \sigma Z_i$  - Data Points: 50

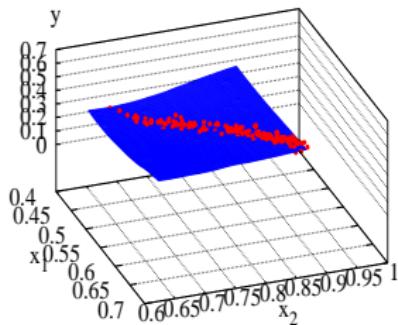


Figure: posterior of  $f(\mathbf{x}; \mathbf{u})$

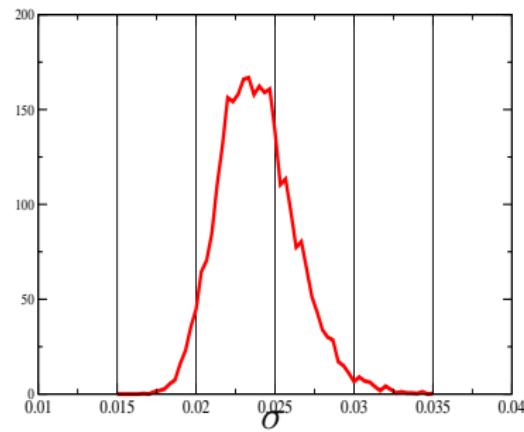
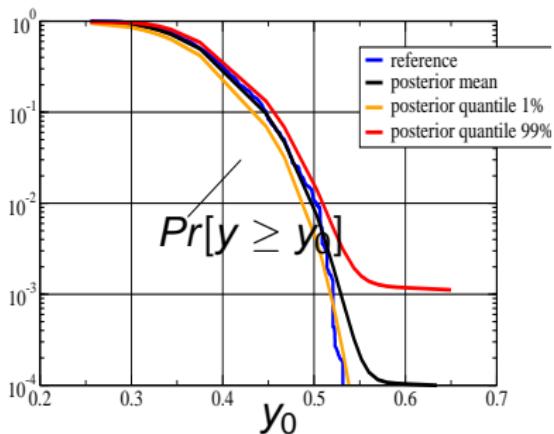
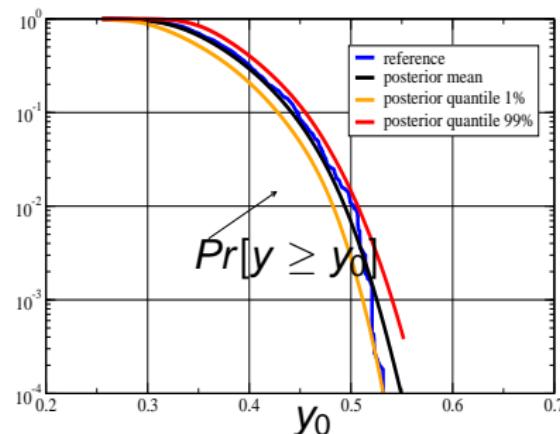


Figure: posterior of  $\sigma$

# Numerical Example



(a) 1 expert - 101 runs of expensive model



(b) 2 experts - 52 runs of expensive model

# Conclusions & Outlook

- Despite progress in Monte Carlo techniques, computational effort can still be significant.
- Significant improvements can be achieved by going beyond the “black-box” and exploit less expensive, *approximate solvers* as **predictors** of the exact response.
- This is combined with a flexible, regression scheme that quantifies relation between approximate and exact solver.
- A Bayesian framework has been proposed for that purpose that all provides confidence intervals for the estimates made.
- One important extension is *active learning*:
  - Select locations of samples in regions that the regression model has high variance  $\text{Var}[p(y | x)]$

# Conclusions & Outlook

## References:

- P.S. Koutsourelakis. Accurate uncertainty quantification using inaccurate models. SIAM Journal of Scientific Computing, 31(5):3274-3300, 2009.
- S.K. Au and J. Beck, Estimation of small failure probabilities in high dimensions by subset simulation, Probabilistic Engineering Mechanics, 16 (2001), pp. 263-277.
- P. Del Moral, A. Doucet, and A. Jasrau, Sequential Monte Carlo Samplers, Journal of the Royal Statistical Society B, 68 (2006), pp. 411-436.
- M.C. Kennedy and A. O'Hagan, Predicting the output from a complex computer code when fast approximations are available, Biometrika, 87 (2000), pp. 1-13.
- C. E. Rasmussen and Z. Ghahramani, Bayesian Monte Carlo, in NIPS 15, MIT Press, 2003.
- J. Vermaak, S.J. Godsill, and Doucet. A. Sequential Bayesian kernel regression. NIPS 16. MIT Press, 2003.