The importance of laboratory proficiency testing schemes in assessing and improving uncertainty

Chris F Wallace[1], *John J M* Powell[1#], and *John A* Reynolds[2]

[1]*Geolabs Limited, Bucknalls Lane, Garston, Watford, Hertfordshire, WD25 9XX, United Kingdom*
[2]*Geolabs Limited, Unit D3 HRS Business Park, Granby Avenue, Birmingham, B33 0SJ*
[#]*jpowell@geolabs.co.uk*

ABSTRACT

Any site characterisation relies on at least some laboratory tests, and some of those test results (often from basic tests such as Liquid Limit and Plastic Limit) are used in initial design calculations based on correlations. However, the reliability of those correlations is heavily dependent on the uncertainty in the results of those laboratory tests.
There is an inter-laboratory proficiency testing scheme that has been running for over fifteen years in the UK, with many worldwide participants. This paper presents a compilation of the scheme's findings to allow an assessment to be made regarding the reliability of different tests. As an example, repeatability within a single laboratory for the Liquid Limit test by cone penetrometer has been shown to be ± 1 %, but between different laboratories this has risen to ± 6 %. Similar ranges have been found in the Plastic Limits which, taken together with the uncertainties from the Liquid Limits, could give rise to significant concerns over using correlations based on the Plasticity Index.
Other examples of results from various test methods are given in this paper and it is argued that much of the uncertainty comes not from the test method itself, but from other factors including basic equipment maintenance, calibration, technician training and competence. It will be seen that laboratory proficiency testing schemes are crucial in highlighting these problems and giving an opportunity to allow better assessment of the quality of both test results and, arguably, the laboratories that produce them.

Keywords: proficiency; uncertainty; laboratory; testing.

1. Introduction

Geotechnical site characterisation will often involve laboratory testing to be scheduled to determine properties of the soil directly, but also indirectly in the form of correlations and designations based on the measured properties.

Whereas inter-laboratory proficiency testing schemes are common for chemical testing, they are less common for physical and mechanical soils testing, despite it being a requirement of ISO/IEC 17025 (2017) that accredited laboratories must ensure that reliable and adequate quality control procedures are in place for monitoring the validity of test results produced by the laboratory.

Without reference to other laboratories' results of tests performed on the same material, individual laboratories – irrespective of whether they are accredited or not – cannot verify that their results are comparable or correct (these are not necessarily the same thing, as will be seen in section 3.9).

To allow individual laboratories to gain a better understanding of the accuracy of their results, Geolabs Limited has been organising for over 15 years an inter-laboratory proficiency testing scheme. This involves sending out identically batched materials for each test to the participating laboratories and collating the returned results.

The tests examined in this paper are:

- Water Content
- Liquid and Plastic Limits (4-Point Cone)
- Particle Density (Gas Jar)
- Particle Density (Pycnometer)
- Particle Size Distribution (Sieve)
- Particle Size Distribution (Pipette)
- Particle Size Distribution (Hydrometer)
- 2.5 kg Compaction Test
- Direct Shear (60 mm small shear box)

Since absolute reference values of the parameters being tested for were not known in advance, for each test parameter being investigated an 'assigned value' was determined. This was calculated as being the mean value once outliers had been excluded. Outliers were considered to be values more than 3 standard deviations from the mean of all values, or where curves had an obviously different shape to the majority of results (this was more subjective but considered justifiable based on experience).

However, outliers are still a valid part of the data set and representative of results a geotechnical engineer might receive, so they are included in the calculation of uncertainty reported for each test. The uncertainty has been calculated based on 2 standard deviations of the

assigned value so the value quoted would apply to approximately 95 % of the results.

Possible reasons for variability within the results are discussed for each test, and there is an implications section showing how the uncertainties could affect some geotechnical parameters and correlations based on them.

2. Test Data

Data for these analyses were selected from various inter-laboratory testing schemes organised by Geolabs Limited between 2012 and 2023. The results were from a total of 62 laboratories, most of which were based in the UK but some from across the world.

The samples that were tested were prepared from materials that Geolabs had collected, thoroughly mixed, then sub-divided by quartering and riffling before being packaged and sent to the individual laboratories. Tests to ensure consistency of material across batches were also carried out at Geolabs.

There is insufficient space here to detail and reference all the standards specified, but more information on these can be requested from Geolabs.

3. Test Results

Unless otherwise specified, uncertainties expressed in percentages are percentage points from the assigned value, not a fraction of the assigned value.

Individual results are displayed in arbitrary laboratory number order.

For all tests, lapsed, drifted or inaccurate calibration of test equipment is a possible component of the uncertainty of the results, as are the weighing uncertainties listed in 3.1, so these are not itemised for subsequent tests in this paper.

3.1. Water Content

Although this was one of the simplest tests – the material only requiring drying to constant mass – there was still a wide range of results as shown in Figs. 1 and 2.

