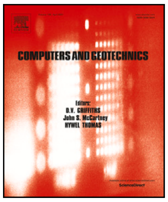
Computers and Geotechnics 141 (2022) 104491

Contents lists available at ScienceDirect

Computers and Geotechnics

journal homepage: www.elsevier.com/locate/compgeo



Research paper

Performance study of iterative Bayesian filtering to develop an efficient calibration framework for DEM

Philipp Hartmann [a], Hongyang Cheng [b,*], Klaus Thoeni [a]

[a] Centre for Geotechnical Science and Engineering, The University of Newcastle, Callaghan, NSW 2308, Australia
[b] Department of Civil Engineering, Faculty of Engineering Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

ARTICLE INFO

Keywords:
Discrete Element Method (DEM)
Convergence
Machine learning
Bayesian calibration
Triaxial compression
Multi-objective optimisation

ABSTRACT

This work presents an efficient probabilistic framework for the Bayesian calibration of micro-mechanical parameters for Discrete Element Method (DEM) modelling. Firstly, the superior behaviour of the iterative Bayesian filter over the sequential Monte Carlo filter for calibrating micro-mechanical parameters is shown. The linear contact model with rolling resistance is used for simulating the triaxial responses of Toyoura sand under different confining pressures. Secondly, synthetic data from DEM simulations of triaxial compression are used to assess the reliability of iterative Bayesian filtering with respect to the user-defined parameters, such as the number of samples and predefined parameter ranges. Excellent calibration results with errors between 1 and 2% are obtained when the number of samples is chosen high enough. It is crucial that the sample size is representative for the distribution of individual parameters within the predefined parameter ranges. The wider the ranges, the more samples are required. The investigation also shows the necessity of including both stress and strain histories, at certain confidence levels, for estimation of the correct mechanical responses, especially the correct fabric responses. Finally, based on the findings of this work a fully-automated open-source calibration tool is developed and demonstrated for selected stress paths.

1. Introduction

The Discrete Element Method (DEM) captures the collective behaviour of a granular assembly by tracking the kinematics of the constituent particles (Cundall and Strack, 1979). In recent years, DEM has been applied to deepen our understanding of the fundamental behaviour of soil, e.g. fabric (Ouadfel and Rothenburg, 2001), particle size distribution (Sufian et al., 2021), and pore-fluid coupling (Cheng et al., 2019a), and to produce accurate predictions for processes where the particulate nature of the material cannot be overlooked such as shear localisation (e.g. Guo and Zhao, 2016). The fundamental issue that prevents DEM from being a practical tool for geotechnical engineers is the challenge in measuring particle-scale properties, kinematics and contact forces at the particle scale (e.g. Fuchs et al., 2014; Hurley et al., 2016; Rorato et al., 2021). Because the micro-mechanical parameters are not always measurable in an experiment, their estimation is mostly formulated as an inverse problem, using statistical inference and optimisation techniques, such as gradient descent (e.g. Gao et al., 2021; Qu et al., 2019) and other ad hoc algorithms (e.g. Richter et al., 2020; Orefice and Khinast, 2020), or simply achieved by “trial and error” as often seen in the literature (e.g. Plassiard et al., 2009; Coetzee, 2017).

The design-of-experiment (DOE) approach has been utilised in the last decades as an initial attempt to identify micro-mechanical parameters with a manageable size of DEM model evaluations (Johnstone, 2010; Coetzee, 2020). Hanley et al. (2011) and Yoon (2007) considered the interactive effects between micro-parameters on characteristic bulk properties in designing their DEM “experiments”. Rackl and Hanley (2017) further extended the idea and developed an optimisation-based calibration approach using Latin hypercube sampling and Kriging. They identified a solution space in which similar results are obtained for a range of parameter combinations.

To further improve the efficiency of these optimisation-based techniques, Machine Learning (ML) algorithms have been applied together with other multi-objective optimisation methods. Do et al. (2018) visualised a Pareto-optimal front between the model accuracy and computational cost with a multi-objective optimisation evolutionary algorithm. The Pareto-optimal front also demonstrates a solution space of feasible parameter combinations that can give similar predictions of the bulk responses, constrained by the desired calibration target. To avoid the high computational cost in sampling the solution (parameter) space, surrogate models, such as support vector machines (Ma et al., 2020), back-propagation neural networks (Ye et al., 2019) and radial

---

* Corresponding author.
E-mail addresses: philipp.hartmann@newcastle.edu.au (P. Hartmann), h.cheng@utwente.nl (H. Cheng), klaus.thoeni@newcastle.edu.au (K. Thoeni).

https://doi.org/10.1016/j.compgeo.2021.104491
Received 28 March 2021; Received in revised form 24 September 2021; Accepted 29 September 2021

basis function (Turkia et al., 2019), can be trained with an initial set of DEM model evaluations using parameters drawn from Latin hypercube sampling. He et al. (2019) combined the back-propagation neural network with a genetic algorithm to reduce the number of DEM model evaluations required for optimisation without compromising the accuracy. Westbrink et al. (2021) trained a multi-objective reinforcement learning agent with multiple calibration test data to find an optimal parameter set at a minimal computational cost.

From these efforts in automating DEM calibration, a common trend that can be observed is the utilisation of micro–micro and micro–macro parameter correlations in the sampling of the solution space. Qu et al. (2020) developed a hybrid calibration approach in which the analytical expression that relates Hertz-type contact parameters to the Young's modulus and Poisson's ratio of isotropic granular packings was used to facilitate the optimisation. Another good example is given by Nguyen et al. (2019), where semi-analytic laws for micro–macro parameter correlations are deduced from an ad hoc constructed numerical database.

In optimisation-based approaches the calibration is usually done for specific characteristic bulk properties, such as Young's modulus, angle of repose, peak- or critical-state macroscopic friction. This is a major limitation because these characteristic properties result from a parameterisation of history-dependent stress and strain responses and may include extra bias and uncertainty. Recursive Bayesian filtering provides a rigorous basis for the calibration of numerical models using the raw time-series data. The framework also allows for quantifying the evolution of model uncertainties over a load history (Oden et al., 2010a,b). Hadjidoukas et al. (2014) developed a versatile computational framework to estimate the posterior distribution of microscopic parameters and quantify the model uncertainties for DEM simulations of granular materials. In addition, the framework also allows for Bayesian model selection (Hadjidoukas et al., 2015), i.e. the evaluation of the robustness of physical models conditioned on reference data (either numerical, experimental, or theoretical).

More recently, Cheng et al. (2018) treated the DEM calibration as an inverse problem and solved it with history-dependent stress and strain data based on sequential Monte Carlo (MC) filtering. The method was further improved by Cheng et al. (2019b) with an ML-based iterative sampling technique and implemented in the Bayesian calibration toolbox `GrainLearning` (GL). Within each iteration, the conditional distribution of model parameters on the reference (experimental) data is updated by the recursive Bayes' rule. This conditional (posterior) distribution is updated based on the complete history of the data and used to train a proposal distribution to draw new samples for the subsequent iteration. As the number of iterations increases, the proposal distribution is progressively narrowed down to local optima with decreasing model uncertainty.

The previous contributions by Cheng et al. (2018, 2019b) focus on the mathematical background and implementation of GL. However, the performance of the algorithm has not yet been discussed in detail, nor has the influence of the user-defined parameters. In order to provide a reliable, robust and fully-automated calibration "black-box", it is crucial to investigate aspects including convergence, computational efficiency, observation versus modelling uncertainties and multiple local optima. The current work focuses on understanding these aspects with several key variations in the DEM model (loading condition and the number of particles) and reference data (experimental and synthetic) for the Bayesian calibration. The numerical framework is briefly introduced in Section 2. The performance of GL is investigated in Section 3 by systematically comparing the GL results to those of MC and experimental data. In Section 4, synthetic reference data generated from a DEM simulation of a triaxial compression test is used for a more detailed investigation of the performance and limitations of GL. Having a numeric solution as reference data for the inference eliminates any measurement errors and reveals the true performance of GL. A final discussion of the obtained results is conducted in Section 5. Finally, the main findings of the study and possible extensions of the work are discussed in Section 6.

2. Numerical framework

2.1. *Discrete element method*

The DEM represents granular materials as assemblies of solid particles with simplified geometries and vanishingly small inter-particle overlaps. Spherical particles are often preferred due to their very high computational efficiency. Governed by springs, dashpots and sliders upon collision, the kinematics of the particles are updated using Newton's equations of motion, based on the net forces and moments resulting from the interactions (Cundall and Strack, 1979). The open-source framework `Yade` (Šmilauer et al., 2015) is used to perform the DEM simulations in this work.

2.1.1. *Preparation of particle packing*

Particle packings of dense granular materials are typically created from a cloud of loose particles which is compacted before undergoing a certain loading path. To remove the computational cost involved in this packing generation process, "stress-free" packings are prepared beforehand so that the cost per DEM model evaluation[1] is kept at a minimum. The packing preparation procedure for the triaxial compression tests conducted in this study can be summarised as follows:

(i) A cloud of randomly distributed spherical particles is generated in a cuboidal periodic cell.
(ii) The cloud is compacted at a reference confining pressure p_{ref}, with a very high normal stiffness to keep the inter-particle overlaps small.
(iii) The inter-particle friction is reduced periodically and the packing is allowed to stabilise (e.g., wait until the ratio between kinetic and elastic potential energy is smaller than 10^{-3}), until a reference void ratio e_{ref} is obtained.
(iv) The packing is unloaded to a confining pressure $p \ll p_{\mathrm{ref}}$ and then allowed to stabilise.

The resulting configuration contains particle positions and radii and the dimensions of the periodic cell. Once generated, the particle configuration, for a given number of particles, will be readily importable for each model evaluation, that is the DEM simulation of isotropic and triaxial compression in which random parameter values will be adopted. As a continuation from the "stress-free" state, the packing with randomised contact properties will be first loaded to a target confining pressure σ_c and then sheared under triaxial compression.

In this work, the reference initial confining pressure and void ratio are taken from a Toyoura sand specimen (see Sun et al., 2007), namely, $p_{\mathrm{ref}} = 0.1\,\mathrm{kPa}$ and $e_{\mathrm{ref}} = 0.68$. Three different confining pressures are considered, i.e. $\sigma_c = 0.2$, 0.5 and $1.0\,\mathrm{MPa}$. For Toyoura sand, these confining pressures may lead to a small degree of particle crushing which in turn can affect the grain morphology and mechanical properties (Sun et al., 2007; Wu et al., 2020). The particle diameters are drawn from a scaled size distribution of Toyoura sand. The density of the sand particles is $\rho_g = 2650\,\mathrm{kg\,m^{-3}}$. Six different initial particle configurations with various numbers of particles (1000, 2000, 5000, 8000, 10000, and 27000) have been generated. Fig. 1 shows an example of a randomly generated packing with 2000 particles.

---

[1] A "model evaluation" is referred to as a single simulation with a certain choice of model parameters. The values chosen for this model evaluation are randomly drawn from a multivariate distribution and referred to as a "sample". A sample should not be confused with a "packing" which is the geometrical configuration of a DEM model.

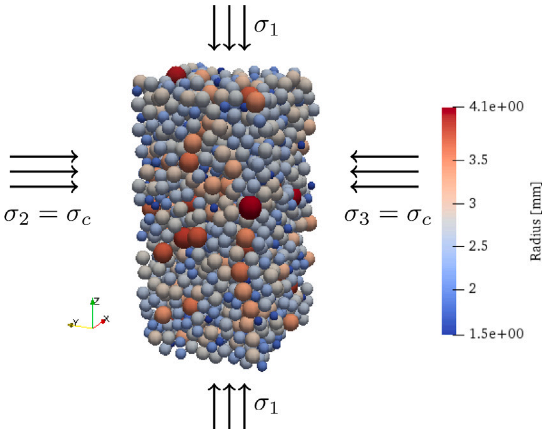


Fig. 1. Illustration of a typical DEM packing consisting of 2000 spherical particles (particles are coloured by size) used in the triaxial compression simulations.

Table 1
Expression for inter-particle normal, tangential, and rolling stiffness and corresponding plastic limit conditions. E and ν are the inter-particle Young's modulus and inter-particle Poisson's ratio, respectively. ϕ is the inter-particle friction angle, r^* is the equivalent radius defined as $r^* = (1/r_1 + 1/r_2)^{-1}$ where r_1 and r_2 are the radii of the two touching particles and η_m is the rolling friction coefficient.

Micro-mechanical stiffness	Plastic limit condition
$k_n = 2Er^*$	
$k_s = 2E\nu r^*$	$\|\mathbf{F}_s\| = \tan\phi \|\mathbf{F}_n\|$
$k_m = \beta_m r_1 r_2 k_s$	$\|\mathbf{M}\| = \eta_m \|\mathbf{F}_n\| \min(r_1, r_2)$

2.1.2. *Contact laws*

The kinematics of each particle is updated from the net force and moment, resulting from contact-level force–displacement and moment–rotation constitutive laws. For two touching spheres the inter-particle normal and tangential forces $\mathbf{F}_n$ and $\mathrm{d}\mathbf{F}_s$ as well as contact moments $\mathbf{M}$ are related to their normal overlap $\mathbf{u}_n$, tangential displacement increment $\mathrm{d}\mathbf{u}_s$ and relative rotation θ via

$$\begin{aligned} \mathbf{F}_n &= k_n \mathbf{u}_n \\ \mathrm{d}\mathbf{F}_s &= k_s\, \mathrm{d}\mathbf{u}_s \\ \mathbf{M} &= k_m \boldsymbol{\theta}. \end{aligned} \tag{1}$$

The parameters k_n, k_s and k_m in Eq. (1) are the stiffnesses in the normal, tangential and bending/twisting directions, respectively. Rolling resistance is typically adopted (through k_m) to handle the fact that soil particles are non-spherical and have a certain surface roughness. This allows keeping computational costs very low compared with simulations with realistic particle shapes. These contact-level springs can be either constant (Cundall and Strack, 1979) or functions of the inter-particle overlap, e.g. according to the Hertz–Mindlin theory. In this work, the linear contact law summarised in Table 1 is used. Both shear and rolling adopt the Mohr–Coulomb type formulation as plastic limit via the inter-particle sliding friction coefficient $\tan(\phi)$ and rolling friction coefficient η_m. The micro-mechanical parameters to be identified in the calibration are the inter-particle Young's modulus E, inter-particle Poisson's ratio ν, inter-particle friction angle ϕ, rolling stiffness coefficient β_m and rolling friction coefficient η_m.

2.2. *Bayesian calibration*

Solving an inverse problem that involves non-linearity and/or discontinuity in the forward model (e.g., DEM) is very challenging. Because of the potentially large computational cost for running the simulations, the samples have to be selected with an optimised strategy in order to ensure high efficiency. In this work, the iterative Bayesian filtering framework `GrainLearning` (GL) developed by Cheng et al. (2019b) is utilised to tackle these challenges.

In Bayesian calibration, the probability distribution of model parameters conditioned on given reference data (termed "posterior distribution") is estimated. This means that not only a perfect match between the reference and prediction data is sought but also the associated uncertainty. The simplest method for this estimation is the sequential MC filter (e.g. Särkkä, 2013). The iterative Bayesian filter implemented in GL allows for adaptive (re)sampling near potential posterior modes in parameter space, until the ensemble predictions of the model parameters converge. In the following, the iterative Bayesian filtering framework will be briefly introduced. Interested readers are referred to Cheng et al. (2019b) for the mathematical background and the numerical implementation.

