Analysis of experiments to validate computer models with binary output

Capt Patrick Chapin

25-26 May 2010

Outline

- Problem Description
- Generalized Linear Models Approach
- Gaussian Stochastic Process Approach
- Examples
- Conclusions
- Future Research
- Questions?
- References

Problem Description

- Origin: Validation study conducted by government agency to assess the suitability of a new computer simulation
- Objective: Verification, Validation and Accreditation of a computer simulation model, specifically the validation portion
- The specifics of the origin and actual computer model are intentionally left vague. The bottom line is we have a computer model that, for each run, will output either "success" or "failure", with the ultimate goal being to check this computer model versus reality.
- We also assume that actual physical testing is expensive, and thus the number of actual, physical tests with which to compare the computer model will be very limited.

Previous Methodology

- Previous researchers conducted tests at 14 design points, with 1 physical trial and 1000 computer trials at each design point. Each design point is a specific set of physical conditions

- Each design point was graded according to how well the results of the physical trial matched the results of the computer experiment, summarizing the results in a table
  - Green Square: Physical trial results occurred in more than 50% of the computer trials.
  - Green Diamond: Physical trial results occurred in between 5% and 50% of the computer trials.
  - Yellow Circle: Physical trial results occurred in less than 5% of computer trials but occurred at least once.
  - Red X: Physical Trial results occurred in none of the computer trials.

Stoplight Chart

Design Point	Variable 1	Variable 2	Variable 3	Variable 4	Variable 5
1	■	■	■	■	■
2	■	■	■	■	■
3	■	■	■	■	■
4	■	■	◆	■	■
5	■	■	■	■	■
6	■	■	■	■	◆
7	■	◆	■	■	■
8	■	■	■	■	●
9	■	■	■	■	■
10	■	◆	■	■	■
11	■	■	■	■	■
12	■	■	■	■	■
13	■	●	■	N/A	N/A
14	■	■	■	N/A	N/A

Can we do better?

- Previous methodology appears very ad-hoc

  1. No clear decision rule on when to reject the adequacy of the computer model
  2. Unknown type-I error probability
  3. Questionable and likely low power to detect true departures
  4. What about failure probabilities at locations not tested?
  5. What about predicting the difference between the computer failure probability and physical failure probability at locations not tested

- Objective: New approach that can provide a partial remedy to points 1-3 and directly address points 4 and 5.

GLM Approach

- One possible approach is to consider the problem as a Generalized Linear Model
- Methodology provides a way to join a random component to a systematic component via a link function
- Key assumptions of this methodology:
  - The form of the systematic component in the link-space can be specified
  - There is enough data to fit the unknown parameters in the systematic component
- Estimation of parameters easy via standard functions in most statistical software

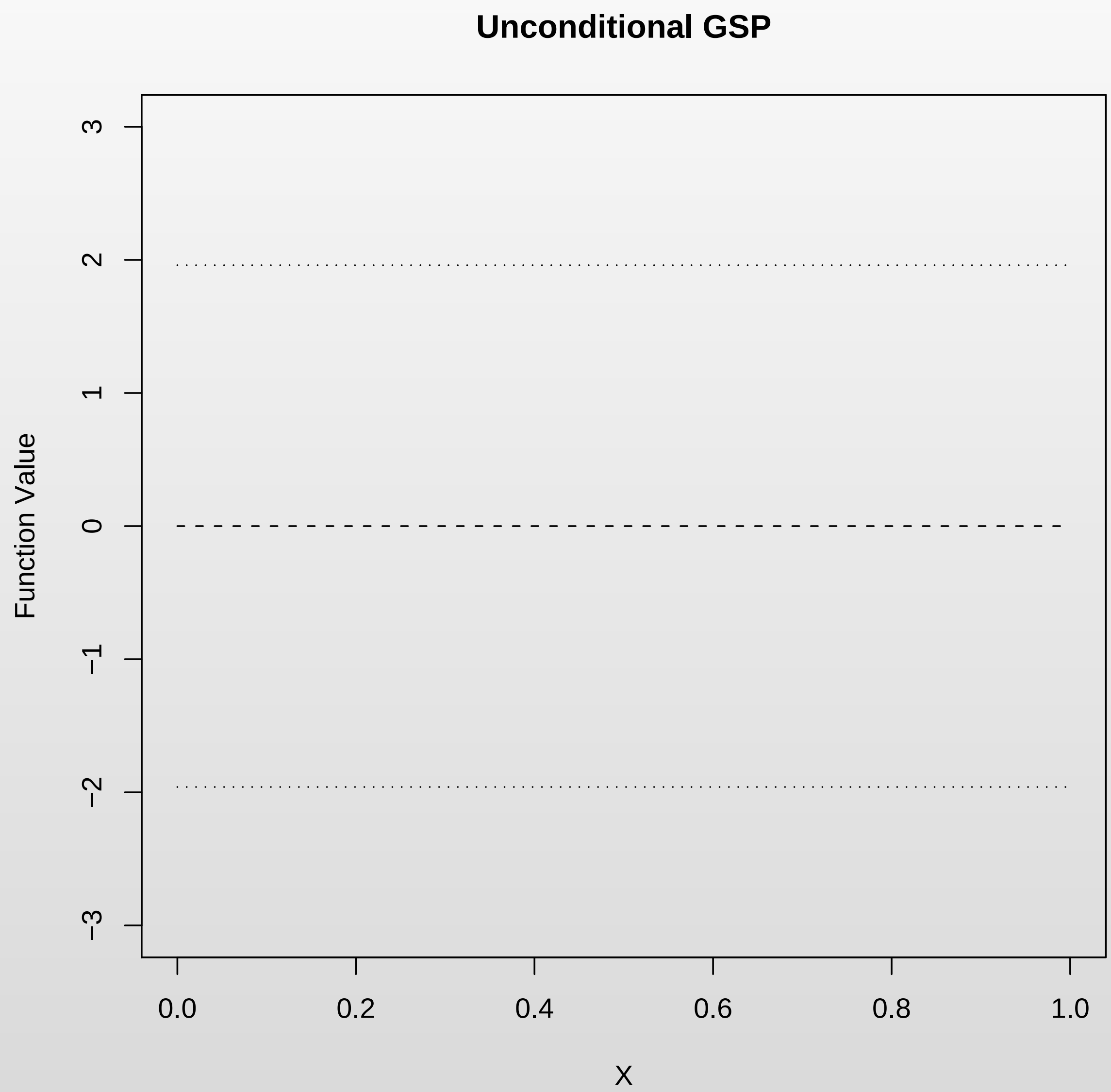
GSP Approach

- The GLM approach depends upon the fact that the functional form of the systematic component can be specified *a priori*.
- This may or may not be a good assumption, especially if we have very few data points and/or failure probability is a complicated function of X.
- One way to relax this assumption is to consider the underlying systematic component function unknown. Sacks *et al.* (23) treat the prediction of unknown functions at unobserved points as a kriging problem while Currin *et al.* (3) use the Gaussian Stochastic Process (GSP) as a prior distribution for the unknown function.
- The benefit of this assumption is that the GSP can be made to fit many different functional forms *that need not be specified in advance*

