

Assessment of Sample Disturbance in Mine Tailings

Jorge Macedo¹, Musaddik Hossain¹, Luis Vergaray¹, Thaleia Travararou¹

Abstract

Obtaining high-quality samples from tailings storage facilities (TSFs) remains a key challenge in tailings engineering. Although samplers such as Dames & Moore (DM) and Gel Push have been evaluated for natural soils, their performance in mine tailings remains less documented. This study evaluates the quality of tailings samples recovered using DM, Gel Push, and Shelby samplers at three TSFs, with direct DM–Gel Push comparisons available at one site. The evaluation combines CPTu data, field shear-wave velocity measurements, laboratory index testing, CRS oedometer tests, bender element tests, controlled disturbance tests on reconstituted specimens, and a qualitative scoring scheme based on tube-level condition, extruded specimen condition, X-ray specimen condition, and prepared specimen condition. Commonly used strain-based consolidation criteria, including reconsolidation volumetric strain and normalized void-ratio change, did not separate the imposed disturbance levels in the tested tailings and showed limited agreement with qualitative scores from retrieved samples. In contrast, recompression-based metrics from consolidation tests provided clearer trends and were more consistent with qualitative scores. Thresholds from previous studies did not transfer directly to the tested tailings, indicating that material-specific properties can influence sample-quality criteria. Laboratory-to-field $V_{s,1}$ ratios provided an independent assessment that was generally consistent with the consolidation-based and qualitative scores, although material variability within a sample could affect interpretation. For the materials tested, DM and Gel Push samplers generally recovered higher-quality specimens than Shelby tubes, with the DM sampler showing slightly better performance than Gel Push in the available comparisons, although the dataset for such comparisons is limited. Overall, the results support using multiple complementary indicators to evaluate sample quality in mine tailings and emphasize that criteria developed for natural soils should be applied carefully and only as reference values.

Keywords:

Mine tailings, field sampling, sample quality

1. Introduction

Tailings sampling is routine in investigations of tailings storage facilities (TSFs), yet studies assessing the quality of recovered samples remain scarce. Most prior evaluations of sampling quality have focused on natural soils (e.g., Umehara et al. (2015); Da Fonseca et al. (2019); DeGroot et al. (2020); Stringer et al. (2015); Landon et al. (2007); Lim et al. (2019)), whereas sample-quality evaluations in mine tailings remain limited (e.g., Jamiolkowski and Masella (2015)). Among previous efforts in natural soils, Umehara et al. (2015) compared conventional thin-wall tube (TB) and Gel-Push–Rotary (GP-Tr) sampling and reported that GP-Tr outperformed tube sampling in non-plastic soils with moderate fines content (FC = 12–24%). Da Fonseca et al. (2019) evaluated Mazier, Gel-Push–Static (GP-S), and Dames & Moore (DM) samplers and found that GP-S and DM generally recovered higher-quality samples than the Mazier sampler in clean sands, silty sands, and sandy silts. DeGroot et al. (2020) investigated low-plasticity silty soils in Halden, Norway, and showed that specimens recovered using the GP-S sampler exhibited behavior indicative of lower disturbance than those obtained with the geometrically poor NGI-54 sampler, which was intentionally included as a reference for poor-quality sampling. In the same study, the Gregory undisturbed sampler (GUS) produced one-dimensional compression curves similar to those obtained from Sherbrooke block samples, suggesting high sample quality. The DM sampler yielded more compressible consolidation responses relative to the block and GUS samples, suggesting somewhat greater disturbance, but the assessments were only qualitative. Other studies comparing block, fixed-piston, and free-piston samplers have shown that fixed-piston systems generally recover higher-quality soft-soil samples than Shelby or free-piston tubes (Landon et al. (2007); Lim et al. (2019)).

GP and DM samplers have been used in Christchurch to recover natural sands, silty sands, and silty soils for liquefaction-related laboratory testing following the Canterbury earthquake sequence (Stringer et al. (2015); Taylor et al. (2013a); Taylor et al. (2013b)). Across this body of work, GP-S, GP-Tr, and DM systems were considered. GP-S recovered high-quality silty sand and silty soil samples at selected sites, whereas GP-Tr trials in clean

sands indicated significant disturbance during sampling and/or transportation. For silty clay materials, GP-S and DM samples yielded similar cyclic resistance, suggesting comparable sample quality (Stringer et al. (2015)).

In contrast to the extensive body of work on natural soils, evaluations of sampler performance for mine tailings remain limited. One notable example is the investigations at the Źelazny tailings deposit, where Gel-Push-Rotary (GP-Tr) sampling produced shear-wave velocity ratios between 80–100% of field values, suggesting good sample quality in silty and silty-sand tailings (Jamiolkowski and Masella (2015)). The gap in limited efforts on mine tailings is significant given the distinct mechanical characteristics of tailings, which often exhibit different scaling of mechanical properties such as strength and dilatancy relative to natural soils (Macedo and Vergaray (2022)).

Among quantitative approaches, sampler performance has commonly been assessed using shear-wave velocity-based indicators and consolidation-based sample-quality metrics. With respect to shear-wave velocity (V_s), it is common practice to normalize V_s to an equivalent reference stress to account for overburden effects. The stress-corrected parameter $V_{s,1}$ is typically defined as $V_{s,1} = V_s (P_a/\sigma'_{v0})^n$ (Robertson et al. (1992); Andrus and Stokoe (2000)), where the exponent is material dependent. Ratios of field to laboratory $V_{s,1}$ have been used to infer the degree of disturbance (Jamiolkowski (2012); Fioravante et al. (2012); Sasitharan et al. (1994)).

Consolidation-based disturbance indices originated with Terzaghi et al. (1996), who proposed that greater volumetric strain upon reconsolidation to in-situ stress indicates greater disturbance. This approach is most appropriate for high-plasticity soils (Hight and Leroueil (2003)). Lunne et al. (1997) later proposed the normalized void ratio change, $\Delta e/e_0$, recognizing that identical volumetric strains may correspond to different disturbance levels depending on initial void ratio. DeJong et al. (2018) highlighted limitations of this criterion for non-clay materials, noting that $\Delta e/e_0$ may decrease in intermediate low-plasticity soils due to differences in compressibility, shallow stress states, or low overconsolidation ratio, without necessarily indicating good sample quality. Recognizing these limitations, DeJong et al. (2018) proposed evaluating disturbance using a work-based metric derived

Table 1: Sample quality evaluation methods typically used in practice

Methods	Indicator	Limitations
<u>Vs based</u>		
V_{s1} based (e.g., Jamiolkowski (2012))	Ratio of field and laboratory estimated V_{s1}	• Potential sampling densification effects not accounted for. • Stress dependence may be site-specific.
<u>Consolidation based</u>		
Terzaghi et al. (1996)	Volumetric strain upon reconsolidation to field stress.	• Highly susceptible to material-specific compressibility and stress effects in low/moderate plasticity soils (DeJong et al. (2018)).
Lunne et al. (1997)	Void ratio change upon reconsolidation to field stress normalized by initial void ratio.	• Highly susceptible to compressibility and stress effects in low/moderate plasticity soils (DeJong et al. (2018)).
DeJong et al. (2018)	Ratio of work-based metrics during recompression and virgin loading	• Reduced performance in low-OCR and low-plasticity soils • Uncertain generalizability of proposed sample quality thresholds (e.g., to mine tailings)

from consolidation tests. Similarly, [Karlsrud and Hernandez-Martinez \(2013\)](#) introduced slope-based indicators obtained directly from consolidation curves. These approaches can be interpreted within a work-based framework ([DeJong et al. \(2018\)](#)). In general, although sampling quality thresholds exist (e.g., [DeJong et al. \(2018\)](#), [Lunne et al. \(1997\)](#)), their applicability to mine tailings remains uncertain. Table 1 summarizes evaluation methods commonly used in practice and their limitations. .

In this study, we evaluate the quality of mine-tailings samples recovered using Gel-Push, DM, and Shelby samplers at three TSFs. Gel-Push was used at all three sites, whereas DM and Shelby samplers were used at selected sites for comparison. We first characterize the sites using cone penetration testing with pore pressure measurements (CPTu) and

field shear-wave velocity measurements, and then describe the sampling systems, field procedures, sample handling, and laboratory preparation protocols. We next discuss different methods for interpreting sampling disturbance from consolidation tests and evaluate them using reconstituted tailings specimens subjected to controlled disturbance levels. We then apply selected metrics to the recovered tailings samples and evaluate the resulting sampler-performance trends against an independent qualitative rating based on visual observations, X-ray imaging, tube condition, and specimen preparation quality. Separately, we also assess sample quality using field and laboratory shear-wave velocity measurements. Finally, we discuss the implications of these findings for selecting and evaluating sampling methods in mine tailings.

2. Geotechnical Conditions of Investigated TSFs and Sampling

2.1. Geotechnical conditions

The geotechnical conditions at the investigated TSFs were defined through CPTu, complemented by measurements of shear (V_s) and body (V_p) wave velocities, and boreholes for sample recovery.

Figure 1 presents the CPTu responses near the sampling location at TSF-1. The data are shown below 12.0 m, which corresponds to the depth below which tailings were deposited. Tip resistance (q_t) ranges from 1-10 Mpa, sleeve friction (f_s) varies from 50–200 kPa, except between 14-20 m and below 40m where it is lower than 50 kPa. Pore pressure (u_2) is generally lower than 300 kPa up to 40 kPa, and at deeper locations it can reach 1500 kPa. The soil behavior index defined by Robertson (2016), I_c , varies from 2.8 to 3.6, indicating soil behaviors from transitional to clay-like. V_s ranges from 230–400 m/s, and V_p ranges from 650-1000 m/s, suggesting partial saturation. Friction ratio (F_r) generally ranges from 1.5 to 4.0.% with a trend to decrease below 35-40 m, suggesting a transition from high compressible slimes to less compressible old tailings. Q_{tn} generally falls between 2 and 30, placing the material in the contractive zone of the Robertson (2016) chart. The normalized pore pressure, B_q , is 0.15 between 14 and 20 m, suggesting a local increase in saturation, and 0.4 below 40 m depth.

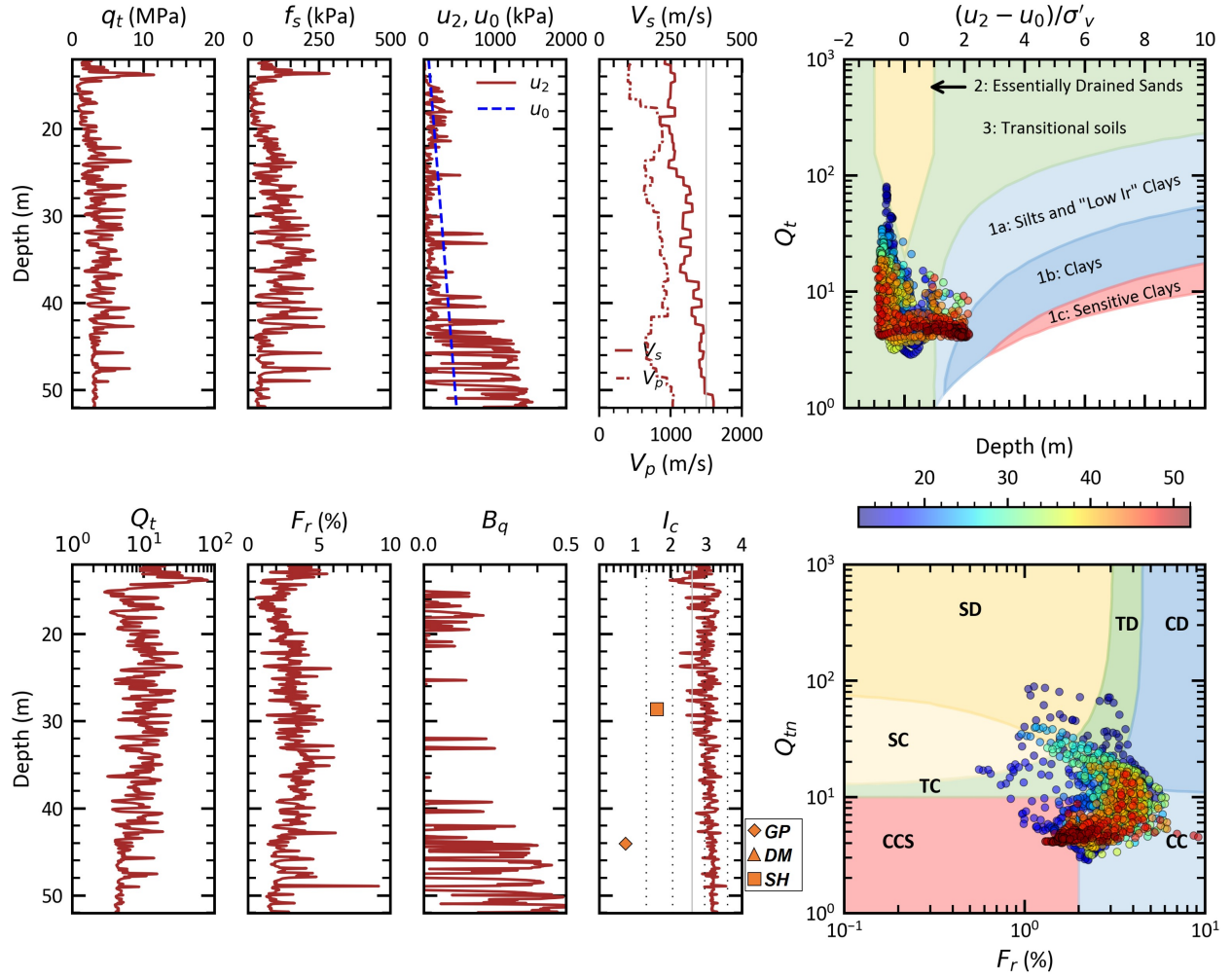


Figure 1: CPT parameters and soil behavior type (Robertson-2016) for site 1, the markers on the I_c chart indicate sampling depth and sampler used.

