

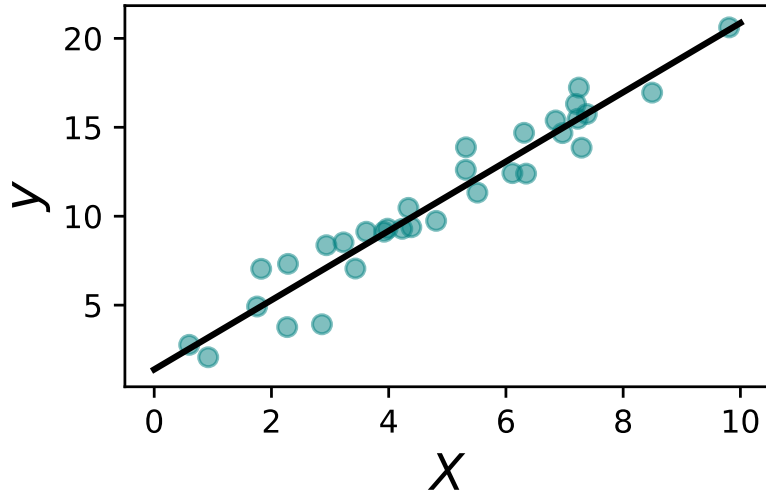
BAYESIAN METHODS FOR MATERIALS MODELING

Logan Ward
Asst. Computational Scientist
Argonne National Laboratory

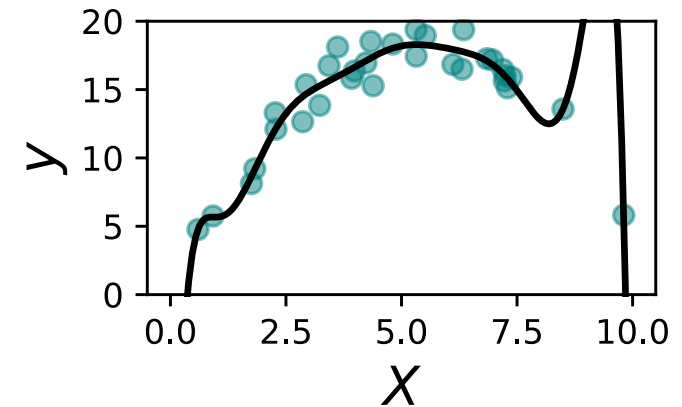
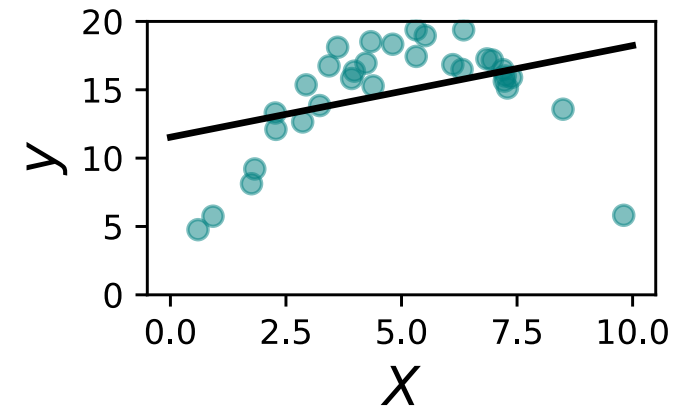
7 February 2022

Why do we need Bayesian modeling?

How certain should I be in my parameters?



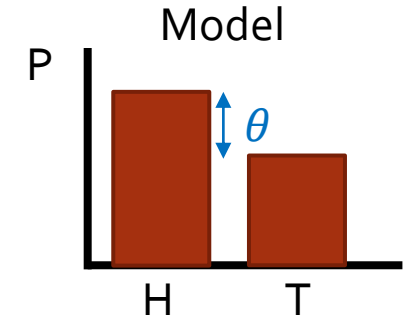
Which model should I choose?