2.2.1. *A non-linear non-Gaussian state-space model*

The forward (DEM) model can be framed within a non-linear, non-Gaussian state-space model with parameters $\boldsymbol{\Theta}$ to be estimated via statistical or Bayesian inference

$$\begin{aligned} \mathbf{x}_t &= \mathbb{F}(\mathbf{x}_{t-1}, \boldsymbol{\Theta}_{t-1}) + \boldsymbol{\nu}_t \\ \mathbf{y}_t &= \mathbb{H}(\mathbf{x}_t, \boldsymbol{\Theta}_t) + \boldsymbol{\omega}_t \\ \boldsymbol{\Theta}_t &= \boldsymbol{\Theta}_{t-1}. \end{aligned} \tag{2}$$

$\mathbb{F}$ can be understood as the process of obtaining the stochastic prediction of the current model state $\mathbf{x}_t$ from the forward model (e.g. DEM), based on the previous model state $\mathbf{x}_{t-1}$ and the parameter $\boldsymbol{\Theta}$ which is also treated as a stochastic variable. If all observations in $\mathbf{y}_t$ are independent and directly comparable to $\mathbf{x}_t$ without any unknown mapping, $\mathbb{H}$ is reduced to the identity matrix $\mathbf{I}_{N_{\mathrm{obs}}}$ withN_{obs} being the number of independent observations. The modelling errors and observation errors $\boldsymbol{\nu}_t$ and $\boldsymbol{\omega}_t$ are assumed to be Gaussian, i.e. that they follow the normal distributions $\mathcal{N}(0, \boldsymbol{\Sigma}_t^M)$ and $\mathcal{N}(0, \boldsymbol{\Sigma}_t^D)$, with a diagonal covariance structure for the sum of both errors $\boldsymbol{\Sigma}_t^M + \boldsymbol{\Sigma}_t^D$.

Eq. (2) is generic for any theoretical, numerical, or empirical models that are formulated to capture time or history dependency. In the context of the present work, the model parameters to be estimated are the contact parameters, i.e. $\boldsymbol{\Theta} = \{\log E, \nu, \phi, \beta_m, \eta_m\}$, and the macroscopic model state and observation vectors $\mathbf{x}_t$ and $\mathbf{y}_t$ at each time or loading step t consist of two independent entries: stress ratio σ_1/σ_c and volumetric strain ε_v, with σ_1 being the major principle stress and σ_c the confining pressure. Both $\mathbf{x}_t$ and $\mathbf{y}_t$ are vectors of size 2. The total covariance matrix $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_t^M + \boldsymbol{\Sigma}_t^D$ is assumed to be $\mathrm{diag}(\mathbf{s}^2)\mathbf{y}_t$ with $\mathbf{s}^2 = s\mathbf{w}^{\mathrm{s}}$ and $\det(\mathbf{w}^{\mathrm{s}}) = 1$, making the normalised covariance coefficient s a user-defined input parameter. When all variable measurements contribute equally to the inference, $\mathbf{w}^{\mathrm{s}}$ is reduced to the identity matrix $\mathbf{I}_{N_{\mathrm{obs}}}$.

2.2.2. *Bayesian filtering*

The aim of Bayesian filtering is to estimate the probability distribution of the augmented model state that consist of parameters (assuming they vary in time) and model state $\hat{\mathbf{x}}_t = (\mathbf{x}_t, \boldsymbol{\Theta}_t)$ conditioned on all available reference data $\mathbf{y}_{1:t}$ until time t. From this conditional probability (termed "posterior") distribution $p(\hat{\mathbf{x}}_t|\mathbf{y}_{1:t})$, the moments such as the means $\hat{\mathrm{E}}$ and variances $\widehat{\mathrm{Var}}$ are computed by

$$\begin{aligned} \hat{\mathrm{E}}[f_t(\hat{\mathbf{x}}_t)|\mathbf{y}_{1:t}] &= \sum_{i=1}^{N_p} w_t^{(i)} f_t(\hat{\mathbf{x}}_t^{(i)}) \\ \widehat{\mathrm{Var}}[f_t(\hat{\mathbf{x}}_t)|\mathbf{y}_{1:t}] &= \sum_{i=1}^{N_p} w_t^{(i)} (f_t(\hat{\mathbf{x}}_t^{(i)}) - \hat{\mathrm{E}}[f_t(\hat{\mathbf{x}}_t)|\mathbf{y}_{1:t}])^2. \end{aligned} \tag{3}$$

In Eq. (3), f_t describes an arbitrary quantity of interest as a function of $\hat{\mathbf{x}}_t$. In the context of DEM modelling, f_t can be any model prediction next to $\mathbf{x}_t = \{\sigma_1/\sigma_c, \varepsilon_v\}$, such as fabric anisotropy, coordination number. N_p is the number of samples drawn from a proposal distribution. Each sampled model state $\hat{\mathbf{x}}_t^{(i)}$ is associated with an importance weight $w_t^{(i)}$, that corrects the Monte Carlo approximation of the full posterior distribution $f_t(\hat{\mathbf{x}}_{0:t})|\mathbf{y}_{1:t}$ from the proposal distribution, using the recursive Bayes' rule. Note that the importance weights $\{w_t^{(i)}; i = 1, \ldots, N_p\}$ are normalised to sum to unity at each time step (cf. Eq. 13–15 in Cheng et al., 2019b).

2.2.3. An efficient multi-level sampling algorithm

Because the posterior distribution of the model parameters is unknown, a non-informative proposal distribution is used to uniformly explore the parameter space in iteration 0. The sampling is done with quasi-random numbers (defined by an initial guess of the parameter ranges), which leads to the so-called sequential quasi-MC filter (Gerber and Chopin, 2015). Although a uniform sampling is unbiased, it is very inefficient because the sampling density is the same everywhere in the parameter space; model evaluations are mostly performed in regions outside the optima. Furthermore, the importance weights for the sequential MC are ensured to degenerate as $t \to \infty$ (Särkkä, 2013) with a vanishing variance in the ensemble, leading to a loss of statistical information.

Weight degeneracy can be circumvented by reintroducing stochasticity in the state-space model (Eq. (2)). This is done by solving the inverse problem again, using resampled parameters $\boldsymbol{\Theta}_k^{(i)}$ and model states $\mathbf{x}_k^{(i)}$. For iteration $k \geq 1$, the resampled parameters are drawn from a more sensible proposal distribution, compared with that for iteration $k - 1$. This proposal distribution is constructed by a non-parametric Gaussian mixture model (Blei and Jordan, 2006) trained with the samples $\boldsymbol{\Theta}_{k-1}^{(i)}$ and the importance weights[2] $w_{T\,k-1}^{(i)}$ from iteration $k - 1$. This means the previous approximation of the posterior distribution $p_{k-1}(\boldsymbol{\Theta}|\mathbf{y}_{1:T})$ is utilised as the proposal distribution for the subsequent iteration. This leads to sampling densities that vary for different modes of the proposal distribution. The sampling densities are determined based on the Gaussian components and their contributions to the mixture (weights): the higher the likelihood, the higher the sampling density. Resampling from this iteratively updated proposal distribution allows zooming into highly probable parameter subspaces in very few iterations as the sampling density progressively increases around the posterior distribution modes. Note that because the number of samples is finite, if one Gaussian component (one mode of the proposal distribution) has a small weight, the sampling density for that mode could be small as well, which can potentially cause local optima to be overlooked after a few iterations.

2.2.4. Numerical aspects

The iterative Bayesian framework contains very few user-defined parameters. The normalised covariance coefficient s (see Section 2.2.1) is chosen such that the effective sample size $ESS = 1/(\sum_1^{N_p}(w_t^{(i)})^2 N_p)$ is larger than $ESS = 30\%$ (Särkkä, 2013; Ruiz and Kappen, 2017). The goal is to have a sufficient number of effective samples for the training of the proposal distribution. The other important parameters are hyperparameters that optimise the number of Gaussian components and mixing proportions. For example, the upper bound of the number of Gaussian components K_{max} is chosen as $N_p/10$, assuming that ten samples are sufficient to estimate one Gaussian component. The Gaussian components can share the same covariance matrix or have different ones. The open-source package `scikit-learn` (Pedregosa et al., 2011) provides the non-parametric Gaussian mixture class and the mean-field variational inference of Blei and Jordan (2006) for the training.

According to Gerber and Chopin (2015), the minimum required sample size N_p scales with the natural logarithm of the number of unknowns n for sequential quasi MC filtering:

$$N_p = \alpha\, n \log(n) \tag{4}$$

where α is a scalar factor to be chosen a priori by the user. From a stochastic point of view, the scaling with $n \log n$ is based on the assumption of n independent unknown model parameters.

2.2.5. Error definitions

For the evaluation of the quality of the individual samples, the absolute error (in percent) between the DEM predictions and the reference data is evaluated in an integral form. The observation sample error $\epsilon^{(i)}(j)$ for sample i and observation j, with $j = 1, \ldots, N_{\text{obs}}$, is defined as

$$\epsilon^{(i)}(j) = \frac{\sum_{t=0}^{T} \|y_t(j) - x_t^{(i)}(j)\|}{\sum_{t=0}^{T} \|y_t(j)\|} \cdot 100\%. \tag{5}$$

As indicated in Section 2.2.1, t in Eq. (5) denotes the t–th data point of observation $y_t(j)$ in time or load history. The associated DEM prediction is denoted by $x_t^{(i)}(j)$. Based on the assumed relative contributions of the individual observations to the inference $\text{diag}(\tilde{\mathbf{w}}^s) = \{\tilde{w}_1^s, \ldots, \tilde{w}_{N_{obs}}^s\}$, the combined sample error for sample i is defined by

$$\tilde{\epsilon}^{(i)} = \sum_{j=1}^{N_{\text{obs}}} \left[\frac{1/\tilde{\mathbf{w}}^s(j,j)}{\sum_{j=1}^{N_{\text{obs}}} 1/\tilde{\mathbf{w}}^s(j,j)} \right] \epsilon^{(i)}(j). \tag{6}$$

For the calculation of the covariance matrix $\boldsymbol{\Sigma}_t$, $\tilde{\mathbf{w}}^s$ is multiplied by $c = \det(\tilde{\mathbf{w}}^s)^{-\frac{1}{N_{obs}}}$ to obtain $\mathbf{w}^{\mathbf{s}}$ such that $\det(\mathbf{w}^{\mathbf{s}}) = 1$. Following Eq. (3), the sum of the combined sample errors $\tilde{\epsilon}^{(i)}$ weighted by the associated importance weights, i.e. weights $w_T^{(i)}$ at the end of the calibration, is used as the measure for the absolute GL error (in percent), namely

$$\hat{\epsilon}_{\text{GL}} = \sum_{i=1}^{N_p} w_T^{(i)} \tilde{\epsilon}^{(i)}. \tag{7}$$

2.2.6. Termination of GL and final sample selection

A stochastic termination criterion in terms of the normalised covariance coefficient s is used and GL is terminated as soon as $s < 0.01$. The outcome of GL in the final iteration are N_p parameter samples with associated probabilities. Eq. (3) gives the mathematical expression for the final set of calibrated parameters when the probability distribution is unimodal. However, the resulting posterior distribution of micromechanical parameters is usually multimodal for DEM models, as will be discussed in Section 3.4. Although it would be stochastically more meaningful to follow Eq. (3) and compute the weighted sum of individual samples, the sample with the highest probability is selected as the final set of calibrated parameters, neglecting those that have smaller weights but still represent other modes or peaks of the distribution. This is because, if the posterior distribution is multimodal, which is often the case for the rolling resistance parameters, a weighted sum could deviate far from each of the peaks. Therefore, the sample with the highest probability in the final iteration is chosen.

2.3. Grainlearning workflow

In the following, the GL workflow is explained on the basis of Fig. 2 for the considered applications presented in this paper. The process refers to the stochastic definitions introduced in Section 2.2 to allow for a smooth transition from theory to application. Overall, the GL workflow can be divided into three steps: initialisation, iterative Bayesian filtering (GL iterations) and postprocessing including sample selection.

During initialisation, the number of samples N_p is defined from the proportionality factor α and the number of unknown parameters n via Eq. (4). The observation vector $\mathbf{y}_{1:T}$ for GL consists of two time series ($N_{\text{obs}} = 2$): one time series of the stress ratio σ_1/σ_c and one of the volumetric strain ε_v. The combined sample errors (Eq. (5)) in stress ratio, $\epsilon_{\text{stress}}^{(i)}$, and in volumetric strain, $\epsilon_{\text{strain}}^{(i)}$ are computed and $\text{diag}(\tilde{\mathbf{w}}^s) = \{\tilde{w}_{\text{stress}}^s, \tilde{w}_{\text{strain}}^s\}$ is set based on the assumed partial contribution of σ_1/σ_c and ε_v to the covariance matrix $\boldsymbol{\Sigma}_t$ in Section 2.2.1.

---

[2] Note, the importance weights updated until the last time step are used for training the Gaussian mixture model.

The final step during initialisation is the generation of N_p parameter samples using a Halton sequence based on an initial guess for the parameter ranges.

The first step of an iteration is the execution of model evaluations with respect to the corresponding N_p samples. In the underlying case, these are N_p triaxial compression simulations using `Yade`, where each model evaluation is randomised by parameter sample (i) that consists of five micro-mechanical parameters $\boldsymbol{\Theta}^{(i)} = \{\log E^{(i)}, \nu^{(i)}, \phi^{(i)}, \beta_m^{(i)}, \eta_m^{(i)}\}$. The outputs of these model evaluations are N_p state vectors $\{\mathbf{x}^{(i)}; i = 1, \ldots, N_p\}$ of model predictions for σ_1/σ_c and ε_v, resulting from the parameter samples. To prevent ambiguities regarding commonly used terminology in engineering, the term "model prediction" is further replaced by "model response". Next, the GL quantities are computed on the basis of the model responses and the reference data. They include the computations of errors, the normalised covariance coefficient and the importance weights $w_T^{(i)}$ at the end of the iteration.

Subsequently, the termination criteria (cf. Section 2.2.6) is evaluated and GL is terminated if $s < 0.01$. When GL is terminated, the sample with the highest probability, i.e. $\max w_T^{(i)}$, in the last iteration is taken as GL result and postprocessing can be performed. Thus, the corresponding parameters of this sample represent the assumed final set of calibrated micro-mechanical parameters. If the threshold of s is not matched, new samples will be drawn from the Gaussian mixture model, cf. Section 2.2.3, that is trained with the previous samples and weights. A new iteration starts thereafter, with the posterior distribution of the parameters to be updated again from $t = 0$ to T.

3. Calibration of micro-mechanical parameters of Toyoura sand

3.1. Simulation set-up

The iterative Bayesian filter, from now on denoted GL, is utilised to calibrate the micro-mechanical parameters of Toyoura sand. The same triaxial compression test with 1000 particles as presented in Cheng et al. (2018) is used for this purpose. Eq. (4) is utilised to determine the number of samples per iteration in GL using $\alpha = 10$. Thus, the sample size applied for the calibrations is $N_p = 80$. The GL results are not only compared against the experimental results by Sun et al. (2007), but also against the results obtained by Cheng et al. (2018) using the sequential quasi-Monte Carlo filter, from now on denoted MC.

As described in Section 2.3, the observation vector for GL consists of the stress ratio σ_1/σ_c and the volumetric strain ε_v. Both measurements are assumed to contribute equally to the covariance matrix, i.e. $w^s_{\text{stress}} = w^s_{\text{strain}} = 0.5$. The predefined parameter ranges used in GL are the same as those used in the MC by Cheng et al. (2018) (Table 2).

The DEM simulations are strain-controlled and conducted in a quasi-static manner: the packing is loaded with a maximum strain rate of $0.1\,\text{s}^{-1}$, followed by a relaxation stage. When the average unbalanced force is lower than 10^{-3}, a subsequent loading continues until the next target axial strain is reached. The strain-controlled procedure to conduct quasi-static simulations ensures that the model state $\mathbf{x}$ is obtained at the exact same strain level as the observation $\mathbf{y}$ and the computational cost is minimised per model evaluation.

