Supplementary Information to the paper

Uncertainty-aware sample mass determination for particle size analyses of gravel-dominated soil

Georg H. Erharter[1] [ORCID: 0000-0002-7793-9994]*, Santiago Quinteros[1] [ORCID: 0000-0002-4895-1580], Diana Cordeiro[1] [ORCID: 0000-0002-9242-4147], Matthias Rebhan[2] [ORCID: 0000-0002-0638-6202], Franz Tschuchnigg[2] [ORCID: 0000-0002-4279-7703]

1) Norwegian Geotechnical Institute, Sandakerveien 140, Oslo, Norway
2) Graz University of Technology, Rechbauerstraße 12, Graz, Austria

* corresponding author: georg.erharter@ngi.no

Application example

For example, one wants to determine the PSD of a coarse-grained fluviatile soil with an estimated D_{max} of 150 mm (there are some cobbles) and an estimated D_{90} of 80 mm. According to ISO 17892-4 the required m_{min} is 225 kg of soil (eq. 1) and it is not clear why so much soil would be required. In contrast to that, the newly developed equation allows setting a desired maximum error / sampling confidence (KS_{p95}) of e.g. 10 %. Based on the estimated D_{90} one can then estimate the required sample mass to be ~63 kg with explicit consideration of that desired sampling confidence (eq. 2). If the total available soil sample mass would, however, only be 20 kg, then equation 10 (of the paper) can be used to determine the error exponent ε (eq. 3) which is 1.44. Substituting this into equation 7 (of the paper) reveals that in this particular soil, one needs to expect that the determined PSD has an error of up to ~20% with respect to the real soil's PSD if only 20 kg of soil sample are available (eq. 4).

$$m_{min}[kg] = 225 = \left(\frac{150}{10}\right)^2$$

<div align="right">*eq. 1*</div>

$$m_{min}[kg] = 63 = \left(\frac{80}{10}\right)^{\frac{\ln(10)-\ln(118.11)}{-1.24}}$$

<div align="right">*eq. 2*</div>

$$\varepsilon = 1.44 = \frac{\ln(20)}{\ln(80) - \ln(10)}$$

<div align="right">*eq. 3*</div>

$$KS_{p95}[m\%] = 19.8 = 118.11 * e^{-1.24*1.44}$$

<div align="right">*eq. 4*</div>

Erharter G., Quinteros S., Cordeiro D., Rebhan M., Tschuchnigg F.

Grain size distribution characterization survey

A survey was conducted to investigate how well operators can visually estimate different parameters that describe the geometry of a sieve curve. The survey was done using Microsoft Forms and responses that were submitted between the start of the survey on 25th of November 2024 until its end on the 9th of December 2024 were included in this analysis.

The following metadata was collected from each participant:

- Name

- Email Address

- Main area of expertise, where participants could choose one of the following answers: Geotechnical engineering, Engineering Geology, Sedimentology, Hydrogeology, Quaternary geology, other (to be specified).

- Current main field of work, where participants could choose one of the following answers: Academia, Industry (consulting, contractors, technology development,...), Other

- Years of experience post master, where participants could choose one of the following answers: 0-5, 5-10, 10-20, 20-30, >30, None (still student or not from this field).

