Emulating computer experiments of transport infrastructure slope stability using Gaussian processes and Bayesian inference

Aleksandra Svalova[1]* , Peter Helm[2] , Dennis Prangle[1], Mohamed Rouainia[2],
Stephanie Glendinning[2] and Darren J. Wilkinson[1]

[1]School of Mathematics, Statistics and Physics, Newcastle University, Newcastle upon Tyne NE1 7RU, United Kingdom
[2]School of Engineering, Newcastle University, Newcastle upon Tyne NE1 7RU, United Kingdom
*Corresponding author. E-mail: alex.svalova@newcastle.ac.uk

Abstract

We propose using fully Bayesian Gaussian process emulation (GPE) as a surrogate for expensive computer experiments of transport infrastructure cut slopes in high-plasticity clay soils that are associated with an increased risk of failure. Our deterioration experiments simulate the dissipation of excess pore water pressure and seasonal pore water pressure cycles to determine slope failure time. It is impractical to perform the number of computer simulations that would be sufficient to make slope stability predictions over a meaningful range of geometries and strength parameters. Therefore, a GPE is used as an interpolator over a set of optimally spaced simulator runs modeling the time to slope failure as a function of geometry, strength, and permeability. Bayesian inference and Markov chain Monte Carlo simulation are used to obtain posterior estimates of the GPE parameters. For the experiments that do not reach failure within model time of 184 years, the time to failure is stochastically imputed by the Bayesian model. The trained GPE has the potential to inform infrastructure slope design, management, and maintenance. The reduction in computational cost compared with the original simulator makes it a highly attractive tool which can be applied to the different spatio-temporal scales of transport networks.

Impact Statement

A significant part of the U.K. rail and highway infrastructure is built on high-plasticity clay soils which have an increased risk of failure. We have performed computer simulations of slope deterioration to determine the relationship between their time to failure and slope geometry, soil strength, and permeability. However, the simulations are very time-consuming. This can be addressed by using statistical surrogate models which are trained to be similar to computer simulator output. The surrogate can perform time to failure predictions for hundreds of slope geometry and strength combinations within hours, compared with years of computational time using the original simulator. The trained surrogate could be used to inform slope design, management, and maintenance on different spatio-temporal scales of transport networks.

This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

*The online version of this article has been updated since original publication. A notice detailing the change has also been published.

1. Introduction

The U.K. transport infrastructure is crucial to the efficient functioning of the economy and society. In particular, rail transportation is crucial to securing efficient and low-carbon logistical services to individuals and businesses; road transport will also continue to play a significant role in the future. A significant part of the U.K. rail infrastructure is nearing 200 years in age while being built on/within high-plasticity soils that are prone to weathering and deterioration. A section of the U.K. Great Western Main Line route connecting London to Bristol is one of the busiest and oldest (ca. 170 years post-construction) rail corridors in the United Kingdom (Skempton, 1996) that has undergone significant deterioration over its life cycle. Earthworks on the highways suffer from similar problems but to a lesser extent because of their relatively recent construction dates (1950s onwards) and shallower slope angles and heights. A lack of understanding of the long-term behavior, performance, and deterioration of infrastructure slopes can lead to uncontrolled deformations and thus reduced service performance (Briggs et al., 2019), negative impact on the economy (Power and Abbott, 2019), and, in the worst case, fatalities (Smith, 2020). Here, we aim to reduce the risks posed to infrastructure systems by better understanding the deterioration of geological assets.

Deterioration processes have been studied through computer experiments (Stirling et al., 2017; Wang et al., 2018) but these are computationally expensive and time-consuming (Maatouk and Bay, 2017). Instead, we use surrogate statistical models, which can be used to approximate computationally burden-some computer code/simulators, based on a small number of simulator runs at well-selected input locations (O'Hagan, 2006).

A simulator can be represented by a function f evaluated at input variables x to produce outputs y, $f(x) = y$. The term "emulator" (O'Hagan, 2006) was originally coined to refer to statistical approximations of a simulator which provide a probability distribution of its output. Gaussian processes (GPs) are one of the most frequently used machine learning (e.g., Seeger, 2004; Rasmussen and Nickisch, 2010; Su et al., 2016) and emulation techniques because of convenient properties such as uncertainty quantification (Maatouk and Bay, 2017), conditional normality, and a flexibility in defining the mean and variance structure of the process. GPs are amenable to statistical inference using Bayesian methods (O'Hagan, 2006). This allows expert opinion to be easily incorporated, which is especially useful when modeling sparse data or computer simulators with high computational cost. Gaussian process emulators (GPEs) have been applied to a variety of problems, including models of microbial communities (Oyebamiji et al., 2017), generic plant functional types (Kennedy et al., 2006), cardiac cells (Chang et al., 2015), and stochastic economic dispatch (Hu et al., 2020). Gaussian process emulation was also used to estimate the probability of soil slope failure in a non-Bayesian setting (Kang et al., 2015).

The following assumptions are made in order for $f(x)$ to be represented by a GPE. It is assumed that $f(x)$ is a smooth, continuous function of its input variables (O'Hagan, 2006) and that outputs of $f(x)$ for any sequence of inputs can be modeled as a multivariate normal distribution with mean and covariance functions depending on the inputs. Although assuming normality is restrictive to some extent, this is offset by the great flexibility in the choice of mean and covariance functions (Gramacy, 2020).

According to O'Hagan (2006), an emulator should satisfy two criteria. First, the emulator has no uncertainty in the output $y_i = f(x_i), i = 1, 2, \ldots, n$ evaluated at a simulator input x_i. Second, elsewhere the emulator gives a distribution of plausible interpolated values. The first condition is often relaxed in practice (Gramacy, 2020) by adding τ to the diagonal of the covariance matrix, where τ is a small number known as the nugget. The latter can account for an uncertainty/error in the simulator output (Cressie, 1993) or improve numerical stability when inverting the covariance matrix of the GPE (Neal, 1997) which may appear numerically ill-conditioned (Gramacy, 2020). Further discussion on the use of the nugget term and its effect can be seen in, for example, Andrianakis and Challenor (2012).

In this paper, we use a GPE to approximate computer experiments with 76 geotechnical simulator runs of infrastructure slope deterioration that would otherwise be infeasibly time-consuming to evaluate. The experimental results presented here are focused on the slopes in high to very high-plasticity soils, such as

slopes in the London Clay (Hight et al., 2003). While over 40 model inputs can be varied (FLAC, 2016; Rouainia et al., 2020; Postill et al., 2021), here we only vary slope geometry (angle and height), soil shear strength (cohesion and friction angle), and permeability. These variables were selected over the remaining ones as previous work has demonstrated their importance in assessing the stability of geotechnical infrastructure (Potts et al., 1997; Ellis and O'Brien, 2007; Rouainia et al., 2020). This allows us to limit the experimental time to weeks while obtaining an informative training data set. To ensure an optimal coverage of the parameter space, a Latin hypercube experimental design is used. The simulator runs were stopped after 184 years of model time, corresponding to the average rail cutting slope age in the London–Bristol corridor (Skempton, 1996). Slopes that have not failed within this time have their time to failure (TTF) censored. In such cases, our GPE imputes the uncensored TTF.

