Bayesian Reflections on Systematics

- Profiling versus marginalizing
- Asymptotics?
- Gaussian processes for emulation

Results from an example involving 9 nuisance parameters and a bias

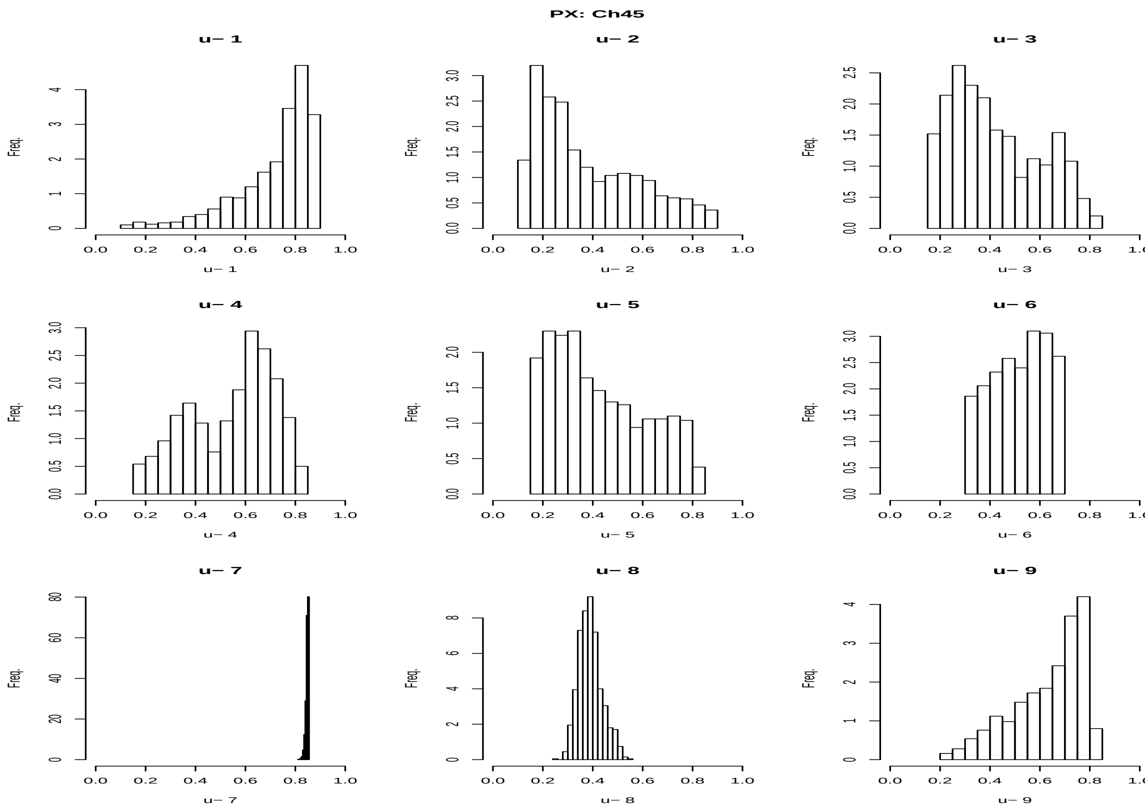
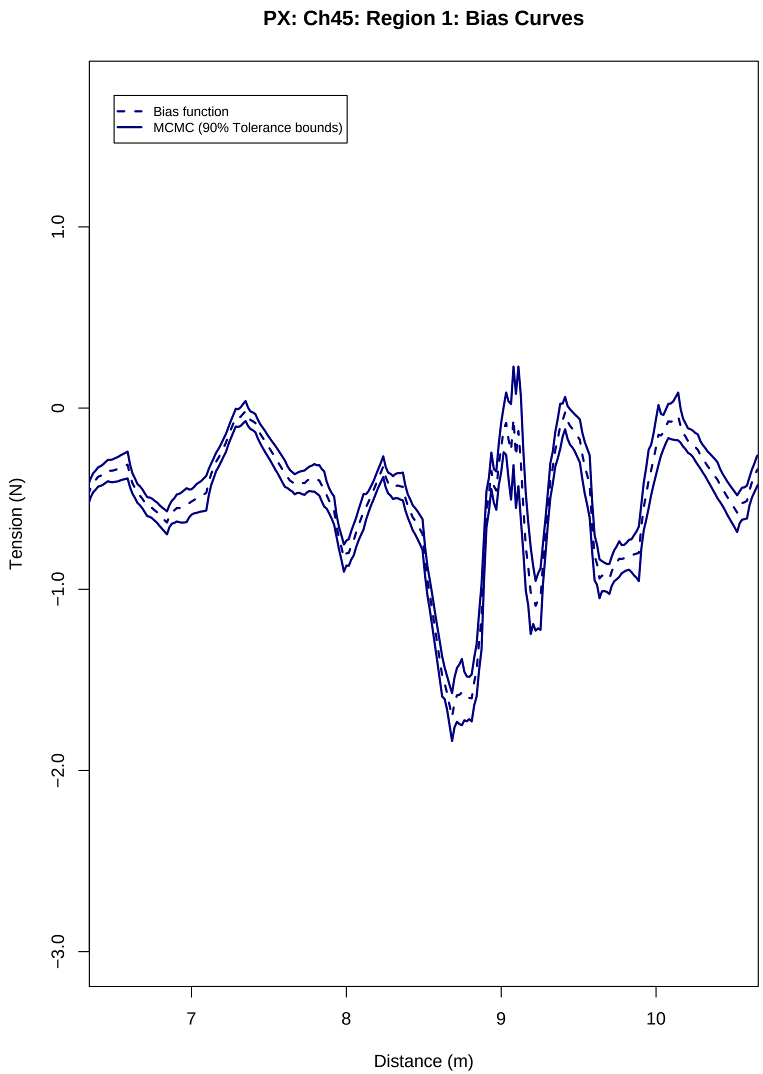


