

# Multiple Model Inference: Calibration, Selection, and Prediction with Multiple Models

Laura P. Swiler<sup>1</sup> and Angel Urbina<sup>2</sup>  
*Sandia National Laboratories, Albuquerque, NM 87185*

Brian J. Williams<sup>3</sup>  
*Los Alamos National Laboratory, NM 87545*

**This paper 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.**

## I. Introduction

Developers and users of models generally want to pick the “best” model for their needs. There are many criteria by which this can be done. Often, there is a notion that a good model will produce predictions that are in good agreement with experimental data but also not be overly complicated (e.g. the “parsimony principle”: if a simpler model with fewer parameters is adequate, we would prefer to use that). A long history exists in the statistical literature about goodness of fit measures.<sup>1</sup> There are also approaches that seek to combine goodness-of-fit with parsimony.<sup>2</sup> Several measures used in model selection are based in information theory and focus on providing relative measures of the information lost when a particular model is used to describe a particular situation or data set. Some of these measures include Akaike’s information criterion (AIC), Bayesian information criterion (BIC), deviance information criterion (DIC), and Kullback-Leibler divergence criterion (K-L). These information theoretic criteria try to balance overfitting with model accuracy by incorporating terms that penalize more parameters (for example) with terms that reflect the goodness of fit.

Bayesian model selection<sup>3-5</sup> involves a different concept. In Bayesian approaches, one specifies a prior probability of a model being correct. Often this is done with the a priori assumption that all models are equally likely to be correct. One also specifies likelihood functions for the data given each model. Then, one uses Bayes’ theorem to calculate the posterior probability that each model is correct, given the data. As part of this calculation, one often calibrates the parameters of each model and this is included in the Bayesian calculations.

In this research, we are concerned with situations where we have a few models, but the physics may be substantially different between models and some parameters are common to all models but some parameters are model unique. The goal of this work is to understand techniques used to perform model selection and model prediction in this context. Specifically, we want to examine model selection and model prediction from an

---

<sup>1</sup> Principal Member of Technical Staff, Optimization and Uncertainty Estimation, MS 1318, Sandia National Laboratories, Albuquerque NM 87185, AIAA member.

<sup>2</sup> Senior Member of Technical Staff, Validation and Uncertainty Quantification, MS 0828, Sandia National Laboratories, Albuquerque NM 87185.

<sup>3</sup> Scientist, Statistical Sciences Group (CCS-6), Los Alamos National Laboratory, PO Box 1663, MS F600 Los Alamos National Laboratory, Los Alamos, NM 87545.

information theoretic viewpoint as well as a Bayesian perspective. We start by providing an overview of model selection work in classical parameter estimation, information theory approaches, and Bayesian methods. Terminology can be an issue, so we define the key concepts and terms used in this manuscript below:

### ***Model***

We define model as a 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, but a model can also include statistical models such as regression models and surrogate models such as neural networks. By model, we are referring to both the model form (sometimes referred to as the “model class”, also referred to as “physics model used”) and the values of parameters associated with the model. In our research, we are concerned with situations where we have a few (2-5) models, but the physics may be substantially different between models (e.g. one-phase vs. two-phase flow, elastic vs. inelastic collisions, low vs. high fidelity, etc.) and some parameters are common to all models but some parameters are model unique.

### ***Calibration***

Calibration is a broad term, and is sometimes called least-squares methods, system identification, parameter estimation, or inverse problems. For this paper, calibration refers to determining optimal parameter settings for a model, so that agreement between model calculations and a set of experimental data is maximized. Often the criteria is to minimize the sum of the squared residuals (where residual is the difference between the model prediction and the experimental data).<sup>6,7</sup> This is typically called minimizing the “sum squared error.” There are more sophisticated objective functions to use as criteria, and this paper will focus on two classes of calibration methods. The first is information-theoretic methods, which seek to optimize a particular criterion which balances model complexity with model precision, such as Akaike’s information criterion. The second is Bayesian methods, which seek to construct posterior distributions on parameters which are consistent with the data and with the assumed form of the likelihood function.

### ***Model Selection***

Model selection is 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 Bayesian measure such as maximum posterior model probability. Note that often we do not just want the “one” best model, but we want to understand how the models compare on the relative measures we are examining, and how sensitive the model goodness is to model form.

### ***Model Inference***

Model inference is sometimes called model prediction: given a model and associated parameter values that have been calibrated in some optimal way, what are the model predictions for various scenarios? Model inference may be based on the best model chosen in a model selection scheme, or it may involve a weighted model average, where the prediction is a weighted sum over all models, where each model prediction is weighted by some type of criterion. In addition, model prediction usually does not just involve one prediction but a set or ensemble of predictions. These predictions typically involve some type of uncertainty quantification, where one wants to quantify the effect that uncertain input variables have on model output.

## **II. Frequentist Approaches**

A common approach to model selection is to calibrate two or more models to the same data set, and pick the one which has a better figure of merit, such as the lowest sum-of-squares error (SSE). Metrics such as  $R^2$  and SSE measure the “goodness of fit” of the model with respect to the training points used to construct the model. Other measures have been developed which balance overfitting with model accuracy by incorporating terms that penalize more parameters (for example) with terms that reflect the goodness of fit. These measures, including Akaike’s information criterion (AIC), Bayesian information criterion (BIC), and the Kullback-Leibler divergence criterion (K-L), are based in information theory. Another approach is stepwise regression, where a large set of candidate models are examined. In stepwise regression, the form of the model is usually a regression model (e.g. a linear or quadratic model), and different models refer to incorporating different numbers of predictor/independent variables. Stepwise methods usually involve forward selection or backward elimination. Forward selection starts with no

variables in the model, considers the variables one at a time, and includes the variable that is the most significant according to an F-test or a t-test. Backward elimination starts with all candidate variables, tests them one by one for statistical significance, and deletes any that are not significant. In this paper, we focus on models that are simulations and not regression models. For this reason, we will limit our discussion of frequentist approaches to nonlinear least squares and to the information theoretic criteria outlined above.

### A. Nonlinear least squares

Nonlinear least squares extend linear regression methods. In nonlinear least squares, there are few limitations on the way parameters may enter into the functional part of a nonlinear regression model. The way in which the unknown parameters in the nonlinear function are estimated, however, is conceptually the same as in linear least squares regression. The nonlinear model of the response  $y$  as a function of the  $n$ -dimensional inputs  $\mathbf{x}$  is given as:

$$y = f(\mathbf{x}; \boldsymbol{\theta}) + \varepsilon \quad (1)$$

where  $f$  is the nonlinear model,  $\boldsymbol{\theta}$  is a vector of parameters to be calibrated,  $\varepsilon$  is a random error term, and we assume that  $E[\varepsilon] = 0$  and  $Var[\varepsilon] = \sigma^2$  and the error terms are independent and identically distributed (iid). Usually  $y$  is a function of  $\mathbf{x}$  but this dependence is often implicit and  $y(\mathbf{x})$  simply written as  $y$ . Note that for nonlinear functions, the derivative of  $f$  with respect to the parameters  $\boldsymbol{\theta}$  depends on at least one of the parameters of the vector  $\boldsymbol{\theta}$ . Given observations of the response  $y$  corresponding to the independent variables  $\mathbf{x}$ , the goal of nonlinear regression is to find the optimal values of  $\boldsymbol{\theta}$  to minimize the error sum of squares function  $S(\boldsymbol{\theta})$ , also referred to as SSE:

$$S(\boldsymbol{\theta}) = \sum_{i=1}^n [(y_i - f(\mathbf{x}_i; \boldsymbol{\theta}))^2] = \sum_{i=1}^n [R_i(\boldsymbol{\theta})]^2 \quad (2)$$

where  $R_i(\boldsymbol{\theta})$  are the residual terms. Nonlinear regression employs an optimization algorithm to find the least squares estimator  $\hat{\boldsymbol{\theta}}$  of the true minimum  $\boldsymbol{\theta}^*$ ; a process that is often difficult<sup>6,7</sup>. Derivative-based nonlinear least squares optimization algorithms exploit the structure of such a sum of squares objective function. If  $S(\boldsymbol{\theta})$  is differentiated twice, terms of residual  $R_i(\boldsymbol{\theta})$ ,  $R_i''(\boldsymbol{\theta})$ , and  $[R_i'(\boldsymbol{\theta})]^2$  result. By assuming that the residuals  $R_i(\boldsymbol{\theta})$  are close to zero near the solution, the Hessian matrix of second derivatives of  $S(\boldsymbol{\theta})$  can be approximated using only first derivatives of  $R_i(\boldsymbol{\theta})$ . An algorithm that is particularly well-suited to the small-residual case and the above formulation is the Gauss-Newton algorithm.

Model selection can be performed by calculating the SSE objective for each particular model, then picking the model that has the smallest SSE. While this may be a good screening technique to eliminate models that cannot be calibrated very accurately, it does not penalize models for complexity. Thus, we have the need for information-theoretic techniques defined below. Note also that people often think of least-squares methods giving a deterministic solution (e.g. point estimates) for the optimal parameter values. While it is true that least-squares methods give particular values (and not distributions) for the parameters, there are ways to calculate individual or joint confidence intervals on those parameter values.<sup>6,8,9</sup> These approaches include Bonferroni intervals, the linear approximation method, the F-test method, and the log-likelihood method.

