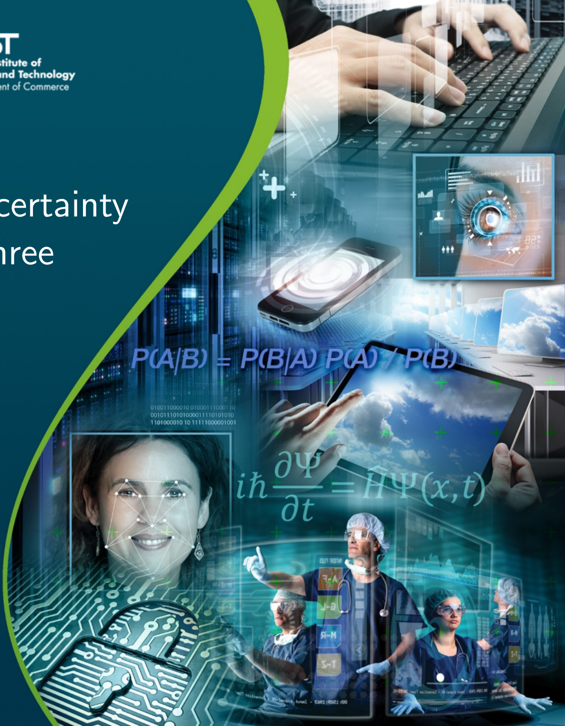
INFORMATION TECHNOLOGY LABORATORY

NIST National Institute of Standards and Technology U.S. Department of Commerce

Statistical Model Uncertainty in Measurements: Three Examples



Adam L. Pintar

October 5, 2017

Background & Disclaimer

- What I mean by model uncertainty
- Recent project to study model uncertainty in NIST work (and more broadly)
- Nothing I consider new or novel
  - Established procedures that I prefer when faced with model uncertainty
- Reflects my own preference for including model uncertainty
- From the Bayesian point of view
- Focused on measurement uncertainty

Options for Model Uncertainty

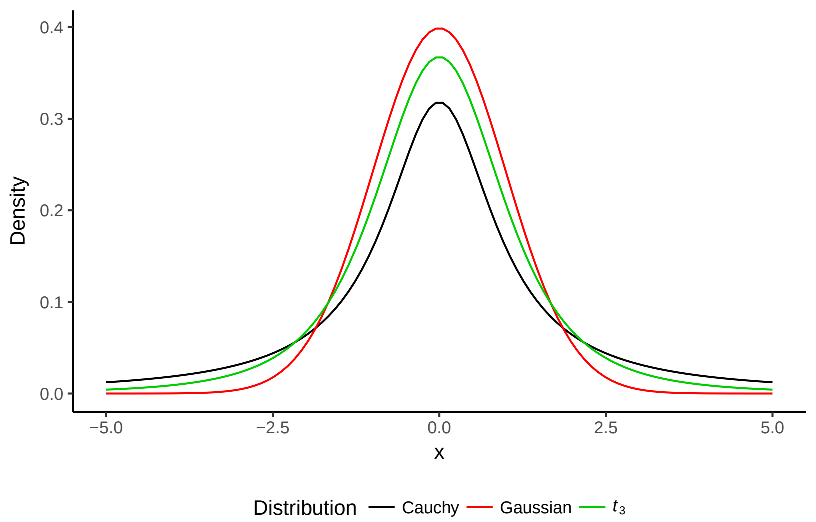
- Model Selection
- Model Averaging
- Non-parametric
- Enveloping
- Anything else?

Continuous Model Expansion

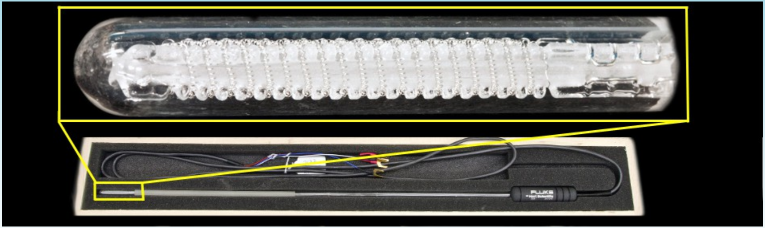
- Bridging model that continuously links each model in a collection
- As opposed to model averaging
  - For improper priors, the posterior model probabilities (weights) are not well defined
  - Computationally daunting with large collections

