Calibration and Validation of Computer Models: a Bayesian Approach Lecture 4: Extensions

Rui Paulo

ISEG/CEMAPRE Technical University of Lisboa, Portugal

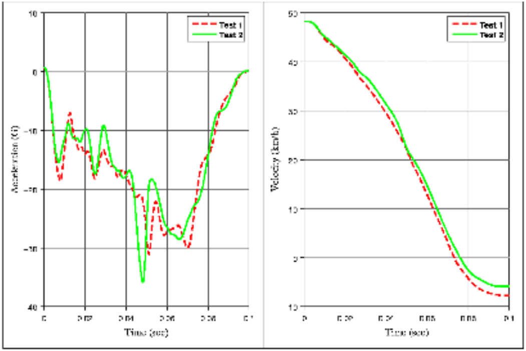
July 1 2011

- The principles of the statistical methodology that we developed are quite general
- The implementation details are, however, problem-specific
- Extensions: functional output, multivariate output
- Applications: Crash and Roadload

Bayarri et al. (2009) JASA

- The Crash computer model simulates the effect of a collision of a vehicle against different types of barriers
- The computer model plays an important role in the design of the vehicle, ensuring it's worth constructing and crashing a prototype
- Main focus: velocity changes after impact at key positions on the vehicle

Acceleration (left) and velocity (right); frontal impact; 30 mph



Finite element

Difficulties

- The methodology we developed was for real-valued output; the output here is a (smooth) function
- Hierarchical modeling: model and field at different conditions not completely quantifiable: different impact barriers; borrowing strength

Inputs and data

- (x_1, x_2)=(impact velocity, barrier type)
- Barrier type can be straight frontal, left angle, right angle, center pole
- Our methodology does not allow for such qualitative inputs so we will resort to a hierarchical analysis
- If we restrict our data to left, right and frontal impact, we can use a quantitative input such as “angle of impact”
- for the most part: $x =$ impact velocity and we only look at straight frontal impact
- Main output: relative velocity of the SDM situated under the driver’s seat: integrate acceleration and subtract impact velocity

Incorporating functional output

- Common approach: represent functions through basis expansion — Roadload
- These velocity curves are quite smooth
- Sample the function at a grid of time points $D^T = \{t_1, \ldots, t_N\}$
- It will turn out important that field and computer model curves are sampled at the same time points
- Input vector is now $\mathbf{z} = (x, t)$
- Similar approaches: Conti and O'Hagan (2010), Rougier (2007) and McFarkand et al. (2008)

- Field and model velocity curves are sampled at the same time points D^T
- Impact velocities for field experiments are in D^F; impact velocities for computer experiments in D^M
- Hence, data are

$$\mathbf{y}^M = (y^M(x,t):\ x \in D^M, t \in D^T)$$
$$\mathbf{y}^F = (y^M(x,t):\ x \in D^F, t \in D^T)$$

- Interesting evaluation criteria: CRITV — Velocity calculated 30ms before the displacement reaches 135 mm

Emulation

- Gaussian process surface approximation as described in Lecture 2
- $\mu(x,t) = \theta^L xt$
- Separable correlation structure
- Problem: dimension of the correlation matrices is $\#D^M \times \#D^T$
- Key simplification: $c^M(D^M \times D^T, D^M \times D^T) = c_x^M(D^M, D^M) \otimes c^T(D^T, D^T)$ where $\otimes$ stands for the Kronecker product:

$$\mathbf{A} \otimes \mathbf{B} = [a_{ij}B]$$

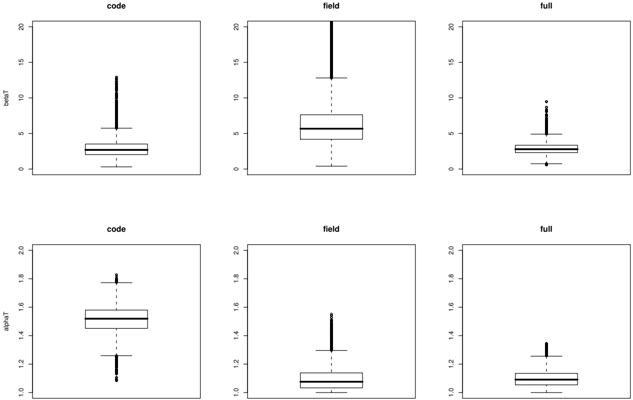
- Facts: $|A \otimes B| = |A|^{n_B}|B|^{n_A}$; $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$
- Exploited in spatiotemporal modeling and in Williams et al. (2006) and Rougier (2007) in computer modeling
- Here we do not resort to an MLE-modular approach