### B. Information Theoretic Criteria

Akaike's information criterion is based on the concept of the Kullback-Leibler (K-L) distance [Reference 2, Burnham and Anderson, Chapter 2]. If  $f$  denotes the "truth model" and  $g$  is an approximating model, the K-L distance or information is defined as:

$$KL(f, g) = \int f(x) \log \left( \frac{f(x)}{g(x | \boldsymbol{\theta})} \right) dx \quad (3)$$

The KL distance refers to the information lost when  $g$  is used to approximate  $f$ . A major issue in practice is how to determine the parameters  $\theta$  for the model  $g$ . Typically, these are obtained using some type of maximum likelihood estimation (MLE). It is improper to just consider the MLE estimator  $\hat{\theta}$  of the true and unknown  $\theta$  value which minimizes the K-L criteria. Instead, Akaike shown that model selection should minimize expected K-L distance. A key part of the AIC measure is to estimate:

$$E_x E_y [\log(g(x|\hat{\theta}(y)))] \quad (4)$$

where the double expectations are taken over independent random samples  $x$  and  $y$  with respect to the truth function. A major contribution of Akaike was to recognize that an approximately unbiased estimator of this double expectation for large samples is:

$$\log(L(\hat{\theta}|data)) - K \quad (5)$$

where  $L$  is the likelihood function,  $\log(L(\hat{\theta}|data))$  is the numerical value of the log-likelihood at its maximum point, and  $K$  is a bias correction term representing the number of parameters that are estimated. According to Burnham and Anderson, "Akaike's finding of a relation between the relative expected K-L distance and the maximized log-likelihood has allowed major practical and theoretical advances in model selection and the analysis of complex data sets." Thus, the AIC was created by multiplying the above relationship by -2 for historical reasons, and AIC is defined as:

$$AIC = -2\log(L(\hat{\theta}|x)) + 2K \quad (6)$$

This is an estimate of the expected distance between the fitted model and the unknown true function. In the case of linear least squares models, if all models in the set assume normally distributed errors with constant variance  $\sigma$ , AIC is given as:

$$AIC = n\log(\hat{\sigma}^2) + 2K \quad (7)$$

$$\hat{\sigma}^2 = \frac{\sum \hat{\varepsilon}_i^2}{n}$$

where  $\hat{\sigma}^2$  is the MLE estimate of  $\sigma^2$  and is computed as the average of the squared estimated residuals  $\varepsilon_i$  from the model.  $K$  here is the total number of estimated regression parameters, including the intercept and  $\sigma^2$ .

The determination of the likelihood function and the maximum likelihood estimate can be challenging. We start with a simple case, where we assume a Gaussian likelihood function for the difference between the model  $g$  and the data  $y$ . Here  $y_i$  is observed data and  $g(\theta)$  is a set of simulator runs. Thus, assuming the case where:

$$y_i = g(\theta) + \varepsilon_i \quad (8)$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

we have an expression for the likelihood:

$$L(\theta) = \prod_{i=1}^n \frac{1}{2\sqrt{\pi\sigma}} e^{-\left[\frac{(y_i - g(\theta))^2}{2\sigma^2}\right]} \quad (9)$$

Optimization algorithms which calculate the maximum likelihood can be derivative or non-derivative based. In our results, we use a global, non-derivative based algorithm called DIRECT.<sup>15</sup>

The log of the likelihood function shown in Equation 9 is often written as:

$$\text{LogLikelihood}(\theta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(\det(C)) - \frac{1}{2}(y_i - g(\theta))'C^{-1}(y_i - g(\theta)) \quad (10)$$

where  $C$  is the  $n*n$  matrix representing the pairwise correlations between the observed data points. However, since we have assumed independent, identically distributed errors as specified in Equation (8), this can be simplified to:

$$\text{LogLikelihood}(\theta) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{(y_i - g(\theta))'(y_i - g(\theta))}{2\sigma^2} \quad (11)$$

Finally, the first term,  $-\frac{n}{2}\log(2\pi)$ , is usually dropped from the likelihood since it is constant. This is the approach we take: we maximize the log-likelihood, or minimize the negative log-likelihood by multiplying Equation 11 by negative one and removing the first term.

### C. Bayesian Approaches

In Bayesian approaches, one specifies a prior probability of a model being correct. Often this is done with the a priori assumption that all models are equally likely to be correct. One also specifies likelihood functions for the data given each model. Then, one uses Bayes' theorem to calculate the posterior probability that each model is correct, given the data. As part of this calculation, one often calibrates the parameters of each model and this is included in the Bayesian calculations.

We start with the framework of Bayesian Modeling Averaging (BMA).<sup>3-5,10-14</sup> In BMA, the posterior distribution of a response quantity  $R$  is given as a weighted average over the  $k$  models under consideration:

$$p(R | D) = \sum_k p(R | M_k, D)p(M_k | D) \quad (12)$$

The first term on the RHS is the average of the posterior distributions of the response for each model under consideration given observational data  $D$ . The second term is the posterior model probability. The posterior probability for model  $M_k$  is given by:

$$p(M_k | D) = \frac{p(D | M_k)p(M_k)}{\sum_{l=1}^K p(D | M_l)p(M_l)} \quad (13)$$

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

$$p(D | M_k) = \int p(D | \theta_k, M_k)p(\theta_k | M_k)d\theta_k \quad (14)$$

where  $\theta_k$  are the parameters to be estimated for model  $M_k$ ,  $p(\theta_k | M_k)$  is the prior density of  $\theta_k$  under model  $M_k$ , and  $p(D | \theta_k, M_k)$  is the likelihood function for model  $M_k$ .

The BMA framework is conceptually simple and appealing: calculate posterior model response as a weighted average of the posterior predictions of each model which are weighted by their posterior model probability. Calculate the posterior model probability using Bayes' theorem, where the likelihood function is integrated over all sets of parameters for that model. However, in practice, calculating these quantities can be difficult. Hoeting et al.<sup>10</sup> list several difficulties: the number of models can be enormous, the integrals involved in the calculation of the

integrated likelihood and posterior model probabilities can be hard to compute, and the specification of the prior probabilities over computing models is challenging. We are not dealing with situations where the number of models is large (which would be the case if we were considering high dimensional problems where “model” referred to different subsets of predictors and different analytic statistical model forms). Instead, we have 2-5 separate simulation models, where each simulation model has somewhat different physics (for example). Beck’s papers<sup>4,11</sup> provide some further extensions to the framework outlined above. In particular, they replace the integral to calculate the integrated likelihood with an approximation:

$$p(D | M_k) \approx p(D | \hat{\theta}_k, M_k) p(\hat{\theta}_k | M_k) (2\pi)^{N_k/2} |H_k(\hat{\theta}_k)|^{-1/2} \quad (15)$$

where  $N_k$  is the number of uncertain parameters for model  $M_k$ ,  $\hat{\theta}_k$  is the parameter vector that maximizes  $p(\theta_k | D, M_k)$ , and  $H_k(\hat{\theta}_k)$  is the Hessian matrix of  $-\log(\Pr(D | \hat{\theta}_k, M_k) \Pr(\hat{\theta}_k | M_k))$ . If one assumes that all models are equally plausible and have equal priors, then instead of selecting a model based on the posterior model probability  $p(M_k | D)$ , one can rank the models based on evidence  $p(D | M_k)$ , or log evidence based on the above approximation:

$$\begin{aligned} \log p(D | M_k) &\approx \\ \log p(D | \hat{\theta}_k, M_k) &+ [\log p(\hat{\theta}_k | M_k) + \frac{N_k}{2} \log(2\pi) - \frac{1}{2} \det |H_k(\hat{\theta}_k)|] \end{aligned} \quad (16)$$

In this formulation, the term in brackets on the right is called the log of Ockham factor, where the Ockham factor is a penalty against overparameterization. This equation means that log evidence = log likelihood + log Ockham factor. This is similar to the AIC and other formulations which include a term such as “data fit + bias against parameterization.” The Bayesian Information Criterion (BIC) uses a similar formulation to Equation 16, with the assumption of equal prior probability on each model and vague priors on the parameters:

$$BIC = \log p(D | \hat{\theta}_k, M_k) - \frac{N_k}{2} \log N \quad (17)$$

There are many ways of computing the evidence in Equation 14 and use that to computer the posterior probability of each model in Equation 13. The simplest way to compute evidence is simply to sample from the prior distribution on the parameters  $p(\theta_k | M_k)$  and estimate the mean<sup>10,16</sup>:

$$p(D | M_k) = \int p(D | \theta_k, M_k) p(\theta_k | M_k) d\theta_k \approx \frac{1}{n} \sum_{i=1}^n p(D | \theta_{i,k}) \quad (18)$$

where the samples  $i=1..n$  of  $\theta_{i,k}$  are taken from the prior distribution of the parameters of model k:  $p(\theta_k | M_k)$ . This estimator often suffers from having a large variance and slow convergence to a Gaussian distribution, however, because most non-informative prior samples have small likelihood values when the data are informative about the parameters. An alternative is to use samples from the posterior distribution, usually obtained from Markov Chain Monte Carlo sampling<sup>17-18</sup>, and compute the harmonic mean of the likelihood values as an estimator:

$$p(D | M_k) \approx \frac{n}{\sum_{i=1}^n [p(D | \theta_{i,k})]^{-1}} \quad (19)$$

where this time, the samples  $i=1..n$  of  $\theta_{i,k}$  are from the posterior distribution of the parameters for model k.

