ORIGINAL PAPER

Mathematical Modelling of Peak and Residual Shear Strength of Rough Rock Discontinuities Using Continued Fractions

Olivier Buzzi[1] · Michael Jeffery[1] · Pablo Moscato[2] · Rafael Bartnik Grebogi[2] · Mohammad Nazmul Haque[2]

Abstract

Estimating the shear strength of large in situ rock discontinuities is often required to assess the stability of rock masses. This estimation is, however, complicated by the well-known scale effect and the fact that the discontinuity surfaces are only partially accessible through traces. A new approach [referred to as the stochastic approach for discontinuity shear strength (StADSS)] was recently presented to address these two points. This approach relies on a random field model and a semi-analytical shear strength model, the latter of which is referred to as the NDSS (Newcastle discontinuity shear strength) model. The NDSS model has to be implemented as a numerical code, and because the StADSS model is a Monte Carlo approach with hundreds if not thousands of simulations, the computational time to obtain a shear strength distribution is not negligible. The objective of this study is to find an efficient alternative to the NDSS model in the form of a continued fraction model that can predict the sheared area within a rough discontinuity subjected to direct shearing under constant normal stress as a function of the material strength, normal stress applied to the discontinuity and the standard deviation of asperity gradients (as the difference in elevation of two points of the surface over the horizontal distance between these points) of the surface. Using a 10/90 training/testing split of the dataset, a memetic algorithm-based truncated continued fraction regression (CFR) model was formulated. The distribution of CFR predictions was found to be very close to that of the dataset used for training. Then, the CFR model was tested against experimental data of the sheared area and shear strength (peak and residual) obtained from small (90 mm per 90 mm) and large (2 m per 2 m) specimens. It was found that 75% of the predictions fall within 20% of the experimental values. The continued fraction regression model can be used as an efficient alternative to the semi-analytical NDSS model, provided that it is used within the bounds of variables used to establish it.

Highlights

- A mathematical model was established to predict shear strength of rock joints
- The model is based on continued fraction regression (CFR) and memetic algorithm
- The CFR model was validated against data of peak and residual shear strength
- 75% of the CFR predictions fall within 10% of the experimental data
- The CFR model can be implemented as a spreadsheet with a low computational cost

✉ Olivier Buzzi
  Olivier.Buzzi@newcastle.edu.au

Michael Jeffery
  Michael.Jeffery@uon.edu.au

Pablo Moscato
  Pablo.Moscato@newcastle.edu.au

Rafael Bartnik Grebogi
  Rafael.BartnikGrebogi@uon.edu.au

Mohammad Nazmul Haque
  Mohammad.Haque@newcastle.edu.au

[1] School of Engineering (Civil Engineering), The University of Newcastle, University Drive, Callaghan, NSW 2308, Australia

[2] School of Information and Physical Sciences, The University of Newcastle, University Drive, Callaghan, NSW 2308, Australia

1 Introduction

Rock fractures, also referred to as discontinuities, constitute zones of weakness in rock masses, whose stability is often governed by the geometrical and mechanical properties of the fractures. Quantifying the shear strength of large in situ discontinuities is hence an essential yet complicated step in the stability analysis of rock masses or rock slopes. The difficulties of such quantification lie in the fact that the information about the discontinuity surface is only partial (visible on a trace) and that there exists a scale effect.

In the past, several models have been proposed to predict the shear strength of large, rough, in situ discontinuities, with the model by Barton (1976) being the most commonly used.

Recently, some of the authors of this paper proposed a new approach, named the StADSS (stochastic approach for discontinuity shear strength), combining a random field model (to create synthetic rock surfaces) and a semi-analytical shear strength model (to obtain shear strength estimates), referred to as the NDSS (Newcastle discontinuity shear strength) model. This approach was validated on a laboratory scale and satisfactorily applied to one rock discontinuity specimen of 2 m per 2 m (Jeffery et al. 2022). This approach is quite versatile in nature, in the sense that any suitable random field model and shear strength model can be used.

Although the NDSS model proved to yield satisfactory shear predictions, it has to be implemented as a numerical code because of the large number of data points on a surface that must be processed. In addition, because the rationale of the StADSS is to test many surfaces, obtaining the distribution of strength requires considerable time, even more so for large surfaces. In this paper, an efficient alternative to the implementation of the NDSS model is presented in the form of a multivariate regression model trained and validated using a dataset of NDSS predictions. More specifically, it is proposed to adopt a continued fraction regression (CFR) approach, which possesses an excellent ability to capture nonlinear regression problems. For example, CFR models were applied to 352 datasets from the physical sciences (Moscato et al. 2020b), and they were used to predict the critical temperature of superconducting materials (Moscato et al. 2020a) and to model the first performance dates of Shakespearean era plays (Moscato et al. 2022). In addition, in a large experiment involving 94 datasets and 15 other regression algorithms, the CFR method provided an analytic mathematical model for each dataset studied and outperformed many state-of-the-art regression approaches (Moscato et al. 2021).

In Moscato et al. (2023), this approach was used to find a mathematical model of the energy of n electrons on a sphere, i.e., a function that closely approximates the value of a solution to the Thomson problem for different values of n.

Unlike more conventional regression analysis approaches, it is possible to rigorously optimise the CFR coefficients to obtain more accurate predictions and owing to the memetic algorithm procedure, it is also possible to eliminate variables that only have a marginal effect on the predictions, hence yielding simpler model formulations. This means that the selection of variables is based on a rigorous and iterative mathematical procedure rather than a user's subjective choice. The objective of this study is to rigorously train and test a CFR model that can replace the NDSS model to avoid the difficulty of numerically implementing the model and the high computational time.

2 Estimation of the Shear Strength and the CFR Model

2.1 Shear Strength Estimation

This study is based on the Newcastle discontinuity shear strength (NDSS) model proposed by Casagrande et al. (2018). The model analyses 3D digitised surfaces, described as a structured dataset of positions (X, Y, Z), with X and Y being the directions of the mean discontinuity plane and Z being the elevation of the data points. For a given surface, the number of data points used to describe the morphology of the surface depends on the spatial resolution, i.e., the spacing between data points in the X and Y directions. To be processed by the NDSS model, the 3D surface needs to be triangulated, with the triangles herein referred to as facets (see "Appendix" section for more information on triangulation and the definition of facets).

The model provides a semi-analytical prediction of the peak and residual shear strength from an estimate of the actual contact area within a discontinuity, quantified by the number of facets that actively contribute to the generation of the shear resistance, denoted N_{CF}. Through research, it has established that only the steepest areas of the surface facing the direction of shearing contribute to the shear response (Grasselli 2001; Grasselli et al. 2002; Jeffery et al. 2022). The main variables and equations of the model are recalled when relevant to this paper, and more detailed information is provided in "Appendix" section. However, the reader should consult Casagrande et al. (2018) for a complete description of the NDSS model.

The objective of this study is to define and validate a multivariate regression (MVR) model to predict the peak and residual shear strength of a rough nonplanar rock discontinuity. Instead of predicting the peak and residual shear strength directly, which would require one MVR model for the peak strength and one for the residual strength, it is proposed to train a single MVR model to predict the number of contributing facets (N_{CF}), from which it is possible to compute both the peak and residual shear strengths as follows (see "Appendix" section for derivations and figures, as well as Casagrande et al. 2018 for more information):

$$\tau_p = \frac{A_{ip} \cdot \left(c + \sigma'_{local\_i} \tan\left(\phi'\right) \right) \cdot N_{CF}}{A_{tot}} \qquad (1)$$

and

$$\tau_r = \tau_p - \frac{A_{ip} \cdot c \cdot N_{CF}}{A_{tot}} \qquad (2)$$

where τ_p and τ_r are the peak and residual shear strength, respectively, and A_{ip} and A_{tot} are the area of a triangular facet projected on the discontinuity plane and the total projected discontinuity area, respectively. For 100 mm per 100 m samples, as tested by Casagrande et al. (2018), A_{tot} is equal to 10,000 mm^2. σ'_{local_i} is the local normal effective stress on a contributing facet, and ϕ' and c are the effective friction angle and cohesion of the intact rock, respectively. See "Appendix" section for the derivation of σ'_{local_i} and Eqs. (1) and (2).

N_{CF} is the number of contributing facets on the surface. The total number of triangular facets on the surface depends on the size of the surface and the spatial resolution, as p.

Note that the material strength is defined here in terms of Coulomb parameters (cohesion and friction angle), but the strength envelope of rocks is often nonlinear and best fitted with a Hoek−Brown criterion. In such cases, the cohesion and friction angle are derived from the Hoek−Brown criterion for a given stress state, as per Hoek (1983):

$$c = \left(\frac{1}{\sin(\phi')} - 1 \right) \frac{\cos(\phi') m_i \sigma_{ci}}{8} - \sigma'_{local\_i} \tan(\phi') \qquad (3)$$

where m_i and σ_{ci} are the material parameter and intact unconfined compressive strength of the Hoek–Brown criterion obtained by fitting experimental triaxial data, and ϕ' is defined as:

$$\phi' = \mathrm{atan}\left(\frac{1}{\sqrt{4h \times \cos^2(\theta) - 1}} \right) \qquad (4)$$

where $h = 1 + \frac{16\left(m_i \sigma'_{local\_i} + \sigma_{ci}\right)}{3\sigma_{ci} m_i^2}$ and $\theta = \frac{1}{3}\left(90 + \mathrm{atan}\left(\frac{1}{\sqrt{h^3 - 1}} \right) \right)$. Casagrande et al. (2018) showed that for a given shearing direction, the shear strength prediction by the NDSS model depends on the material strength (m_i, σ_{ci} and the basic friction angle ϕ_b), the effective normal stress applied to the discontinuity (σ'_n) and the morphology of the discontinuity surface. In research conducted by Jeffery et al. (2021), it was further shown that for a stationary surface, the shear strengths (peak and residual) depend on the standard deviation of asperity gradients rather than on the standard deviation of asperity heights. Consequently, it is proposed that a

multivariate regression model that relates N_{CF} to m_i, σ_{ci}, ϕ_b, σ'_n and sd_i be established.