Statistical model

- Reality and bias: $y^R(x,t) = y^M(x,t) + b(x,t)$
- Field output functions:

$$y^F(x^{\star i}, t) = y^R(x^{\star i}, t) + \varepsilon_i(t)$$

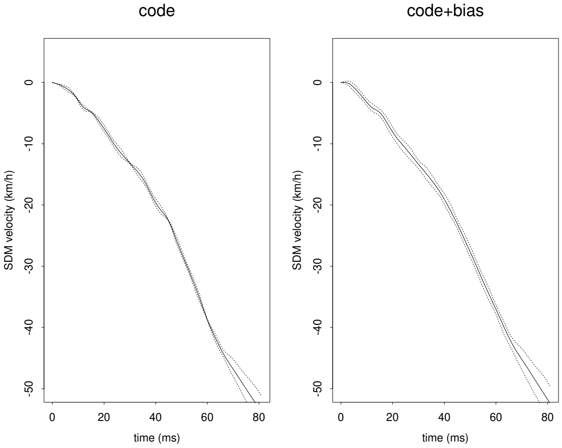
  where $\varepsilon_i(t)$ are independent realizations from a GP prior with mean zero, precision λ^F and correlation function $c^T(t,s) = \exp(-\beta^T |t-s|^{\alpha_T})$
- Bias: GP with $\alpha_x^b = 2$ and mean μ^b
- Key assumption: GP correlation parameters associated with time — α^T and β^T — are the *same* for the computer model, field error, and bias term


code
field
full
betaT
0
5
10
15
20
code
field
full
alphaT
1.0
1.2
1.4
1.6
1.8
2.0

Full Bayesian analysis

- The data is $\mathbf{y} = (\mathbf{y}^F, \mathbf{y}^M)$ which is Gaussian with covariance matrix $\boldsymbol{\Sigma} \otimes c^T(D^T, D^T)$ and mean $\mathbf{X}'\boldsymbol{\theta}$, where $\boldsymbol{\theta} = (\mu^M, \mu^b)$
- $\pi(\theta) \propto 1$
- smoothness parameters α are independent uniform on $(1, 2)$
- precision and range parameters: independent exponentials centered at (a multiple) of the marginal maximum likelihood estimates

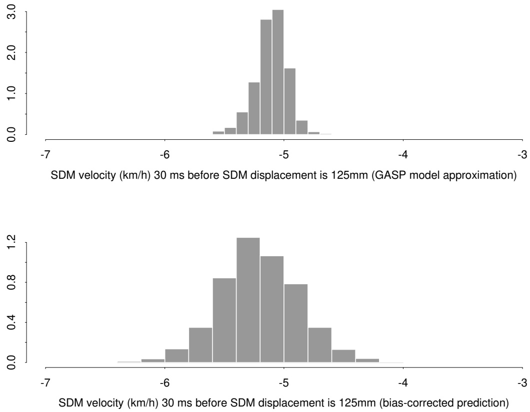
SDM velocity for straight frontal; impact velocity 56.3 km/h



Prediction of CRITV

- Simulation based inference makes it easy to predict complicated functions of $y^R(x, t)$
- $\mathrm{DISP}(t) = -\int_0^t y^R(x, \nu)\ d\nu$
- $\mathrm{CRITV} = y^R(x, \mathrm{DISP}^{-1}(125) - 30)$
- We can produce a pure-model prediction and a bias-corrected prediction of CRITV

Prediction of CRITV for impact velocity 56.3 km/h (top is pure model; lower is bias-corrected)



