Dear Editor and Reviewer,

Please find below our point-by-point response to reviewer's comments. We are still convinced that the 1D directional variography provides an important information about the anisotropy and heterogeneity of the investigated samples (which has not been used by previous studies and which in fact could presents the "novel" component of our approach), and can assist in characterizing the flow-determined REV substantially. We think that the problem was that our former charts did not include all the necessary details that misled the reviewer. Nevertheless, **we decided to drop it completely**, following the reviewer and editor suggestions (below). We plan to address it systematically and in full detail in a new paper.

In the last sentence of his Comment 6 the reviewer indicates: "*Given the flaws of this variographic analysis it's not surprising to me that the "apparent", interpreted range is ten times smaller than the classical REV determination. The authors should consider either doing it properly, possibly involving some experienced geostatistician in the analysis, or* ***else leave it all together***."

The issue also appears in editor's decision: "*However, the reviewer also notes that you should undertake a careful analysis if the geostatistical investigations are indeed necessary. As they stand now, the results and corresponding interpretation are questionable.* ***Should you choose to drop the variogram analysis****, the reviewer recommends that the new simulations of permeability within the whole of sample S3 be presented and discussed as part of the main text.*".

We present our detailed response to the reviewer comments point-by-point with additional charts and insets. Our first author performed the additional quite extensive coding and calculations of 3D variography to address their comment #5. We address the simulations in S3 sub-volumes and in the entire sample in our response to comment #7. Figure numbers addressed below in our responses to comments #1-6 correspond to the previous version of our manuscript.

Authors response to reviewer's comments:

Reviewer's Comment 1*:*

> In the variograms of figure 12(a) and (b) and all of Fig. 14 I can only recognize pure nugget effect combined with high-frequency fluctuations. The computation of the experimental variograms in terms of retained "bins" and "lags" must be probably adjusted (here I mean: increased to supports larger than these high frequencies fluctuations, and not following an incremental pixel by pixel lags. In other words: the original image should probably be "regolarized" in order to apply variographic analysis) in order to smooth them out. Furthermore, it's not clear which are the calculated experimental values near the origin, since they are covered by the model line. In any case, relying on the fitted range of the gaussian models as measure of correlation length for the porosity in these samples is in my opinion not supported by the experimental evidences. First of all, the gaussian model is always a suspect one to the geostatistician, since realizations of the corresponding random function are infinitely often differentiable (they are analytic functions): this is contradictory with their randomness (see Matheron 1972, C-53, p73-74; available at http://cg.ensmp.fr/bibliotheque/public). Secondly, a practical argument for rejecting this variogram model in the presented cases is the fact that at the origin it is supposed to have near-horizontal behavior (roughly, an S-shaped function). The calculated experimental variograms do not display such behavior, with possible exception of figure 18(b).

Authors' Response 1:

