This document contains supplementary material for the paper '*mlegp*: statistical analysis for computer models of biological systems using R ', by Garrett M. Dancik and Karin S. Dorman.

1 METHODS

Only methods not discussed in the main paper are described here. A cross-validated prediction of the observation $z(x^{(i)})$ is the prediction obtained following the removal of all observations at $x^{(i)}$ from the training set z_{obs}. Denoting $z_{\text{obs,-i}}$ as z_{obs} without observations at $x^{(i)}$, the cross-validated prediction is $\text{E}\left[z(x^{(i)}) \mid z_{\text{obs,-i}}\right]$. Plotting cross-validated predictions $\text{E}\left[z(x^{(i)}) \mid z_{\text{obs,-i}}\right]$ against observation $z(x^{(i)})$ produces diagnostic plots, with predicted variances $\text{Var}\left[z(x^{(i)}) \mid z_{\text{obs,-i}}\right]$ used to add confidence bounds.

To evaluate the benefit of accounting for heterscedasticity in the data, we analyze pathogen load at five days post infection. We fit one Gaussian process (GP) using a constant nugget term and another GP with a diagonal nugget matrix whose ith element is proportional to the sample variance estimated from two independent computer runs at parameter setting $x^{(i)}$. Diagnostic plots are used to compare the performance of these two methods.

2 RESULTS AND DISCUSSION

In Table 1 we describe the five parameters $x_1, \ldots, x_5$ varied during simulation of the computer model. After fitting a GP to the model, we examine diagnostic plots comparing cross-validated predictions from the GP predictors with observed values for these responses (Fig. 1). The early time point is better predicted by the GP than the late time point; nevertheless, 93% of true observations at the late time point fall in the 95% confidence intervals (data not shown).

We report the Functional Analysis of Variance (FANOVA) decomposition contributions for all main and two-way factor interaction effects for pathogen load at five and 18 days post infection (dpi) in Table 1. Main effects of all model parameters on pathogen load at five and 18 dpi are given in Fig. 2, and interaction effects for the most important parameter interactions at each time point are provided in Fig. 3. The results from *mlegp* show that the growth rate parameter x_1 is the most influential parameter for pathogen load at both five and 18 dpi (Table 1), but the effect of this parameter on pathogen load changes over time (Fig. 2). Two other important parameters are the carrying capacity of infected cells (x_2), which is positively related to pathogen load at five dpi, and the threshold level of pathogen that triggers the adaptive immune response (x_5), which is positively related to pathogen load at 18 dpi (Table 1, Fig. 2). However, both of these parameters have large interaction effects with x_1 at their respective time points (Table 1), indicating that the nature of their effects depends on the value of x_1. Based on the interaction effect plots (Fig. 3), x_2 has little effect on pathogen load at five dpi when x_1 is low, while x_5 has little effect on pathogen load at 18 dpi when x_1 is high.

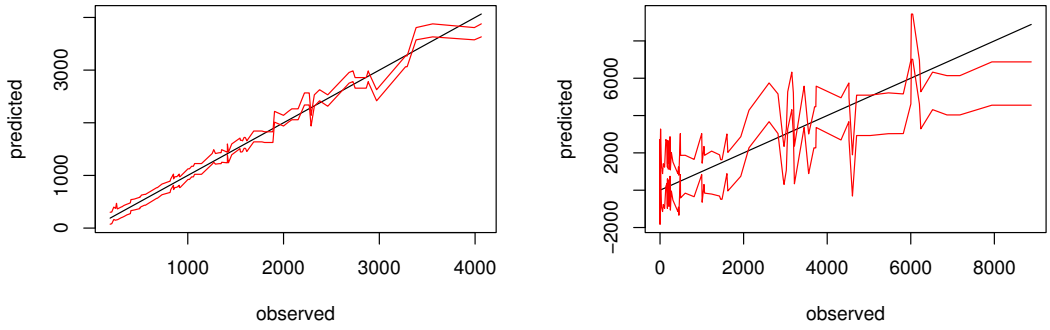
Lastly, we demonstrate the advantage of using an arbitrary diagonal nugget matrix, rather than a constant nugget term, in a GP model of pathogen load at five dpi, where output is heteroscedastic. Cross-validated diagnostic plots for both GPs are provided in Fig. 4, where we see the GP with diagonal nugget matrix is a more precise model. In particular, confidence bands of the constant nugget GP are overly

Table 1. FANOVA Decomposition of GP predictors for pathogen load at five and 18 dpi. The contribution of each effect is reported as a percentage of the total functional variance of each GP predictor.

description	parameter	5 dpi	18 dpi
pathogen growth rate	x_1	85.64	56.54
carrying capacity of infected host cell (macrophage)	x_2	6.89	2.79
Macrophage recruitment parameter	x_3	0.00	0.00
T cell recruitment rate	x_4	0.01	2.20
Threshold level of pathogen that triggers adaptive immune response	x_5	0.00	26.47
	x_1:x_2	7.40	0.39
	x_1:x_3	0.00	0.00
	x_1:x_4	0.00	0.08
	x_1:x_5	0.00	11.38
	x_2:x_3	0.00	0.00
	x_2:x_4	0.00	0.03
	x_2:x_5	0.00	0.06
	x_3:x_4	0.00	0.00
	x_3:x_5	0.00	0.00
	x_4:x_5	0.00	0.06
	All main and 2-way interactions	99.94	100.00

conservative for low values of the response, with 100% of the data points falling in the 95% confidence intervals when pathogen load is less than 1900, compared to 94.7% of the data points for the GP with a diagonal nugget matrix.