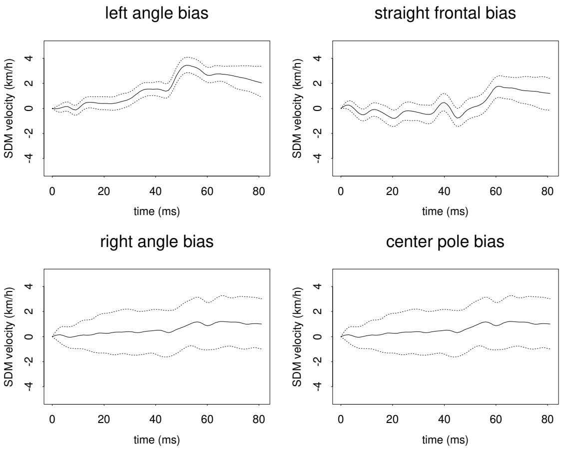
Hierarchical Modeling

- Jointly modeling the different impact barriers
- Alternative would be to use multivariate GP — more later
- Requires assumptions about how different output components are related: y_i^M and b_i, $i = 1, \ldots, K$ will be modeled using distributions drawn from common populations
- Combine information between different models, borrow strength, sharpen the individual analysis

Assumptions

1. Few model data for some configurations: common correlation parameters for the GP prior across the K configurations
2. Precisions λ^M and λ^F are also common
3. μ_i^M are assumed to arise from a two-stage hierarchical model
4. μ^b common
5. $\log(\lambda_i^b) \sim \mathrm{N}(\eta, 4q^2)$ — $q = 0.1$ means that the biases are expected to vary about 10%
6. correlation parameters for the bias processes are assumed to be common

Bias under 4 different barrier types, impact velocity 56.3 km/h



Roadload problem (Bayarri et al. 2007)

- Car is driven along a road seven times; time history of loads in the suspension system is recorded
- Car is specified by 7 inputs, subject to manufacturing variability

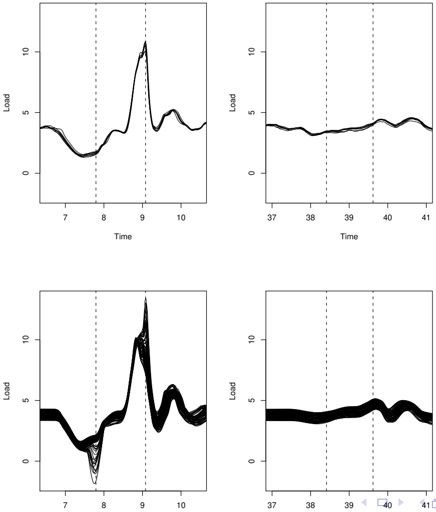
$$\mathbf{x}^* = \mathbf{x}_{\text{nom}} + \boldsymbol{\delta}^* \qquad y_r^F(\mathbf{x}^*; t), \quad r = 1, \ldots, 7$$

- Computer model has 2 additional calibration inputs, u_1 and u_2

$$y^M(\mathbf{x}_k, \mathbf{u}_k; t)$$

- Irregular functional output, one single uncertain controllable vector of inputs in the field
- In addition to 7 field curves, 65 output curves (LHD)

Field output (top) and model output (bottom)



Statistical modeling

- $y_r^F(\mathbf{x}_{\text{nom}} + \boldsymbol{\delta}^\star; t) = y^R(\mathbf{x}_{\text{nom}} + \boldsymbol{\delta}^\star; t) + \varepsilon_r(t)$; $\varepsilon_r(\cdot)$ independent zero mean GP
- $y^R(\mathbf{x}; t) = y^M(\mathbf{x}, \mathbf{u}^\star; t) + b(\mathbf{x}; t)$
- unknowns are $(y^M, \mathbf{u}^\star, \boldsymbol{\delta}^\star, b, V_\varepsilon)$ where V_ε is the covariance function of ε

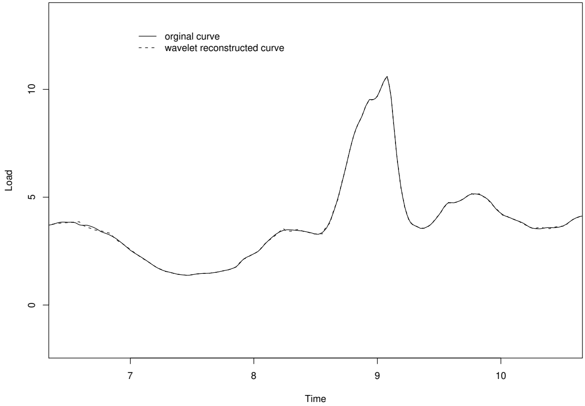
Wavelet Decomposition

$$y^M(\mathbf{x},\mathbf{u};t) = \sum_i w_i^M(\mathbf{x},\mathbf{u})\Psi_i(t)$$
$$y_r^F(\mathbf{x};t) = \sum_i w_{ir}^F(\mathbf{x})\Psi_i(t)$$

