

Multiple Model Inference: Calibration and Selection with Multiple Models

Laura P. Swiler, Angel Urbina

Sandia National Laboratories

P.O. Box 5800, Albuquerque, NM, 87185-1318

lpswile@sandia.gov, 505-844-8093; aurbina@sandia.gov, 505-844-4988

Brian Williams

Statistical Sciences Group

Los Alamos National Laboratory,

PO Box 1663, MS F600

Los Alamos, NM 87545

brianw@lanl.gov; 505-667-2331

Uncertainty Quantification in Industrial and Energy Applications:
Experiences and Challenges
IMA Workshop, June 02-04, 2011

Sandia National Laboratories is a multi-program laboratory operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin company, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.



Abstract

This talk compares three approaches for model selection: classical least squares methods, information theoretic criteria, and Bayesian approaches. Least squares methods are not model selection methods although one can select the model that yields the smallest sum-of-squared error function. Information theoretic approaches balance overfitting with model accuracy by incorporating terms that penalize more parameters with a log-likelihood term to reflect goodness of fit. Bayesian model selection involves calculating the posterior probability that each model is correct, given experimental data and prior probabilities that each model is correct. As part of this calculation, one often calibrates the parameters of each model and this is included in the Bayesian calculations. Our approach is demonstrated on a structural dynamics example with models for energy dissipation and peak force across a bolted joint. The three approaches are compared and the influence of the log-likelihood term in all approaches is discussed.



Overview

- Introduction of methods
 - Calibration
 - Model Selection
 - Inference
- Case study
- Results
- Summary



Definitions

- Model: mathematical abstraction used to represent the mappings between input parameters and output responses. Most often, we refer to “model” as a computational simulation code used to predict physical phenomena of interest.
- Calibration: the identification of optimal parameter settings for a model, so that agreement between model calculations and a set of experimental data is maximized. Calibration is sometimes called least-squares methods, system identification, parameter estimation, or inverse problems.
- Model selection: the process of determining the best model out of N available models, according to some criterion. The criterion may be a goodness-of-fit measure, maximum likelihood, an information theoretic measure, or a maximum posterior model probability in a Bayesian approach.
- Model inference: the process of using calibrated model parameters to predict a response at new input settings. Model prediction usually does not just involve one prediction but a set or ensemble of predictions, based on uncertainty quantification of parameters and/or model averaging.

Model Calibration: Frequentist Approaches

- Nonlinear least squares: find the optimal values of θ to minimize the error sum of squares function $S(\theta)$

$$S(\theta) = \sum_{i=1}^n [(y_i(x_i) - g(x_i; \theta))]^2 = \sum_{i=1}^n [R_i(\theta)]^2$$

Experimental data

Simulation output that depends on x and θ

- Maximum Likelihood: find the optimal values of θ to maximize the likelihood of the parameters, given the data:

$$\left. \begin{aligned} y_i &= g(x_i; \theta) + \varepsilon_i \\ \varepsilon_i &\sim N(0, \sigma^2) \end{aligned} \right\} \text{Assuming i.i.d. error}$$

Gaussian Likelihood Function

$$L(\theta) = \prod_{i=1}^n \frac{1}{2\sqrt{\pi}\sigma} e^{-\left[\frac{(y_i - g(\theta))^2}{2\sigma^2}\right]}$$

$$\text{LogLikelihood}(\theta) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{(y_i - g(\theta))'(y_i - g(\theta))}{2\sigma^2}$$



Model Calibration: Bayesian Approaches

Given data D and a prior distribution on parameters to be calibrated, $p(\theta)$, find the posterior distribution of the parameters given the data

$$p(\theta | D) = \frac{p(D, \theta)}{p(D)} = \frac{p(D | \theta)p(\theta)}{p(D)} = \frac{p(D | \theta)p(\theta)}{\int_{\theta} p(D | \theta)p(\theta)d\theta}$$

This is often written as: $p(\theta | D) \propto p(D | \theta)p(\theta)$

posterior \propto *likelihood* * *prior*

Again, we assume a Gaussian likelihood:

$$p(D | \theta) = L(\theta) = \prod_{i=1}^n \frac{1}{2\sqrt{\pi}\sigma} e^{-\left[\frac{(D_i - g(\theta))^2}{2\sigma^2}\right]}$$

In practice, Monte-Carlo Markov Chain methods (MCMC) are used to generate posterior distributions.



