Disclaimer: we totally understand that the manuscript under consideration is overloaded with details, and referees most likely were not able to process most of them within reasonable time frame. Those specific details originate from very diverse disciplines:

- *granular media for sample preparation,*
- *rheology for viscosity measurements,*
- *temperature effects for permeability stabilization,*
- *computed tomography for image acquisition,*
- *image processing for contrast issues and segmentation,*
- *lattice Boltzmann method for flow simulations (dealing with solution convergence and truncation error),*
- *supercomputing for enabling the high-res 3D pore-scale simulations, which we did not mention at all.*

Many details may seem trivial — but this is mainly due to the proper examination, execution, and explanation of the entire workflow. All presented details are already a shortened version, and all were necessary to capture the scope of the claims made as well as specifics of the work. The latter is a tight mix of fundamental science and engineering, while connecting reality with its computer model twin with no free parameters, and striving to stay within accuracy limits of the available equipment and methods.

========== Referee: 1 ==========

Comments to the Author

The present study revisits a very old issue of porous media flow that goes back to the seminal and pioneering work of Henry Darcy (1856), which is the correct determination of permeability of a porous medium. As a matter of fact, the manuscript makes the very strong claim that many if not all measurements have been incorrect. Well, we do not say that all measurements are incorrect. Quantitative picture in Fig. 6D,E suggests that sometimes the flaw magnitude is small or absent. The machinery behind this analysis impressive ranging from high-precision measurement stands that appear to be state-of-the-art to highly resolved numerical simulation that resolve the spatial geometry of the pores measured by CT scanner.

Having said this, I have problems with the conclusions drawn in this manuscript, as it seems to be not well founded enough to support the strong claim made in the abstract. We now have better explanation of the observed flaw, paragraphs “The sharp pressure drop…” and “Let us qualitatively assess…” in Section 3.4. Additional experiments and more thorough review of existing procedures as well as suitable suggestions to improve those flawed measurements are urgently needed to turn this study into a useful exercise.

My major concerns are:

1.1) I think that the authors should either be much more specific on the experimental apparatus that they wish to criticize and improve, or they should do a more comprehensive study. It seems like the authors have focused on constant head experiments with a certain geometry of two tubes that guide the flow in and out of a porous flow cell. However, there are many other methods that are currently being used to determine permeability, such as very long porous flow cells, falling head setups and pumping tests in field studies. As the manuscript reads right now, it seems like the authors engineered their setup so that they would inevitably encounter the deficiency of their measurements.

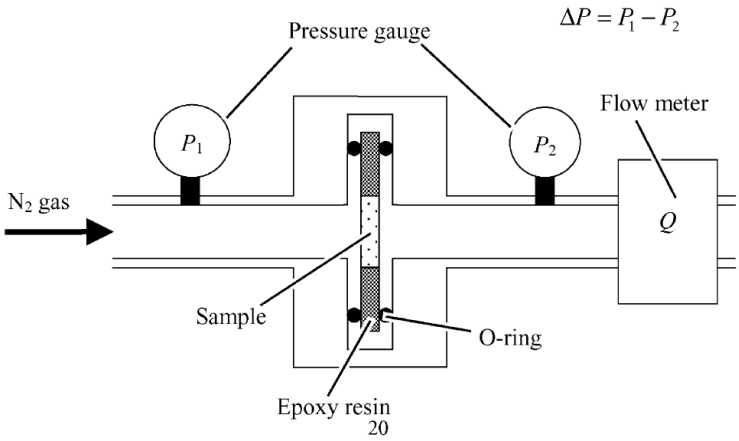
Permeability, *k*, defined from Darcy’s law, assumes that for a given incompressible, isothermal fluid there is a linear relation between applied (or measured) pressure drop-over-length, *dP*, and the corresponding flow rate, *f*. *dP* can be applied to the sample and the corresponding *f* estimated, or conversely, *f* can be created and *dP* measured. *dP* can originate from an external pump or from a body force (gravity, specifically falling head), which at the end will lead to the same result. In principle all experiments we found in the literature are based on the logic above. What is varied is the location of pressure ports (in- and out-of-sample) and the sample geometry/dimensions. We do not exactly understand

what Referee1 means under “very long porous flow cells”, but it seems that they also follow the description above, either with external ports (then Lsmpl >> Ltubes) or internal ports (then there are no tube-sample contacts within hydrodynamic region and the system is not affected by the reported flaw).

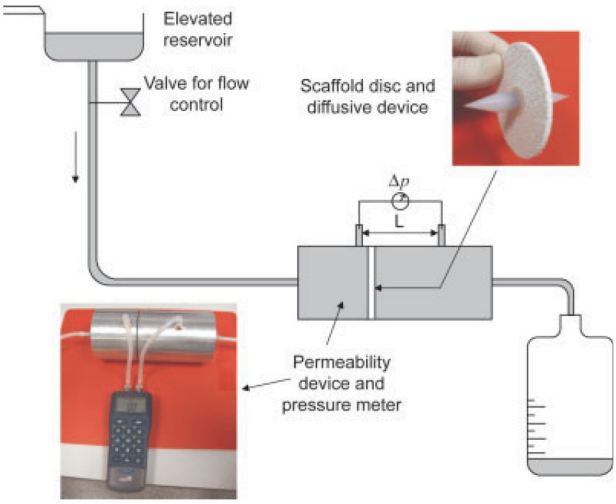
What is related to the field-scale studies (“pumping tests in field studies”), the Darcy law physics (*dP* vs *f)* is also there, but most likely such measurements are impacted by additional phenomena such as compressibility effects (“pressure diffusion”), Klinkenberg effect (if there is gas flow in nano-pores), thermal disequilibrium, unclear boundary condition impact (what happens at the opposite boundary of the fluid plume injected through the well? If such a boundary can in principle be defined), capillary forces, etc.

Below we show experimental setups from variety of fields, where *k* or similar quantity is estimated with a *dP* vs *f* relation one way or another.

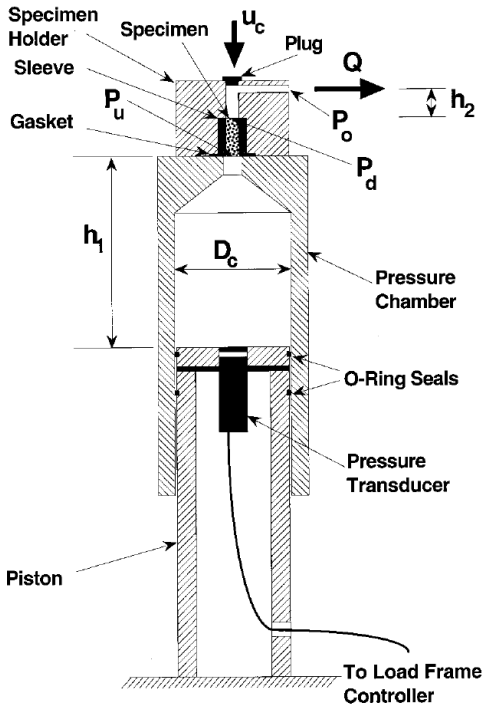
Porous alumina ceramics http://dx.doi.org/10.1016/j.jeurceramsoc.2006.02.030 :



Bioceramic scaffolds https://doi.org/10.1002/jbm.b.31065 :



Trabecular bone https://doi.org/10.1114/1.195 :



Some filters: in https://dx.doi.org/10.1039/c5ta06543g the authors do not say “permeability” but actually discuss it comparing the slopes for two filters (membranes):

Fig1.

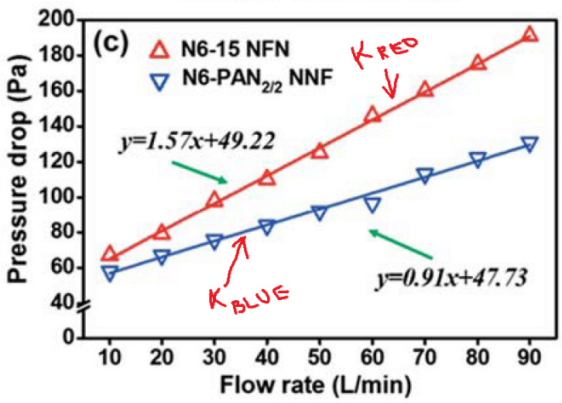
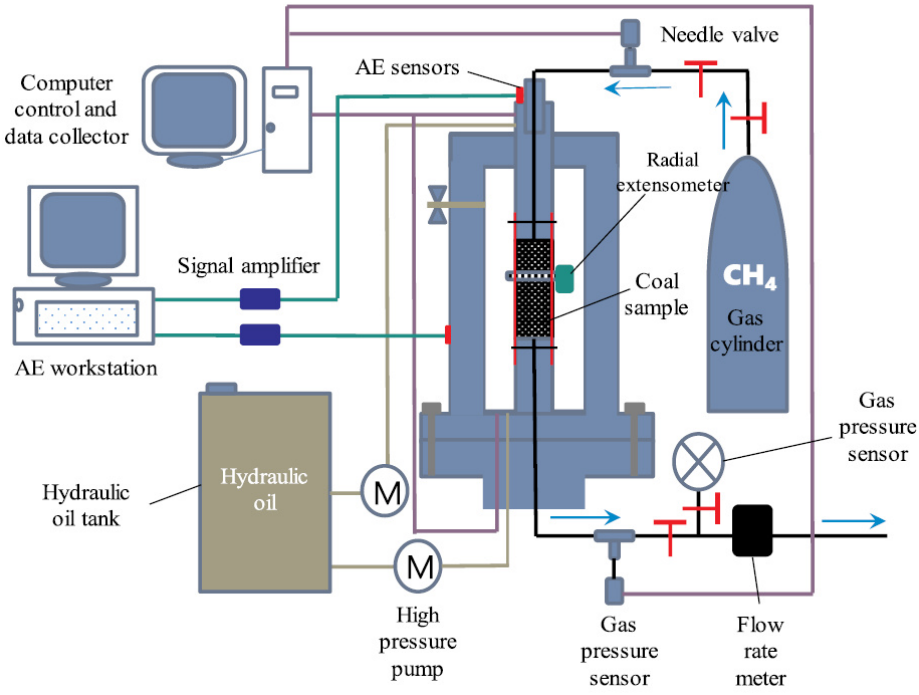
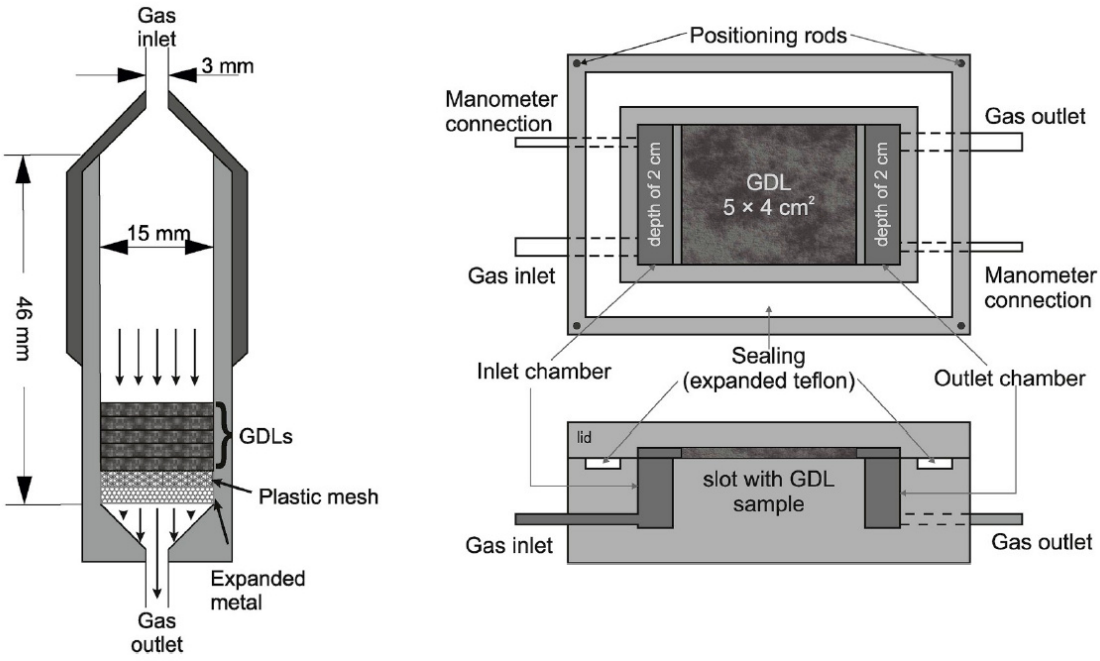


Fig.8

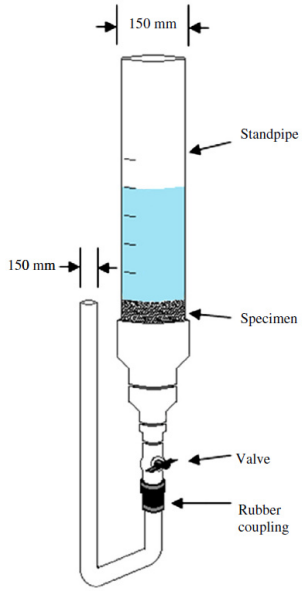
Coal https://doi.org/10.1016/j.jngse.2019.103054 :



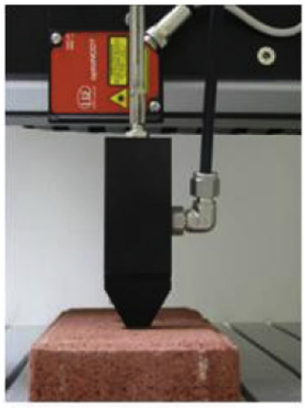
Gas diffusion layer (GDL), fuel cells https://doi.org/10.1016/j.jpowsour.2020.227828 :

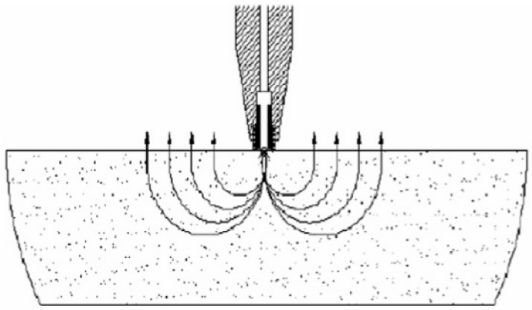


Pervious concrete http://dx.doi.org/10.1016/j.conbuildmat.2014.02.034 By the way, here we have a falling head permeameter, but again, nothing else than *dP* (body force or gravity) vs *f* can be measured here:

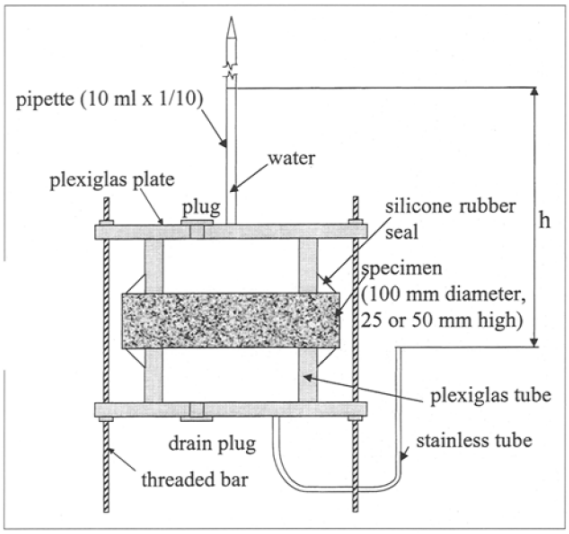


Porous building materials http://dx.doi.org/10.1016/j.conbuildmat.2016.02.193 :

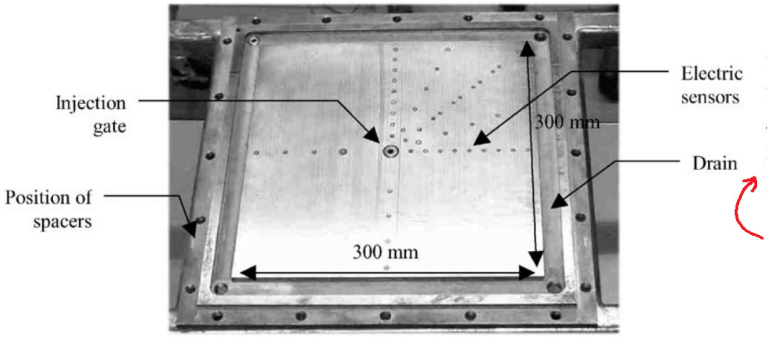




Cracked concrete https://doi.org/10.1016/S0008-8846(97)00031-8 (falling head permeameter is here again):



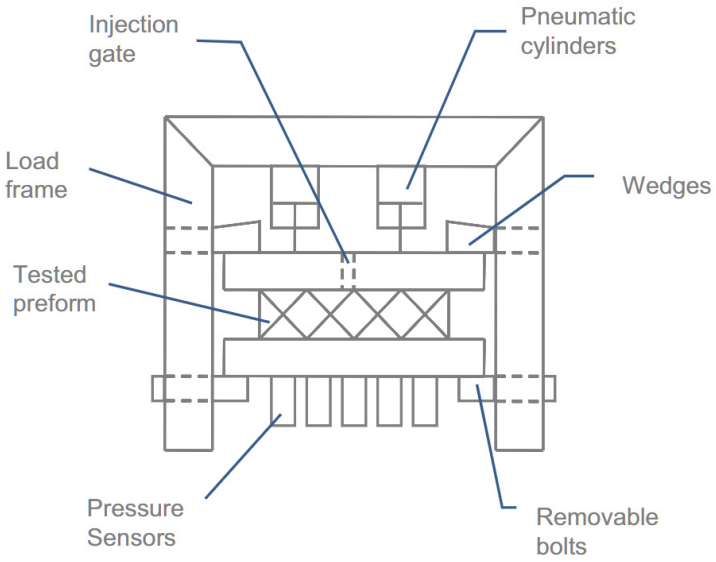
Fibrous reinforcements https://doi.org/10.1016/S1359-835X(02)00035-0. For background please see this video https://www.youtube.com/watch?v=1u-2GvhghQA . Here it is interesting to note that with injection of a fluid into the sample from top, and propagation of the fluid front towards the mold edges, the tube-to-sample diameter ratio (in our terminology, sample = fluid mass inside the fluid front) will decrease until reaching the edges:



underneath the plate. The fluid is injected into the mold through a 6 mm-diameter hole in the middle of the plate. Just below the surface of the plate a pressure transducer is installed in the injection tube to measure the actual pressure at which the test fluid enters the cavity.

