Sandia National Laboratories

SAND2016-2559C

Model Selection Under Uncertainty: Coarse-Graining Atomistic Models

Kathryn Farrell-Maupin

J. Tinsley Oden and Danial Faghihi

SIAM UQ 2016
Lausanne, Switzerland

April 5-8, 2016

U.S. DEPARTMENT OF ENERGY

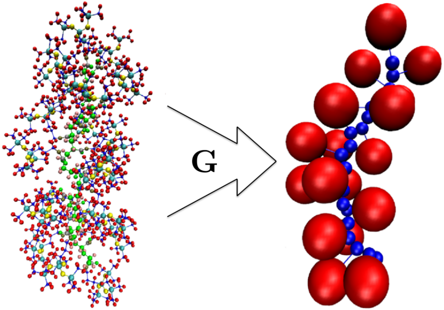
National Nuclear Security Administration

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND NO. 2011-XXXXP

Outline

1. Coarse-Graining Atomistic Systems
   - Motivation
   - The AA System
   - The CG System
   - Uncertainties in the CG model
2. Bayesian Framework for Predictive Models
   - Bayes' Rule and the Prediction Pyramid
   - Model Plausibility and Model Selection
3. OPAL: The Occam-Plausibility ALgorithm
   - The Algorithm
   - Example: Polyethylene

Motivation: What is coarse-graining and why do it?



Coarse-graining is the process of aggregating atoms into larger representative particles

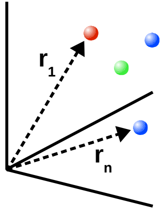
This reduces the number of degrees of freedom needed for simulating the molecular system

Consequential questions:

- What is the model?
- Is the model valid for predicting specific QoIs?
- How do we cope with and quantify uncertainties?

The AA and CG Systems

AA System



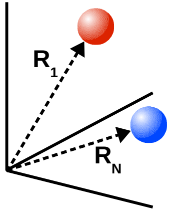
$\mathbf{r}^n$ = configuration, $\mathbf{p}^n$ = momenta

$$H(\mathbf{r}^n, \mathbf{p}^n) = K(\mathbf{p}^n) + V(\mathbf{r}^n) = \sum_{i=1}^{n} \frac{\mathbf{p}_i \cdot \mathbf{p}_i}{2m_i} + V(\mathbf{r}^n)$$

If q is a phase function describing the quantity of interest,

$$Q = \langle q \rangle \qquad or \qquad Q = \pi(q)$$

CG System



$\mathbf{R}^N$ = configuration, $\mathbf{P}^N$ = momenta

$$H_{CG}(\mathbf{R}^N, \mathbf{P}^N) = K(\mathbf{P}^N) + V_{CG}(\mathbf{R}^N) = \sum_{i=1}^{N} \frac{\mathbf{P}_i \cdot \mathbf{P}_i}{2M_i} + V_{CG}(\mathbf{R}^N)$$

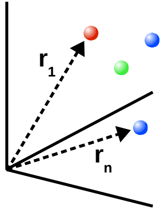
If q_{CG} is the corresponding phase function,

$$Q_{CG} = \langle q_{CG} \rangle \qquad or \qquad Q_{CG} = \pi(q_{CG})$$

Determining V_{CG} so that the CG model adequately represents, and therefore may be used as a surrogate model for, the all-atom model is a main goal of this work

The AA and CG Systems

AA System



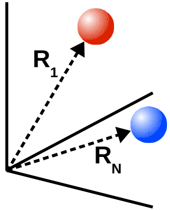
$\mathbf{r}^n$ = configuration, $\mathbf{p}^n$ = momenta

$$H(\mathbf{r}^n, \mathbf{p}^n) = K(\mathbf{p}^n) + V(\mathbf{r}^n) = \sum_{i=1}^{n} \frac{\mathbf{p}_i \cdot \mathbf{p}_i}{2m_i} + V(\mathbf{r}^n)$$

If q is a phase function describing the quantity of interest,

$$Q = \langle q \rangle \qquad or \qquad Q = \pi(q)$$

CG System



$\mathbf{R}^N$ = configuration, $\mathbf{P}^N$ = momenta

$$H_{CG}(\mathbf{R}^N, \mathbf{P}^N) = K(\mathbf{P}^N) + V_{CG}(\mathbf{R}^N) = \sum_{i=1}^{N} \frac{\mathbf{P}_i \cdot \mathbf{P}_i}{2M_i} + V_{CG}(\mathbf{R}^N)$$

If q_{CG} is the corresponding phase function,

$$Q_{CG} = \langle q_{CG} \rangle \qquad or \qquad Q_{CG} = \pi(q_{CG})$$

Determining V_{CG} so that the CG model adequately represents, and therefore may be used as a surrogate model for, the all-atom model is a main goal of this work

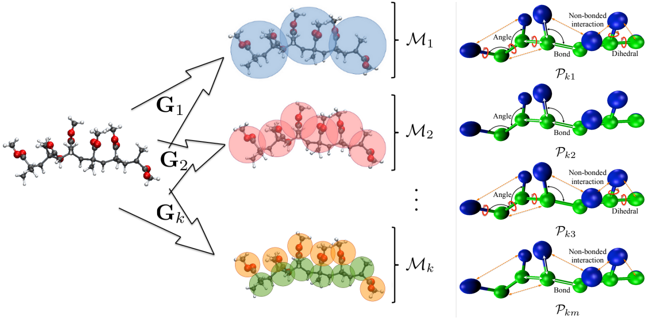
The AA-to-CG Map

The QoI $Q_{CG} = Q_{CG}(G, \boldsymbol{\theta})$ depends on the definition of V_{CG}, with parameters $\boldsymbol{\theta}$, which depends on the AA-to-CG map G

Choice of G may be influenced by

- Chemical intuition
- Computational limitations

For each G_i, the set of possible representations of V_{CG} is denoted $\mathcal{M}_i$



Uncertainties in the CG Model

- The choice of G is not well-defined
- Once G is specified, V_{CG} must be determined. Each G_i yields a set of possible model classes representing V_{CG},

$$\mathcal{M}_i = \{\mathcal{P}_{i1}(\boldsymbol{\theta}_{i1}), \mathcal{P}_{i2}(\boldsymbol{\theta}_{i2}), \ldots, \mathcal{P}_{im}(\boldsymbol{\theta}_{im})\}, \quad i = 1, 2, \ldots, k$$

- The parameters $\boldsymbol{\theta}_{ij}$ for each model $\mathcal{P}_{ij}$ are unknown and are uncertain, random vectors

Bayes' Rule

Cox's Theorem $\Rightarrow$ Every natural extension of Aristotelian logic with uncertainties is Bayesian

Bayes' Rule

$$\underbrace{\pi(\boldsymbol{\theta}|\mathbf{y})}_{\text{Posterior}} = \frac{\overbrace{\pi(\mathbf{y}|\boldsymbol{\theta})}^{\text{Likelihood}} \times \overbrace{\pi(\boldsymbol{\theta})}^{\text{Prior}}}{\underbrace{\pi(\mathbf{y})}_{\text{Evidence}}}$$

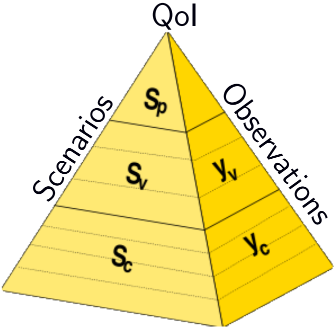
$\pi(\boldsymbol{\theta})$ captures any information that is known about the parameters before calibration

