Proc. of the 8th International Symposium on Geotechnical Safety and Risk (ISGSR)
Edited by Jinsong Huang, D.V. Griffiths, Shui-Hua Jiang, Anna Giacomini and Richard Kelly
 Published by Research Publishing, Singapore.
doi: 10.3850/978-981-18-5182-7_07-011-cd

Bayesian Emulation of Computer Experiments of Infrastructure Slope Stability Models

Aleksandra Svalova[1*], Peter Helm[2], Dennis Prangle[3], Mohamed Rouainia[2], Stephanie Glendinning[2], and Darren Wilkinson[4]

[1]School of Mathematics, Statistics and Physics, Newcastle University, Newcastle-upon-Tyne, NE1 7RU, United Kingdom.
[*]E-mail: alex.svalova@newcastle.ac.uk
[2]School of Engineering, Newcastle University, Newcastle-upon-Tyne, NE1 7RU, United Kingdom.
[3]Institute for Statistical Science, University of Bristol, Bristol, BS8 1TH, United Kingdom.
[4]Department of Mathematical Sciences, Durham University, Durham, DH1 3LE, United Kingdom.

Abstract: We performed a fully-Bayesian Gaussian process emulation and sensitivity analysis of a numerical model that simulates transport cutting slope deterioration. In the southern UK, a significant proportion of transport infrastructure is built in overconsolidated high-plasticity clay that is prone to deterioration due to seasonal wetting-drying cycles and weather extremes (Stirling 2021; Postill et al. 2021). Geotechnical modelling software (FLAC) was used to simulate the dissipation of excess pore water pressure and seasonal pore water pressure cycles in cuttings (Rouainia et al. 2020). However, due to their high computational expense, it was impractical to perform the number of computer simulations that would be sufficient to understand deterioration behaviour over a range of cutting geometries and soil strength parameters. To address this, we used Gaussian processes and Bayesian inference to emulate the relation between deterioration factors and slope properties (Bastos and O'Hagan 2009). These factors include time to failure (Svalova et al. 2021), failure area, and factor of safety. For our training data, we used a Latin hypercube design to create a computer experiment of 76 numerical models whereby we varied slope height, angle, peak cohesion, peak friction, and permeability. Some of the runs did not reach ultimate limit state failure, resulting in censored times to failure and failure areas. We used Markov chain Monte Carlo sampling to obtain posterior distributions of the emulator parameters, as well as the censored times to failure (Brooks et al. 2010; Kyzyurova 2017). Our emulator could be used to inform slope design, management, and maintenance on different spatio-temporal scales of transport networks.

Keywords: emulation; Gaussian processes; infrastructure slopes; Bayesian inference.

1 Introduction

A large part of the UK geotechnical rail infrastructure is approaching 200 years of age whilst being built on/within soils that are vulnerable to weathering and deterioration. Here, we focus on cuttings in high-plasticity soils which were identified as being of interest by stakeholders and have been shown to be vulnerable to delayed failure on transport networks (Perry 1989). Such high-plasticity earthworks form a significant proportion of the southern UK railway network and have undergone significant deterioration over its working life. A lack of understanding of the long-term deterioration of infrastructure slopes can lead to uncontrolled deformations, thus reduced service performance (Briggs et al. 2019), negative impact on the economy (Power and Abbott 2019), and fatalities (Smith 2020).

The present study focuses on obtaining a better understanding of the deterioration of infrastructure assets, and thus reduces the risks posed to infrastructure systems. We study deterioration through emulating a set computer experiment runs (Rouainia et al. 2020; Postill et al. 2021) using Gaussian processes (Rasmussen and Nickisch 2010; Su et al. 2016) and Bayesian inference (O'Hagan 2006). The latter allows the use of expert opinion, which is very useful when experimental data is computationally expensive to obtain. Our simulator is based on a geotechnical model (GM) deploying a coupled hydro-mechanical strain-softening Mohr-Coulomb constitutive model and is implemented using the Fast Lagrangian Analysis of Continua with Two-Phase Flow software (FLAC-TP; Itasca, 2016). Overconsolidated clays undergo strain-softening during swelling, as observed following the excavation of cut slopes. Softening also occurs during seasonal ratchetting driven by seasonal shrink-swell cycles. The adopted softening behaviour is derived from previous modeling studies (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 stability is modeled through a coupled fluid-mechanical approach (Rouainia et al. 2020) utilizing a non-local strain-softening model (Summersgill et al. 2018; Postill et al. 2021). This allows a detailed assessment of weather-driven deterioration.

