

Article

Not peer-reviewed version

---

# Parameter Variance of the Duncan Formula for Nonlinear Shear Strength of Coarse-Grained Soil

---

[Heng Chi](#)<sup>\*</sup>, Hengdong Wang, Yufeng Jia, [Degao Zou](#)

Posted Date: 9 July 2025

doi: 10.20944/preprints202507.0820.v1

Keywords: coarse-grained soil; shear strength; variance



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Parameter Variance of the Duncan Formula for Nonlinear Shear Strength of Coarse-Grained Soil

Chi Heng <sup>1,2,\*</sup>, Wang Heng Dong <sup>2</sup>, Jia Yu Feng <sup>1</sup> and Zou De Gao <sup>1</sup>

<sup>1</sup> Institute of Earthquake Engineering, School of Infrastructure Engineering, Dalian University of Technology, Dalian, China

<sup>2</sup> Shanghai Municipal Engineering Design Institute (Group) CO., LTD, Shanghai, China

\* Correspondence: chiheng@mail.dlut.edu.cn.

## Abstract

Currently, the shear strength of coarse-grained soil is commonly determined using the Duncan formula, which establishes a relationship between the shear strength and confining pressure. Specifically, the parameters of the Duncan formula are estimated based on available test data using least squares regression, which mitigates the limitations associated with small sample sizes and significant errors in the classic grouping method. However, hypothesis testing in mathematical statistics reveals that the residual variances from the classic least squares estimation exhibit heteroscedasticity and autocorrelation, violating the fundamental assumptions required for least squares regression. Consequently, parameter estimates obtained through classic least squares regression have large variances, leading to unreliable statistical inferences. To address this issue, we propose a generalized least squares method that corrects for heteroscedasticity and autocorrelation of the residuals. Triaxial test data for five different coarse-grained soil materials are analyzed using the proposed method. The results show that the mean values of the estimated parameters of the Duncan formula are close to those obtained using the classic method, while their variances are significantly reduced, demonstrating the effectiveness of the proposed approach.

**Keywords:** coarse-grained soil; shear strength; variance

## 1. Introduction

Coarse-grained soil has the advantages of easy accessibility, low cost, and high permeability and is therefore widely utilized in engineering projects such as roads, earth-rock dams, land reclamation, and high-fill airports. The shear strength of coarse-grained soil is a critical technical parameter as it directly affects both project safety and investment benefits.

The guidelines for determining the standard soil shear strength vary among different standards. For instance, the *Design Code for Rolled Earth-Rock Fill Dams* (SL274-2020) and standard NB/T10872-2021 stipulate that the standard value should be the average of the minimum shear strength. These specifications also define the required number of test groups: for anti-seepage soil materials, at least 11 groups are required, and each group contains 4–6 confining pressure levels. For coarse-grained materials, NB/T10872-2021 mandates a minimum of six test groups, whereas SL274-2020 requires at least 11 groups for both fine-grained soil and coarse-grained soil. The *Unified Standard for Reliability Design of Hydraulic Engineering Structures* (GB5099-2013) recommends using the 0.1 quantile of the probability distribution as the standard value for the strength of geotechnical materials and artificial foundations. In addition, the *Code for Investigation of Geotechnical Engineering* (GB 50021-2001, 2009 Edition) specifies that the standard value of a geotechnical parameter should be based on the one-sided confidence limit of the confidence interval for the mean value of the parameter matrix, which is derived using statistical interval estimation theory. By simplifying this approach, the relationship between the standard value  $\phi_k$  and the mean value  $\phi_m$  can be expressed as follows:

$$\phi_k = \gamma_s \phi_m \gamma_s = 1 \pm \left( \frac{1.704}{n^{1/2}} + \frac{4.678}{n^2} \right) \delta \quad (1)$$

where  $\delta$  is the coefficient of variation, which is the ratio of the standard deviation to the mean.  $n$  is the number of groups.

Apparently, it is equally important to determine the mean and variance (or standard deviation) of the soil shear strength for engineering practices.

Currently, the shear strength of coarse-grained soil is predominantly represented by the Duncan formula. In reliability analyses of the anti-sliding stability and foundation bearing capacity of coarse-grained soil, it is essential to predefine the variance of the parameters. Therefore, it is crucial to investigate the variance of the variables in the Duncan formula for determining the nonlinear shear strength of coarse-grained soil (Aghchai et al., 2020).

## 2. Nonlinear Shear Strength of Coarse-Grained Soil

The shear strength of soil is generally described by the Mohr-Coulomb theory:

$$\tau_f = \sigma_n \tan \varphi + c \quad \sigma_1 = \sigma_3 \tan^2(\varphi/2 + 45^\circ) + 2c \tan(\varphi/2 + 45^\circ), \quad (2)$$

where  $\tau_f$  is the shear strength of the soil,  $\sigma_n$  is the normal stress on the shear plane,  $c$  and  $\varphi$  are the cohesion (or interlocking force) and internal friction angle of the soil, and  $\sigma_1$  and  $\sigma_3$  are the major and minor principal stresses at the limit equilibrium failure point in the triaxial test.