The structure of this article is as follows. Section 2 describes the computer experiments and geotechnical model (GM) that simulate deterioration processes in cut slopes. Section 3 outlines the background related to the emulator, Bayesian inference, and emulating the censored computer output. Section 4 presents the results of Markov chain Monte Carlo (MCMC) inference of the GPE posterior distribution and related sensitivity analysis. Section 5 discusses the TTF predictions for a range of cutting slope geometry and soil strength scenarios as well as for London–Bristol corridor railway and M4 motorway cuttings. Finally, Section 6 concludes the study and outlines future directions for research.

2. Application: Computer Experiments of Cut Slope Deterioration

2.1. Introduction

The Great Western Main Line (GWML) and the M4 motorway corridors are constructed through a number of geological formations, a significant proportion of which are comprised of over-consolidated and high-plasticity clays (Charlesworth, 1984; Skempton, 1996). These materials are prone to seasonal shrink–swell movements as they wet and dry, along with downslope ratcheting (Take and Bolton, 2011), both of which contribute to strain softening (Rouainia et al., 2020; Postill et al., 2021) reducing their strength over time. With the GWML and M4 earthworks that are of ca. 175 and 65 years of age, respectively, the soils from which they are formed will have been subject to a varying number of these seasonal cycles and undergone an indeterminate magnitude of deterioration from their initial state. Here, we focus on replicating these deterioration processes in high-plasticity, over-consolidated clays, particularly London Clay (Hight et al., 2003; Dixon et al., 2019; Rouainia et al., 2020).

The time-dependent stability of cut slopes has been the subject of previous numerical modeling work (Potts et al., 1997; Tsiampousi et al., 2017; Summersgill et al., 2018). Capturing the influence of weather and seasonal cycles, as well as the strain-softening behavior is crucial to the accurate modeling of over-consolidated high-plasticity clays.

2.2. The GM

This subsection is a summary of the key mechanisms and properties that influence the slope deterioration behavior in the GM used in our simulator. For more detail, see Rouainia et al. (2020) and Postill et al. (2021).

The GM is implemented using the Fast Lagrangian Analysis of Continua with Two-Phase Flow software (FLAC-TP; FLAC, 2016), making use of a strain-softening Mohr–Coulomb constitutive model (FLAC, 2016). The parameters adopted in the modeling have been informed by the strain-softening behavior from previous modeling studies in over-consolidated clays (Potts et al., 1997; Ellis and O'Brien, 2007; Summersgill et al., 2018) as well as laboratory and field data (Bromhead and Dixon, 1986). The impact of weather and climate on cut slopes is modeled through a coupled fluid-mechanical approach (Rouainia et al., 2020) utilizing a nonlocal strain-softening model (Summersgill et al., 2018; Postill et al., 2021). This allows a detailed assessment of weather-driven deterioration.

The soil is modeled as a porous medium with variable saturation and depth-dependent saturated permeability (Postill et al., 2021). A two-phase flow approach is adopted where the pore phases are

Table 1. *Material and geometry input variables used in the computer experiments.*

Property	h_s (m)	$\cot\theta_s$	ϕ'_p (°)	c'_p (kPa)	k_1^w (m/s)
Range	$[4, 20]$	$[0.5, 7.5]$	$[18.5, 25]$	$[3, 10]$	$\left[1.45 \times 10^{-9}, 2.5 \times 10^{-8}\right]$

assumed to be air and water and are treated as immiscible fluids with differing density and viscosity. Fluid flow is Darcian, whereby water and air flow velocity is a function of the respective pore fluid pressures and the relative permeability of the soil. The latter, in turn, is a function of the degree of water and air saturation, derived using the van Genuchten–Maulem relation (van Genuchten, 1980).

The constitutive model is nonlinear elastic where the bulk and shear moduli behave as a function of the mean effective stress. These parameters were adopted from prior modeling studies (Potts et al., 1997; Jurečič et al., 2013) and the stiffness behavior is validated in Postill et al. (2021). The shear strength of high-plasticity clays undergoes a reduction as shear strain increases during yielding. This post-failure strength decrease is replicated in the model based on a reduction in the Mohr–Coulomb shear strength parameters (effective peak cohesion c' and friction angle ϕ') as a function of increasing plastic shear strains. More details about the GM can be found in Rouainia et al. (2020) and Postill et al. (2021).

2.2.1. Input variables

The five input variables in this study are the slope angle cotangent $\cot\theta_s$ and height h_s (converted to a single geometry variable—see Supplementary Figure S1) derived from light detection and ranging (LiDAR) survey data provided by project stakeholders (Mott MacDonald and Network Rail), the peak shear strength parameters (c'_p and ϕ'_p) before the material has undergone any strength reduction derived from previous laboratory data and modeling studies (Apted, 1977; Ellis and O'Brien, 2007), and the reference coefficient of permeability of the soil at 1 m depth with respect to water (k_1^w) derived from field measurements (Dixon et al., 2019). For more information on the adopted GM and the values adopted for the remaining input parameters, see FLAC (2016), Rouainia et al. (2020), and Postill et al. (2021).

Table 1 summarizes ranges of the parameters used in computer simulations which were selected based on previous studies (Rouainia et al., 2020; Postill et al., 2021) and expert opinion from partners in the ACHILLES project. During emulation, permeability was scaled by 10^8 to bring the explanatory variables to a common scale. Supplementary Figure S1 illustrates that we have chosen geometries that are concentrated within the 4–12 m height and 1–4 angle cotangent ranges. This corresponds to the most common geometries in the London–Bristol corridor (Network Rail, 2017).

2.2.2. Model run time

The duration of the geotechnical modeling runs varied from approximately 75 min to 10 days. The latter depends on a number of factors, including the total number of elements in the model, model geometry, permeability, and the adopted strength parameters. For example, a model with the longest run time would correspond to a slope that is shallow enough to be stable for the full potential duration of the boundary time series and which includes a high near surface permeability yielding small flow time step. Conversely, the highest/steepest slope models tended to fail very quickly after modeled excavation.

3. Background

Here we review the methods which will be used for our GPE. Section 3.1 reviews GPEs, based on Bastos and O'Hagan (2009), Section 3.2 introduces the censored emulator model, Section 3.3 describes Bayesian inference and prediction for GPEs, Section 3.4 outlines our sensitivity analysis methods and Section 3.5 reviews Latin hypercube experimental design that we adopted for this study.

3.1. Gaussian process emulator

A GP can be understood as an infinite-dimensional multivariate normal distribution for functions (Bastos and O'Hagan, 2009). A GPE $\eta(\cdot)$ is a GP conditioned on experimental simulations that can be used to predict simulator output under other conditions, and quantify its uncertainty. Assume $\eta(\cdot)$ takes a generic input $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)$ where $x_i \in \chi_i \subset \mathbb{R}$. A scalar-valued GP is fully defined by its mean and covariance functions m and V, $\eta(\cdot)|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}, \tau \sim \mathrm{GP}(m(\cdot), V(\cdot, \cdot))$ (Bastos and O'Hagan, 2009).

