A Framework for Validation of Computer Models*

M.J. Bayarri, J.O. Berger, D. Higdon, M.C. Kennedy, A. Kottas, R. Paulo, J. Sacks

National Institute of Statistical Sciences

J.A. Cafeo, J. Cavendish, C.H. Lin, J. Tu

General Motors

*Supported by grants from General Motors and the National Science Foundation to the National Institute of Statistical Sciences.

Outline

- Validation via *tolerance bounds*
- Features of Bayesian model validation
- The six steps of the framework
- Generalizations
- Illustrations using computer models for spot welding and for vehicle crashes

Tolerance Bounds

Instead of asking "Is the model correct?*", we recommend routinely reporting tolerance bounds for predictions.

Example 1: With probability 80%, the model prediction 5.17 (at specified input x) will be within ± 0.44 of the true process value (at input x).

Example 2: With probability 80%, the model prediction 6.28 (at specified input x^*) will be within ± 1.6 of the true process value (at input x^*).

*a question that can be addressed, but which is usually irrelevant.

This simple reporting device overcomes the difficulties that

- it is often impossible to adequately characterize the regions of model accuracy and inaccuracy;
- the degree of accuracy that is needed can vary from one application of the computer model to another;
- it is usually crucial to incorporate *model bias*, as well as variance, in assessment of accuracy.

Features of Bayesian Model Validation

- It can incorporate all types of uncertainty; from uncertainty in model inputs or parameters to uncertainty in the data to uncertain expert opinion.
- It can be implemented even if data (model-run data and/or field data) is very limited.
- The model-run and field data can be observed at different input values.
- One can 'tune' unknown parameters of the computer model based on field data, and at the same time apply the validation methodology.

- Analysis of model bias is included – indeed, is central.
- The analysis is naturally sequential, allowing updating as new information arrives.
- Optimal predictions involve a synthesis of data-driven and model-based prediction, with each contributing most in the domain where it is most powerful.
- The methodology includes development of a fast 'response surface' approximation to the computer model, with computed accuracy.
- Prediction for a 'related situation' is possible.

Philosophy and Caveats

- Validation is a hard problem which requires sophisticated statistical methodology.
- Incorporation of physical knowledge is often crucial in model validation but, without field data, validation will always be suspect.
- The methodology we propose needs extension to
  - stochastic computer models
  - large dimensional input spaces
  - large numbers of unknown model parameters
  - large data sets

Sketch of the Framework

Step 1. Specify model inputs and uncertain model parameters with associated uncertainties or ranges – this is called the Input/Uncertainty (I/U) map.

Step 2. Determine evaluation criteria.

Step 3. Design experiments and collect data (field and computer-run).

Step 4. If the computer model is slow, develop a fast (response surface) approximation.

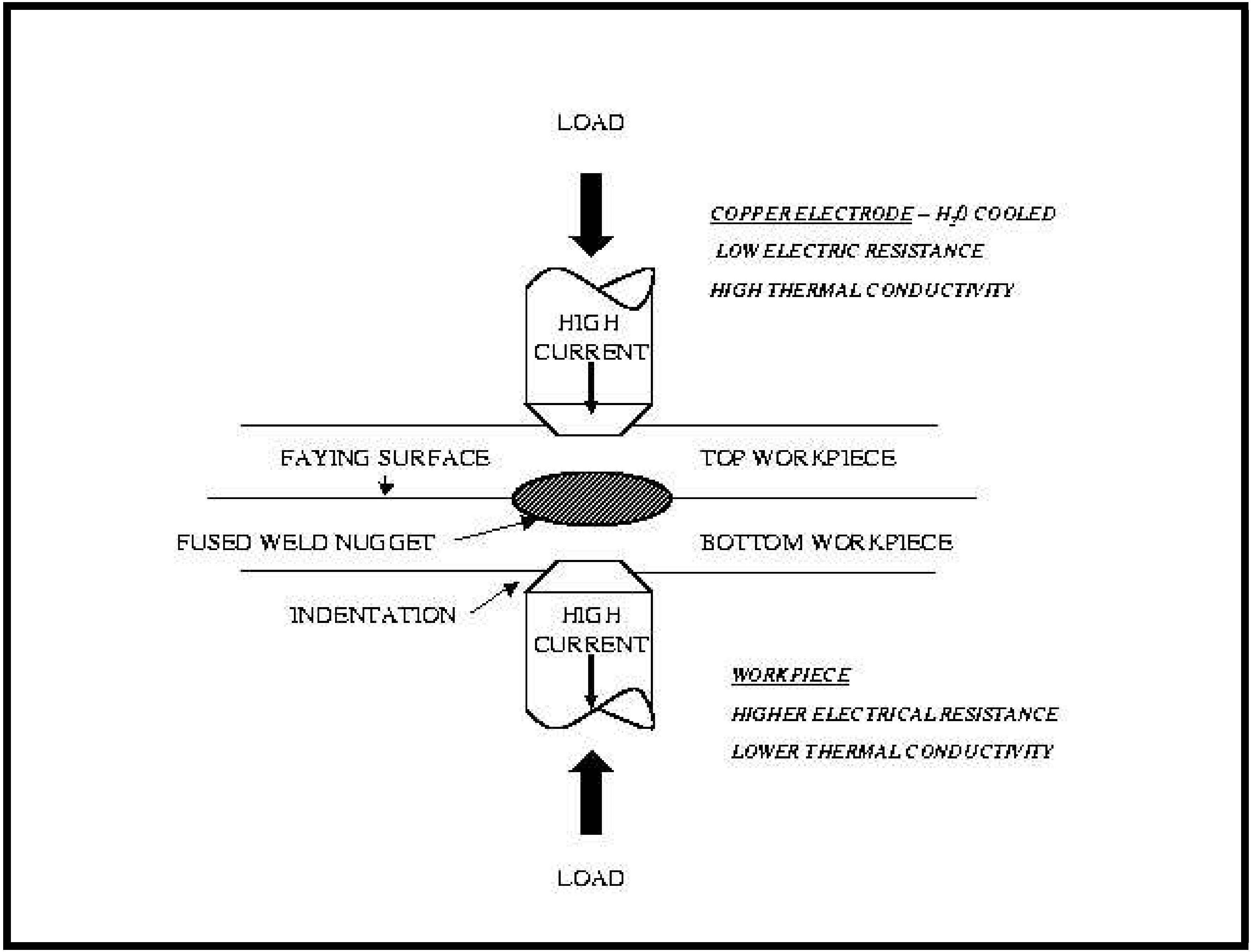
Step 5. Analyze and compare computer model output and field data, involving

- statistical modelling of field data error;
- tuning/calibrating model input parameters based on the field data;
- updating uncertainties in parameters, given the data;
- assessment of model bias;
- computing the tolerance bounds for prediction.