Bayes Rule: Infer parameters using evidence

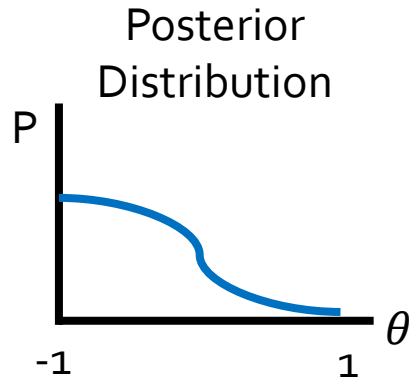
“Posterior”: Probability of parameters (θ) given observations (D) and model (M)

“Likelihood”: Probability of observations given parameters and model

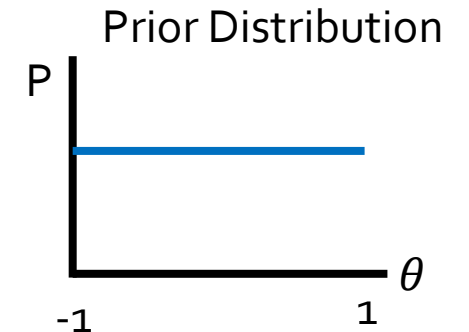


$$P(\theta|D, M) = \frac{P(D|\theta, M)P(\theta|M)}{P(D|M)}$$

“Prior”: Initial assumption of parameters given model



“Evidence”: Likelihood of observations given model



Using Bayes' rule: Integrals over the posterior

You can use integration to solve for probabilistic measures of the integral

Math: $E(h(\theta)|D) = \int h(\theta)p(\theta|D)d\theta$

English: Expected value of $h(\theta)$ by average over all possible θ weighted by the probability of θ

Solving that integral analytically is often impossible. Can always solve numerically:

Math: $\int h(\theta)p(\theta|D)d\theta \approx \frac{1}{S} \sum_{s=1}^S h(\theta_s)$

English: Compute the average of $h(\theta)$ for many samples of θ taken from the posterior distribution

Problem: How do you generate samples from the posterior!?

Markov-Chain Monte Carlo: Sampling S with balance

Algorithm:

1. Select initial point, θ_0 , from prior distributions

2. Generate new point: $\theta_i = \mathcal{N}(\theta_{i-1}, \sigma)$ ← Pick a new point from Normal Distribution

3. Compute posterior probability for θ_i and θ_{i-1}

$$P(\theta_n|D, M) = P(D|\theta_n, M)P(\theta_n|M) \leftarrow \text{See practical exercises}$$

4. Compute “acceptance probability”

$$P(\text{accept}) = \min \left(1, \frac{P(\theta_i|D, M)}{P(\theta_{i-1}|D, M)} \right) \leftarrow \begin{array}{l} \text{This assures samples will follow} \\ \text{posterior distribution} \\ \text{“detailed balance”} \end{array}$$

5. With probability $P(\text{accept})$, keep θ_i as the current point

6. Repeat from 2

Run many iterations of “MCMC” and you will get samples that follow posterior

There are better methods for doing this

Kombine ([Farr and Farr, 2015](#))

KOMBINE

Table Of Contents

- [kombine](#)
- [Example Usage](#)
- [User Guide](#)
- [API Documentation](#)
- [Contributors](#)

This Page

[Show Source](#)

Quick search

kombine

kombine is an ensemble sampler built for efficiently exploring multimodal distributions. By using estimates of ensemble's instantaneous distribution as a proposal, it achieves very fast burnin, followed by sampling with very short autocorrelation times.

Example Usage

Construct an 8-D bimodal target distribution:

```
import numpy as np

class Target(object):
    def __init__(self, ndim, nmodes):
        # Generate random inverse variances for each dimension
        self.ivar = 1. / np.random.rand(ndim)

        # Space modes 5-sigma apart
        std = np.sqrt(1/self.ivar)
        self.means = 5 * std * np.arange(nmodes)[: , np.newaxis]
```

MultiNest ([Feroz et al., 2013](#))

Table of Contents

Welcome to pymultinest's documentation!