Given no prior knowledge about the structure of $m : \mathbb{R}^p \to \mathbb{R}$, it is often assumed to be a linear transformation of the input variables, that is, $m(\boldsymbol{x}) = h(\boldsymbol{x})^T \boldsymbol{\beta}$, where $h(\cdot) : \mathbb{R}^p \to \mathbb{R}^q$ is a function mapping x to a vector of linear regressors and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_q)$. To include a constant term in $m(\boldsymbol{x})$, the first element of $h(\boldsymbol{x})$ can be set to 1. We assume the covariance function $V : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ has the form $V(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2[C(\boldsymbol{x}, \boldsymbol{x}', \theta) + \tau \mathbb{I}(\boldsymbol{x}, \boldsymbol{x}')]$, where σ is a scale parameter, $C(\boldsymbol{x}, \boldsymbol{x}', \theta)$ is a correlation function, τ is the nugget (Andrianakis and Challenor, 2012), and $\mathbb{I}(\boldsymbol{x}, \boldsymbol{x}')$ is an indicator function for the event $\boldsymbol{x} = \boldsymbol{x}'$.

In principle, C can be any function that is smooth, continuous, and positive semidefinite. We consider two popular correlation functions. The first is the Gaussian function $C_G(\boldsymbol{x}, \boldsymbol{x}', \boldsymbol{\theta}) = \exp\left\{ -\sum_{i=1}^{p} \frac{(x_i - x_i')^2}{\theta_i^2} \right\}$, where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$ is a vector of correlation lengths. The correlation lengths represent the "radius of association" and larger $\boldsymbol{\theta}$ values typically lead to smoother/flatter GPs. Second, we consider the Matérn correlation family (Rasmussen and Williams, 2006):

$$C_M(\boldsymbol{x}, \boldsymbol{x}', \boldsymbol{\theta}, v) = \prod_{i=1}^{p} \frac{1}{\Gamma(v) 2^{v-1}} \left(\frac{\sqrt{2v}|x_i - x_i'|}{\theta_i} \right)^v K_v \left(\frac{\sqrt{2v}|x_i - x_i'|}{\theta_i} \right), \tag{1}$$

where K_v is the modified Bessel function of second kind of order v.

As GPs are closed under conditioning, it is possible to derive an analytical expression for the GPE conditioned on a set of simulator runs. Assume a (finite) collection of n observed experimental outputs $\boldsymbol{y} = (\eta(x_1), \eta(x_2), \ldots, \eta(x_n))$ performed on the inputs $x_1, x_2, \ldots, x_n$, which comprise the training data. We assume no repeated inputs, so $\boldsymbol{x}_i = \boldsymbol{x}_j$ if and only if $i = j$. The n-vector $\boldsymbol{y}$ follows a multivariate normal distribution, $\boldsymbol{y}|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}, \tau \sim \mathrm{N}(H_x \boldsymbol{\beta}, \sigma^2 \Sigma_x)$, where H_x is a matrix of regressors whose ith row is $h(\boldsymbol{x}_i)$ and $\Sigma_{x(i,j)} = C(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{\theta}) + \tau \mathbb{I}(i, j)$. Using standard rules for conditioning on a subset of observations (Gramacy, 2020),

$$\eta(\cdot)|\boldsymbol{y}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}, \tau \sim \mathrm{GP}(m^*(\cdot), V^*(\cdot, \cdot)), \text{ where}$$
$$m^*(\boldsymbol{x}) = h(\boldsymbol{x})^T \boldsymbol{\beta} + t(\boldsymbol{x})^T \Sigma_x^{-1}(\boldsymbol{y} - H_x \boldsymbol{\beta}), \ V^*(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2 \left(C(\boldsymbol{x}, \boldsymbol{x}', \boldsymbol{\theta}) - t(\boldsymbol{x})^T \Sigma_x^{-1} t(\boldsymbol{x}') \right), \tag{2}$$

where $t(\boldsymbol{x}) = (C(\boldsymbol{x}, \boldsymbol{x}_1, \boldsymbol{\theta}), C(\boldsymbol{x}, \boldsymbol{x}_2, \boldsymbol{\theta}), \ldots, C(\boldsymbol{x}, \boldsymbol{x}_n, \boldsymbol{\theta}))^T$ is a column vector of correlations between the (generic) emulator input $\boldsymbol{x}$ and training inputs $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$.

3.2. Censored computer output

Censoring is a common phenomenon where only partial information is available on some observations, that is, it is only known that they are outside a specified range. A GPE can be conditioned on both censored and uncensored observations. Previous applications of censored GPs include hydrologic system response (Wani et al., 2017) and pyroclastic volcano flow (Kyzyurova, 2017).

Censored values can be estimated together with $\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}$ and τ using data augmentation during MCMC sampling. For our application, suppose n experiments at the set of inputs $\boldsymbol{x}_o = (\boldsymbol{x}_{o,1}, \boldsymbol{x}_{o,2}, \ldots, \boldsymbol{x}_{o,n})$ produced uncensored observations $\boldsymbol{y}_o$. Also, suppose n_c experiments evaluated at the set of inputs $\boldsymbol{x}_c = (\boldsymbol{x}_{c,1}, \ldots, \boldsymbol{x}_{c,n_c})$ produced censored observations. Here, this means that the experiment reached the end of model time of 184 years without reaching the ultimate failure state. Let $\boldsymbol{y}_c$ be the corresponding vector of (hypothetical) uncensored times to failure. We then aim to infer $\boldsymbol{y}_c$ together with the parameters $\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}$ and τ. Define a new process $\eta_c(\cdot)$ where (Kyzyurova, 2017):

$$\eta_c(\mathbf{x}) = \begin{cases} \eta(\mathbf{x}), & \text{if } \eta(\boldsymbol{x}) < c, \\ c & \text{otherwise} \end{cases} \tag{3}$$

where $\eta(\cdot)$ has been defined earlier. The distribution of $\eta_c(\mathbf{x})$ at design points $\mathbf{x} = (\boldsymbol{x}_c, \boldsymbol{x}_o)$ is then (Kyzyurova, 2017):

$$\eta(\boldsymbol{x}_o)|\boldsymbol{\beta},\sigma^2,\boldsymbol{\theta},\tau \sim \mathrm{N}\big(H_o\boldsymbol{\beta},\sigma^2\Sigma_o\big),\ \eta_c(\boldsymbol{x}_c)|\eta(\boldsymbol{x}_o),\boldsymbol{\beta},\sigma^2,\boldsymbol{\theta},\tau \sim \mathrm{TN}_{(c,\infty)}(m_c, V_c),$$

$$m_c = H_c\boldsymbol{\beta} + \Sigma_{c,o}\Sigma_o^{-1}(\eta(\boldsymbol{x_o}) - H_o\boldsymbol{\beta}) \text{ and } V_c = \sigma^2\big(\Sigma_c - \Sigma_{c,o}\Sigma_o^{-1}\Sigma_{o,c}\big), \tag{4}$$

where $\mathrm{TN}_{(c,\infty)}$ denotes a normal distribution truncated below c. In (4), H_o and Σ_o are equivalent to H_x and Σ_x as defined as in Section 3.1, here using the inputs $\boldsymbol{x}_o$. Similarly, H_c is a matrix of regressors associated with $\boldsymbol{x}_c$, that is, its ith row is $h(\boldsymbol{x}_{c,i})$, and $\Sigma_{c(i,j)} = C\big(\boldsymbol{x}_{c,i},\boldsymbol{x}_{c,j},\boldsymbol{\theta}\big) + \tau\mathbb{I}(i,j)$. Also, $\Sigma_{c,x(i,j)} = C\big(\boldsymbol{x}_{c,i},\boldsymbol{x}_{o,j},\boldsymbol{\theta}\big)$ and $\Sigma_{o,c} = \Sigma_{c,o}^T$. To obtain truncated multivariate normal samples from Equation (4) (used later in a Gibbs sampler), we used the "TruncatedNormal" package in R (Botev and Belzile, 2020). The package uses a minimax tilting method for exact i.i.d. data simulation from the truncated multivariate normal distribution. More details can be found in Botev (2017).