- Experimental semivariogram values near the origin: we now added to the semivariogram of samples S1-S3 (Figures 12_new, 14_new, 18_new, below) insets that focus on the short lag distances (subplots d-f in the Figures 12_new, 14_new, 18_new below). Now, **it can be seen that near the origin a nugget effect is not observed and the experimental semivariograms are characterized by a parabolic behavior (red dots), which implies a continuity (due to the smoothing occurred when averaging over a slice) of the investigated property and with a continuous slope**. We remind that the data investigated in the semivariograms are an averaged porosity along incremental slices (whereas the averaging is applied on binarized voxel value: '1' – pore or '0' – grain) with a single voxel width along (x-,y-,z-) directions (Figure 11,13,17 in the manuscript). The parabolic behavior near the origin is fitted well by a Gaussian model (blue dashed lines). **Therefore, calculating the experimental semivariogram based on slice-by-slice porosity is an appropriate tool for variogram interpretation, a tool that for the best of our knowledge was not used in previous studies using CT data**.
- Furthermore, it was stated by the reviewer that "The computation of the experimental variograms in terms of retained "bins" and "lags" must be probably adjusted". We followed the reviewer's suggestion to plot semivariograms with a lag distance of larger support volumes (Figure 12**aa** below). When using lag separations at the scale size of the fluctuations (125 µm, in black), the fluctuations are only dampened and smoothed. **We think, therefore, that regularizing is not needed for semivariogram calculation based on slice-by-slice porosity of CT data**.
- Concerning the "High-frequency fluctuations": Due to the absence of a nugget effect, the high-frequency fluctuations are assumed to be related to a true geological-derived spatial variability in the sample, rather than to a non-regularized image (see above). The average distance from peak-to-trough in the semivariogram is about 0.1 mm, with a similar value to the range of correlation that was calibrated with the Gaussian model for all samples and in all directions. From petrographic imaging of the sandstone sample, we relate these fluctuations in the semivariogram to changes in porosity influenced by the grain packing / sorting. Within the semivariograms, an additional structure with high variability can be observed (e.g. 12_new_c, 18_new_a-c, below), causing the scale of fluctuations in the semivariograms (produced by the grain packing) to be "overshadowed". The structure that overshadows has a range of correlation on the mm-scale and relates to the more "macroscopic" inhomogeneity in each direction. The analysis and our interpretation for each sample are the following:
  - For S1, Figure *12_new_c* shows mm structure only along z-direction. In this sample that structure relates to the mm-scale layering that was observed from thin section microscopy. The mm-structure that is derived in z-direction and its absence in x- and y- directions implies a distinct anisotropy.
  - For S2 Figure *18_new_c* mm-scale structures are observed in x-, y- and z- directions. However, in the z-direction the calibrated range of correlation of the hole-effect model is 2 mm, whereas in x- and y- directions the range is 1 mm. This sample contains a clayey matrix, which is distributed in the sample in a patchy behavior. Accordingly, the structures observed and calculated from the semivariogram are related to this geological phenomenon. The larger heterogeneity in z-direction again implies a distinct anisotropy. The longer range in z-direction is related to layering, as derived from the high (negative-) correlation between porosity and clay matrix in z-direction and no correlation in the x- and y- directions (Figures 15-16 in the manuscript).
  - For S3 *Figure 14*_new, the semivariograms do not show a mm-scale structure in any direction. Results are thus solely related to the grain packing. It implies that no other structure exists in the domain under investigation, and that this sample is relatively homogenous compared to the others.