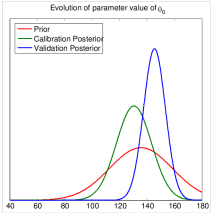
$\pi(\mathbf{y}|\boldsymbol{\theta})$ is the probability of seeing the data $\mathbf{y}$ given parameters $\boldsymbol{\theta}$

$\pi(\mathbf{y})$ measures the evidence of the model

Cox (1946), van Horn (2003), Jaynes (2003)

Bayes' Rule and the Prediction Pyramid





Prior

$$\pi(\boldsymbol{\theta})$$

Calibration $(S_c, \mathbf{y}_c)$

$$\pi(\boldsymbol{\theta}|\mathbf{y}_c) = \frac{\pi(\mathbf{y}_c|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\mathbf{y}_c)}$$

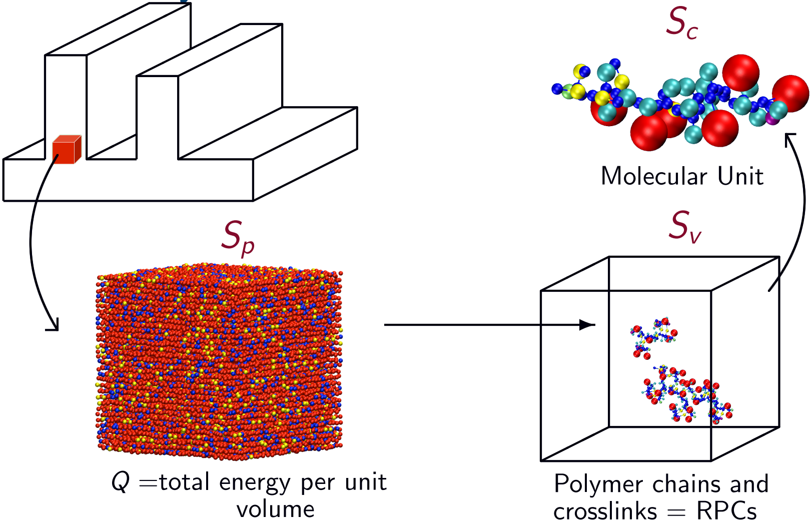
Validation $(S_v, \mathbf{y}_v)$

$$\pi(\boldsymbol{\theta}|\mathbf{y}_v, \mathbf{y}_c) = \frac{\pi(\mathbf{y}_v|\boldsymbol{\theta}, \mathbf{y}_c)\pi(\boldsymbol{\theta}|\mathbf{y}_c)}{\pi(\mathbf{y}_v|\mathbf{y}_c)}$$

Prediction $(S_p,$ QoI$)$

$$\pi(Q) = \pi(Q|\boldsymbol{\theta}, S_v, S_c)$$

The Prediction Process: Traveling up the Prediction Pyramid



Bayesian Model Validation

If the observable value is also a pdf, $\pi(q)$, validity is determined by

$$D_{KL}(\pi(q)\|\pi(Q|\boldsymbol{\theta})) < \gamma_{tol,1}$$

where

$$D_{KL}\left(\pi\left(q\right)\|\pi\left(Q|\boldsymbol{\theta}\right)\right) = \int \pi\left(q\left(\omega\right)\right)\log\frac{\pi\left(q\left(\omega\right)\right)}{\pi\left(Q\left(\omega\right)|\boldsymbol{\theta}\right)}d\omega.$$

If the observable value is a scalar, q, validity is determined by

$$|q - Q| < \gamma_{tol,2}$$

where

$$Q = \mathbb{E}_{\pi_v}\left[\pi(Q|\boldsymbol{\theta})\right] = \int_{\Theta}\pi(Q|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{y}_v, \mathbf{y}_c)\, d\boldsymbol{\theta}$$

A sequence of validation scenarios may be considered

Model Plausibility and Model Selection

$\mathcal{M}$ = set of parametric model classes = $\{\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_m\}$

Each $\mathcal{P}$ has its own likelihood and parameters $\boldsymbol{\theta}_j$

Bayes' rule in expanded form:

$$\pi(\boldsymbol{\theta}_j|\mathbf{y}, \mathcal{P}_j, \mathcal{M}) = \frac{\pi(\mathbf{y}|\boldsymbol{\theta}_j, \mathcal{P}_j, \mathcal{M})\pi(\boldsymbol{\theta}_j|\mathcal{P}_j, \mathcal{M})}{\pi(\mathbf{y}|\mathcal{P}_j, \mathcal{M})}, \quad 1 \leq j \leq m$$

where

$$\pi(\mathbf{y}|\mathcal{P}_j, \mathcal{M}) = \int \pi(\mathbf{y}|\boldsymbol{\theta}_j, \mathcal{P}_j, \mathcal{M})\pi(\boldsymbol{\theta}_j|\mathcal{P}_j, \mathcal{M})\, d\boldsymbol{\theta}_j$$

Now apply Bayes' Rule to the evidence:

$$\begin{aligned} \rho_j &= \pi(\mathcal{P}_j|\mathbf{y}, \mathcal{M}) = \frac{\pi(\mathbf{y}|\mathcal{P}_j, \mathcal{M})\pi(\mathcal{P}_j|\mathcal{M})}{\pi(\mathbf{y}|\mathcal{M})} \\ &= \text{model plausibility} \end{aligned}$$

$$\sum_{j=1}^{m} \rho_j = 1$$

Sensitivity Analysis

S_{T_i} measures the total contribution from parameter θ_i to the variance in the output Y and indicates the importance of θ_i

$Y(\boldsymbol{\theta})$ = model output
$V(Y)$ = variance in Y
$V(Y|\boldsymbol{\theta}_{\sim i})$ = variance in Y when all parameters except θ_i are fixed

Total Sensitivity Index

$$S_{T_i} = \frac{\mathbb{E}\left[V(Y|\boldsymbol{\theta}_{\sim i})\right]}{V(Y)}$$

S_{T_i} indicates which parameters are informed by the observables in the calibration and validation scenarios

A. Saltelli, *et al* (2008)

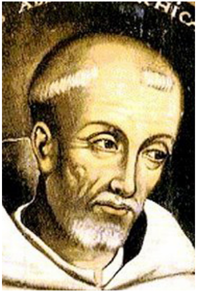
Occam's Razor

Occam's Razor

Among competing theories that lead to the same prediction, the one that relies on the fewest assumptions is the best.

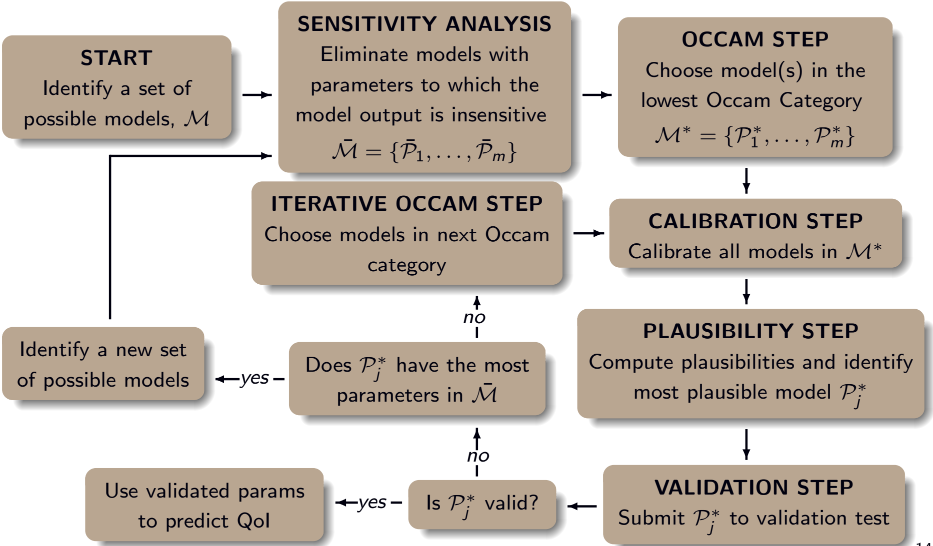
When choosing among a set of models:
The simplest valid model is the best choice.