3.3. Bayesian inference and prediction

Inference for the GPE $\eta(\cdot)$ and its parameters $\boldsymbol{\beta},\sigma^2,\boldsymbol{\theta}$ and τ above could be obtained by maximum likelihood using numerical optimization. While this easily produces point estimates, it is hard to take into account all sources of uncertainty (O'Hagan, 2006). Instead, we follow a Bayesian approach which can naturally incorporate parameter uncertainty, using MCMC (Brooks et al., 2010) for inference. We used a Metropolis-within-Gibbs sampler whereby $\boldsymbol{\beta},\sigma^2,\boldsymbol{\theta}$ and τ were updated using Metropolis–Hastings steps and $\boldsymbol{y}_c$ was updated using a Gibbs step using Equation (4). Section 4.3 has more details regarding our application.

Parameters inferred using MCMC can then be used for predicting the simulator output over a grid of inputs. Using a Bayesian predictive distribution (BPD) accounts for parameter uncertainty. One can sample from the marginal BPD for input $\boldsymbol{v} = (v_1, v_2, \ldots, v_p)$ as follows. First, randomly select an iteration of the MCMC output and use the corresponding parameters. Second, use Equation (2) to sample from the GPE for input $\boldsymbol{v}$ given the selected parameters (taking $\boldsymbol{x} = (\boldsymbol{x}_o, \boldsymbol{x}_c)$ and $\boldsymbol{y} = (\boldsymbol{y}_o, \boldsymbol{y}_c)$ where $\boldsymbol{y}_c$ is part of the MCMC output). Repeated sampling produces a Monte Carlo approximation to the marginal BPD at this location. This was used to produce Figure 3 and Supplementary Figures S2, S3, and S7.

3.4. Sensitivity analysis

It is often important to quantify the effect of different input variables on the response of a GPE. This can be obtained through sensitivity analysis (SA) (O'Hagan, 2006; Farah and Kottas, 2014). Sensitivity indices explain the proportion of variation in the mean response of the emulator because of an individual or a combination of input variable(s) (Gramacy, 2020).

Assume that U represents the input density. For independent input variables x_i, $U(x) = \prod_{i=1}^{p} u_i(x_i)$, where u_i are the x_i marginal densities. The simplest form of sensitivity indices are the main effects (ME) (Gramacy, 2020) of an explanatory variable x_i, $i = 1, 2, \ldots, p$:

$$me(x_i) \equiv E_{U_{-i}}\{\eta|x_i\} = \int \int_{-\chi_{-i}} \eta p(\eta|x_1, x_2, \ldots, x_p) \cdot u_{-i}(x_{-i}) dx_{-i} d\eta. \tag{5}$$

Here, $u_{-j} = \prod_{i \neq j} u_i(x_i)$, and χ_{-i} and x_{-i} are defined similarly (Gramacy, 2020). The formalism (5) deterministically varies x_i while averaging over the input densities U_{-i}, assuming that all inputs are uncorrelated. In other words, $me(x_i)$ is the expected value of the response given x_i while averaging over x_{-i}. The distributions of the main effects are calculated by calculating the expectation in Equation (5) at every MCMC sample of the $\eta(\cdot)$ posterior distributions.

In our application, we also perform a fully Bayesian SA using first-order and total sensitivity indices following the method in Farah and Kottas (2014) and Gramacy (2020): the derivation of the following formalisms can be found therein. The first-order sensitivity index $S_{1,i}, i = 1, 2, \ldots, p$ evaluates the fractional contribution of the x_i input variable to the variance of the output (Farah and Kottas, 2014):

$$S_{1,i} = \frac{E\left[E^2[\eta|x_i]\right] - E^2[\eta]}{Var(\eta)}. \tag{6}$$

In other words, $S_{1,i}$ is the response sensitivity to main variable effects (Gramacy, 2020). The total sensitivity index (Homma and Saltelli, 1996; Farah and Kottas, 2014) $S_{T,i}$ is a measure of the entire influence attributable to a given variable (Gramacy, 2020):

$$S_{T,i} = 1 - \frac{E\left[E^2[\eta|x_{-i}] - E^2[\eta]\right]}{Var(\eta)}. \tag{7}$$

A large difference between the distributions of $S_{1,i}$ and $S_{T,i}$ would indicate that the interactions between the x_i and the remaining input variables are important to explaining the output variation (Farah and Kottas, 2014). The posterior distributions of the sensitivity indices can be sampled by computing the expectations in Equations (6) and (7) at every MCMC sample of the emulator parameters (Farah and Kottas, 2014).

3.5. Experimental design

The reliability of the emulator very strongly depends on the experimental design (Busby, 2009), that is, the choice of input variables. Space-filling designs aim to spread out the input variables and produce a diversity of data once responses have been observed (Gramacy, 2020). We focus on a particular space-filling approach: Latin hypercube designs (LHD). Here each of the p input variables are divided into N equally sized intervals and one point is selected randomly per interval in each dimension (Santner, 2018; Gramacy, 2020).

Although LHDs are space-filling in each of the input coordinates, however, they are not necessarily space filling in the p-dimensional hypercube space (Dette and Pepelyshev, 2021). An alternative approach is to search for a design optimizing the *maximin* criterion: maximizing the minimum distance between pairs of points (Santner, 2018; Gramacy, 2020). We used a hybrid *maximin LHD* approach in which five LHDs are generated, and the optimal maximin choice is selected. This was implemented using the package "pyDOE" (Baudin et al., 2012) in Python.

4. Application Methods

4.1. Model, priors, and design

We deployed a GP emulator for GM experiments of rail cut slope stability. We emulated the output variable of time to failure (TTF), that is, the time to reaching the slope's ultimate failure state. As input variables, we used slope height, angle cotangent, (effective) peak cohesion, (effective) peak friction angle, and permeability (hydraulic conductivity). In our earlier notation, these variables are, respectively, $h_s, \cot\theta_s, c_p', \phi_p', k_1^w$. However, for ease of reference, we refer to these as $x_1, x_2, \ldots, x_5$ below.