Step 6. Feedback information into the current validation exercise and feed-forward information into future validation activities.

Testbed 1. Resistance Spot Welding

- Key process inputs include
  - the load, L, applied to copper electrodes;
  - a direct current of magnitude C;
  - material type and surface.
- The thermal/electrical/mechanical physics of the spot weld process is modeled by a coupling of partial differential equations, implemented in a finite element computer model using ANSYS.
- The key unknown in the computer model is u, the resistance of the faying surface.


LOAD
COPPER ELECTRODE – H_2O COOLED
LOW ELECTRIC RESISTANCE
HIGH THERMAL CONDUCTIVITY
HIGH CURRENT
FAYING SURFACE
TOP WORKPIECE
FUSED WELD NUGGET
BOTTOM WORKPIECE
INDENTATION
HIGH CURRENT
WORKPIECE
HIGHER ELECTRICAL RESISTANCE
LOWER THERMAL CONDUCTIVITY
LOAD

Step 1. Specify the Input/Uncertainty (I/U) Map

- The I/U map has four attributes:
  1) a list of key model features or inputs;
  2) a ranking (1-5) of the importance of each input;
  3) uncertainties, either distributions or ranges of possible values, for each input;
  4) current status of each input describing how the input is currently treated in the model.
- The I/U map is ideally created during the model development process.
- The I/U map is dynamic, and frequently updated.

INPUT		IMPACT	UNCERTAINTY	CURRENT STATUS
Geometry	electrode symmetry-2d	3	unspecified	fixed
	cooling channel	1	unspecified	fixed
	gauge	unclear	unspecified	1, 2mm
materials		unclear	Aluminum (2 types $\times$ 2 surfaces)	fixed
Stress/ strain	piecewise linear	4	unspecified (worse at high T)	fixed
	C_0, C_1, σ_s	3	unspecified	fixed
contact resistance	$1/\sigma = u \cdot f$; f fixed	3	unspecified	fixed by modeler
	$u = 0$ (electrode/sheet) u =tuning (faying)	5	$u \in [0.8, 8.0]$	tuned to data for 1 metal
thermal conductivity		2	unspecified	fixed
current		5	no uncertainty	controllable
load		5	no uncertainty	controllable
mass density		1	unspecified	fixed
specific heat		1	unspecified	fixed
numerical parameters	mesh	1	unspecified	convergence/speed compromise
	M/E coupling time	1	unspecified	
	boundary conditions	1	unspecified	fixed
	initial conditions	1	unspecified	fixed

Table 1: The I/U map for the spot weld model

Step 2. Determine Evaluation Criteria, including

- specification of an evaluation criterion: for SPOT WELD, the size of the nugget after 8-cycles;
- specification of the relevant domain of input variables: for SPOT WELD,
  - Material: Aluminum 5182-O and Aluminum 6111-T4
  - Surface: treated or untreated
  - Gauge (mm): 1 or 2
  - Current (kA): 21 to 26 for 1mm aluminum; 24 to 29 for 2mm aluminum
  - Load (kN): 4.0 to 5.3
  - Resistivity u: 0.8 to 8.0

Step 3. Design of Experiments and Data

Both computer and field experiments are typically crucial in validation. For SPOT WELD:

- The computer experiments were designed to cover the domains of the four key inputs, C = current, L = load, G = gauge, and u = resistivity.
- The cost – thirty minutes per computer run – resulted in a limited number, 52, of possible runs.
- A 'space filling' Latin Hypercube design was used.
- Field data consisted of previous lab experiments, 5 replicates at each of 12 input values.

Step 4. Develop a Fast Model Approximation

- for employment in the field;
- for optimization over input variables;
- to find optimal designs for additional validation or model-development experiments;
- for use in the Bayesian calibration and validation methodology.

Notation for model output: $y^M(\boldsymbol{x}, \boldsymbol{u})$, where
$\boldsymbol{x}$ is a vector of controllable inputs
$\boldsymbol{u}$ is a vector of calibration or tuning parameters.

The GASP Model Approximation to $y^M(\boldsymbol{x}, \boldsymbol{u})$

- Fit a Gaussian spatial model to the model-run data.
- One can then compute
  $\hat{y}^M(\boldsymbol{x}, \boldsymbol{u})$, the *GASP model approximation*
  $V^M(\boldsymbol{x},\boldsymbol{u})$, the *variance* of the approximation.

Example: For SPOT WELD, at input $(C, L, G, u) = (26, 5, 2, 4)$, the GASP approximation is $\hat{y}^M(26, 5, 2, 4) = 6.12$, and its variance is $V^M(26, 5, 2, 4) = 0.0046$.

- The computations are of a Kalman-filter type.
- Fitting can be done by maximum likelihood (using the GASP code of Walsh), or by Bayesian analysis.