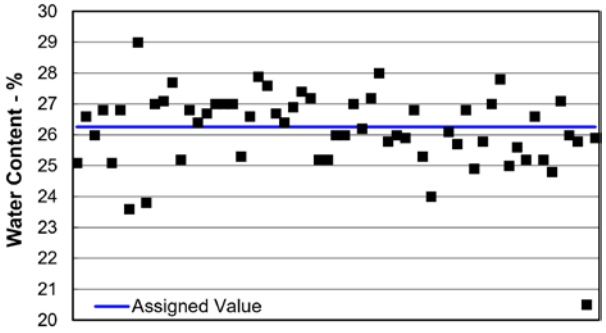

Figure 1. Water Content – Reported values

The range of the 63 reported values was 20.5 % to 29.0 % with an assigned value of 26.3 % and an uncertainty of ± 2.1 %. One very low result was excluded from the assigned value calculations.

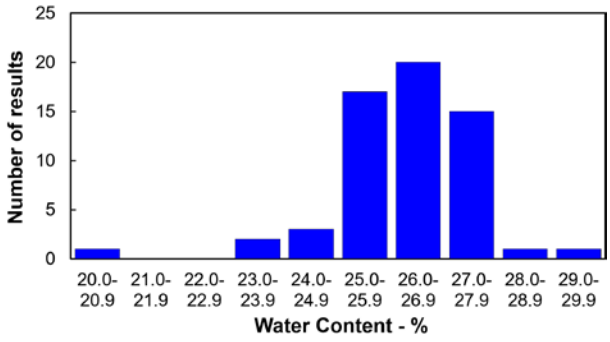

Figure 2. Water Content – Summary

The major components of the uncertainty could include:

- Rounding errors (the standard requested specified reporting to 1 decimal place, but 15 of the 61 results were whole numbers, which implies several of the results had likely been rounded).

Weighing uncertainties:

- Balance calibration.
- Balance drift.
- Incorrect drying temperature or duration.
- Absorption of water vapour during cooling.

3.2. Liquid & Plastic Limits

The four point Liquid Limit test involves mixing soil to different water contents to find what water content gives 20 mm penetration with a cone penetrometer having a 30° tip angle.

The Plastic Limit test involves finding the water content at which a thread of soil starts to crack when rolled to 3 mm in diameter.

The four penetration versus water content points for each Liquid Limit test, together with their best-fit trendlines, are shown in Fig. 3.

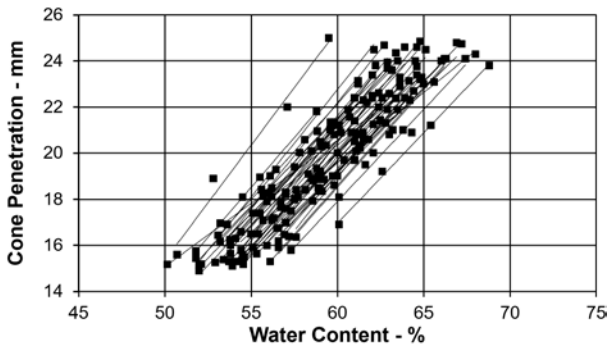

Figure 3. LL Cone Penetrations (4-pt cone)

The range of the 48 reported Liquid Limit values was 55 % to 64 % as shown in Fig. 4 with an assigned value of 59.6 % and an uncertainty of ± 3.4 %. No results were excluded from the assigned value calculations. The results from the previous year's proficiency testing scheme had a larger uncertainty of ± 6 %.

The major components of the uncertainty for the Liquid Limits could include:

- Air pockets incorporated into material in the cup.
- Dirty cone.
- Material not mixed to homogeneity.
- Blunt, scratched or rough cone.

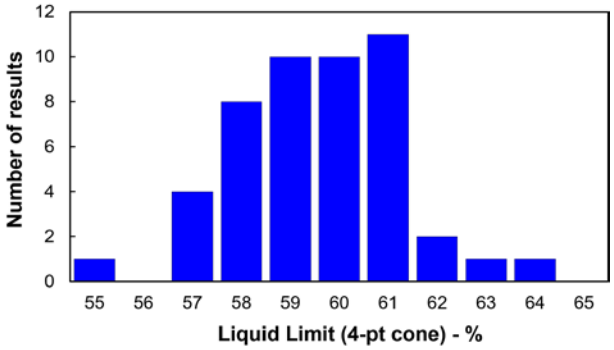

Figure 4. Liquid Limits (4-pt cone) – Summary

Although not discussed here, it should be noted that results from Liquid Limit tests performed using the 1-point cone method showed an even greater scatter than those performed using the 4-point cone method.

The range of the reported Plastic Limit values accompanying the 4-point cone Liquid Limits was 19 % to 26 % as shown in Fig. 5 with an assigned value of 22.2 % and an uncertainty of ± 3.8 %.

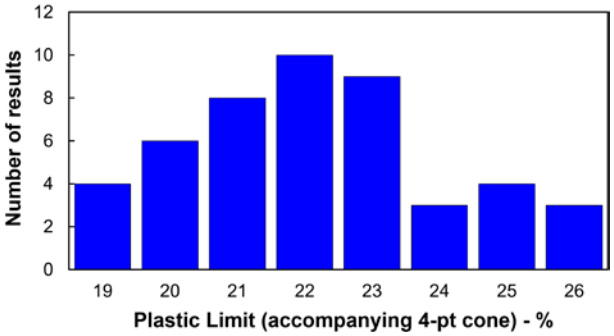

Figure 5. PL (accompanying 4-pt cone) – Summary

The calculated Plasticity Indexes from these tests had a range of 32 % to 43 % as shown in Fig. 6 with an assigned value of 37.4 % and an uncertainty of ± 5.0 %.

