Another take on the bias term

SAMSI programme on Complex Computer models

Methodology Subprogramme

February 2007

The Idea

- Math/physics model is

$$\frac{\partial y}{\partial x} = f(u, x) \tag{1}$$

- Computer model computes solution to (1):

$$y^M(u, x) \tag{2}$$

- Investigate whether source of bias is the fact that $u = u(x)$
- Problem: Solution to

$$\frac{\partial y}{\partial x} = f(u(x), x) \tag{3}$$

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:

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

- 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*, $y^M(\cdot) \sim \mathrm{GP}(\mu(\cdot), \frac{1}{\lambda^M} c^M(\cdot,\cdot))$

$$\mu(x,u) = \Psi(x)'\theta^L$$
$$c^M((u,x),(v,z)) = \exp(\beta_1|u-v|^{\alpha_1})\exp(\beta_2|x-z|^{\alpha_2})$$

- $\partial y^M(u,x) \equiv \frac{\partial y^M}{\partial u}(u,x)$
- ∂y^M is still a Gaussian process if $\alpha_1 = 2$ and

$$\mathsf{E}(\partial y^M(u,x)) = \frac{\partial \mu}{\partial u}(u,x)$$
$$\mathsf{Cov}(Dy^M(u,x), y^M(v,z)) = \frac{1}{\lambda^M}\frac{\partial c^M}{\partial u}((u,x),(v,z))$$
$$\mathsf{Cov}(\partial y^M(u,x), \partial y^M(v,z)) = \frac{1}{\lambda^M}\frac{\partial^2 c^M}{\partial u \partial v}((u,x),(v,z))$$

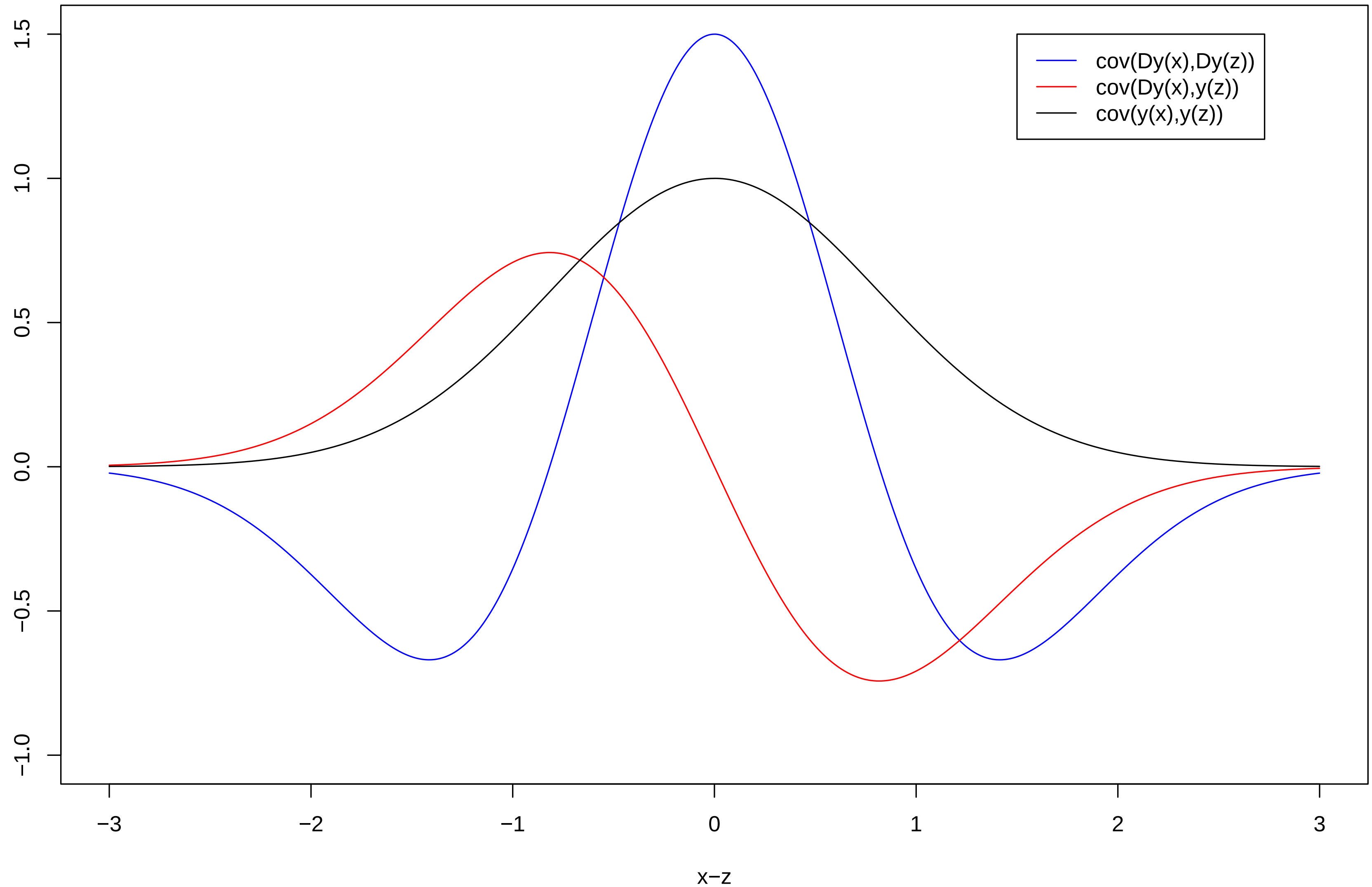
- which in the case at hand turn out to be

