Bayesian Process Engineering

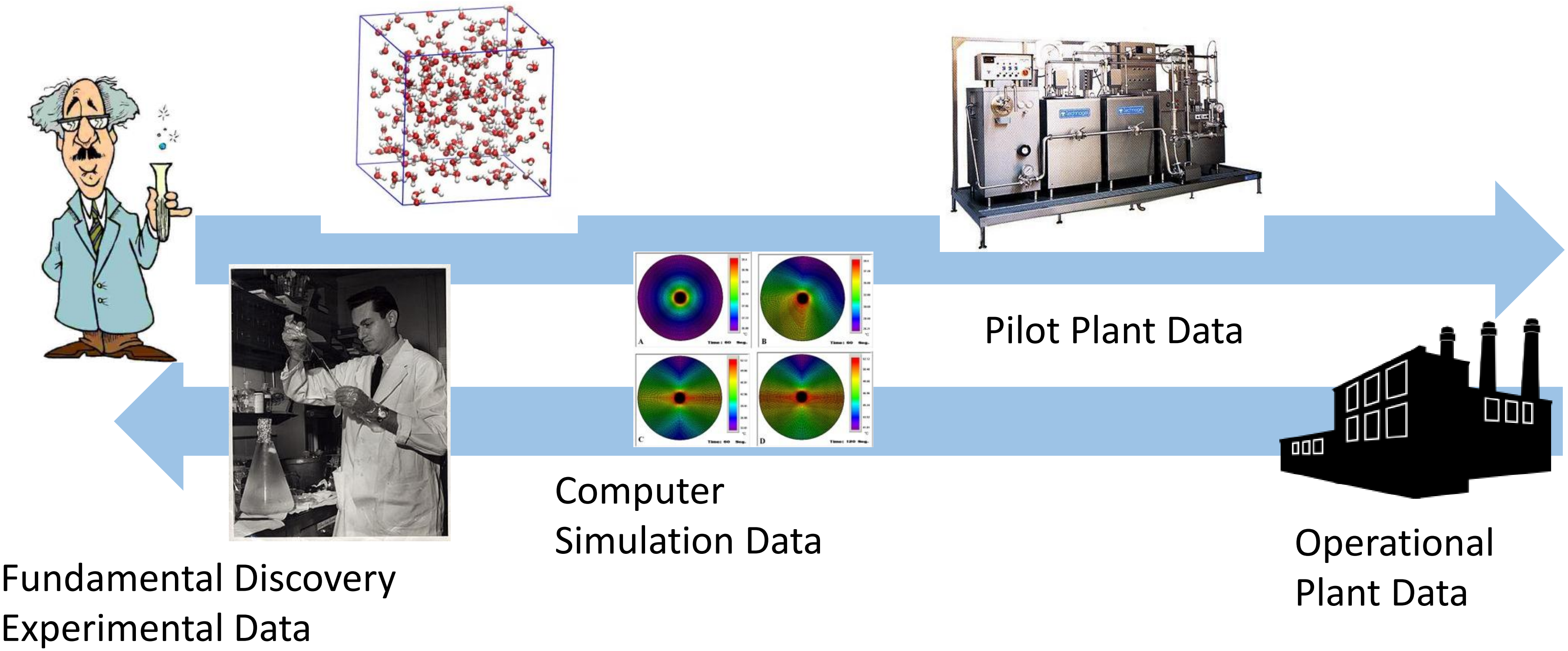
Dr. Matthew J Realff

School of Chemical & Biomolecular Engineering

Georgia Tech

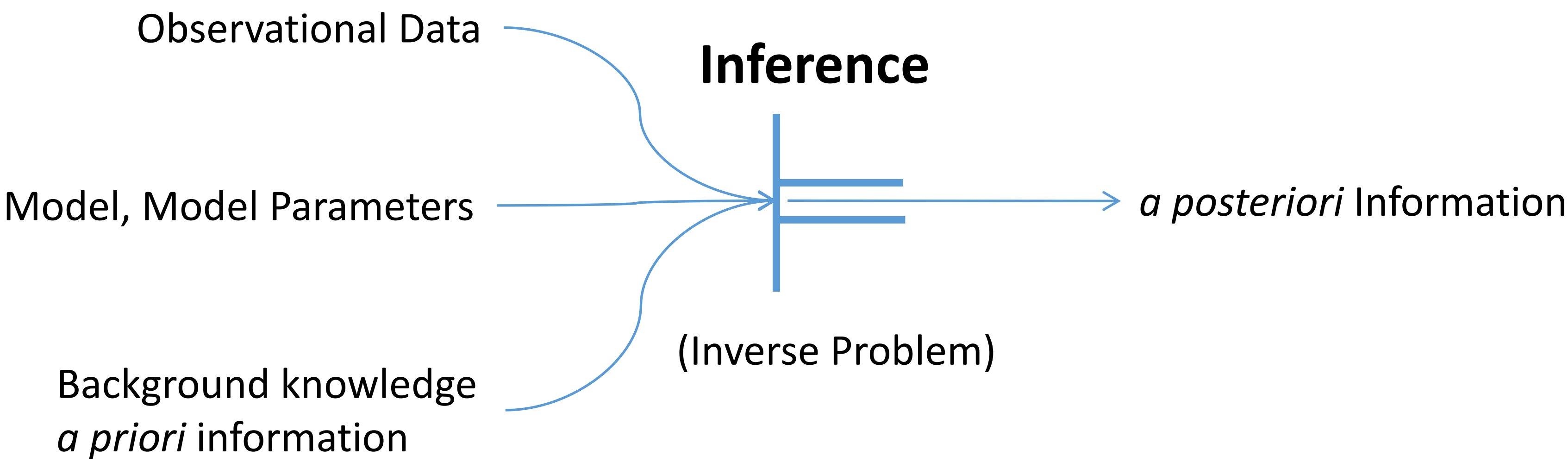
Process Systems Engineering

Process Systems Engineering as an integrative discipline along the data supply chain.



- How to fuse the data from different sources with different fidelity across time?
- How to value the information and guide new information acquisition?
- How to support decision-making in uncertain and evolving environments?

Inference as the central task



Deterministic Inference: Seeks single value of the model parameter to explain the data. Regularization used to enforce smoothness conditions around solution and handle outliers, closeness to a priori model.

$$S(m) = \left\| d^{obs} - d^{th}(m) \right\|_D + \left\| m - m^{apr} \right\|_M$$

Misfit function

Probabilistic Inference: process to infer the probability distribution of a random variable.

Posterior of hidden random variables of the model from the data, given the data and background information. Quantifies the “chance” that a given model is the true one.

W. Debski, Probabilistic Inverse Theory, Advances in Geophysics, 52, 2010, Pages 1–102,

Why Probability and Bayesian?

Calculus is the language for reasoning about rates of change
Probability is language for reasoning about uncertainty

Given limited assumptions about rational beliefs, a Bayesian update is demonstrated to be the optimal way to update prior beliefs with new information.

Posterior parametric distribution

Likelihood distribution

Parametric prior distribution

$$P(\theta|y_{meas}) = \frac{\mathcal{L}(y_{meas}|\theta)P(\theta)}{\int \mathcal{L}(y_{meas}|\theta)P(\theta)d\theta}$$

Performing this integration step is computationally challenging

Exact evaluation of the posterior or expectation is intractable, approximate inference proceeds either by deterministic or stochastic approaches.

