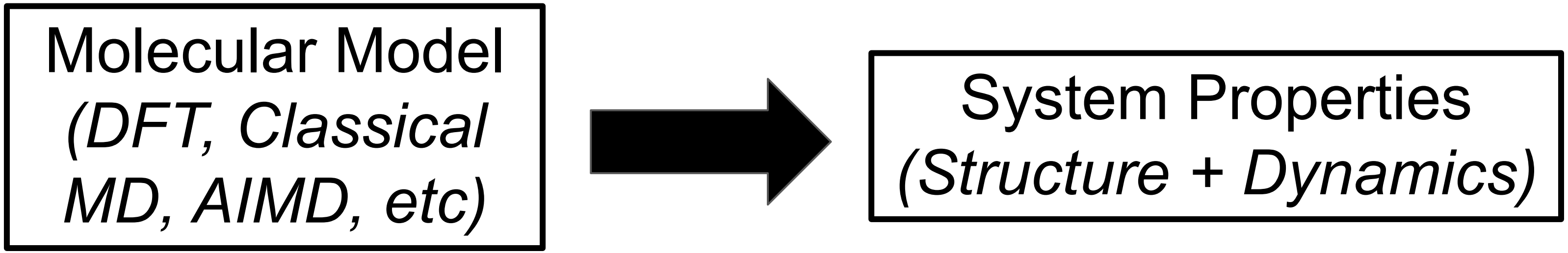
Machine learning accelerated methods to predict interatomic forces from experimental structure measurements

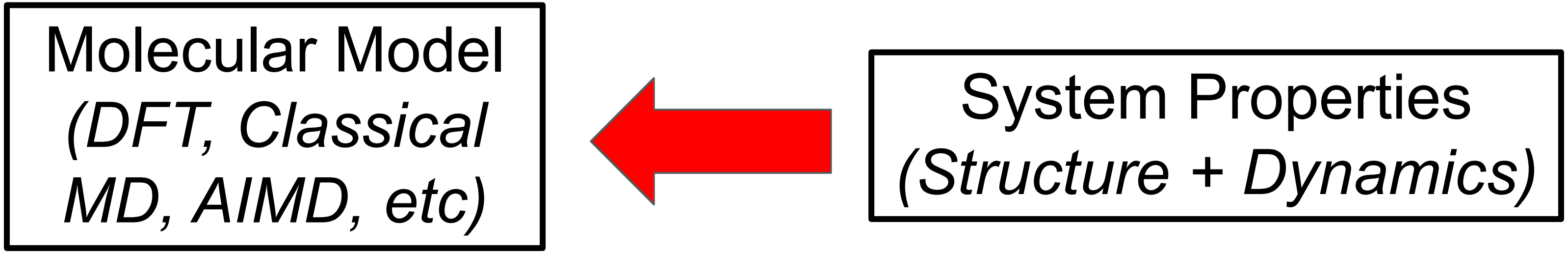
Presenter: **Brennon L. Shanks**
Harry W. Sullivan
Abdur R. Shazed
Michael P. Hoepfner

University of Utah, Department of Chemical Engineering

The Standard Paradigm for Atomistic Modeling

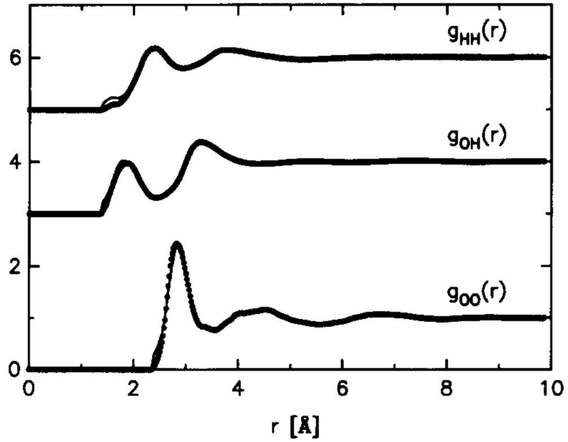


The Inverse Problem for Model Design



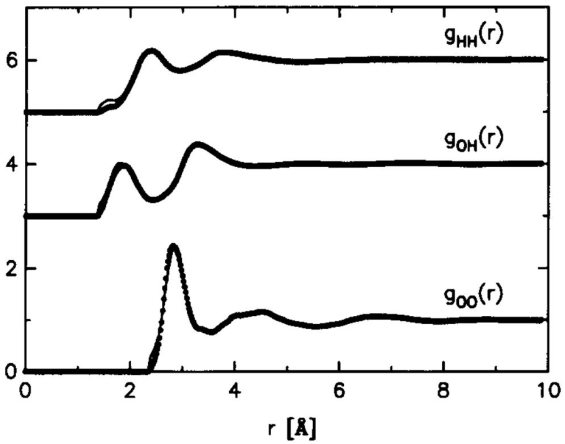
Applications of the Inverse Problem for Interesting Chemistry