- simple $\Rightarrow$ number of parameters
- valid $\Rightarrow$ passes Bayesian validation test

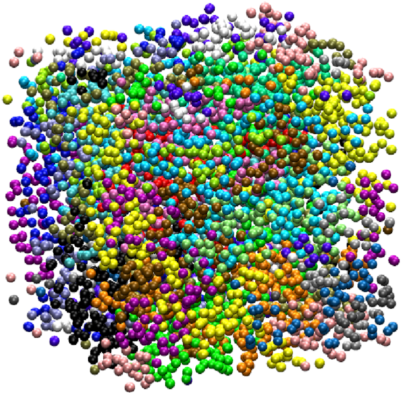


How do we choose a model that adheres to this principle?

The Occam-Plausibility Algorithm

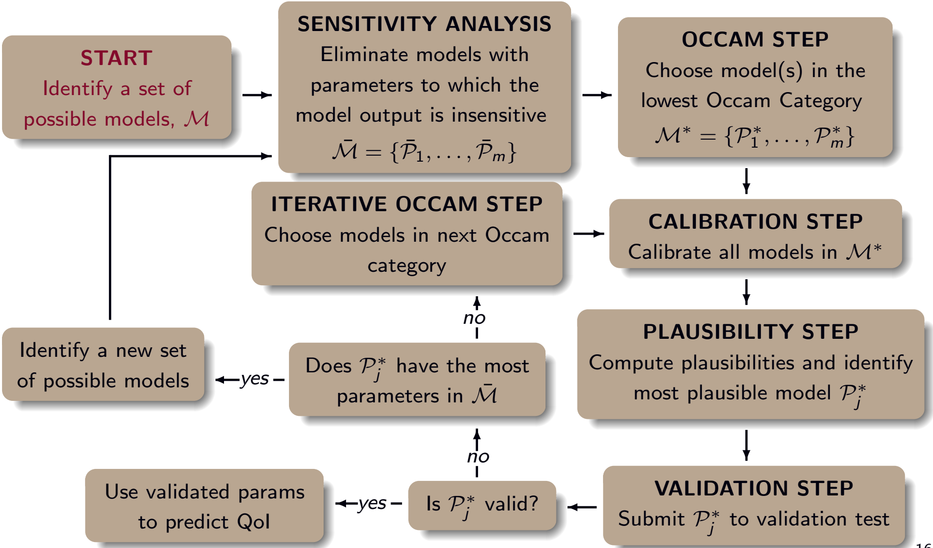


Example Application: Polyethylene



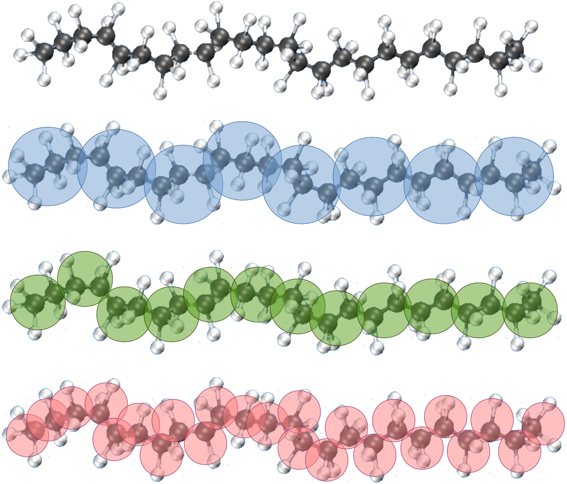
Prediction scenario contains 25 chains of $C_{80}H_{162}$ and the QoI is the potential energy of the system

The Occam-Plausibility Algorithm



OPAL Step 1: Initialization

What is the CG map?



OPAL Step 1: Initialization

How is the potential energy represented?

Assume the OPLS functional form,

$$
\begin{aligned}
V(\mathbf{R}^n) \quad = \quad & \sum_{bonds} K_R(R - R_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 \\
& + \sum_{dihedrals} \sum_{n=1}^{4} \frac{V_n}{2}[1 + (-1)^{n-1} \cos(n\psi)] \\
& + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] f
\end{aligned}
$$

Alternatively, we may use

$$
\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{9} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] f
$$

OPAL Step 1: Initialization

How is the potential energy represented?

Assume the OPLS functional form,

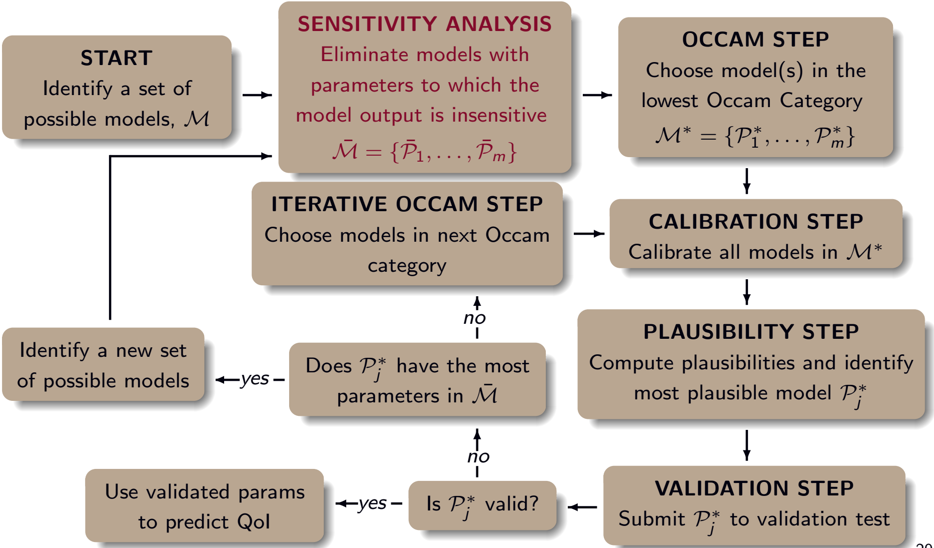
$$
\begin{aligned}
V(\mathbf{R}^n) \;=\; & \sum_{bonds} K_R(R - R_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 \\
& + \sum_{dihedrals} \sum_{n=1}^{4} \frac{V_n}{2}\left[1 + (-1)^{n-1}\cos(n\psi)\right] \\
& + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right] f
\end{aligned}
$$

Alternatively, we may use

$$
\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{9} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right] f
$$

