

Data-Driven, Bayesian Uncertainty Quantification for Large-Scale Simulations

Lina Kulakova

WITH: P. Angelikopoulos, P. Chatzidoukas, C. Papadimitriou, P. Koumoutsakos

SIAM UQ, Lausanne

5-8 April, 2016

CSElab

Computational Science & Engineering Laboratory
<http://www.cse-lab.ethz.ch>

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

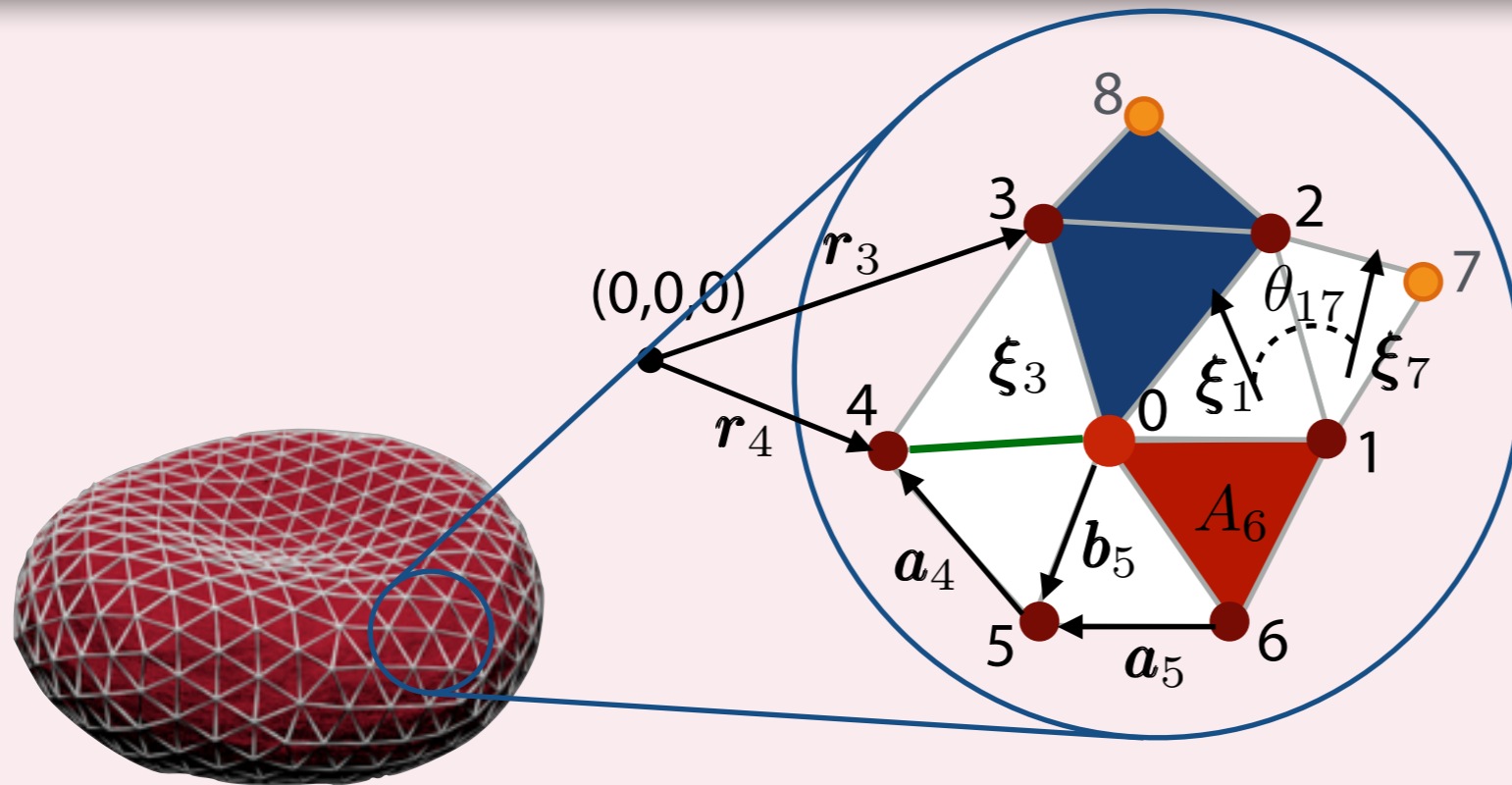
Motivation

- ▶ Simulations in Engineering and Life Sciences usually involve computationally intensive models (e.g. Molecular Dynamics)
- ▶ Bayesian inference => large numbers of model evaluations
- ▶ High Performance Computing (HPC) is a must!
- ▶ How to exploit HPC architectures for Bayesian UQ+P?

Examples of Models

Example 1: Red Blood Cell model (1/3)

J. Li et al., 2005



Calibrate for:

maximum spring extension x_0
persistence length p

$$\mathbf{F}^{\text{cell}} = \sum_{n=1}^N \mathbf{F}_{0,n-1,n,n+1}^{\text{dihedral},1} + \mathbf{F}_{0,n,N+n,n+1}^{\text{dihedral},2} + \mathbf{F}_{0,n,n+1}^{\text{triangle}} + \mathbf{F}_{0,n}^{\text{bond}}$$

$$\mathbf{F}_{0,n}^{\text{bond}} = -\frac{k_B T}{p} \left(\frac{1}{4(1-x_n)^2} - \frac{1}{4} + x_n \right) + \frac{k_p}{x_0 l_0} + 2\sqrt{k_B T \gamma^T} \overline{d\mathbf{W}_{ij}^S}$$

$x_n = \frac{b_n}{l_0/x_0}$

[spring-like force]

[dissipative force]

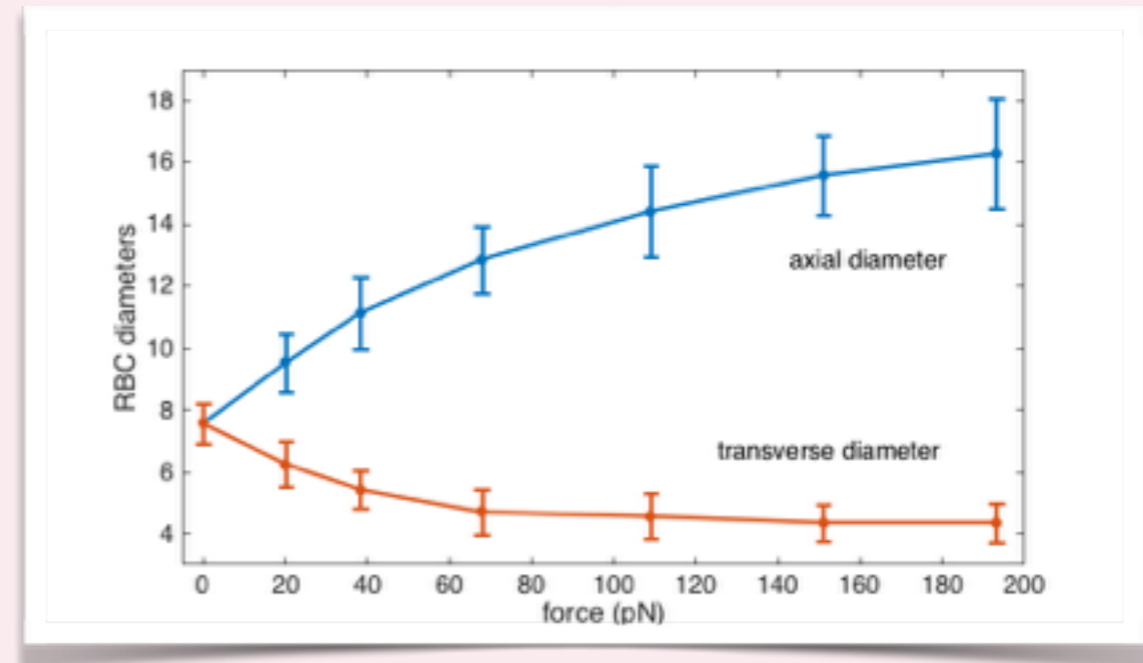
$$\overline{d\mathbf{W}_{ij}^S} = d\mathbf{W}_{ij}^S - \text{tr}[d\mathbf{W}_{ij}^S] \mathbf{1}/3$$

Example 1: Red Blood Cell model (2/3)

