Article:
Van Lerberghe, A., Pasquale, A., Rodriguez, S. et al. (2025) Data-driven parametric modelling of split-Hopkinson pressure bar tests on cohesive soils. International Journal of Impact Engineering, 198. 105218. ISSN: 0734-743X

https://doi.org/10.1016/j.ijimpeng.2024.105218

International Journal of Impact Engineering 198 (2025) 105218

Contents lists available at ScienceDirect

International Journal of Impact Engineering

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

Data-driven parametric modelling of split-Hopkinson pressure bar tests on cohesive soils

Arthur Van Lerberghe [a,*], Angelo Pasquale [b], Sebastian Rodriguez [b], Andrew D. Barr [a], Sam D. Clarke [a], Dominique Baillargeat [c], Francisco Chinesta [b,c]

[a] Department of Civil & Structural Engineering, University of Sheffield, Mappin Street, Sheffield, S1 3JD, UK
[b] PIMM Lab, Arts et Métiers Institute of Technology, 151 Boulevard de l'Hôpital, Paris, 75013, France
[c] CNRS@CREATE Ltd, 1 Create Way, #08-01 CREATE Tower, Singapore, 138602, Singapore

ARTICLE INFO

Keywords:
Data-driven parametric modelling
Physics informed machine learning
Curve metamodeling
High-strain-rate testing
Split-Hopkinson pressure bar
Cohesive soils

ABSTRACT

Soil-filled wire and geotextile gabions stand as vital bulwarks in military bases, harnessing soil's innate capacity to absorb shock and safeguard both personnel and critical assets from blast and fragmentation effects. Yet, the dynamic response of cohesive soils under extreme loads remains largely unexplored, leaving engineers grappling with a significant void in knowledge as they strive to fortify structures against emerging threats. This paper considers the high-strain-rate behaviour of kaolin clay using the split Hopkinson pressure bar in both confined and unconfined configurations, with a range of moisture contents representing dry, partially-saturated and saturated conditions. Analysis of the results indicates distinct phase behaviours in transmitted and radial stress based on strain rate, moisture content and confinement. Leveraging cutting-edge machine learning models such as the Proper Orthogonal Decomposition (POD) and sparse Proper Generalised Decomposition (sPGD), data-driven parametric models were developed based on the experimental data. These models enable the prediction of cohesive soil behaviour at specified strain rate and moisture content, enabling engineers to rapidly predict soil behaviour in response to new threats and ground conditions.

1. Introduction

Fortification engineers rely on soil-filled barriers like Hesco Concertainers for blast protection, yet cohesive soil's high-strain-rate behaviour remains elusive despite its widespread use. Understanding cohesive soils such as clay and silt is crucial for the design of fortifications, due to their global presence and potential as alternatives where sandy soils are less common.

Precise data on local soils are vital for fortification engineers to adapt designs, urging comprehensive studies on soil behaviour under extreme loading conditions. Focusing on kaolin clay offers a foundation for understanding its response to strain rate, moisture, and confinement variations through SHPB tests. The complexities of cohesive soils, including their undrained behaviour and fine particle size, pose challenges compared to cohesionless soils such as sand [1–6], necessitating new research to inform resilient fortification designs and bridge existing knowledge gaps.

Experimental research on kaolin clay has explored its behaviour under varied confinement, moisture content and strain rate conditions [7]. In summary, experimental testing on kaolin clay demonstrated that fully saturated cohesive soil samples exhibited fluid-like behaviour and showed a clear dependence on strain rate. The study also highlighted the significant effects of strain rate and moisture content on the high-strain-rate response of these soils. Additionally, phase behaviour was observed under both confinement conditions, reflecting categorical variations in axial and radial stress responses. On the other hand, numerical modelling demonstrated that a valid SHPB model design could represent both testing conditions. However, the LS-DYNA models created showed significant limitations in modelling cohesive soils: they could not capture phase behaviour, and the Mohr–Coulomb model proved largely ineffective in modelling cohesive soils under high-strain-rate conditions. The model was speed-driven and struggled to accurately represent the material's cohesive properties. Additionally, the perfect seal of the confining ring in the numerical model affected the amplitude of the results. Overcoming the limitations of time-consuming high-strain rate testing, and inaccuracies in numerical modelling, machine learning methods leverage collected test data to construct data-driven parametric models capable of predicting cohesive soil behaviour.

* Corresponding author.
E-mail addresses: avanlerberghe1@sheffield.ac.uk (A. Van Lerberghe), angelo.pasquale@ensam.eu (A. Pasquale), sebastian.rodriguez_iturra@ensam.eu (S. Rodriguez), a.barr@sheffield.ac.uk (A.D. Barr), francisco.chinesta@ensam.eu (F. Chinesta).

https://doi.org/10.1016/j.ijimpeng.2024.105218
Received 7 May 2024; Received in revised form 7 November 2024; Accepted 31 December 2024
Available online 16 January 2025

Various machine learning algorithms can be utilised for regression tasks. Linear regression assumes a relationship between input features and the target variable, while polynomial regression extends this concept by considering polynomial relationships. Regularised regressions, such as Ridge and Lasso, add penalty terms to prevent overfitting and induce sparsity in models [8]. Support vector Regression (SVR) [9] extends support vector machines (SVMs) for regression tasks, while decision trees [10] and random forest regression [11] split the data and combine trees to improve accuracy. Gradient boosting [12] builds weak learners sequentially, correcting errors, and deep learning techniques [13], like artificial neural networks, learn complex relationships for regression.

Polynomial regression techniques, employed in computational mechanics [14] and structural deformation [15], have yet to be applied in the field of blast and impact dynamics. This study aims to utilise SHPB test data and machine learning tools to rapidly predict the high-strain-rate behaviour of cohesive soils under diverse loading conditions and material properties.

2. Parametric modelling: background, method & theory

Model Order Reduction (MOR) is a field within computational science focused on simplifying complex mathematical models, such as differential equations or transfer functions, while preserving their essential dynamics. By reducing degrees of freedom, MOR creates reduced-order models (ROMs) that enable faster simulations, optimisation, and control design with high fidelity. This approach is particularly useful for high-dimensional or parametric systems commonly encountered in scientific applications.

Now, a general procedure for constructing parametric surrogates of curves is presented, using extensive literature and detailed reviews on state-of-the-art MOR technologies [16–18]:

1. Consider a scenario where experimental data is collected, comprising of input parameters $\mathbf{p}_i$ and corresponding output curves $g(x;\mathbf{p}_i)$. Each curve represents the system's behaviour under various conditions, with $\mathbf{p}_i$ representing geometrical or material parameters. In simulation-based engineering, data $g(x;\mathbf{p}_i)$ is typically obtained through simulation software runs, with the parameters of interest $\mathbf{p}_i$ potentially including modelling features.
2. Each experimental data point can be viewed as a snapshot $(\mathbf{p}_i, g(x;\mathbf{p}_i))$, where $i = 1, 2, \dots, n_s$, (n_s is the number of sampling points used for training) depicting parameter combinations and their corresponding output curves.
3. Dimensionality reduction techniques such as Principal Component Analysis (PCA) or Proper Orthogonal Decomposition (POD) are applied directly to test data to extract dominant modes of variability in the output curves $g(x;\mathbf{p}_i)$.
4. The reduced basis functions $\phi_j(x)$ are constructed by identifying the dominant modes through dimensionality reduction. These basis functions effectively capture the essential features of the output curves.
5. Using the reduced basis functions, a surrogate model $\hat{g}(x,\mathbf{p})$ is constructed to approximate the output curves $g(x;\mathbf{p}_i)$ based on the input parameters $\mathbf{p}_i$ and the domain x. This surrogate model can be formulated as:
$$\hat{g}(x,\mathbf{p}) = \sum_{j=1}^{m} \lambda_j(\mathbf{p})\phi_j(x)$$
where $\lambda_j(\mathbf{p})$ are the coefficients of the surrogate model, which depend on the parameters $\mathbf{p}$.
6. Surrogate models for $\lambda_j(\mathbf{p})$ can be built by training a regression algorithm on the available dataset, to establish links between input parameters and measured output.

In polynomial regressions, addressing high-dimensional parametric challenges has led to techniques like Proper Generalised Decomposition (PGD) [19–21]. PGD, a tensor-based method, uses separable representations and a greedy iterative algorithm for adaptive basis construction. It is applied across engineering and scientific fields, including blast and impact dynamics.