Typically, the indices  $c$  and  $\varphi$  are constants. However, for coarse-grained soil, the shear strength exhibits strong nonlinearity; hence, it is necessary to define these indices within a limited confining pressure range. In contrast to the linear relationship observed under low confining pressure conditions, under high confining pressure conditions, the strength envelope deviates from a straight line. To characterize the shear strength of coarse-grained soil across a wide confining pressure range, the following approaches have been employed in engineering practices. The first method involves a segmented description, in which different values of  $c$  and  $\varphi$  are used for low (typically 100–800 kPa) and high confining pressure conditions. Another approach directly models the curved envelope of the shear strength, such as the Demeler formula  $\tau_f = A\sigma^m$  and the Duncan formula

$$\varphi = \varphi_0 - \Delta\varphi \lg(\sigma_3/P_a), \quad (3)$$

where  $A$  and  $m$  are the shear strength coefficient and exponent.  $\varphi_0$  is the internal friction angle when the confining pressure is equal to the standard atmosphere, and  $\Delta\varphi$  is the decrease in the friction angle when the confining pressure  $\sigma_3$  increases by 10 times. According to the Duncan formula, the friction angle is expressed as follows:  $\varphi = \sin^{-1} \left( \frac{\sigma_1 - \sigma_3}{\sigma_1 + \sigma_3} \right)$ .

There are two methods for determining nonlinear strength indices. In the first method, the strength indices  $\varphi_0$  and  $\Delta\varphi$  are determined based on the test groups, and then, their minimum average values are determined. In the second method, the minimum average values of  $(\sigma_1 - \sigma_3)$  and  $(\sigma_1 + \sigma_3)$  under each confining pressure are determined first, and then, the friction angle is calculated based on these minimum average values. Finally, the strength indices are determined based on the Duncan formula. The parameters derived using these two methods differ and both methods suffer from an insufficient number of grouped samples, leading to large errors and a low accuracy for intermediate transition variables. To address this issue, all test data should be utilized, and the least squares method should be used to directly estimate the nonlinear strength parameters (Chen et al., 2019; Cao et al., 2007).

Linear regression for estimating Duncan's nonlinear shear strength parameters of coarse-grained soil.

$$\varphi = \varphi_0 - \Delta\varphi \lg \left( \frac{\sigma_3}{P_a} \right) \quad (4)$$

where  $\lg(\sigma_3/P_a)$  is the independent variable, which is a non-random variable.  $\Phi$  is the dependent variable, which is a random variable.  $\varphi_0$  and  $-\Delta\varphi$  are the parameters to be determined.

The prerequisite of this method is that the regression residuals meet the requirements of classic least squares estimation.

## 3. Classic Least Squares Estimation

The linear regression model is as follows:

$$(\varphi)_i = \varphi_0 - \Delta\varphi[lg(\sigma_3/P_a)]_i + e_i, (i = 1, 2, \dots, n), (5)$$

where  $e_i$  is the residual term or the error term, which represents the influences of factors other than  $[lg(\sigma_3/P_a)]$  on the friction angle  $\varphi$  and the measurement error.

If the residual term  $e_i$  ( $i = 1, 2, \dots, n$ ) satisfies the following three conditions, then the least squares regression is valid: (a) the mathematical expectation is  $E(e_i) = 0$ , (b) the residuals under different confining pressures have equal variances, that is,  $\text{Var}(e_i) = \sigma^2$ , ( $i = 1, 2, \dots, n$ ), and (c) the residuals under different confining pressures are uncorrelated, that is,  $\text{Cov}(e_i, e_j) = 0$ ,  $i \neq j$ . The above is the Gauss-Markov hypothesis.

The least squares method finds estimates of  $\varphi_0$  and  $-\Delta\varphi$  that minimize the square of the length of the residual vector  $e = (\varphi)_i - \{\varphi_0 - \Delta\varphi[lg(\sigma_3/P_a)]_i\}$ , i.e.,  $\|(\varphi)_i - \{\varphi_0 - \Delta\varphi[lg(\sigma_3/P_a)]_i\}\|^2$ .

By expanding this formula and taking partial derivatives with respect to  $\varphi_0$  and  $-\Delta\varphi$  and setting them equal to zero,  $\varphi_0$  and  $-\Delta\varphi$  can be obtained.

$$\begin{aligned} [lg(\sigma_3/P_a)]_i &= x_i, (\varphi)_i = y_i, \frac{1}{n}\sum_{i=1}^n x_i = \bar{x}, \frac{1}{n}\sum_{i=1}^n y_i = \bar{y}, \\ -\Delta\varphi &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, (6) \\ \varphi_0 &= \bar{y} + \Delta\varphi\bar{x}. \end{aligned}$$

By substituting  $n$  sets of observation data  $([lg(\sigma_3/P_a)]_i, (\varphi)_i)$  into the above equation, we can obtain estimated values of  $\varphi_0$  and  $-\Delta\varphi$ .

