Draft benchmark description: FluidFlower international benchmark study

Jan M. Nordbotten[1], Martin Fernø[2], Bernd Flemisch[3], Ruben Juanes[4], Magne Jørgensen[5]

--- DRAFT VERSION 2.1 ---

Abstract

Numerical benchmark studies of multi-phase flow in porous media generally lack a clearly defined ground truth. In this benchmark study, we consider a problem inspired by CO_2 storage, where the dominant processes are associated with multiphase flows, capillarity, dissolution, and convective mixing. Supporting the numerical benchmark study, we conduct a series of physical realizations in the FluidFlower experimental setup.

1. Overview and vision

The group of Professor Fernø at the University of Bergen has constructed a new experimental rig, named "FluidFlower", as shown in Figure 1. The rig allows for relatively large-scale multiphase 2D flow experiments (almost 3 by 2 meters) on model geological geometries, with an unprecedented data acquisition in terms of a pressure sensor network (seen in the background) together with optical registration of both saturation and acidity (pH). Moreover, the design allows for repeated experiments within the exact same experimental setup, allowing physical uncertainty to be addressed (see also: https://fluidflower.w.uib.no/).

[1] Department of Mathematics, University of Bergen
[2] Department of Physics and Technology, University of Bergen
[3] Institut für Wasser- und Umweltsystemmodellierung, Universität Stuttgart
[4] Department of Civil and Environmental Engineering, Massachusetts Institute of Technology,
[5] Simula Research Laboratory

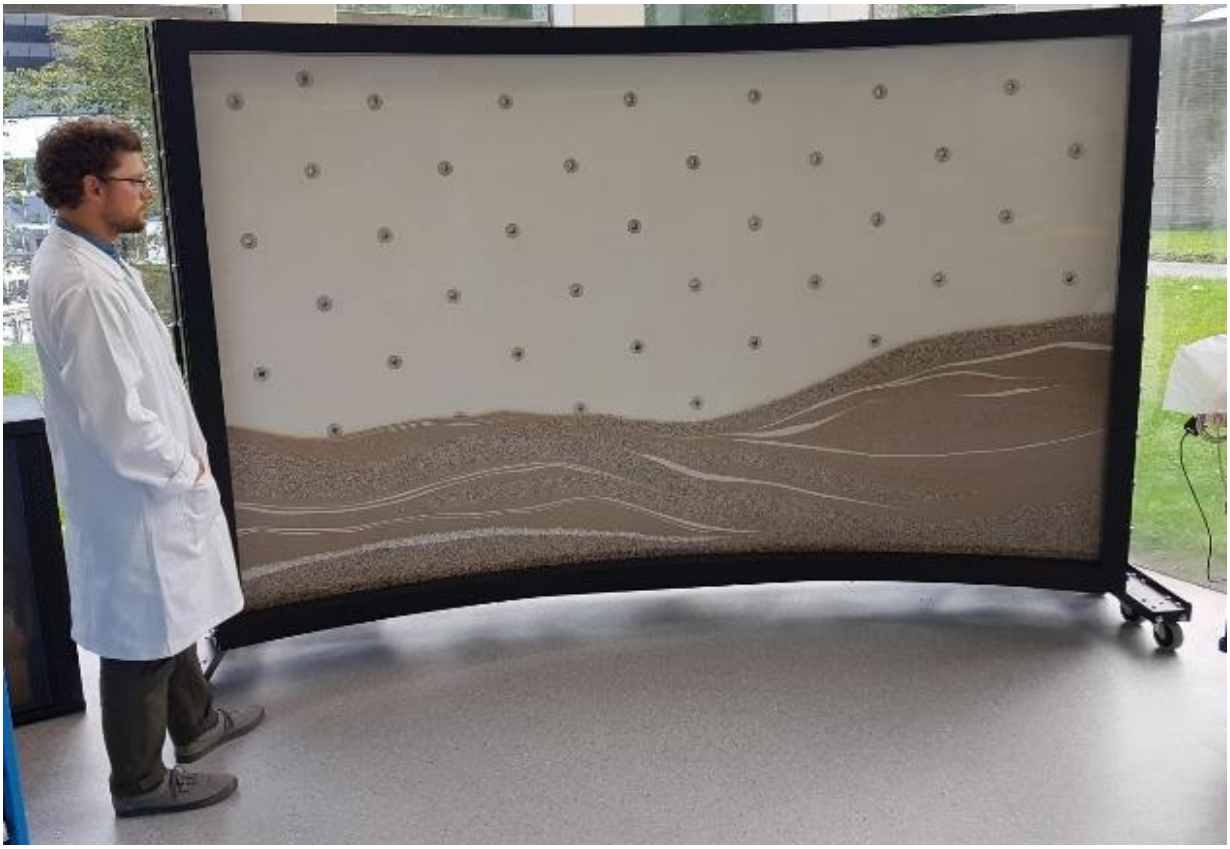


Figure 1. Full-scale FluidFlower rig, half-filled with geological layers (lower part). In the upper part, which is not yet filled in this photo, the pressure sensor network is visible.

1.1 Benchmark study

To leverage these new experimental capabilities, we are conducting this international benchmarking study for modeling, simulation and forecasting of multiphase, multicomponent flows in porous media. Several multiphase flow phenomena are present in the experimental setup, including both heterogeneous and hysteretic two-phase flow properties (leading to both structural and residual gas trapping) and two-component mixing with concomitant development of gravity fingers, as illustrated in a sample geometry in Figure 2.

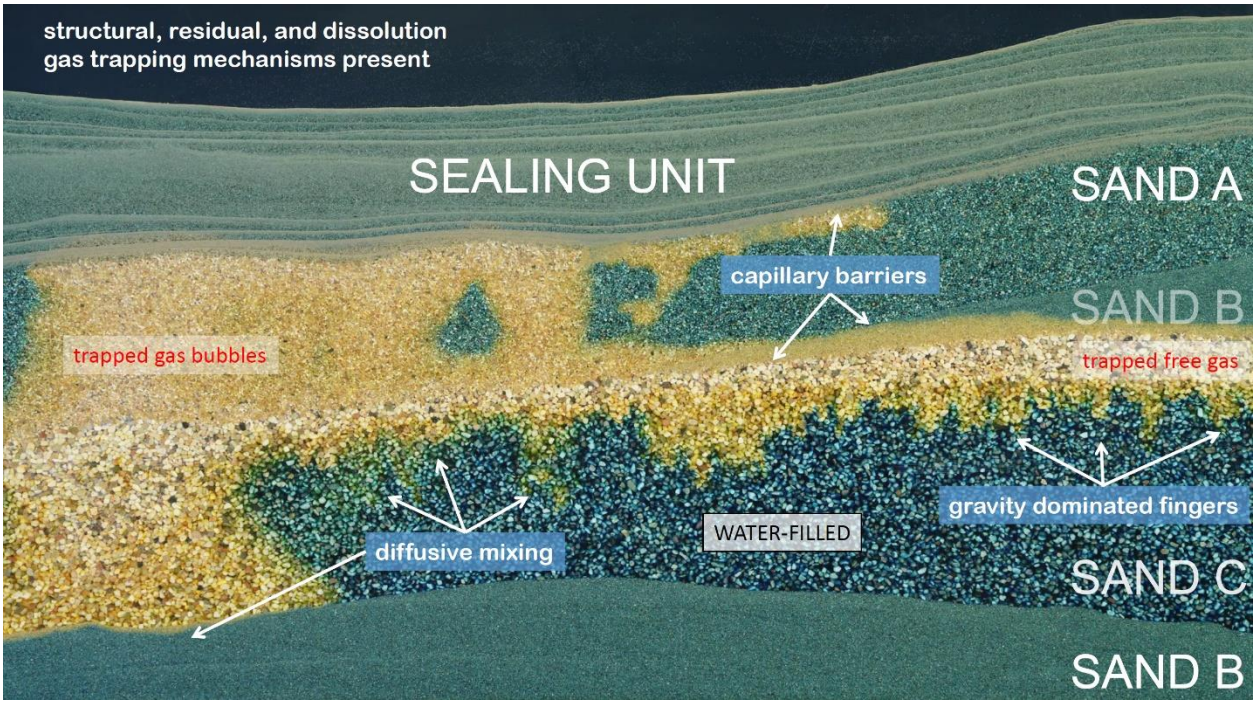


Figure 2: Sample experiment in a downscaled FluidFlower rig. Visible here are multiple fluid-filled gas layers, free-phase CO_2, and variable concentrations of dissolved CO_2 in the water phase. Important processes for CO_2 storage, such as capillary entry effects, hysteresis, gravity fingering, are all apparent.

The main goal of the benchmark is to provide a full-physics validation of the state-of-the-art simulation capabilities within the international porous media community. To this aim we envision the following benchmarking process:

International numerical benchmark study:

- A full geometric, operational, and petrophysical description is hereby[6] provided to all participants, including commonly measured multiphase flow parameters for porous media simulation based on *ex situ* measurements (grain size distributions, porosities, permeabilities, end-point relative permeabilities, entry pressures). All parameters are given with our best estimates of tolerances.
- Additionally, experimental single-phase flow data (well tests and tracer tests) are provided[7] to the participants for model calibration.
- Participants must themselves provide equation-of-state data, based on the liquid and gas used.
- All participants are asked to provide simulation output in terms of target quantities, together with both quantitative and qualitative confidence intervals, according to a common protocol as specified in Section 3.

Physical experiments:

[6] In this draft version 2.0, some quantities are preliminary, or tabulated from literature. These are marked with purple color, and will be replaced by measured values in the final version of the benchmark description.
[7] The single phase flow data will be available in the final version of the benchmark description.

- On the exact geometry provided to the participants, we will at the UoB conduct at least 5 identical multi-phase flow experiments.

In-house parameter estimation at UoB and MIT:

- Complementing the benchmarking study, UoB will conduct an in-house numerical parameter estimation study based on the full set of multi-phase flow experiments.
- Also complementing the benchmarking study, MIT and UoB will collaborate on assessing the value of using analog system (here in the form of a smaller version of the FluidFlower) for calibrating numerical simulations.

Together, the physical experiments and numerical simulation will allow us to address the confidence with which the main physical processes associated with subsurface CO_2 storage can be modeled:

A) The physical experiments will provide both a "ground truth", and will also allow us to quantify to some extent the inherent uncertainties associated with the porous media flow itself by means of repeated experiments on the same geometry.
B) The in-house parameter estimation study will allow us to address any discrepancies between *ex situ* measured parameters and *in situ* properties, as well as to provide a sense of the extent the standard model equations for porous media can capture the observed flows.
C) The numerical simulations reported by each research group will allow us to quantify the precision and intra-group uncertainty as assessed by each of the research groups individually.
D) The total panel of research groups will allow us to quantify the inter-group variability in the results (and thus to some extent the reproducibility of numerical modeling), and the relationship between the ensemble precision and uncertainty.

The combination of A), B), C) and D) within the same benchmarking study is to our knowledge unprecedented within the porous media community. This will allow us, within the context of this setup, to address for the first time the correlation between the numerical uncertainty quantification and a measure of real system uncertainty.

1.2 Participants

Following a call for participation, the following 10 research groups have tentatively joined the benchmark study:

Institution	Contact person 1	Contact person 2
CSIRO	Jonathan Ennis-King	
Herriot-Watt	Florian Doster	Sebastian Geiger
Imperial College	Matthew Jackson	
LANL	Satish Karra	Viswanathan, Hari S
Melbourne	Stephan Matthai	
Stanford	Hamdi Tchelepi	
Stuttgart	Holger Class	Dennis Gläser
TU Delft	Denis Voskov	Hadi Hajibeygi
UT Austin	Nick Espinosa	Mary Wheeler

In addition, Equinor ASA (represented by Ali Mojaddam Zadeh and Bamshad Nazarian), have expressed interest in participating in the benchmark through their collaboration with LANL.

1.3 Time frame

- 30. April, 2021: Closed call for participation opens.
- 15. June, 2021: Call closes.
- 15. July, 2021: Preliminary benchmark description supplied to participants.
- 16. July – 19. August, 2021: Preparation phase, discussion possible among all participants and the experimental team.
- 20. August, 2021: Deadline for feedback on preliminary benchmark description.
- 16. September, 2021: Kick-off Zoom meeting, second iteration of benchmark description distributed.
- 17. September – 14. November, 2021: Blind phase, no direct communication between different participants or with the experimental group.
- 30. September, 2021: Submission of signed participation agreements.
- 1. October, 2021: Final benchmark description circulated to participants.
- 14. November, 2021: **Deadline for submitting blind benchmark data**.
- 17. November, 2021: Virtual workshop and comparison of "fully blind" simulation forecasts.
- 16. November 2021 – 31. January 2022: Synchronization phase, communication between all participants enabled, but not with the experimental group.
- 27. January, 2022: **Deadline for submitting final benchmark data**.
- 1-2 February, 2022: Real workshop in Norway (location to be determined) with presentation of final simulation forecasts, experimental results, model calibration study, and synthesis of results.
- Spring 2022: Writing and submitting papers based on the findings of the study.

2. Benchmark description

The geometric description is motivated by typical North Sea reservoirs, and has been developed in consultation with faculty and researchers at the Department of Geology, UoB. *A special thanks goes to Robert Gawthorpe, Atle Rotevatn and Casey Nixon for their helpful comments*. The physical and petrophysical properties are based on unconsolidated sands, as measured by the group of Professor Martin Fernø at the University of Bergen.