A variable time step is used by applying `Yade`'s `GlobalStiffnessTimeStepper` with a time step safety coefficient of 0.8. Thus, the time step is automatically calculated based on the micro-mechanical parameters and the model configuration and, hence, will vary between simulations. All simulations are performed on a high-performance computing cluster with Intel Xeon Gold 6150 (2.70 GHz) processors. For the sake of repeatability each simulation is run on a single thread.

Table 2
Means and standard deviations of the micro-mechanical parameters used in the calibration process (after Cheng et al., 2018).

	$\log E$ [GPa]	ν [–]	ϕ [°]	β_m [N m rad^{-1}]	η_m [–]
Mean	9	0.25	30	0.5	0.5
Std. dev.	2	0.25	30	0.5	0.5

3.2. Comparison between GL, MC and experimental data

The sets of micro-mechanical parameters estimated by the calibration procedures for varying confining pressures σ_c are listed in Table 3. Values estimated with both, GL and MC, are provided in the same table for a better comparison. The values in parenthesis are those obtained with MC by Cheng et al. (2018).

Comparing the MC and GL solutions reveal a similar trend for $\log E$ and ϕ with respect to the applied confining pressures. Generally, the higher the pressure, the higher $\log E$ and the lower ϕ. Exceptions of these trends are only observable for the MC results, where the resulting micro-mechanical parameters are exactly the same for $\sigma_c = 0.2\,\text{MPa}$ and $\sigma_c = 0.5\,\text{MPa}$. The stiffness ratio k_s/k_n, denoted as ν, shows an opposite trend with increasing pressure. The reason for the variation of $\log E$, ν and ϕ as functions of σ_c can be attributed to mild-to-moderate grain crushing, which leads to the change of grain morphology (e.g., surface roughness and asperity), as the confining pressure increases. Nevertheless, the isotropic and deviatoric loads are not large enough to cause an evolution of the particle size distribution as in Wu et al. (2020). Therefore, non-crushable DEM particles with the same particle size distribution still suffice for the confining pressures considered.

No clear trend can be observed for the rolling resistance parameters. Moreover, when comparing the same calibrations for the two methods, the calibration with $\sigma_c = 0.5\,\text{MPa}$ (cf. Table 3) stands out. Even though $\log E$ and ϕ vary only slightly between GL and MC, the remaining three parameters vary significantly. For MC $\nu = 0.080$, $\beta_m = 0.564\,\text{N m rad}^{-1}$ and $\eta_m = 0.194$ are obtained, whereas for GL $\nu = 0.161$, $\beta_m = 0.071\,\text{N m rad}^{-1}$ and $\eta_m = 0.885$ are obtained. It is assumed that two model evaluations with small variations in micro-mechanical parameters are related to the same solution (local minimum). In contrast, two evaluations with significant parameter variations indicate that the solutions do not belong to the same local minimum. This indicates that different solutions are found using GL.

For a further qualitative evaluation of the results, the model responses based on the most probable identified parameters are plotted against the reference data. Thus, the variation of the stress ratio σ_1/σ_c and volumetric strain ε_v are plotted against the strain in the normal direction ε_1 (Fig. 3). Comparing the model responses of GL and MC with the experimental data clearly shows the superior behaviour of GL over MC. All curves related to GL fit the experimental data better than the curves associated with the MC calibration. For the previously mentioned calibrations with $\sigma_c = 0.5\,\text{MPa}$ (Fig. 3), where the calibrated micro-mechanical parameters considerably differ between MC and GL, the fit in stress ratio and volumetric strain is significantly better for the GL calibration. Because of the marked difference in the associated curves, the assumption that a different and better solution is found using GL is confirmed.

To confirm the qualitative results, a quantitative evaluation is performed. The observation sample errors in stress ratio, ϵ_{stress}, and in volumetric strain, ϵ_{strain}, are evaluated using Eq. (5). In Fig. 4, ϵ_{stress} is plotted against ϵ_{strain}, where arrows are used to indicate the difference between MC and GL results. The heads are related to the GL observation sample errors and the tails to the MC observation sample errors. Thus, the arrow lengths and directions indicate the differences between the two approaches. It can be seen that ϵ_{stress} and ϵ_{strain} are always smaller for the GL calibrations. ϵ_{stress} ranges from 2.75 to 7.53% and ϵ_{strain} from 4.77 to 21.41% for the MC calibrations. The related error ranges are with $\epsilon_{\text{stress}} \in [1.65\%, 2.18\%]$ and $\epsilon_{\text{strain}} \in [3.48\%, 7.91\%]$ significantly smaller

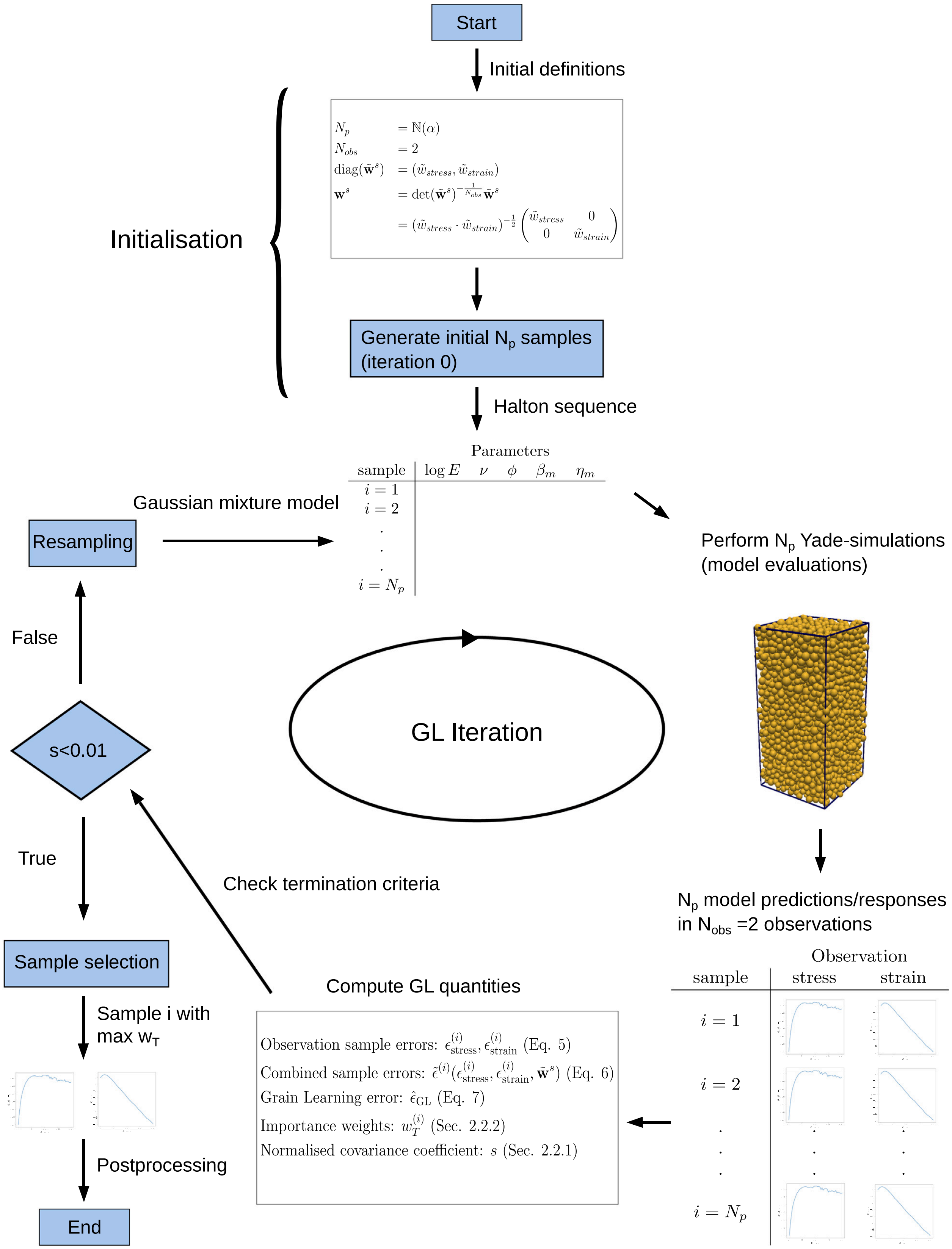
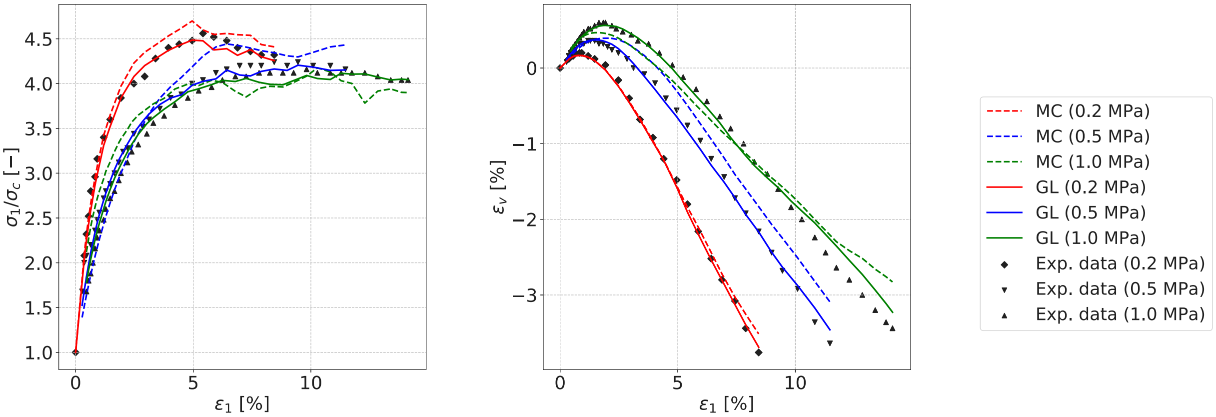


Fig. 2. GrainLearning workflow.

Table 3
Micro-mechanical parameters obtained with GL and MC using 1000 particles. The values in parentheses correspond to those obtained with MC (see Cheng et al., 2018).

σ_c [MPa]	$\log E$ [GPa]	ν [–]	ϕ [°]	β_m $[\mathrm{N\,m\,rad^{-1}}]$	η_m [–]
0.2	8.722 (8.736)	0.078 (0.080)	31.263 (30.203)	0.436 (0.564)	0.178 (0.194)
0.5	8.782 (8.736)	0.161 (0.080)	30.626 (30.203)	0.071 (0.564)	0.885 (0.194)
1.0	8.961 (9.098)	0.241 (0.439)	25.441 (22.043)	0.086 (0.076)	0.944 (0.726)



(a) (b)

Fig. 3. Comparison of experimental data (Exp. data) by Sun et al. (2007) with GL and MC model responses in (a) stress ratio and (b) volumetric strain for simulations with 1000 particles and varying confining pressures.

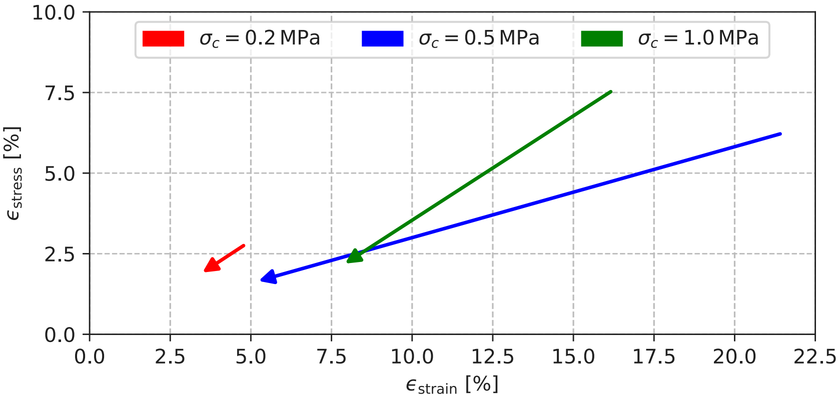


Fig. 4. Error in stress ratio $\epsilon_{\mathrm{stress}}$ against error in volumetric strain $\epsilon_{\mathrm{strain}}$ for simulations with 1000 particles and varying confining pressures. Arrow heads are related to GL errors and tails to MC errors.

for GL. The biggest improvement in errors is obtained for the confining pressure of $\sigma_c = 0.5\,\mathrm{MPa}$ (blue arrow in Fig. 4) with a reduction from 6.22 to 1.65% in $\epsilon_{\mathrm{stress}}$ and from 21.41 to 5.22% in $\epsilon_{\mathrm{strain}}$. Thus, the better performance of GL is also quantitatively confirmed.

3.3. Convergence

3.3.1. Evolution of GL error over number of GL iterations

The convergence properties of the GL calibration framework are investigated by calculating the GL error $\hat{\epsilon}_{\mathrm{GL}}$ defined in Eq. (7) by taking into account the weighted mean of the combined sample errors $\tilde{\epsilon}$ from all samples within an iteration.

The evolution of $\hat{\epsilon}_{\mathrm{GL}}$ over the number of GL iterations for the previous simulations using 1000 particles is depicted in Fig. 5. It can clearly be seen that $\hat{\epsilon}_{\mathrm{GL}}$ decreases with an increasing number of iterations. This oscillation-free behaviour is desired since it indicates a good convergence behaviour of GL. Convergence is reached between iteration 6 and 10 with $\hat{\epsilon}_{\mathrm{GL}}$ between 3.58 and 5.98%.

3.3.2. Influence of number of particles

In DEM simulation the particle size distribution is generally scaled up and sometimes truncated. This is necessary to keep computational costs at an acceptable level. The resulting number of particles is generally less compared to the physical one. Hence, the influence of the number of particles on the GL calibration framework is investigated next. Considered is the confining pressure of $\sigma_c = 0.5\,\mathrm{MPa}$. GL is performed for 1000, 2000, 5000, 8000, 10000 and 27000 particles. Since the confidence in the volumetric strain reference data is significantly smaller than in the data of the stress ratio (cf. Section 3.2), their relative contributions to the covariance matrix are adjusted to 0.2 for the stress ratio and 0.8 for the volumetric strain. This leads to a higher stochastic weighting on the stress ratio.

The estimated micro-mechanical parameters are listed in Table 4. No clear trend can be seen for any of the micro-mechanical parameter with respect to the number of particles. The ranges are relatively wide with $\log E \in [8.683\,\mathrm{GPa}, 8.973\,\mathrm{GPa}]$, $\nu \in [0.079, 0.296]$, $\beta_m \in [0.036\,\mathrm{N\,m\,rad^{-1}}, 0.352\,\mathrm{N\,m\,rad^{-1}}]$, $\eta_m \in [0.284, 0.576]$ and $\phi \in [24.923\,°, 34.327\,°]$. Nevertheless, the curves in Fig. 6 show all a very good agreement with the experimental data.

Table 4
Identified micro-mechanical parameters and corresponding errors for GL with varying number of particles.