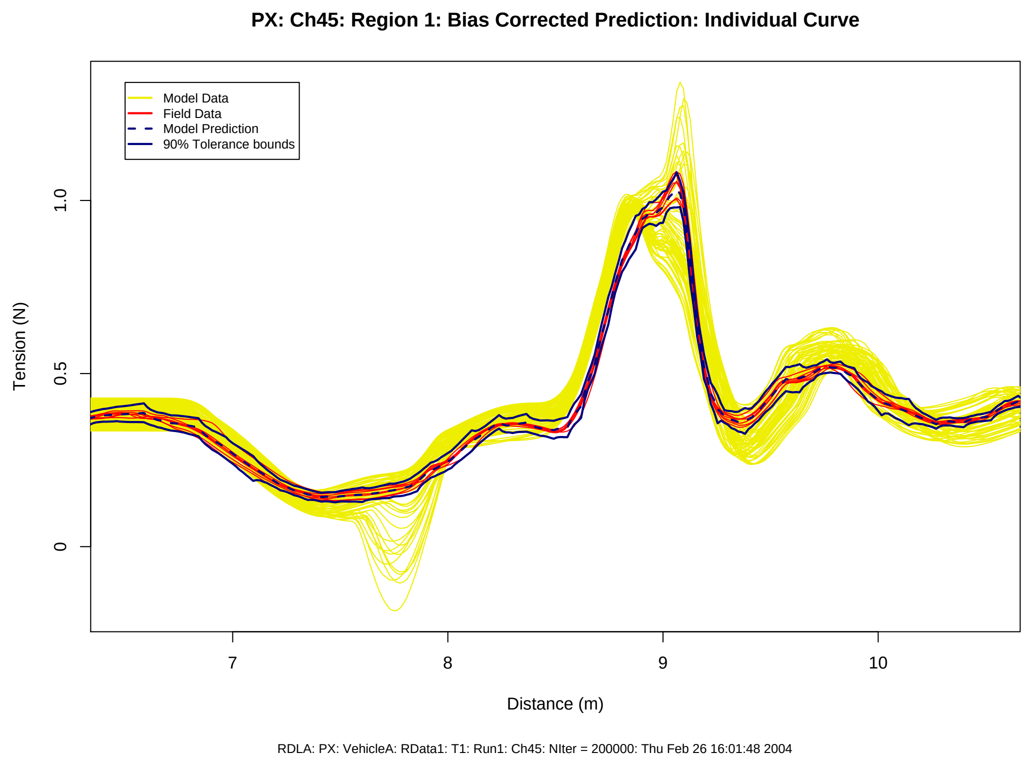
Figure 1: Posterior distributions for the nuisance parameters: uniform priors for the first two and Gaussian priors for the others.

