

# USING KNOWN BOUNDARIES TO IMPROVE BAYESIAN EMULATION

REBEKAH FEARNHEAD DEPARTMENT OF MATHEMATICAL SCIENCES, DURHAM UNIVERSITY

## INTRODUCTION

Complex Mathematical Models are being used in many scientific areas to help describe complex physical systems. For example, in systems biology, predator prey models, or to simulate galaxy formation. One problem with this is that it could take from a week to a month to do a single evaluation. One way to solve this problem is to use an emulator which mimics the computer model function that we want to evaluate, and does it much faster than the full model for the complex system. An emulator provides a prediction and uncertainty statement for the outputs of the function,  $f(x)$ .

## EMULATION IN 2D

To emulate in two dimensions, the simplest model that can be used is  $f(x) = u(x)$ , where  $u(x)$  is a weakly stationary process. Once runs,  $D$ , of the model have been performed,  $f(x)$  needs to be updated, using:

$$E_D[f(x)] = E[f(x)] + \text{Cov}[f(x), D]\text{Var}[D]^{-1}(D - E[D]),$$

$$\text{Var}_D[f(x)] = \text{Var}[f(x)] - \text{Cov}[f(x), D]\text{Var}[D]^{-1}\text{Cov}[D, f(x)],$$

where  $E_D[f(x)]$  and  $\text{Var}_D[f(x)]$  are the adjusted expectation and variance of  $f(x)$ .

For example, using the function,

$$f(x) = -\sin(2\pi x_2) + 0.9 \sin(2\pi(1-x_1)(1-x_2)),$$

and evaluating runs at 9 design points, the following results are produced:

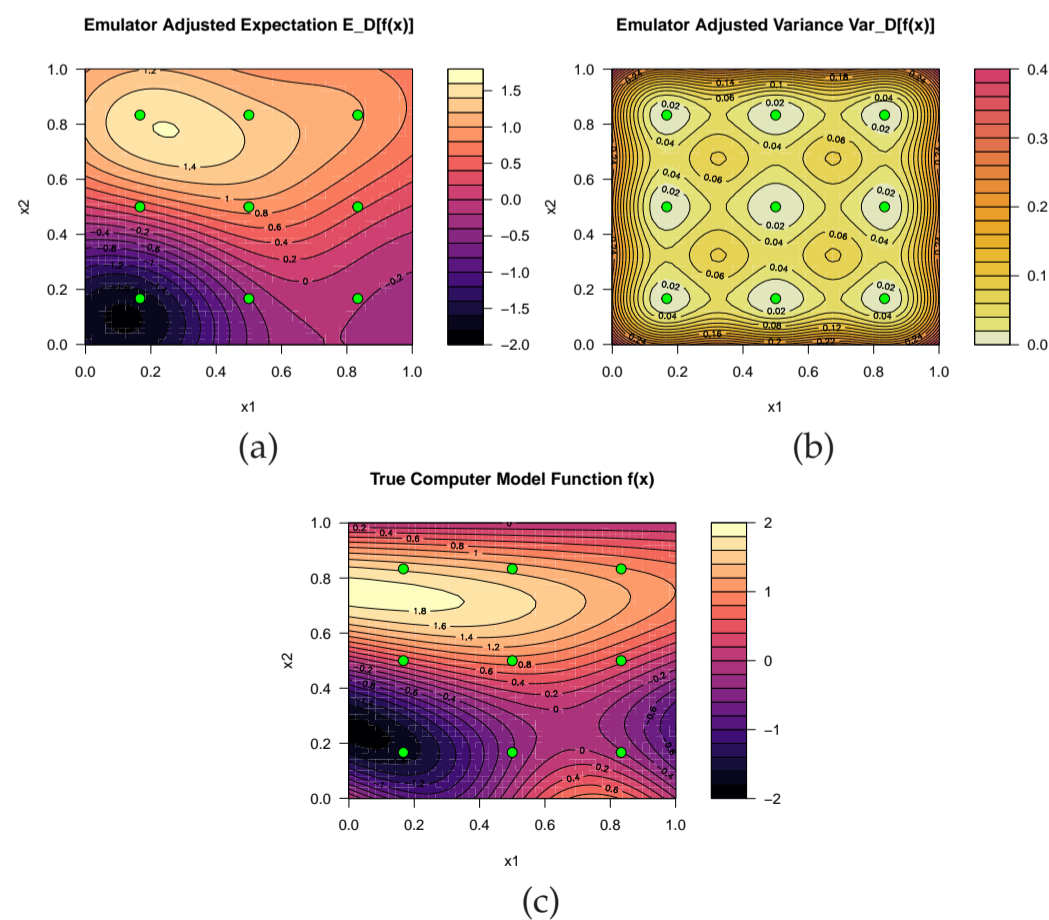


Figure 2: The a) Expectation, and b) Variance, of the Emulator given 9 design points, and c) the True Function value.

## IMPROVING DESIGN

Whilst the emulator in the previous example performs well and models the characteristics of the function well, sometimes it struggles to model some details, for example periodic behaviour.

To improve this, a Latin Hypercube Design can be used to pick the points to evaluate at, by splitting each axis into  $n$  sections and making sure that there is one point in each of the  $n$  sections for each axis.

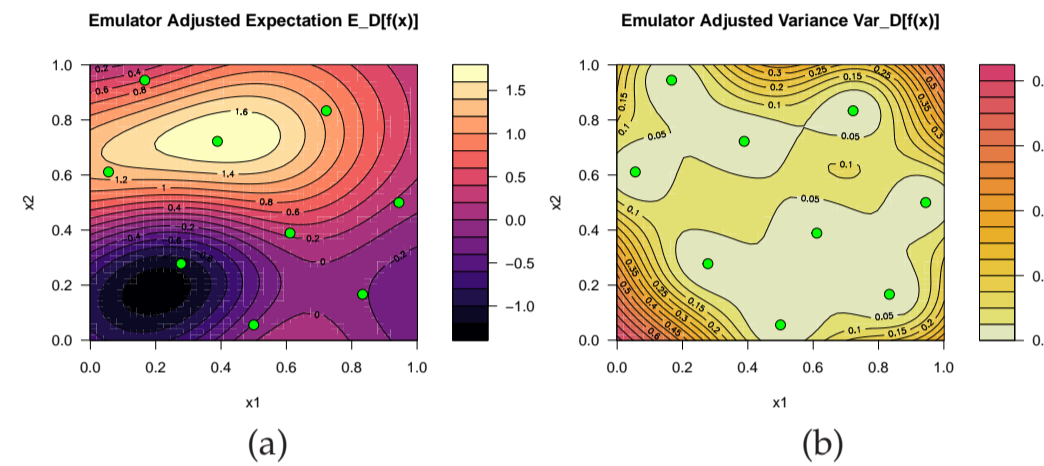


Figure 3: The a) Expectation, and b) Variance, of the Emulator on the example function, given 9 design points in a Latin Hypercube Design.

## ADDING A KNOWN BOUNDARY

Even in complicated models, sometimes setting a variable to a given value will simplify its equations, possibly to a constant. Therefore, one way to improve the performance of an emulator, without having to evaluate more runs, is to add a known boundary to it which can be easily calculated.

Given a known boundary,  $\mathcal{K}$ , the orthogonal projection,  $x^K$ , of point  $x$ , onto the boundary at distance  $a$ , and  $r_1(a)$ , the correlation structure:

$$E_{\mathcal{K}}[f(x)] = E[f(x)] + r_1(a)(f(x^K) - E[f(x^K)]),$$

$$\text{Cov}_{\mathcal{K}}[f(x), f(x')] = \sigma^2(r_1(a-a') - r_1(a)r_1(a'))r_{-1}(x^K - x'^K),$$

$$\text{Var}_{\mathcal{K}}[f(x)] = \sigma^2(1 - r_1(a)^2).$$

Adding the known boundary at  $x_1 = 0$  on the example function, without any design points can already capture some of its behaviour.