out-of-sample pressure port location

The same application as above, but with internal pressure ports, https://dx.doi.org/10.3929/ethz-a-005730612 (p. 210):



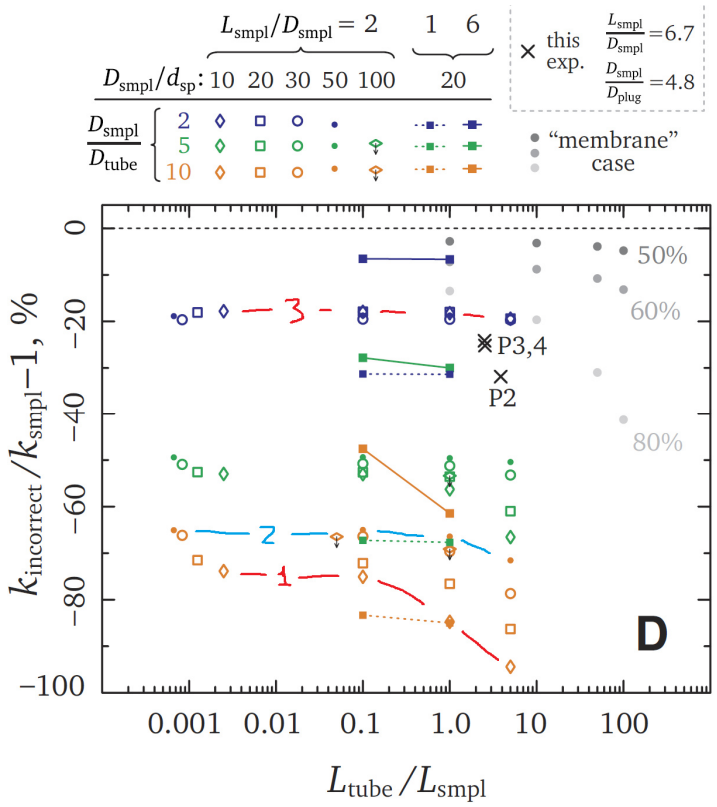
1.2) The authors should take a better effort in suggesting more sophisticated ways to correct permeability based on their observation. Clearly, doing a CT-scan and carry out geometrically resolving numerical simulations as suggested on page 10 will not be an option for regular engineering practice.

We provide more insight into the observed flaw (paragraphs “The sharp pressure drop…” and “Let us qualitatively assess…” in Sec. 3.4), replaced panel 6E with a new rough and simple approximation of the permeability error via geometrical parameters, also extended the paragraph “The flaw can be addressed in several ways…” in Section 4.

1.3) It seems like a lot of the misconception originates from the sample geometry, namely that the length of the inlet and outlet tubes, where pressure is measured, is of similar length than the flow cell. This seems to violate the concept of a representative element volume (REV) as it is commonly used in microscale experiments to derive continuum parameters such as permeability. Hence, I recommend to do a more thorough study on the ratio of tube length to flow cell length, where the error should reduce as the ratio decreases.

1.3.1 In our understanding, the concept of REV is used in situations when the entire system cannot be studied for some reason. In our case, we *designed* the experiment–simulation system to avoid REV-related ambiguity, and to have the full geometry at a pore-scale in computer memory and to exclude geometric features we cannot scan (i.e., beads are non-porous/totally impermeable to flow). Originally, we considered including only lateral confining wall of the glass tube. After first results (P2 with nylon mesh) we realized that the longitudinal boundaries had to be added too (P3, P4 with silicon mesh) as well as tubing.

1.3.2 (Please also see our reply to 3.14) Contrary to our initial plan, in experiments, the tubing length was varied but non-systematically. Originally P2 had a longer tube (~25cm, look closer at Fig. A9,A) compared with P3,P4 (~15cm, Fig. A9,D): for the latter, the tube was cut after we realized that the full system needs to be re-simulated, and the longer tubing would dramatically increase (4x) the number of LBM iterations. But we varied the tube length systematically in our model system (Fig. 6A). Here we observe quite logical results: a longer tubing of smaller radius has the strongest impact on permeability for the largest sample pore sizes (or the sample permeability or grain size, "1" + orange diamonds in Figure below) while reducing the *sample* permeability reduces tubing length impact ("2" + orange dots below). When the tube diameter approaches the sample diameter (blue symbols), the tubing length impact decreases ("3" + all blue symbols below).



1.4) Authors make various strong claims that are not backed up by references:

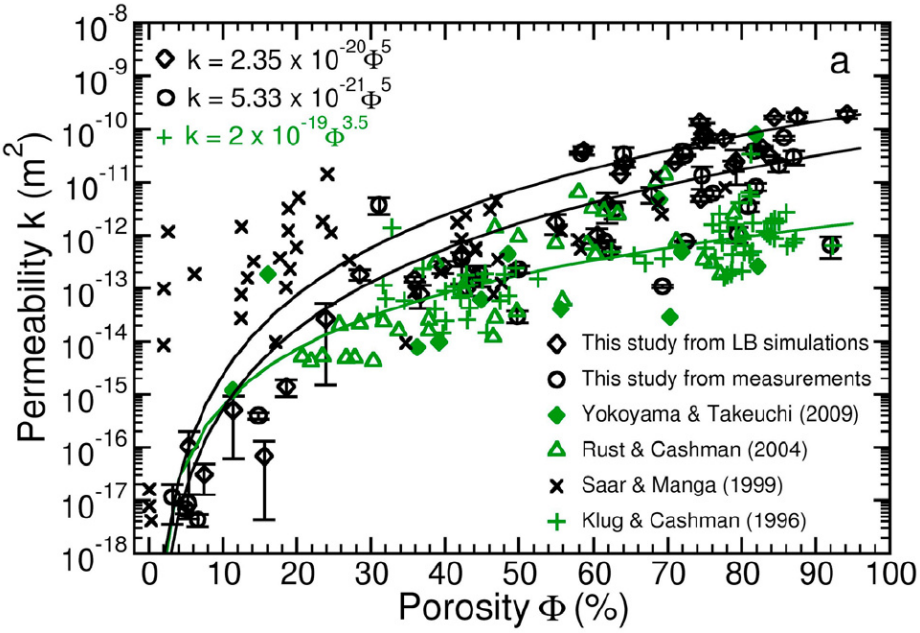
1.4.1 page 2, second paragraph: what are examples where large differences are not uncommon?

Please see reply to 1.4.2

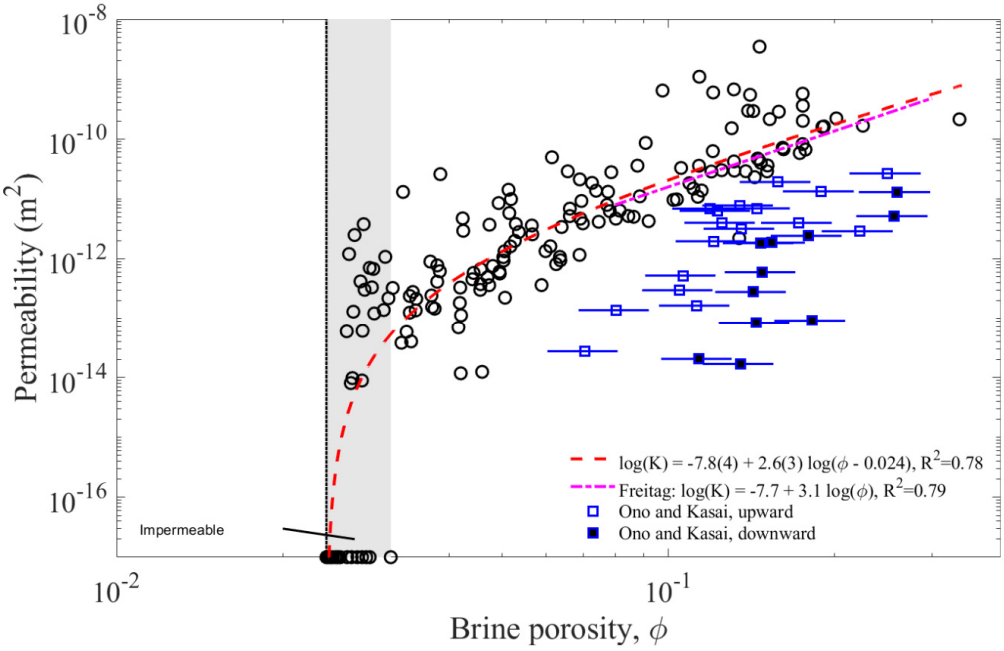
1.4.2 Page 2, second paragraph: what are the studies, where differences seem to acceptable in some fields?

First, we would like to reply to only 1.4.2 because it follows from 1.4.1, and therefore we added citations only at the end of the second paragraph. Our logic is that if the decade differences were uncommon (rare) they would not be generally accepted.

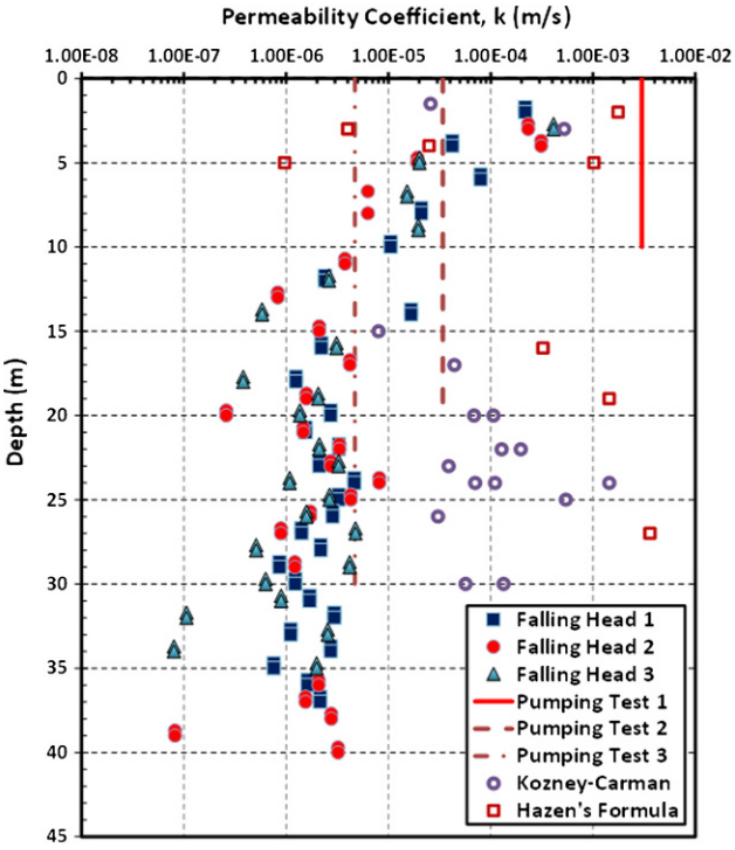
Now, back to replying to 1.4.2. In the figure below (from https://doi.org/10.1029/2009JB007047, also added as a citation in the 1.4.3 context), one sees different studies on similar samples (i.e., of basaltic glass) presented by different symbols: black crosses, green triangles, green diamonds, green crosses. Each study spans 1-2 decades in permeability. In this figure, the reader is naturally drawn to compare a decade-high cloud of one symbol against the cloud of other symbols, and a kind of "cloud war" emerges, i.e., the authors suggest and accept such a comparison:



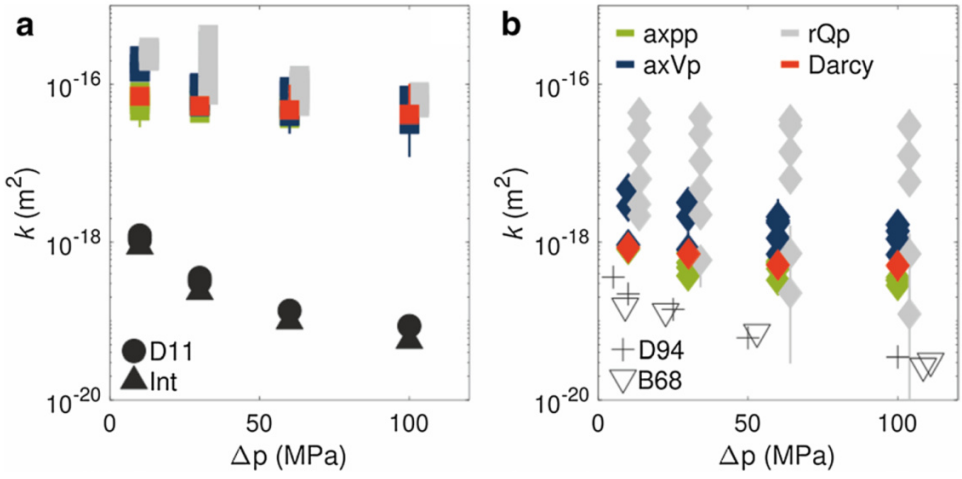
The same is here https://doi.org/10.5194/tc-15-4047-2021 (was cited in different context but now added to the 1.4.2 context):



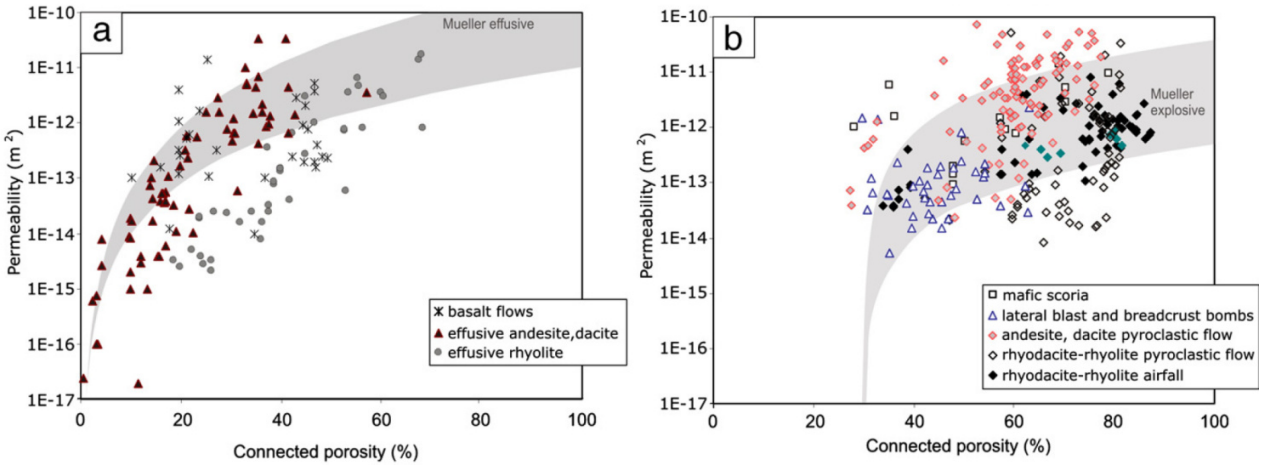
Or here http://dx.doi.org/10.1016/j.aej.2016.07.034 (new citation):



Or here https://dx.doi.org/10.1007/s11242-020-01508-8 (new citation):



Or here https://dx.doi.org/10.1016/j.epsl.2009.01.023 (new citation):



==========

1.4.3 Page 2,third paragraph: what are the studies, where several-fold differences between simulations and experiments are not uncommon?

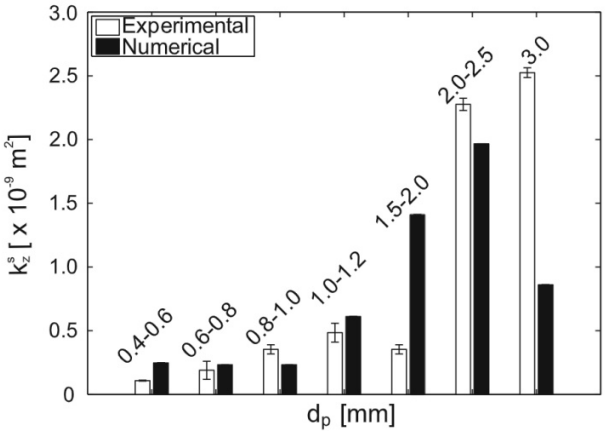
Here http://dx.doi.org/10.1002/2014WR015959 (new citation)

Table 1. Summary of Porosity and Hydraulic Conductivity—Experimental and Simulated

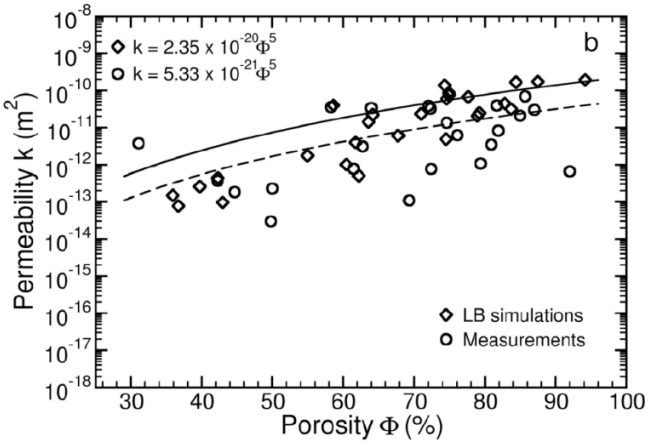
Case	Porosity (Percent)	Hydraulic Conductivity (cm/min)
Experiment	17.8	1.57
Binary stokes	17.0	21.10[a]
Ternary stokes	3.6	3.39[b]/3.61[a]
Ternary stokes-Darcy	17.8	4.75[b]

[a]TETHYS.
[b]STAR-CCM.