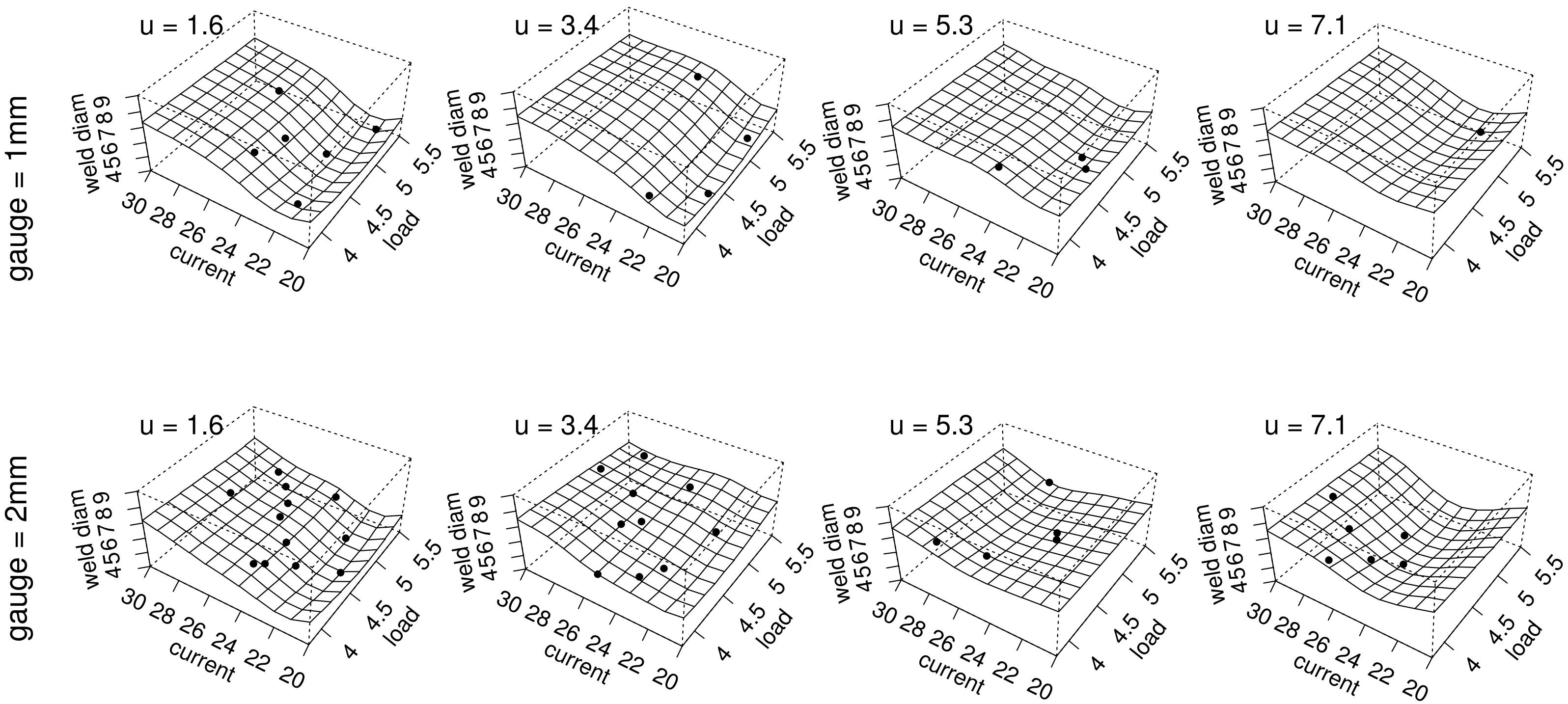


Figure 1. The GASP approximation, $\hat{y}^M(\boldsymbol{x}, \boldsymbol{u})$, to the spot weld model $y^M(\boldsymbol{x}, \boldsymbol{u})$. The dots are the model-run data.

Some Technical Details

Defining $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{u})$, GASP has

- *mean function* $\boldsymbol{\Psi}(\boldsymbol{z})\,\boldsymbol{\theta}$, where
  - the *basis functions* $\boldsymbol{\Psi} = (\Psi_1, \ldots, \Psi_k)$ are specified
  - $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)'$ is unknown (and to be estimated)

  Note: A constant mean is often satisfactory.

- *covariance function*, for the d-dimensional $\boldsymbol{z}$,
  $\mathrm{Cov}(\boldsymbol{z}, \boldsymbol{z}^*) = \frac{1}{\lambda} \prod_{j=1}^{d} \exp\left(-\beta_j |z_j - {z_j}^*|^{\alpha_j}\right)$.
  - The covariance 'separability' can greatly speed computation and 'fitting' in high dimensions.
  - Stochastic inputs can be handled very easily.

Step 5. Comparison of model and field data: calibration/tuning; estimation of bias; and development of tolerance bands for prediction

The computer model is related to reality via

$$y^R(\boldsymbol{x}) = y^M(\boldsymbol{x}, \boldsymbol{u}^*) + b(\boldsymbol{x})\,,$$

where $\boldsymbol{u}^*$ is the true (but unknown) value of $\boldsymbol{u}$ and $b(\boldsymbol{x})$ is the model bias. Field data at inputs $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ are obtained, and modeled as

$$y^F(\boldsymbol{x}_i) = y^R(\boldsymbol{x}_i) + \epsilon_i^F\,,$$

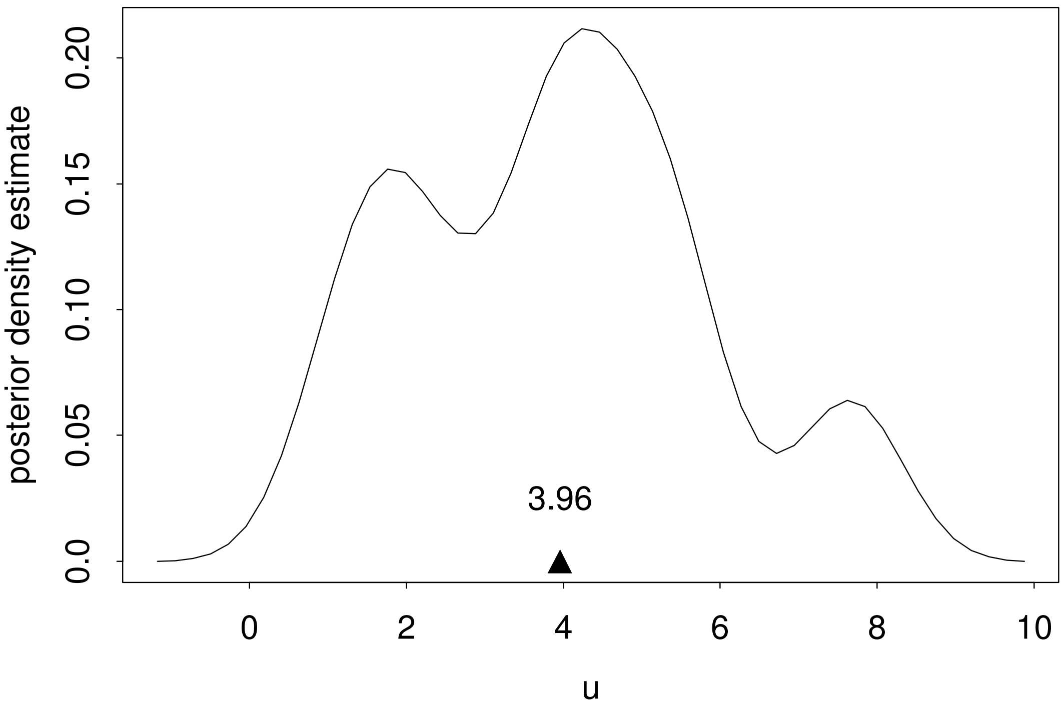
where the ϵ_i^F are i.i.d. Normal$(0, 1/\lambda^F)$ random errors.