- Profiling would seem fine for parameters 7 and 8.

- Profiling for the others?



- Clear and well estimated bias (Gaussian process prior).



- Obtaining the correct answer for the function of interest required use of the *joint* posterior distribution of the bias and the nuisance parameters.

Constrained nuisance parameters

The problem: We only know that $\theta_i \in (a_i, b_i)$, $i = 1, \ldots, m$.

Standard conservative answer: Uncertainty is
$\tilde{V} = \sup_{\{\theta_1,\ldots,\theta_m\}} V(\theta_1, \ldots, \theta_m)$.

- $\tilde{V}$ is typically much too large to be useful.

Anti conservative answer: Report $V(\frac{a_1+b_1}{2}, \ldots, \frac{a_m+b_m}{2})$.

Reasonable alternative: Use Gaussian distributions fitted to the intervals.

Nice alternative: Use uniform distributions on the intervals.

- This has the nice property of giving the largest uncertainty V over a wide choice of reasonable distributions on the intervals (under independence).

And then there are all sorts of versions with correlations.

Asymptotics?

Hartigan (1966): for any prior having full support and data $\boldsymbol{x}$ of sample size n,

- form the $100(1-\alpha)\%$ equal-tailed posterior credible interval $C(\boldsymbol{x})$;
- consider $C(\boldsymbol{x})$ as a frequentist confidence procedure;
- astonishingly, as the sample size $n \to \infty$, the frequentist coverage of the Bayesian credible sets is $1-\alpha$, up to an error of order C/n for some constant C.
  - This is astonishing because achieving frequentist coverage up to an error of C/n is noteworthy (achieving $C/\sqrt{n}$ is easy - e.g. log-odds method or delta method), and yet Hartigan's result holds for essentially any prior distribution.
  - Small sample frequentist performance of the Bayes credible sets is only good for decent objective priors.

Use of Gaussian processes for emulation (approximation) of simulation models

Simulators (complex models of processes) often take hours to weeks for a single run. One often needs fast simulator approximations (emulators, surrogate models, meta models, response surface approximations, . . .) .

Statistical Design of Runs of the Simulator

(needed to create the emulator) McKay et al. (1979); Sacks et al. (1989); Welch et al. (1992); Bates et al. (1996); Lim et al. (2002); Santner et al. (2003)

Notation: $\boldsymbol{x} \in \boldsymbol{X}$ will denote the d-dimensional vector of inputs (nuisance parameters and parameters of interest) to the simulator. These could be initial conditions, control inputs, model parameters, ...

The simulator output is denoted by $y^M(\boldsymbol{x})$.

Goal for design: Choose m points $D = \{\boldsymbol{x}_1, .., \boldsymbol{x}_m\}$ at which the simulator is to be evaluated, yielding $\boldsymbol{y}^D = (y^M(\boldsymbol{x}_1), \ldots, y^M(\boldsymbol{x}_m))'$. From these, the emulator (approximation) to the simulator will be constructed.

– Folklore says that m should be at least $10d$ although many more runs are often needed (but often not available).

Criterion: In general, should be problem specific. General purpose criteria involve finding "space-filling" designs, such as *Latin Hypercube Designs*.