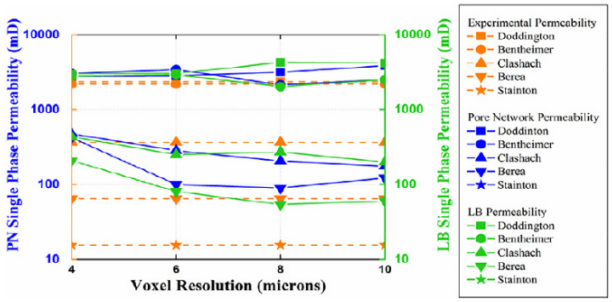
Or here http://dx.doi.org/10.1007/s10035-017-0705-x (was cited in different context but now added to the 1.4.3 context):



Or here https://doi.org/10.1029/2009JB007047 (new citation, also appears in the 1.4.2 context):



Or here http://dx.doi.org/10.1016/j.advwatres.2015.07.012 (new citation):



. 5. Predicted PN and LB single-phase permeability as a function of scanned voxel olution for sandstones: Doddington (rectangles), Bentheimer (circles), Clashach per triangles), Berea (lower triangles) and Stainton (stars). The measured experi-

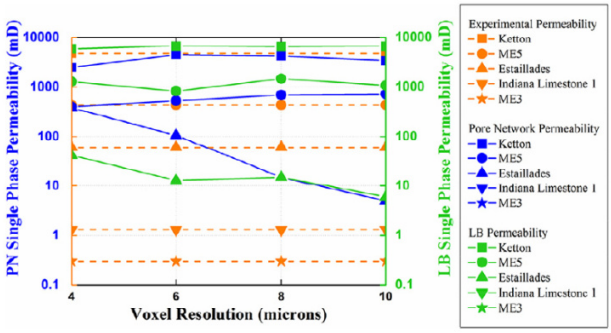
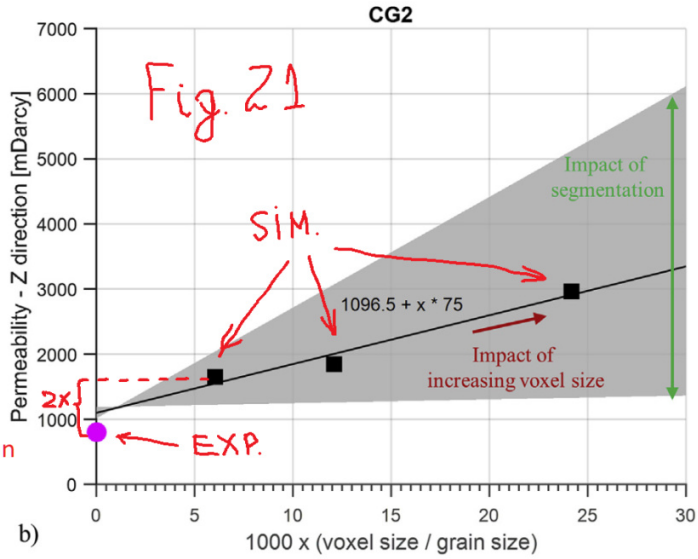


Fig. 6. Predicted PN and LB single-phase permeability as a function of scanned v

Or here http://dx.doi.org/10.1016/j.fuel.2014.12.079 (new citation):

Table 1, row 2 (0.045Darcy, exp) vs Table 4, row 1 (0.12e-11m^2 or 1.2Darcy, sim, pore-network).

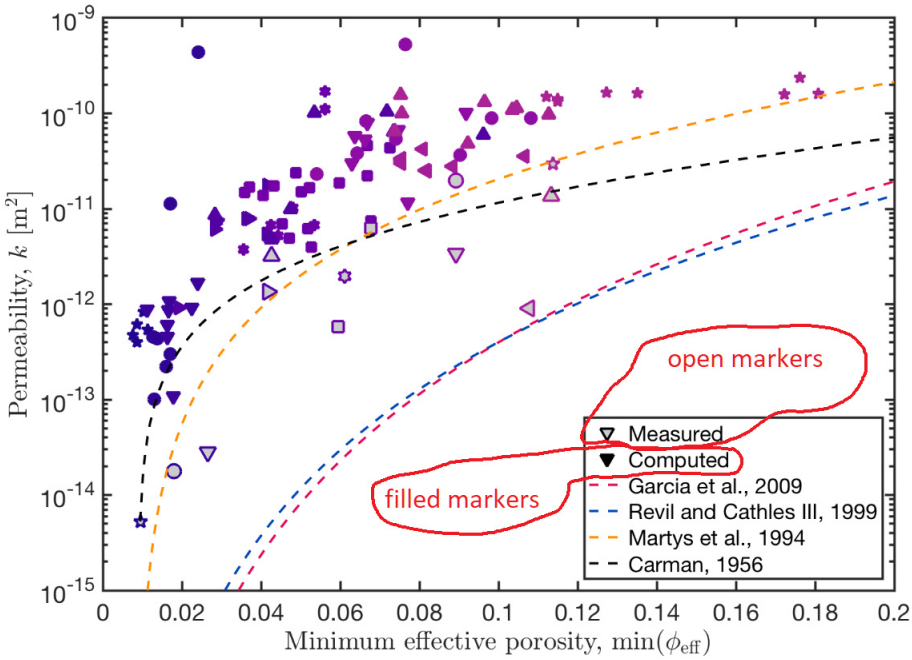
Or here http://dx.doi.org/10.1016/j.marpetgeo.2017.07.004 (new citation, one also check their Fig. 11a):



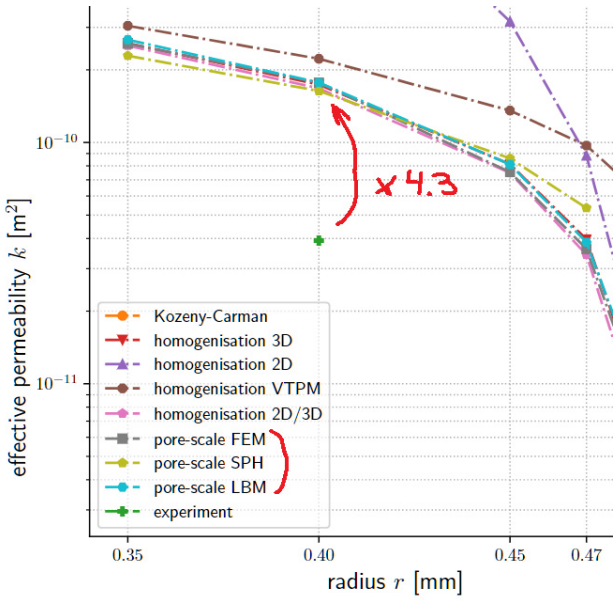
It is funny to see such a kind of extraploation indicated by the black line. Compare it and its scale to what we do in Fig. 3C.

Sorry, but I see here overestimation by factor of 2. And nothing else.

Or here https://doi.org/10.5194/se-11-1079-2020 (was cited in different context but now added to the 1.4.3 context):



Or here https://doi.org/10.1007/s11242-021-01586-2 (was cited in different context but now added to the 1.4.3 context, please also see the discussion in our reply to Referee 2):



==========

1.4.4 Page 5, bottom paragraph: How do you conclude that <10% of pressure loss occur in the empty sample holder and tubes?

Good question. We followed the logic described in ASTM D5084 – 16a standard, which also seems to be a general approach to estimate the so-called “parasitic head losses”:

5.2.3 *Head Losses*—Head losses in the tubes, valves, porous end pieces, and filter paper may lead to error. To guard against such errors, the permeameter shall be assembled with no specimen inside and then the hydraulic system filled.

5.2.3.2 *Constant Rate of Flow*—If a constant rate of flow test is to be used, the rate of flow to be used in testing a specimen shall be supplied to the permeameter and the head loss measured. The head loss without a specimen shall be less than 0.1 times the head loss when a specimen is present.

"0.1 times" is some ad hoc number here, and it is just a coincidence with the estimates we got for our system

We performed LBM simulations on empty hollow cylinders, which are tubes + plugs + sample holder. This is a flow through 3 coaxial cylinders installed in a series. We calculated the dimensionless permeability for such a system, k_empty, using equal cross-section area of the sample holder over the full system, and compared it to the permeability of the full system k_system with the sample (tubes + plugs + sample inside holder). k_system/k_empty ~= 6.2%. That is., the pressure loss without a sample is about 16 times smaller compared with the pressure loss with the sample installed. Remarks: r1) The use of "equal cross-section area" is the knowledge from the second part of the manuscript (after Fig. 4), but permeability of such a virtual system is exactly "what is measured". r2) we used S3 system (with P3 sample) for such a calculation, because the system S2 with P2 sample (P2 perm is shown in Fig. 4) has 74% longer tubing (plugs+tubes) and we simply cannot simulate it. Sample P2 is 15% longer than P3 (6.8cm vs 5.9cm) and the corresponding system S2 has 74% longer tubing compared to S3 (26.5cm vs 15.2cm). Based on these numbers we expect similar ratio for k_system/k_empty (~<10%) for S2 because the tube length variation has a weaker impact on the system permeability compared with the sample variation.

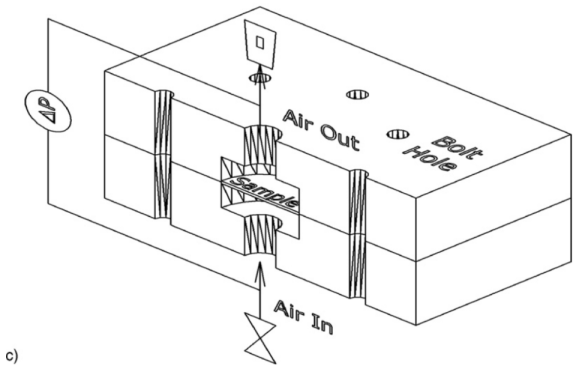
Later we found that there is another way to perform such an analysis. We checked our estimations using online Poiseuille's flow calculator, for example this one: https://www.calctool.org/CALC/eng/fluid/hagen-poiseuille, and got identical results with the analysis above, while comparing pressure drops from the calculator (empty system, using d ~= 1.8mm and L ~= 15cm or 26cm) and our experimental pressure drops (with the sample installed). More specifically, for the system S2 we got k_system/k_empty ~= 9%.

The text in Section 3.1 is updated accordingly.

==========

1.4.5 p. 8, lower third: what are the studies, where such ratios are common for the permeability measurements of membranes?

Here https://dx.doi.org/10.1016/j.jpowsour.2006.06.096 (was cited in different context, but now added to the 1.4.5 context):



Or here http://dx.doi.org/10.1016/j.memsci.2017.02.051 (new citation):

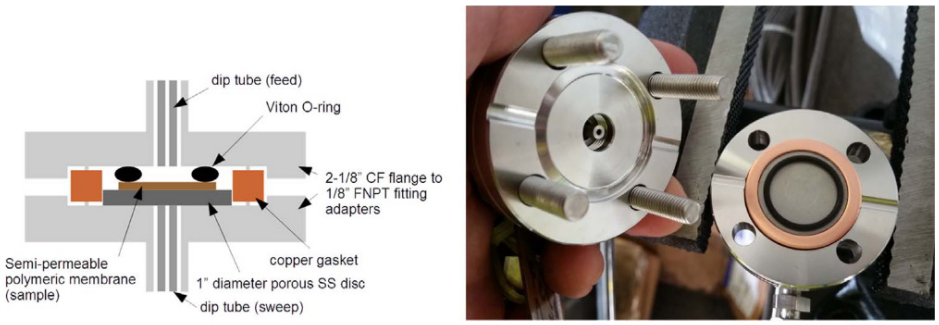
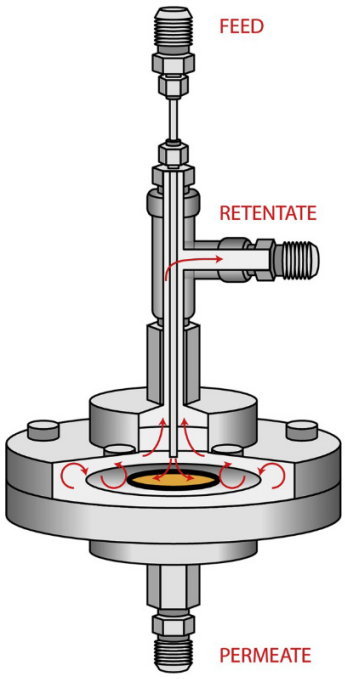


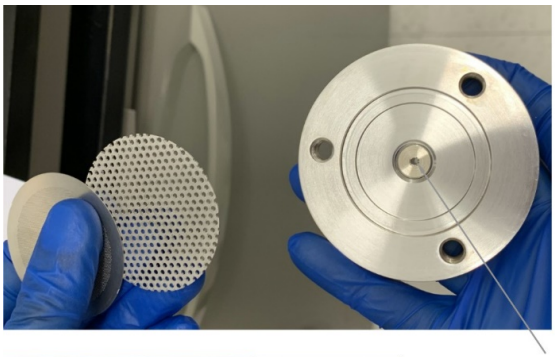
Fig. 3. a) Schematic and b) picture of the CF flange flat sheet membrane cell assembly.

Or here https://doi.org/10.1016/j.memsci.2019.117497 (new citation):



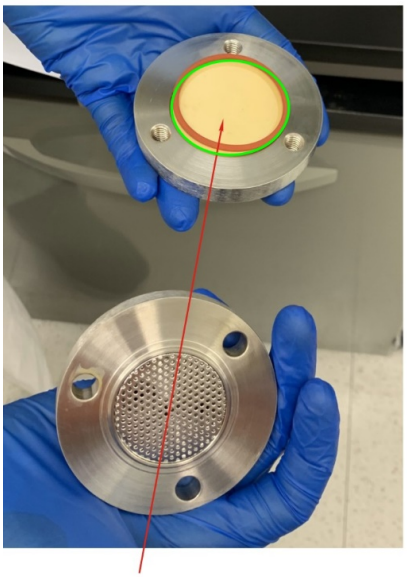
Or here https://www.con-vergence.com/product/gas-permeation-test-cell/ (a commercial product):

Or here, our personal experience:

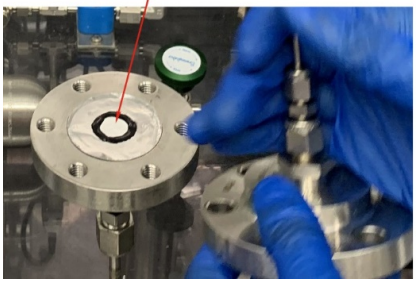


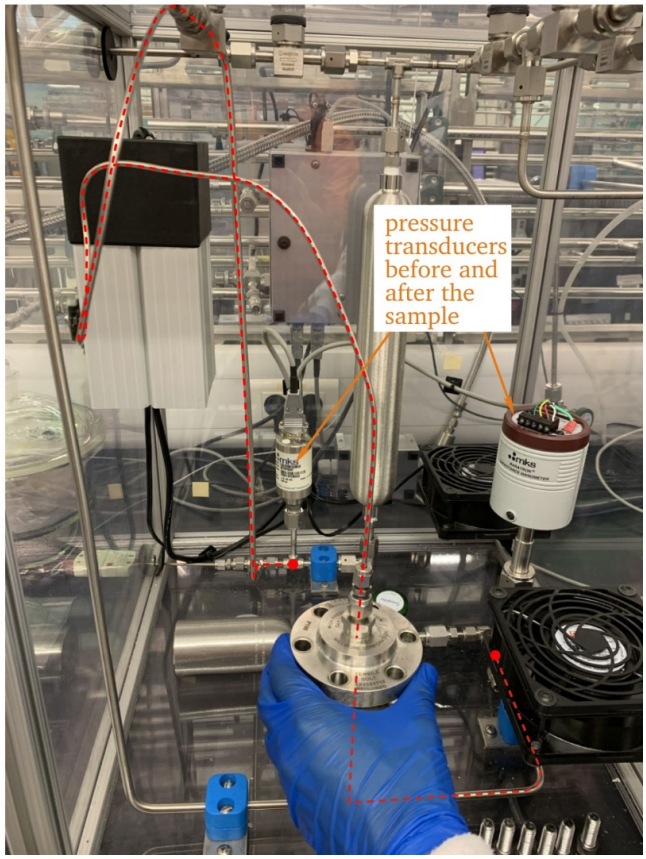
These pictures are from another center at our institution. They design various membranes and also estimate their permeability. The flaw is deeply rooted in their equipment.

tube–sample contact



membrane samples





In addition, I found the style of writing hard to comprehend. The continuous referring to the extensive appendices (that are three times longer than the original article) was very distracting. On the other hand, some of those appended information were essential to understand the article. I recommend rewriting to create a self-contained article. Here the problem is that we have two messages: 1) how exp-sim permeability match was obtained, and 2) what are the implications (i.e., the flaw). From our experience, while presenting this study to others, the first part can take, for example, 45 min out of 60 min presentation, and after that the audience (or readership) is already exhausted when we start with the second part. But the second part seems to be of more practical importance and the main message of this work. That is why we organized our work in such a fashion. Please also keep in mind that during submission we had to split our work into two different .pdf files due to the very strict (and unnecessary) 6MB file size limitation. Such a split broke the internal .pdf hyperlinks. The final paper .pdf should be one file, and we believe that working clickable hyperlinks within that single .pdf will make navigation through the paper content significantly easier. At a risk of being repetitive, we wanted to show the entire story, warts-and-all. This is difficult and tedious. It took us at least 12 different paper iterations to arrive at the manuscript you read. Another, expedient solution would be to gloss over most details, and present the results and conclusions that are not supported by our own experimental evidence. In doing so, we would make this paper similar to the papers cited above; cute and substantially indefensible.