Bayesian Analysis

- Specify *prior* probability densities for unknown elements of the model,
  - the density $p(\boldsymbol{u})$ for $\boldsymbol{u}$ (from the I/U map)
  - a density $p(\lambda^F)$ (precision of the measurement error)
  - a prior density for the bias function $b(\boldsymbol{x})$ (chosen to be a smooth Gaussian process)
  - prior densities for the unknown parameters in GASP.
- Utilize Bayes theorem to obtain the *posterior* density of all unknowns, given the data. (Note that this automatically incorporates all uncertainties in the problem.)
- Predictions and tolerance bounds follow easily.

Calibration and Tuning

For SPOT WELD, Bayesian analysis yields the following posterior density for u. The optimal 'tuned' value of u is 3.96, but it is much better to utilize the entire density.



Predicting Reality Using a New Model Run

To predict the real process at some (new) input $\boldsymbol{x}$, first run the computer model, obtaining $y^M(\boldsymbol{x}, \hat{\boldsymbol{u}})$.

Model Prediction: If $y^M(\boldsymbol{x}, \hat{\boldsymbol{u}})$ is used directly as the prediction, its variance, $V_{\hat{\boldsymbol{u}}}(\boldsymbol{x})$, is available.

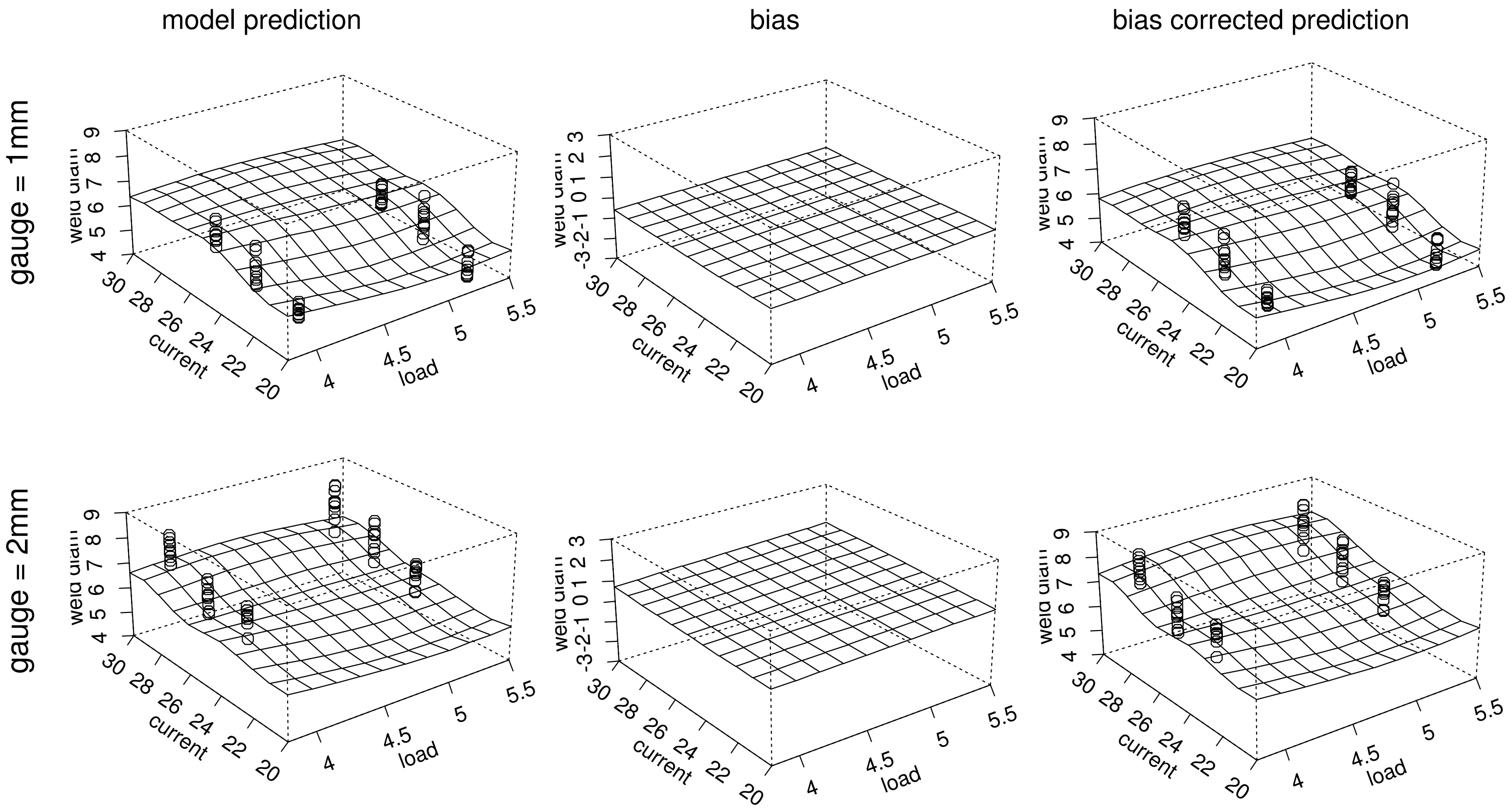
Bias-Corrected Prediction: Better is to estimate the bias $\hat{b}_{\hat{\boldsymbol{u}}}(\boldsymbol{x})$, and use the bias-corrected prediction

$$\hat{y}^R(\boldsymbol{x}) = y^M(\boldsymbol{x}, \hat{\boldsymbol{u}}) + \hat{b}_{\hat{\boldsymbol{u}}}(\boldsymbol{x}).$$

Example: In SPOT WELD, at input G=2, L=4.888, C=29.44, and $\hat{u} = 3.96$, the pure model prediction is $\hat{y}^M = 7.16$, with variance $V_{3.96} = 0.628$. The estimated bias is $\hat{b}_{3.96} = 0.342$, so the bias-corrected prediction is $\hat{y}^R(4.888, 29.44, 2) = 7.50$.

