A comparison of probabilistic distributions of undrained shear strength of soils in Nipigon River, Canada

N S Kanwar\textsuperscript{1,2,3} and J Deng\textsuperscript{1,2}

\textsuperscript{1}Department of Civil engineering, Lakehead University, Thunder Bay, Ontario, P7B 5E1, Canada
\textsuperscript{2}Centre of Excellence for Sustainable Mining and Exploration, Lakehead University, Thunder Bay, Ontario, P7B 5E1, Canada
\textsuperscript{3}Email: nkanwar@lakeheadu.ca

Abstract. In probabilistic reliability analysis and design, critical geotechnical variables such as soil shear strength are usually regarded as random variables with a probability distribution rather than deterministic values or constants. In this paper, the vane shear test is briefly introduced and used to obtain undrained shear strength of soil in the area of Nipigon river landslide, Ontario, Canada. Then the maximum entropy method is presented to generate an unbiased probabilistic distribution for soil properties based on optimal-order moments from observed soil samples. A comparative study between maximum entropy distributions and traditional lognormal and normal distributions is conducted to evaluate the performance of fitted probabilistic distributions. Kolmogorov-Smirnov goodness of fit test shows that the maximum entropy distribution with four order moments fit the undrained shear strength best. The analytical entropy distribution obtained can be used in probabilistic reliability analysis.

1. Introduction

Soil properties play a critical role in analysis, design and construction of geotechnical slope engineering.

In order to determine its engineering properties quantitatively, soil must be measured, tested, and evaluated by laboratory testing, in situ testing, or monitoring during construction. Vane shear test is one the most important in-situ testing, because it has some advantages: (a) It is a simple standardized yet robust test and easy to carry out; (b) Shear strength is directly measured; (c) It is the only in situ test by which residual strength is directly measured; (d) Sometimes borehole is not needed; but rod friction needs to be eliminated or measured [1]. However, cohesionless soils, such as sandy and gravel soil, are not able to maintain the undrained conditions, thus cannot be tested with the vane shear test.

It is well known that soil properties in slope engineering, similar to other geotechnical parameters, are bound to uncertainties, such as inherent variability in soil properties, geological incongruities, climatic and environmental conditions, analytical and computational errors, etc. In probabilistic approach, modelling and characterization of uncertainties in random variables are the first and foremost step [2]. This is because subsequent reliability analysis of structures is dependent upon the characterization of random variables. Uncertainties associated with random variables are usually quantified by probability curves, mostly by probability distribution curves and its parameters. The potential distribution can be determined by using one of the following methods: (a) Plotting the sample data on a probability paper; (b) Constructing a frequency diagram; and (c) Conducting
goodness-of-fit tests for distribution, such as Kolmogorov-Smirnov (K-S) test and Chi-square test. To uniquely define a distribution, its parameters need to be estimated using available data by the method of moments or method of maximum likelihood. In geotechnical engineering, Normal distribution and lognormal distribution are the most commonly used two probabilistic distributions [3].

An alternative approach for probabilistic distribution fitting stems from maximum entropy principle. The maximum entropy is based on Shannon's entropy which is defined as a measure of uncertainty in observed sample information. The maximum entropy method has been adopted by various researchers to estimate the probability density function [4,5]. Deng et al. developed rigorous quantile functions being exceptionally fit for a small sample size by using the maximum entropy principle [6-8].

This paper focuses on a comparison of probabilistic distributions of undrained shear strength of soils in Nipigon river area, north-western Ontario, Canada. The vane shear test is used to obtain the soil sample of undrained shear strength, and entropy distributions of various orders are determined from the test sample data and compared to normal distribution and lognormal distribution, with the purpose of finding the best probabilistic density function for the undrained shear strength of soils.

2. Vane shear test

2.1. Nipigon river landslide

The research is aimed at soil investigation in the area of Nipigon river landslide. The landslide occurred in the morning of April 23rd, 1990, at the north of the town of Nipigon, Ontario, Canada (Figure 1). It involved almost 300,000 m$^3$ of soil, and extended almost 350 m inshore of the river with the maximum width of approximately 290 m. This resulted in substantial property damage and significant long term environmental and economic consequences. Since then, there have been frequent failures of both natural and artificial slopes in the area [9].

![Figure 1. Location of vane shear soil testing.](image-url)

The land comprises of mostly glacioclustrine plain and pockets of sand silt in the delta. The drainage conditions are relatively poor in the local relief. The aerial photography report after the landslide revealed that the area had many bank failures. The land is mostly wet with poor drainage
conditions adding instability to the banks of the region. The recent investigation in 2018 of the Nipigon slope by our Lakehead University graduate students, clearly indicates that the area is prone to frequent slope failures of various scales.