Scattering Analysis
Alan Soper 1996, *Chem. Phys.*

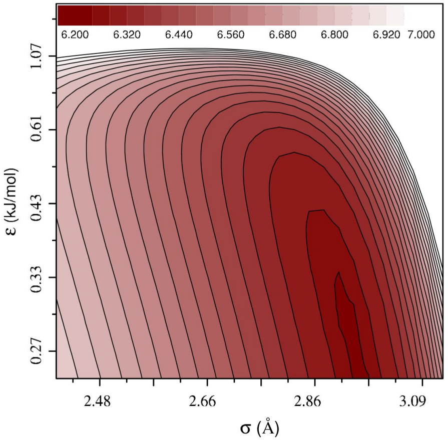


Applications of the Inverse Problem for Interesting Chemistry

Scattering Analysis
Alan Soper 1996, *Chem. Phys.*

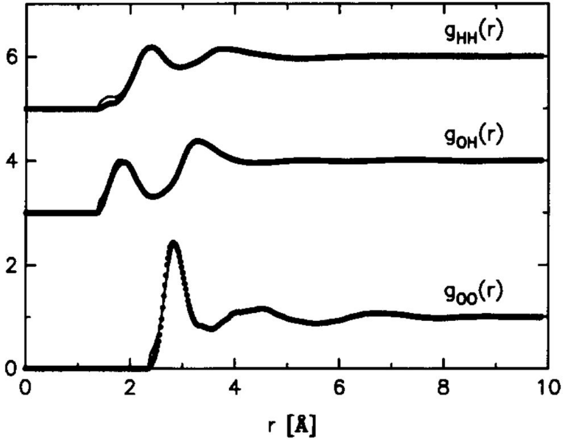


Coarse-Graining
Scott Shell 2013,
J. Chem. Phys.

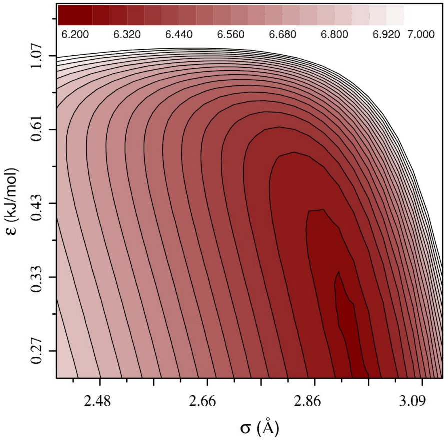


Applications of the Inverse Problem for Interesting Chemistry

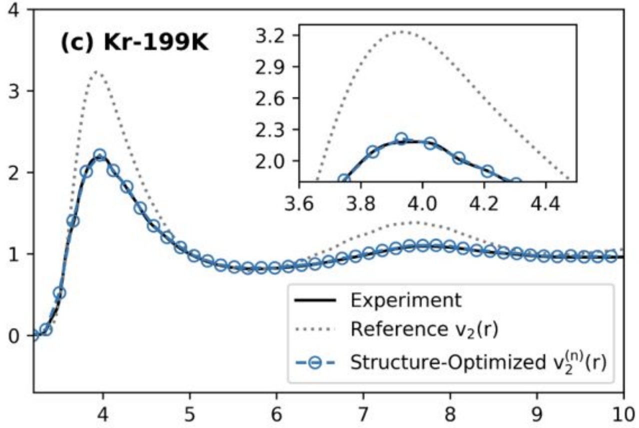
Scattering Analysis
Alan Soper 1996, *Chem. Phys.*



Coarse-Graining
Scott Shell 2013,
J. Chem. Phys.

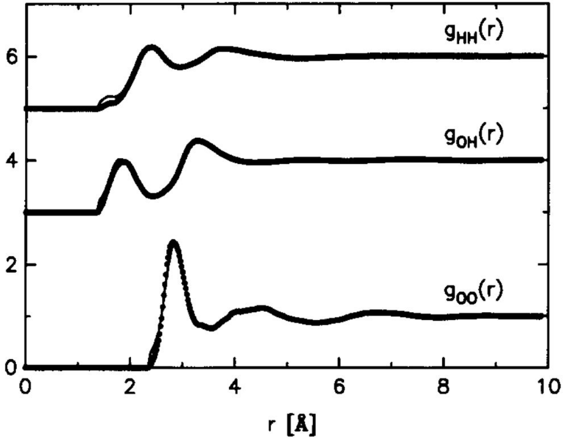


Force Field Development in Simple Liquids
Brennon Shanks 2022, *J. Phys. Chem. Lett.*

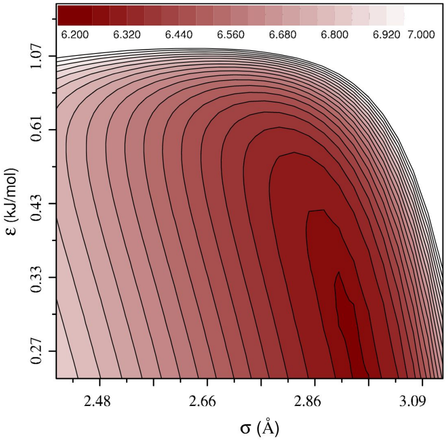


Applications of the Inverse Problem for Interesting Chemistry

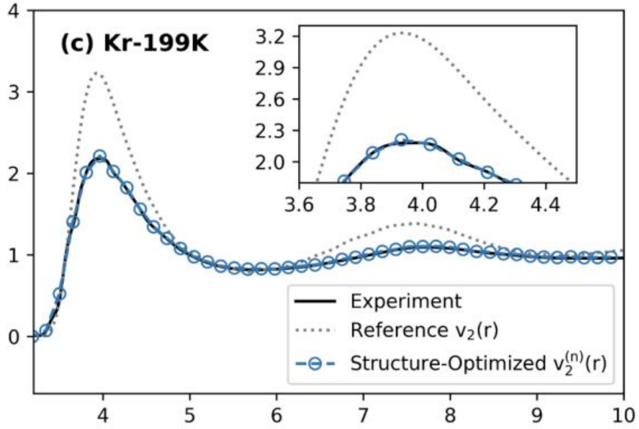
Scattering Analysis
Alan Soper 1996, *Chem. Phys.*