Predicting Reality Without a New Model Run

For SPOT WELD, and using *only* the GASP model approximation.



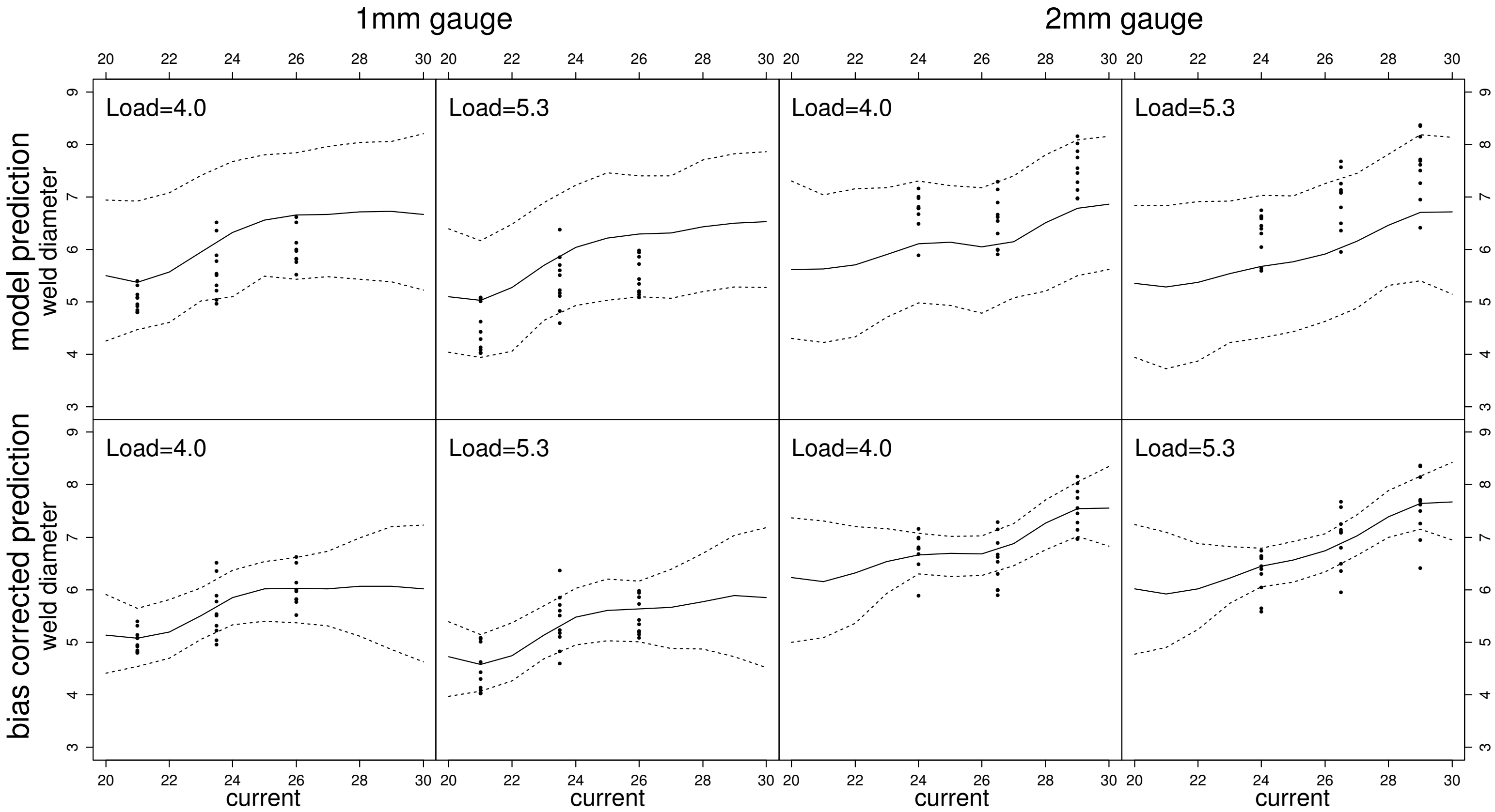
Tolerance Bounds With a New Model Run

For SPOT WELD, at inputs G=2, L=4.888, C=29.44, and $\hat{u} = 3.96$,

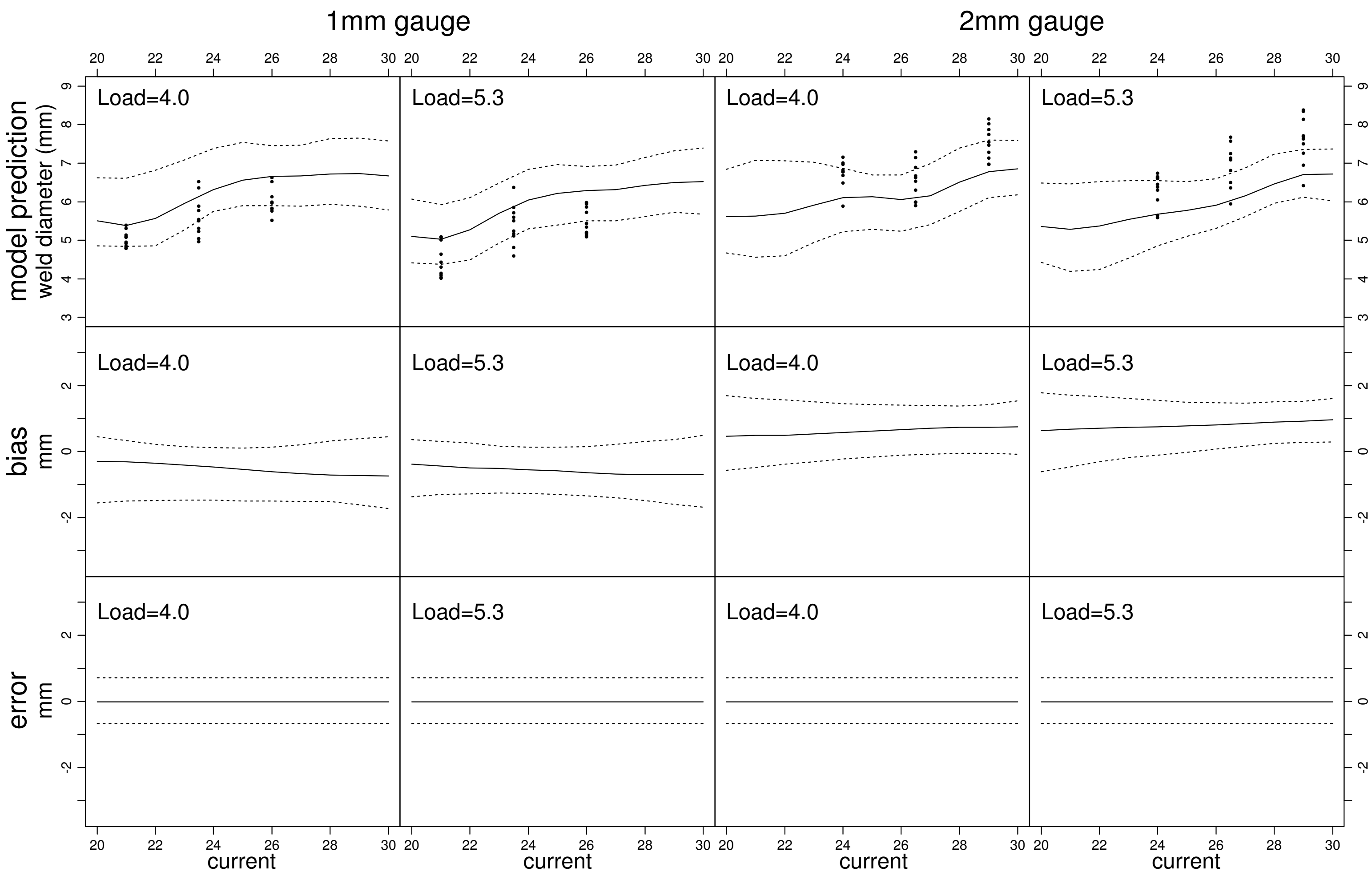
- The pure model prediction was $\hat{y}^M = 7.16$, and the corresponding 90% tolerance bounds are $(6.02, 8.30)$.
- The bias-corrected prediction was 7.50, and the corresponding 90% tolerance bounds are $(6.15, 8.30)$.

Note: In bias-corrected predictions, both the new model-run and the field data can be important. Indeed, model-runs dominate locally (if, e.g., the effect of local changes are being evaluated), while field data often dominates globally.

Tolerance Bounds Without a New Model Run