In addition, while the sub-panels are nicely done in general, the numbering of the panels was not done consecutively and it was hard to intuitively identify the corresponding sub panels that are being referred to in the text. First, we have only two 1) there are only two figures out of 20+ with non-straightforward enumeration (Figs. 1 and 6). 2) This has its own reason, too. Each figure occupies a rectangular frame. In principle, this frame can either be filled with useful information (i.e., panels, legend) or with gaps. Each sub-panel is also a rectangle of a fixed aspect-ratio. Our approach was first to place and scale the sub-panels to minimize gaps/maximize size of each sub-panel, and only then to enumerate them. In Fig. 1 we renumbered two panels to better follow the sequence of employed procedures (experiment -> imaging -> simulation), but still they are enumerated from the bottom to top. In Fig. B1 we moved panel “i” to the bottom to follow top-to-bottom enumeration.

Specific comments:

1.1s) Why enumerate as 1a) and 1b) in the abstract?

We renumbered the identified issues as 1), 2), 3). The original intention was to group issue 1) (incorrect pressure gradient + omission of the virtual sample part) and to distinguish it from issue 2) (pressure loss at sample–tube contact).

1.2s) p. 3: How are P2, P3, P4 defined?

We added the following text: “Beads1 were packed inside the samples P2 and P3, while beads2 — inside P4. Samples P2, P3, P4 have the length (porosity) of 6.87 cm (35.7%), 5.96 cm (35.4%), 6.03 cm (35.7%), respectively. The tubing length, including plugs, is 26.5\,cm for P2 and 15.2\,cm for P3 and P4. Please see Appendix A.3 for more details on the samples and packing procedure.”

1.3s) p. 4: properly cite Lo1952, Ja1975

Done.

1.4s) p. 6, figure 4: author should demonstrate repeatability and introduce error bars

Experimental values scatter as in Fig. 2, i.e., within 1-2%. Simulation values have similar accuracy if we do not consider image contrast, which shifts black cross down by ~10% we mention in the main text. We added the following text to the

figure caption: “Uncertainty of our experimental and simulated values is within symbol size of both black and red crosses, respectively. However, this uncertainty does not include the impact of pressure losses in the system without the sample installed or due to the image contrast (see main text).”

1.5s) The numerical code needs to be thoroughly validated as the authors claim that this is their tool of choice to validate the experiments. The employed code was validated in detail with basic tests such as flow in horizontal and diagonal channels (http://dx.doi.org/10.1016/j.jcp.2014.10.038), where it agrees with analytical solution within machine accuracy while using Λ = 3/16 and 3/8, respectively. The same code reproduced the reference permeability values for BCC, FCC, SC packings obtained with LBM+multi-reflection boundary condition (~10^{-4} accurate, resolution and Λ-independent, tables 7,8,9 in 10.1016/j.jcp.2014.10.038) or from Larson&Higdon 1989 (https://doi.org/10.1063/1.857545). The same code allowed us to identify a flaw in the reference solutions of Larson&Higdon 1989 for overlapping BCC, FCC, SC packs of low porosities, reproducing also original flawed friction coefficients (https://doi.org/10.1063/1.5116700). Additionally, the flow fields produced by our code were validated by another group against a Palabos lbm implementation, and agreement within machine accuracy was observed for the same lattice type (D3Q19) and identical choice of LBM collision parameters.

========== Referee: 2 ==========

Comments to the Author

The manuscript presents a numerical and experimental study to estimate the permeability of a porous medium.

The idea is to demonstrate that the usual measurement of permeability is generally incorrect due to the pressure drop between the pressure port and the sample.

Although I am not an experimentalist, I am extremely skeptical of this manuscript.

The problem is that I simply cannot believe that all researchers and engineers who measure permeability forget about the presence of tubes. This is something that can be estimated quickly by a graduate student using the analogy of hydraulic conductances in serie.

Of course, this analogy is not exact because the pressure is not perfectly homogeneous everywhere, but it is unlikely to lead to errors "within a decade".

Even if the authors are right and no one thought about it, I don't think Journal of Fluid Mechanics is the appropriate journal. It would be better suited to a journal specialized in experimental techniques in porous media.

I am very sorry because the authors have put a lot of effort into these experiments and numerical simulations, but my conclusion is that the manuscript is not suitable for publication in JFM.

We appreciate the enormous effort of Referee2 to evaluate our work. Nevertheless, Referee2 recognizes that if, and only if, our finding is valid then its scale is significant and using “major flaw” in our abstract makes sense. We appreciate that, too.

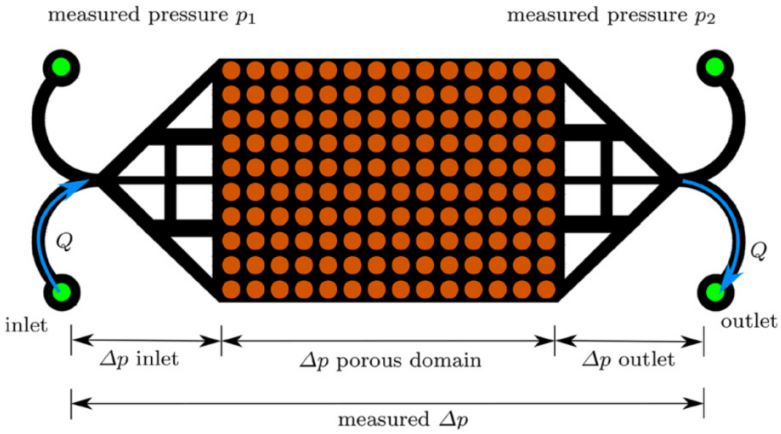
More specifically:

Well, we do not state that “all researchers and engineers who measure permeability forget about the presence of tubes”. Common understanding is that large internal tube diameter (say, 1mm) has negligible impact on the sample with small pores (say, micrometers). We *demonstrate* that this is not always the case, and what the implications are. Sometimes the permeability error can be below equipment accuracy. But the most important part is the presence of the virtual geometry while connecting pressure ports to another sample from outside. This allows to accurately connect real-life systems with their virtual counterparts (aka digital twins or models).

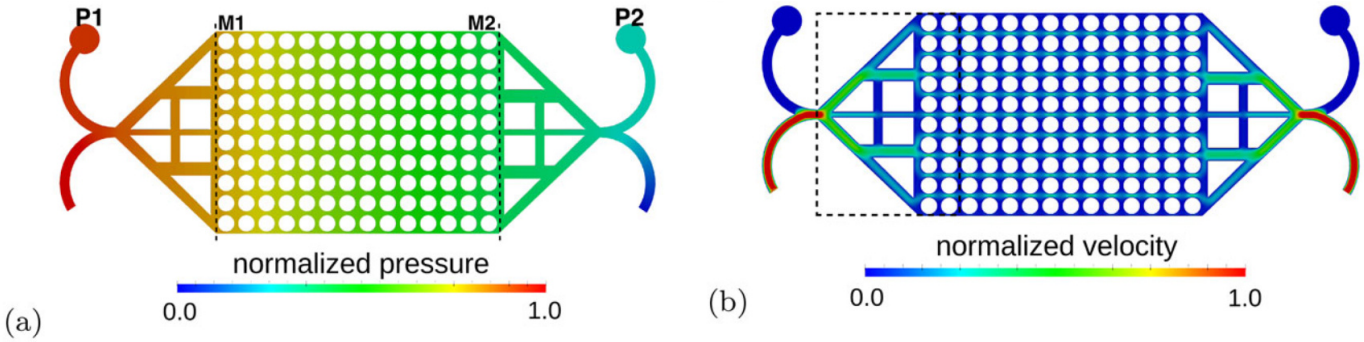
Concerning the statement “can be estimated quickly by a graduate student” — first, there are no real-life samples with accurate -reference- permeability values. (In principle, this phrase is enough to terminate further discussion, but we continue.) Therefore, when another permeability value is measured, with what can it be compared? Experienced scientists may know that densely packed spherical beads should follow the Kozeny-Carman relation with constant k_kc ~= 180, the fact we also employed in Fig. 4, but even here sometimes k_kc = 150 is used. And, second, from an extensive internet and literature survey we got the impression that when experimentalists see some anomaly in the measured values with external pressure ports, they simply switch to internal ports. Unfortunately, we had found before but were unable to find again the particular text supporting this statement. Originally this text was found while surveying studies by (or related to) Timothy Scheibe from the Pacific Northwest National Laboratory.

We also have a very good recent example from a 2021 article https://doi.org/10.1007/s11242-021-01586-2 authored by 14 mature scientists. In this work, the authors 3D-printed the following micromodel:

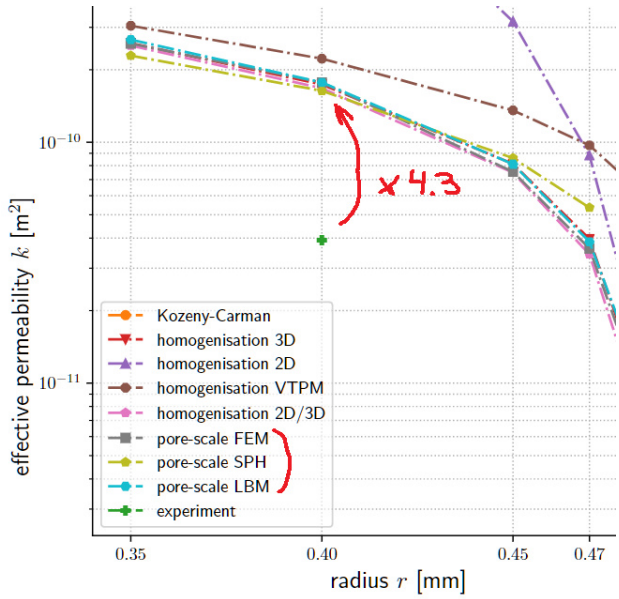
Fig. 6 Design of the PDMS micromodel



which is composed of cylinders confined from top and bottom by walls. Exactly this geometry, together with inlet and outlet ports, was used in the 3D pore-scale lattice Boltzmann simulations:



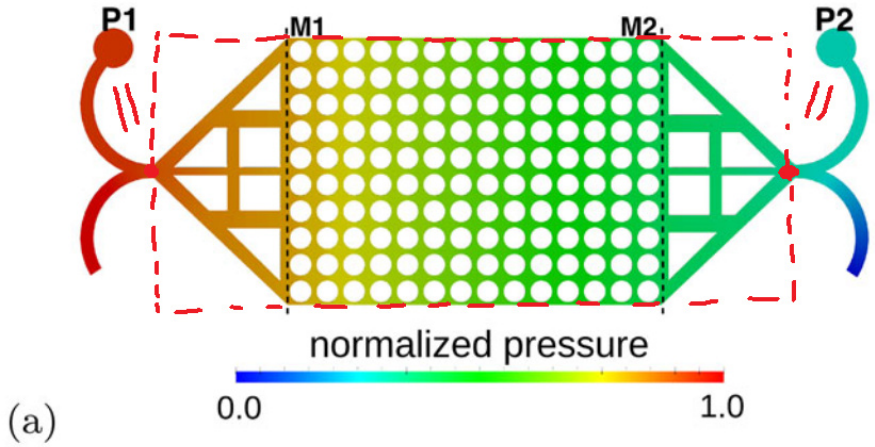
Later, the authors took one cylinder, confined from top and bottom in 3D, assumed it to be an REV for this system, applied three pore-scale methods (finite element, smooth-particle hydrodynamics and lattice Boltzmann), and obtained identical simulation results which all together are a factor of 4.3 (or +330%) off from the experimental value:



And this was their best -explicitly stated- exp-sim permeability match for the geometry which is not even scanned but known. A more careful examination of their work revealed the following:

The pressure drop across the full device ($p_2 - p_1$) was calculated as $\Delta p^{\mathrm{LBM}} = 3.96$ mbar from LBM simulations, which is slightly different from the experimental measurement $\Delta p^{\mathrm{exp}} = 4.2 \pm 1.2$ mbar, but well within the error margin. This indicates that the inclusion of the inlet and outlet in the numerical simulations results in a better agreement with the experimental values, compared to the LBM simulations without the inlet and outlet (Fig. 5).

where they matched very well the simulated and experimental pressure drops. The pressure drop is *the only* output quantity they measured (with the flow rate and viscosity as input quantities). In other words, they matched the simulated and experimental *permeabilities*, but simply did not realize this fact. In order to obtain the match, they needed to use what we report in our work, namely 1) correct system length for the pressure-drop over length calculation (between red dots, not between M1 and M2) and 2) virtual geometry of -equal cross-section- for the full system (the red dashed rectangle) to get correct porosity and obtain correct superficial velocity from the simulated flow field:

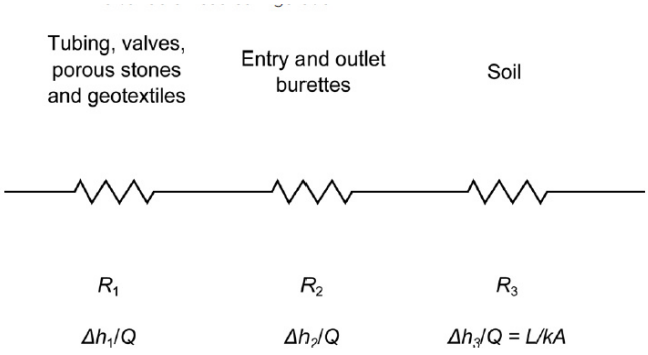


Therefore, the flaw we report looks trivial once stated and understood. But it becomes trivial only *after it is stated, not before*. Referee2 states that this is a simple issue which "can be estimated quickly by a graduate student using the analogy of hydraulic conductances in serie[s]". It seems to us that the students cannot ☺

==========

Here is one more example from a different field, where the authors actually took the "analogy of hydraulic conductances in serie" and estimated what they call parasitic head losses. About the authors, they are not graduate students, and one of them (Chapuis, R. P.) seems to be in charge of maintaining ASTM standards related to permeability measurements. Moreover, he is the author of a review (https://dx.doi.org/10.1007/s10064-012-0418-7) listing 14 common mistakes in the permeability measurements of soils. In other words, these authors have decades of experience in permeability measurements, and it seems that they understand what they are doing while also teaching others, at least in their field, what should and should not do during permeability measurements.

Now, here are the conductances–resistances (we attach the full paper "conductances-resistances.pdf" to our reply because it may be quite tricky to get free access to it):



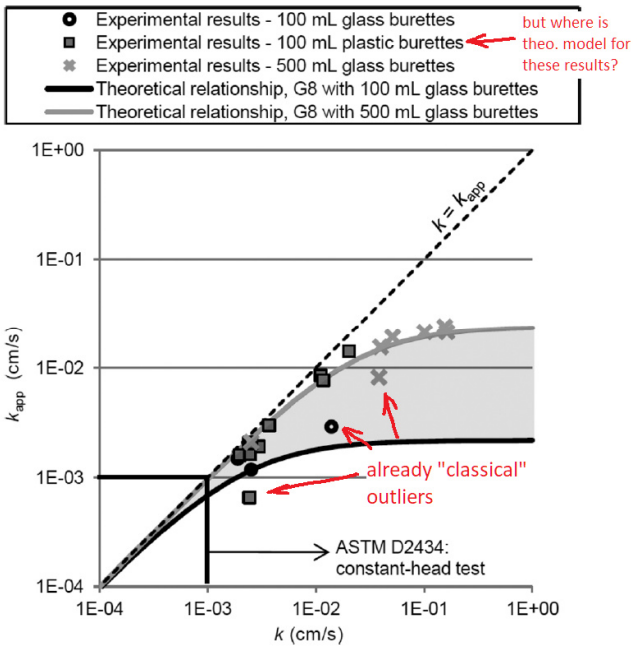
Based on the electric analogy and Eq 4, the relationship between k_{app} and the real soil hydraulic conductivity k, the one obtained with lateral piezometers during constant-head tests, is given by the following equation:

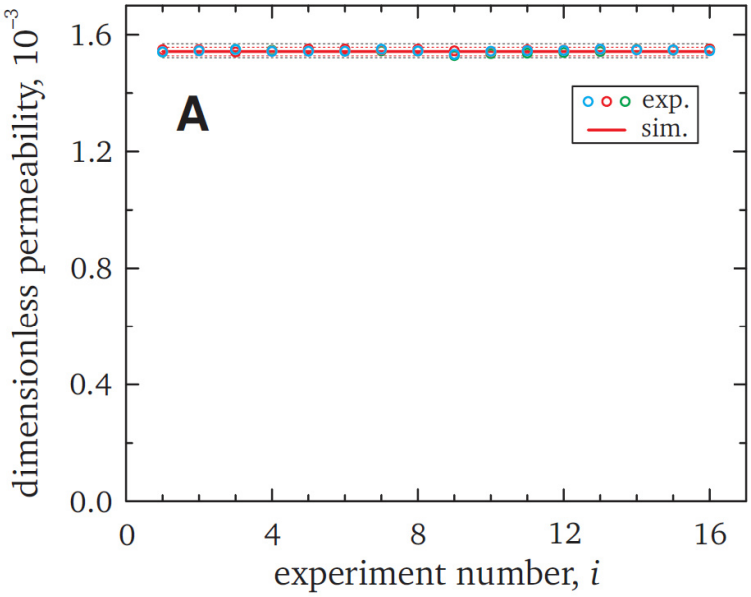
$$k_{app} = \frac{1}{\frac{1}{k} + \frac{A}{L}\left(\frac{\Delta h_1}{Q} + \frac{\Delta h_2}{Q}\right)} \quad (5)$$

where:

$\Delta h_1/Q$ and $\Delta h_2/Q$ = the two parasitic head-loss components that are presented on **Fig. 2**.

And here are their results or predictions (left panel) and our counterpart (right panel):





In principle, yes, their model offsets something. But summarizing this study, to what extent the robust quantitative conclusions can be made from the left panel as opposed to the right one? These are two different realities.