Using a tresholding procedure, leads to the approximations

$$y^M(\mathbf{x},\mathbf{u};t) = \sum_{i\in I} w_i^M(\mathbf{x},\mathbf{u})\Psi_i(t)$$
$$y_r^F(\mathbf{x};t) = \sum_{i\in I} w_{ir}^F(\mathbf{x})\Psi_i(t)$$

Wavelet reconstruction and original curve



Statistical model

Matching coefficients in the statistical model, we get

$$
\begin{aligned}
w_i^R(\mathbf{x}) &= w_i^M(\mathbf{x}, \mathbf{u}^*) + w_i^b(\mathbf{x}) \quad \forall i \in I \\
w_{ir}^F(\mathbf{x}) &= w_i^R(\mathbf{x}) + \varepsilon_{ir} \quad \forall i \in I \,.
\end{aligned}
$$

- ε_{ir} are Gaussian with mean zero and independent across r
- we assume that they are also independent across i with possibly different variance σ_i^2

Emulation

- Now the *wavelet coefficients* get a Gaussian process prior: for each $i \in I$,

$$w_i^M(\cdot) \sim \mathrm{GP}(\mu_i, \frac{1}{\lambda_i^M} c_i^M(\cdot, \cdot))$$

- For each i, find the wavelet coefficients of the model data $\{w_i(\mathbf{x}_k, \mathbf{u}_k)\}$ and compute the corresponding posterior $w_i(\cdot, \cdot) \mid \{w_i(\mathbf{x}_k, \mathbf{u}_k)\}$
- $\# I = 289$, which leads to $[2 \times 9 + 2] \times 289 = 5780$ parameters
- Fix parameters at MLE $\hat{\boldsymbol{\theta}}_i = (\hat{\lambda}_i^M, \hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)$

$$w_i(\mathbf{z}) \mid \mathbf{w}_i^M, \hat{\boldsymbol{\theta}}_i \sim \mathrm{N}(\hat{m}_i(\mathbf{z}), \hat{V}_i^M(\mathbf{z}))$$

Priors

- Because only one value of $\mathbf{x}_{\rm nom}$ is evaluated in the field, w_i^b are constant
- Each wavelet coefficient w_i^b belongs to a resolution level j. Those are modeled as

$$w_i^b \sim \mathrm{N}(0, \tau_j^2)$$

- $\pi(\tau_j^2 \mid \{\sigma_i^2\}) \propto (\tau_j^2 + \bar{\sigma}_j^2/7)^{-1}$

MCMC

$$\pi(w^M(\boldsymbol{\delta}^\star, \mathbf{u}^\star), \mathbf{w}^b, \boldsymbol{\delta}^\star, \mathbf{u}^\star, \boldsymbol{\sigma}^2, \boldsymbol{\tau}^2 \mid D) =$$
$$\pi(w^M(\boldsymbol{\delta}^\star, \mathbf{u}^\star) \mid \mathbf{w}^b, \boldsymbol{\delta}^\star, \mathbf{u}^\star, \boldsymbol{\sigma}^2, \boldsymbol{\tau}^2 D)$$
$$\pi(\mathbf{w}^b \mid \boldsymbol{\delta}^\star, \mathbf{u}^\star, \boldsymbol{\sigma}^2, \boldsymbol{\tau}^2, D)$$
$$\pi(\boldsymbol{\delta}^\star, \mathbf{u}^\star, \boldsymbol{\tau}^2 \mid \boldsymbol{\sigma}^2, D)$$
$$\pi(\boldsymbol{\sigma}^2 \mid D)$$

