Gaussian Stochastic Processes for Hazard Mapping

Researchers: Tao Cui   Mentor: Elaine Spiller, PhD

Marquette University – MSCS Department

Introduction

Hazard mapping is an essential tool used to estimate the risk faced by residents living in areas susceptible to natural disasters. Computer models are often used in situations where experimental data is costly or impossible to obtain. However, accurate computer models can take hours or even days to compute, which is problematic when attempting to validate the model, which requires running the model hundreds or thousands of times.

We analyzed statistical surrogates using Gaussian stochastic process (GaSP) models that can be computed more efficiently than a computer model, and aim to apply this technique to validate landslide models using data gathered from previously conducted simulations.

Fig 1: A Landslide at the No. 3 Freeway between Taipei and Keelung,Taiwan[1].

Objectives

- Use Gaussian stochastic processing (GaSP) to approximate models
- Incorporate experimental data to account for bias in approximation
- Utilize Markov chain Monte Carlo (MCMC) techniques to select parameters for the computer model that best match field data

Approximating Computer Models

A Gaussian process approximation can be used as a statistical surrogate for a computer model. The surrogate takes the form of a multivariate normal distribution with conditional mean **m(x*)** described below[2,3]:

$$m(x^*) = \Psi(x^*) \times \hat{\theta} + R(x^*) \times R(\beta)^{-1} \times (Y - X\hat{\theta})$$

where:

X – *inputs (design points);* Y – *output*

$\Psi(x^*)$ – *linear regression of* x^*

$$\hat{\theta} = (X^T R(\beta)^{-1} X)^{-1} X^T R(\beta)^{-1} Y$$

$R(\beta)$ – *covariance matrix with elements defined by*

$$r(x_i, x_j) = \prod_{k=1}^{p} \exp(-\beta_k |x_i - x_j|^2)$$

$R(x^*)$ – *vector with elements* $r(x_i, x^*)$

Finding Parameters for Computer Model

In order to compute **m(x*)**, we must select values for the correlation parameters **β.** One selection method is to choose the parameters that optimize the following maximum likelihood equation:

$$L(\beta) \propto |R(\beta)|^{-1/2} |X^T R(\beta)^{-1} X|^{-1/2} (S^2(\beta))^{-(n-p)/2}$$

with $S^2(\beta) = (Y - X\hat{\theta})^T R(\beta)^{-1} (Y - X\hat{\theta})$

Further, the selection process can be improved by incorporating a reference prior into the optimization equation. Below is an example where a GaSP is used to approximate an arbitrary function.

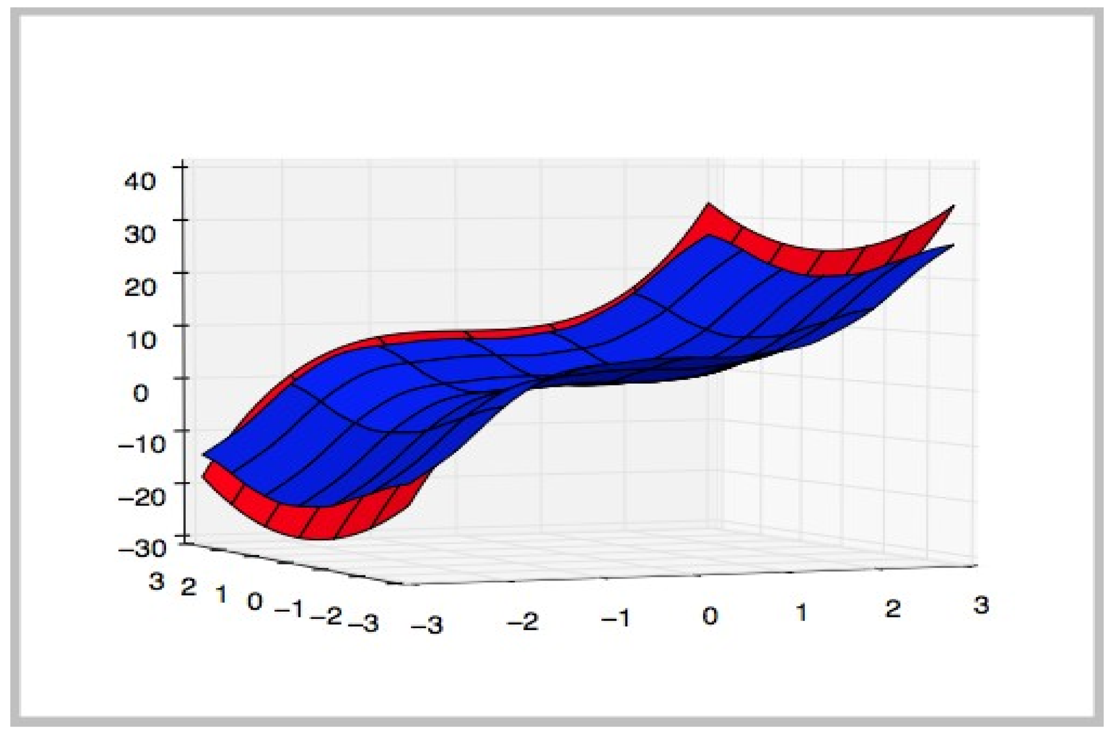

Fig 2: 3D Approximation using 10 design points.