We used LHD to create 76 vectors of input variables to use as the training suite of the GM experiments. This was based on a four-dimensional design, with the first dimension corresponding to a combination of height and angle cotangent which we refer to as *geometry*; see Supplementary Figure S1 for details. Based on exploratory analysis, detailed in Section 4.2, we used the full linear form of the regressor function $h(x)$, including an intercept, $h(x) = (1, x_1, x_2, \ldots, x_5)$. Thus, our mean function was $m(x) = \beta_0 + \sum_{i=1}^{5} \beta_i x_i$.

The prior distributions for β, θ and σ^2 were chosen to reflect beliefs of our geotechnical and statistical experts about $\eta(\cdot)$:

$$\beta_0 \sim \mathrm{N}(0, 10^2), \ \beta_i \sim \mathrm{N}(0, 4^2), \ \sigma^2 \sim \mathrm{IGa}(10, 100),$$
$$\theta_i \sim \mathrm{Exp}(0.2), \ \tau \sim \mathrm{IGa}(3, 1), \ i = 1, 2, \ldots, 5.$$

Also note that it is important to select informative priors for the covariance function parameters, otherwise identifiability issues are common (see, e.g., Zhang, 2004).

4.2. Modeling and tuning choices

4.2.1. Alternative regressor choices

We compared our full linear choice of $h(\boldsymbol{x})$ to other choices, in particular, $h(\boldsymbol{x}) = 0$ (zero mean) and $h(\boldsymbol{x}) = 1$ (constant mean). Note that in the geostatistical literature, the full linear and constant mean forms of $h(\boldsymbol{x})$ correspond to universal and simple forms of kriging (Oliver and Webster, 2015). Supplementary Figure S2 shows results comparing choices of $h(\boldsymbol{x})$. It illustrates that the full linear form of $h(\boldsymbol{x})$ appears most appropriate, as the resulting TTF contours are near-linear and are consistent with the literature (Ellis and O'Brien, 2007; Gao et al., 2015). Using only a constant mean term for $h(\boldsymbol{x})$ led to nonmonotonic dependence of TTF on slope angle, which was deemed to be unrealistic. The full linear form also allows both the mean and the covariance functions to contribute to explaining variation in the relationship between the input and response, as opposed to the covariance function only when using a constant or zero mean.

4.2.2. Transformation of response and choice of kernel

Various transformations of the TTF were considered to avoid predicting negative values and to minimize prediction errors. The discussion and illustration of their predictive properties for a set of validation runs are shown in section "Output transformation validation" and Supplementary Figure S3. Supplementary Figure S3 illustrates the goodness of predictions for the different transformations by plotting them against the observed times to failure. Square root transformation appears to give the highest agreement with the observed TTFs, illustrated by the lowest mean square error. Additionally, the $C_M(\boldsymbol{x}, \boldsymbol{x}', \boldsymbol{\theta}, 5/2)$ (abbreviated as $C_{M,5/2}$) correlation function gave the optimal R^2 and MSE. The cube root transformation also appears plausible. However, it over-predicts the TTF in its upper distribution tail. Therefore, the square root transformation of TTF and $C_{M,5/2}$ was used for this study.

4.3. MCMC details

The regressor coefficients β_0 and β_4 were updated jointly using a multivariate normal proposal distribution centered at the current chain value. The variance matrix of the proposal distribution was set to approximate the posterior correlation that was observed in a pilot run. Component-wise updates were used for $\beta_1, \beta_2, \beta_3$, and β_5, using an adaptive proposal variance following the procedure by Xiang and Neal (2014). The parameters σ^2, τ and $\theta_i, i = 1, 2, \ldots, 5$ were updated using a log-normal proposal distribution centered at the natural logarithm of the current chain value. The proposal variances of σ^2, θ_i and τ were also tuned using the above method. The censored observations $\eta(\boldsymbol{x}_c)$ were updated using a Gibbs step following Equation (4).

The chain burn-in and variance adaptive period was 10^5 iterations which were discarded and the scheme was then run for a further 2×10^5 iterations. Supplementary Figure S4 illustrates the trace plots of the MCMC draws. Good mixing can be observed for all parameters, especially the censored observations. Supplementary Figure S5 illustrates effective sample sizes (ESSs) for the three correlation functions, which is a diagnostic for posterior autocorrelation (Brooks et al., 2010). All of the ESSs are above 3,000 with censored observations having the highest values.

5. Application Results

5.1. Parameter inference

Figure 1 illustrates the posterior densities of the GPE parameters, all of which appear unimodal. Posterior density plots for all of the censored values can be found in Supplementary Figure S6. For most parameters, the posterior variance is significantly reduced compared with the prior distribution. This is especially the
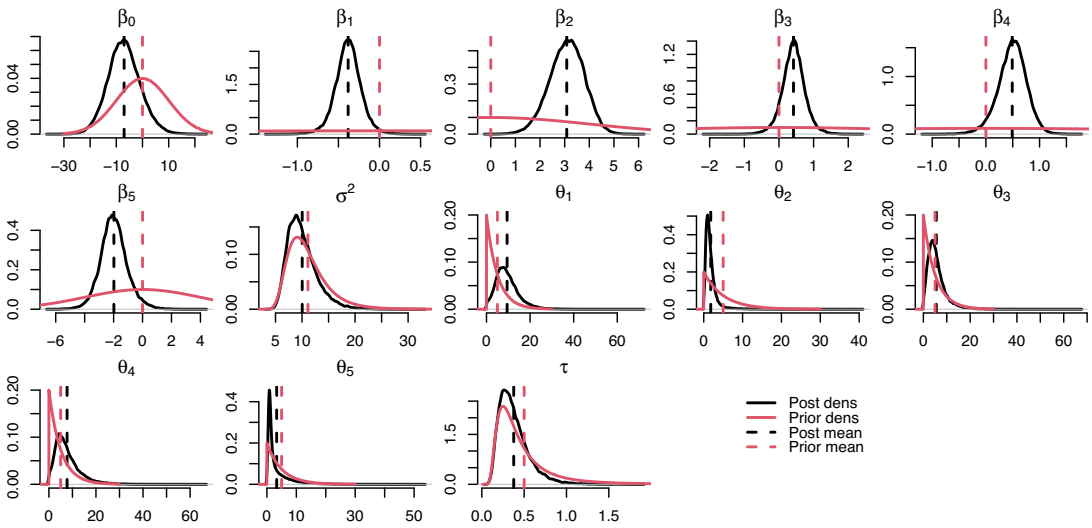
Figure 1. *Density plots of the posterior draws of the GPE model. Density plots of the y_c posterior distributions are in Supplementary Figure S6.*
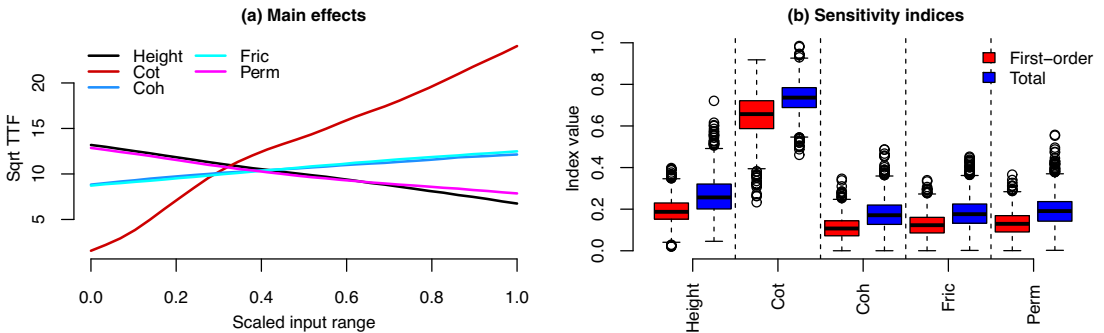
Figure 2. *Sensitivity analysis of the GPE. Plot (a) illustrates the main effects sensitivity analysis corresponding to Equation 5, and plot (b) illustrates first-order and total sensitivity indices corresponding to Equations 6 and 7. Slope angle cotangent, cohesion, friction angle, and permeability are abbreviated "Cot", "Coh", "Fric", and "Perm". In plot (a), square root of time to failure is abbreviated "Sqrt TTF".*