$$\mathsf{E}(\partial y^M(u,x)) = 0$$

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

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

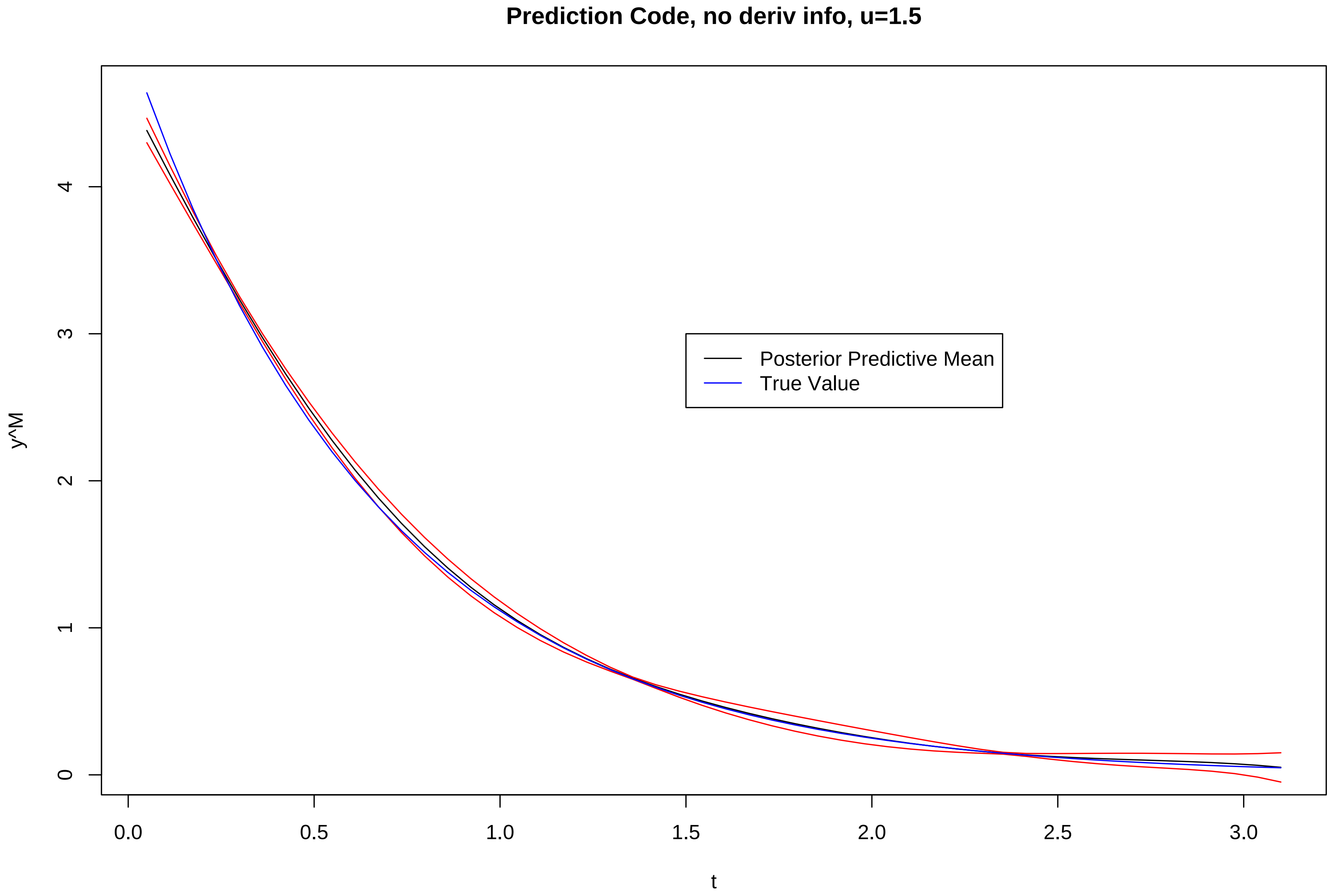
- We could fix α_2 at its mle in $(1,2]$, but GaSP does not allow for fixing only a subset of the roughness parameters, so all roughness parameters are fixed at 2.