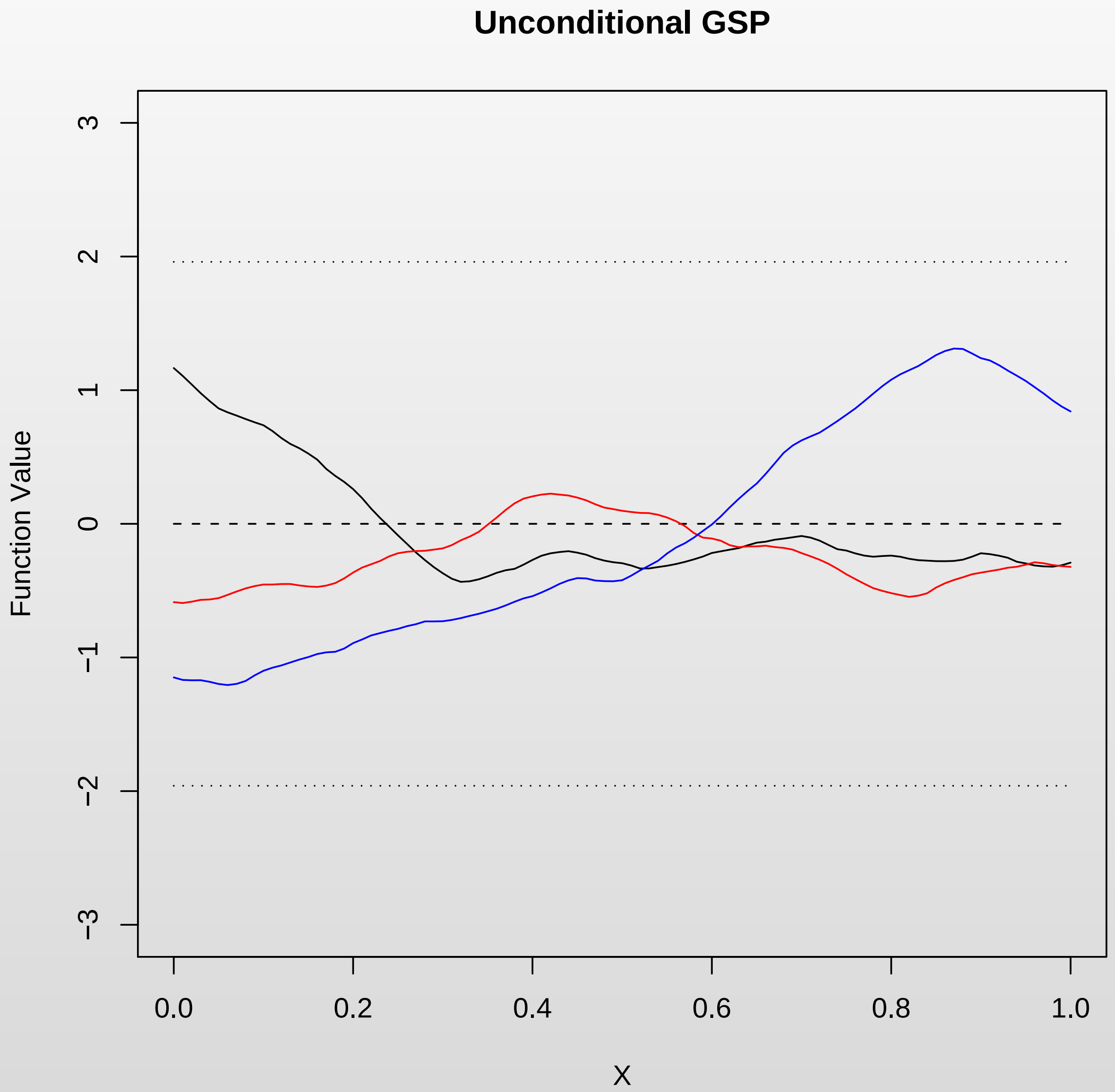
What is a GSP?

- A GSP is a random process where any finite vector has a joint multivariate Gaussian (Normal) distribution. For our purposes, the GSP exists over the k-dimensional covariate space.
- We constrain the GSP to be stationary:
  - Constant mean, $E[T(x_i)] = \mu, \forall x_i$
  - Constant variance, $V[T(x_i)] = \sigma^2, \forall x_i$
  - Covariance dependent only on difference between points, i.e. $\mathrm{Cov}[T(x_i), T(x_j)] = \sigma^2 R(x_i - x_j, \boldsymbol{\theta}), \forall x_i, x_j$
- The functional form of the correlation function, R, must be specified up to parameter $\boldsymbol{\theta}$. The choice of R dictates continuity and smoothness properties of the possible sample paths of the GSP.