2.1 External Geometry

The porous media in the FluidFlower is contained between an optically transparent front panel and a sealed back panel with perforations for fluid injection/production and pressure measurements. The length of the porous media is 2840 mm and height is 1300 ± 30 mm (note large ± due to non-horizontal top porous media surface, see Figure 6 for exact height. *Note that the visible length of the porous media is 2800 mm, but the active porous media extends 3 cm behind black metal frame on each side*). The porous media is curved (both the front and back panels are curved to be able to sustain internal forces of sand + liquids); the curvature is 1/8 of a circle with radius 3.6 m. The nominal width of the porous media between the front and back panel is 19 mm, but varies both from left-to-right and top-to-bottom due to internal forces when filled with sand and water. Width at each side (far left and far right) is fixed at 19 mm, and increases towards the center to 28 mm, see Figure 3. Note that the widths have been measured after initial sand filling of the rig by removing the front panel. The height of the

coarse-grained sand used (not used in benchmark) was measured point-wise along the left-to-right and top-to-bottom axes.

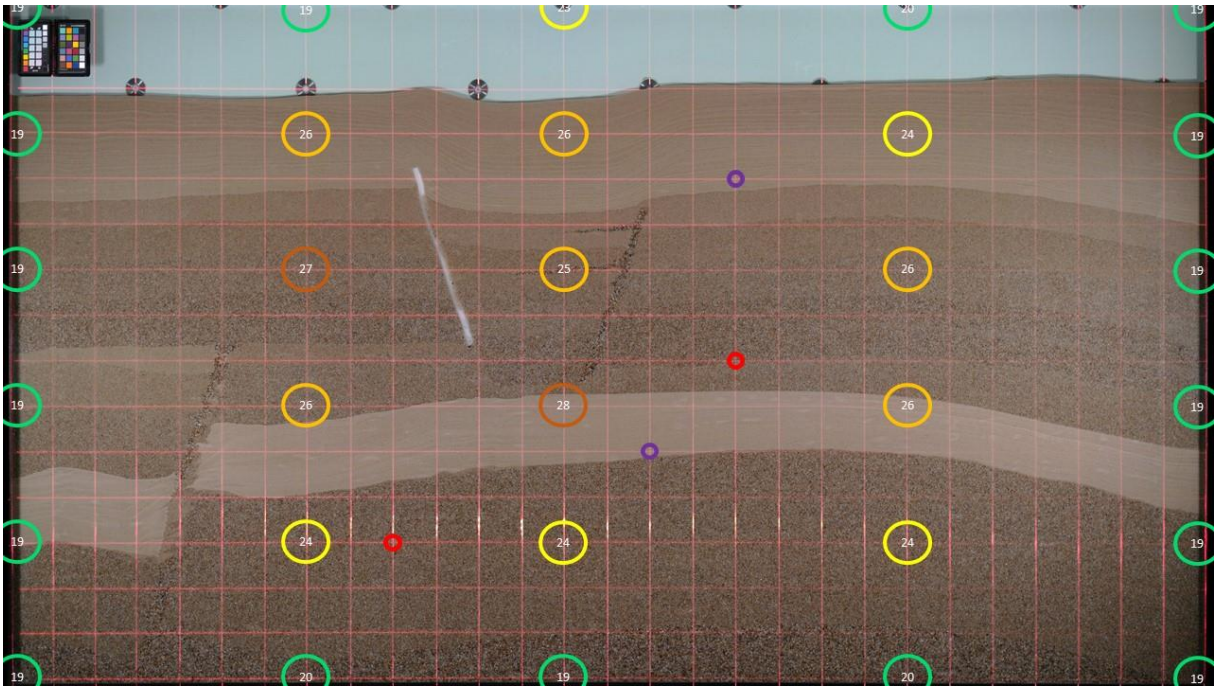


Figure 3: Porous media width variation (in mm) within the curved flow rig, using an image corrected curvature. The width at each side (far left and far right) is fixed at 19 mm. The laser grid (red squares) is 100 x 100 mm. Injection ports (in red circles) and pressure measurement ports (purple) are described in more detail below.

The FluidFlower has 56 perforations (see Figure 4 for description) for fluid injection and pressure measurements. During the benchmark study there will be two injection ports and two pressure ports. The remaining perforations will be closed (some are permanently plugged, some are able to serve as additional pressure/injection/outlet ports). The four open ports (pressure measurements and fluid injection) have a nominal outer diameter of 1/8 in (3.2 mm) and inner diameter of 1.8 mm. The two injection ports extend 10 mm out from the back panel into the sand to reduce gas flow along the back. All other ports (except bottom row) are flush with the back panel, and are sealed with a custom-made plastic (POM) port. A thin layer of silicone is used to maintain the seal to avoid leakage. Upon injection of CO_2 gas, excess fluids will be produced from the top row from several ports with 10+ mm diameter. The bottom row (blue ports, see Figure 4) are technical ports for re-setting fluids between CO_2 injection tests, and will remain closed during CO_2 injections. Each of the seven blue ports have stainless steel tubes (same inner and outer diameter as above) that point downwards into the lowest (high-permeable) sand layer and stop approximately 5 mm above the bottom of the porous media. The steel tubes extend 8 mm out from the back panel, and the space between the tube and back panel is filled with silicone.

The FluidFlower boundaries are closed at the bottom and both left and right sides (no-flow boundaries). The top is open and in contact with (fluctuating) atmospheric pressure, with a free water table fixed at a constant elevation (constant hydraulic head) of 1500 mm above the bottom. Ports in the top row are alternating production/injection ports to ensure a fixed water table.

2.2 Geology

Figure 4 provides a sketch of the intended geometry, whereas Figure 5 is the actual geometry. To avoid discrepancies between the intended and actual geometries, the actual geometry is provided as a high-resolution photograph, shown here in reduced resolution as Figure 5.

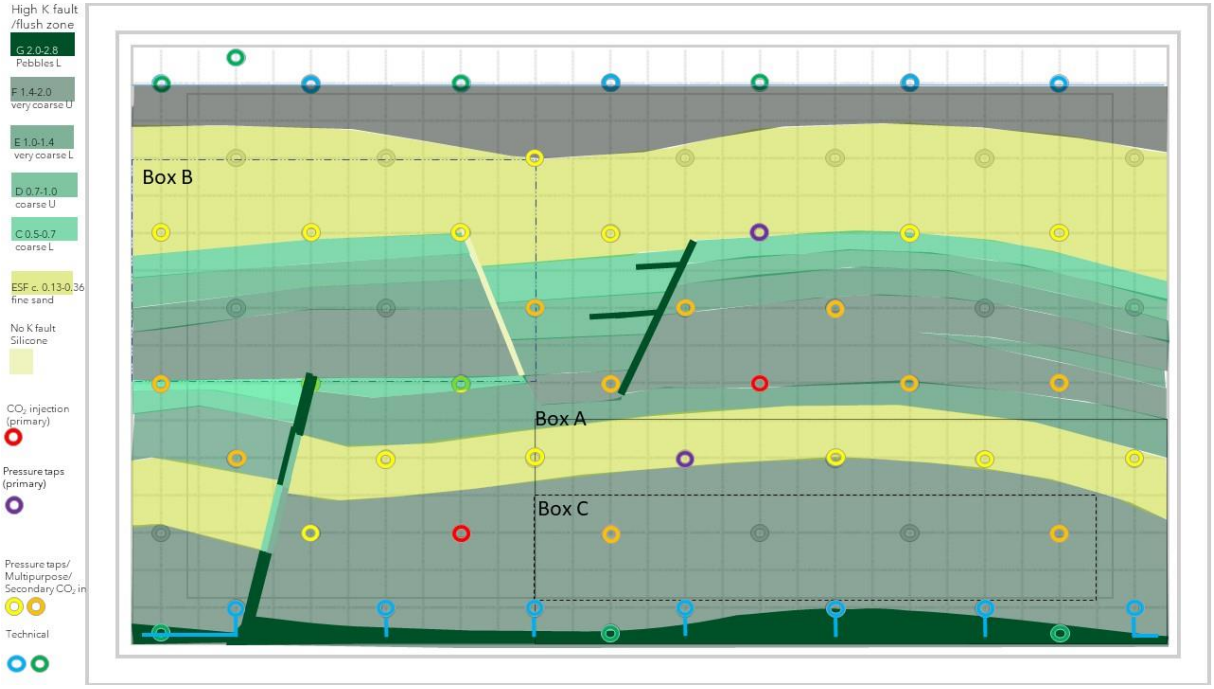


Figure 4: A sketch of the intended benchmark geometry. The geometry includes an anticline (right side) where CO_2 accumulation is anticipated. There are three fault-like structures in the geometry, with different permeability (two high and one low zero). The lower fault (left side) is ”heterogeneous” and consists of several sand types. This approach is used to reduce smaller grained sands leaking into large-grained sands. The upper, left fault is sealed (using silicone) and may be regarded as a sealing fault. The upper right fault is high-permeable and built using a single sand type. There are two CO_2 injection points (red circles) and two pressure taps (purple circles). Box A, B and C are indicated in the figure, and described in more detail below. Each gray square in the grid in the background is 10 x 10 cm. *A high-resolution version of sketch is available for the benchmark study, and is available on GitHub (https://github.com/fluidflower).*

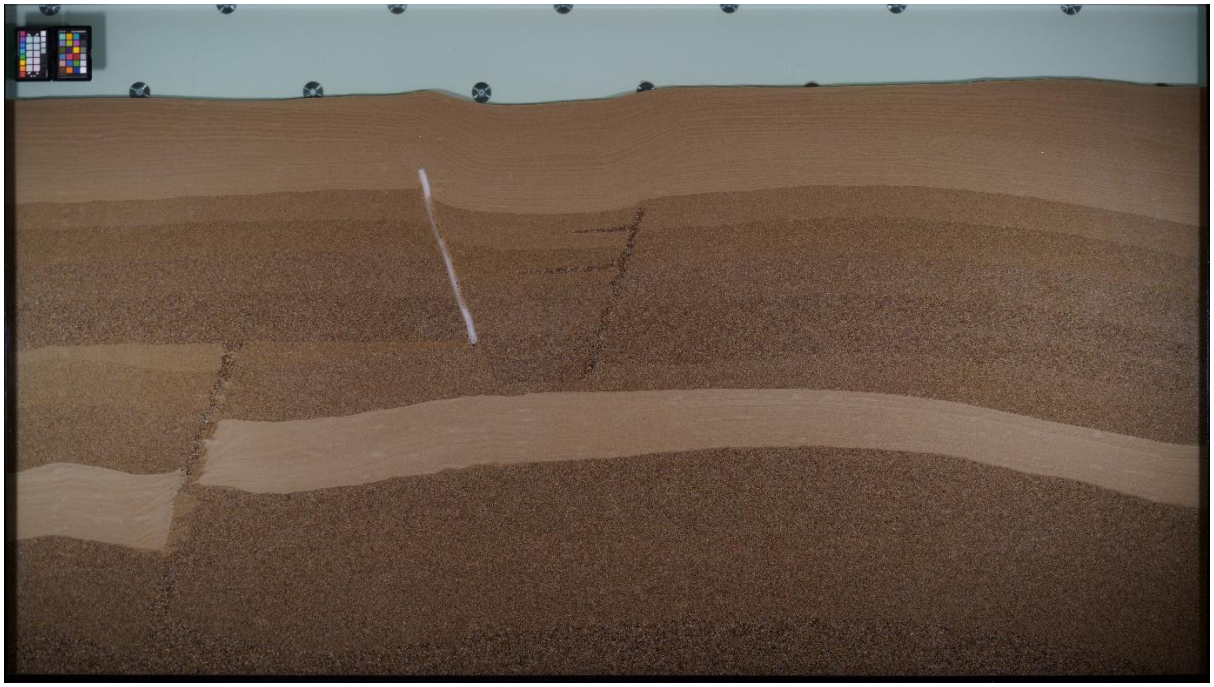


Figure 5: A photograph image of the actual benchmark geometry in the FluidFlower rig. The image is corrected for curvature using Matlab. *The high-resolution version of this data is considered the most accurate geological description available for the benchmark study, and will be published on* (https://github.com/fluidflower).

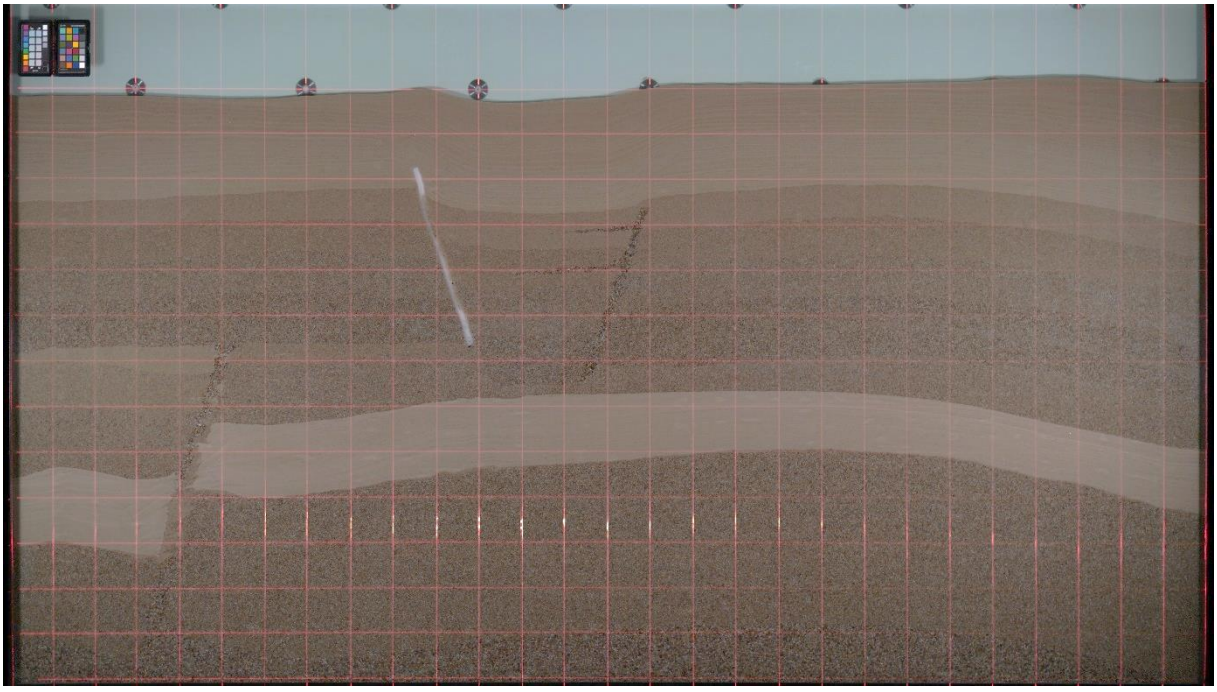


Figure 6: A photograph image of the benchmark geometry (identical to Figure 5, expect brightness and contrast) with laser grid (each square is 100 x 100 mm). The image is corrected for curvature using Matlab. LINK?