Picard et al. [16] present more advanced methods for calculating model evidence, the posterior model probabilities, and Bayesian model averages. Two methods they highlight are the Wolpert method, which is an improvement on the harmonic mean estimate, and reversible jump MCMC. In the reversible jump approach, one calculates the posterior model probabilities for all models “simultaneously” within one Monte Carlo Markov Chain, and the

sampling within the chain involves jumping from one model to another with certain probabilities and sampling the parameters of that model accordingly. Many more details about reversible jump MCMC are provided in Reference 16.

In summary, the different model selection criteria are compared in Table 1.

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.

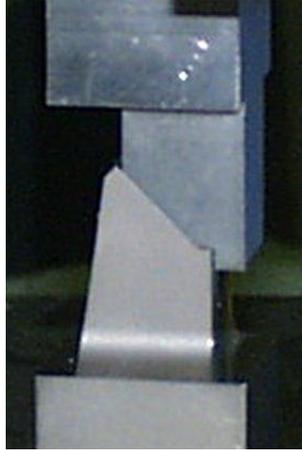
**Table 1: Model Selection Criteria, both Information Theory (AIC, AIC-C, BIC) and Bayesian**

Note that we have formulated these information criteria to be in a format so that the goal is to MAXIMIZE them all (for example, AIC is usually expressed with a -2 multiplying the formulation for historical reasons:

$AIC = -2 * \log p(D | \hat{\theta}_k, M_k) + 2N_k$ . But the formulation in the table allows easy comparison and the goal in all cases is to find a model and associated set of parameters to maximize the information criterion.

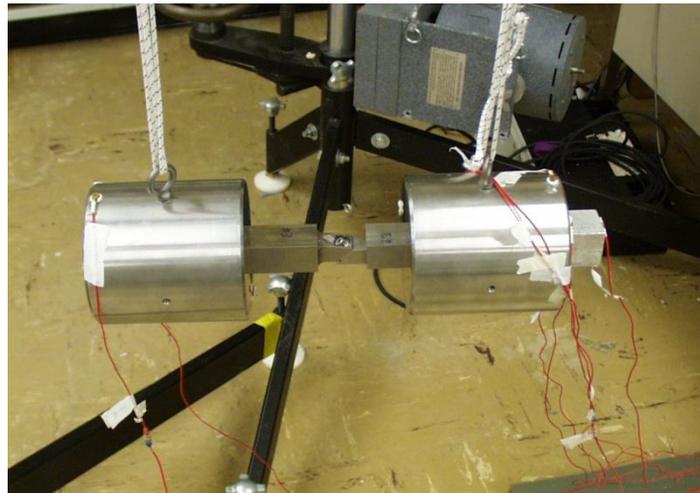
### III. Examples

Our example focuses on the mechanical properties of a bolted joint system, specifically the microslip of the system. The physical system is shown below in Figure 1. The support element is a short leg consisting of two sub-elements as shown in Figure 1. The upper sub-element is solid and fabricated from stainless steel; the lower sub-element is a machined shell structure fabricated also from stainless steel. The two sub-elements are joined with a screw. The screw passes freely through a hole in the mating surface of the lower sub-element (though, under some circumstances, it may bear against one edge of the hole) and it attaches into a threaded hole in the mating surface of the upper sub-element. The torque on the screw is specified so that gross slippage is improbable along the mating surfaces. That is, it should only very rarely happen that the entire mating surface of the upper sub-element is in motion relative to the entire mating surface of the lower sub-element.

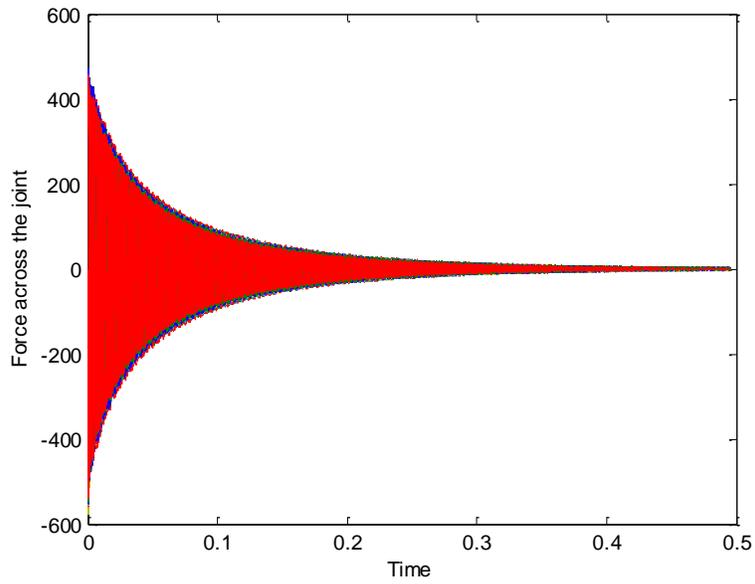


**Figure 1: Support element**

The support element shown in **Figure 1** was tested in a dumbbell configuration as shown in **Figure 2**. The end masses were 30 lbs each and the system was hung from a fixed support using bungee cords. An impulse load was applied to the system by striking one of the end masses with an instrumented hammer and acceleration responses were recorded using accelerometers at either side of the bolted interface. Acceleration responses were obtained from 45 different tests (consisting of 9 different pieces of hardware and repeating each individual test 5 times on each of the 9 pieces). The resulting force time histories (obtained from  $F=ma$ ) are shown in **Figure 3**.

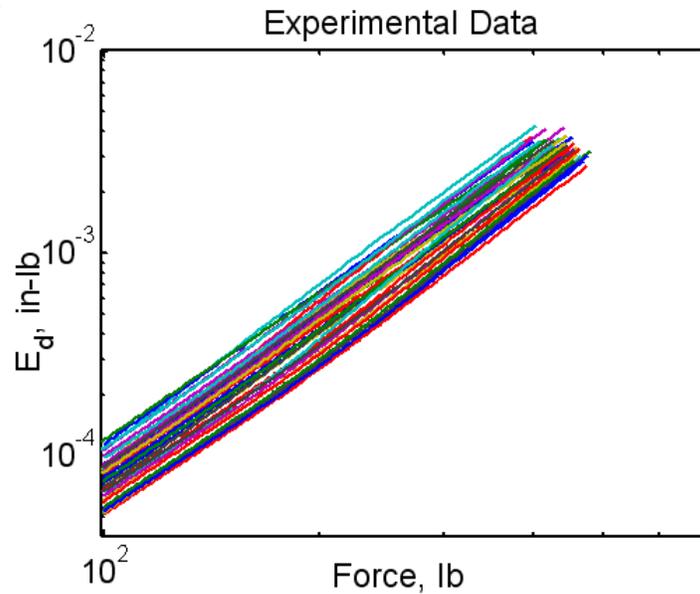


**Figure 2: Dumbbell test configuration**



**Figure 3: Force time histories from 45 tests of hardware shown in Figure 2**

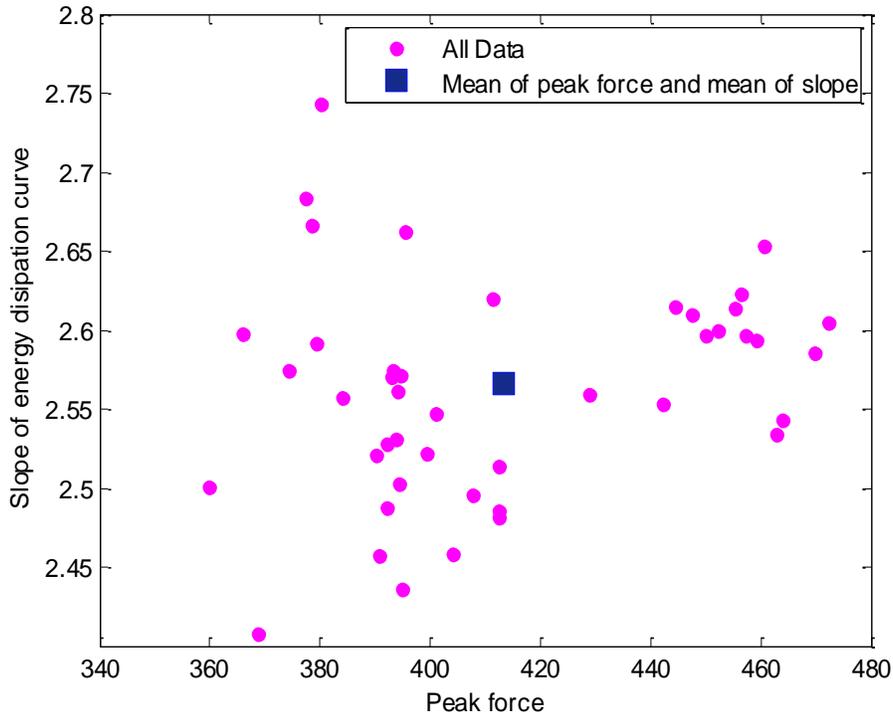
Using a post-processing technique that involves estimating the decay envelope for each of the force time histories shown in **Figure 3**, an estimate of the energy dissipated per cycle as a function of force can be computed. These are shown in **Figure 4**.