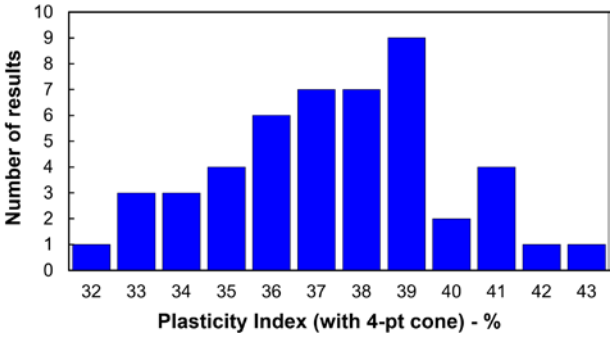

Figure 6. PI (with 4-pt cone) – Summary

The major components of the uncertainty for the Plastic Limits could include:

- Threads rolled to the wrong diameter.
- Poor rolling technique.
- Threads allowed to partially dry before weighing.
- Worn glass rolling plate.

3.3. Particle Density (Gas Jar)

This test involves weighing a gas jar with ground-glass lid empty, with water, with dry coarse particles, and with coarse particles and water.

The range of the 37 reported values was 2.58 Mg/m³ to 2.66 Mg/m³ as shown in Fig. 7 with an assigned value of 2.619 Mg/m³ and an uncertainty of ± 0.038 Mg/m³.

The major components of the uncertainty for the Particle Density by Gas Jar method could include:

- Trapped air between the particles.
- Results not corrected for the actual temperature the test was performed at.
- Water remaining on the outside of the gas jar.

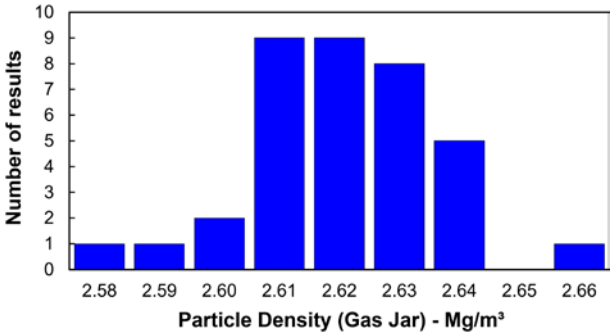

Figure 7. Particle Density (Gas Jar) – Summary

3.4. Particle Density (Pycnometer)

This test involved weighing a glass pycnometer with tapered ventilated stopper empty, with water, with dry fine particles, and with fine particles and water.

The range of the 36 reported values was 2.50 Mg/m³ to 2.73 Mg/m³ as shown in Fig. 8 with an assigned value of 2.656 Mg/m³ and uncertainty of ± 0.12 Mg/m³.

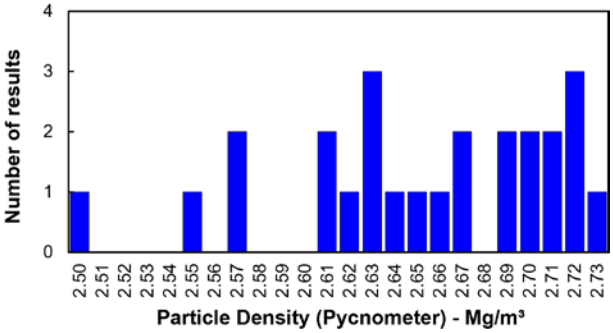

Figure 8. Particle Density (Pycnometer) - Summary

The major components of the uncertainty for the Particle Density by Pycnometer method could include:

- Trapped air between the particles.
- Material lost during stirring, possibly stuck to the stirrer.
- Results not corrected for the actual temperature the test was performed at.
- Water remaining on the outside of the pycnometer.
- Water not topped-up to completely fill the ventilation hole.

3.5. Particle Size Distribution (Sieve)

For this test the material is wet-sieved to remove the fines before the coarser material is dry-sieved. The percentage passing curves from the 56 tests are shown in Fig. 9. One outlier result was excluded from the assigned value calculations as it showed much less coarse material than the other results.

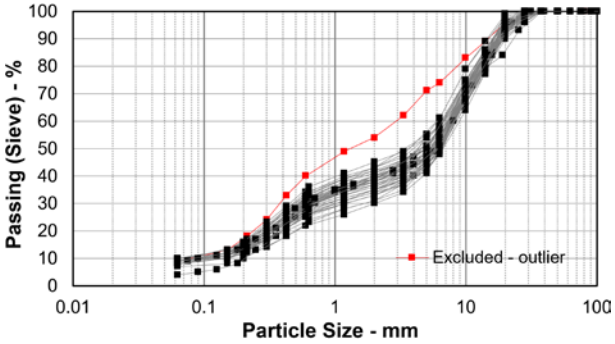

Figure 9. PSD (Sieve) - percentage passing curves

To aid comparison of the results, the percentage of fines (material <63 µm) and the percentage of gravel were assessed. The percentage of fines ranged from 4 % to 10 % as shown in Fig. 10 with an assigned value of 8.9 % and an uncertainty of ± 1.8 %.