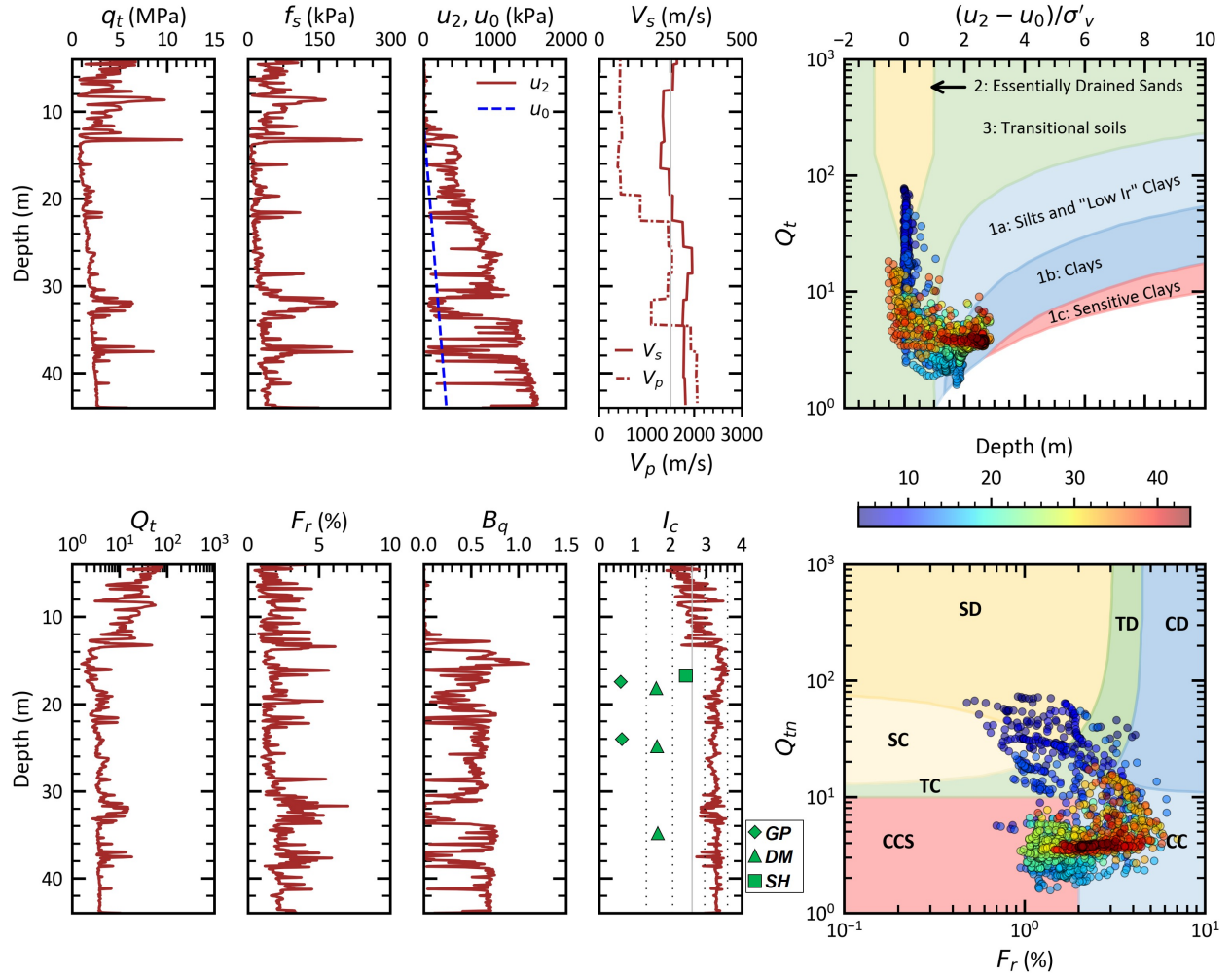


Figure 2: CPT parameters and soil behavior type (Robertson-2016) for site 2, the markers on the I_c chart indicate sampling depth and sampler used.

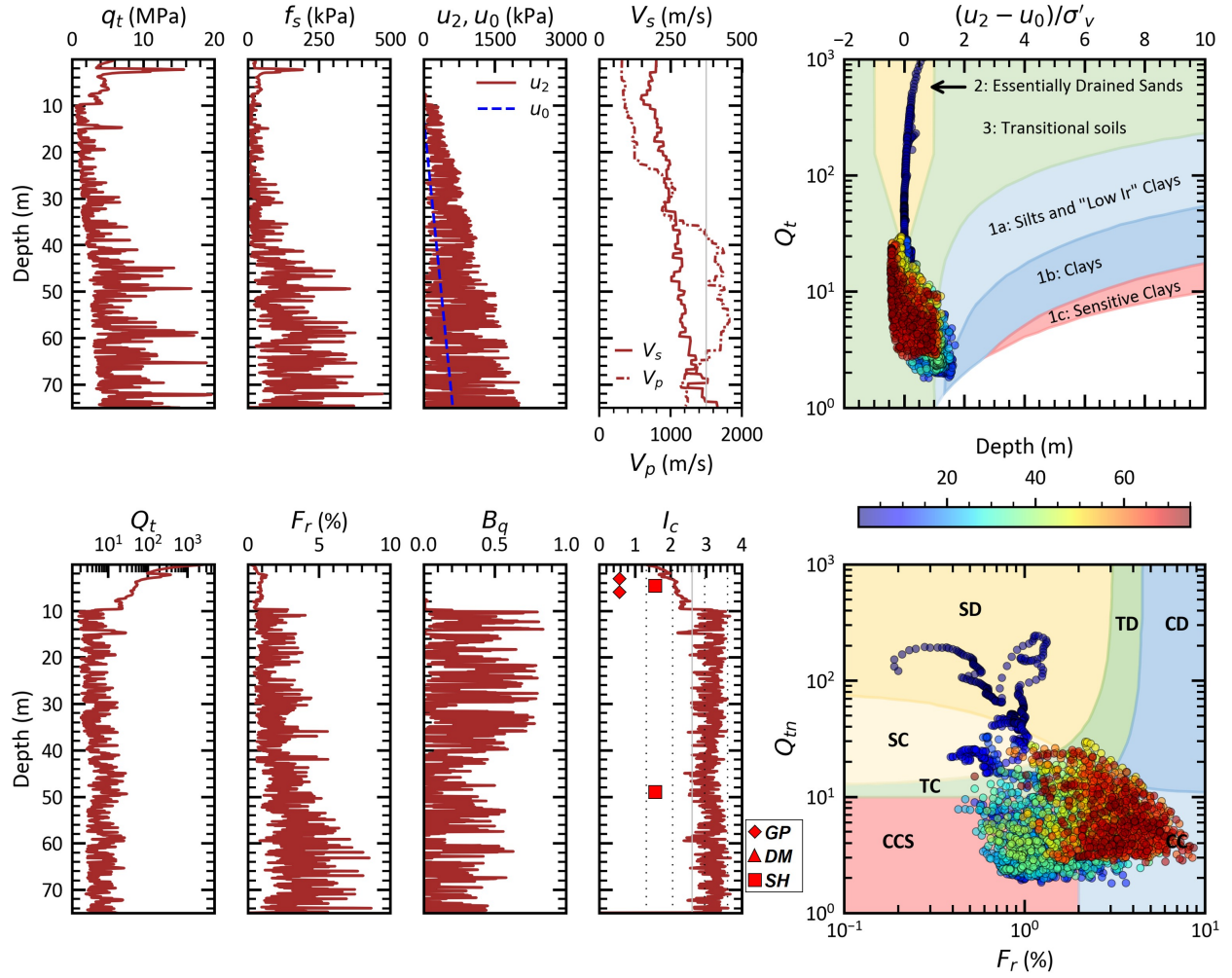


Figure 3: CPT parameters and soil behavior type (Robertson-2016) for site 3, the markers on the I_c chart indicate sampling depth and sampler used.

Figure 2 shows the CPTu responses near the sampling location at TSF-2. q_t is higher in the first 14m (2 to 8MPa) corresponding to underflow tailings, and lower (generally <3 Mpa) at deeper locations, consistent with a transition to slimes. This is also consistent with the lower I_c (2.0 to 3.0) in the first 14m compared to the larger values (3.0-3.2) at deeper locations. f_s also shows higher values (30-100 kPa) in the first 14 m compared to the deeper layers, where f_s is generally below 50 kPa except for a localized layer at 33m. u_2 starts increasing at about 12.0 m, where the water table is located. Q_{tn} is generally between 10 and 100 on the first 14 m and then below 10 for deeper locations, indicating a transition to a more contractive response with increasing depth in the Robertson (2016) CPTu chart. F_r is generally between 1% and 4% with significant variations at some depths, indicating compressible materials with layering. B_q is generally between 0.3 and 0.6 below the water table, consistent with fine-grained materials. There are some layers (e.g., 33 m, 38m) where B_q drops to zero, suggesting interlayering with coarse layers. V_s ranges from 200–300 m/s, and V_p is around 450 m/s up to 20 meters, suggesting partial saturation. At deeper depths, V_p ranges from 1450- 2050 m/s, indicating full saturation.

Figure 3 shows the CPTu responses near the sampling location at TSF-3. The first 10 m corresponds to sandy tailings with q_t of 4MPa, f_s of 40 kPa, null u_2 , an I_c between 2.0 and 2.6, F_r of about 0.5, and null B_q . The CPTu indicates a transition to intermediate and clayey materials below 10 m (I_c between 2.8 and 3.2) with q_t less than 3 MPa to 40m and q_t between 3 and 10 MPa at deeper depths. f_s also increases below 40 m from values that are less than 40 kPa to values up to 200 kPa at deeper parts in the profile. Below 10 m, Q_{tn} is lower than 10 and F_r varies between 1% and 6% with larger values and fluctuations at the deeper part of the profile below 40m. B_q is generally between 0.3 and 0.5, with higher values from 10 to 40m. This information collectively suggests a clayey contractive response with the presence of sensitive materials between 10 and 40 m, and a predominant clayey contractive response at deeper depths. The V_s shows a general increase with depth, ranging from 150 m/s to 400 m/s. V_p increasing from 300-800 m/s up to 25 m depth, suggesting partial saturation. Below 30 m, V_p is generally higher than 1500 m/s, indicating full saturation.

2.2. Sampling

Sampling depths are superimposed on the I_c profiles in Figures 1, 2, and 3 for the three TSF sites. For the Gel-Push system, most samples were recovered using the GP-S sampler, except for the second-from-the-top sample at TSF3, which was collected using a GP-Tr sampler. TSF2 considered sampling with all sampler types. At TSF1 and TSF2, sampling targeted predominantly fine-grained slimes, whereas at TSF3, sampling focused mainly on coarse-grained tailings. Tables A1, A2, and figure A1 in Appendix A summarize the main features of the considered samplers; a brief summary with relevant references is provided next.

The DM hydraulic fixed-piston sampler (Figure A1(a)) follows the Osterberg-type hydraulically activated fixed piston principle. The device incorporates several refinements, including a 450 mm-long thin-walled smooth brass liner with a constant inside diameter, a 60° external cutting edge angle, and a low area ratio. A neoprene skirt seal at the transition between the pressure cylinder and liner enhances sealing. The smooth brass liner reduces soil–wall interface friction during penetration, while the skirt seal prevents entry of disturbed material from the borehole base and helps maintain suction during retrieval. Sampling proceeds in three stages: borehole preparation, hydraulic advancement of the liner at controlled pressure to full penetration (450 mm), and extraction after a short dwell period. Bray and Sancio (2006), Viana da Fonseca and Pineda (2017), and Viana da Fonseca et al. (2019) provide more details on the DM sampler and its operation.

The GP-S system (Figure A1 (b)) uses hydraulic pressure to advance the sampling system and incorporates three pistons: a stationary piston, a piston that advances the sampling tube, and a piston that activates the core catcher. During sampling, the stationary piston remains at the borehole base, while the advancing piston drives the shoe, sampling tube, and liner downward. A key modification relative to conventional Osterberg samplers is the hydraulically activated core catcher, which retains the specimen during withdrawal and dispenses a viscous polymer solution to coat the sample surface. The polymer is driven downward by the advancing piston, flows through the annulus between the sampling and liner tubes, and coats the specimen through the core-catcher fins, while excess polymer is

vented through a pressure control valve. Upon reaching full penetration, hydraulic flow is redirected to activate the core catcher, which extends its fins to constrict and secure the sample. [Mori and Sakai \(2016\)](#) provides additional details on the GP-S sampler. The GP-Tr sampler (Figure A1(c)) is a rotary triple-tube system that employs polymer-assisted lubrication. The outer barrel rotates to advance the sampler, while a non-rotating inner liner captures the specimen. A free piston, initially seated near the cutting shoe within the liner, is displaced upward as soil enters the tube, driving a viscous polymer into the annulus between the liner and the inner barrel. The polymer is discharged near the cutting shoe to coat the specimen and reduce interface friction, while excess polymer is vented through a pressure-control valve to prevent polymer infiltration into the soil. The sampler is advanced under controlled rotation and penetration rate, and the specimen is retained within the liner by the combined action of the cutting shoe geometry and free-piston system during retrieval. Additional details are provided by [Mori and Sakai \(2016\)](#). Sampling with Shelby tubes was conducted in accordance with [ASTM D1587/D1587M-15](#) using a standard Shelby tube.

Following sampling, the sample tubes were sealed, placed upright in a specially designed box, and then carefully transported by car from the testing sites to Georgia Tech laboratories, where a majority of the soils were tested. The samples handling and transportation followed the recommendations by [Stringer et al. \(2020\)](#) and [Bray et al. \(2016\)](#).