2.1. POD-based modes extraction

To construct the snapshots matrix for training data $\{g_i(x)\}_{i=1}^{n_s}$, for $x \in X = \{x_j\}_{j=1}^{n_x}$ utilised in this study, consider the following procedure:

$$\mathbf{S} = \begin{bmatrix}\mathbf{g}_1 & \mathbf{g}_2 & \dots & \mathbf{g}_{n_s}\end{bmatrix} \in \mathbb{R}^{n_x \times n_s},$$

where $\mathbf{g} \in \mathbb{R}^{n_x \times 1}$ contains the evaluations of $g(x)$ over the discrete ensemble X.

Next, a reduced factorisation of the snapshots matrix is achieved through a standard truncated POD of rank r:

$$\mathbf{S} \approx \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$$

where $\mathbf{U} \in \mathbb{R}^{n_x \times r}$, $\boldsymbol{\Sigma} \in \mathbb{R}^{r \times r}$, $\mathbf{V} \in \mathbb{R}^{n_s \times r}$. From these, we can define the matrices of POD modes and coefficients, as follows:

$$\boldsymbol{\Phi} := \mathbf{U} = \begin{bmatrix}\boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \dots & \boldsymbol{\phi}_r\end{bmatrix}, \quad \boldsymbol{\Lambda} := \mathbf{V}\boldsymbol{\Sigma} = \begin{bmatrix}\lambda_1 & \lambda_2 & \dots & \lambda_r.\end{bmatrix}$$

The matrix $\boldsymbol{\Phi}$ contains, by columns, the functions of the reduced POD basis $\{\boldsymbol{\phi}_i(x)\}_{i=1}^{r}$ evaluated at points in X, while $\boldsymbol{\Lambda}$ collects the projection coefficients into the reduced basis. For a generic curve $g_k(x)$ belonging to the training dataset, where $k = 1, \dots, n_s$ and $x \in X$, its reduced counterpart is given by:

$$g_k^{(r)}(x) = \sum_{i=1}^{r} \lambda_{k,i}\phi_i(x), \tag{1}$$

and, in particular, its discrete form reads

$$\mathbf{g}_k^{(r)} = \boldsymbol{\Lambda}_{k,\bullet}\boldsymbol{\Phi}^T,$$

where $\boldsymbol{\Lambda}_{k,\bullet}$ denotes the kth row of the matrix $\boldsymbol{\Lambda}$.

Now, let us consider a parametric curve dependent on d features $\bar{\mathbf{p}} \in \Omega$, denoted as $g(x;\bar{\mathbf{p}})$, where $x \in X$. From Eq. (1) it is evident that once the reduced basis matrix $\boldsymbol{\Phi}$ is available, this function is projected onto this basis solely through the POD (parametric) coefficients $\{\lambda_i(\mathbf{p})\}_{i=1}^{r}$:

$$g^{(r)}(x;\bar{\mathbf{p}}) = \sum_{i=1}^{r} \lambda_i(\bar{\mathbf{p}})\phi_i(x).$$

The equation above indicates that a reduced-order parametric metamodel for the curves can be constructed using only the set of coefficients $\{\lambda_i(\mathbf{p})\}_{i=1}^{r}$. Specifically, the following parametric function shall be constructed:

$$\mathbf{f}(\mathbf{p}) = \begin{bmatrix}\lambda_1(\mathbf{p}) \\ \lambda_2(\mathbf{p}) \\ \vdots \\ \lambda_r(\mathbf{p})\end{bmatrix} : \Omega \subset \mathbb{R}^d \to \mathbb{R}^r,$$

from the training dataset available as $\{\mathbf{p}_k, \boldsymbol{\Lambda}_{k,\bullet} = (\lambda_{k,1}, \lambda_{k,2}, \dots, \lambda_{k,r})\}_{k=1}^{n_s}$, obtained after the POD.

2.2. Advanced PGD-based sparse nonlinear regressions

Nonlinear regression methods such as the sparse Proper Generalised Decomposition (sPGD) are increasingly vital for managing complex parameters efficiently while preserving accuracy, complementing MOR techniques for real-world engineering problems [19–21]. These methods have recently gained significant attraction in industry, with various approaches expanding their applicability for approximating parametric curves [15,22–24].

In this section, the focus shifts to the concept behind the PGD-based regression methods for constructing metamodels dependent on d

features, which are used in this study [15]. This setup enables the prediction of parametric curves using the coefficients $\lambda_i(\boldsymbol{p})$ corresponding to the POD modes, as previously suggested.

For each coefficient $\lambda_i(\mathbf{p})$, the challenge lies in constructing the function

$$f(p^1, \dots, p^d) : \Omega \subset \mathbb{R}^d \to \mathbb{R},$$

which depends on d features (parameters) p^k, $k = 1, \dots, d$, within the parametric space Ω, given a sparse sample of n_s points and their corresponding outputs.

The sparse PGD (sPGD) represents the function f using a low-rank separated representation

$$f(p^1, \dots, p^d) \approx \tilde{f}^M(p^1, \dots, p^d) = \sum_{m=1}^{M} \prod_{k=1}^{d} \psi_m^k(p^k), \tag{2}$$

constructed from rank-one updates within a greedy constructor. Here, $\tilde{f}^M$ denotes the approximation, M the number of employed modes (sums), and ψ_m^k the one-dimensional functions pertaining to mode m and dimension k.

The functions ψ_m^k, $m = 1, \dots, M$ and $k = 1, \dots, d$, are derived from a standard approximation basis $\mathbf{N}_m^k$ using coefficients $\mathbf{a}_m^k$:

$$\psi_m^k(p^k) = \sum_{j=1}^{D} N_{j,m}^k(p^k) a_{j,m}^k = (\mathbf{N}_m^k)^T \mathbf{a}_m^k,$$

where D represents the number of degrees of freedom (nodes) of the chosen approximation and $\mathbf{N}_m^k$ is the vector collecting the shape functions.

In standard regression, the approximation $\tilde{f}^M$ is obtained by minimising the error function

$$\tilde{f}^M = \arg\min_{f^*} \|f - f^*\|_2^2 = \arg\min_{f^*} \sum_{i=1}^{n_s} |f(\mathbf{p}_i) - f^*(\mathbf{p}_i)|^2,$$

where $\tilde{f}^M$ takes the separated form of Eq. (2), n_s is the number of sampling points used for training, and $\mathbf{p}_i$ are the vectors containing the input data points of the training set. It is important to note that to avoid overfitting, the number of basis functions D must satisfy $D < n_s$.

The approximation coefficients for each one-dimensional function are computed using a greedy algorithm. Given the approximation up to order $M-1$, the Mth order term is determined as

$$\tilde{f}^M = \sum_{m=1}^{M-1} \prod_{k=1}^{d} \psi_m^k(p^k) + \prod_{k=1}^{d} \psi_M^k(p^k)$$

The resulting function is expected to approximate f not only in the training set but also at any point $\mathbf{p} \in \Omega$.

The main challenge lies in balancing detailed approximations with limited data and avoiding overfitting. To address this, a modal adaptivity strategy (MAS) was introduced alongside sPGD. However, achieving desired accuracy with MAS can lead to overfitting or premature termination, resulting in low-order PGD solutions that lack sufficient detail [19]. Additionally, in cases with sparse non-zero elements in the interpolation basis, MAS may struggle to accurately capture the true model, reducing accuracy. Various regularisation methods combining L^1 and L^2 norms affecting coefficients $\mathbf{a}_m^k$ have been proposed to enhance predictive performance beyond sPGD or create parsimonious models with improved accuracy [20].

3. Split-Hopkinson pressure bar testing of cohesive soils

This is a concise overview of the Split-Hopkinson pressure bar testing of cohesive soils conducted in [7], highlighting all the key testing parameters. For further details, please refer to [7].

Table 1
Overview of the kaolin clay material properties.

Soil properties	Units	Value
Primary mineral	–	Kaolinite
Particle density, ρ_s	Mg m^{-3}	2.65
Liquid limit, LL	%	40
Plastic limit, PL	%	25
Plastic Index, PI	%	15
D_{50}	μm	0.74
Particle sphericity	–	Low – Medium
Angularity	–	Subrounded – Subangular
Surface texture	–	Smooth

3.1. Material characterisation

The soil chosen is defined as white fine CLAY (CL) according to EN ISO 14688–1:2002 [25], and referred to as 'kaolin clay' for brevity. Table 1 outlines the material properties of the kaolin clay, derived using the methods detailed below.

The kaolin clay's particle size distribution (PSD) was assessed using IMERYS' supplier-provided data sheet. Fig. 1 shows the cumulative PSD, revealing a D_{50} of 0.74 μm. With 80% clay and 20% silt, the soil is classified as CLAY (CL).