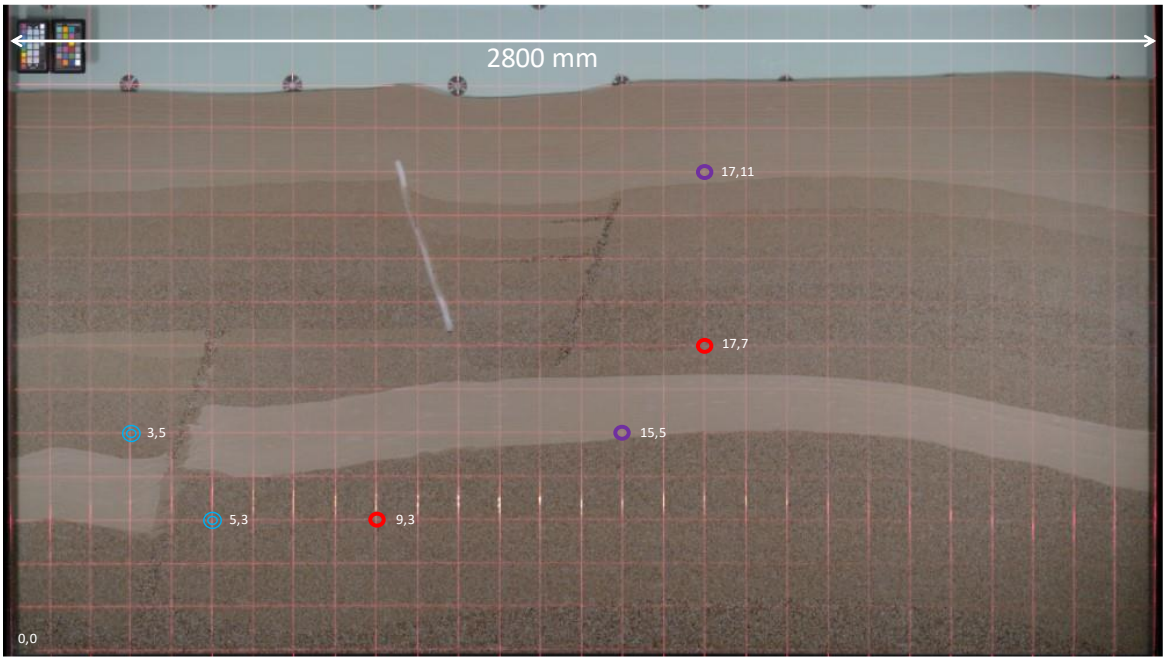


Figure 7: Port locations on a laser grid (each square is 100 x 100 mm). Horizontal length between far left and far right vertical lasers is 2800 mm.
Note 1 *The visible length of the porous media is 2800 mm, but the active porous media extends 3 cm behind the black metal frame on each side.*
Note 2 *Shadows from the metal frame present on the left and right edge. No shadow at the bottom, but porous media extends 3 cm below bottom laser line.*
Note 3. *Port locations (circles) indicated. Color code: red is injection points, purple is pressure ports and open blue circles are additional well injection tests in the vicinity of the fault.*

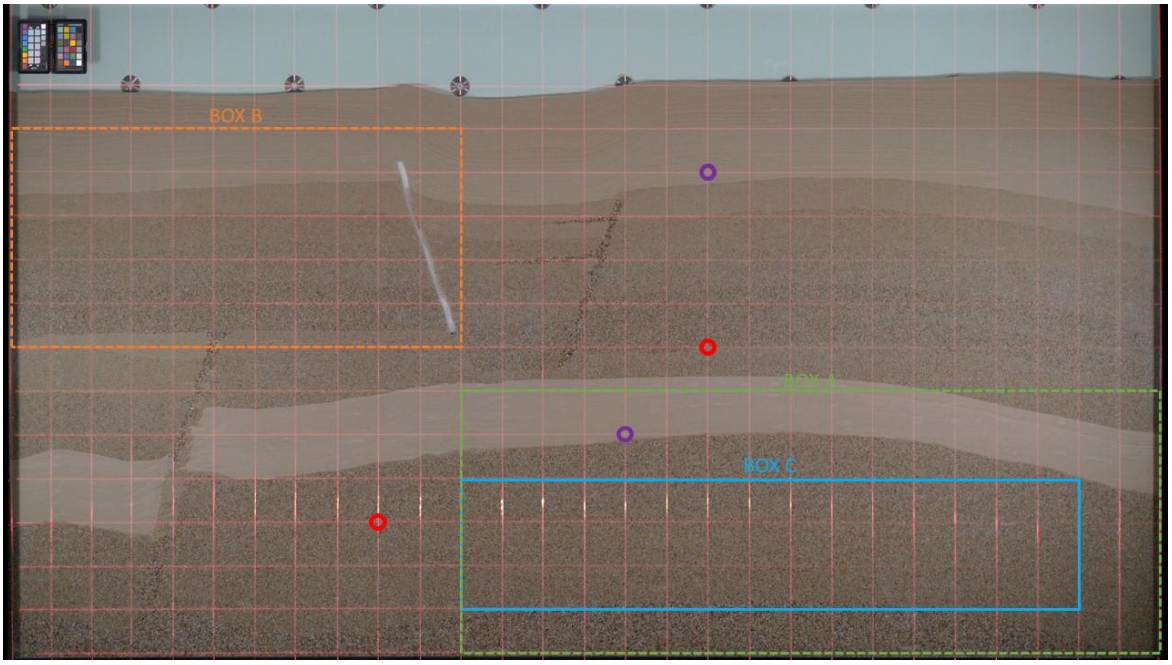


Figure 8: Box coordinates. Top right (TR), top left (TL), bottom right (BR), bottom left (BL)
Box A: TL (11,6) -> TR (28,6); BL (11,0) -> BR (28,0)
Box B: TL (0,12) ->TR (11, 12); BL (0,7) -> BR (11,7)
Box C: TL (11, 4) ->TR (26,4); BL (11,1) ->BR (26,1)

2.2.1 Sedimentation process

The dry, unconsolidated sand types were separately poured from the top into the water-filled (distilled water) void between the front and back panels. Each sand layer was built from the bottom and upwards, and features (predominantly change in dipping angle) were created by manipulating the layer during pouring, using guiding Styrofoam rectangles, funnels and plastic hoses. Mechanical manipulation (raking/scratching) was kept to a minimum and only in some areas in the vicinity of the faults. Faults (high permeable narrow bands of sand, at a steep angle + one sealed fault with silicone) were constructed though an iterative process, explained more detail below.

The hydrostatic pressure during geometry assembly was 10 cm higher than at operating conditions. Hence, once the geometry was constructed, the water-level in the FluidFlower was lowered to operating water-level (kept constant during all injections). This led to an initial, pre-injection compaction of the sand. Further compaction was observed during subsequent days due to temperature fluctuations in the room where the rig is located, and during water (degassed distilled water) injection in all injection ports and technical ports in the bottom row.

2.2.2 Pre-injection flushing

Prior to CO_2 injection, the geometry is flushed with different aqueous fluids to further clean the sand and prepare for CO_2 injection experiments. The injection rates used (and associated pressures) are higher than during CO_2 injections. Hence, the geometry has experienced multiple flushing sequences, to satisfy compaction. Possible further compaction due to daily temperature fluctuations may occur, and will be documented. All flushing sequences will be monitored with time-lapsed imaging, using an image resolution capable of detecting single sand grains (sands E to G in Table 1 below).

2.3 Facies properties

The data for all sand types are reported in detail in Appendix A. Below follows a general description.
Sand labeled “Danish quartz sand” was purchased (totally 3.5 tons) from a commercial Norwegian supplier. Sand types used in the benchmark study were sieved from the supplied sand stock by the experimental team in Bergen - then washed (tap water) and acid washed until carbonate impurities were removed (time for each acid wash varied between 24 h to 72 h for the different sands type). The sands were then dried at 60 ºC for at least 24 hours, and stored in 15 L plastic containers (washed and dried) until use. The different sand grain distributions are detailed in Table 1, along with sieve sizes, classification scheme. More information of each sand type in Appendix A.

Table 1: Six different sand grain distributions and corresponding to the coloring in Figure 4.

	Sieved grain size range [mm]	Grade	Class
ESF	*	Fine	Sand
C	0.5 - 0.71	Coarse (lower)	
D	0.71 - 1.0	Coarse (upper)	
E	1.0 - 1.41	Very coarse (lower)	
F	1.41 - 2.0	Very coarse (upper)	
G	2.0 - 2.8	Granules	Gravel

** ESF sand not sieved (only washed, acid washed and dried) to maintain fine particles. Supplier provided 0.13 – 0.36 mm range; this was not accurate and grain sizes below and above cited range was observed.*

The absolute permeability of each sand type listed in Table 1 was measured prior to filling the FluidFlower rig. Measurements were performed using a glass cylinder (nominal dimensions: length 450 mm, diameter 20 mm). Each sand type was poured into a vertically positioned glass tube filled with DI water with bottom endpiece attached. Care was taken to not assert force to the sand when attaching the top endpiece to avoid compaction beyond what would be expected in the FluidFlower rig. The glass tube was then placed horizontally and flushed overnight with degassed DI water at a high rate to remove potential trapped air bubbles during assembly. Two absolute pressure transducers were placed at inlet and outlet, in addition to a differential pressure transducer. A fourth pressure transducer recorded changes in atmospheric pressure and temperature fluctuations. The permeability was measured in both directions, using > 5 cycles of 6 ascending and descending constant volumetric injection rates until a stable differential pressure was achieved. The flow rate was then reversed and the 7 cycles repeated. The sand was then removed from the glass tube, weighted and dried at 60ºC for several days before repeating a second round of measurements. Average permeability for each sand type with standard deviation will be reported, along with measured porosity values.

The unsteady-state end-point relative permeabilities during both drainage and imbibition will be measured using CO_2-saturated water and CO_2. The average end-point fluid saturation will be measured from volumetric and weight measurements and reported with uncertainty.

The capillary entry pressure to gas will be reported for each sand type, based on gas column break-through experiments in a down-scaled FluidFlower under equivalent sedimentary protocol as for the experiment. The porous media will be filled with pH-sensitive dye and CO_2 flow will be monitored with time-lapsed imaging to detect the maximum gas column height under the sealing layer (ESF sand, see Table 1) and between sand types. The gas column height will be converted to pressure to provide the entry pressure for each sand type. Several measurements for each sand are performed, providing the entry pressure value with standard deviation.

2.4 Fault Facies

The upper, high permeable fault (see Figure 4) was created with sand G (Table 1); the upper left fault is sealing (with silicone) and the lower left fault is characterized as heterogeneous. The faults (expect the sealing one) were created by using Styrofoam shapes, funnels and plastic hoses as mentioned above. Fault sand(s) were filled simultaneously as the adjacent layers, with the sand layer within the fault always above (in the order of 5 cm) the adjacent layer. The Styrofoam shapes were lifted step-wise during filling, and sand in the adjacent layer moved next to the fault sand.

2.5 Fluid properties

The following fluids are used:

- Water composition: distilled water + pH sensitive salts
- Gaseous CO_2: high purity (99.999% - 5.0 purity)

All experiments are conducted at room temperature (fluctuating, will be measured and reported, aim for approximately 20°C) and atmospheric pressure at the free water table at the top of the experiment. We will attempt to keep temperature fluctuations limited, however due to the available laboratory space, a constant temperature cannot be guaranteed. Variations in atmospheric pressure and temperature will be supplied to the benchmark participants at the completion of the experiments.

2.6 Operational conditions

The geometry will initially be 100% water-filled. Gaseous CO_2 will be injected into the lower injection point (see Figure 3) with a constant mass flow (actual flow rate to be specified; unit standard milliliters per minute – sccm, will be specified). After some delay (to be specified, but on the order of a few hours), CO_2 will be injected into the upper injection point with a constant mass flow.

2.7 Well test data

To allow for calibration of the numerical models, and in line with real field operations, we provide 2 sets of calibration data.

2.7.1 Pressure response

We will supply well-test data associated with active injection wells (the two CO_2 injection points) used in this benchmark. For each of the well tests, one well will be rate-controlled, and pressure data will be measured at the other injection point, as well as the two pressure measurement points. Data will be supplied as a text-file with the following data-structure:

Time-stamp	Injection point ID	Rate (injector)	Pressure 1	Pressure 2	Pressure 3

In addition, we will perform well test as described above in vicinity of the lower left fault using ports in the reservoir sands. Hence, we will measure pressure response across the fault to below and above the main sealing unit in the bottom of the model.

2.7.2 Tracer flow

We will supply one set of tracer flow data. For the tracer flow data, tracer (to be specified) will be injected using the same operational conditions as set out in Section 2.6. Data will be supplied as image files.

2.8 Measurables

The following measurables shall be reported by the experiment, and predicted by the benchmark participants, details of which, including time resolution, are described in Section 3.