2.3. Handling of Samples in the Laboratory

Upon arrival at the laboratory, sample tubes were inventoried and extruded in accordance with the procedures recommended by [Bray et al. \(2016\)](#) for DM and [Stringer et al. \(2015\)](#), [Stringer et al. \(2020\)](#) for Gel-Push (GP) samplers. Tubes were unsealed and mounted in the extrusion frame while preserving their original vertical orientation. The extrusion system (Figure 4a, 4b, 4c) followed the recommendations by [Stringer et al. \(2015\)](#). The assembly was rigidly wall-mounted and comprised a hydraulically driven piston, a rigid top plate, intermediate support plates, and support collars to ensure vertical alignment and uniform load transfer. A guiding collar was installed at the tube head to facilitate a clean horizontal cut using a sharp knife or wire. GP samples were extruded at full, sequentially obtaining the

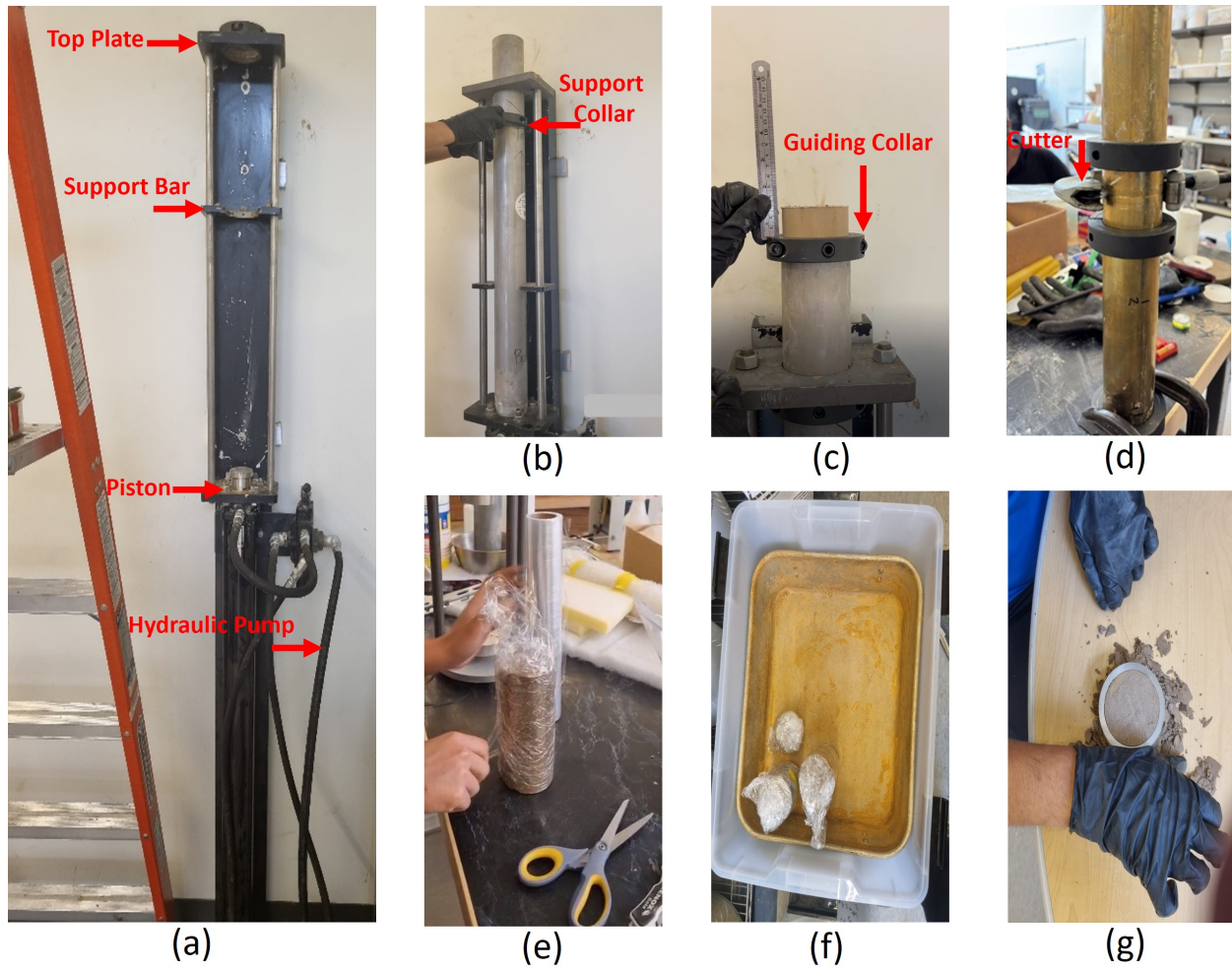


Figure 4: Laboratory handling of samplers- (a) Components of the extruder assembly, (b) Support collar placed under top plate to resist extrusion force and ensure alignment, (c) Guiding collar on top of the sampler to facilitate horizontal cutting of sample, (d) Cutting of a DM tube before extrusion (e) Wrapping of sample following extrusion, (f) Storage of sample in a moist box, (g) Sample preparation for testing

desired specimen height, as described in [Stringer et al. \(2015\)](#). The DM tubes were sectioned into shorter lengths before extrusion to reduce wall friction following [Bray et al. \(2016\)](#), (Figure 4d). The DM tube was clamped to a rigid, vibration-free surface using collars, and the cutting location was supported on both sides to limit shear-induced deformation before cutting with a sharp metal cutter. After this process, samples were extruded.

For all samplers, material near the top and bottom of each tube was discarded due to anticipated disturbance from sealing and extrusion, and was used primarily for index testing. After extrusion, specimens were either sealed with cling film to minimize moisture loss (Fig. 4e) and stored on a raised platen within a humidified container (Fig. 4f), or tested immediately (Fig. 4g). These handling procedures align with the recommendations of [Stringer et al. \(2015\)](#) and [Stringer et al. \(2020\)](#).

3. Testing Program

The testing program comprised grain-size analysis, Atterberg limit tests, moisture content measurements, constant rate of strain (CRS) oedometer tests, and bender element measurements on recovered samples. In addition, CRS tests were performed on reconstituted specimens targeting different levels of disturbance. [Appendix B- Table B1](#) summarizes the testing program.

Figure 5 summarizes key properties, including grain-size distribution, fines content, plasticity index (PI), and initial void ratio, across sites and sampling methods. As noted earlier, sampling at TSF-1 and TSF-2 targeted finer soils than at TSF-3, as reflected in the grain-size distributions, which show coarser materials at TSF-3. The fines content and PI data are consistent with this trend. The samples from TSF-3 collected from the first 10 m were non-plastic and are plotted as $PI = 0$. Excluding these, PI values are generally consistent across sites and samplers, ranging between 7 and 15. One exception is a DM sample from TSF-2. The top half has a PI of 15, similar to other materials retrieved from the site, while the bottom half has higher fines content ($FC = 100\%$) and a PI of 25.

Initial void ratios for most samples range from 0.70 to 0.90. A notable exception is a GP-Tr sample from TSF-3, which exhibits a higher void ratio. Prior to presenting the

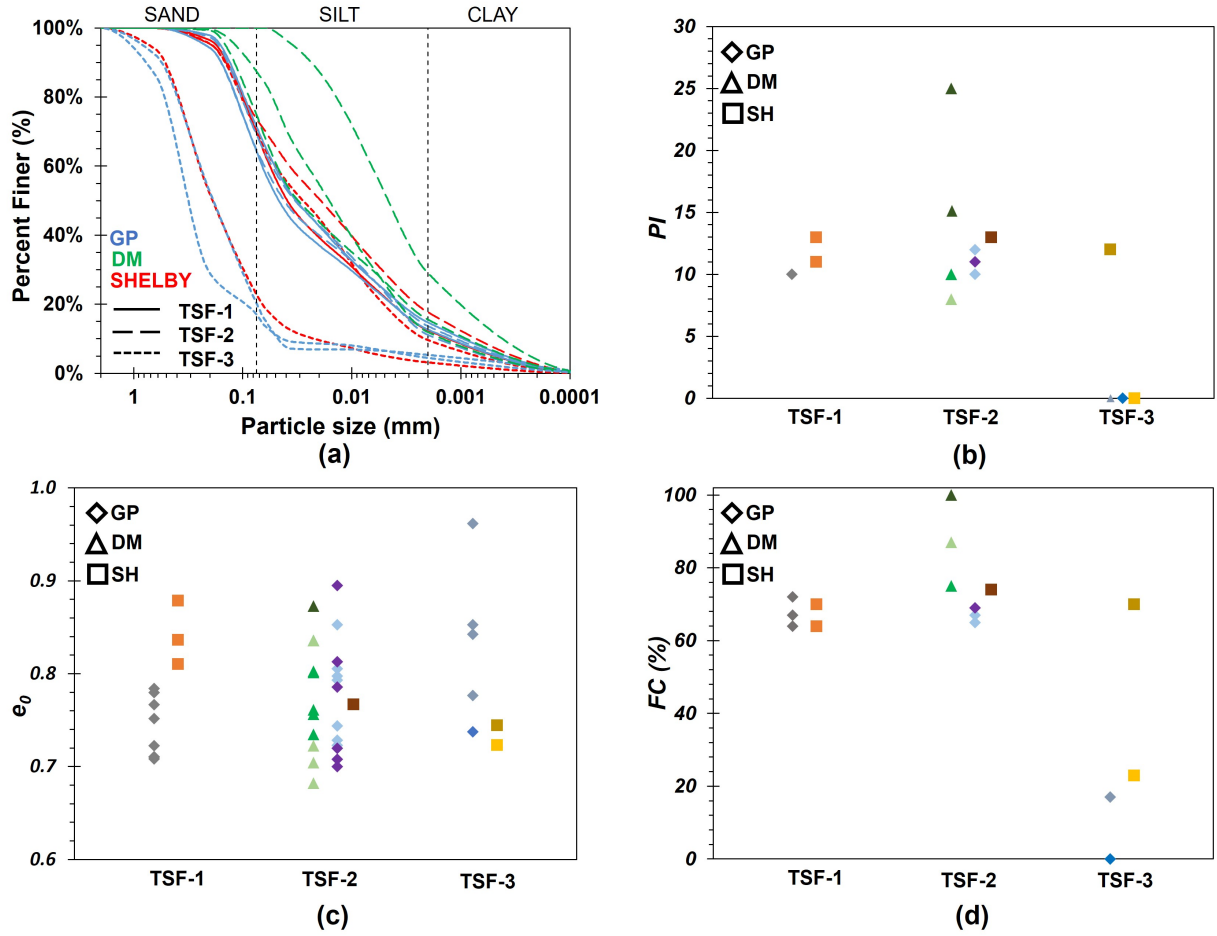


Figure 5: Tested material properties of different sites (TSF-1, TSF-2, and TSF-3) (a) Grain size analysis (b) Plasticity index (c) Initial void ratio (d) Fines Content. Markers represent sampler tube type, and Similar color represents similar sampler tube

consolidation and bender element results for recovered and reconstituted samples, the basis for disturbance assessment is discussed in the next section.

4. Methods for Evaluating Sample Disturbance

Sampling disturbance was evaluated using multiple complementary metrics. The intent was not to assume that any single metric provides an absolute measure of sample quality in mine tailings, because most existing criteria were developed for natural clays or low-plasticity silty soils. Instead, the metrics were used comparatively to assess whether they provide consistent trends for: (1) laboratory specimens subjected to controlled disturbance levels,

and (2) field samples recovered using different samplers. The metrics considered include reconsolidation strain indices (Terzaghi et al., 1996; Lunne et al., 1997), compression-index ratios (DeJong et al., 2018), work-based indices (DeJong et al., 2018), oedometer-based modulus indicators (Karlsrud and Hernandez-Martinez, 2013), void-index-based indicators (Cai et al., 2026), and exploratory swelling recompression metrics. Recommended disturbance thresholds for the considered metrics were also used, but only as reference values because they were not developed for tailings. The following subsections define the consolidation-based metrics; the basis for the shear-wave velocity metrics was presented in the Introduction.

4.1. Reconsolidation strain and normalized void-ratio change

Reconsolidation-based metrics were computed from CRS oedometer tests by evaluating the deformation required to reload each specimen to the estimated in-situ vertical effective stress, σ'_{v0} . The reconsolidation volumetric strain, $\epsilon_{v,r}$ (Terzaghi et al. (1996)), was calculated as the vertical strain accumulated during loading from the seating stress to σ'_{v0} . The normalized void-ratio change (Lunne et al. (1997)) was also computed as $\Delta e/e_0 = (e_i - e_{\sigma'_{v0}})/e_0$, where e_i is the void ratio at the beginning of reconsolidation, $e_{\sigma'_{v0}}$ is the void ratio after reconsolidation to σ'_{v0} , and e_0 is the reference initial void ratio.

4.2. Compression-index ratios

The compression-index ratio, $C_{r,i}/C_c$, used by DeJong et al. (2018), was considered to quantify changes in the shape of the consolidation response. The initial recompression index, $C_{r,i}$, is defined as the slope of the first-loading recompression segment in $e - \log \sigma'_v$ space, measured from the seating stress to σ'_{v0} , unless otherwise noted. The virgin compression index, C_c , was measured over a post-yield stress interval representative of normally consolidated compression. According to DeJong et al. (2018), disturbance increases as $C_{r,i}/C_c$ increases.

4.3. Work-based recompression index

Following DeJong et al. (2018) one-dimensional work per unit volume was computed as $W = \int \sigma'_v d\epsilon_v$ directly from the consolidation curves, and used to estimate the $C_{rw,i}/C_{cw}$

ratio defined by DeJong et al. (2018), where $C_{rw,i}$ is the work-based initial recompression index and C_{cw} is the work-based virgin compression index. See DeJong et al. (2018) for additional details on these metrics. Stress intervals similar to those used for $C_{r,i}$ and C_c were used in the estimations. Disturbance increases as $C_{rw,i}/C_{cw}$ increases; this metric provides a stress-strain-work-based measure of recompression behavior, rather than relying only on slopes in $e - \log \sigma'_v$ space (DeJong et al. (2018)).

4.4. Oedometer modulus-shape indicator

The oedometer modulus-shape indicator, M_0/M_L , follows the sample-quality interpretation proposed by Karlsrud and Hernandez-Martinez (2013). The tangent constrained modulus is calculated from a consolidation curve as $M = d\sigma'_v/d\epsilon_v$. Two characteristic values are identified. The maximum initial reloading modulus, M_0 , defined during loading toward the yield region, after the initial seating portion of the curve. The minimum modulus, M_L , defined as the minimum tangent modulus reached after the modulus decreases during yielding and before it increases again along the virgin compression range. The ratio M_0/M_L is used as an indicator of disturbance, with higher values associated with lower disturbance. Karlsrud and Hernandez-Martinez (2013) provides additional details.