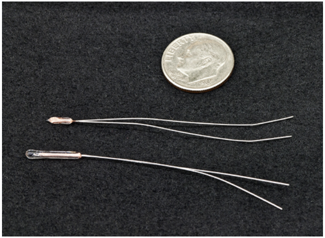
Students-t Distribution

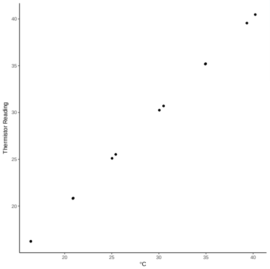
- Bridges the Cauchy and Gaussian distributions



Thermistor Calibration







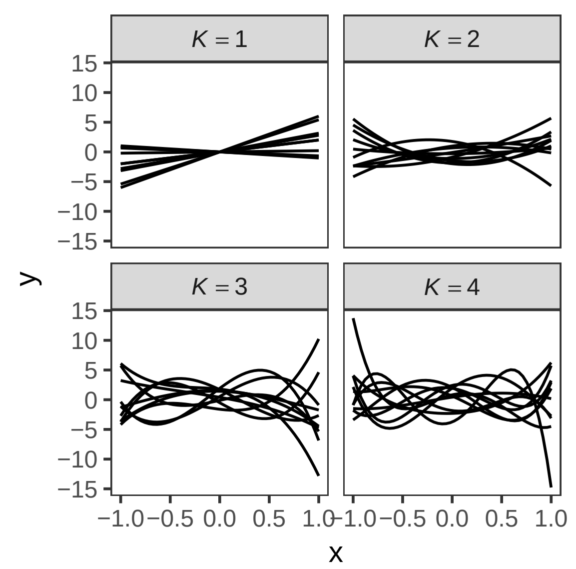
Polynomial Models

$$y = \alpha + \sum_{k=1}^{K} \beta_k P_k(x) + \epsilon$$

- $P_k(x)$ - orthogonal polynomial of order k
- β_k - coefficients to be estimated
- K controls the bias-variance trade-off

Simulated Polynomials

- $\alpha = 0$



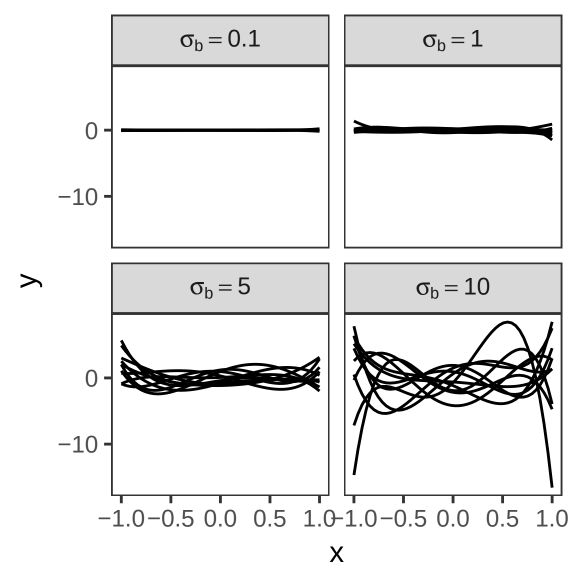
Alternative to Choosing K

$$y = \alpha + \sum_{k=1}^{K} \beta_k P_k(x) + \epsilon$$
$$\beta_k \sim \text{Gaussian}(0, \sigma_b^2)$$

- Set large K
- "Shrink" coefficients toward zero
- Ridge regression
- σ_b controls bias-variance trade-off
- Uncertainty in σ_b may be interpreted as model uncertainty