Coarse-Graining
Scott Shell 2013,
J. Chem. Phys.



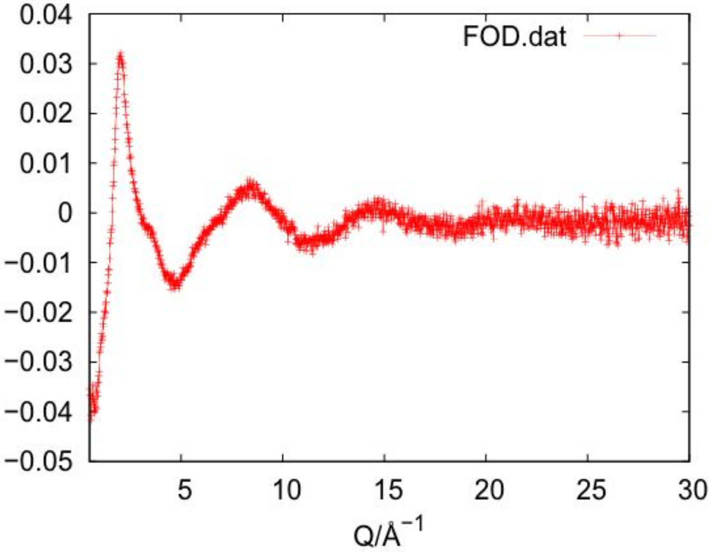
Force Field Development in Simple Liquids
Brennon Shanks 2022, *J. Phys. Chem. Lett.*



Inverse methods are widely used for scattering analysis and coarse-graining **but** <u>not to build force fields from experimental data</u>… **why?**

Unique Challenges of Inverting Experimental Scattering Data

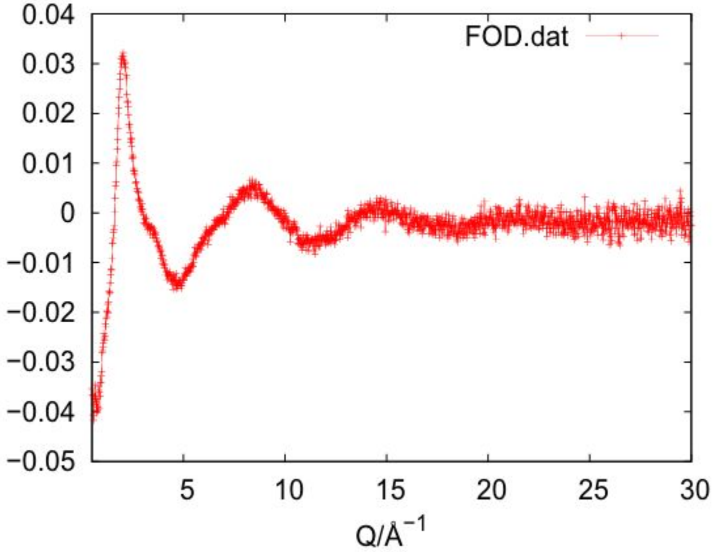
Measurement Uncertainty



Noise in the Structure Factor of Water
Neuefiend 2012, *Nuc. Inst. Methods.*

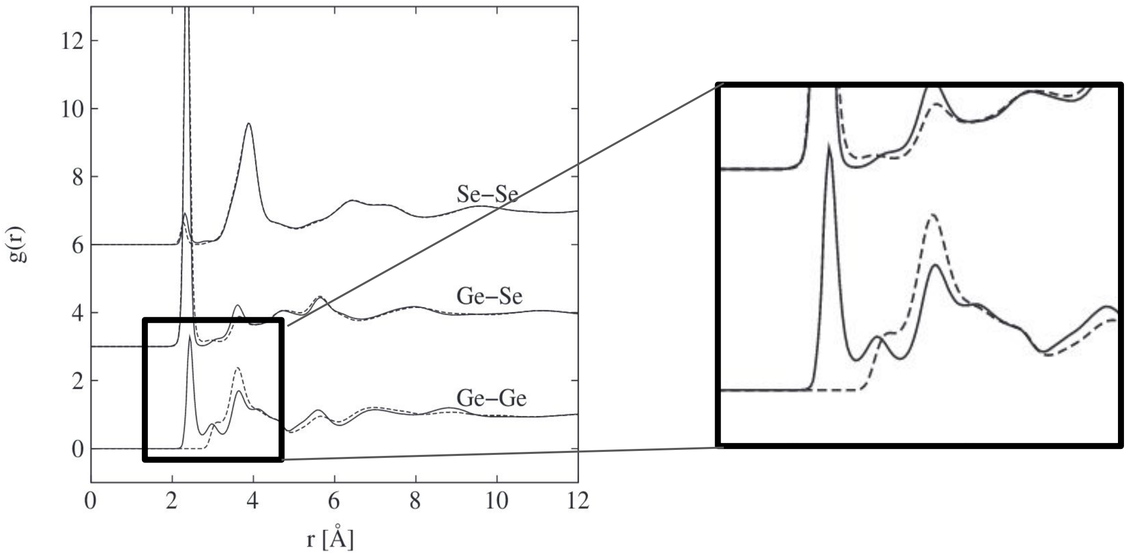
Unique Challenges of Inverting Experimental Scattering Data

