An introduction to model calibration
... and a some incoherent thoughts

Derek Bingham
Statistics and Actuarial Science
Simon Fraser University
Outline

• Radiative shock example

• Model calibration with limited model evaluations ... and Gaussian processes

• Model calibration with discrepancy: CRASH example
Many processes are investigated using computational models

- Many scientific applications use mathematical models to describe physical systems

- The computer models frequently:
  1. require solutions to PDEs or use finite element analysis
  2. have high dimensional inputs
  3. have outputs which are complex functions of the input factors
  4. require a large amounts of computing time
  5. have features from some of the above
Center for Radiative shock Hydrodynamics (CRASH)

A conceptually simple experiment
- Launch a thin Be plasma down a shock tube at \( \sim 200 \text{ km/s} \)

A radiative shock is a wave in which both hydrodynamic and radiation transport physics play a significant role in the shock’s propagation and structure
Final goal is to predict elliptical tube quantities of interest and uncertainty, without using any data from elliptical tube experiments.
Will make predictions in extrapolative regime

- **Initial experiments:**
  - 1 ns, 3.8 kJ laser irradiates Be disk → plasma down Xe gas filled shock tube at ~ 200 km/s
  - Circular tube; diam = 575μm
  - Timing 13-14 ns

- **Wide tube experiments**
  - Laser energy ~ 3.8kJ
  - Circular tube; diam = 575, 1150μm
  - Timing 13, 20, 26 ns

- **Nozzle experiments**
  - Laser energy ~ 3.8kJ
  - nozzle length = taper length = 500μm
  - Circular tube; diam = 575μm
  - Timing 26 ns

- **Extrapolation experiments**
  - Laser energy ~ 3.8kJ
  - **Elliptical** tube; diam = 575-1150μm
  - Aspect ratio = 2
  - Includes nozzle in shock tube
  - Timing 26 ns
Have several outputs & inputs

- **Outputs** ($y$)
  - Shock location
  - Shock breakout time
  - Wall shock location
  - Axial centroid of Xe
  - Area of dense Xe

- **Inputs** ($x$)
  - Observation time
  - Laser energy
  - Be disk thickness
  - Xe gas pressure

- **Calibration parameters** ($\theta$)
  - Electron flux limiter
  - Laser scale factor

We measured and computed shocks at 13 ns.
Common thread: combining simulations, field observations for prediction, calibration and uncertainty quantification

1 ns, 3.8 kJ laser irradiates Be disk

Calibration: finding input parameter settings consistent with observations

Prediction of new observations with uncertainty
Basic calibration problem

• Have a computer model that is a function of inputs, $x$ and $t$
  – Design inputs ($x$): inputs that can be measured or adjusted in the field
  – Calibration inputs ($t$): inputs that are needed to run the model, may govern the behaviour of the physical system, but whose value is not known in the field

• The computer model represents the mean of a physical system

• Have noisy observations from the field that have been observed at $t = \theta$ and, possibly different values of $x$

• Unfortunately, $\theta$ is not known to the experimenters
Basic inverse problem

• **Goals:**
  – estimate the calibration parameters, $\theta$, using field observations and computational model
  – Construct predictive model for the physical system that accounts for all sources of uncertainty
Model calibration – Statistical formulation

\[ y_s(x, t) = \eta(x, t) \]
\[ y_f(x, \theta) = \eta(x, \theta) + \delta(x) + \epsilon \]

- Where,
  - \( x \) model or system inputs;
  - \( y_f \) system response
  - \( y_s \) simulator response
  - \( \theta \) calibration parameters
  - \( \epsilon \) random error

Have data from 2 separate sources – field observations and computer model outputs

Also have a model for systematic discrepancy, \( \delta(x) \)
Model calibration – Statistical formulation

\[ y_s(x, t) = \eta(x, t) \]
\[ y_f(x, \theta) = \eta(x, \theta) + \delta(x) + \epsilon \]

Gaussian process models
Aside: Use Gaussian processes for emulating computer model output

- GP’s have proven effective for emulating computer model output (Sacks et al., 1989)

- Emulating computer model output
  - output varies smoothly with input changes
  - output is noise free
  - passes through the observed response
  - GP’s outperform other modeling approaches in this arena
Aside: Why a statistical emulator?

- Can only run the code a limited number of times
  - where to run the code
  - how many times do you run the code

- Want to predict output with uncertainty at un-observed inputs... need foundation for statistical inference
Aside: Statistical formulation for GP emulation of computer models

- Computer model: \( \eta : \mathbb{R}^d \rightarrow \mathbb{R} \) Usually scale inputs to unit cube

- Function is expensive so, get to observe a sample of \( n \) runs from the computer model

- Specify a set of inputs where we will run the code \( x_1, x_2, \ldots, x_n \)

- Run the code and get the outputs \( y^T = (y_1, y_2, \ldots, y_n) \)
Aside: Statistical formulation for GP emulation of computer models

• Will view the computer code as a single realization of a Gaussian process:

\[ y(x) = \mu + z(x) \]

where,

\[ E(z(x)) = 0 \]
\[ \text{Var}(z(x)) = \sigma^2 \]
\[ z(x) \sim N(0, \sigma^2) \]
\[ \text{Corr}(z(x), z(x')) = \prod_{i=1}^{d} e^{-\phi_i(x_i - x'_i)^2} \]