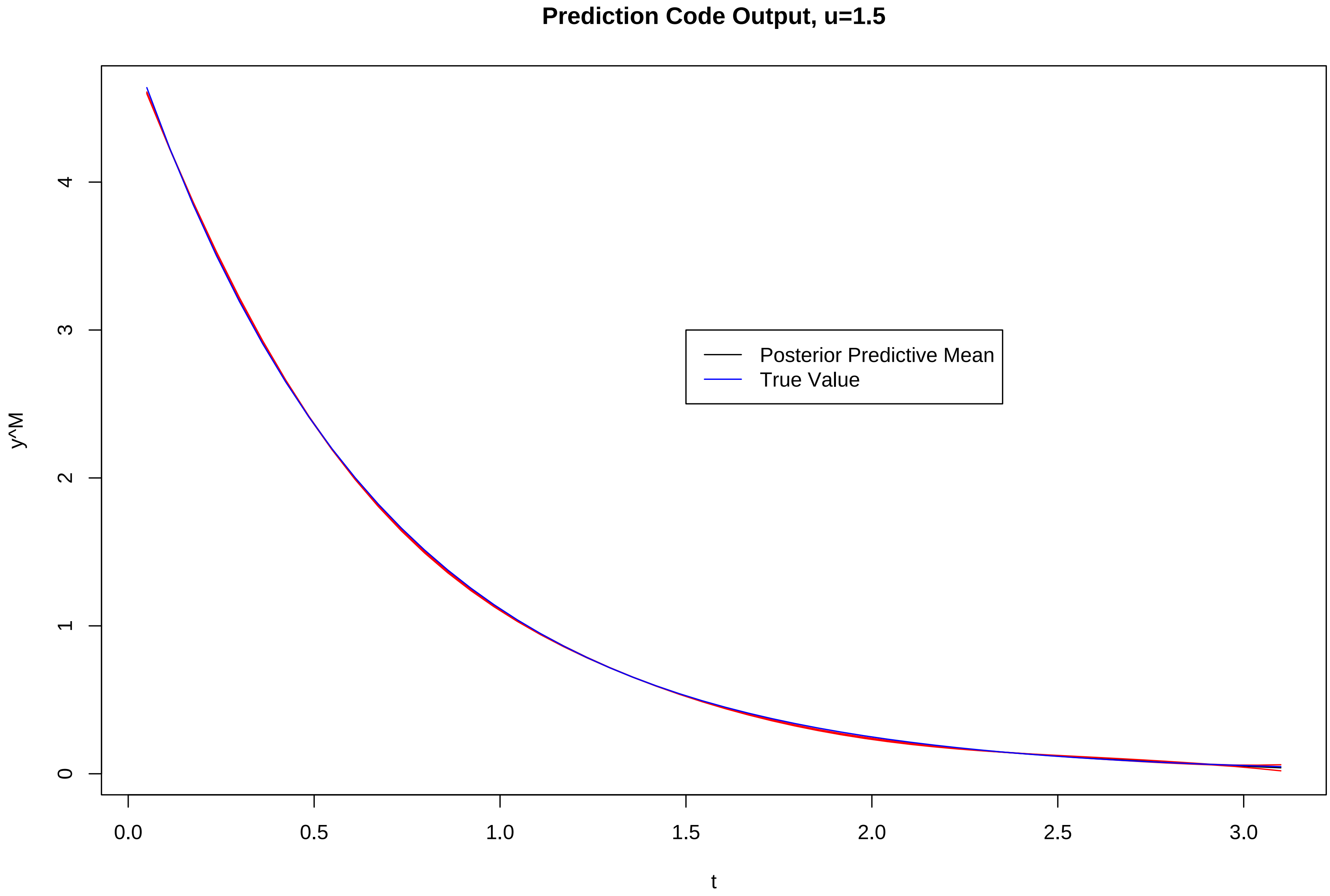