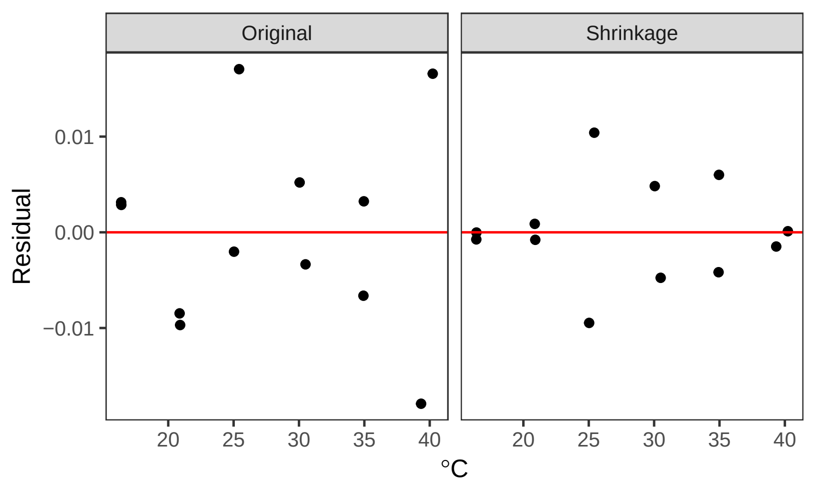
Simulated Polynomials

- $K = 4$ and $\alpha = 0$

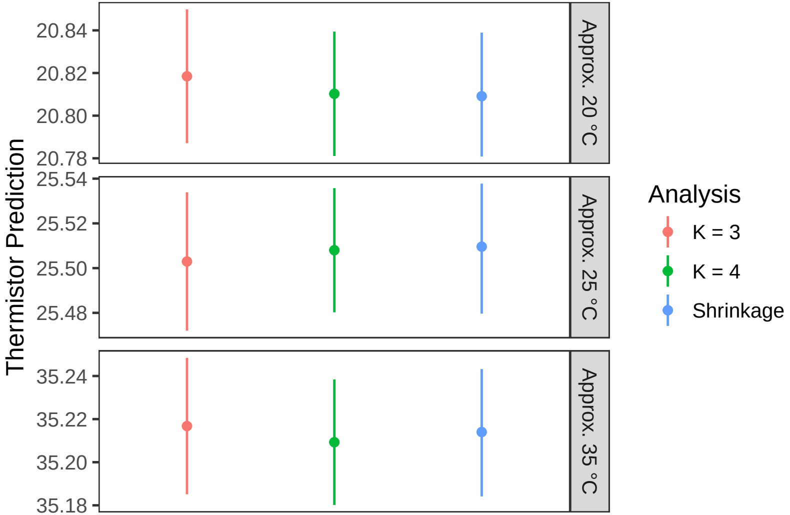


Calibration Curve Residuals

- Original: $K = 3$ selected from $K \leq 6$
- Shrinkage: $K = 6$



Calibration Curves up Close



- The lines depict 95% prediction intervals

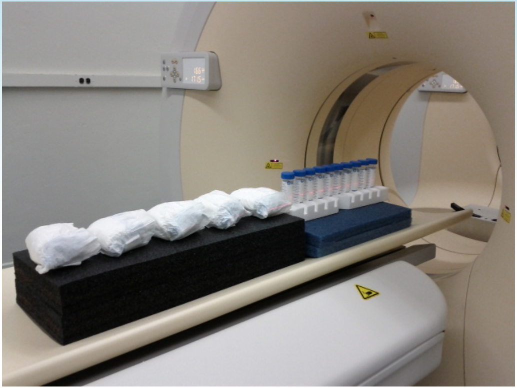
Calibration Example Summary

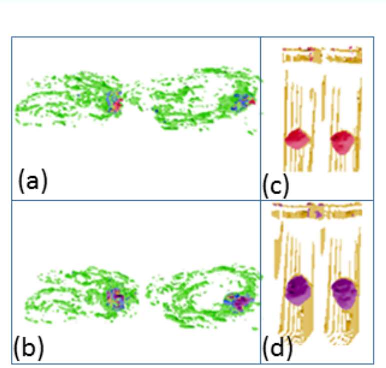
- Choice of polynomial degree
- Original analysis selected the “best” degree
- Shrinkage
    - Fit one model instead of six, but with an extra shrinkage parameter
    - Uncertainty in the shrinkage parameter may be interpreted as model uncertainty