Model	Bonds	Angles	Dihedrals	LJ 12-6	LJ 9-6	Params
$\mathcal{P}_1$	✓					3
$\mathcal{P}_2$		✓				3
$\mathcal{P}_3$				✓		3
$\mathcal{P}_4$					✓	3
$\mathcal{P}_5$	✓	✓				5
$\mathcal{P}_6$	✓			✓		5
$\mathcal{P}_7$	✓				✓	5
$\mathcal{P}_8$		✓		✓		5
$\mathcal{P}_9$		✓			✓	5
$\mathcal{P}_{10}$			✓			7
$\mathcal{P}_{11}$	✓	✓		✓		7
$\mathcal{P}_{12}$	✓	✓			✓	7
$\mathcal{P}_{13}$	✓		✓			7
$\mathcal{P}_{14}$		✓	✓			7
$\mathcal{P}_{15}$			✓	✓		7
$\mathcal{P}_{16}$			✓		✓	7
$\mathcal{P}_{17}$	✓	✓	✓			9
$\mathcal{P}_{18}$	✓		✓	✓		9
$\mathcal{P}_{19}$	✓		✓		✓	9
$\mathcal{P}_{20}$		✓	✓	✓		9
$\mathcal{P}_{21}$		✓	✓		✓	9
$\mathcal{P}_{22}$	✓	✓	✓	✓		11
$\mathcal{P}_{23}$	✓	✓	✓		✓	11

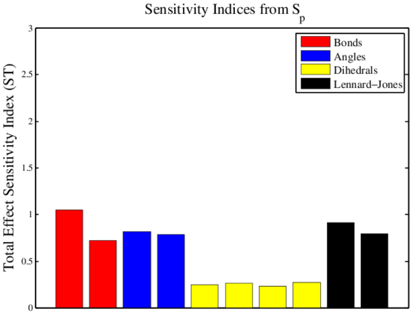
The Occam-Plausibility Algorithm



OPAL Step 2: Sensitivity Analysis

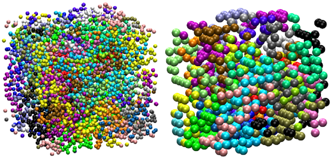
Sensitivity analysis is performed in the prediction scenario with the QoI as the observable

Output $Y = \langle V_{CG}(\boldsymbol{\theta}) \rangle$

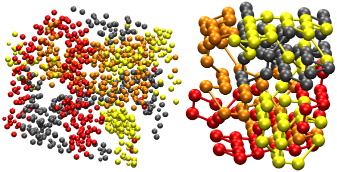


Scenarios

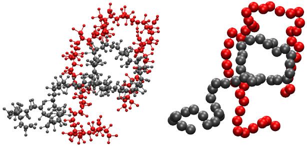
Prediction Scenario



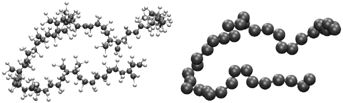
Validation Scenario # 2



Validation Scenario # 1

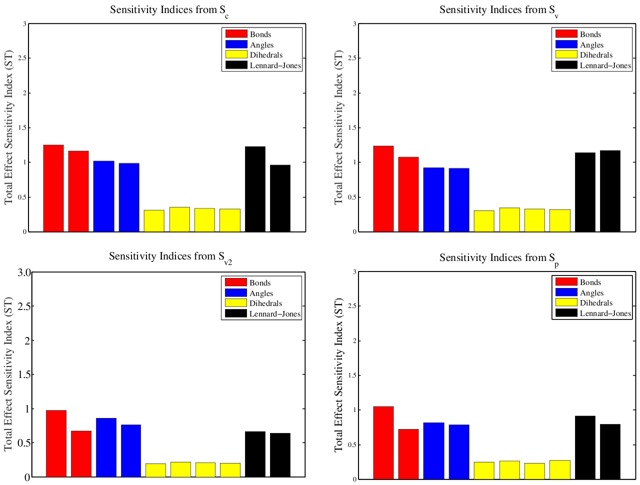


Calibration Scenario



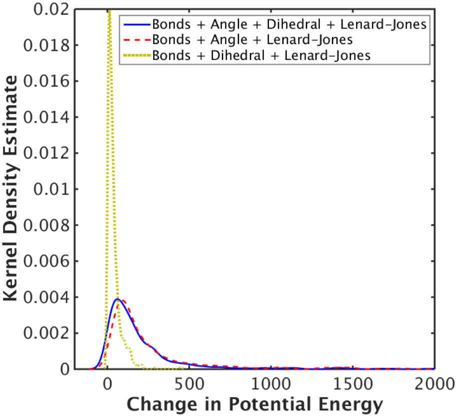
OPAL Step 2: Sensitivity Analysis

Output: $Y = \langle V_{CG}(\boldsymbol{\theta}) \rangle$



OPAL Step 2: Sensitivity Analysis

The sensitivity indices show that the dihedral parameters are unimportant, but how important are they?



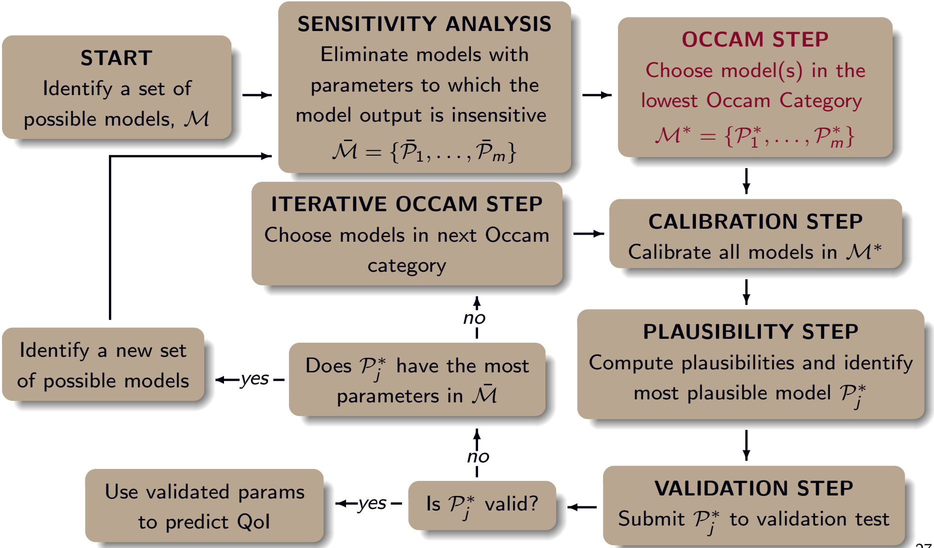
Model	Bonds	Angles	Dihedrals	LJ 12-6	LJ 9-6	Params
$\mathcal{P}_1$	✓					3
$\mathcal{P}_2$		✓				3
$\mathcal{P}_3$				✓		3
$\mathcal{P}_4$					✓	3
$\mathcal{P}_5$	✓	✓				5
$\mathcal{P}_6$	✓			✓		5
$\mathcal{P}_7$	✓				✓	5
$\mathcal{P}_8$		✓		✓		5
$\mathcal{P}_9$		✓			✓	5
$\mathcal{P}_{10}$			✓			7
$\mathcal{P}_{11}$	✓	✓		✓		7
$\mathcal{P}_{12}$	✓	✓			✓	7
$\mathcal{P}_{13}$	✓		✓			7
$\mathcal{P}_{14}$		✓	✓			7
$\mathcal{P}_{15}$			✓	✓		7
$\mathcal{P}_{16}$			✓		✓	7
$\mathcal{P}_{17}$	✓	✓	✓			9
$\mathcal{P}_{18}$	✓		✓	✓		9
$\mathcal{P}_{19}$	✓		✓		✓	9
$\mathcal{P}_{20}$		✓	✓	✓		9
$\mathcal{P}_{21}$		✓	✓		✓	9
$\mathcal{P}_{22}$	✓	✓	✓	✓		11
$\mathcal{P}_{23}$	✓	✓	✓		✓	11