**Figure 4: Energy dissipated per cycle as a function of time for each test configuration**

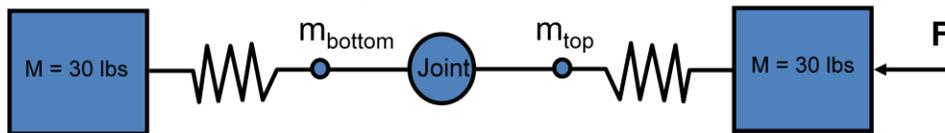
From the curves shown in **Figure 4**, we can calculate the slope of each one of these and this is a representation of the energy dissipation behavior of the system (note that for a linear system, the slope will be equal to 2.0). Also, from **Figure 3**, we can obtain the absolute peak force across the joint. This will be representative of the stiffness in the system. Both quantities are plotted in

**Figure 5** and are used to calibrate the candidate models of the energy dissipation mechanism of the bolted interface. Note that the data points are in pink and the means of the two quantities are shown by the blue point, at a slope of 2.56 and a peak force of 416g.



**Figure 5: Slope of energy dissipated and peak force across the joint**

A simple finite element model (shown in **Figure 6**) was constructed to represent the experimental system shown in **Figure 2**.



**Figure 6: Model of test hardware**

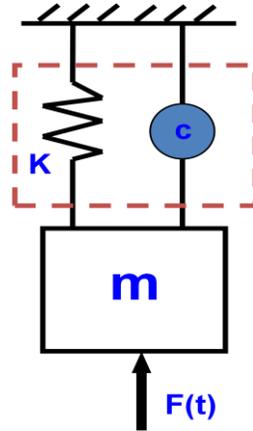
To represent the effects of microslip at a joint, the following mathematical model formulations are explored:

- 1) a simple dashpot and spring model
- 2) a Smallwood model
- 3) an Iwan model.

Each will be briefly described next.

### A. Dashpot & Spring Model

A dashpot represents a damping term proportional to velocity. Dashpot elements combine a viscous friction damper with a simple linear spring (see **Figure 7**).



**Figure 7: Simple schematic of a dashpot and spring model**

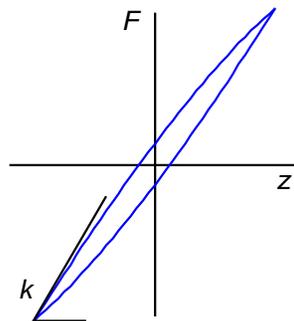
The spring is included to avoid singular stiffness matrices when dashpots are connected without springs. The damping factor is the damping matrix entry. It has units of force·time/length. For a single degree of freedom system with a mass= $M$ , the following equation is satisfied.

$$M\ddot{u} + C\dot{u} + Ku = F(t) \quad (20)$$

Given the mass and  $F(t)$ , the parameters  $C$  and  $K$  need to be estimated. Since this is a linear model of the energy dissipation of a system, it is known that the slope of the energy dissipation versus force curve will always be equal to 2.0.

### B. Smallwood's Model

This model was developed by Smallwood, Gregory and Coleman (2000)<sup>19</sup>, and they used the observed experimental response of joints in systems excited by structural dynamic excitation to justify its form. When a joint in a mechanical system is excited in such a manner that the force across the joint is a harmonic function and the relative displacement across the joint is also a harmonic function, it is observed that the graph of force versus relative displacement always resembles the curve shown in **Figure 8**.



**Figure 8: Force versus relative displacement curve during one cycle of response for system that dissipates energy.**

Smallwood observed that two simple, symmetric mathematical forms govern the variation in force across a joint as a function of relative displacement across the joint during the application of increasing force and the application of

decreasing force. These mathematical forms correspond to the regimes in which the outer segments of the curve in Figure 8 are generated. The expression governing force as a function of displacement while the force and displacement increase (the upper segment of the curve in Figure 8) is:

$$F_u(z) = k(z - z_i) - k_n(z - z_i)^n + F_i \quad (21)$$

where  $z$  is the relative displacement across the joint,  $k$  is the tangent slope of the upper curve at the point where force and displacement are minima,  $k_n$  is the coefficient of the nonlinear component of the restoring force,  $n$  is the exponent of the nonlinear term, and  $z_i$  and  $F_i$  are the relative displacement and force at the lower left point of the restoring force curve. The expression governing force as a function of displacement while the force and displacement decrease (the lower segment of the curve in Figure 8) is:

$$F_d(z) = -k(z_j - z) + k_n(z_j - z)^n + F_j \quad (22)$$

where the terms that appear in Eq. (21) also appear here, except that  $z_j$  and  $F_j$  are the displacement and force coordinates at the upper right point on the restoring force curve. The parameter  $k$  is also the tangent slope of the lower curve at the point where force and displacement are maxima. There is no mass associated with Smallwood's joint model.

The expressions in Eqs. (21) and (22) model the upper and lower segments of the restoring force versus relative displacement curves as a constant, plus a linear term, plus a single nonlinear term. The use of common terms in Eqs. (21) and (22) enables continuity at the end points of the upper and lower curves, and it enables a symmetry between the nonlinear components of the upper and lower curves about a line that joins the end points at the lower left and upper right. The parameters of the upper and lower curves are related by the expression:

$$F_j = k(z_j - z_i) - k_n(z_j - z_i)^n + F_i \quad (23)$$

The relations in Eqs. (21) and (22) provide an accurate representation of the force across the mechanical joint as a function of the relative displacement across the joint for environments that excite narrowband response, i.e., environments for which there is one and only one positive local maximum or minimum of the force between each pair of zero crossings of the force time history, and every extremum of one sign is followed by an extremum with the opposite sign between the next two zero crossings of the force curve. The model defined in Eqs. (21) through (23) can be used to approximate the energy dissipated in a joint during one cycle of system response. The energy dissipated is established by integrating Eq. (21) from  $z_i$  to  $z_j$ , and adding to that the integral of Eq. (22) from  $z_j$  to  $z_i$ . That quantity is:

$$E = k_n \left( \frac{n-1}{n+1} \right) (z_j - z_i)^{(n+1)} \quad (24)$$

Given  $n$ , Eq. (21) can be inverted to solve for the parameter  $k_n$  as:

$$k_n = E \left( \frac{n+1}{n-1} \right) (z_j - z_i)^{-(n+1)} \quad (25)$$

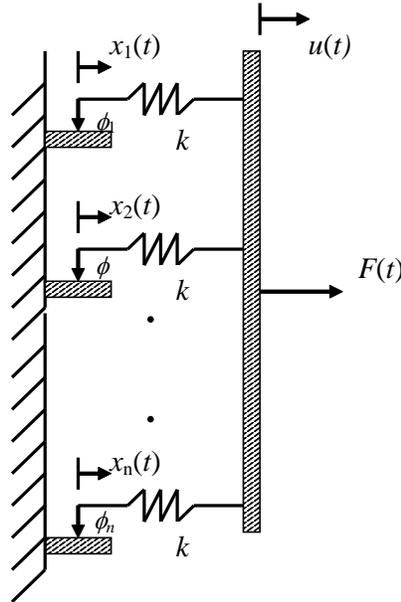
A value of the parameter  $k_n$  can be identified for each input force level corresponding to a difference  $(z_j - z_i)$  and the value of energy dissipated at that force level. Each test was run at five force levels, therefore, five  $k_n$ 's can be identified for each experimental setup. Given  $k_n$ , Eq. (23) can be inverted to establish an estimate for the parameter  $k$ . The formula for  $k$  is:

$$k = \frac{(F_j - F_i) + k_n(z_j - z_i)^n}{(z_j - z_i)} \quad (26)$$

The three parameters,  $n$ ,  $k$ , and  $k_n$ , completely characterize Smallwood's model for restoring force across a mechanical joint. Therefore, when experimental data are available, if the three parameters can be inferred from the data, then the Smallwood model representing the joint can be calibrated to the experiment.

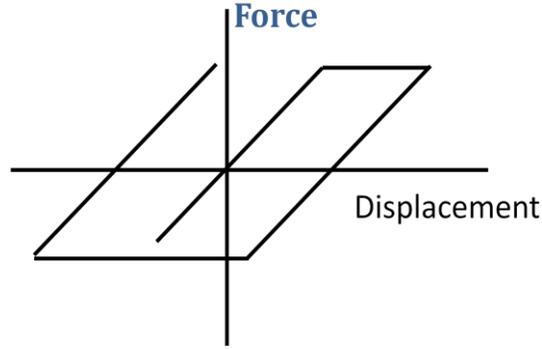
### C. Iwan's Model

The model considered here was first used to simulate the behavior of hysteretic systems. It was proposed by Iwan (1966)<sup>20</sup>, and he cited experimental evidence to justify its use. The model schematic is shown in Figure 9. It is a parallel sequence of sub-elements, each of which is a linear elastic spring in series with a rigid, perfectly plastic slider element. All the springs have common spring constant  $k$ . The  $(i)$ th slider has slipping strength  $\phi_i$ . The accumulated deformation of the  $(i)$ th slider element is denoted  $x_i(t)$ . The external force applied to the Iwan element is  $F(t)$  and its external deformation is  $u(t)$ . There is no mass associated with the Iwan element. Each sub-element is also known as a Jenkins element.



