SAND 2011-XXXX C

SAND2011-7204C

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

Laura P. Swiler

Sandia National Laboratories
P.O. Box 5800, Albuquerque, NM, 87185-1318
lpswile@sandia.gov, 505-844-8093

Brian Williams
Statistical Sciences Group
Los Alamos National Laboratory,
PO Box 1663, MS F600
Los Alamos, NM 87545
brianw@lanl.gov;505-667-2331

Presentation at the NEAMS PI Meeting
Argonne National Labs, October 11-13, 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.

Overview

- Introduction of methods
  - Calibration
  - Model Selection
  - Inference
- Case study with R7
- Results
- Summary

Definitions

- Model: 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 multiple models.

Model Calibration: Frequentist Approaches

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

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

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:

Gaussian Likelihood Function

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

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

Assuming i.i.d. error:

$$y_i = g(x_i;\theta) + \varepsilon_i$$
$$\varepsilon_i \sim N(0,\sigma^2)$$

Sandia National Laboratories

Model Calibration: Bayesian Approaches

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

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

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

$$posterior \propto likelihood * prior$$

Again, we assume a Gaussian likelihood:

$$p(D \mid \theta) = L(\theta) = \prod_{i=1}^{n} \frac{1}{2\sqrt{\pi\sigma}} e^{-\left[\frac{(D_i - g(\theta)^2}{2\sigma^2}\right]} \qquad y_i = g(x_i;\theta) + \delta(x_i) + \varepsilon_i$$

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 over-fitting
- Information theory based on Kullback-Leibler distance.

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

- 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:

$$AIC = -2\log(Lik(\hat{\theta} \mid x)) + 2K$$

Goodness of fit

Penalty for number of model parameters K

Sandia National Laboratories

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:

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

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:

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

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.

Sandia National Laboratories

Model Selection Criteria

Model Selection Criterion	Expression
Akaike (AIC)	$AIC = -2\log p(D \mid \hat{\theta}_k, M_k) + 2N_k$
Akaike corrected (AIC-c)	$AIC_c = -2\log p(D \mid \hat{\theta}_k, M_k) + 2\left(N_k + \frac{N_k(N_k+1)}{N - N_k - 1}\right)$
Akaike-Schwarz Bayesian (BIC)	$BIC = -2\log p(D \mid \hat{\theta}_k, M_k) + N_k \log N$
Deviance (DIC): Hierarchical Modeling Generalization of AIC	$D(\theta) = -2\log p(D \mid \theta_k, M_k) + C$ $\overline{D} = E_\theta[D(\theta)]$ $effective\_parms = \overline{D} - D(\overline{\theta})$ $DIC = \overline{D} + effective\_parms$
Bayesian model selection	Log evidence = $\log p(D \mid M_k)$ 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

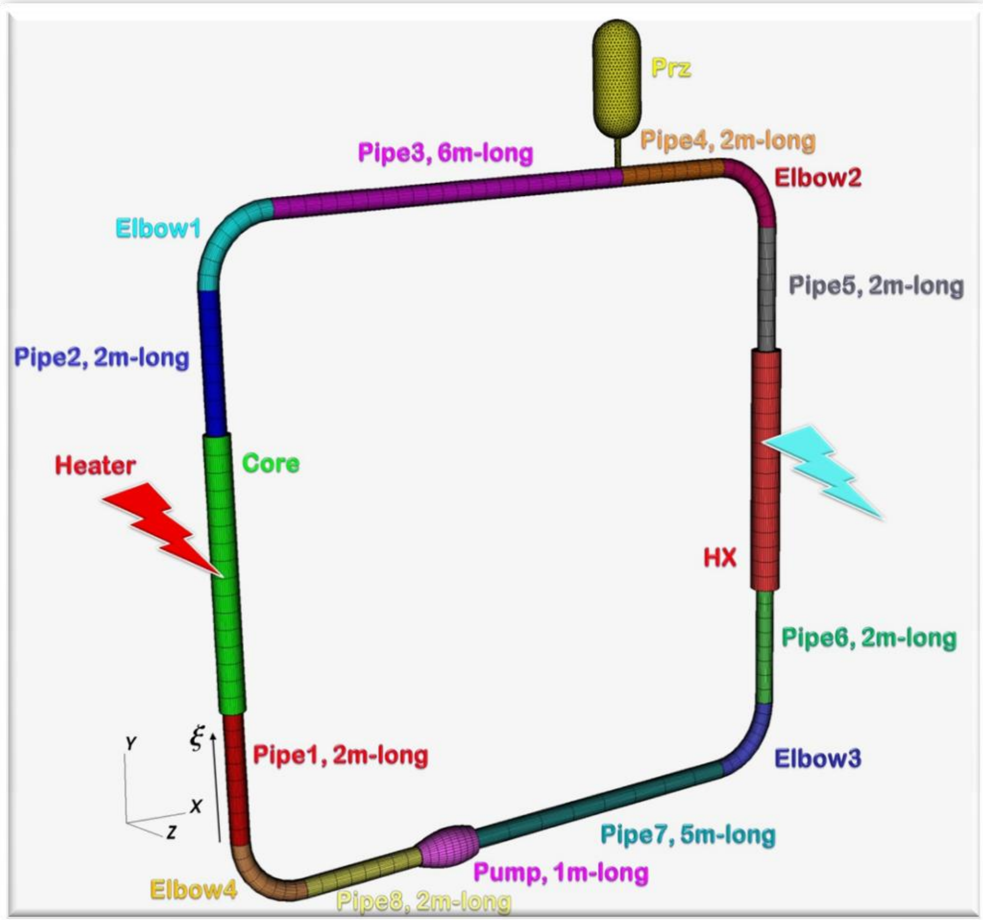
Bayesian approach: For a particular model, one has an entire posterior distribution on parameters θ which can be propagated through the simulation model to generate posterior realizations of the responses.