case for most $\boldsymbol{\beta}$ coefficients and the censored values. Most of the posterior densities of the censored observations (Supplementary Figure S6) also show a decrease in variance compared with that of their prior distributions. The larger posterior variances can be observed for the observations with larger posterior means, for example, $y_{c,1}$ and $y_{c,3}$, as those estimates have the largest distance to the training data. Conversely, the uncertainty around the posterior estimations of, for example, $y_{c,53}$ and $y_{c,68}$ is very small. There is relatively little info about $\boldsymbol{\theta}$ in the data; however, their posterior distributions are relatively close to their priors compared with those of $\boldsymbol{\beta}$.

Supplementary Figure S7 illustrates the posterior marginal distributions of the TTF versus each of the input variables. The majority of the data is captured in the 95% confidence interval bands. While all of the relations have a trend, slope angle cotangent appears to have the strongest relation with the TTF.

5.2. Sensitivity analysis

Figure 2a illustrates a main effects sensitivity analysis of the input variables, which are equivalent to the mean trends in Supplementary Figure S7. All of the main effects are monotonic, and the most influential

variable appears to be slope angle cotangent, in this case because of the steepest gradient. Cohesion and friction angle have very similar main effects as they soften at the same rate as a function of plastic strain (e.g., Labuz and Zang, 2012).

Figure 2b illustrates the first-order and total sensitivity index distributions. Again, angle cotangent has the greatest contribution to the variability in TTF, followed by height. Cohesion, friction angle, and permeability explain a similar proportion of the TTF variability. All of the input variables have similar distributions of the first-order and total sensitivity indices, indicating an absence of strong interaction effects (Farah and Kottas, 2014).

### 5.3. TTF posterior predictive maps

Plots of TTF versus slope geometry can be very informative in illustrating deterioration given fixed soil strength and permeability. Plots of TTF versus slope geometry allow infrastructure owners and designers to estimate likely design life of geotechnical infrastructure where previously this was not possible. Figure 3 illustrates the distribution of TTF for slope height and angle in four soil strength scenarios (cohesion, friction angle, and permeability values are shown in the legend). All of the negative predicted values were set to zero. Supplementary Figure S8 illustrates equivalent maps of posterior predictive variance. Low- and high-strength soils reflect the range of shear strengths for over-consolidated high-plasticity clays based on laboratory and filed data for London Clay strength and permeability (Bromhead and Dixon, 1986; Potts et al., 1997).

The change in TTF is approximately linear with respect to geometry, and appears most sensitive to changes in the slope angle. This is consistent with our sensitivity analysis and failure profiles reported earlier in the literature (Ellis and O'Brien, 2007; Gao et al., 2015). For the low-strength soil example, a significant proportion of slope geometries fail in less than 10 years. Fortunately, the scenario is unlikely be representative of average in situ peak strength conditions for the whole soil mass (e.g., Kovacevic et al., 2007), especially where vegetation roots can influence soil strength at the near surface (Woodman et al., 2020). For London Clay examples, the TTF contours are consistent with the literature (Ellis and O'Brien, 2007; Gao et al., 2015). The plots in the lower panel of Figure 3 illustrating TTF for the cuttings in the London–Bristol corridor assumes strength and permeability in agreement with published results (Hight et al., 2003; Dixon et al., 2019). Note that the training data do not cover the upper ranges of slope height and angle (i.e., upper triangle of the plots), thus the TTF predictions made in that region are an extrapolation.

Supplementary Figure S8 illustrates that the posterior variance is minimized for London Clay-type soils which are away from the boundaries of our Latin Hypercube design. Therefore, extrapolating further away from the experimental data leads to a rapid increase in posterior variance which is undesirable. In conditions of low data availability, it would be advisable to avoid construction of slopes which are too steep or too shallow. Also, it would be advisable to perform another numerical experiment and emulation for materials which are toward or outside the boundary of our design, for example, medium-plasticity clays.