	Identified parameters					Errors [%]		
Particles	$\log E$ [GPa]	ν [−]	ϕ [°]	β_m [N m rad^{-1}]	η_m [−]	ϵ_{stress}	ϵ_{strain}	$\tilde{\epsilon}$
1000	8.753	0.152	34.327	0.065	0.324	3.39	7.58	4.23
2000	8.727	0.226	28.479	0.076	0.576	3.11	6.62	3.81
5000	8.973	0.079	24.923	0.352	0.284	2.41	11.56	4.24
8000	8.812	0.131	29.733	0.095	0.707	2.01	6.82	2.97
10000	8.800	0.126	26.960	0.153	0.409	2.51	5.82	3.17
27000	8.683	0.296	33.055	0.036	0.567	2.69	10.65	4.28

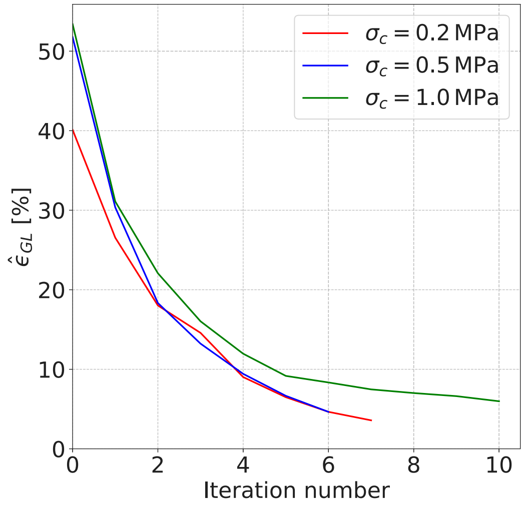


Fig. 5. Evolution of GL error $\hat{\epsilon}_{\mathrm{GL}}$ over number of iterations for simulations with 1000 particles and varying confining pressures.

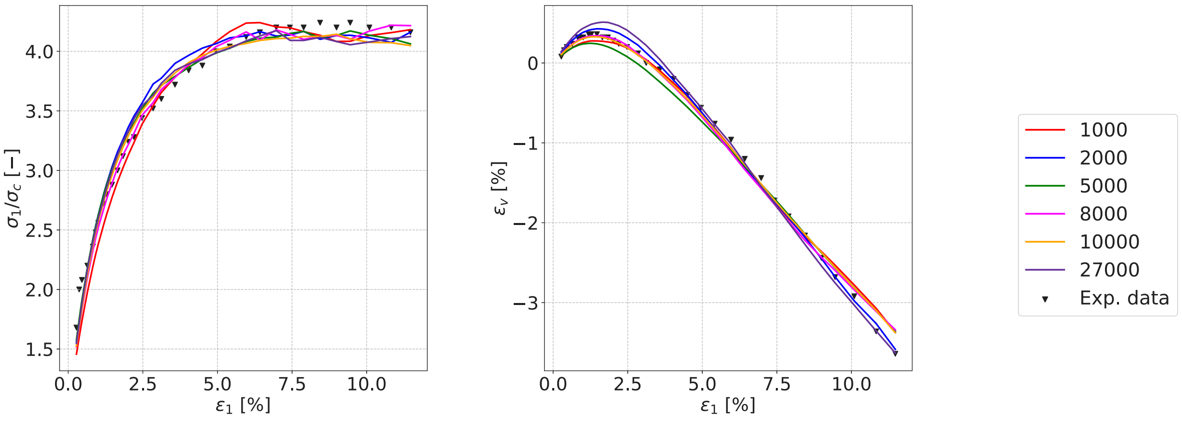
ϵ_{stress} and ϵ_{strain} as well as the combined sample error $\tilde{\epsilon}$ are considered for a quantitative evaluation (Table 4). The smallest ϵ_{stress} of 2.01% is obtained for 8000 particles and the smallest ϵ_{strain} of 5.82% for 10000 particles. In contrast, the highest $\epsilon_{\text{stress}} = 3.39\%$ and $\epsilon_{\text{strain}} = 11.56\%$ are obtained for 1000 and 5000 particles, respectively.

Via Eq. (6), the combined sample error $\tilde{\epsilon}$ can be computed with $\tilde{\mathbf{w}}^s = \{0.2, 0.8\}$. In ascending order of the number of particles $\tilde{\epsilon}$ is 4.23, 3.81, 4.24, 2.97, 3.17 and 4.28%. Thus, increasing the number of particles does not automatically lead to a decreasing error, i.e. better model responses. It seems that $\tilde{\epsilon}$ does not converge with an increasing number of particles. Different solutions, i.e. combinations of stochastically dependent micro-mechanical parameters, are obtained.

3.4. Multiple solutions

The advantage of training a Gaussian mixture model with the updated statistics and using it for resampling (i.e. selecting the parameters for the next iteration) is that the correlation structure among the parameters is used. In the following, the performance of the GL calibration with 2000 particles and $\sigma_c = 0.5$ MPa is investigated in more detail. The resampled micro-mechanical parameters in five-dimensional space are visualised by ten two-dimensional planes (Fig. 7). The converging trends are clearly indicated by the clustering of the resampled parameters, as the number of iterations increases. GL needs a total of 12 iterations (iteration 0 to iteration 11) in order to reach below the target error in the ensemble prediction. Fig. 7 shows the resampled parameters of iterations 0 to 6. The parameter values for iteration 0 are uniformly sampled from quasi-random numbers between the predefined ranges. At iteration 5 an excellent agreement is already achieved between one numerical data set and the experimental data, but the distribution of the parameter samples still shows multi-modal features, resulting in a rather large variance in the ensemble predictions. Fig. 8 shows the stress–strain responses obtained with the top ten most probable parameter samples. Table 5 summarises the correlation coefficients of the samples for iteration 5. They are based on the scalar Pearson product-moment correlation coefficient (Rodgers and Nicewander, 1988), which is a measurement for the linear relation between two variables. The coefficient is defined between $-1 \le \rho \le 1$, whereby 1 corresponds to a perfect positive linear relation and -1 to a perfect negative relation. Absolute values around 0.4 are considered as moderate linear relation and values around 0.8 as strong linear relation. The `NumPy` library (Oliphant, 2006) is used for the computation of the Pearson coefficients. Note, the coefficients after iteration 5 are not given because the samples are already localised to one of the clusters and the correlation structures are no longer visible. In the following, multiple solutions and statistical correlations between the parameters will be discussed based on the samples of iteration 5.

Fig. 7 and Table 5 show that the inter-particle Young's modulus E is moderately to strongly correlated with the parameters relevant to



(a)

(b)

Fig. 6. Comparison of experimental data (Exp. data) by Sun et al. (2007) with GL model responses in (a) stress ratio and (b) volumetric strain for varying number of particles.

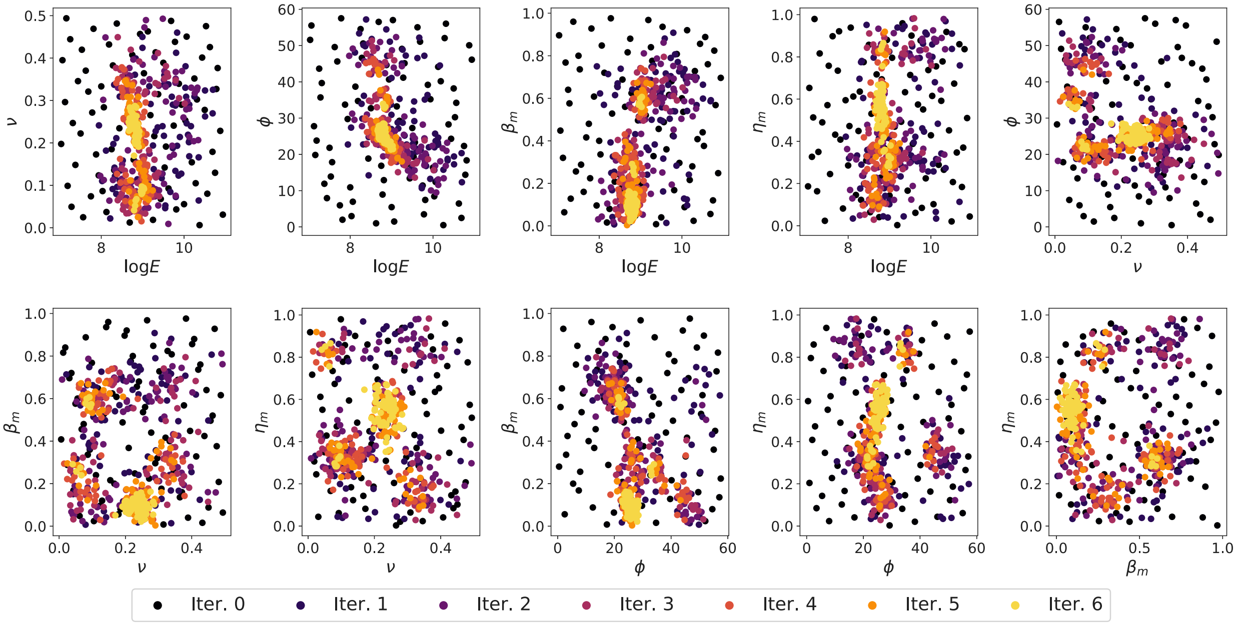
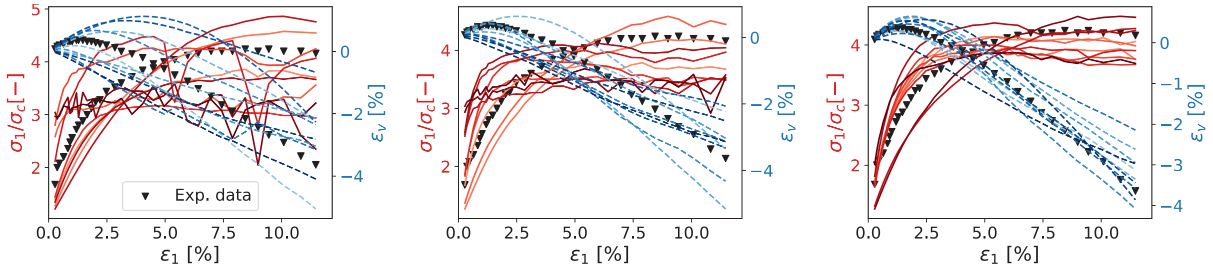
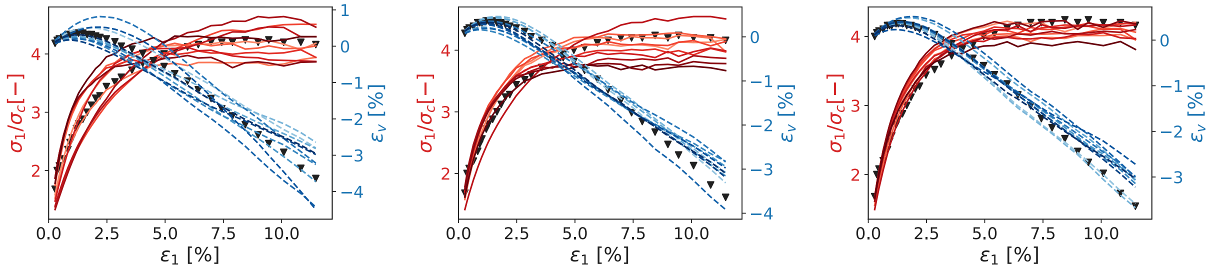


Fig. 7. Resampled micro-mechanical parameters over consecutive iterations. The proposal distribution, which the samples are drawn from, is clearly multi-modal.



(a) Iteration 0   (b) Iteration 1   (c) Iteration 2



(d) Iteration 3   (e) Iteration 4   (f) Iteration 5

Fig. 8. Stress–strain responses obtained with the top ten most probable parameter samples for iterations 0 to 5.

Table 5
Correlation coefficients between the five micro-mechanical parameters obtained at iteration 5.

	$\log E$	ν	ϕ	β_m	η_m
$\log E$	1.000	−0.676	−0.479	0.660	−0.133
ν		1.000	−0.173	−0.599	−0.231
ϕ			1.000	−0.332	0.392
β_m				1.000	−0.483
η_m	sym.				1.000

tangential and rolling stiffnesses (ν and β_m), with $|\rho| = 0.676$ and 0.66. E is almost independent of the rolling friction coefficient η_m ($|\rho| = 0.133$) and no multiple optima seem to exist in the $\log E - \eta_m$ plane. ν and ϕ seem to be independent of each other at iteration 5 ($|\rho| = 0.173$), although some clusters exist at $\phi \approx 35°$. The correlation coefficients become unreliable when several isolated clusters of samples emerge from the resampling algorithm. This happens most significantly to the rolling parameters β_m and η_m. The samples in the $\beta_m - \eta_m$ plane show a curved, non-monotonic trend along which all combinations of β_m and η_m can potentially be a good solution. Although multiple solutions exist for β_m and η_m, they appear to show moderate correlations with the inter-particle friction angle ϕ, with $|\rho| = 0.332$ and 0.392. E and ϕ seems to

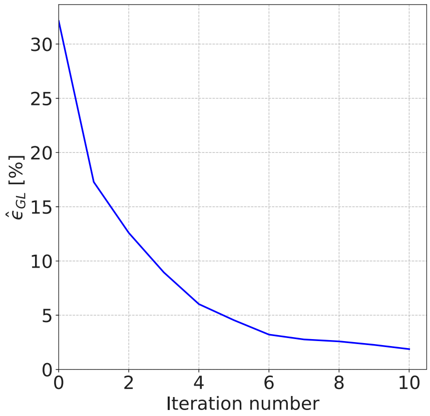


Fig. 9. GL error $\hat{\epsilon}_{\mathrm{GL}}$ over iteration number using synthetic reference data.

be moderately correlated as well, particularly in iteration 5, with $|\rho| = 0.479$. As shown in Fig. 8(f), at iteration 5 the accuracy is acceptable for a standard calibration. The negative outcome of continuing the iterations (after iteration 5) is that some local optima will tend to be overlooked as the iterations continue, even though they are identified in several previous iterations.

4. GL using synthetic reference data

As shown in Section 3, the outcome of a GL calibration depends on simulation parameters (e.g. boundary conditions, number of particles) as well as on calibration specific parameters. Thus, it is possible that for different parameter combinations the model responses are accurate to a similar degree. This opens a question: which combination is the correct one? For a proper parametric study of GL-specific parameters, it is beneficial to compare the model responses against an analytical solution. Since there is no analytical solution for the problem at hand, synthetic data generated using a DEM simulation can be used. The benefit of this approach is that not only the confidence in model responses is increased, but the micro-mechanical parameters to be calibrated are known a priori.

4.1. Numerical set-up and proof of concept

As a proof of concept, synthetic reference data is generated using a DEM simulation with 2000 particles. The choice for 2000 particles is arbitrary and based on the main objective to investigate the performance of GL and not to perform detailed DEM analyses. Using 2000 particles allows keeping the computational costs of the following study at an acceptable level.

A similar simulation set-up as in Section 3.1 is used for the synthetic data generation and the subsequent GL calibrations. Thus, similar to the previous calibrations, all simulations start with the exact same prepared initial packing and hence the same initial porosity. The same linear contact law with a confining pressure of $\sigma_c = 0.5\,\mathrm{MPa}$ is considered. The packings are loaded incrementally for every 0.4% axial strain, followed by a relaxation stage (see Section 3.1), until the total axial strain reaches 20%. Note, that the considered strain window (i.e. the history) has a direct impact on the calibration results. The longer the history, in this case, the history from zero to maximum axial strain, the more information is available for calibrating the parameters and, hence, the more representative the calibrated parameters will be. All micro-mechanical parameter samples from the last iteration step of the GL result for 2000 particles are used to define the target synthetic micro-mechanical parameters. They are defined as the weighted sum of all parameters of the final samples with associated posterior probabilities. If not explicitly stated otherwise, the predefined parameter ranges indicated in Table 2 and a sample size based on $\alpha = 10$ (cf. Eq. (4)) are used. The synthetic micro-mechanical parameters for the synthetic reference data generation are listed in Table 6.

The convergence of the GL error $\hat{\epsilon}_{\mathrm{GL}}$ over the number of iterations is plotted in Fig. 9. It can be seen that $\hat{\epsilon}_{\mathrm{GL}}$ converges to 1.88% at iteration 10. A significant reduction in $\hat{\epsilon}_{\mathrm{GL}}$ is observed within the first six iterations, with a reduced convergence rate afterwards. This indicates that a good fit has already been found after iteration 6 and slightly better solutions close to this fit are found in the subsequent iterations. This is in line with the findings of Section 3.4, where an acceptable fit is found after iteration 5.