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):

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

Sandia National Laboratories

Case study: R7 Virtual Reactor



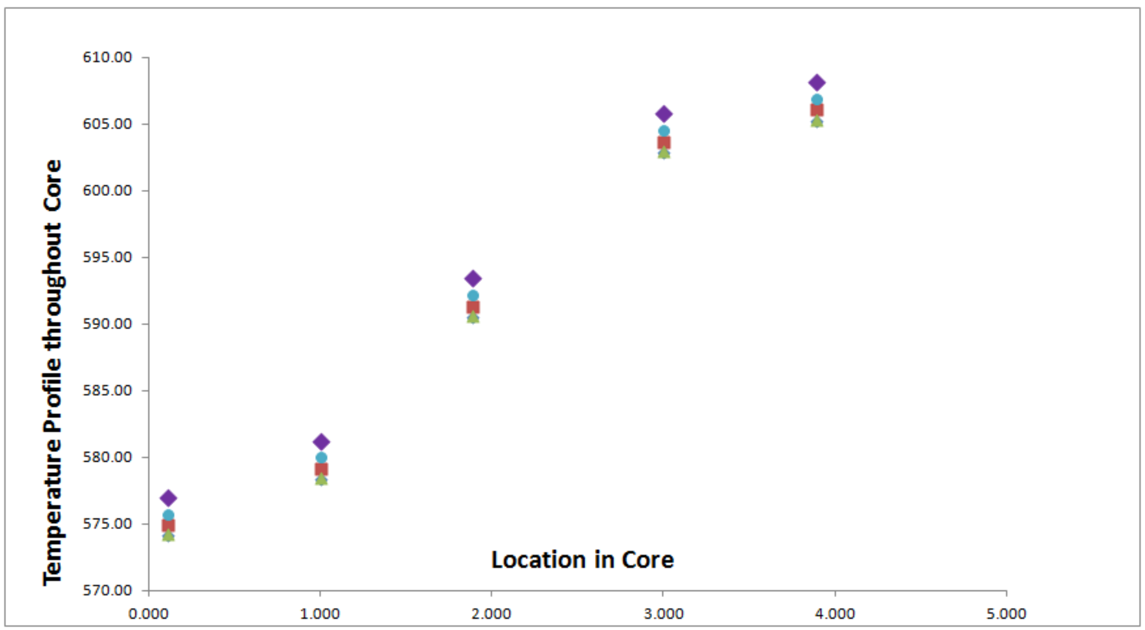
- Simple thermal-hydraulics loop that represents a simplified plant
- The loop is 10m tall and 10m wide.
- The loop has 8 pipes, 4 elbows, a pump, the core, a heater and heat exchanger (HX), a pressurizer, etc.
- The working fluid is water at high temperature and pressure, using single phase flow.
- The power output of this reactor is set to a nominal value of 1.25 MW.

The purpose of this model is to **demonstrate** the calibration and model selection methods.