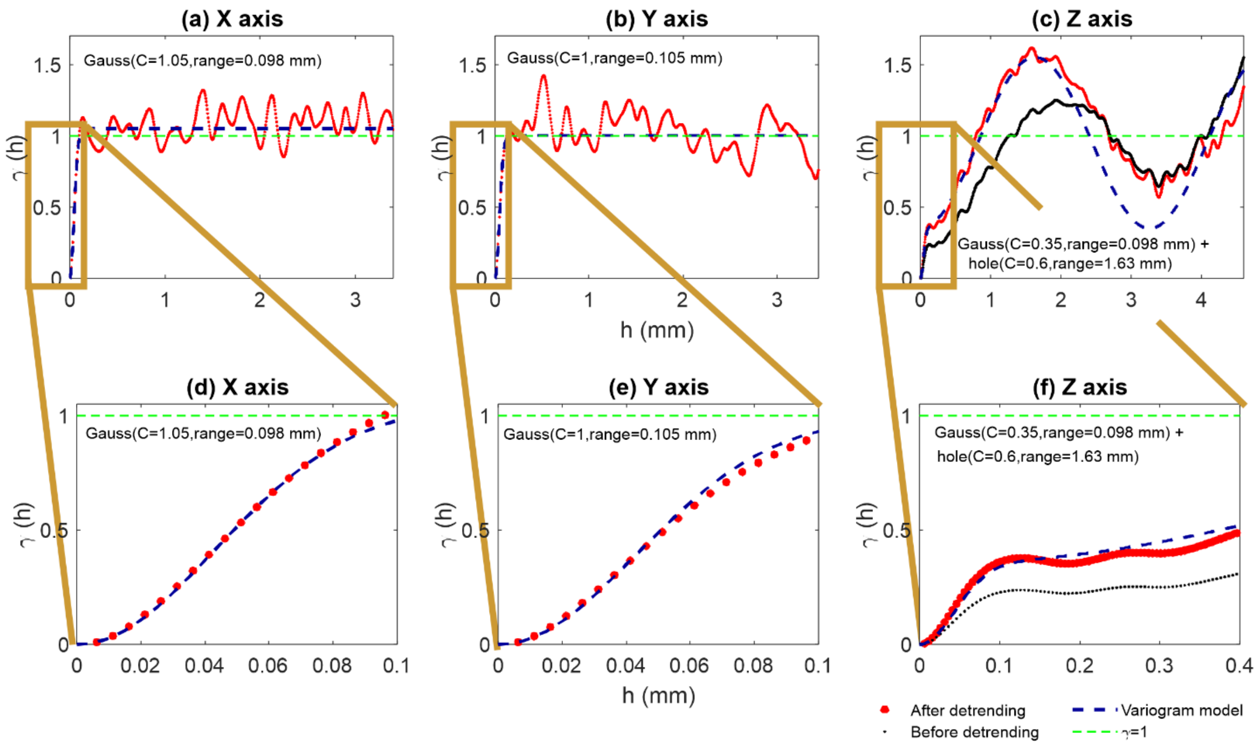

Figure 12_new: Semivariograms of S1 in a) x-direction, b) y-direction and c) z-direction. z-score transform was applied to the histograms (Figure 11d). For z-direction it was applied before (in black) and after (in red) the trend removal (Figure 11c). The variogram analytical models (Gaussian and hole-effect) are also displayed (dashed blue line), modelled with nugget set to zero. d-f) zoom-in on short lag distances to assess the behavior near the origin and the goodness of the fit.

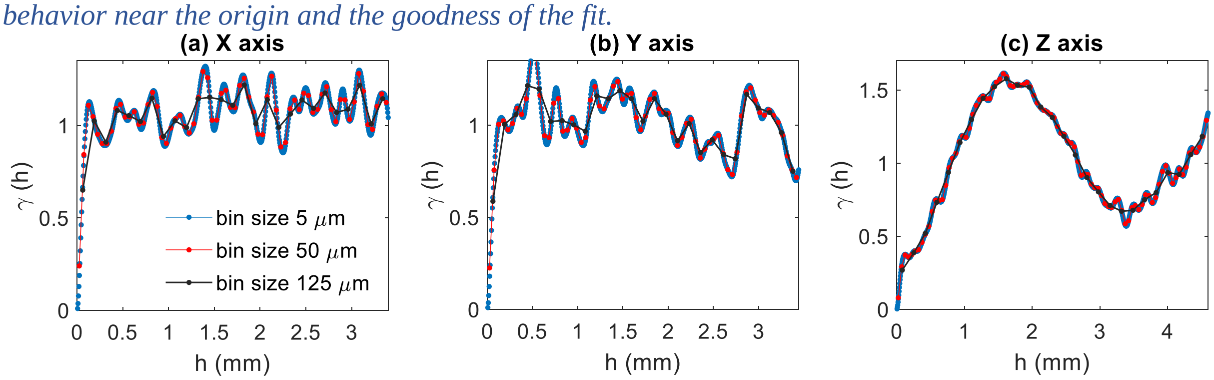

Figure 12aa: Experimental semivariograms calculated with 3 different bin sizes: 5, 50 and 125 µm, in a) X-, b) Y- and c) Z- axis, for sample S1.