TITAN2D simulating pyroclastic flow on Montserrat (Bayarri et al. (2009); Dalbey et al. (2012))

Inputs: *Initial conditions:* x_1 = flow volume (V); x_2 = flow direction (φ); *Model parameters:* x_3 = basal friction (δ_{bed}); x_4 = internal friction (δ_{int}).

Background of the application: The simulator, TITAN2D, for given inputs V, φ, δ_{bed}, and δ_{int}, yields a description of the pyroclastic flow over a large space-time grid of points. Each run of TITAN2D takes two hours.

Of primary interest is the maximum (over time) flow height, $y^M(V, \varphi, \delta_{bed}, \delta_{int})$, at k spatial locations over the island.

- Flow heights $y^M(V, \varphi, \delta_{bed}, \delta_{int}) > 1$m are deemed to be catastrophic.

The analysis begins

- by choosing a Latin hypercube design to select $m = 2048$ design points in the feasible input region $\boldsymbol{X} = [10^5\text{m}^3, 10^{9.5}\text{m}^3] \times [0, 2\pi] \times [5, 18] \times [15, 35]$;
- running TITAN2D at these preliminary points, yielding the 'data'

$$\boldsymbol{y}^D = \begin{pmatrix} \boldsymbol{y}^M(\boldsymbol{x}_1) \\ \boldsymbol{y}^M(\boldsymbol{x}_2) \\ \vdots \\ \boldsymbol{y}^M(\boldsymbol{x}_m) \end{pmatrix} = \begin{pmatrix} (y_1^M(\boldsymbol{x}_1), & y_2^M(\boldsymbol{x}_1), & \dots & y_k^M(\boldsymbol{x}_1)) \\ (y_1^M(\boldsymbol{x}_2), & y_2^M(\boldsymbol{x}_2), & \dots & y_k^M(\boldsymbol{x}_2)) \\ & & \vdots & \\ (y_1^M(\boldsymbol{x}_m), & y_2^M(\boldsymbol{x}_m), & \dots & y_k^M(\boldsymbol{x}_m)) \end{pmatrix};$$

- constructing the emulator from $\boldsymbol{y}^D$ (in general, a matrix of size 2048×10^9).

GaSP Emulators

Model the *real-valued for now* simulator output $y^M(\boldsymbol{x})$ as an *unknown* function via a *Gaussian stochastic process*:

$$y^M(\cdot) \sim \mathrm{GaSP}(\,\mu(\cdot),\, \sigma^2\, c\,(\cdot,\cdot)),$$

with mean function $\mu(\boldsymbol{x})$, variance σ^2, and correlation function $c\,(\cdot,\cdot)$ if, for any inputs $\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_m\}$ from $\boldsymbol{X}$,

$$(y^M(\boldsymbol{x}_1),\ldots,y^M(\boldsymbol{x}_m)) \sim MVN\left((\mu(\boldsymbol{x}_1),\ldots,\mu(\boldsymbol{x}_m)),\sigma^2\boldsymbol{C}\right)\;, \tag{1}$$

where $\boldsymbol{C}$ is the $m \times m$ correlation matrix with (i,j) element $c_{i,j} = c\,(\boldsymbol{x}_i,\boldsymbol{x}_j)$, with the separable Matérn correlation function

$$c\,(\boldsymbol{x},\boldsymbol{x}^*) = \prod_{j=1}^{4}\left[\left(1 + \frac{\sqrt{5}\;|x_j - x_j^*|}{\gamma_j} + \frac{5\;|x_j - x_j^*|^2}{3\gamma_j^2}\right)\exp\left(-\frac{\sqrt{5}\;|x_j - x_j^*|}{\gamma_j}\right)\right].$$

The simplest imaginable GaSP for the k-dimensional $\boldsymbol{y}^M(\boldsymbol{x}) = (y_1^M(\boldsymbol{x}), y_2^M(\boldsymbol{x}), \ldots y_k^M(\boldsymbol{x}))$ (Gu et al. (2016)) :