Bayesian Machine Learning

Machine Learning: Making inferences about missing or latent data from the observed data

A well defined model: Make predictions about unobserved data

$$P(\theta|D,m) = \frac{P(D|\theta,m)P(\theta|m)}{P(D|m)}$$

m = model structure
θ = model parameters

Make a prediction

Using the posterior as the prior

$$P(D_{test}|D,m) = \int P(D_{test}|\theta,D,m)P(\theta|D,m)d\theta$$

Compare models

$$P(m|D) = \frac{P(D|\mathrm{m})P(m)}{P(D)} \qquad P(D|m) = \int P(D|\theta,m)P(\theta|m)d\theta$$

Example of Bayesian analysis in Process Systems Engineering: An adsorption process

- Data Fusion for isotherm data – multiple sources can be integrated seamlessly in estimating parameters

- Uncertainty Quantification for cyclic process performance – determine the reliability of model prediction based on uncertainties in data and model

- Design of experiments – optimally design experiments to reduce the uncertainty in model prediction and process performance

Bayesian data fusion

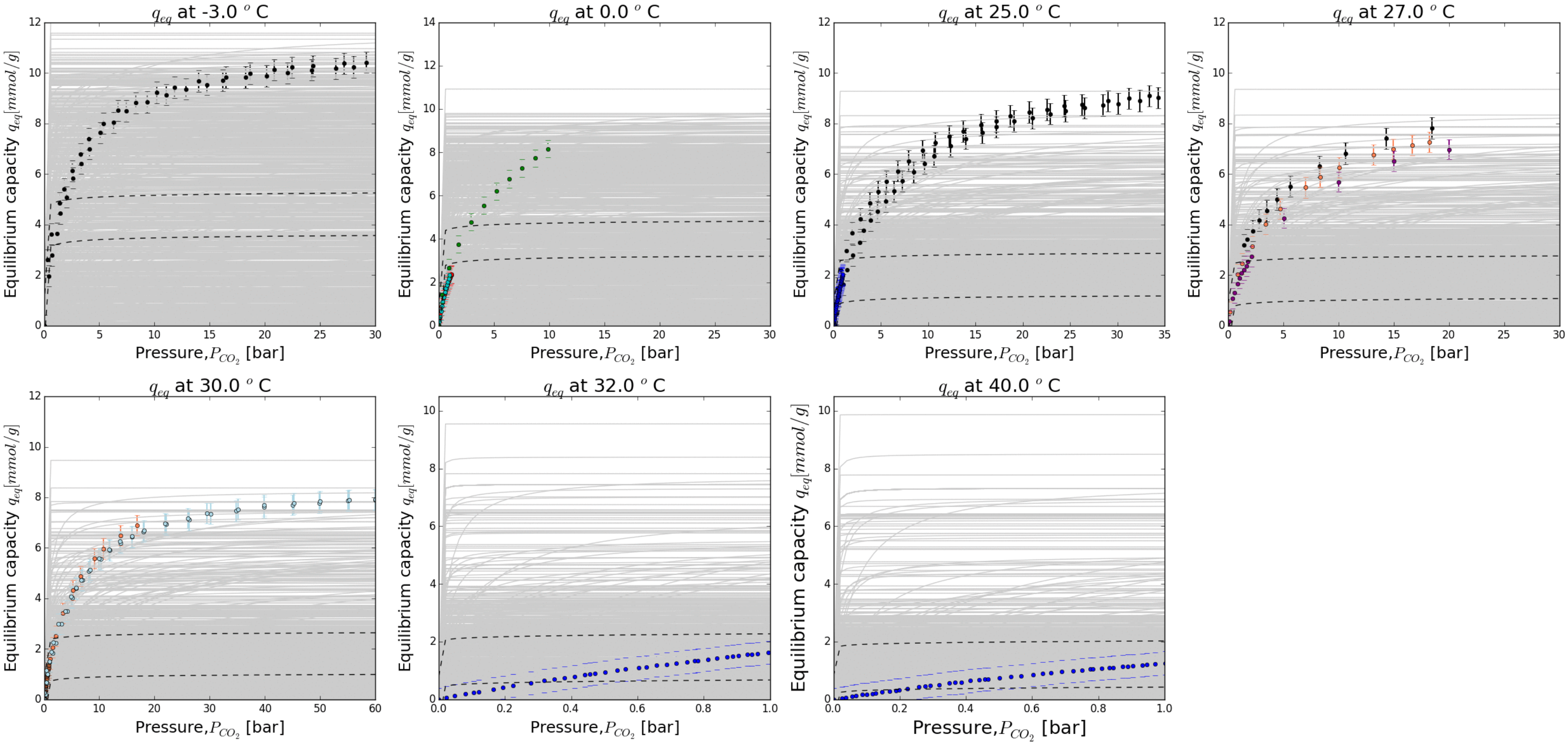
- Adsorption isotherm equilibrium data of Uio66 (MOF adsorbent) on CO_2 collected from NIST from *9 different experimental groups*
  - *at 7 different temperatures at various ranges of CO_2 partial pressure.*
- Langmuir isotherm model used to fit the data and estimate **four** unknown parameters

CO_2 Adsorption equilibrium capacity : $q_{eq} = q_m \frac{bP}{1+bP}$

Maximum CO_2 saturation capacity : $q_m = q_{m0} \exp\left(\eta\left(1 - \frac{T}{T_0}\right)\right)$

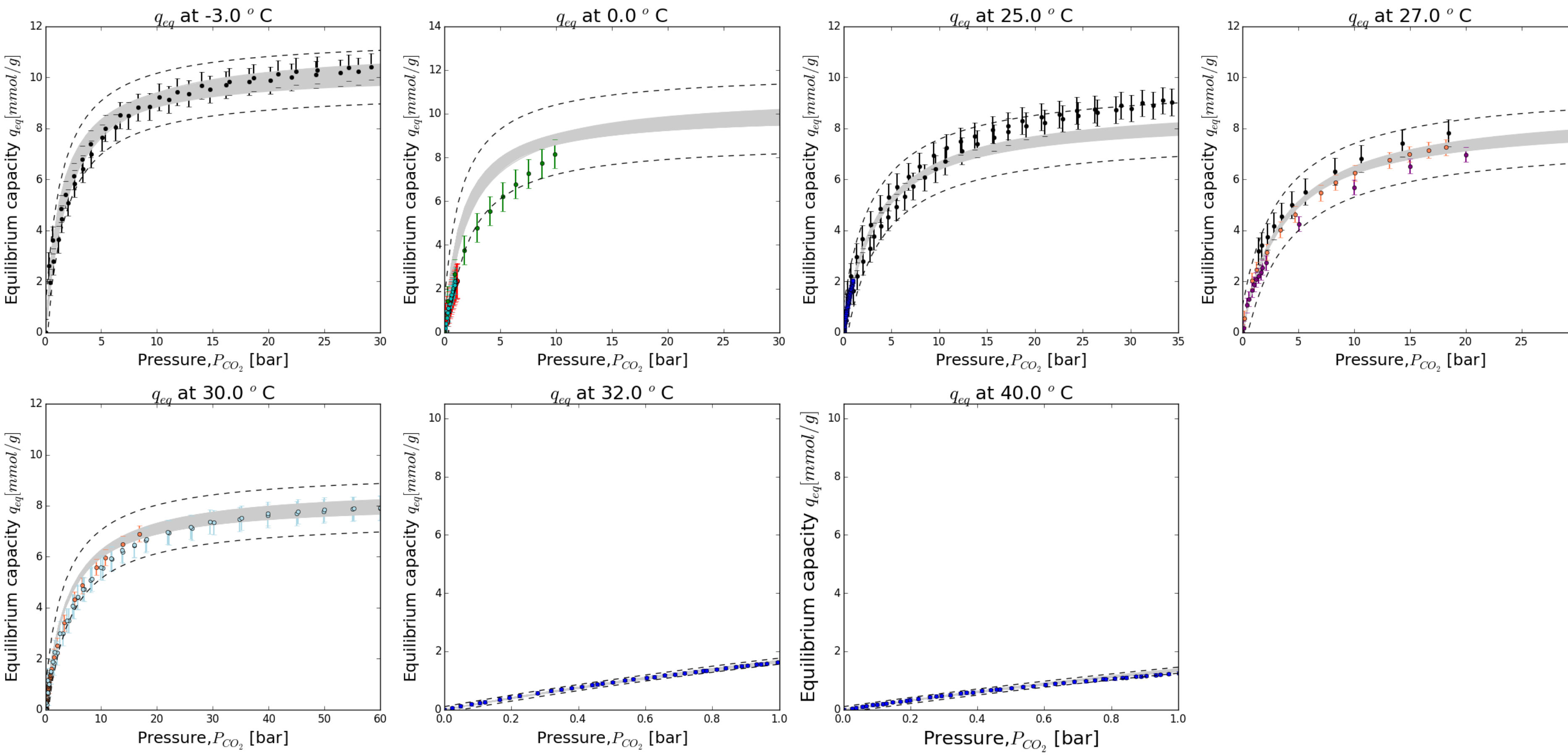
Langmuir affinity constant: $b = b_0 \exp\left(-\frac{\Delta H_0}{RT_0}\left(\frac{T_0}{T} - 1\right)\right)$