The calibrated synthetic micro-mechanical parameters are listed in Table 6. Comparing the calibrated with the synthetic micro-mechanical parameters reveals that the parameters found by GL are very close to the parameters of the synthetic data. The differences in $\log E$ and ϕ are within 0.15 and 1.53%, respectively. The differences in the remaining three parameters are slightly more and can be attributed to the randomness in the microstructure.

Fig. 10 shows the model responses and the synthetic data. An excellent agreement can be observed for both stress ratio and volumetric strain. This can also be seen in the observation sample errors $\epsilon_{\mathrm{stress}}$ and $\epsilon_{\mathrm{strain}}$ (Table 6). They are 1.06 and 1.20%, respectively, which again highlights the excellent fit. Moreover, the combined sample error $\tilde{\epsilon}$ is with 1.13% smaller than all $\tilde{\epsilon}$ obtained for the calibrations using experimental data (cf. Section 3).

4.2. Combined influence of predefined parameter ranges and number of samples

For the investigation of the influence of the predefined parameter range on the performance of GL, only the range for $\log E$ is varied according to Table 7 and the default ranges as indicated in Table 2 are used for the remaining parameters. The reason why only $\log E$ is taken into account is that the default ranges of β_m and η_m already cover their full range and with $\nu \in [0, 0.5]$ and $\phi \in [0^\circ, 60^\circ]$ the majority of associated possible ranges is covered. In contrast, the possible range for E is considerably wider and can vary in magnitude depending on the material.

GL is performed for the given ranges for $\alpha = 5$, 10 and 20 in order to investigate the combined influence of the predefined parameter range with the number of samples. The goal is to investigate if there is a dependence of the predefined parameter ranges on the specific number of required samples.

Fig. 11 shows the evolution of $\hat{\epsilon}_{\mathrm{GL}}$ over the iteration number for all combinations. It should be noted that the results for the wide range with $\alpha = 5$ are truncated and that GL was terminated after 100 iterations. Even though $\hat{\epsilon}_{\mathrm{GL}}$ converges around 10.00%, the applied termination criteria are never fulfilled. Moreover, $\hat{\epsilon}_{\mathrm{GL}}$ is significantly higher in comparison to all other combinations. Thus, using $\alpha = 5$ and the wide range leads to an unsuccessful GL calibration. Similarly, GL is not immediately terminated for the mid range with $\alpha = 5$ and $\alpha = 20$, even though there are no noticeable changes in $\hat{\epsilon}_{\mathrm{GL}}$ after iteration 7 and it takes additional iterations before the GL termination criteria are fulfilled. It seems that the performance of GL depends on the predefined parameter ranges and the number of samples. Nevertheless, most of the combinations converge within less than 10 iterations.

The associated calibrated micro-mechanical parameters and the corresponding errors $\epsilon_{\mathrm{stress}}$, $\epsilon_{\mathrm{strain}}$ and $\tilde{\epsilon}$ are listed in Table 8. In addition, Fig. 12 depicts the associated model responses compared with the synthetic data. Generally a very good agreement can be observed between model responses and synthetic data. The combinations with the smallest sample size ($\alpha = 5$) and the small and wide ranges deviate the most. This is also confirmed by the errors listed in Table 8.

In the following, the combined sample errors $\tilde{\epsilon}$ are used to evaluate and compare the GL model responses. The errors are between 1.13 and 6.36%. Five out of nine errors are below 2% and the associated

Table 6
Synthetic and calibrated micro-mechanical parameters.

	Identified parameters					Errors [%]		
	$\log E$ [GPa]	ν [−]	ϕ [°]	β_m $\left[\mathrm{N\,m\,rad^{-1}}\right]$	η_m [−]	$\epsilon_{\mathrm{stress}}$	$\epsilon_{\mathrm{strain}}$	$\tilde{\epsilon}$
Synthetic data	8.720	0.227	28.099	0.079	0.569	–	–	–
Calibrated	8.733	0.186	28.530	0.095	0.797	1.06	1.20	1.13

Table 7
Ranges for $\log E$ used in GL for studying the influence of the predefined parameter range.

	Small range	Mid range	Wide range
$\log E$ [GPa]	8–10	7–11	6–12

calibration results are classified as optimal. Independent of the range, $\tilde{\epsilon}$ is always higher than 2% for $\alpha = 5$. Thus, a value of $\alpha = 5$ turns out to be sub-optimal. A reason behind this is that for $\alpha = 5$ the highest $\log E$ values are obtained, all overestimating the synthetic solution.

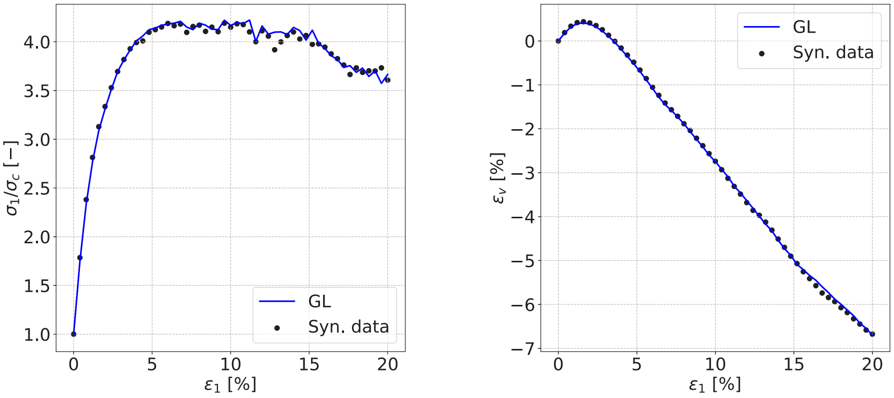
The remaining GL calibrations with $\tilde{\epsilon}$ higher than 2% are for the wide range with $\alpha = 10$, resulting in $\tilde{\epsilon} = 2.63\%$. In contrast, the error is 1.63% for the small range and 1.13% for the mid range. Since $\tilde{\epsilon}$ is only 1.28% for the wide range and $\alpha = 20$, the number of samples with $\alpha = 10$ is not sufficient for the wide range. Furthermore, the results indicate that there is a dependency between the number of samples and the predefined parameter ranges. The wider the ranges, the more samples should be used. Additionally, the error for the small ranges is always higher than for the associated mid ranges. The reason behind this is that, for small ranges, $\log E$ tends to be overestimated and, hence, higher than for the mid ranges.

The most important finding is that $\alpha = 10$ is sufficient when the range of predefined inter-particle Young's modulus is not excessively wide. The wider the range, the more samples are required for a successful calibration. To confirm this hypothesis, two additional GL calibrations are performed. Considered is an extra wide range of $\log E \in [5\,\mathrm{GPa}, 13\,\mathrm{GPa}]$ with $\alpha = 20$ and $\alpha = 30$. Based on the previous results, a better GL result for $\alpha = 30$ is expected. Table 9 summarises the final estimated micro-mechanical parameters. It should be noted that GL was terminated manually after step 100 for $\alpha = 20$. The resulting micro-mechanical parameters differ clearly from each other, whereby the parameters for $\alpha = 30$ are overall closer to the synthetic micro-mechanical parameters (cf. Table 6).

Fig. 13 compares the stress–strain responses with the synthetic data. An excellent fit can be observed for $\alpha = 30$, whereas the GL model response for $\alpha = 20$ only fits the stress ratio reasonably well. The combined sample error $\tilde{\epsilon}$ is 8.52% for $\alpha = 20$, indicating that the terminated GL model responses are not very accurate. On the contrary, GL converges for $\alpha = 30$ with $\tilde{\epsilon} = 1.28\%$ (Table 9).

4.3. Prediction for uncalibrated stress paths

In a next step, the synthetic micro-mechanical parameters and the calibrated parameters obtained for the proof-of-concept (Table 6) are used for the sake of validation. Therefore, model responses for the uncalibrated stress paths with confining pressures $\boldsymbol{\sigma}_c = 0.2\,\mathrm{MPa}$ and $1.0\,\mathrm{MPa}$ are considered. Note that in contrast to the experimental reference data in Section 3, particle crushing or change of grain morphology are irrelevant (i.e. spheres will stay spheres) and the previously mentioned correlations of the micro-mechanical parameters and the applied confining pressure are non-existing.



(a) (b)

Fig. 10. GL Model responses in (a) stress ratio and (b) volumetric strain compared with synthetic reference data.

Table 8
Identified micro-mechanical parameters for all combinations and corresponding errors.

Combination		Identified parameters					Errors [%]		
Range	α	$\log E$ [GPa]	ν [−]	ϕ [°]	β_m $\left[\mathrm{N\,m\,rad^{-1}}\right]$	η_m [−]	$\epsilon_{\mathrm{stress}}$	$\epsilon_{\mathrm{strain}}$	$\tilde{\epsilon}$
Small	5	8.950	0.063	26.937	0.466	0.217	1.45	6.80	4.13
Mid	5	8.805	0.184	26.406	0.112	0.554	1.76	2.88	2.32
Wide	5	9.090	0.381	22.721	0.068	0.432	4.19	8.52	6.36
Small	10	8.785	0.122	27.285	0.222	0.236	1.08	2.18	1.63
Mid	10	8.733	0.186	28.530	0.095	0.797	1.06	1.20	1.13
Wide	10	8.736	0.300	26.625	0.078	0.767	2.71	2.55	2.63
Small	20	8.778	0.107	26.736	0.741	0.175	1.82	1.94	1.88
Mid	20	8.706	0.286	27.928	0.069	0.493	1.44	1.33	1.38
Wide	20	8.715	0.198	27.635	0.115	0.260	0.95	1.61	1.28

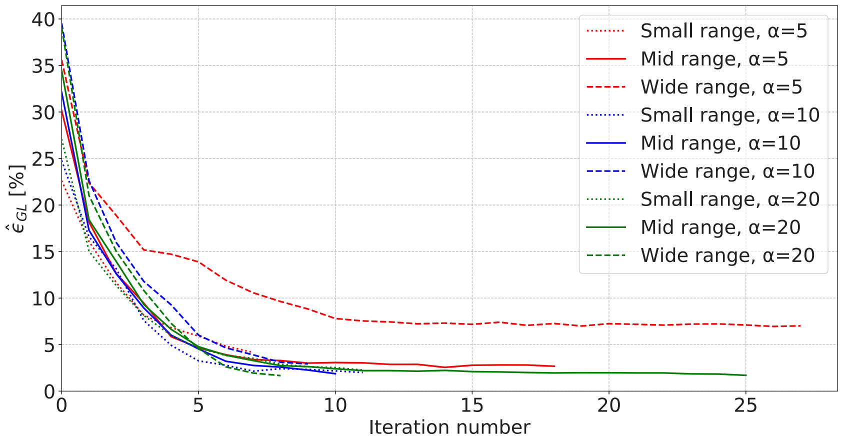
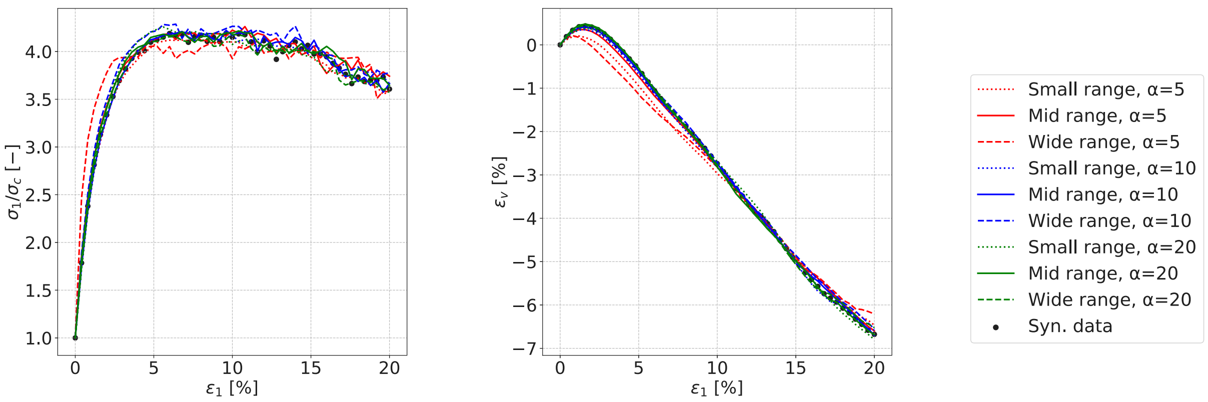


Fig. 11. GL error $\hat{\epsilon}_{\mathrm{GL}}$ over iteration number for all combinations of α values and predefined parameter ranges.

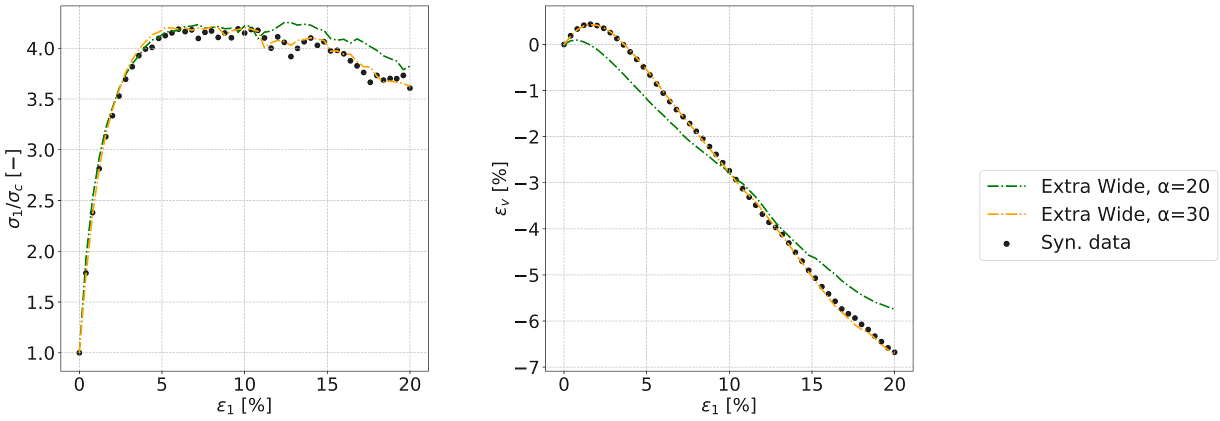
Table 9
Identified micro-mechanical parameters for extra wide parameter ranges and corresponding errors.

Combination		Identified parameters					Errors [%]		
Range	α	$\log E$ [GPa]	ν [–]	ϕ [°]	β_m $\left[\mathrm{N\,m\,rad^{-1}}\right]$	η_m [–]	$\epsilon_{\mathrm{stress}}$	$\epsilon_{\mathrm{strain}}$	$\tilde{\epsilon}$
Extra wide	20	9.219	0.037	24.365	0.637	0.520	3.03	14.01	8.52
Extra wide	30	8.738	0.169	27.129	0.174	0.233	1.15	1.41	1.28



(a) (b)

Fig. 12. Model responses in (a) stress ratio and (b) volumetric strain for varying ranges and α values.



(a) (b)

Fig. 13. Model responses in (a) stress ratio and (b) volumetric strain for extra wide range and different α values.

Table 10
Resulting errors for the model responses under uncalibrated confining pressures.

Confining pressure	Errors [%]		
σ_c [MPa]	ϵ_{stress}	ϵ_{strain}	$\tilde{\epsilon}$
0.2	2.68	2.60	2.64
1.0	1.11	2.26	1.69

Table 11
Errors in coordination number and anisotropic indicators.