2.8.1 Pressure

Pressure shall be reported at each of the two pressure sensors in [N/m^2].

2.8.2 Phase composition

The distribution of the CO_2-phase shall be reported in boxes labeled A and B in Figure 4. The phase distribution, in $\mathrm{kg/m^2}$, shall be reported in the following categories: Mobile free phase (CO_2 at saturations for which the relative permeability exceeds 0), immobile free phase (CO_2 at saturations for which the relative permeability equals 0), dissolved (CO_2 in water phase) and seal (CO_2 in any form in

yellow sand regions in Figure 4). The sum of the three first categories, mobile, immobile and dissolved, shall equal the total mass of CO_2 per area in the box.

2.8.3 Convection

For the box labeled Box C in Figure 3, the convective mixing shall be reported as the integral of the magnitude of the gradient in relative concentration of dissolved CO_2. In other words, if mass-fraction of CO_2 in water is denoted χ_c^w, and the dissolution limit is denoted $\chi_{c,\max}^w$, then the following quantity M shall be reported

$$M(t) \equiv \int_C \left| \nabla \left(\frac{\chi_c^w}{\chi_{c,\max}^w} \right) \right| dx$$

3. Data reporting and interaction

All result data will be uploaded by the participants to Git repositories within the GitHub organization https://github.com/fluidflower. Each participant will get write access to a dedicated repository named after his/her institution, e.g., https://github.com/fluidflower/stuttgart. During the blind phase, only the participants themselves will have access to their respective repositories. For the synchronization phase, read access to all participant repositories will be granted for all participants. After the workshop in February, the repositories will be further opened to include also the results from the physical experiments. Upon submission of the papers, the relevant repositories will be turned public.

The reported data will be analyzed in two respects: Both in terms of an intercomparison of numerical simulation capability, but also in terms of our ability to correctly assess key properties of the system. Consequently, we establish both a dense and a sparse reporting protocol.

3.1 Dense data – basis for numerical intercomparison study

The reported dense data should be the best numerical data available, in the sense of finest grid resolution, most representative physics, and so forth.

All measurables identified in section 2.7 shall be reported at 10-minute intervals starting at the initial injection and lasting 120 hours. The data is expected in csv format in the repository in a file `time_series.csv` of the form

```
# t, p_1, p_2, mob_A, imm_A, diss_A, seal_A, <same for B>, M_C
0.000e+00, 1.234e+56, 1.234e+56, <…>
6.000e+02, 1.234e+56, 1.234e+56, <…>
...
```

according to the measurables defined in section 2.7.

Additionally, a map of the phase compositions, as detailed in Section 2.7.2, as shall be reported for each 24 hours after injection starts. In order to facilitate comparisons, these maps shall be reported on a uniform Cartesian grid of 284 by 120 cells (1 cm by 1 cm grid cells from the bottom of the domain). For each temporal snapshot indicated by X hours, X = 24, 48, …, cell values should be provided in csv format in a file `spatial_map_<X>h.csv` of the form

```
# x, y, saturation, concentration
5.000e-03, 5.000e-03, 1.234e+56, 1.234e+56
1.500e-02, 5.000e-03, 1.234e+56, 1.234e+56
…
2.835e+00, 5.000e-03, 1.234e+56, 1.234e+56
5.000e-03, 1.500e-02, 1.234e+56, 1.234e+56
1.500e-02, 1.500e-02, 1.234e+56, 1.234e+56
… …
2.825e+00, 1.645e+00, 1.234e+56, 1.234e+56
2.835e+00, 1.645e+00, 1.234e+56, 1.234e+56
```

The origin of the coordinate system should be located in the lower left corner with the x-axis positively oriented towards the right and the y-axis positively oriented towards the top.

3.2 Sparse data – basis for predictive capability assessments

The reported data should be given to the best ability of each research groups, and need not be direct simulation output. Indeed, it is permitted (encouraged) to use one's own experience to temper the numerical predictions, as if this was an analysis of a real field operation.

The following sparse data is requested:

1. *As a proxy for assessing risk of mechanical disturbance of the overburden*: Maximum pressure at sensor number 1 and 2.
2. *As a proxy for when leakage risk starts declining*: Time of maximum mobile free phase in Box A.
3. *As a proxy for our ability to accurately predict near well phase partitioning*: All quantities defined in Section 2.7.2 in Box A at 72 hours after injection starts.
4. *As a proxy for our ability to handle uncertain geological features*: All quantities defined in Section 2.7.2 in Box B at 72 hours after injection starts.
5. *As a proxy for our ability to capture onset of convective mixing*: Time for which the quantity M defined in Section 2.7.3 first exceeds 110% of the width of Box C.
6. *As a proxy for our ability to capture migration into low-permeable seals*: Total mass of CO_2 in the top seal facies (areas marked yellow in the sketch) at final time within Box A.

Each of the sparse data shall be reported as six numbers, representing the prediction of the mean quantity as obtained by the experiments (stated in terms of P10, P50 and P90 values), as well as the prediction in the standard deviation of the quantity over the ensemble of experiments (again stated as P10, P50, and P90 values). Loosely speaking the predictions of mean values assess the capability of predicting the various measurables, while the predictions of the standard deviations address the extent to which these quantities are deterministic.

As basis for generating the predictions and uncertainties, any preferred methodology may be chosen (ensemble runs, some methods from uncertainty quantification, human intuition from experience or any combination of these).

The quantities are expected to be uploaded to the repository in form of a csv file `sparse_data.csv` of the form

```
# idx, p10_mean, p50_mean, p90_mean, p10_dev, p50_dev, p90_dev
1a, 1.234e+56, <…>, # pressure at sensor 1 [N/m2]
1b, 1.234e+56, <…>, # pressure at sensor 2 [N/m2]
```

```
2, 1.234e+56, <…>, # time of max mobile free phase in Box A [s]
3a, 1.234e+56, <…>, # mobile free phase in Box A at 72h [kg/m2]
3b, 1.234e+56, <…>, # immobile free phase in Box A at 72h [kg/m2]
3c, 1.234e+56, <…>, # dissolved in water in Box A at 72h [kg/m2]
3d, 1.234e+56, <…>, # seal in Box A at 72h [kg/m2]
4a-d <same for Box B>
5, 1.234e+56, <…>, # time when M exceeds 110% of Box C's width [s]
6, 1.234e+56, <…>, # total mass of CO2 in the top sel facies [kg]
```

3.3 Qualitative data – questionnaires

During the benchmark period, and in particular during the workshops, questionnaires will be provided to gather contextual information. These will address both issues of numerical modeling and simulation, as well as issues of choices and judgements made for assessing the confidence intervals for the sparse data.

3.4 Reporting deadlines

The blind data shall be uploaded to the respective repository by November 14th , 2021. Revised data, if desired, can be submitted in the same format by January 27st, 2022.

Each participating group will report on their blind data at the virtual workshop on November 17, 2021, and at the real-life workshop February 1-2, 2022. The workshops will be structured so that our mutual judgement of the quality of both your own results, as well as those of your peers, can be assessed.

3.5 Limited interaction

In order to protect the integrity of the results, we ask for dedicated communication rules during the different phases of the benchmarking process. To facilitate remote communication between participants, and also to store this communication for evaluating the benchmarking process, a Discord server has been set up at https://discord.gg/8Q5fZS3T47. Apart from a general channel that is initially open to everyone involved, a private channel is installed for each participant which should be used for communicating with the benchmark organizers.

For the preparation phase before the kick-off meeting, no restrictions apply. This phase should be used by all participants and the experimental group to shape this benchmark description. During the blind phase, each participating group acts on its own and we ask that no communication on the subject takes place between different participating groups and with the experimental group. For the participants, the general Discord channel will be read-only. Any questions of clarification should be directed to Bernd Flemisch by the respective private Discord channel, who will post the replies on the general channel if they are of common interest. After the virtual workshop in November, the synchronization phase starts, during which the participating benchmarking groups can communicate freely, preferably over the Discord channels for monitoring and archival of the process. Still, no communication with the experimental group will be allowed before the real workshop in February.

Appendix A: Facies descriptions

This appendix gives detailed information of the six sand types used in the geologic setup.

Multi phase flow properties

The multiphase flow properties will be measured and reported for each facies in the final version of the document. As placeholder values, participants are suggested to use:

- Entry pressure: $10\frac{\gamma}{D}$, where $\gamma = 72\ \mathrm{e}^{-3}\ N/m$ and D is the largest facies grain size in meters.
- End-point water saturation during drainage: 0.1, with gas relative permeability $k_g = 0.9$.
- End-point gas saturation during imbibition: 0.3, with water relative permeability $k_w = 0.5$.

Facies 1 – Sand F (Main Reservoir)

Grain size distribution

- Sieve sizes used: 1.41 mm – 2.0 mm

Figure A1 - A microscope image of **Sand F** and measured values for each of the 100 sand grains. Because the sand grains are not perfect spheres, two lengths were measured and labelled “long” and “short”.

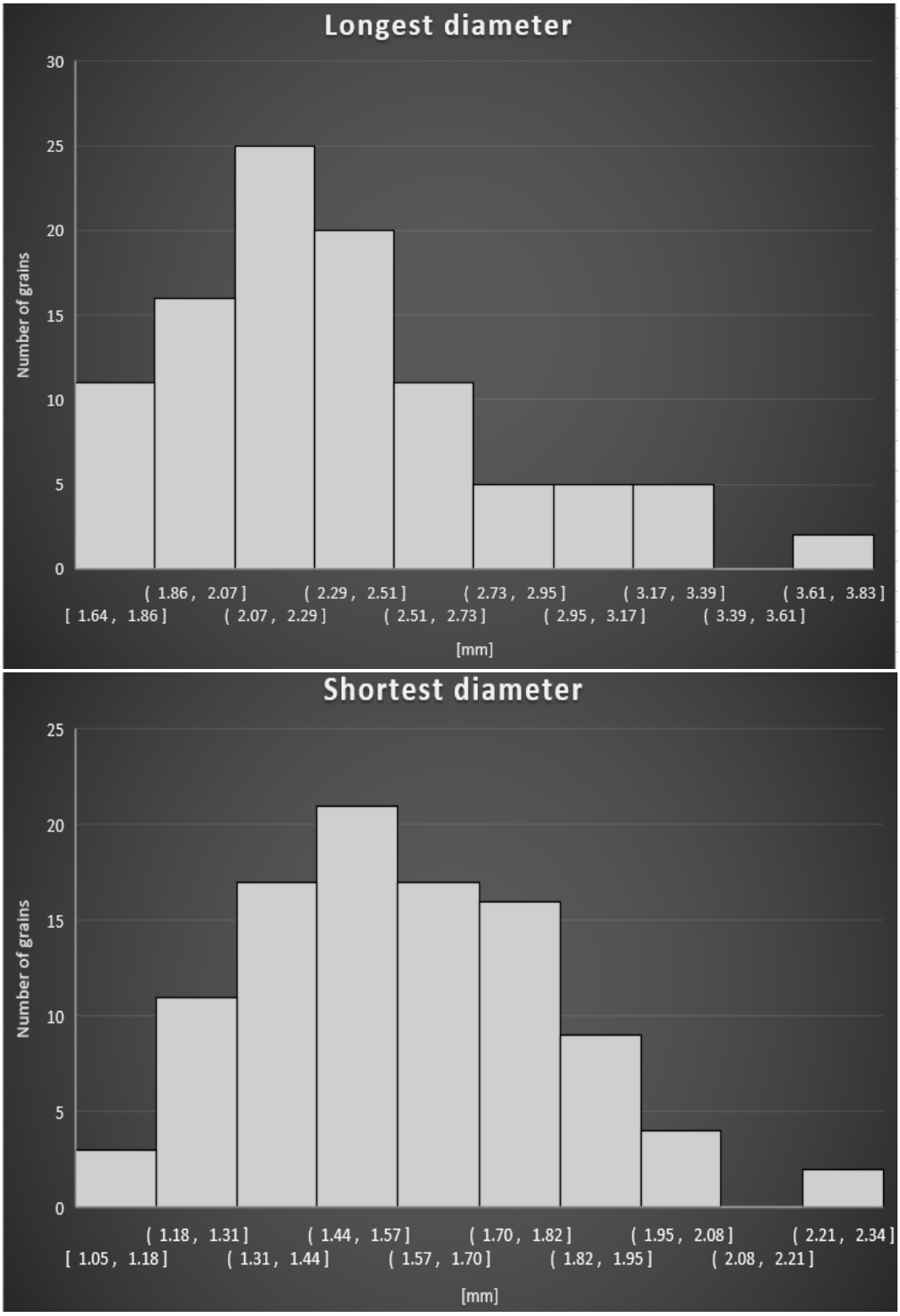


Figure A2 - Histograms for long (top) and short (bottom) diameters measured for **Sand F**. Each plot has 100 measurements distributed in 10 intervals along the x-axis and the number of sand-grains along the y-axis

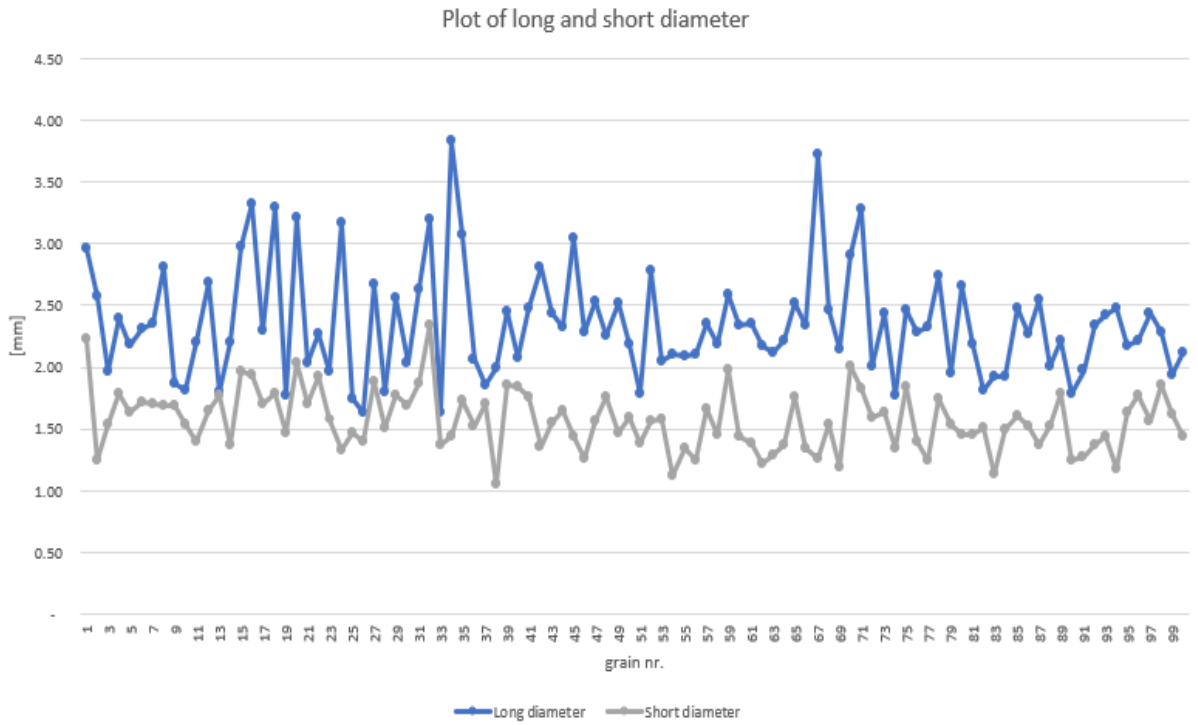


Figure A3 – Graphs for long (top, blue) and short (bottom, gray) grain diameters measured for **Sand F** using image in Figure A1. Each graph has 100 measurements.