Uncertainty Decomposition

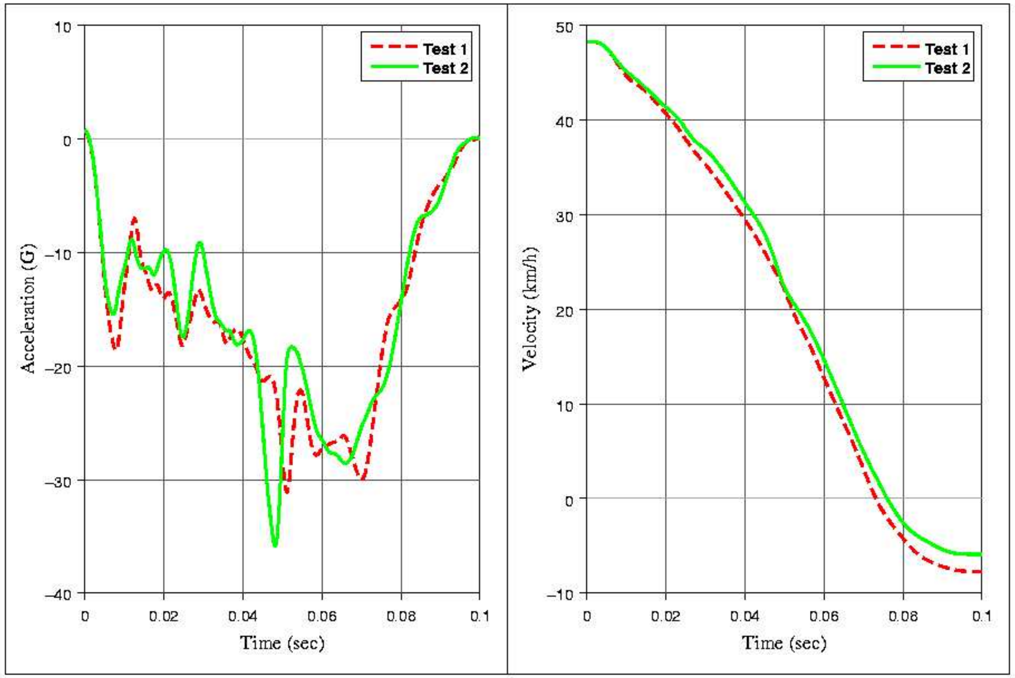


The Bayesian Computation

- Utilizes a simulation method called Markov Chain Monte Carlo.
- It is intensive, requiring thousands of iterations.
  - If thousands of model runs are not feasible, the GASP model approximation must be used.
  - The simplifications arising from the specified form of the covariance function can be important.
- Software for this approach to model validation is not currently available.

Generalization to Functional Data

Testbed 2. A Vehicle Crash Model. Collision of a vehicle with a barrier is implemented as a non-linear dynamic analysis code using a finite element representation of the vehicle. The focus is on velocity changes at the driver seat, as in the following 30mph crash.