Combination			Errors [%]			
Range	α	Grading	ϵ_Z	ϵ_{a^c}	ϵ_{a^n}	ϵ_{a^t}
Wide	5	sub-optimal	5.16	7.33	4.13	15.32
Mid	10	optimal	0.43	2.02	1.31	2.15
Extra Wide	20	sub-optimal	2.02	6.01	3.63	3.27
Extra Wide	30	optimal	0.49	1.83	1.50	2.99

The model responses in stress ratio and volumetric strain for the mentioned sets of micro-mechanical parameters under $\sigma_c = 0.2$ MPa and 1.0 MPa are depicted in Fig. 14. The discrete points belong to the model responses obtained with the synthetic parameters. The solid lines correspond to the predictions using the parameters calibrated for $\sigma_c = 0.5$ MPa. Overall, a very good agreement between the curves of the synthetic and calibrated parameters is obtained for both confining pressures. The stress responses for $\sigma_c = 1.0$ MPa are almost identical, whereas for $\sigma_c = 0.2$ MPa some differences are observable for strains greater than 15%. This is also reflected by the associated errors ϵ_{stress}, ϵ_{strain} and $\tilde{\epsilon}$, depicted in Table 10.

Altogether, satisfactory results are obtained when comparing the model responses under these two confining pressures for the synthetic and calibrated micro-mechanical material parameters. This suggests that the calibrated parameters also give reliable results for predictions under non-calibrated confining pressures.

4.4. *Micro-mechanical predictions*

So far, all investigations have been done with respect to the macroscopic model responses and successful calibration was achieved at the macroscale. An open question is how well the fabric is captured using the micro-mechanical parameters calibrated from the macroscopic responses. Since the synthetic data is generated by a DEM simulation, all required micro-mechanical information of the reference solution is available. This allows an accurate comparison of the micro-structure over time. This can also be seen as an additional validation of the GL results since no micro-mechanical data is used for the calibration.

For the micro-mechanical consideration, the coordination number Z and fabric are considered. Neglecting particles which do not contribute to a stable state (Thornton, 2000), the coordination number is computed as

$$Z = \frac{2C - N_1^c}{N^c - N_0^c - N_1^c} \tag{8}$$

where C is the accumulated number of contacts, N^c the number of particles and N_0^c and N_1^c the number of particles with zero and one contact, respectively. The fabric and mechanical anisotropy is evaluated on the basis of the fabric tensors (Guo and Zhao, 2013; Rothenburg and Bathurst, 1989)

$$\begin{aligned}
\boldsymbol{\varphi}^c &= \frac{1}{N^c} \sum_{N^c} \mathbf{n}^c \otimes \mathbf{n}^c \\
\chi^n &= \frac{1}{N^c} \sum_{N^c} \frac{f_n^c \mathbf{n}^c \otimes \mathbf{n}^c}{1 + \mathbf{a}^c : (\mathbf{n}^c \otimes \mathbf{n}^c)} \\
\chi^t &= \frac{1}{N^c} \sum_{N^c} \frac{f_t^c \mathbf{t}^c \otimes \mathbf{n}^c}{1 + \mathbf{a}^c : (\mathbf{n}^c \otimes \mathbf{n}^c)}.
\end{aligned} \tag{9}$$

The fabric tensors are an indication of the spatial distribution of the contact normals $\boldsymbol{\varphi}^c$, the fabric of the normal force chains χ^n and the fabric of the tangential force chains χ^t. In Eq. (9), $\mathbf{n}^c$ and $\mathbf{t}^c$ are the normal and tangential unit vectors; f_n^c and f_t^c are the magnitudes of the inter-particle contact forces in the normal and tangential directions. For the definition of anisotropies, the deviatoric parts of the fabric tensors are used. Further, the anisotropic tensor due to geometrical change $\mathbf{a}^c$ as well as the mechanical anisotropy tensors corresponding to the normal and tangential parts of force chains $\mathbf{a}^n$ and $\mathbf{a}^t$ are used. They are defined as

$$\begin{aligned}
\mathbf{a}^c &= \frac{15}{2} \operatorname{dev}(\boldsymbol{\varphi}^c) \\
\mathbf{a}^n &= \frac{15}{2} \frac{\operatorname{dev}(\chi^n)}{\operatorname{tr}(\chi^n)} \\
\mathbf{a}^t &= \frac{15}{3} \frac{\operatorname{dev}(\chi^t)}{\operatorname{tr}(\chi^n)}.
\end{aligned} \tag{10}$$

Based on Sr, the inclination between the respective anisotropy tensor and the stress deviator $\operatorname{dev}(\boldsymbol{\sigma})$, the anisotropic indicators a^* ($*$ stands for c, n and t) are computed as

$$a^* = \operatorname{sign}(Sr) \sqrt{\frac{3}{2} \mathbf{a}^* : \mathbf{a}^*}. \tag{11}$$

Thus, the second invariant of the anisotropic tensor is used to measure the intensity of the anisotropy. In the following analysis a^c, a^n and a^t, calculated according to Eq. (11), are used in the comparisons.

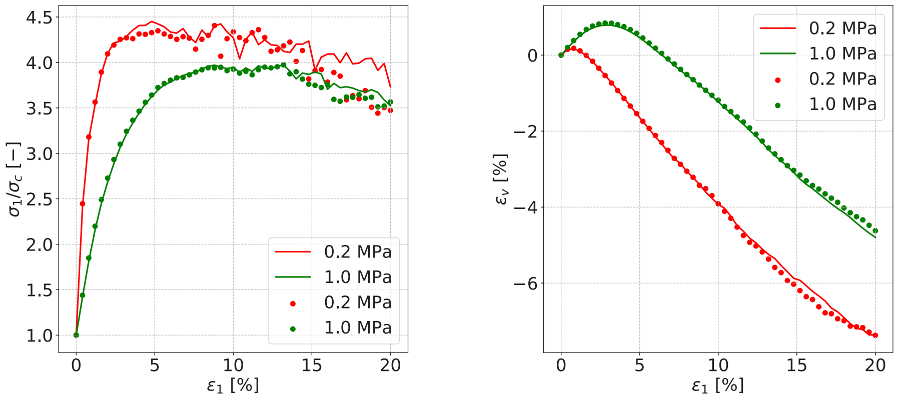
Four GL results are considered in the analysis, two optimal and two sub-optimal solutions. The classification is done with respect to the combined sample errors $\tilde{\epsilon}$. The two best GL results, mid range with $\alpha = 10$ and extra wide range with $\alpha = 30$ are used as optimal solutions. Their corresponding errors are with $\tilde{\epsilon} = 1.13\%$ and $\tilde{\epsilon} = 1.28\%$ the smallest. In contrast, the two worst GL results, extra wide range with $\alpha = 20$ and wide range with $\alpha = 5$, are used as sub-optimal solutions. Their corresponding errors are with $\tilde{\epsilon} = 8.52\%$ and $\tilde{\epsilon} = 6.36\%$ the highest. As described before, the associated parameters and model responses for the mid and wide range are depicted in Table 8 and Fig. 12, respectively. The same is depicted for the extra wide range in Table 9 and Fig. 13 respectively.

For a better understanding on how a GL calibration influences the micro-structure, the coordination number and the anisotropic indicators of the chosen calibrations are compared with those of the synthetic data. The relevant differences (Δ) are plotted in Fig. 15. Additionally, the observation sample errors (calculated according to Eq. (5)) are used as quantitative evaluation. They are shown in Table 11.

For the differences in both coordination number and anisotropic indicators, a clear tendency can be observed. The maximum difference in each plot belongs to the wide range with $\alpha = 5$ and is already observable within 2% axial strain. Compared with the other GL results, the deviations of $\alpha = 5$ from the reference solution are the highest over the whole strain range. The differences in the coordination number and the anisotropic indicator in the tangential direction are significantly higher over the entire strain range, compared with the other GL solutions. Accordingly, all corresponding observation sample errors ϵ_* ($* = Z, a^c, a^n, a^t$) are the highest for this calibration.

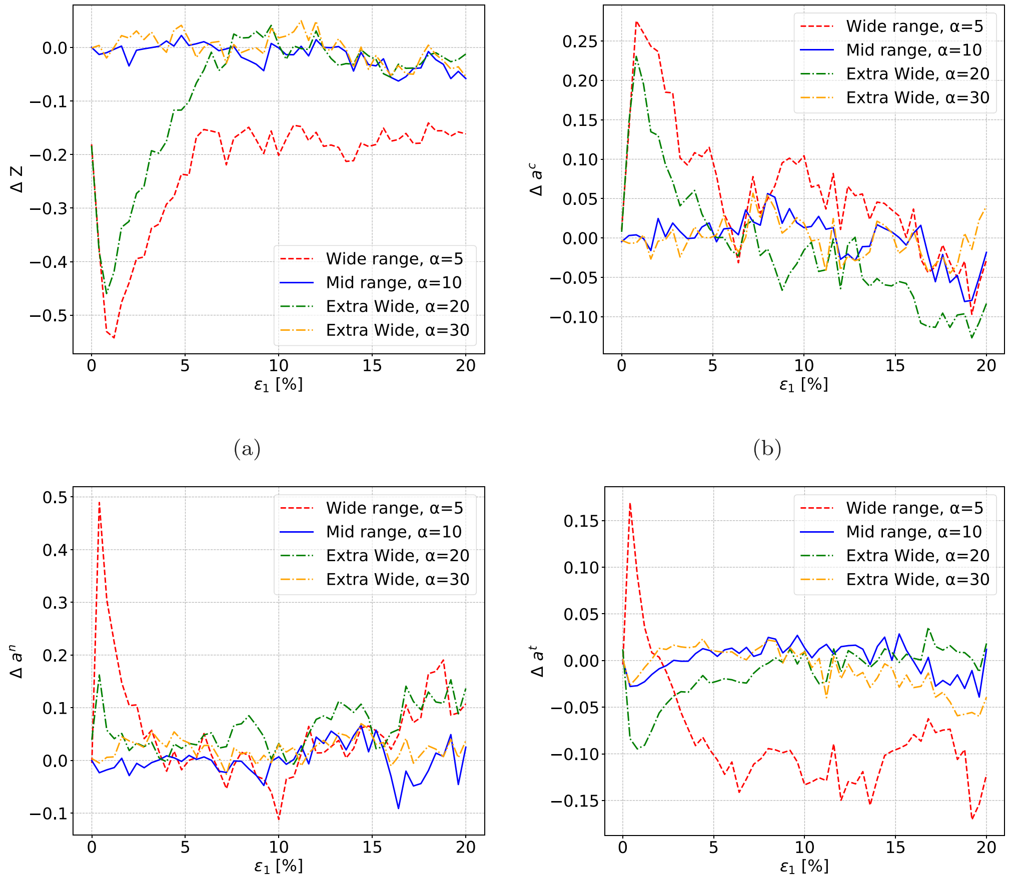
Large differences are also observed for the second sub-optimal GL result (extra wide with $\alpha = 20$). Similar to the other sub-optimal result discussed previously, large differences are obtained within the first part of the applied strain range. From a quantitative point of view the corresponding ϵ_* are relatively high as well.

In contrast to the two sub-optimal GL results, significantly smaller differences are obtained for the two optimal GL results. This is also reflected in the observation sample errors, between 0.43% and 2.99%. As a consequence, the micro-mechanical behaviour in terms of fabric is captured significantly better, leading to the macroscopic responses almost identical to the reference data.



(a) (b)

Fig. 14. Model responses in (a) stress ratio and (b) volumetric strain for the confining pressures of $\sigma_c = 0.2\,\mathrm{MPa}$ and $1.0\,\mathrm{MPa}$. The discrete points show the model responses using the synthetic parameter set. The solid lines correspond to the predictions using the parameter set calibrated for $\sigma_c = 0.5\,\mathrm{MPa}$ (Table 6).



(a) (b)

(c) (d)

Fig. 15. Differences (Δ) of (a) coordination number Z, and anisotropic indicators (b) a^c, (c) a^n and (d) a^t with respect to solutions of the synthetic data.

Table 12
Correlations coefficients between errors in model response and errors in fabric.

	ϵ_Z	ϵ_{a^c}	ϵ_{a^n}	ϵ_{a^t}
$\epsilon_{\mathrm{stress}}$	0.82	0.79	0.95	0.86
$\epsilon_{\mathrm{strain}}$	0.69	0.86	0.73	0.44

The result indicates that there are inherent relationships between the errors in the macroscopic and microscopic model responses. As a measurement for the correlation, the correlation coefficient introduced in Section 3.4 is used. The observation sample errors of the macroscopic model response and the fabric are used as the variables. Table 12 summarises the resulting correlation coefficients between errors in the macroscopic model prediction and errors in the fabric. With respect to the observation sample error in stress ratio, the correlation coefficients are 0.82, 0.79, 0.95 and 0.86. That means there is at least a strong linear correlation between ϵ_{stress} and all observation sample errors in the fabric.

4.5. Computation times

Computation time might be a limiting factor for the practical application of GL. Hence, the computation times of the synthetic GL calibrations are analysed in more detail. Considered are the maximum computation time per iteration, the accumulated maximum computation time and the accumulated total computation time (Fig. 16). Note, that an iteration always consists of a batch of N_p DEM simulations (model evaluations). If necessary, the results are truncated after iteration 15. At this point, the GL error $\hat{\epsilon}_{\text{GL}}$ has generally converged (cf. Fig. 11) and no noticeable changes are expected.

From a computational resource point of view, it is essential to investigate the accumulated maximum computation time as well as the accumulated total computation time. Since the time of sample generation and resampling is negligible in comparison to the DEM model evaluations, the accumulated simulation times represent how long the GL calibrations can take. The accumulated maximum computation time represents a GL calibration under perfect conditions where all DEM simulations of a batch, i.e. within an iteration, are performed in parallel (e.g. 80 threads would be required for 80 model evaluations). In this case the maximum computation time per iteration determines the time required for the corresponding iteration. The accumulated total computation time represents a GL calibration under the worst possible conditions. All DEM model evaluations are performed in series and thus the sum of simulation times within a batch must be considered. As a consequence, the accumulated maximum and accumulated total computation times in the final iteration represent the times required for a complete GL calibration. The latter one is under the worst possible and the first-mentioned under the best possible conditions.

The highest maximum computation times per iteration are obtained in iterations 0 and 1 for all GL calibrations (Fig. 16(a)). The peaks are in the range of 0.52–4.08 h. Thus, a high variation between the different GL calibrations is observable. The three highest peaks correspond to the three GL calibrations with a wide range and the two smallest peaks to small ranges. Generally, the maximum computation time reduces over the number of iterations and at iteration 6 the highest maximum computation time reduces to 0.29 h. Afterwards, only minor changes in maximum computation times are observed. This is in line with the finding of Section 3.4. An acceptable accuracy for standard calibrations is achieved and only minor changes of micro-mechanical parameters within the samples take place. Thus, there is no big difference in computation times anymore. Furthermore, the high maximum computation times in the beginning (i.e. iteration 0 and 1) are related to samples with high $\log E$. The higher $\log E$, the smaller the time step and the higher the computation time. Consequently, the highest maximum computation times in the beginning are obtained with the widest range.

In addition to the maximum computation times per iteration, the computation time histograms of iterations 0 and 1 for the calibration with the wide range and $\alpha = 20$ are shown as an inset in Fig. 16(a). It can be seen that in iteration 0 seven DEM simulations take between 1–2 h, two simulations between 2 and 3 h and a single simulation slightly more than 4 h (160 simulations in total). In contrast, in iteration 1 only five simulations take between 1 and 2 h with a maximum computation time of 1.41 h. The reason for this is that the samples in iteration 0 are quasi-randomly generated by a Halton number generator with respect to the pre-defined parameters ranges (cf. Section 2.2). This includes the generation of samples with high stiffness as well as samples with unfavourable parameter combinations. Hence, the high computation times are not only related to the required time step, but also the number of time steps needed to reach stability for certain unfavourable parameter combinations, such as ϕ close to zero. Since samples with unfavourable parameter combinations are generally not giving good model responses, they are automatically eliminated after iteration 0.