• For \( n \) data points, will have the covariance matrix, \( \Sigma = \sigma^2 R \)

\[ y \sim N(\mu 1_n, \Sigma) \]
Aside: Realizations of a GP for a fixed model

\[ \mu = 0 ; \sigma^2 = 25 ; \phi = 52 \]
Aside: The parameters have meaning

- The mean, $\mu$, is the mean over all realizations
- Making the variance, $\sigma^2$, larger re-scales the vertical axis
- If $\phi_i = 0$, the function does not vary with respect to this input
- When $\phi_i$ is big, the function will be wigglier (a technical term?)
- Response where the inputs are close together will be more highly correlated than inputs that are far apart
Aside: Can emulate computer model with uncertainty

True function and observations

True function, emulated mean function and 95% prediction intervals
Hierarchical model is used to combine simulations and observations with discrepancy

- View computational model as a draw of a random process (again!!)

- Denote vectors of simulation trials as \( \mathbf{y}_s \) and field measurements as \( \mathbf{y}_f \) respectively

- Suppose that these are \( n \) and \( m \) -vectors respectively

- Can combine sources of information using a single GP

\[
\begin{align*}
    y_s(x, t) &= \eta(x, t) \\
    y_f(x, \theta) &= \eta(x, \theta) + \delta(x) + \epsilon
\end{align*}
\]

\[
\begin{pmatrix} 
    \mathbf{y}_s \\
    \mathbf{y}_f
\end{pmatrix} \sim MVN \left( \mu, \Sigma_\eta + \Sigma_\delta + \Sigma_\epsilon \right)
\]
Estimation of model parameters

• Approach is Bayesian so need prior distributions for model parameters needed
  – Inverted-gamma priors for variance components
  – Beta priors for the correlation parameters
  – Log-normal priors for the calibration parameters

• Samples from the posterior are generated through MCMC (have used single site Metropolis and Hamiltonian MCMC)
Calibration idea: Discrepancy model

Adams et al., 2012
**CRASH: Have observations and simulations**

- **Observations:**
  - 20 experiments

- **CRASH Simulations from run-sets 12 and 13**
  - **Experiment variables:** Be thickness, Laser energy, Xe fill pressure, Observation time, Aspect ratio, Nozzle length, Tube diameter and Taper length
  - **Calibration parameters:** Electron flux limiter, Laser energy scale factor

- **Code results and previous calibration studies indicated outputs insensitive to the electron flux limiter**

- Electron flux limiter constant in run-set 13 simulations
Simulations and Experimental Data

8 X’s and 1 θ; y is shock location (for a specific time)

- Run-set 12 sims
- Run-set 13 sims
- Observations
- Year 5 experiments
One-dimensional sensitivity plots
(prediction is an extrapolation in aspect ratio)
Sample inputs

Sample parameters from \( \pi(\text{parameters} | \text{data}) \)

Sample emulator
\( \eta(x, \theta) \)

Adjust for discrepancy
\( \eta(x, \theta) + \delta(x) \)

Include observation error
\( \eta(x, \theta) + \delta(x) + \varepsilon \)

Prediction strategy
## Input sources of uncertainty

<table>
<thead>
<tr>
<th>Variables</th>
<th>Nominal value</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Design Variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laser energy</td>
<td>3800 J</td>
<td>Normal Day to day stdev = 65.0 J</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Within day stdev = 49.4 J</td>
</tr>
<tr>
<td>Be thickness</td>
<td>21 microns</td>
<td>Uniform[20.5,21.5] microns</td>
</tr>
<tr>
<td>Xe Pressure</td>
<td>1.15 atm</td>
<td>Normal stdev = 0.1 atm</td>
</tr>
<tr>
<td>Tube diameter</td>
<td>1150 microns</td>
<td></td>
</tr>
<tr>
<td>Taper length</td>
<td>500 microns</td>
<td></td>
</tr>
<tr>
<td>Nozzle length</td>
<td>500 microns</td>
<td></td>
</tr>
<tr>
<td>Aspect ratio</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td><strong>Calibration Variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laser energy scale factor</td>
<td>Found via calibration. Range [0.6,1]</td>
<td></td>
</tr>
<tr>
<td>Electron flux limiter</td>
<td>Fixed for RS 13. Assumed constant (0.06)</td>
<td></td>
</tr>
</tbody>
</table>
Calibration model successfully predicts experiments with aspect ratio = 1

Parameter and input uncertainty
Parameter, input, discrepancy and observational uncertainty
Calibrated predictions
(aspect ratio = 2)
Post-Mortem

- Median prediction = 3401.3
- All but one observation is within 10% of median prediction
- Largest value differs from median by 10.92%
Post-Mortem

- Model clearly under-predicts shock location when aspect ratio = 2
  - Discrepancy model did not anticipate
  - Response does show a small change as aspect ratio grows from 1 to 2
  - We have only model calculations and judgment to support this idea
  - No experiments to assess model ability to account for aspect ratios > 1
Issues – searching for deficiencies

- Discrepancy reaches a limit in terms of predictive uncertainty

- The discrepancy process is a GP, but is informed by comparing code to data... is likely too ambitious a model for the amount of data that we had

- In the extrapolative regime, there is no information about how large the discrepancy should grow above and beyond the statistical model

- No obvious ways to include other variables not measured or modified in the training experiments
Thanks for your time