Regarding notations, in previous publications by the authors (Casagrande et al. 2018; Buzzi and Casagrande 2018; Jeffery et al. 2021, 2022, 2023), the standard deviation of gradients is denoted σ_i. The notation is here changed to sd_i to avoid confusion with stress variables, i.e., σ_{ci} and σ'_n.

2.2 Continued Fraction Regression (CFR) Model

Continued fraction regression (CFR) is a multivariate regression method that was first introduced by Sun and Moscato (2019) and uses an analytic continued fraction representation for the unknown target function one is trying to model. Mathematically, an analytic function is defined as a function approximated as a power series expansion. One of the reasons why the use of continued fraction regression has been investigated is that it can provide a simple analytical model with low complexity for predicting a target function. To define an analytic continued fraction, consider an array $\boldsymbol{x}$ of n variables x_i:

$$\boldsymbol{x} = \{x_1, x_2, \dots, x_n\} \qquad (5)$$

The multivariate function $f(\boldsymbol{x})$ can be represented as an analytic continued fraction that reads:

$$f(\boldsymbol{x}) = g_o(\boldsymbol{x}) + \cfrac{h_o(\boldsymbol{x})}{g_1(\boldsymbol{x}) + \cfrac{h_1(\boldsymbol{x})}{g_2(\boldsymbol{x}) + \cfrac{h_2(\boldsymbol{x})}{g_3(\boldsymbol{x}) + \dots}}} \qquad (6)$$

where $g_i(\boldsymbol{x})$ and $h_i(\boldsymbol{x})$ are $g_i(x) = \boldsymbol{a}_i^T \boldsymbol{x} + \alpha_i$ and $h_i(x) = \boldsymbol{b}_i^T \boldsymbol{x} + \beta_i$, respectively. Each function $g_i(\boldsymbol{x})$ is associated with a vector $\boldsymbol{a}_i \in \mathbb{R}^n$ and a constant $\alpha_i \in \mathbb{R}$. The same applies to each function $h_i(\boldsymbol{x})$ associated with a vector $\boldsymbol{b}_i \in \mathbb{R}^n$ and a constant $\beta_i \in \mathbb{R}$. The depth of the CFR function, denoted d, refers to the largest subscript of function $g(\boldsymbol{x})$, i.e., $g_d(\boldsymbol{x})$. Note that at depth d, without losing generality, $h_d(\boldsymbol{x})$ is assumed to be equal to 0.

To define the CFR function, one needs (1) to identify the variables (x_i) that influence the output and (2) to define the coefficients a_i and b_i for the functions $g_i(\boldsymbol{x})$ and $h_i(\boldsymbol{x})$ that yield the best goodness of fit [quantified via the mean square error (MSE)] between the prediction and training datasets. The aim of continued fraction regression is to provide a simple and relatively interpretable analytical model of low complexity for predicting a variable that is a nonlinear function of several variables. Defining the final CFR function also referred to as the CFR model, is achieved by applying a search and optimisation procedure (a memetic algorithm) to training datasets. All technical aspects of the memetic
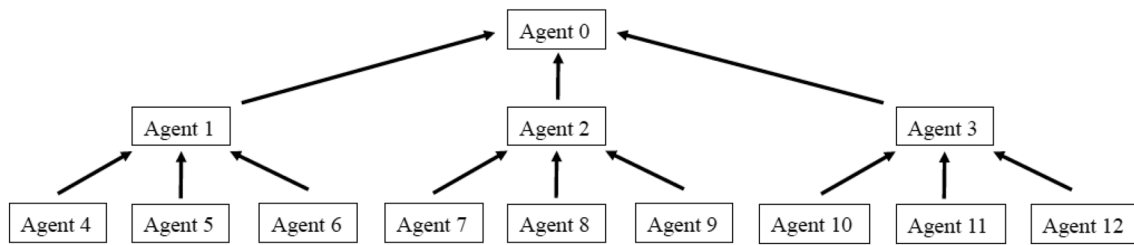
Fig. 1 Memetic algorithm represented as a hierarchical tree (redrawn from Moscato et al. 2021)

algorithm are fully detailed in Moscato et al. (2021), and only the general principle is recalled here for the benefit of the reader.

The memetic algorithm is based on the concept of a population structure of optimising agents, which in this case contains three levels and 13 agents, numbered 0–12 (see Fig. 1). A feature of this method is that mathematical models (solutions of a mixed combinatorial and nonlinear optimisation problem) that have initially been randomly created can be progressively selected and improved. The optimisation, evolution and selection of the models occur via rigorous mathematical algorithms as described in Moscato et al. (2021).

In this algorithm, two solutions are created for each of agents 4–12. The variables of each solution are randomly selected from the list of possible input variables, and the coefficient assigned to each variable is randomly selected from the range $(-3; 3)$. For each solution, the coefficients identified from the memetic algorithm are optimised using a variant of the Nelder−Mead nonlinear optimisation technique (Fajfar et al. 2017) as a local search operation. As a result, each agent has two solutions, the performance of which is assessed in terms of the mean squared error (MSE) using the training dataset. The solution with the best performance is referred to as a "pocket" solution, and the other solution is referred to as the "current" solution. Each agent has one pocket solution and one current solution. In the evolutionary step of the algorithm, some current solutions are then modified using variation operations named recombination and mutation. The objective is to create a new offspring as a current solution. If that offspring performs better than the best solution, it becomes the new pocket solution of the agent.

A mutation operation consists of making random changes to the current solution. It can either remove or add a variable from the model. If a variable is added, its coefficient is randomly selected in the range $(-3; 3)$, and a local search is applied to attempt to improve the solution in terms of its performance. As indicated by its name, a recombination operation consists of combining two existing solutions, i.e., two different mathematical models, referred to as 'parent' solutions, from two different agents. To create a new offspring solution to replace the current solution, variables are randomly selected from the union, intersection, or symmetric

difference of the variables of the parents. If a variable is only present in one of the parents, its coefficient is equal to the coefficient in the parent solution. If a variable is present in both parent solutions, its coefficient is a function of both parent coefficients. The Nelder–Mead algorithm is then applied to all offspring coefficients for local optimisation. In this study, mutation is applied to 10% of the solutions, and all current solutions are used as a parent solution for recombination.

The best solution or model of agents 4–6 is moved to agent 1, that of agents 7–9 is moved to agent 2, and finally, the best solution for agents 10–12 is moved to agent 3. The best model of agents 1, 2, and 3 is moved to agent 0.

The performance of the final model of agent 0 is then reevaluated to account for its complexity. Indeed, the objective is not only to obtain a model with good performance but also to formulate a simple model (for computational efficiency), i.e., a model with the lowest depth and least number of variables (the variables that have no strong influence are discarded). To achieve this, a delta penalty method is used (Sun and Moscato 2019), where the performance of a model is evaluated through a corrected mean square error:

$$\text{MSE}_{\text{corrected}} = \text{MSE} \times (1 + \delta \times N) \tag{7}$$

where δ is arbitrarily selected as 0.1 and N is the number of variables used in the model. The more complex the formulation is, the more penalised the model, and the higher the corrected MSE. The process of identifying the best-performing model via the memetic algorithm is repeated 200 times until 200 best-performing models are produced. All 200 best-performing models are then evaluated in terms of MSE_corrected, and the model with the lowest MSE_corrected is retained as the trained model.

Both memetic algorithms and the CFR approach are becoming very popular for data science applications (Cotta et al. 2018; Moscato and Mathieson 2019; Moscato et al. 2022). The memetic algorithm was implemented in C++ and executed using a computer equipped with a 6 Core Intel(R) Core(TM) i7-9750H CPU with a 2.60 GHz clock speed and 16.0 GB of RAM running on the Ubuntu 20.04.1 LTS operating system.

2.3 Correction for Surface Size and Resolution

As will be detailed in Sect. 3, the synthetic dataset used to train and validate the regression model was obtained by applying the NDSS model to 2 m per 2 m rough surfaces, referred to as reference surfaces (as per Jeffery et al. 2021). In the remainder of this paper, the spatial resolution, surface area and number of contributing facets of the reference surfaces are identified with a subscript "o", i.e., Res_o (equal to 1 mm), A_o (equal to 4 m^2) and N_{CF_o}, respectively. The notations Res, A and N_{CF} refer to surfaces with either another size and/or another spatial resolution.