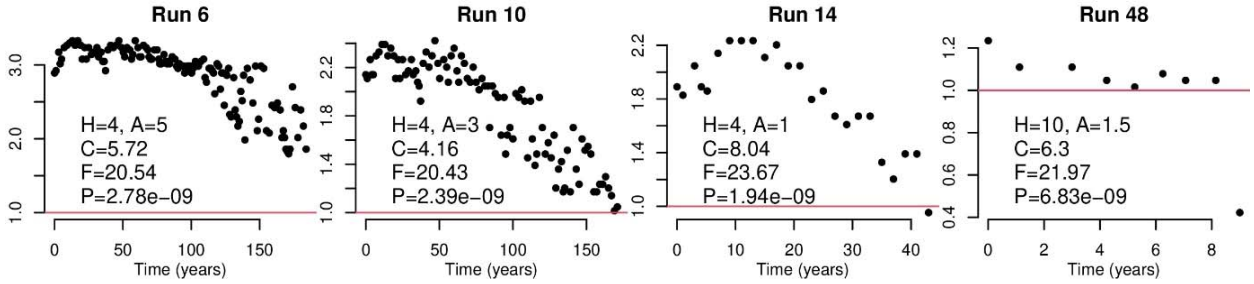

Figure 1. Examples of winter FOS measurements, against time since construction. Red line indicates an FOS of unity (ultimate limit state failure). Legends: height (H), angle cotangent (A), cohesion (C), friction angle (F), and permeability (P).

In this work, we use a Gaussian process emulator (GPE) to approximate computer experiments with 76 simulator runs. The simulator can be represented by a function $f(\boldsymbol{x}) = \boldsymbol{y}$, where $\boldsymbol{y}$ is the observed output and $\boldsymbol{x}$ is the corresponding input. It is assumed that $f(\boldsymbol{x})$ is a smooth, continuous function of its input variables (O'Hagan 2006) and that the outputs of $f(\boldsymbol{x})$ can be modeled as a multivariate normal distribution for any $\boldsymbol{x}$. The training output $\boldsymbol{y}_i = f(\boldsymbol{x}_i), i = 1,2,\dots,n$ is evaluated at a simulator input $\boldsymbol{x}_i$, and for any inputs that do not belong to $\boldsymbol{x}_i$, the emulator produces a distribution of interpolated values. Our inputs $\boldsymbol{x}$ are slope geometry (angle and height), peak soil shear strength (cohesion and friction angle) and near surface reference permeability; their ranges are summarized in Table 1. These variables were selected due to their importance in assessing the stability of geotechnical infrastructure (Potts et al. 1997; Ellis and O'Brien 2007; Rouainia et al. 2020). We monitored a number of outputs, including time to failure (TTF), failure area, and factor of safety (FOS). Figure 1 illustrates some examples of the FOS curves. Evidently, there is a large variation in the starting FOS values, TTF, and noise along the main trend. We deployed a Latin hypercube experimental design to ensure an optimal coverage of the parameter space. The computer experiment simulator runs were stopped after 184 years of model time, corresponding to the maximum rail cutting slope age in the London-Bristol corridor (Skempton 1996). Slopes that have not failed within this time have their TTF and failure area censored and are imputed by the GPE as well.