Now one can ask a few simple questions: why the left panel shows permeability-vs-permeability, k_{app} vs k, but not *the relative error* in predicted and apparent permeabilities? Why does the left panel use log scale? I.e., are the sample dimensions, viscosity, flow rate (time, water volume), etc. measured so inaccurately? Why there are outliers and what causes them? Why the left panel uses dimensional but not dimensionless permeabilities? Removing dimensionality by normalizing with (grain diameter)^2 would reveal a lot of interesting (and/or ugly) aspects, also helping to avoid log-scale to hide things.

In the left panel, the authors wanted to demonstrate the statement from ASTM standards that for low-permeability samples (<10^{-3} cm/s) parasitic head loses are unimportant and external pressure ports can be used. That panel was designed to support this point of view.

We added the following paragraph to highlight the difference between our work and the study above:

"For the sake of completeness we note that in geotechnical community porous disks (end pieces) are commonly used to confine a granular sample inside sample holder. The experimental system without the sample but with the porous disks

is used to estimate parasitic head losses. It appears that such porous disks may mostly or partially offset the pressure loss at the sample--tube contact we report here. But such estimations will be impacted by the thickness of porous disks (a mesh confining the sample from both sides is actually a thin porous disk), as well as the ratio of the disk pore size to the sample pore size. Also, a tight placement of each porous disk against the corresponding tube end plays a crucial role."

==========

Our study contains one interesting reference, which probably was not noticed by Referee2. Namely, "Experimental determination of the permeability of textiles: A benchmark exercise", http://dx.doi.org/10.1016/j.compositesa.2011.04.021

A B S T R A C T

In this international permeability benchmark exercise, in-plane permeability data for two reinforcement fabrics, obtained using a total of 16 different experimental procedures, were compared. Although, for each procedure, the results appear consistent, different procedures result in a scatter of up to one order of magnitude in principal permeability values for each fabric at any given fibre volume fraction. The ratio of the principal permeability values varies by factors of up to 2. While experimental uncertainties and variability of the specimens affect the scatter in results for any single series of experiments, it is suspected that the main source of scatter in results from different procedures is related to human factors. Aiming at standardisation of measurement methods and interchangeability of results, "good practice" guidelines will be formulated in order to eliminate sources of scatter.

(with a follow up 3 years later, "Experimental determination of the permeability of engineering textiles: Benchmark **II**", http://dx.doi.org/10.1016/j.compositesa.2014.02.010 , where the scatter in permeability was reduced). In this study, the permeability of fibrous reinforcements was investigated, which may sound very specific, but it appears that such materials are of enormous industrial importance. The authors sent identical -artificial- samples (i.e., of small/controllable variability) to different groups worldwide, asked them to measure permeability, collected the results, and produced the manuscript. Citation from this study (left):

2011

most frequently processed in thin shell-like structures. However, there is a complete lack of standardisation for measurement of the in-plane permeability of fabrics, and it is well known that permeability data obtained using different methods are not necessarily consistent. In 1995, Parnas et al. [12] proposed use of a reference fabric for standardisation of permeability measurement methods, but to date no standards or guidelines have been put in place. Lundström et al. [13] report on a small-scale benchmark

1995

Given the conflicting data in the literature, and the critical need for high quality permeability data, programs were initiated at NIST to develop standard reference materials (SRM) for permeability characterization and standard reference data (SRD) for the composites industry. In this work, an SRM is proposed and initial characterization data reported. The

The very simple and basic question we ask, which also could be addressed to Referee2: why do such studies still exist?

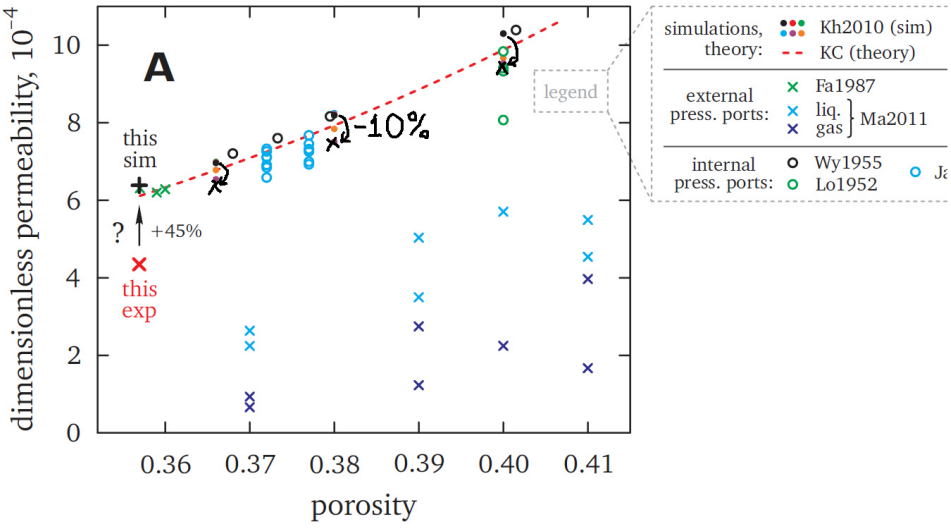
========== Referee: 3 ==========

In this manuscript, authors are proposing an in-depth analysis of the intrinsic permeability measurement of a porous sample, assessed by flow numerical simulations.

The manuscript is very well (sometimes overly) documented (see remarks below). My major comments on this work are the following.

3.1. From the beads sizes and sample holder (inner tube) diameter, it is understood that roughly 18 beads distributed over a sample diameter are expected. This is probably too small to avoid significant effect of the tube wall i) on the packing itself, ii) in terms of the boundary condition (zero velocity) impacting the the core flow (see comment #9 below).

We are not afraid of the wall. Yes, the wall impacts the arrangement of beads inside the tube, and its impact on porosity distribution propagates from the wall across 3-5 bead diameters for monodisperse spheres. But we do not operate with "bulk packing/wall-affected region", and use a "sample", just a sample composed of glass beads. And we perform simulations on the whole sample (+tubing, plugs), which is in fact just a set of `solid` and `fluid` voxels. Wall impact may be discussed in Fig. 4A, where, for example, data from Kh2010 (colored dots) are for the computer-generated bulk (periodic) packings without the wall. At the same time, Table 1.1 in Kh2010 suggests that adding the wall to these bulk packs reduces permeability by ~10% (see also our reply to 3.9). Will this change any conclusions for Fig. 4A? This knowledge was already kept in mind by us once we observed 45% difference for strictly no reason, btw where both simulated and experimental geometries already had a cylindrical wall included.



Additionally, for the employed simulation approach, there are only two voxel types, `solid` and `fluid`. `wall` voxels mentioned in the CT section are treated as `solid` by the flow solver. (But they are intentionally kept in supplementary material geometries to make pore-space-only analysis of the sample more convenient.) From this point of view, the same bounce-back boundary condition is applied at all `solid` voxels mimicking beads, plugs, meshes, wall, and tubing.

3.2. Many pictures are not necessary to a comprehensive reading of the manuscript, nor to justify the results. Authors should consider cautiously reducing these materials (and letting them available upon request as supporting data, for instance). Appendices are 31 pages long and should be shortened as they look like a report more than a set of essential information. Moreover, some repetitions can be noticed (in particular in Appendix A) and must be avoided.

We believe that "available upon request" option will significantly reduce actual availability of the material, taking into account busyness of modern scientists. From our personal experience, we were sending emails to various authors asking

for experimental details. Some of them did not reply, some promised to reply but never did, and some of them actually replied. Also we do not know for how long some of us will stay in what is called “science”, taking into account modern academic practices in promoting paper spam rather than studies like this. We would prefer to keep the figures as they are now, because behind each figure there was a story (or reason) to create and describe it.

For the sake of completeness we mention that a few panels (not whole figures) are in principle optional, like B1,A, but if we remove it, there will be an empty space left in the rectangular frame already dedicated to that figure. We also assume that a picture is worth a thousand words.

“Moreover, some repetitions can be noticed (in particular in Appendix A) and must be avoided.” The only possible repetitions we could identify in Apx. A is the first paragraph where we mention main practical limitations (bead size, viscosity, resolution) from experimental and scanning procedures. But this paragraph is a summary, justifying most of our choices and summarizing key details, such as flow regime (Re values), characteristic pore size, scanning resolution.

3.3. The packing procedure that makes use of an ultrasonic bath does not seem appropriate to ensure mechanically stable beads arrangement. There is no proof provided in the manuscript that this is indeed the case during flushing for permeability measurement. This remains a conjecture.

Well, we only state “which supports the assumption about mechanical stability of individual beads. From the operational experience, there was no evidence of bead displacement within the sample”. I.e. we do not state any proofs but just our assumption, and we share experience to the best of our knowledge. Second, ultrasound vibration by itself does not guarantee anything, but if it results in lower porosity values (~<36%), then it is already a sign of a more stable packing compared to the higher porosity packings. See our arguments in the reply to 3.7. Also, please take a look at this figure:

During packing process the welded tube end was at the bottom and it was tightly closed with the bottom plug (right on the figure). The number of beads to be packed was chosen such that the top plug (left on the figure) does not fit tightly, and leaves a gap. Flow experiments and flushing were performed with inflow from the top plug side. This was done expecting possible packing compaction. If such a compaction would occur, we would move the top plug inside the sample, reducing the gap, available void space, and the sample porosity. But this never was needed, and any attempt to move the plug inside the sample was prevented by the skeleton of the tightly packed beads.

3.4. Nothing is said about the origin of pressure fluctuations observed during the flow experiment that are smoothed by averaging. I suspect this is due to outlet liquid dropping (see next comment). This must be carefully elucidated. It appears that during the multiple paper iterations the corresponding phrase was lost. We recovered it: “The pressure oscillations in Fig. 1E are related to the drop formation and detachment at the outlet. See a more detailed analysis in Fig. A6.”

3.5. The capillary effects at the outlet are artifacts that can (and should) be eliminated. In principle the second pressure transducer could be installed downstream, and pressure oscillations could be suppressed by, for example, adding the second tank. But we decided simply to analyze the capillary effect in detail (Section A.5, Fig. A.6). Raw pressure logs are available as supporting information, and the relative impact of capillary effect reduces with the increase of flow rate

(which we varied 4-fold). According to Fig. 5C,D, there is no systematic difference between minimum and maximum flow rates, which again supports the assumption that the capillary effect was accurately taken into account via P0 analysis.

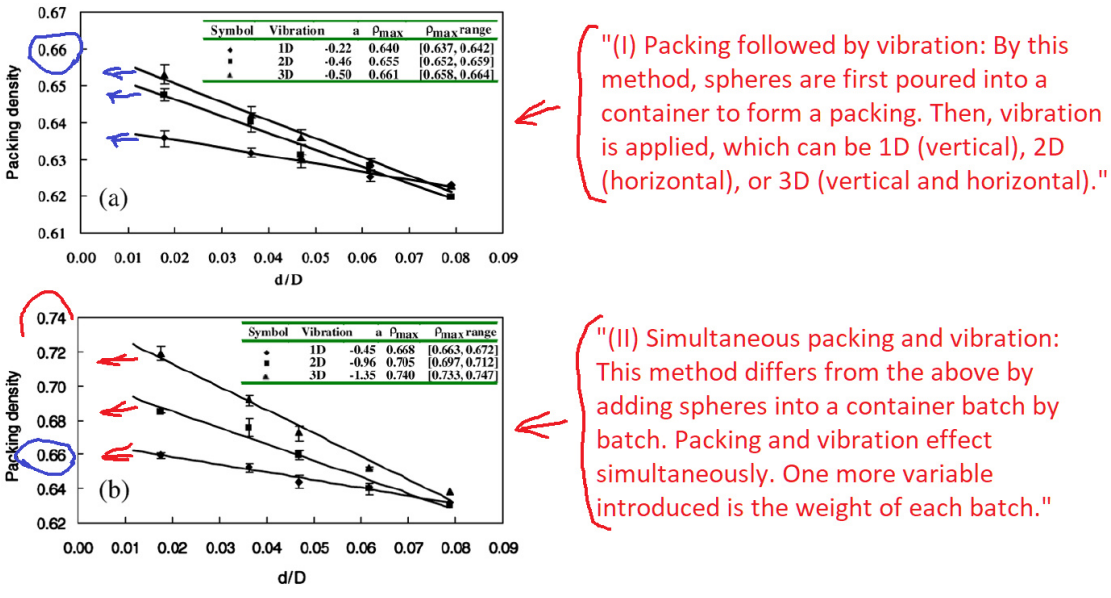
3.6. It would be necessary to provide the range of shear rate expected within the pores during measurement and provide a comparison with the shear rate used during viscosity measurements.

Despite that, we do not understand the reasoning behind such estimation (everything is in linear regime, either during the viscosity or permeability measurements), we added the following text at the end of Apx. A.6: For the sake of completeness, we estimate the shear rate within sample pores, γ_{pores}, during permeability measurements using equation for the shear rate in pipe flow ($\gamma_{pipe} = 8v/d$, where v is the linear fluid velocity and d is the pipe diameter). If we assume the pore diameter to be 1/3 of a bead diameter or 150 microns, and the linear flow velocity of 1mm/s, $\gamma_{pores} \sim = 50/s$.

3.7. Regarding image segmentation, the constraining parameter is a target porosity that is assumed to be uniform, corresponding to the global porosity measured over the whole sample. However, due to the beads packing procedure (batch after batch), it is likely that porosity is not uniform from one image slice to another and therefore, the uniform porosity constraint may introduce some artifact. Authors must address and comment this feature. Figure C1 seems to indicate that such a batch effect is not noticeable. *(Please notice that in the second paragraph at the beginning of Review3 and in 3.2 Referee3 notice overdocumentation and excessive number of figures (details), while one of those optional details, Fig. C1, allows Referee3 immediately test the hypothesis about sample heterogeneity due to batch-by-batch packing)* However, this is biased since the values reported in this figure result from application of filters based on the above mentioned constraint having a smoothing effect.

"it is likely that porosity is not uniform from one image slice to another and therefore, the uniform porosity constraint may introduce some artifact. Authors must address and comment this feature." For this purpose we dedicated the whole subsection B2, analyzing the impact of contrast on the final image. The segmentation rule is "blindly" applied to *each voxel*, and any large batch-scale heterogeneity would be seen immediately, especially at high resolutions, r~=60. Contrast loss is not uniform, and, unsurprisingly, found to be concentrated near bead contacts. But we also analyzed the limiting behavior, when the contrast loss disappears, and segmented real images start matching computer-generated images which has no contrast loss by definition. This match by itself validates the employed segmentation procedure. This is discussed in Apx. C3 and briefly touched in the text related to Fig 3. By the way, this is *extremely* non-trivial moment.

We disagree with the statement "due to the beads packing procedure (batch after batch), it is likely that porosity is not uniform from one image slice to another". Namely, referring to 10.1103/PhysRevLett.97.265501:



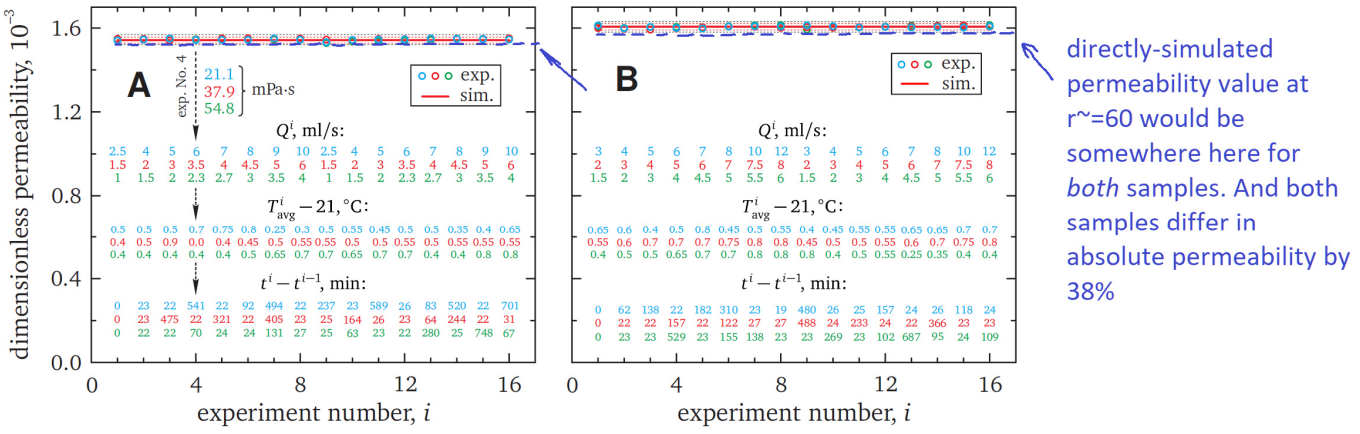
where d/D is bead-to-confining cylinder ratio. We can see that batch-by-batch packing procedure leads to higher packing densities (lower porosities). This is only possible when void space is redistributed more uniformly (and densely) inside the sample during packing process. Our sample preparation procedure was originally designed to be packed batch-by-batch. From our experimental experience we observed that during ultrasound application (by the way, vibrating water was inside the sample at this moment, too), *all* beads start to rotate, vibrate, and settle. And this settling is quite fast, literally within first second (i.e., there was no need to wait 30 sec between batches as we did). After settling, the history of voids due to adding another batch is literally "erased". Moreover, we *assume* that applying vibration for very long time could lead to size segregation of beads.

If one looks at Fig. 4A, it is interesting to note that our samples are the densest among other studies. Denser packing means more uniform packing: there is simply no space available for having (continuous) void gaps here or there. There can be some exceptions from this rule, but they are related to very low porosities, significantly below random-close-packing limit (~0.366 for uniform spheres), where BCC or FCC inclusions have to occur. Otherwise, again, there is no space available to fit beads. But creating such packings is a non-trivial task and not for our procedure.

"Figure C1 seems to indicate that such a batch effect is not noticeable. However, this is biased since the values reported in this figure result from application of filters based on the above mentioned constraint having a smoothing effect" We do not understand which filters are implied here, but each porosity value in Fig. C1 is the averaging over each transverse lattice slice, calculating the fraction of (`fluid` voxels)/(`fluid+solid` voxels). Such averaging in principle can be considered as "filtering". But in any case, all 3D images for all resolutions are available as supporting information, and can be analyzed by any other, "filter-free", method.