The number of contributing facets predicted by the CFR model is only meaningful for the reference surfaces that contain 8×10^6 facets. For example, let us consider a fixed set of input variables and a fixed set of surface references. Assume that the model predicts that 16×10^4 facets are sheared. Now consider the same inputs (and hence the same model output) but for a surface that contains 8×10^4 facets (for example, a 1 m per 1 m surface with a 5 mm spatial resolution), the direct application of the CFR model would return that there are more sheared facets than the surface contains, which is a nonsensical output. The proper way to analyse this result is to consider that 16×10^4 facets represent 2% of the total number of facets of the reference surface and the smaller surface with a different spatial resolution would also have 2% of its facets sheared. Hence, the proper number of contributing facets for the surface to consider is $N_{CF} = 0.02 \times 8 \times 10^4 = 1600$ facets. From this number and with the projected area of each facet, it is possible to compute the shear strength of the surface. To express this transformation formally, two corrections are presented in Eqs. (8) and (9).

First, the number of contributing facets for the reference surface N_{CF_o} needs to be multiplied by the surface ratio (A/A_o) and resolution ratio ($\text{Res}_o/\text{Res}_a$)^2. Given that $A_o = 4$ m^2 and $\text{Res}_o = 1$ mm, we obtain:

$$N_{CF} = N_{CF_o}\left(\frac{A}{A_o}\right)\left(\frac{\text{Res}_o}{\text{Res}}\right)^2 = N_{CF_o}\left(\frac{A}{4\text{Res}^2}\right) \tag{8}$$

where N_{CF_o} is the number of contributing facets predicted by the CFR model and N_{CF}, A and Res are the number of contributing facets, the total area and the spatial resolution of the surface being considered, respectively.

In addition, assuming that the spatial resolution is the same in the X and Y directions, the projected facet area A_{ip} is equal to (see Fig. 9 in "Appendix" section):

$$A_{ip} = \frac{\text{Res}^2}{2} \tag{9}$$

Equations (8) and (9) can now be combined with Eqs. (1)–(4) to estimate the shear strength of a discontinuity surface with different dimensions and resolutions than those of the reference surfaces.

Note that an attempt was made to directly predict the percentage of facets that have been sheared using continuous fraction regression, but the prediction accuracy was not as good as that presented in the remainder of this paper.

3 Training and Validation Datasets

A preliminary study revealed that better predictions can be obtained when using $\ln(N_{CF_o})$ and $\tan(\phi_b)$ instead of N_{CF_o} and ϕ_b. Therefore, the tangent of ϕ_b was considered in the set of independent variables $\mathbf{x}$ to predict $\ln(N_{CF_o})$.

The process of obtaining a dataset of $\{\ln(N_{CF_o}), sd_i, \sigma_{ci}, \tan(\phi_b), m_i, \sigma'_n\}$ to train and test the CFR model is described in detail in the next sections and is summarised as follows:

- First, 474 synthetic rock surfaces were created using the 2D LAS random field model used by Casagrande et al. (2018). The peak and residual shear strengths of these surfaces were estimated using the NDSS model with 1000 different combinations $\{\sigma_{ci}, \tan(\phi_b), m_i, \sigma'_n\}$ as inputs. This 474,000-point dataset is referred to as the raw dataset.

- The original dataset was then reduced to 14,000 data points with a specific filtering process to obtain a dataset with specific values of sd_i. This dataset is referred to as the reduced dataset, which is split, and its smaller part, comprising 10% of the reduced dataset, is used for the training task, and the remaining 90% of the data is used for the testing task.

- The 14,000-point dataset was then enriched using meta features, a mathematical approach used to increase the ability of a model to introduce some powers and roots of the input variables and capture nonlinearities. This dataset is referred to as the enriched dataset.

- The CFR model was trained with 10% of the samples from the enriched dataset, and the remaining 90% of the samples were used to test the generalisation performance.

3.1 Raw Dataset

Before elaborating on how the data were created by the combination of the random field model and the NDSS model, it is important to specify that research by Jeffery et al. (2021) showed that a natural rough rock surface can be decomposed into three levels of roughness, namely, large scale (*L*), intermediate scale (*I*) and small scale (*S*) roughness (see the example in Fig. 2a), which is consistent with the rock mechanics literature. The NDSS prediction is strongly
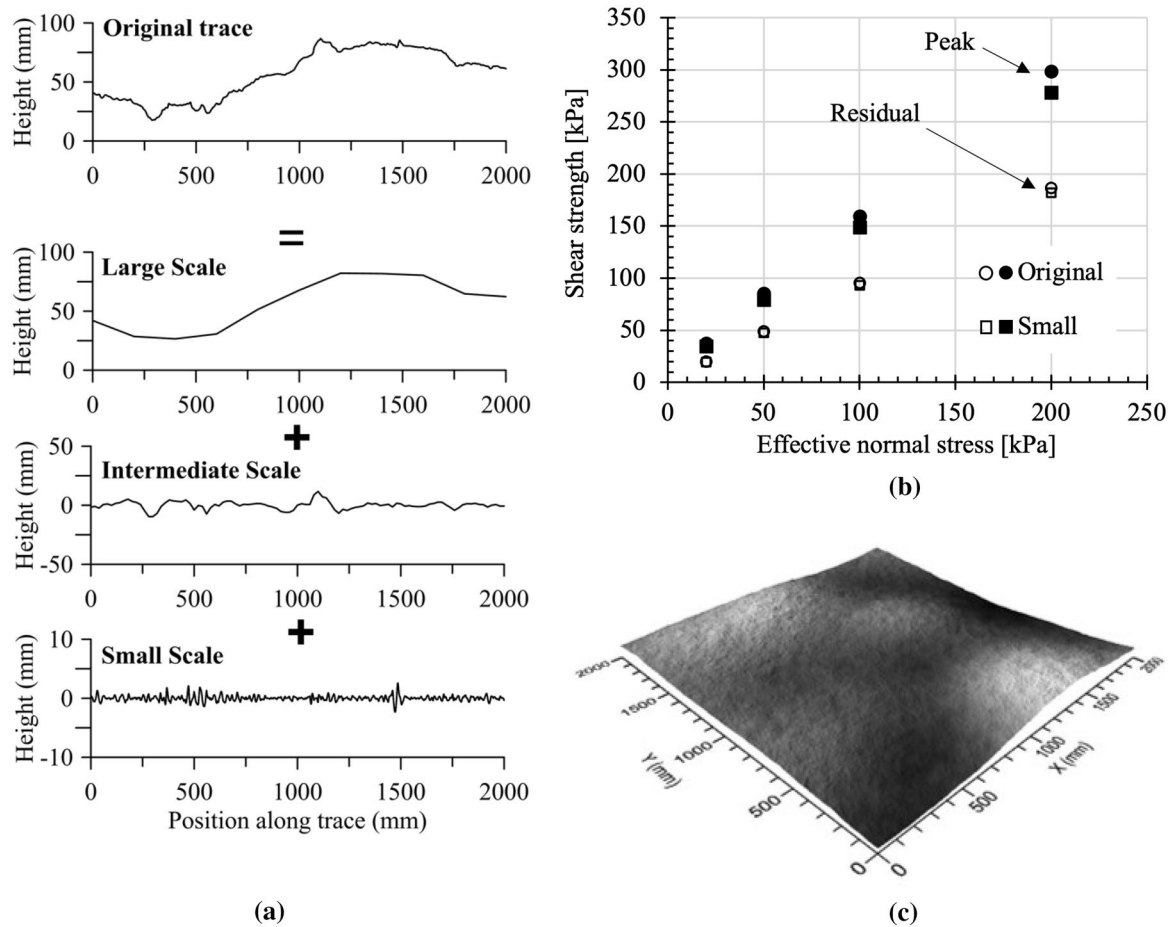
(a)

(b)

(c)

Fig. 2 **a** Example of the decomposition of a trace profile into three subprofiles on a large scale, an intermediate scale and a small scale (from Jeffery et al. 2021). **b** Examples of the peak (full symbols) and residual (hollow symbols) shear strength envelope obtained from the

NDSS model for the rough surface shown in **c** and its small scale component. The circle and square symbols correspond to the original surface and small-scale subsurface, respectively. **c** View of the original surface

governed by the inclination of asperities of the small-scale roughness (*S*). The inclination of asperities is referred to as gradients and is calculated as the difference in elevation between two adjacent points over the distance between the two points parallel to the mean shear plane.

Figure 2, adapted from Jeffery et al. (2021), illustrates the key contribution of asperity gradients on a small scale. Figure 2a first shows how a trace profile can be decoupled into three subprofiles at large, intermediate and small scales. The superposition of these three subprofiles gives the original profile. Similarly, discontinuity surfaces can be decomposed into three levels of roughness (small, intermediate and large).

Figure 2b shows the peak and residual failure criteria for a rough surface (referred to as "original", shown in Fig. 2c) and the criteria of the corresponding small-scale subsurface (referred to as "small"). The strength envelope of the small-scale surface is only marginally lower than that of the original surface, which suggests that most of the shear strength, as predicted by the NDSS model,

comes from the small-scale roughness. Consequently, in the remainder of this paper, given the large amount of data needed to train and test the regression model, only surfaces with small-scale roughness were created.