Bayesian Updating of Isotherm Parameters



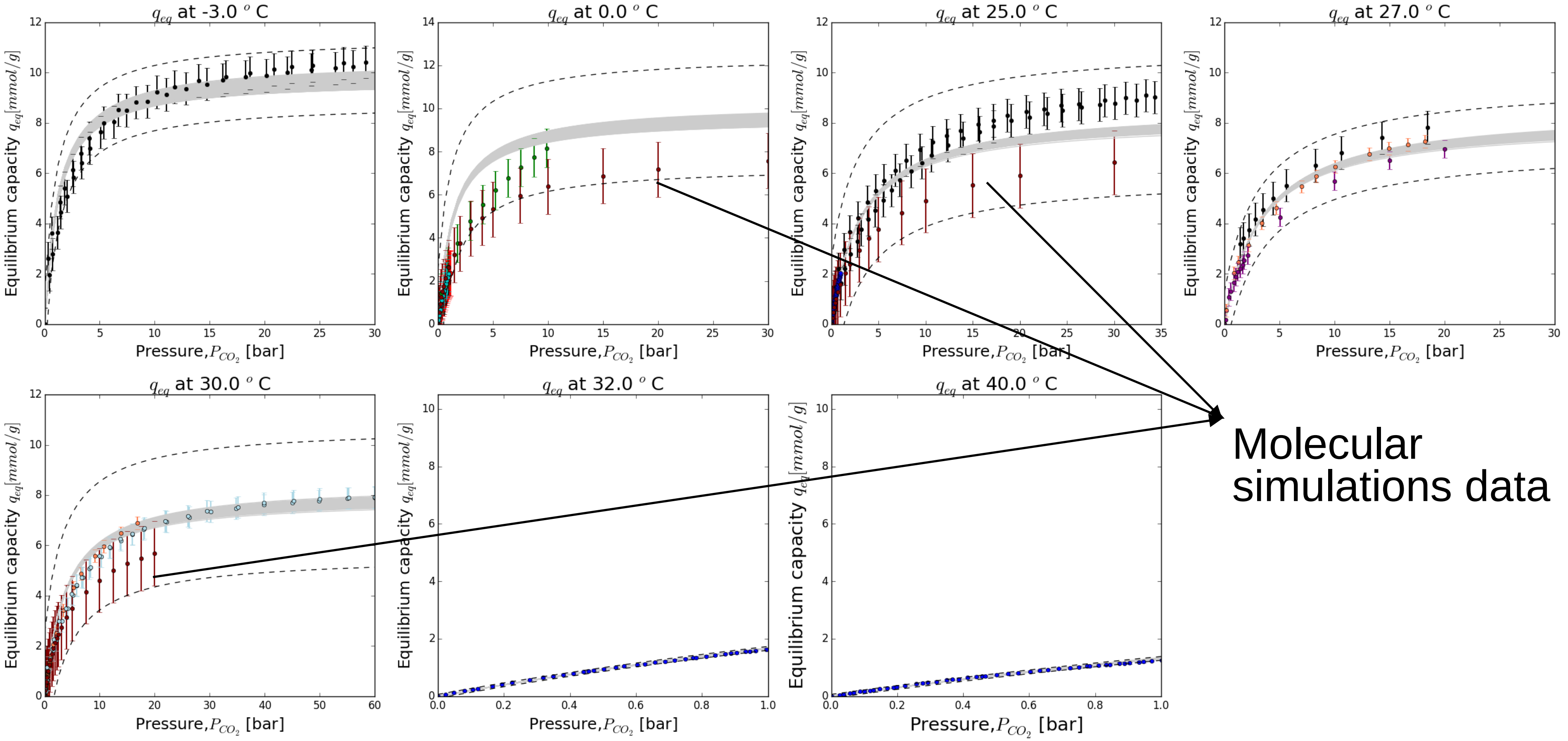
- Dashed lines represent 95% credible intervals

Bayesian fit of all experimental data



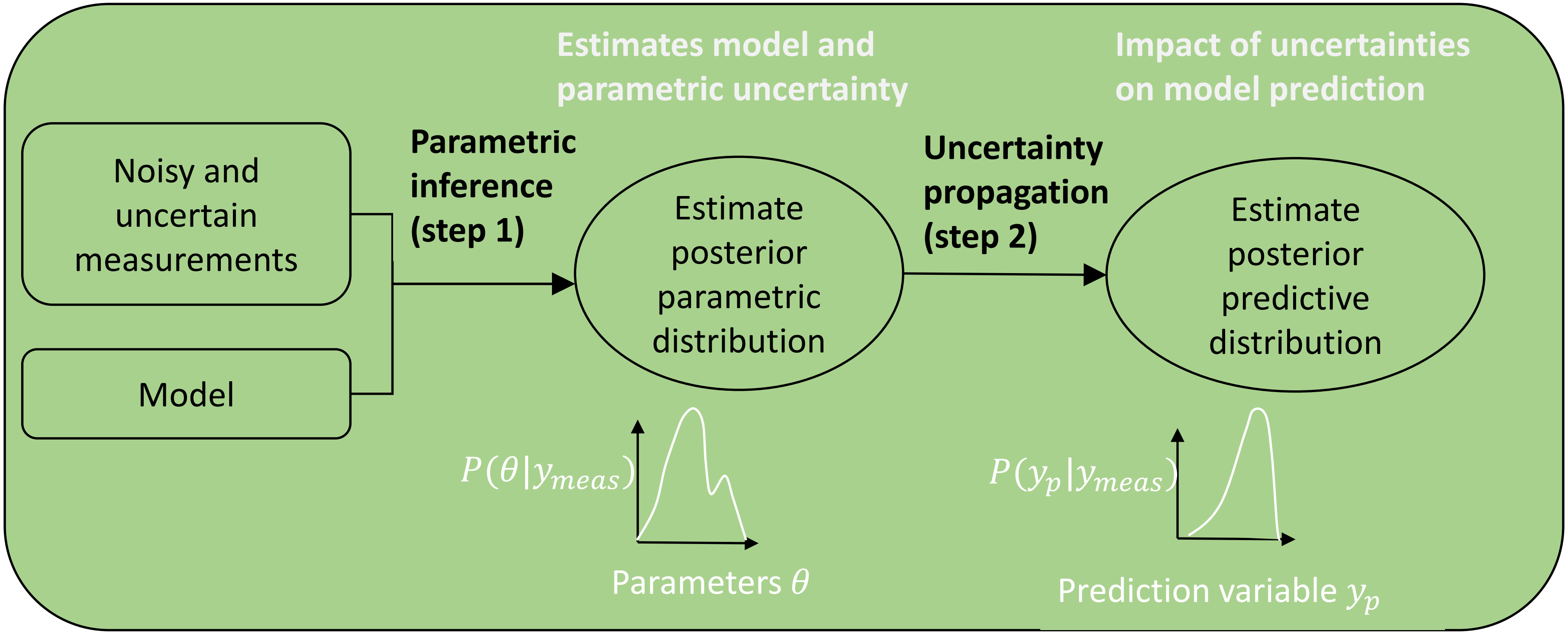
- Each group's data colored separately
- Different error measure for each group estimated and shown
  - Some groups more accurate than others

