Another take on the bias term

SAMSI programme on Complex Computer models
Methodology Subprogramme
February 2007
The Idea

• Math/physics model is

\[ \frac{\partial y}{\partial x} = f(u, x) \]  
(1)

• Computer model computes solution to (1):

\[ y^M(u, x) \]  
(2)

• Investigate whether source of bias is the fact that \( u = u(x) \)

• Problem: Solution to

\[ \frac{\partial y}{\partial x} = f(u(x), x) \]  
(3)

is in general not \( y^M(u(x), x) \).
- When code is slow, and it is not feasible to construct a fast approximation to the solution of (3),
  - Let $u(x) = u + b(x)$
  - $y^M(u + b(x), x) = y^M(u, x) + b(x) \frac{\partial y^M}{\partial u}(u, x)$

- Statistical model:

$$y^F(x_i) = y^M(u^*, x_i) + b(x_i) \frac{\partial y^M}{\partial u}(u^*, x_i) + \varepsilon_i$$  (4)

- Model (4) is intended as an approximation to Peter’s approach (Tomassini et al.,’07), and can also be seen as a version of the Kennedy and O’Hagan model with a more structured bias term.

- Computer model and its derivatives are used in the statistical model.

- Fast approximation to the output of the code but also to its derivative are needed.
Derivatives of Gaussian Processes

• a priori, \( y^M(\cdot) \sim \text{GP}(\mu(\cdot), \frac{1}{\lambda M} c^M(\cdot, \cdot)) \)

\[
\mu(x, u) = \Psi(x)' \theta^L
\]

\[
c^M((u, x), (v, z)) = \exp(\beta_1 |u - v|^\alpha_1) \exp(\beta_2 |x - z|^\alpha_2)
\]

• \( \partial y^M(u, x) \equiv \frac{\partial y^M}{\partial u}(u, x) \)

• \( \partial y^M \) is still a Gaussian process if \( \alpha_1 = 2 \) and

\[
E(\partial y^M(u, x)) = \frac{\partial \mu}{\partial u}(u, x)
\]

\[
\text{Cov}(\partial y^M(u, x), y^M(v, z)) = \frac{1}{\lambda M} \frac{\partial c^M}{\partial u}((u, x), (v, z))
\]

\[
\text{Cov}(\partial y^M(u, x), \partial y^M(v, z)) = \frac{1}{\lambda M} \frac{\partial^2 c^M}{\partial u \partial v}((u, x), (v, z))
\]
• which in the case at hand turn out to be

\[ E(\partial y^M(u, x)) = 0 \]

\[ \text{Cov}(\partial y^M(u, x), y^M(v, z)) = -\frac{2}{\lambda M^2} \beta_1 (u - v) c^M((u, x), (v, z)) \]

\[ \text{Cov}(\partial y^M(u, x), \partial y^M(v, z)) = \frac{2}{\lambda M} \beta_1 [1 - 2\beta_1 |u - v|^2] c^M((u, x), (v, z)) \]

• We could fix \( \alpha_2 \) at its mle in \((1, 2]\), but GaSP does not allow for fixing only a subset of the roughness parameters, so all roughness parameters are fixed at 2.
(Similar plot in Solak et al. 2003)
Toy Example

- \( \frac{dy(t)}{dt} = -uy(t); \ y(0) = y_0 \)

- Solution \( y^M(u, t) = y_0 \exp(-ut) \) is treated as an expensive computer model.

- Model and its derivatives with \( y_0 \equiv 5 \) are exercised at a 15-point Latin hypercube design in \([0.5, 2] \times [0.1, 3.0]\) in the \((u, t)\) space.

- The plots that follow have been produced by computing estimates of the parameters of the model using the code data only (not the derivative data)
Prediction Code, no deriv info, u=1.5

- Posterior Predictive Mean
- True Value
Prediction Code Output, $u=1.5$

![Graph showing prediction code output with posterior predictive mean and true value.](image)
Toy Example – Field data

- Field data simulated from

\[ y^F(t_i) = (y_0 - c) \exp(-u_* t_i) + c + \varepsilon_i \]

with \( y_0 = 5, \ c = 1.5, \ u_* = 1.7 \) and \( \varepsilon \sim \mathcal{N}(0, \sigma^2) \), \( \sigma = 0.3 \). Three replicates at each of 10 \( t_i \) time points.