Model Selection: Frequentist Approaches

- Nonlinear least squares: Pick the model with the smallest error sum of squares function $S(\theta)$
- Maximum Likelihood: There are several criteria based on information theory. They all seek to maximize goodness-of-fit while penalizing for overfitting (e.g. having a very complicated model with many parameters may give a great match to the data but will not extrapolate).
- Information theory based on Kullback-Leibler distance.

$$KL(f, g) = \int f(x) \log\left(\frac{f(x)}{g(x|\theta)}\right) dx$$

- The KL distance refers to the information lost when g is used to approximate f . Akaike showed that model selection should minimize expected K-L distance, and he found a relationship between the relative expected K-L distance and the maximized log-likelihood. His measure, AIC, is:

$$AIC = -2 \log \mathcal{L}(\hat{\theta} | x) + 2K$$

Goodness of fit

Penalty for number
of model
parameters K



Model Selection: Bayesian Approaches

Given K models, data D and a prior distribution on parameters for each model, $p(\theta_k)$, choose the model with the highest posterior probability of being the true model given the data:

$$p(M_k | D) = \frac{p(D | M_k)p(M_k)}{\sum_{l=1}^K p(D | M_l)p(M_l)}$$

Where $p(M_k)$ is the prior probability that model M_k is the true model, and $p(D|M_k)$ the integrated likelihood function of model M_k . This is also referred to as the evidence for model M_k :

$$p(D | M_k) = \int p(D | \theta_k, M_k) p(\theta_k | M_k) d\theta_k$$

In practice, there do not exist robust methods to calculate this (estimators of the integrated likelihood tend to have high variance). Some approaches are the Wolpert method, the harmonic mean, and reversible jump MCMC. Cite Fall 2010 paper.



Model Selection Criteria

Model Selection Criterion	Expression
Akaike (AIC)	$AIC = \log p(D \hat{\theta}_k, M_k) - N_k$
Akaike corrected (AIC-c)	$AIC_c = \log p(D \hat{\theta}_k, M_k) - \left(N_k + \frac{N_k(N_k + 1)}{N - N_k - 1} \right)$
Akaike-Schwarz Bayesian (BIC)	$BIC = \log p(D \hat{\theta}_k, M_k) - \frac{N_k}{2} \log N$
Bayesian model selection	Log evidence = $\log p(D M_k)$ This may be calculated via simple mean, harmonic mean, or use of MCMC methods.



Model Inference

Nonlinear least squares and Maximum likelihood approaches: Given calibrated parameters $\hat{\theta}$, evaluate the model at these parameter values. Use standard approaches to generate confidence intervals on parameters (e.g. linear approximation, F-test, etc.) Samples can be taken from bounds on parameters to generate ensembles of predictions.

Bayesian approach: For a particular model, one has an entire posterior distribution on parameters θ which can be propagated through the simulation model to generate posterior realizations of the responses.

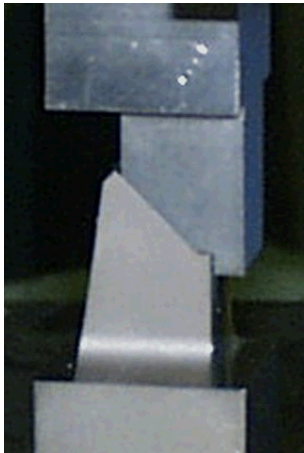
In the case of multiple models, Bayesian Model Averaging can also be used to generate a weighted response (weighted by the posterior probabilities on the models):

$$p(R | D) = \sum_k p(R | M_k, D) p(M_k | D)$$

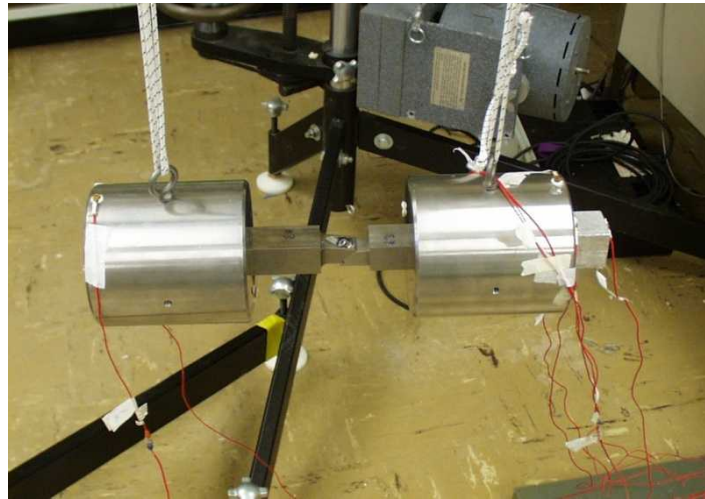
In this work, we focused on model calibration and selection, not inference.

