

Discussion: Refining shear strength characteristic value using experience

David J. Baxter MEng, MSc, CEng, MICE

Research Engineer, Centre for Innovative and Collaborative Engineering, Loughborough University, UK

Neil Dixon BSc, PhD, FGS

Professor of Geotechnical Engineering, Department of Civil and Building Engineering, Loughborough University, UK

Paul R. Fleming BEng, PhD, MIHT

Senior Lecturer in Geotechnical Engineering, Department of Civil and Building Engineering, Loughborough University, UK

Ken Cromwell BEng, CEng, MICE

Managing Director, KC Piling Consultants; formerly Managing Director, Rock and Alluvium Ltd

Robert A. M. Watkins BSc, CEng, FICE

FKW Ltd, Consulting Engineers, Bristol, UK

Contribution by Robert A. M. Watkins

The authors have presented an interesting paper (Baxter *et al.*, 2008) on the adoption of statistical techniques to provide estimates for soil parameters. The use of a database of prior results and new values to refine soil shear strength values is particularly interesting, since the technique may apply to parameters for many materials, and not just the example given for soil shear strength.

The authors suggest the use of Bayes' theorem for this problem. This approach appears to follow the basic approach set out in Appendix D to BS EN 1990 (BSI, 2002).

To understand the data processing suggested by the authors, the discussor attempted to recreate the processing of the Woolwich sample database from the information contained within the paper. In so doing, a number of observations were made. It would be helpful if the authors could comment on these observations.

The Woolwich sample (and the prior database) contained values that at first sight appear either very high or quite low. Did the authors consider applying the use of statistical techniques, such as the analysis of residuals, to the original database or the Woolwich sample to identify possible outliers? The prior database was obtained from many different sources (31 different contractors). It is well known that the presence of outliers in a dataset can affect the slope of a regression line, even if the mean value of the data is not much affected. As an example, if the grouped data from the original dataset are analysed with and without possible outliers (which occur at depths below 25 m), the resulting regression equations are

$$5. \quad y = 6.438x + 61.56 \quad (R^2 = 0.8658) \text{ with outliers}$$

$$6. \quad y = 6.048x + 64.25 \quad (R^2 = 0.9524) \text{ without outliers}$$

where y is the shear strength, x is the depth below ground level and R is Pearson's correlation coefficient.

The Woolwich sample data, if analysed for error residuals, appear to contain two outliers. These might, in any case, have been identified from the data plot in Figure 4. It seems likely that there are also three or four outliers in the original dataset. There are physical reasons why such results might have occurred, and these include London Clay containing a sandy or gravelly matrix sometimes found at the base of the stratum, the presence of nodules within the clay matrix and clay with variable moisture contents.

The authors give a value for the coefficient of variation for the Woolwich sample of 0.3. It is not clear how this is arrived at. Observed values of shear strength, admittedly scaled from Figure 4, provide a raw coefficient of variation of the sample of 0.44. However, calculating the coefficient of variation based on the residuals from the linear regression appears to give a coefficient of variation of 0.36, and it is this latter value that the discussor considers might be appropriate for use with the analysis presented by the authors.

For a parameter that cannot take a negative value, as the coefficient of variation becomes larger, it becomes more likely that the distribution is skewed. Where the coefficient of variation is 0.5 or greater, it is quite likely that the distribution is skewed. Although part of the sample distribution may closely resemble the normal distribution, the log-normal distribution may provide a better fit. Although this involves additional computation, did the authors consider formally testing the distribution? The discussor has previously used the Kolmogorov–Smirnov test for goodness of fit when assessing results for parameters of other materials.

It is not immediately clear from the paper what hypothesis is being tested when the Woolwich sample is being compared with the prior database. Are the authors suggesting that the least squares regression on the prior database represents the best estimate of the relationship between shear strength with depth for

London Clay generally, and that the Woolwich sample should be a subset of that prior database? If so, then the use of the grouped data (by depth) from that prior database to adjust the sample data represents a divergence from that hypothesis. This is because the grouped data do not provide a straight-line relationship between shear strength and depth. Would the authors not agree that the use of the regression line values directly at each grouped depth would lead to a more appropriate refinement of the sample? Moreover, the use of the coefficient of variation (or standard deviation) for each grouped value for shear strength loses the greater reliability obtained by using the coefficient of variation from the prior database as a whole.

The authors have adopted the use of Schneider's approximation given in Equation 3 of the paper to obtain the characteristic value for shear strength. However, it would appear to the discussor that this approximation provides a cautious estimate of the mean shear strength rather than a true characteristic value (or worst credible value). The true characteristic value, below which 5% of the population should lie, should formally be defined, for a normal distribution, as

$$7. \quad x_k = \mu - k_\infty \sigma$$

where x_k is the characteristic value, μ is the population mean value, σ is the population standard deviation and k_∞ is a constant that takes the value of 1.645.

The best estimate of the characteristic value, based on the sample, is

$$8. \quad x_k = m - k_s s$$

where m is the sample mean, s is the sample standard deviation and k_s is a constant that depends on the number of observed values in the sample, but will always be larger than 1.645.