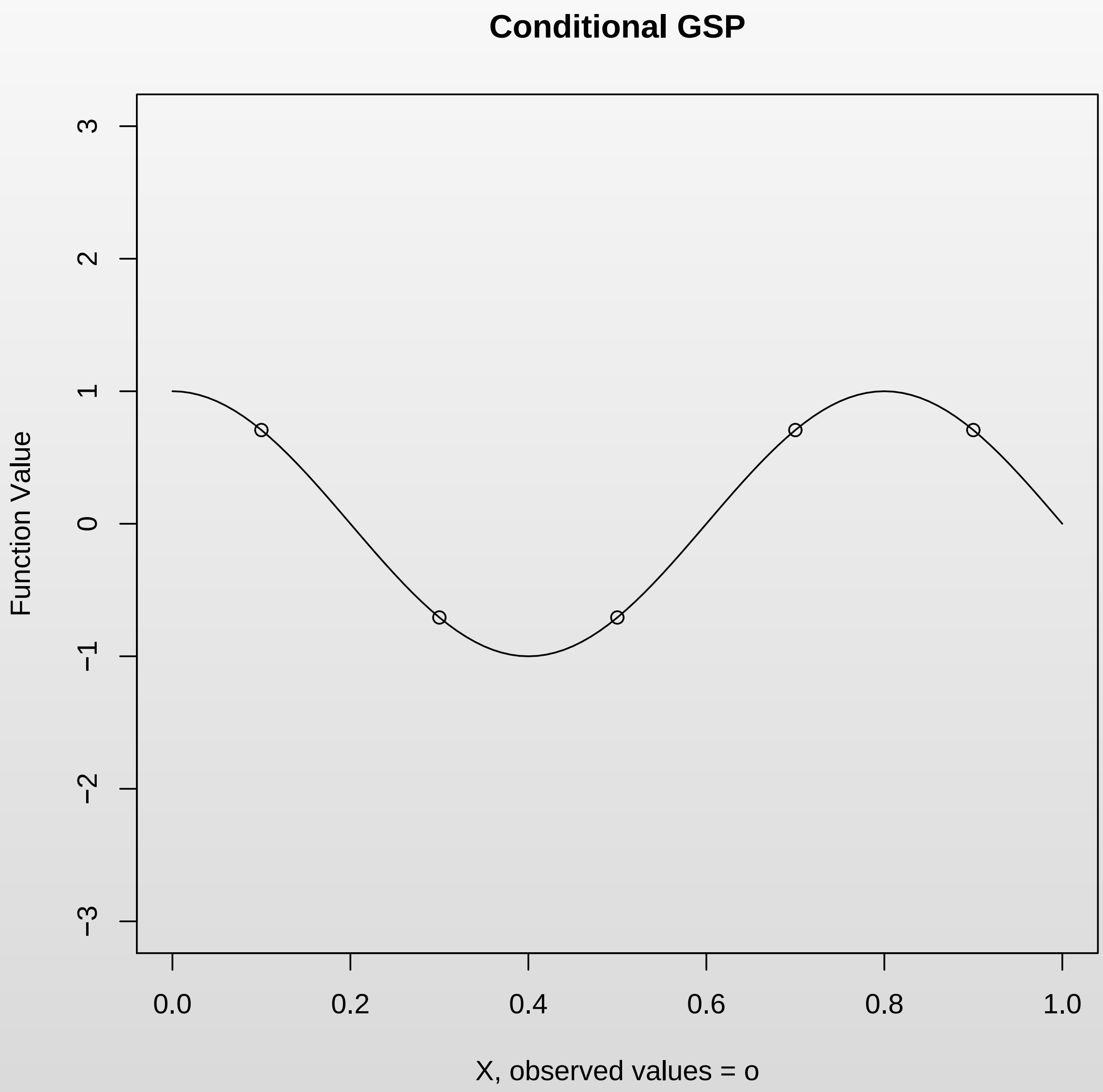
How a GSP Works 1



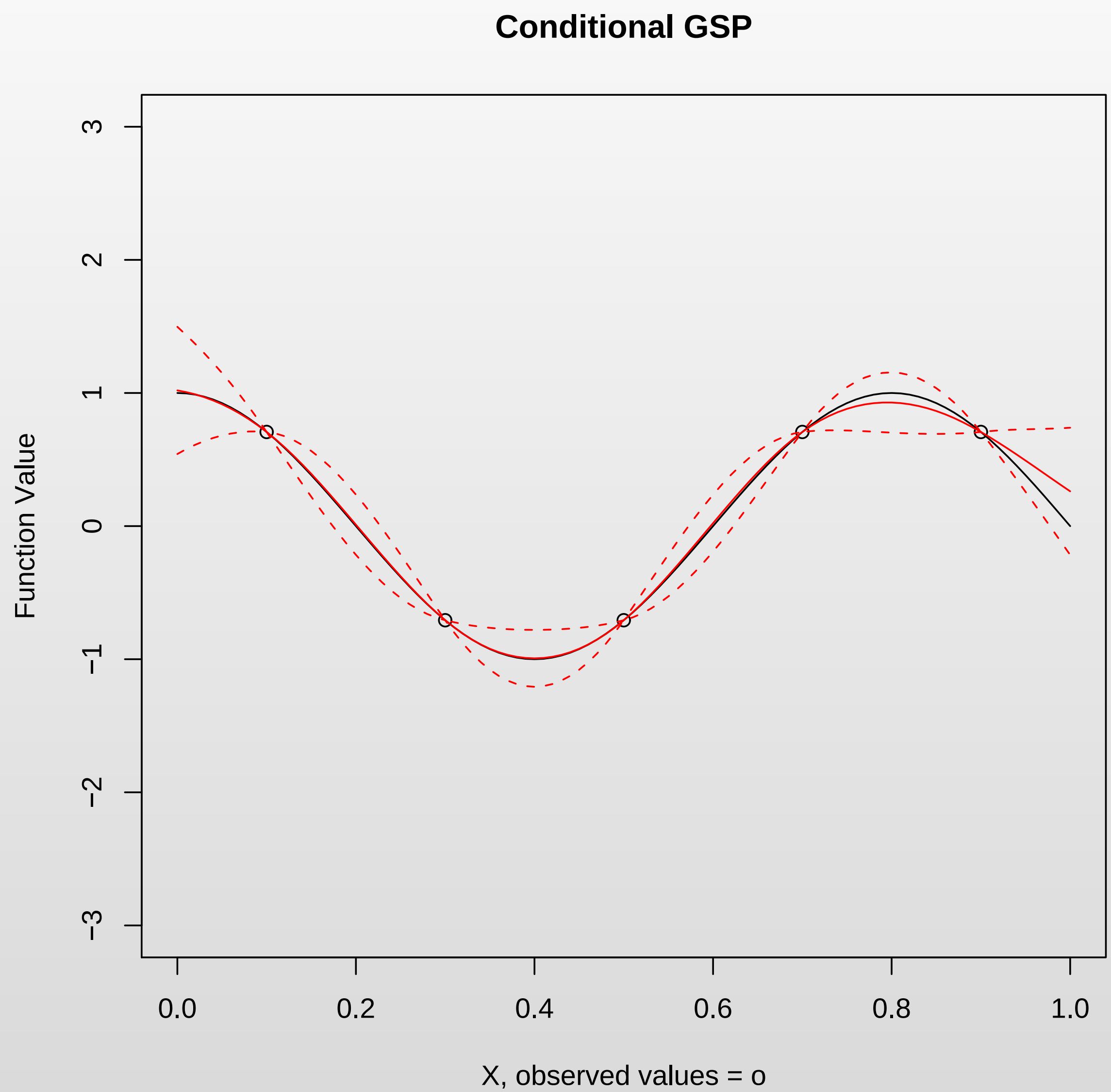
How a GSP Works 2



How a GSP Works 3



How a GSP Works 4



Setting up the GSP

- Here, we are primarily following the work of Currin *et al.* (3) and Kennedy and O'Hagan (13)
- In this case, the GSP is treated as a prior distribution on the unknown systematic function in the "link space"
- So, for each covariate location x_i and x_j^* we observe Y_i and Y_j^* failures, which we assume to be distributed as binomial variables with parameters n_i, p_i, and n_j^*, p_j^* respectively.
- The GSPs serve as prior distributions on the values of $p_1, p_2, \ldots, p_l$ and $p_1^*, p_2^*, \ldots, p_{l^*}^*$.

Combining Computer and Physical Trials

- First, assume that the unknown systematic function for the computer trials can be approximated by a GSP, say T^*.
- Next, assume that the unknown systematic function for the physical trials is approximated by a GSP, say T.
- In the manner suggested by Kennedy & O'Hagan (13), combine the two by assuming that $T = T^* + \delta$. That is, for any given value x in the covariate space, $T(x) = T^*(x) + \delta(x)$ where T^* and δ are independent processes
- δ represents the difference between unknown computer function and the true unknown physical function, and the primary goal here is to estimate $g^{-1}(T) - g^{-1}(T^*)$ (the probability scale difference function).

Definitions and Implications 1

- Let X denote the collection of physical design points, and X^* denote the collection of computer design points
- $\begin{pmatrix} T^*(X^*) \\ T^*(X) \end{pmatrix} \sim MVN\left\{ \begin{pmatrix} \mu_1 J_{l^*} \\ \mu_1 J_l \end{pmatrix}, \begin{pmatrix} \Sigma_1^* & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_1 \end{pmatrix} \right\}$
  - $\{\Sigma_1^*\}_{i,j} = \sigma_1^2 R_C(x_i^* - x_j^*, \boldsymbol{\theta}_1)$
  - $\{\Sigma_1\}_{i,j} = \sigma_1^2 R_C(x_i - x_j, \boldsymbol{\theta}_1)$
  - $\{\Sigma_{12}\}_{i,j} = \sigma_1^2 R_C(x_i - x_j^*, \boldsymbol{\theta}_1)$
- $\delta(X) \sim MVN\left(\mu_2 J_l, \Sigma_2\right)$
  - $\{\Sigma_2\}_{i,j} = \sigma_2^2 R_\delta(x_i - x_j, \boldsymbol{\theta}_2)$
- Then $T(X) = T^*(X) + \delta(X)$ is MVN with mean vector $(\mu_1 + \mu_2) J_{l \times 1}$ and covariance matrix $\Sigma_1 + \Sigma_2$.

Definitions and Implications 2

- $\text{Cov}(T(x_i), T^*(x_j^*)) = \text{Cov}(T^*(x_i) + \delta(x_i), T^*(x_j^*)) = \text{Cov}(T^*(x_i), T^*(x_j^*)) = \Sigma_{12}$
- Then $[T(X), T^*(X^*)]^T$ ~MVN, with $E\{[T(X), T^*(X^*)]^T\} = [(\mu_1 + \mu_2)J_{l\times 1}, \mu_1 J_{l^*\times 1}]^T$ and $V[T(X), T^*(X^*)]^T = \begin{pmatrix} \Sigma_1 + \Sigma_2 & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_1^* \end{pmatrix}$, given $(\mu_1, \sigma_1^2, \boldsymbol{\theta}_1, \mu_2, \sigma_2^2, \boldsymbol{\theta}_2)$
- Let $t_i = \text{logit}(T(x_i))$ and $t_j^* = \text{logit}(T^*(x_j^*))$, then unscaled posterior can then be written as

$$\exp\left\{\sum_{i=1}^{l} (n_i[Y_i t_i - \log(1 + e^{t_i})])\right\}$$
$$\times \exp\left\{\sum_{i=1}^{l^*} (n_i^*[Y_i^* t_i^* - \log(1 + e^{t_i^*})])\right\}$$
$$\times \phi_{l+l^*}\left(\begin{array}{c} T(X) \\ T^*(X^*) \end{array} \middle| \mu_1, \sigma_1^2, \boldsymbol{\theta}_1, \mu_2, \sigma_2^2, \boldsymbol{\theta}_2\right)$$

Solution Method

- With the unscaled posterior, sample from this distribution using Markov Chain Monte Carlo (MCMC) simulation on the unknown parameters $T(X), T^*(X^*)$. Still need to deal with GSP parameters $\mu_1, \sigma_1^2, \boldsymbol{\theta}_1, \mu_2, \sigma_2^2, \boldsymbol{\theta}_2$.
- Direct estimation is difficult, number of unknowns $\geq l + l^* + 6$ which is more than the number of data points.
- Use a hierarchial model, assign priors to GSP parameters.
- The prior distributions of the hyper parameters of the GSP are:
  - $\mu_1, \mu_2 \sim$Normal$(0, 25)$
  - $\sigma_1^2, \sigma_2^2 \sim \Gamma^{-1}(1, 1)$
  - $\theta \sim \mathrm{uniform}(0, k+2)$ for all theta parameters

Prediction Methods 1

- Still need a method to predict T, T^* and/or δ at locations not in the design
- Use the BLUP developed in Sacks *et al.* (23) and Santner *et al.* (24), which borrows from kriging
- let x_0 be a point at which we want to predict the physical, computer and error processes, let $T'(x_0) = [T(x_0), T^*(x_0)]^T$, and $\Sigma_0 = Cov(T'(x_0), [T(X), T^*(X^*)]^T)$.
- Point Estimates: for each value from the posterior sample of $T(X)$, $T^*(X^*)$ and if applicable $\mu_1, \sigma_1^2, \theta_1, \mu_2, \sigma_2^2, \theta_2$:
- $E\left(\widehat{T'(x_0)}\right) =$ $\begin{pmatrix} \mu_1 + \mu_2 \\ \mu_1 \end{pmatrix} + \Sigma_0 \left(\Sigma\right)^{-1} \begin{pmatrix} T(X) - (\mu_1 + \mu_2) J_l \\ T^*(X^*) - \mu_1 J_{l^*} \end{pmatrix}$