Case study: Bolted Joint

- Want to understand the mechanical properties of a bolted joint system, specifically microslip
- Two response measures of interest: peak force and slope of energy dissipated

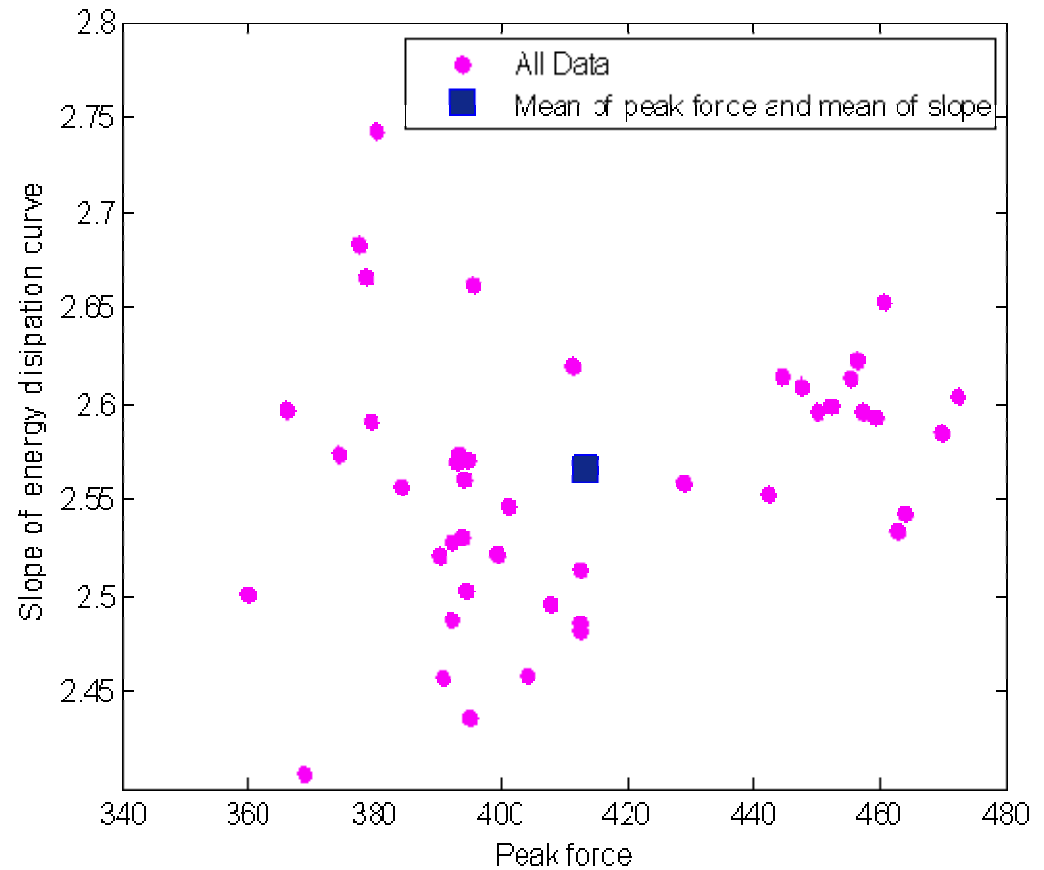
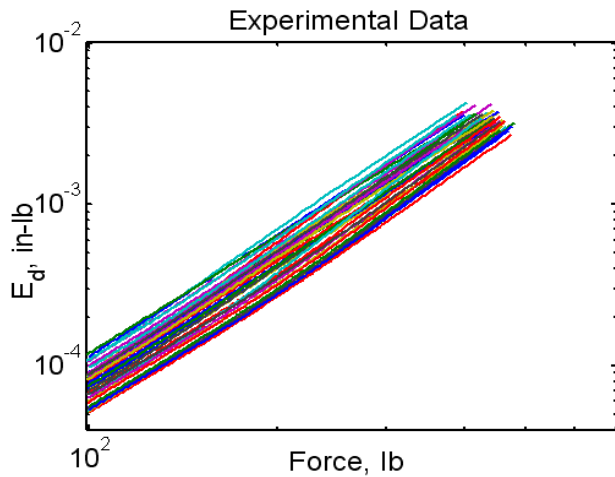
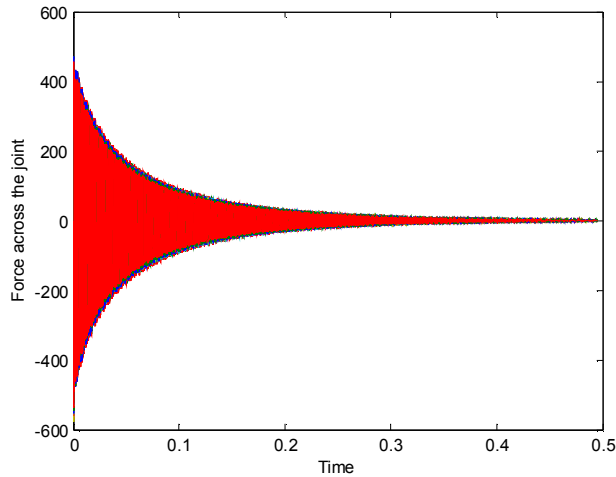