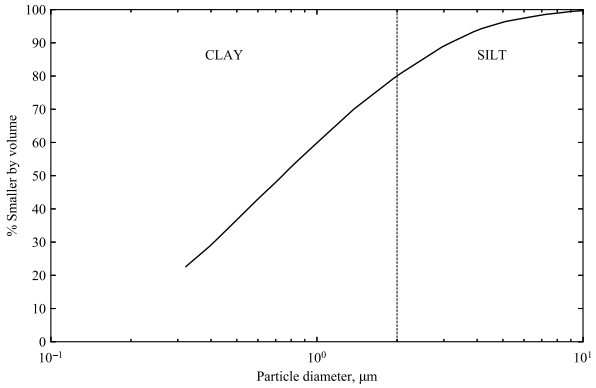

Fig. 1. Cumulative particle size distribution of the kaolin clay soil.

The particle density, ρ_s, represents the density of the solid mineral particles. Together with the bulk dry density, ρ_d, it determines the soil's void ratio using the relationship:

$$e = \frac{\rho_s}{\rho_d} - 1$$

The kaolin clay's particle density, ρ_s, was determined to be 2.65 Mg m^{-3}, following the method outlined in BS 1377–2:1990 §8.2 [26], which is the density of kaolinite.

X-ray diffraction (XRD) was conducted on the kaolin clay to identify its constituent minerals. In XRD, an incident X-ray beam diffracts due to the specimen's regular atomic structure. By measuring diffraction intensity at various incident angles, θ, a unique diffraction pattern is obtained. This pattern was compared to a database of known patterns for phase identification. Phase analysis utilised a PANalytical Aeris diffractometer and the ICDD's Diffraction File (PDF-4+). Fig. 2 displays the diffraction pattern of the kaolin clay, revealing primarily kaolinite with some quartz present.

Sample consistency is vital, hence Atterberg limits must be consistent across all soil samples. The fall cone test was used to obtain the Atterberg limits of kaolin clay [26]. The liquid limit (LL), plastic limit (PL), and plastic index (PI) are 40%, 25% and 15%, respectively [26]. These values surpass the A-line, confirming the soil's classification as CLAY (CL) [27].

Kaolinite-sized particles in the soil were qualitatively assessed using a scanning electron microscope (SEM), employing descriptors from EN

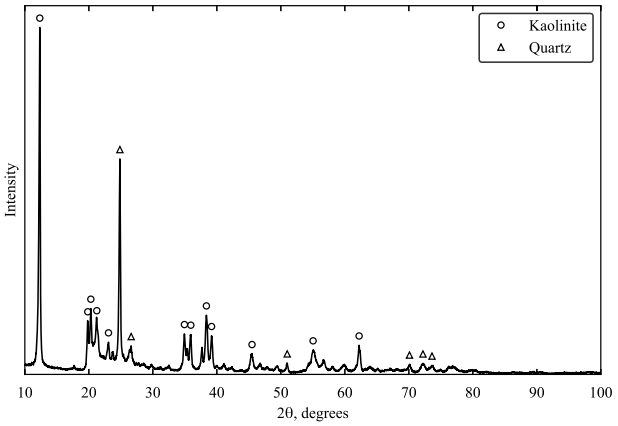

Fig. 2. X-ray diffraction data of the kaolin clay soil.

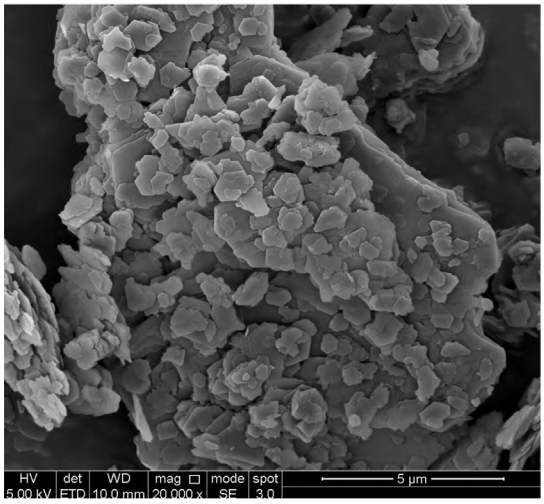

Fig. 3. SEM imagery of kaolin clay at 20,000x magnification.

ISO 14688–1:2002 [25]. Before insertion into the SEM, the kaolin clay soil undegoes gold (Au) coating. Fig. 3 illustrates the kaolin clay particles, revealing variations from low to medium sphericity, with subrounded to subangular shape, and smooth surface texture.

3.2. Specimen preparation

Kaolin clay samples were prepared using powdered speswhite kaolin clay, mixed with water at a 1:1 ratio to form a slurry. The slurry was consolidated in a pressurised cylindrical Rowe cell to 600 kPa, resulting in a fully saturated kaolin clay wheel with a 44% moisture content. Controlled drying was employed to study the effect of moisture content on the high-strain-rate behaviour of kaolin clay. Specimens with moisture content levels ranging from 0 to 44% were prepared, covering all saturation levels, from unsaturated to fully saturated [7]. For both unconfined and confined SHPB testing, the following procedure was followed to prepare the specimens:

1. Cylindrical kaolin clay samples with varying moisture content are made using a 25 mm stainless-steel cylinder slicer. The specimens have a nominal length of 5 mm and a diameter of 25 mm. The initial weight of the kaolin clay specimen is recorded immediately after it has been sliced.

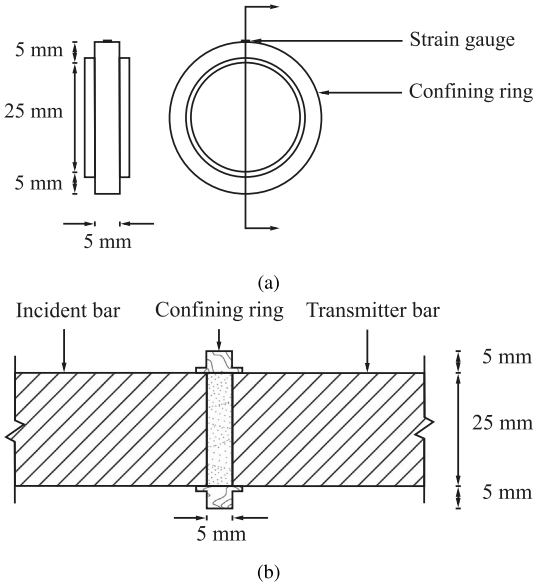

Fig. 5. Diagrams showing (a) the confining ring for confined SHPB testing and (b) the confining ring with the sample inside, ready for testing.

2. Samples are air dried in a temperature-controlled setting at 20 degrees Celsius, and weighed at regular intervals to measure their current moisture content based on their initial wet weight and current weight.
3. Cut and air-dried samples are then wrapped in polyvinylidene chloride to minimise changes in moisture content between sample preparation and testing. The samples are prepared and tested in different laboratories, hence they are stored in sealed plastic bags until required for testing.

3.3. Experimental setup

The SHPB experimental setup features a conventional pressure bar arrangement comprising a striker bar (350 mm), an incident bar

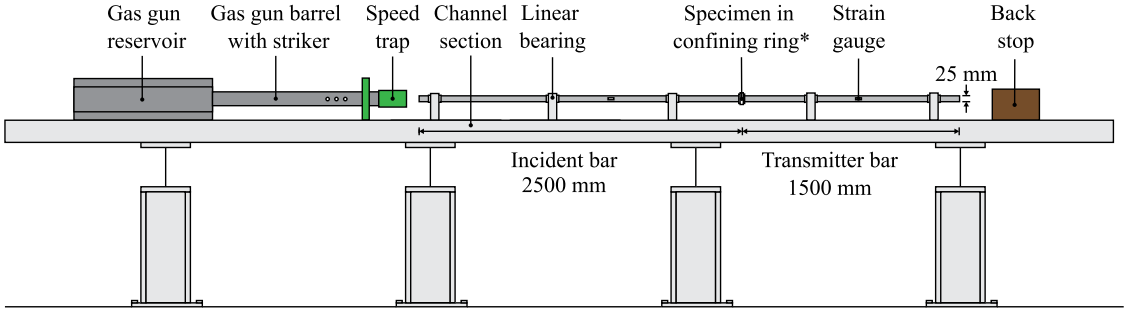

Fig. 4. Schematic diagram of the SHPB experimental setup with the confining ring [* removed for unconfined SHPB tests].