A total of 474 3D synthetic surfaces (size of 2 m per 2 m, 1 mm spatial resolution) were generated by the

Table 1 Values of input variables used to establish the raw dataset of N_{CFo}

	Variable	Unit	Values considered
Surface statistics	σ_z^2	mm^2	0.15–0.5 in 0.025 increments
	CL	mm	2.5–50 mm (32 values of CL)
	sd_i	mm/mm	~0.02 to ~0.62
Loading	σ_n'	MPa	0.005, 0.02, 0.05, 0.08, 0.1, 0.3, 0.5, 0.8, 1, 2
Rock strength	σ_{ci}	MPa	20, 40, 60, 80, 100
	ϕ_b	$^{\mathrm{o}}$	20, 25, 30, 35, 40
	m_i	No unit	5, 10, 20, 30

2-dimensional local average subdivision (2D LAS) random field model (Fenton and Vanmarcke 1990). The inputs of the 2D LAS (variance of height σ_z^2 and correlation length CL) were varied to obtain a wide range of roughness values (see Table 1). The mean value of asperity height is 0. The standard deviation of the gradient ranges from approximately 0.02–0.62.

The peak and residual shear strengths of the 474 synthetic surfaces were estimated using the NDSS model for 1000 combinations of σ'_n, σ_{ci}, ϕ_b and m_i resulting from the values given in Table 1. The raw dataset hence contains 474,000 data points.

The range of values given in Table 1 reflects typical ranges reported in the literature for sedimentary, igneous, and metamorphic rocks (see ISRM R4 Strong Rock field estimate strength class, Hoek and Brown 1997; Barton 1973, 1976; Alejano et al. 2012).

Although it is well recognised in the literature that shear strength is related to normal effective stress and rock strength parameters (see Barton 1976), roughness is usually described using the joint roughness coefficient, not the standard deviation of gradients. Figure 3a clearly demonstrates that a strong relation exists between N_{CFo} and sd_i for a given combination of material parameters. Such a degree of correlation was observed for all combinations of input variables. From this figure, it is suggested that the shear response of rough nonplanar discontinuities is strongly controlled by the distribution of asperity gradients, and consequently, the
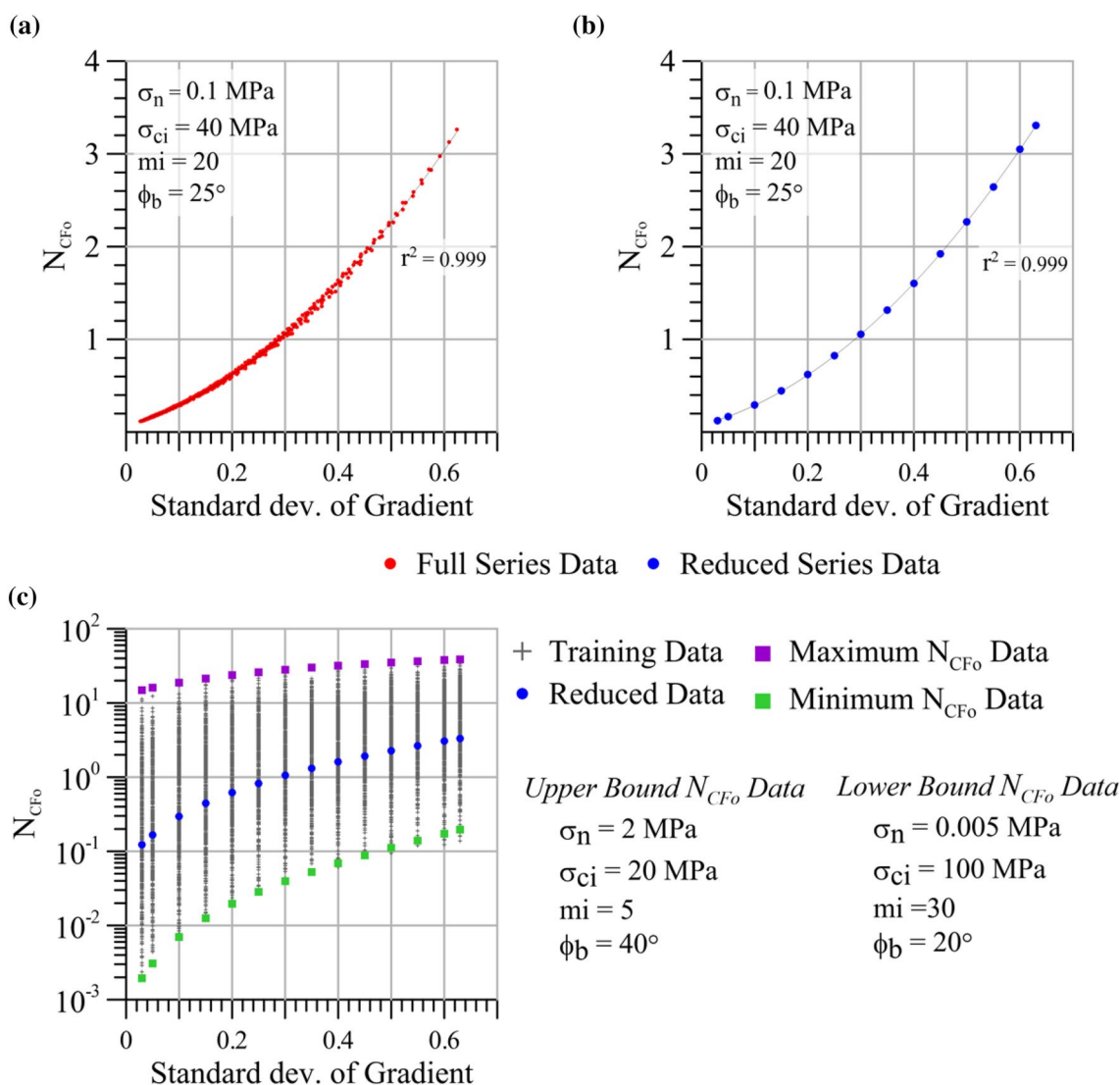

Fig. 3 Evolution of the number of contributing facets (N_{CFo}) with the standard deviation of gradients, sd_i. **a** N_{CFo} values associated with 474 synthetic surfaces, with a 3rd order polynomial trend line fitting and for one combination of variables. **b** A total of 14 N_{CFo} values were obtained by the postprocessing procedure, sampled along the fitted trend line at equal increments of sd_i and forming the reduced dataset. **c** All N_{CFo} values of the reduced dataset. The upper and lower bounds are highlighted

Table 2 Example of one original combination of inputs for the following output: $\ln(NCF_o)=5.690359$. sd_i, $\tan(\phi_b)$ and m_i have no unit, σ'_n in MPa, σ_{ci} in MPa

Standard deviation of gradients sd_i	0.03
Effective normal stress σ'_n	0.005
Tangent of basic friction angle $\tan(\phi_b)$	0.466308
Unconfined compressive strength σ_{ci}	100
Material parameter m_i	20

standard deviation of gradients can be considered an adequate roughness descriptor.

3.2 Reduced Dataset

To train and validate the CFR model, it is preferable to have a dataset of manageable size with controlled values of standard deviation of gradients, given its significance, which is not the case for the raw dataset. For each combination of (σ'_n, σ_{ci}, ϕ_b and m_i), the evolution of $\ln(N_{CFo})$ with sd_i was plotted and fitted with a 3rd-order polynomial (see Fig. 3a). The polynomial was then used to extract 14 ($\ln(N_{CFo})$, sd_i) points, including the minimum and maximum values of sd_i and 12 equally spaced (increments of 0.5) values in between (see Fig. 3b). This process led to a dataset of 14,000 points (14 standard deviations and 1000 material parameters).

Figure 3c presents the values of N_{CFo} as a function of sd_i of the reduced dataset and shows a large range of variation, from several hundred facets to several million facets. To highlight the influence that other variables have on N_{CFo}, the upper and lower bounds of N_{CFo} are plotted with sd_i. These bounds were obtained using extreme values (within the range of Table 1) of σ'_n, σ_{ci}, ϕ_b and m_i.

3.3 Enriched Dataset

The original dataset contains an output ($\ln(N_{CFo})$) and five input variables (σ'_n, σ_{ci}, sd_i, $\tan(\phi_b)$ and m_i). For simplicity of formulation and training efficiency, the model does not test the powers or roots of the input variables (e.g., m_i^2, $m_i^{0.5}$), as per Eq. (6). To overcome this limitation, the powers and roots of the variables are introduced as extra input variables associated with the same output. These extra variables are generally referred to as "feature engineering" in the area of machine learning and data mining. In this study, for each independent variable x, the following array $\boldsymbol{x'}$ of data associated with the same output ($\ln(N_{CFo})$) was used:

$$\boldsymbol{x'} = \left\{ x, x^{-1}, x^{\frac{1}{2}}, x^{-\frac{1}{2}}, x^{\frac{1}{3}}, x^{-\frac{1}{3}}, x^2, x^{-2}, x^3, x^{-3} \right\}$$

Tables 2 and 3 show examples of the original dataset and enriched dataset.