Single phase flow properties

- Porosity: To be reported
- Permeability: To be reported

Facies 2 – Sand E

Grain size distribution

- Sieve sizes used: 1.0 mm – 1.41 mm

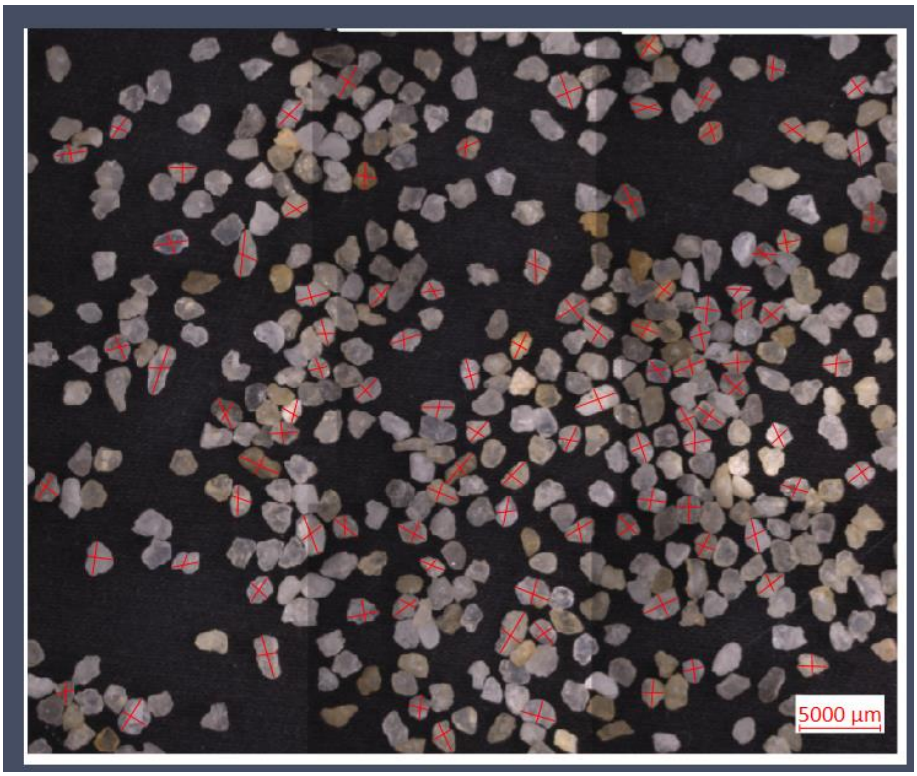
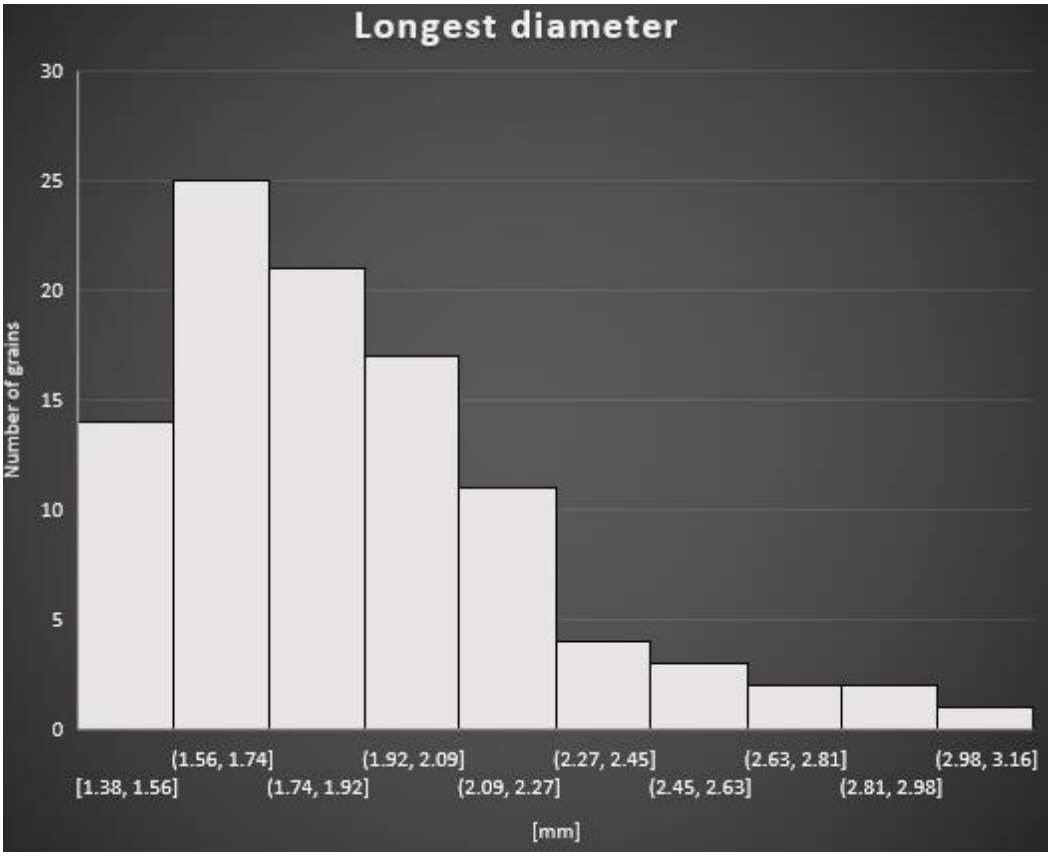


Figure A4 - A microscope image of **Sand E** and measured values for each of the 100 sand grains. Because the sand grains are not perfect spheres, two lengths were measured and labelled “long” and “short”.



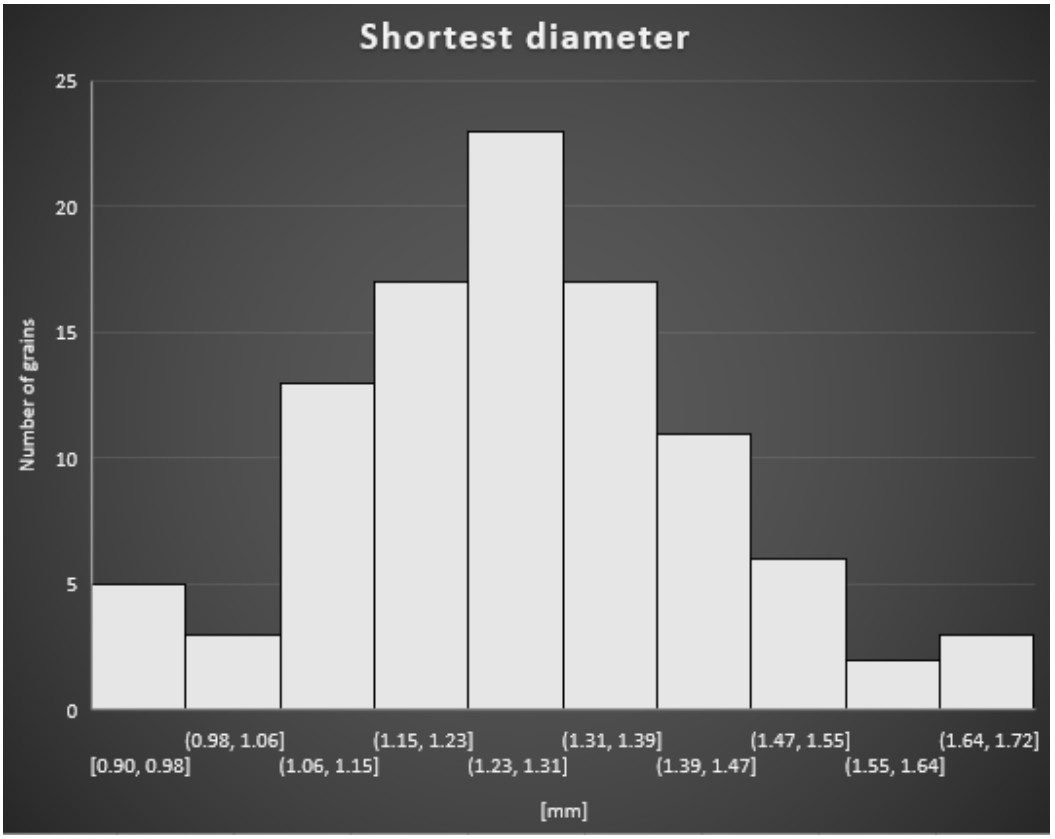


Figure A5 - Histograms for long (top) and short (bottom) diameters measured for **Sand E**. Each plot has 100 measurements distributed in 10 intervals along the x-axis and the number of sand-grains along the y-axis

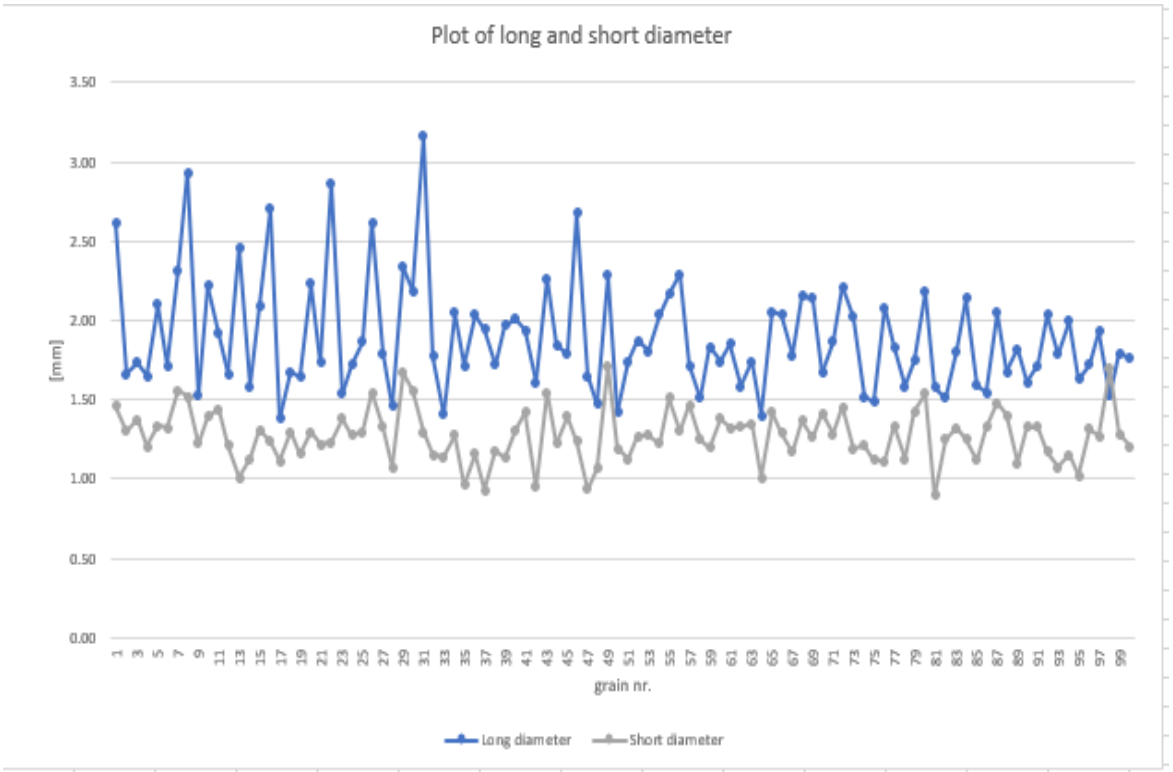


Figure A6 – Graphs for long (top, blue) and short (bottom, gray) grain diameters measured for **Sand E** using image in Figure A4. Each graph has 100 measurements.