5.3.1. Comparison to previous work

Our results in Figure 3 are consistent with those of the Global Stability and Resilience Appraisal (GSRA) performed by Mott Macdonald and Network Rail (Network Rail, 2017; Abbott, 2018). They derived contours of vulnerability to deep-seated rotational failure for soils with different plasticities. The bottom plots illustrate the cutting slope geometry data in the London–Bristol corridor that are built in London Clay-type soil (data provided by Mott MacDonald and Network Rail). The bottom-right plot in Figure 3 overlays the vulnerability contours for cuttings in cohesive, high-plasticity soils over the TTF predictions for cuttings in the London–Bristol corridor. The high GSRA failure potential (FP)corresponds to TTFs of 50 years or less. The low FP corresponds to TTFs of 150 years and above. At low slope heights (< 8 m), the vulnerability contours are determined by a combination of height and angle. For other heights, they are almost entirely determined by the angle. In our predictions, the relation between TTF and slope geometry
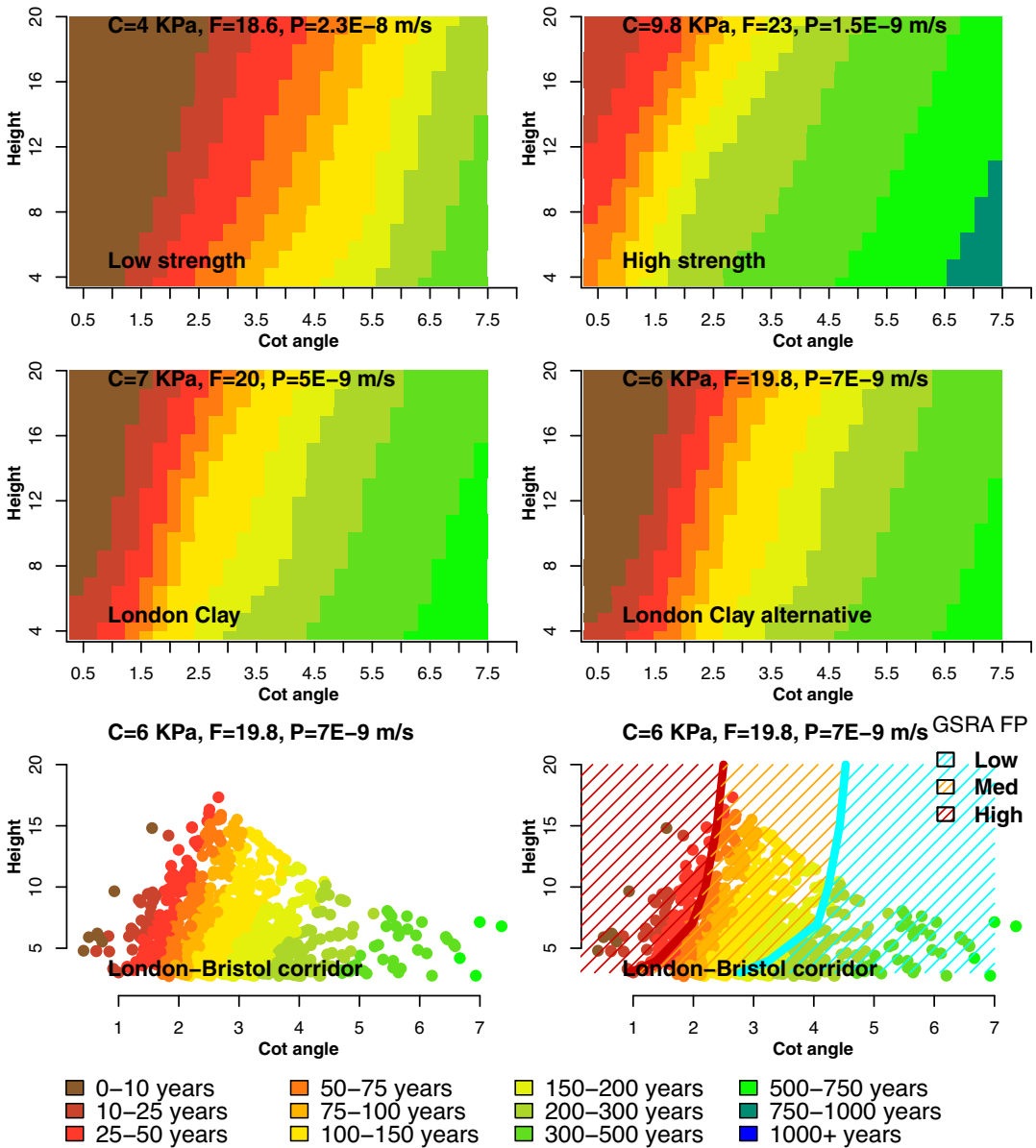
Figure 3. *Expected time to failure for the GPE model. Cohesion kPa (C), friction angle (F), and permeability m/s (P) are fixed for every scenario and shown in the top left. The bottom-right plot illustrates the GSRA (Network Rail, 2017) failure potential (GSRA FP) contours. Posterior predictive variance maps are shown in Supplementary Figure S8.*

is more linear, although the gradient is steeper in the slope angle direction, which is consistent with our sensitivity analysis.

6. Conclusion

We have used GP emulation to produce a surrogate for the cutting slope stability GM. The high computational expense of the latter renders it impractical to evaluate for the number of experiments that

is required for understanding the input–output relationships. Latin hypercube sampling was used to design an optimally space-filling experiment to train the emulator. Bayesian inference and MCMC simulation were used to estimate the GPE parameters. For a number of experiments, the ultimate failure state was not achieved during the 184 years of experimental time. For such models, the TTF was unobserved and was later also estimated using MCMC. The trained GPE was then used to produce TTF maps illustrating the relationship between slope geometry and FP. The computational expense involved in producing such inference is on the order of hours, which makes the GPE a highly attractive tool that can be used in real-time for rapid characterization of slope stability and deterioration. For instance, every plot in Figure 3 has 900 points and takes 3–4 hr to produce once the emulator has been trained, and its computation could be easily parallelized if necessary. Conversely, 900 runs of the GM simulator would take years of CPU time to evaluate, thus would be unfeasible. As the TTF maps can be provided for any combination of input parameters, the emulator has an outstanding computational advantage over the GM simulator and has a great potential to inform infrastructure slope design and maintenance in engineering practice. Future work in developing the emulator includes incorporating weather and climate change variables, for example, discrete extreme rainfall events along with the changing patterns and magnitude of seasonal weather thought likely to occur because of a changing future climate.

References

Abbott S (2018) *Earthworks Technical Strategy.* Technical Report, Network Rail.

Andrianakis I and Challenor PG (2012) The effect of the nugget on Gaussian process emulators of computer models. *Computational Statistics & Data Analysis 56*, 4215–4228.

Apted JP (1977) *Effects of Weathering on some Geotechnical Properties of London Clay.* PhD thesis, Imperial College, University of London.

Bastos LS and O'Hagan A (2009) Diagnostics for Gaussian process emulators. *Technometrics 51*, 425–438.

Baudin M, Christopoulou M, Colette Y and Martinez J-M (2012) *pyDOE: The Experimental Design Package for Python.* Available at https://pythonhosted.org/pyDOE/#credits, (accessed 20 October 2019).

Botev Z and Belzile L (2020) TruncatedNormal: Truncated Multivariate Normal and Student Distributions. Available at https://CRAN.R-project.org/package=TruncatedNormal, (accessed 10 December 2020).

Botev ZI (2017) The normal law under linear restrictions: Simulation and estimation via minimax tilting. *Journal of the Royal Statistical Society B 79*, 125–148.

Briggs KM, Dijkstra TA and Glendinning S (2019) Evaluating the deterioration of geotechnical infrastructure assets using performance curves. In *International Conference on Smart Infrastructure and Construction 2019 (ICSIC)*. London: ICE, pp. 429–435.

Bromhead EN and Dixon N (1986) The field residual strength of London Clay and its correlation with laboratory measurements, especially ring shear tests. *Géotechnique 36*, 449–452.

Brooks S, **Gelman A**, **Jones GL and Meng X-L** (2010) *Handbook of Markov Chain Monte Carlo*. Boston: Chapman & Hall/CRC.

Busby D (2009) Hierarchical adaptive experimental design for Gaussian process emulators. *Reliability Engineering & System Safety 94*, 1183–1193.

Chang EY, **Strong M and Clayton RH** (2015) Bayesian sensitivity analysis of a cardiac cell model using a Gaussian process emulator. *PLoS One 10*, 1–20.

Charlesworth G (1984) Chapter 7: Trunk road motorways in England. In *A History of British Motorways*. London: Thomas Telford, pp. 94–157.

Cressie N (1993) *Statistics for Spatial Data*. New York: John Wiley & Sons.

Dette H and Pepelyshev A (2021) Generalized latin hypercube design for computer experiments. *Technometrics 52*, 421–429.

Dixon N, **Crosby CJ**, **Stirling R**, **Hughes PN**, **Smethurst J**, **Briggs K**, **Hughes D**, **Gunn D**, **Hobbs P**, **Loveridge F**, **Glendinning S**, **Dijkstra T and Hudson A** (2019) In situ measurements of near-surface hydraulic conductivity in engineered clay slopes. *Quarterly Journal of Engineering Geology and Hydrogeology 52*, 123–135.