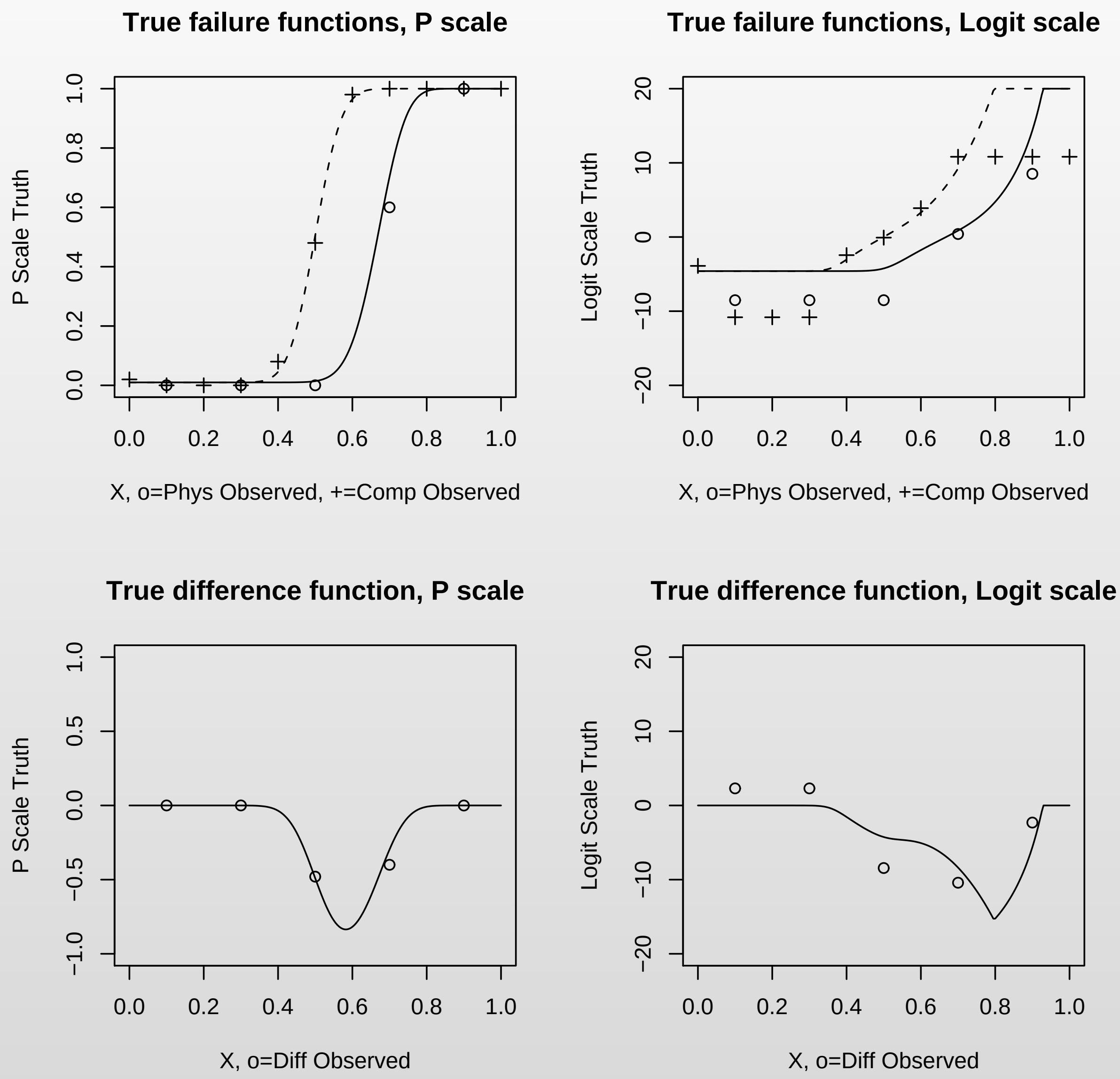
Prediction Methods 2

- $$V\left(\widehat{T'(x_0)}\right) = \begin{pmatrix} \sigma_1^2+\sigma_2^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 \end{pmatrix} - \begin{pmatrix} f_0 & \Sigma_0 \end{pmatrix} \begin{pmatrix} \mathbf{0}_{2\times 2} & F_{0,P}^T & F_{0,C}^T \\ F_{0,P} & \Sigma_1+\Sigma_2 & \Sigma_{12}^T \\ F_{0,C} & \Sigma_{12} & \Sigma_1^* \end{pmatrix}^{-1} \begin{pmatrix} f_0 & \Sigma_0 \end{pmatrix}^T$$
- where $f_0 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$, $F_{0,P} = (1,1) \otimes J_l$ and $F_{0,C} = (1,0) \otimes J_{l^*}$
- Using this information, we can construct a predictive sample of $T(x_0)$ and $T^*(x_0)$ from our posterior sample, and use this sample to calculate predictors and credible intervals.
- Also note that we have all we need to construct inferences about the error process as well.

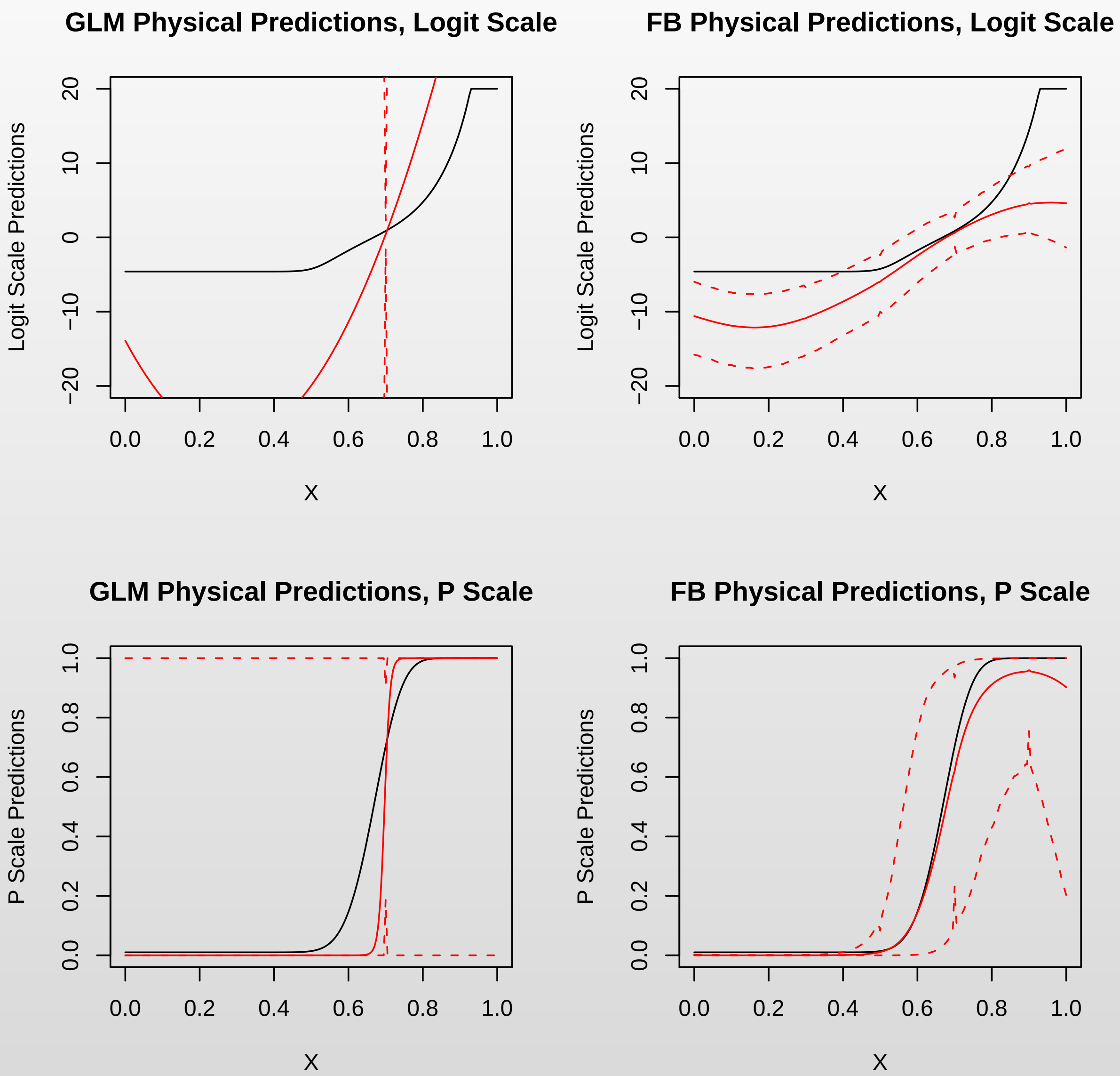
Examples: Preliminaries

- GLM: 1D Models have full quadratic model. Higher dimension are linear with all two way interactions. GLMs fitted using *glm* function in R.
- Full Bayesian method also implemented. 2D models have 2-dimensional correlation parameters
- The GSP is analyzed via MCMC simulation using a Metropolis-Hastings within Gibbs Algorithm using a the Adaptive Metropolis Algorithm of Haario *et al.* (8).
- Experimental design for higher dimensions chosen as Maximin Latin Hypercubes as described by Morris and Mitchell (16).