Length vs. Volume for CT Tumor Measurements

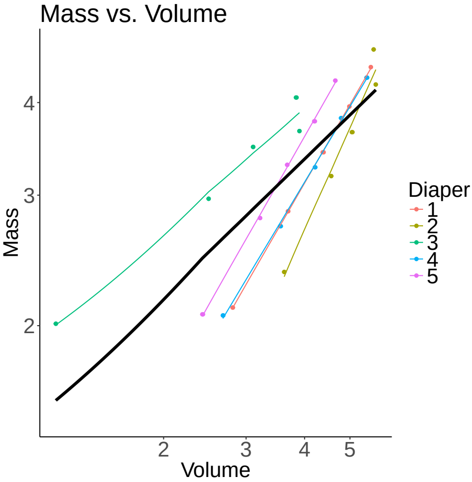
- Clinical goal is detecting tumor growth
- Dimensional measurements using CT images may face difficulties
  - Soft edges
  - Irregular shapes
- Mass can serve as a gold standard in phantom studies
- But mass is unavailable in clinical settings
- How good is CT measured volume as a surrogate?
- How good is RECIST (length) as a surrogate?

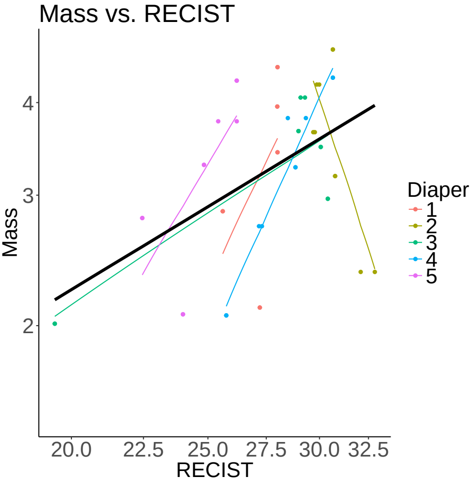
Tumor Phantom Experiment





Diaper Measurements





Two Models

- Homogeneous diapers

$$\log\{\text{mass}\}_{ij} = \beta_0 + \beta_1 \log\{\text{volume}\}_{ij} + \epsilon_{ij}$$

- Heterogeneous diapers

$$\log\{\text{mass}\}_{ij} = \beta_{0,i} + \beta_{1,i} \log\{\text{volume}\}_{ij} + \epsilon_{ij}$$

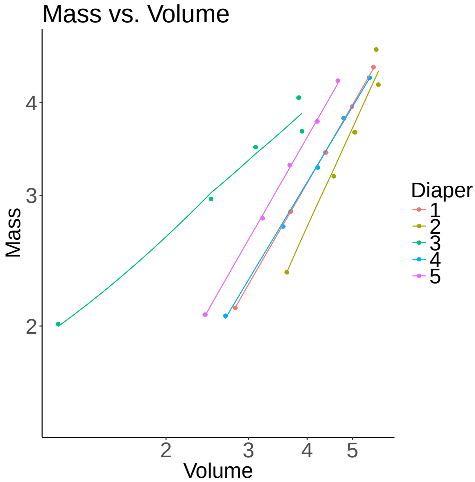
- Similar for RECIST

Multi-level Model (Bridging Model)

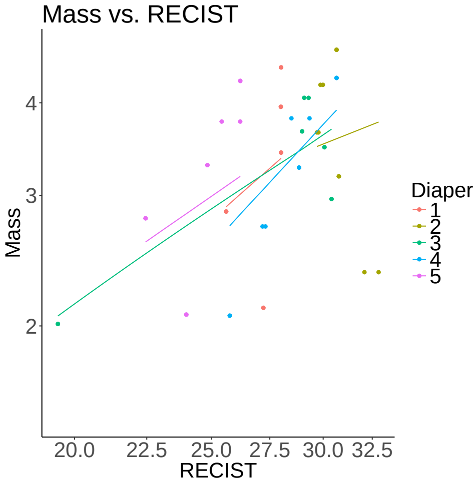
$$
\begin{aligned}
\log\{\text{mass}\}_{ij} &= \beta_{0,i} + \beta_{1,i}\log\{\text{volume}\}_{ij} + \epsilon_{ij} \\
\beta_{0,i} &\sim \text{Gaussian}(\beta_0, \sigma_0^2) \\
\beta_{1,i} &\sim \text{Gaussian}(\beta_1, \sigma_1^2)
\end{aligned}
$$