After collecting this information, the participants were presented with a series of four synthetic sediment samples that were generated with the code framework of this project that is provided in the supplementary information section of the paper. Each sample shows spherical black grains in a 500 by 500 mm large field on white ground. A measuring scale is given on the border of the field with 50 mm spaced ticks and some reference grains are given below the sample with sizes between Ø=100 to Ø=2mm. The samples are shown in Figure A 1 to Figure A 4.
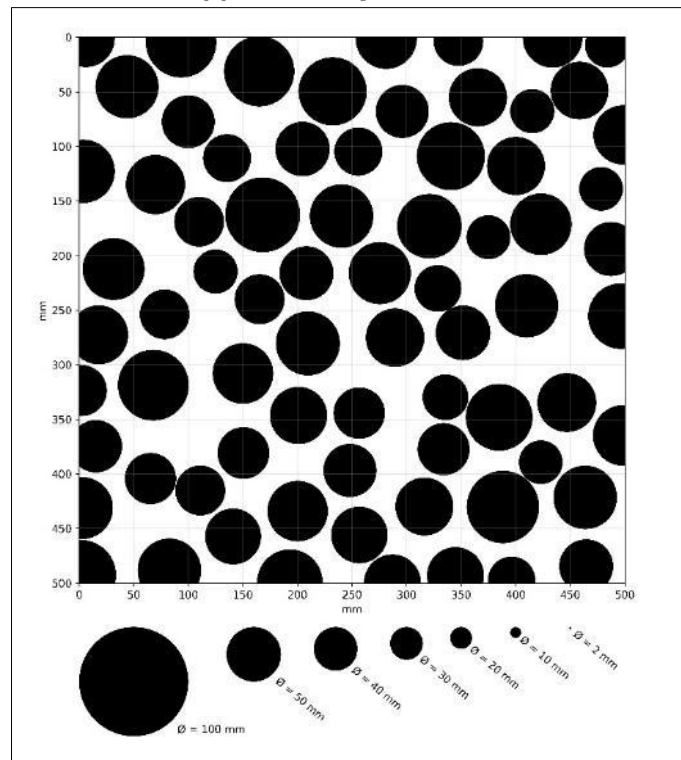
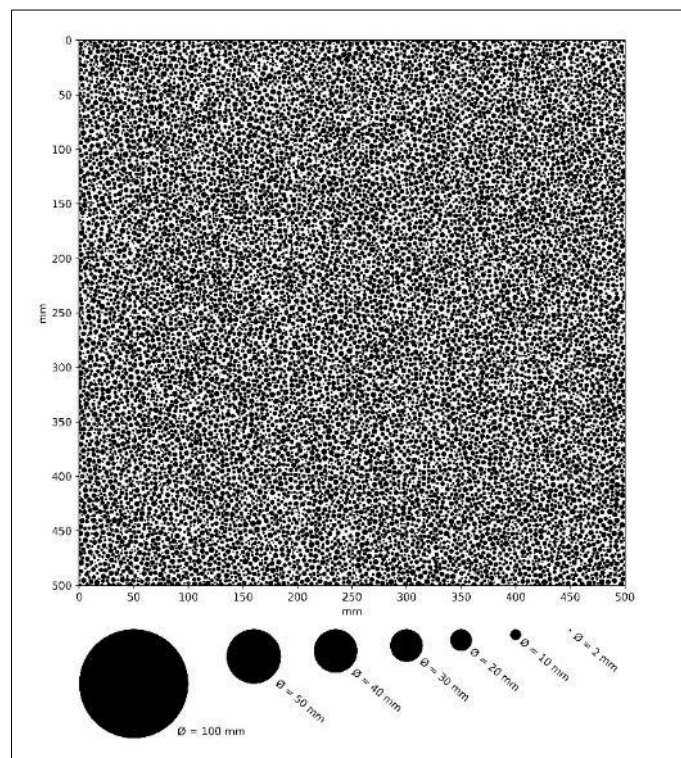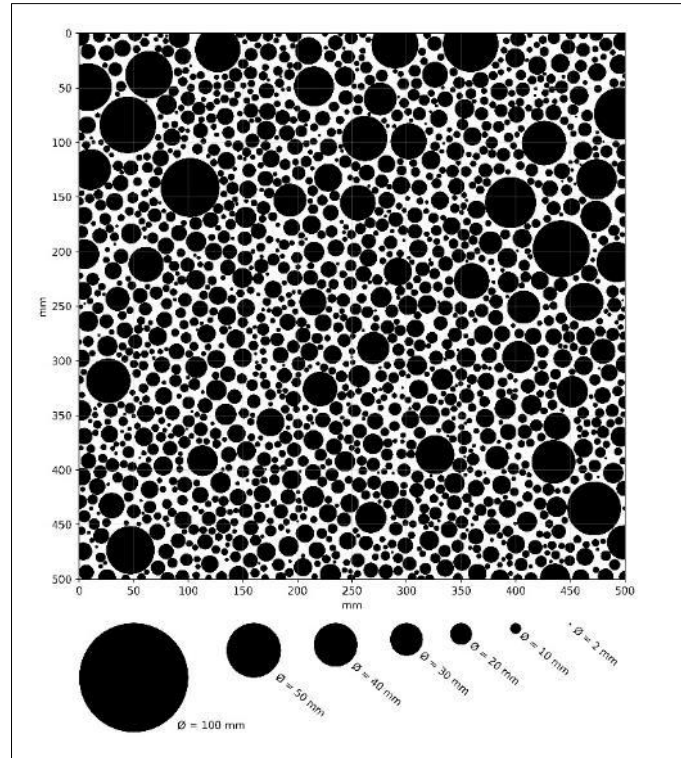*Figure A 1: Sample 1.*
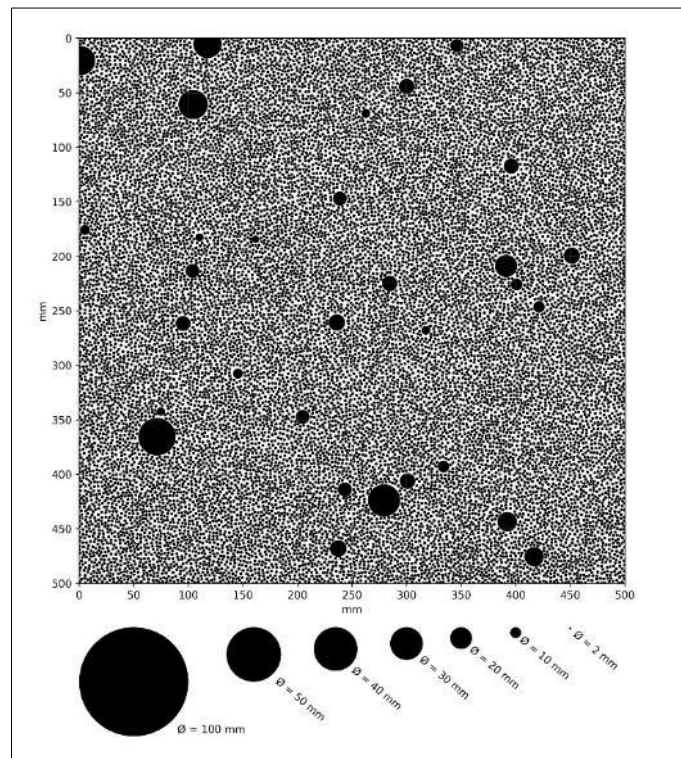