Support Element

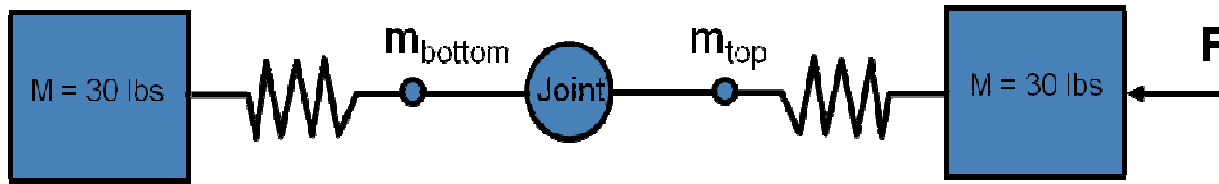


Dumbbell test configuration

Case study: Experimental Data

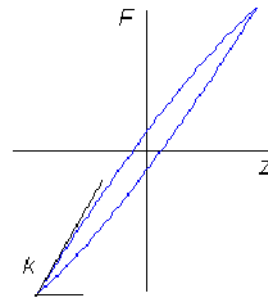


Case Study: Three models

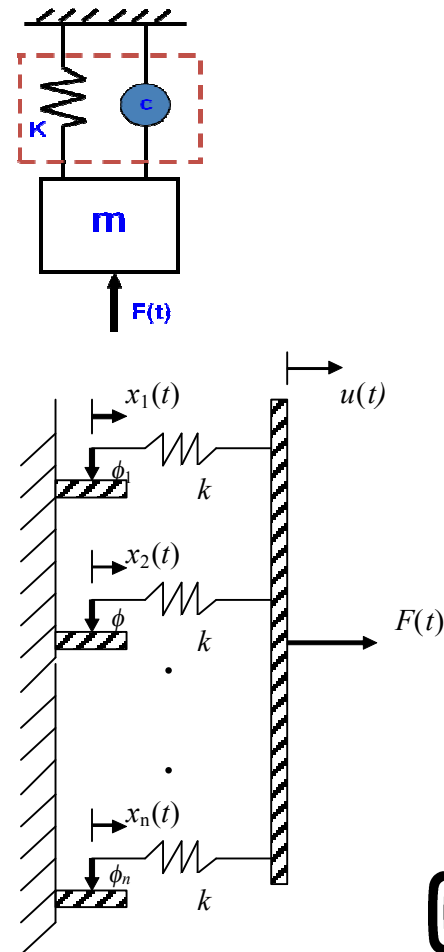


Dashpot: A dashpot represents a damping term proportional to velocity. Dashpot elements combine a viscous friction damper with a simple linear spring.

Smallwood: Model the force vs. displacement curves assuming the force across the joint and relative displacement are harmonic functions



Iwan: Simulate the behavior of hysteretic systems by using parallel sequence of sub-elements, each of which is a linear elastic spring in series with a rigid, perfectly plastic slider element.



Calibration Results: Least Squares

- Used nl2sol in DAKOTA with 20 starting points.
- NI2sol is a Levenberg-Marquardt algorithm with Gauss-Newton approximations to the Hessian matrix. It is especially designed for NLLS.
- Note that Smallwood was very consistent, always going to the same optimal point regardless of initial point.