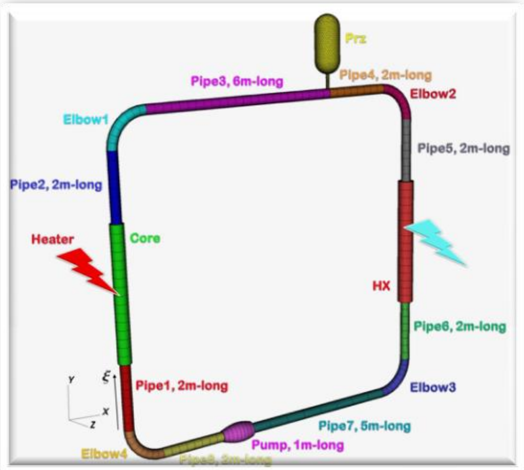
Sandia National Laboratories

Virtual Reactor Data

- For the purposes of this demo, we had no data.
- We simulated data, varying two parameters (nominal power output and wall temperature) that are not part of the calibration parameters.
- The plan is to get this methodology integrated and ready for the data from the APEX facility at Oregon State. This facility will have a ¼ scale reactor whose purpose is to provide V&V data on PWR reactors, specifically separate and integral effects for thermal hydraulic modeling.
- Two outputs of interest: peak temperature profile in the core, and flow rate.
- We simulated five data points. The temperature spread is shown below, the flow rate was 684 for all the five runs.



Different Models in the VR Loop



Heat Transfer

- Single model form
- Dittus-Boelter
- $Nu = NussCoeff * ReD^{NussRePower} * Pr^{n}$
- Two calibration parameters: NussCoeff and NussRePower

Wall Friction

- Two models: Blasius or Filonenko
- $\xi_{Blasius} = BlasiusCoeff * (ReD)^{-1/4}$
- $\xi_{Filonenko} = (FiloCoeff1 * \log_{10}ReD – FiloCoeff2)^{2}$
- One calibration parameter in Blasius; two in Filonenko

Local Friction in Junction

- Two models: Constant (ζ) or Expansion-Contraction
- $\zeta_e = K_e(1-S_a/S_b)^2$
- $\zeta_c = K_c(1-S_b/S_a)$
- One calibration parameter in constant model; two in exp-contr

Case Study

- We generated data from model (1,1) which is the "Truth" data
- We want to see if we can calibrate the models to match the truth data
- We also want to see if the model selection criteria give us information about the goodness of the models
- Nominal runs of the models are shown below:

	Temp1	Temp2	Temp3	Temp4	Temp5	Flow Rate
Model (1,1)	574.10	578.36	590.47	602.85	605.21	684.52
	574.90	579.17	591.31	603.69	606.05	684.52
	574.22	578.48	590.60	602.97	605.34	684.52
	576.95	581.22	593.43	605.83	608.17	685.30
	575.71	579.98	592.15	604.54	606.89	684.59
Model(1,2)	581.58	586.11	598.72	611.17	613.47	623.09
	582.36	586.90	599.51	611.95	614.23	623.09
	581.69	586.23	598.83	611.29	613.59	623.09
	584.36	588.89	601.52	613.88	616.12	623.09
	583.15	587.69	600.31	612.72	614.98	623.09
Model(2,1)	542.50	545.32	554.09	563.85	565.82	1085.10
	543.24	546.08	554.91	564.73	566.71	1085.31
	542.61	545.43	554.21	563.98	565.95	1085.13
	545.15	548.03	557.01	566.99	569.00	1085.91
	543.99	546.84	555.73	565.62	567.61	1085.54
Model(2,2)	559.13	562.69	573.30	584.67	586.93	851.85
	559.93	563.50	574.16	585.59	587.85	851.85
	559.25	562.81	573.43	584.81	587.07	851.85
	561.97	565.58	576.37	587.91	590.19	851.86
	560.73	564.32	575.03	586.50	588.77	851.86