OPAL Step 2: Sensitivity Analysis

Models with insensitive parameters may be eliminated, yielding a new set of possible models

Model	Bonds	Angles	Dihedrals	LJ 12-6	LJ 9-6	Params	Category
$\bar{\mathcal{P}}_1$	✓					3	1
$\bar{\mathcal{P}}_2$		✓				3	
$\bar{\mathcal{P}}_3$				✓		3	
$\bar{\mathcal{P}}_4$					✓	3	
$\bar{\mathcal{P}}_5$	✓	✓				5	2
$\bar{\mathcal{P}}_6$	✓			✓		5	
$\bar{\mathcal{P}}_7$	✓				✓	5	
$\bar{\mathcal{P}}_8$		✓		✓		5	
$\bar{\mathcal{P}}_9$		✓			✓	5	
$\bar{\mathcal{P}}_{10}$	✓	✓		✓		7	3
$\bar{\mathcal{P}}_{11}$	✓			✓	✓	7	

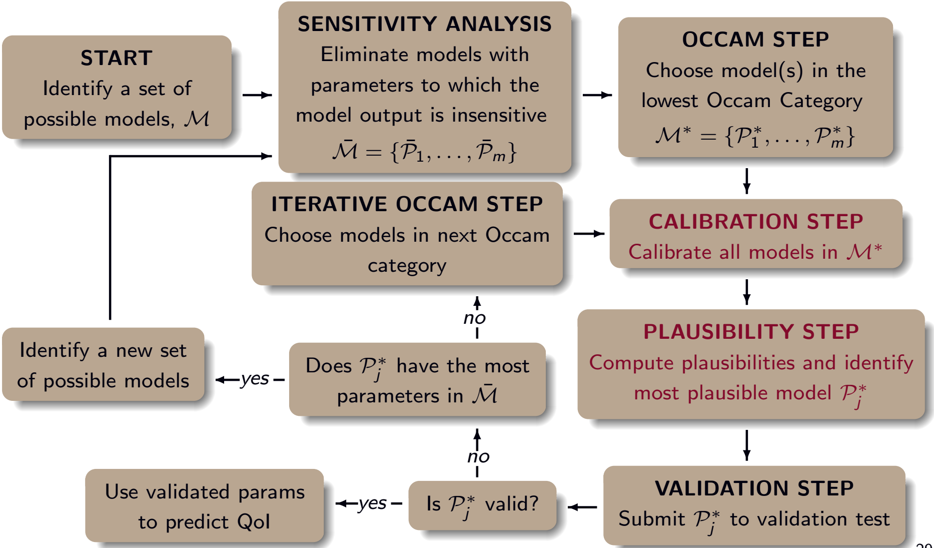
The Occam-Plausibility Algorithm



OPAL Step 3: Occam Step

Model	Bonds	Angles	Dihedrals	LJ 12-6	LJ 9-6	Params	Category
$\bar{\mathcal{P}}_1$	✓					3	1
$\bar{\mathcal{P}}_2$		✓				3	
$\bar{\mathcal{P}}_3$				✓		3	
$\bar{\mathcal{P}}_4$					✓	3	
$\bar{\mathcal{P}}_5$	✓	✓				5	2
$\bar{\mathcal{P}}_6$	✓			✓		5	
$\bar{\mathcal{P}}_7$	✓				✓	5	
$\bar{\mathcal{P}}_8$		✓		✓		5	
$\bar{\mathcal{P}}_9$		✓			✓	5	
$\bar{\mathcal{P}}_{10}$	✓	✓		✓		7	3
$\bar{\mathcal{P}}_{11}$	✓			✓	✓	7	

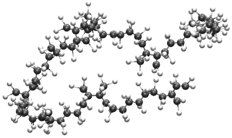
The Occam-Plausibility Algorithm



OPAL Step 4: Calibration

Calibration

$$\pi(\boldsymbol{\theta}_j^* | \mathbf{y}, \mathcal{P}_j^*, \mathcal{M}^*) = \frac{\pi(\mathbf{y} | \boldsymbol{\theta}_j^*, \mathcal{P}_j^*, \mathcal{M}^*) \pi(\boldsymbol{\theta}_j^* | \mathcal{P}_j^*, \mathcal{M}^*)}{\pi(\mathbf{y} | \mathcal{P}_j^*, \mathcal{M}^*)}$$



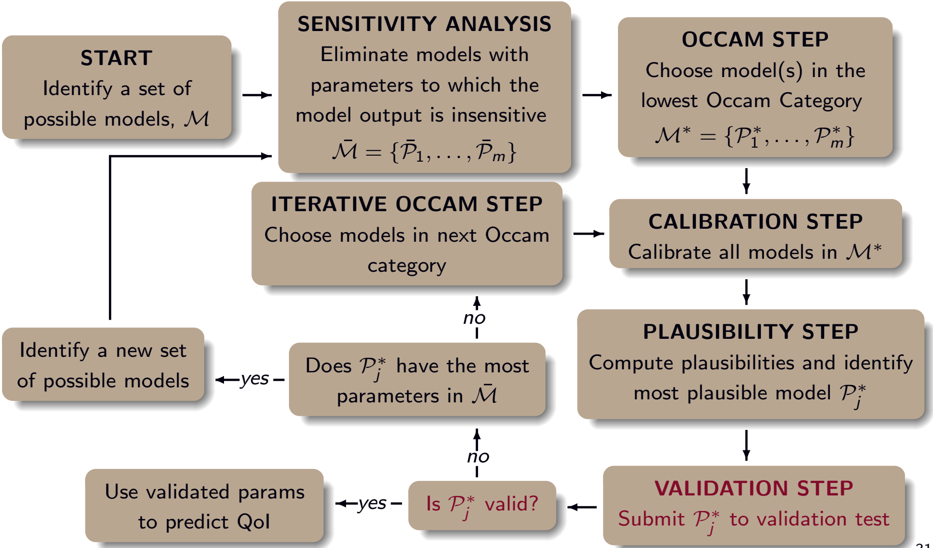
Here, $\mathbf{y}$ = potential energy of $C_{80}H_{162}$

Plausibility

$$\rho_j^* = \pi(\mathcal{P}_j^* | \mathbf{y}, \mathcal{M}^*) = \frac{\pi(\mathbf{y} | \mathcal{P}_j^*, \mathcal{M}^*) \pi(\mathcal{P}_j^* | \mathcal{M}^*)}{\pi(\mathbf{y} | \mathcal{M}^*)}$$

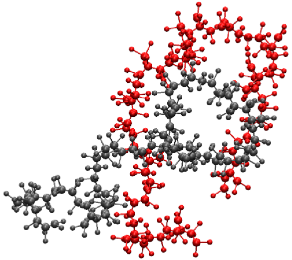
Model	Bonds	Angles	Dihedrals	LJ 12-6	LJ 9-6	Params	Plausibility
$\mathcal{P}_1^*$	✓					3	1
$\mathcal{P}_2^*$		✓				3	0
$\mathcal{P}_3^*$				✓		3	0
$\mathcal{P}_4^*$					✓	3	0

The Occam-Plausibility Algorithm



OPAL Step 5: Validation

As a validation scenario, we consider 2 chains at T = 300K in a canonical ensemble.