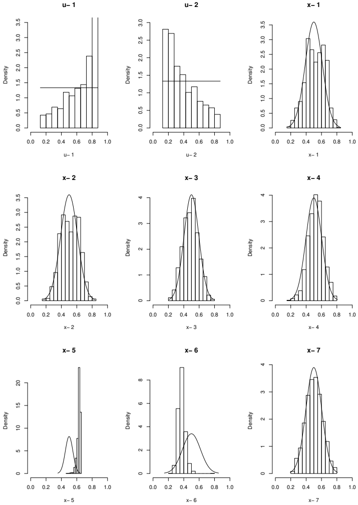
where

$$\pi(\boldsymbol{\sigma}^2 \mid D) \propto \prod_{i \in I} \frac{1}{(\sigma_i^2)^3} \exp[-s_i^2/(2\sigma_i^2)] \times \int L(\bar{\mathbf{w}}^F, \mathbf{s}^2 \mid \boldsymbol{\delta}^\star, \mathbf{u}^\star, \boldsymbol{\sigma}^2, \boldsymbol{\tau}^2)\, d\boldsymbol{\delta}^\star\, d\mathbf{u}^\star\, d\boldsymbol{\sigma}^2\, d\boldsymbol{\tau}^2$$

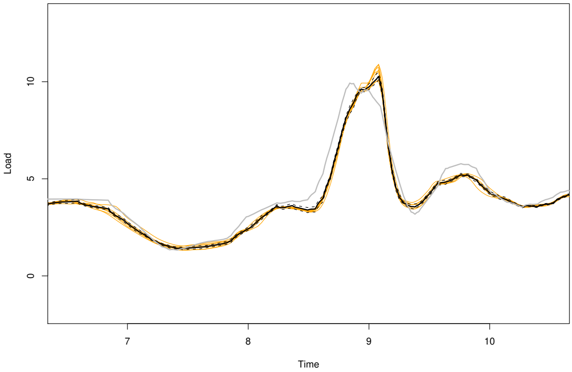
Modularization

- All except $\pi(\boldsymbol{\delta}^\star, \mathbf{u}^\star, \boldsymbol{\tau}^2 \mid \boldsymbol{\sigma}^2, D)$ and $\pi(\boldsymbol{\sigma}^2 \mid D)$ are standard form
- We ignore the integral in $\pi(\boldsymbol{\sigma}^2 \mid D)$ and estimate the σ_i^2 using replicate information
- bias can be replaced by larger σ_i^2 and this prevents this from happening
- Modularization techniques — Liu et al. (2009)

Posterior of $\boldsymbol{\delta}^\star$ and $\mathbf{u}^\star$

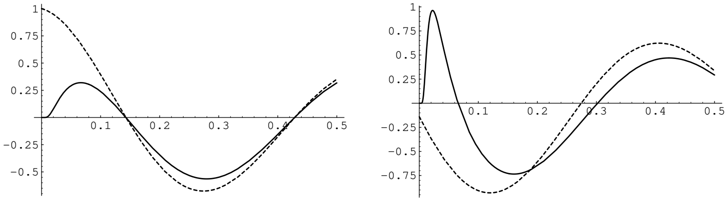


Bias-corrected prediction



Example

- $\mathbf{y}^R(u) = (y_1^R(u), y_2^R(u))'$; $\mathbf{y}^M(u) = (y_1^M(u), y_2^M(u))'$
- Reality: solid; Model: dashed (Santner et al 2003)



Example

- Model observed at 5 equally spaced points
- "True value": $u^\star = 0.2$; $\mathbf{y}^R(u^\star) = (-0.346, -0.649)$
- Simulate 7 replicates of $\mathbf{y}^F$: $\mathbf{y}_k^F(u^\star) \sim N(\mathbf{y}^R(u^\star), \mathbf{\Sigma}^F)$ with

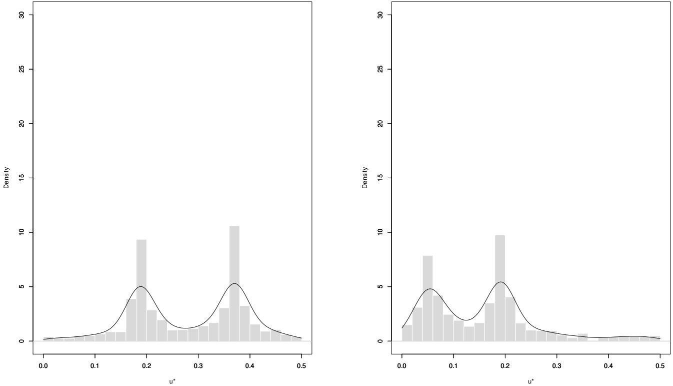
$$\mathbf{\Sigma}^F = \begin{pmatrix} 1/400 & 0.5/400 \\ 0.5/400 & 1/400 \end{pmatrix} .$$