==========

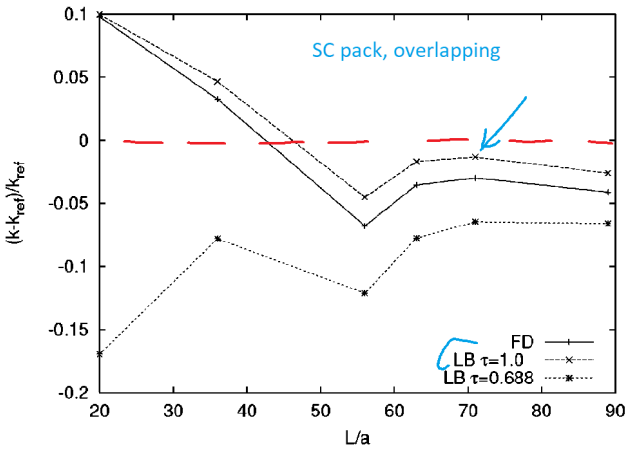
General reply to 3.8–3.11: We could avoid using any extrapolation/error correction approaches and use just the highest available resolution, r ~= 60 voxels/sphere. In this case, we would need about 3^5~=250 times more computational time for each simulation (compared to r~=20, due to O(r^5) complexity scaling), and the (tubing+plugs+sample) system would require about 24h on 500k (half-a-million) CPU cores on available supercomputer (while it has only 200k cores). What would we get at the end? Underestimation of the experiment by ~2% for both samples (not even by ~2.5% because added tubing would slightly reduce the permeability error estimated from the 80%-part of the sample):



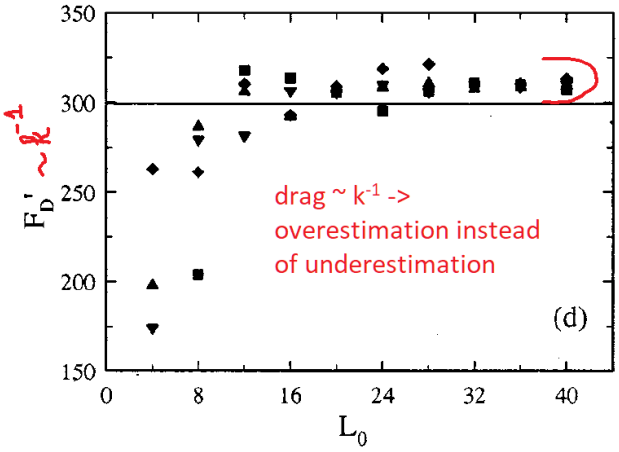
directly-simulated permeability value at r~=60 would be somewhere here for *both* samples. And both samples differ in absolute permeability by 38%

And this underestimation is literally anywhere, where the accurate reference exists:

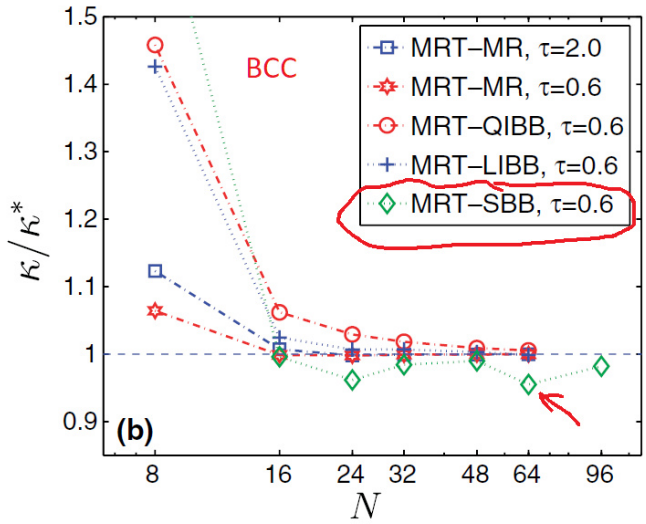
C. Manwart, … R. Hilfer, …, 10.1103/PhysRevE.66.016702

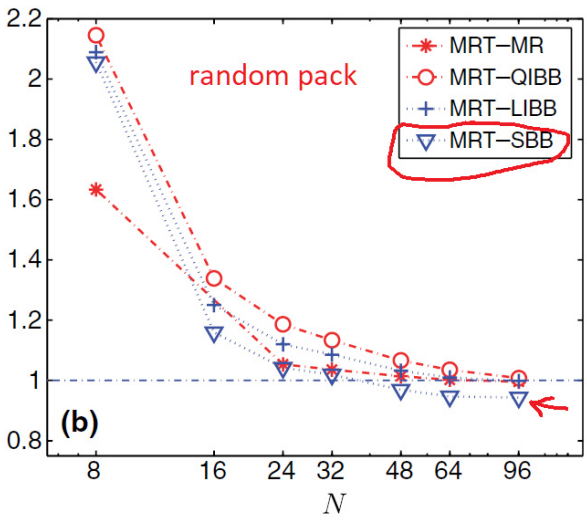


R. Verberg, Ladd AJC, 10.1103/PhysRevE.65.016701

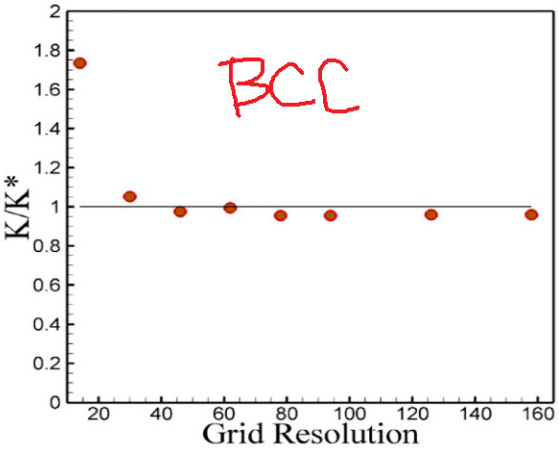


C. Pan, Li-Shi Luo, C. T. Miller, 10.1016/j.compfluid.2005.03.008

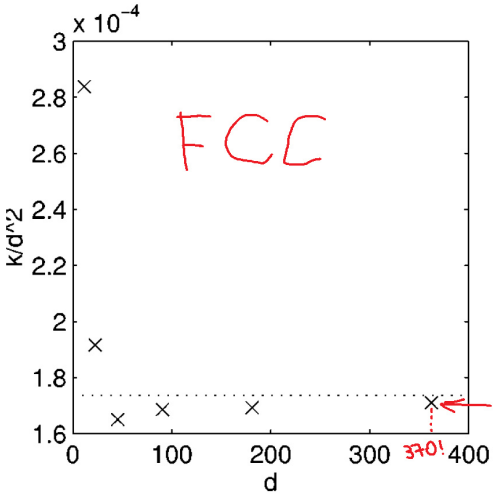




10.1016/j.compfluid.2018.09.020

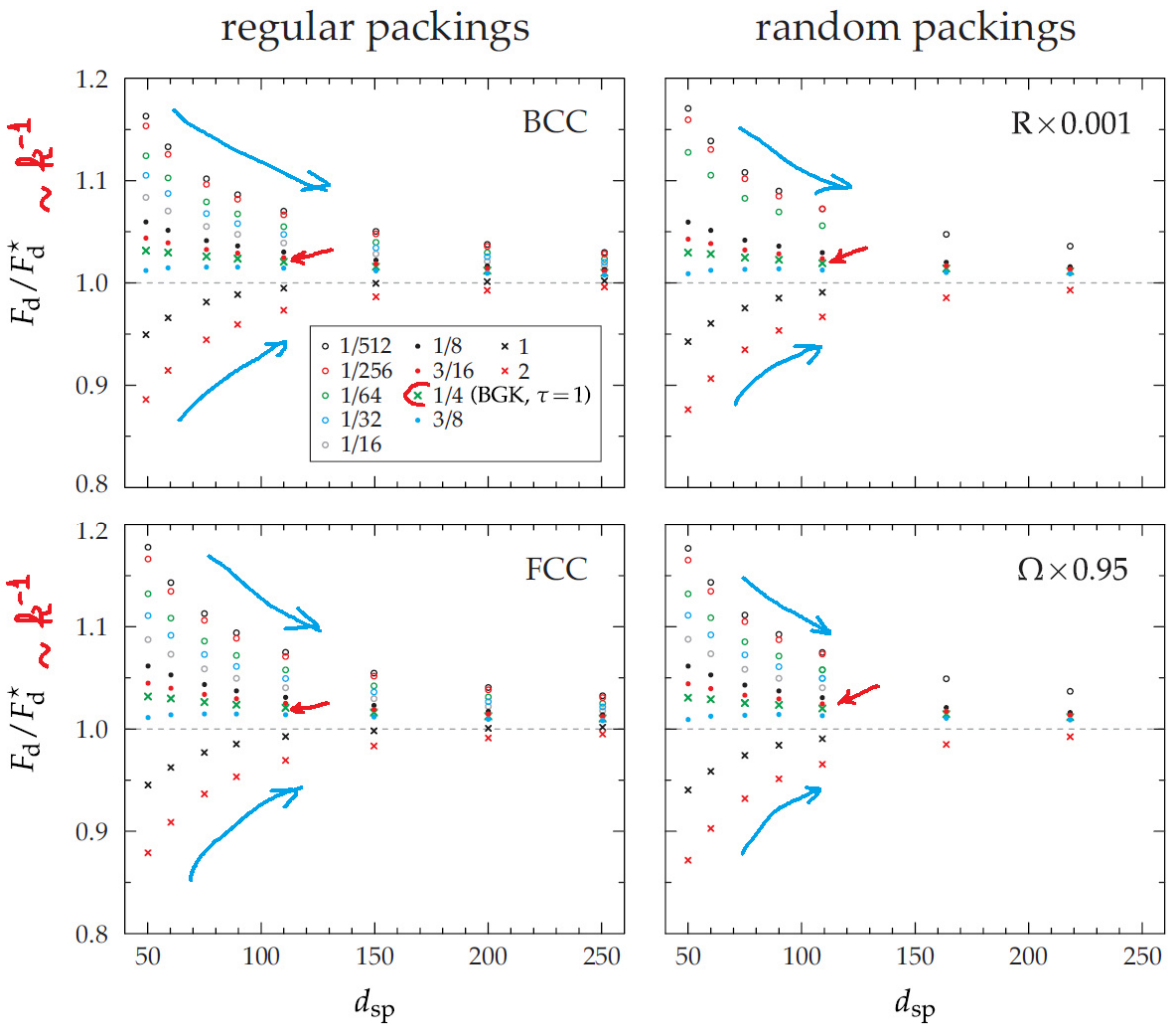


Maier, Bob, … Davis, H. T., … 10.1063/1.869550



If we were reviewing this work with r ~= 60 at Λ = 0.25, we would also ask: and what about the magic number, Λ? Your permeability is Λ-dependent. Λ controls truncation errors and Λ is a free parameter, which has to be chosen, but there is no clear guidance. Therefore your simulated permeability is also a free number. A proper reply could be like this: “Well, yes, it is free, but its impact reduces with the resolution because Λ controls the location of zero-velocity boundary

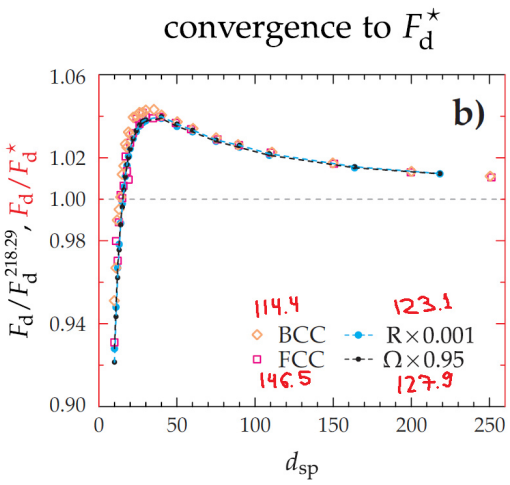
between solid and adjacent fluid node: higher resolution -> fraction of fluid relative to the boundary nodes increases -> less impact of boundary nodes or Λ on permeability, as the blue arrows in the figure below show:"



Then we would continue asking: okay, in the figure above you show that the blue dots (Λ=3/8 providing exact velocity profile for diagonal channel) show smaller error, why did you take Λ=1/4 (green crosses indicated in red) then instead? A proper reply: "Well, it depends on the resolution. For the shown resolution region of quite high resolutions of r>50, Λ=3/8 is a better choice, but if we go towards r~=20, Λ=1/4 shows smaller error for all considered geometries (Fig. 15 in 10.1016/j.jcp.2014.10.038)"

And, therefore, if we want to target sub-percentage permeability error, we will end up with a Λ-resolution *pair*. What we actually did with Λ=0.25 and r=20. To be honest, we could use r<20 or r>20 with another Λ, but this study was designed from the beginning for popular/commonly accepted Λ=0.25 (τ = 1) to avoid unnecessary (at this point) discussions. Adding tubing to the simulation domain was not included in our initial plan. But once the most complex geometry part — porous sample — is clearly understood, dealing with additional over-resolved hollow cylinders is pretty straightforward.

3.8. The second method of permeability estimation (Λ = 0.25) is based on an empirical observation and is not really convincing. Well, please look at this figure from 10.1016/j.jcp.2014.10.038:



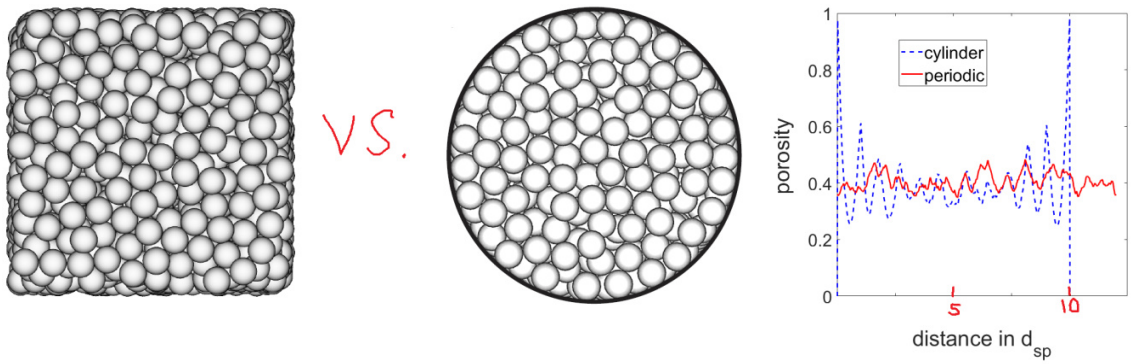
with the reference drag (~k^{-1}) values indicated in red. Notice the Y-axis scale. Permeability (drag) values are 28% different there, but relative errors — almost identical. This cannot be a simple coincidence. But please also note that those geometries have equal porosity, which is important too.

An identical relative error behavior is not surprising. Λ values like 3/16, 3/8, 1/8 having special meaning (exact profiles and exact average flow rate) in open horizontal or diagonal channels geometry become "nothing special" in bead pack geometries (see the 4-panel figure above with the blue arrows). The principal difference between these two geometries is the presence of a spherical grain, which is an object 1) with boundaries not-aligned with the cubic lattice (degrading bounce-back boundary condition accuracy from -2 to -1), and 2) forcing fluid flow into convergence-divergence, which seems to be related to 1). These two facts are the essence of flow in porous media. Then there is no surprise that it is less important how we arrange those grains in space, as the similarity in the relative error above indicates.

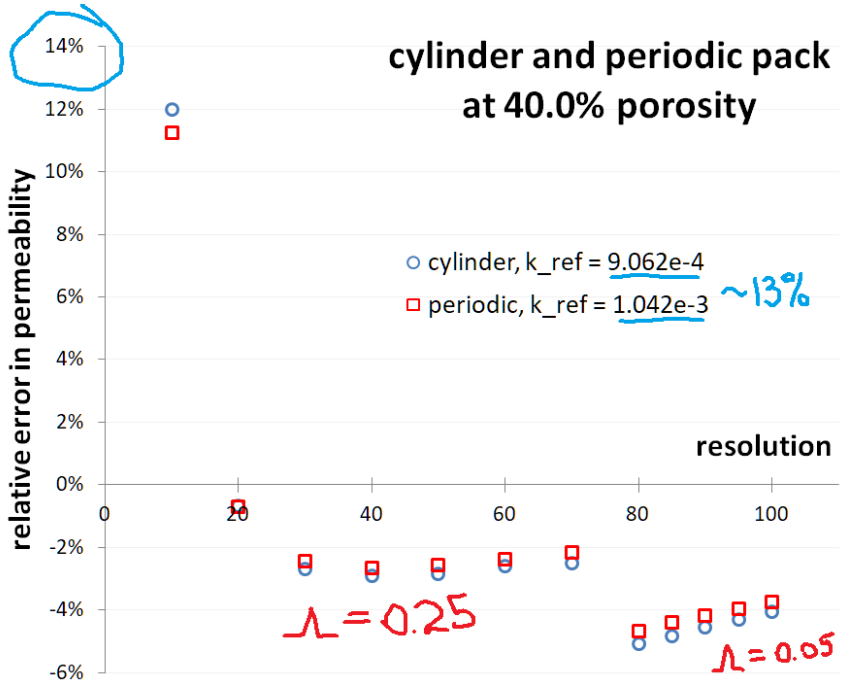
Again, the second method by itself does not guarantee anything. But when two different methods (second and extrapolation) indicate similar permeability values, this is already food for thought.

3.9. The procedure to estimate the permeability on the real structures, consisting in applying the same relative error observed on generated bead packs is subject to a flaw. Indeed, the error estimate carried out on generated media results from periodic structures. Periodicity is not ensured for the real systems and, to my opinion, the fact that the real beads packs only involve less then 20 beads over a tube diameter may induce an effect leading to a different behavior compared to an extremely large domain in which periodic conditions would be reasonable. Consequently, conditions are not the same between the generated and real beds.