Table 3 Example of one enriched combination of inputs based on the inputs of Table 2 for the following output: $\ln(NCF_o)=5.690359$. sd_i, $\tan(\phi_b)$ and m_i have no unit, σ'_n in MPa, σ_{ci} in MPa

Power	1/3	1/2	1	2	3	-1/3	-1/2	-1	-2	-3
Standard deviation of gradients	$sd_i^{1/3}=0.310723$	$sd_i^{1/2}=0.173205$	$sd_i=0.0300$	$sd_i^2=0.0009$	$sd_i^3=0.000027$	$sd_i^{-1/3}=3.21897$	$sd_i^{-1/2}=5.773502$	$sd_i^{-1}=33.3333$	$sd_i^{-2}=1111.111$	$sd_i^{-3}=37,037.037$
Effective normal stress	$\sigma'^{1/3}_n=0.170997$	$\sigma'^{1/2}_n=0.070710$	$\sigma'_n=0.00500$	$\sigma'^2_n=0.000025$	$\sigma'^3_n=1.25\text{E}{-}07$	$\sigma'^{-1/3}_n=5.848035$	$\sigma'^{-1/2}_n=14.14213$	$\sigma'^{-1}_n=200$	$\sigma'^{-2}_n=40,000$	$\sigma'^{-3}_n=8,000,000$
Tangent of basic friction angle	$\tan^{1/3}(\phi_b)=0.775456$	$\tan^{1/2}(\phi_b)=0.682867$	$\tan(\phi_b)=0.466308$	$\tan^2(\phi_b)=0.217443$	$\tan^3(\phi_b)=0.101395$	$\tan^{-1/3}(\phi_b)=1.28956$	$\tan^{-1/2}(\phi_b)=1.46441$	$\tan^{-1}(\phi_b)=2.144505$	$\tan^{-2}(\phi_b)=4.598903$	$\tan^{-3}(\phi_b)=9.862372$
Unconfined compressive strength	$\sigma_{ci}^{1/3}=4.641588$	$\sigma_{ci}^{1/2}=10$	$\sigma_{ci}=100$	$\sigma_{ci}^2=10,000$	$\sigma_{ci}^3=1,000,000$	$\sigma_{ci}^{-1/3}=0.215443$	$\sigma_{ci}^{-1/2}=0.1$	$\sigma_{ci}^{-1}=0.01$	$\sigma_{ci}^{-2}=0.0001$	$\sigma_{ci}^{-3}=0.000001$
Material parameter m_i	$m_i^{1/3}=2.714417$	$m_i^{1/2}=4.472135$	$m_i=20$	$m_i^2=400$	$m_i^3=8000$	$m_i^{-1/3}=0.368403$	$m_i^{-1/2}=0.22360$	$m_i^{-1}=0.05$	$m_i^{-2}=0.0025$	$m_i^{-3}=0.000125$

Using the powers and roots of variables in feature engineering can provide several advantages, including nonlinear transformations, scaling, feature selection, and feature interpretation.

For example, the powers and roots of variables can introduce nonlinear transformations that can help capture nonlinear relationships between the independent input variables and the target variable; the square of an input variable may be more relevant to the target variable than the variable itself, especially if there is a quadratic relationship between them. The use of meta features has also been employed in classification problems (e.g., Rocha de Paula et al. 2011; Johnstone et al. 2012). As such, it is a more powerful approach than simply considering more data from the original dataset.

3.4 Data Sampling for Training and Testing

To create a reduced-size dataset to guide the training phase of the CFR method, 10% of the original dataset (a total of 1400 samples from the enriched reduced dataset) was selected using the following process: the variance of the enriched dataset was first computed. Then, 10% of the data were randomly sampled, and the variance was computed. The sampling process was iterated until the variance difference between the original dataset and the subset was less than a defined threshold (here 0.01). The process iterates until the new dataset meets the variance threshold condition. Therefore, it contains 10% of the samples, and it is used as the training dataset, while the remaining 90% of the samples are used to test the generalisability of the models.

4 Results

4.1 Training and Testing of the CFR Model

A subset of 1400 samples was used to train the continued fraction regression model with depth $= \{0, 1, 2, 3, 4\}$, and this process was repeated for each depth 25 times. The continued fraction regression method executed with the training set identified the best model when depth $= 1$. In the following, the CFR model is hence referred to as CFR-d1, and it is given as follows:

$$\ln \left(N_{\text{CFo}} \right) = g_{\text{o}}(\boldsymbol{x}') + \frac{h_{\text{o}}(\boldsymbol{x}')}{g_1(\boldsymbol{x}')} \tag{10}$$

with

$$g_{\text{o}}(\boldsymbol{x}') = 0.497568 \cdot sd_i^{\frac{1}{2}} + 0.051503 \cdot \sigma_n'^{\frac{1}{3}} - 0.604231$$
$$\cdot \sigma_{\text{ci}}^{\frac{1}{3}} - 0.125605$$

$$h_{\text{o}}(\boldsymbol{x}') = 0.001748 \cdot \tan^2 \left(\phi_{\text{b}} \right) + 0.017096 \cdot \sigma_n'^{\frac{1}{3}}$$
$$+ 0.000020 \cdot \sigma_{\text{ci}}^{\frac{1}{3}} + 0.002485$$

$$g_1(\boldsymbol{x}') = -0.000307 \cdot sd_i^{\frac{1}{2}} + 0.000973 \cdot \sigma_n'^{\frac{1}{3}} + 0.000544$$

with σ_n' in MPa, σ_{ci} in MPa, and sd_i in m/m.

Note that m_i does not appear in the CFR-d1 model, which implies that this variable only has a marginal influence on $\ln(N_{\text{CFo}})$. However, m_i is a relevant parameter that contributes to the shear strength of the discontinuity, as captured in Eqs. (3) and (4), once the number of contributing facets is obtained.

The mean squared error (MSE) score of the CFR-d1 model obtained for the training set was 0.1136, and for the test set, it was 0.0424. The performance of the model is illustrated in Fig. 4 in terms of the distribution of relative error between the prediction from the CFR-d1 model and the dataset. The relative error, in percent, is computed as:

$$\text{error}\,[\%] = \frac{100(\text{CFR} - \text{Data})}{\text{Data}} \tag{11}$$

where CFR is the value of prediction by the CFR model and Data is the value of $\ln(N_{\text{CFo}})$ from the training dataset for the same set of input variables. A positive error value reflects an overestimation, while a negative error corresponds to an underestimation. Figure 4 shows that $\sim 95\%$ of the CFR predictions fall within a $\pm 20\%$ error band, which is an excellent result.
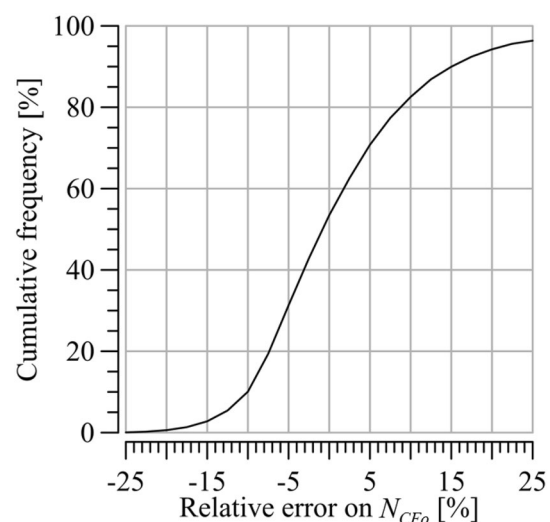

Fig. 4 Distribution of relative error between the training dataset and CFR-d1 prediction, computed from Eq. (11)

4.2 Comparison Between CFR-d1 Predictions and Experimental Values of the Sheared Area

In this section, the predictions of the trained CFR-d1 model are compared to the experimental values of the sheared area obtained by Jeffery et al. (2022). The authors conducted a series of low normal stress, large scale direct shear tests on a series of 2 m per 2 m perfectly matching mortar discontinuity replicates with a 1 mm spatial resolution (total number of facets of 8×10^6). After each test, estimates of the contact area that contributed to the development of peak shear strength were obtained via digital analysis of the damaged area.

The direct shear tests were conducted at normal stresses σ'_n of 5, 14, 22 and 31 kPa. In the direction of shearing, the whole discontinuity surface was observed to have a standard deviation of gradients sd_i equal to 0.143. The material properties of the mortar used to cast the replicas were an intact compressive strength (σ_{ci}), a material constant (m_i) and a basic friction angle (ϕ_b) of 68.7 MPa, 8.5 and 36.2°, respectively. The CFR-d1 model [Eq. (10)] was applied using values of normal stress and material parameters

Table 4 Values of N_{CFo} from the CFR-d1 prediction for the 2 m per 2 m surface tested by Jeffery et al. (2022). Input parameters: $sd_i = 0.143$, $\sigma_{ci} = 68.7$ MPa, $m_i = 8.5$, and $\tan(\phi_b) = 0.732$

	$\sigma'_n = 5$ kPa	$\sigma'_n = 5$ kPa	$\sigma'_n = 14$ kPa	$\sigma'_n = 22$ kPa	$\sigma'_n = 31$ kPa
N_{CFo}	4542	4542	9132	12,727	16,502

consistent with the experimental values. The sheared area was computed by multiplying the value of N_{CFo} given by the CFR-d1 by the area of a facet (0.5 mm^2). The values of N_{CFo} predicted by the CFR-d1 model are given in Table 4.

Figure 5a presents a comparison of the experimental sheared area and that estimated from the CFR-d1 model. The results in this figure confirm the observations presented in Sect. 4.1, in that the predictive ability of the CFR-d1 model is excellent. Figure 5b reveals that CFR-d1 predictions lie within ±10% of the experimental data, with only one data point outside that range, with an error of − 17% (identified by the arrow in Fig. 5b). Such a good agreement is a very strong indication that the CFR-d1 model is capable of adequately predicting shear strength, a hypothesis that will be tested in the next section.

4.3 Comparison Between CFR-d1 Predictions and Experimental Values of Shear Strength

Following the comparison of the CFR-d1 predictions with the experimental values of the sheared area obtained by Jeffery et al. (2022), testing the predictive ability of the CFR-d1 model further is proposed by comparing the experimental values of peak and residual shear strength obtained by Jeffery et al. (2022) for the same surface and those derived from the predicted values of $\ln(N_{CFo})$ and Eqs. (1)–(5). The predicted shear strength values are given in Table 5.