- About
- Get PyMultiNest
- Citing PyMultiNest
- Documentation:
- Indices and tables

Next topic

Installing PyMultiNest and PyCuba

This Page

[Show Source](#)

Quick search

Welcome to pymultinest's documentation!

About

This is the documentation for two python modules for Bayesian analysis, parameter estimation and model selection: pymultinest and pycuba.

- PyMultiNest** interacts with **MultiNest**, a Nested Sampling Monte Carlo library.
- PyCuba** interacts with **cuba**, a library for various numerical integration methods.

Get PyMultiNest

- Follow the instructions in the installation guide.
- You can download the source code from the [GitHub code repository](#).

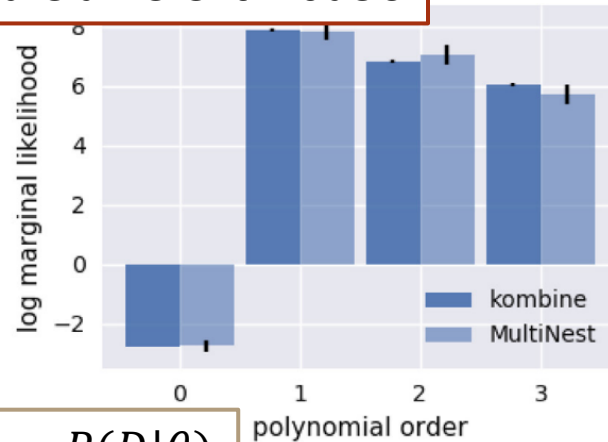
Citing PyMultiNest

Many great things once you can compute integrals

Sample a many points from the posterior, then you can...

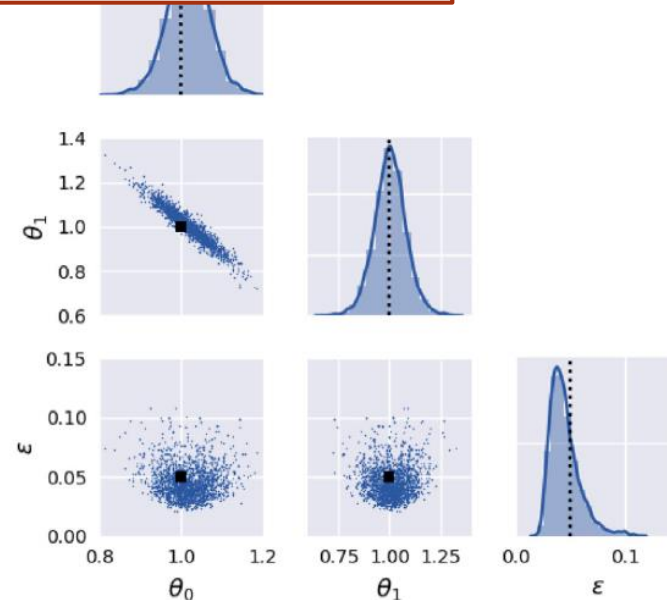
$$E(h(\theta)|D) = \int h(\theta)p(\theta|D)d\theta$$