Good question. As we replied to 3.1, we are not afraid of the wall. To demonstrate our point, we performed additional simulations for computer-generated periodic and confined packings at even smaller aspect ratio of 10 beads per cylinder diameter, both at fixed porosity of 40.0%, with the corresponding porosity profiles:



We use here monosized spheres making the porosity wall oscillations even stronger while in the manuscript beads are distributed by size which relaxes the wall impact. To find the reference k-values, we blindly apply extrapolation using r = 80…100 and Λ = 0.05 for both systems. Then we perform simulations at lower resolutions at Λ = 0.25 to check the relative error in k. Here is what we got:



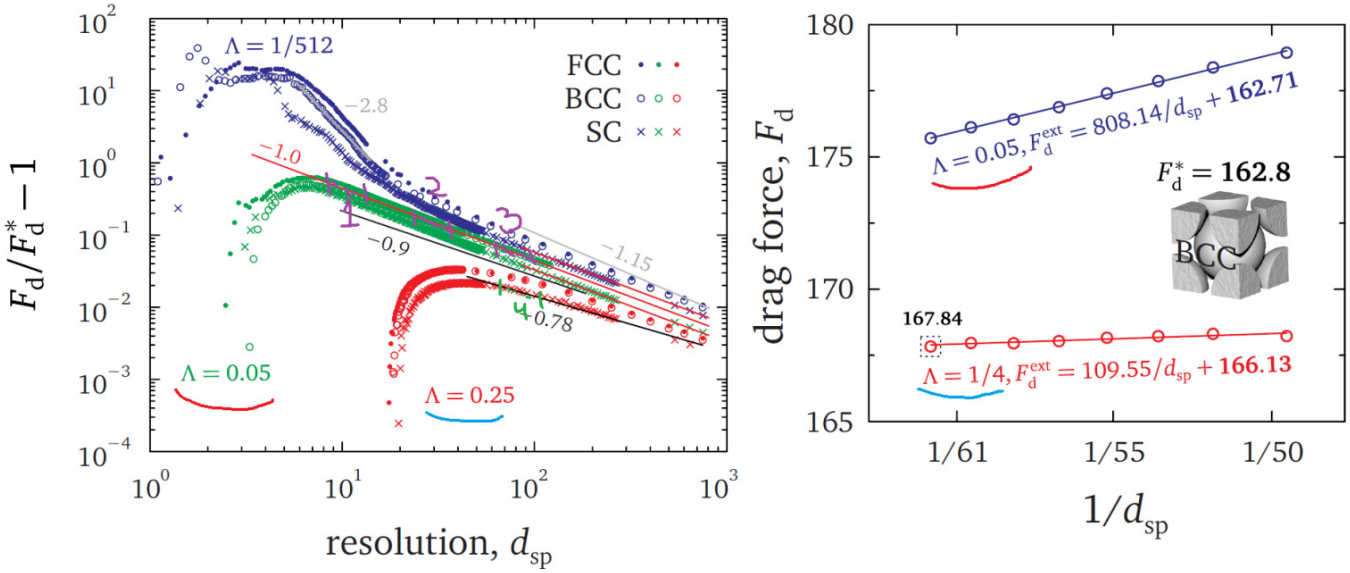
The relative difference in errors for Λ = 0.25 between periodic and confined packs does not exceed ~0.3% (only at resolution 10 it is 0.7%, but we corrected error at r~=60), while their permeabilities differ by ~13% and both absolute k values agree well with Fig. 4A. It seems that adding the wall does not change things much, at least for the considered geometries and porosities. By the way, the error magnitude at r=50 and r=60 is about -2.5% for the two simulated geometries with the porosity = 40% is very similar to our error corrections in the manuscript of -2.7% (Fig. C3) for the porosity of ~36%. Again a coincidence? Is seems that this study is full of coincidences.

In our study we originally planned to generate confined packings, but our algorithm was not able to reach such low porosities as in experiment even for periodic packings (we got only 36.2% against 35.6% in experiment), not even speaking about the confined packings where the wall adds additional restrictions on bead placement.

Hence, even if we do not use the relative error from computer-generated packings at Λ = 0.25, and rely solely on Λ = 0.05 extrapolated values, we will get 5.772e-4 instead of 5.821e-4 for P3 and 6.01e-4 instead of 6.027e-4 for P4 (Fig. C3,E-H), or -0.8% and -0.3%, respectively, becoming even smaller after adding tubing. And please do not forget about the standard brute force approach described in the “General reply to 3.8–3.11”, which always exists.

3.10. Why are the linear fits on figs C3B,D carried out in a piece-wise manner? This seems to mean that a linear regression on the whole range of r is not appropriate. Then I do not understand how the linear extrapolation can be reliable.

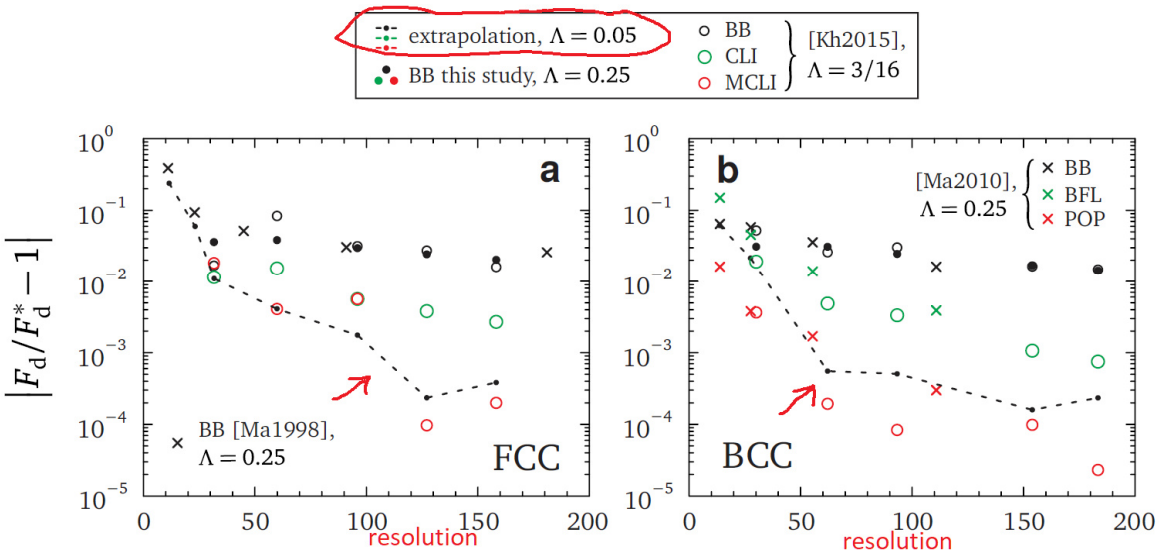
As we state in Apx. C.3, “Taking Λ ~= 0.05 greatly accelerates transition to first-order-only convergence with increasing r”. Strictly speaking, exactly linear convergence is never observed, but k always approaches it with r increasing, sometimes non-monotonically. Please take a look at this figure from 10.1063/1.5042229 :



On the left, one can see that Λ = 0.05 (green symbols) places all geometries in close-to-linear convergence regime for r~>10. However, using strictly linear extrapolation from range1 of the resolution values (the violet “1” range on the left) will result in a larger error in extrapolated permeability (or drag) compared to range2, while permeability from range3 (the violet “3”) will be even more accurate. One can blindly apply the same linear extrapolation to Λ = 0.25 values (the green “4” on the left), but this will result in a much larger permeability error compared with Λ = 0.05. But the evidence of a larger error cannot be seen from just placing a line over several simulated permeability values, as shown in the right panel. Both the blue line with blue circles (Λ = 0.05) and the red line with circles (Λ = 0.25) are excellent fits. But extrapolated value for Λ = 0.25 has clearly larger error (166.13 against 162.8 reference) compared with Λ = 0.05 (162.71 against 162.8). If we increase the resolution significantly, say to r = 1000, Λ = 0.25 extrapolation will also start showing values much closer to 162.8. Because, as we stated, the error always approaches the linear regime originating from the slowest contribution due to the bounce-back boundary condition.

We use piecewise ranges and do not take very wide resolution range (say, r from 10 to 100) for a *single* linear extrapolation because permeabilities at higher resolutions always behave “more linearly” compared to the lower resolutions, and employing small resolutions will introduce unnecessary shift in the extrapolated values. But still, for linear extrapolation, we need to have several points in addition to r_max, and our suggestion was to use the extrapolation range [r_min, r_max] with r_min ~= 0.8 * r_max.

Below is the result of a piecewise linear extrapolation from the same study 10.1063/1.5042229 (see original document for 4 additional panels). Each black dot on a dashed line is a result of another piecewise extrapolation. As one sees, increasing r (which is r_max) and using piecewise extrapolation reduces the error in extrapolated k or drag:



One also sees that for FCC with porosity of 0.26 we need r>100 to reach 10^{-3} accuracy while for BCC with 0.32 the value of r = 60 is already enough. Using this logic, we checked the sensitivity of the extrapolation approach for our particular geometries in C3B,D. But all this was done just to improve sub-percent accuracy. There is always an alternative, less accurate option as described in General reply to 3.8–3.11.

3.11. The result obtained from a down-sampled image (red cross in Fig. 3) is just ad hoc and I am afraid it does not really provide any informative recommendation for other systems. Yes, because there is a fundamental question about how to introduce characteristic length scale, which is the mean bead diameter in our case, to determine the discretization resolution for arbitrary porous media geometry. We do not answer this question (in this study ☺). We designed this study by using spherical grains to avoid getting trapped by this question.

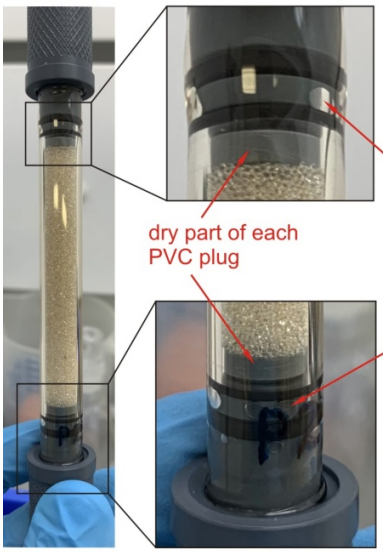
3.12. Some other issues are not considered in this work, including potential trapped air bubbles inside the porous sample when liquid permeability is performed as in the reported work.

It is interesting to note that in the second paragraph at the beginning of Review3 and in 3.2 Referee3 notice overdocumentation and excessive number of figures (details), whereas here (3.12) Referee3 asks for more details ☺

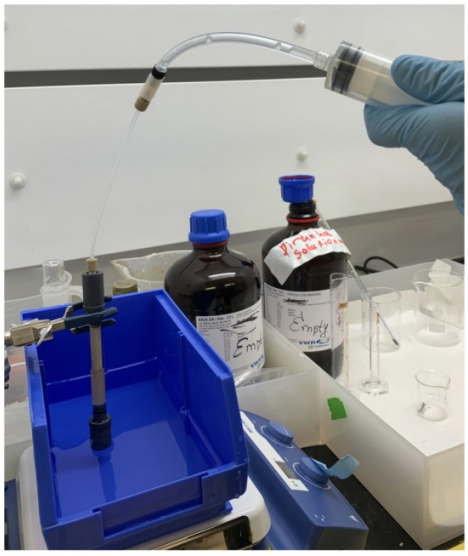
Good practical question. We added the following text: “Despite the packing process with immersion, all three samples contained some air bubbles inside, visible through the transparent confining wall. Placing the sample under vacuum allowed dissolved air and bubbles to leave the sample through the holes in plugs.”

Additionally, we illustrate this degassing procedure only here, to avoid adding more figures to the manuscript. In the figure below we zoom only on the near-plug regions, because it is difficult to picture bubbles in the sample’s center. The largest bubbles always were trapped near the plugs, while degassing approach removes air from the whole sample.

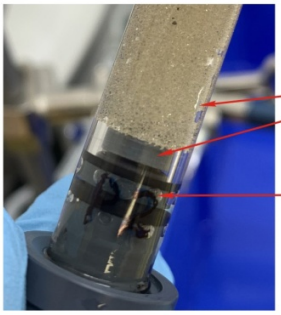
1. Originally dry sample P2

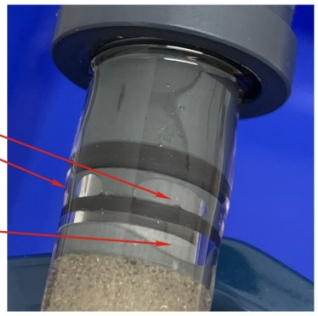


2. Sample saturation with water

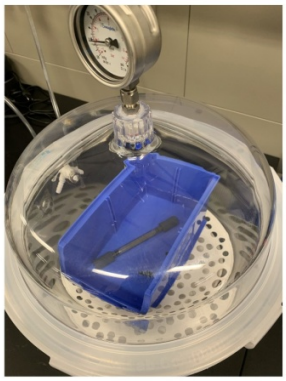


3. Trapped bubbles



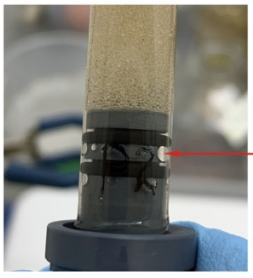


4. Sample is under vacuum

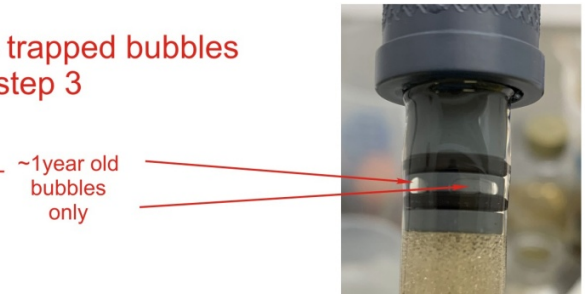


5. After 1h the whole sample is full of bubbles inside. There are also bubbles on the surface of caps — probably that is the trapped air film also affected our density measurements (Apx. A.1)

6. Vacuum release — some magic happens and there are no bubbles anymore at normal pressure



7. No trapped bubbles from step 3



3.13. It is not reasonable to envisage numerical simulations to assess permeability estimate from measurement the way they are conducted in this work in a routine manner. This must be discussed.

Ha! According to this comment it appears that we have overdone some steps (procedures) in our work. Initially our goal was to obtain sub-percent exp-sim agreement needed for another project. The only warning we had in mind, while planning this project, was that nobody before achieved the sub-percent agreement in a clear way, i.e., without free parameter speculation. On the other hand, we did not see any logical reason (but not someone's opinion or belief) why this could not be done. This whole study actually was an iterative approach to the desired sub-percent disagreement, looking for any practical limitation why this could not be done.

The specifics of the problem under investigation are that this problem is unforgiving. Its multidisciplinary nature imposes prohibitively steep entry requirements. Bead density measurements are inaccurate? Sorry, but then it does not matter how accurately pressure was measured or how high the scanning resolution was. Inappropriate sample dimensions choice? Sorry, but then it cannot fit into the scanner field of view at a desired resolution, and others steps are just unimportant. Everything was accurately measured, and, by chance, scanned at appropriate resolution? Then the operator would ruin all the accuracy by subjective judgment, while using operator-dependent segmentation and various image processing routines (to make image "cool"). And the rest does not matter again. Everything was properly scanned and measured, but the flow solver did not converge? Sorry, but the rest does not matter. Again.

May be the problem is not with the execution of our study but with similar type of projects which became "routine" for no reason? The only *practical* explanation we can find here is the blind planning and promising to accomplish just another project within some, strictly speaking, random deadline.

Nevertheless, we added the following text to the Outlook section:

Connecting the reality with its virtual counter-part using only first principles is a very time-consuming process due to its multidisciplinary nature, which is a tight mix of scientific and engineering practices. But at the same time, it allows to reveal the full predictive power of computer models in order to understand and control real-life processes around us.

3.14. The parametric analysis provided in section 3.4 is a bit confusing. Indeed, the relevant dimensionless parameters must be identified in order to make the analysis "universal". I am not sure that the 4 parameters used in this section are the most appropriate ones. This is of importance while expecting that the information provided at this point would serve as a reference for readers to judge on the quality of a measurement performed with a similar setting.

Originally, we considered only Ltube/Lsmpl and, later, Dsmpl/Dtube parameters. But we observed the disagreement in errors for our experimental and model systems (the model suggested below -50% in Fig. 6D for Dsmpl/Dtube=5 while experiment with similar Dsmpl/Dtube showed only ~-30% error relative to the accurate simulations on 3D image). Then we realized that the sample length is the key missing parameter: the very long sample will simply steal most of the pressure losses from the rest of the system, and Dsmpl/Dtube will become unimportant. Afterward, we added Lsmpl/Dsmpl to mimic the relative sample length. And this set of parameters appears to be sufficient to estimate the relative error due to the pressure loss at tube–sample contact, as our experiments in the manuscript suggest, as well as the bonus section at the end of this rebuttal.

We added the following text:

Let us assess the choice of the parameters describing the geometry in Fig.6A,B. Normalized pressure in Fig.6F always drops from 1 to 0 while fluid passes the tube, the tube--sample contact and sample (and the tube--sample contact with the tubing again). Altering the geometrical configuration redistributes the pressure losses between those individual

parts. Namely, an increase of the tube length relative to the rest of the geometry (L_{tube}/L_{smpl}) increases its relative contribution to the total pressure loss. If the sample has higher permeability, the relative contribution of the tube to the total pressure loss also increases (orange diamonds vs. orange dots in Fig.6D). Similarly, larger tubing diameter decreases the corresponding relative pressure loss: compare the slope of the orange and blue symbols in Fig.6D. Once the tubing and sample diameters are fixed (D_{tube}/D_{smpl}), their ratio defines the magnitude of the pressure loss at the tube--sample contact (orange vs. green vs. blue symbols in Fig.6D). After that, the increase of the sample length (L_{smpl}/D_{smpl}) reduces the relative contribution of the tube--sample contact pressure loss (as well as tubing) to the total pressure loss, and vice versa. This is noticeable comparing the filled markers with the dashed line vs. the open markers without line vs. the filled markers with solid line, all in identical color.