- The model above can be rewritten as

\[ \frac{dy(t)}{dt} = -u(1 - c/y(t))y(t) \]

and so \( u = u(t) \).

- Hyperparameters for the prior on \( b \)?
Notation

- $D^M = \{(u_i, x_i)\} = \text{code design};$  $D^F = \{x_j^*\} = \text{field design}$;
- $y^M = y^M(D^M); \partial y^M = \partial y^M(D^M)$;
- $\bar{y}^F = \bar{y}^F(D^F), s_F^2 = \sum (y_{ij}^F - \bar{y}_i^F)^2, n_i \text{ replicates at each } x_i^*$
- $D_u^F = \{(u, x_j^*)\}$
- $y_*^M = y^M(D_u^F); \partial y_*^M = \partial y^M(D_u^F); b = (b(x_j^*))$
- \textit{a priori}, $b(\cdot) | \lambda^b, \beta^b \sim \text{GP}(0, \frac{1}{\lambda^b} \exp(-\beta^b|x - x^*|^2))$
(Augmented) Likelihood

\[
f(y^F, y^M, \partial y^M, y^*_M, \partial y^*_M, b \mid \theta^L, \theta^M, \lambda, \beta, \lambda^F, u) = \\
\lambda^F \chi^2 (\lambda^F s^2_F \mid \sum (n_i - 1)) \times \\
N(\bar{y}^F \mid y^*_M + b \circ \partial y^*_M, \text{diag } n^{-1} / \lambda^F) \times \\
f(y^M_M, \partial y^M_M \mid y^M, \partial y^M, \theta^L, \theta^M, u) \times \\
f(y^M, \partial y^M_M \mid \theta^L, \theta^M) \times \\
N(b \mid 0, c^b(D^F) / \lambda^b)
\]

where \( \circ \) stands for the Hadamard product of matrices, i.e., entry-wise product.
\[ \lambda^F | - \sim \Gamma(\lambda^F | a_1, a_2) \text{ where} \]

\[
a_1 = N_F/2 + \alpha_F
\]

\[
a_2 = r_F + s_F^2/2 + ||\bar{y}^F - y^M_\ast - b \circ \partial y^M_\ast||^2/2
\]

and, \textit{a priori}, \( \lambda^F \sim \Gamma(\alpha_F, r_F) \)

\[ \lambda^b | - \sim \Gamma(\lambda^b | a_1, a_2) \text{ where} \]

\[
a_1 = N_F/2 + \alpha_b
\]

\[
a_2 = r_F + b' \left[ c^b(D^F) \right]^{-1} b / 2
\]

and, \textit{a priori}, \( \lambda^b \sim \Gamma(\alpha_b, r_b) \)
• $b \mid - \sim N(m, V)$ where

$$V^{-1} = \lambda^F \text{diag}(\partial y^M_\star \circ \partial y^M_\star \circ n) + \lambda^b [c^b(D^F)]^{-1}$$

$$m = \lambda^F V \left[ \partial y^M_\star \circ n \circ (\bar{y}^F - y^M_\star) \right]$$

• $y^M_\star, \partial y^M_\star \mid - \sim N(m, V)$ where

$$V^{-1} = \lambda^F \begin{pmatrix} \text{diag } n & \text{diag}(n \circ b) \\ \text{diag}(n \circ b) & \text{diag}(n \circ b \circ b) \end{pmatrix} + \Sigma^{-1}_{*|\cdot}$$

$$m = V \begin{bmatrix} \lambda^F \begin{pmatrix} n \circ \bar{y}^F \\ n \circ b \circ \bar{y}^F \end{pmatrix} + \Sigma^{-1}_{*|\cdot} \mu_{*|\cdot} \end{bmatrix}$$

with $\Sigma_{*|\cdot}$ and $\mu_{*|\cdot}$ representing, respectively, the conditional covariance and mean of $(y^M_\star, \partial y^M_\star)$ given $(y^M, \partial y^M)$.
• To draw from the full conditional of $u$, the Metropolis-Hastings ratio involves

$$\frac{f(y^*_M, \partial y^*_M \mid y^M, \partial y^M, u)}{f(y^*_M, \partial y^*_M \mid y^M, \partial y^M, u)}$$