Data: stretching experiment (Suresh et al., 2005)



Credit: Suresh et al., 2005



Simulation: *uDeviceX* <https://github.com/uDeviceX/uDeviceX>

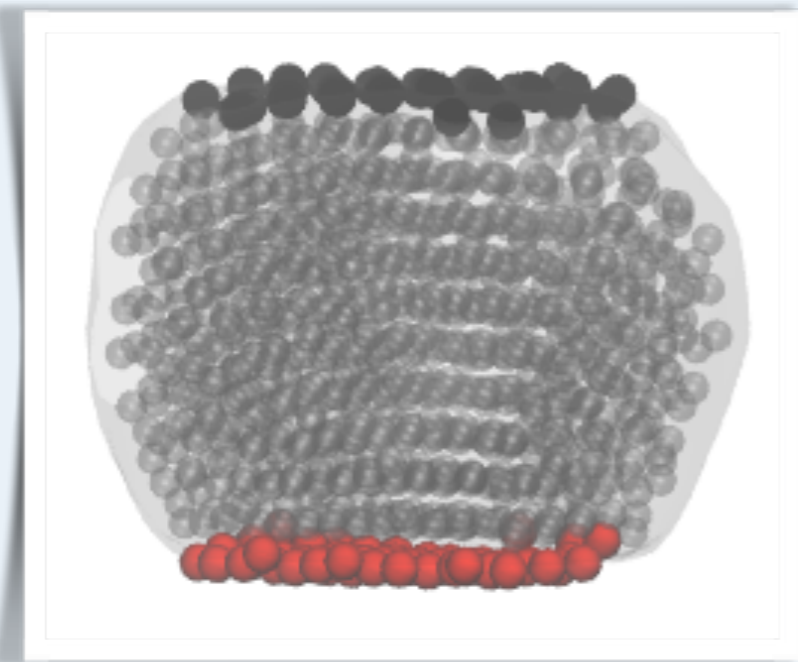
Stochastic forward model:

$$\mathbf{D} = M(\boldsymbol{\theta}) + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim \mathcal{N}(0, \boldsymbol{\Sigma}),$$

$$\boldsymbol{\Sigma} = \text{diag}(\underbrace{\sigma_1^2 + \tau_1^2 + \omega_1^2, \dots, \sigma_2^2 + \tau_1^2 + \omega_1^2}_N, \underbrace{\sigma_2^2 + \tau_2^2 + \omega_2^2, \dots, \sigma_2^2 + \tau_2^2 + \omega_2^2}_N)$$

Example 2: Subcellular Element Model (1/3)

T. J. Newman, 2005



Cell with N subcellular elements

$$\eta \dot{\mathbf{y}}_i = \mathbf{F}_i + \mathbf{F}_i^R$$

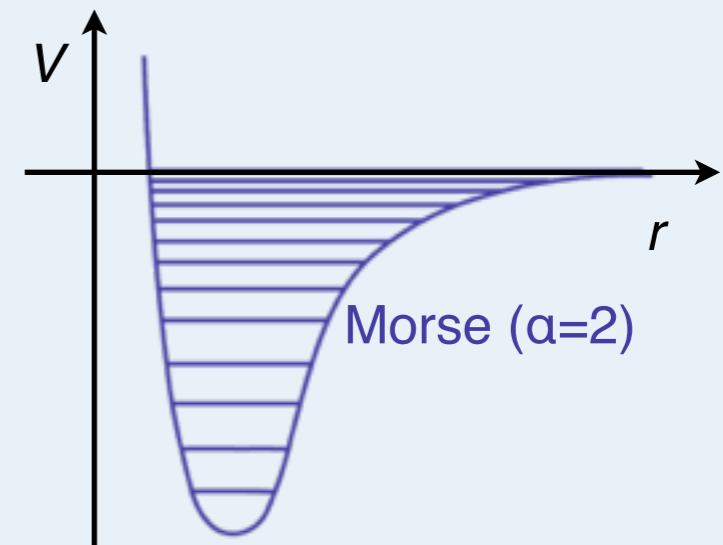
conservative force

random force

Parameters of \mathbf{F}_i :

Shape of the Morse-like potential:

$$V(r, \varphi) = u_0 \left(e^{2\rho(1-r^2/(\varphi^2 r_0^2))} - \alpha e^{\rho(1-r^2/(\varphi^2 r_0^2))} \right) \varphi^3$$

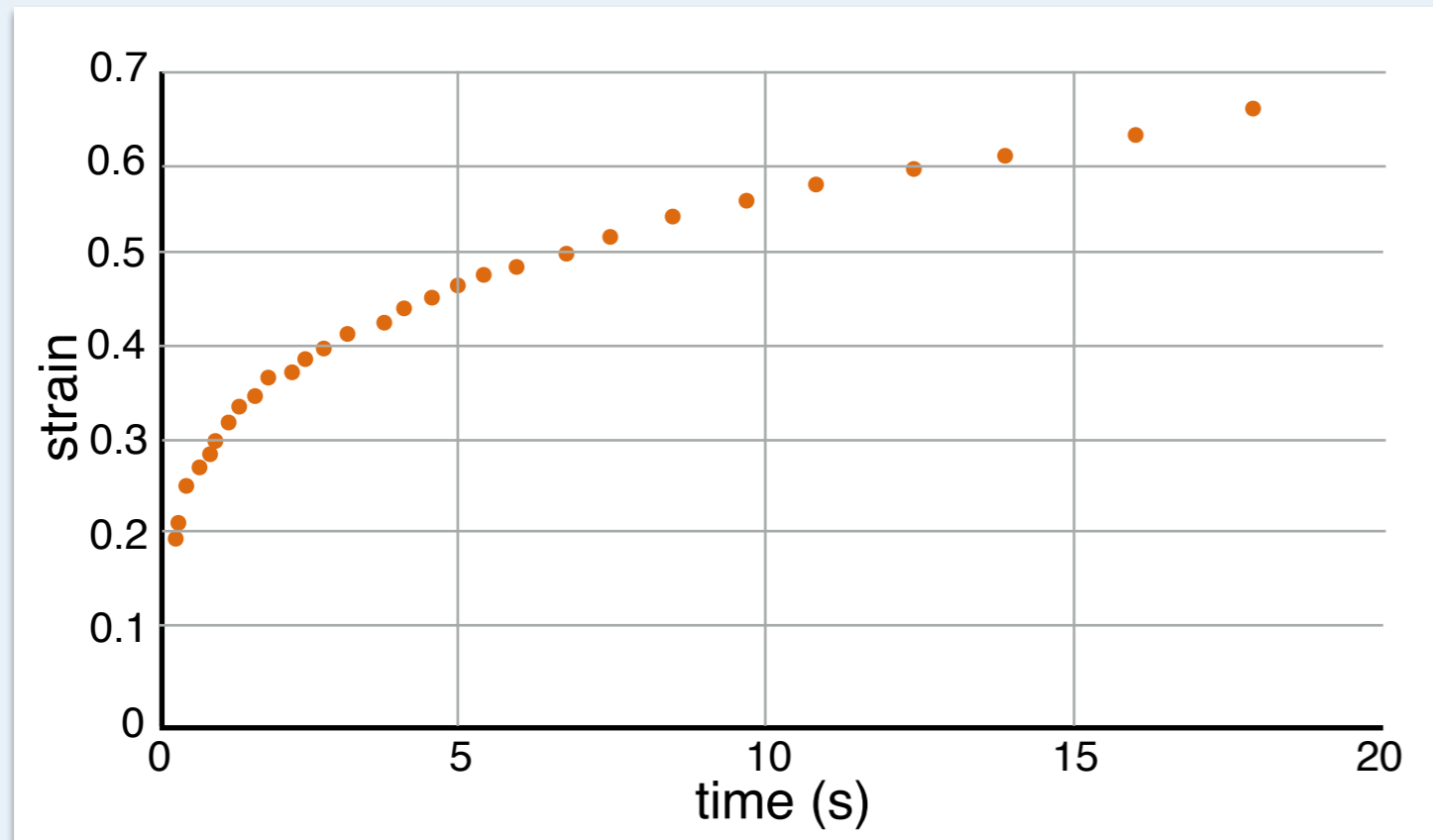


Stiffness $\kappa = \kappa_0 N^{-1/3} (1 - \lambda N^{-1/3})$

Viscosity $\eta = \eta_0 / N$

Example 2: Subcellular Element Model (2/3)