Figure A 2: Sample 2.

Figure A 3: Sample 3.

Figure A 4: Sample 4.

For each sample, the participants were asked to estimate the D_{min}, D_{10}, D_{30}, D_{50}, D_{60}, D_{90} and D_{max}. The participants were told not to be too precise and to take not more than 3 minutes per sample. A total number of 95 responses were collected. From these 95 responses, 14 had to be completely removed because the

participants gave consistently not credible responses that indicated a misunderstanding of the survey (e.g. always the same number, decreasing grain sizes from D_{min} to D_{max}, etc.). Furthermore, single results for samples had to be removed for similar reasons but it can be observed that there are more erroneous submissions for sample 1 than for the others, thus indicating that some participants needed the first sample to get used to the task. After response cleaning, a total of 71, 81, 80 and 80 responses were left for the samples 1-4 respectively. A visualization of the collected participant metainformation is shown in Figure A 5.
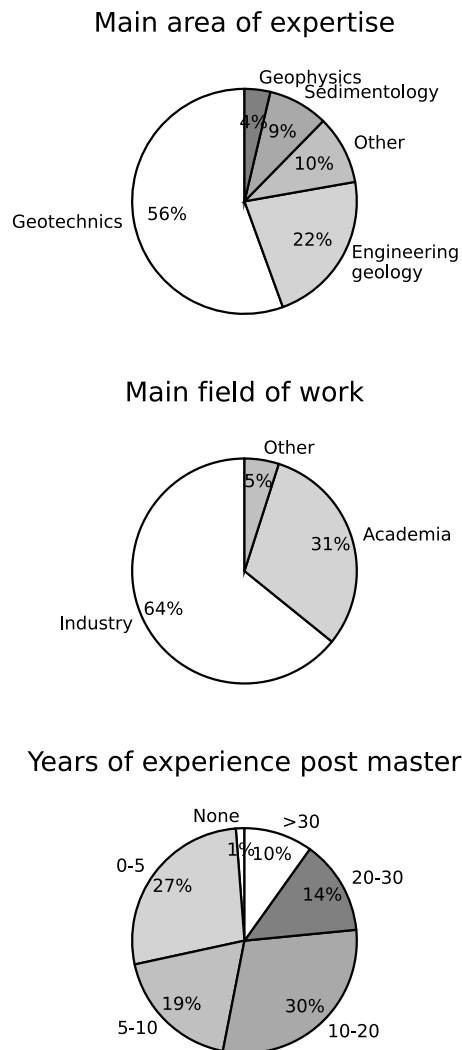
Main area of expertise