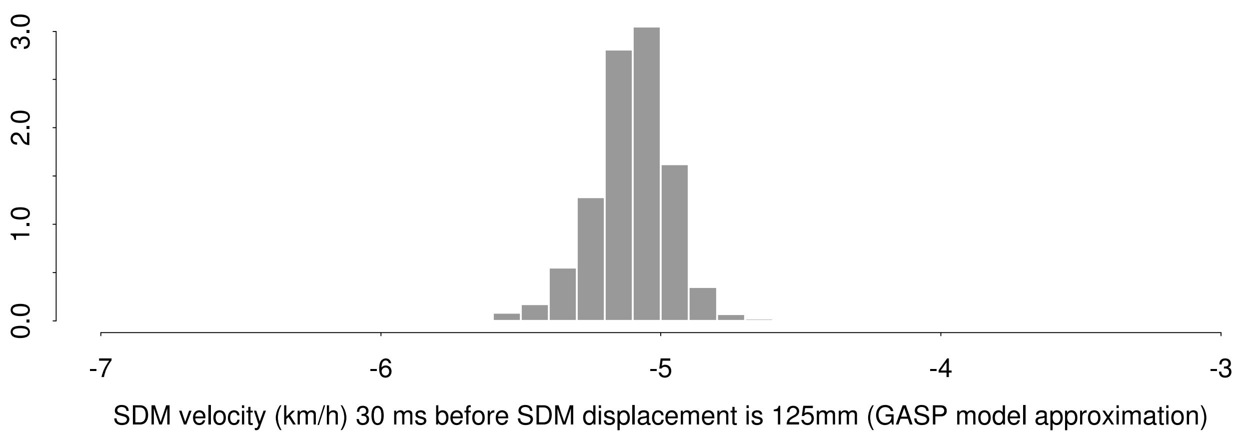


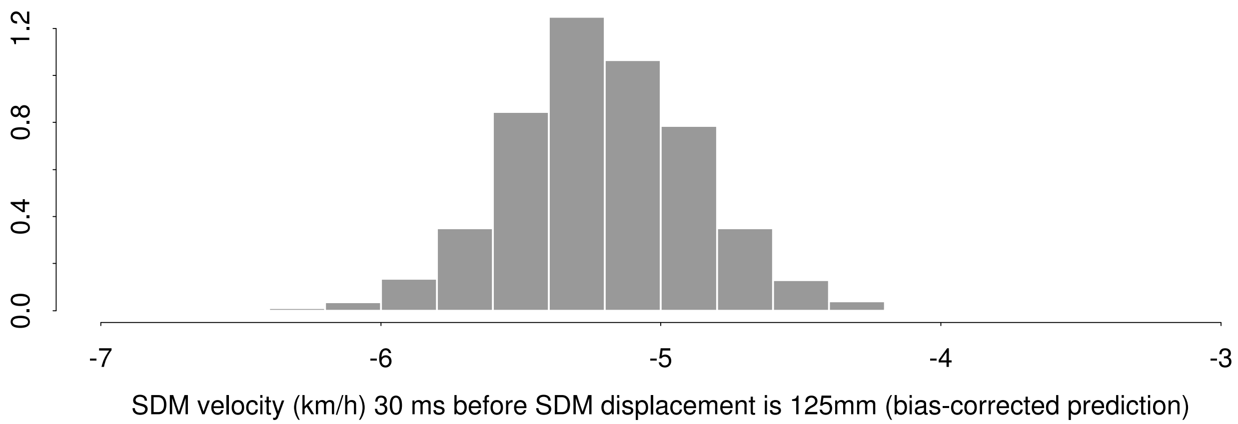
Representation and Analysis of Functions

- Discretize t, and include as another model input in $\boldsymbol{x}$.
- Assume all Gaussian process correlations involving t have the same form.
- Then the separable form of the correlation functions allows a major Kronecker product simplification in the computation (because $\boldsymbol{C}_{x,t}^{-1} = (\boldsymbol{C}_x \otimes \boldsymbol{C}_t)^{-1} = \boldsymbol{C}_x^{-1} \otimes \boldsymbol{C}_t^{-1}$).

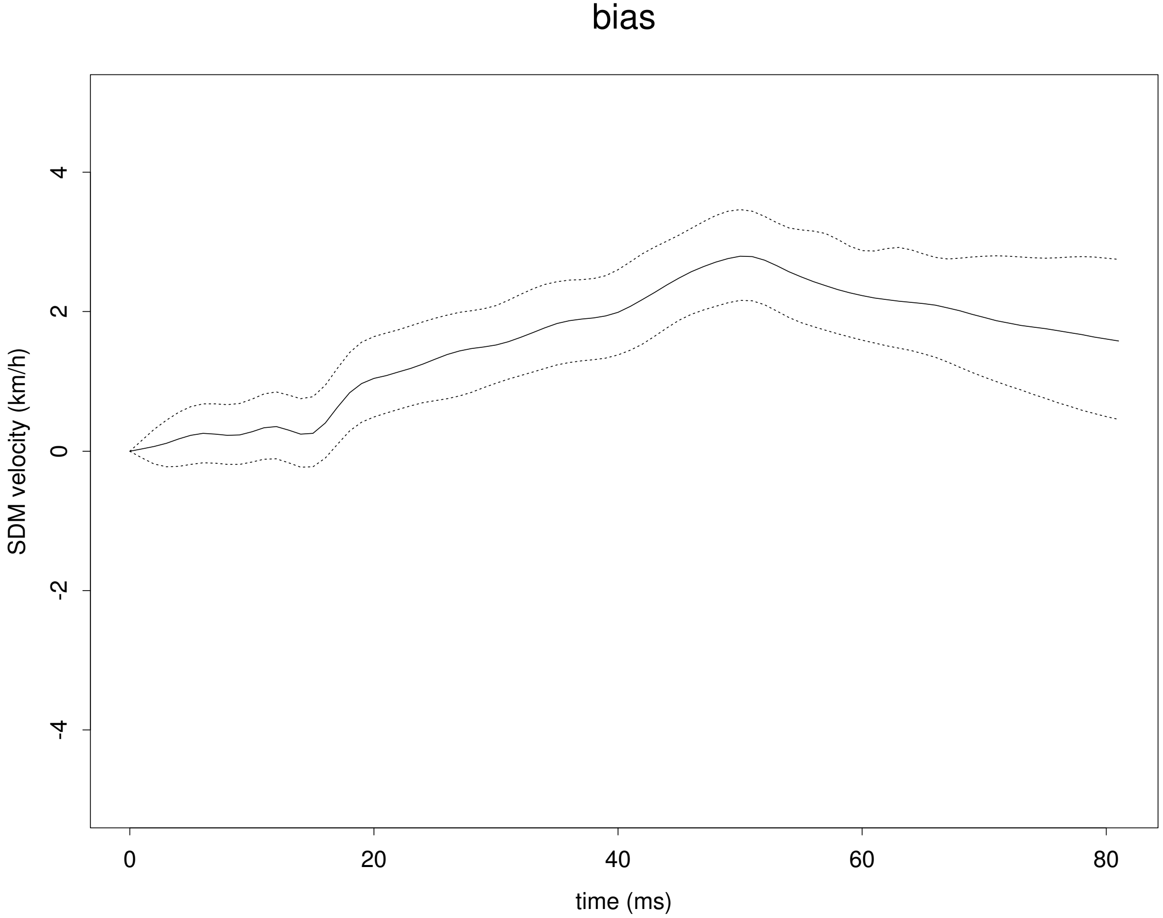
Example: For CRASH, a key evaluation criterion, CRITV, is "driver seat velocity calculated 30ms before its displacement reaches 125mm." Discretize to the 19 time points $t = 1, 3, \ldots, 15, 17, 20, 25, \ldots, 65$ms, with more smaller times because of their importance in estimating CRITV.