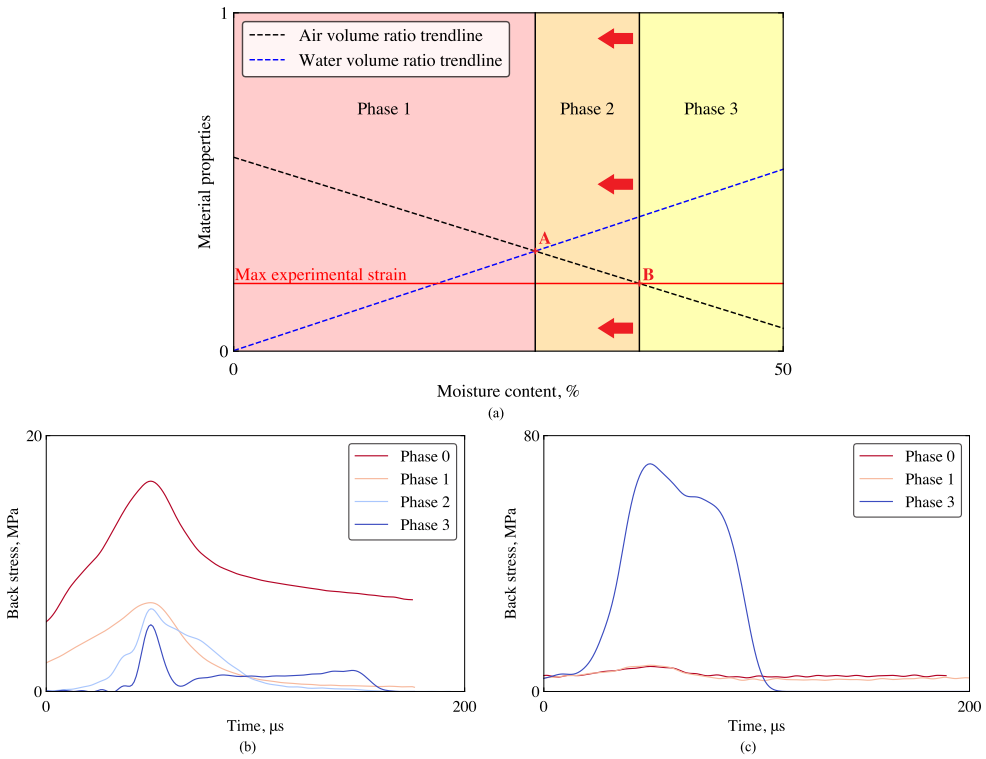

Fig. 6. (a) Phase behaviour observed in SHPB testing of kaolin clay, defined by air and water volume ratios and maximum experimental strain. The indicative stress transmission through the clay in each phase at 2800 s^{-1} is depicted for (b) unconfined and (c) fully confined test conditions.

(2500 mm) and a transmitter bar (1500 mm), each with a diameter of 25 mm (Fig. 4). For confined SHPB testing, a steel confining ring (Fig. 5a) was used to encase the sample between the incident and transmitter bars (Fig. 5b).

Pressure bar strain gauges signals were captured using a TiePie Handyscope four-channel digital oscilloscope with a 14-bit A-D resolution, a sampling frequency of 1 MHz, and a record length of 131.072 kSa. The material's axial stress response was monitored using a pair of Kyowa KSP-2-120-E4 strain gauges mounted on the incident and transmitter bars, while the radial stress was measured using a single strain gauge mounted on the outside of the confining ring. Strain gauge signals were collected from the incident and transmitter bars using a half Wheatstone bridge configuration, and from the confining ring using a quarter Wheatstone bridge configuration [7]. These signals were then used to calculate the stresses in the kaolin clay specimen using the conventional three-wave method [3,7].

3.4. Test programme

For unconfined SHPB testing, tests commenced at 8, 12 and 16 m/s on kaolin clay samples with moisture content ranging from 0 to 44%. Subsequent tests were conducted at 18, 20 and 22 m/s to assess the impact of higher strain rates. Correspondingly, peak average strain rates of 1200, 1900 and 2800 s^{-1}, were achieved for 8, 12 and 16 m/s, respectively [7].

For confined SHPB testing, tests started at 12 m/s on kaolin clay samples with moisture content ranging from 0 to 41% for behavioural comparison with the unconfined SHPB tests. Then, tests at 18, 20 and 22 m/s were conduced. Under confined conditions, the average peak strain rates were 2600, 2800 and 3100 s^{-1}, for 18, 20 and 22 m/s, respectively [7].

Under both testing conditions, the specimen's axial and radial stresses and strains are measured.

It is important to note that the test speed in these SHPB experiments is setup-specific, while the strain rate reflects what the sample experiences during testing.

3.5. Testing results

The SHPB test signals for the incident bar, transmitter bar and confining ring were processed using the open-source Python algorithm `SHPB_Processing.py` [28]. This algorithm incorporates the subroutine `dispersion.py`, which implements Tyas and Pope's dispersion-correction approach [29]. This ensures accurate representation of the sample's axial and radial stresses and strains during testing.

The behaviour of the tested kaolin clay samples can be categorised into four distinct "phases", each defined by its stress transmission characteristics, termed the "back stress" in SHPB tests [7]. These phases are delineated by the air volume ratio V_{air}/V, water volume ratio V_{water}/V, and the maximum experimental strain attained during the experiment, as illustrated in Fig. 6a:

- **Phase 0:** Encompasses all dry specimens, positioned on the y-axis in Fig. 6a.
- **Phase 1:** Comprises partially-saturated specimens where soil pores are primarily filled with air. The maximum moisture content for phase 1 is defined by the boundary formed by the intersection of the air and water ratio trendlines, denoted as Point A in Fig. 6a.
- **Phase 2:** Comprises partially-saturated specimens where soil pores are primarily filled with water. The upper limit of moisture content is defined by the intersection of the air volume ratio with the max experimental strain, marked as Point B in Fig. 6a.
- **Phase 3:** Defines experiments which begin partially-saturated, but reach full saturation during testing.

The behaviour of soil specimens vary across the four phases, depending on whether testing is unconfined or fully confined.

In unconfined conditions (Fig. 6b), each phase exhibits a distinct stress transmission pattern, with increased moisture content correlating with reduced peak stress due to enhanced lateral movement and specimen extrusion. While moisture content minimally impacts stress transmission within each phase, back stress transmission increases with

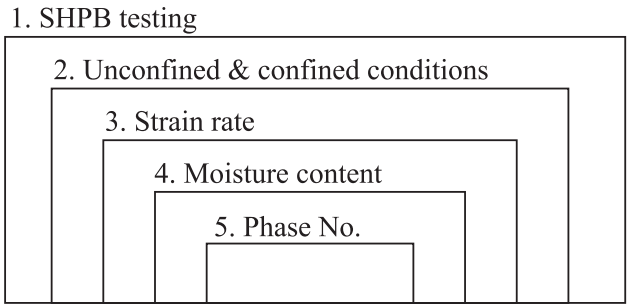

Fig. 7. Diagram of the classification of experimental SHPB test data.

rising strain rates.

In fully confined conditions (Figs. 6c), phases 0 and 1 primarily involve the compaction of the dry soil fraction, showcasing a similar shock absorption effect on transmitted stress. The attainment of high strains in confined SHPB tests eliminates phases 2 at higher strain rates, leading to an abrupt transition between phases 1 and 3. Upon reaching saturation, clay samples assume fluid-like behaviour with $\nu = 0.5$, resulting in a significantly heightened stress transmission.

4. Development of data-driven parametric models

4.1. Data-driven classification

The classification of the data was conducted based on both the experimental test results and the initial sample parameters.

The data obtained from the SHPB tests on the kaolin clay samples, along with their initial parameters, can be classified into five categories represented by a Russian doll model, as illustrated in Fig. 7: SHPB testing, unconfined and confined conditions, strain rate, moisture content, and phase number.

These parameters are ranked from 1 to 5, each denoting a different level of specificity in data classification. Parameter 1 encompasses all the tests, while parameter 5 represents a more specific subset. The parameters are defined as follows:

1. "SHPB testing" encompasses all tests conducted using the SHPB apparatus.
2. "Unconfined & confined conditions" categorises the tests which were performed under "unconfined" or "confined" conditions.
3. "Strain rate" indicates the tests were conducted at different strain rates: 1200, 1900, 2800 s^{-1} for unconfined and 2600, 2800 and 3100 s^{-1} for confined.
4. "Moisture content", reflects the varying moisture levels of the kaolin clay samples tested, ranging from 0 to 44%.
5. "Phase No". assigns a specific phase number (0, 1, 2 or 3) to each moisture content, as detailed in Section 3.5.

A total of 144 SHPB experimental tests were carried out on kaolin clay samples spanning moisture contents from 0 to 44%. Among these, 98 tests were performed under unconfined conditions, while 46 tests were conducted under fully confined conditions. The majority of the test data was utilised to develop the parametric models, with the exception of one test from each condition, which was reserved for subsequent model validation.