After obtaining an approximation for the computer model, the model can be compared to data obtained through experiments or simulations. This field data is used to determine the bias in the model, which reflects the difference between the model and "reality" (as indicated by the field data). The bias function can be calculated using another GaSP, adjusting the covariance matrix to include measurement error. Below is a example of a bias-corrected prediction compared with a prediction solely using the model approximation.

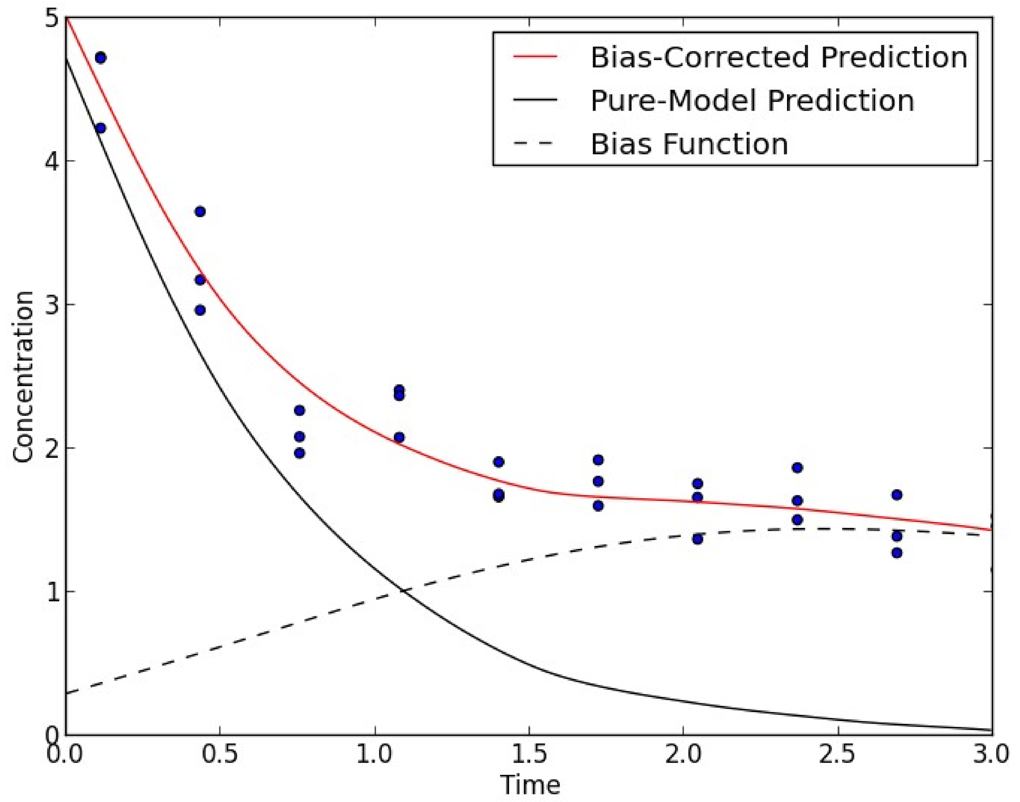

Fig 3: Model approximation adjusted for bias.

Markov Chain Monte Carlo (MCMC) Process

Markov chain Monte Carlo (MCMC) processes can be used in situations where computing a density function is otherwise computationally difficult. We aim to model the stationary distribution π(x) using the following process:

1. Select the first element in the Markov chain x_0
2. To find x_n, sample a new value z from the sampling distribution $q(x_0,z)$
3. Set x_{n+1} = z with probability $\pi(x) * q(x_0,z) / q(z,x_0) * \pi(x)$. Otherwise, set $x_{n+1} = x_n$

A demonstration of this process is shown below.