2 Gaussian Process Emulator

2.1 Introduction

A GPE $\eta(\cdot)$ is a Gaussian process conditioned on observations (i.e. simulator runs) producing an interpolator for predictions and uncertainty quantification. This introduction to GPEs is borrowed from our earlier work on emulation using Gaussian processes (Svalova et al. 2021). The emulator $\eta(\cdot)$ takes a generic input $\boldsymbol{x} = (x_1, x_2, \dots, x_p)$ where $x_i \in \chi_i \subset \mathbb{R}$. A Gaussian process is fully defined by its mean and covariance functions m and V, $\eta(\cdot) \sim \mathrm{GP}(m(\cdot), V(\cdot,\cdot))$. Here, $m: \mathbb{R}^p \to \mathbb{R}$ is a linear map of the inputs, $m(\boldsymbol{x}) = h(\boldsymbol{x})^T\boldsymbol{\beta}$, where $h(\cdot): \mathbb{R}^p \to \mathbb{R}^q$ maps $\boldsymbol{x}$ to a vector of linear regressors and $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_q)$ is a coefficient vector. We assign $V(\boldsymbol{x}, \boldsymbol{x}') = \tau[C(x, x', \boldsymbol{\theta}) + \epsilon\mathbb{I}(\boldsymbol{x}, \boldsymbol{x}')]$, where τ is the scale parameter, ϵ is a nugget for improving numerical stability (Andrianakis and Challenor 2012), and $\mathbb{I}$ is an indicator function for the event $\boldsymbol{x} = \boldsymbol{x}'$. The correlation function C is frequently a Gaussian or a Matérn type (Rasmussen and Williams 2006), and $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)$ is a vector of correlation lengths.

The analytical expression for a GPE conditioned on a set of runs is defined as follows. Given a collection of n observed experimental outputs $\boldsymbol{y} = (\eta(\boldsymbol{x}_1), \eta(\boldsymbol{x}_2), \dots, \eta(\boldsymbol{x}_n)), n < \infty$ performed on the inputs $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n$, the n-vector $\boldsymbol{y}$ follows a multivariate normal distribution, $\boldsymbol{y}|\boldsymbol{\beta}, \tau, \boldsymbol{\theta}, \epsilon \sim N(H_x\boldsymbol{\beta}, \tau\Sigma_x)$, where H_x is a matrix of regressors whose ith row is $h(\boldsymbol{x}_i)$ and $\Sigma_{x(i,j)} = C(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{\theta}) + \epsilon\mathbb{I}(i,j)$. Using standard rules for conditioning on a set of observations (Gramacy 2020), $\eta(\cdot)|y, \boldsymbol{\beta}, \tau, \boldsymbol{\theta}, \epsilon \sim GP(m^*(\cdot), V^*(\cdot,\cdot))$, where

$$m^*(\boldsymbol{x}) = h(\boldsymbol{x})^T\boldsymbol{\beta} + t(\boldsymbol{x})^T\Sigma_x^{-1}(\boldsymbol{y} - H_x\boldsymbol{\beta}),\ V^*(\boldsymbol{x}, \boldsymbol{x}') = \tau\big(C(\boldsymbol{x}, \boldsymbol{x}', \boldsymbol{\theta}) - t(\boldsymbol{x})^T\Sigma_x^{-1}t(\boldsymbol{x}')\big), \quad (1)$$

where $t(\boldsymbol{x}) = \big(C(\boldsymbol{x}, \boldsymbol{x}_1, \boldsymbol{\theta}), C(\boldsymbol{x}, \boldsymbol{x}_2, \boldsymbol{\theta}), \dots, C(\boldsymbol{x}, \boldsymbol{x}_n, \boldsymbol{\theta})\big)^T$ is a column vector of correlations between the (generic) emulator input $\boldsymbol{x}$ and training inputs $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n$. The above formalism applies to single-output processes, however it can be extended to multiple-output problems straightforwardly using multivariate normal theory.