4.5. Void-index-based indicator

The void-index-based indicators follow the framework proposed by Cai et al. (2026). The void index was defined by Burland (1990) as $I_v = (e - e_{100}^*)/(e_{100}^* - e_{1000}^*)$, where e_{100}^* and e_{1000}^* are reference void ratios on the intrinsic compression line at vertical effective stresses of 100 and 1000 kPa, respectively. For natural soils, Cai et al. (2026) defined the disturbance indicator as $D_s = (\Delta I_{v1} + \Delta I_{v2})/I_{v0}$, where $\Delta I_{v1} = I_{v0} - I_{v1}$ captures the reduction in void index during reconsolidation to σ'_{v0} , and ΔI_{v2} is the distance between the compression curve and the sedimentation compression line at a reference stress of 1000 kPa and I_{v0} is the void index at the start of the consolidation test Cai et al. (2026). For reconstituted soils, they proposed the alternative metric $D_{re} = \Delta I_{v1}/I_{v0}$. See Cai et al. (2026) for additional details.

4.6. Exploratory swelling recompression metrics

In addition to the established metrics described above, two exploratory metrics were evaluated to assess whether the swelling response provides a useful reference for interpreting recompression behavior in the mine tailings we tested. The motivation for these metrics is discussed later in the "Discussion" section. The swelling index, C_s , was measured during unloading following virgin compression. The first exploratory metric is the angular difference between the initial recompression and swelling paths in $e - \log \sigma'_v$ space, defined as $\Delta\theta = \tan^{-1}(C_{r,i}) - \tan^{-1}(C_s)$. Larger positive values of $\Delta\theta$ indicate a greater difference between the recompression and swelling path slopes. The metric is therefore interpreted as a geometric slope-deviation index. The second exploratory metric is $C_{r,i}/C_s$, which compares the initial recompression slope with the swelling slope rather than with the virgin compression slope.

The metrics described above emphasize different aspects of disturbance. Reconsolidation strain and $\Delta e/e_0$ quantify volume change during reloading to σ'_{v0} . Compression index and work-based ratios characterize the relative recompression response. The M_0/M_L ratio captures the shape of the tangent modulus curve. Void-index metrics attempt to account for intrinsic compressibility. The exploratory C_s metrics evaluate whether the swelling response provides an additional material-specific reference for interpreting recompression, and the V_s ratio metric discussed in the introduction provides an independent small-strain stiffness comparison. Because these metrics are material- and protocol-dependent, they were interpreted collectively rather than independently. Their performance is first evaluated by comparing their trends with controlled disturbance levels in reconstituted specimens, which is discussed next.

5. Simulation of Disturbance and Assessments

The simulation of sampling-induced disturbance was conducted using CRS consolidation tests, following procedures similar to those of [DeJong et al. \(2018\)](#). Specimens were prepared using slurry deposition with the tailings previously described, characterized by a plasticity index (PI) of 7–15. The strain rate was selected in accordance with [ASTM D4186-06](#). Three



(a)



(b)



(c)



(d)

Figure 6: (a) LD specimen after the test (b) MD specimen- trimming and placing inside a smaller ring (c) HD specimen upon freezing (d) HD specimen upon thawing

disturbance levels were considered: low disturbance (LD), moderate disturbance (MD), and high disturbance (HD).

The LD condition represents sampling under minimal disturbance. Specimens were trimmed, placed in a 63.5-mm-diameter oedometer ring, saturated, and loaded to the virgin compression line. After confirming entry into the virgin compression region, the specimen was allowed to creep for at least 2 hours at a constant vertical effective stress, σ'_{v0} . The specimen was then unloaded to a low stress level (10–20 kPa) and subsequently reloaded along the virgin compression line. The recompression slope, C_r , was determined from this reloading path. The specimen was further loaded to σ'_{vu} , followed by a second unloading stage to evaluate the swelling index, C_s . The MD condition represents a disturbance induced by extrusion and specimen preparation (e.g., trimming). As in the LD procedure, a stress history was first established at σ'_{v0} on the virgin compression line, followed by unloading. The specimen was then extruded, trimmed, and placed into a smaller oedometer ring (50.8 mm diameter), after which it was saturated and reloaded into the virgin recompression line. C_r was obtained from this reloading path. The sample was then further loaded to σ'_{vu} , with a subsequent unloading stage used to estimate C_s . This MD condition is analogous to the best-practice disturbance (BPD) condition described in [DeJong et al. \(2018\)](#). The HD condition represents severe disturbance. After establishing the stress history at σ'_{v0} , the specimen was unloaded, extruded, wrapped, frozen for 24 hours, and subsequently thawed in air for 24 hours. In addition to extrusion and handling effects, this procedure introduces disturbance due to freeze–thaw cycling. To enhance the visible disturbance while minimizing moisture loss, specimens were wrapped and lightly sprayed with water before freezing. The freezing of pore water and external moisture produced pronounced additional disturbance. Subsequent testing followed the same procedure as for the MD condition. Figure 6a illustrates an LD specimen after testing. The disturbance induced by MD trimming is shown in Figure 6b, while the disturbances induced by HD freezing and thawing are illustrated in Figures 6c and 6d.

Figure 7 illustrates the performance of the disturbance metrics for the reconstituted specimens subjected to controlled disturbance levels. Figures 7a and 7b show the results

obtained using the strain-based criteria of Terzaghi et al. (1996) and Lunne et al. (1997). Both metrics show substantial overlap among LD, MD, and HD specimens and do not clearly separate the imposed disturbance levels. This result is consistent with DeJong et al. (2018), who showed that strain-based metrics, including volumetric strain upon reconsolidation and $\Delta e/e_0$, do not reliably quantify disturbance in intermediate natural soils. These results indicate that strain-based criteria alone are insufficient for classifying disturbance in the tailings considered in this study.

Figure 7c presents the performance of the work-based quality metric $C_{rw,i}/C_{cw}$ proposed by DeJong et al. (2018), whereas Figure 7d shows the corresponding results for the slope-based metric $C_{r,i}/C_c$. Both metrics provide better separation of the imposed disturbance levels than the strain-based criteria. DeJong et al. (2018) suggested that common threshold values could be applied to $C_{rw,i}/C_{cw}$ and $C_{r,i}/C_c$. However, the present results show that the two metrics do not yield equivalent thresholds, particularly for separating MD and HD specimens. For instance, a threshold of about 0.5 would separate the MD from HD tests better in the case of $C_{rw,i}/C_{cw}$. In addition, the threshold of 0.15 proposed by DeJong et al. (2018), expected to provide a separation between LD and MD specimens, does not apply for the tested tailings; thresholds on the order of 0.25–0.30 and 0.20–0.25 are more representative for $C_{rw,i}/C_{cw}$ and $C_{r,i}/C_c$, respectively. This suggests that universal thresholds should not be assumed; instead, threshold values may be material dependent and, in this case, may be affected by differences in compressibility between the present tailings and the natural silts examined by DeJong et al. (2018), as discussed later (see "Discussion" section). Also, the $C_{rw,i}/C_{cw}$ and $C_{r,i}/C_c$ metrics are clearer in separating LD from MD specimens than MD from HD specimens. This observation is consistent with DeJong et al. (2018), who reported similar limitations for natural soils with $PI < 15$, a range also representative of the materials examined in this study.

Figures 7e and 7f illustrate the performance of the exploratory swelling–recompression metrics $\Delta\theta$ and $C_{r,i}/C_s$, respectively. Both metrics separate the induced disturbance levels reasonably well for the tests considered. For $\Delta\theta$, values below 0.5 correspond to LD specimens, values between 0.5 and 2.0 correspond to MD specimens, and values greater than

2.0 correspond to HD specimens. For $C_{r,i}/C_s$, values below 1.3 generally correspond to LD specimens, values between 1.3 and 2.3 correspond to MD specimens, and values greater than 2.3 correspond to HD specimens.

Figure 7g shows the results for the M_0/M_L metric. The characteristic tangent-modulus response used to define M_0 and M_L was identifiable only for the LD specimens (Appendix C Figure C1). For visualization, the MD and HD specimens are plotted with $M_0/M_L = 1$, representing the absence of a distinct peak–minimum modulus response. The M_0/M_L criterion was proposed by Karlsrud and Hernandez-Martinez (2013) for structured Norwegian clays, for which high-quality block samples typically exhibit a pronounced contrast between the initial reloading modulus and the post-yield minimum modulus. Therefore, the threshold values proposed for these clays should not be assumed to apply directly to other materials. For the tailings tested herein, the LD specimens generally show M_0/M_L values greater than approximately 1.5, differing from the $M_0/M_L > 2.0$ threshold proposed by Karlsrud and Hernandez-Martinez (2013) for very good to excellent clay samples. Overall, for the examined tailings, M_0/M_L provides a qualitative indication of the LD response but has limited applicability to assessing MD and HD specimens.

Figure 7h presents the results obtained using the void-index-based metric D_{re} proposed by Cai et al. (2026). The metric was evaluated following the recommendations of Cai et al. (2026), which adopt the void-index normalization of Burland (1990). The D_{re} metric did not clearly separate LD, MD, and HD specimens. This may be because D_{re} relies on an intrinsic compression framework originally formulated for clays; therefore, its applicability to mine tailings may require material-specific calibration, which is beyond the scope of this study. In addition, D_{re} is based only on the change in void index during reconsolidation to the reference stress state and therefore retains the pre-yield focus of conventional reconsolidation-based metrics. Although Cai et al. (2026) showed that D_{re} was useful for a Shanghai soil, this material is a low-compressibility clay with a significant PI (17), which may partly explain the reduced performance observed in this study.

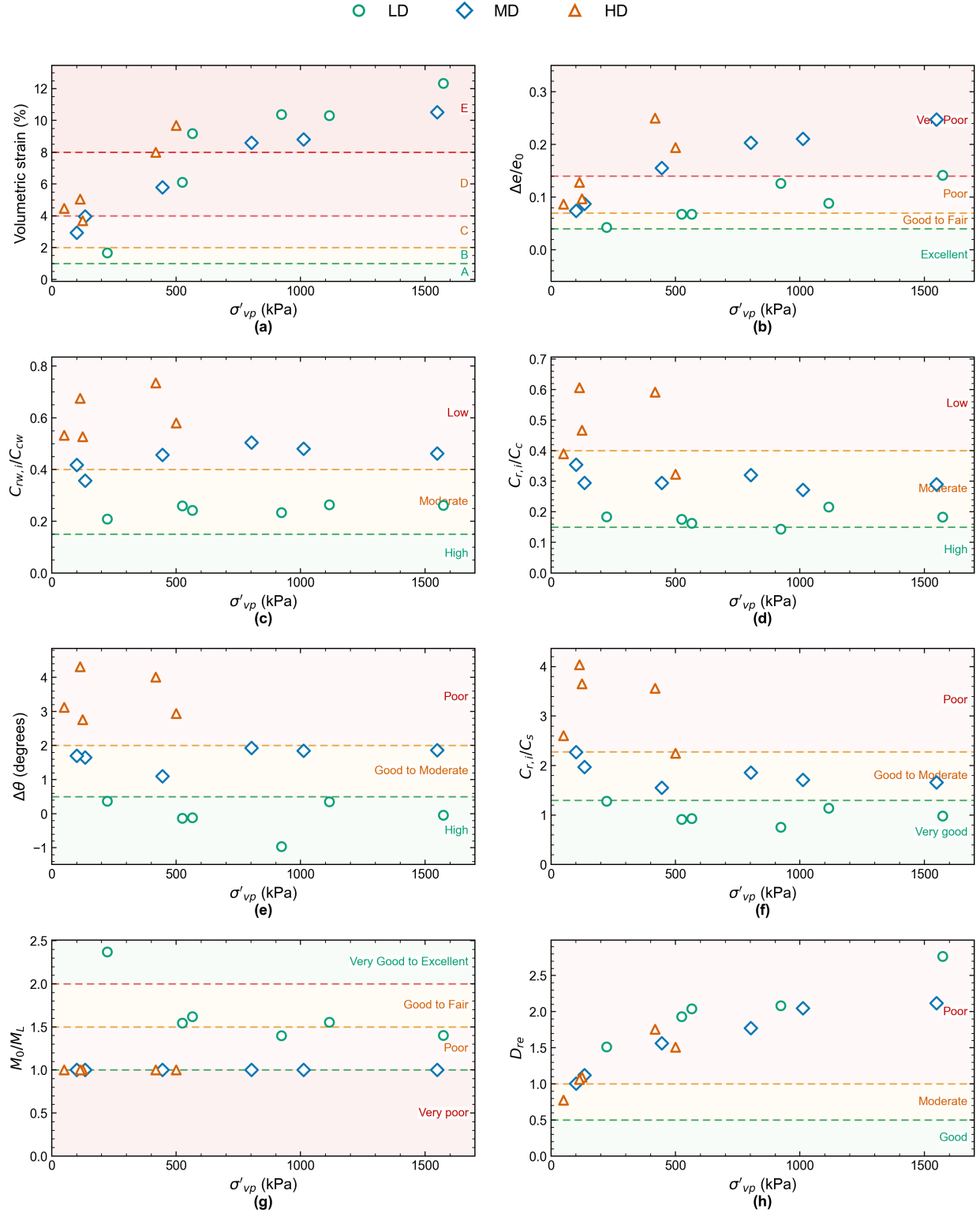


Figure 7: Evaluation of induced disturbance by methodical rating criteria with the thresholds (a) Terzaghi et al. (1996): Volumetric strain (%) (b) Lunne et al. (1997): $\Delta e/e_0$ (c) DeJong et al. (2018): $C_{rw,i}/C_{cw}$ (d) DeJong et al. (2018): $C_{r,i}/C_c$ (e) Proposed $\Delta\theta$ (degrees) (f) Proposed $C_{r,i}/C_s$ (g) Karlsrud and Hernandez-Martinez (2013): M_0/M_L (h) Cai et al. (2026): D_{re}

6. Quantitative versus Qualitative Assessment of Retrieved Samples

The metrics that provided useful separation among the LD, MD, and HD specimens in the controlled disturbance tests were used to assess the quality of retrieved field samples. These metrics include $C_{r,i}/C_c$, $C_{rw,i}/C_{cw}$, $\Delta\theta$, and $C_{r,i}/C_s$. $\epsilon_{v,r}$, and $\Delta e/e_0$, were also considered because of their common use in tailings engineering. In addition, a qualitative visual-based scoring scheme was developed to assess the condition of the retrieved samples and to compare these ratings with the quantitative metrics.