Validation

$$\pi(\boldsymbol{\theta}_1^*|\mathbf{y}_v, \mathbf{y}_c) = \frac{\pi(\mathbf{y}_v|\boldsymbol{\theta}_1^*, \mathbf{y}_c)\pi(\boldsymbol{\theta}_1^*|\mathbf{y}_c)}{\pi(\mathbf{y}_v)}$$

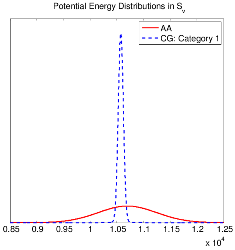
Here, $\mathbf{y}_v$ is the potential energy

How well does this updated model reproduce the desired observable?

Let

- $\pi(Q) = \pi(V_{AA}) \quad \Rightarrow \quad \pi(V_{CG}|\boldsymbol{\theta}^*) = \pi(V_{CG}(\boldsymbol{\theta}^*))$
  $\gamma_{tol,1} = 0.15\sigma_{AA}^2 \mathcal{O}\left(\mathbb{E}\left[\pi(V_{AA})\right]\right)$
- $Q = \langle V_{AA} \rangle \quad \Rightarrow \quad \mathbb{E}\left[\pi(V_{CG}|\boldsymbol{\theta}^*)\right] = \langle V_{CG}(\boldsymbol{\theta}^*) \rangle$
  $\gamma_{tol,2} = 0.1Q$

OPAL Step 5: Validation

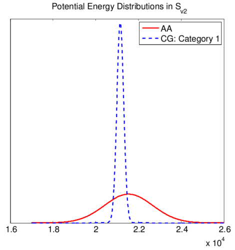


If we compare the distributions,

$$D_{KL}(\pi(V_{AA})\|\pi(V_{CG}|\boldsymbol{\theta}^*)) = 0.0622\sigma_{AA}^2\mathcal{O}\left(\mathbb{E}\left[\pi(V_{AA})\right]\right) < \gamma_{1,tol}$$

If we compare the ensemble average,

$$|\langle V_{AA}\rangle - \langle V_{CG}(\boldsymbol{\theta}^*)\rangle| = 0.0118\,\langle V_{AA}\rangle < \gamma_{2,tol}$$



If we compare the distributions,

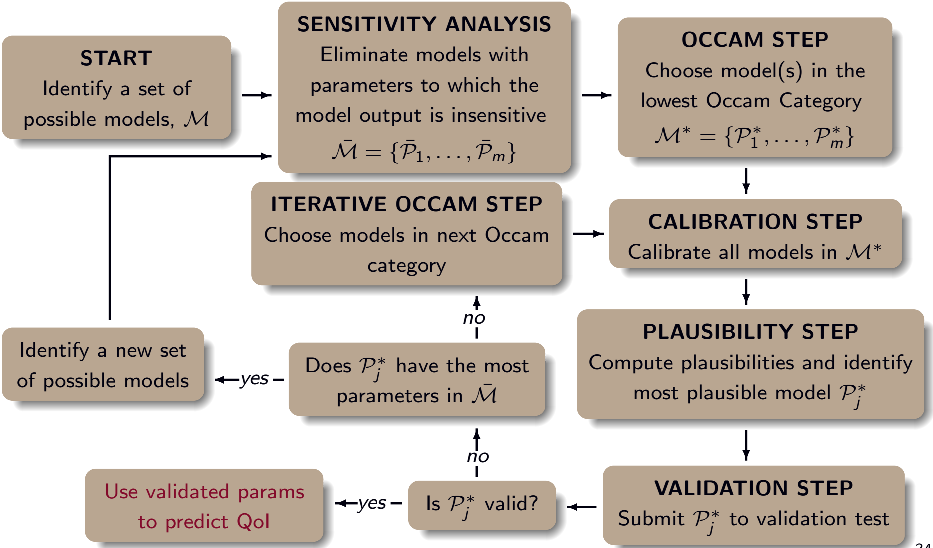
$$D_{KL}(\pi(V_{AA})\|\pi(V_{CG}|\boldsymbol{\theta}^*)) = 0.0826\sigma_{AA}^2\mathcal{O}\left(\mathbb{E}\left[\pi(V_{AA})\right]\right) < \gamma_{1,tol}$$

If we compare the ensemble average,

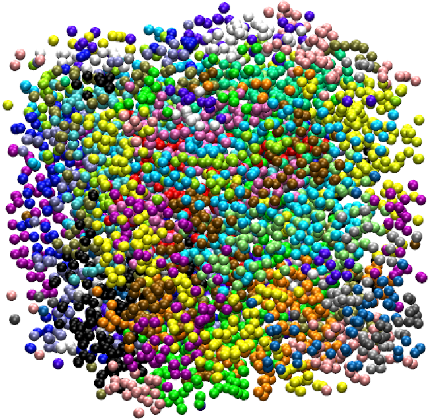
$$|\langle V_{AA}\rangle - \langle V_{CG}(\boldsymbol{\theta}^*)\rangle| = 0.0181\,\langle V_{AA}\rangle < \gamma_{2,tol}$$

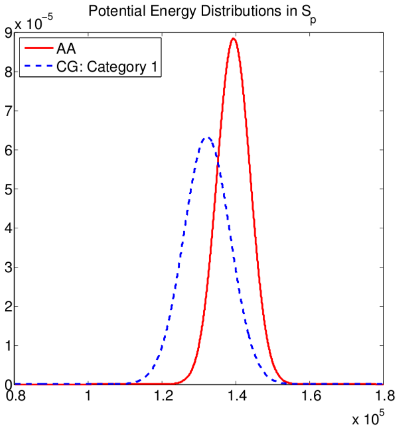
Model is NOT invalid

The Occam-Plausibility Algorithm



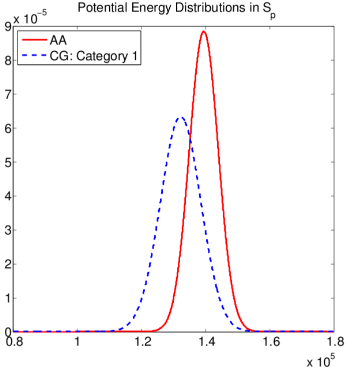
Prediction





Potential Energy Distributions in S_p
x 10^{-5}
AA
CG: Category 1
0
1
2
3
4
5
6
7
8
9
0.8
1
1.2
1.4
1.6
1.8
x 10^5