2.2 Censored observations

Some of the experimental runs have not reached failure in 184 years of model time, thus their time to failure and failure area are censored. This is not the case for models which undergo rapid failure resulting in very short time series. The censoring occurring in the latter is currently interpreted as failure immediately after construction, and modelling time series which have fewer than three FOS measurements are ignored. What follows only applies to the censored area and TTF. Suppose n experiments at $\boldsymbol{x}_o = (\boldsymbol{x}_{o,1}, \boldsymbol{x}_{o,2}, \dots, \boldsymbol{x}_{o,n})$ produced uncensored

observations $\boldsymbol{y}_o$. Also, suppose that n_c experiments at $\boldsymbol{x}_c = (\boldsymbol{x}_{o,1}, \boldsymbol{x}_{o,2}, \ldots, \boldsymbol{x}_{o,n})$ produced censored observations, and assume that $\boldsymbol{y}_c$ is the corresponding vector of (hypothetical) uncensored times to failure. Together with $\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}$ and τ, $\boldsymbol{y}_c$ can also be inferred using the GPE. Define a new process $\eta_c(\cdot)$ where (Kyzyurova 2017)

$$\eta_c(\boldsymbol{x}) = \begin{cases} \eta(\boldsymbol{x}), & \text{if } \eta(\boldsymbol{x}) < c, \\ c, & \text{otherwise,} \end{cases} \tag{2}$$

where $\eta(\cdot)$ is as above. Therefore, at design points $\mathrm{X} = (\boldsymbol{x}_c, \boldsymbol{x}_o)$ $\eta_c(\boldsymbol{x})$ is distributed as follows

$$\eta(\boldsymbol{x}_o)|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}, \tau \sim \mathrm{N}(H_o\boldsymbol{\beta}, \sigma^2\Sigma_o), \ \eta_c(\boldsymbol{x}_c)|\eta(\boldsymbol{x}_o), \beta, \sigma^2, \boldsymbol{\theta}, \tau \sim \mathrm{TN}_{(c,\infty)}(m_c, V_c), \tag{3}$$

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

where $\mathrm{TN}_{c,\infty}$ denotes a normal distribution truncated below c. In the above, H_o and Σ_o are equivalent to H_x and Σ_x as defined earlier, here using inputs $\boldsymbol{x}_o$. Similarly, H_c is a regressor matrix associated with $\boldsymbol{x}_c$, and $\Sigma_{c(i,j)} = C\left(\boldsymbol{x}_{c,i}, \boldsymbol{x}_{c,j}, \boldsymbol{\theta}\right) + \tau\mathbb{I}(i,j)$ and $\Sigma_{o,c} = \Sigma_{c,o}^T$. We used the package TruncatedNormal in R (Botev and Belzile 2020) to obtain truncated multivariate normal samples.

2.3 Experimental design

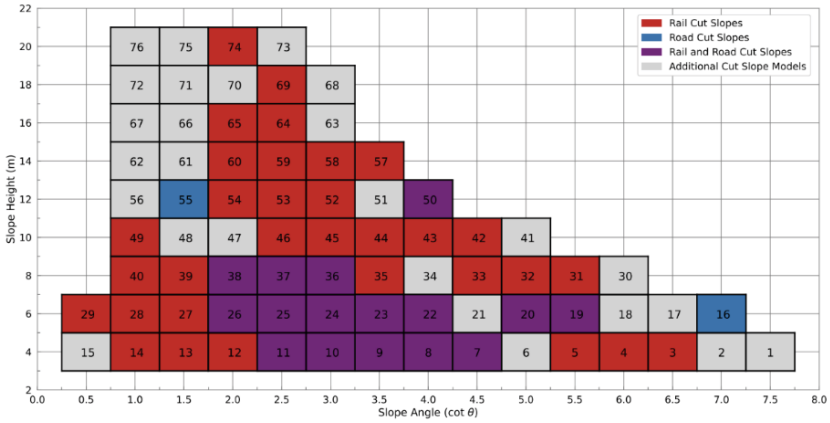

Figure 2. Geometry array of the training data used in our study. The numbers indicate identifiers of the training data runs. Adapted from (Svalova, et al. 2021) Supplementary Material.

A set of 76 runs of the geotechnical simulator was created to train the GPE. To obtain an optimal coverage of the parameter space we used a Latin hypercube design (LHD) with a maximin criterion (Santner 2018). Table 1 illustrates the variable ranges used in experimental design, which were chosen to represent railway and highway cuttings in high-plasticity soils (Cripps and Taylor 1986). Figure 2 illustrates the geometries which are covered in our design, which were chosen based on slope survey data provided by Network Rail and Mott MacDonald. In particular, the training run identifiers were sampled in the LHD, and then converted into height and angle.