Dashpot				Smallwood				Iwan				
Iteration	SSE	k	mu	SSE	klin	knon	n	SSE	chi	phi_max	R	S
1	7.86E+04	4.15E+06	7.00E+01	4.90E+04	2.16E+06	1.59E+06	1.47	3.01E+05	-6.14E-01	1.49E-04	3.87E+06	8.74E+06
2	6.44E+04	3.81E+06	7.00E+01					1.76E+05	-6.45E-01	1.56E-04	2.85E+07	4.38E+06
3	4.75E+04	2.97E+06	6.19E+01					4.74E+04	-6.50E-01	6.99E-05	1.07E+07	5.64E+06
4	4.96E+05	5.89E+07	7.00E+01					3.55E+05	-7.33E-01	8.91E-05	3.41E+07	1.03E+07
5	4.75E+04	2.74E+06	4.00E+01					1.55E+05	-6.61E-01	9.86E-05	3.38E+07	4.38E+06
6	8.47E+04	4.29E+06	7.00E+01					2.43E+05	-5.14E-01	1.47E-04	3.74E+07	6.48E+06
7	8.38E+04	4.27E+06	7.00E+01					4.74E+04	-5.20E-01	8.67E-05	3.53E+07	4.50E+06
8	4.47E+05	2.71E+07	7.00E+01					3.10E+05	-5.69E-01	9.66E-05	1.11E+07	9.23E+06
9	4.75E+04	3.05E+06	6.99E+01					4.74E+04	-6.83E-01	7.27E-05	2.78E+07	4.66E+06
10	4.75E+04	2.71E+06	3.72E+01					4.74E+04	-5.14E-01	7.56E-05	2.06E+07	5.33E+06
11	4.75E+04	2.84E+06	4.97E+01					7.47E+04	-5.17E-01	5.77E-05	5.20E+06	7.55E+06
12	4.75E+04	2.71E+06	3.73E+01					1.90E+05	-7.30E-01	1.11E-04	1.72E+07	4.96E+06
13	4.75E+04	2.88E+06	5.29E+01					2.31E+05	-5.94E-01	1.53E-04	2.39E+07	6.05E+06
14	7.32E+04	4.03E+06	7.00E+01					2.56E+05	-6.29E-01	1.19E-04	7.36E+06	7.04E+06
15	5.02E+05	6.68E+07	7.00E+01					4.74E+04	-6.05E-01	6.41E-05	4.06E+07	5.81E+06
16	4.75E+04	3.05E+06	7.00E+01					4.74E+04	-5.14E-01	6.07E-05	1.49E+07	6.72E+06
17	4.92E+05	5.38E+07	7.00E+01					4.74E+04	-7.43E-01	5.77E-05	2.17E+07	5.76E+06
18	4.75E+04	2.89E+06	5.47E+01					4.75E+04	-5.96E-01	5.77E-05	1.14E+07	7.00E+06
19	4.75E+04	2.71E+06	3.79E+01					2.17E+05	-5.98E-01	1.25E-04	3.98E+07	5.60E+06
20	4.75E+04	2.78E+06	4.41E+01					4.74E+04	-7.32E-01	5.77E-05	2.59E+07	5.66E+06

Calibration Results: Maximum Likelihood

- Augmented the standard MLE to incorporate “matching” of two response functions:

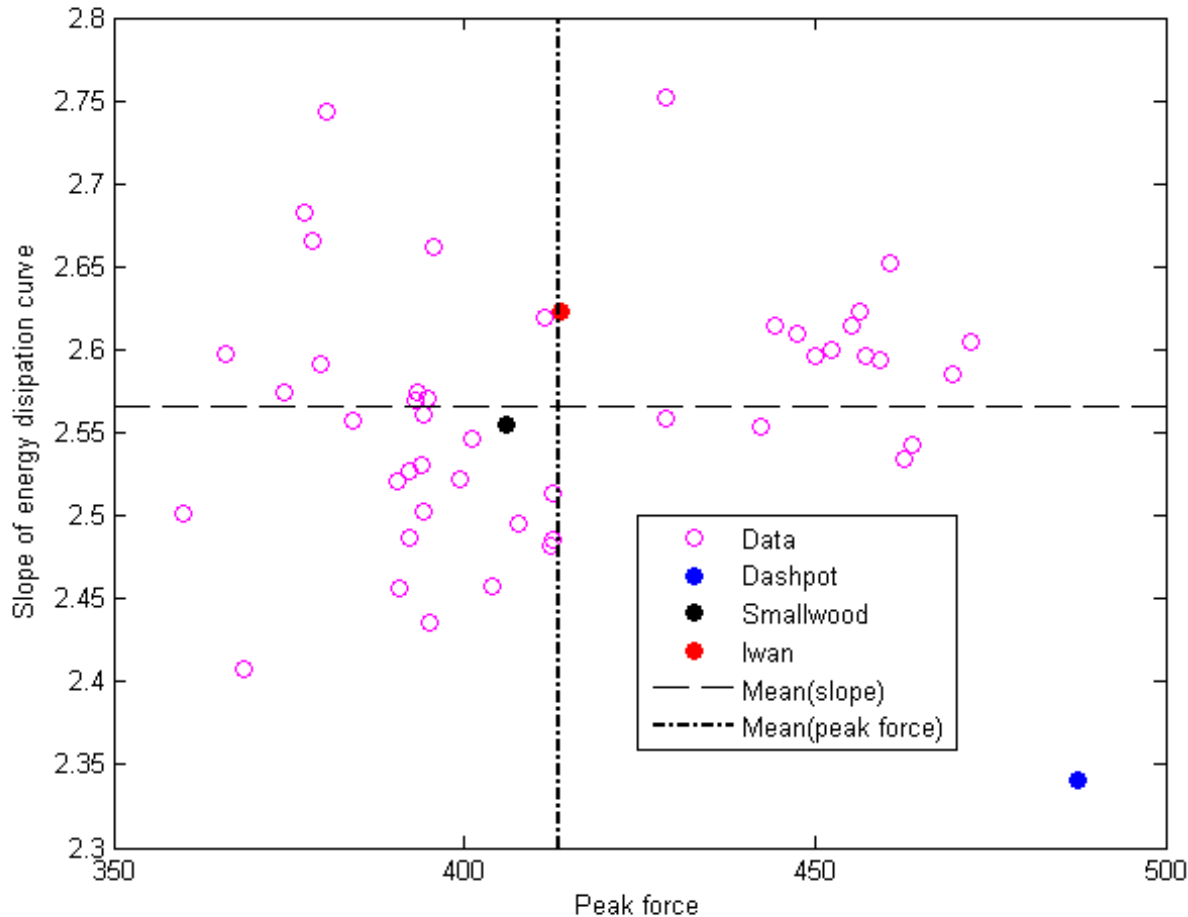
$$LogLikelihood_{Peak}(\theta) = -\frac{45}{2} \log(\sigma_{Peak}^2) - \frac{(y_{peak,i} - g(\theta))'(y_{peak,i} - g(\theta))}{2\sigma_{Peak}^2}$$