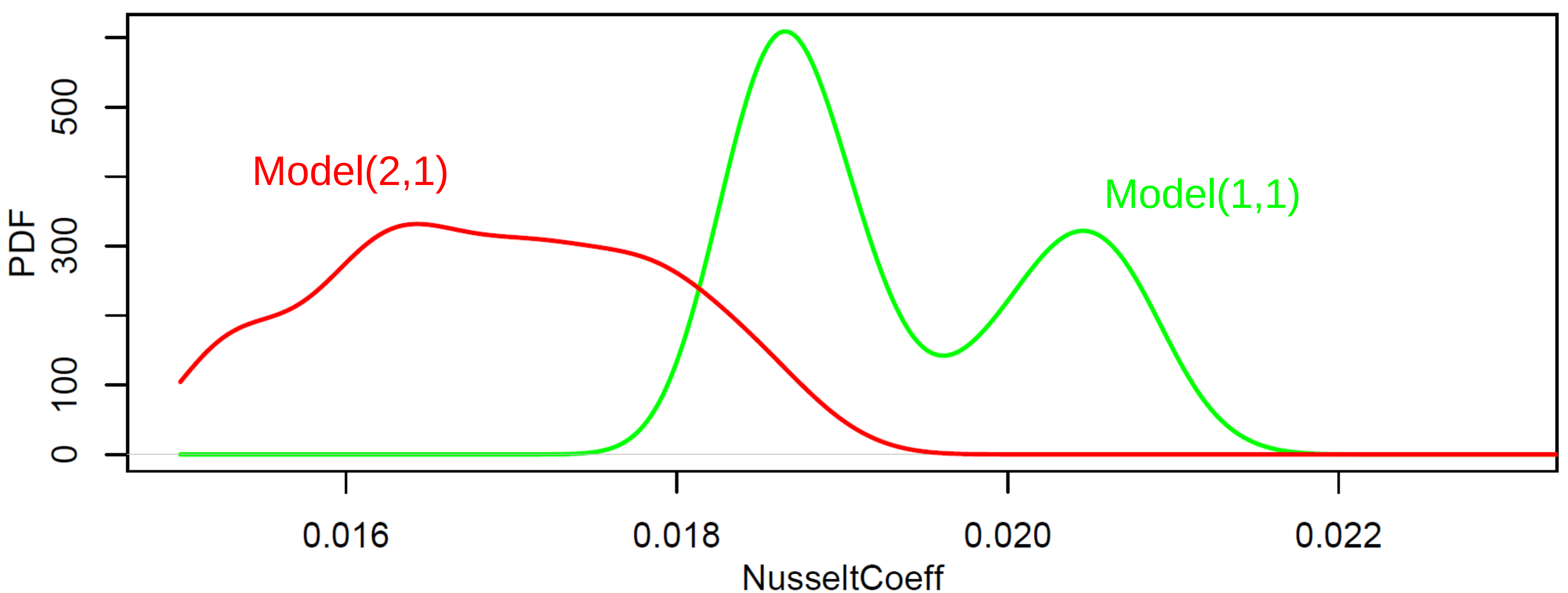
> Flow rate is quite different across models, will be harder to match

Step 1: Calibration

Table 3: Parameters to be calibrated for each model combination

Model(1,1) Variables	Lower Bound	Upper Bound
NusseltCoeff	0.015	0.025
NusseltRePower	0.79	0.81
BlasCoeff	0.25	0.36
Zeta	0.95	1.05
Model(1,2) Variables	Lower Bound	Upper Bound
NusseltCoeff	0.015	0.025
NusseltRePower	0.79	0.81
FiloCoeff1	1.75	1.85
FiloCoeff2	1.6	1.7
Zeta	0.95	1.05
Model(2,1) Variables	Lower Bound	Upper Bound
NusseltCoeff	0.015	0.025
NusseltRePower	0.79	0.81
BlasCoeff	0.25	0.36
Ke	0.95	1.05
Kc	0.35	0.45
Model(2,2) Variables	Lower Bound	Upper Bound
NusseltCoeff	0.015	0.025
NusseltRePower	0.79	0.81
FiloCoeff1	1.75	1.85
FiloCoeff2	1.6	1.7
Ke	0.95	1.05
Kc	0.35	0.45

Step 1: Calibration



The interpretation of model parameters having vague prior information can be dependent on the context in which the parameters are being calibrated.

The bimodality in the marginal distribution of the NusseltCoeff for Model (1,1) is due to a compensating error with the NusseltRePower parameter (not shown).

A more informative prior on the NusseltCoeff would be useful.

Model Selection

TEMPERATURE ONLY								
	With Discrepancy				No Discrepancy			
Model	ML	AIC	AICC	BIC	ML	AIC	AICC	BIC
(1,1)	-14.50	43	51.47	51.54	-14.50	39	43.42	45.1
(1,2)	-14.51	45.02	56.27	54.77	-14.51	41.02	47.24	48.33
(2,1)	-23.41	62.83	74.08	72.58	-47.42	106.84	113.06	114.15
(2,2)	-17.06	52.12	66.79	63.09	-17.01	48.02	56.49	56.55