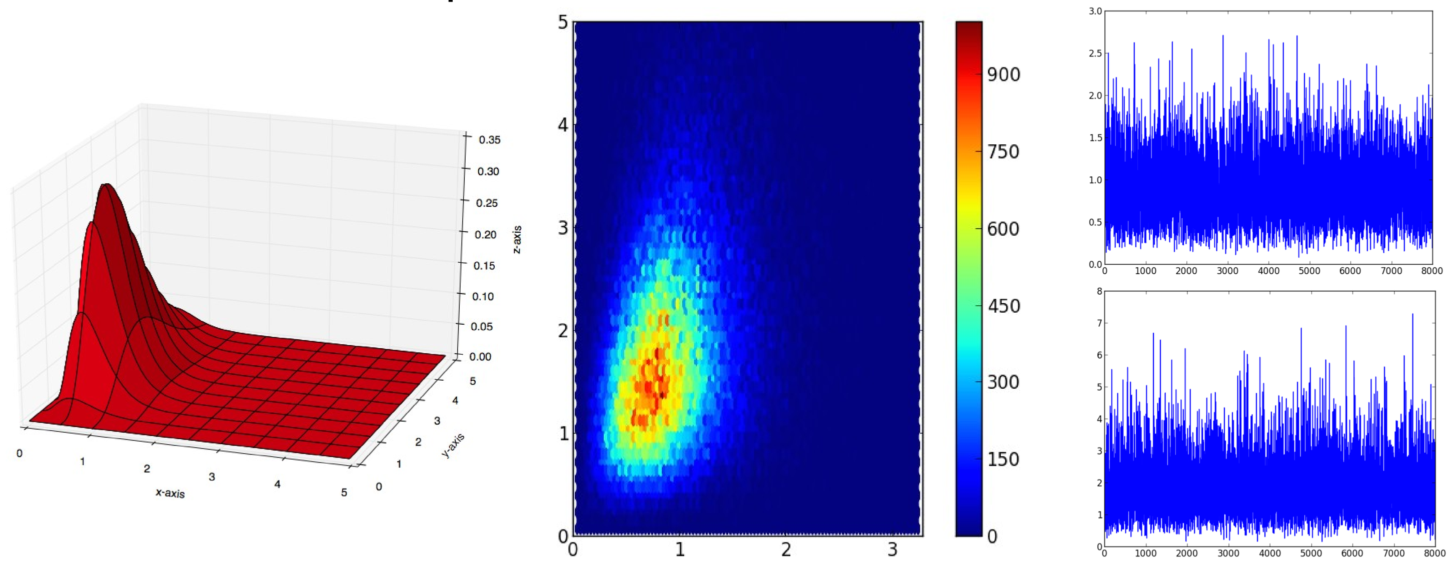

Fig 4: The target distribution (left), coordinates from MCMC samples as a histogram (center) and as a plot over time (right).

Conclusion

The model that we have created allows the user to approximate and map hazard functions in a way that was previously not possible. We intend to move forward by creating an approximation using real world landslide data to validate our model.

With this approximation, we can adjust for the bias in our model by utilizing data obtained through simulations, as shown in Figure 5. Using MCMC techniques, we can also calculate a range of values for the parameters in our computer model that best fit the field data.

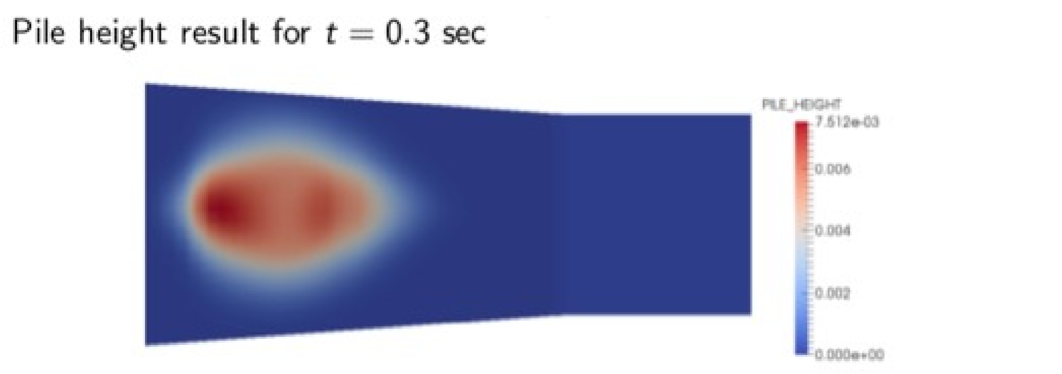

Fig 5: Experimental data from landslide simulation[2].

Acknowledgements

We would like to thank Elaine Spiller for her guidance throughout our research. We also wish to thank Dennis Brylow and Kim Factor, who organized the Research Experience for Undergraduates (REU) at Marquette University. Lastly, we wish to acknowledge the MSCS department and the National Science Foundation (NSF) for their financial support.

Contact

Tao Cui
Department of Mathematics, Statistics, and Computer Science
Marquette University
Tao.Cui@marquette.edu

References

1. Feww (September 1, 2013) "Car caught in sudden landslide," [Digital image]. Retrieved from https://feww.wordpress.com/tag/taiwan-landslide/
2. Sacks, J., and Welch, W. (2010). "Design and Analysis of Computer Experiments," [Course notes]. *National Institute of Statistical Sciences.*
3. Bayarri, M., et al. (2007). "A Framework for Validation of Computer Models," *Technometrics*, **49:2**, 138-154.
4. Vidakovic, B. (2004). "MCMC Methodology," [Course notes]. *Georgia Tech University.*
5. Experimental Data from Marcus Bursik
6. Paulo, R. (2005). "Default priors for Gaussian processes," *Annals of Statistics*, **33:2,** 556-582.
7. John Bihn, Dakota Sullivan, Dr. Elaine Spiller