$$LogLikelihood_{Slope}(\theta) = -\frac{45}{2} \log(\sigma_{Slope}^2) - \frac{(y_{Slope,i} - g(\theta))'(y_{Slope,i} - g(\theta))}{2\sigma_{Slope}^2}$$

$$NegLogLikelihoodTotal(\theta) = -1 * [LogLikelihood_{Peak}(\theta) + LogLikelihood_{Slope}(\theta)]$$

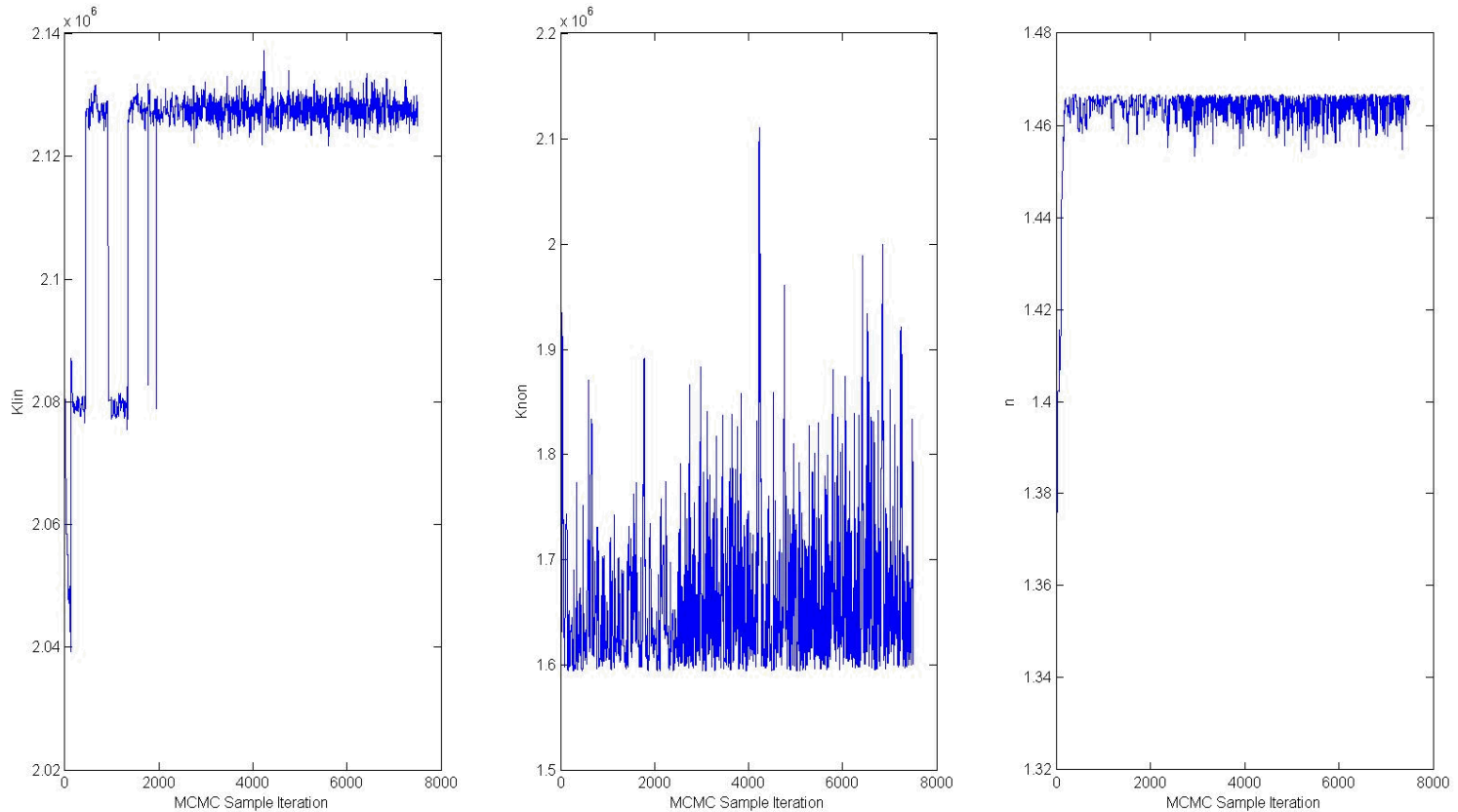
MAXIMUM LIKELIHOOD RESULTS SOLVING FOR ERROR TERMS AND PARAMETERS					
DASHPOT		SMALLWOOD		IWAN	
Parameter	Optimal Value	Parameter	Optimal Value	Parameter	Optimal Value
k	1.01E+07	klin	2.13E+06	chi	-5.35E-01
mu	69.78	knon	1.71E+06	phi_max	7.41E-05
sigma_peak	39.88	n	1.465	R	1.95E+07
sigma_slope	0.10	sigma_peak	33.704	S	5.43E+06
		sigma_slope	0.078	sigma_peak	33.704
				sigma_slope	0.075
NEG LOG LIKE.	281.86		85.42		84.74