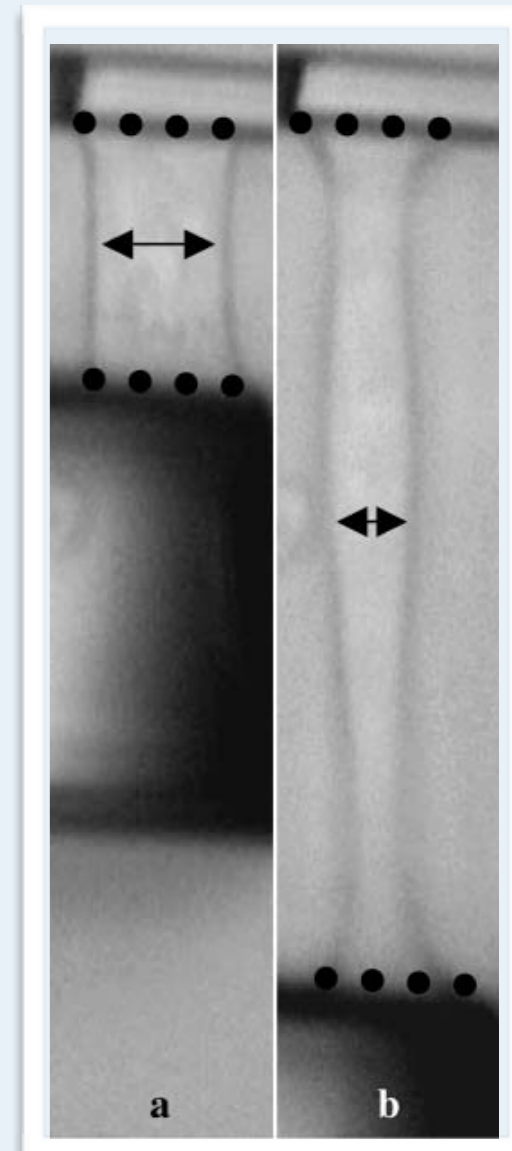
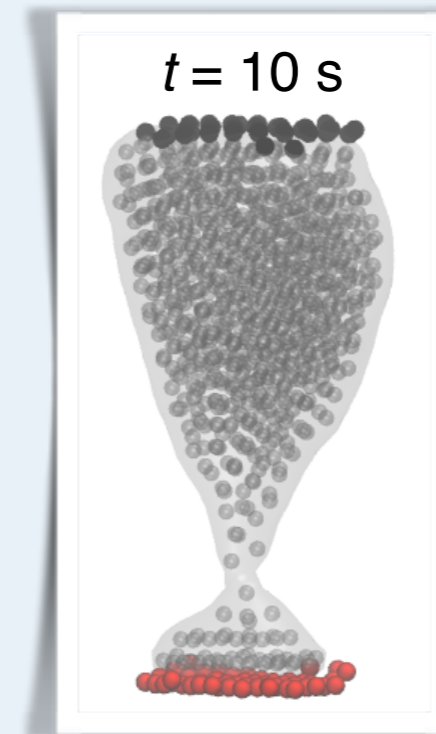
Data: strain vs time (Desprat et al., 2005)



Stochastic forward model:

$$\mathbf{D} = M(\boldsymbol{\theta})$$

$$\text{discrepancy: } SSE = \sum_{i=1}^{N_D} (\varepsilon_S(t_i) - \varepsilon_D(t_i))^2$$

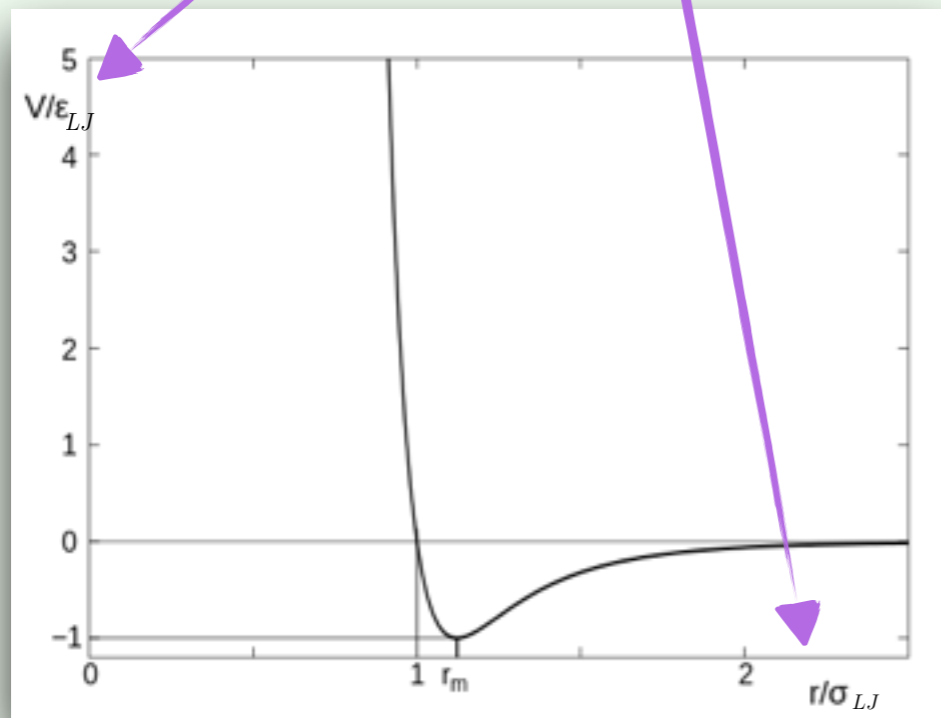


Desprat et al., 2005

Example 3: Lennard-Jones for Helium (1/3)

Lennard-Jones potential:

$$V_{LJ}(r; \sigma_{LJ}, \epsilon_{LJ}) = 4\epsilon_{LJ} \left[\left(\frac{\sigma_{LJ}}{r} \right)^{12} - \left(\frac{\sigma_{LJ}}{r} \right)^6 \right]$$