An *independent* GaSP is assigned to each coordinate $y_j^M(\boldsymbol{x})$, with

- prior mean functions of the regression form $\mu_j(\boldsymbol{x}) = \Psi(\boldsymbol{x})\boldsymbol{\theta}_j$, where $\Psi(\boldsymbol{x})$ is a *common* l-vector of given basis functions (equal to $(1, x_1) = (1, V)$ in the example) and the $\boldsymbol{\theta}_j$ are *differing* unknown regression coefficients;
- *differing* unknown prior variances σ_j^2;
- *common* estimated correlation parameters $\hat{\boldsymbol{\gamma}}$.

In the analysis,

- The $\boldsymbol{\theta}_j$ and σ_j^2 are dealt with in a fully Bayesian manner using $\pi(\boldsymbol{\theta}_j, \sigma_j^2) = 1/\sigma_j^2$.
- $\hat{\boldsymbol{\gamma}}$ is the marginal posterior mode arising from use of the reference prior for correlation parameters, as implemented in **RobustGaSP: Robust Gaussian Stochastic Process Emulation in R** by Gu and Palomo.
  - Note that maximum likelihood can give very poor performance here.

The mean function of the posterior GaSP for $y_j^M(\boldsymbol{x}^*)$ at new input $\boldsymbol{x}^*$ is

$$\hat{\mu}_j(\boldsymbol{x}^*) = \Psi(\boldsymbol{x}^*)\hat{\boldsymbol{\theta}}_j + C(\boldsymbol{x}^*)\boldsymbol{C}^{-1}(\boldsymbol{y}_j^D - \boldsymbol{\Psi}\hat{\boldsymbol{\theta}}_j)\,,$$

where $\boldsymbol{y}_j^D$ is the j^{th} column of $\boldsymbol{y}^D$ and $\hat{\boldsymbol{\theta}}_j = (\boldsymbol{\Psi}'\boldsymbol{C}^{-1}\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}'\boldsymbol{C}^{-1}\boldsymbol{y}_j^D$, with $\boldsymbol{\Psi}$ being the $m \times l$ matrix whose rows are the bases evaluated at each design point, $\boldsymbol{C}$ being the earlier specified $m \times m$ correlation matrix, $\Psi(\boldsymbol{x}^*) = (\Psi_1(\boldsymbol{x}^*), \ldots, \Psi_l(\boldsymbol{x}^*))$, and $C(\boldsymbol{x}^*) = (c(\boldsymbol{x}_1, \boldsymbol{x}^*), \ldots, c(\boldsymbol{x}_m, \boldsymbol{x}^*))$.

This can be rewritten

$$\hat{\mu}_j(\boldsymbol{x}^*) = \sum_{i=1}^{m} h_i(\boldsymbol{x}^*) y_{ij}^D\,,$$

where $h_i(\boldsymbol{x}^*)$ is the i^{th} element of the m-vector

$$\boldsymbol{h}(\boldsymbol{x}^*) = \left[(\Psi(\boldsymbol{x}^*) - C(\boldsymbol{x}^*)\boldsymbol{C}^{-1}\boldsymbol{\Psi})(\boldsymbol{\Psi}'\boldsymbol{C}^{-1}\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}'\boldsymbol{C}^{-1} + C(\boldsymbol{x}^*)\boldsymbol{C}^{-1}\right]\,.$$

As $\boldsymbol{\Psi}$ and $\boldsymbol{C}$ (and the functions of them) can be pre-computed, computing $\boldsymbol{h}(\boldsymbol{x}^*)$ (at a new $\boldsymbol{x}^*$) requires roughly m^2 numerical operations.