Calibration Results: Maximum Likelihood



Model Results using Optimal MLE parameters		
	Peak Force	Slope
Dashpot	487.50	2.34
Smallwood	406.10	2.55
Iwan	413.60	2.62
Mean of Exp. Data	415.90	2.56

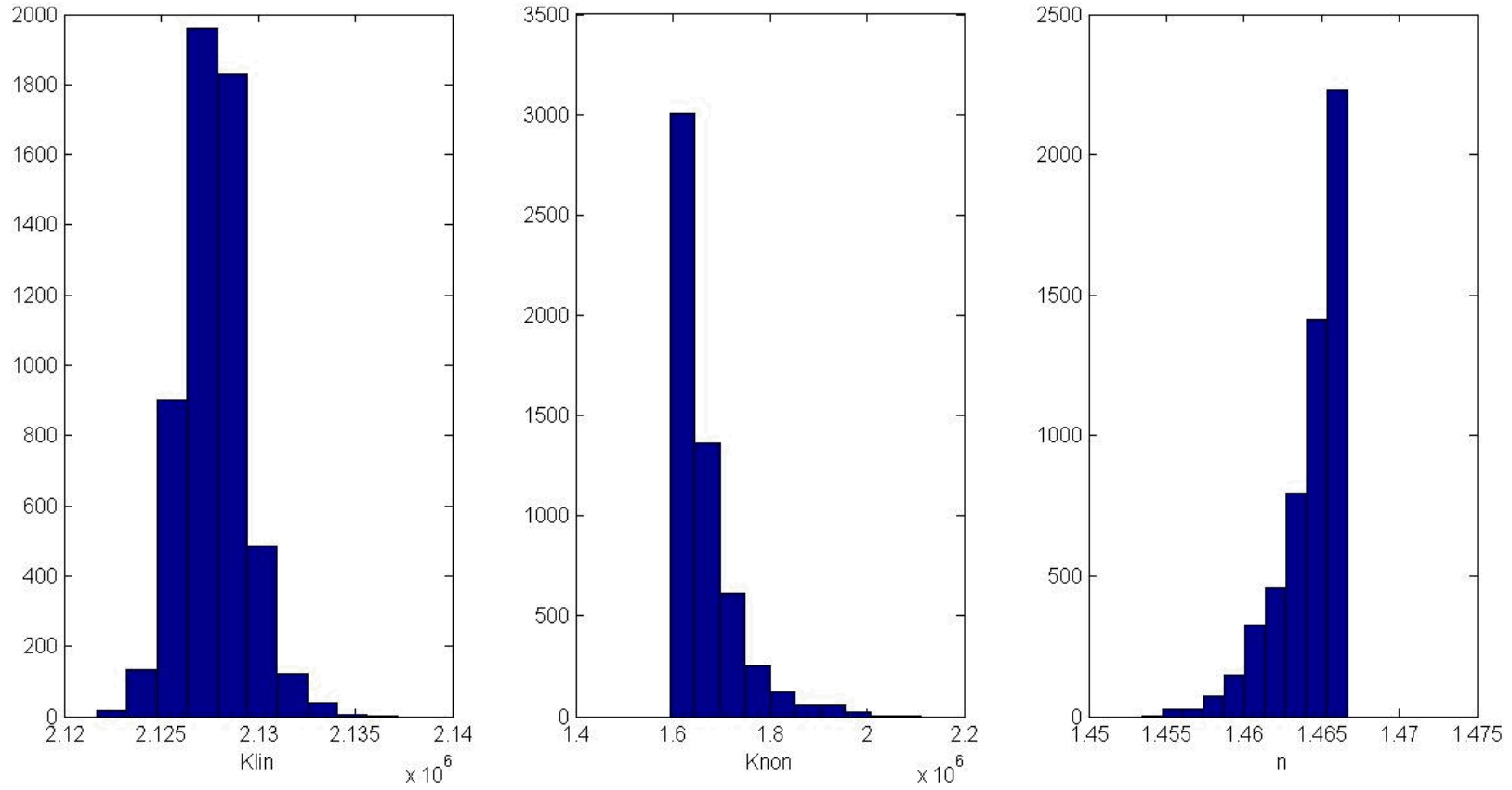
Calibration Results: Bayesian Approach



Posterior MCMC Chain for Smallwood Model Parameters



Calibration Results: Bayesian Approach



Histograms for Posterior estimates of Smallwood Model Parameters

Overall Calibration Comparison

	Dashpot			
	k	mu		
NLLS	1.29E+07	59.28		
MLE	1.01E+07	69.78		
MCMC 1	5.71E+06	35.74		
MCMC 2	7.89E+06	56.32		
	Smallwood			
	klin	knon	n	
NLLS	2.16E+06	1.59E+06	1.47	
MLE	2.13E+06	1.71E+06	1.46	
MCMC 1	2.13E+06	1.66E+06	1.46	
	Iwan			
	Chi	Phi Max	R	S
NLLS	-0.62	9.52E-05	2.26E+07	6.29E+06
MLE	-0.53	7.41E-05	1.95E+07	5.43E+06
MCMC 1	-0.54	7.31E-05	1.37E+07	5.56E+06

- Very good agreement between optimal parameters determined by MLE and mean of MCMC chain for Smallwood and Iwan.
- Dashpot results more varied, due to difficulty of matching slope.



Model Selection

Nonlinear Least Squares: Pick Smallwood according to smallest average SSE, pick Iwan for smallest minimum SSE (over the 20 starting points)

	Dashpot	Smallwood	Iwan
AVERAGE SSE	1.56E+05	4.90E+04	1.36E+05
MINIMUM SSE	4.75E+04	4.90E+04	4.74E+04
MAXIMUM SSE	5.02E+05	4.90E+04	3.55E+05