Measurement Uncertainty



Noise in the Structure Factor of Water
Neuefiend 2012, *Nuc. Inst. Methods.*

"Experimental" Model Uncertainty



Real-space structure is non-unique
Alan Soper 2007, *Condens. Matt.*

Bayesian Methods Can Quantify Parameter Uncertainty Given Uncertain Experimental Data

The Basic Outline of Bayesian Approaches

Bayesian Methods Can Quantify Parameter Uncertainty Given Uncertain Experimental Data

The Basic Outline of Bayesian Approaches

(1) Define **'prior'** distributions

$$p(\theta)$$

Bayesian Methods Can Quantify Parameter Uncertainty Given Uncertain Experimental Data

The Basic Outline of Bayesian Approaches

(1) Define **'prior'** distributions

(2) Define and evaluate a **'likelihood'** function

$$p(\mathscr{Y}|\theta)p(\theta)$$

Bayesian Methods Can Quantify Parameter Uncertainty Given Uncertain Experimental Data

The Basic Outline of Bayesian Approaches

(1) Define **'prior'** distributions

(2) Define and evaluate a **'likelihood'** function

(3) Solve for the **'posterior'** distribution

$$p(\theta|\mathscr{Y}) = \frac{p(\mathscr{Y}|\theta)p(\theta)}{p(\mathscr{Y})}$$

The posterior is a direct quantification of parameter uncertainty based on your experimental data, Y.

Bayesian Methods Can Quantify Parameter Uncertainty Given Uncertain Experimental Data

The Basic Outline of Bayesian Approaches

(1) Define **'prior'** distributions

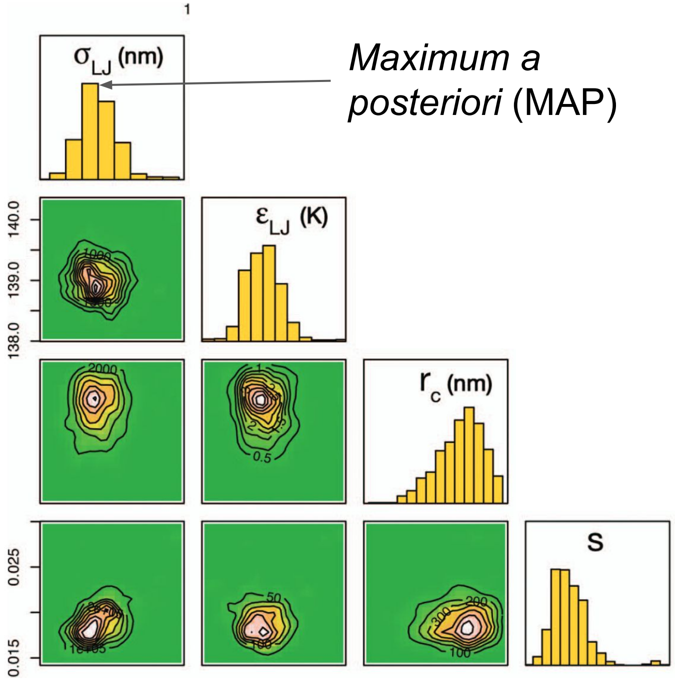
(2) Define and evaluate a **'likelihood'** function

(3) Solve for the **'posterior'** distribution

$$p(\theta|\mathscr{Y}) = \frac{p(\mathscr{Y}|\theta)p(\theta)}{p(\mathscr{Y})}$$

The posterior is a direct quantification of parameter uncertainty based on your experimental data, Y.

Parameter Uncertainty



Marginal Posteriors on LJ Parameters
Koumoutsakos 2015, *J. Chem. Phys.*

Accelerated Bayesian Inference with Gaussian Process Surrogates

Evaluating the Bayesian likelihood is easy! Just run ~1 million molecular simulations to populate the model parameter space and you're done!

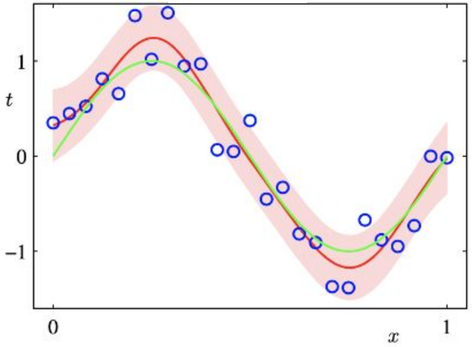
Accelerated Bayesian Inference with Gaussian Process Surrogates

Evaluating the Bayesian likelihood is easy! Just run ~1 million molecular simulations to populate the model parameter space and you're done!