TEMPERATURE and FLOW RATE								
	With Discrepancy				No Discrepancy			
Model	ML	AIC	AICC	BIC	ML	AIC	AICC	BIC
(1,1)	-11.08	40.17	51.17	52.78	-11.08	34.17	39.04	42.58
(1,2)	-17.14	54.29	68.18	68.3	-29.15	72.29	78.84	82.1
(2,1)	-28.25	76.5	90.4	90.51	-78.56	171.11	177.66	180.92
(2,2)	-20.78	63.56	80.89	78.97	-43.43	102.87	111.44	114.08

Information theoretic criteria:

- All cases rank model(1,1) first, followed by model(1,2), model(2,2), and model(2,1)
- "With discrepancy" shows fewer differences across models: discrepancy term compensates for the model inadequacy
- Temperature only is "easier" to fit, has better information criteria values EXCEPT for model(1,1)

Model Selection: Bayesian Approach

Recall posterior model probability:

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

Where the evidence for model M_k is: $$p(D \mid M_k) = \int p(D \mid \theta_k, M_k)\, p(\theta_k \mid M_k)\, d\theta_k$$

Calculation methods are not robust: estimators of the integrated likelihood tend to have high variance.

1. Simple mean of likelihood: $$p(D \mid M_k) \approx \frac{1}{N} \sum_{i=1}^{N} p(D \mid \theta_{k_i}, M_k)$$
2. Harmonic mean: $$p(D \mid M_k) \approx \frac{N}{\sum_{i=1}^{N} \left[p(D \mid \theta_{k_i}, M_k) \right]^{-1}}$$
3. Wolpert method: $$\frac{\sum_{i=1}^{N} \left[p(D \mid \theta_{k_i}, M_k) \right]^{-1}}{N} \approx \frac{1}{p(D \mid M_k)} + Z N^{1/\alpha - 1}$$

Model Selection: Bayesian Approach

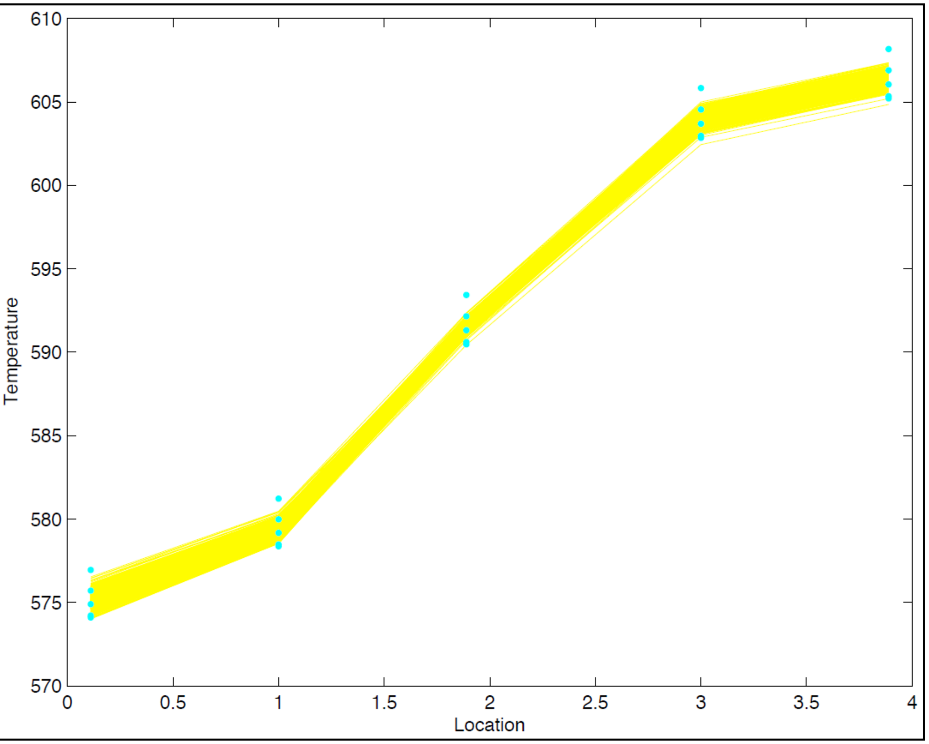
DEVIANCE INFORMATION CRITERION (DIC)				
	With Discrepancy		No Discrepancy	
Model	Temp. Only	Temp + Flow Rate	Temp. Only	Temp + Flow Rate
(1,1)	33.4	30.53	32.92	30.51
(1,2)	37.39	46.84	37.75	66.61
(2,1)	56	69.62	101.05	165.66
(2,2)	45.97	57.39	45.98	100.85

- The ranking of the models is the SAME regardless of case:
  - Matching temperature only vs. temperature and flow rate
  - Models with and without discrepancy
- These rankings are also consistent with the information theoretic rankings