Prediction

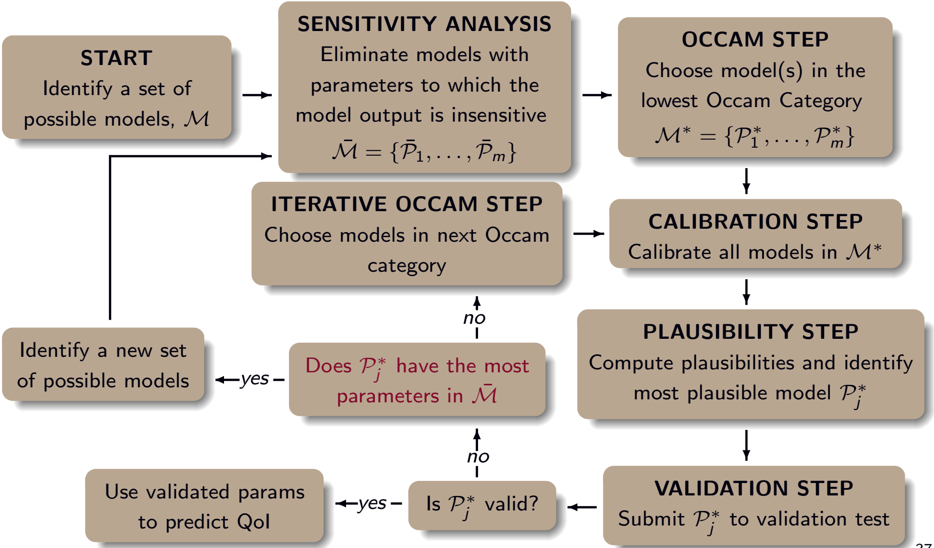


What if we set

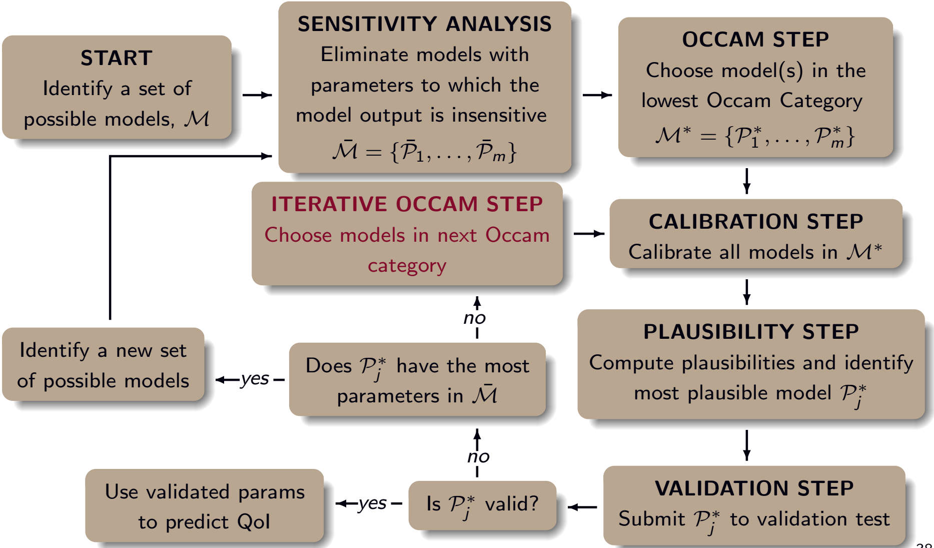
$$\gamma_{1,tol} = 0.06\sigma_{AA}^2\mathcal{O}\left(\mathbb{E}\left[\pi(V_{AA})\right]\right)?$$

Then this model is also *invalid*

The Occam-Plausibility Algorithm



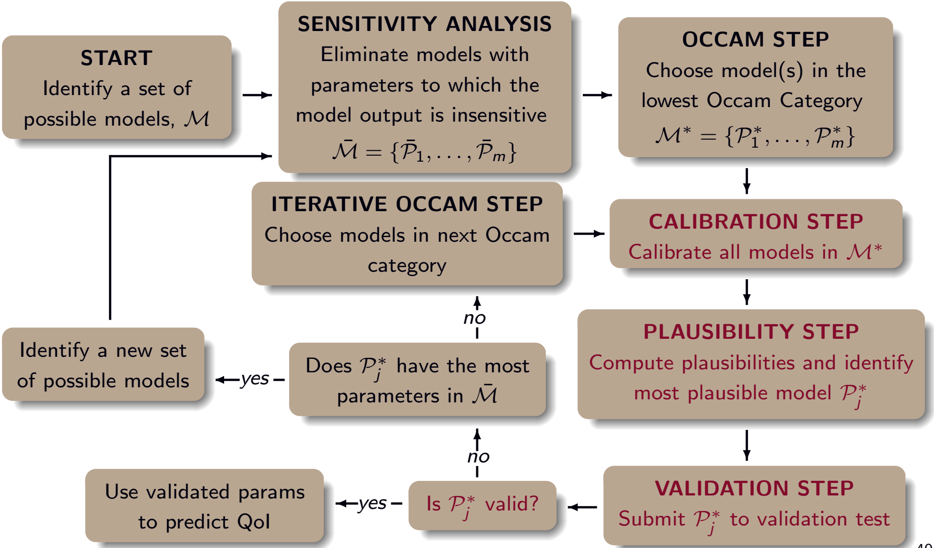
The Occam-Plausibility Algorithm



OPAL Step 3: Occam Step

Model	Bonds	Angles	Dihedrals	LJ 12-6	LJ 9-6	Params	Category
$\bar{\mathcal{P}}_1$	✓					3	1
$\bar{\mathcal{P}}_2$		✓				3	
$\bar{\mathcal{P}}_3$				✓		3	
$\bar{\mathcal{P}}_4$					✓	3	
$\bar{\mathcal{P}}_5$	✓	✓				5	2
$\bar{\mathcal{P}}_6$	✓			✓		5	
$\bar{\mathcal{P}}_7$	✓				✓	5	
$\bar{\mathcal{P}}_8$		✓		✓		5	
$\bar{\mathcal{P}}_9$		✓			✓	5	
$\bar{\mathcal{P}}_{10}$	✓	✓		✓		7	3
$\bar{\mathcal{P}}_{11}$	✓			✓	✓	7	

The Occam-Plausibility Algorithm



OPAL Steps 4 & 5: Calibration, Plausibility, and Validation

Sandia National Laboratories

Model	Bonds	Angles	Dihedrals	LJ 12-6	LJ 9-6	Params	Plausibility
$\mathcal{P}_1^*$	✓	✓				5	1
$\mathcal{P}_2^*$	✓			✓		5	0
$\mathcal{P}_3^*$	✓				✓	5	0
$\mathcal{P}_4^*$		✓		✓		5	0
$\mathcal{P}_5^*$		✓			✓	5	0

How well does this updated model reproduce the desired observable?

Let

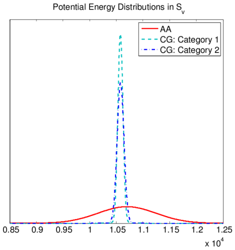
- $\pi(Q) = \pi(V_{AA}) \quad \Rightarrow \quad \pi(V_{CG}|\boldsymbol{\theta}^*) = \pi(V_{CG}(\boldsymbol{\theta}^*))$

  $$\gamma_{tol,1} = 0.06\sigma_{AA}^2 \mathcal{O}\left(\mathbb{E}\left[\pi(V_{AA})\right]\right)$$

- $Q = \langle V_{AA} \rangle \quad \Rightarrow \quad \mathbb{E}\left[\pi(V_{CG}|\boldsymbol{\theta}^*)\right] = \langle V_{CG}(\boldsymbol{\theta}^*) \rangle$

  $$\gamma_{tol,2} = 0.1Q$$

OPAL Step 5: Validation

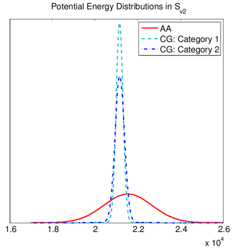


If we compare the distributions,

$$D_{KL}(\pi(V_{AA})\|\pi(V_{CG}|\boldsymbol{\theta}^*)) = 0.044\sigma_{AA}^2\mathcal{O}\left(\mathbb{E}\left[\pi(V_{AA})\right]\right) < \gamma_{1,tol}$$