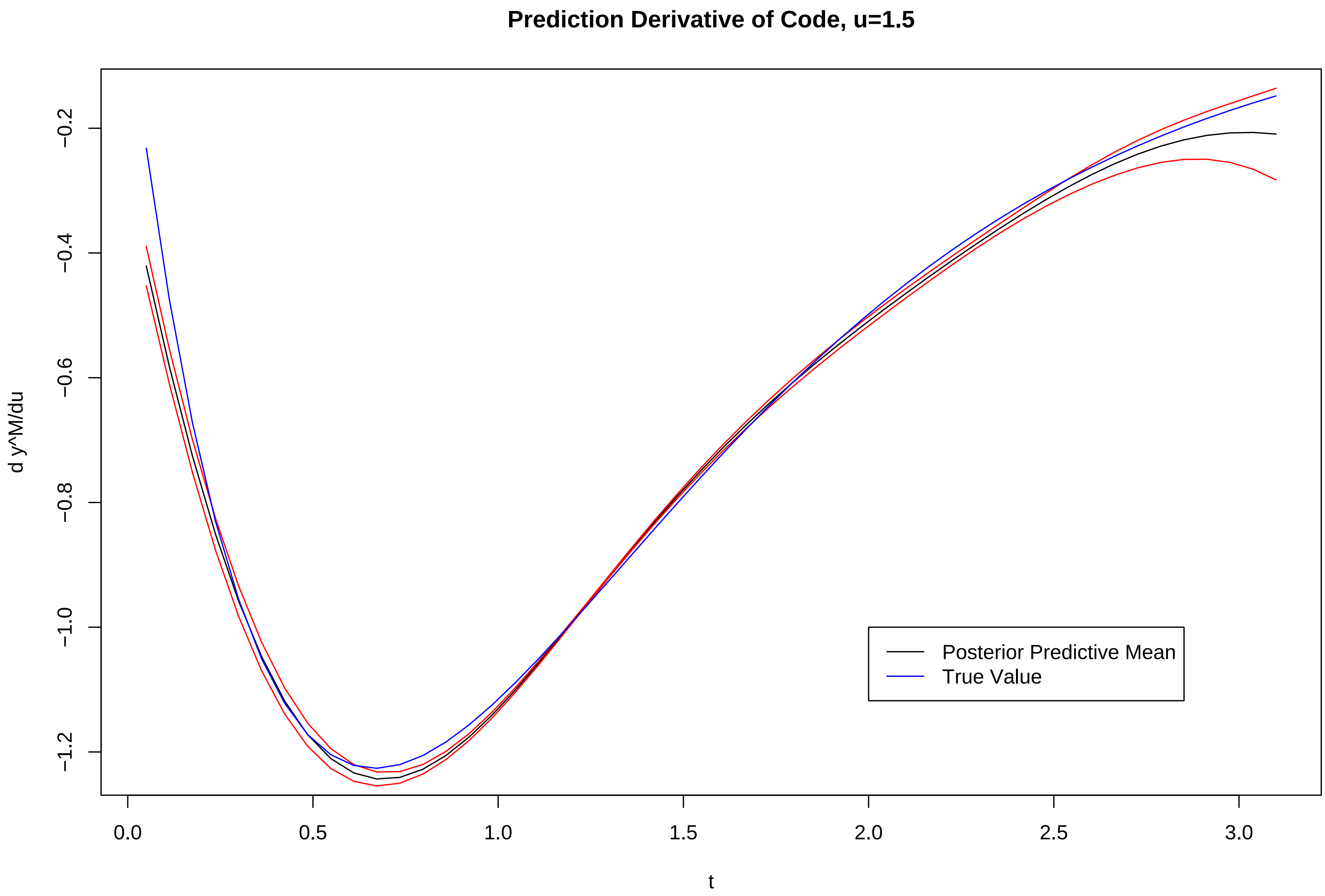
(Similar plot in Solak et al. 2003)