- Goal: estimate u and predict $\mathbf{y}^R$

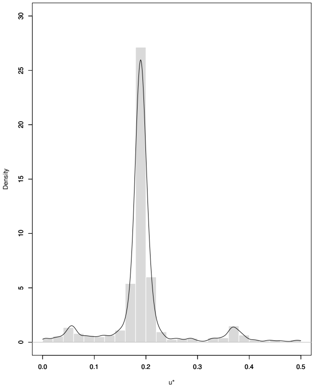
Example

- $y_1^M(u) = y_1^R(u^\star) \Leftrightarrow u \in \{0.185, 0.372\}$
- $y_2^M(u) = y_2^R(u^\star) \Leftrightarrow u \in \{0.050, 0.194\}$
- The posterior for u using data from one of the components only is potentially bimodal
- The two analysis are probably not compatible
- Added uncertainty

Results I


Density
0
5
10
15
20
25
30
0.0
0.1
0.2
0.3
0.4
0.5
u*
Density
0
5
10
15
20
25
30
0.0
0.1
0.2
0.3
0.4
0.5
u*

Results II



Results III

	$\hat{y}_1^R(u^\star)$		$\hat{y}_2^R(u^\star)$	
Approach	Mean	0.95% CI	Mean	95% CI
Separate	-0.372	(-0.621, -0.285)	-0.673	(-1.030, -0.417)
Combined	-0.354	(-0.401, -0.309)	-0.663	(-0.736, -0.593)

Highlights

- Single calibration strategy which results from combining all the available information
- Common features are reinforced; others are smoothed out
- Reduced uncertainty in posterior predictions

Possible approaches

- Multivariate GP: cross-covariance functions are difficult to specify
- Separability often assumed, otherwise computationally too demanding
- Sample size issues
- Qian et al. (2008), Higdon et al. (2008), Conti and O'Hagan (2010), Fricker et al. (2010)

Linear models of Coregionalization

- Gelfand et al (2004)
- p-dimensional Gaussian process:

$$\mathbf{y}^M(\mathbf{z}) = \boldsymbol{\mu}(\mathbf{z}) + \mathbf{A}\mathbf{w}(\mathbf{z}) + \varepsilon$$

- $\boldsymbol{\mu}(\mathbf{z}) = \left(\mathbf{I}_p \otimes \mathbf{h}(\mathbf{z})^t\right)\boldsymbol{\eta}, \quad \mathbf{h}(\mathbf{z}) \in \mathbb{R}^s$
- $\mathbf{A}$ is $p \times r$ full column rank, $r \leq p$
- components $w_i(\mathbf{z})$ of $\mathbf{w}(\mathbf{z})$ are independent GP with mean zero, precision λ_i^M and correlation structure $c_i^M(\cdot, \cdot)$

LMC emulators

- Conti and O'Hagan: $c_i^M \equiv c^M$ which implies separability but allows for integrating out everything except β^M
- Fricker et al: c_i^M not equal: no longer separable; only mean can be integrated out
- Higdon et al: High dimensional output, dimensionality reduction since $r << p$; singular value decomposition of transformed model data

Our proposal

- If matrix $\mathbf{A}$ is known, we transform model data as $\mathbf{w}_D^M = \mathbf{A}^{-1}\mathbf{y}_D^M$
- $\mathbf{w}_D^M$ come from independent GP, so software for univariate GP can be used to estimate parameters
- Estimate matrix $\mathbf{A}$ via singular value decomposition
- With these estimates,

$$\mathbf{y}^M(\mathbf{z}) \mid \mathbf{y}_D^M, \widehat{\boldsymbol{\theta}}^M \sim \mathrm{N}_p(\mathbf{A}\,\widehat{\mathbf{m}}(\mathbf{z}), \mathbf{A}\,\widehat{\mathbf{V}}(\mathbf{z})\,\mathbf{A}^t)$$

where $\widehat{\mathbf{m}}(\mathbf{z}) = (\widehat{m}_i(\mathbf{z})', i = 1, \ldots, p)'$,
$\widehat{\mathbf{V}}(\mathbf{z}) = \mathrm{diag}(\widehat{V}_i(\mathbf{z}), i = 1, \ldots, p)$