Table 1. Input variables to the computer experiment.

Property	Height (m)	Angle cotangent	Peak friction (˚)	Peak cohesion (kPa)	Nr. Surface Reference Permeability (m/s)
Range	[4, 20]	[0.5, 7.5]	[18.5, 25]	[3, 10]	[1.45×10^{-9}, 2.5×10^{-8}]

3 Motivating Applications in Slope Stability Analysis

The GPE described above can be applied to a variety of practical scenarios. Here, we have studied cuttings in London Clay and applied emulation to estimate TTF, failure area, and FOS. A trained emulator could rapidly obtain Bayesian estimates for these deterioration indicators in a matter of seconds for a single earthwork or hours for several hundreds of scenarios. This is in contrast to days/weeks of computation time for a single earthwork using the original geotechnical simulator.

3.1 Time to failure

Emulation was applied to TTF observations on 50 out of the 76 simulator runs. The remaining 26 runs did not reach failure state within 184 years of model time. Their TTF was, thus, censored and estimated conditionally on the observations. An asset's predicted TTF can be used as a rapid assessment of stability and level of deterioration.

As TTF is strictly positive, square root of TTF was emulated by assigning it a Gaussian process prior and formulating the likelihood on the square root of the TTF observations.

3.2 Failure area

An algorithm was created that identified the bounds of the failed soil mass and exported the spatial coordinates (Figure 3). As the failure surface in the modelled slopes is dominated by plastic shearing, the algorithm identified model zones that were undergoing plastic shear failure and that were also experiencing shear strain rates greater than a prescribed minimum. The coordinates of the shear surface were exported from the model and a polynomial curve was fitted to the data and the area constrained by the ground surface and the shear surface was calculated using numerical integration. Similarly to TTF, the failure area (at the ultimate limit state) was only observed for the 50 runs that reached failure during computation, and the remaining 26 areas were estimated using the emulator. As with TTF, the square root of area was emulated by assigning it a Gaussian process prior.

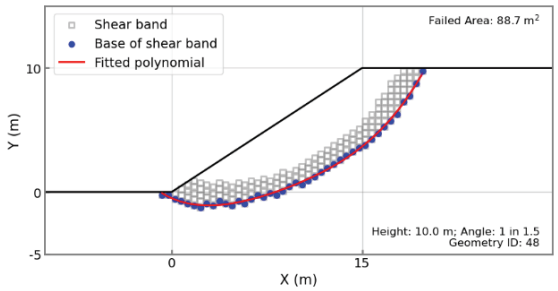

Figure 3. Example shear surface extracted from the FLAC model (grey squares) of **Run 48** (Figure 1), python algorithm identifies base of surface (blue) and a polynomial is fitted. The area between the shear surface and the ground surface is derived using numerical integration.

3.3 Factor of safety

Factor of safety is frequently used in design (Matthews et al. 2014), and factor of safety deterioration curves could be used in the maintenance and management of infrastructure earthworks and in the design of new build earthworks (Postill et al. 2021). To emulate FOS change with time without the high computational expense incurred in the emulation of dynamic models we assumed that the exponent of FOS follows linear regression. For a computer run $i \in \{1, \dots, N\}$, FOS $y_{i,j}$ changes with time t_i via the following relation:

$$z_{i,j} = \exp\{y_{i,j}\} = e + \alpha_i - \gamma_i t_{i,j} + \varepsilon_{i,j},\ \varepsilon_{i,j} \sim \mathrm{N}(0, \sigma_i^2),\ j = 1, \dots, n_i. \tag{5}$$