• The current implementation of this MCMC strategy is such that $u$ is not moving. The fact that there is very little uncertainty in the approximation to $y^M$ (due to the inclusion of the derivative information) is a likely explanation for this phenomenon. (Note also that the formula above does not involve the field data $\overline{y}^F$ explicitly.)
Integrate out $y^M_*$, $\partial y^M_*$ from the (augmented) likelihood to obtain

$$f(y^F, y^M, \partial y^M, b | \theta^L, \theta^M, \lambda^b, \beta^b, \lambda^F, u) = \lambda^F \chi^2 (\lambda^F s^2_F | \sum (n_i - 1)) \times$$

$$N(\bar{y}^F | m, V + \text{diag } n^{-1}/\lambda^F) \times$$

$$N(b | 0, c^b(D^F)/\lambda^b)$$

where, with $\Sigma_*$ partitioned as $(\Sigma_{ij})$, and $\mu_* = (\mu_1, \mu_2)$

$$m = \mu_1 + b \circ \mu_2$$

$$V = \Sigma_{11} + \text{diag } b \Sigma_{21} + (\text{diag } b \Sigma_{21})' + \text{diag } b \Sigma_{22} \text{ diag } b$$
Motivated by Jim’s example, an interesting possibility which is feasible and may alleviate the problem is to iteratively sample from

1. \( \lambda^b, \lambda^F, b \mid y^*_M, \partial y^*_M, u, y^F, y^M, \partial y^M \)

2. \( y^*_M, \partial y^*_M \mid \lambda^b, \lambda^F, b, u, y^F, y^M, \partial y^M \)

3. \( u \mid \lambda^b, \lambda^F, b, y^F, y^M, \partial y^M \)

To sample from 1, one uses a Gibbs sampler and all full-conditionals are closed form and given before; from 2, it’s direct; from 3, one proceeds as before but now we need to evaluate the ratio of Normal densities like

\[
N(\bar{y}^F \mid m, V + \text{diag } n^{-1}/\lambda^F)
\]

as specified in the previous slide.
Remarks:

- The full conditional of $\lambda^F$ is not closed for anymore, and evaluating it requires added matrix computations

$$[\lambda^F \mid -] \propto (\lambda^F)^{\sum(n_i-1)/2} \exp(-\lambda^F s^2_F/2) \left| V + \text{diag } n^{-1}/\lambda^F \right|^{-1/2} \times \exp \left\{ -(\bar{y}^F - m)'[V + \text{diag } n^{-1}/\lambda^F]^{-1}(\bar{y}^F - m)/\lambda^F \right\} \times \Gamma(\lambda^F \mid \alpha_F, r_F)$$

with $m = \mu_\star \cdot (1 + b)$

- The full conditional of $\lambda^F$ in the previous set up is most likely a good proposal, especially if there is reasonable number of replicates.

- Perhaps most importantly, the full conditional of $b$ is not closed form either, and finding a good proposal distribution here should be a more delicate problem.
Remarks:

- One can integrate out $b$ from the augmented likelihood, but that does not alleviate the problem of the full conditional of $u$ not depending on the field data:

$$
\begin{align*}
&f(y^F, y^M, \partial y^M, y_\star^M, \partial y_\star^M, | \theta^L, \theta^M, \lambda^b, \beta^b, \lambda^F, u) = \\
&\lambda^F \chi^2(\lambda^F s^2_F | \sum(n_i - 1)) \times \\
&N(\bar{y}^F | y^M_\star, \text{diag } \partial y^M c^b(D^F) \text{ diag } \partial y^M / \lambda^b + \text{diag } n^{-1} / \lambda^F) \times \\
&f(y^M_\star, \partial y^M_\star | y^M, \partial y^M, \theta^L, \theta^M, u) \times \\
&f(y^M, \partial y^M | \theta^L, \theta^M)
\end{align*}
$$

- In the case of an additive bias as in the Kennedy and O’Hagan model, one can simultaneously integrate out $b$ and $y^M_\star$. That does not seem to be the case here.