The insights gained into the high-strain-rate behaviour of cohesive soils, as discussed in Section 3.5, enabled the categorisation of the test data into separate confinement conditions and different phase behaviours, based on the sample's moisture content and strain rate. This physics-informed classification, was conducted manually, guided by test findings, with the strain rate identified as the key parameter influencing the material's response. The model simplifies from a polynomial regression to a linear analysis, with "strain rate", parameter 3, as the key parameter. The analysis within each material phase number showed that variations in moisture content had no discernible impact on the overall response. Consequently, the mean response was used for the parametric models. In summary, for unconfined conditions, the strain rate was 1200, 1900, 2800 s^{-1}, with four material phases, while for confined conditions, it was 2700, 2800 and 3100 s^{-1} with 3 material phases, with phase 2 disappearing at 2700 s^{-1} under confined conditions.

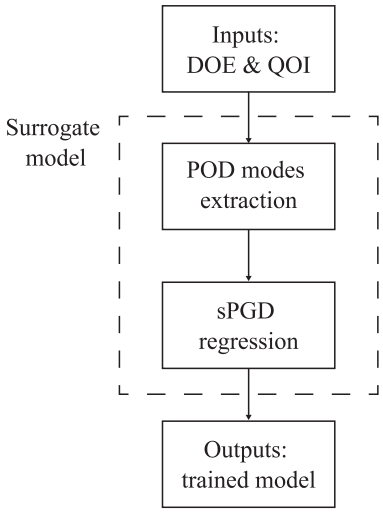

Fig. 8. POD_sPGD flowchart.

4.2. Parametric model

With the data classified into corresponding SHPB test conditions and phase numbers based on strain rate and moisture content, attention shifts to the parametric model. Fig. 8 details the structure of the parametric model used for training the test data, comprising inputs, a surrogate model, and outputs.

The input section focuses on the model's input parameters: the Design of Experiments (DOE) and Quantity of Interest (QOI). The DOE represents the parameter the model trains on, which in this case, is the key parameter: strain rate. Three strain rate values are utilised for each confinement condition. The QOI refers to the mean data properties of interest used for training against the DOE. For unconfined SHPB testing, there are five QOIs: strain, front stress, back stress, mid stress and time. Confined SHPB testing includes six QOIs: strain, front stress, mid stress, back stress and radial stress and time.

Data training for each model is facilitated by the surrogate model, made of two sections: POD mode extraction and sPGD regression. Since there is only one parameter, strain rate, POD mode extraction focuses on a single mode, mapping the data's curve behaviour along this parameter. Subsequently, sPGD regression utilises this information to predict the data's behaviour based on this relationship along the set parameter.

After passing through the surrogate model, the training data is saved, resulting in the creation of four different models for each phase number, as illustrated in Fig. 9.

4.3. Visualisation model

After training the tests data using the POD_sPGD model, the trained model results are saved according to their respective phase number. This process is illustrated in Fig. 9, termed as the test data structuring and model training flowchart.

This trained data is utilised for the visualisation model, as depicted in the visualisation flowchart shown in Fig. 9. The visualisation model

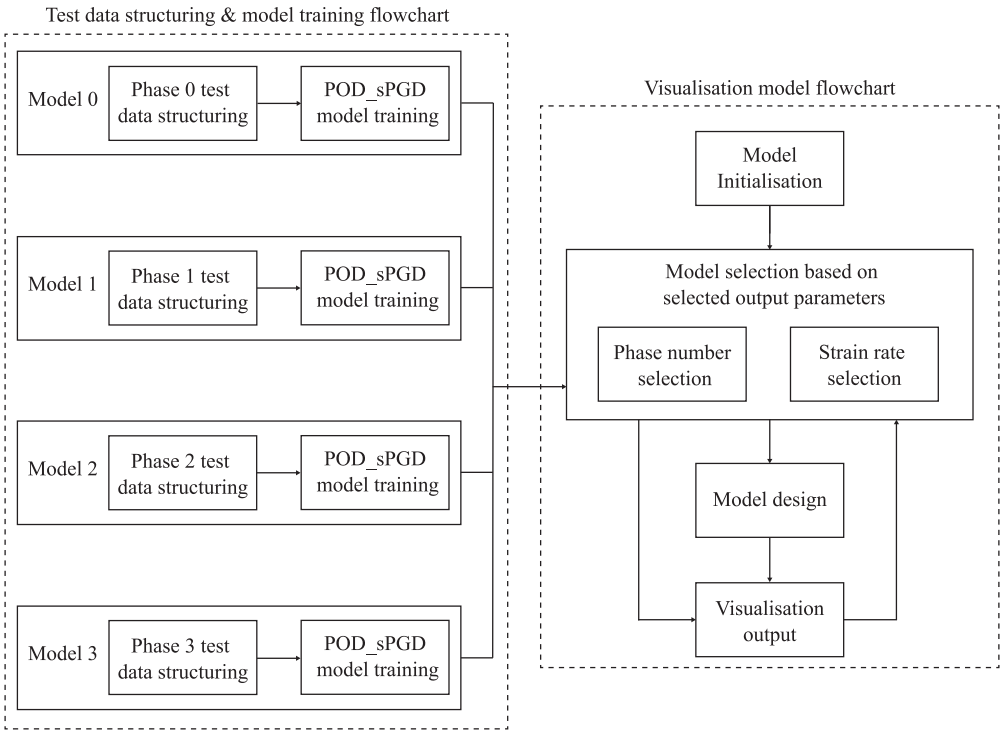

Fig. 9. Overall design flowchart of the data-driven parametric models.

consists of a four sections: model initialisation, model selection based on selected output parameters, model design and visualisation output. Model initialisation establishes the input parameters and widget architecture. The model selection section is made of the phase number and strain rate selection functions. It established the relationship between the trained POD_sPGD models for each phases number and the corresponding moisture content ranges obtained from testing. The model design encompasses interactive widget functionalities, while the visualisation output updates the visual plots based on selected moisture content and strain rate. As the selected strain rate changes, the model representing the data remains the same. However, altering moisture content prompts a transition between different material phase number, resulting in a change of the model number selected to represent the data.

The flowchart depicted in Fig. 9 guides the creation of two widgets for each SHPB test condition, using the trained mean data. These final interactive widgets, are displayed in Figs. 10 and 11. The unconfined SHPB test model features four plots, representing the four QOIs: strain, front stress, back stress and mid stress over time. Conversely, the confined SHPB test model includes five plots, representing the five QOIs: strain, front stress, mid stress, back stress and radial stress over time.

The plot axes in Figs. 10 and 11 adjust responsively based on the selected moisture content, corresponding to a specific phase at a given strain rate. The y-axis dynamically adjusts to accommodate the maximum potential value.

Confidence patches, depicting the 90% confidence interval based on the standard deviation of the test data, are included in Figs. 10 and 11 in light grey for both unconfined and confined SHPB test conditions. They are especially crucial for unconfined SHPB tests due to potential errors associated with testing.

The widgets in Figs. 10 and 11, feature sliders for strain rate and moisture content, facilitating adjustments between the different phase behaviours of cohesive soil. The sliders are user-friendly, allowing for quick and interactive changes. The selected values for moisture content and strain rate are displayed on the right hand side of each slider. Additionally, a point picking option is available above the sliders, enabling selection of specific moisture content and strain rate values, with the complete selected ranges next to them. Both options provide immediate visualisation of cohesive soil responses under the selected parameters, offering instant information.

4.4. Model validation

To validate the data-driven parametric models, a comparative evaluation against a random SHPB test is imperative. This involves testing a kaolin clay sample with a specific moisture content under both unconfined and confined conditions using the SHPB apparatus, at a specific strain rate. The test results in red are compared against the model predictions.

The additional unconfined SHPB test was conducted on a kaolin clay sample with a moisture content of 11.32% at a speed of 12.0 m/s, corresponding to a strain rate of 1996 s^{-1}. Similarly, the confined SHPB test involved a kaolin clay sample with a moisture content of 21.36% at a speed of 20.0 m/s, resulting in a strain rate of 2976 s^{-1}.

The comparative results between the data-driven parametric models and the additional SHPB tests are depicted in Figs. 12 and 13, where the supplementary tests are plotted in red on the interactive visual widgets. When comparing the newly performed tests and the data-driven parametric models on the visual widgets, the predicted representation exhibits remarkable accuracy, closely resembling the expected material response seen in testing. In both test conditions, the maximum amplitude is slightly higher, which is expected as the model represents the mean behaviour. The confidence interval around the black line is crucial, indicating the possible range of sample behaviour. Therefore, it is prudent to base designs on the maximum value.