The final accumulated maximum computation times range from 1.96 h for the small range with $\alpha = 5$ to 8.29 h for the wide range with $\alpha = 20$ (Fig. 16(b)). For the default range (mid range) and the advised $\alpha = 10$ the final accumulated maximum computation time is 3.91 h about twice as high as the minimum time. However, the associated combined sample error of 1.13% is clearly smaller than the 4.13% which is obtained for the small range with $\alpha = 5$ (cf. Table 8).

Considering the accumulated total computation times (Fig. 16(c)), the final values are drastically higher than for the accumulated maximum computation times. For the default range (mid range) and the advised $\alpha = 10$ the final accumulated total computation time is 119.24 h. The minimum of 44.15 h is obtained for the small range with $\alpha = 5$. Due to the serial approach, the three lowest final accumulated total computation times correspond to the GL runs with $\alpha = 5$. In contrast, the three highest times, each greater than 250 h, are obtained for the GL runs with $\alpha = 20$. This reveals the importance of using a platform which allows parallel execution of the DEM simulations.

5. Discussion

In the previous sections, a comprehensive study on the performance of iterative Bayesian filtering for the calibration of DEM models based on triaxial compression tests was presented. In a first step, the superior behaviour of the iterative approach compared to the non-iterative approach was shown by performing Bayesian calibrations of micro-mechanical parameters of Toyoura sand (Section 3). The errors in stress ratio and volumetric strain for the confining pressure of $\sigma_c = 0.5\,\text{MPa}$ are reduced by factors of 3.77 and 4.10, respectively. The associated calibration was automatically terminated after iteration 6, i.e. a total of seven simulation batches were necessary to reach convergence. With 80 samples per batch (cf. Eq. (4) for $\alpha = 10$ and five fitting parameters), 560 model evaluations were performed to obtain the optimal solution.

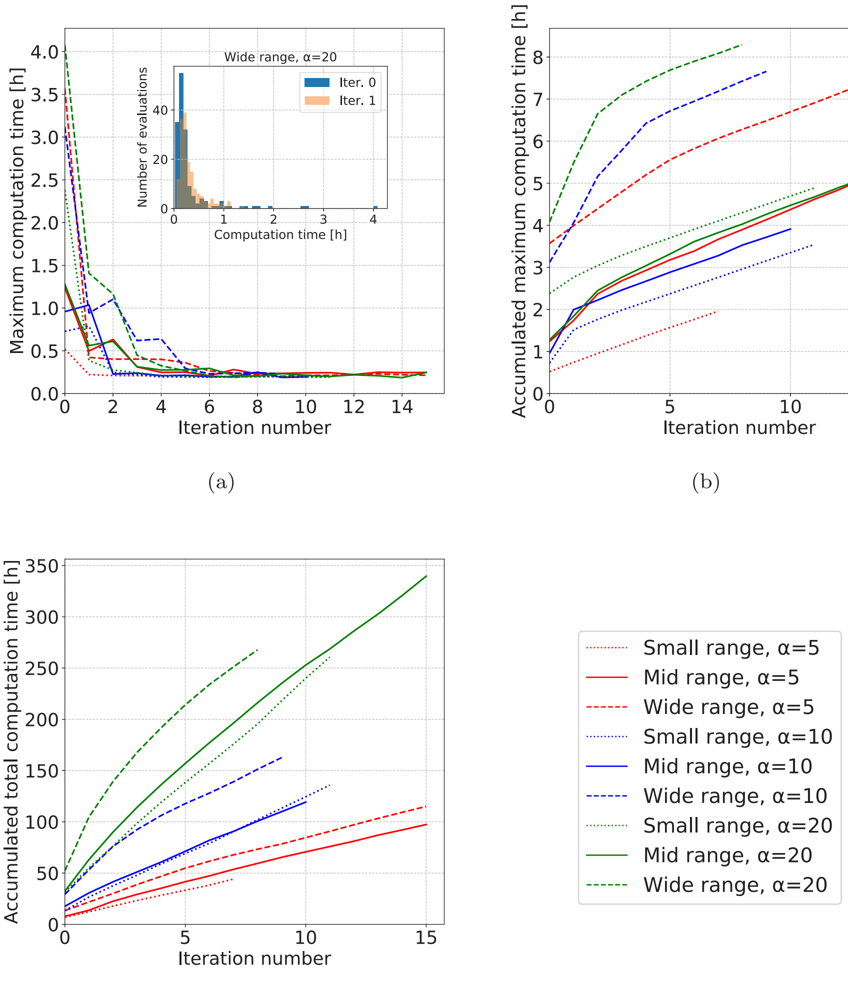
In a second step, synthetic reference data was used as a quasi-analytic solution to further the validation in Section 4. Based on 11 GL calibrations, it was shown that GL leads to outstanding accuracy when a sufficient number of samples was used. Generally, the number of required samples was estimated with respect to the unknown scalar α via Eq. (4). However, it was discovered that α, thus the optimal number of samples, depends on the predefined parameter ranges. Only a varying range of inter-particle Young's modulus, $\Delta \log E$, is considered since the ranges of the remaining parameters were already well covered.

The variation of the combined sample error with respect to $\log E$ and α is now shown in Fig. 17. $\Delta \log E$ is plotted against α for all synthetic GL calibrations. The corresponding combined sample errors $\tilde{\epsilon}$ are classified and indicated by different colours. Green is related to $\tilde{\epsilon} < 2\%$ and classified as optimal. Red and orange correspond to $\tilde{\epsilon} > 5\%$ and $2\% \le \tilde{\epsilon} \le 5\%$, respectively, and represent sub-optimal solutions.

Sub-optimal solutions are obtained for all simulations with $\alpha = 5$ and the first optimal solutions (green dots) are obtained for $\alpha = 10$. Consequently, there is a threshold for α below which no optimal GL solutions are obtained, independent of the range $\Delta \log E$. The reason behind this is that the number of samples is not representative for the distribution of individual micro-mechanical parameters within the predefined parameter ranges. The same holds for the wide range ($\Delta \log E = 6\,\text{GPa}$) with $\alpha = 10$ and the extra wide range ($\Delta \log E = 8\,\text{GPa}$) with $\alpha = 20$.

Based on the regime map for the calibration quality, it is advised to use $\alpha = 10$ as the default value to determine the number of samples whilst applying a maximum predefined range of $\Delta \log E = 4\,\text{GPa}$ for the inter-particle Young's modulus. If it is impossible to pre-estimate E in the range of $\Delta \log E = 4\,\text{GPa}$ it is advised to increase α by 10 for each widening of $\Delta(\Delta \log) E = 2\,\text{GPa}$.

Generally, GL results with acceptable accuracy are obtained at iteration 6. Up to this iteration, significant changes within the samples are



(a)

(b)

(c)

Fig. 16. (a) Maximum computation time per iteration with inlet showing computation time histograms of iterations 0 and 1 for wide range and $\alpha = 20$, (b) accumulated maximum (parallel) computation time and (c) accumulated total (serial) computation time. Note that all simulations have been carried out using one thread only.

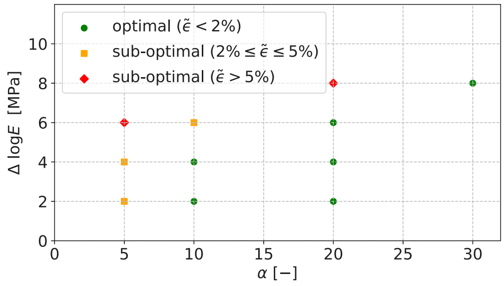


Fig. 17. Colour coded combined sample errors $\tilde{\epsilon}$ for all synthetic GL calibrations with respect to the predefined range $\Delta \log E$ and sample size defining factor α.

created by the resampling algorithm. Afterwards only minor changes in micro-mechanical parameters within the samples are observed and local minima with low probabilities are overlooked for the sake of the more probable local minima. To investigate this behaviour further, the evolution of observation sample errors in stress ratio ϵ_{stress}, volumetric strain ϵ_{strain} and the combined sample error $\tilde{\epsilon}$ are considered for the sample with the highest probability in each iteration. They are plotted for the synthetic GL calibration with $\alpha = 10$ and the mid range ($\Delta \log E = 4\,\text{GPa}$) in Fig. 18.

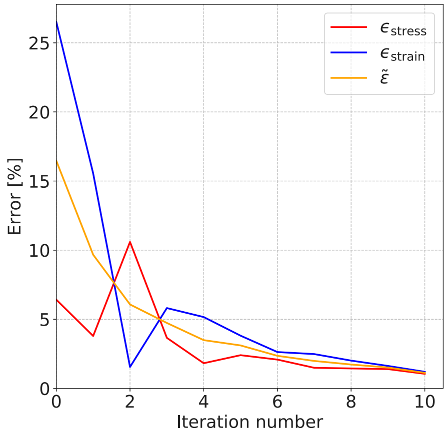


Fig. 18. Evolution of observation sample errors in stress ratio ϵ_{stress}, volumetric strain ϵ_{strain} and combined sample error $\tilde{\epsilon}$ for synthetic GL calibration with $\alpha = 10$ and mid range ($\Delta \log E = 4\,\text{GPa}$).

From Fig. 18 it can be clearly seen that $\tilde{\epsilon}$ decreases with an increasing number of iterations. In contrast, there is no oscillation free behaviour for ϵ_{stress} and ϵ_{strain}. On the one hand, ϵ_{stress} increases from 3.79 to 10.60% after iteration 1 before decreasing again to 3.66% at iteration 3. On the other hand, ϵ_{strain} decreases from 15.55 to 1.55% after iteration 1 before increasing again to 5.81% at iteration 3. The reason behind this oscillatory behaviour is that the local minima for the sample with the highest probability changes. Thus, it is not a convergence problem of GL. This is seen in the combined sample error $\tilde{\epsilon}$ which gradually decreases from 9.67 to 6.07 and 4.73% from iteration 1 to iteration 3. Starting from iteration 5 both, ϵ_{stress} and ϵ_{strain} decrease with an increasing number of iterations and with $\tilde{\epsilon} = 3.11\%$ an acceptable accuracy is obtained. With respect to the previous classification the result is still sub-optimal. However, after iteration 5 $\tilde{\epsilon}$ decreases slowly to 2.36, 1.99, 1.73, 1.51 and finally to 1.13% in iteration 10. Thus, $\tilde{\epsilon}$ is only reduced by 1.98% within the last five iterations. In contrast, $\tilde{\epsilon}$ is reduced by 13.35% from iteration 0 to iteration 5. As a consequence, the question arises if the further reduction of errors after iteration 5 is in proportion to the additional computational effort. This question is problem dependent and cannot be answered in a general manner.

As shown in Section 4.2, it is not sufficient to only use the normalised covariance coefficient s for the termination of GL. More sophisticated termination criteria are obtained when additionally including the increment of $\hat{\epsilon}_{\text{GL}}$, as well as the increment of $\tilde{\epsilon}$. Including the increment of $\hat{\epsilon}_{\text{GL}}$ guarantees an automatic termination of the calibration in case of convergence to a normalised covariance coefficient higher than the applied threshold. This is the case when resampling does not lead to noticeable changes in the parameter samples anymore. A terminated calibration with a normalised covariance coefficient significantly higher than the threshold can then be used as an indicator for sub-optimal results. The inclusion of $\tilde{\epsilon}$ in the termination criteria allows the possible reduction of GL iterations when the desired accuracy is reached. Thus, it represents a trade-off between accuracy and computational effort. Even if the desired accuracy is not reached it might be beneficial to automatically terminate GL when the increment of $\tilde{\epsilon}$ undercuts a predefined threshold. This is required for the calibration with highly disturbed experimental data where an excellent fit, independent of the applied optimisation scheme, is impossible. Thus, it might be possible that the desired $\tilde{\epsilon}$ is never reached. Terminating GL when the increment of $\tilde{\epsilon}$ is sufficiently small guarantees a solution close to the best possible solution with a reduced computational effort for the calibration. Based on the GL results using synthetic data, $s = 0.01$, $\Delta\hat{\epsilon}_{GL} = 0.2\%$, $\tilde{\epsilon} = 2\%$ and $\Delta\tilde{\epsilon} = 0.2\%$ are proposed as default values for the termination criteria. Consequently, new termination criteria are implemented in GL:

$$s \leq 0.01 \quad \text{or} \quad \Delta\hat{\epsilon}_{GL} \leq 0.2\% \quad \text{or} \quad \tilde{\epsilon} \leq 2\% \quad \text{or} \quad \Delta\tilde{\epsilon} \leq 0.2\%. \tag{12}$$

Furthermore, Section 4.2 shows that a reasonable, but still sub-optimal, fit for the stress is possible while the error for the volumetric strain can be quite high (e.g. Fig. 13), even though both are used as reference data in the calibration. This indicates that the volumetric strain should not be neglected in the reference data. Otherwise, significant errors in the corresponding model response in volumetric strain are expected. To proof this hypothesis, the GL calibration with $\alpha = 10$ and the mid range is run again, but without the volumetric strain in the reference data.

Fig. 19 shows the associated model response in stress ratio and model prediction in volumetric strain, obtained in iteration 7. As expected, an excellent fit for the stress ratio is obtained whereas the prediction of the volumetric strain is less satisfactory. A quantitative confirmation is obtained by the corresponding observation sample errors. The error in stress ratio, $\epsilon_{\text{stress}} = 1.84\%$, indicates an optimal GL result. In contrast, the error in volumetric strain, $\epsilon_{\text{strain}} = 19.67\%$, is more than ten times higher than ϵ_{stress}. This reveals that it is not sufficient to consider the stress response only. Moreover, having only a single observation in the reference data reduces the constraints within the Bayesian filtering and this increases the risk to fall into sub-optimal local minima. A benefit of GL is its inherent multi-objectivity. The approach is not limited to a single observation vector. This allows the calibration of parameters with respect to stress and volumetric strain response at the same time, as clearly shown in this work.

Micro-mechanical predictions reveal that the error in the fabric is strongly related to the errors in the macroscopic model response, independent of the associated micro-mechanical parameters. As a consequence, it is crucial to reduce the error in model responses as much as possible within the calibration, even if it comes with higher computational costs. Otherwise, errors in the fabric are expected and the micro-mechanical behaviour cannot be capture correctly in more complex configurations. This is also seen for the differences (Δ) of coordination number Z and anisotropic indicator a^c with respect to the two calibrations with $\alpha = 10$ and the mid range (Fig. 20). For both Z and a^c the resulting differences with respect to the synthetic solution are significantly higher when only the stress response is used in the reference data. This outcome was expected due to the high error in volumetric strain.

Summarising, the micro-mechanical material behaviour cannot be captured accurately when a contact model with sub-optimal calibrated parameters is used for more complex simulations. Thus, it is not advised to terminate GL for the sake of computational costs before the proposed threshold of $\tilde{\epsilon} = 2\%$ or a sufficiently large strain range is reached.