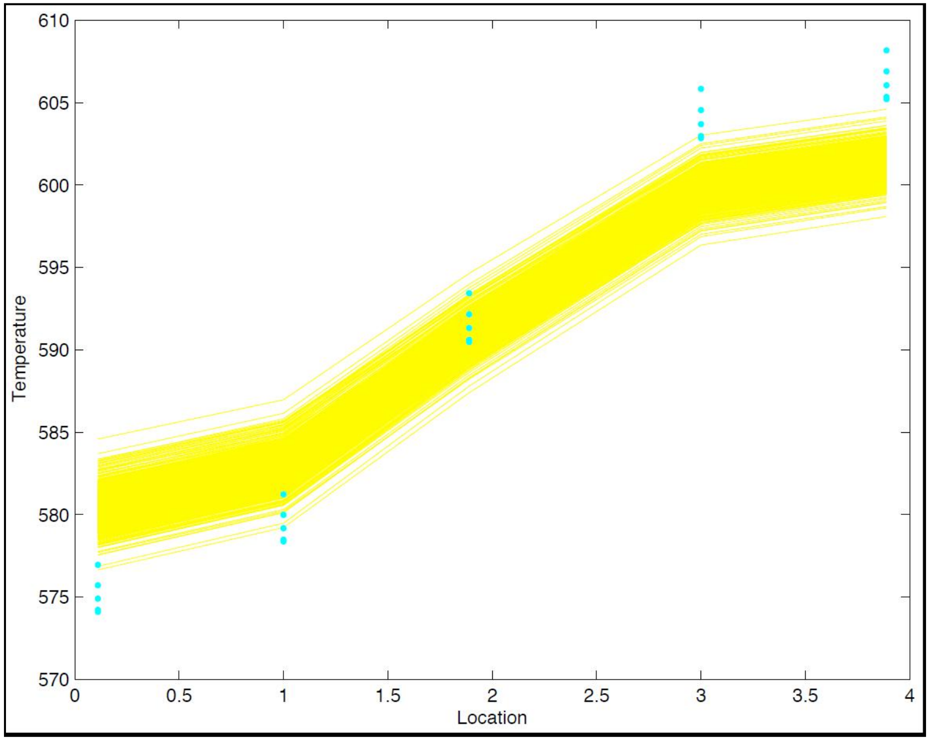
Sandia National Laboratories

Model Prediction: Bayesian Approach

without discrepancy term, calibration to temp only



MODEL (1,1)

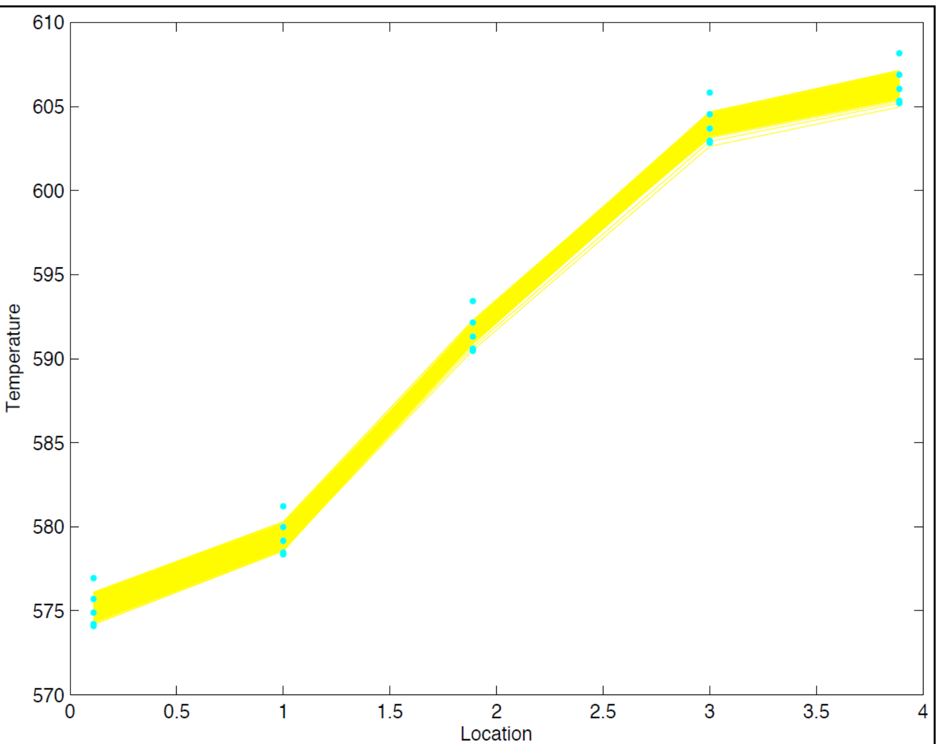


MODEL(2,1)