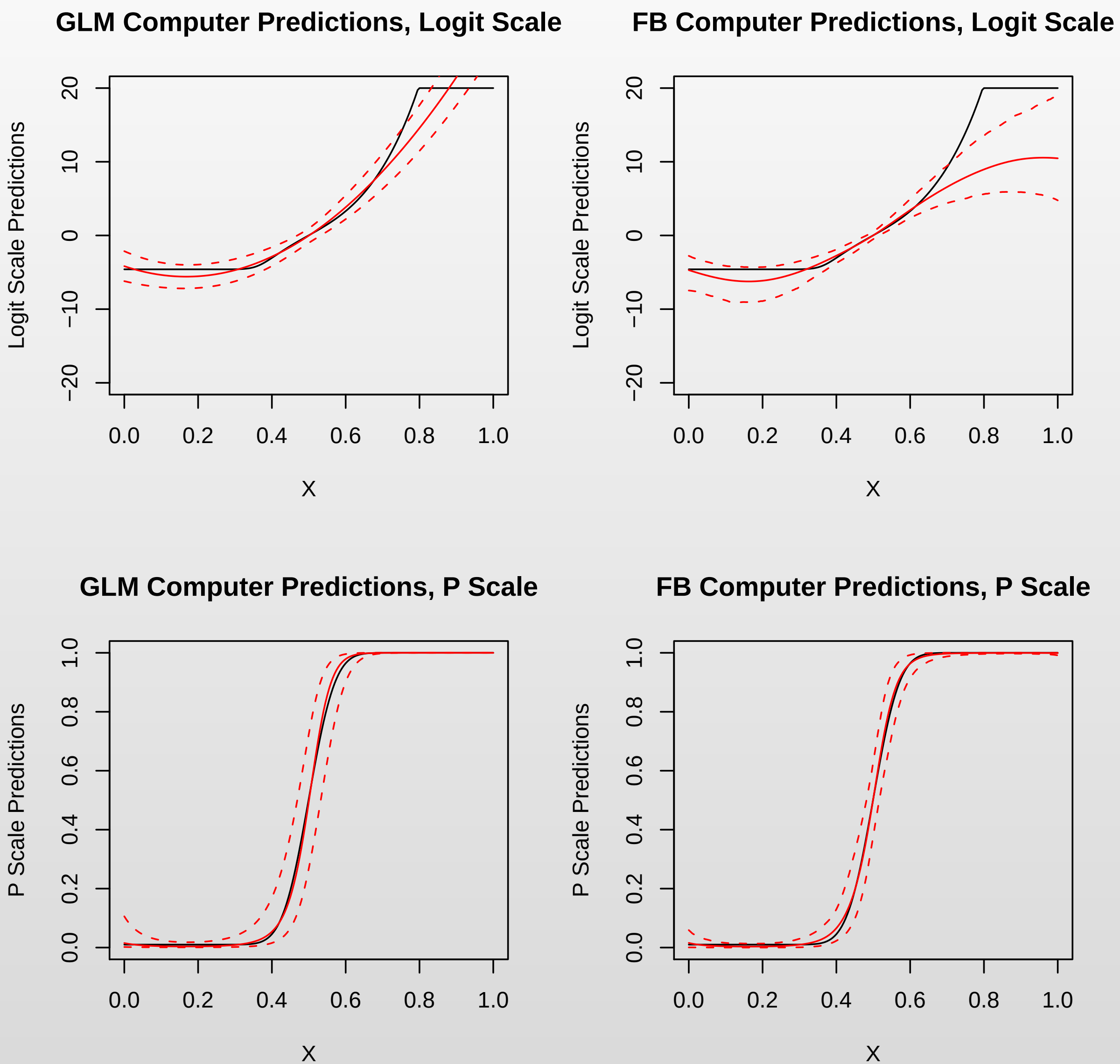
Example 1


True failure functions, P scale
P Scale Truth
0.0
0.2
0.4
0.6
0.8
1.0
X, o=Phys Observed, +=Comp Observed
True failure functions, Logit scale
Logit Scale Truth
−20
−10
0
10
20
X, o=Phys Observed, +=Comp Observed
True difference function, P scale
P Scale Truth
−1.0
−0.5
0.0
0.5
1.0
X, o=Diff Observed
True difference function, Logit scale
Logit Scale Truth
X, o=Diff Observed