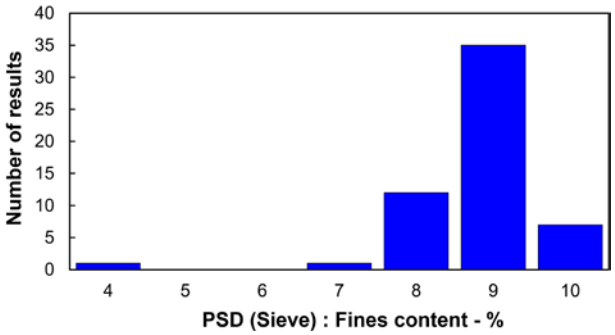

Figure 10. PSD (Sieve) - percentage fines summary

For a relatively simple test – washing the fines out and then passing the retained material through sieves – the percentage of gravel had a wide range from 46 % to 70 % as shown in Fig. 11 with an assigned value of 62.3 % and an uncertainty of ± 7.6 %. Considering the relatively tight grouping of the fines content percentages, this would imply that the variability in the gravel content was caused by factors affecting the dry sieving part of the test rather than the washing part of the test.

The major components of the uncertainty for the Particle Size Distribution by Sieving could include:

- Insufficient time sieving the material so finer material still remained on the sieve.
- Insufficient manual manipulation of larger particles to try to fit them through the sieve apertures so finer material remained on the sieve.
- Ineffective washing leading to finer material sticking to coarser particles.
- Worn sieves having larger apertures than stated
- Material lost through carelessness or particles becoming lodged in apertures.

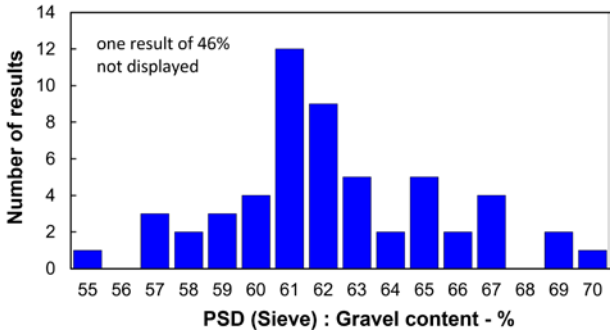

Figure 11. PSD (Sieve) - percentage gravel summary

3.6. Particle Size Distribution (Pipette)

For this sedimentation test a sample is added to water, agitated to form a suspension, then small sub-samples of the suspension are extracted at set times with a pipette as the suspension settles, then the solids content of each sub-sample is determined.

The percentage passing curves from the 25 tests are shown in Fig. 12. Two outlier results were excluded from the assigned value calculations as they showed much less clay content than the other results, and 3 results were excluded as they appeared to be missing a substantial proportion of the coarser material > 0.1 mm.

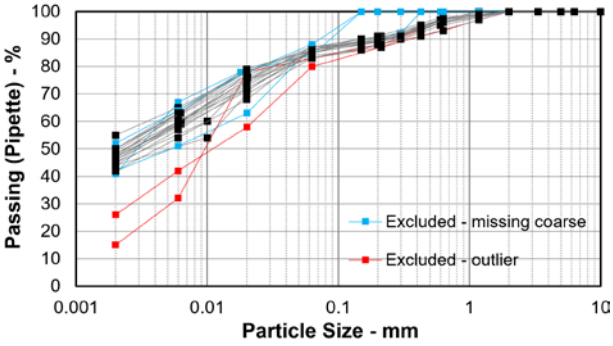

Figure 12. PSD (Pipette) - % passing curves

Considering this test is specifically used to measure the finer fractions of a soil, there was a comparatively wide range of clay fractions from 15% to 55 % as seen in Fig. 13 with an assigned value of 47.3 % and an uncertainty of ± 16.0 %, although this dropped to ± 5.5 % if the excluded values were removed.

The major components of the uncertainty for the Particle Size Distribution by pipette could include:

- Insufficient or ineffective deflocculant used leading to clumping of finer particles.
- Insufficient or ineffective end-over-end agitation leading to clumping of finer particles.
- Pipetted lowered or retracted too quickly leading to turbulence.
- Remnants of sampled liquid left in pipette due to insufficient rinsing.

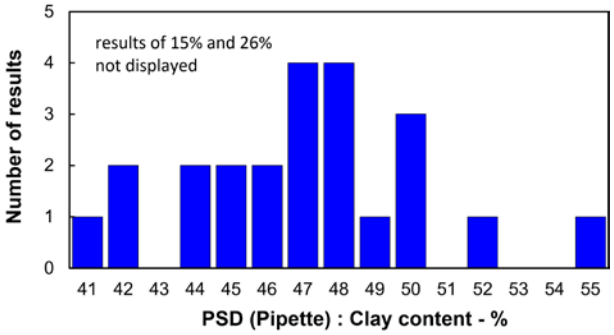

Figure 13. PSD (Pipette) - percentage clay summary

3.7. Particle Size Distribution (Hydrometer)

For this sedimentation test a sample is added to water, agitated to form a suspension, then its density measured using a hydrometer at set times as the suspension settles.