Compare different models



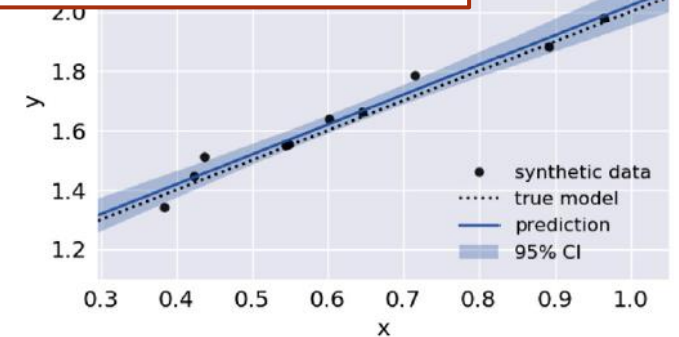
$$h(\theta) = P(D|\theta)$$

Establish confidence intervals of parameters



$$h(\theta) = \theta_0$$

Establish confidence intervals of predictions

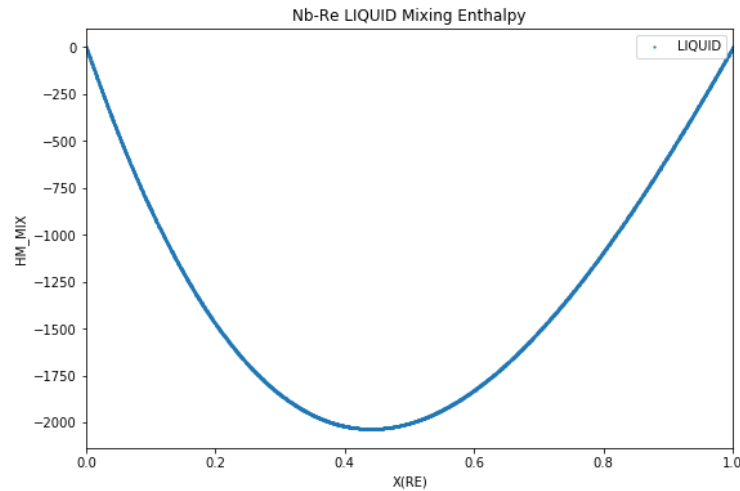


$$h(\theta) = \theta_0 + \theta_1 x$$

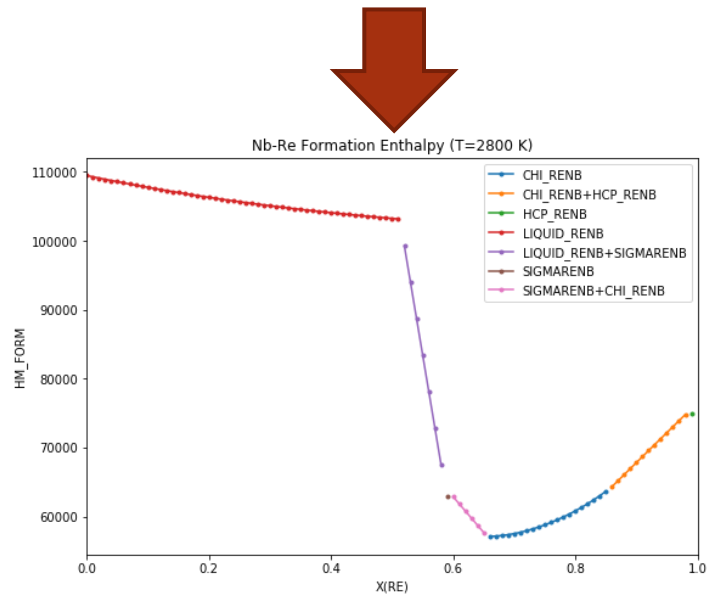
MATERIALS ENGINEERING EXAMPLE

Courtesy of [Paulson et al.](#)