Ellis EA and O'Brien AS (2007) Effect of height on delayed collapse of cuttings in stiff clay. *Proceedings of the Institution of Civil Engineers - Geotechnical Engineering 160*, 73–84.

Farah M and Kottas A (2014) Bayesian inference for sensitivity analysis of computer simulations, with an application to radiative transfer models. *Technometrics 56*, 159–173.

FLAC (2016) *Fast Lagrangian Analysis of Continua, Ver. 8.0*. Minneapolis, MN: Itasca Consulting Group.

Gao Y, **Wu D and Zhang F** (2015) Effects of nonlinear failure criterion on the three-dimensional stability analysis of uniform slopes. *Engineering Geology 198*, 87–93.

Gramacy R (2020) *Surrogates*. Boca Raton: CRC Press.

Hight DW, **McMillan F**, **Powell JM**, **Jardine RJ and Allenou CP** (2003) Some characteristics of London Clay. In *Characterisation and Engineering Properties of Natural Soils*. Singapore: National University of Singapore, pp. 851–907.

Homma T and Saltelli A (1996) Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering and System Safety 52*, 1–17.

Hu Z, **Xu Y**, **Korkali M**, **Chen X**, **Mili L and Tong CH** (2020) Uncertainty quantification in stochastic economic dispatch using Gaussian process emulation. In *2020 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. Washington, DC: IEEE, pp. 1–5.

Jurečič N, **Zdravković L and Jovičić V** (2013) Predicting ground movements in London Clay. *Proceedings of the Institution of Civil Engineers - Geotechnical Engineering 166*, 466–482.

Kang F, **Han S**, **Salgado R and Li J** (2015) System probabilistic stability analysis of soil slopes using Gaussian process regression with Latin hypercube sampling. *Computers and Geotechnics 63*, 13–25.

Kennedy MC, **Anderson CW**, **Conti S and O'Hagan A** (2006) Case studies in Gaussian process modelling of computer codes. *Reliability Engineering & System Safety 91*, 1301–1309.

Kovacevic N, **Hight DW and Potts DM** (2007) Predicting the stand-up time of temporary London Clay slopes at Terminal 5, Heathrow Airport. *Géotechnique 57*, 63–74.

Kyzyurova KN (2017) *On Uncertainty Quantification for Systems of Computer Models*. PhD thesis, Department of Statistical Science, Duke University.

Labuz JF and Zang A (2012) Mohr–Coulomb failure criterion. *Rock Mechanics and Rock Engineering 45*, 975–979.

Maatouk H and Bay X (2017) Gaussian process emulators for computer experiments with inequality constraints. *Mathematical Geosciences 49*, 557–582.

Neal RM (1997) *Monte Carlo Implementation of Gaussian Process Models for Bayesian Regression and Classification*. Technical Report 9702. Department of Statistics, University of Toronto, Toronto, ON.

Network Rail (2017) *CP6 Earthworks Asset Policy Development. Task 63 - Global Stability and Resilience Appraisal*. Interim Report. Technical Report.

O'Hagan A (2006) Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering & System Safety 91*, 1290–1300.

Oliver MA and Webster R (2015) *Basic Steps in Geostatistics: The Variogram and Kriging*. Cham, Switzerland: Springer.

Oyebamiji OK, **Wilkinson DJ**, **Jayathilake PG**, **Curtis TP**, **Rushton SP**, **Li B and Gupta P** (2017) Gaussian process emulation of an individual-based model simulation of microbial communities. *Journal of Computational Science 22*, 69–84.

Postill H, **Helm PR**, **Dixon N**, **Glendinning S**, **Smethurst JA**, **Rouainia M**, **Briggs KM**, **El-Hamalawi A and Blake AP** (2021) Forecasting the long-term deterioration of a cut slope in high-plasticity clay using a numerical model. *Engineering Geology 280*, 1–19.

Potts DM, **Kovacevic N and Vaughan PR** (1997) Delayed collapse of cut slopes in stiff clay. *Géotechnique 47*, 953–982.

Power CM and Abbott S (2019) Introduction to ground-related risk to transportation infrastructure. *Quarterly Journal of Engineering Geology and Hydrogeology 52*, 280–285.

Rasmussen CE and Nickisch H (2010) Gaussian processes for machine learning (GPML) toolbox. *Journal of Machine Mearning Research 11*, 3011–3015.

Rasmussen CE and Williams CI (2006) *Gaussian Processes for Machine Learning*. Cambridge, MA: The MIT Press.

Rouainia M, **Helm P**, **Davies O and Glendinning S** (2020) Deterioration of an infrastructure cutting subjected to climate change. *Acta Geotechnica 15*, 2997–3016.

Santner TJ (2018) *The Design and Analysis of Computer Experiments. Springer Series in Statistics*, 2nd Edn. New York: Springer.

Seeger A (2004) Gaussian processes for machine learning. *International Journal of Neural Systems 14*, 69–106.

Skempton AW (1996) Embankments and cuttings on the early railways. *Construction History 11*, 33–49.

Smith C (2020) Landslide Confirmed as Trigger for Fatal Stonehaven Derailment. Available at https://www.newcivilengineer.com/latest/landslide-confirmed-as-trigger-for-fatal-stonehaven-derailment-14-08-2020/, (accessed 20 August 2020).

Stirling RA, **Glendinning S and Davie CT** (2017) Modelling the deterioration of the near surface caused by drying induced cracking. *Applied Clay Science 146*, 176–185.

Su G, **Yu B**, **Xiao Y and Yan L** (2016) Gaussian process machine-learning method for structural reliability analysis. *Advances in Structural Engineering 7*, 1257–1270.

Summersgill FC, **Kontoe S and Potts DM** (2018) Stabilisation of excavated slopes in strain-softening materials with piles. *Géotechnique 68*, 1–14.

Take W and Bolton M (2011) Seasonal ratcheting and softening in clay slopes, leading to first-time failure. *Géotechnique 61*, 757–769.

Tsiampousi A, **Zdravkovic L and Potts D** (2017) Numerical study of the effect of soil–atmosphere interaction on the stability and serviceability of cut slopes in London Clay. *Canadian Geotechnical Journal 54*, 405–418.

van Genuchten MT (1980) A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal 44*, 892–898.

Wang R, **Chong L**, **Xu J and Pan L** (2018) Development and verification of large deformation model considering stiffness deterioration and shear dilation effect in FLAC3D. *International Journal of Mining Science and Technology 28*, 959–967.

Wani O, **Scheidegger A**, **Carbajal JP**, **Rieckermann J and Blumensaat F** (2017) Parameter estimation of hydrologic models using a likelihood function for censored and binary observations. *Water Research 121*, 290–301.

Woodman ND, **Smethurst JA**, **Roose T**, **Powrie W**, **Meijer GJ**, **Knappett JA and Dias T** (2020) Mathematical and computational modelling of vegetated soil incorporating hydraulically-driven finite strain deformation. *Computers and Geotechnics 127*, 1–12.

Xiang F and Neal P (2014) Efficient MCMC for temporal epidemics via parameter reduction. *Computational Statistics and Data Analysis 80*, 240–250.

Zhang H (2004) Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of American Statistical Association 99*, 250–261.