- Interpret uncertainty in β_0, β_1, σ_0, and σ_1 as model uncertainty
- Error in volume measurement
- Similar for RECIST

Results



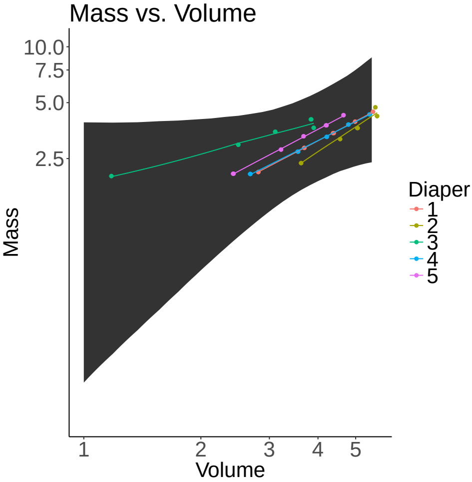
Mass vs. Volume
Mass
Volume
4
3
2
2
3
4
5
Diaper
1
2
3
4
5

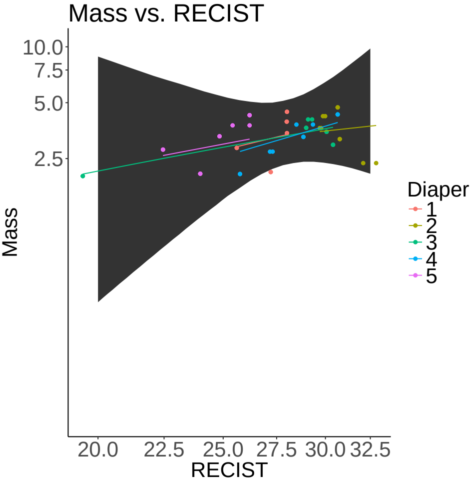


Mass vs. RECIST
Mass
RECIST
4
3
2
20.0
22.5
25.0
27.5
30.0
32.5
Diaper
1
2
3
4
5

Added Benefit

- Prediction of results from new experiment
- The grey regions depict 95% pointwise prediction intervals





Tumor Phantom Example Summary

- Two similar data sets
- Separate lines for one data set
- Single line for the other
- Random coefficients model forms a bridge
- “Pooling” information across diapers
- Interpret uncertainty in the amount of pooling as model uncertainty

Standard Reference Materials (SRMs)

Heterogeneity vs. Homogeneity

- Homogeneous model

$$y_{ij} = \mu + \epsilon_{ij}$$

- Heterogeneous model

$$y_{ij} = \mu_i + \epsilon_{ij}$$

Bridging Model

$$
\begin{aligned}
y_{ij} &= \mu_i + \epsilon_{ij} \\
\mu_i &\sim \text{Gaussian}(\mu, \sigma_\mu^2)
\end{aligned}
$$

- In small samples the homogeneous model may not recieve much weight

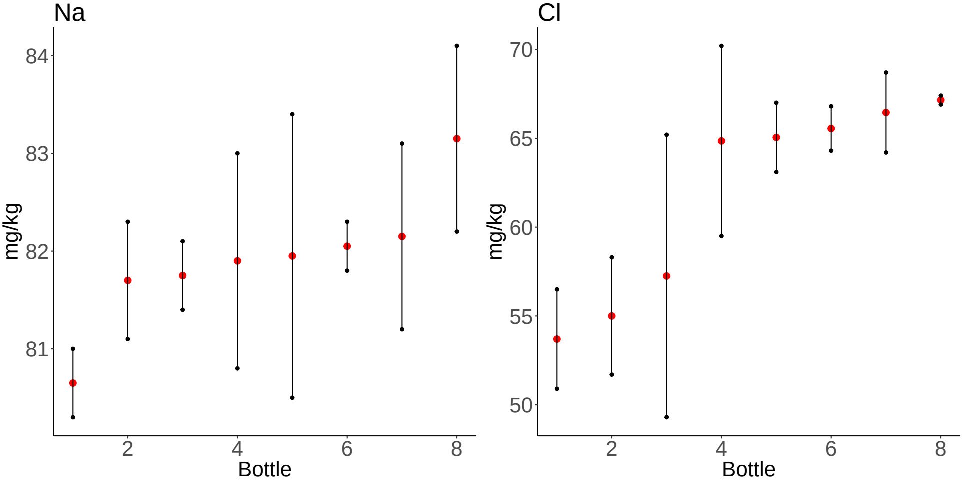
Spike and Slab Prior

$$\pi(\sigma_\mu | p) = \begin{cases} p & \text{if } \sigma_\mu = 0 \\ (1-p)f(\sigma_\mu) & \text{otherwise} \end{cases}$$