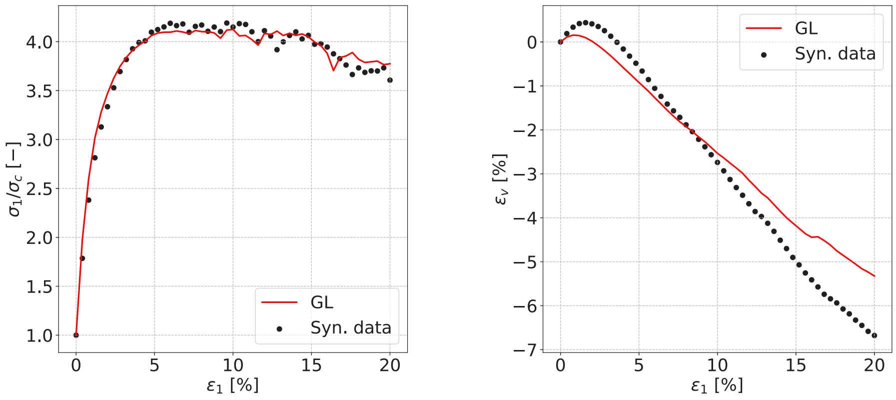
The investigation of computation times for GL in Section 4.5 reveals that the efficiency of GL highly depends on computational resources. The required computation time for the calibration with $\alpha = 10$ and the mid range under best possible (all model evaluations in parallel) and worst possible (all model evaluation in series) conditions vary significantly. Under the worst possible conditions (119.24 h) the total computation time is more than 30 times higher than under best possible conditions (3.91 h). Consequently, it is advised to perform GL on computing platforms that allow parallelisation.

Based on the findings of this work, a more sophisticated and completely automated version of GL is developed. Within the calibration toolbox, the predefined parameter ranges of Table 2 ($\log E \in [7\,\text{GPa}, 11\,\text{GPa}]$) and $\alpha = 10$ are set as defaults to guarantee optimal GL results. Furthermore, the extended termination criteria (Eq. (12)) with corresponding default values is implemented. Moreover, the toolbox is designed to run DEM simulations in parallel on several independent platforms by implementing `Python`–`Shell` interfaces for several platforms. The resampling of parameters is performed in `Python`, the DEM simulations are started via `Shell`-script on the chosen platform and as soon as all simulations have finished the Bayesian filtering is performed in `Python`.

6. Conclusions

The work at hand constitutes a major step to a fully-automated and reliable calibration toolbox for DEM simulations, based on an iterative Bayesian filtering framework. The presented framework features excellent performance in terms of model responses of a triaxial compression test using the stress as well as the volumetric strain as reference data. In comparison to the sequential Monte Carlo method, the errors in stress ratio and volumetric strain are reduced by a factor of about 4. In addition, with respect to synthetic data, optimal results with errors between 1 and 2% are obtained. Optimal calibration results are guaranteed when the number of samples is high enough. With the proposed default ranges of micro-mechanical parameters, the sample size is reliably estimated on the basis of the previously unknown scalar factor α by setting a default value of $\alpha = 10$. Thus, for five micro-mechanical parameters 80 samples are representative for the distribution of individual parameters within their default ranges.

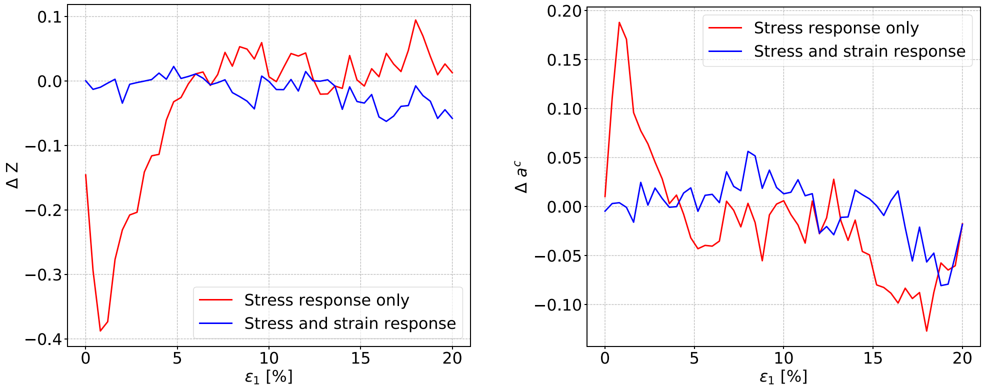
It is further revealed that the calibration of micro-mechanical parameters for the DEM is clearly a multi-objective problem. The error in volumetric strain skyrockets to 19.67% when the volumetric strain is removed from the reference data. This has also a negative impact on the prediction in the fabric since it is demonstrated that the errors in the fabric are correlated to the errors in stress and volumetric strain



(a)

(b)

Fig. 19. (a) Model response in stress ratio and (b) model prediction in volumetric strain for synthetic reference data neglecting the volumetric strain in the reference data.



(a)

(b)

Fig. 20. Differences (Δ) of (a) coordination number Z and (b) anisotropic indicator a^c with respect to the calibrations with $\alpha = 10$ and the mid range.

predictions. Thus, it is not sufficient to calibrate with respect to the stress response only.

Moreover, it is shown that the calibrated micro-mechanical parameters can be used to accurately predict the behaviour under non-calibrated stress paths provided no particle crushing or change of grain morphology is involved. Nevertheless, it is also pointed out that particle crushing or change of grain morphology has an influence on the micro-mechanical parameters, and if relevant, the correct stress regime should be used in the calibration procedure. The current study focuses on drained triaxial compression, with the initial micro-structure originating from a dense Toyoura specimen. Whether the micro-mechanical parameters calibrated in the dense state can give accurate prediction for non-calibrated loose states or undrained stress paths is not tested. The capability of the calibration framework to estimate micro-parameters from the whole spectrum of soil volumetric behaviour (i.e., compressive, dilative and volume-conserved) will be investigated in future work.

The fully-automated Bayesian calibration toolbox `GrainLearning 1.0`[3] is released based on the current findings. Default values for the predefined parameter ranges as well as for the sample size (determined on the basis of $\alpha = 10$) are implemented. More complete termination criteria are formulated and implemented. The criteria include the normalised covariance coefficient and the increment of the GL error as collective measures of all DEM model responses within an iteration. If the combined sample error (or its increment) is sufficiently small, GL will be terminated to avoid redundant iterations that do not improve the model accuracy further.

A huge benefit of the developed toolbox is its flexible implementation in terms of `Python–Shell` interfaces for various computing platforms as well as its modular expandability. For example, replacing the simulation framework (model evaluations) in the currently implemented wrapper is straightforward, which means `GrainLearning 1.0` is not limited by the `Yade` software. This also means that `GrainLearning` can be readily used for calibrating the micro-mechanical parameters of irregularly-shaped particles, as well as any arbitrary multi-objective optimisation problems.

Future work involves implementing and integrating more efficient sampling and optimisation techniques. Currently, the computational cost is spent mostly on re-evaluating the DEM model with different parameter samples. One possibility to reduce the computational cost is to “emulate” model evaluations in the parameter subspace that has previously been visited with a data-driven surrogate model. Moreover, an extension to additional observation vectors in the reference data in terms of multiple stress paths would be beneficial. The implementation is straightforward since `GrainLearning 1.0` is designed to accept an arbitrary number of observations.

---

[3] https://github.com/chyalexcheng/grainLearning

CRediT authorship contribution statement

Philipp Hartmann: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Hongyang Cheng:** Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing, Visualization, Supervision. **Klaus Thoeni:** Conceptualization, Methodology, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to acknowledge the financial support of the Australian Research Council (DP190102407). The support of the HPC facilities of the University of Newcastle is also gratefully acknowledged. In particular, the first author would like to thank Mr Aaron Scott for his technical assistance with the usage of the HPC facilities and AWS.

References

Blei, D.M., Jordan, M.I., 2006. Variational inference for dirichlet process mixtures. Bayesian Anal. 1 (1), 121–143.

Cheng, H., Luding, S., Rivas, N., Harting, J., Magnanimo, V., 2019a. Hydro-micromechanical modeling of wave propagation in saturated granular crystals. Int. J. Numer. Anal. Methods Geomech. 43 (5), 1115–1139.

Cheng, H., Shuku, T., Thoeni, K., Tempone, P., Luding, S., Magnanimo, V., 2019b. An iterative Bayesian filtering framework for fast and automated calibration of DEM models. Comput. Methods Appl. Mech. Engrg. 350, 268–294.

Cheng, H., Shuku, T., Thoeni, K., Yamamoto, H., 2018. Probabilistic calibration of discrete element simulations using the sequential quasi-Monte Carlo filter. Granul. Matter 20 (1), 11.

Coetzee, C.J., 2017. Review: Calibration of the discrete element method. Powder Technol. 310, 104–142.

Coetzee, C., 2020. Calibration of the discrete element method: Strategies for spherical and non-spherical particles. Powder Technol. 364, 851–878.

Cundall, P.A., Strack, O.D.L., 1979. A discrete numerical model for granular assemblies. Geotechnique 29 (1), 47–65.

Do, H.Q., Aragón, A.M., Schott, D.L., 2018. A calibration framework for discrete element model parameters using genetic algorithms. Adv. Powder Technol. 29 (6), 1393–1403.

Fuchs, R., Weinhart, T., Meyer, J., Zhuang, H., Staedler, T., Jiang, X., Luding, S., 2014. Rolling, sliding and torsion of micron-sized silica particles: Experimental, numerical and theoretical analysis. Granul. Matter 16 (3), 281–297.

Gao, Y., De Simone, G., Koorapaty, M., 2021. Calibration and verification of dem parameters for the quantitative simulation of pharmaceutical powder compression process. Powder Technol. 378, 160–171.

Gerber, M., Chopin, N., 2015. Sequential quasi Monte Carlo. J. R. Stat. Soc. Ser. B Stat. Methodol. 77 (3), 509–579.

Guo, N., Zhao, J., 2013. The signature of shear-induced anisotropy in granular media. Comput. Geotech. 47, 1–15.

Guo, N., Zhao, J., 2016. 3D Multiscale modeling of strain localization in granular media. Comput. Geotech. 80, 360–372.

Hadjidoukas, P., Angelikopoulos, P., Papadimitriou, C., Koumoutsakos, P., 2015. Π4U: A high performance computing framework for Bayesian uncertainty quantification of complex models. J. Comput. Phys. 284, 1–21.

Hadjidoukas, P., Angelikopoulos, P., Rossinelli, D., Alexeev, D., Papadimitriou, C., Koumoutsakos, P., 2014. Bayesian Uncertainty quantification and propagation for discrete element simulations of granular materials. Comput. Methods Appl. Mech. Engrg. 282, 218–238.

Hanley, K.J., O'Sullivan, C., Oliveira, J.C., Cronin, K., Byrne, E.P., 2011. Application of taguchi methods to DEM calibration of bonded agglomerates. Powder Technol. 210 (3), 230–240.

He, P., Fan, Y., Pan, B., Zhu, Y., Liu, J., Zhu, D., 2019. Calibration and verification of dynamic particle flow parameters by the back-propagation neural network based on the genetic algorithm: Recycled polyurethane powder. Materials 12 (20).

Hurley, R.C., Hall, S.A., Andrade, J.E., Wright, J., 2016. Quantifying interparticle forces and heterogeneity in 3D granular materials. Phys. Rev. Lett. 117, 098005.

Johnstone, M.W., 2010. Calibration of DEM Models for Granular Materials using Bulk Physical Tests (Ph.D. thesis). The University of Edinburgh.

Ma, C., Yang, J., Zenz, G., Staudacher, E.J., Cheng, L., 2020. Calibration of the microparameters of the discrete element method using a relevance vector machine and its application to rockfill materials. Adv. Eng. Softw. 147, 102833.

Nguyen, T.-T., André, D., Huger, M., 2019. Analytic laws for direct calibration of discrete element modeling of brittle elastic media using cohesive beam model. Comput. Part. Mech. 6 (3), 393–409.

Oden, J.T., Moser, R., Ghattas, O., 2010a. Computer predictions with quantified uncertainty, Part I. SIAM News 43 (9), 2008–2011.

Oden, J.T., Moser, R., Ghattas, O., 2010b. Computer predictions with quantified uncertainty, Part II. SIAM News 43 (10), 2008–2011.

Oliphant, T.E., 2006. A Guide to NumPy. Vol. 1. Trelgol Publishing USA.

Orefice, L., Khinast, J.G., 2020. A novel framework for a rational, fully-automatised calibration routine for dem models of cohesive powders. Powder Technol. 361, 687–703.

Ouadfel, H., Rothenburg, L., 2001. 'Stress–force–fabric' relationship for assemblies of ellipsoids. Mech. Mater. 33 (4), 201–221.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Plassiard, J., Belheine, N., Donzé, F., 2009. A spherical discrete element model: Calibration procedure and incremental response. Granul. Matter 11, 293–306.

Qu, T., Feng, Y., Zhao, T., Wang, M., 2019. Calibration of linear contact stiffnesses in discrete element models using a hybrid analytical-computational framework. Powder Technol. 356, 795–807.

Qu, T., Feng, Y., Zhao, T., Wang, M., 2020. A hybrid calibration approach to hertz-type contact parameters for discrete element models. Int. J. Numer. Anal. Methods Geomech. 44 (9), 1281–1300.

Rackl, M., Hanley, K.J., 2017. A methodical calibration procedure for discrete element models. Powder Technol. 307, 73–83.

Richter, C., Rößler, T., Kunze, G., Katterfeld, A., Will, F., 2020. Development of a standard calibration procedure for the dem parameters of cohesionless bulk materials – part II: Efficient optimization-based calibration. Powder Technol. 360, 967–976.

Rodgers, J.L., Nicewander, W.A., 1988. Thirteen ways to look at the correlation coefficient. Amer. Statist. 42 (1), 59–66.

Rorato, R., Arroyo, M., Gens, A., Andò, E., Viggiani, G., 2021. Image-based calibration of rolling resistance in discrete element models of sand. Comput. Geotech. 131, 103929.

Rothenburg, L., Bathurst, R.J., 1989. Analytical study of induced anisotropy in idealized granular materials. Geotechnique 39 (4), 601–614.

Ruiz, H., Kappen, H.J., 2017. Particle smoothing for hidden diffusion processes: Adaptive path integral smoother. IEEE Trans. Signal Process. 65 (12), 3191–3203.

Särkkä, S., 2013. Bayesian Filtering and Smoothing. Institute of Mathematical Statistics Textbooks. Cambridge University Press.

Sufian, A., Artigaut, M., Shire, T., O'Sullivan, C., 2021. Influence of fabric on stress distribution in gap-graded soil. J. Geotech. Geoenviron. Eng. 147 (5), 04021016.

Sun, D.A., Huang, W.X., Sheng, D.C., Yamamoto, H., 2007. An elastoplastic model for granular materials exhibiting particle crushing. In: Key Engineering Materials, vol. 340, Trans Tech Publ, pp. 1273–1278.

Thornton, C., 2000. Numerical simulations of deviatoric shear deformation of granular media. Geotechnique 50 (1), 43–53.

Turkia, S.Ben, Wilke, D.N., Pizette, P., Govender, N., Abriak, N.-E., 2019. Benefits of virtual calibration for discrete element parameter estimation from bulk experiments. Granul. Matter 21 (4), 110.

Šmilauer, V., et al., 2015. Yade Documentation, second ed. http://dx.doi.org/10.5281/zenodo.34073, URL http://yade-dem.org/doc/.

Westbrink, F., Elbel, A., Schwung, A., Ding, S.X., 2021. Optimization of DEM parameters using multi-objective reinforcement learning. Powder Technol. 379, 602–616.

Wu, Y., Yamamoto, H., Cui, J., Cheng, H., 2020. Influence of load mode on particle crushing characteristics of silica sand at high stresses. Int. J. Geomech. 20 (3), 04019194.

Ye, F., Wheeler, C., Chen, B., Hu, J., Chen, K., Chen, W., 2019. Calibration and verification of dem parameters for dynamic particle flow conditions using a backpropagation neural network. Adv. Powder Technol. 30 (2), 292–301.

Yoon, J., 2007. Application of experimental design and optimization to PFC model calibration in uniaxial compression simulation. Int. J. Rock Mech. Min. Sci. 44 (6), 871–889.