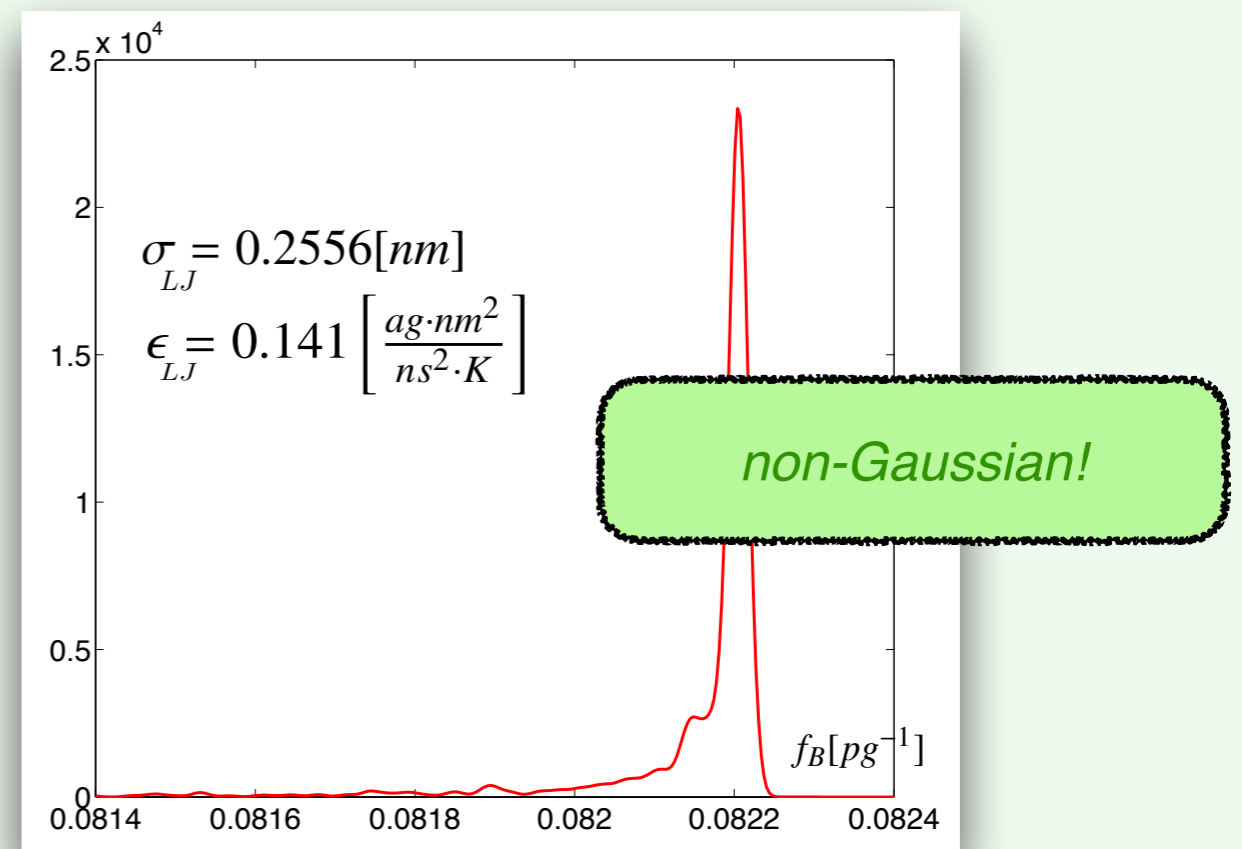


Data:

Boltzmann factor:

$$f_B = \left\langle \exp\left(-\frac{H}{Tk_B}\right) \right\rangle$$

relative probability of a particular arrangement with a given energy



PDF of the Boltzmann factor for a system with 1000 atoms

Example 3: Lennard-Jones for Helium (2/3)

Discrepancies:

Gaussian Setting

$$\rho(\overset{\text{simulation outcome}}{x}, \overset{\text{data}}{y}) = \sqrt{\left(\frac{\mu_x - \mu_y}{\mu_x}\right)^2 + \left(\frac{\sigma_x - \sigma_y}{\sigma_x}\right)^2} \quad \text{mean and standard deviation}$$

Quantile Setting

$$\rho(\overset{\text{simulation outcome}}{x}, \overset{\text{data}}{y}) = \left(\sum_{k=1}^4 \left(\frac{q_k(x) - q_k(y)}{q_k(x)} \right)^2 \right)^{1/2} \quad q = (0.2, 0.4, 0.6, 0.8) \\ \text{quantiles}$$

Kullback-Leibler Setting

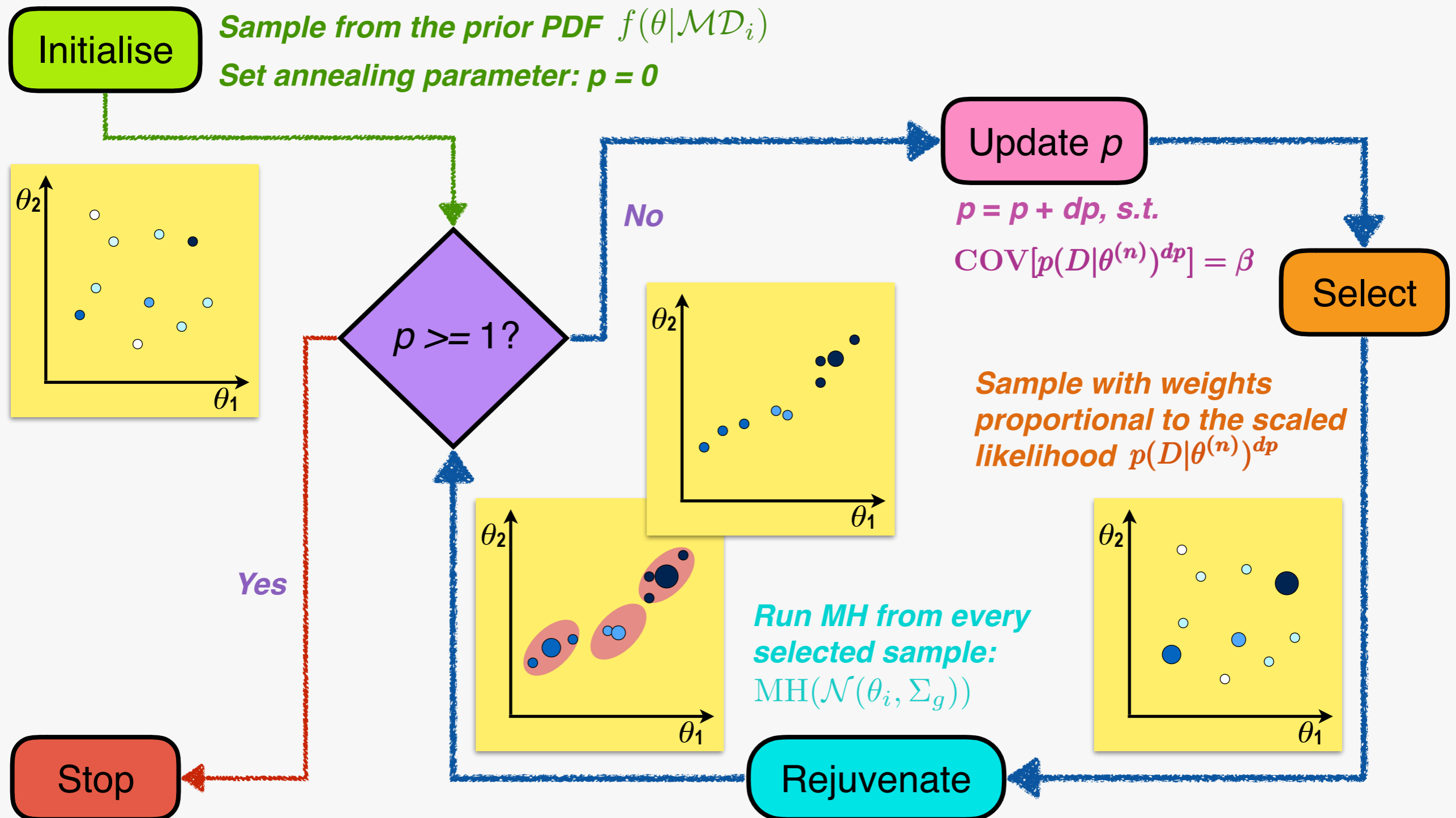
$$\rho(\overset{\text{simulation outcome}}{x}, \overset{\text{data}}{y}) = D_{KL}(P||Q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} dx$$