**Figure 9: Schematic of Iwan's parallel-series model**

When a force is applied to a single Jenkins element the deformation is linear, with spring constant  $k$ , until the strength of the rigid, perfectly plastic slider is achieved, then sliding commences and continues until the direction of the force is reversed. The force versus displacement history of a single Jenkins element might resemble the curve shown in **Figure 10**.



**Figure 10: A possible force versus displacement history for a single Jenkins element, a sub-element of the Iwan model**

The equation relating force to deformation across an Iwan element is:

$$F(t) = \sum_{i=1}^n q_i(t) \quad (27)$$

where  $q_i(t)$ ,  $i = 1, \dots, n$ , is the force in the ( $i$ )th Jenkins element. Each sub-element force,  $q_i(t)$ , is governed by:

$$q_i(t) = \begin{cases} k(u(t) - x_i(t)) & |u(t) - x_i(t)| < \phi_i / k \\ \phi_i \operatorname{sgn}(\dot{u}) & |u(t) - x_i(t)| = \phi_i / k \end{cases} \quad (28)$$

where  $\operatorname{sgn}(\cdot)$  denotes the signum function. The quantity  $x_i(t)$  is the absolute slippage that has accumulated in the ( $i$ )th Jenkins element, and it is governed by:

$$\dot{x}_i(t) = \begin{cases} 0 & |u(t) - x_i(t)| < \phi_i / k \\ \dot{u}(t) & |u(t) - x_i(t)| = \phi_i / k \end{cases} \quad (29)$$

This equation reflects the rigid, perfectly plastic characteristic of the slider. That is, slippage is identically zero when the load on the slider is below a threshold, and the slippage rate of the slider equals the externally imposed deformation rate once slippage commences. In Segalman<sup>21</sup>, there is a description of a plausible form for the distribution of slippage strengths of the Jenkins elements in an Iwan element used to simulate the behavior of a lap joint. The form of the distribution is:

$$\rho(\phi) = R\phi^\chi \quad \phi \geq 0 \quad (30)$$

where  $R$  is a positive constant, and  $\chi$  is a constant in the interval  $(-1, 0)$ . In terms of the parameter  $\chi$ , the slope of the energy dissipated per cycle versus external force curve plotted on a log-log graph equals  $3 - \chi$ . An argument regarding the practicality of this expression led Segalman to modify it, slightly. The form used is:

$$\rho(\phi) = R\phi^\chi + S\delta(\phi - \phi_{\max}) \quad \phi \geq 0 \quad (31)$$

where  $S$  is a positive constant and  $\phi_{\max}$  is a practical limit to the strength of any Jenkins element. The value of  $S$  is selected to insure that the slope of the restoring force versus deformation curve achieves a value of zero when global slippage occurs at the joint. The four parameters,  $R$ ,  $S$ ,  $\chi$ , and  $\phi_{\max}$ , enable the specification of the Iwan model. (When the number of Jenkins elements to be used in computations is specified, these parameters imply a value for the Jenkins element stiffness,  $k$ .) Therefore, when experimental data are available, if the four parameters can be inferred from the data, then the Iwan model representing the joint can be calibrated to the experiment.

## IV. Results

The following are preliminary results of this analysis.

### A. Frequentist Results

This section presents the results of nonlinear least squares and maximum likelihood. Based on the maximum likelihood result, we were able to calculate some of the information theoretic criteria such as AIC.

#### Nonlinear least squares

The nonlinear least squares optimization, as described in section II.A, was performed using the nl2sol optimization method in DAKOTA<sup>22</sup>. The nl2sol method is a Levenberg-Marquardt algorithm with Gauss-Newton approximations to the Hessian matrix. It treats all of the residuals separately. Note that in this case, we have 45 residuals from 45 experimental data points, where each data point has two response quantities: peak force across the joint and slope of energy dissipated. The results are shown in Tables 2-3 below:

Iteration	Dashpot			Smallwood				Iwan				
	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

**Table 2. Optimal parameter results using least squares methods**

Table 2 shows 20 separate least squares optimization runs for each model, where each of the 20 starting points was a randomly chosen location in the parameter space. The reason we performed 20 optimizations was that the Dashpot and Iwan models exhibited multiple local minima, depending on the starting point. This is to be expected with gradient-based methods which are local optimizers. There is some interesting behavior here: starting from different initial points often resulted in SSE estimates which were quite different, and the resulting optimal parameters were also quite different. It is also interesting to note that the behavior of the Smallwood model was quite stable: it converged to the same point all twenty times, regardless of the initial starting point of the optimization.

To compare the performance across all twenty of these results, we looked at a few statistics on the population of twenty solutions for each model, as shown in Table 3. In Table 3, we can see that if we were comparing the three models based on the average SSE from the 20 runs, the Smallwood model is the best performer by far. Similarly, the Smallwood model has the smallest maximum SSE. However, both the Dashpot and Iwan models were able to find a minimum SSE over the twenty runs that is better than the Smallwood model SSE. Thus, if we were to take the “best” SSE as the minimum SSE from the twenty runs, we would say that the Iwan and Dashpot models are comparable, but the Smallwood model is a little worse.

	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

**Table 3. Summary statistics using least squares methods**

Finally, in the case of the Dashpot and Smallwood models, we looked at the correlations of the optimal parameter values and the SSE. We found some results as expected, for example the “k” value in the Dashpot model was highly correlated with the SSE, in part because k influences the peak force. Note that in the least squares formulation, we had a total of 90 residual terms, 45 for peak force and 45 for the slope of the dissipated energy. These were not weighted (although that is something we would like to investigate), so in these results, the peak force residuals had a much higher magnitude (in the hundreds) vs. the residuals from the slope terms (which were usually less than one). Thus, the influence of the peak force terms was higher in the optimization.

### Maximum Likelihood Estimation

In contrast to the least squares formulation where the peak force residuals were added to the slope residuals for a total error sum of squares, in the likelihood formulation we multiply the likelihood obtained for the peak force objective with the likelihood obtained for the slope objective to obtain a total likelihood. That is, in log space, we add the log likelihood from the peak force with the log likelihood for the slope:

$$\begin{aligned}
LogLikelihood_{peak}(\theta) &= -\frac{45}{2} \log(\sigma_{Peak}^2) - \frac{(y_{peak,i} - g(\theta))(y_{peak,i} - g(\theta))}{2\sigma_{Peak}^2} \\
LogLikelihood_{slope}(\theta) &= -\frac{45}{2} \log(\sigma_{Slope}^2) - \frac{(y_{Slope,i} - g(\theta))(y_{Slope,i} - g(\theta))}{2\sigma_{Slope}^2} \\
NegLogLikelihoodTotal(\theta) &= -1 * [LogLikelihood_{peak}(\theta) + LogLikelihood_{slope}(\theta)]
\end{aligned} \tag{32}$$

In the past, we have had difficulty properly calibrating the error terms, in this case,  $\sigma_{peak}$  and  $\sigma_{slope}$ . So, we first approximated  $\sigma_{peak}$  and  $\sigma_{slope}$  by the standard deviations of the spread of the actual data shown in Figure 5. We fixed  $\sigma_{peak}$  and  $\sigma_{slope}$  at the values of 32.8 and 0.0734 according to the experimental data, and we only optimized the likelihood for the parameters (this is the case of constant sigma in Table 4 below). Then, we performed the full optimization, where the optimization parameters included all of the model parameters  $\theta$  and  $\sigma_{peak}$  and  $\sigma_{slope}$ . The results of these optimizations are shown in Table 4 below. The results are very similar and consistent between the constant error term case and the case where we optimized the sigma terms: we were pleased to see that the optimization resulted in  $\sigma_{peak}$  and  $\sigma_{slope}$  that were very close the variation in the experimental data.

MAXIMUM LIKELIHOOD RESULTS ASSUMING CONSTANT ERROR TERMS					
DASHPOT		SMALLWOOD		IWAN	
Parameter	Optimal Value	Parameter	Optimal Value	Parameter	Optimal Value
k	6.05E+06	klin	2.13E+06	chi	-5.35E-01
mu	35.00	knon	1.70E+06	phi_max	7.45E-05
		n	1.466	R	1.95E+07
				S	5.43E+06
NEG LOG LIKE	339.90		85.05		83.84
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

**Table 4: Maximum likelihood parameter estimates obtained in two cases: fixing the error terms and optimizing the parameters (upper) and optimizing both error terms and parameters (lower).**

We used a global, non-derivative based method called DIRECT<sup>15</sup> to optimize the likelihood. The optimization method worked well. Even in the 6-dimensional problem of solving for the four parameters for the Iwan model and the two error terms, the global optimization converged in 163 function evaluations. The bounds for the  $\sigma_{peak}$  and  $\sigma_{slope}$  terms were [20,40] and [0,0.1], respectively, in the optimization. The bounds for the parameter values were determined by expert judgment and were fairly large.