New data addition to the fit



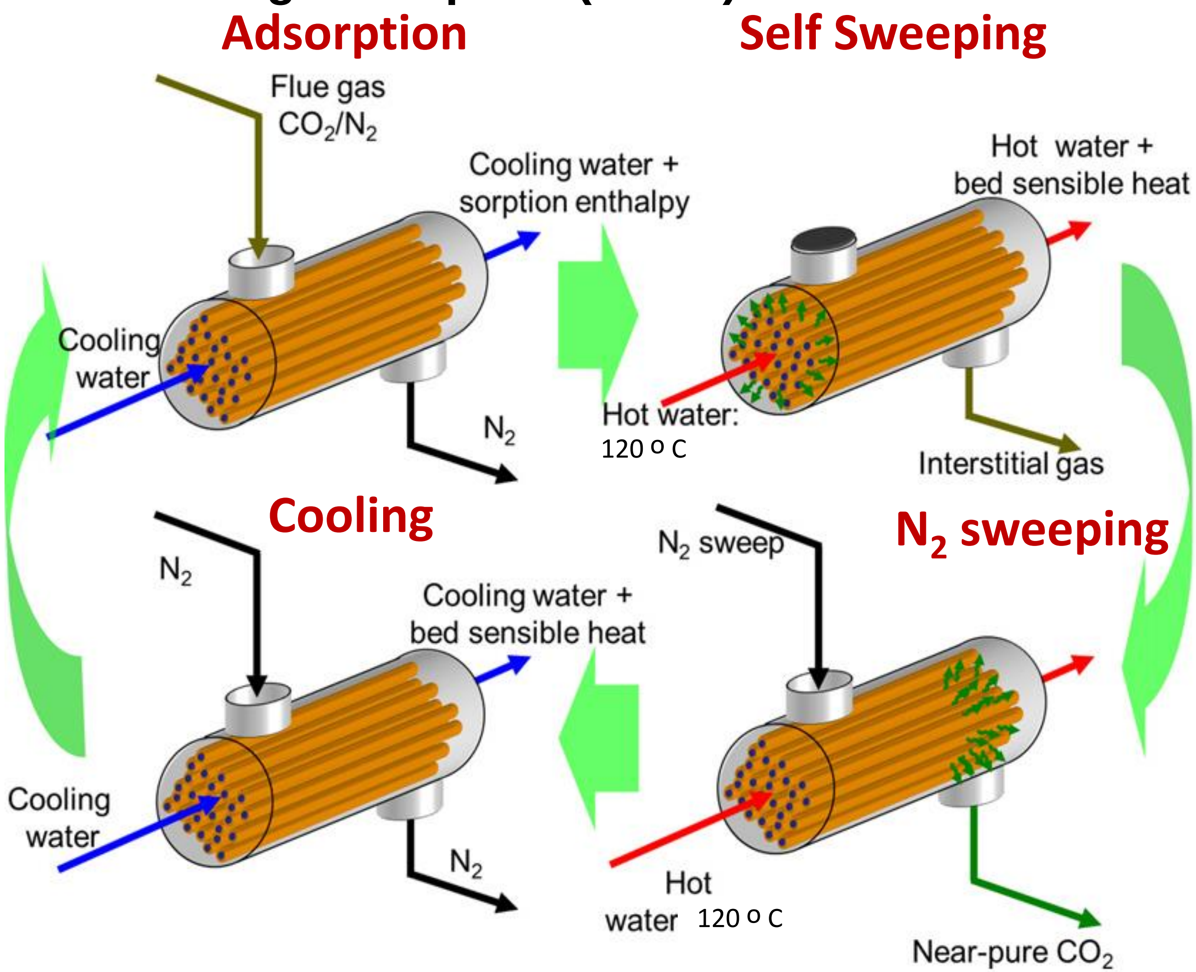
- New data from computational (molecular) simulations added.
  - Increased uncertainty level because of error (std dev = 1.25) in computational data (model uncertainties)

Bayesian uncertainty quantification (UQ)



Application study : Cyclic adsorption process to capture CO_2 from flue gas

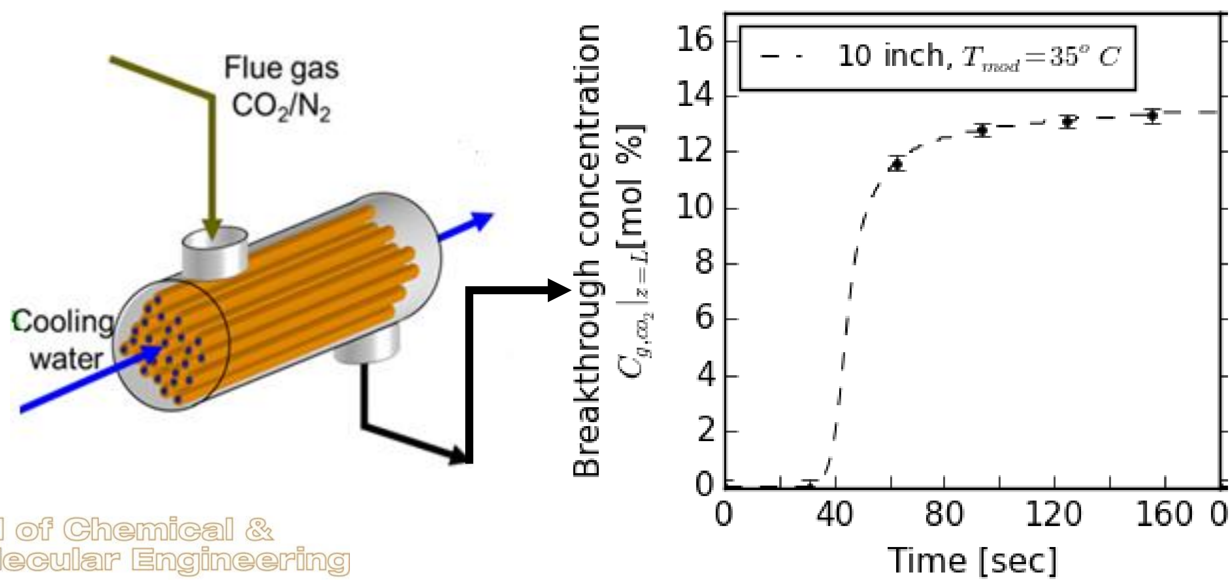
Rapid Thermal Swing Adsorption (RTSA) with a hollow fiber module



1. Kalyanaraman J et al.. Modeling and experimental validation of carbon dioxide sorption on hollow fibers loaded with silica-supported poly(ethylenimine). *Chem.Eng.J.* 2015;259:737-751.