shows how much information was lost when approximating P with Q

Bayesian Inference Algorithms

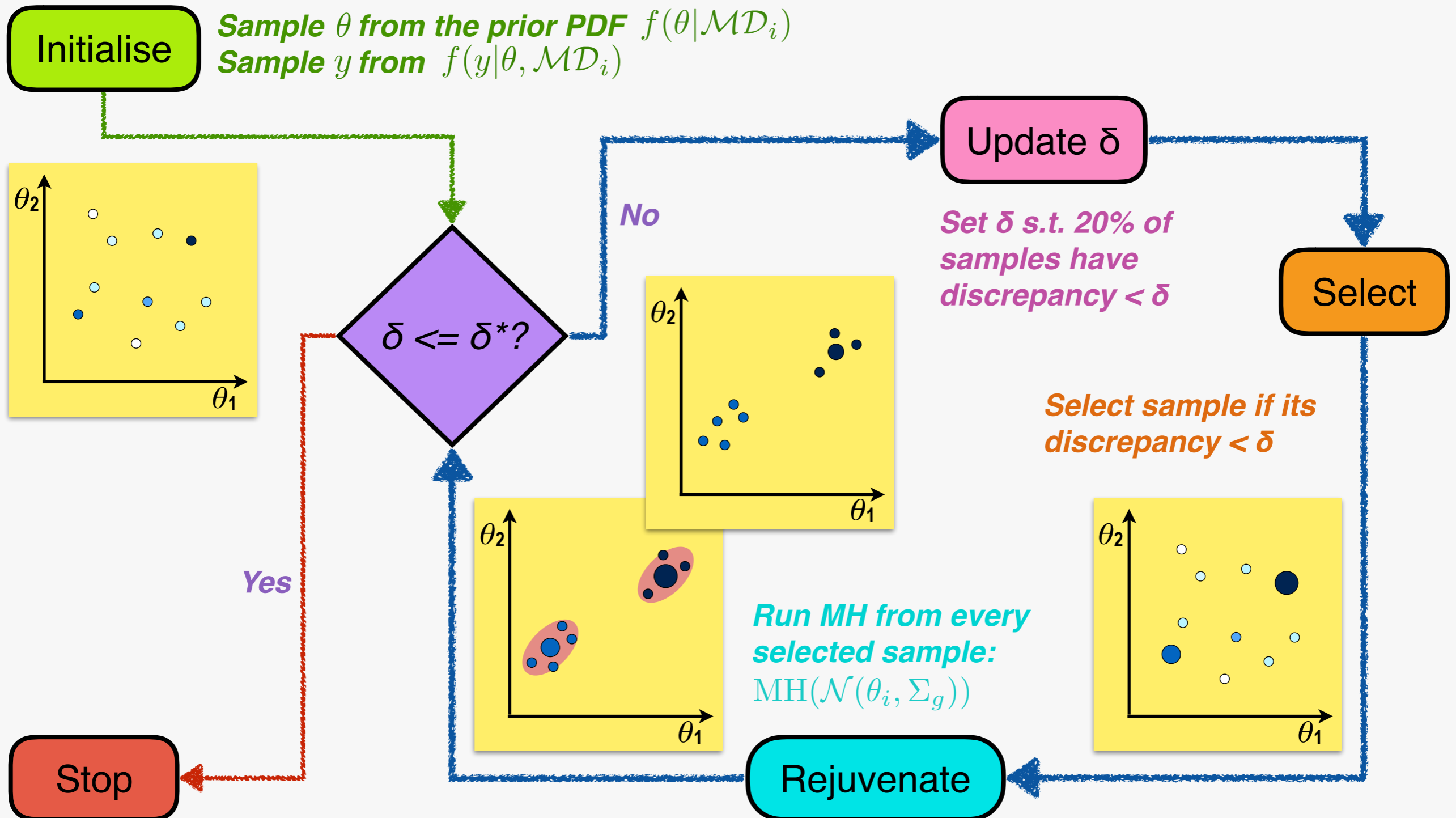
TMCMC

Ching and Chen, J. Eng. Mech., 133 (7), 2007



ABC-SubSim

Chiachio et al., SIAM J. Sci. Comput., 36 (3), 2014



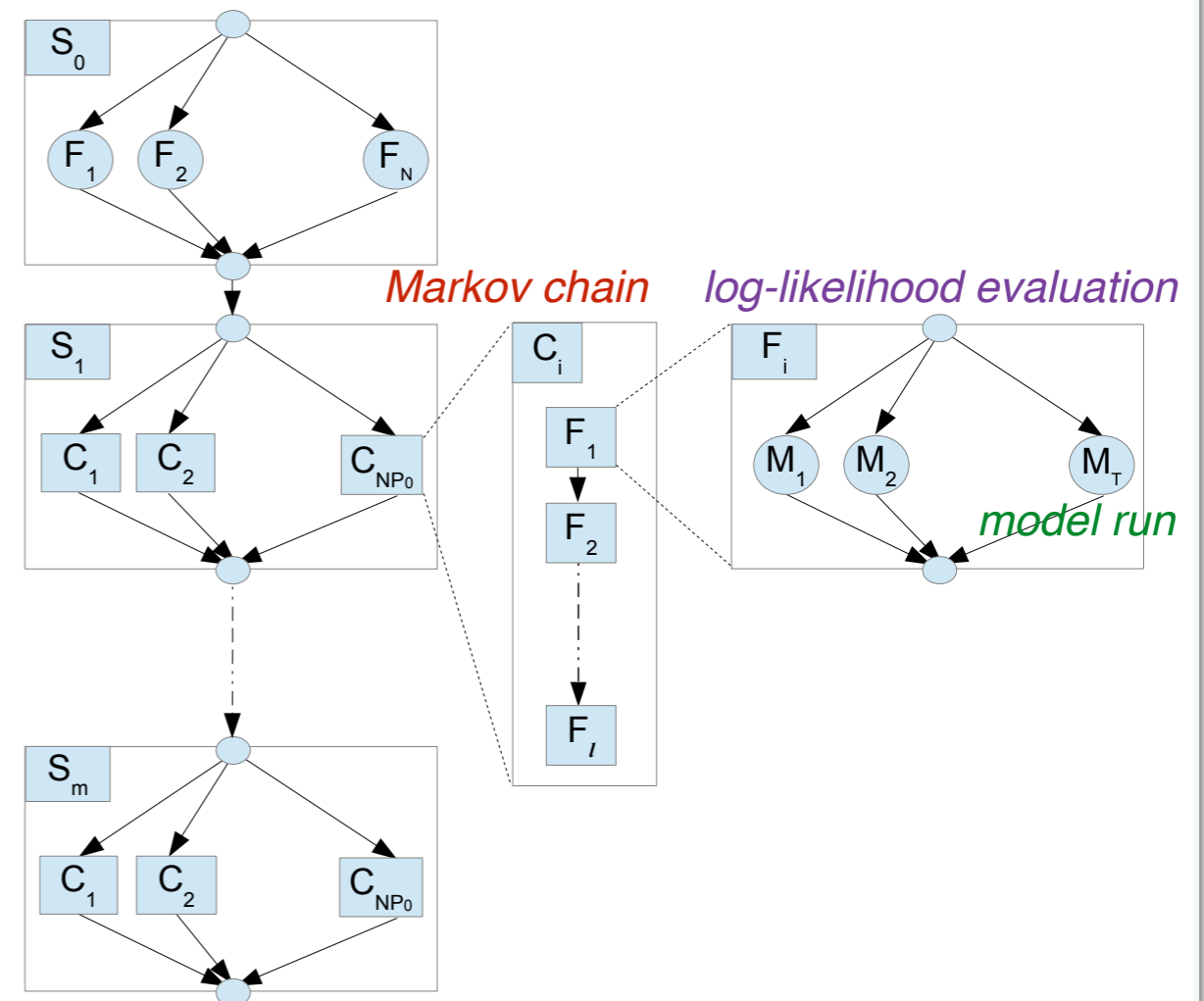
HPC Tools

HPC Approach: TORC

P. Hadjidoukas et al., 20th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP), 2012

- ▶ Runs on various architectures ranging from laptops to clusters
- ▶ Automatic load-balancing
- ▶ Integrable with external software used for model evaluation (e.g. LAMMPS)
- ▶ Task-based parallelism
- ▶ Easy to write parallel code

Task graph of a population-based sampling algorithm



TORC: code example

serial code

```
void task(double *x, double *y)
{
    *y = x[0] + x[1];
}

int main(int argc, char *argv[])
{
    double result[100];

    for (int i=0; i<100; i++)
    {
        double d[2] = {drand48(), drand48()};
        task(d, &result[i]);
    }

    return 0;
}
```

TORC code

```
void task(double *x, double *y)
{
    *y = x[0] + x[1];
}

int main(int argc, char *argv[])
{
    double result[100];

    torc_register_task(task);
    torc_init(argc, argv, MODE_MW);

    for (int i=0; i<100; i++)
    {
        double d[2] = {drand48(), drand48()};
        torc_task(-1, task, 2,
                2, MPI_DOUBLE, CALL_BY_COP,
                1, MPI_DOUBLE, CALL_BY_RES,
                &d, &result[i]);
    }
    torc_waitall();
    return 0;
}
```

Pi4U: UQ Library on top of TORC

- ▶ Open-source library distributed under LGPL licence
- ▶ Available at <http://www.cse-lab.ethz.ch/software/Pi4U>
- ▶ Algorithms:
 - *TMCMC* (for exact Bayesian inference)
 - *ABC-SubSim* (for approximate Bayesian inference)
 - *CMA-ES* (for optimisation)
 - *Subset Simulation* (for rare events sampling)
 - *A-PNDL* (for adaptive parallel numerical differentiation)