Finally, we can write the complete parallel partial posterior (PP) mean vector (the emulator of the full simulator output at a new input $\boldsymbol{x}^*$) as

$$\hat{\boldsymbol{\mu}}(\boldsymbol{x}^*) = (\hat{\mu}_1(\boldsymbol{x}^*), \ldots, \hat{\mu}_k(\boldsymbol{x}^*)) = \boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{y}^D \,.$$

- Computation of $\boldsymbol{h}(\boldsymbol{x}^*)$ requires m^2 computations.
- Computation requires only $2mk$ arithmetic operations if $k >> m$.
  - It is crucially important to have differing $\boldsymbol{\theta}_j$ and σ_j^2 at each coordinate, but this comes with essentially no computational cost.
- The emulator is an interpolator so, when $\boldsymbol{x}^*$ equals one of the runs $\boldsymbol{x}_i$, the emulator will return the exact values from the computer run.
- As the emulator mean is just a weighted average of the actual simulator runs, it hopefully captures some of the dynamics of the process.
- Using the posterior GaSP mean and variance to construct confidence intervals seems to yield reasonable coverage in almost all applications.

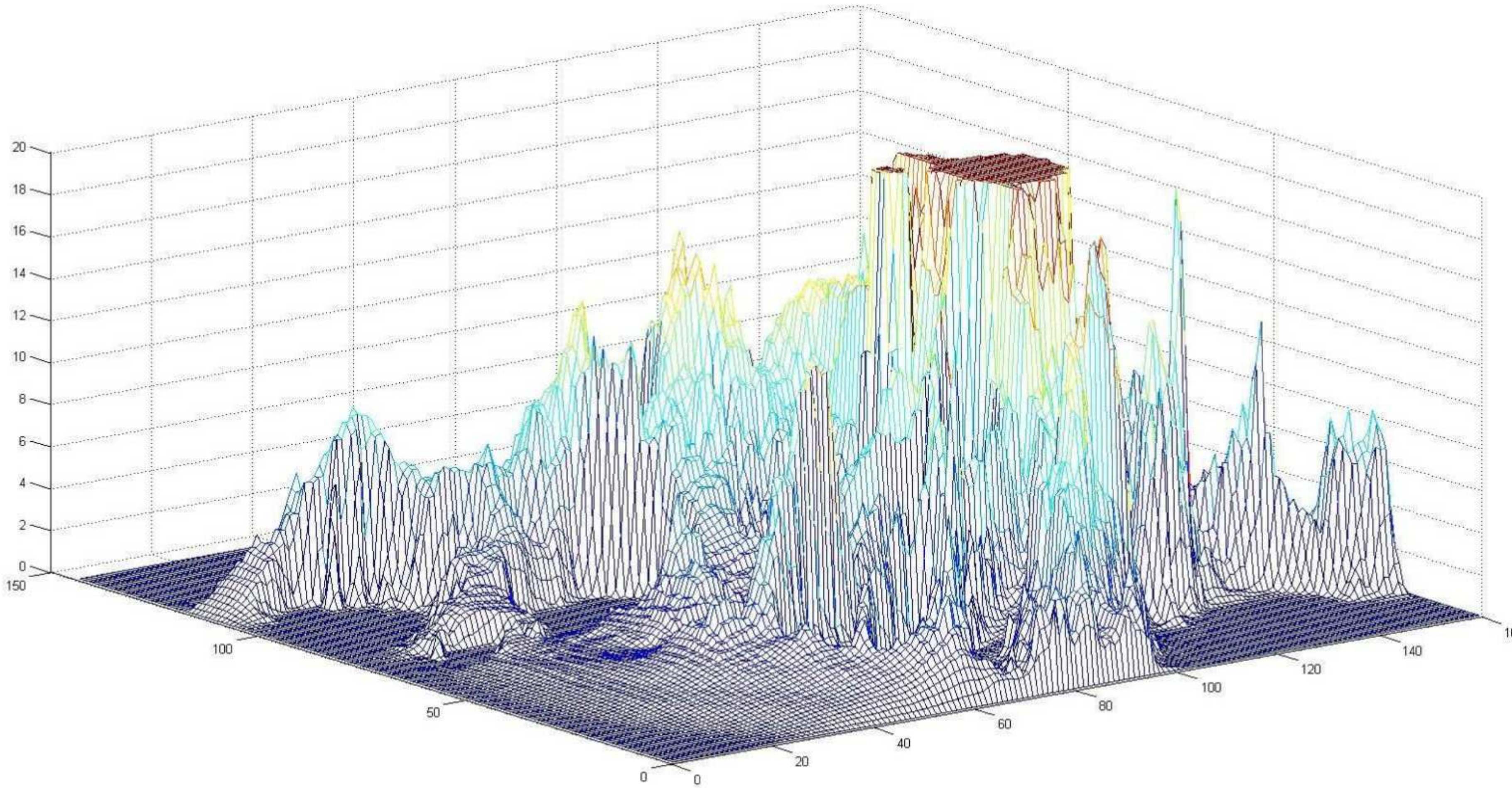


Figure 2: The mean of the emulator of 'maximum flow height over time' from TITAN2D, at 24,000 spatial locations over Montserrat and for new input values $V = 10^{7.462}$, $\varphi = 2.827$, $\delta_{bed} = 11.111$, and $\delta_{int} = 27.7373$.