The percentage passing curves from the 31 tests are shown in Fig. 14. Two outlier results were excluded from the assigned value calculations as they showed much less and much greater clay content than the other results, and 5 results were excluded as they appeared to be missing a substantial proportion of the coarser material > 0.1 mm.

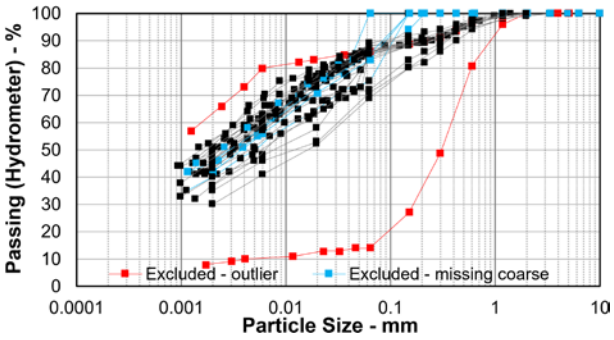

Figure 14. PSD (Hydrometer) - percentage passing curves

This test method for measuring the fine fractions of a soil had a larger range of clay fractions than for the pipette method: from 8% to 63 % as seen in Fig. 15 with an assigned value of 44.9 % and a higher uncertainty of ± 18.0 %, dropping only to ± 11.6 % once the excluded values were removed.

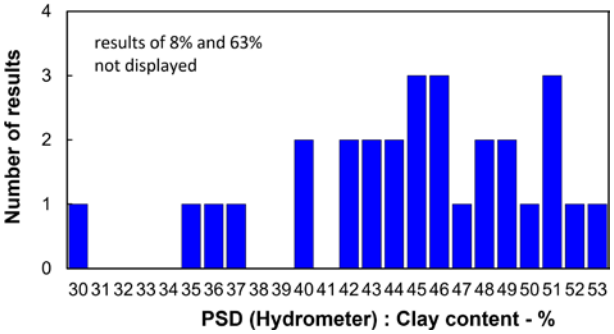

Figure 15. PSD (Hydrometer) - percentage clay summary

The major components of the uncertainty for the Particle Size Distribution by hydrometer could include:

- Insufficient or ineffective deflocculant used leading to clumping of finer particles.
- Insufficient or ineffective end-over-end agitation leading to clumping of finer particles.
- Hydrometer placed or removed too quickly leading to turbulence.
- Hydrometer left in suspension between readings allowing material to settle on its bulb.
- Misreading of hydrometer stem.

The pipette and hydrometer tests were both performed on the same material. Even though the assigned values of clay percentages were similar for both methods (47 % for the pipette and 45 % for the hydrometer), the higher uncertainty for the hydrometer and its much flatter spread of results as seen in Fig. 16 suggest the pipette is a far more reliable method than the hydrometer.

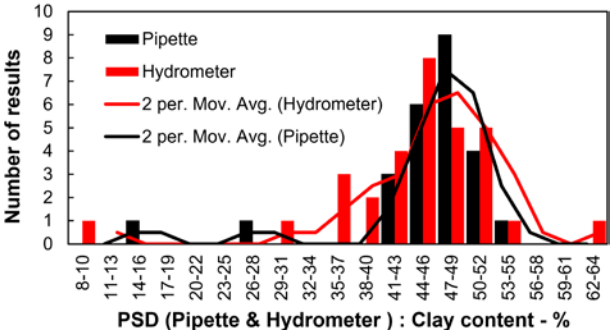

Figure 16. PSD - Pipette and Hydrometer compared

3.8. 2.5 kg Compaction Test (dry density/water content relationship)

For this test a sample is remoulded using a fixed compactive effort from a free-falling rammer within a guide at several different water contents and the resulting dry densities determined.

The data points from the 34 compaction curves (with straight lines drawn between them to show their relationships) together with points showing their reported Maximum Dry Densities (MDD) and Optimum Water Contents (OWC) are shown in Fig.17.

The range of the reported maximum dry density values was 1.36 Mg/m³ to 1.68 Mg/m³ as shown in Fig. 18 with an assigned value of 1.548 Mg/m³ and an uncertainty of ± 0.10 Mg/m³.

The range of the reported optimum water content values covered nearly the whole range of water contents tested: from 8.6 % to 33 % as shown in Fig. 19 with an assigned value of 23.3 % and an uncertainty of ± 8.0 %.

The major components of the uncertainty for a compaction test could include:

- Insufficient standing time after adding water to allow water content to become homogenous rather than clumps of clay being wetter on the outside and drier on the inside.
- Ineffective or insufficient mixing of added water to sample.
- Height of compacted material being below the top of the mould or more than 6 mm above the top of the mould.
- Material being dried-back at too high a temperature before testing.