2.2 Vane shear testing of soils in Nipigon River, Canada

The vane shear test is one of the most commonly adopted tests to obtain accurate undrained shear strength of in-situ saturated cohesive soils. The soils considered are weak and compressible, having the properties of soft to firm clayey soils. The method is a robust and time saving test as compared to other soil field tests, because it creates less disturbance of the soil while implementing the test on the soil.

American Society for Testing and Materials (ASTM D2573-08) standards were followed to complete the vane shear test [10]. The main procedure of the VST is: (a) Place the vane shear equipment at the borehole. According to the required depth, the test can be carried out in the pre-drilled borehole or the rod can be inserted without drilling any hole before the test. (b) Push the vane slowly into the borehole or vane housing with a single thrust to the required depth. (c) Apply torque at a rate of 0.1 deg/sec and record the maximum torque at the failure. (d) Rotate the vane continuously 7 to 10 times and record the residual torque at the end.

The testing was carried out at the Nipigon slope where the failure occurred in 1992, and the coordinates of one testing site are 49°4'33"N, 88°18'34"W. The vane shear tests were performed at various bore holes on the top 1 meter to 3 meters of the soil layer, generally in the firm clayey silt layer, to obtain 121 values of the undrained shear strength, which are listed in Table 1.

| Table 1. Undrained shear strength (kPa) values from VST at the Nipigon river slope (121 values). |
|---|---|---|---|---|---|---|---|
| 11 | 27.55 | 35.15 | 39.9 | 51.3 | 57 | 65.55 | 71.25 |
| 18.05 | 28.5 | 35.15 | 39.9 | 52.25 | 57 | 65.55 | 74.1 |
| 18.05 | 29.45 | 36.1 | 39.9 | 52.25 | 57.95 | 66.5 | 77.9 |
| 19 | 29.45 | 37.05 | 40.85 | 52.25 | 57.95 | 66.5 | 80.75 |
| 19 | 30.4 | 37.05 | 40.85 | 52.25 | 57.95 | 66.5 | 80.75 |
| 19 | 32.3 | 38 | 41.8 | 53.2 | 58.9 | 66.5 | 80.75 |
| 20.9 | 32.3 | 38 | 41.8 | 53.2 | 60.8 | 67.45 | 81.7 |
| 20.9 | 33.25 | 38 | 42.75 | 54.15 | 61.75 | 67.45 | 84.55 |
| 22.8 | 33.25 | 38 | 44.65 | 54.15 | 61.75 | 68.4 | 85.5 |
| 22.8 | 33.25 | 38 | 44.65 | 55.1 | 62.7 | 68.4 | 87.4 |
| 23.75 | 33.25 | 38 | 45.6 | 55.1 | 63.65 | 68.4 | 95 |
| 23.75 | 33.25 | 38.95 | 46.55 | 56.05 | 64.6 | 68.4 | 95 |
| 25.65 | 33.25 | 39.9 | 47.5 | 56.05 | 64.6 | 68.4 | 95 |
| 26.6 | 33.25 | 39.9 | 49.4 | 57 | 64.6 | 68.4 | 96.9 |
| 26.6 | 34.2 | 39.9 | 49.4 | 57 | 64.6 | 71.25 | 104.5 |
| 104.5 |

3. Probabilistic modelling of undrained shear strength of soil

The undrained shear strength is considered as a random variable. Maximum entropy distributions are developed and compared to normal distribution and lognormal distribution.

3.1 Maximum entropy distribution

Mathematically, entropy is given as

$$ H = - \int_{D} f(x) \ln [ f(x) ] dx $$

(1)
where $X$ is a continuous random variable in the domain $D$ and $f(x)$ is the probability density function. A function is defined to measure the entropy difference between two probability assignments $p_1(x)$ and $p_2(x)$.

$$H[p_1(x), p_2(x)] = \int_D p_1(x) \ln\left(\frac{p_1(x)}{p_2(x)}\right) dx$$

(2)

Jayne states that "the minimally prejudiced assignment of probabilities is one which minimizes the entropy subject to the satisfaction of the constraints imposed by the available information". Minimize

$$H[p(x), p_0(x)] = \int_D p(x) \ln\left(\frac{p(x)}{p_0(x)}\right) dx$$

(3)

subject to satisfaction of constraints:

$$p(x) \geq 0, \int_D p(x) \, dx = 1,$$

(4)