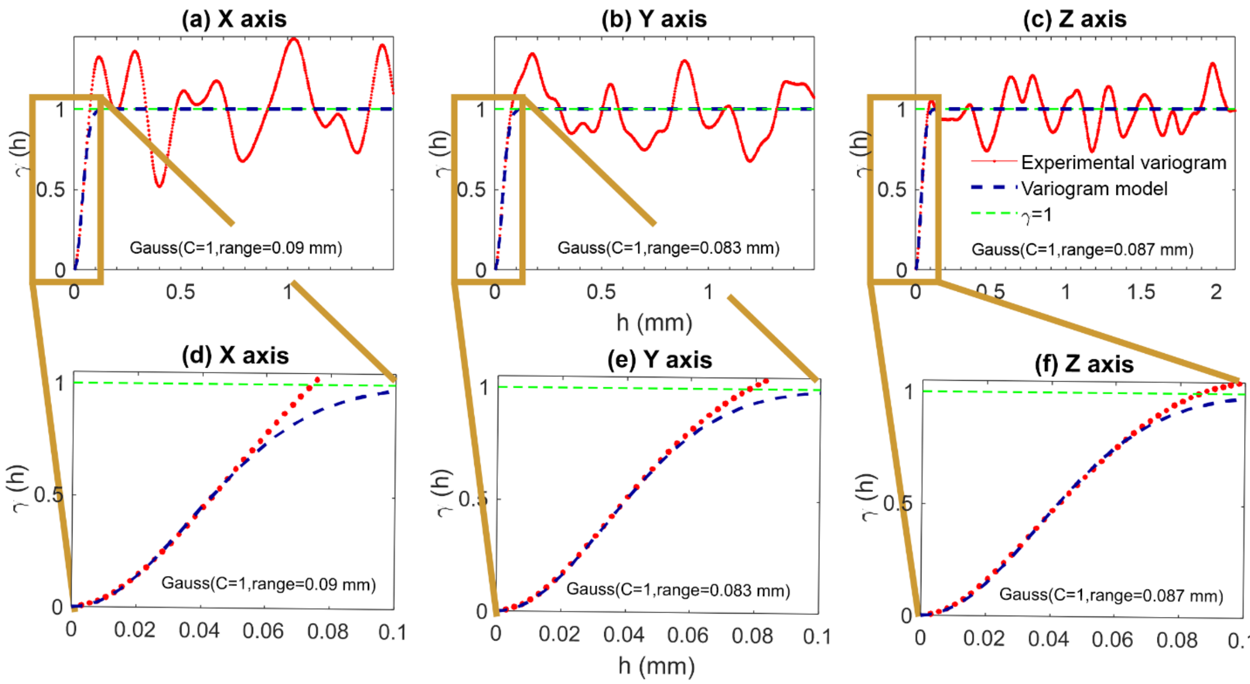

Figure 14_new: Semivariograms of S3 in a) x-direction, b) y-direction and c) z-direction. The experimental semivariogram was modelled using Gaussian model (dashed blue line). The semivariogram models are also displayed, modelled with nugget set to zero d-f) zoom-in on short lag distances to assess the behavior near the origin and the goodness of the fit.