Note that the negative log likelihood values in Table 4 are positive. This is due in part to the fact that partly to the fact that the first term in the log likelihood of the peak force:

$$-\frac{45}{2} \log(\sigma_{Peak}^2) \approx -\frac{45}{2} \log(32.8^2) \approx -157$$

This term is then added to the residual squared term to get the log likelihood of the peak force, and the log likelihood of the peak force is added to the log likelihood of the slope. However, the magnitude of this first term does dominate some calculations. When it is added to the others, the -157 makes the sum of the likelihoods negative, but since we are minimizing the negative log likelihood, the negative log likelihood then becomes positive.

Based on the results in Table 4, if we were selecting a model purely based on the one that maximized the likelihood of the data given a model choice, we would choose the models in the following order: Iwan, Smallwood, and Dashpot. The likelihood values of these models (for the case where we are optimizing the error terms) are approximately -84.7, -85.4, and -281.9, respectively. Thus, the Iwan model has the largest likelihood value.

However, if we consider the information theoretic criteria which penalize for overfitting and value model parsimony (see Table 1), we get the following results as shown in Table 5:

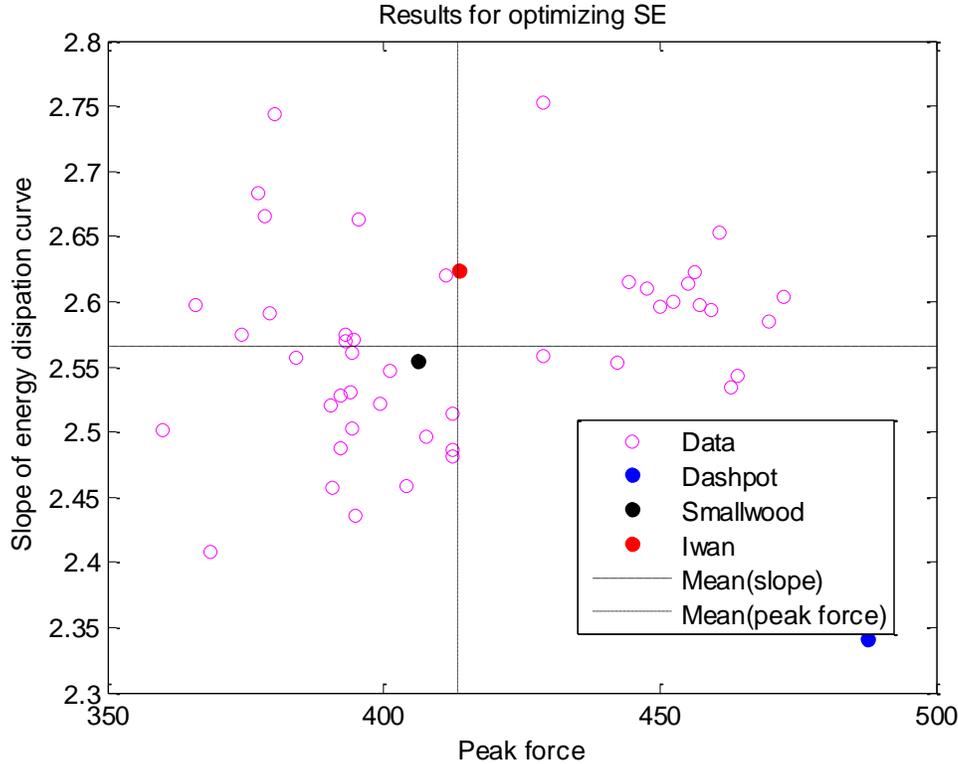
	Dashpot	Smallwood	Iwan
<b>Log Likelihood</b>	-281.86	-85.42	-84.74
<b>AIC</b>	-285.86	-90.42	-90.74
<b>AIC-C</b>	-286.10	-90.78	-91.19
<b>BIC</b>	-290.86	-96.67	-98.24

**Table 5: Preferred Models, shown in yellow, according to various Information Criteria**

As mentioned, if we are considering the log likelihood as the sole criterion, we would choose the Iwan model. However, since the Smallwood model has nearly comparable performance as the Iwan model and has one less parameter, it does a bit better according to the information theoretic metrics of AIC, AIC-C, and BIC, since it is more parsimonious. These models provide a nice example where penalizing for overfitting does make a difference in model selection choice. As a final verification of the MLE process, we ran the models with the optimal parameter settings shown in Table 4, and determined the peak force across the joint and the slope of the energy dissipated. These quantities are displayed in Table 6 and Figure 10, and again show that the Smallwood and Iwan model yield predictions closest to the data.

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

**Table 6. Model Performance using Optimal MLE Parameters**



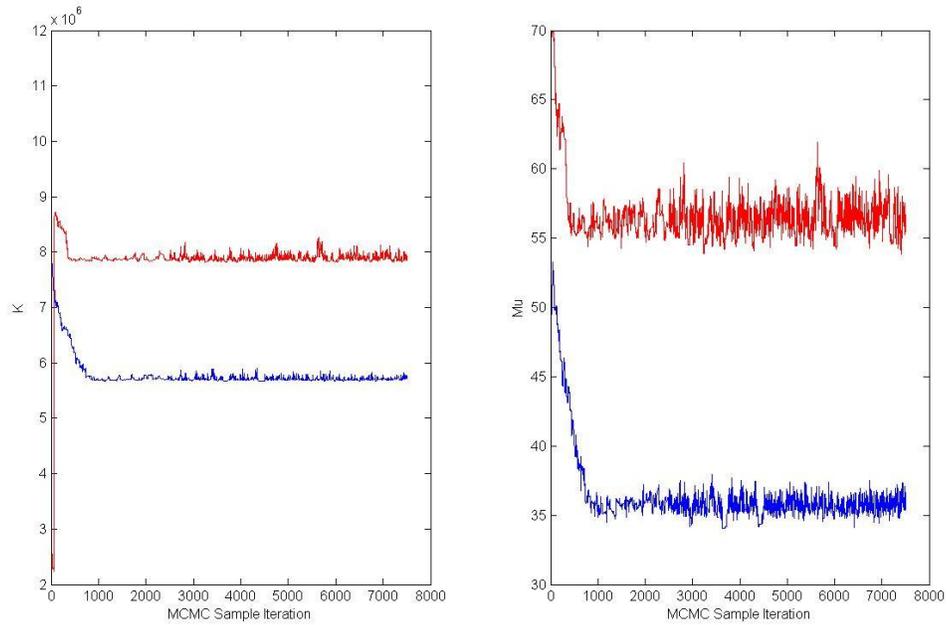
**Figure 10: Results of models at MLE values for optimal parameters**

## B. Bayesian Results

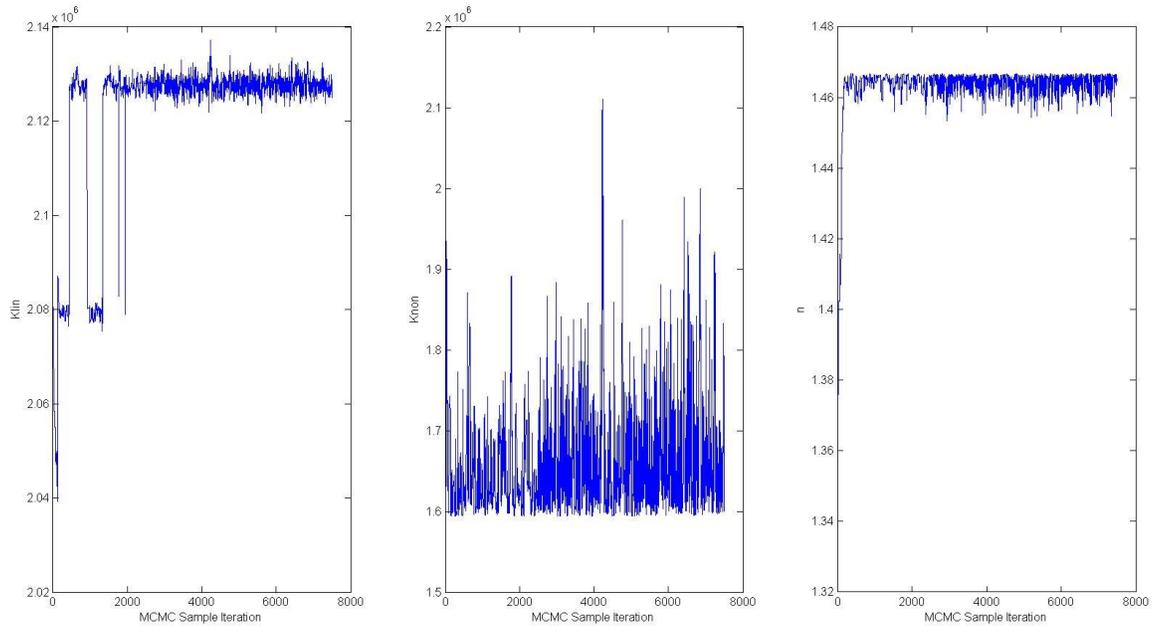
This section presents the results of Bayesian calibration. We were not able to generate the posterior model probabilities and evidence (Equations 13-14), but we hope to do that soon. At this point, our results demonstrate the calculation of the posterior estimates on the parameters, given the priors and the likelihood. Note that similar to the MLE approach presented above, the likelihood formulation had to be extended to account for two objectives: peak force and slope of the energy dissipated.