Thermodynamic modeling: CALPHAD

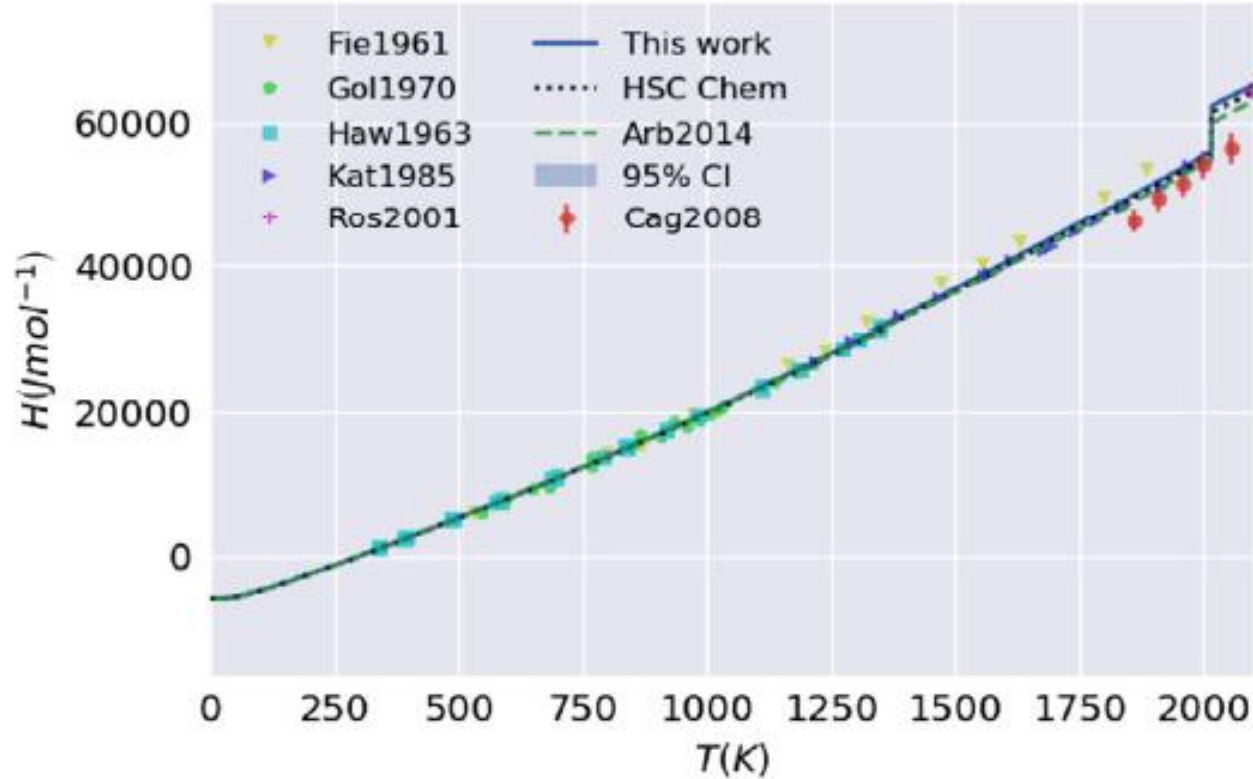


Step 1: Fit thermodynamic models of materials



Step 2: Use them to compute relative stability

Key issue: Fitting thermodynamic models



Many problems:

1. Uncertainty in model form
2. Noise in data collection
3. Disagreement between sources

Solution: Bayesian methods!

Fitting Thermodynamic Models

Innovation 1: Choosing different thermodynamic models

$$H(T) - H(298.15K) = \frac{3R\theta}{e^{\frac{T}{\theta}} - 1} + a\frac{T^2}{2} + \frac{bT^3}{2} \quad C_p(T) = \frac{d}{dT} H(T)$$

Innovation 2: Learning which data sources are more reliable

$$P(D|\theta, \alpha, M) = \prod_i \prod_j \mathcal{N}(y_j^i | M(x_j^i, \theta), \frac{1}{\alpha^i})$$

Include a "dataset uncertainty" (α)

Loop over datasets (i) and points within dataset (j)

Larger α_i means more weight

Innovation 3: Fitting to more than one property at a time

$$P(D|\theta, M) = \prod_i \mathcal{N}(D_{H,i} | H(T_i, \theta), \epsilon_i) \prod_j \mathcal{N}(D_{C_p,j} | C_p(T_j, \theta), \epsilon_j)$$

Different terms for each type of data

Same parameters (θ) in both types

Result: Can fit models, parameters, datasets

Can rule out certain models

Table 1

Potential models, their log Marginal-Likelihoods and Bayes' Factor with respect to the selected model (marked with an asterisk) are presented for each phase.