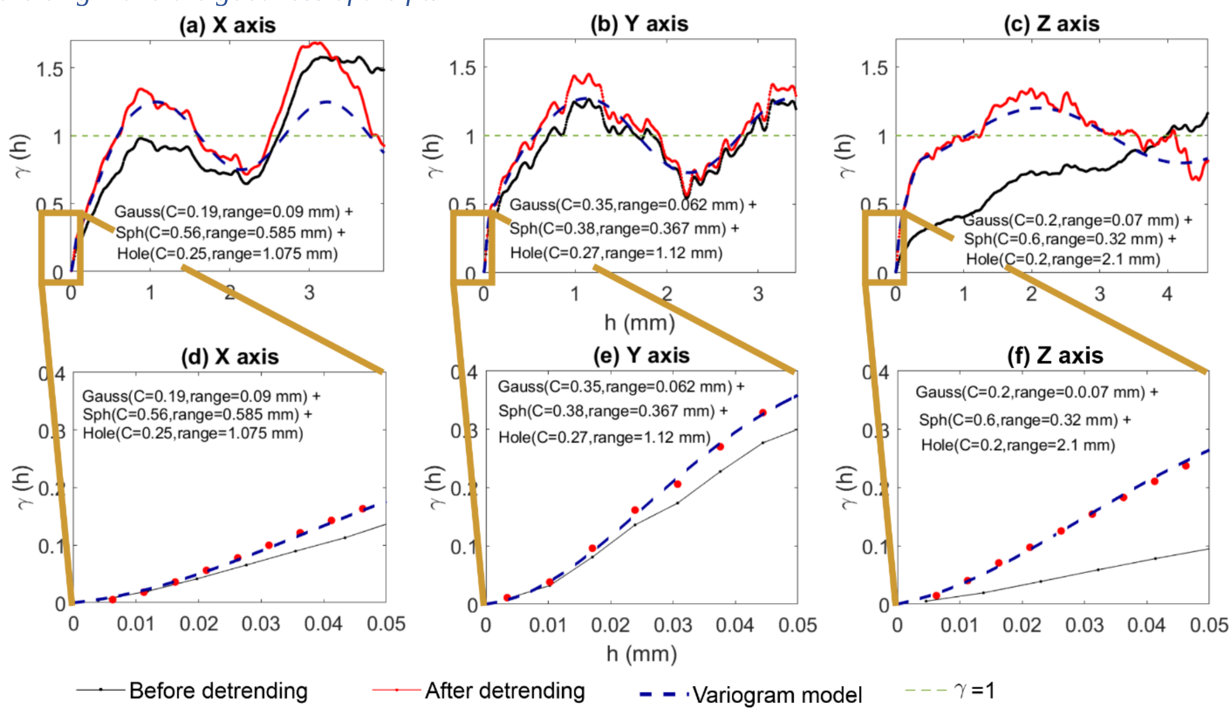

Figure 18_new: Semivariograms of S2 in a) x-direction, b) y-direction and c) a-direction. z-score transformation was applied to the histogram, before (black) and after (red) the trend removal. The experimental semivariogram was also modelled using nested Gaussian, spherical and hole-effect models

(dashed blue line) with nugget set to zero. d-f) zoom-in on short lag distances to assess the behavior near the origin and the goodness of the fit.

Reviewer's Comment 2:

For the same graphs, it would help to have the experimental variograms plotted as points at the centres of each actually computed bin instead of (what it seems) regularly spaced points, which I can't tell if they correspond to the actual computed bins or if it's a graphical artifact.

Authors' Response 2:

Actually, the calculated values are now displayed by points (in red, Figure 12_new, 14_new, 18_new). The reason it seems like a line is that the lag separations (or bin size) is the size of a voxel 5 µm. Within the range of between the minimal and maximal lags (5 µm and 3000) there are 600 points, thus look continuous and seemed like a line demonstrated previously. The points can be seen when using the zoom-in tool. Thus, one can see that semivariogram values refer to a lag distance which excludes any graphical artifacts.

Reviewer's Comment 3

The z-direction in fig. 12(c) shows a clear nugget effect, I don't see why the authors don't acknowledge that. The same applies to fig. 18(b).

Authors' Response 3:

Please see Figure *12_new_f*, which zooms-in to assess the behavior near the origin and shows that there is no nugget effect, but a parabolic behavior (S-shaped) near the origin.

Reviewer's Comment 4:

The de-trending of the sample S1 if I understand correctly, has been done just subtracting the average of the porosity, and not using the "external drift" represented by the cement vol. percent, which is actually provided by figure 20 and is an excellent observation. So, when doing a linear detrend the authors should use a linear model based on the cement volume-% in direction z. Even better however, since it's clear from figure 11(c) that we are in presence of a clear stratification with a period of little less than 4 mm in z-direction (corresponding indeed to the minimum value of the variogram for distances at little less than 4 mm), it's clear from the beginning that a linear de-trend would not be sufficient in this case. A solution would be to apply here a higher order de-trend (polynomial but also smoothing splines or loess...), so to focus on the fluctuations around the middle and eliminate the hole effect all together.