Main field of work

Years of experience post master

Figure A 5: Statistics of the metainformation that was collected from the participants in the survey.

A visualization of the participants' responses in relation to the true values (assessed based on the simulated grain distribution) for every sample is given in figure Figure A 6. While the average estimated parameters are close to the true values, it can be seen that all parameters show substantial variability. There are no generally observable trends, and it is not observable that the D_{max} is, for example, significantly easier to assess than other

D-values. The only exception is sample 4 which has a pronounced gap graded distribution, and it is visible that participants alternate between assigning the D_{90} to the small or the large grain sizes. Analyzing these results also must consider the logarithmic scale of the problem where e.g. overestimating the size of a 4 mm grain by 100% is less severe than overestimating the size of a 40 mm grain by 50%.
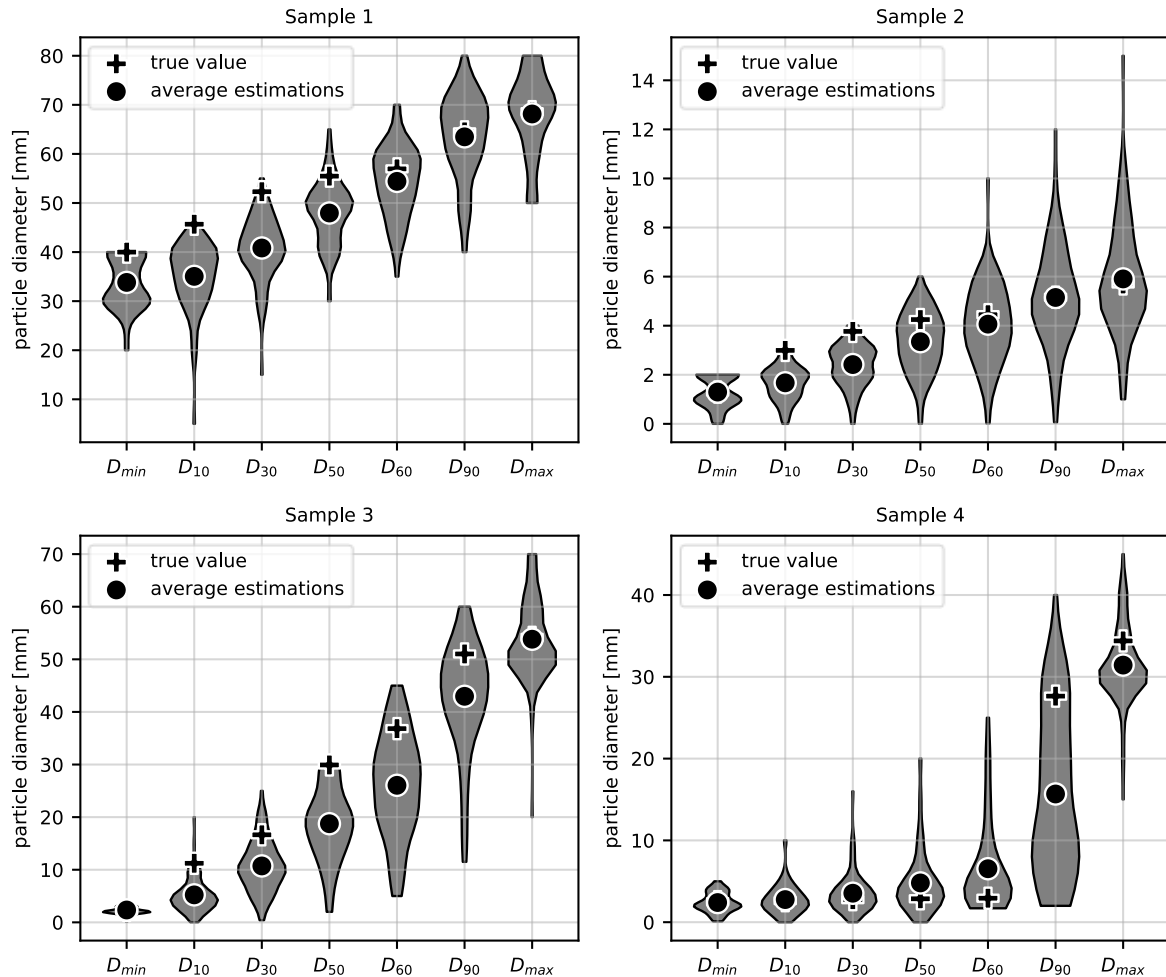