Results

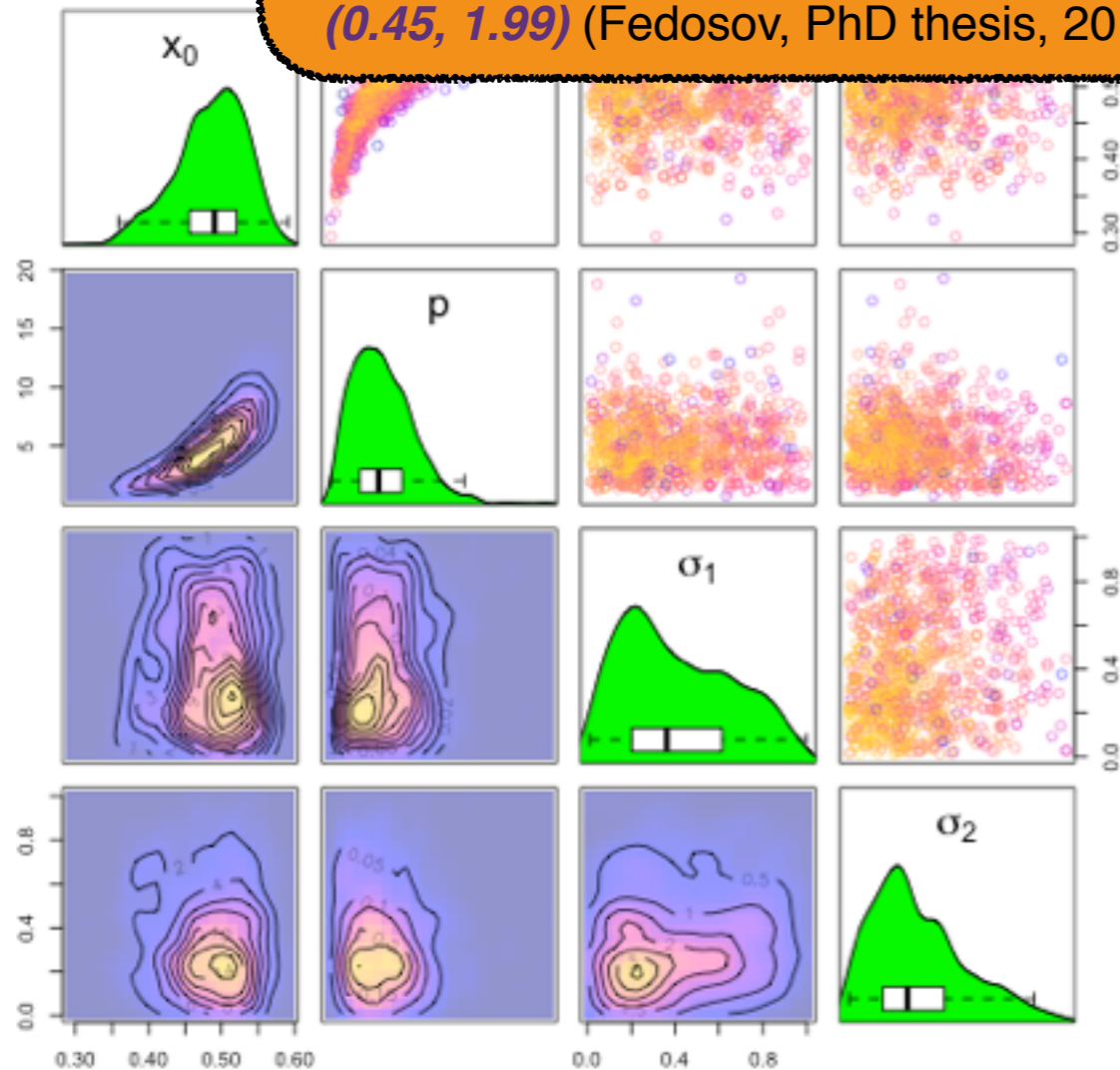
Example 1: Red Blood Cell model (3/3)

Algorithm: TMCMC

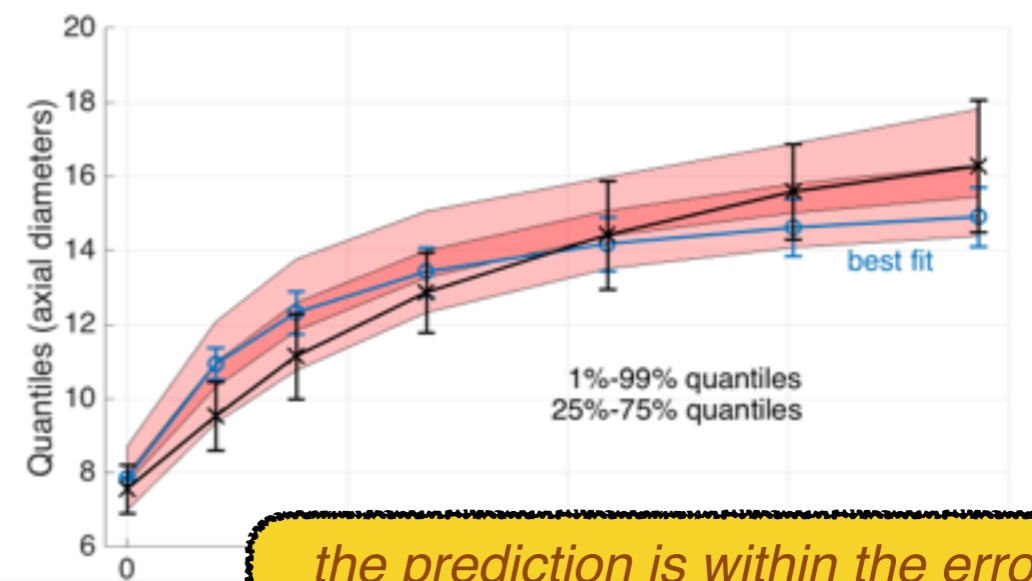
Done in collaboration with S. Litvinov, D. Alexeev

1024 samples per stage, 128 GPU nodes on Piz Daint (CSCS),
5 hours of wall-clock time per stage. CUDA + TORC workers.

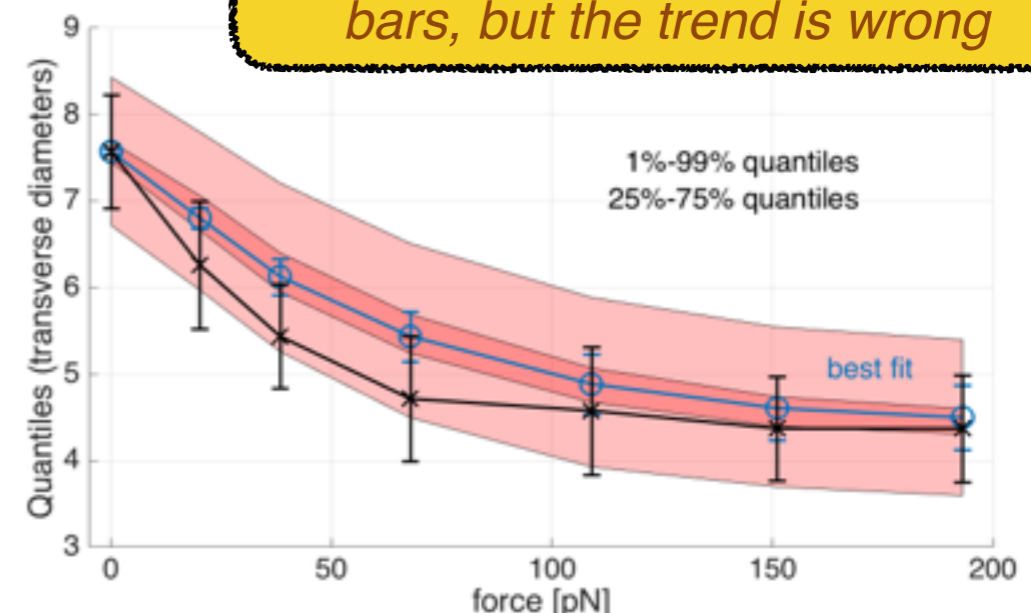
MPV: **(0.49, 4.86)** (our calibration) vs
(0.45, 1.99) (Fedosov, PhD thesis, 2010)