The experimental and predicted strength values, for both peak and residual shear strength, are plotted in Fig. 6, and
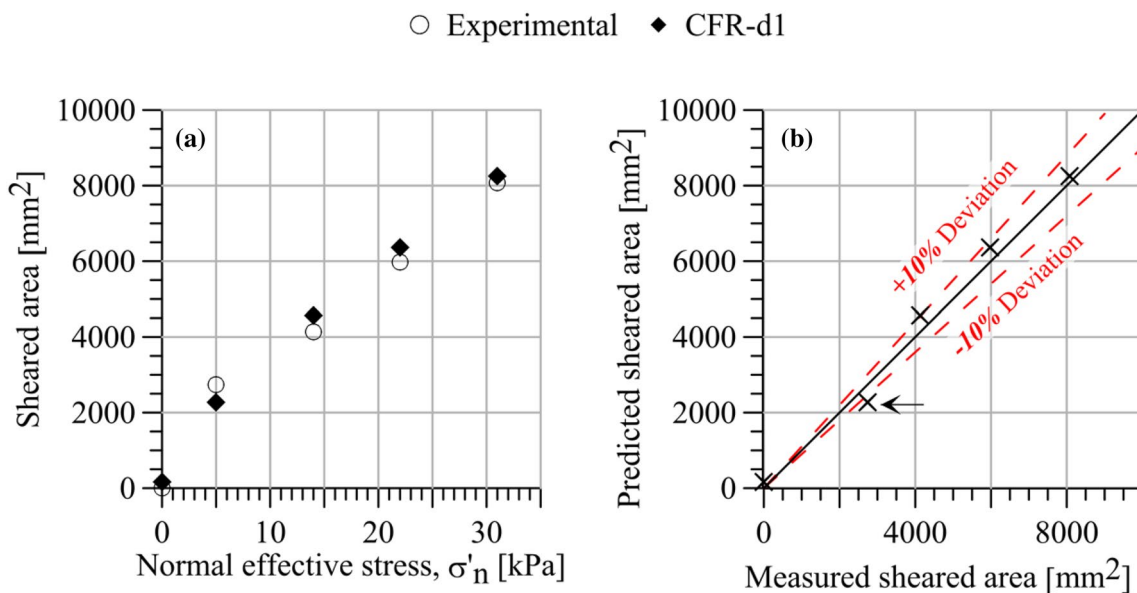

Fig. 5 **a** Evolution of the experimental and CFR-d1 predicted values of a sheared area with normal effective stress. **b** Comparison of the experimental and predicted values of the sheared area with an indica-tion of ±10% relative error (red dashed line). The black arrow indi-cates the point with a − 17% relative error

Table 5 Peak and residual shear strength derived from the values of N_{CFo} predicted by the CFR-d1 model (see Table 4) for the 2 m per 2 m surface tested by Jeffery et al. (2022). Input parameters: $sd_i = 0.143$, $\sigma_{ci} = 68.7$ MPa, $m_i = 8.5$, and $\tan(\phi_b) = 0.732$

	$\sigma_n' = 5$ kPa	$\sigma_n' = 5$ kPa	$\sigma_n' = 14$ kPa	$\sigma_n' = 22$ kPa	$\sigma_n' = 31$ kPa
τ_p (kPa)	12.4	12.4	28.6	42.9	56.8
τ_r (kPa)	4.7	4.7	12.5	19.5	26.5

once again, excellent agreement can be seen between the predicted and experimental data.

The predictive ability of the CFR-d1 model was further tested using direct shear test results obtained on the same three surfaces tested by Casagrande et al. (2018). The mortar replicas of the different surfaces were tested under normal stress ranging from 20 to 500 kPa. The surface parameters relevant to making a CFR-d1 prediction are given in Table 6. Note that the same mortar mixture used by Casagrande et al. (2018) was used to make new replicas that were cured in water for four weeks and new data were obtained. A series of triaxial tests were conducted to characterise the material strength under a level of confinement consistent with high-stress values expected at the few contact points within a rough discontinuity (on the order of 30–50 MPa). From these tests, an m_i value of approximately 13 was obtained (see Table 6), as opposed to approximately 35 (Casagrande et al. 2018), which seems too high for a mortar. Except for this difference in material characterisation, the shear tests were conducted as per Casagrande et al. (2018), and in the interest of conciseness, the experimental methods are not repeated here.

Figure 7 shows that most of the CFR-d1 predictions are relatively close to the experimental values of peak and shear strength.

Finally, all the results are summarised in Fig. 8, which also shows a distribution of relative errors. In 80% of the cases, the discontinuity shear strength can be predicted within ±10% of the experimental value, which can be considered an excellent result.

5 Conclusions

This paper focused on establishing a continued fraction regression (CFR) model to use as an alternative to the Newcastle discontinuity shear strength (NDSS) model and to address issues of numerical implementation and computational cost. The model was trained and tested using an extensive dataset of sheared facet data.

Table 6 Parameters of the S, M, and R surfaces (see Casagrande et al. 2018) tested in the direct shear test under constant normal stress

	Surface		
	S	M	R
Dimensions (mm)	91*91	91*90	88*91
Spatial resolution (mm)	0.5	0.5	0.5
Number of facets	66,248	66,248	66,248
sd_i for the whole surface in the direction of shearing (rad)	0.107	0.138	0.232
σ_{ci} (MPa)	49.7	49.7	60.1
m_i	13.6	13.6	13.9
$\tan(\phi_b)$	0.66	0.66	0.75

Fig. 6 Evolution of experimental and CFR-d1 predicted values of shear strength (peak and residual) with normal effective stress for the 2 m per 2 m surface tested by Jeffery et al. (2022)
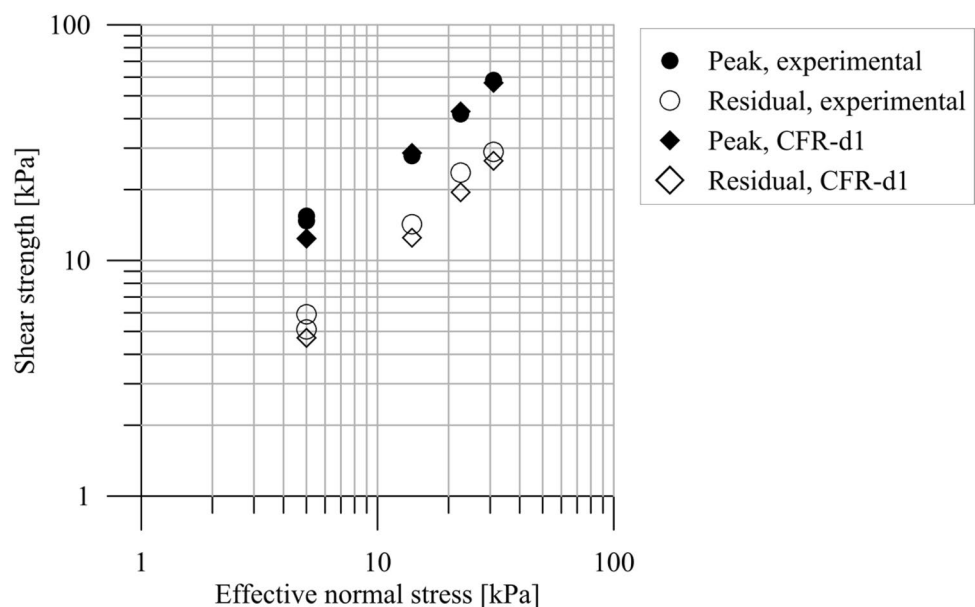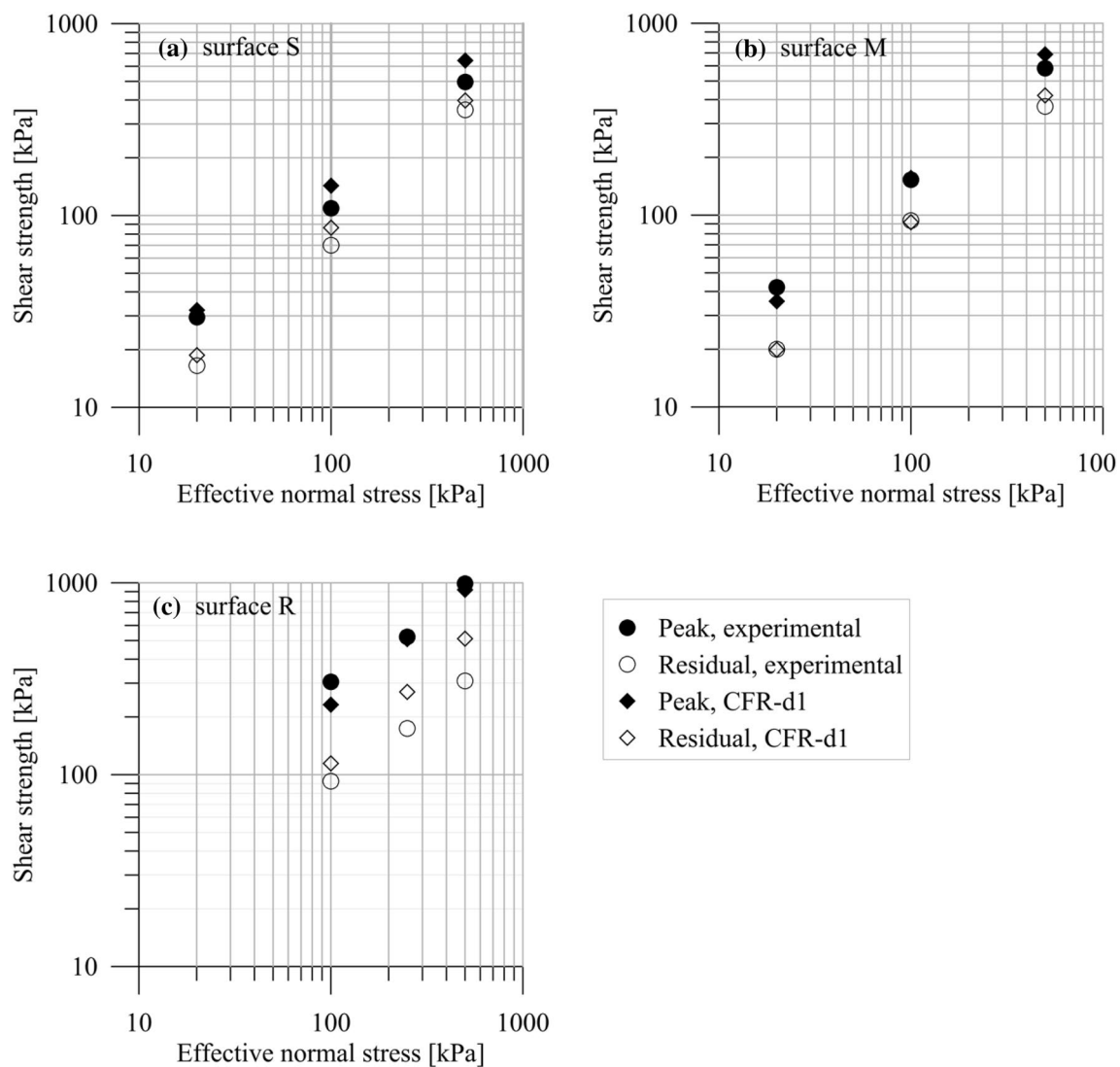
Fig. 7 Evolution of the experimental and CFR-d1 predicted values of shear strength (peak and residual) with normal effective stress for the *S*, *M*, *R* surfaces tested by Casagrande et al. (2018)