To enhance the model validation process, a quantitative comparison was conducted between the experimental data and the model predictions for both unconfined and confined SHPB test conditions, as shown in Figs. 12 and 13. Metrics such as the coefficient of determination (R^2) and root mean square error (RMSE) were computed for each plot, comparing the model's predictions to the experimental results (Table 2). This evaluation provides objective evidence of the model's accuracy and reliability, enabling a thorough assessment of its performance in

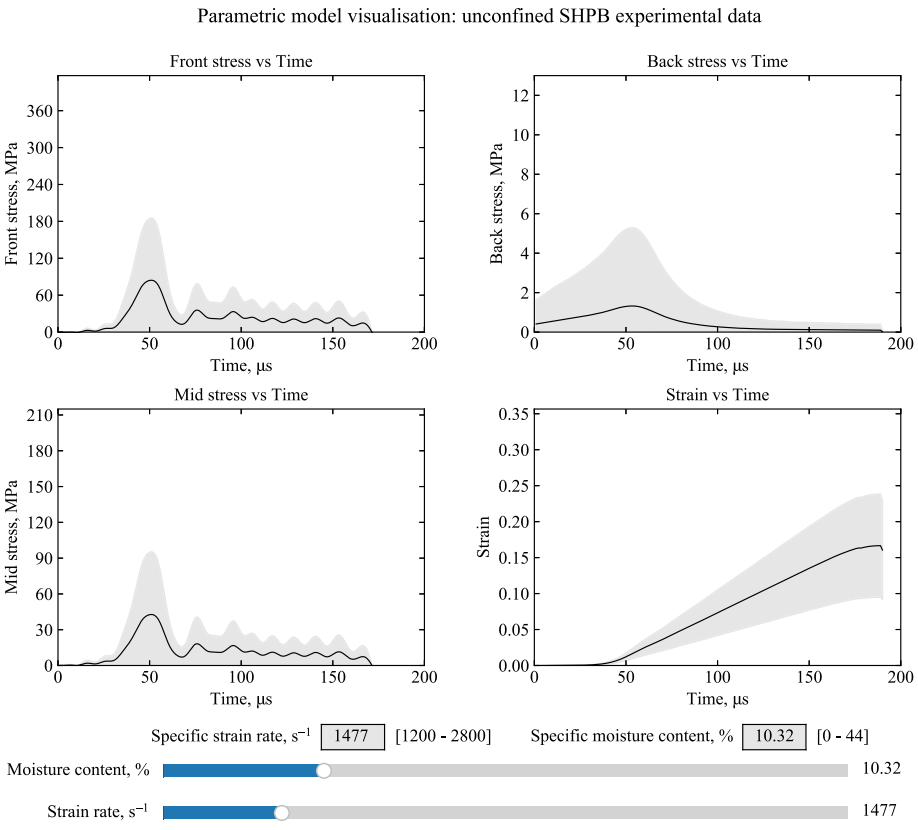

Fig. 10. Parametric model visualisation widget for unconfined SHPB experimental data.

predicting the high-strain-rate behaviour of cohesive soils.

As depicted in Table 2, R^2 values demonstrate a high level of agreement, indicating a strong correlation between the model and experimental data. Meanwhile, RMSE values are approaching 0 and are less than 10%–20% of the data range, further confirming the model's accuracy against the validation test data. This underscores the model's reliability and minimal variance between the predictions and the actual data. Notably, a single R^2 value is measured, as the comparison is made along the y-axis, with both curves sharing the same x-range.

This consistency across various test conditions reinforces reliability, enabling engineers to make informed decisions based on accurate assessments of high-strain-rate behaviour. The precision not only validates the models but also enhances their applicability in real-world scenarios, ensuring they effectively guide the design and implementation of solutions to complex engineering challenges.

5. Discussion

Experimental SHPB tests were performed on kaolin clay samples, covering a range of moisture contents from 0 to 44%. These tests were conducted under both unconfined and confined conditions, totalling 144 experiments, with 2 tests reserved for model validation. Of these, 98 tests were completed under unconfined conditions, while 46 tests were carried out under confined conditions. Analysis of the results revealed distinct phase behaviours of cohesive soils under high-strain-rate conditions, guiding the development of the parametric models.

Numerous constitutive models in LS-DYNA have been evaluated for modelling the high-strain-rate behaviour of cohesive soils, but they showed significant limitations [7]. As an alternative, parametric modelling options were explored, with the POD-sPGD method chosen to build these data-driven models. While these models are often viewed as simple predictive tools, their true benefit lies in supplementing existing experimental data, providing a more comprehensive dataset for calibrating and assessing future numerical models. This approach bridges the gap between Split-Hopkinson Pressure Bar (SHPB) testing and numerical simulations, emphasising the need for both. Parametric modelling offers a robust method to better integrate SHPB test data with numerical models, improving our ability to interpret and calibrate high-strain-rate behaviour.

In contrast to traditional machine learning approaches, such as artificial neural networks (ANNs), the parametric modelling approach using Proper Orthogonal Decomposition (POD) combined with the sPGD method does not require intensive training cycles, validation processes, or hyper parameter tuning. Unlike ANNs, which rely on iterative training over multiple epochs and the optimisation of loss functions such as mean squared error (MSE), the POD-sPGD method directly leverages experimental data to construct a reduced-order model without the need for cumbersome training schemes. This provides a clear advantage in terms of efficiency, making it more suitable for applications where large-scale training datasets or computationally intensive processes are impractical. This distinction underscores the simplicity and computational efficiency of POD-based approaches in comparison to traditional machine learning techniques, offering a streamlined yet powerful alternative for generating accurate models from limited experimental data.

The parametric models developed in this study were designed without reliance on implemented machine learning Python libraries, ensuring transparency and avoiding a black box scenario. Instead, the machine learning model was developed based on mathematical principles, with the coupling of POD modes extraction and sPGD regression methods proving invaluable in creating these parametric models. The foundation for building these models was laid upon novel SHPB test

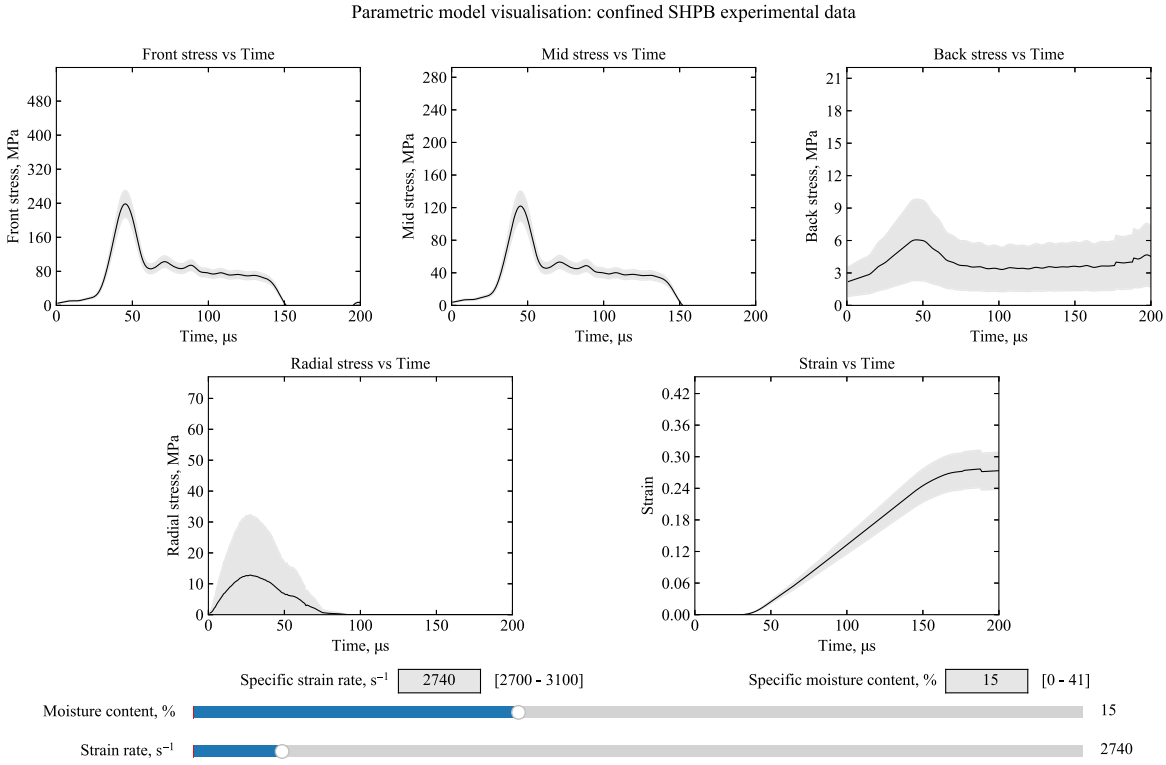

Fig. 11. Parametric model visualisation widget for confined SHPB experimental data.