In the above, the expected value of $z_{i,j}$ immediately after construction is at least $e = \exp\{1\}$, $\mathrm{E}[z_{i,j} | t_{i,j} = 0] \geq \exp\{1\}$. Further, α_i is the initial "distance from failure" and γ_i is the decline rate, n_i is the number of FOS measurements in a run i. Note that α_i, γ_i and σ_i^2 are strictly positive $\forall\ i$ and j. It follows that $\boldsymbol{z}_i = (z_{i,1}, \dots, z_{i,n_i})$ has a Gaussian likelihood

$$\mathrm{L}(\alpha_i, \gamma_i, \sigma_i^2 | \boldsymbol{z}_i, \boldsymbol{t}_i) \equiv f(\boldsymbol{z}_i | \alpha_i, \gamma_i, \sigma_i^2, \boldsymbol{t}_i) = \prod_{j=1}^{n_i} \frac{1}{2\pi\sigma_i^2} \exp\left\{ -\frac{(z_{i,j} - e - \alpha_i + \gamma_i t_{i,j})^2}{2\sigma_i^2} \right\}. \tag{6}$$

Natural logarithms of α_i, γ_i, and σ_i^2 are assigned independent Gaussian process priors and then emulated. The latter parameters were estimated in one MCMC step together with the remaining model parameters $\boldsymbol{\beta}, \tau, \boldsymbol{\theta}$, and ϵ.

4 Bayesian Parameter Inference

We use Bayesian inference with Markov chain Monte Carlo simulation (e.g. Brooks et al. 2010) in order to obtain parameter uncertainty estimates in the form of posterior distributions. Whilst it is possible to obtain the estimates using e.g. numerical optimization, the Bayesian approach allows us to incorporate expert knowledge into estimating parameter values. A Metropolis-within-Gibbs sampler was used for $\boldsymbol{\beta}, \tau, \boldsymbol{\theta}$, and ϵ, wherease $\boldsymbol{y}_c$ was updated using Eq. (1) through a Gibbs step. The code was written in R statistical software (R Core Team 2021). In the case of TTF and failure area, the coefficients were given prior distributions as follows:

$$\beta_0 \sim \mathrm{N}(0,100), \beta_i \sim \mathrm{N}(0,16), \theta_i \sim \mathrm{Exp}(0.2), \tau \sim \mathrm{IGa}(10,100), \epsilon \sim \mathrm{IGa}(3,1), i = 1,2, \dots,5. \tag{7}$$

In the case of the FOS modelling, the following priors were used

$$\beta_{0,p} \sim \mathrm{N}(0,100), \beta_{i,p} \sim \mathrm{N}(0,16), \theta_i \sim \mathrm{Exp}(0.2), \tau_p \sim \mathrm{IGa}(10,100), \epsilon \sim \mathrm{IGa}(3,1), i = 1,2, \dots,5, p = 1,2,3. \tag{8}$$

We also used $\log(\boldsymbol{\alpha}) \sim \mathrm{N}(h(\boldsymbol{x})^T\boldsymbol{\beta_1}, \tau_1\Sigma_x), \log(\boldsymbol{\gamma}) \sim \mathrm{N}(h(\boldsymbol{x})^T\boldsymbol{\beta_2}, \tau_2\Sigma_x)$, and $\log(\boldsymbol{\sigma}^2) \sim \mathrm{N}(h(\boldsymbol{x})^T\boldsymbol{\beta_3}, \tau_3\Sigma_x)$ prior distributions, where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)^T, \boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_n)^T$, and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)^T$.

5 Results

5.1 Time to failure

The results regarding MCMC efficiency can be found in our earlier publication (Svalova et al. 2021). Figure 4 illustrates the posterior distributions of TTF for different soil scenarios. The values increase linearly with angle cotangent, which is reasonable as a decrease in slope steepness leads to an increase in TTF. It is also evident that TTF is, on average, higher for the high-strength soil example. For the London Clay-type soil, an earthwork with a height of 6 m and base of 15 m (angle cotangent of 2.5) has an estimated failure time of 100-150 years.