posterior distribution of the parameters



the prediction is within the error bars, but the trend is wrong



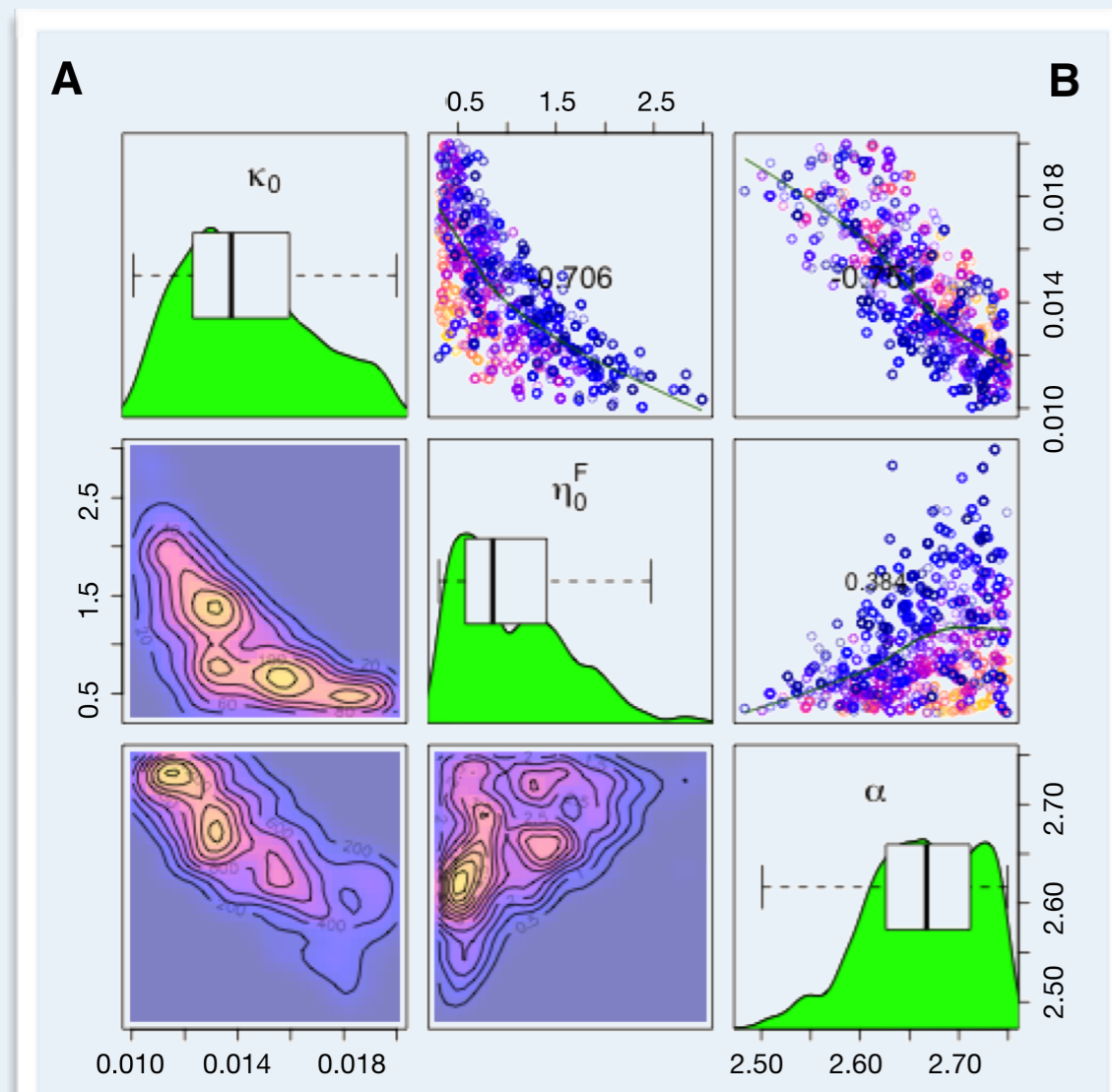
robust prediction

Example 2: Subcellular Element Model (3/3)

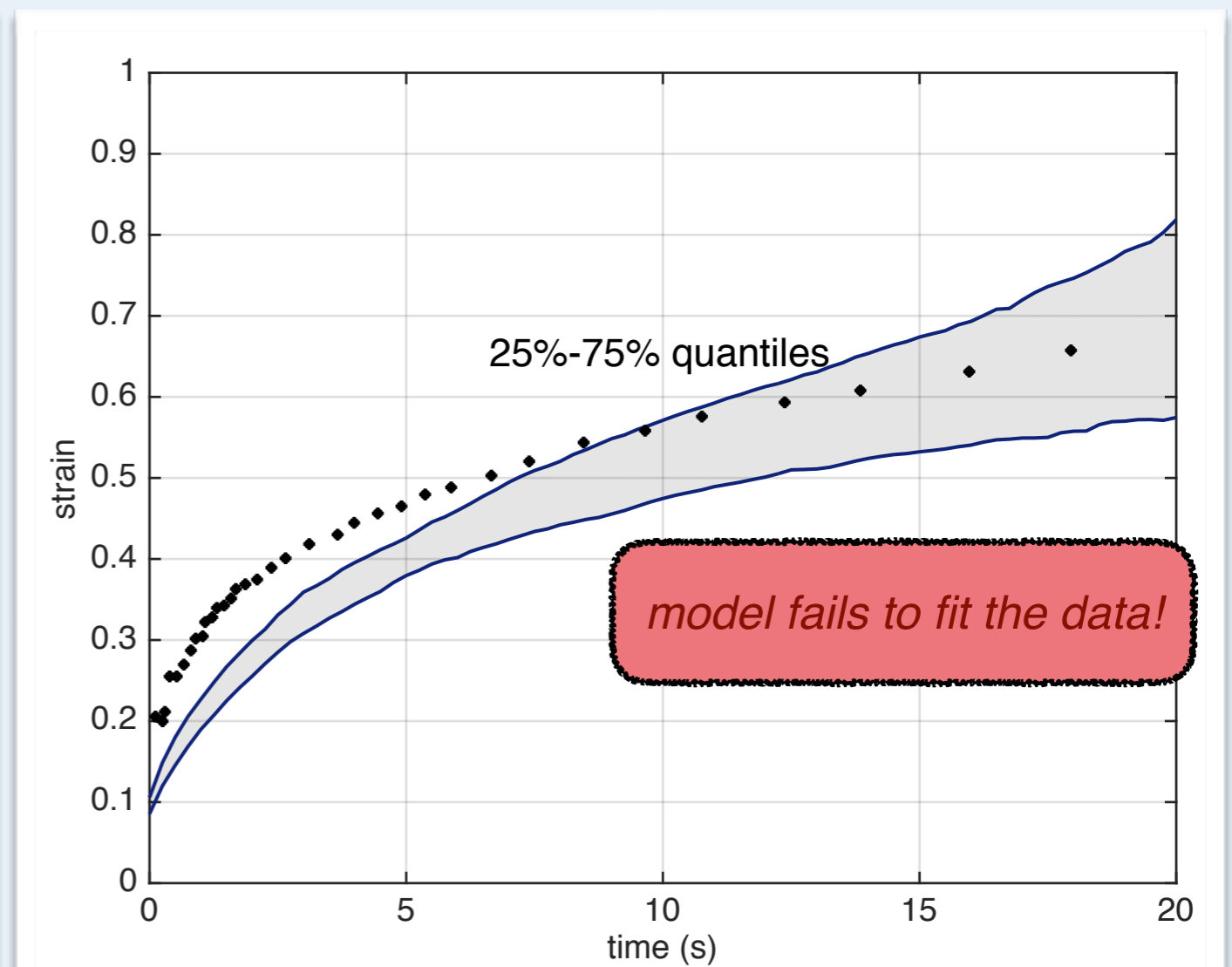
Algorithm: ABC-SubSim

Credit: A. Economides, G. Tauriello, 2015

2000 samples per stage, 384 CPUs on Brutus cluster (ETHZ),
2.5 hours of wall-clock time per stage. TORC workers.