The qualitative scoring scheme combines observations from four stages : tube-level condition, extruded specimen condition, X-ray specimen condition, and prepared specimen condition. Tube-level condition was assessed from the condition of the tube and recovered sample during unpacking, unsealing, and extrusion. Extruded specimen condition was assessed from visual observations of the specific specimen selected for testing after extrusion and sectioning. X-ray specimen condition was assessed from the portion of the X-ray image corresponding to the tested specimen. Prepared specimen condition was assessed from issues observed during trimming, placement in the test device, and final specimen preparation. For each category, one of three ratings was assigned: good, moderate, or poor, corresponding to scores of 3, 2, and 1, respectively. Table 2 summarizes the scoring criteria used in the qualitative assessment.

For each test specimen, the four category scores were averaged to obtain an overall quality score. When multiple specimens were obtained from the same tube, the tube-level condition score was kept constant, whereas the other category scores were assigned independently for each specimen. Figure 8 presents representative examples of specimens with good, moderate, and poor overall quality scores. Specific category ratings are also shown for illustration.

Table 2: Description of qualitative sample-quality scoring criteria.

Parameter	Description	Basis for scoring	Quality assignment
Tube-level condition	Condition of the tube and recovered sample observed during unpacking, unsealing, and extrusion. This score is assigned once per tube and kept constant for specimens obtained from the same tube.	Sealing condition, sliding of the core during unsealing or extrusion, low recovery, overall sample issues, visible disturbance observed during extrusion while the sample is still in the tube.	Good = 3: no issues observed. Moderate = 2: one or two issues observed and not expected to strongly affect the specimens. Poor = 1: more than two issues observed, or one severe issue such as significant visible disturbance of the tube, significant sample issues etc.
Extruded specimen condition	External condition of the specific specimen selected for testing after extrusion and sectioning and before final trimming or placement in the test device.	Visible cracks, shear marks, piston marks, soft or squishy zones, dried/excess gel, chipping, spalling, or other externally visible defects on the extruded specimen.	Good = 3: no visible defects observed. Moderate = 2: minor or localized chipping, dents that do not significantly extended in the selected specimen. Poor = 1: severe or continuous defects, including major cracks, heavy shear marks, spalling, chipping etc.
X-ray specimen condition	Internal condition of the specific portion of the X-ray image corresponding to the specimen selected for testing.	Magnitude, continuity, and number of internal defects, including boundary shear marks, cracks, dents, voids, or disturbed zones visible in the X-ray image.	Good = 3: no internal defects observed in the selected region. Moderate = 2: isolated, minor, or localized defects not expected to dominate the test specimen. Poor = 1: multiple, severe, or continuous defects affecting the selected region.
Prepared specimen condition	Condition of the specimen after final trimming ¹ , placement in the oedometer ring or bender element setup, and preparation immediately prior to testing.	Damage introduced or observed during specimen trimming, transfer, seating, ring placement, surface leveling, bender-element preparation, or final setup.	Good = 3: no preparation-related issues observed. Moderate = 2: minor surface disturbance, small gaps, localized trimming marks, or limited spalling/chipping. Poor = 1: significant preparation-related disturbance, including major spalling, severe chippings, crackings, large boundary gaps, or severe disturbance during placement etc..

Figure 9 compares the quantitative metrics defined in the previous section with the qualitative score introduced herein for Gel Push, DM, and Shelby samples across all sites. For $C_{r,i}/C_c$, $C_{rw,i}/C_{cw}$, $\Delta\theta$, and $C_{r,i}/C_s$, the LD, MD, and HD scenarios and metric thresholds discussed previously are used to define “Very good,” “Good to moderate,” and “Poor” categories, respectively. The quantitative metric thresholds are not intended to be universal; rather, they are used as reference values to help interpret the quality of the recovered samples. For $\epsilon_{v,r}$ and $\Delta e/e_0$, the quality categories commonly used in practice were adopted.

Figures 9e and 9f show that the correlation between the qualitative scores and the strain-based metrics, $\epsilon_{v,r}$ and $\Delta e/e_0$, is poor. These metrics do not vary systematically with the qualitative scores and classify most samples as poor to very poor, even when the qualitative scores are high. This result is consistent with the limited performance of $\epsilon_{v,r}$ and $\Delta e/e_0$ observed in the controlled disturbance tests and with the limitations of these metrics reported for low- to moderate-plasticity natural soils (e.g., DeJong et al., 2018).

The other metrics, $C_{r,i}/C_c$, $C_{rw,i}/C_{cw}$, $\Delta\theta$, and $C_{r,i}/C_s$, show a comparable performance and clearer relationship with the qualitative score: lower metric values generally correspond to higher qualitative scores. Based on the global R^2 values, $\Delta\theta$ shows a slightly better performance and overall agreement with the qualitative score, followed by $C_{rw,i}/C_{cw}$ and $C_{r,i}/C_c$, whereas $C_{r,i}/C_s$ shows slightly weaker agreement. These metrics consistently identify the two Shelby samples recovered from TSF-1 at approximately 29 m as the lowest-quality samples (highest metric values), which is consistent with their qualitative score of 1. The thresholds discussed in the previous section classify these Shelby samples as poor, except for $C_{r,i}/C_c$, which places them in the moderate category.

The same metrics also consistently identify the DM samples from TSF-2 at approximately 18 m (qualitative score 2.3) and 25 m (qualitative score 2.5), the GP sample from TSF-2 at approximately 24 m (qualitative score 2.5), and the GP-Tr sample from TSF-3 at approximately 5.5 m (qualitative score 2.5) as among the highest-quality samples, consistent

¹Bender specimens obtained from DM samplers were tested without prior trimming. Therefore, preparation quality was assessed based on any issues observed while installing the membrane, O-rings, and other test components.

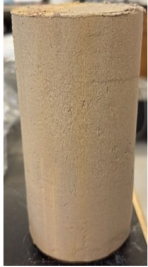
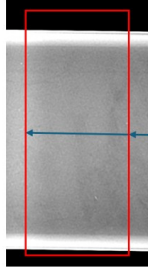
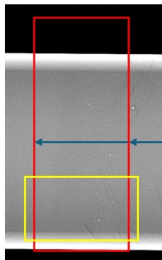
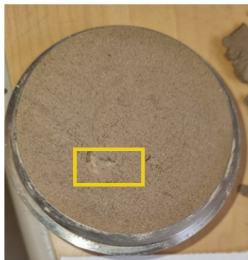
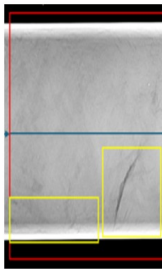
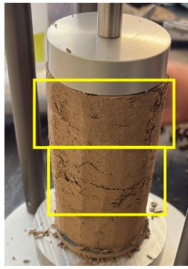
	Tube-level condition	Extruded specimen condition	X-ray specimen condition	Prepared specimen condition
Good	 (a)	 (b)	 (c)	 (d)
Moderate	 (e)	 (f)	 (g)	 (h)
Poor	 (i)	 (j)	 (k)	 (l)

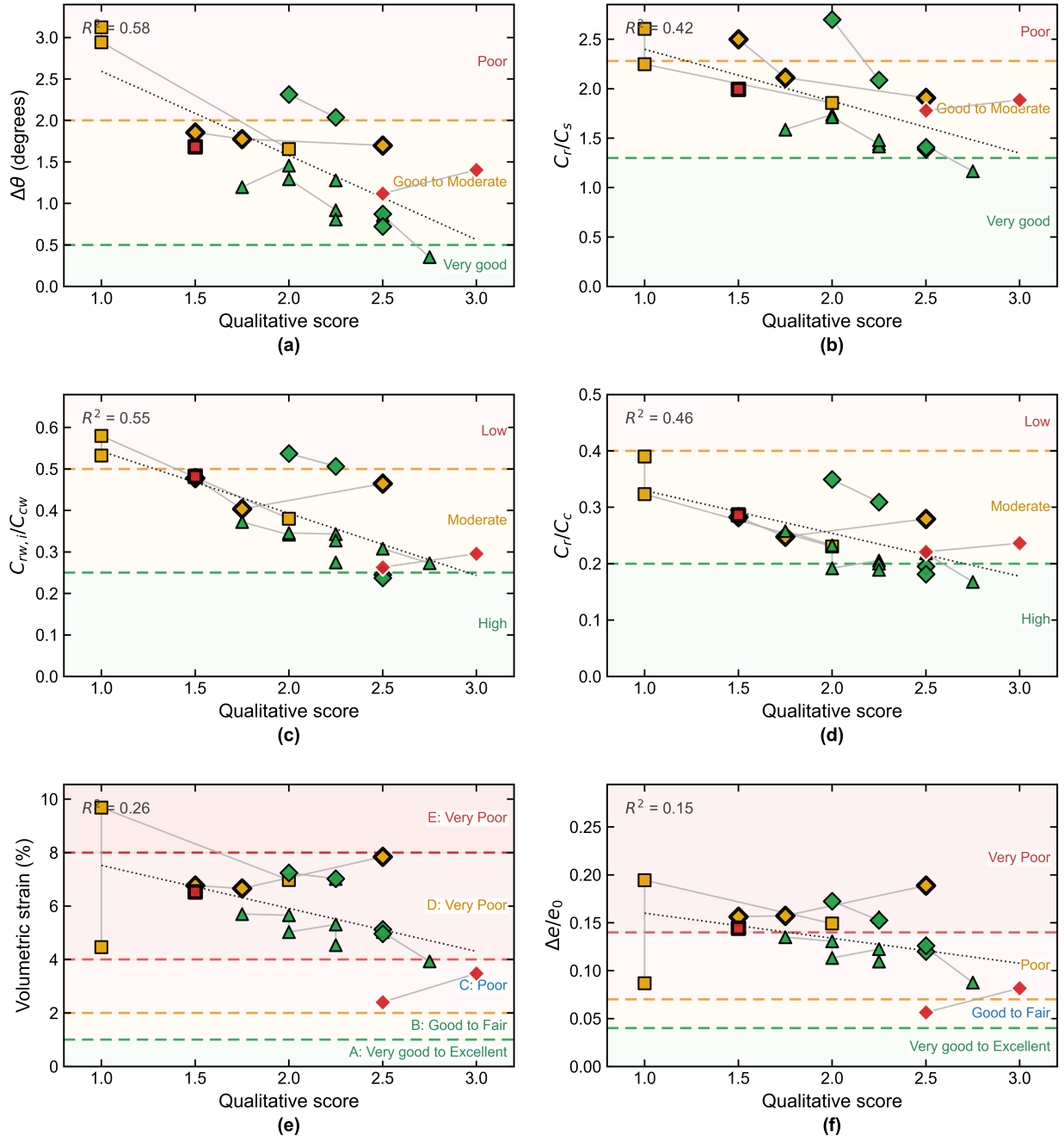
Figure 8: Examples of: (a) a good tube-level condition: a DM sample collected from TSF-2 at a depth of 24.99 m; (b) a good extruded specimen condition: a bender specimen retrieved from a GP sampler; (c) a good X-ray condition: an X-ray image of an oedometer specimen retrieved from a GP sampler; (d) a good prepared specimen condition: an oedometer specimen retrieved from a DM sampler; (e) a moderate tube-level condition: a GP sample collected from TSF-2 at a depth of 23.77 m with only noted issue of sliding between unsealing and placement in the extrusion assembly (f) a moderate extruded specimen condition: an oedometer specimen with dried gel marks on its surface, retrieved from a GP sampler; (g) a moderate X-ray condition: an X-ray image of an oedometer specimen showing several small boundary shear marks, retrieved from a DM sampler; (h) a moderate prepared specimen condition: an oedometer specimen showing surface chipping during trimming, retrieved from a Shelby sampler; (i) a poor tube-level condition: a GP sample collected from TSF-1 at a depth of 43.89 m, with no sealing, sliding issues, and overall bad quality of the retrieved specimens; (j) a poor extruded specimen condition: an oedometer specimen with heavy shear marks, retrieved from a GP sampler; (k) a poor X-ray condition: an X-ray image of an oedometer specimen showing a heavy shear/crack mark and several small boundary shear marks, retrieved from a Shelby sampler; and (l) a poor prepared specimen condition: a bender specimen showing heavy preparation marks and spalling across the sample due to the challenging material condition, retrieved from a Shelby sampler.

with their relatively high qualitative scores. The remaining DM and GP samples have metric values and qualitative scores that fall in between, with most classified as “Good to moderate”. The two remaining Shelby samples from TSF-1 (qualitative score 2.0) and TSF-3 (qualitative score 1.5) show quantitative metrics comparable to those of remaining GP samples, but indicate a lower quality (higher values for the metrics) than the remaining DM samples.

Notably, the GP-Tr samples collected from approximately 5.5 m at TSF-3 (qualitative scores of 2.5 and 3.0), which were composed predominantly of sandy material, consistently plot near the lower bound of the metric values observed for the Gel Push samples. This suggests comparatively high quality relative to most Gel Push samples, except for the GP-s sample from TSF-2 at approximately 24 m, which shows comparable values. The GP-Tr sample was transported frozen, which may have helped preserve sample quality; however, no definitive conclusion can be drawn because this was the only sample with these characteristics.

TSF-2 is the only site where multiple DM and Gel Push samples were collected; therefore, it enables direct comparisons that may be obscured in the aggregate analysis across all sites. Figure 10 presents the quantitative quality metrics for TSF-2 samples recovered using GP-s and DM samplers, considering $C_{r,i}/C_c$, $C_{rw,i}/C_{cw}$, $\Delta\theta$, and $C_{r,i}/C_s$. At depths of 17–19 m, all criteria consistently indicate lower metric values, and therefore higher sample quality, for the DM sampler. At depths of 23–25 m, the pattern is less distinct. The metrics are generally only slightly lower for DM than for GP-s, suggesting slightly better DM performance, except for $C_{rw,i}/C_{cw}$, which indicates slightly better performance for GP-s. The deeper DM sample at approximately 34 m has no corresponding GP-s sample. This sample exhibits metric values comparable to the lowest values (i.e., highest quality) estimated for the GP-s samples.