- Sliding rammer clogged with trapped/adhering material.
- Inadequate control of the compactive effort and the compaction procedure (drop height, rammer mass etc.).

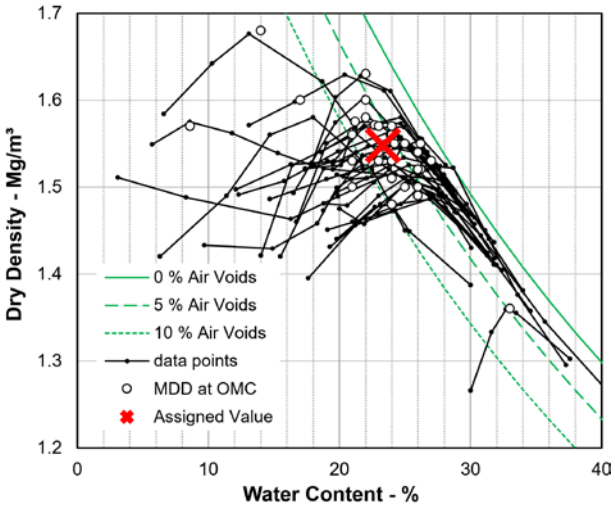

Figure 17. Compaction Test – DD/WC curves

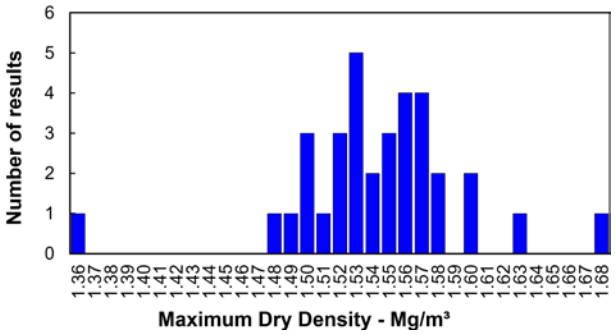

Figure 18. Compaction Test – MDD summary

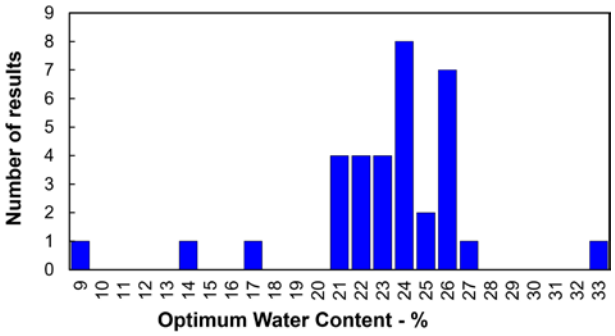

Figure 19. Compaction Test – OWC summary

3.9. Direct Shear (60 mm small shear box)

For this test specimens are remoulded into a horizontally split box, consolidated, then sheared by laterally displacing the top and bottom halves of the box. Best fit lines applied to three results having different normal stresses allow the angle of friction and apparent cohesion to be determined.

A laboratory would typically want their result to be in the centre cluster of results, as near as possible to the assigned value. At first glance, for the eight reported angle of friction values ranging from 17.5° to 23°, that would be around 20.3° as seen in Figs. 20 and 21.

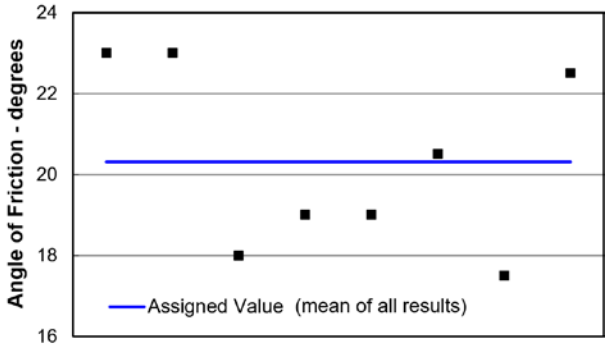

Figure 20. Angle of Friction (based on all results)

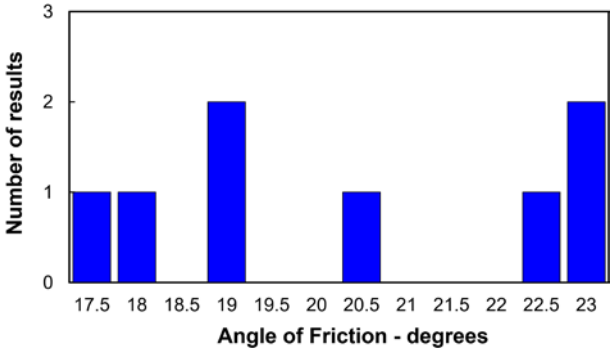

Figure 21. Angle of Friction - summary

The mean error for these values was ± 1.9°. However, on inspecting the rates of displacement used it became apparent that the highest angles of friction were obtained when shearing at the slowest rates (0.00069 mm/min to 0.00156 mm/min), whereas the lowest angles of friction came from tests sheared at much faster rates (0.023 mm/min to 0.077 mm/min), so between 14 and 111 times faster!