We used the Gaussian Process Models for Simulation Analysis (GPMSA) code developed at Los Alamos National Laboratory to calculate the posterior parameter estimates. GPMSA is a code that provides the capability for Bayesian calibration: given a computational simulation model with unknown parameters and prior distributions on those parameters, along with observational data (e.g. from experiments), GPMSA uses a Monte Carlo Markov Chain (MCMC) algorithm to produce posterior distributions on the parameters. When the computational simulation is then executed using samples from the posterior distributions of the parameters, the simulation model results that are produced are consistent with (“agree with”) the experimental data. Note that typically, GPMSA constructs an emulator from simulation runs collected at various settings of input parameters. The emulator is a statistical model of the system response, and it is used in the MCMC process. However, in this analysis, the MCMC ran the simulation models directly: there is no Gaussian Process emulation.

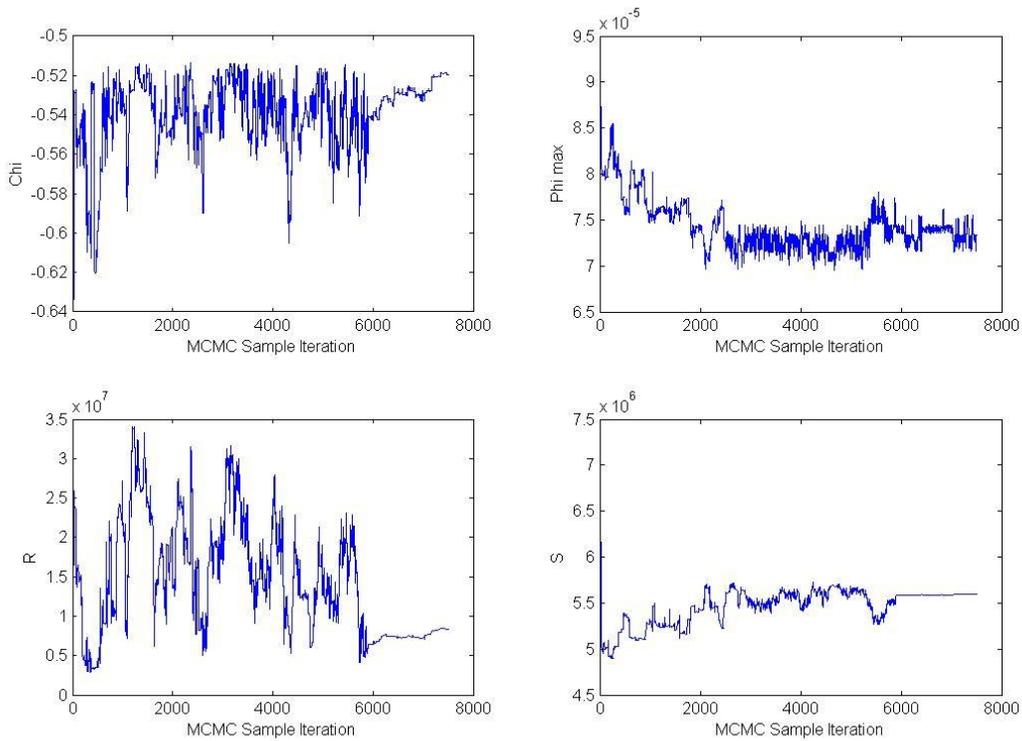
The posterior MCMC samples, based on a chain of 7500 samples, are shown below in Figure 11-13 for the Dashpot model, the Smallwood model, and the Iwan model. Note that we performed two runs of the Dashpot model with different starting points, thus there are two sets of MCMC results.



**Figure 11: MCMC History of Parameter Samples for the Dashpot model, based on two starting points (red vs. blue)**

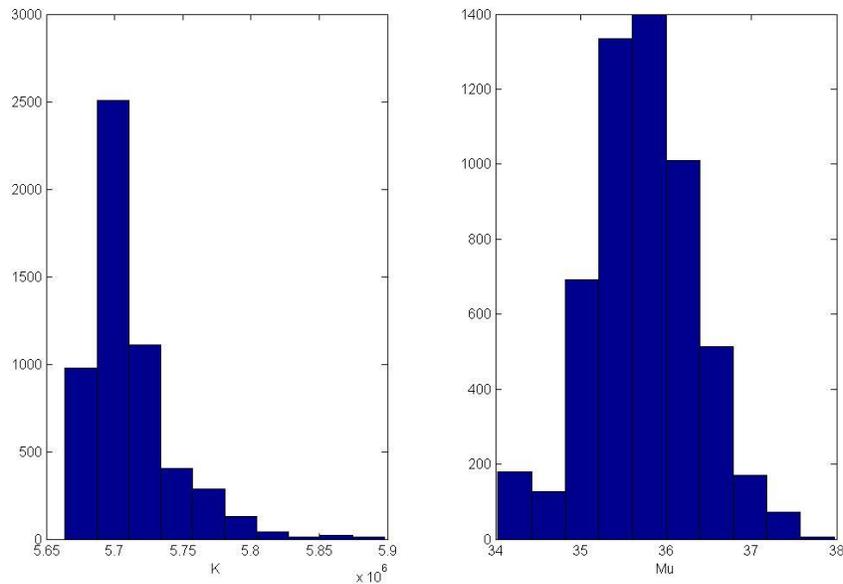


**Figure 12: MCMC History of Parameter Samples for the Smallwood model**

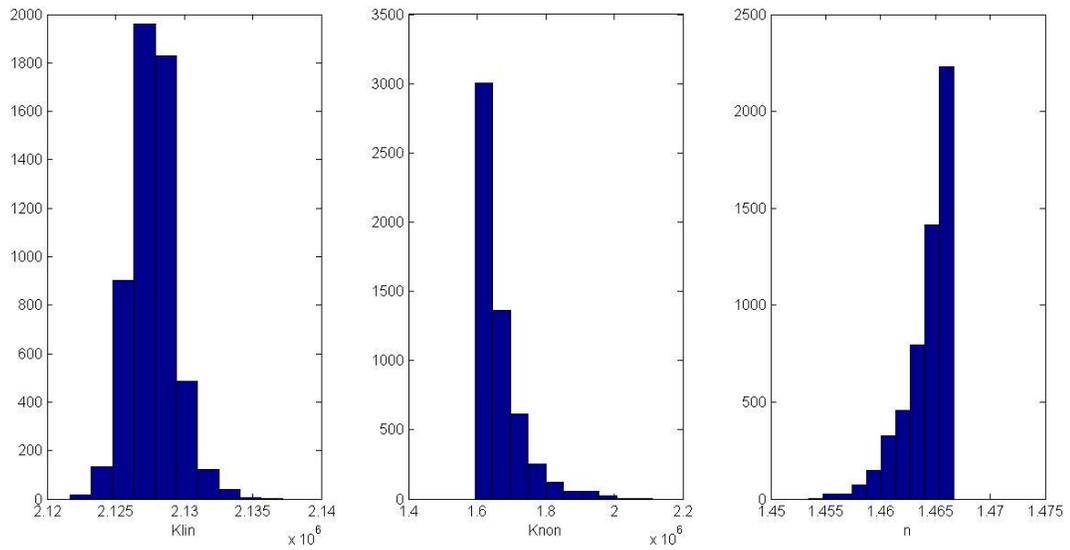


**Figure 13: MCMC History of Parameter Samples for the Iwan model**

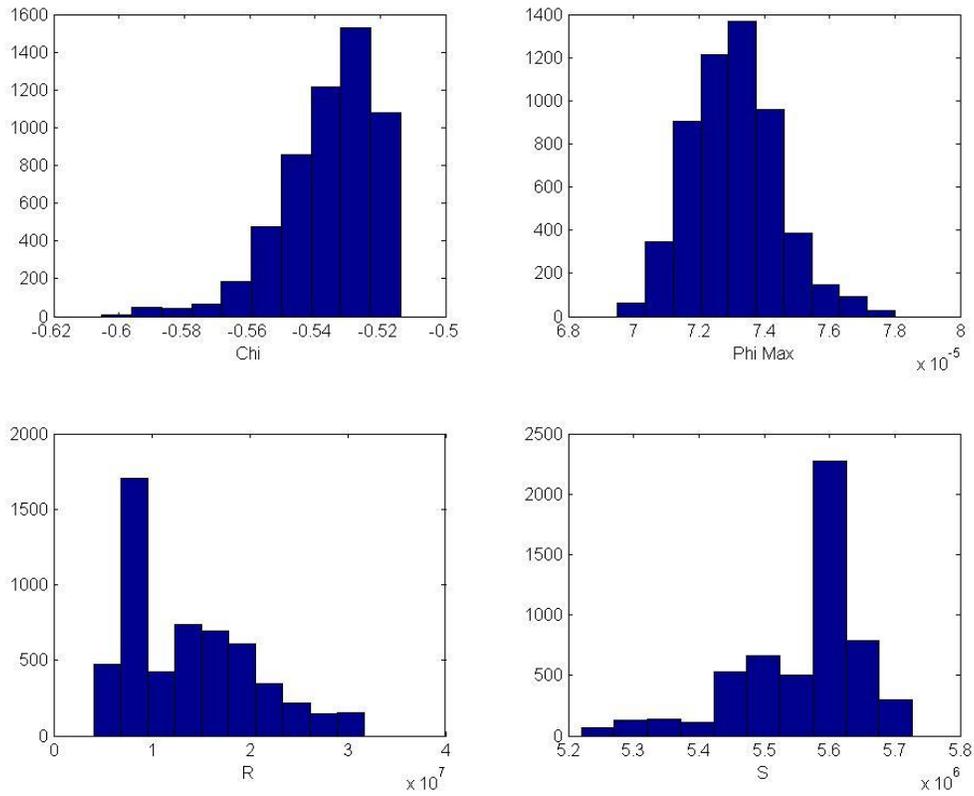
We truncated the first 2000 samples from the data sets due to a “burn-in” period. Then, we took the resulting means of the remaining samples, as shown in Table 7. The posterior means are fairly consistent with the NLLS and MLE approaches, especially for the Smallwood and Iwan models. Finally, histograms of the posterior samples are shown in Figures 14-16.