Single phase flow properties

- Porosity: To be reported
- Permeability: To be reported

Facies 3 – Sand D

Grain size distribution

- Sieve sizes used: 0.71 mm – 1.00 mm

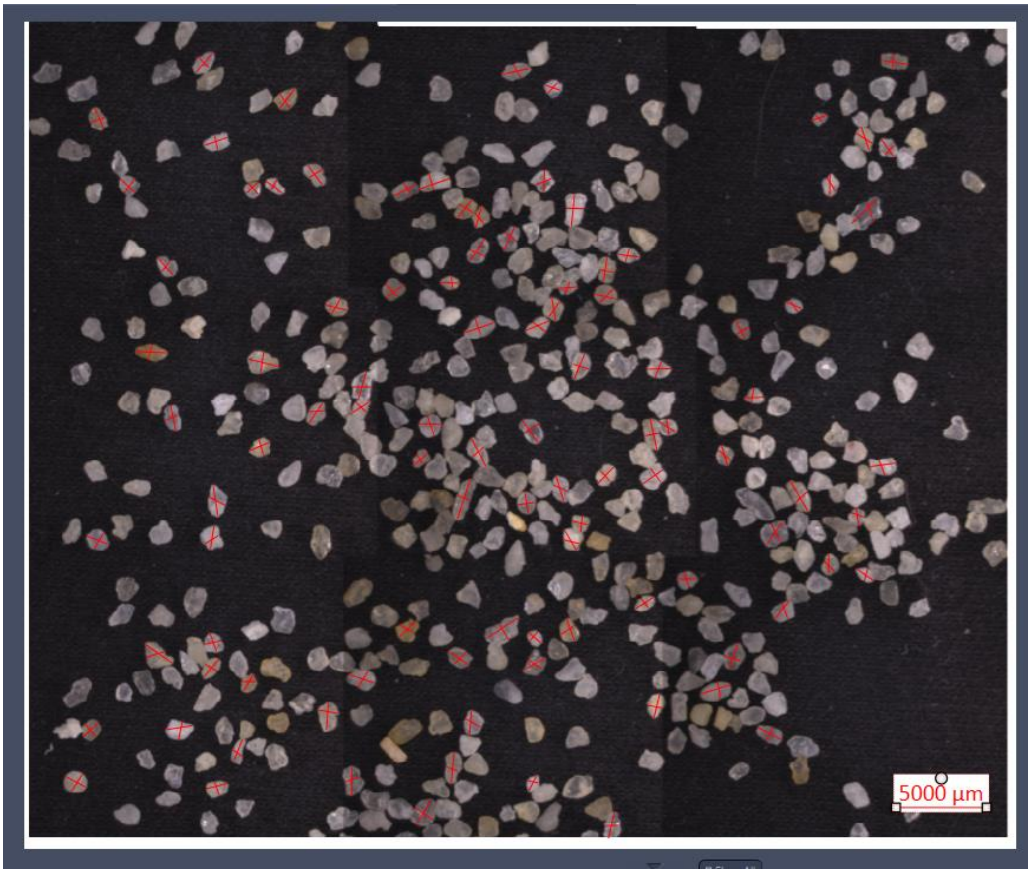
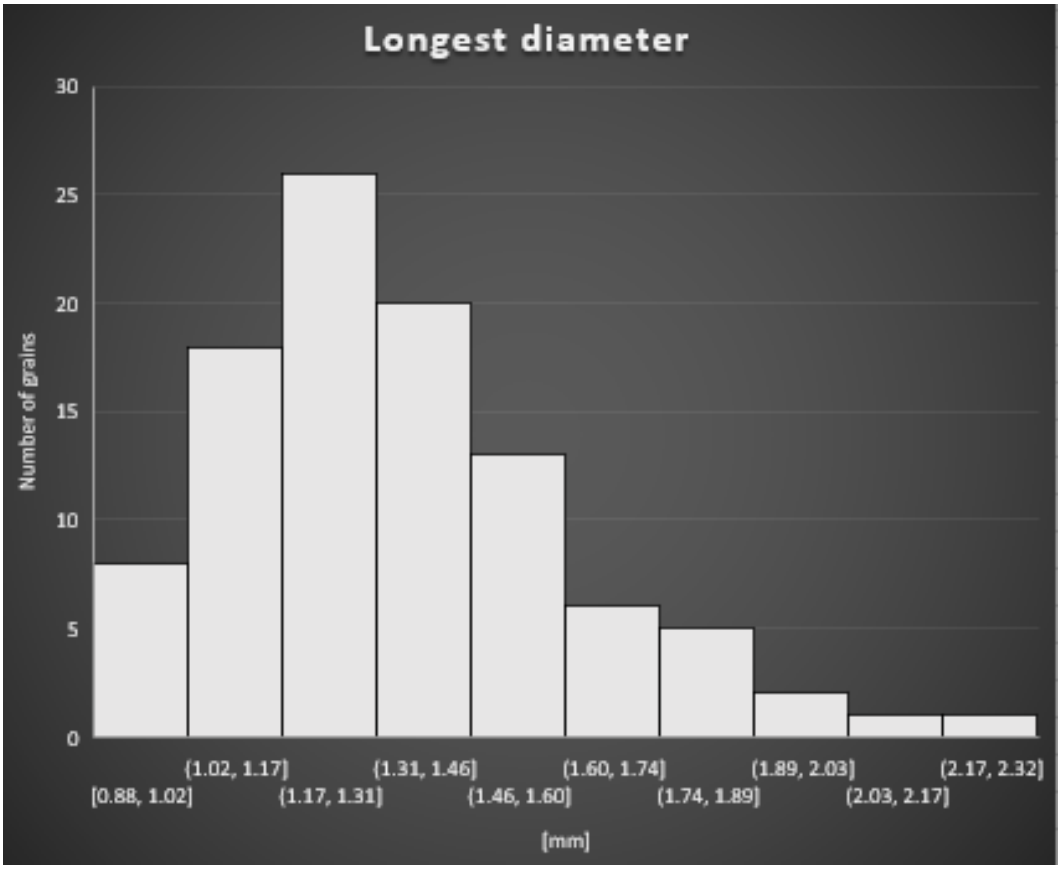


Figure A7 - A microscope image of **Sand D** and measured values for each of the 100 sand grains. Because the sand grains are not perfect spheres, two lengths were measured and labelled “long” and “short”.



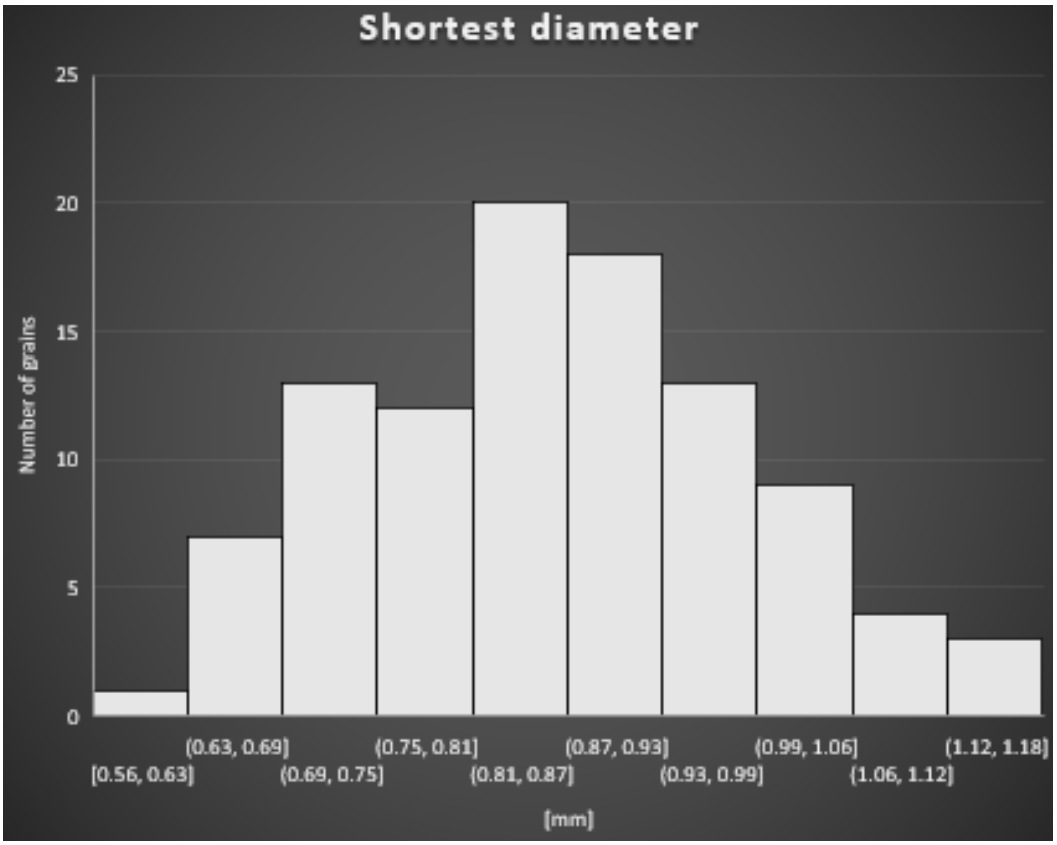


Figure A8 - Histograms for long (top) and short (bottom) diameters measured for **Sand D**. Each plot has 100 measurements distributed in 10 intervals along the x-axis and the number of sand-grains along the y-axis

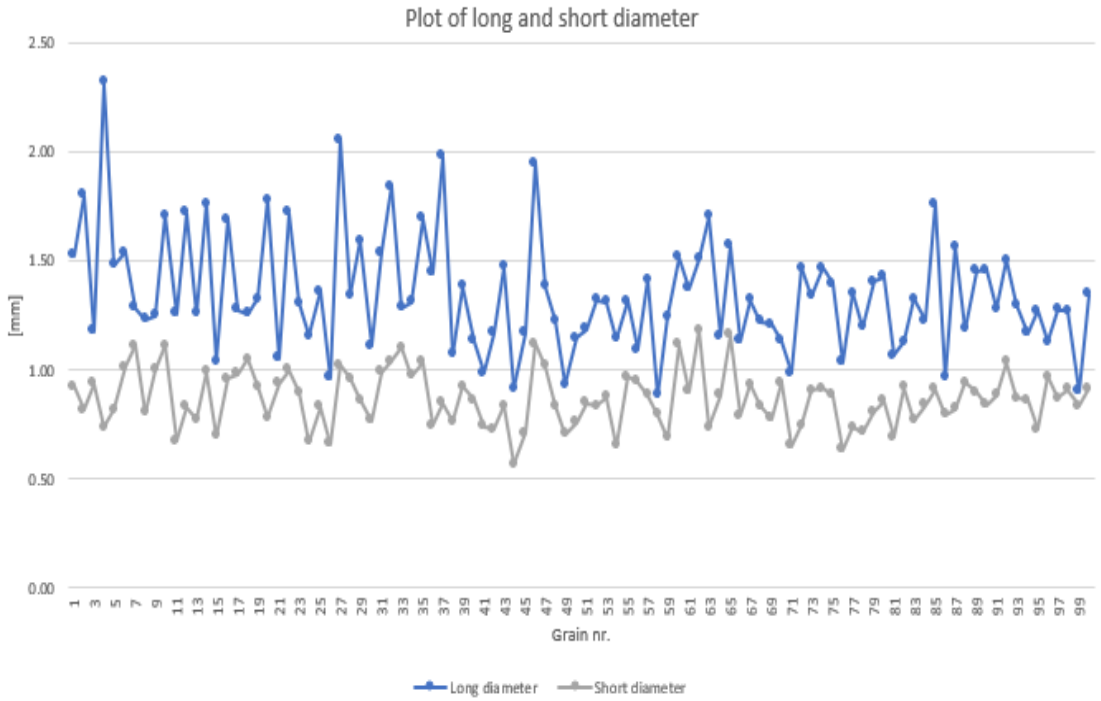


Figure A9 – Graphs for long (top, blue) and short (bottom, gray) grain diameters measured for **Sand D** using image in Figure A7. Each graph has 100 measurements.