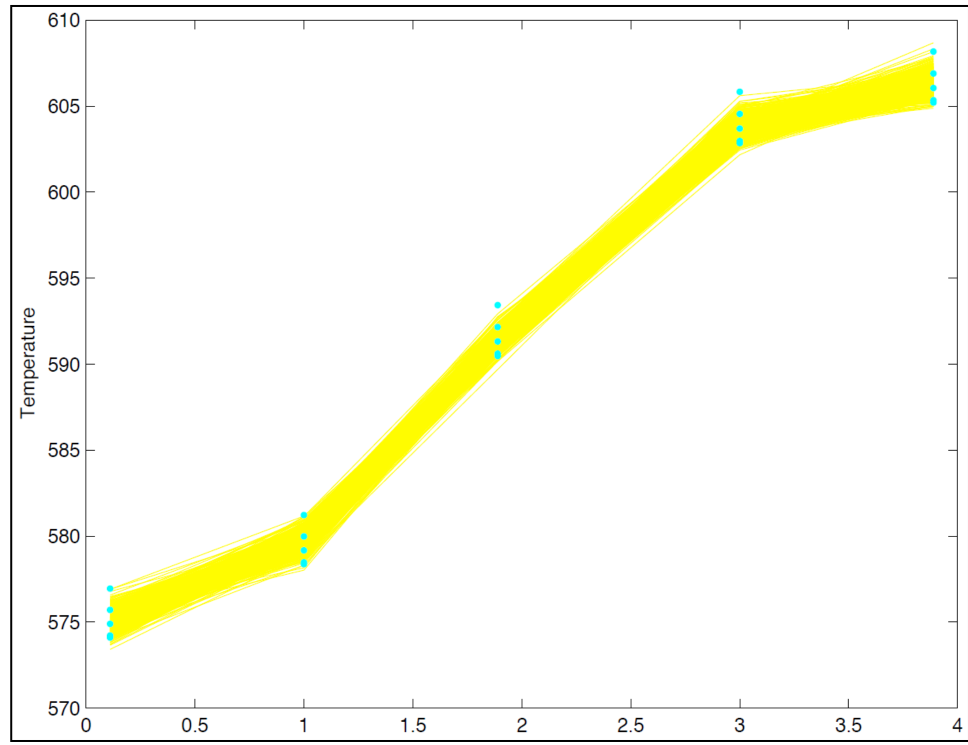
Sandia National Laboratories

Model Prediction: Bayesian Approach

with discrepancy term, calibration to temp only



MODEL (1,1)