Cyclic process model and the measured experimental data

- Totally *10 unknown model parameters* to be estimated.
- Model composed for eight coupled PDEs - takes around 5 minutes to simulate the adsorption step alone and 40-50 minutes to reach cyclic steady state.
- CO_2 breakthrough profile in adsorption step used to estimate unknown parameters (8 experimental data at varying conditions)



Likelihood distribution $\mathcal{L}(\boldsymbol{y_{meas}}|\boldsymbol{\theta})$

- Measure of how likely the model with parameters θ, fits the data y_{meas}

- Assumes a Gaussian distribution for errors (model mismatch)

$$y_{meas,i} = y_{model,i}(\theta) + \epsilon$$

$$\epsilon \sim \mathcal{N}(0, \sigma_e^2); i = 1..N \text{ (no of experiments)}$$

$$\mathcal{L}(y_{meas}|\theta) = \Pi_i^N \frac{1}{\sqrt{2\pi}\sigma_e} e^{-\frac{(y_{meas,i} - y_{model,i})^2}{2\sigma_e^2}}$$

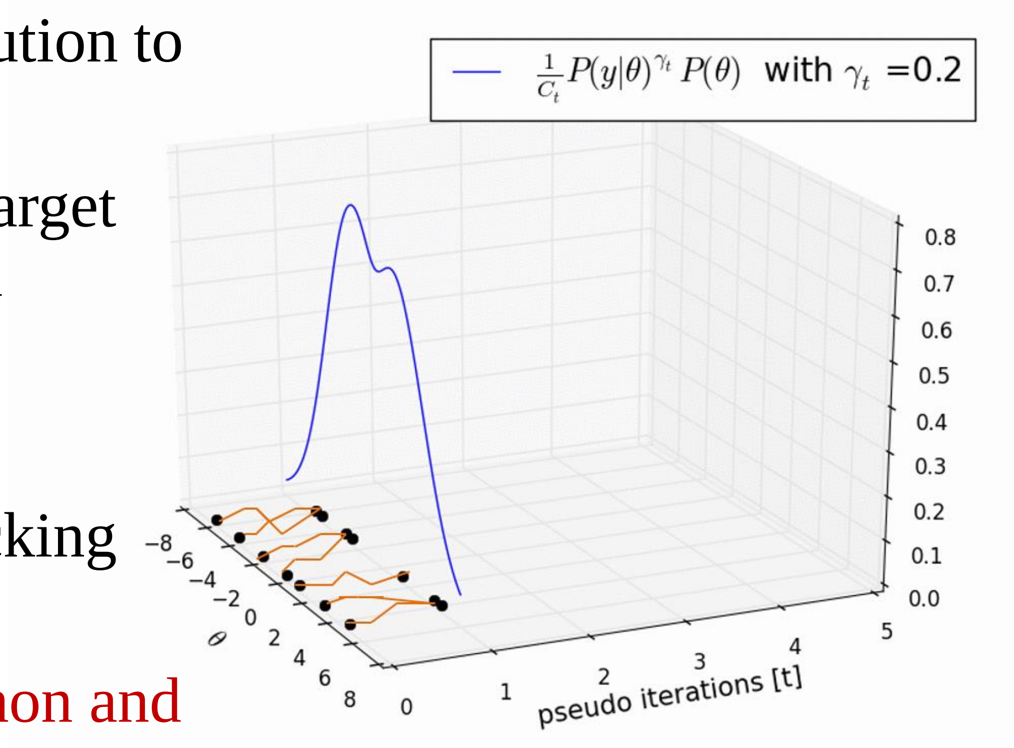
- For every sample value θ, full forward simulation is run for each experiment to calculate likelihood

- Computationally expensive
  - ❖ With CPU time of single simulation $\approx 5 - 6\ min$, SMC requires at least

2 weeks (8 experiments x 12 MCMC runs per iteration x 30 time iterations x 6mins)

Sequential Monte Carlo to determine $P(\theta|y_{meas})$

- Uses samples across the distribution to track
- Makes incremental updates to target distribution with γ_t exponent in likelihood
- Faster convergence
- Parallelizable – one sample tracking per processor
- In our case custom-built in Python and used in parametric inference for our example breakthrough data.