#### 4. Homogeneity of Variance and Correlation Test of the Residuals of the Duncan Formula Estimated VIA Least Squares Regression

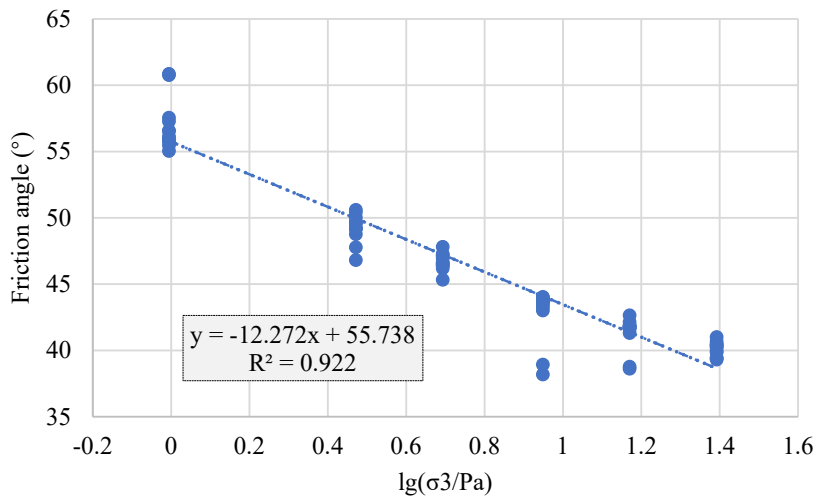
Currently, the prerequisite of the least squares method is generally not tested when determining the shear strength parameters of soil. Chen et al. (2005) noticed the variance of the regression residuals and proposed that the weighted least squares method be applied to improve the regression. The weighting coefficient is  $w_i = 1/(\alpha + \beta x_i)^2$ , and  $\alpha$  and  $\beta$  are obtained via linear regression of the residuals and independent variables. This approach partially eliminates the variance of the residuals, but it does not fundamentally solve the problem. In the regression method recommended by the specification, i.e.,  $q = \frac{\sigma_1 - \sigma_3}{2}$ , and  $p = \frac{\sigma_1 + \sigma_3}{2}$ , the independent variable  $p = \frac{\sigma_1 + \sigma_3}{2}$  is also a random variable. Based on this characteristic, Yu et al. (2012) proposed that the orthogonal least squares method be applied to correct the shear strength estimated based on regression using  $p$ - $q$ . The above methods only involve regressions of the parameters of the Mohr-Coulomb formula, which only partially eliminates the variance of the regression residuals. To the best of our knowledge, the variance and correlation of the residuals in nonlinear regression of the shear strength of coarse-grained soil has not been studied (Chen et al., 2007; Gong et al., 2020; Gong et al., 2018).

##### 4.1. Regression Residuals and the Covariance Matrix

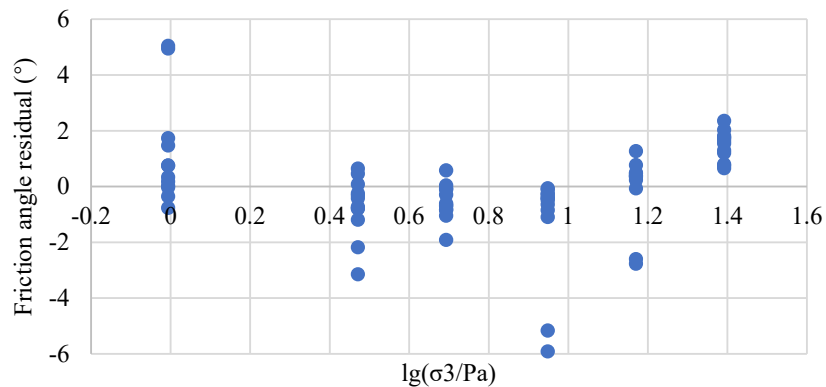
Based on the test results for several coarse-grained materials from the sites of engineering projects, the nonlinear shear strength was determined to examine the variance and correlation of the residuals in the nonlinear strength regression. Specifically, in the triaxial test on coarse-grained granite from Area I of earth-rock dam project A, twelve sets of parallel tests were conducted under six confining pressure conditions (i.e., 100, 500, 1000, 1500, 2000, and 2500 kPa). The regression equation is shown in Figure 1, and the regression residuals are shown in Figure 2. As can be seen from Figure 2, the distribution of the residuals varies across the different confining pressures, indicating that the variance of the residuals is not uniform. To determine whether the residuals under the six confining pressures are correlated, it is necessary to test the covariance matrix of the residuals. The regression residuals of the friction angle from the triaxial tests were calculated, and the residual covariance matrix was obtained:

$$\text{Cov}(e_i, e_j) = \begin{bmatrix} 3.685 & 0.969 & 0.331 & 1.591 & 0.82 & 0.688 \\ 0.969 & 1.147 & 0.308 & 0.747 & 0.438 & 0.353 \\ 0.331 & 0.308 & 0.402 & 0.641 & 0.432 & 0.188 \\ 1.591 & 0.747 & 0.641 & 4.02 & 2.432 & 0.754 \\ 0.82 & 0.438 & 0.432 & 2.432 & 1.607 & 0.468 \\ 0.688 & 0.353 & 0.188 & 0.754 & 0.468 & 0.286 \end{bmatrix} \quad (7)$$

It can be seen that the main diagonal elements of the residual covariance matrix are not equal, and the non-diagonal elements are not zero, suggesting variance homogeneity and correlation of the residuals. In order to accurately determine the homogeneity and correlation of the regression residuals, hypothesis testing is required.



**Figure 1.** Classic least squares regression of nonlinear shear strength and friction angle of coarse-grained soil from Area I of engineering project A.



**Figure 2.** Regression residuals of nonlinear shear strength and friction angle of coarse-grained soil from Area I of engineering project A from large-scale triaxial tests.

First, the Kolmogorov–Smirnov (K-S) method was applied to test whether the residuals as a whole satisfy the assumption of a normal distribution. Next, assuming that the ratio of the sample variances of the regression residuals at each confining pressure level follows the F-distribution, the sample variance at each confining pressure was calculated to test whether the residual variances were consistent across different confining pressures under a certain probability level (quantile). This step corresponds to the variance homogeneity test. Finally, the correlation of the regression residuals under different confining pressure levels was examined.

#### 4.2. K-S Test for Normal Distribution of Regression Residuals

(1) Null hypothesis  $H_0: F(x) = F_0(x)$ ; and alternative hypothesis  $H_1: F(x) \neq F_0(x)$ , where  $F_0(x)$  is the given normal distribution function. (2) The test statistic  $D_n = \sup|F_n(x) - F(x)|$  is selected as the principal variable. When the null hypothesis is true, the test statistic is  $D_n = \sup|F_n(x) - F_0(x)|$ . At the significance level  $\alpha$ , if  $\sqrt{n}D_n > k$ , then the null hypothesis is rejected, where  $k$  is an appropriate constant. Kolmogorov derived the limiting distribution of  $\sqrt{n}D_n$  and tabulated it. (3) The rejection region is determined. For a given significance level  $\alpha$  that satisfies  $P(\sqrt{n}D_n \geq t_\alpha) = \alpha$ , the quantile  $t_\alpha$  is the critical value, and the rejection region is  $[t_\alpha, +\infty)$ .

Table 1 presents the regression residuals for the coarse-grained granite from Area I of project A under various confining pressure conditions. The mean and standard deviation of all of the residuals are  $2.6659 \times 10^{-4}$  and 1.6573. For a significance level of 0.01,  $P(\sqrt{n}D_n \geq t_{0.01}) = 0.01$ , and  $t = 1.63$ , which is taken as an approximate value of  $t_{0.01}$ . Then, the rejection region is  $[1.63, +\infty)$ .  $D_n = \max(d_{01}, d_{02}) = 0.1782$ . Because  $\sqrt{n}D_n = \sqrt{72} \times 0.1782 = 1.5119$ , and it falls outside the rejection region  $[1.63, +\infty)$ , we accept the hypothesis that the regression residuals of the shear strength are normally distributed.

**Table 1.** Least squares regression residuals and the residual variances.

Confining pressure (kPa)	Friction angle residual (°)												Variance
100	0.3	-0.	0.1	0.0	0.7	4.9	-0.	5.0	1.7	1.4	0.7	-0.	3.685
	32	77	75	19	49	45	356	4	3	66	49	033	
300	-0.	-2.	-1.	-0.	-0.	0.0	-3.	-0.	0.6	-0.	-0.	0.4	1.147
	798	182	196	318	421	79	15	251	39	771	728	59	
500	-0.	-1.	0.5	-0.	-0.	-0.	-0.	-0.	0.0	-1.	-0.	-0.	0.402
	669	916	78	296	108	071	833	612	38	05	84	707	
900	-5.	-5.	-0.	-1.	-0.	-0.	-0.	-0.	-0.	-0.	-0.	-0.	4.02
	914	168	487	096	419	259	848	064	406	285	114	644	
1500	-2.	-2.	0.7	0.4	0.4	0.4	0.2	0.3	0.3	1.2	-0.	0.3	1.607
	77	604	63	69	03	87	33	66	84	7	071	35	
2500	0.6	0.7	1.7	1.2	2.0	2.3	0.7	1.8	1.5	1.7	1.2	1.6	0.286
	53	94	57		26	55	46	14	44	28	93	29	

#### 4.3. Homogeneity Test of Regression Residuals Under Different Confining Pressure Conditions

Theoretical studies have shown that the F-test is sensitive to deviations from homogeneity of variance, making it suitable for testing the variance of residuals under different confining pressures. The null hypothesis  $H_0$  and alternative hypothesis  $H_1$  are

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 \text{ and } H_1: \sigma_1^2, \sigma_2^2, \dots, \sigma_k^2, \text{ and they are not all equal.}$$

Determine the rejection region.