This is the expression adopted by BS EN 1990 for the characteristic value. Rearranging Equation 8 in terms of the coefficient of variation, the best estimate for the characteristic value becomes

$$9. \quad x_k = m(1 - k_s V)$$

where V is the coefficient of variation of the sample.

As will become immediately obvious, calculation of the characteristic value using Equation 9 may lead to negative values or very low values of shear strength where the coefficient of variation is large and the characteristic is estimated on the basis of a small sample. It is for this reason that refining the sample on the basis of the prior database becomes desirable if unrealistically

low values of shear strength are not to result. The use of a log-normal distribution, where appropriate, avoids this difficulty.

There would appear to be a typographical error in the equation for the linear trend for the characteristic strength for the Woolwich sample, stated within the paper as $c_u = 65.7d + 51.2 \text{ kN/m}^2$.

If the linear trend line for the characteristic value for the shear strength of the Woolwich sample (shown in Figure 11) is superimposed upon the sample data, it can be seen that a very significant number of the observed values for the Woolwich site (approximately 16 out of 37) fall below the characteristic line. For a characteristic trend line, it might be expected that not more than 2 out of a sample of 37 (about 5%) should fall below this line. If the characteristic line is intended to provide the 'worst credible' value for shear strength when checking pile performance at the ultimate limit state, then it may be difficult to substantiate the reliability of the estimate provided by the authors' linear trend line, and hence the safety of the system.

Authors' reply

The authors would like to thank Mr Watkins for his comments and his interest in the paper. He suggests that the combination of a database of prior results with new values to refine characterisation of soil parameters may be extended to other materials, and that the approach is compatible with BS EN 1990 (BSI, 2002). The authors agree, and indeed it was intended that the paper would not only introduce this concept but also provide an example application of simple tools that could be easily utilised by the practising engineer to achieve this. In re-creating the analysis, Mr Watkins has made a number of observations, and has suggested extension of the analysis to incorporate other statistical distributions or to further refine the estimate of the mean and characteristic values.

It was observed that the sample information (and the prior database) contains some values that appear either high or low, and are potentially statistical outliers. In current practice such values are often dealt with subjectively; an engineer may choose to include or exclude such values based on experience, on statistical analyses, or for other reasons. The inclusion of all values in the database and in the analysis was intentional, to provide a wholly objective approach. There are valid reasons why outliers may occur, and some of these have been suggested by the discussor. The occurrence of such conditions may have an effect on pile behaviour, and therefore cannot be ignored. The database of prior information, if it is considered representative of the population, needs to contain such values. The inclusion of these values in the sample data is therefore equally important.

The discussor has suggested a number of methods of calculating the coefficient of variation, which differ slightly from that which has been employed. The value used has been calculated as the mean of the individual coefficients of variation calculated from

each depth band. The degree of de-trending that occurs in this process explains why the value given is closer to the value of those suggested by the discussor, based on the residuals from the regression line.

It is suggested that the assumption of normal distribution be tested. This has been conducted, and is reported by Baxter (2009). Where a log-normal distribution is encountered, then this can be used along with the appropriate analysis tools.

There is indeed an error in the stated equation for the linear trend for the characteristic strength for the Woolwich sample. This should have read $c_u = 6.57d + 51.2 \text{ kN/m}^2$

The discussor has questioned the selection of characteristic value. It is important to consider the engineering application of the processes described in the paper, and not just the statistics in isolation. The capacity of a pile in a cohesive material such as London Clay is dominated by the friction on the shaft of the pile. In essence this is proportional to the average of the undrained shear strength over the length of the pile. The correct definition of the characteristic value is therefore crucial to the process. It is the mean value of shear strength (represented by the regression line) that is of interest; the characteristic value becomes that value for which there is a 5% chance that the mean value falls below it. This is a subtle but important difference.

The authors concur that there are further analyses that may be conducted and tools that may be employed as part of the updating process, and these may produce a refined output. The authors' intention was to introduce simple tools that a practising engineer could adopt without the need for specialist additional knowledge. There is no reason why this cannot be used as a building block onto which other tools may be added. It is important, however, to retain an overview based on engineering judgement, and not to elevate the importance of the statistical manipulation too high. It must also be remembered that site investigations often provide fairly crude results with less than ideal frequency of data, and that high-powered statistics should not be used to hide the uncertainty associated with this.

REFERENCES

- Baxter DJ (2009) *Innovation in the Design of Continuous Flight Auger and Bored Displacement Piles*. EngD thesis, Loughborough University.
- Baxter DJ, Dixon N, Fleming PR and Cromwell K (2008) Refining shear strength characteristic value using experience. *Proceedings of the Institution of Civil Engineers – Geotechnical Engineering* **161(5)**: 247–257.
- BSI (British Standards Institution) (2002) BS EN 1990: 2002: Eurocode: Basis of Structural Design. BSI, Milton Keynes.