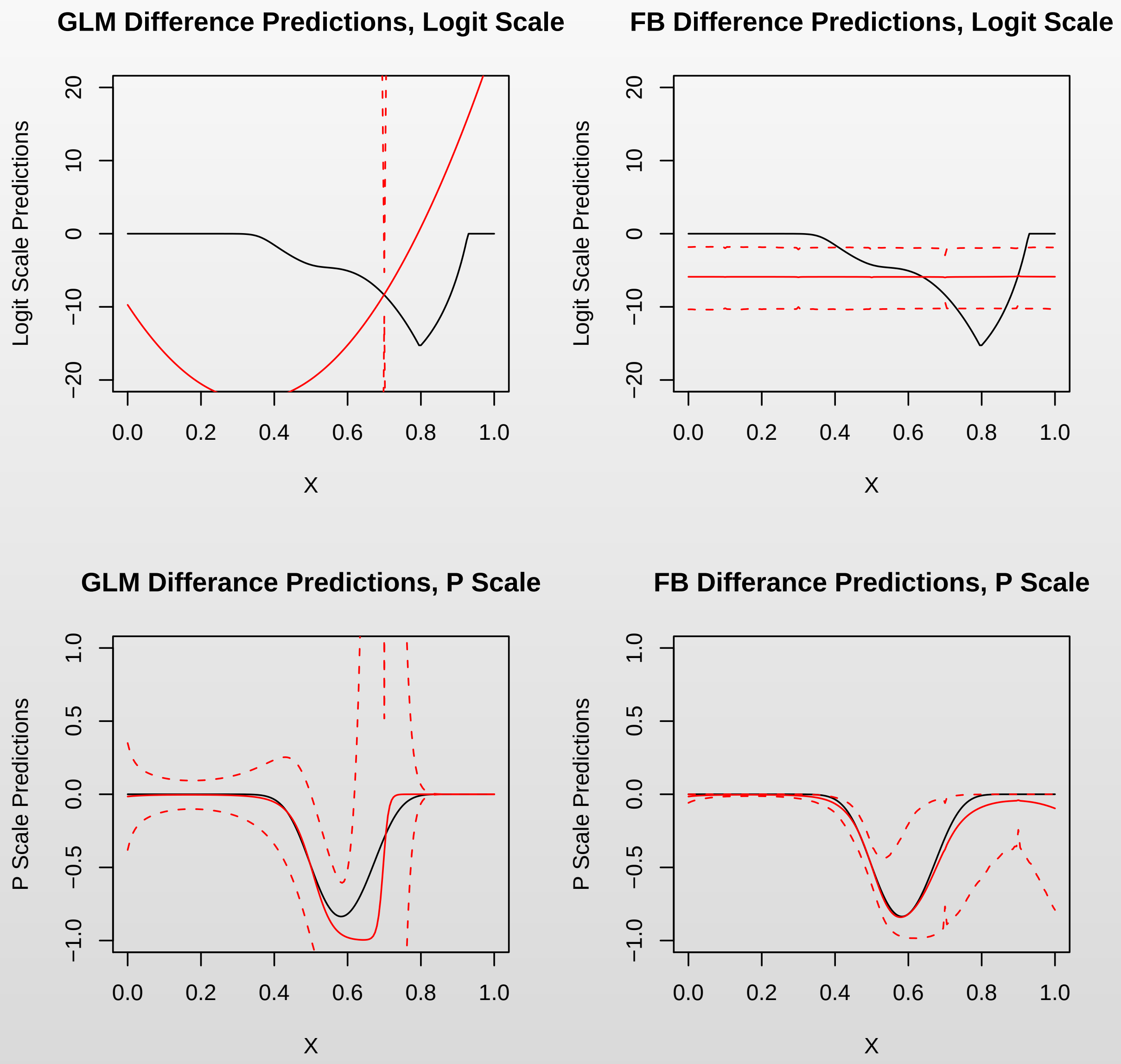
Example 1, Physical Predictions



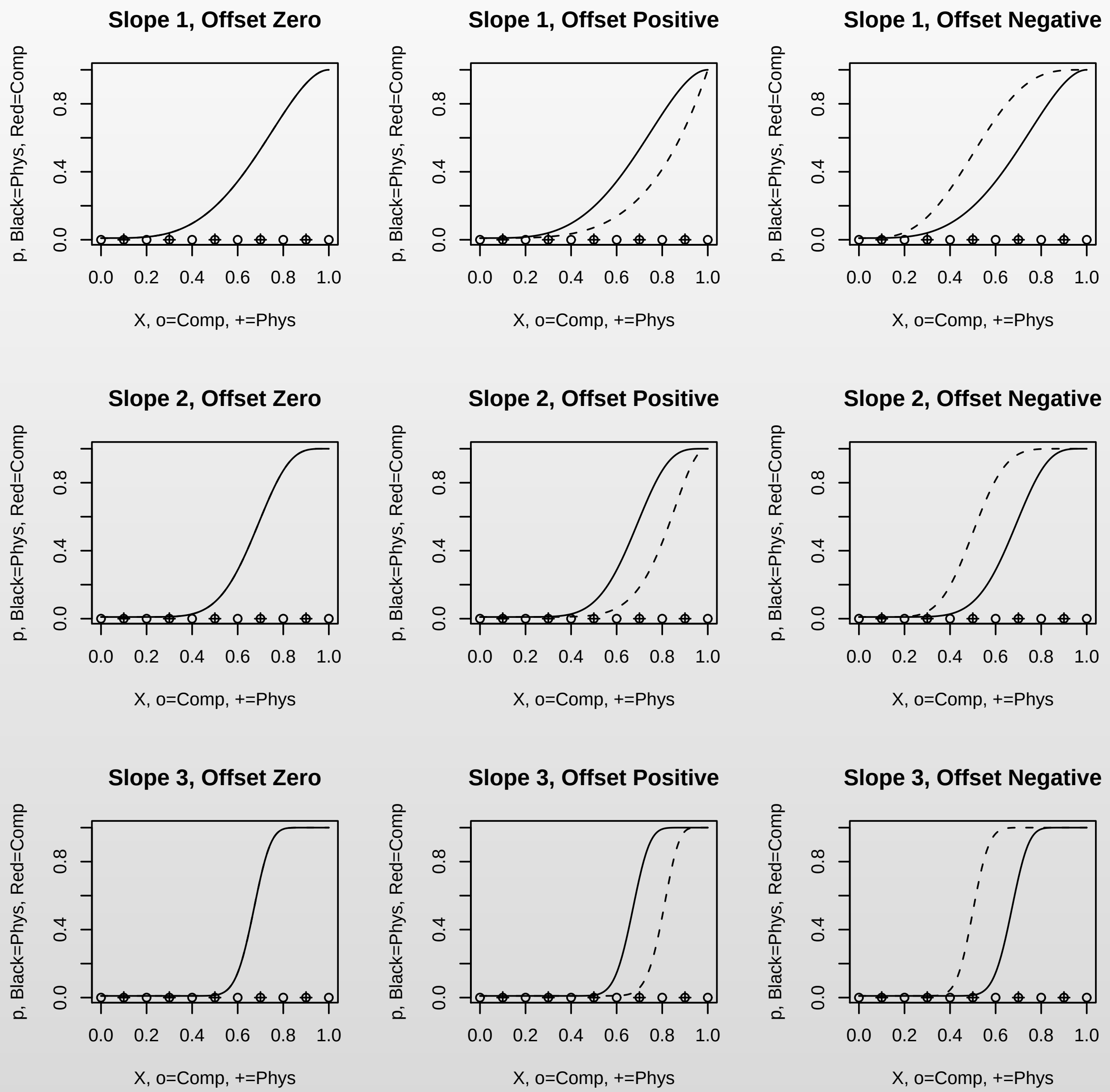
Example 1, Computer Predictions



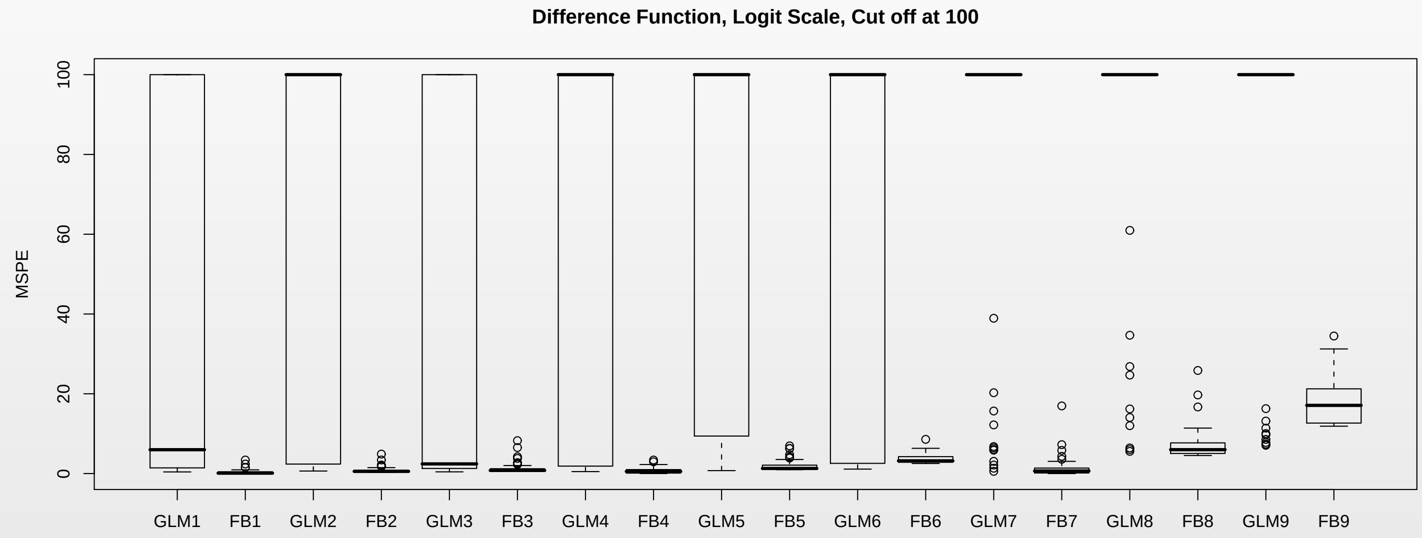
Example 1, Difference Predictions

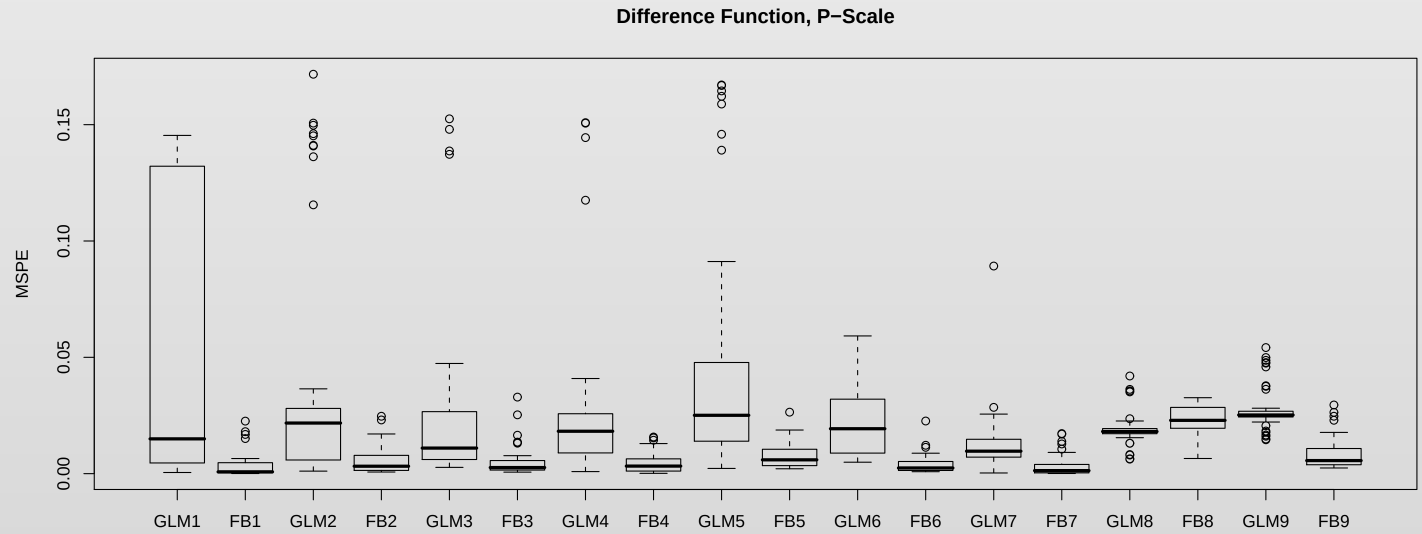


Example 2

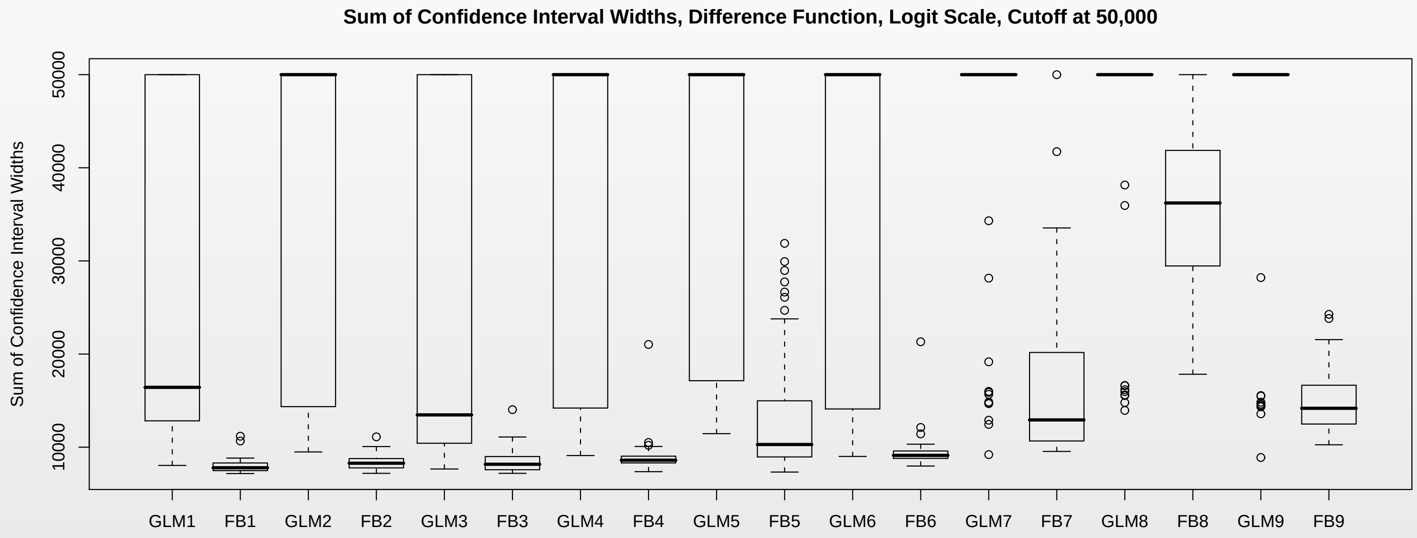

Slope 1, Offset Zero
Slope 1, Offset Positive
Slope 1, Offset Negative
Slope 2, Offset Zero
Slope 2, Offset Positive
Slope 2, Offset Negative
Slope 3, Offset Zero
Slope 3, Offset Positive
Slope 3, Offset Negative
p, Black=Phys, Red=Comp
X, o=Comp, +=Phys
0.0
0.4
0.8
0.0 0.2 0.4 0.6 0.8 1.0