An additional sample-quality evaluation was conducted by comparing stress-normalized shear-wave velocities, $V_{s,1}$, measured in the field and in the laboratory. Specifically, the laboratory-to-field $V_{s,1}$ ratio was used. Field $V_{s,1}$ values were estimated as discussed in the “Introduction” section, using the CPTu-based shear-wave velocity measurements presented in the “Geotechnical Conditions” section. Laboratory $V_{s,1}$ values were estimated from bender element tests by measuring shear-wave velocity at different mean effective stresses. Appendix D figure D1 provides the bender element data and interpretation. Figure 11 shows the



Site:	Sampler:	Depth / origin:	Connection:
TSF-1	DM	Shallow (<15 m)	— Same tube
TSF-2	GP	Mid (15-35 m)	
TSF-3	SH	Deep (>35 m)	

Figure 9: Quantitative metrics vs Qualitative scores of retrieved specimens by (a) $\Delta\theta$ (degrees), (b) $C_{r,i}/C_s$, (c) DeJong et al. (2018): $C_{rw,i}/C_{cw}$ (d) DeJong et al. (2018): $C_{r,i}/C_c$ (e) Terzaghi et al. (1996): Volumetric strain (%) (f) Lunne et al. (1997): $\Delta e/e_0$

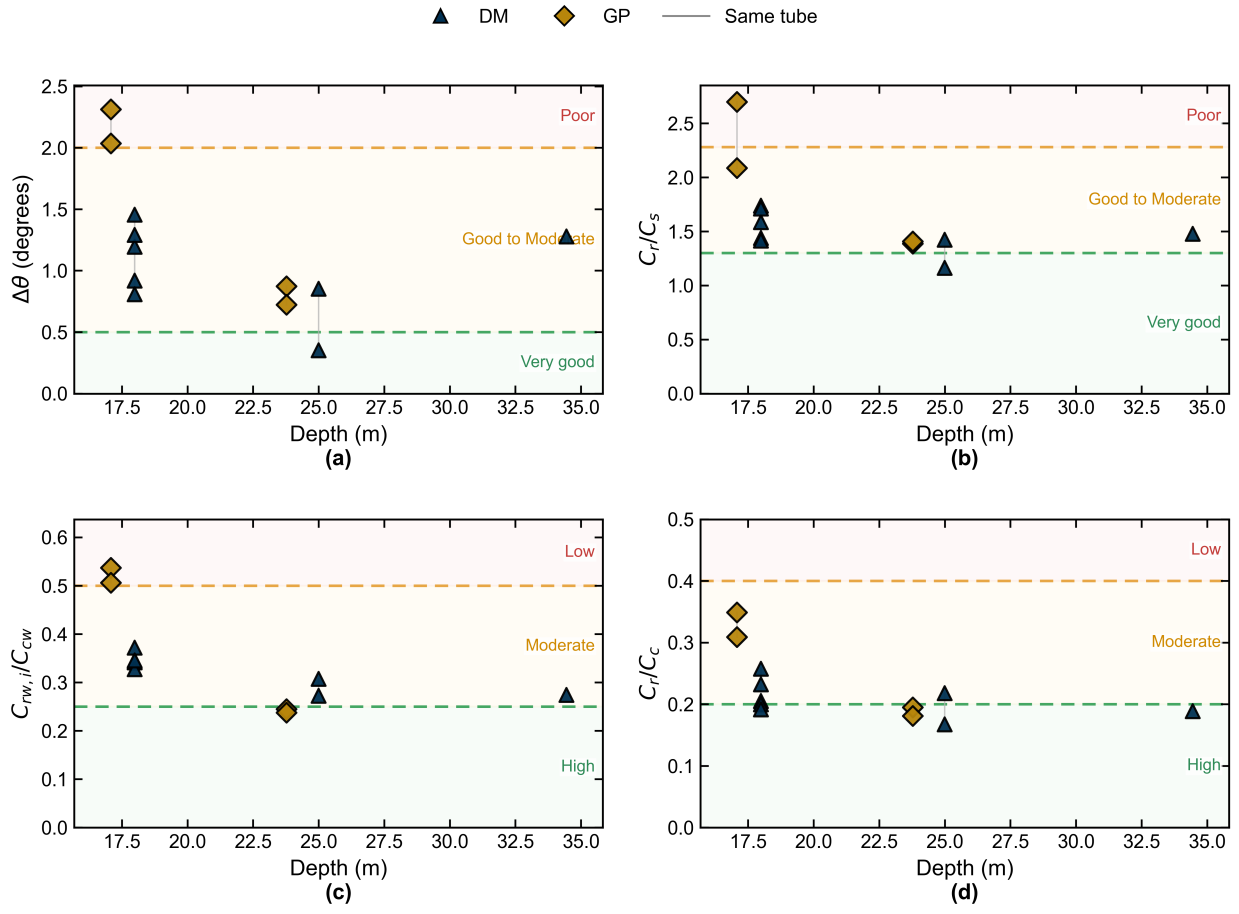


Figure 10: Depth vs quantitative scores of retrieved specimens in TSF-2 by (a) $\Delta\theta$ (degrees), (b) $C_{r,i}/C_s$, (c) DeJong et al. (2018): $C_{rw,i}/C_{cw}$ (d) DeJong et al. (2018): $C_{r,i}/C_c$

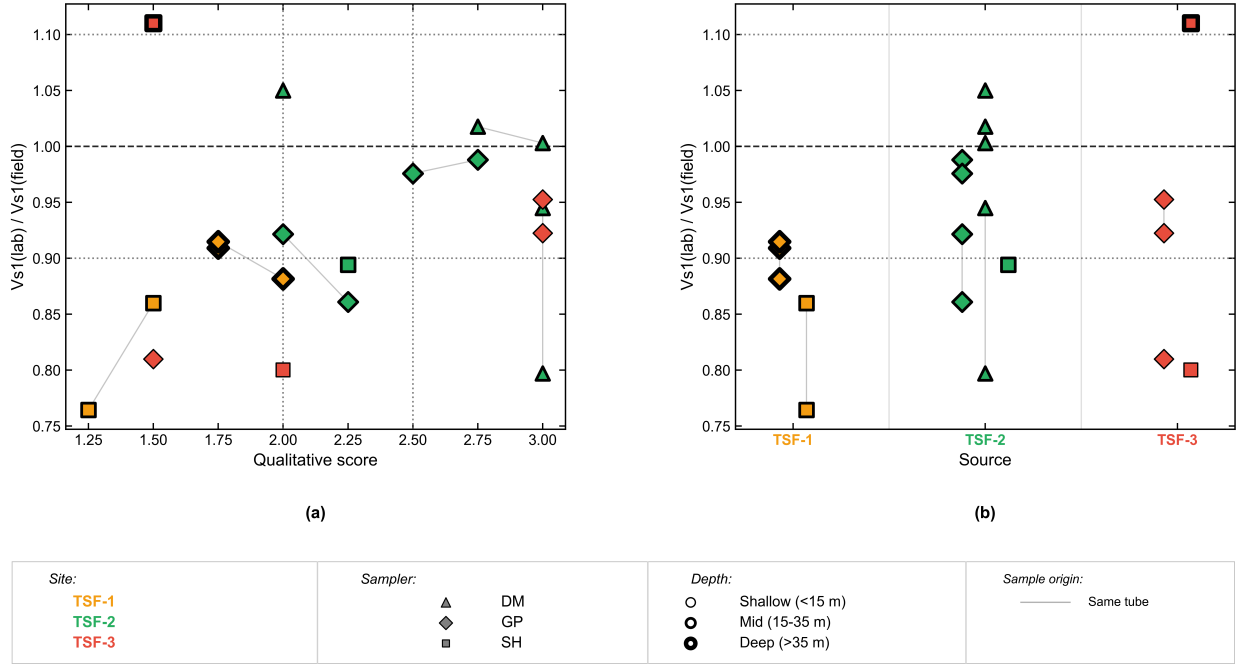


Figure 11: V_{s1} based sample quality evaluation: (a) V_{s1} ratio against qualitative score and (b) site-wise V_{s1} ratio for different samplers

laboratory-to-field $V_{s,1}$ ratios for Gel Push, DM, and Shelby samples from the different sites. Qualitative scores, based on the same framework used for the consolidation specimens, were also assigned to the samples for which bender element measurements were performed.

Samples with $V_{s,1}$ ratios below approximately 0.85 generally correspond to lower qualitative scores. For example, samples from the Shelby tube recovered at TSF-1 at approximately 29 m, which showed the lowest quality in the consolidation-based assessment, have $V_{s,1}$ ratios between approximately 0.75 and 0.85 and qualitative scores lower than 1.5. Therefore, the consolidation and bender element results are consistent in identifying low sample quality for this Shelby tube. Additional low-ratio cases include samples from TSF-3 for which only bender element tests were performed, including a GP-S sample at approximately 2.4 m and a Shelby sample at approximately 3.4 m. These specimens have $V_{s,1}$ ratios on the order of 0.8 and qualitative scores lower than 2.0. Overall, for the tested samples, $V_{s,1}$ ratios below approximately 0.85 are generally associated with qualitative scores lower than 2.0, suggesting lower sample quality.

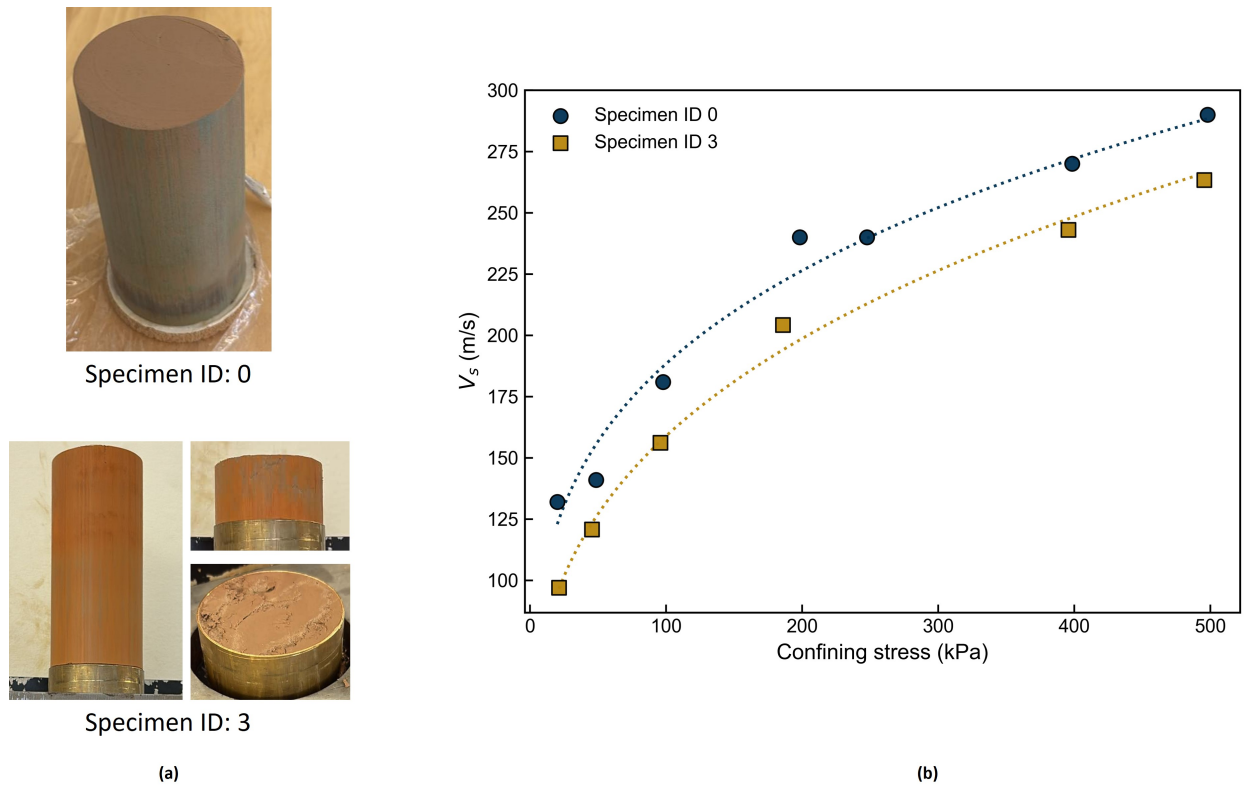


Figure 12: DM sampler collected from TSF-2 at depth 34.99 meter showing (a) different material properties for two different specimens (ID 0 and 3); (b) difference in measured shear wave velocity for those specimens at different confining stresses

Samples with $V_{s,1}$ ratios close to 1.0 generally correspond to higher-qualitative scores. For example, Bender element specimens from DM tubes for which consolidation tests showed low metric values, and therefore higher inferred quality, at approximately 18 m and 24 m exhibit $V_{s,1}$ ratios between approximately 1.00 and 1.05. The specimens with $V_{s,1}$ ratios closest to 1.0 have qualitative scores greater than 2.5, whereas the specimen with a $V_{s,1}$ ratio of 1.05 has a qualitative score of 2.0. Similarly, specimens from Gel Push tubes for which consolidation tests showed low metric values, including the GP-s tube from TSF-2 at approximately 24 m and the GP-Tr tube from TSF-3 at approximately 5.5 m, exhibit $V_{s,1}$ ratios near 1.0 for the TSF-2 samples and between approximately 0.90 and 1.0 for the TSF-3 samples. The corresponding qualitative scores are greater than 2.5. Thus, the V_s -based assessment is generally consistent with the consolidation-based assessment in identifying higher-quality samples as those with high qualitative scores, laboratory-to-field $V_{s,1}$ ratios close to 1.0, and low consolidation-based metric values.