Model α -phase	Number of model parameters	Log marginal likelihood	Bayes' factor
Einstein	1	-1744.2	~ 0
Debye	1	-1262.9	~ 0
Debye + Linear	2	-1072.6	~ 0
Debye + Quadratic	3	-813.2	~ 0
Debye + Cubic	4	-640.2	~ 0
Debye + Quartic*	5	-623.1	1
Debye + Quintic	6	-627.4	1.4×10^{-2}
Debye + SR	5	-629.7	1.4×10^{-3}
β-phase			
Constant	2	-534.2	~ 0
Linear	3	-511.1	3.0×10^{-3}
Quadratic*	4	-505.3	1
Cubic	5	-518.5	1.9×10^{-6}
Liquid-phase			
Constant	2	-491.4	~ 0
Linear*	3	-471.0	1
Quadratic	4	-476.0	6.7×10^{-3}

Parameters and errors for models

Table 2

The mean, standard deviation, 2.5th percentile bound and 97.5th percentile bound are presented for each parameter for the best model for each phase.

Parameter α -phase	Mean	Std. dev.	2.5% CI	97.5% CI
θ_D	206.45	0.28	205.93	207.02
a_2	1.41×10^{-3}	7.58×10^{-5}	1.26×10^{-3}	1.56×10^{-3}
a_3	9.71×10^{-6}	2.29×10^{-7}	9.26×10^{-6}	1.02×10^{-5}
a_4	-5.61×10^{-9}	2.44×10^{-10}	-6.07×10^{-9}	-5.12×10^{-9}
a_5	1.10×10^{-12}	7.63×10^{-14}	9.41×10^{-13}	1.24×10^{-12}
$\alpha_{Ade1952}$ Adenstedt (1952)	1.111	0.532	0.348	2.402
$\alpha_{Anu1972}$ Arutyunov, Banchila, and Filippov (1972)	0.569	0.138	0.340	0.876
$\alpha_{Bur1958}$ Burk, Estermann, and Friedberg (1958)	0.188	0.023	0.147	0.237
$\alpha_{Cag2008}$ Cagran, Hüpf, Wilthan, and Pottlacher (2008)	0.533	0.217	0.210	1.046
$\alpha_{Cez1974}$ Cezairliyan and McClure (1975)	4.137	0.959	2.613	6.343
$\alpha_{Col1971}$ Collings and Ho (1971)	0.209	0.040	0.140	0.295
$\alpha_{Fie1961}$ Fieldhouse and Lang (1961)	0.444	0.118	0.250	0.709
$\alpha_{Fil1971}$ Filippov and Yurchak (1971)	0.640	0.163	0.367	1.002

Relative importance of data (α)

Take away points

Bayesian method lets you combine:

1. Initial guesses about parameters “priors”
2. Experimental evidence

to generate a new distribution of parameters “**posterior**”

Compute properties by “**sampling posterior**”

Use Bayesian to get better **confidence intervals**
on physics models

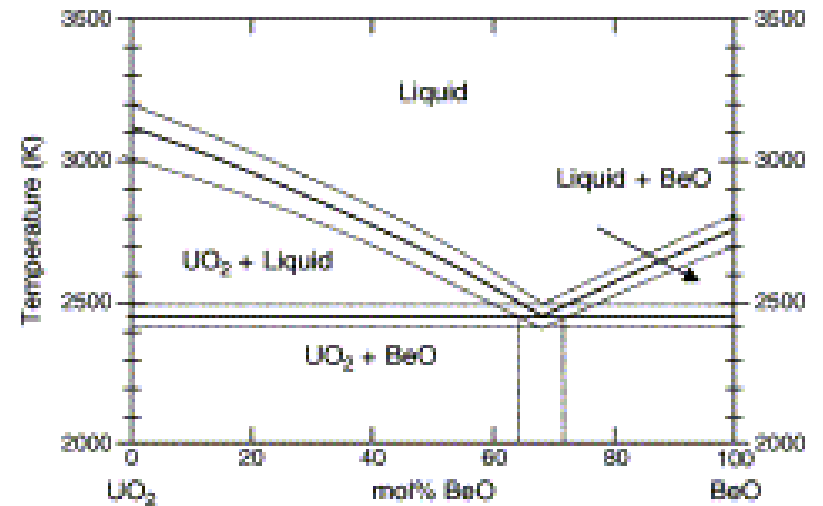


Figure: [Stan, Reardon. Calphad \(2003\)](#)