And

$$\int_D x^k p(x) \, dx = \mu_k, \quad k = 1, 2, ..., K,$$

(5)

where $p_0(x)$ and $p(x)$ are the prior distribution function and true distribution function of the random variable $X$, respectively. $K$ is the highest order of moments of the random variable considered, $\mu_k$ is the $k$th moment about the origin. The solution of the minimization problem is

$$p(x|\mu) = p_0(x) \exp\left[\lambda_0 + \sum_{k=1}^{K} \lambda_k x^k\right],$$

(6)

where $\lambda_k$ is a set of Lagrangian multipliers associated with the physical constraints Eq. (5), $\lambda_0$ is the multiplier associated with normalization constraint in Eq. (4). Substituting Eq. (6) into Eq. (4) yields

$$\lambda_0 = -\ln\left(\int_D p_0(x) \exp\left[\sum_{j=1}^{K} \lambda_j x^j\right] \, dx\right).$$

(7)

Substituting Eqs. (6) and (7) into Eq. (5), one can compute the values of $\lambda$ from the following equation

$$\frac{\int_D x^k p_0(x) \exp\left[\sum_{j=1}^{K} \lambda_j x^j\right] \, dx}{\int_D p_0(x) \exp\left[\sum_{j=1}^{K} \lambda_j x^j\right] \, dx} = \mu_k, \quad k = 1, 2, ..., K $$

(8)

If an observed sample, $x_1, x_2, ..., x_n$ is given, then the $k$th moment about the origin, $\mu_k$, can be estimated as

$$\mu_k \approx \hat{\mu}_k = \frac{1}{n} \sum_{j=1}^{n} x_j^k$$

(9)

### 3.2 Goodness-of-fit test

The Kolmogorov-Smirnov test compares the cumulative distribution function of an assumed probabilistic distribution and the observed cumulative frequencies from a sample data [2]. The first step is to rearrange the sample data in increasing order. Then the maximum difference between the two cumulative distribution functions of the ordered data can be calculated as

$$D_n = \max|F_X(x_i) - S_n(x_i)|,$$

(10)

where $F_X(x_i)$ is the theoretical cumulative distribution function of the assumed distribution for the $i$th data, and $S_n(x_i)$ is the corresponding stepwise cumulative distribution function of the observed ordered samples, which can be determined as

$$S_n(x_i) = \begin{cases} 0 & x_i < x_1 \\ \frac{m}{n} & x_m \leq x_i \leq x_{m+1}, \\ 1 & x_i \geq x_n \end{cases}$$

(11)
if the maximum difference $D_n$ is equal to or less than the tabulated value $D_n^{\alpha}$, the assumed distribution is acceptable at the significance level $\alpha$.

The advantage of the Kolmogorov-Smirnov test is that there is no need to divide the data into intervals; thus the determination of the number or size of the interval is avoided.

3.3 Results and discussion

From Table 1, the mathematical moments can be calculated, and the results are listed in Table 2, where central moments, transformed central moments on [0,1], and moments about the origin on [0,1] are shown. The moments have been transformed to the domain [0,1] in order to increase the accuracy.

Based on these moments, the maximum entropy distribution can be determined by using the method in Section 3.1. It is found the 4th order entropy distribution is the best, where the parameters are shown in Table 3, i.e.,

$$f_1(x) = \exp(-10.16360285 + 0.40671402x - 0.0092029036x^2 + 8.5959908 \times 10^{-5}x^3 - \frac{3.04649777 \times 10^{-7}}{x^4}),$$

where $f_1(x)$ is the maximum entropy density function, $X$ is the random variable for soil undrained shear strength. For comparison, normal distribution and lognormal distribution are also fitted to the soil sample data and their parameters are estimated by using the method of maximum likelihood, as shown in Table 3 with the probability density distributions as

$$f_2(x) = \frac{1}{21.1043 \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{x-51.6847}{21.1043} \right)^2 \right],$$

$$f_3(x) = \frac{1}{0.4298 \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{\ln x - 3.8582}{0.4298} \right)^2 \right],$$

where $f_2(x)$ is the normal probability density function and $f_3(x)$ is the lognormal probability density function.