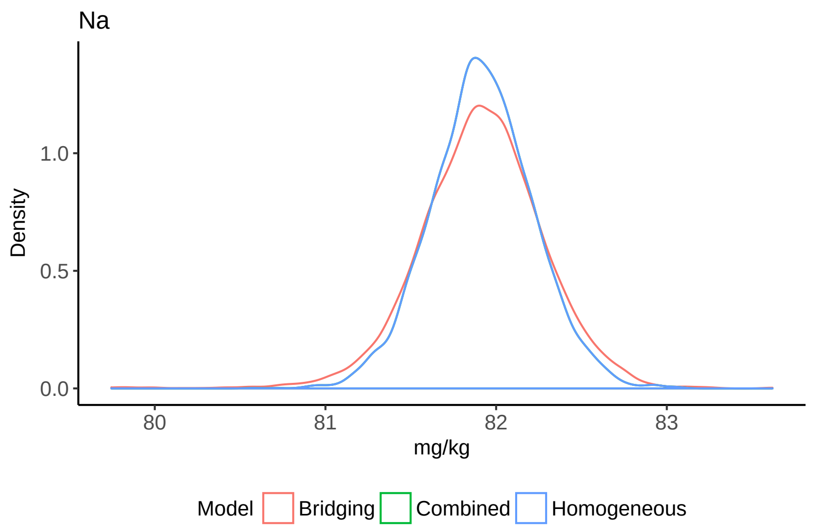
$$p \sim \text{Uniform}(0, 1)$$

- Equivalent to averaging the homogeneity and bridging models
- Computation
  - Full Bayesian – MCMC
  - Approximate – BIC or Stacking

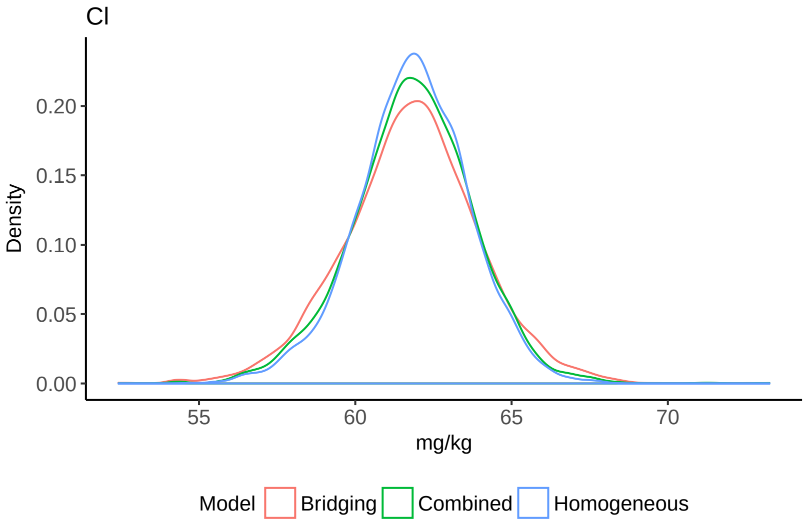
SRM 2718a Green Petroleum Coke



Results



Na
Density
1.0
0.5
0.0
80
81
82
83
mg/kg
Model
Bridging
Combined
Homogeneous


Cl
Density
0.20
0.15
0.10
0.05
0.00
55
60
65
70
mg/kg
Model
Bridging
Combined
Homogeneous

SRM Example Summary

- Chemists strive to achieve homogeneous units
- Bridging model may down weight homogeneity model in small samples
- Emphasize it with spike and slab prior
- Model uncertainty captured in spike weight and σ_μ