Toy Example

- $dy(t)/dt = -uy(t)$; $y(0) = y_0$
- 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
y^M
t
0
1
2
3
4
0.0
0.5
1.0
1.5
2.0
2.5
3.0


Prediction Code Output, u=1.5
Posterior Predictive Mean
True Value
y^M
t
0
1
2
3
4
0.0
0.5
1.0
1.5
2.0
2.5
3.0


Prediction Derivative of Code, u=1.5
d y^M/du
−0.2
−0.4
−0.6
−0.8
−1.0
−1.2
0.0
0.5
1.0
1.5
2.0
2.5
3.0
t
Posterior Predictive Mean
True Value

Toy Example – Field data

- Field data simulated from

$$y^F(t_i) = (y_0 - c)\exp(-u_\star t_i) + c + \varepsilon_i$$

with $y_0 = 5$, $c = 1.5$, $u_\star = 1.7$ and $\varepsilon \sim \mathrm{N}(0, \sigma^2)$, $\sigma = 0.3$. 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

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

(Augmented) Likelihood

$$
\begin{aligned}
f(\boldsymbol{y}^F, \boldsymbol{y}^M, \boldsymbol{\partial y}^M, \boldsymbol{y}^M_\star, \boldsymbol{\partial y}^M_\star, \boldsymbol{b} \mid \boldsymbol{\theta}^L, \boldsymbol{\theta}^M, \lambda^b, \beta^b, \lambda^F, u) = \\
\lambda^F \chi^2(\lambda^F s_F^2 \mid \textstyle\sum(n_i - 1)) \times \\
N(\bar{\boldsymbol{y}}^F \mid \boldsymbol{y}^M_\star + \boldsymbol{b} \circ \boldsymbol{\partial y}^M_\star, \operatorname{diag} \boldsymbol{n}^{-1}/\lambda^F) \times \\
f(\boldsymbol{y}^M_\star, \boldsymbol{\partial y}^M_\star \mid \boldsymbol{y}^M, \boldsymbol{\partial y}^M, \boldsymbol{\theta}^L, \boldsymbol{\theta}^M, u) \times \\
f(\boldsymbol{y}^M, \boldsymbol{\partial y}^M \mid \boldsymbol{\theta}^L, \boldsymbol{\theta}^M) \times \\
N(\boldsymbol{b} \mid 0, c^b(D^F)/\lambda^b)
\end{aligned}
$$

where $\circ$ stands for the Hadamard product of matrices, ie, entry-wise product.