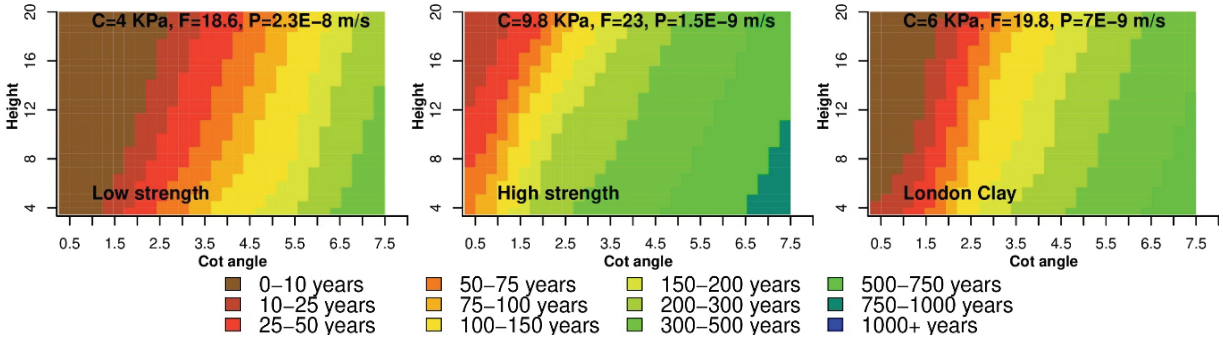

Figure 4. Maps of posterior mean values of TTF versus slope geometry. The legends in the top-left corner of the plots show cohesion (C), friction angle (F), and permeability (P). Image adapted from Svalova et al (2021).

5.2 Failure area

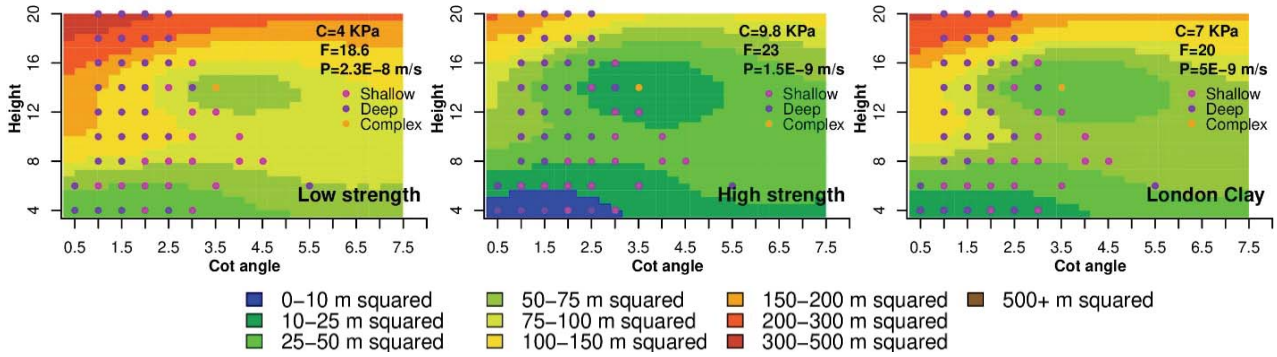

Figure 5. Map of posterior mean values of failure area versus slope geometry. The dots indicate experimental data, categorized according to the estimated failure mechanism.

Figure 5 illustrates posterior mean values of the failure area versus slope geometry. Evidently, the change in area is more complex than that of TTF. There appears to be a change of behaviour starting from around 8 m slope height. We hypothesized that this might be caused by the relationship between failure area and failure type (shallow vs deep), thus we categorized failures based on the failure surface geometry. It is evident that with the change in failure mechanism, the dependency of failure area on geometry changes. We did not attempt to emulate the relationship between failure type and experiment inputs, however this could be a subject of future investigation.