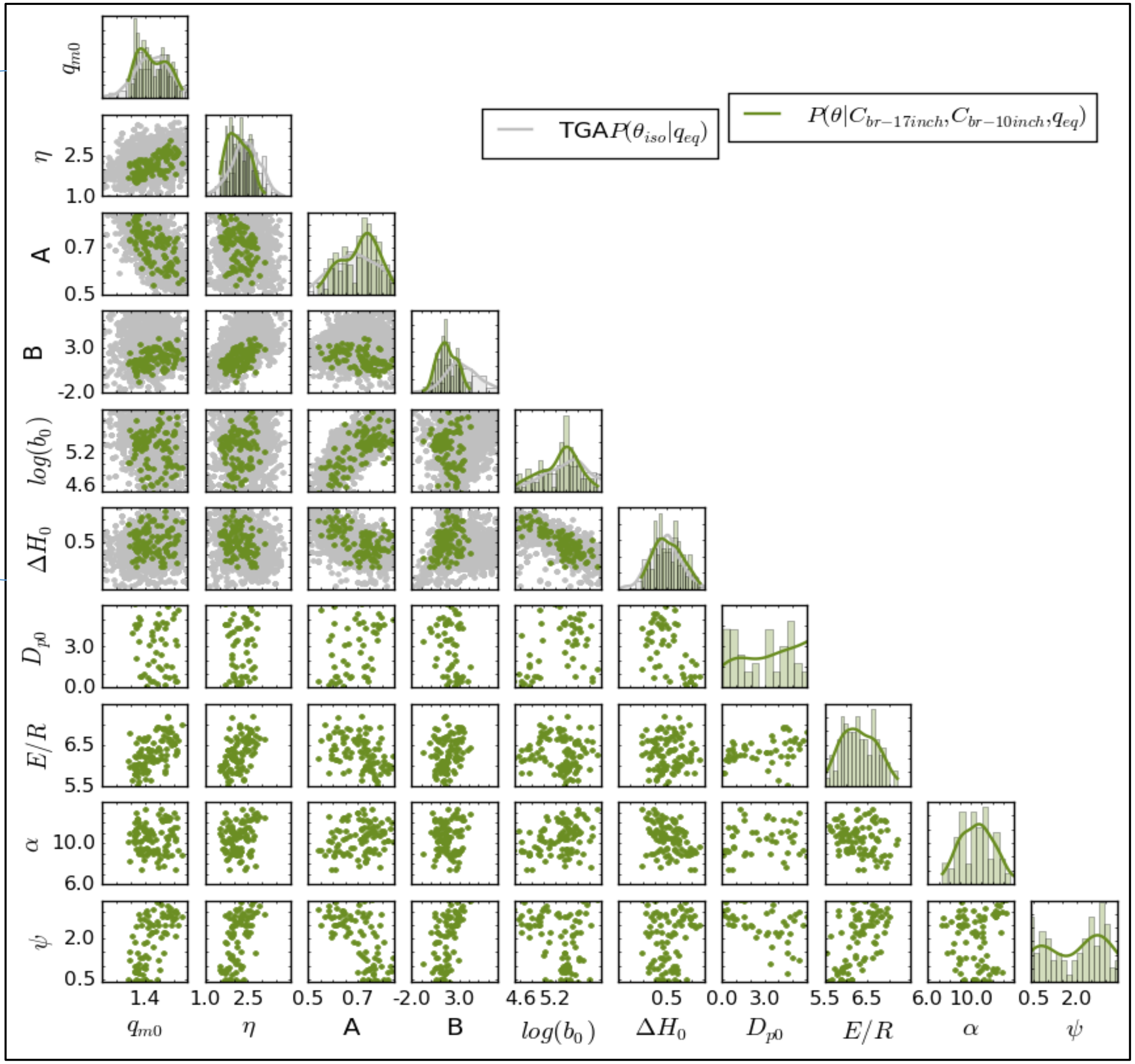


Jeremiah, E. et al, Water Resour. Res, 47,2011

Results on Parameter Estimation of Posterior distribution

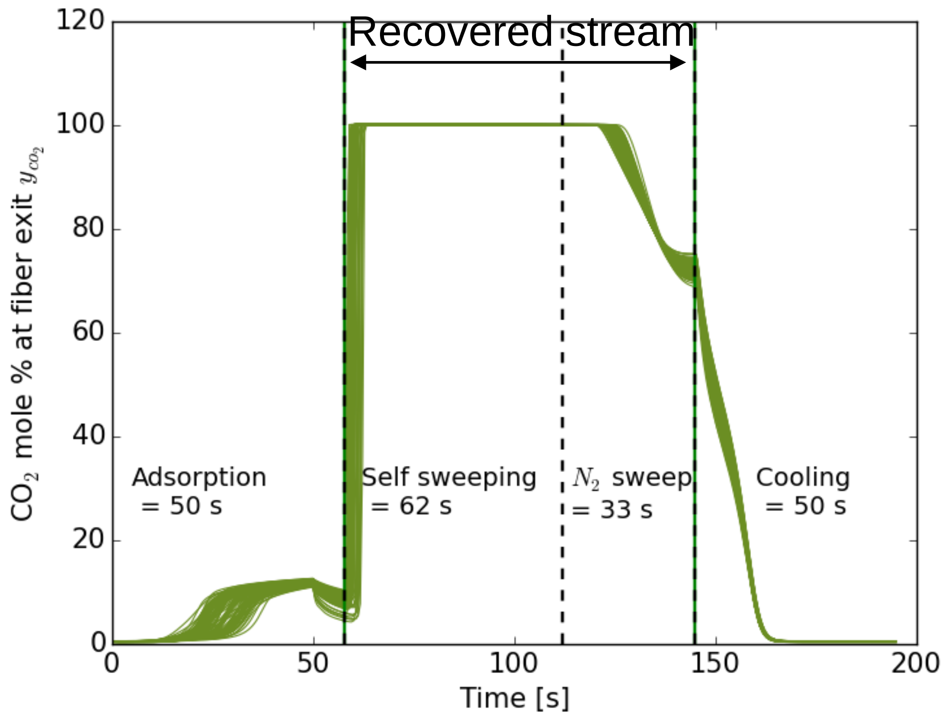
Adsorption isotherm parameters

Mass transfer resistance parameters

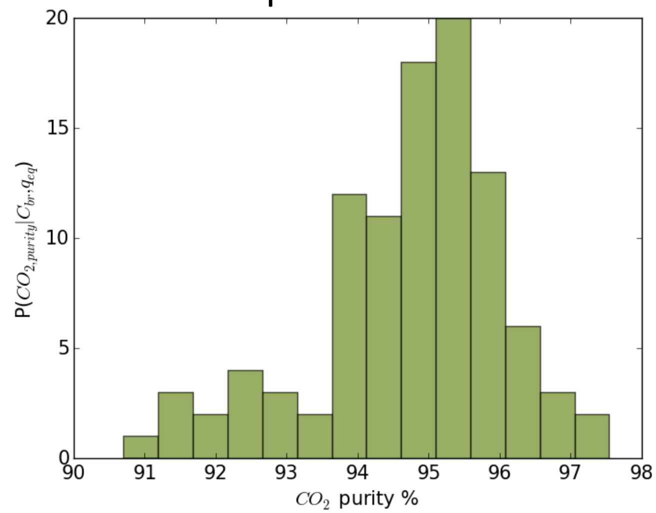


Propagating the uncertain parameters through the cyclic model

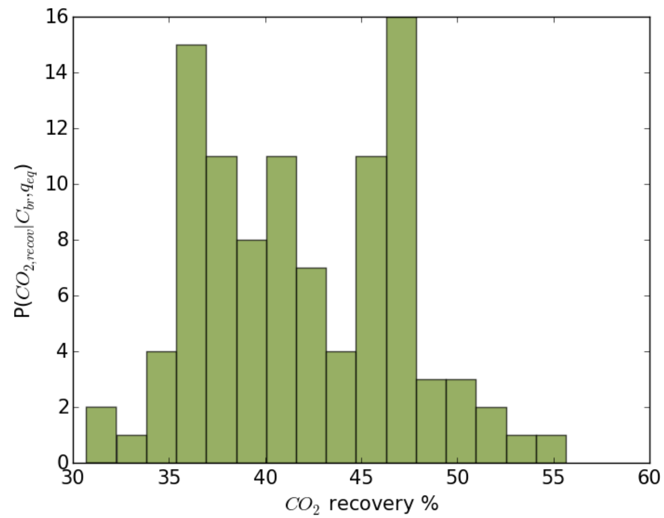
CO$_2$ mole fraction at the exit over the cycle without prediction error



CO$_2$ purity of product stream without prediction error



CO$_2$ recovery of product stream without prediction error



Georgia Tech | School of Chemical & Biomolecular Engineering

Bayesian experimental design

- ❑ Choose most efficient experiment that minimize time and resources
- ❑ Need to quantify uncertainty of process economic metric of y_p
  - ➢ Challenge : Computationally very expensive

Utility function : Value of information at $\boldsymbol{d}$

$$U(d) = \int \log\left(\frac{P(y_p|\, y_{orig}, d)}{P(y_p|y_{orig})}\right) P(y_p|y_{orig}, d) P(y_{orig}|d)\, dy_p\, dy_{orig}$$

Log of pdf – Information content

Relative improvement in distribution of y_p after experiment at d

$$U'(d) = \int \log\left(\frac{P(\theta_s|y_{orig}, d)}{P(\theta_s|y_{orig})}\right) P(\theta|y_{orig}, d) P(y_{orig}|d)\, d\theta\, dy_{orig}$$

Parameters with high sensitivity for y_p- found using Global Sensitivity Analysis

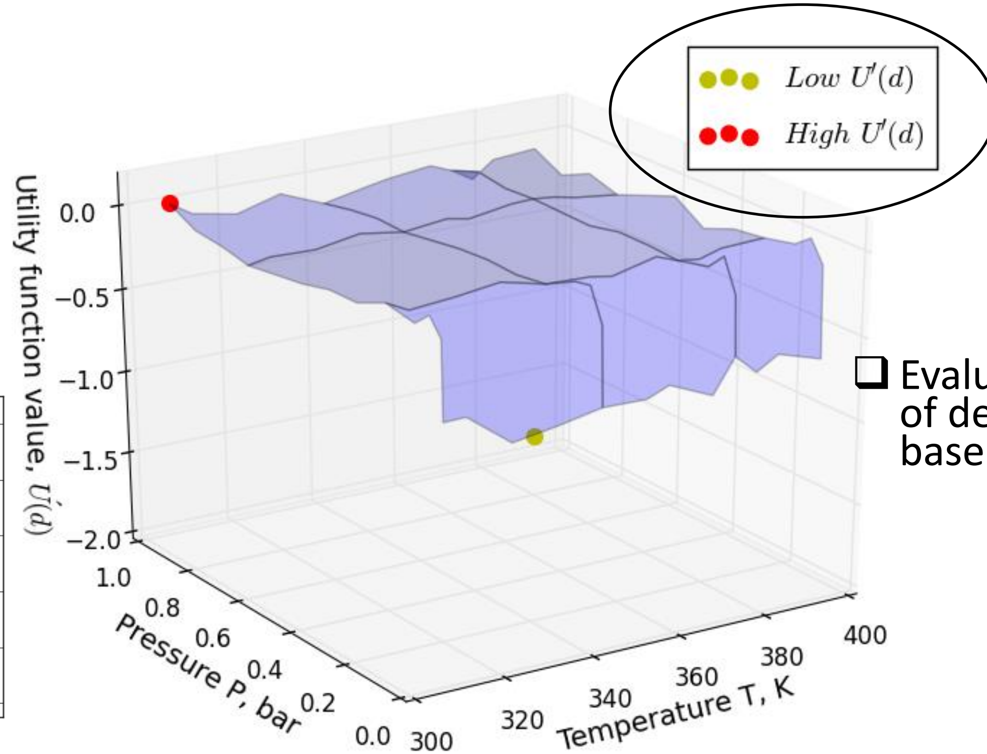
Relative improvement in distribution of θ after experiment at d

Global Sensitivity Analysis – Link between the parametric information to predictive information