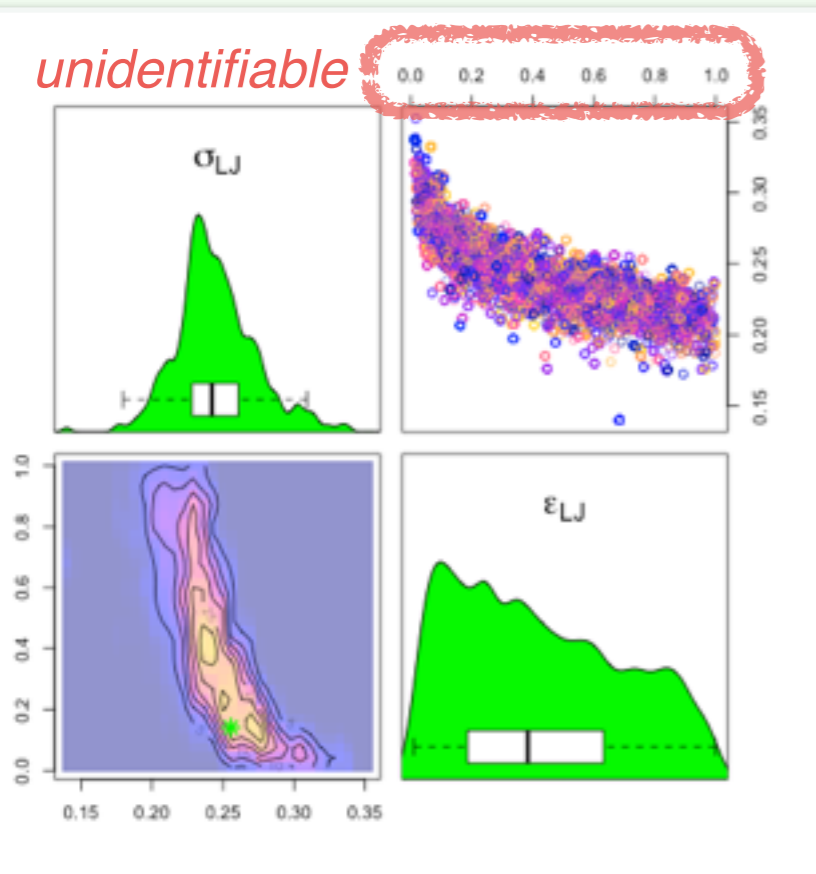
posterior distribution of the parameters



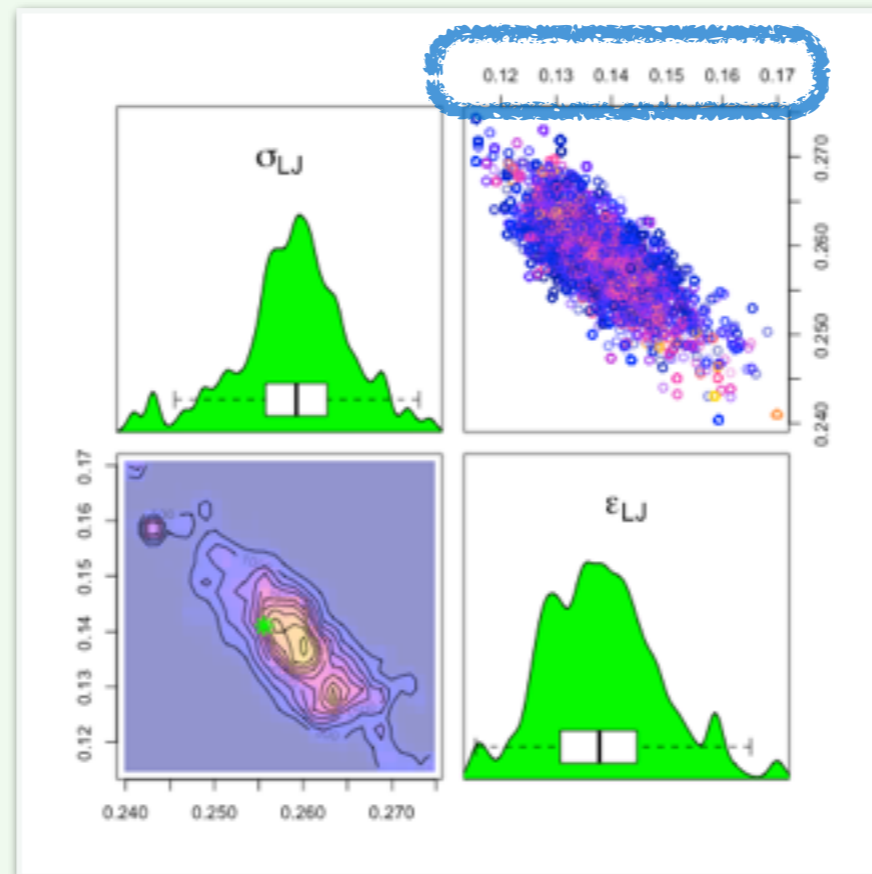
robust prediction

Example 3: Lennard-Jones for Helium (3/3)

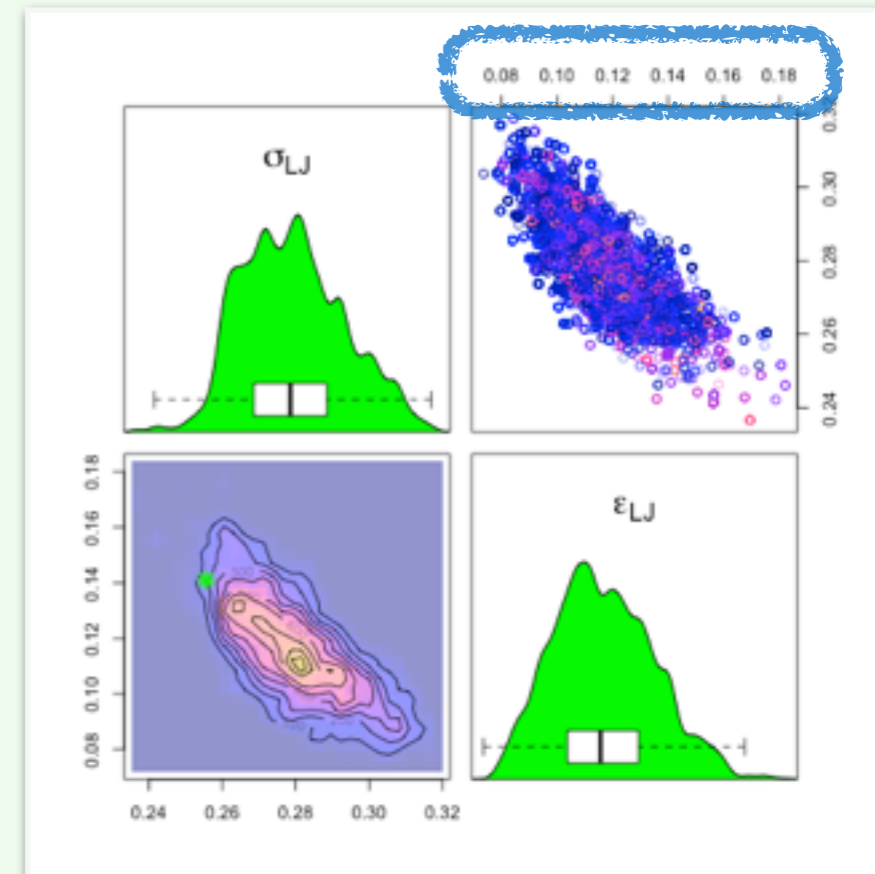
Gaussian Setting



Quantile Setting



Kullback-Leibler Setting



“true” parameters:

$$\sigma = 0.2556$$

$$\epsilon = 0.141$$

| Model | $[\sigma_l, \sigma_r]$ | $\bar{\sigma}$ | u_σ | $[\epsilon_l, \epsilon_r]$ | $\bar{\epsilon}$ | u_ϵ | N_{gen} | δ |
|----------|------------------------|----------------|------------|----------------------------|------------------|--------------|-----------|----------|
| M_G | [0.1,0.8] | 0.2437 | 12.7% | [0.01,1.0] | 0.443 | 60.5% | 4 | 0.02 |
| M_Q | [0.1,0.8] | 0.2591 | 2.1% | [0.01,1.0] | 0.136 | 6.3% | 7 | 2e-5 |
| M_{KL} | [0.1,0.8] | 0.2737 | 6.2% | [0.01,1.0] | 0.128 | 18.4% | 6 | 0.09 |

15360 samples per stage, 4096 CPUs on Piz Daint (CSCS),
0.4 hours of wall-clock time per stage. MPI (LAMMPS)+TORC workers.

Summary

- ▶ Pi4U allows to exploit efficiently HPC architectures for Bayesian inference in large-scale models
- ▶ We must re-examine the validation of many classical simulation models in Science and Engineering
- ▶ Bayesian inference offers a way to do this systematically

Thank you!