**Figure 14: Histograms of Posterior Distributions on K and Mu parameters for Dashpot model**



**Figure 15: Histograms of Posterior Distributions on  $K_{lin}$ ,  $K_{non}$ , and  $n$  parameters for Smallwood model**



**Figure 16: Histograms of Posterior Distributions on  $\chi$ ,  $\Phi_{max}$ ,  $R$ , and  $S$  parameters for Iwan model**

	<b>Dashpot</b>			
	<b>k</b>	<b>mu</b>		
MCMC 1	5.71E+06	35.74		
MCMC 2	7.89E+06	56.32		
	<b>Smallwood</b>			
	<b>klin</b>	<b>knon</b>	<b>n</b>	
MCMC 1	2.13E+06	1.66E+06	1.46	
	<b>lwan</b>			
	<b>Chi</b>	<b>Phi Max</b>	<b>R</b>	<b>S</b>
MCMC 1	-0.54	7.31E-05	1.37E+07	5.56E+06

**Table 7. Means of Posterior Parameter Samples, using Bayesian calibration**

## V. Conclusion

This paper has presented three main approaches to model calibration and model selection: classical least squares methods, information theoretic criteria, and Bayesian approaches. These approaches are well-documented in the literature, however we found them non-trivial to apply to a moderately complex structural dynamics model of a bolted joint where we were trying to match to two sets of experimental data: peak force across the joint and slope of the energy dissipated. For the nonlinear least squares approach, two of the models had multiple optima that were not similar. A weighted least squares is probably the best method to ensure the optimization is trying to reduce the SSE contributed by both objectives relatively equally. For the maximum likelihood approach, the likelihood had to be carefully formulated to include both objectives. We had good results with a global optimization method to determine both optimal model parameter values and to determine the values of the error terms. Using optimization to solve for the error terms gave results very close to the experimental variation we saw in the data, confirming that our process was working well. The information criteria demonstrated that we had a situation where selecting a model based purely on likelihood could result in a different model choice than selecting a model according to an information criterion, for example. Finally, the Bayesian calibration appeared to work reasonably well. An overall comparison of the calibrated parameters using the three methods is shown in Table 8:

	<b>Dashpot</b>			
	<b>k</b>	<b>mu</b>		
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		
	<b>Smallwood</b>			
	<b>klin</b>	<b>knon</b>	<b>n</b>	
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	
	<b>lwan</b>			
	<b>Chi</b>	<b>Phi Max</b>	<b>R</b>	<b>S</b>
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

**Table 8. Optimal parameter values as calculated by nonlinear least squares (NLLS), maximum Likelihood estimation (MLE), and Bayesian calibration (MCMC)**

Note that the comparison in Table 8 is just a summary, and doesn't include details. For example, in Table 8, we report the mean of the 20 optimization runs we performed for the least squares method, and we only report the mean of the posterior samples calculated by MCMC in the Bayesian approach. However, we note the similarity in the results, partly due to the central and critical role that the likelihood plays in all of these calibration approaches. We are still in the process of post-processing the posterior parameter estimates to calculate evidence and posterior model probabilities to compare with the least squares and maximum likelihood approaches. Our concern is that all of the model selection criteria are essentially based on the likelihood, and they will tend to prefer the same model unless the models are substantially different in number of parameters and complexity with very limited data.

### Acknowledgments

The authors would like to thank the Validation, Verification, and Uncertainty Quantification (VU) program under the Department of Energy's Nuclear Energy Advanced Modeling and Simulation (NEAMS) program. We would also like to acknowledge discussion and collaboration with Rick Picard at Los Alamos National Laboratory.

### References

- <sup>1</sup>Draper, N. R. and H. Smith, *Applied Regression Analysis*, John Wiley & Sons, NY, NY 1998.
- <sup>2</sup>Burnam, K. P. and D. Anderson. *Model Selection and Multi-Model Inference*. Springer, 2<sup>nd</sup> edition, 2002.
- <sup>3</sup>Wasserman, L. *Bayesian Model Averaging and Model Selection*. Carnegie Mellon University Technical Report 666, 1997.
- <sup>4</sup>Beck, J. L. and K.V. Yuen. "Model Selection using Response Measurements: Bayesian Probabilistic Approach," *Journal of Engineering Mechanics*, 130, 192-203, February 2004.
- <sup>5</sup>Raftery, A.E. (1995). *Bayesian model selection in social research (with Discussion)*. *Sociological Methodology*, 25, 111-196.
- <sup>6</sup>Seber, G. A. F. and C. J. Wild, *Nonlinear Regression*, Wiley & Sons, 2003.
- <sup>7</sup>Aster, R. C., Borchers, B., and C. H. Thurber. *Parameter Estimation and Inverse Problems*. Elsevier Academic Press, 2005.
- <sup>8</sup>Rooney, W. C. and L. T. Biegler, "Design for Model Parameter Uncertainty Using Nonlinear Confidence Regions", *Process Systems Engineering*, Vol. 47, pp. 1794—1804, 2001.
- <sup>9</sup>Vugrin, K. W, L. P. Swiler, R. M. Roberts, N. Stuckey-Mack, and S.P. Sullivan. "Confidence Region Estimation Techniques for Nonlinear Regression: Three Case Studies." *Water Resources Research*, Vol. 43, WO3423, 2006.
- <sup>10</sup>Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999), "Bayesian Model Averaging: A Tutorial," *Statistical Science* 14 382-417.
- <sup>11</sup>Cheung, S.H. and J.L. Beck. "Bayesian Model Updating using Hybrid Monte Carlo Simulation with Application to Structural Dynamics Models with Many Uncertain Parameters," *Journal of Engineering Mechanics*, 135, 243-255, April 2009.
- <sup>12</sup>Claeskens, G. and N. Hjort. *Model Selection and Model Averaging*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2008.
- <sup>13</sup>Raftery, A. and Zheng, Y. *Long-Run Performance of Bayesian Model Averaging*. Technical Report no. 433, Department of Statistics, University of Washington.
- <sup>14</sup>Raftery, A.E. and Zheng, Y. (2003). *Discussion: Performance of Bayesian Model Averaging*. *Journal of the American Statistical Association*, 98, 931-938.
- <sup>15</sup>Gablonsky, J. M., Aug. 1998. An implementation of the direct algorithm. Tech. Rep. Technical Report CRSCTR98-29, Center for Research in Scientific Computation, North Carolina State University, Chapel Hill, NC.
- <sup>16</sup>Picard, R.R., Williams, B.J., Swiler, L.P., Urbina, A., and R.L. Warr. *Multiple Model Inference With Application to Uncertainty Quantification for Complex Codes*. Los Alamos Laboratory Technical Report LA-UR- 10-06382, Sept. 2010.

<sup>17</sup>Gamerman, D. (1997), *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Chapman and Hall/CRC, Boca Raton.

<sup>18</sup>Gilks, W.R., S. Richardson, and D.J. Spiegelhalter (1996). *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC, Boca Raton, 1996.

<sup>19</sup>Smallwood, D., Gregory, D., Coleman, R., 2000, "Damping Investigations of a Simplified Frictional Shear Joint," Proceedings of the 71st Shock and Vibration Symposium, SAVIAC, The Shock and Vibration Information Analysis Center.

<sup>20</sup>Iwan, W.D. A distributed-element model for hysteresis and its steady-state dynamic response, *Journal of Applied Mechanics*, Vol. 33, 4, pp. 893-900, 1966.

<sup>21</sup>Segalman, D. J. *A Four-parameter Iwan Model for Lap-type Joints*. Technical Report SAND2002-3828, Sandia National Laboratories, Albuquerque, New Mexico, USA, 2002.

<sup>22</sup>Adams, B.M., Bohnhoff, W.J., Dalbey, K.R., Eddy, J.P., Eldred, M.S., Gay, D.M., Haskell, K., Hough, P.D., and Swiler, L.P., "DAKOTA, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 5.0 User's Manual," Sandia Technical Report SAND2010-2183, December 2009. Updated December 2010 (Version 5.1)