5.3 Factor of safety

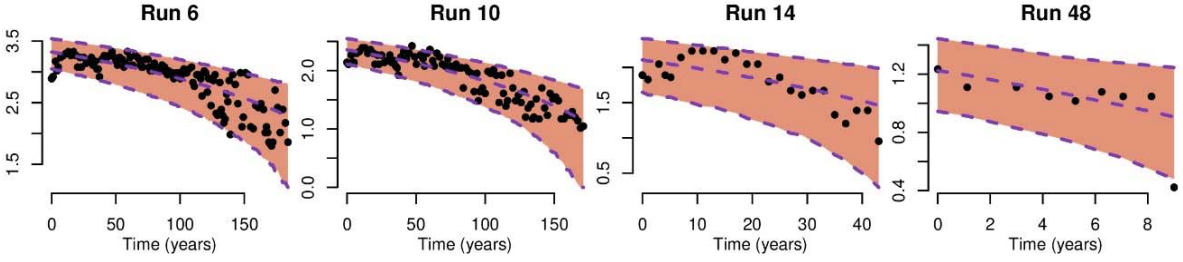

Figure 6. Posterior distributions of the FOS, mean values and 95% Bayesian predictive intervals are in dashed lines.

Figure 6 illustrates the posterior distributions of FOS for a selection of simulator runs. It is evident that the mean trajectory is covering the general decline in the FOS reasonably well, and the 95% posterior regions capture the variation in the majority of cases. The posterior estimates error variance σ_i^2 appear to capture the spread of the measurements well. However, it might be the case that for e.g. run 14, the proposed exponential regression is not concave enough. We will, therefore, be trying different parametric models, as well as dynamical modelling, to find one (class of models) that fits most optimally.

6 Conclusions

We had developed a GPE of the geotechnical simulator in a Bayesian setting in order to improve computational efficiency, whilst providing uncertainty estimates for the parameters controlling our emulator. This emulator is suitable for single- and multiple-output scenarios and can be used for a rapid assessment of earthwork deterioration state, as well as asset management and maintenance. Whilst, currently, our study only covers cuttings in high-plasticity soils that have not undergone interventions, we plan to emulate deterioration of earthworks in medium-plasticity soils, embankments, the impact of climate change, and the use of interventions (e.g. soil nails).

References

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

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

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

Bromhead, E.N. 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, G.L., and Meng, X.-L. (2010). *Handbook of Markov Chain Monte Carlo,* Boston: Chapman & Hall/CRC.

Burland, J.B., Rampello, S., Georgiannou, V. N., and Calabresi, G. (1996). A laboratory study of the strength of four stiff clays. *Géotechnique*, 46(3), 491–514.

Cripps, J.C. and Taylor, R.K. (1986). Engineering characteristics of british over-consolidated clays and mudrocks I. Tertiary deposits. *Engineering Geology*, 22(4), 349–376. doi:https://doi.org/10.1016/0013-7952(86)90004-9.

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

FLAC. 2016. Fast Lagrangian Analysis of Continua. Minneapolis: Itasca Consulting Group.

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

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

Matthews, C., Farook, Z., and Helm, P.R. (2014). Slope stability analysis – limit equilibrium or the finite element method? *Ground Engineering*, May: 22-28.

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

Perry, J. (1989). *A survey of slope condition on motorway earthworks in England and Wales.* Transport Research Laboratory.

Postill, H., Helm, P.R., Dixon, N., El-Hamalawi, A., Glendinning, S., and Take, W.A. (2021). Strength parameter selection framework for evaluating the design life of clay cut slopes. *Proceedings og the ICE - Geotech. Engineering. Published online ahead of print.* doi:10.1680/jgeen.21.00125.

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

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

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

R Core Team. (2021). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Rasmussen, C.E. and Williams, C.I. (2006). *Gaussian Processes for Machine Learning*, Cambridge: The MIT Press.

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

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, T.J. (2018). The Design and Analysis of Computer Experiments. In *Springer Series in Statistics*, New York: Springer.

Skempton, A.W. (1996). Embankments and cuttings on the early railways. *Construction History*, 11, 33-49.

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

Stirling, R.A., Glendinning, S., and Davie, C.T. (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, F.C., Kontoe, S., and Potts, D.M. (2018). Stabilisation of excavated slopes in strain-softening materials with piles. *Géotechnique*, 68, 1–14.

Svalova, A., Helm, P., Prangle, D., Rouainia, M., Glendinning, S., and Wilkinson, D.J. (2021). Emulating computer experiments of transport infrastructure slope stability using Gaussian processes and Bayesian inference. *Data-Centric Engineering*, 2, E12.

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.