If we compare the ensemble average,

$$|\langle V_{AA}\rangle - \langle V_{CG}(\boldsymbol{\theta}^*)\rangle| = 0.0115\,\langle V_{AA}\rangle < \gamma_{2,tol}$$



If we compare the distributions,

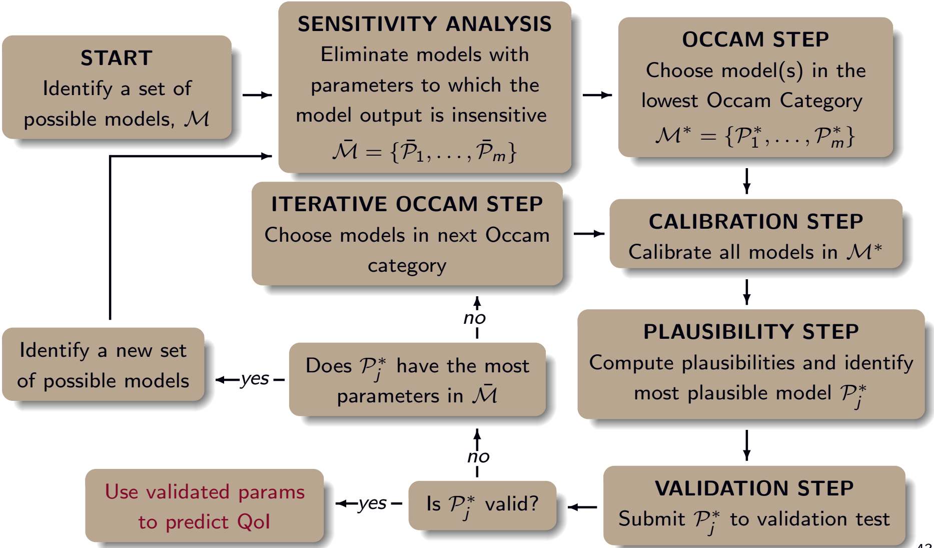
$$D_{KL}(\pi(V_{AA})\|\pi(V_{CG}|\boldsymbol{\theta}^*)) = 0.0587\sigma_{AA}^2\mathcal{O}\left(\mathbb{E}\left[\pi(V_{AA})\right]\right) < \gamma_{1,tol}$$

If we compare the ensemble average,

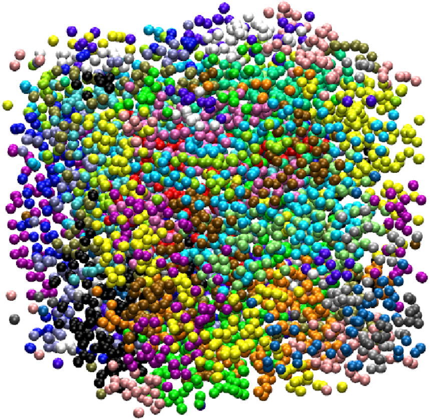
$$|\langle V_{AA}\rangle - \langle V_{CG}(\boldsymbol{\theta}^*)\rangle| = 0.0178\,\langle V_{AA}\rangle < \gamma_{2,tol}$$

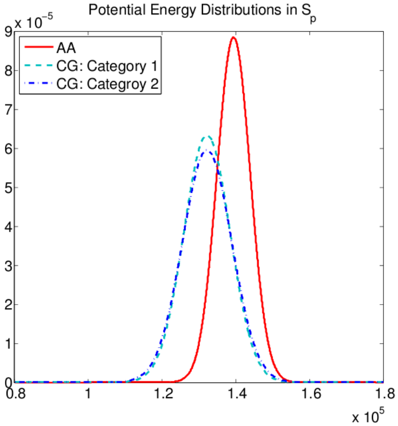
Model is NOT invalid

The Occam-Plausibility Algorithm



Prediction




Potential Energy Distributions in S_p
x 10^{-5}
AA
CG: Category 1
CG: Categroy 2
0
1
2
3
4
5
6
7
8
9
0.8
1
1.2
1.4
1.6
1.8
x 10^5

Deterministic Model Selection

In the deterministic setting, we seek $\mathcal{P}_i$ such that

$$D_{KL}(g\|\pi(\mathbf{y}|\boldsymbol{\theta}_i^{\dagger}, \mathcal{P}_i, \mathcal{M})) < D_{KL}(g\|\pi(\mathbf{y}|\boldsymbol{\theta}_j^{\dagger}, \mathcal{P}_j, \mathcal{M})) \quad \forall \mathcal{P}_j \in \mathcal{M}$$

However, the true value of $D_{KL}(\cdot\|\cdot)$ is usually impossible to compute, but can be approximated by the Akaike Information Criterion,

$$AIC_i = -2 \log \pi(\mathbf{y}|\boldsymbol{\theta}_i, \mathcal{P}_i, \mathcal{M}) + 2K_i$$

where

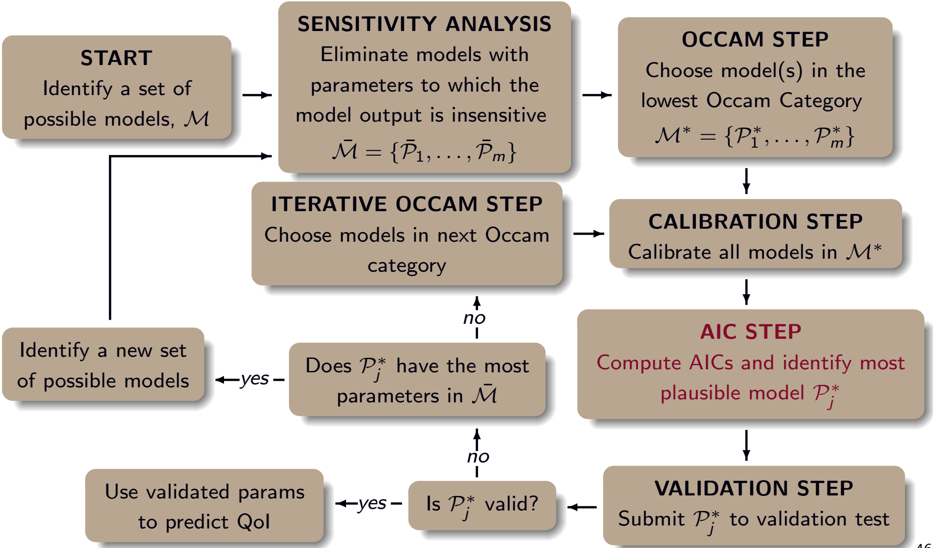
K_i = number of parameters in $\mathcal{P}_i$

$\boldsymbol{\theta}_i^{\dagger}$ = MLE

This can be used as a model selection criterion in OPAL

Akaike (1973, 1974)

The Occam-Plausibility Algorithm



Conclusions

- OPAL adaptively selects the simplest valid model in the presence of uncertainties by combining the notions of Occam's Razor and Bayesian calibration, validation, and selection
- Sensitivity analysis can be used to determine appropriate scenario/observable pairs for calibration and validation
- OPAL can also be paired with other methods of model selection and calibration/validation

Thank you!

UT Funding Through:
DOE-DE-SC009286 MMICC
AFOSR: FA9550-12-1-0484
DOE Eureka: DE-SC0010518