Figure A 6: Results of the survey. The distribution and bandwidth of participants' responses is shown with grey violin plots.

Lastly, the participant assessed values were used to compute C_u and C_c for the samples and their respective distribution based on the participants feedback variability (Figure A 7). It can be seen that the variability for these computed values is substantial but it also must be considered that these are calculated values and not directly estimated values. The ground truth values for the parameters under investigation of the survey are given in Table A 1.
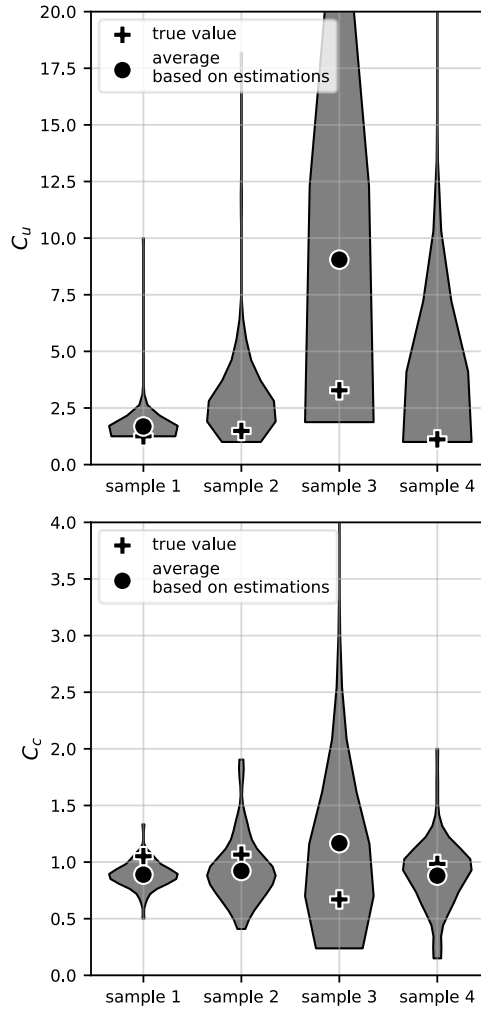
Figure A 7: Variability of C_u and C_c computed from the participants responses.

Table A 1: Ground truth values for the parameters of the survey.

Sample	D_{min} [mm]	D_{10} [mm]	D_{30} [mm]	D_{50} [mm]	D_{60} [mm]	D_{90} [mm]	D_{max} [mm]	C_u	C_c
1	40.0	45.7	52.3	55.5	57.0	64.5	68.6	1.2	1.1
2	1.3	3.0	3.8	4.2	4.4	5.2	5.7	1.5	1.1
3	2.3	11.2	16.6	29.9	36.8	51.0	54.2	3.3	0.7
4	2.5	2.6	2.8	2.9	2.9	27.6	34.4	1.1	1.0