Let  $\alpha = 0.05$ ,

$$P(F \leq F_{1-\frac{\alpha}{2}}(n_1 - 1, n_2 - 1)) = \frac{\alpha}{2}, P\left(F \geq F_{\frac{\alpha}{2}}(n_1 - 1, n_2 - 1)\right) = \frac{\alpha}{2}, (8)$$

$$P\left(\frac{S_1^2}{S_2^2} \leq F_{0.975}(11,11)\right) = 0.025,$$

$$F_{0.975}(11,11) = \frac{1}{F_{0.025}(11,11)} = \frac{1}{3.48} = 0.287356,$$

and  $F_{0.025}(11,11) = 3.48$ .

Then, the rejection region is  $(0, 0.2874]$  and  $[3.48, +\infty)$ .

Table 2 shows the sample variance ratios of the regression residuals under various confining pressures and whether they are in the rejection region.  $\sqrt{\phantom{x}}$  indicates that the variance ratio is in the acceptance region and the null hypothesis is accepted, while  $\times$  indicates that the variance ratio is in the rejection region and the null hypothesis is not accepted.

**Table 2.** Ratio of variance of least squares regression residuals under different confining pressures.

Confining pressure (kPa)	100	300	500	900	1500	2500
100	1					
300	0.311/ $\sqrt{\phantom{x}}$	1				
500	0.109/ $\times$	0.351/ $\sqrt{\phantom{x}}$	1			
900	1.091/ $\sqrt{\phantom{x}}$	3.503/ $\times$	9.991/ $\times$	1		
1500	0.436/ $\sqrt{\phantom{x}}$	1.4/ $\sqrt{\phantom{x}}$	3.994/ $\times$	0.4/ $\sqrt{\phantom{x}}$	1	
2500	0.078/ $\times$	0.249/ $\times$	0.71/ $\sqrt{\phantom{x}}$	0.071/ $\times$	0.178/ $\times$	1

Note:  $\sqrt{\phantom{x}}$  denotes acceptance of the null hypothesis,  $\times$  denotes rejection of the null hypothesis.

It can be seen that except for the diagonal elements where the variance ratio is equal to 1, seven of the sample variance ratios of the residuals under different confining pressures are in the acceptance region and eight are in the rejection region. Therefore, the variance homogeneity hypothesis is not accepted.

#### 4.4. Residual Independence Test Under Different Confining Pressure Conditions

Assuming that the regression residuals under different confining pressures are two different random variables U and V, the independence test is equivalent to testing their correlation coefficient  $\rho$ . The null hypothesis  $H_0$  and alternative hypothesis  $H_1$  are  $H_0: \rho = 0$ , and  $H_1: \rho \neq 0$ . The sample correlation coefficient  $R$  is used as the point estimator of  $\rho$ .  $R$  is shown in Table 3.

**Table 3.** Calculation of  $R$ .

Confining pressure (kPa)	100	300	500	900	1500	2500
100	1					
300	0.471	1				
500	0.272	0.453	1			

900	0.413	0.348	0.504	1		
1500	0.337	0.323	0.537	0.957	1	
2500	0.67	0.616	0.555	0.704	0.69	1

$$R = \frac{\sum_{i=1}^n (U_i - \bar{U})(V_i - \bar{V})}{\sqrt{\sum_{i=1}^n (U_i - \bar{U})^2} \sqrt{\sum_{i=1}^n (V_i - \bar{V})^2}} \quad (9)$$

When  $H_0$  is true,

$$t = \frac{R}{\sqrt{1-R^2}} \sqrt{n-2}. \quad (10)$$

It follows a  $t$ -distribution with degrees of freedom of  $n-2$ .  $t$  can be found in Table 4. Given a significance level  $\alpha$ ,

$$P(|t| \geq t_{\frac{\alpha}{2}}(n-2)) = \alpha. \quad (11)$$

If the observed value of  $t$  satisfies  $|t| \geq t_{\frac{\alpha}{2}}(n-2)$ , then the null hypothesis  $H_0$  is rejected.

$$\alpha = 0.1.$$

$P(|t| \geq t_{0.05}(10)) = 0.1$ , and  $t_{0.05}(10) = 1.812$ .