Instead, we train a GP on N ~ 1000 simulations
For data containing η independent variables.

~200 fold speed up

$GP(\boldsymbol{\theta}^*)$



99% LOOCV Score

$(N\eta \text{ x } \dim(\boldsymbol{\theta}) + 1)$

$$\hat{\mathbf{X}} = \begin{bmatrix} \theta_{1,1} & \theta_{2,1} & \dots & r_1 \\ \theta_{1,1} & \theta_{2,1} & \dots & r_2 \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{1,1} & \theta_{2,1} & \dots & r_\eta \\ \theta_{1,2} & \theta_{2,2} & \dots & r_1 \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{1,N} & \theta_{2,N} & \dots & r_\eta \end{bmatrix}$$

Accelerated Bayesian Inference with Gaussian Process Surrogates

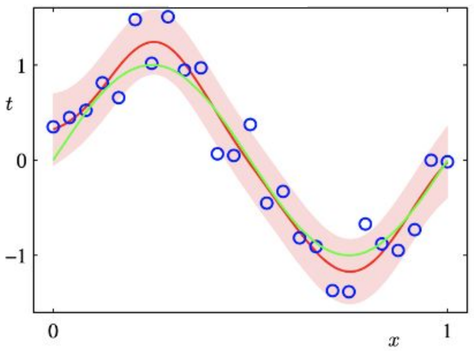
Evaluating the Bayesian likelihood is easy! Just run ~1 million molecular simulations to populate the model parameter space and you're done!

Instead, we train a GP on N ~ 1000 simulations For data containing η independent variables.

~200 fold speed up

$(N\eta \text{ x } \dim(\boldsymbol{\theta}) + 1)$

$GP(\boldsymbol{\theta}^*)$



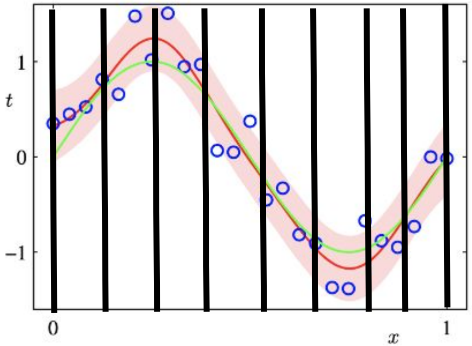
99% LOOCV Score

$$\hat{\mathbf{X}} = \begin{bmatrix} \theta_{1,1} & \theta_{2,1} & \dots & r_1 \\ \theta_{1,1} & \theta_{2,1} & \dots & r_2 \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{1,1} & \theta_{2,1} & \dots & r_\eta \\ \theta_{1,2} & \theta_{2,2} & \dots & r_1 \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{1,N} & \theta_{2,N} & \dots & r_\eta \end{bmatrix}$$

Local GPs reduce matrix size and are about 600 fold faster than full GPs

$GP_k(\boldsymbol{\theta}^*)$

$\eta \quad (N \text{ x } \dim(\boldsymbol{\theta}))$

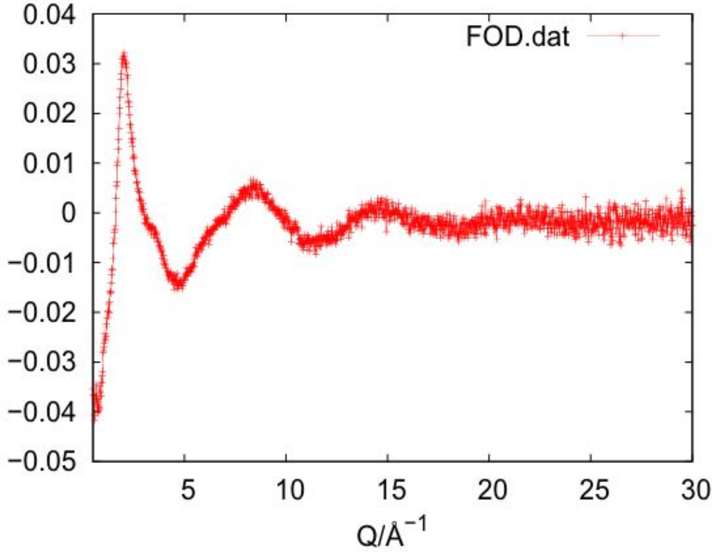


99% LOOCV Score

$$\hat{\mathbf{X}}' = \begin{bmatrix} \theta_{1,1} & \theta_{1,2} & \dots \\ \theta_{1,2} & \theta_{2,2} & \dots \\ \vdots & \vdots & \vdots \\ \theta_{1,N} & \theta_{2,N} & \dots \end{bmatrix}$$