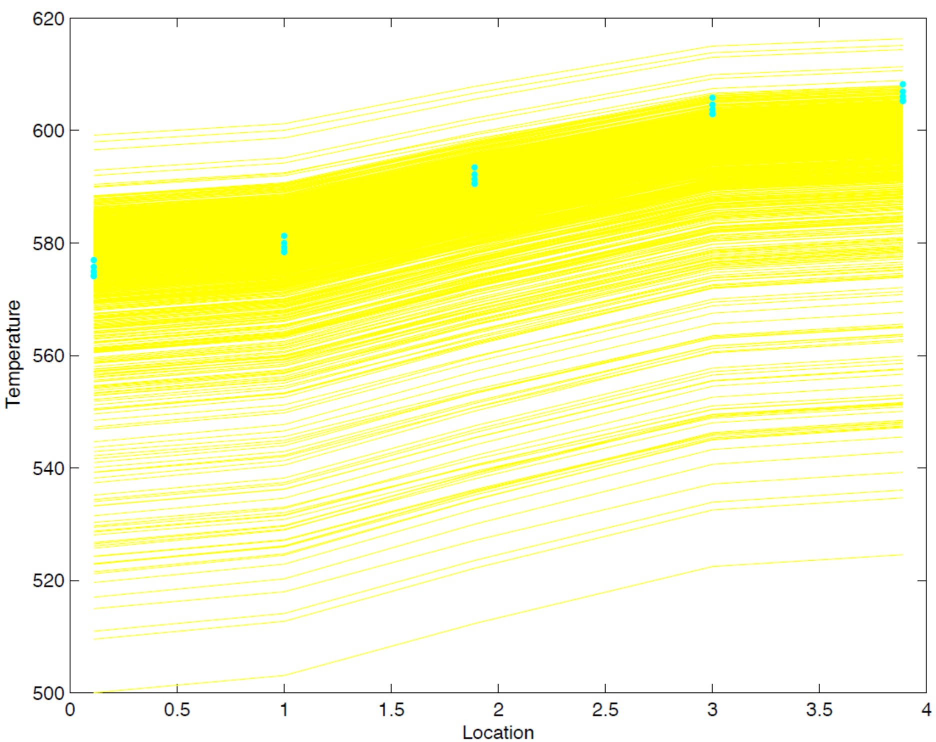


MODEL(2,1)

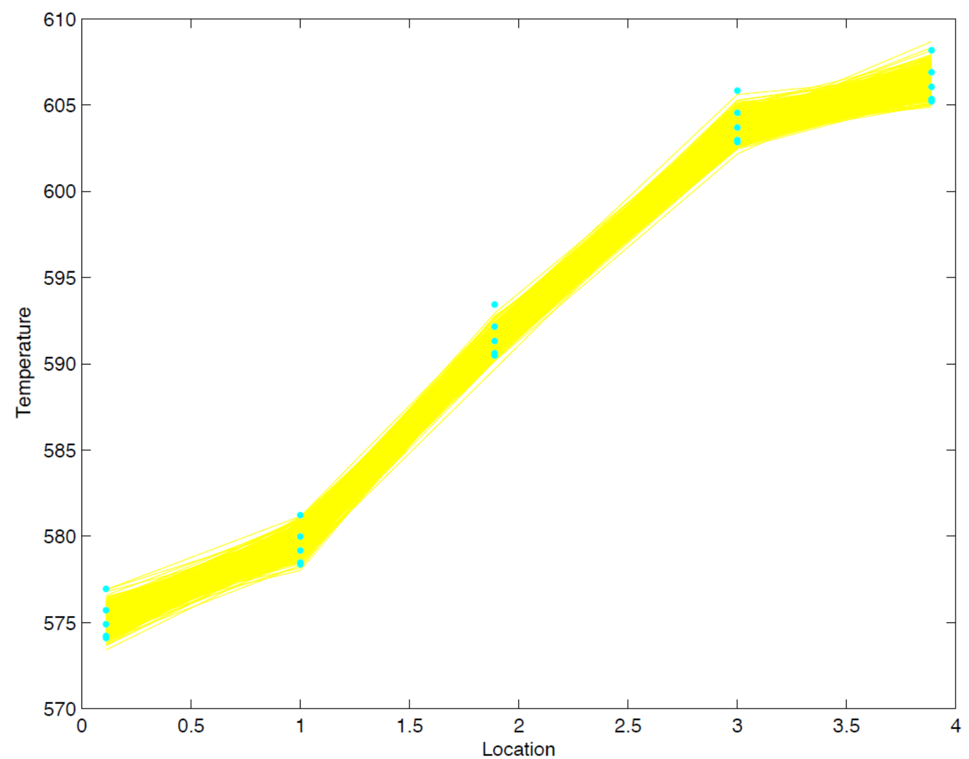
Sandia National Laboratories

Model Prediction: Bayesian Approach

with discrepancy term, calibration to temp and flow rate



MODEL (2,1) before discrepancy



MODEL(2,1) after discrepancy

Summary

- We tried a variety of calibration and model selection methods
- All of the model selection criteria ranked the models in the same order
- Models had an easier time calibrating to temperature, harder to match both calibration and flow rate: the addition of another quantity of interest (and its associated data) increases the difficulty of calibrating to all data sources simultaneously.
- The presence of a discrepancy term significantly improved the performance of some models (especially very poor models) relative to the best performing model, in terms of increasing their information content by empirically correcting inadequacy in direct model predictions.
- Nevertheless, discrepancy is not able to improve the performance of a poor model to the extent it would outrank a good model because empirical corrections of model predictions tend to increase their uncertainty relative to the prediction uncertainty arising from models not requiring adjustment.

Sandia National Laboratories

Summary

- The model selection methods were designed for large numbers of statistical models, not a few substantively different physics models.
- The model selection criteria do not incorporate:
  - Complexity of the models (except through number of parameter term)
  - Mesh complexity
  - Computational cost
  - Order of physics
  - Time step
  - Etc.
- We want to extend the model selection methods to the case of computational physics models
- We want to apply the framework of model calibration/selection/prediction to other NEAMS problems.