The rejection regions are  $(0, -1.812]$  and  $[1.812, +\infty)$ .

**Table 4.** Calculation of  $t$ .

Confining pressure (kPa)	100	300	500	900	1500	2500
100						
300	1.688/√					
500	0.894/√	1.607/√				
900	1.434/√	1.174/√	1.845/×			
1500	1.132/√	1.079/√	2.013/×	10.432/×		
2500	2.854/×	2.473/×	2.11/√	3.135/×	3.015/×	

Note: √ denotes acceptance of the null hypothesis, × denotes rejection of the null hypothesis.

It can be seen that seven of the above  $t$  values are within the rejection region. Therefore, the independence hypothesis of the residuals under each confining pressure is rejected.

The regression residual analysis of the nonlinear shear strength based on the Duncan formula for the coarse-grained soils from three engineering projects (projects A, B, and C) indicates that the diagonal elements of the covariance matrix of the least squares regression residuals for the nonlinear shear strength friction angle are not equal, and the non-diagonal elements are not zero. Although the probability distribution of the residuals passes the normality test, the residuals under different confining pressures exhibit homogeneity of variance and irrelevance. This confirms the occurrence of residual variance heteroscedasticity and autocorrelation across the different confining pressures.

## 5. Elimination of Residual Variance Heteroscedasticity and Autocorrelation

If the residuals exhibit heteroscedasticity and autocorrelation, the regression equation no longer meets the prerequisites of least squares regression. In this case, if the least squares method is used, although the estimated values of the parameters are unbiased, the variances of the parameters could



deviate from the actual value. A solution is to replace the regression variables. For the linear regression equation

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \mathbf{E}(\mathbf{e}) = \mathbf{0}, \mathbf{Cov}(\mathbf{e}) = \sigma^2\mathbf{V}, \quad (12)$$

where  $\mathbf{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix}'$ ,  $\mathbf{y} = [\varphi_1 \ \varphi_2 \ \cdots \ \varphi_n]'$ .  $\boldsymbol{\beta} = [\varphi_0 \ -\Delta\varphi]$ ,  $x_1 = \lg(\sigma_3^1/P_a)$ ,  $\cdots$ ,  $x_i = \lg(\frac{\sigma_3^i}{P_a})$ ,  $\cdots$ , and  $x_n = \lg(\sigma_3^n/P_a)$ ,  $\sigma_3^i$  and  $\varphi_i$  are the confining pressure and friction angle of the  $i$ th test. The symbol  $'$  denotes matrix transposition.  $\mathbf{V}$  is the covariance matrix of the residual vector under different confining pressures.  $\sigma^2$  is the variance of the residual vector.

It can be seen that the covariance matrix of the residual is not an identity matrix.  $\mathbf{V}$  is nonsingular and positive definite, and there exists a nonsingular symmetric matrix  $\mathbf{K}$  such that  $\mathbf{K}'\mathbf{K} = \mathbf{K}^{-1}\mathbf{K} = \mathbf{V}$ , that is, the matrix  $\mathbf{K}$  is the square root matrix of  $\mathbf{V}$ .

By defining new variables (Wang et al., 2021; Douglas et al., 2022), namely,

$$\mathbf{z} = \mathbf{K}^{-1}\mathbf{y}, \mathbf{B} = \mathbf{K}^{-1}\mathbf{X}, \text{ and } \mathbf{g} = \mathbf{K}^{-1}\mathbf{e} \quad (13)$$

the regression equation  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$  becomes  $\mathbf{K}^{-1}\mathbf{y} = \mathbf{K}^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{K}^{-1}\mathbf{e}$ , i.e.,

$$\mathbf{z} = \mathbf{B}\boldsymbol{\beta} + \mathbf{g}. \quad (14)$$

After this transformation, the model residuals have a zero mean, i.e.,  $\mathbf{E}(\mathbf{g}) = \mathbf{K}^{-1}\mathbf{E}(\mathbf{e}) = \mathbf{0}$ , and the covariance matrix of  $\mathbf{g}$  is

$$\begin{aligned} \mathbf{Var}(\mathbf{g}) &= \{[\mathbf{g} - \mathbf{E}(\mathbf{g})][\mathbf{g} - \mathbf{E}(\mathbf{g})]'\} = \mathbf{E}(\mathbf{g}\mathbf{g}') \\ &= \mathbf{E}(\mathbf{K}^{-1}\mathbf{e}\mathbf{e}'\mathbf{K}^{-1}) = \mathbf{K}^{-1}\mathbf{E}(\mathbf{e}\mathbf{e}')\mathbf{K}^{-1} \\ &= \sigma^2\mathbf{K}^{-1}\mathbf{V}\mathbf{K}^{-1} = \sigma^2\mathbf{K}^{-1}\mathbf{K}\mathbf{K}\mathbf{K}^{-1} \\ &= \sigma^2\mathbf{I}. \end{aligned}$$