MCMC (1)

- $\lambda^F \mid - \sim \Gamma(\lambda^F \mid a_1, a_2)$ where

$$a_1 = N_F/2 + \alpha_F$$
$$a_2 = r_F + s_F^2/2 + ||\bar{\boldsymbol{y}}^F - \boldsymbol{y}_\star^M - \boldsymbol{b} \circ \boldsymbol{\partial y}_\star^M||^2/2$$

and, *a priori*, $\lambda^F \sim \Gamma(\alpha_F, r_F)$

- $\lambda^b \mid - \sim \Gamma(\lambda^b \mid a_1, a_2)$ where

$$a_1 = N_F/2 + \alpha_b$$
$$a_2 = r_F + \boldsymbol{b}' \; [c^b(D^F)]^{-1} \; \boldsymbol{b}/2$$

and, *a priori*, $\lambda^b \sim \Gamma(\alpha_b, r_b)$

- $\boldsymbol{b} \mid - \sim N(\boldsymbol{m}, \boldsymbol{V})$ where

$$\boldsymbol{V}^{-1} = \lambda^F \operatorname{diag}(\boldsymbol{\partial y}_\star^M \circ \boldsymbol{\partial y}_\star^M \circ \boldsymbol{n}) + \lambda^b [c^b(D^F)]^{-1}$$
$$\boldsymbol{m} = \lambda^F \boldsymbol{V} \, [\boldsymbol{\partial y}_\star^M \circ \boldsymbol{n} \circ (\bar{\boldsymbol{y}}^F - \boldsymbol{y}_\star^M)]$$

- $\boldsymbol{y}_\star^M, \boldsymbol{\partial y}_\star^M \mid - \sim N(\boldsymbol{m}, \boldsymbol{V})$ where

$$\boldsymbol{V}^{-1} = \lambda^F \begin{pmatrix} \operatorname{diag} \boldsymbol{n} & \operatorname{diag}(\boldsymbol{n} \circ \boldsymbol{b}) \\ \operatorname{diag}(\boldsymbol{n} \circ \boldsymbol{b}) & \operatorname{diag}(\boldsymbol{n} \circ \boldsymbol{b} \circ \boldsymbol{b}) \end{pmatrix} + \boldsymbol{\Sigma}_{\star|\cdot}^{-1}$$
$$\boldsymbol{m} = \boldsymbol{V} \left[\lambda^F \begin{pmatrix} \boldsymbol{n} \circ \bar{\boldsymbol{y}}^F \\ \boldsymbol{n} \circ \boldsymbol{b} \circ \bar{\boldsymbol{y}}^F \end{pmatrix} + \boldsymbol{\Sigma}_{\star|\cdot}^{-1} \boldsymbol{\mu}_{\star|\cdot} \right]$$

with $\boldsymbol{\Sigma}_{\star|\cdot}$ and $\boldsymbol{\mu}_{\star|\cdot}$ representing, respectively, the conditional covariance and mean of $(\boldsymbol{y}_\star^M, \boldsymbol{\partial y}_\star^M)$ given $(\boldsymbol{y}^M, \boldsymbol{\partial y}^M)$.

- To draw from the full conditional of u, the Metropolis-Hastings ratio involves

$$\frac{f(\boldsymbol{y}_\star^M, \boldsymbol{\partial y}_\star^M \mid \boldsymbol{y}^M, \boldsymbol{\partial y}^M, v)}{f(\boldsymbol{y}_\star^M, \boldsymbol{\partial y}_\star^M \mid \boldsymbol{y}^M, \boldsymbol{\partial y}^M, u)}$$

- 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 $\bar{\boldsymbol{y}}^F$ explicitly.)