Other samples show intermediate responses. For example, the $V_{s,1}$ ratios for samples from Gel Push tubes at TSF-1 at approximately 44 m and TSF-2 at approximately 17 m range from approximately 0.85 to 0.93, whereas samples from the same tubes were classified by consolidation-based metrics as ranging from “good to moderate” to “poor.” These values are consistent with lower performance relative to the samples identified as having higher quality. A similar observation applies to a sample from a Shelby tube at TSF-3 at approximately 49 m, which has a $V_{s,1}$ ratio of approximately 1.13, whereas a sample from the same tube shows consolidation-based metrics in the “good to moderate” range and close to the “poor” category.

An interesting case is the DM tube collected from TSF-2 at approximately 34 m. Two bender element tests on samples recovered from this tube produced $V_{s,1}$ ratios of approximately 0.80 and 0.95, whereas the qualitative scores for both bender specimens are 3.0 and the consolidation-based metrics for a sample from the same tube indicate a “good to moderate” condition. This may suggest an inconsistency between the lower $V_{s,1}$ ratio and its corresponding qualitative score. A closer inspection indicates a material change within the tube (figure 12a). The material in the upper half of the tube has a PI of approximately

15 and corresponds to the first bender specimen, whereas the material in the lower half has a PI of approximately 25 and corresponds to the second bender specimen. This transition is consistent with the shear-wave velocities measured at different mean effective stresses (figure 12b), which show lower velocities, and therefore lower $V_{s,1}$, for the higher-plasticity specimen. In contrast, the field shear-wave velocity profile does not resolve this material transition and effectively smooths the measured velocity over the interval. This produces a lower laboratory-to-field $V_{s,1}$ ratio for the higher-plasticity material and could be misinterpreted as lower sample quality. This case highlights the need for careful interpretation of laboratory-to-field $V_{s,1}$ ratios.

7. Discussion

7.1. Exploratory disturbance metrics in consolidation tests

In the previous assessments, two exploratory swelling recompression metrics, $\Delta\theta$ and $C_{r,i}/C_s$, were considered to evaluate sample disturbance. Here, we discuss the motivation for these metrics. Figure 13a illustrates the conceptual basis. Consider an initially normally consolidated material at void ratio e_0 and vertical effective stress σ'_{v0} . Path 0–1 represents an ideal sampling scenario, in which stress relief occurs along the swelling/unloading line, and no additional disturbance is imposed. After ideal sampling, the remaining effective stress is mainly associated with suction or negative pore pressure. In intermediate soils, including low-plasticity tailings, the capacity to retain negative pore pressure after stress relief may be lower than in plastic clays (Zapata et al. (2000); Sadrekarimi and Zehforoosh (2025)). Therefore, ideal unloading may move the specimen to a low-stress, highly overconsolidated state that may lie on the dense/dilative side of the critical state line (CSL). This high-OCR idealization is consistent with DeJong et al. (2018), who used OCR values of up to about 5 to reproduce near-ideal sampling conditions in low-plasticity silty soils.

Path 1–2 represents one possible manifestation of additional disturbance during non-ideal sampling. In a subsequent consolidation test, the specimen recompresses from the disturbed state toward the normally consolidated line. The upward shift to state 2 implies a greater

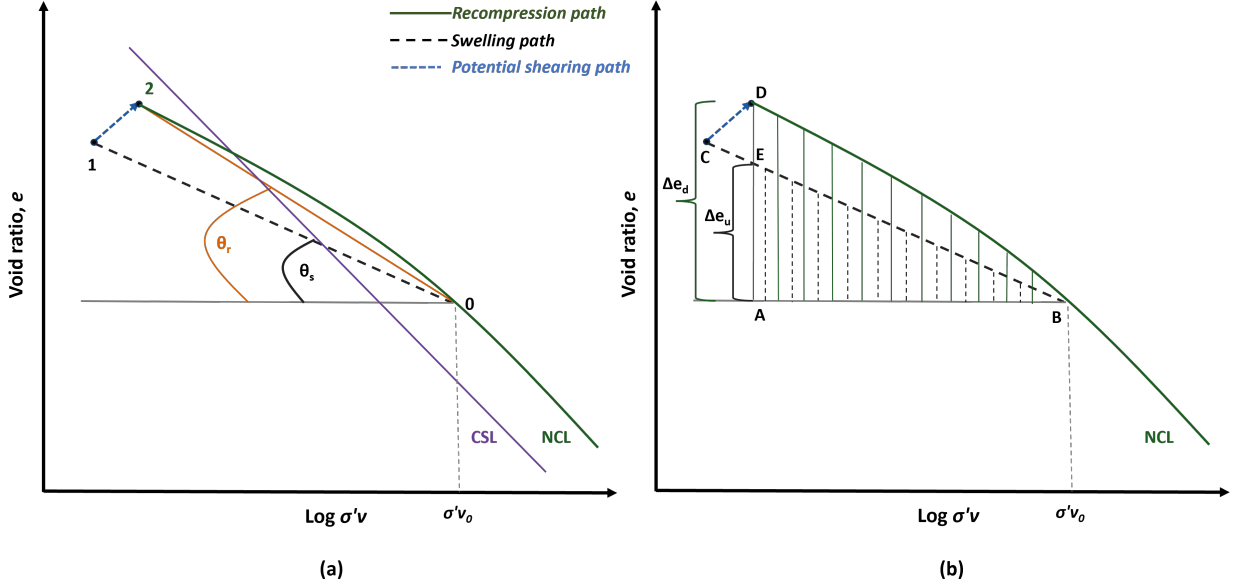


Figure 13: Evaluation of disturbance in intermediate soil

557 departure between disturbed recompression and ideal rebound to return the specimen toward
 558 the pre-disturbance yield or virgin-compression state, consistent with the broader work-based
 559 interpretation of sampling disturbance discussed by Becker et al. (2011) and DeJong et al.
 560 (2018).

561 In this context, sampling disturbance may be reflected by a recompression path that devi-
 562 ates from the ideal swelling/rebound path. This deviation can be represented geometrically
 563 by the difference between the slopes of the recompression and swelling paths in $e - \log \sigma'_v$
 564 space. The $\Delta\theta$ index is defined as:

$$\begin{aligned}\Delta\theta &= \theta_r - \theta_s \\ &= \tan^{-1}(C_r) - \tan^{-1}(C_s),\end{aligned}$$

565 where C_r and C_s are the recompression and swelling indices, respectively. Larger positive
 566 values of $\Delta\theta$ indicate a greater departure of the recompression path from the ideal swelling
 567 path and are interpreted as greater sampling-induced disturbance.

Figure 13b provides an alternative geometric interpretation of disturbance. The area enclosed by ABD is interpreted as a geometric proxy for the additional recompression demand associated with the upward shift of the path caused by sampling-induced disturbance. To evaluate this disturbance relative to an ideal sampling condition, an ideal unloading path over the same stress interval is required. This reference path, denoted as BE , represents the ideal swelling/rebound response. The area ABE therefore provides a geometric reference associated with ideal unloading over the same stress interval.

If the curved paths are approximated by straight-line segments over the interval AB , the ratio of the triangular areas can be written as

$$\begin{aligned}\frac{\Delta ABD}{\Delta ABE} &= \frac{0.5 AB AD}{0.5 AB AE} \\ &= \frac{\Delta e_d}{\Delta e_u},\end{aligned}$$

where Δe_d is the void-ratio change associated with the disturbed recompression path over the selected stress interval, and Δe_u is the void-ratio change associated with ideal unloading/rebound over the same interval. This expression resembles Lunne's $\Delta e/e_0$ criterion, but it differs from it because the denominator is not the initial void ratio, but the ideal rebound void ratio change over the selected stress interval. Because both quantities are evaluated over the same log-stress interval, this ratio can also be expressed using secant slopes as

$$\begin{aligned}\frac{\Delta e_0}{\Delta e_u} &= \frac{\Delta e_0/AB}{\Delta e_u/AB} \\ &= \frac{C_r}{C_s},\end{aligned}$$

where C_r and C_s are the recompression and swelling indices interpreted over the corresponding stress interval. The two interpretations shown in Figure 13 therefore motivate related but distinct exploratory metrics: $\Delta\theta$ captures the angular departure between the recompression and swelling paths, whereas $C_{r,i}/C_s$ follows from a slope-ratio interpretation. These interpretations rely on simplifying assumptions; therefore, the resulting metrics should

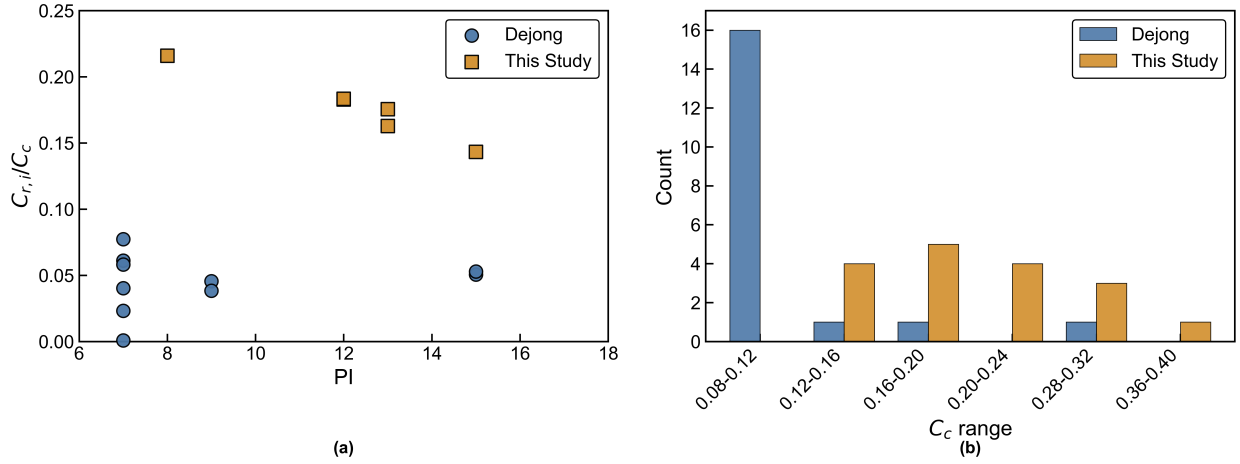


Figure 14: Comparison of (a) $C_{r,i}/C_c$ against PI, and (b) distribution of C_c with DeJong et al. (2018) presented data

be considered only exploratory.

7.2. Compressibility and quality thresholds

The controlled disturbance tests showed that the quality thresholds recommended by DeJong et al. (2018) for natural silts differ from those identified in this study for the examined mine tailings. To investigate this difference, Figure 14a compares selected data from DeJong et al. (2018) for silty soils with the tested tailings for specimens with plasticity indices between 7 and 17, which is comparable to the range of the materials tested herein. For the purposes of this discussion, the comparison focuses on high-quality specimens and $C_{r,i}/C_c$. The threshold of 0.15 recommended by DeJong et al. (2018) is consistent with the high-quality range of their data; however, it does not apply to the tested tailings, for which a threshold in the order of 0.20–0.25 appears more representative.

A closer comparison of the datasets suggests that this difference may be related, at least in part, to differences in material compressibility. Figure 14b shows that the tailings tested in this study have generally higher C_c values than the natural silts evaluated by DeJong et al. (2018). For the tested tailings, the mean C_c is approximately 0.21, with most values ranging from 0.16 to 0.24. In contrast, most of the DeJong et al. (2018) data have C_c values lower than 0.14, with a mean value of approximately 0.13 and only a few specimens reaching

values comparable to those of the tested tailings. Because $C_{r,i}/C_c$ is already normalized by C_c , this comparison should not be interpreted as showing that C_c alone controls the threshold. Rather, C_c is used here as an indicator of broader differences in compressibility and mineralogy that may influence the thresholds. Although uncertainties remain, these comparisons suggest that material-specific properties can influence the threshold values used to interpret sample-quality metrics. Additional future work is needed in this direction.

8. Conclusions

This study evaluated the quality of mine-tailings samples recovered using DM, Gel Push, and Shelby samplers at three TSFs. Gel Push was used at all three sites, whereas DM and Shelby samplers were used at selected sites for comparison. Samples from TSF-2 provided the basis for direct comparison between DM and Gel Push samplers. Geotechnical conditions were characterized using CPTu tests with shear-wave velocity measurements and laboratory index testing. The sample-quality evaluation combined CRS oedometer tests, bender element tests, controlled disturbance tests on reconstituted specimens, and a qualitative scoring scheme based on tube-level condition, extruded specimen condition, X-ray specimen condition, and prepared specimen condition.

The controlled disturbance tests showed that commonly used strain-based consolidation criteria, including $\epsilon_{v,r}$, and $\Delta e/e_0$, did not separate the imposed disturbance levels for the tested tailings. These metrics also showed limited agreement with qualitative scores for retrieved field samples and tended to classify many specimens as poor to very poor, even when other evidence indicated better sample condition. These results, consistent with previous studies on natural silts, indicate that strain-based criteria alone are not sufficient for evaluating sample disturbance in the tailings considered in this study.

In contrast, $C_{rw,i}/C_{cw}$, $C_{r,i}/C_c$, $\Delta\theta$, and $C_{r,i}/C_s$ provided more informative trends. These metrics separated the controlled disturbance levels more clearly than $\epsilon_{v,r}$ and $\Delta e/e_0$ and were generally consistent with qualitative scores for field-retrieved specimens. The M_0/M_L metric was useful only as a qualitative indicator of the LD response, showing a limited applicability for assessing MD and HD specimens. The void-index-based metric D_{re} also did not clearly

separate the imposed disturbance levels, suggesting that void-index frameworks developed primarily for clays may require additional calibration before being applied to mine tailings.

Threshold values proposed in previous studies for $C_{rw,i}/C_{cw}$ and $C_{r,i}/C_c$ did not transfer directly to the tested tailings. This result suggests that disturbance thresholds developed for natural soils should not be assumed to be directly transferable to mine tailings. Instead, threshold values appear to be material dependent and should be interpreted in relation to material properties. An initial comparison with published data suggests that compressibility may influence threshold values, although additional work is needed to evaluate this effect more broadly. Given these uncertainties, multiple quantitative metrics should be used collectively and interpreted in a referential manner. For the tested tailings, lower values of $C_{rw,i}/C_{cw}$, $C_{r,i}/C_c$, $\Delta\theta$, and $C_{r,i}/C_s$ generally indicate higher sample quality.