Therefore,  $\mathbf{g}$  has a zero mean, constant variance, and uncorrelated variance. Equation (14) satisfies the general assumptions of the classic least squares estimation, and the least squares function is

$$\mathbf{Q}(\boldsymbol{\beta}) = \mathbf{g}'\mathbf{g} = \mathbf{e}'\mathbf{V}^{-1}\mathbf{e} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

The canonical equation for least squares is

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}}^* = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}.$$

The solution to the equation is

$$\hat{\boldsymbol{\beta}}^* = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}. \quad (15)$$

## 6. Examples

The triaxial test data for five types of coarse-grained materials used in three engineering projects (A, B, and C) were used to perform classic least squares regression and regression after eliminating the heteroscedasticity and autocorrelation of the residual variance. After obtaining the regression equation, the covariance matrix of the parameters was calculated (Table 5).

**Table 5.** Regression parameters of the Duncan formula for nonlinear shear strength of coarse-grained materials.

	Regression method	Regression equation	Covariance matrix
Slightly weathered coarse-grained granite from	Classic least squares method	$\varphi = 55.738 - 12.272\lg(\sigma_3/P_a)$	$\begin{bmatrix} 2.054 & -1.954 \\ -1.954 & 2.510 \end{bmatrix}$
	Generalized least squares method	$\varphi = 54.321 - 10.227\lg(\sigma_3/P_a)$	$\begin{bmatrix} 0.748 & -0.402 \\ -0.402 & 0.266 \end{bmatrix}$

Area I of the engineering project A			
Coarse aggregate from Area II of engineering project A	Classic least squares method	$\varphi = 50.309 - 11.680\lg(\sigma_3/P_a)$	$\begin{bmatrix} 1.854 & -1.763 \\ -1.763 & 2.265 \end{bmatrix}$
Area II of engineering project A	Generalized least squares method	$\varphi = 50.709 - 11.721\lg(\sigma_3/P_a)$	$\begin{bmatrix} 0.569 & -0.312 \\ -0.312 & 0.309 \end{bmatrix}$
Weathered coarse-grained granite from Area II of engineering project B	Classic least squares method	$\varphi = 54.234 - 12.033\lg(\sigma_3/P_a)$	$\begin{bmatrix} 1.939 & -1.844 \\ -1.844 & 2.369 \end{bmatrix}$
Area II of engineering project B	Generalized least squares method	$\varphi = 54.004 - 12.131\lg(\sigma_3/P_a)$	$\begin{bmatrix} 1.137 & -1.396 \\ -1.396 & 1.884 \end{bmatrix}$
Coarse aggregate from Area II of engineering project B	Classic least squares method	$\varphi = 53.190 - 13.415\lg(\sigma_3/P_a)$	$\begin{bmatrix} 2.477 & -2.355 \\ -2.355 & 3.026 \end{bmatrix}$
Area II of engineering project B	Generalized least squares method	$\varphi = 52.824 - 12.415\lg(\sigma_3/P_a)$	$\begin{bmatrix} 1.504 & -1.499 \\ -1.499 & 1.807 \end{bmatrix}$
Gravel clay material from engineering project C	Classic least squares method	$\varphi = 39.977 - 12.374\lg(\sigma_3/P_a)$	$\begin{bmatrix} 2.200 & -2.090 \\ -2.090 & 2.685 \end{bmatrix}$
Area II of engineering project C	Generalized least squares method	$\varphi = 41.75 - 11.18\lg(\sigma_3/P_a)$	$\begin{bmatrix} 2.181 & 0.920 \\ 0.920 & 1.152 \end{bmatrix}$

Note: The generalized least squares method can eliminate residual variance heteroscedasticity and autocorrelation. Intermediate variables are used to ensure that the regression meets the prerequisites of the least squares method.

It can be seen that for each coarse-grained material, the regression parameters obtained using the two methods are similar, yet the variances of the parameters are reduced when the proposed method is applied.

For example, for the coarse-grained material from Area II of Project A, the regression equation from the classic least squares method and the covariance matrix between  $\varphi_0$  and  $\Delta\varphi$  are

$$\varphi = 50.309 - 11.680 \lg\left(\frac{\sigma_3}{P_a}\right) \quad (16)$$

$$\text{and Cov}(\varphi_0, \Delta\varphi) = \begin{bmatrix} 1.854 & -1.763 \\ -1.763 & 2.265 \end{bmatrix} \quad (17)$$

After eliminating the residual variance heteroscedasticity and autocorrelation, the regression equation and covariance matrix are

$$\varphi = 50.709 - 11.721 \lg(\sigma_3/P_a) \quad (18)$$

$$\text{and Cov}(\varphi_0, \Delta\varphi) = \begin{bmatrix} 0.569 & -0.312 \\ -0.312 & 0.309 \end{bmatrix} \quad (19)$$

It can be seen that after eliminating the heteroscedasticity and autocorrelation, the parameter variance is reduced, providing more accurate estimates of the shear strength parameters of the coarse-grained soil.