Although the direct shear test is a total stress test since pore pressure is not measured, it is intended to be sheared at a rate slow enough that drained conditions prevail. The shearing rate is calculated using the t_{90} value derived from the settlement versus square root of time plot of the consolidation stage. For the tests run at high speeds we suspected that the laboratories concerned used the initial steeper part of the consolidation curve (where the material of the remoulded specimens was being rearranged in essentially undrained conditions via the compression of air voids) rather than the slightly later shallower part of the curve where water had time to drain and true consolidation occurred. Deriving the t_{90} from the shallower portion would give a much larger time and a correspondingly slower shearing rate. This principle is shown with example settlement data in Fig. 22.

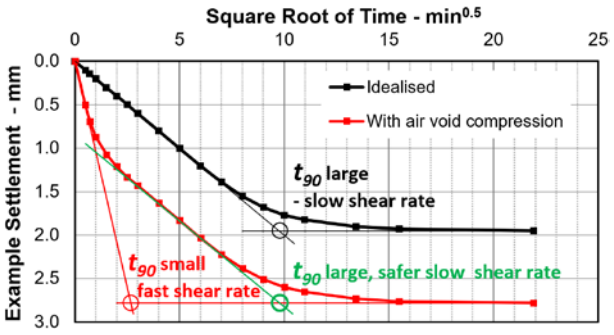

Figure 22. Effect of air void compression on t_{90}

In consequence, although the faster shearing rate tests had top and bottom drainage via the porous plates, it is likely that the excess pore pressures did not have sufficient time to dissipate from the clayey material, hence pore pressures likely rose during the shearing stage, so reducing the actual (but unmeasured) effective stress resulting in a corresponding loss in strength and a lowering of the angle of friction.

Taking these pore pressure effects into account, we considered only the tests that were sheared at the slowest rates to have been valid, with the other tests having had the systematic error of being sheared too quickly. Consequently, as shown in Fig.23, we considered the correct assigned value to have been 23° rather than 20.3°, which increased the mean error from 1.9° to 2.7°.

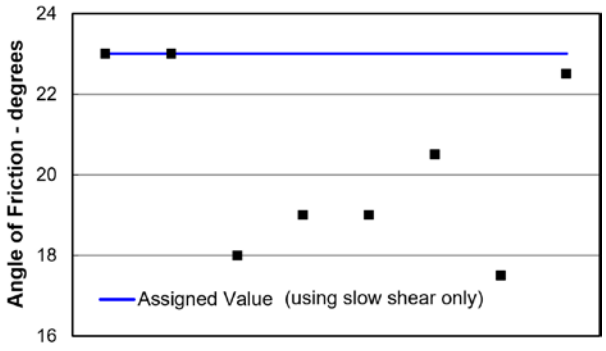

Figure 23. Angle of Friction (based on slow shear)

The major components of the uncertainty for the Direct Shear test could include:

- Shearing the test too quickly (a relatively rapid shearing rate calculated according to the specification for a known clay material should be treated with caution and instead a safer, slower rate considered to avoid unmeasured pore pressure changes being developed).
- Clogged filter plates preventing pore pressure dissipation.
- Not allowing consolidation to finish before shearing, so leading to higher pore pressures and lower shear strengths.
- Not raising the top half of the shear box a little before shearing; this can lead to increased friction and higher reported shear strengths.

4. Some Implications

4.1. Uncertainty and Range of Values

Although it may be tempting to allow for the quoted uncertainty in your geotechnical designs, there is the not insignificant 5 % chance that the result from **your** chosen laboratory is one of the outliers. Taking the percentage gravel from the PSD test as an example, if your laboratory's result was the 46 % gravel outlier, then instead of being within ± 7.3 % of the assigned value, it would have been 16.2 % away!

Only by ensuring your chosen laboratory takes part in a proficiency testing scheme – and performs **well** in that proficiency testing scheme – can you gain confidence in their results.

4.2. Plasticity Index and Soil Volume Change Potential

The ± 5 % uncertainty of the Plasticity Index (PI) could easily result in a misclassification of soil volume change potential as described in "Section 4.2 – Building near trees" (NHBC, 2024). A PI of 37 % would classify a soil as medium volume change potential, whereas the same soil with a PI of supposedly 42 % would be classified as having a high volume change potential. If a site is misclassified as high volume change potential when it is medium, foundations for, say, a housing development may be taken down further than necessary with a subsequent cost implication. Conversely, if a high volume change potential site is misclassified as medium, the foundation depth, where trees are present, may be inadequate with a potential structural problem arising in the future.

4.3. Particle Density and Degree of Saturation

The wide range of particle density values could cause problems to engineers analysing other tests since the particle density is often used with offshore projects to calculate the degree of saturation. Assuming a sample with a particle density of 2.65 Mg/m³ was 100 % saturated at a dry density of 1.80 Mg/m³, then using the range of particle densities in this paper its calculated degree of saturation would vary between 93 % and 118 %. A degree of saturation greater than 100 % would alert the user of this information that something was inconsistent, but they would not know if it was the dry density, the water content or the particle density that was amiss. However, a degree of saturation less than 100 % is certainly possible and might lead the user to wonder if the laboratory had not fully saturated the test specimen, or worry that an offshore sample had partially desiccated during transport or storage before being tested.