Utility Function for Optimal Experiment Design

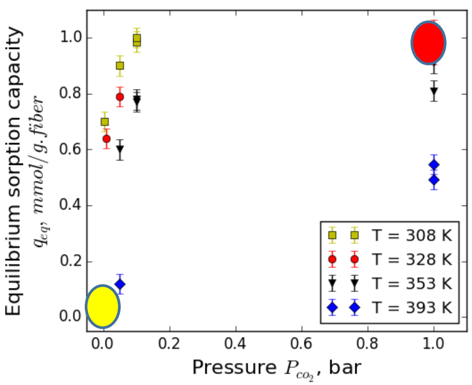
Modified Utility function surface $U'(d)$

$d = \{T, P_{co_2}\}$- Design conditions for experiment



Two extreme cases chosen to verify $U'(d)$

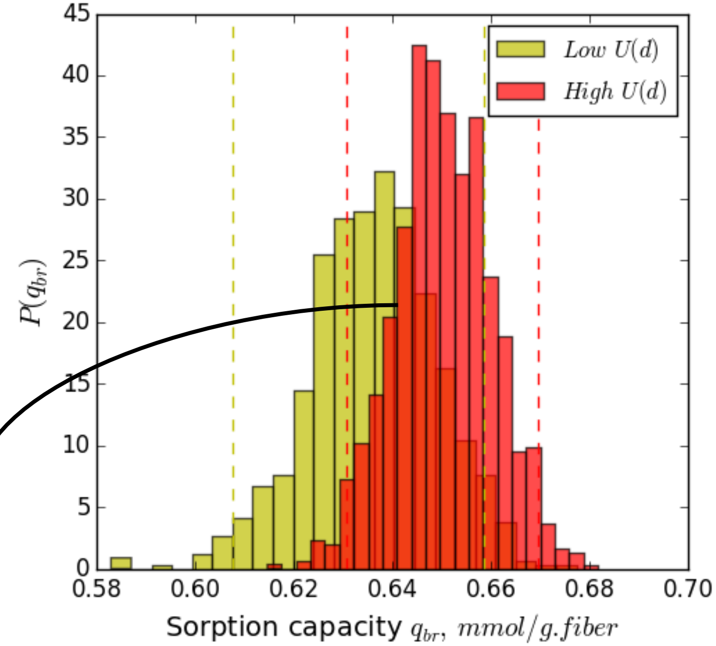
- ❑ Evaluated by enumeration of design space –Simulation based approach



* Rezeai, F. et al. Aminosilane-Grafted Polymer/Silica Hollow Fiber Adsorbents for CO2 Capture from Flue Gas, ACS 2013, 5, 3921-3931

Uncertainty Reduction by Optimal Experimental Design

Uncertainty comparison in Breakthrough capacity (design variable) prediction



Dashed lines indicate 95% credible intervals

- ❑ Leads to ~ 5% reduction of net CO_2 capture cost *

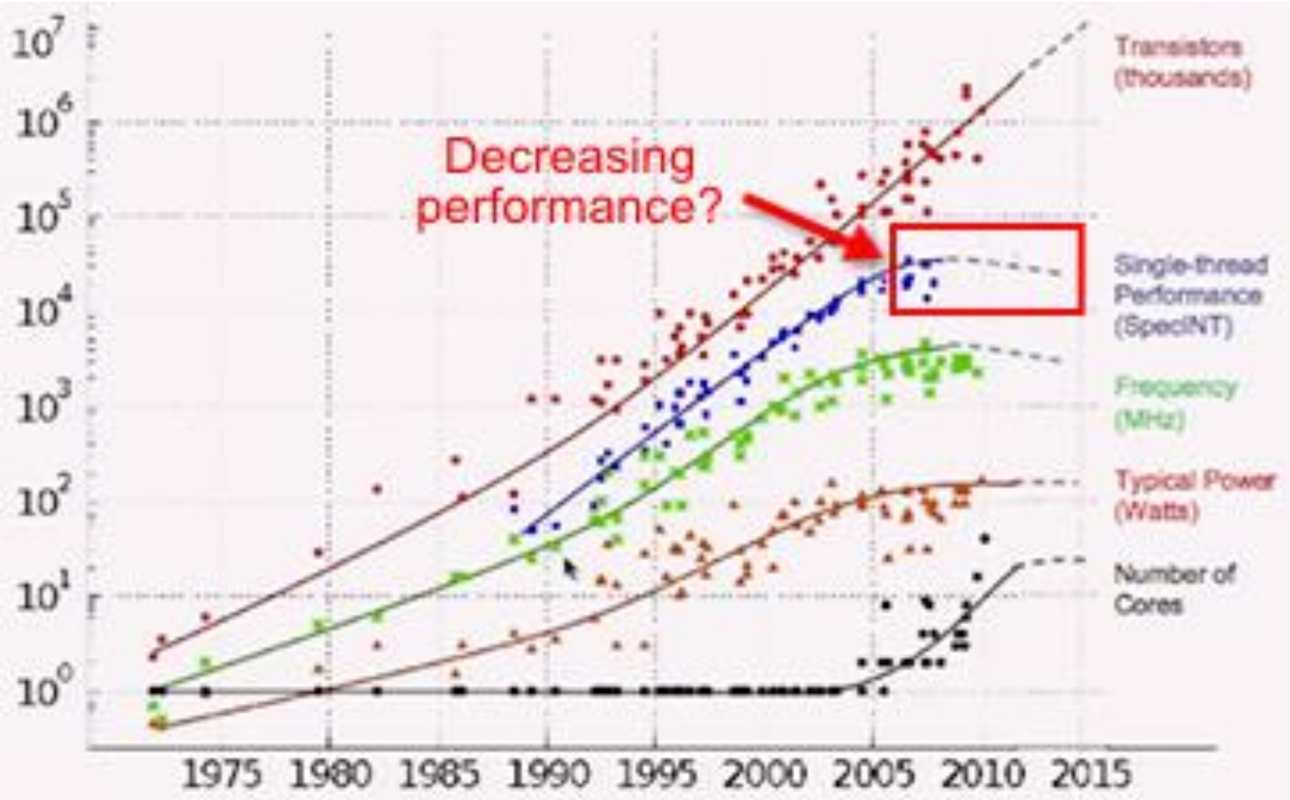
* Kulkarni and Sholl, Analysis of equilibrium based TSA processes for direct CO_2 capture from air IECR (51), 2012

- ❑ Breakthrough capacity considering uncertainty increased from 0.605 to 0.63 mmol/g fiber(5% increase) with one added data

MILP Computing Performance 1988-2017

Algorithm Performance 147,650x
Machine Performance 17,120x
Combined improvement = 2,527,768,000x = 2.5 billion times more effective

Overview of Mixed-integer Programming: Recent Advances, and Future Research Directions , Jeff Linderoth, University Wisconsin-Madison, FOCAPO 2017, Tuscon, Arizona, January 8th-12th.