Table 2. Mathematical moments of the sample of Undrained shear strength (kPa).

| Central moments            | Mean          | Variance    | 3rd order | 4th order |
|----------------------------|---------------|-------------|-----------|-----------|
| Transformed central moments on [0,1] | 0.3973        | 0.02616     | 0.002436  | 0.002003  |
| Moments about the origin on [0,1] | 0.3973        | 0.1840      | 0.09635   | 0.0555796 |

Table 3. Parameters of entropy, Normal, and lognormal distributions.

| Entropy distribution | $\lambda_0$ | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ |
|----------------------|-------------|-------------|-------------|-------------|-------------|
| Entropy distribution | 10.16360285 | -0.40671402 | 0.0092029036 | -8.5959908e-5 | 3.04649777e-7 |
| Normal distribution  | $\mu$=51.6847 | $\sigma$=21.1043 |
| Lognormal distribution | $\lambda$=3.8582 | $\zeta$=0.4298 |

Table 4. K-S test for Normal, lognormal, and entropy distributions.

| Normal distribution Entropy distribution (K=2) | Entropy distribution (K=3) | Entropy distribution (K=4) | Entropy distribution (K=5) | Lognormal distribution |
|-----------------------------------------------|-----------------------------|-----------------------------|-----------------------------|------------------------|

5
Figure 2 compares the histogram for soil undrained shear strength and normal, lognormal, and entropy distributions. It clearly depicts that the entropy model fits the histogram best. The cumulative distribution functions are compared in Figure 3. All three distributions seem to fit the stairstep graph well.

To determine whether the fitted models are good enough over the sample data or not, Kolmogorov-Smirnov goodness-of-fit test is conducted. The value of $D^\alpha_{121}$ at the significance level $\alpha = 0.05$ for a sample of 121 elements is $D_{121}^{0.05} = \frac{1.35}{\sqrt{121}} = 0.1227$ [2]. All the $D_n$ in Table 4 are less than this $D_{121}^{0.05}$, which suggests that all three distributions are acceptable. However, the entropy distribution with order $K=4$ seems the best, because $D_n = 0.06267$ is the smallest among all the distributions.

| $D_n$   | 0.1017 | 0.09442 | 0.08576 | 0.06267 | 0.06799 | 0.08588 |

Figure 2. Histogram for soil undrained shear strength and Normal, lognormal, and entropy distributions.

Figure 3. K-S test on soil undrained shear strength for Normal, lognormal, and entropy distributions.
4. Conclusions
The vane shear test is one of the most commonly used soil tests to determine the undrained shear strength of the in-situ saturated cohesive soil. American Society for Testing and Materials (ASTM D2573-08) standards were followed to conduct the vane shear test in Nipigon river landslide area, northwestern Ontario, Canada. The soil sample is comprised of 121 values of undrained shear strength. Maximum entropy method is presented to generate an unbiased probabilistic distribution for soil variables based on optimal-order moments from soil samples. A comparative study between maximum entropy models with various orders and traditional lognormal and normal distributions is conducted to evaluate the performance of fitted probabilistic distributions. Kolmogorov-Smirnov goodness of fit test shows that the entropy distribution with four order moments fits the undrained shear strength data best. The analytical entropy distribution can then be utilized in future probabilistic reliability analysis. One of the advantages of the maximum entropy method is that sophistication of the entropy distribution can be judiciously adjusted such that the most unbiased and optimal model of distribution can be obtained. In comparative study depicts that the traditional determination of distributions and parameters from observed sample data lags behind due to the restriction on the family of presumed standard classical distributions. There exist no solid grounds that all geotechnical variables should be confined to presumed standard classical distributions (such as normal, lognormal, exponential distribution, etc).

References
[1] Ameratunga J, Sivakugan N and Das B M 2016 Correlations of Soil and Rock Properties in Geotechnical Engineering (New Delhi, India: Springer India)
[2] Haldar A and Mahadevan S 2000 Probability, Reliability and Statistical Methods in Engineering Design (New York: John Wiley)
[3] Cao Z, Wang Y and Li D 2017 Probabilistic Approaches for Geotechnical Site Characterization and Slope Stability Analysis (Berlin: Springer-Verlag Berlin Heidelberg)
[4] Kapur J N 1989 Maximum Entropy Models in Science and Engineering (New York: JohnWiley & Sons)
[5] Singh V P 2013 Entropy Theory and Its Applications in Environmental and Water Engineering (New York: John Wiley)
[6] Deng J, Pandey M D and Xie W C 2012 Maximum entropy principle and partial probability weighted moments AIP Conference Proceedings 1443 190–7
[7] Deng J and Pandey M D 2009 Using partial probability weighted moments and partial maximum entropy to estimate quantiles from censored samples Probabilistic Eng. Mech. 24 407–17
[8] Deng J, Pandey M D and Gu D 2009 Extreme quantile estimation from censored sample using partial cross-entropy and fractional partial probability weighted moments Struct. Saf. 31 43–54
[9] Dodds R B, Burak J P and Eigenbrod K D 1993 Nipigon River Landslide Third Int. Conf. on Case Histories in Geotechnical Engineering (St. Louis, Missouri) pp 517-23
[10] ASTM 2008 Standard Test Method for Field Vane Shear Test in Cohesive Soil D2573-08 (United States of America: ASTM International)