In establishing the dataset, clear correlations were observed between the peak shear strength and standard deviation of gradients of the surfaces, all other variables being constant. This observation suggests that the standard deviation of gradients is a reliable indicator of surface roughness.

The training and testing of the CFR model returned low values of mean squared error, indicating that the model can satisfactorily predict the number of contributing facets for given input combinations. Furthermore, the CFR model predictive capability was tested against experimental data of shear area and shear strength, and 75% of the predictions fell within 10% of experimental values, which confirms the excellent performance of the CFR model.

The findings of this study suggest that the trained CFR model can efficiently be used as an alternative to the NDSS model, offering a much simpler implementation and much lower computational times.

Appendix: Details About the NDSS Model (Casagrande et al. 2018)

In this appendix, some information about the Newcastle discontinuity shear strength (NDSS) model, originally published in Casagrande et al. (2018) and reproduced here for clarity, is presented. The original model was derived in terms of total stress because the authors did not consider the contribution of fluid pressure within the discontinuity. In the rest of this appendix, the model is presented in terms of
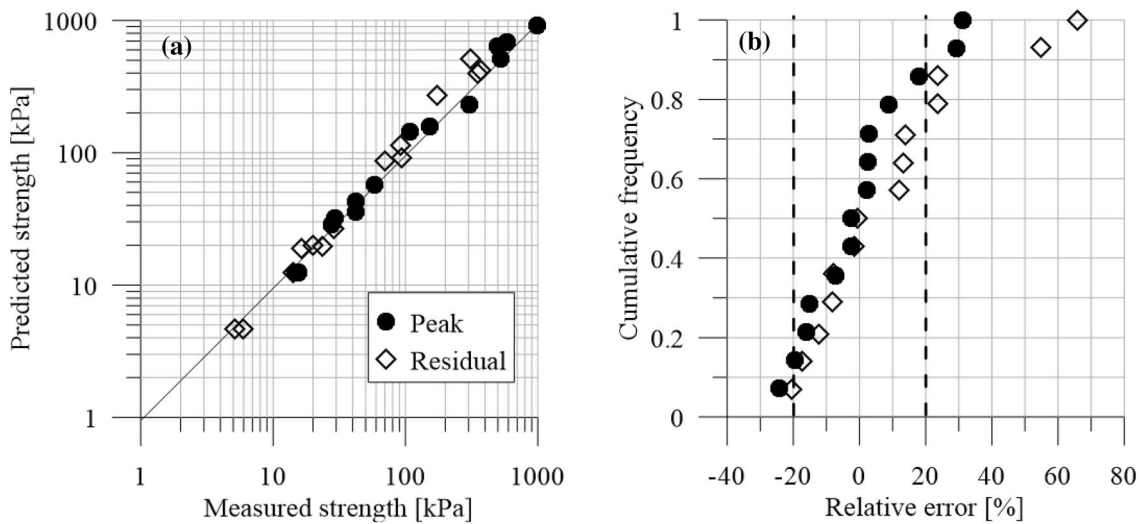
Fig. 8 **a** Comparison of the experimental (measured) and CFR-d1 predicted shear strength (peak and residual) data from Figs. 6 and 7. The continuous line has a one-to-one gradient. **b** Distribution of relative errors for all data from Figs. 6 and 7

effective stress and effective parameters, as it is well recognised that the hydromechanical response of rock discontinuities is governed by effective stress (Brady and Brown 2004).

Figure 9a shows an example of a 110 mm per 110 mm triangulated surface (a) with a data point every 0.5 mm in both the X and Y directions. Upon shearing, the discontinuity dilates along the steepest asperities, referred to as active facets or contributing facets (see the sketch in Fig. 9b). The NDSS model assumes that only active facets are in contact because of dilation upon shearing. As a result, the vertical effective force exerted on the whole discontinuity (noted F'_{macro}) is uniformly distributed among all active facets that are subjected to a local effective vertical force f'_{local_i} equal to:

$$f'_{local\_i} = \frac{F'_{macro}}{N_{CF}} \tag{12}$$

where N_{CF} is the total number of contributing facets (see Fig. 9b). The local effective stress, normal to the shear plane of the discontinuity, is then estimated for each contributing facet as:

$$\sigma'_{local\_i} = \frac{f'_{local\_i}}{A_{ip}} \tag{13}$$

where A_{ip} is the area of the facet projected on the XY plane (see Fig. 9c), which depends on the spatial resolution. Given the spatial resolutions along X and Y, Res_X and Res_Y, the projected area is equal to $A_{ip} = \frac{Res_X \times Res_Y}{2}$. In this study,

$Res_X = Res_Y$ and A_{ip} is simplified as $A_{ip} = \frac{Res^2}{2}$, where Res is the spatial resolution.

In the dataset used for this study, the digitised surface has a spatial resolution of 1 mm in both the X and Y directions, so $A_{ip} = 0.5$ mm^2.

The NDSS model considers two possible asperity interaction mechanisms: a facet of the top wall can slide on its bottom counterpart along surface A_i, or the facets (top and bottom) can be sheared along a horizontal plane, i.e., along surface A_{ip} (see Fig. 9c). Assuming that the material shear strength arises from a cohesive term (c) and a frictional component, the force required to shear each contributing facet, denoted f_{shear_i}, is simply the shear strength for a given normal stress multiplied by the sheared area of each facet A_{ip}:

$$f_{shear\_i} = A_{ip} \cdot \left(c + \sigma'_{local\_i} \cdot \tan(\phi') \right) \tag{14}$$

where ϕ' is the effective Coulomb friction angle of the material, c is the cohesion, A_{ip} is the area of the facet projected on the XY plane (see Fig. 9) and σ'_{local_i} is the local vertical effective stress acting on facet i [see Eq. (13)]. The peak shear force f_{peak} is the sum of the contribution of all active facets:

$$f_{peak} = \sum_{i=1}^{N_{cf}} f_{shear\_i} = \sum_{i=1}^{N_{cf}} A_{ip} \cdot \left(c + \sigma'_{local\_i} \cdot \tan(\phi') \right) \tag{15}$$