Application: Understanding Experimental Uncertainty Under a Known Model

Measurement Uncertainty



Noise in the Structure Factor of Water

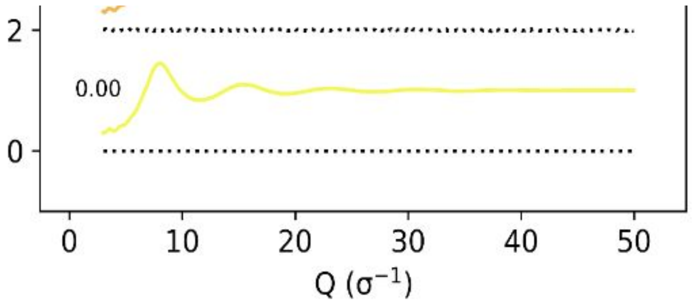
Neuefiend 2012, *Nuc. Inst. Methods.*

Investigating the impact of measurement uncertainty in Mie fluids

Mie Fluid Interaction Potential

$$v_2^{\mathrm{Mie}}(r) = \frac{\lambda}{\lambda - 6}\left(\frac{\lambda}{6}\right)^{6/\lambda - 6} \varepsilon \left[\left(\frac{\sigma}{r}\right)^{\lambda} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

Run Model Simulation

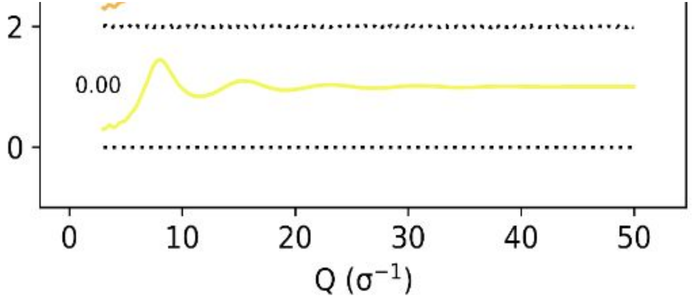


Investigating the impact of measurement uncertainty in Mie fluids

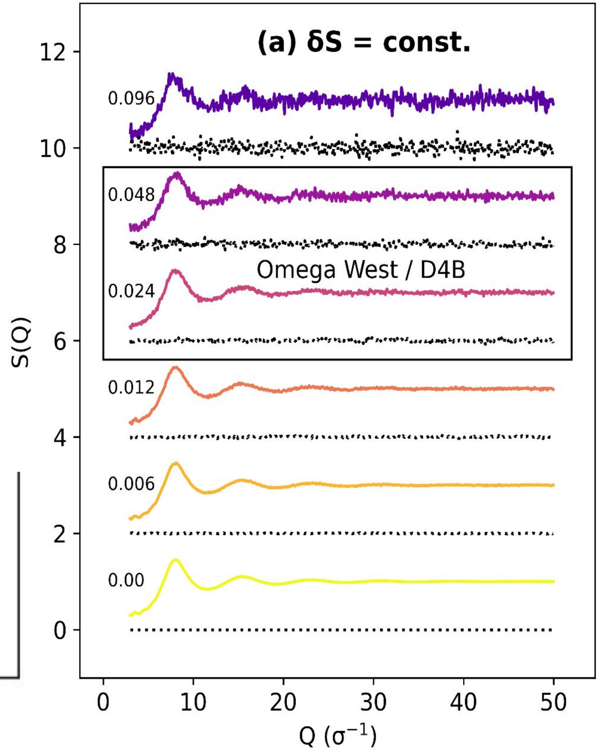
Mie Fluid Interaction Potential

$$v_2^{\mathrm{Mie}}(r) = \frac{\lambda}{\lambda-6}\left(\frac{\lambda}{6}\right)^{6/\lambda-6} \varepsilon \left[\left(\frac{\sigma}{r}\right)^{\lambda} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