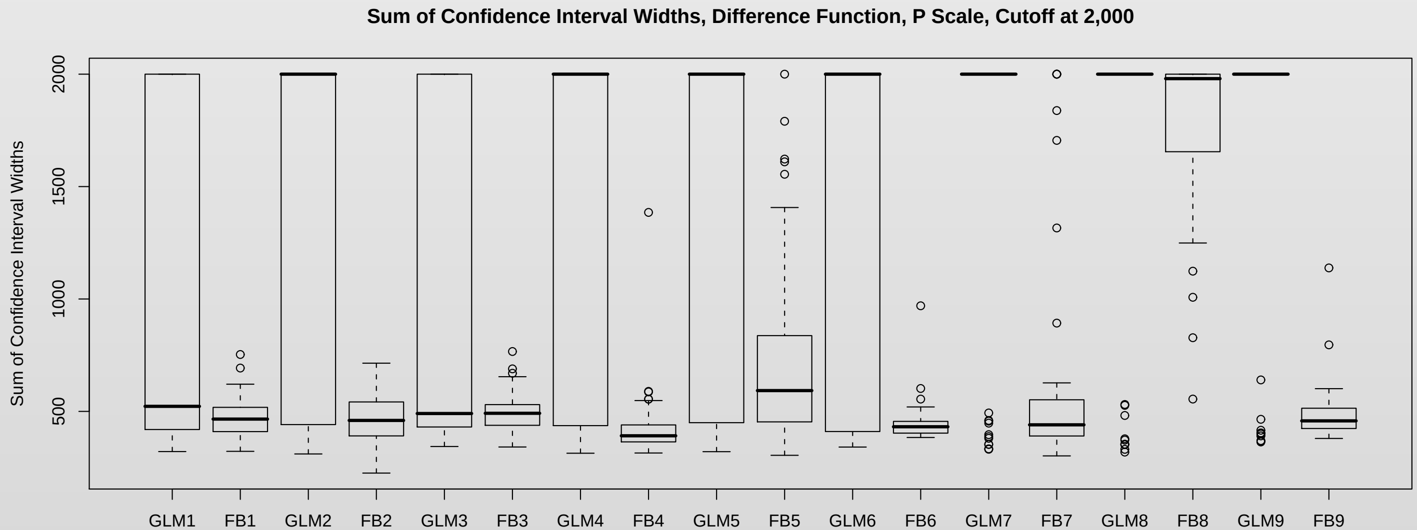
Example 2, Difference MSPE



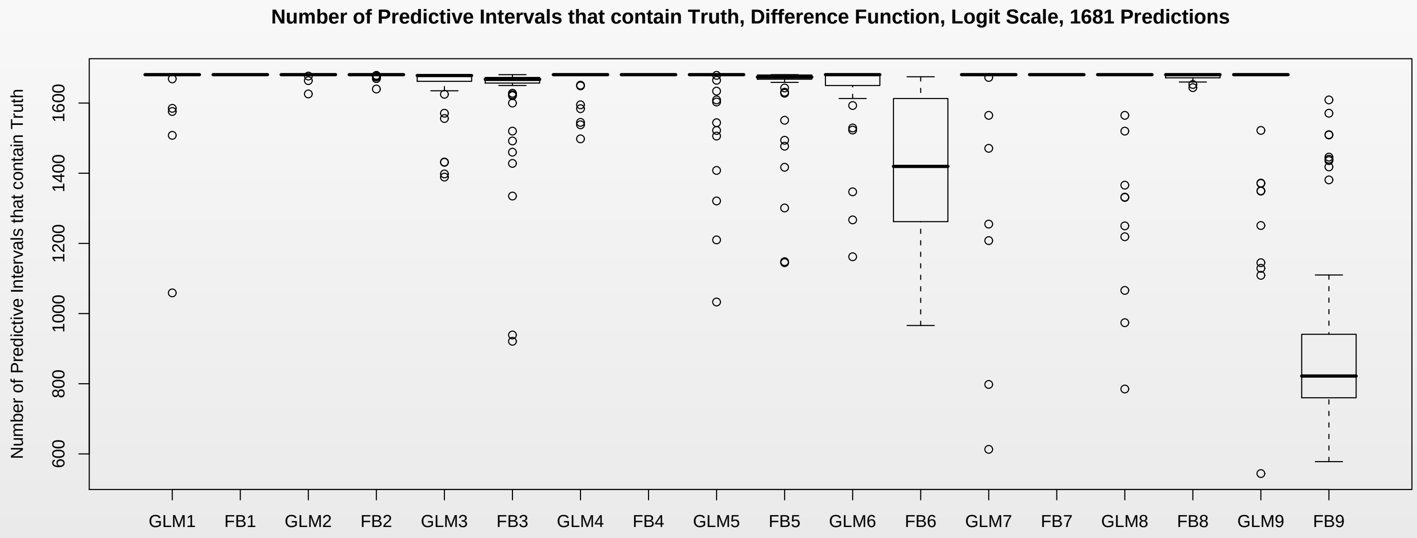


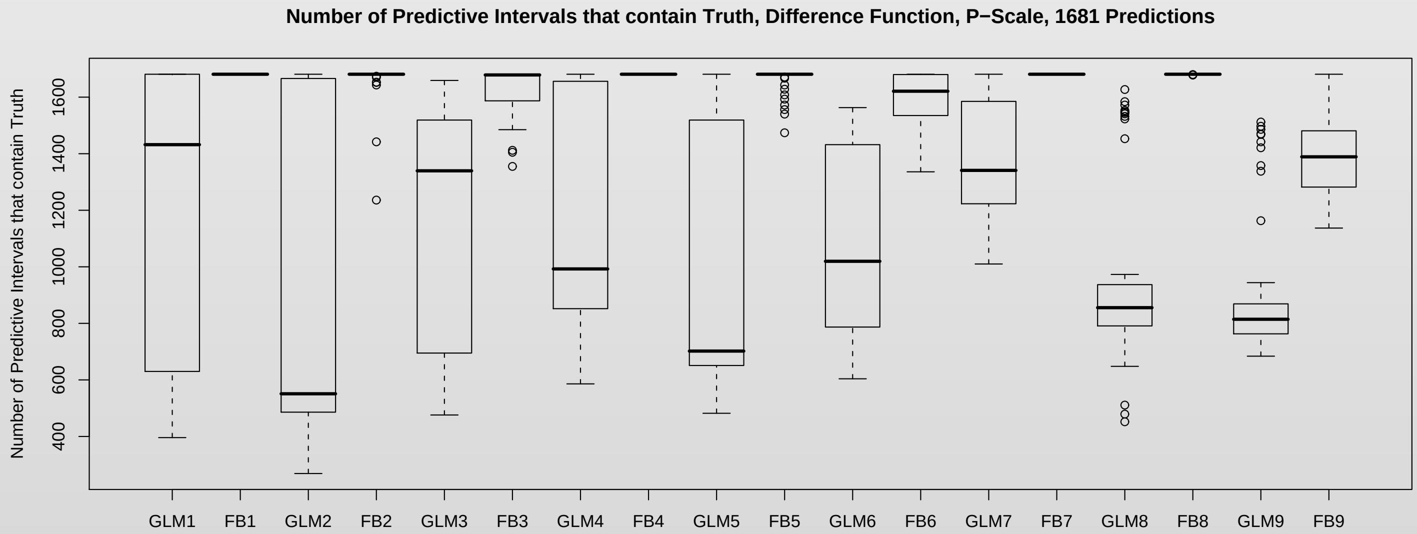
Example 2, Difference sum of confidence / credible interval widths