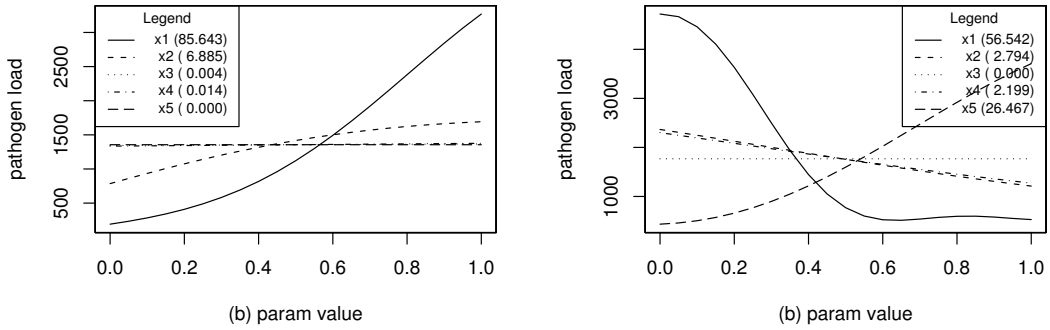
Fig. 1. GP diagnostics plots: solid black line, observed output vs. observed output; solid red lines, cross-validated prediction $\pm$ 1 standard deviation.



(a) GP emulator for pathogen load at 5 dpi

(b) GP emulator for pathogen load at 18 dpi

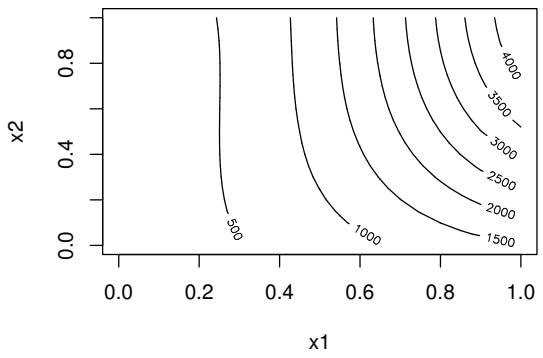
Fig. 2. Main effects of all parameters using the Guassian process predictors for pathogen load at five and 18 dpi.



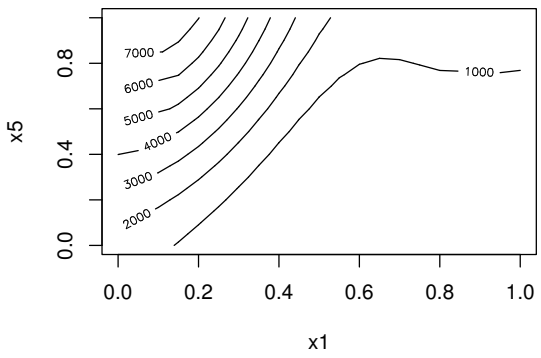
(a) 5 dpi

(b) 18 dpi

Fig. 3. Interaction effects of the most important interactions for pathogen load at five and 18 dpi.

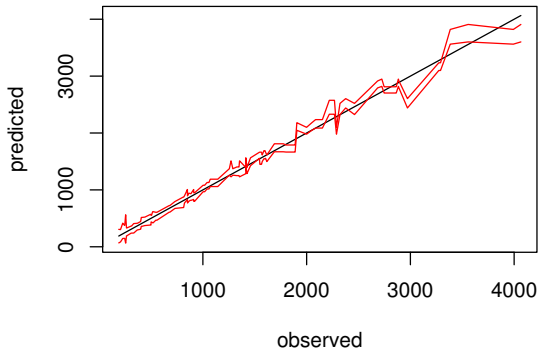


(a) x_1 and x_2 at 5 dpi

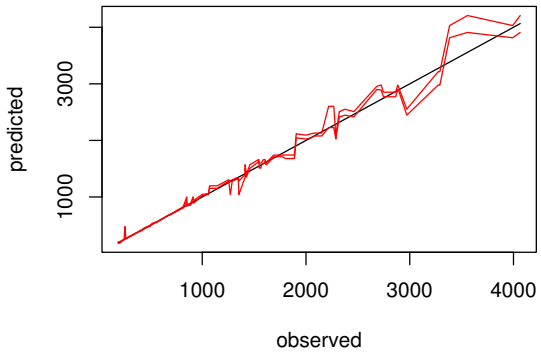


(b) x_1 and x_5 at 18 dpi

Fig. 4. GP diagnostic plots for pathogen load at 5 dpi, for (a) GP with constant nugget term and (b) GP with a diagonal nugget matrix whose i^{th} element is proportional to the sample variance of two indepent computer model runs evaluated at $x^{(i)}$; solid black lines, observed output vs. observed output; solid red lines, predicted mean output $\pm$ 2 standard deviations.



(a) GP with constant nugget term



(b) GP with user-defined diagonal nugget matrix