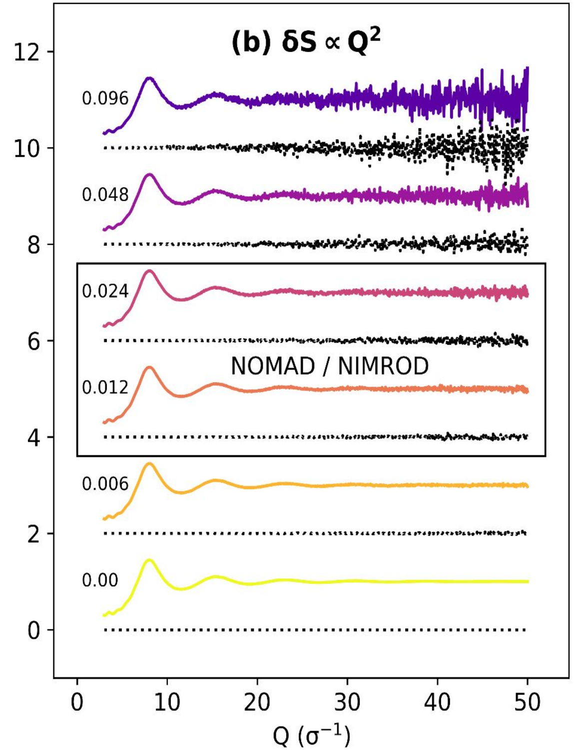
Run Model
Simulation



Reactor Source

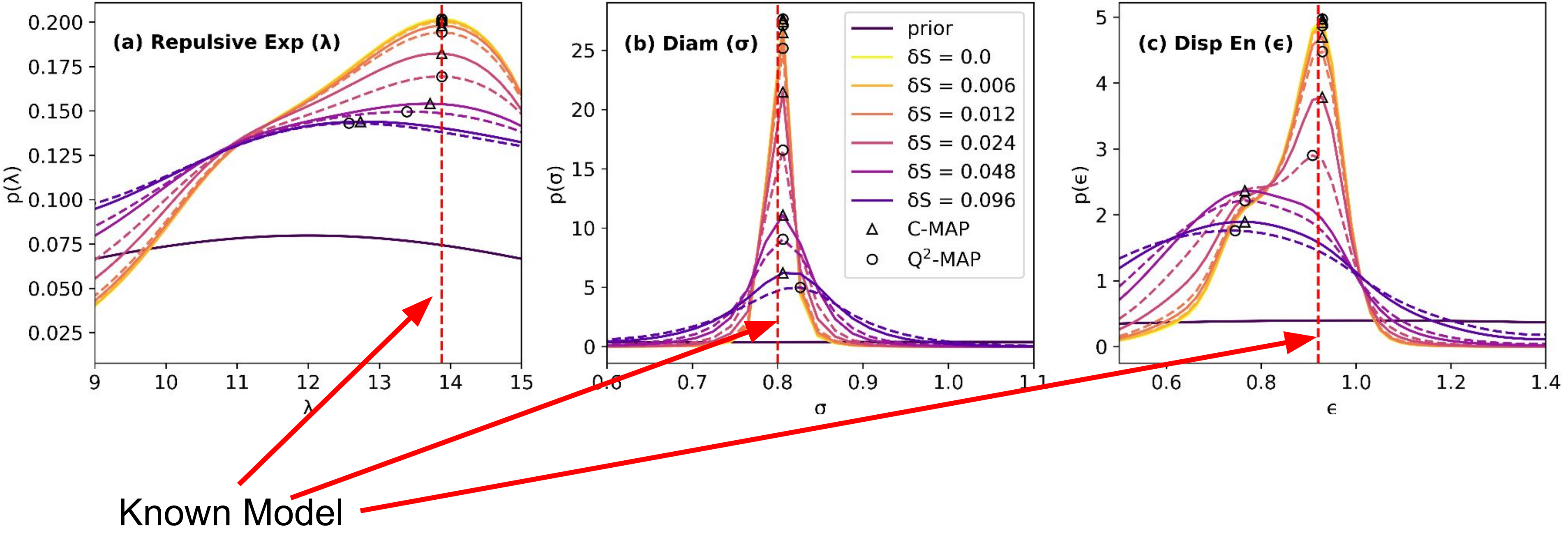


Spallation Source



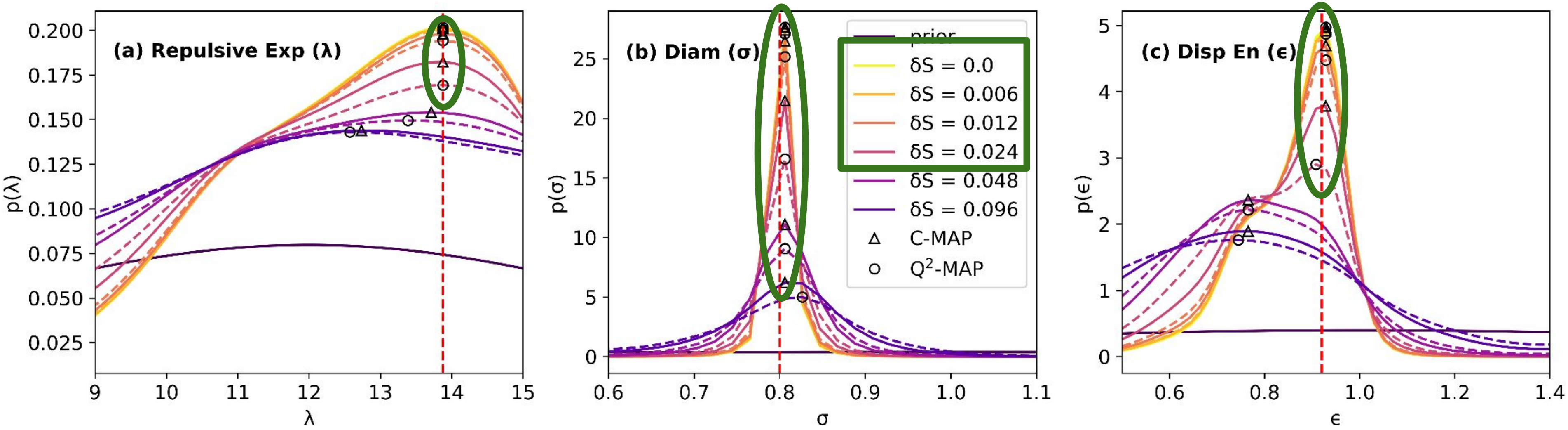
Can we recover our original model from the structure?

Bayesian Marginal Probability Distribution on Model Parameters



Can we recover our original model from the structure?