The shear stress is computed as the peak force divided by the total projected area of the surface:
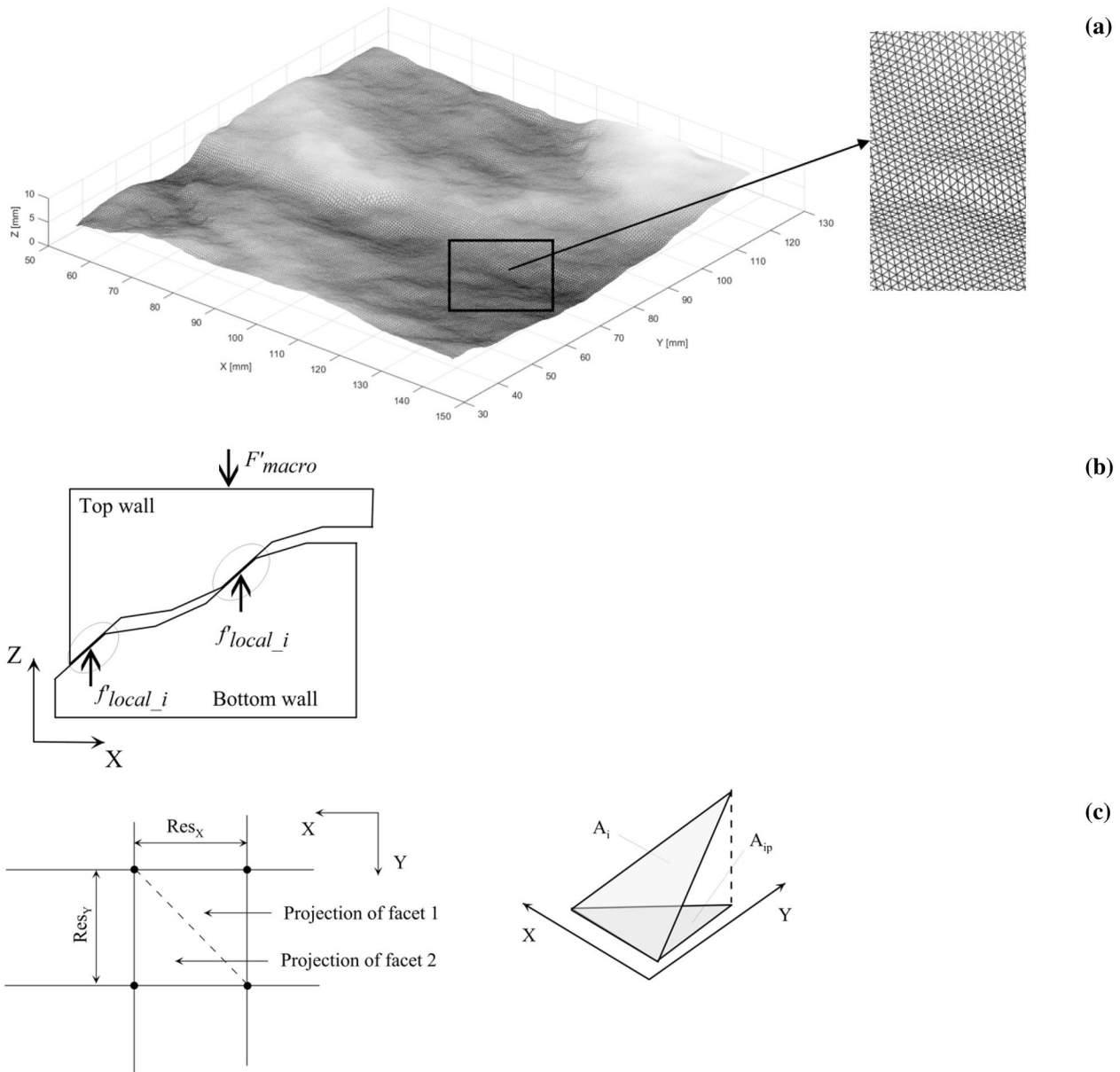
(a)

(b)

(c)

Fig. 9 a Example of a triangulated rock surface. **b** A 2D schematic representation of discontinuity dilation upon shearing with a redistribution of the load applied to the discontinuity on the steepest, i.e., contributing, facets. (c, left): Representation of four adjacent data points in the XY plane on which triangular facets are defined and schematic representation of a facet of surface A_i and projected surface A_{ip} (c, right). Res_X and Res_Y are the spatial resolution in the X and Y directions, respectively

$$\tau_p = \frac{f_{\text{peak}}}{A_{\text{tot}}} = \sum_{i=1}^{N_{cf}} \frac{A_{ip} \cdot \left(c + \sigma'_{\text{local\_}i} \cdot \tan(\phi') \right)}{A_{\text{tot}}}$$
$$= \frac{A_{ip} \cdot \left(c + \sigma'_{\text{local\_}i} \cdot \tan(\phi') \right) N_{cf}}{A_{\text{tot}}}$$

For 100 mm per 100 m square samples, as tested by Casagrande et al. (2018), A_{tot} is equal to 10,000 mm^2.

The residual shear strength is calculated from the peak shear force by considering that the difference between the peak and residual forces is the cohesive contribution of all sheared asperities:

$$f_{\text{residual}} = f_{\text{peak}} - c \cdot N_{\text{cf}} \cdot A_{\text{ip}} \tag{16}$$

where f_{peak} is estimated as per Eq. (15), c is the material cohesion, N_{cf} is the total number of contributing facets and A_{ip} is the facet area projected on the XY plane (see Fig. 9). The residual shear strength is then expressed as:

$$\tau_r = \frac{f_{\text{residual}}}{A_{\text{tot}}} = \tau_p - \frac{c \cdot N_{\text{cf}} \cdot A_{\text{ip}}}{A_{\text{tot}}}. \tag{17}$$

Data availability statement Data used to establish the CFR model can be provided upon request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Alejano LR, Gonzalez J, Muralha J (2012) Comparison of different techniques of tilt testing and basic friction angle variability assessment. Rock Mech Rock Eng 45:1023–1035

Barton N (1973) Review of a new shear-strength criterion for rock joints. Eng Geol 7:287–332

Barton N (1976) The shear strength of rock and rock joints. Int J Rock Mech Min Sci Geomech Abstracts 13:255–279

Brady BHG, Brown ET (2004) Rock mechanics for underground mining, 3rd edn. Kluwer Academic Publishers, London, p 628

Buzzi O, Casagrande D (2018) A step towards the end of the scale effect conundrum when predicting the shear strength of large in situ discontinuities. Int J Rock Mech Min Sci 105:210–219

Casagrande D, Buzzi O, Giacomini A, Lambert C, Gordon (2018) A new stochastic approach to predict peak and residual shear strength of natural rock discontinuities. Rock Mech Rock Eng 51:69–99

Cotta C, Mathieson L, Moscato P (2018) Memetic algorithms. In Handbook of heuristics. Springer, London

Fajfar I, Puhan J, Bűrmen Á (2017) Evolving a Nelder–Mead algorithm for optimization with genetic programming. Evol Comput 25(3):351–373

Fenton GA, Vanmarcke EH (1990) Simulation of random fields via local average subdivision. J Eng Mech 116:1733–1749

Grasselli G (2001) Shear strength of rock joints based on quantified surface description. Rock Mech Rock Eng 39:295–314

Grasselli G, Wirth J, Egger P (2002) Quantitative three-dimensional description of a rough surface and parameter evolution with shearing. Int J Rock Mech Min Sci 39:789–800

Hoek E (1983) Strength of jointed rock masses, 23rd Rankine Lecture. Géotechnique 33(3):187–223

Hoek E, Brown E (1997) Practical estimates of rock mass strength. Int J Rock Mech Min Sci 34:1165–1186

Jeffery M, Huang J, Fityus SG, Giacomini A, Buzzi O (2021) A rigorous multiscale random field approach to generate large scale rough rock surfaces. Int J Rock Mech Min 142:104716

Jeffery M, Crumpton M, Fityus SG, Huang J, Giacomini A, Buzzi O (2022) A shear device with controlled boundary conditions for very large nonplanar rock discontinuities. Geotech Test J 45:725–752

Jeffery M, Huang J, Fityus S, Giacomini A, Buzzi O (2023) A Large-scale application of the stochastic approach for estimating the shear strength of natural rock discontinuities. Rock Mech Rock Eng. https://doi.org/10.1007/s00603-023-03393-1

Johnstone D, Milward EA, Berretta R, Moscato P, Initiative ADN (2012) Multivariate protein signatures of pre-clinical Alzheimer's disease in the Alzheimer's disease neuroimaging initiative (ADNI) plasma proteome dataset. PLoS ONE 7(4):e34341

Moscato P, Mathieson L (2019) Memetic algorithms for business analytics and data science: a brief survey. In: Moscato P, de Vries NJ (eds) Business and consumer analytics: new ideas. Springer, Cham

Moscato P, Nazmul Haque M, Huang K, Sloan J, de Oliveira JC (2020a) Learning to extrapolate using continued fractions: predicting the critical temperature of superconductor materials. arXiv e-prints. arXiv:2012.03774

Moscato P, Sun H, Haque MN (2020b) Analytic continued fractions for regression: results on 352 datasets from the physical sciences. In: IEEE congress on evolutionary computation, CEC 2020b, Glasgow, United Kingdom, July 19–24, 2020b. IEEE, New York, pp 1–8

Moscato P, Sun H, Haque MN (2021) Analytic continued fractions for regression: a memetic algorithm approach. Expert Syst Appl 179:115018

Moscato P, Craig H, Egan G, Haque MN, Huang K, Sloan J, de Oliveira JC (2022) Multiple regression techniques for modelling dates of first performances of shakespeare-era plays. Expert Syst Appl 200:116903

Moscato P, Haque MN, Moscato A (2023) Continued fractions and the Thomson problem. Sci Rep 13:7272. https://doi.org/10.1038/s41598-023-33744-5

Rocha de Paula M, Gómez Ravetti M, Berretta R, Moscato P (2011) Differences in abundances of cell-signalling proteins in blood reveal novel biomarkers for early detection of clinical Alzheimer's disease. PLoS ONE 6(3):e17481

Sun H, Moscato P (2019) A memetic algorithm for symbolic regression. In: IEEE congress on evolutionary computation, CEC 2019, Wellington, New Zealand, June 10–13. IEEE, New York, pp 2167–2174

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.