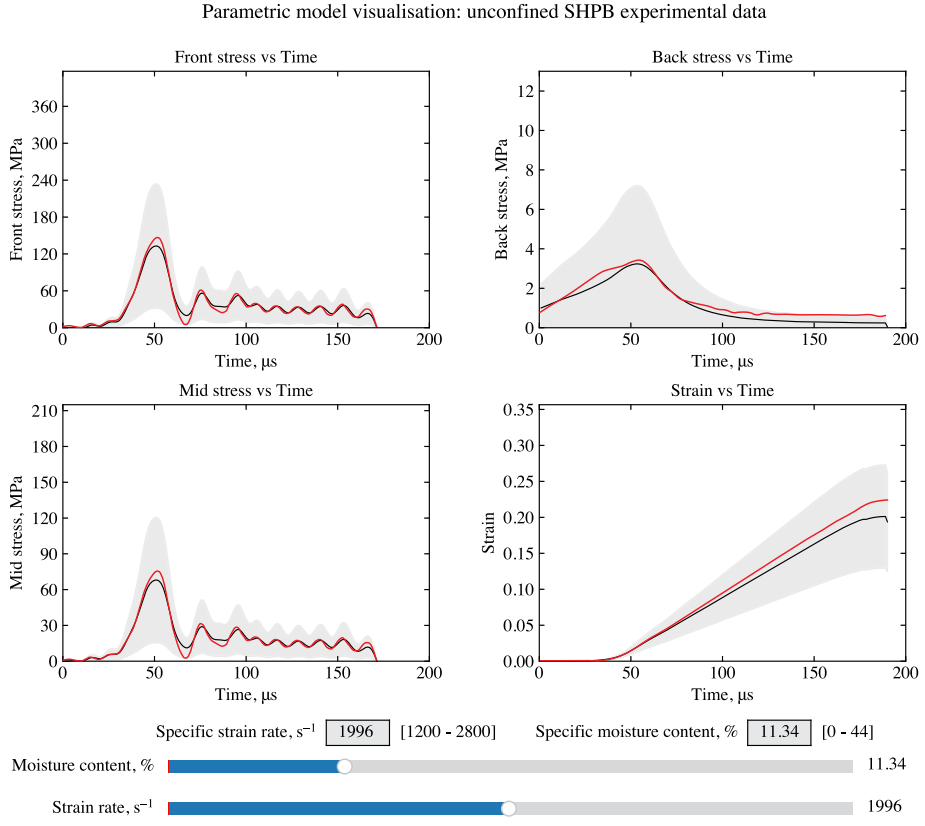

Fig. 12. Parametric model validation: comparison of an unconfined SHPB test (in red) and its data-driven parametric model.

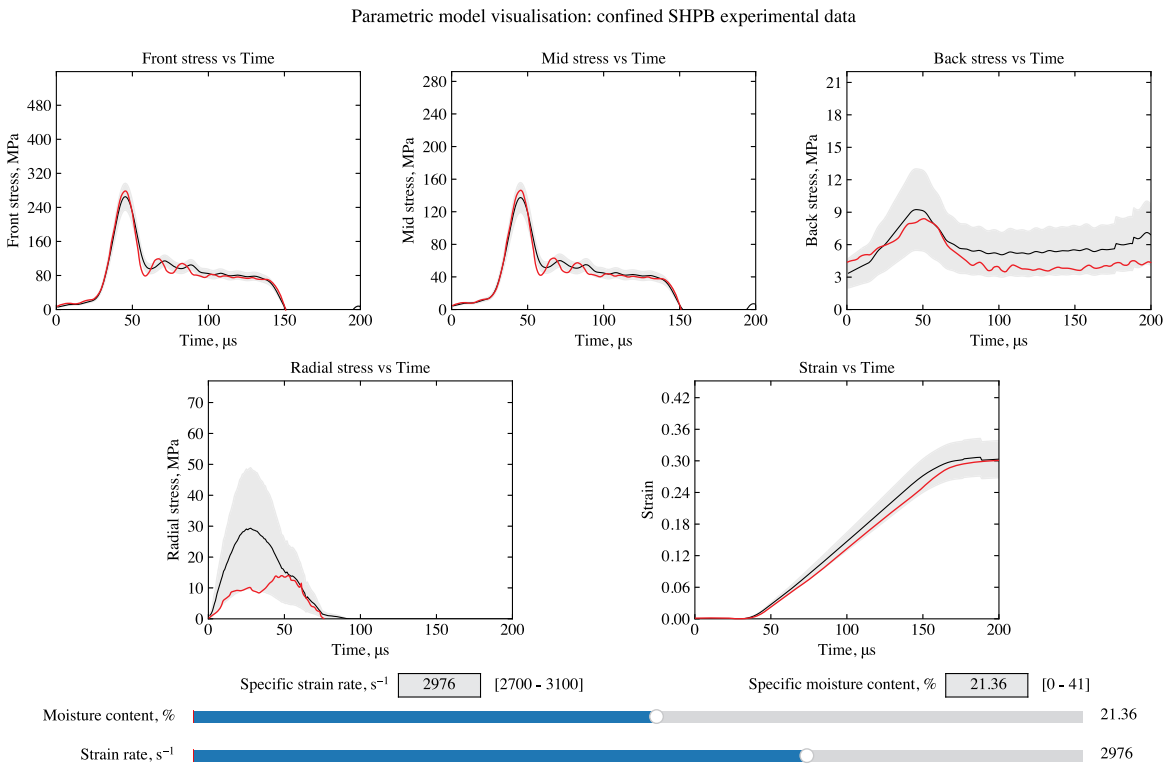

Fig. 13. Parametric model validation: comparison of a confined SHPB test (in red) and its data-driven parametric model.

Table 2
Comparison of experimental and parametric model results (versus time): R^2 and RMSE values.

	Unconfined SHPB data: model vs experimental results				
	Front stress	*Back stress*	*Mid stress*	*Strain*	*Radial stress*
R^2	0.73	0.87	0.70	0.97	–
RMSE	9.4	12.5	10.2	4.4	–
	Confined SHPB data: model vs experimental results				
	Front stress	*Back stress*	*Mid stress*	*Strain*	*Radial stress*
R^2	0.71	0.76	0.68	0.99	0.59
RMSE	11.1	12.8	11.5	2.2	26.8

data.

The development of the models was guided by physics-informed classification of test data. Clear delineation of test data into different phase numbers based on confinement, moisture content, and strain rate behaviour eliminated the need for classification algorithms. Constructed through POD modes extraction and sPGD regression algorithms, the surrogate model played a vital role in producing trained data representing each phase number observed in material testing. The model's architecture was predominantly shaped by the strain rate, identified as the key parameter.

A vital aspect for understanding the trained data from the parametric models is a clear visual representation. To fulfil this need, interactive widgets were crafted for both testing scenarios, offering high responsiveness and user-friendliness. These widgets offer instantaneous responses and harness the trained data from the parametric model.

In the current models, stress-time plots were generated, as this approach is more aligned with practical, on-site applications and better suited for evaluating the model's effectiveness. However, if needed, the validation model can also be adapted to use strain, instead of time, allowing for the generation of stress–strain curves.

Model validation involved comparing the trained model data, depicted in black, with a distinct SHPB test, highlighted in red. Unconfined tests demonstrated notable accuracy, with confined tests, though fewer, also displaying considerable precision. Unconfined tests featured larger confidence intervals due to their susceptibility to testing errors, whereas confined tests showcased smaller confidence intervals owing to their high consistency. Further refinement of the confined parametric model can be achieved through additional testing. Nonetheless, at this stage, the model proves sufficiently precise, offering significant time and cost savings compared to traditional testing. Overall, the parametric model effectively predicts the high-strain-rate behaviour of cohesive soils under unconfined and confined conditions within the tested range.

The machine learning tools (i.e. the POD and sPGD) used in the development of these data-driven parametric models face several limitations. Their performance is highly dependent on the quality of the data collected, and their efficiency is influenced by the quantity of available data based on the chosen experimental approach. In this case, the limited number of tests conducted for each moisture content and strain rate under different confinement conditions has an impact on the accuracy and robustness of the models. Additional testing would further improve the quality of these models.

The data-driven methods employed in this paper can be applied to a wide range of scenarios, provided that data has been collected and key parameters influencing the collected outputs have been identified. Naturally, the more comprehensive the dataset and the greater the number of influential parameters, the more fine-tuning the model will require for optimal performance.