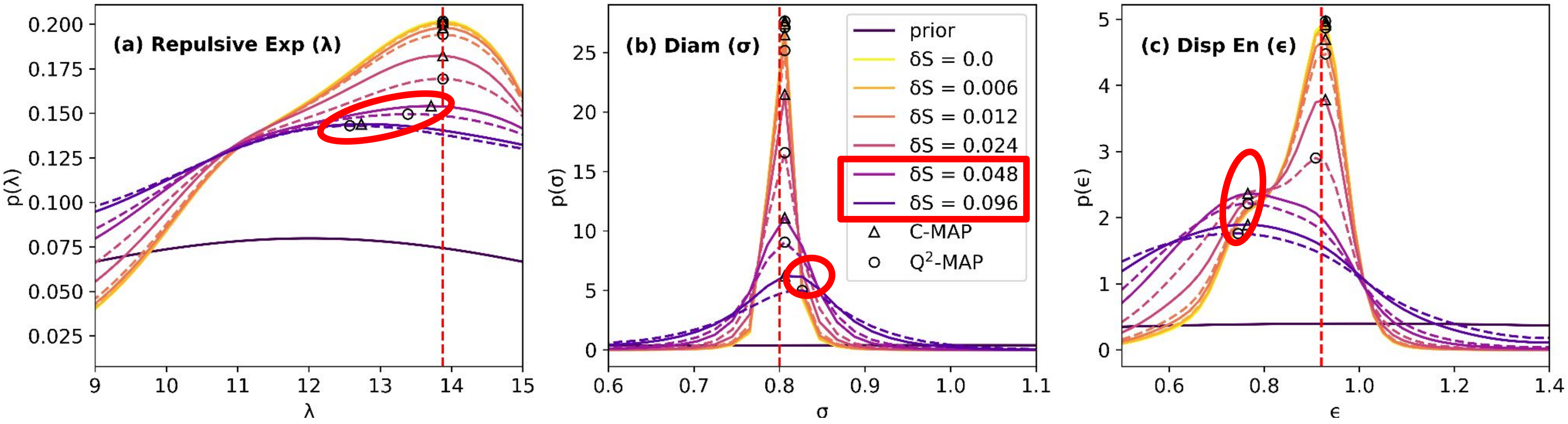
Bayesian Marginal Probability Distribution on Model Parameters



Bayesian optimization recovers force field parameters with high-accuracy for low uncertainty structure factor measurements.

Can we recover our original model from the structure?

Marginal Probability Distribution on Model Parameters

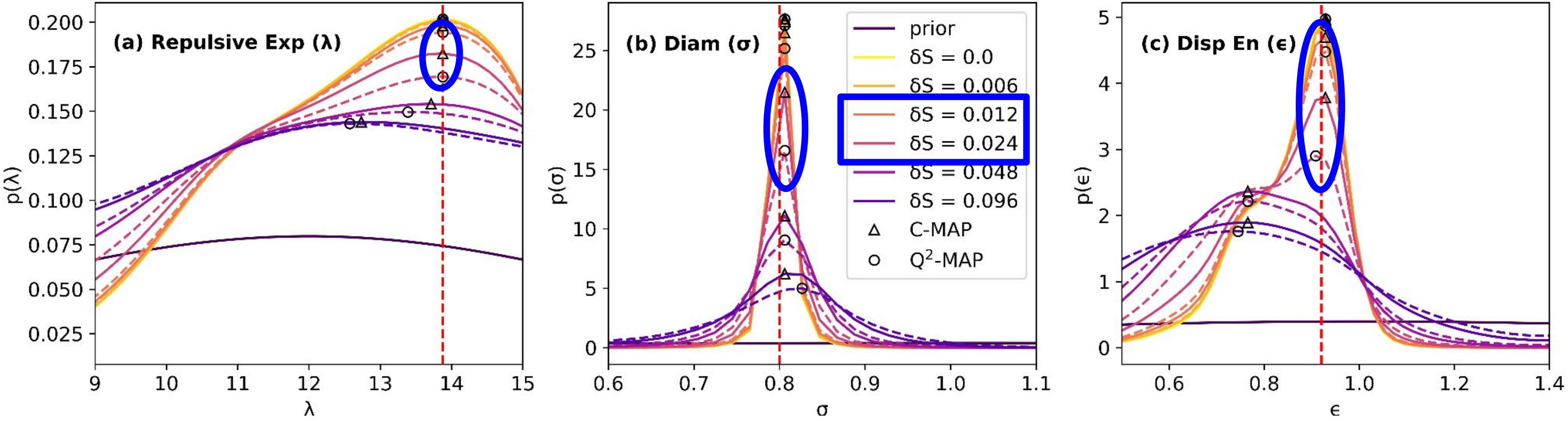


Uncertainty increases and accuracy declines rapidly below a 0.024 variance.

This data quality is representative of the 1960s-1980s neutron sources.

Can we recover our original model from the structure?

Marginal Probability Distribution on Model Parameters



Existing instruments (NOMAD/NIMROD) can provide measurements below the precision threshold.

Summary and Key Takeaways

- Inverse problems are useful for interesting chemistry, including scattering analysis, coarse-graining, and atomistic force field development.
- Bayesian uncertainty quantification is a rigorous framework to understand our confidence in our models.
- We need ML accelerated approaches to populate Bayesian likelihood distributions.
- Bayesian UQ can help answer fundamental questions around interatomic forces, enable active learning approaches using decision theory, and rigorously incorporate experimental data into atomistic models.