Posterior Densities for CRITV





Estimates and 80% error bands for SDM-velocity bias at 30km/h impact



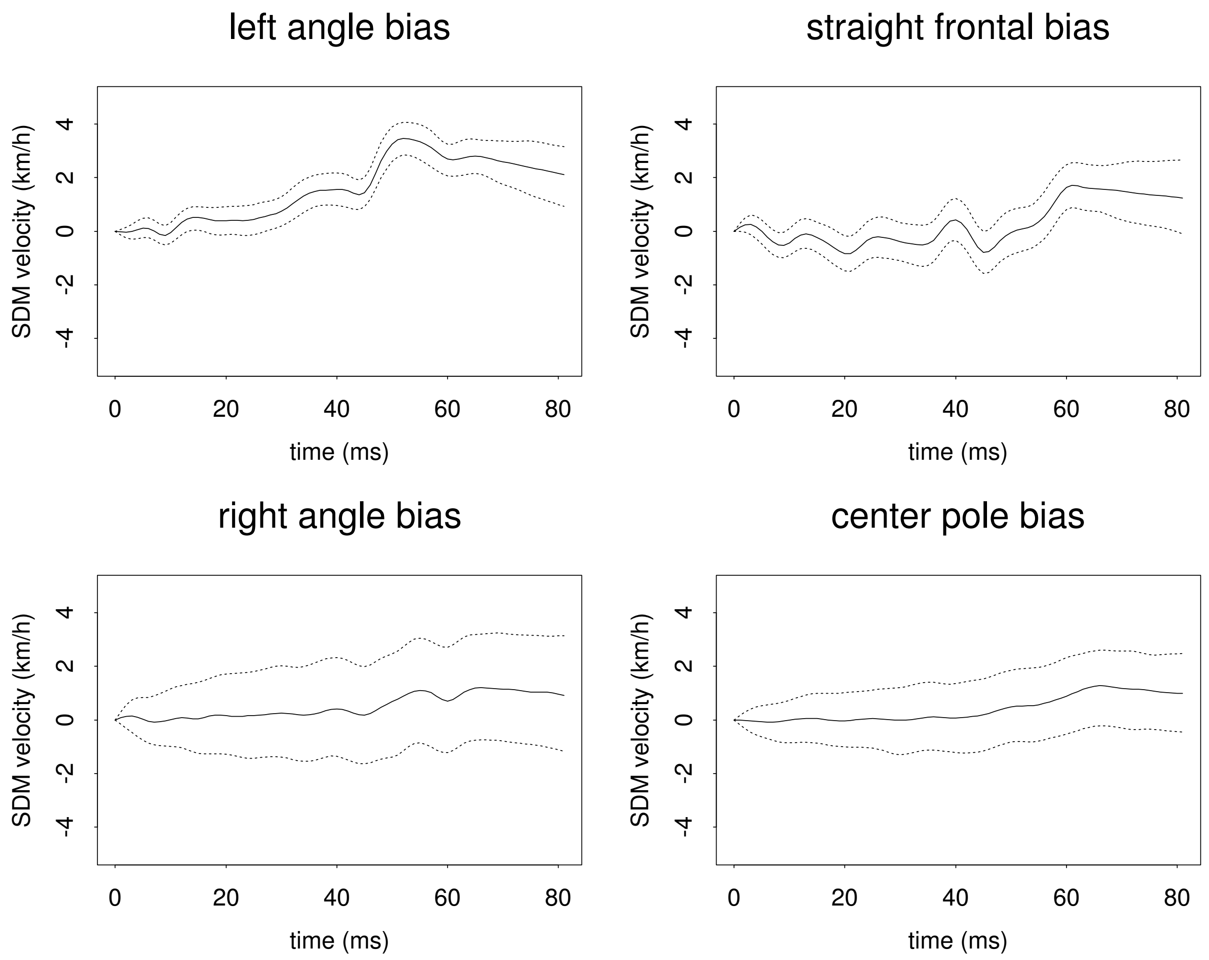
Extrapolating Past the Range of the Data

- The above methodology will tend to return large tolerance bands in extrapolation, unless one just 'assumes' that bias estimates extrapolate.
- Less dogmatic is to model the new scenario as being related – but not identical – to already studied situations. This is done by *hierarchical Bayesian analysis*, which is done here by assuming
  - common smoothness and variances for the various GASP approximations;
  - mean structures arising from a common density;
  - biases are related up to a 'proportional variation' q.

For CRASH, the earlier analysis was for straight frontal impact. Hierarchical modeling allows treatment of left angle, right angle and center pole impacts. Here are the posterior mean (standard deviation) for CRITV at a 56.3km/h impact and with $q = 0.1$ (i.e., up to 10% variation in biases is expected apriori).

Barrier type	$CRITV$ (model only)	$CRITV$ (bias-corrected)
left angle	-6.08 (0.34)	-6.34 (0.49)
straight frontal	-5.13 (0.13)	-5.22 (0.30)
right angle	-6.89 (0.65)	-6.80 (0.96)
center pole	-6.55 (0.74)	-6.54 (0.91)

Pointwise 80% posterior intervals for bias, based on the hierarchical model, when only previous model runs are utilized.



Bayesian Determination of the Probability that the Computer Model Is Correct

- Specify a prior probability, π_0, that the computer model, $\mathcal{M}_0$, is correct.
- Select an alternative model $\mathcal{M}_1$ (which here, will be 'model + bias').
- Specify prior densities for unknown parameters of $\mathcal{M}_0$ and $\mathcal{M}_1$ (already done).
- Compute the *posterior probability that* $\mathcal{M}_0$ *is correct* (which for CRASH, is essentially zero).