Single phase flow properties

- Porosity: To be reported
- Permeability: To be reported

Facies 4 – Sand C

Grain size distribution

Sieve sizes used: 0.50 mm – 0.71 mm

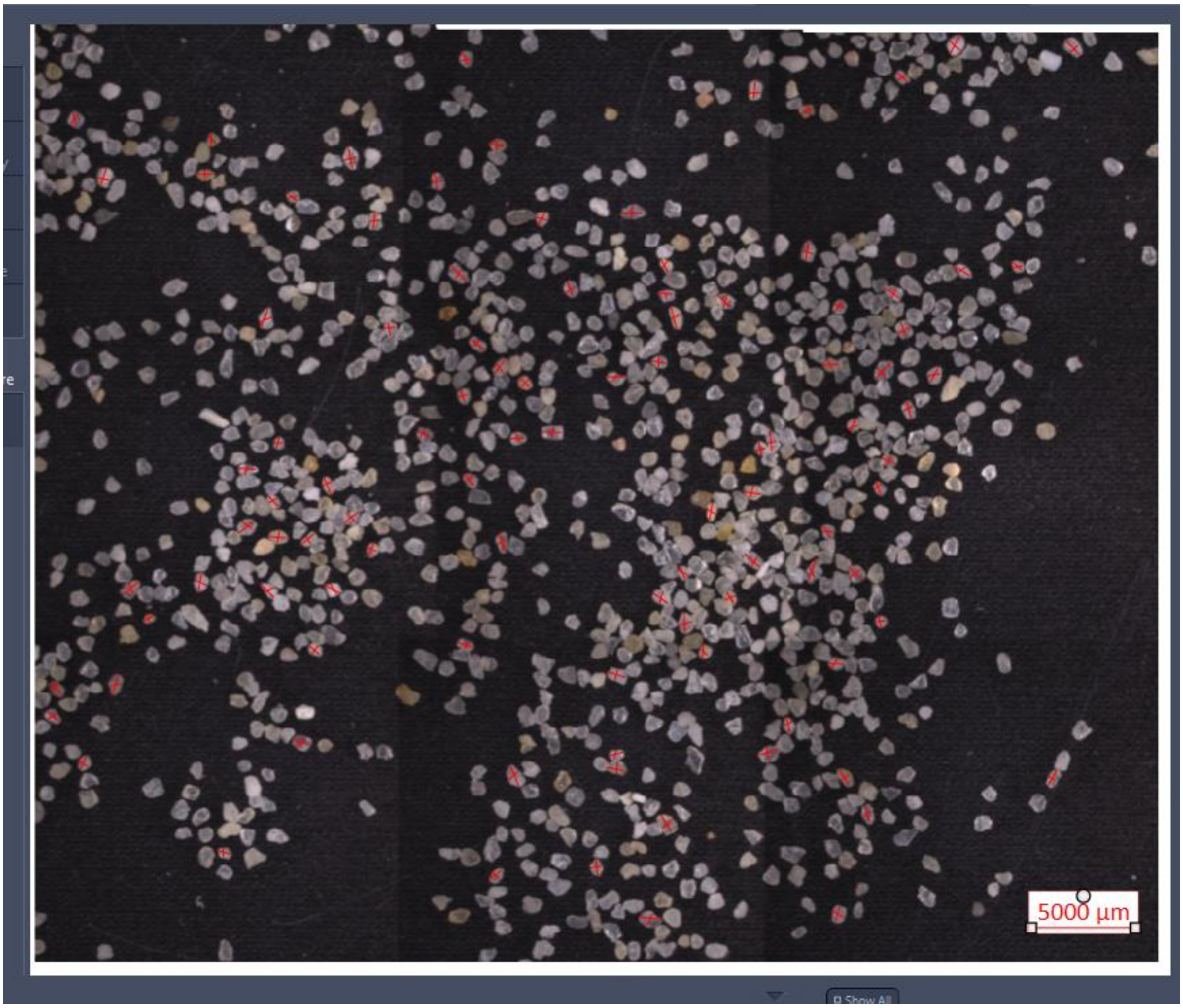


Figure A10 - A microscope image of **Sand C** and measured values for each of the 100 sand grains. Because the sand grains are not perfect spheres, two lengths were measured and labelled “long” and “short”.

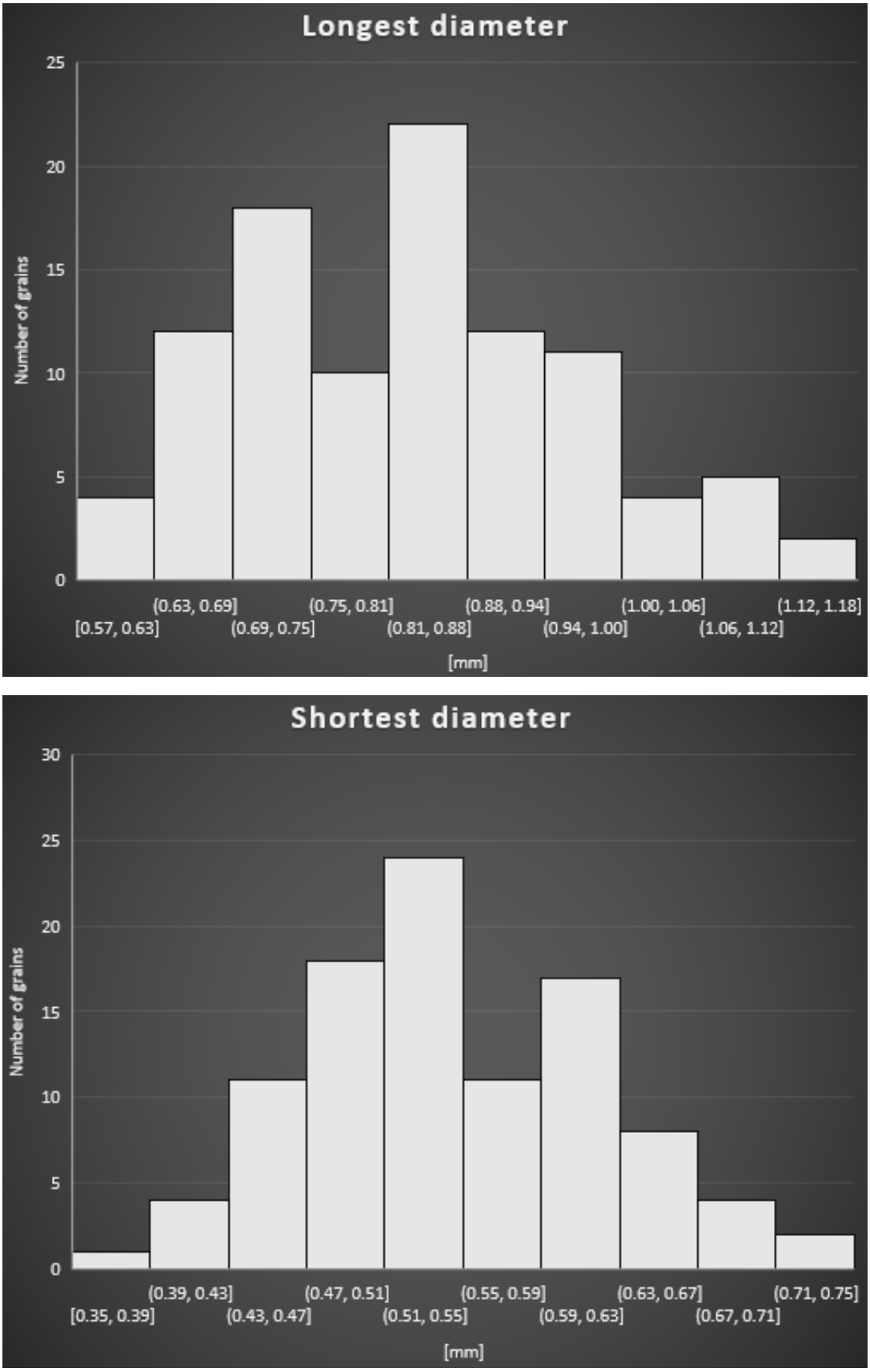


Figure A11 - Histograms for long (top) and short (bottom) diameters measured for **Sand C**. Each plot has 100 measurements distributed in 10 intervals along the x-axis and the number of sand-grains along the y-axis

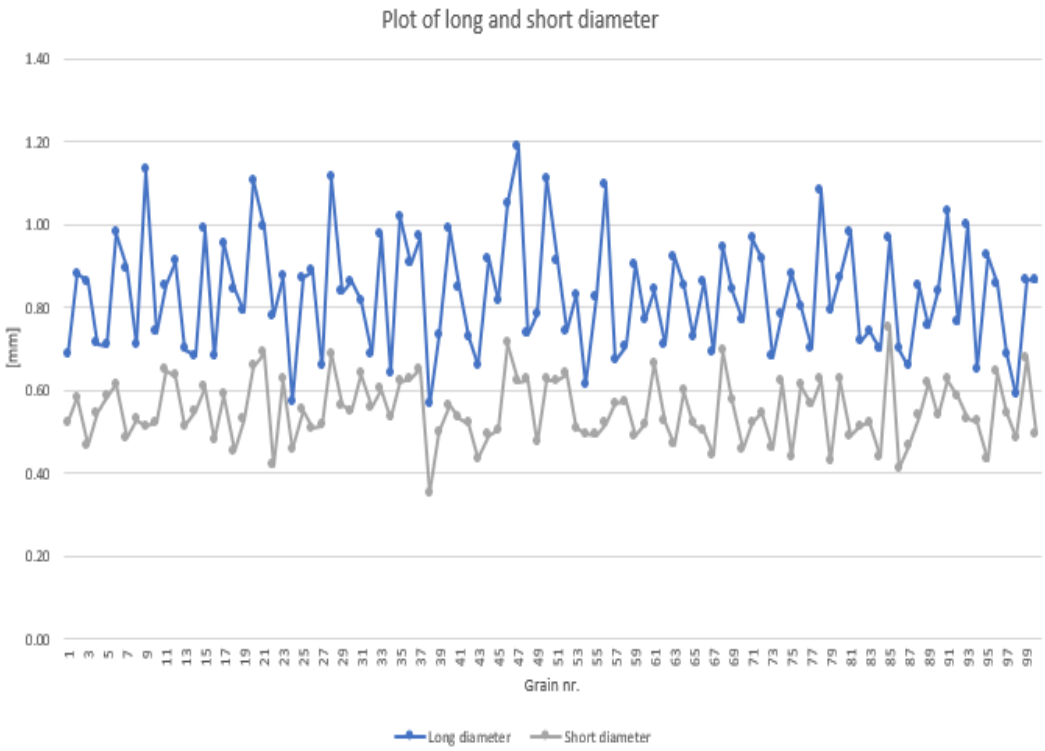


Figure A12 – Graphs for long (top, blue) and short (bottom, gray) grain diameters measured for **Sand C** using image in Figure A10. Each graph has 100 measurements.

Single phase flow properties

- Porosity: To be reported
- Permeability: To be reported

Facies 5 – Sand ESF

Grain size distribution

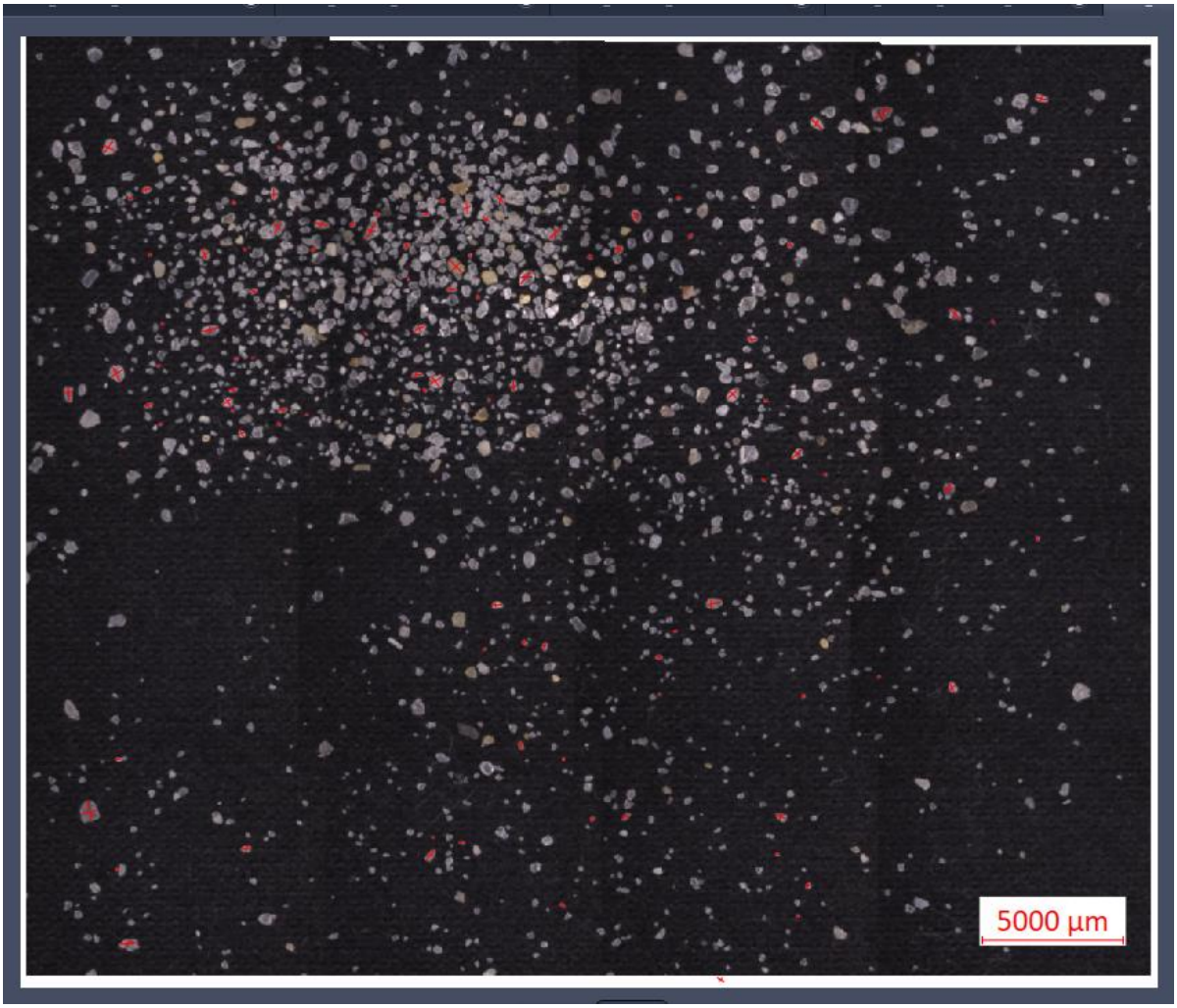


Figure A13 - A microscope image of **Sand ESF** and measured values for each of the 100 sand grains. Because the sand grains are not perfect spheres, two lengths were measured and labelled “long” and “short”.