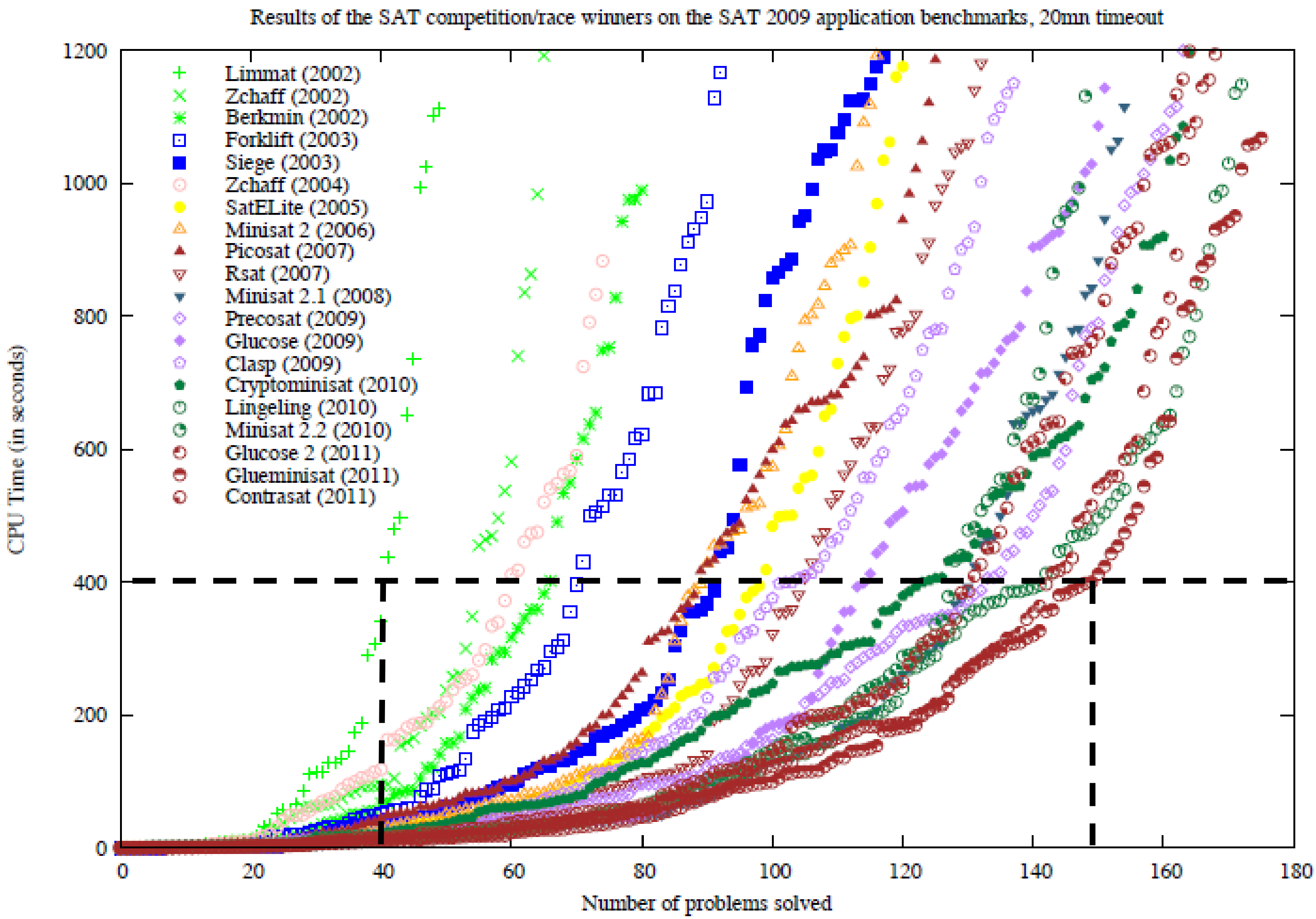
Are we reaching the limits of single-core performance improvements?

Shift towards increased number of cores.

http://www.lanl.gov/orgs/hpc/salishan/salishan2011/3moore.pdf

Logical Reasoning and Inference

E.g. Boolean Satisfiability --SAT



Efficient inference for large scale problems within 20 years.

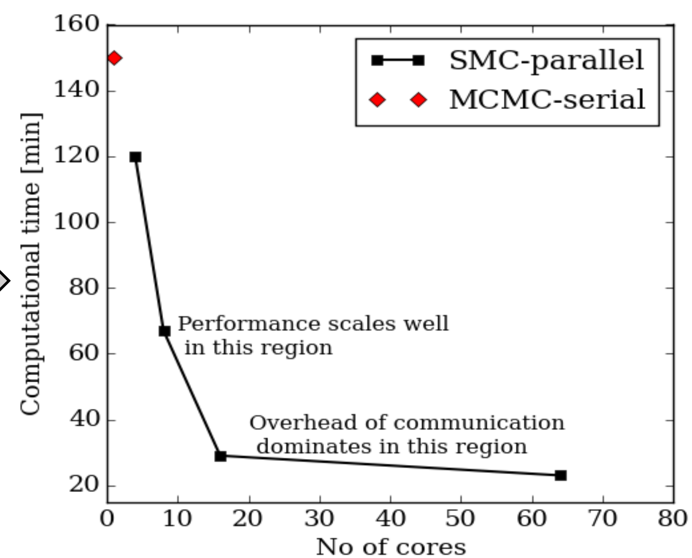
In 2011, the biggest application instance solved by at least one solver contained **10M variables, 32M clauses, and a total of 76M literals.**

https://www.cs.helsinki.fi/u/mjarvisa/papers/jarvisalo-leberre-roussel-simon.aimag.pdf

Computational advances in MCMC and Bayesian analysis in last forty years

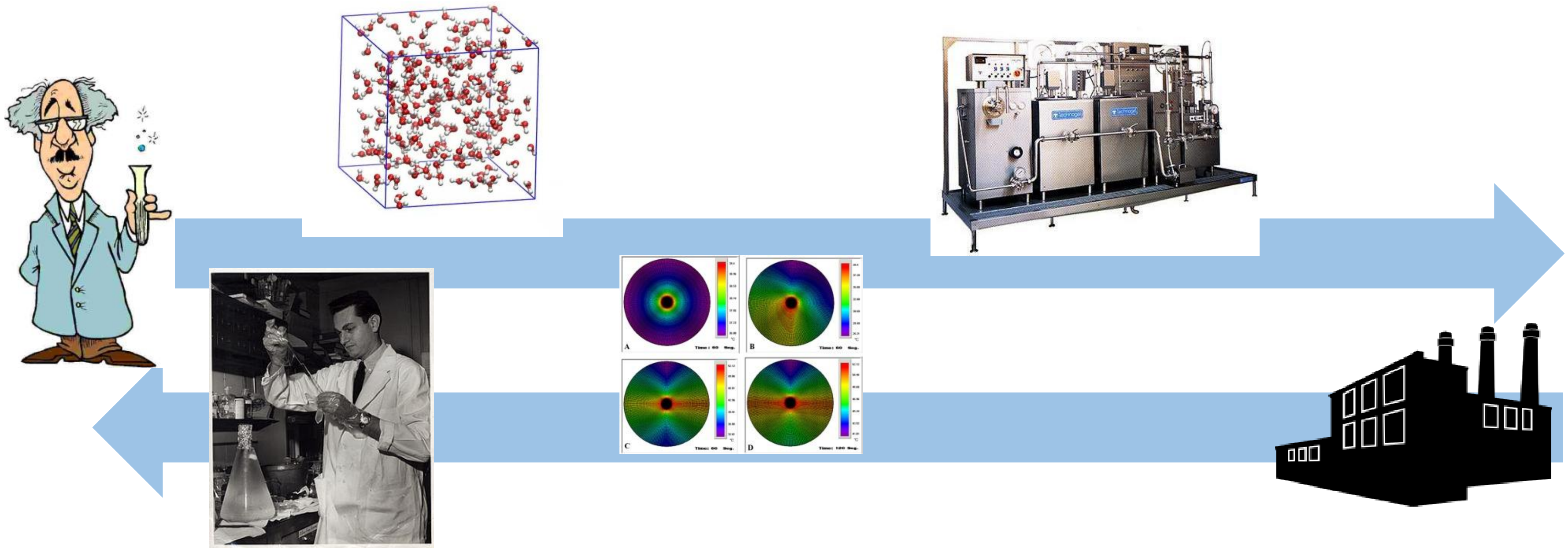
Algorithm Performance in terms of parallel algorithms 500x[1]
Complete parallelism of model evaluations 20x[2]
Machine Performance 17,120x
Combined improvement = 1,712,000,000x = 1.7 billion times more effective

- Parallel computation with SMC vs serial single threaded MCMC computation for Bayesian inference for a small isotherm model (8x speedup)
- For larger complex model speedup upto 500x



1. Workshop on recent advances in Bayesian computation, 2010
2. Amdahl's law

Bayesian Process Systems Engineering 2040



Ambient large datasets AND carefully designed small datasets

Chemical engineering models driven by fundamentals

Massively parallel probabilistic inference

$\models$

?