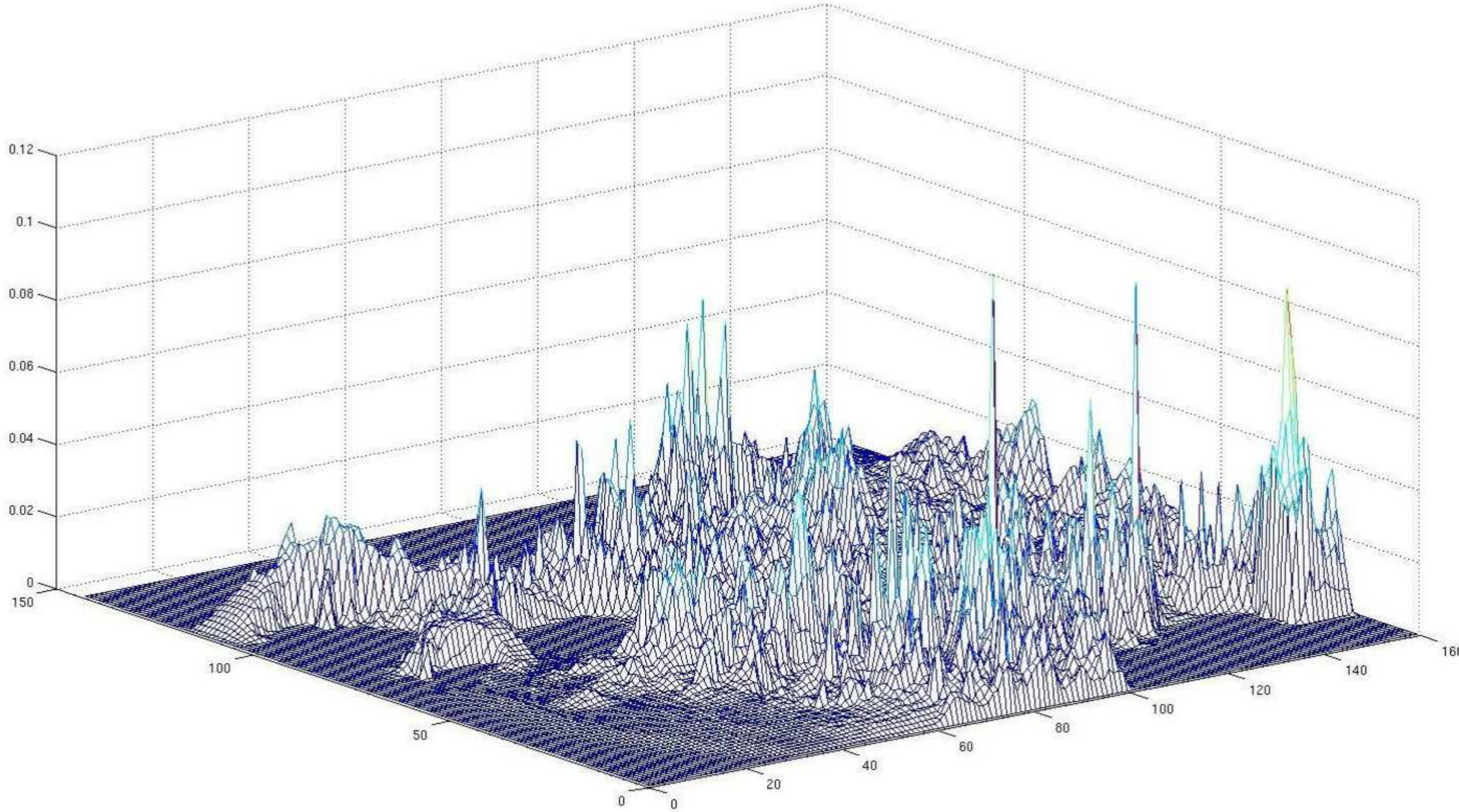


Figure 3: Variance of the emulator of 'maximum flow height over time' from TITAN2D, at 24,000 spatial locations over Montserrat and for new input values $V = 10^{7.462}$, $\varphi = 2.827$, $\delta_{bed} = 11.111$, and $\delta_{int} = 27.7373$.

Movie Time

Determination of 1m contours of maximum flow over time at $k = 23,040$ spatial locations at various inputs to develop the emulator.

Generalizations

- When the simulator output is a distribution:
  - if the distribution has a known form (up to hyperparameters), use GaSP's to model the hyperparameters;
  - nonparametric things can also be done.
- If input parameters have Gaussian distributions, closed form expressions for the mean and covariance of the resulting posterior process are available.
- Differing GaSP simulation models can be linked analytically.
- Discrete input parameters can be handled by
  - developing separate GaSPs for each discrete value;
  - treating the parameter as continuous if it has a number of values.

References

Bates, R. A., R. J. Buck, E. Riccomagno, and H. P. Wynn (1996). Experimental design and observation for large systems (Disc: p95–111). *Journal of the Royal Statistical Society, Series B, Methodological 58*, 77–94.

Bayarri, M. J., J. O. Berger, E. S. Calder, K. Dalbey, S. Lunagomez, A. K. Patra, E. B. Pitman, E. T. Spiller, and R. L. Wolpert (2009). Using statistical and computer models to quantify volcanic hazards. *Technometrics 51*, 402–413.