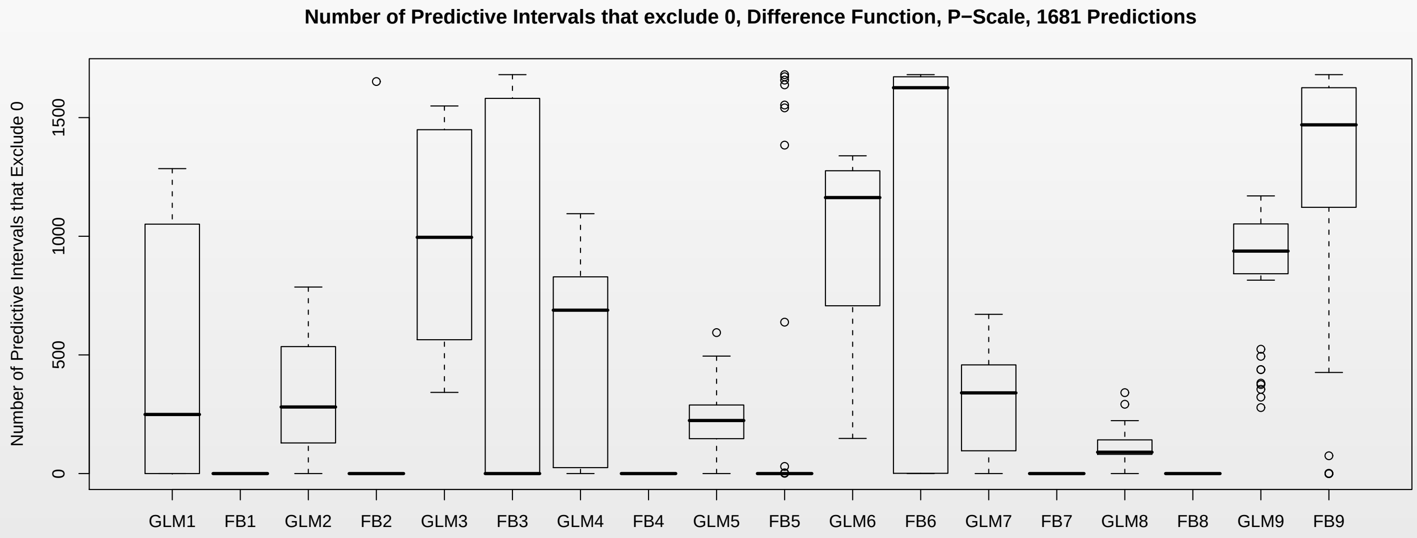


Example 2, Number of difference intervals that contain the true function value





Example 2, Number of probability-scale difference intervals that exclude zero



Truth	Max Diff	% Sig Diff	Truth	Max Diff	% Sig Diff
1	0.0000	0.0000	6	0.6507	0.3510
2	0.4268	0.2421	7	0.0000	0.0000
3	0.4820	0.4295	8	0.8411	0.1452
4	0.0000	0.0000	9	0.9168	0.2415
5	0.5599	0.2064			

Conclusions

- Generally, the Full Bayesian method performs better than the GLM, in terms of MSPE and coverage
- The Full Bayesian method is also much more conservative, i.e. produces p-scale difference intervals that exclude zero much less often than the GLM method
- Full Bayesian method is also much more computationally complex, at least 2 orders of magnitude.
- In cases where lots of data is available, the GLM may be a plausible alternative

Future Research

- Design of Experiments: How to choose covariate locations to test.
- Incorporate messy, i.e. real world, data.
- GLM-GSP Hybrid, computer simulation function modeled as a GLM, physical process = computer simulation + error process.
- Account for the multivariate nature of the motivating example.

Questions?

Questions?

References 1

[1] CRAINICEANU, C. M., DIGGLE, P. J., AND ROWLINGSON, B. Bivariate binomial spatial modeling of loa loa prevalence in tropical africa. *Journal of the American Statistical Association 103* (March 2008), 21–37(17)

[2] CRESSIE, N. A. C. *Statistics for Spatial Data*, revised ed. John Wiley and Sons, Inc, New York, 1993

[3] CURRIN, C., MITCHELL, T., MORRIS, M., AND YLVISAKER, D. Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association 86*, 416 (1991), 953–963

References 2

[4] DIGGLE, P., THOMSON, M., CHRISTENSEN, O., ROWLINGSON, B., OBSOMER, V., GARDON, J., WANJI, S., TAKOUGANG, I., ENYONG, P., KAMGNO, J., REMME, J., BOUSSINESQ, M., AND MOLYNEUX, D. Spatial modelling and the prediction of loa loa risk: decision making under uncertainty. *Annals of Tropical Medicine and Parasitology 101* (September 2007), 499–509(11)

[5] GELMAN, A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis 1* (2006), 515–533

[6] GELMAN, A., CARLIN, H. B., STERN, H. S., AND RUBIN, D. B. *Bayesian Data Analysis*, second ed. Chapman & Hall/CRC, Boca Raton, Florida, 2004

[7] HAARIO, H., SAKSMAN, E., AND TAMMINEN, J. An adaptive metropolis algorithm. *Bernoulli 7*, 2 (2001), 223–242

References 3

[8] Higdon, D., Gattiker, J., Williams, B., and Rightley, M. Computer model calibration using high-dimensional output. *Journal of the American Statistical Association 103*, 482 (2008), 570–583

[9] Johnson, M. E., Moore, L. M., and Ylvisaker, D. Minimax and maximin distance designs. *Journal of Statistical Planning and Inference 26*, 2 (1990), 131 – 148

[10] Kelsall, J. E., and Diggle, P. J. Spatial variation in risk of disease: A nonparametric binary regression approach. *Applied Statistics 47*, 4 (1998), 559–573

[11] Kennedy, M. C., and O'Hagan, A. Bayesian calibration of computer models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 63*, 3 (2001), 425–464

References 4

[12] Kennedy, M. C., and O'Hagan, A. Supplementary details on bayesian calibration of computer models. Supplementary material for "Bayesian Calibration of Computer Models" published in Journal of the Royal Statistical Society. Series B (Statistical Methodology), volume 63, number 3., 2000. Available from http://shef.ac.uk/ st1ao/ps/calsup.ps

[13] Martin, A. D., Quinn, K. M., and Park, J. H. *MCMCPack: Markov chain Monte Carlo (MCMC) Package*, 2008. R package version 0.9-4

[14] Mitchell, T. J., and Morris, M. D. Bayesian design and analysis of computer experiments: Two examples. *Stat. Sin. 2*, 2 (1992), 359–379

[15] Morris, M. D., and Mitchell, T. J. Exploratory designs for computational experiments. *Journal of Statistical Planning and Inference 43*, 3 (1995), 381 – 402

References 5

[16] MORRIS, M. D., MITCHELL, T. J., AND YLVISAKER, D. Bayesian design and analysis of computer experiments: Use of derivatives in surface prediction. *Technometrics 35*, 3 (1993), 243–255

[17] NELDER, J. A., AND MEAD, R. A Simplex Method for Function Minimization. *The Computer Journal 7*, 4 (1965), 308–313

[18] PLUMMER, M., BEST, N., COWLES, K., AND VINES, K. *coda: Output analysis and diagnostics for MCMC*, 2008. R package version 0.13-3

[19] R DEVELOPMENT CORE TEAM. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0

References 6

[20] RAO, C. R., AND TOUTENBURG, H. *Linear Models: Least Squares and Alternatives*, second ed. Springer-Verlag, New York, 1999

[21] SACKS, J., SCHILLER, S. B., AND WELCH, W. J. Designs for computer experiments. *Technometrics 31*, 1 (1989), 41–47

[22] SACKS, J., WELCH, W. J., MITCHELL, T. J., AND WYNN, H. P. Design and analysis of computer experiments. With comments and a rejoinder by the authors. *Stat. Sci. 4*, 4 (1989), 409–435

[23] SANTNER, T. J., WILLIAMS, B. J., AND NOTZ, W. I. *The Design and Analysis of Computer Experiments*. Springer-Verlag, New York, 2003

[24] SHEWRY, M. C., AND WYNN, H. P. Maximum entropy sampling. *Journal of Applied Statistics 14*, 2 (1987), 165–170

Example Correlation Functions

- Define d to be the distance between x_i and x_j.
- Linear Correlation Function: For $1/2 < \theta < \infty$,
$$R(d,\theta) = \begin{cases} 1 - d/\theta, & d < \theta; \\ 0, & d \geq \theta. \end{cases}$$
- Nonnegative Cubic Correlation Function: For $\theta > 0$,
$$R(d,\theta) = \begin{cases} 1 - 6(d/\theta)^2 + 6(d/\theta)^3, & d < \theta/2\ ; \\ 2(1 - d/\theta)^3, & \theta/2 \leq d < \theta; \\ 0, & d \geq \theta. \end{cases}$$
- Power Exponential Family: For $\theta > 0$ and $0 < p \leq 2$
$$R(d,\theta) = \exp\{-(d/\theta)^p\}$$
- Multi-dimensional Product Correlation Functions:
$$R(\mathbf{d},\boldsymbol{\theta}) = \prod_{i=1}^{k} R(d_i,\theta_i)$$
- Throughout, we use the nonnegative cubic correlation function