## 7. Conclusions

The shear strength of coarse-grained soil has strong non-linearity and is commonly determined using the Duncan formula. However, the classic grouping method for estimating the parameters has the limitations of small sample sizes and large errors. When the classic least squares method is applied to estimate the Duncan shear strength parameters based on triaxial test data, it has been found that the residual variance exhibited heteroscedasticity and autocorrelation, which violates the prerequisites of the least squares method. In this study, we introduced a generalized least squares method to eliminate these issues. By analyzing triaxial test data from five types of coarse-grained soil materials used in three engineering projects, it was found that while the mean values of the shear strength parameters remained relatively unchanged compared to those obtained using the classic least squares method, the variance was significantly reduced, demonstrating the effectiveness of the proposed method. These improvements enable a more accurate characterization of the nonlinear shear strength of coarse-grained soil materials, enhancing the precision of geotechnical calculations.

**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 supported by National Natural Science Foundation of China 52379116. We thank LetPub (www.letpub.com.cn) for its linguistic assistance during the preparation of this manuscript.

## References

1. SL274-2020, Design Code for Rolled Earth-rock Fill Dams [S], DOI: 10.13722/j.cnki.jrme.2022.1202
2. NB/T10872-2021, Design Code for Rolled Earth-rock Fill Dams [S], DOI: 10.3876/j.issn.1000-1980.2025.02.005
3. GB5099-2013, Unified standard for reliability design of hydraulic engineering structures [S].
4. GB 50021-2001, Code for investigation of geotechnical engineering (2009 Edition) [S].
5. CHEN Li-hong, CHEN Zu-yu, LI Guang-xin. Discussion of linear regression method to estimate shear strength parameters from results of triaxial tests. *Rock and Soil Mechanics*. 2005, 26(11): 1785–1789, DOI: 10.16285/j.rsm.2005.11.011.
6. YU Dong-ming, YAO Hai-lin, WU Shao-feng. Difference and modification of regression analysis methods to estimate shear strength parameters obtained by triaxial test. *Rock and Soil Mechanics*. 2012, 33 (10) : 3037-3042, DOI: 10.16285/j.rsm.2012.10.019.
7. CHEN Li-hong, CHEN Zu-yu, LI Guang-xin. A modified linear regression method to estimate shear strength parameters. *Rock and Soil Mechanics*. 2007, 28 (7) : 1421-1426, DOI: 10.16285/j.rsm.2007.07.010
8. WANG Gui-song, CHEN Min, CHEN Li-ping. Linear statistical model - linear regression and variance analysis. Beijing: Higher Education Press, 2021, ISBN: 9787040076059.
9. DOUGLAS C. Montgomery, Elizabeth A. Peck, G Geoffrey Vining. Introduction to Linear Regression Analysis (Fifth Edition), Beijing: China machine Press, 2022, ISBN: 9787111532828.

10. GONG Feng-qiang, LUO Song, GE Lin, et al. Evaluation of shear strength parameters of rocks by preset angle shear, direct shear and triaxial compression tests[J]. *Rock Mechanics and Rock Engineering*, 2020, 53: 2505-2519, DOI: 10.1007/s11440-020-00908-1.
11. SHEN Jia-yi, SHU Zheng, CAI Ming, et al. A shear strength model for anisotropic blocky rock masses with persistent joints[J]. *International Journal of Rock Mechanics and Mining Sciences*, 2020, 134: 1-11, DOI: 10.1016/j.ijrmms.2020.104441.
12. TANG Xiao-song, LI Dian-qing, WANG Xiao-gang, et al. Statistical characterization of shear strength parameters of rock mass for hydropower projects in China [J]. *Engineering Geology*, 2018, 245: 258-265, DOI: 10.1016/j.enggeo.2018.07.014.
13. AGHCHAI M H, MAAREFVAND P, RAD H S. Analytically determining bond shear strength of fully grouted rock bolt based on pullout test results[J]. *Periodica Polytechnica Civil Engineering*, 2020, 64(1): 212-222, DOI: 10.3311/PPci.14411.
14. CHEN Li-hong, LI Xu, XU Yao, et al. Accurate estimate of soil shear strength parameters[J]. *Journal of Central South University*, 2019, 26: 1000-1010, DOI: 10.1007/s11771-019-4264-4
15. CAO Wen-gui, ZHANG Yong-jie. Study on determining method for parameters of rock's shear strength based on asymmetric triangular fuzzy numbers[J]. *Chinese Journal of Rock Mechanics and Engineering*, 2007, 26(7): 1340-1346, DOI: 10.13722/j.cnki.jrme.2007.07.010

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.