"Hierarchical" Scenarios

- Output of model and field observations can be partitioned into groups that, given $\mathbf{u}$ can be considered independent:

$$\mathbf{y}^M(\mathbf{u}) = (\mathbf{y}_1^M(\mathbf{u}), \ldots, \mathbf{y}_L^M(\mathbf{u}))'$$

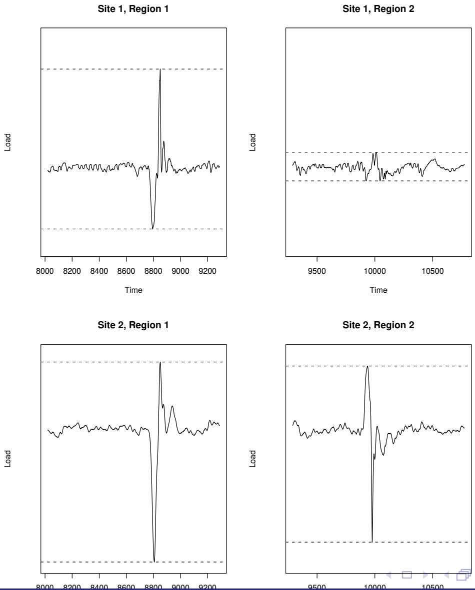
and

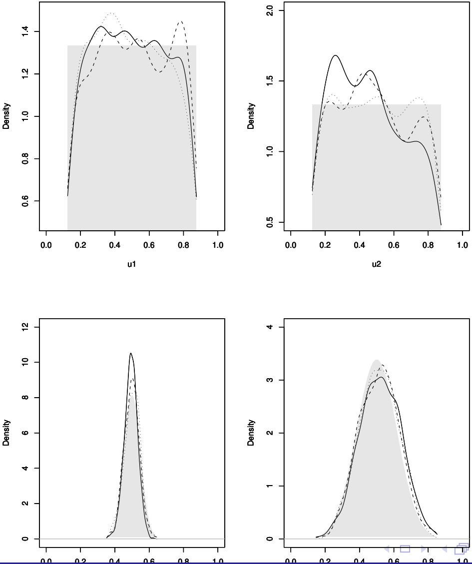
$$\mathbf{y}_k^F = (\mathbf{y}_{1k}^F(\mathbf{u}), \ldots, \mathbf{y}_{Lk}^F(\mathbf{u}))'$$

- If it were not for the fact that each of these components share the same $\mathbf{u}$, we could perform separate analysis
- Everything naturally extends to this setting just by adding a subscript and performing a product; $L = 1$ brings us back to the original formulation

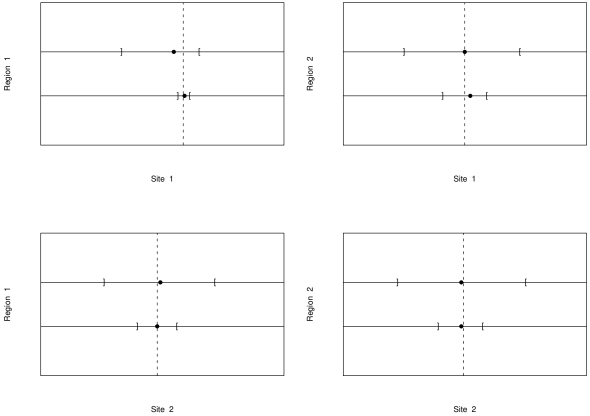
Application

- Prototype vehicle is driven along a road that has two potholes
- There are two sensors on two different sites on the vehicle registering the history of load — sites 1 and 2
- $\mathbf{x} = (x_1, x_2)'$; $\mathbf{u} = (u_1, u_2)'$; $K = 7$
- We are interested in the range of the load history at both sites when the car hits the two potholes
- Two different potholes are modeled as independent but the two sensors for the same pothole are not


Site 1, Region 1
Site 1, Region 2
Load
Load
8000 8200 8400 8600 8800 9000 9200
9500 10000 10500
Time
Time
Site 2, Region 1
Site 2, Region 2
Load
Load


Density
u1
Density
u2
Density
Density

Bias-corrected prediction of ranges: separate (top) and combined



Discussion

- Separately analyzing different components of the output may lead to conflicting calibration strategies and added uncertainty
- Methodology combines the separate analysis taking advantage of existing software
- Computationally less demanding; potentially scales up
- Combining possibly very different marginals into joint: copulas