Dalbey, K., M. Jones, E. B. Pitman, E. S. Calder, M. Bursik, and A. K. Patra (2012). Hazard risk analysis using computer models of physical phenomena and surrogate statistical models. *Int. J. for Unceratainty Quantification*. To appear.

Gu, M., J. O. Berger, et al. (2016). Parallel partial gaussian process emulation for computer models with massive output. *The Annals of Applied Statistics 10*(3), 1317–1347.

Hartigan, J. (1966). Note on the confidence-prior of welch and peers. *Journal of the Royal Statistical Society: Series B (Methodological) 28*(1), 55–56.

Lim, Y. B., J. Sacks, W. Studden, and W. J. Welch (2002). Design and analysis of computer experiments when the output is highly correlated over the input space. *Canadian Journal of Statistics 30*(1), 109–126.

McKay, M. D., W. J. Conover, and R. J. Beckman (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics 21*, 239–245.

Sacks, J., W. J. Welch, T. J. Mitchell, and H. P. Wynn (1989). Design and analysis of computer experiments (C/R: p423–435). *Statistical Science 4*, 409–423.

Santner, T. J., B. Williams, and W. Notz (2003). *The Design and Analysis of Computer Experiments*. Springer-Verlag.

Welch, W. J., R. J. Buck, J. Sacks, H. P. Wynn, T. J. Mitchell, and M. D. Morris (1992). Screening, predicting, and computer experiments. *Technometrics 34*, 15–25.