Additional comments:

3.1a. A decade of variation is not a relative difference of a 100%, the latter only representing a factor of 2! We were puzzled by this comment. The only explanation we found is that Fig. 6D has the lowest possible value of –100% while we mention “decade variation”. Well, in Fig. 6D even the underestimation of permeability by a factor of 100 will result in –99%. I.e., k_incorrect/k_sample – 1 can never be below –100%. Otherwise, we do not know how to reply to this comment. We totally understand the difference between a decade variation (within 1000%) and two times (+100% or –50%). We also replaced the original (and not very useful) panel E in Fig. 6 to demonstrate the magnitude of reciprocal quantity k_sample/k_incorrect – 1, in %.

3.2a. Viscous dissipation inside the sample should be evaluated in order to check that the 0.2-0.3◦C temperature increase is plausible. We see the only way to estimate the temperature increase on the sample surface due to viscous dissipation is to solve heat equation on a domain mimicking our experimental system using thermal material properties. But even if we performed such a kind of estimation, how would we use this information? In the paper, we just report our laboratory observations: +0.2…0.3 reading of a digital thermometer at the sample surface in all experiments, and T1>=T2. We have no idea what happens inside the sample from the thermal point of view. Simple waiting until cooling down the sample was enough to stabilize the permeability measurements, which was our primary goal.

3.3a. Authors should also consider glycerol adsorption at the glass surface that can potentially create some viscosity gradient inside the pores.

We disagree with Referee3 that such consideration is necessary. Each measurement preceded ~30 pore volumes preflush that would establish equilibrium concentration of glycerol at beads surface. This experimental detail is specified Appendix A.6. Despite this argument, we continue and estimate the potentially adsorbed amount of glycerol and its impact on the injected solution viscosity. Our arguments are the following. Average bead diameter is Dp=0.5 mm. For spheres, specific surface area (or surface-to-volume ratio, Av) is related to bead diameter as Dp = 6/Av. Thus, specific surface area is 1.2e4 m^{-1}. Glycerol molecule diameter is about 10 angstroms and cross section is about ~1e-18 m^2. To populate the surface of all the beads it would require ~3e-3 mol of glycerol per m3 of the packing. The actual packing volume is 3.8 mL. Therefore, it requires ~1e-8 mol of glycerol to cover glass bead surfaces with a monolayer of glycerol (there is no reason to think it would be more than a monolayer coverage, but even 10 monolayers would not make any difference). If all this glycerol is consumed during 1 pore-volume injection, it would result in the concentration change on the order of 1e-6 mol/L. Such a concentration change would induce viscosity variation of about ~1e-6 Pa*s (http://www.met.reading.ac.uk/~sws04cdw/viscosity_calc.html) for our experimental conditions. This accuracy is inaccessible for our measurement equipment.

3.4a. The sample length is never clearly indicated in the manuscript (only in section B.2 of Appendix B). This information should be explicitly provided, together with many other dimensions of importance in the experimental setup.

Please see our reply to 1.2s.

3.5a. Readability of some figures must be improved (e.g., dashed green circle in Fig. B3C can not be identified, drawings on Fig. D1 (Ma2011) are not distinguishable, most of the graphs (in particular Fig. 5) are barely readable).

We appreciate a careful inspection of our figures. We updated Figures B3C and D1. It appears that Referee3 might have used a printed version of our article, which makes it difficult to visualize color drawings over photographs. We added the following text to D1 (and C1 with porosity profiles overlap): “For better experience, please use online version.”

To summarize, I acknowledge the careful inspection of many sources of bias in permeability measurement experiments and their interpretation to determine a porous sample permeability, as well as the accurate report provided in this work.

Nevertheless, it must be admitted that the identified sources of errors are quite trivial (viscosity effects induced by temperature fluctuations, additional pressure drops in core holder system and tubing and hence importance of pressure point measurement placement, singular pressure drops in constriction/enlargements of the connecting system, implying that the overall pressure drop is not just the sum of the regular pressure drops in connecting tubes and sample, etc.). Again, the quantification of these effects is very valuable and they certainly need to be appreciated, depending on the level of permeability, sample size and experimental configuration. Evidently, all the artifacts in experimental settings pointed out by the authors are significant here due to the huge permeability value of the porous bed under consideration ($\simeq$ 170D. We operate with dimensionless permeability which makes our analysis *universal* as long as the Stokes physics is valid), making the additional pressure drops not negligible compared to that induced by the porous medium which is, indeed, the one of interest — please see reply to 1.4.4 as well as reply to Referee2. In the latter reply, their error of 4.3 times most likely originates not from the pressure loss, but from the incorrectly calculated pressure gradient. All the issue is there: a correct estimation of the former which must remain negligible with respect to the later. And that is why we vary in our model Dsample/d_sp, by packing smaller and smaller spheres inside the fixed holder+tubes. What Referee3 implies here is the “parasitic head losses” vs. sample permeability, which is addressed in Fig 6B,D (holder+tubes are fixed -> parasitic head losses are fixed, reduction in bead size in such system -> pressure drop due to the sample increases relative to the parasitic head losses). I think every experimentalist, attentive to produce results of unquestionable quality, are very well aware of this (in particular, the measurement of the sample inlet/outlet pressures (or pressure difference between the two) and potential singular effects at entrance/exit).

The english language of the paper is fine.

On the whole, this is an interesting paper. I can recommend publication in Journal of Fluid Mechanics once the above comments are carefully addressed.

========== BONUS ==========

========== BONUS ==========

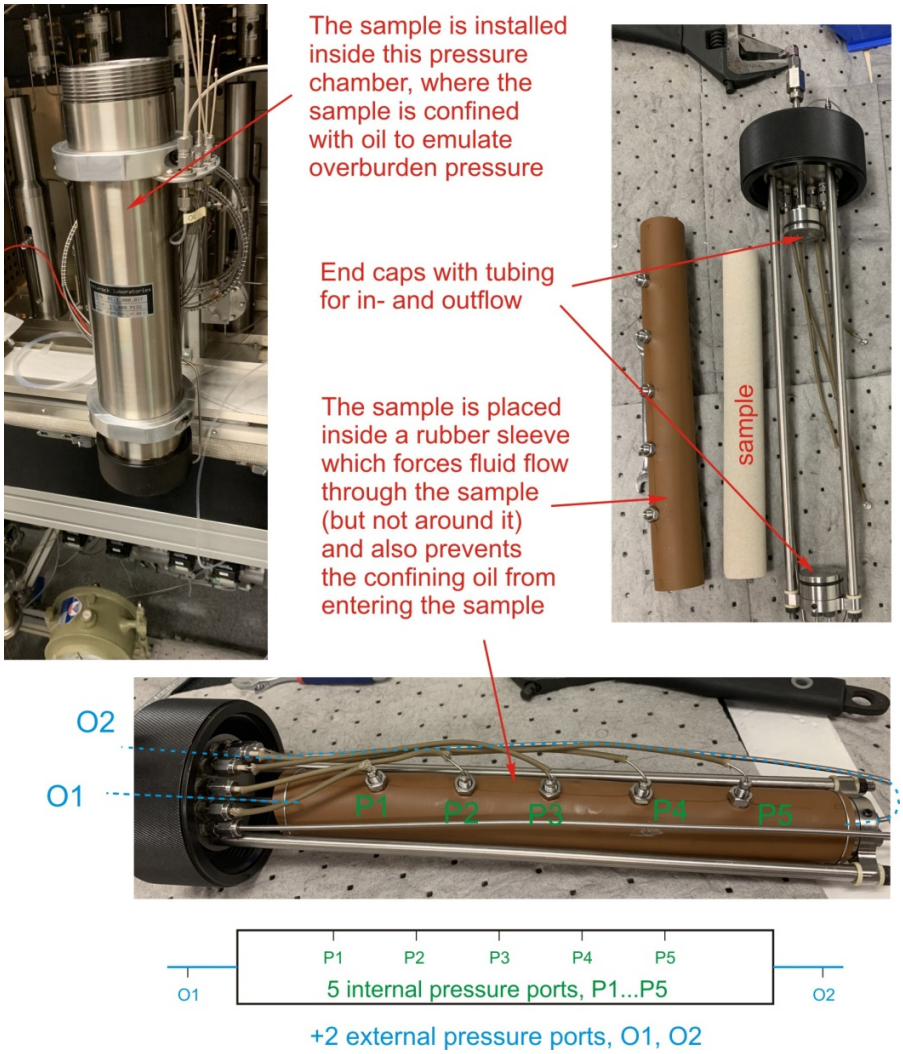
========== BONUS ==========

To demonstrate the reported flaw, we used our Center equipment for flow studies in geological samples. The high-pressure coreflood system described below allows measuring pressure simultaneously by internal and external ports.

Experiment description.

The equipment with internal and external ports operates with significantly longer samples compared to the typical sample dimensions (~30cm or 12”, 6 times longer than the one shown Fig. 6G, while diameter of both samples is the same, 1.5” or 38mm). For the sample we took a sandstone with approximate permeability of 0.5D, targeting the sample homogeneity. The system without the sample was tested first, and the corresponding pressure drop (“parasitic head losses”) was subtracted from the pressure measurements with the sample later.

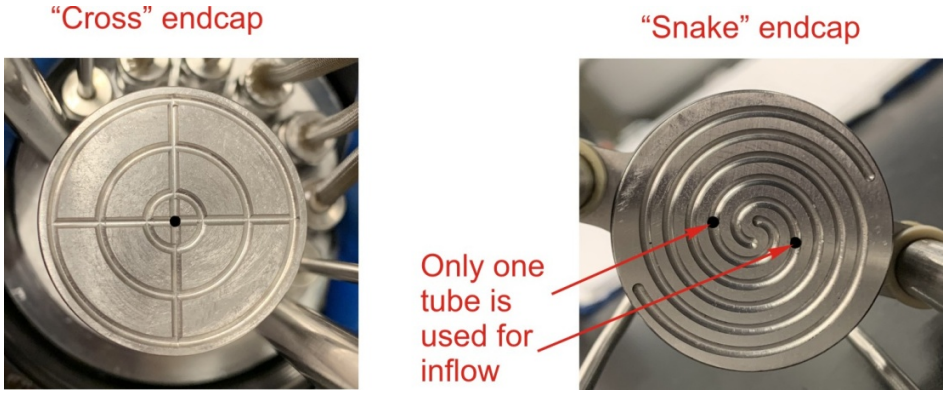
The sample is placed inside the rubber sleeve and is subject to confining pressure:



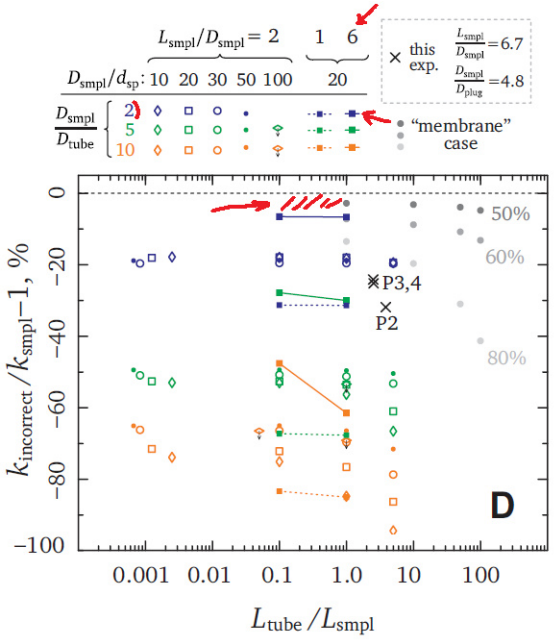
Pressure drops can be calculated between O1 and O2 as well as between P1-P2, P2-P3,…, also P1-P5, and compared, using the distance between P1..P5 pressure ports or the full sample length for O1-O2 pressure drop.

The first permeability test using all 7 ports was done by manufacturer during installation and tuning the equipment. They observed a small, about 5%, underestimation in permeability obtained from external ports in comparison to internal ports.

Later after more careful analysis of the equipment we revealed the following details. The fluid is injected via endcaps, with “cross-” and “snake-”shaped grooves:



These endcaps have deeper and wider grooves compared to Fig. 6G. Let us assume that the pressure loss at the tube-sample contact is proportional to the area over which fluid enters the sample, at least when the sample pores are significantly smaller than the tube diameter. I.e., for the fluid, it is significantly “easier” to flow along the grooves first, and only then to enter the sample while using full area of the grooves. Using the images of both grooves, we estimated their area to be about 25% of the end cap area on average. (The “snake” endcap has two separate grooves, one for each tube, and ~25% area is only for one groove). This translates to D_smpl/D_tube ~= 2, i.e. a tiny smaller tube with the original ratio of D_smpl/D_tube~=24 acts as a much larger tube due to the grooves. Taking into account the sample length and diameter, L_{smpl}/D_{smpl} = 8, and using Fig. 6D from the submitted manuscript, we also get a smaller underestimation from the model (while very approximate empirical estimation from Fig. 6E suggests -11%):

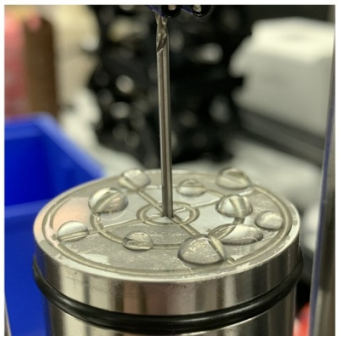


But in an attempt to reproduce better the model system, we decided to block the grooves with epoxy:

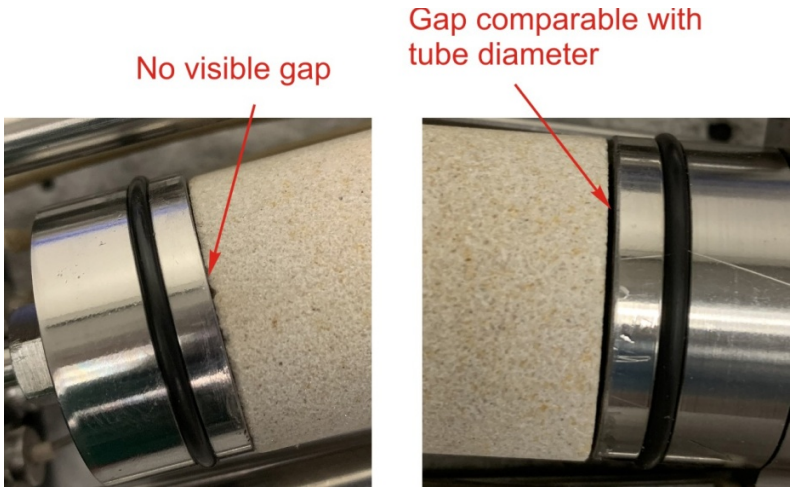
snake endcap with epoxy but before polishing

cross endcap with epoxy, after polishing

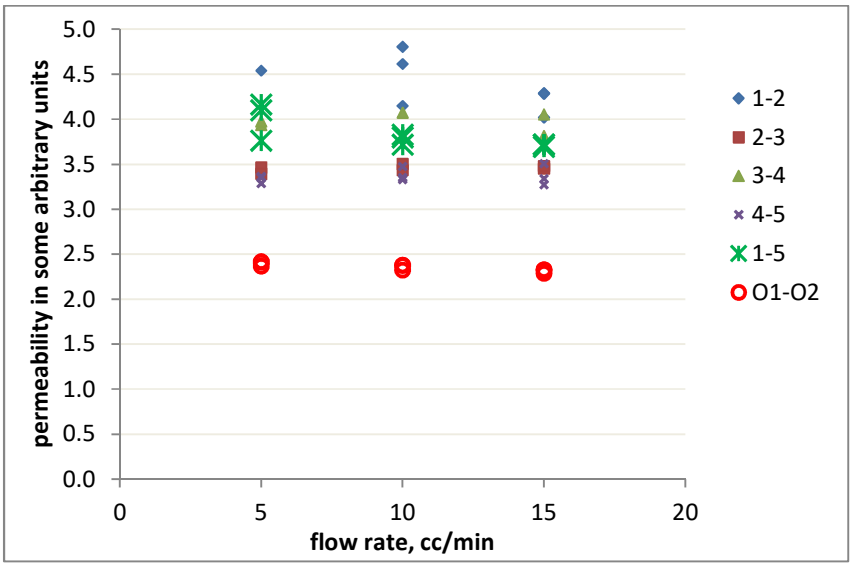
Drill bit was used to prevent epoxy from entering the tube. Drops of water were placed to (unsuccessfully) demonstrate flatness of the surface after epoxy application and polishing



After polishing, the resulting endcap surface became flat. During sample installation we realized that there is a gap between one of the sample ends and the snake endcap:



It appears that both the sample end and endcap normal vectors are not properly aligned with the sample axis. Here is what we got from the permeability measurements while using 1-mol brine as a working fluid and repeating experiments 3 times during 3 days:



Using our model predictions from Fig 6E, we expected $k_{smpl}/k_{incorrect}$-1 for D_smpl/D_tube = 24 and L_smpl/D_smpl = 8 to be about +150% (0.5*24/8 = 1.5). But in fact we got +63% on average while calculating permeability from the P1-P5 and O1-O2 pressure ports.

This experiment on the solid (but not granular) sample suggests that the gap between sample end and endcap surfaces reduces the pressure drop at tube-sample contact. At the same time, this fact has the opposite effect: taking another sample with ends more parallel to the endcaps will show lower permeability values due to the pressure loss at the tube-sample contact but not due to the inherent geometrical properties. Also, this fact suggests the existence of a simple reason for several-fold scatter in permeability measurements performed by the same group on the same equipment for similar samples.