

Linking Statistical Emulators

Ksenia Kyzyurova; Jim Berger and Robert Wolpert (advisors)

Department of Statistical Science, Duke University

Motivation

- Gaussian processes (GaSPs), together with an objective Bayesian implementation, are a common tool for emulating complex computer models of processes.[1]
- More than one computer model may need to be utilized for the predictive goal or inferential purpose.
- This project *goal* is to couple a computer model of a volcanic pyroclastic flow with a model of basal friction (the friction of the pyroclastic flow with the surface).

Simulation

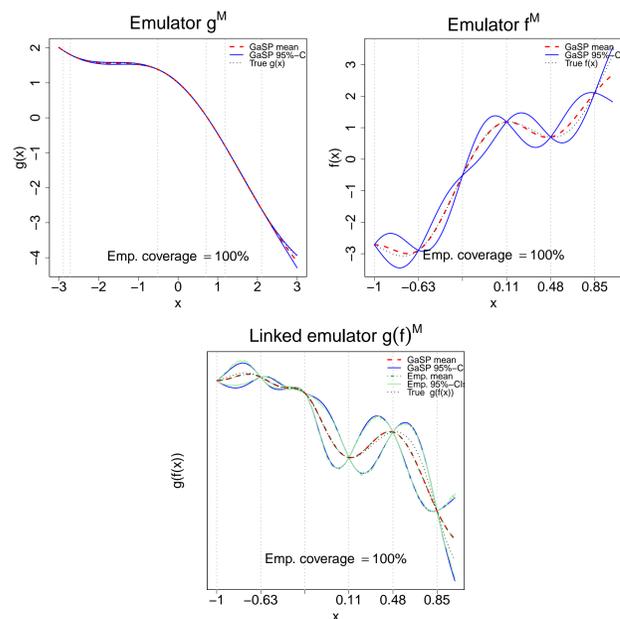


Figure 1: The results of applying the methodology to two test functions:

$g(x) = \cos(x) - x$ in the range $x \in [-3, 3]$ and $f(x) = 3x + \cos(5x)$ in the range $x \in [-1, 1]$ to emulate $g \circ f(x)$.

Dotted vertical lines correspond to training inputs which were used to fit emulators.

The coupled emulator is doing exceptionally well, acting as an interpolator at the design points and providing coverage much better than the nominal coverage.

Methodology

Model likelihood for the data g^M resulting from the runs of the computer model g at computer model input x_g , given parameters θ_g , is

$$g^M(x_g)|\theta_g \sim \text{MVN}(\mu(x_g), \sigma^2 C)$$

where σ^2 is the unknown variance and C is the correlation matrix whose (k, l) element is given by $c(x_g^k, x_g^l)$. $c(\cdot, \cdot)$ is a power exponential correlation function with smoothness parameter $\alpha \in (0, 2]$.

Objective Bayesian implementation allows learning about the model parameters θ_g . *GaSP emulator* is the predictive posterior distribution, given parameters θ_g , at a new input x'_g :

$$g^M(x'_g)|g^M(x_g), \theta_g \sim \text{GaSP}(\mu_{\theta_g}^*(x'_g), \sigma_{\theta_g}^*).$$

Now, suppose that a new input to the computer model g is obtained as *another GaSP emulator* f^M of another computer model f . Analogously, this emulator is of the form

$$f^M(x'_f)|f^M(x_f), \theta_f \sim \text{GaSP}(\mu_{\theta_f}^*(x'_f), \sigma_{\theta_f}^*).$$

We are interested in the marginal distribution of an emulator of a composite computer model $g \circ f$ at a new input x'_f , that is

$$\begin{aligned} & p((g \circ f)^M(x'_f)|g^M(x_g), f^M(x_f), \theta_f, \theta_g, x'_f) \\ &= \int p(g^M(f^M(x'_f))|g^M(x_g), f^M(x'_f), \theta_g) p(f^M(x'_f)|f^M(x_f), \theta_f) df^M(x'_f). \end{aligned}$$

We call $\xi = (g \circ f)^M(x'_f)|g^M(x_g), f^M(x_f), \theta_f, \theta_g, x'_f$ the *linked emulator*. It does not have a closed form distribution, but

If the emulator g^M has correlation function smoothness parameter $\alpha = 2$ the mean $E\xi$ and the variance $V\xi$ of the linked emulator are available in closed form.