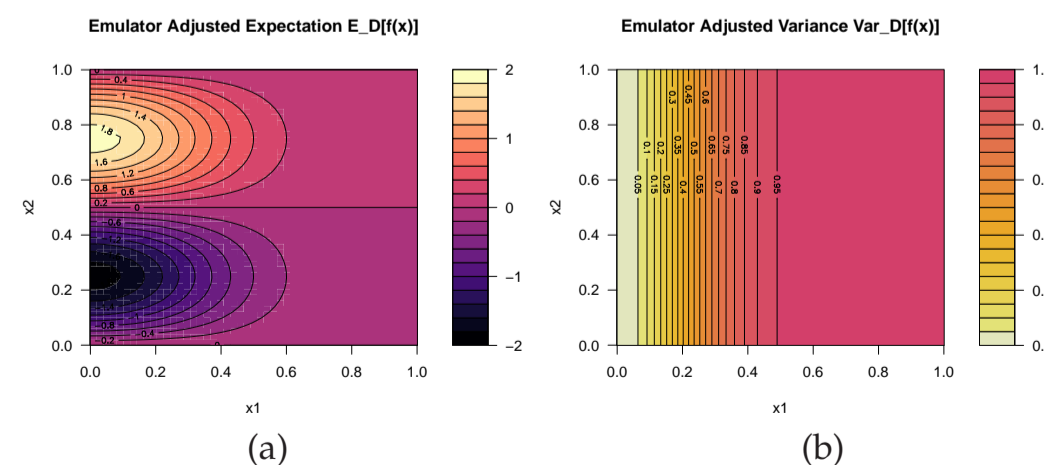


Figure 4: The a) Expectation, and b) Variance, of the Emulator with a known boundary.

## LOTKA-VOLTERRA MODEL

The Lotka-Volterra Predator-Prey model uses deterministic differential equations to describe the populations of prey ( $g_1$ ) and predators ( $g_2$ ) over a time series,  $t$ :

$$\frac{dg_1}{dt} = x_1 g_1 - x_2 g_1 g_2,$$

$$\frac{dg_2}{dt} = x_2 g_1 g_2 - x_3 g_2,$$

where the inputs  $x_1$ ,  $x_2$  and  $x_3$  represent the reproduction speed of the prey, the predator-prey interaction, and the death rate of the predators respectively.

Using suggested input values  $x_1 = 1$ ,  $x_2 = 0.00044$ , and  $x_3 = 1.8$ , and starting populations of 2000 prey and 800 predators produces the following behaviour for the populations over time:

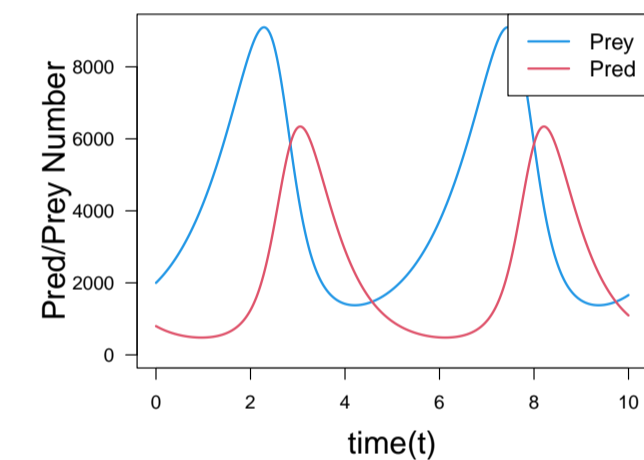


Figure 1: Predator and Prey Population over time.

## FUTURE RESEARCH

- Applying the known boundary technique to the Lotka-Volterra model would work well, since, setting any one of the input variables to 0 would simplify the equations, making them much faster to evaluate. For example, setting  $x_2$ , the predator-prey interaction to 0 will cause the prey population to continuously grow, and the predators to eventually die out.
- More outputs or variables could be added to the Lotka-Volterra model, for example, additional animals to interact with the predator and prey, or adding a natural death rate for the prey.
- Additionally, a second known boundary could be added to the emulator, either perpendicular or parallel to the first known boundary.

## REFERENCES

- [1] Vernon, I., Jackson, S. E., and Cumming, J. A. Known boundary emulation of complex computer models. *SIAM/ASA Journal on Uncertainty Quantification*, 7(3):838–876, 2019.