Information theoretic criteria: Pick Iwan strictly on the basis of Maximum Likelihood, but pick Smallwood according to the other information criteria. The fact that Iwan and Smallwood are comparable but Smallwood has one fewer parameters makes Smallwood the better model in this case.

	Dashpot	Smallwood	Iwan
Log Likelihood	-281.86	-85.42	-84.74
AIC	-285.86	-90.42	-90.74
AIC-C	-286.10	-90.78	-91.19
BIC	-290.86	-96.67	-98.24

Note that the criteria were formulated so maximum is better.



Model Selection: Bayesian approach

- Add Harmonic means, with uncertainty from Brian
- Add discussion of discrepancy



Summary

- We tried a variety of calibration and model selection methods
- For models that were ill-posed or had some parameters that interacted significantly, nonlinear least squares found multiple optimal solutions.
- We saw that MLE estimation and the means of the Bayesian posterior estimates gave very similar results for Smallwood and Iwan.
- Although Iwan might be a slightly better model in terms of a slightly better likelihood and SSE, Smallwood is the best model according to the information theoretic criteria such as AIC.
- Implementing and debugging the methods proved more challenging than we expected, especially since we were trying to match two disparate sets of data.
- Information criteria appears to be a more robust, less expensive approach for model selection than Bayesian methods.
- Bayesian methods buy you more in the model inference step.