MCMC (2)

Integrate out $\boldsymbol{y}_\star^M, \boldsymbol{\partial y}_\star^M$ from the (augmented) likelihood to obtain

$$f(\boldsymbol{y}^F, \boldsymbol{y}^M, \boldsymbol{\partial y}^M, \boldsymbol{b} \mid \boldsymbol{\theta}^L, \boldsymbol{\theta}^M, \lambda^b, \beta^b, \lambda^F, u) = \lambda^F \chi^2(\lambda^F s_F^2 \mid \sum (n_i - 1)) \times$$
$$N(\bar{\boldsymbol{y}}^F \mid \boldsymbol{m}, \boldsymbol{V} + \operatorname{diag} \boldsymbol{n}^{-1}/\lambda^F) \times$$
$$N(\boldsymbol{b} \mid 0, c^b(D^F)/\lambda^b)$$

where, with $\boldsymbol{\Sigma}_{\star|\cdot}$ partitioned as $(\boldsymbol{\Sigma}_{ij})$, and $\boldsymbol{\mu}_{\star|\cdot} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$

$$\boldsymbol{m} = \boldsymbol{\mu}_1 + \boldsymbol{b} \circ \boldsymbol{\mu}_2$$
$$\boldsymbol{V} = \boldsymbol{\Sigma}_{11} + \operatorname{diag} \boldsymbol{b}\ \boldsymbol{\Sigma}_{21} + (\operatorname{diag} \boldsymbol{b}\ \boldsymbol{\Sigma}_{21})' + \operatorname{diag} \boldsymbol{b}\ \boldsymbol{\Sigma}_{22} \operatorname{diag} \boldsymbol{b}$$

Motivated by Jim's example, an interesting possibility which is feasible and may alleviate the problem is to iteratively sample from

1. $\lambda^b, \lambda^F, \boldsymbol{b} \mid \boldsymbol{y}_\star^M, \boldsymbol{\partial y}_\star^M, u, \boldsymbol{y}^F, \boldsymbol{y}^M, \boldsymbol{\partial y}^M$
2. $\boldsymbol{y}_\star^M, \boldsymbol{\partial y}_\star^M \mid \lambda^b, \lambda^F, \boldsymbol{b}, u, \boldsymbol{y}^F, \boldsymbol{y}^M, \boldsymbol{\partial y}^M$
3. $u \mid \lambda^b, \lambda^F, \boldsymbol{b}, \boldsymbol{y}^F, \boldsymbol{y}^M, \boldsymbol{\partial y}^M$

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

$$N(\bar{\boldsymbol{y}}^F \mid \boldsymbol{m}, \boldsymbol{V} + \operatorname{diag} \boldsymbol{n}^{-1}/\lambda^F)$$

as specified in the previous slide.

Remarks:

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

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

  with $\boldsymbol{m} = \boldsymbol{\mu}_{\star|\cdot} \circ (\mathbf{1} + \boldsymbol{b})$

- The full conditional of λ^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 $\boldsymbol{b}$ is not closed form either, and finding a good proposal distribution here should be a more delicate problem.

Remarks:

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

$$\begin{aligned}
& f(\boldsymbol{y}^F, \boldsymbol{y}^M, \boldsymbol{\partial y}^M, \boldsymbol{y}^M_\star, \boldsymbol{\partial y}^M_\star, \mid \boldsymbol{\theta}^L, \boldsymbol{\theta}^M, \lambda^b, \beta^b, \lambda^F, u) = \\
& \quad \lambda^F \chi^2(\lambda^F s_F^2 \mid \textstyle\sum (n_i - 1)) \times \\
& \quad N(\bar{\boldsymbol{y}}^F \mid \boldsymbol{y}^M_\star, \operatorname{diag} \boldsymbol{\partial y}^M \; c^b(D^F) \; \operatorname{diag} \boldsymbol{\partial y}^M / \lambda^b + \operatorname{diag} \boldsymbol{n}^{-1}/\lambda^F) \times \\
& \quad f(\boldsymbol{y}^M_\star, \boldsymbol{\partial y}^M_\star \mid \boldsymbol{y}^M, \boldsymbol{\partial y}^M, \boldsymbol{\theta}^L, \boldsymbol{\theta}^M, u) \times \\
& \quad f(\boldsymbol{y}^M, \boldsymbol{\partial y}^M \mid \boldsymbol{\theta}^L, \boldsymbol{\theta}^M)
\end{aligned}$$

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