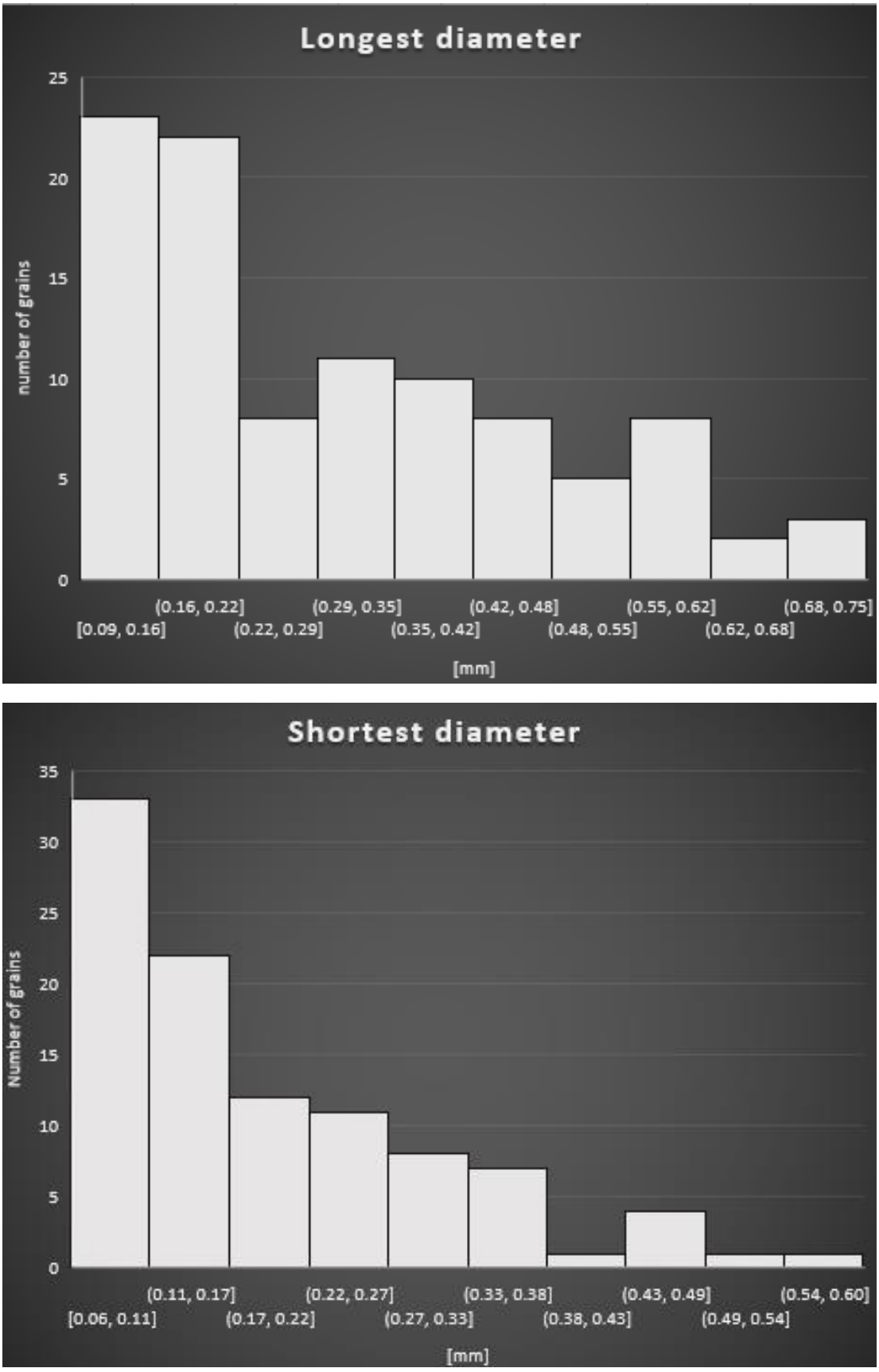


Figure A14 - Histograms for long (top) and short (bottom) diameters measured for **Sand ESF**. Each plot has 100 measurements distributed in 10 intervals along the x-axis and the number of sand-grains along the y-axis

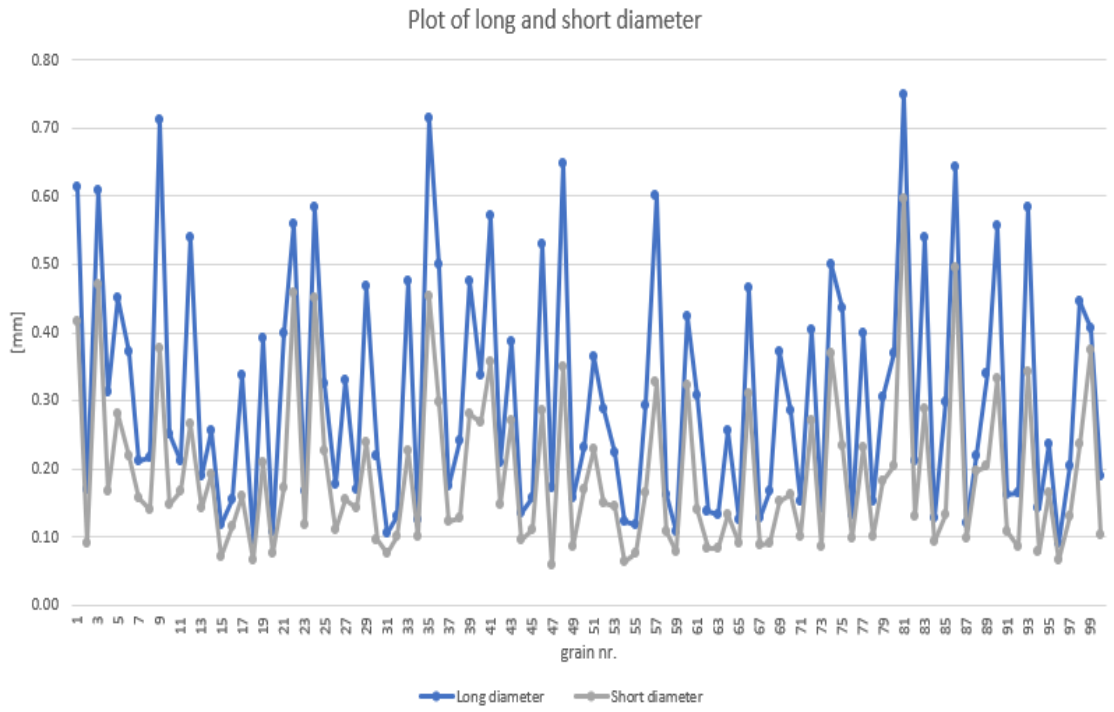


Figure A15 – Graphs for long (top, blue) and short (bottom, gray) grain diameters measured for **Sand ESF** using image in Figure A13. Each graph has 100 measurements.

Single phase flow properties

- Porosity: To be reported
- Permeability: To be reported

Facies 6 – Sand G

Grain size distribution

- Sieve sizes used: 2.0 mm – 2.6 mm

Figure A16 - A microscope image of **Sand G** and measured values for each of the 100 sand grains. Because the sand grains are not perfect spheres, two lengths were measured and labelled “long” and “short”.

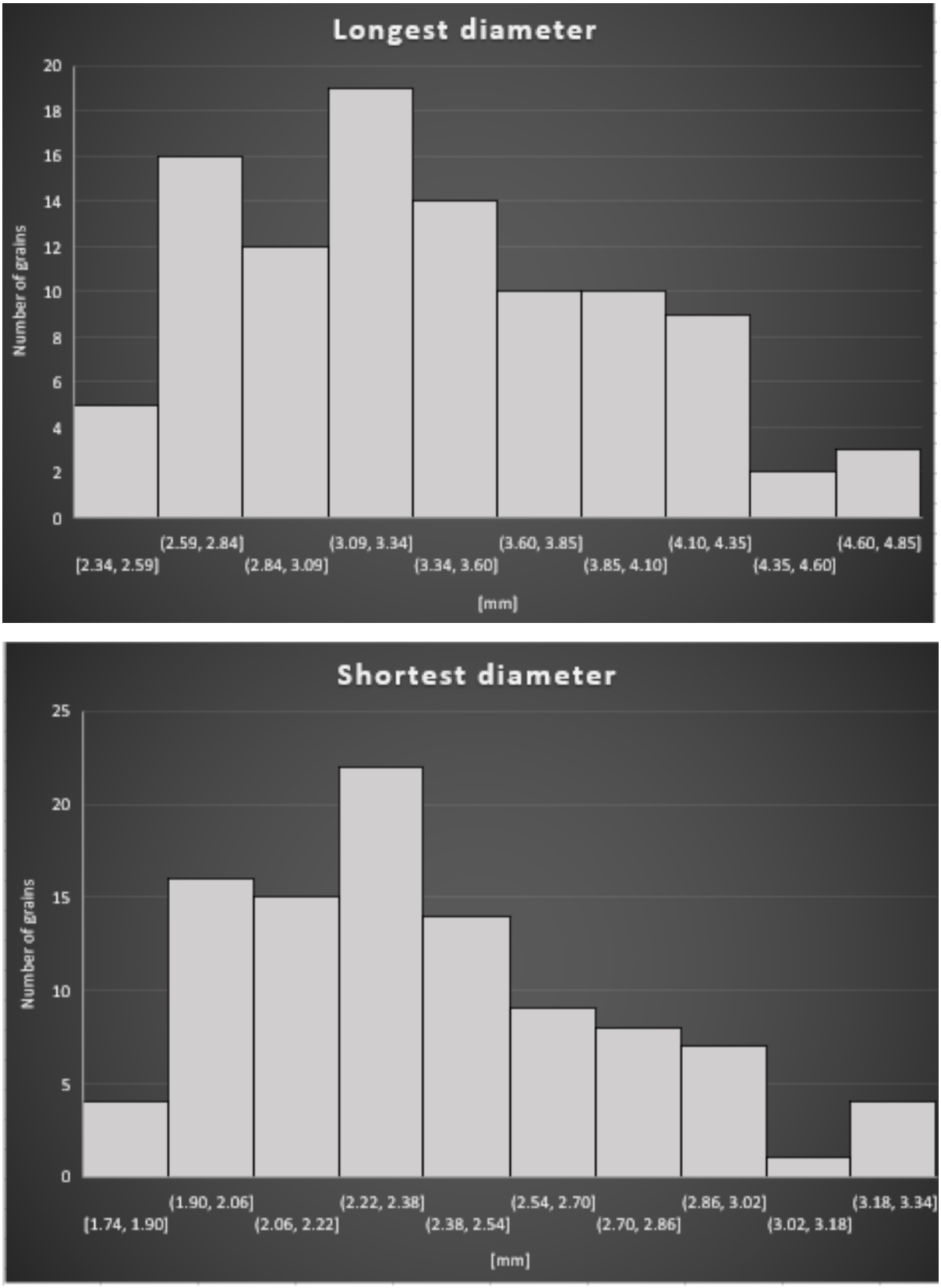


Figure A17 - Histograms for long (top) and short (bottom) diameters measured for **Sand G**. Each plot has 100 measurements distributed in 10 intervals along the x-axis and the number of sand-grains along the y-axis

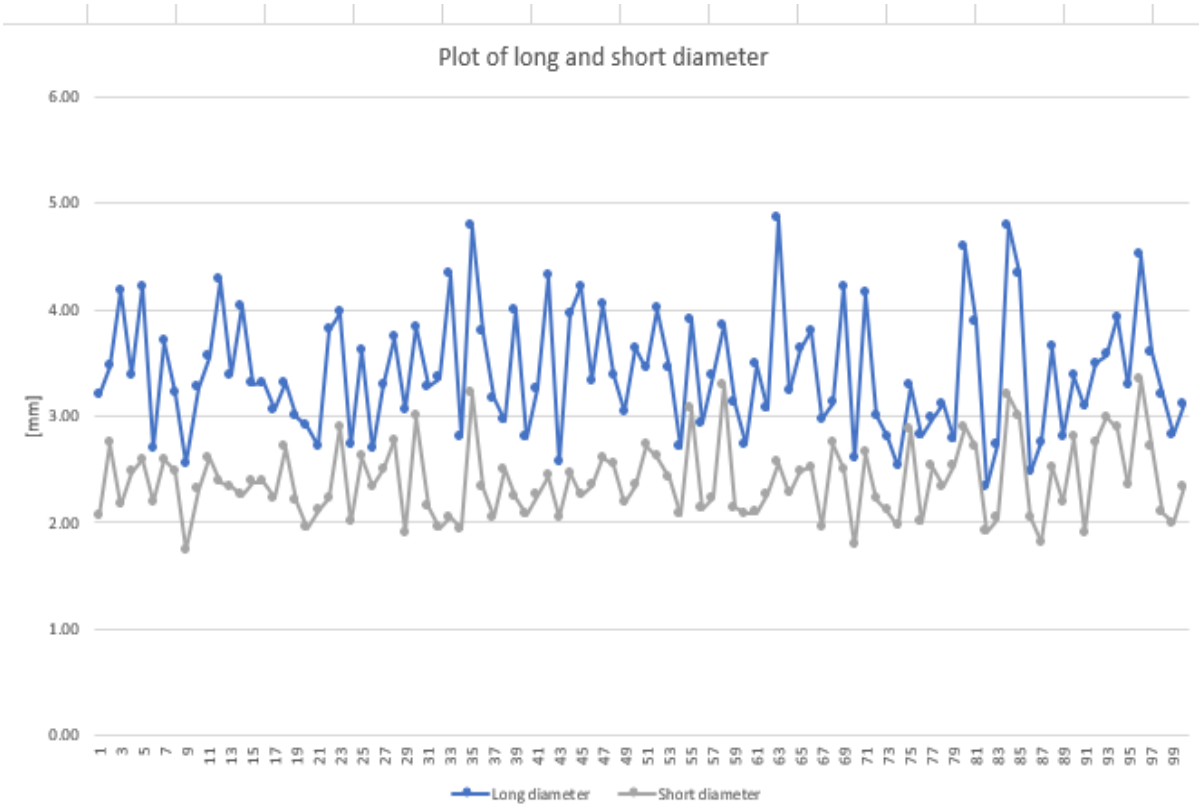


Figure A18 – Graphs for long (top, blue) and short (bottom, gray) grain diameters measured for **Sand G** using image in Figure A16. Each graph has 100 measurements.