4.4. Water Content, Plasticity Index and Angle of Friction

Geotechnical engineers sometimes use correlations with parameters from lower cost, more readily available tests to predict other parameters that would otherwise be derived from higher cost tests.

An example of this is a formula from Göktepe et al (2008) which predicts the angle of friction (Φ) based on water content (ω) and Plasticity Index (PI) as shown in Eq. 1.

$$\Phi = -0.638 + 0.58\ \omega + 0.05\ PI \quad (1)$$

When using the assigned values of 26.3 % for the water content and 37.4 % for the PI the predicted angle of friction is 16.5°. However, allowing for an uncertainty of ± 2.1 % for the water content and ± 5.0 % for the PI gives the predicted angle of friction a range of 15° to 18°. While this is not a large range, it could have significant cost implications in embankment design.

Despite the relatively narrow range of predicted angles of friction, engineers should keep in mind that these are only correlations, albeit based on large numbers of results, and that individual results could be affected by macro features in the soil fabric - such as fissures, pockets and laminations - that could lower the angle.

4.5. Direct Shear - Rate of Strain and Angle of Friction

The results of the direct shear tests showed that the rate of strain used during the shearing stage can significantly change the measured angle of friction. As seen in this paper with a clay material, shearing the test too fast lowered the angle of friction, likely due to a rise in undrained pore pressure reducing the effective stress and hence the angle of friction. If this was for an actual project rather than a proficiency scheme, then the angle of friction would have been under-reported and led to a safe, but possibly overly expensive, geotechnical design.

However, if a silty or sandy material had been similarly sheared too quickly, then the material might have been trying to dilate around its peak shear strength, in which case the pore pressure could have significantly dropped with a corresponding increase in effective stress and peak shear strength, so leading to an over-reporting of the angle of friction. This could have led to an unsafe geotechnical design.

5. Conclusions

5.1. Laboratory Improvement

Since most of the laboratories that participated in the proficiency schemes run by Geolabs Limited were accredited laboratories, it is obvious that accreditation on its own is not sufficient to ensure good results. It is only by participating in a proficiency scheme that laboratories can identify if their results fall outside the expected norms and that they need to investigate what they are doing differently to other laboratories that needs rectifying or improving.

The components of uncertainty listed for each test would be a good starting point when looking for potential problems.

5.2. Choice of Laboratory

This paper has shown that for all the test types examined there was a wide range of results from the laboratories that participated. If you have an influence on which laboratory is used, we recommend that you ask to see that laboratory's proficiency test results: are they near the assigned value, giving you confidence that their results can be relied on, or are they one of the outliers? If they are an outlier, is there a good reason for it, such as seen with the slower shearing rate direct shear tests?

We have seen that choosing a laboratory that has been accredited by a recognized and respected body – such as the United Kingdom Accreditation Service (UKAS) - is not sufficient on its own to ensure good results. Both UKAS and the Association of Geotechnical Specialists (AGS) support proficiency schemes, UKAS even mandating participation if an appropriate scheme exists. A client needs to investigate and build up trust in their chosen laboratory to ensure that their laboratory understands the nuances of the test method and of the soil's behaviour, rather than the laboratory merely following test specifications without understanding their ramifications.

5.3. Choice of Test

Where there is more than one method to determine a soil property, consideration should be given to which method gives the lowest uncertainty in order to get the most reliable results.

We saw in this paper that the pipette method for particle size distribution by sedimentation gave a significantly lower uncertainty than the hydrometer method. Unless there is a good reason for specifying the hydrometer method, such as requiring more points on the PSD curve to allow making a finer discrimination of the sizes, such as might be wanted for tailings testing, then there is strong evidence that the pipette should be the preferred method.

Although not compared in this paper, there is also strong evidence from proficiency schemes and other sources that the cone penetrometer is a more reliable method than the Casagrande method for measuring the Liquid Limit. This was discussed at length in Powell et al (2015).

References

Göktepe, B., Altun S., Altintas, G., Tan, Ö. 2008. "Shear strength estimation of plastic clays with statistical and neural approaches", Building and Environment, 43(5), pp. 849-860, https://doi.org/10.1016/j.buildenv.2007.01.022

International Organization for Standardization "ISO/IEC 17025:2017 General requirements for the competence of testing and calibration laboratories",,ISO, 2017, https://www.iso.org/standard/66912.html

NHBC (National House-Building Council) "NHBC Standards 2024", [online] Available at: https://nhbc-standards.co.uk/4-foundations/4-2-building-near-trees/, accessed: 25/04/2024.

Powell, J. J. M., Shields, C. H., Wallace, C. F. 2015. "Liquid Limit testing - only use the Cone Penetrometer!", XVI European Conference on Soil Mechanics and Geotechnical Engineering, Edinburgh, United Kingdom, 2015, Vol 6, pp. 3305-3310.