The laboratory-to-field $V_{s,1}$ ratio provided an independent assessment that was generally consistent with the consolidation-based metrics. Ratios below approximately 0.85 were generally associated with lower qualitative scores, whereas ratios close to 1.0 generally corresponded to higher qualitative scores and lower consolidation-based disturbance metrics (higher quality) for specimens from the same tube. However, the results also showed that $V_{s,1}$ ratios can be affected by material variability within the sampled interval. For example, one DM tube from TSF-2 showed that changes in the plasticity index and shear-wave velocity within the tube can produce laboratory-to-field $V_{s,1}$ ratios that may be misinterpreted as lower sample quality.

For the materials evaluated, DM and Gel Push samplers generally recovered higher-quality specimens than Shelby tubes. Shelby samples from TSF-1 at approximately 29 m were consistently identified as low quality by qualitative scores, consolidation-based metrics, and $V_{s,1}$ ratios. In contrast, DM samples from TSF-2 and selected GP-S and GP-Tr samples showed comparatively higher quality. Comparisons at TSF-2 between DM and Gel Push samples suggest that DM performed better than Gel Push at depths of approximately 17–19 m, whereas the difference between DM and GP-S was less distinct at approximately 23–25 m, with slightly better performance for DM depending on the metric considered.

Overall, the results support the use of multiple complementary indicators for evaluating

sample quality in mine tailings. Qualitative scores, consolidation-based metrics, and shear-wave velocity-based measurements capture different aspects of sampling disturbance, and no single metric should be used in isolation. The findings highlight the need for material-specific interpretation of disturbance metrics in tailings, particularly when applying criteria originally developed for natural clays or silts. The $C_{rw,i}/C_{cw}$, $C_{r,i}/C_c$, $\Delta\theta$, and $C_{r,i}/C_s$ metrics provided useful interpretations for the tested tailings and offer a promising basis for future calibration using additional tailings materials and sampling protocols.

9. Acknowledgment

We thank AECOM for helpful discussions during data collection and Fugro for assisting with field sample collection.

10. Competing interests statement

The authors declare that there are no competing interests.

11. Data availability statement

Data generated or analyzed during this study are available from the corresponding author upon reasonable request.

12. Funding statement

This study was funded by internal funds at the Georgia Institute of Technology. The first author also thanks the TAILENG network for numerous discussions on sample quality.

References

Andrus, R.D., Stokoe, K.H., 2000. Liquefaction Resistance of Soils from Shear-Wave Velocity. *Journal of Geotechnical and Geoenvironmental Engineering* 126, 1015–1025. doi:[10.1061/\(ASCE\)1090-0241\(2000\)126:11\(1015\)](https://doi.org/10.1061/(ASCE)1090-0241(2000)126:11(1015)).

- ASTM D1587/D1587M-15, . Standard Practice for Thin-Walled Tube Sampling of Fine-Grained Soils for Geotechnical Purposes (Withdrawn 2024). Technical Report. ASTM International.
- ASTM D4186-06, . Test Method for One-Dimensional Consolidation Properties of Saturated Cohesive Soils Using Controlled-Strain Loading. URL: <https://store.astm.org/d4186-06.html>, doi:10.1520/D4186-06.
- Becker, D.E., Crooks, J.H., Been, K., Jefferies, M.G., 2011. Work as a criterion for determining in situ and yield stresses in clays. <https://doi.org/10.1139/t87-070> 24, 549–564. URL: <https://doi.org/10.1139/t87-070?download=true>, doi:10.1139/T87-070.
- Bray, J.D., Beyzaei, C.Z., Cubrinovski, M., Riemer, M., Markham, C., Stringer, M.E., Jacka, M., Wentz, F.J., Haycock, I., 2016. Recommendations for High-Quality Field Sampling Procedures: Using the Dames and Moore (DM) Hydraulic Fixed-Piston Thin-Walled Tube Sampler. Technical Report. University of Canterbury. URL: <https://hdl.handle.net/10092/101428>.
- Bray, J.D., Sancio, R.B., 2006. Assessment of the Liquefaction Susceptibility of Fine-Grained Soils. *Journal of Geotechnical and Geoenvironmental Engineering* 132, 1165–1177. URL: <https://doi.org/10.1061/%28ASCE%291090-0241%282006%29132%3A9%281165%29?download=true><https://ascelibrary.org/doi/10.1061/%28ASCE%291090-0241%282006%29132%3A9%281165%29>, doi:10.1061/(ASCE)1090-0241(2006)132:9(1165).
- Burland, J.B., 1990. On the compressibility and shear strength of natural clays. *Géotechnique* 40, 329–378. URL: <https://dx.doi.org/10.1680/geot.1990.40.3.329>, doi:10.1680/GEOT.1990.40.3.329.
- Cai, H., Ye, G., Feng, J., Xie, W., Fan, K., Shen, Y., 2026. A void-index-based framework for sample quality evaluation in low-plasticity silty clays. *Transportation Geotechnics* 61. doi:10.1016/j.trgeo.2026.102104.

- 710 Da Fonseca, A., Ferreira, C., Ramos, C., Molina Gómez, F., 2019. The geotechnical test site
711 in the greater Lisbon area for liquefaction characterisation and sample quality control of
712 cohesionless soils. AIMS Geosciences 5, 325–343. doi:[10.3934/geosci.2019.2.325](https://doi.org/10.3934/geosci.2019.2.325).
- 713 DeGroot, D.J., DeJong, T.J., Lukas, G.W., Krage, P.C., Blaker, O., 2020. Re-
714 cent advances in sampling and laboratory characterization of intermediate soils,
715 in: 6th International Conference on Geotechnical and Geophysical Site Charac-
716 terization (ISC2020). URL: [https://www.issmge.org/publications/publication/
717 recent-advances-in-sampling-and-laboratory-characterization-of-intermediate-soils](https://www.issmge.org/publications/publication/recent-advances-in-sampling-and-laboratory-characterization-of-intermediate-soils),
718 doi:<https://doi.org/10.53243/ISC2020-509>.
- 719 DeJong, J.T., Krage, C.P., Albin, B.M., DeGroot, D.J., 2018. Work-Based Framework
720 for Sample Quality Evaluation of Low Plasticity Soils. Journal of Geotechnical and
721 Geoenvironmental Engineering 144. doi:[10.1061/\(asce\)gt.1943-5606.0001941](https://doi.org/10.1061/(asce)gt.1943-5606.0001941).
- 722 Fioravante, V., Giretti, D., Jamiolkowski, M., Rocchi, G., 2012. Triaxial tests on undisturbed
723 gravelly soils from the Sicilian shore of the Messina Strait. Bulletin of Earthquake
724 Engineering 2012 10:6 10, 1717–1744. URL: [https://link.springer.com/article/10.
725 1007/s10518-012-9374-7](https://link.springer.com/article/10.1007/s10518-012-9374-7), doi:[10.1007/S10518-012-9374-7](https://doi.org/10.1007/S10518-012-9374-7).
- 726 Viana da Fonseca, A., Ferreira, C., Ramos, C., Molina-Gómez, F., 2019. The geotechnical
727 test site in the greater Lisbon area for liquefaction characterisation and sample quality
728 control of cohesionless soils. AIMS Geosciences 5, 325–343. URL: [http://www.aimspress.
729 com/article/10.3934/geosci.2019.2.325](http://www.aimspress.com/article/10.3934/geosci.2019.2.325), doi:[10.3934/geosci.2019.2.325](https://doi.org/10.3934/geosci.2019.2.325).
- 730 Viana da Fonseca, A., Pineda, J., 2017. Getting high-quality samples in ‘sensitive’
731 soils for advanced laboratory tests. Innovative Infrastructure Solutions 2017 2:1
732 2. URL: <https://link.springer.com/article/10.1007/s41062-017-0086-3>, doi:[10.
733 1007/S41062-017-0086-3](https://doi.org/10.1007/S41062-017-0086-3).
- 734 Hight, D.W., Leroueil, S., 2003. Characterization of soils for engineering purposes, in: Tan,

735 T.S., Phoon, K.K., Hight, D.W., Leroueil, S. (Eds.), Characterisation and Engineering
736 Properties of Natural Soils. A.A. Balkema, Lisse, Netherlands. volume 1.

737 Jamiolkowski, M., 2012. Role of geophysical testing in geotechnical site characterization.
738 Soils and Rocks 35, 117–137. doi:[10.28927/SR.352117](https://doi.org/10.28927/SR.352117).

739 Jamiolkowski, M., Masella, A., 2015. Geotechnical Characterization of Copper Tailings at
740 Zelazny Most Site. Technical Report.

741 Karlsrud, K., Hernandez-Martinez, F.G., 2013. Strength and deformation properties of Nor-
742 wegian clays from laboratory tests on high-quality block samples. Canadian Geotechnical
743 Journal 50, 1273–1293. doi:[10.1139/cgj-2013-0298](https://doi.org/10.1139/cgj-2013-0298).

744 Landon, M.M., DeGroot, D.J., Sheahan, T.C., 2007. Nondestructive Sample Quality Assess-
745 ment of a Soft Clay Using Shear Wave Velocity. Journal of Geotechnical and Geoenviron-
746 mental Engineering 133, 424–432. doi:[10.1061/\(ASCE\)1090-0241\(2007\)133:4\(424\)](https://doi.org/10.1061/(ASCE)1090-0241(2007)133:4(424)).

747 Lim, G.T., Pineda, J., Boukpeti, N., Carraro, J.A.H., Fourie, A., 2019. Effects of sampling
748 disturbance in geotechnical design. Canadian Geotechnical Journal 56, 275–289. doi:[10.1139/cgj-2018-0016](https://doi.org/10.1139/cgj-2018-0016).

749

750 Lunne, T., Berre, T., Strandvik, S., 1997. Sample disturbance effects in soft low plastic
751 Norwegian clay, in: Proceedings of the International Symposium on Recent Developments
752 in Soil and Pavement, A.A. Balkema, Leiden, Netherlands. pp. 81–103.

753 Macedo, J., Vergaray, L., 2022. Properties of mine tailings for static liquefaction assess-
754 ment. Canadian Geotechnical Journal 59, 667–687. URL: [https://doi.org/10.1139/](https://doi.org/10.1139/cgj-2020-0600)
755 [cgj-2020-0600](https://doi.org/10.1139/cgj-2020-0600), doi:[10.1139/cgj-2020-0600](https://doi.org/10.1139/cgj-2020-0600).

756 Mori, K., Sakai, K., 2016. The GP sampler: a new innovation in core sampling, in:
757 Lehane, B.M., Acosta-Martínez, H.E., Kelly, R. (Eds.), Geotechnical and Geophysical Site
758 Characterisation 5, Australian Geomechanics Society, Sydney, Australia.

Robertson, P.K., 2016. Cone penetration test (CPT)-based soil behaviour type (SBT) classification system — an update. Canadian Geotechnical Journal 53. URL: [/doi/pdf/10.1139/cgj-2016-0044?download=true](#), doi:10.1139/CGJ-2016-0044.

Robertson, P.K., Woeller, D.J., Finn, W.D., 1992. Seismic cone penetration test for evaluating liquefaction potential under cyclic loading. Canadian Geotechnical Journal 29, 686–695. URL: [/doi/pdf/10.1139/t92-075?download=true](#), doi:10.1139/T92-075.

Sadrekarami, A., Zehforoosh, F., 2025. Effect of Fines Content on the Saturation and the Matric Suction of Oil Sand Tailings | ISS-MGE, in: 4th Pan-American Conference on Unsaturated Soils, Ottawa, Canada. URL: <https://www.issmge.org/publications/publication/effect-of-fines-content-on-the-saturation-and-the-matric-suction-of-oil-sand-tailings> doi:<https://doi.org/10.53243/PANAM-UNSAT-2025-264>.

Sasitharan, S., Robertson, P.K., Sego, D.C., 1994. Sample disturbance from shear wave velocity measurements. Canadian Geotechnical Journal 31, 119–124. URL: [/doi/pdf/10.1139/t94-013?download=true](#), doi:<https://doi.org/10.1139/t94-013>.

Stringer, M., Cubrinovski, M., Bray, J.D., Haycock, I., Beyzaei, C.Z., Jacka, M., Wentz, F.J., 2020. Recommendations for High Quality Field Sampling using the Gel-Push Type S sampler New Zealand Ministry of Business, Innovation and Employment (MBIE) DISCLAIMER. Technical Report.

Stringer, M., Taylor, M.L., Cubrinovski, M., 2015. Advanced Soil Sampling of Silty Sands in Christchurch. URL: <https://api.semanticscholar.org/CorpusID:56321378>.

Taylor, M., Cubrinovski, M., Bradley, B., 2013a. Cyclic strength of Christchurch sands with undisturbed samples, in: New Zealand Society for Earthquake Engineering Annual Conference (NZSEE2013), University of Canterbury. Civil and Natural Resources Engineering, Wellington, New Zealand. URL: <http://hdl.handle.net/10092/8191>.

- 784 Taylor, M., Cubrinovski, M., Bradley, B., Horikoshi, K., 2013b. The undrained cyclic strength
785 of undisturbed and reconstituted Christchurch sands, in: 19th New Zealand Geotechnical
786 Society (NZGS) Symposium, University of Canterbury. Civil and Natural Resources
787 Engineering, Queenstown, New Zealand. URL: <http://hdl.handle.net/10092/9037>.
- 788 Terzaghi, K., Peck, R.B., Mesri, G., 1996. Soil Mechanics in Engineering Practice. WILEY.
- 789 Umehara, Y., Chiaro, G., Kiyota, T., Hosono, Y., Yagiura, Y., Chiba, H., 2015. Effective-
790 ness of 'Gel-Push' Sampling Technique to Retrieve Undisturbed Sandy Specimens for
791 Liquefaction Test.
- 792 Zapata, C., Houston, W., Houston, S., Walsh, K., 2000. Soil–Water Characteristic Curve
793 Variability, in: Geotechnical Special Publication, pp. 84–124. doi:[10.1061/40510\(287\)7](https://doi.org/10.1061/40510(287)7).