Simulation and examples show that the linked emulator is well approximated by $N(E\xi, V\xi)$. This is also shown analytically with a Taylor's series approximation.

Pyroclastic flow at Soufrière Hills Volcano

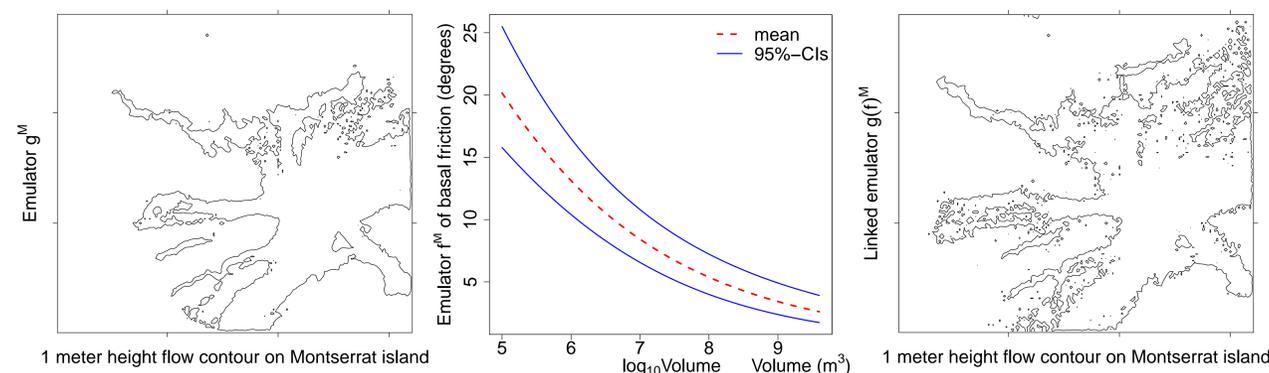


Figure 2: Left: Emulator g^M 1 meter height flow spatial contours on Montserrat island for input values: volume of a pyroclastic flow $V = 10^{7.25}$, initiation angle $\phi = -31.57^\circ$ and basal friction $\delta = 11.12^\circ$.

Center: Emulator f^M of uncertain basal friction.

Right: Resulting linked emulator. The pyroclastic flow has changed its contour given the basal friction uncertainty.

Conclusions

- Linking statistical emulators is a very promising alternative to direct coupling of corresponding computer models.
- The method provides a framework for developing an emulator of a system of emulators of multiple computer models.

Future work

- The ultimate goal of this project is to assess probabilities of geophysical hazardous events on Montserrat island and to identify conditions which lead to these events.

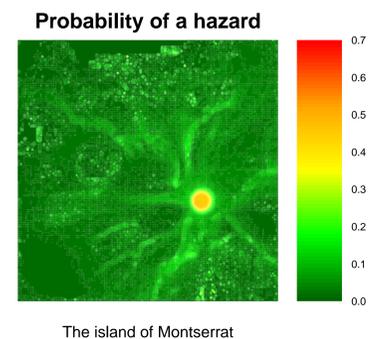


Figure 3: The preliminary map of probabilities of a hazard obtained at more than 20 thousand locations on the island of Montserrat. These probability calculations are based on the distribution of the linked emulator of the computer model for the height of a pyroclastic flow from Soufrière Hills Volcano at various locations on the island and on specification that the volume of the pyroclastic flow is $V \sim \text{Pareto}(5 \cdot 10^4, .64)$ and the initiation angle is $\theta \sim U(0, 2\pi)$. [1]

References

- [1] M.J. Bayarri and J.O. Berger and E.S. Calder and K. Dalbey and S. Lunagomez and A.K. Patra and E.B. Pitman and E.T. Spiller and R.L. Wolpert (2009). Using Statistical and Computer Models to Quantify Volcanic Hazards. *Technometrics*, 51(4), 402-413.

Contact

E-mail ksenia@stat.duke.edu

Web <http://stat.duke.edu/~kk194/>