6. Summary

Utilising cutting-edge machine learning polynomial regression techniques such as the POD and sPGD, engineers can now access two separate data-driven parametric models, offering rapid predictive capabilities for the behaviour of cohesive soils under various loading conditions and material parameters.

Experimental SHPB tests were conducted on kaolin clay samples, covering a range of moisture contents from 0 to 44%, under both unconfined and fully confined conditions. Through these tests, the degree of confinement, moisture content and strain rate were identified as the significant factors influencing the specimen's behaviour. This dataset served as the foundation for testing and training the parametric models.

The work undertaken in classifying the data was instrumental, as it demonstrated the feasibility of organising the findings within each phase number under specific confinement conditions and strain rates. Moreover, given the similar behaviour observed among test results within each phase number, the mean behaviour was used and represented. Strain rate emerged as the key parameter within each individual phase, effectively simplifying the complexity of the problem from nonlinear to linear. The POD analysis was therefore conducted along a single mode with strain rate as the key parameter, and sPGD modelling was executed within each distinct phase number, for each confinement state.

All tests, except for one from each test condition, were utilised for model construction, while the reserved tests were set aside for validation, allowing for comparisons with actual experiments, underscoring the accuracy of the two data-driven parametric models.

Interactive visual widgets were tailored for each test condition, enriching the user experience and streamlining the utilisation of these data-driven parametric models to empower engineers to respond promptly and effectively to threats.

For future endeavours, expanding the scope to include lower and higher strain rates would be beneficial, providing a more comprehensive understanding of the behaviour of cohesive soils at high-strain-rates. Nonetheless, the current results already offer substantial time and cost savings. This would include applications such as designing effective soil-filled barriers without the need for additional extensive experimental testing.

It is envisioned that these models will empower engineers to safeguard personnel and infrastructure from a variety of threats through the rapid assessment of cohesive soil properties. The POD-sPGD approach offers remarkable versatility and adaptability compared to traditional machine learning models such as ANNs, allowing for quick adjustments to reflect new data without extensive training or hyper parameter tuning. This efficiency makes the models ideal for practical applications, such as designing soil-filled barriers and enhancing infrastructure protection.

CRediT authorship contribution statement

Arthur Van Lerberghe: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Angelo Pasquale:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Conceptualization. **Sebastian Rodriguez:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **Andrew D. Barr:** Writing – review & editing, Writing – original draft. **Sam D. Clarke:** Writing – review & editing, Supervision. **Dominique Baillargeat:** Writing – review & editing, Supervision, Conceptualization. **Francisco Chinesta:** Writing – review & editing, Supervision, Conceptualization.

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

This research was funded by the Engineering and Physical Sciences Research Council (EPSRC), and the Defence Science and Technology Laboratory (Dstl).

Authors also acknowledge the support of the ESI Group through its research chair CREATE-ID at Arts et Métiers ParisTech. This research is part of the program DesCartes and is supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) program.

Data availability

Data will be made available on request.

References

[1] Ross C Allen, Nash Phillip T, Friesenhahn Gerard J. Pressure waves in soils using a split-Hopkinson pressure bar. Technical report, Southwest Research Institute San Antonio TX; 1986.
[2] Veyera George E. Uniaxial stress-strain behavior of unsaturated soils at high strain rates. Technical report, Wright Lab Tyndall Afb Fl; 1994.
[3] Gray III George T. Classic split Hopkinson pressure bar testing. ASM Handb 2000;8:462–76.
[4] Bragov AM, Lomunov AK, Sergeichev IV, Tsembelis K, Proud WG. Determination of physicomechanical properties of soft soils from medium to high strain rates. Int J Impact Eng 2008;35(9):967–76.
[5] Luo Huiyang, Cooper William L, Lu Hongbing. Effects of particle size and moisture on the compressive behavior of dense eglin sand under confinement at high strain rates. Int J Impact Eng 2014;65:40–55.
[6] Barr Andrew D. Strain-rate effects in quartz sand [Ph.D. thesis], University of Sheffield; 2016.
[7] Van Lerberghe Arthur, Li Kin Shing O, Barr Andrew D, Clarke Sam D. High strain rate behaviour of cohesive soils. International Journal of Impact Engineering 2025;198:105189.
[8] Hastie Trevor, Tibshirani Robert, Friedman Jerome. The elements of statistical learning. Springer series in statistics, New York, NY, USA: Springer New York Inc; 2001.
[9] Smola AJ, Schölkopf B. A tutorial on support vector regression. Stat Comput 2004;14(3):199–222.
[10] Breiman Leo, Friedman Jerome, Stone Charles J, Olshen RA. Classification and regression trees. Chapman and Hall/CRC; 1984.
[11] Breiman L. Random forests. Mach Learn 2001;45:5–32.
[12] Friedman Jerome H. Greedy function approximation: A gradient boosting machine. Ann Statist 2001;29(5):1189–232.
[13] Goodfellow Ian, Bengio Yoshua, Courville Aaron. Deep learning. MIT Press; 2016, http://www.deeplearningbook.org.
[14] Pasquale Angelo, Champaney Victor, Kim Youngtae, Hascoët Nicolas, Ammar Amine, Chinesta Francisco. A parametric metamodel of the vehicle frontal structure accounting for material properties and strain-rate effect: application to full frontal rigid barrier crash test. Heliyon 2022;8(12).
[15] Champaney Victor, Pasquale Angelo, Ammar Amine, Chinesta Francisco. Parametric curves metamodelling based on data clustering, data alignment, POD-based modes extraction and PGD-based nonlinear regressions. Front Mater 2022;9:904707.
[16] Benner Peter, Ohlberger Mario, Cohen Albert, Willcox Karen. In: Benner Peter, Ohlberger Mario, Cohen Albert, Willcox Karen, editors. Model reduction and approximation. Philadelphia, PA: Society for Industrial and Applied Mathematics; 2017.
[17] Rozza Gianluigi, Hess Martin, Stabile Giovanni, Tezzele Marco, Ballarin Francesco, Gräßle Carmen, et al. Model order reduction: Volume 3: Applications. De Gruyter; 2020.
[18] Rozza Gianluigi, Hess Martin, Stabile Giovanni, Tezzele Marco, Ballarin Francesco, Gräßle Carmen, et al. Volume 2 snapshot-based methods and algorithms. De Gruyter; 2020.

[19] Ibáñez Pinillo Rubén, Abisset-Chavanne Emmanuelle, Ammar Amine, González David, Cueto Elías, et al. A multidimensional data-driven sparse identification technique: The sparse proper generalized decomposition. Complexity 2018;2018:1–11.
[20] Sancarlos Abel, Champaney Victor, Cueto Elias, Chinesta Francisco. Regularized regressions for parametric models based on separated representations. Adv Model Simul Eng Sci 2023;10.
[21] Champaney Victor, Chinesta Francisco, Cueto Elias. Engineering empowered by physics-based and data-driven hybrid models: A methodological overview. Int J Mater Form 2022;15.
[22] Pasquale Angelo, Champaney Victor, Kim Youngtae, Hascoët Nicolas, Ammar Amine, Chinesta Francisco. A parametric metamodel of the vehicle frontal structure accounting for material properties and strain-rate effect: application to full frontal rigid barrier crash test. Heliyon 2022;8:e12397.
[23] Schmid Alexander, Pasquale Angelo, Ellersdorfer Christian, Champaney Victor, Raffler Marco, Guévelou Simon, et al. PGD based meta modelling of a lithium-ion battery for real time prediction. Front Mater (Comput Mater Sci) 2023;10.
[24] Schmid Alexander, Pasquale Angelo, Ellersdorfer Christian, Raffler Marco, Champaney Victor, Ziane Mustapha, et al. Mechanical characterization of Li-ion cells and the calibration of numerical models using proper generalized decomposition. Proc ASME 2023 Int Mech Eng Congr Expo IMECE2023 2023.
[25] ISO. 14688-1: 2002: Geotechnical investigation and testing – identification and classification of soil – part 1: Identification and description. Br Stand Inst 2002.
[26] BSI. BS 1377-2: 1990: Methods of test for soils for civil engineering purposes – Part 2: Classification tests. UK: BSI London; 1990.
[27] Casagrande Arthur. Classification and identification of soils. Trans Am Soc Civ Eng 1948;113(1):901–30.
[28] Van Lerberghe Arthur, Li Kin Shing O, Barr Andrew D, Clarke Sam D. An open-source algorithm for correcting stress wave dispersion in split-hopkinson pressure bar experiments. Sensors 2025;25(1):281.
[29] Tyas Andrew, Pope Dan J. Full correction of first-mode Pochammer–Chree dispersion effects in experimental pressure bar signals. Meas Sci Technol 2005;16(3):642.