Authors' Response 4:

- The de-trending in z-axis was performed by subtracting the "external drift" as shown in the histogram 11d (in the former version of the manuscript). Figure 11C (red, right y-axis) shows the de-trended residual porosity, slice-by-slice profile. The de-trending was performed based on the trend observed in slice-by-slice porosity (black, left y-axis). The de-trend functions chosen were either 1st or 2nd order polynomial fit, when the 2nd order one was chosen if the improvement compared to the 1st order fit was significantly better.

- The reviewer suggests using the cement percentage in S1 for de-trending the slice-by-slice porosity, by calculating the cement trend. The cement in S1 is observed by microscopy as meniscus bridging and grain coating made of flakes at 1-10 µm scale. This size is in the order of the imaging resolution (scans of 2.5 and 5 µm voxels grid), not high enough for quantification of the exact characteristics (volume, type, grain sizes), therefore relying on cement content as a property which is correlated with porosity encompasses much uncertainty and therefore should not be considered. Sample S2 with a prevalence of clay matrix (Figure 15 in the manuscript), however, does show a negative correlation in z-direction (Figure 16c), which relates to the deposition processes, but no correlation in x- and y- directions. Therefore, we think that doing the de-trending for porosity based on the clay or cement is not the way, but using the porosity for de-trending, as we previously used.

Reviewer's Comment 5:

Concerning figure 19 and the whole study about it, it seems to me that it just refers to subsets of the original sample, in which the variograms were still computed slice by slice, whereas I initially suggested to use 3D "supports" for the computation of the experimental variograms, and not restrictions of the domain. I am not convinced that this part adds valuable informations to the analysis at the moment.

Authors' Response 5:

As we understand, the reviewer suggests performing a 3D variography based on single pixels and not slice by slice porosity. We tested the 3D variography for the CT data performing the additional coding, for a histogram used for computation is binary – voxel values are either '1' (pore) or '0' (grain/matrix). We are interested in calculating the correlation length of the pores, therefore only pore voxels are considered when voxels were chosen randomly. Because the entire dataset computationally is too heavy (more than 1000x1000x1000 voxels, spatially continuous data), we used in the experimental semivariogram computation only a part of the total possible pairs (we used 10^6 pairs in each direction). The semivariogram results for S1 are shown in Figure 21 below, zooming in at different lag distances.

- For short distances (Figure 21a) for all directions the range of correlation is approximated to about 0.1 mm, similar to range calibrated from the semivariogram computed from slice-by-slice porosity (Figure 12_new above). This structure is related to the grain packing.
- For intermediate maximal lag distances (Figures 21 b,c), no structure is observed, in contrast to the layering structure of mm-scale that was observed from the slice-by-slice porosity in z-direction (Figure 12c). Figure 21 is presented just as a clarification for the reviewer.

- Figure 19 in the manuscript is removed (along with other figures related to the variography).

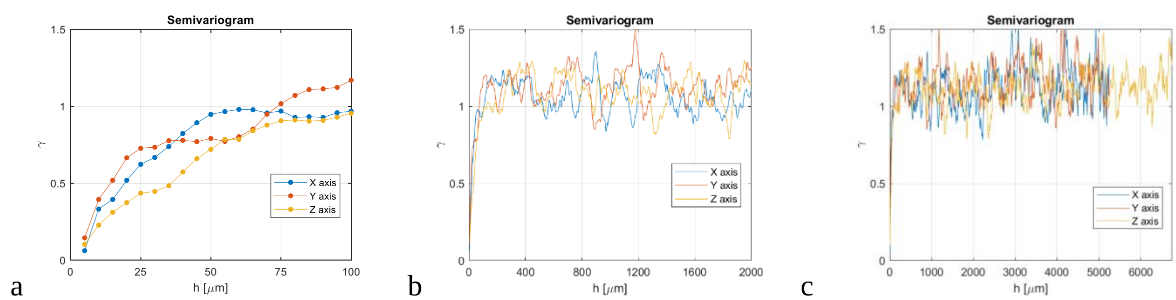

Figure 21: Semivariogram for S1 measured on voxel-based binarized data. Randomly chosen voxels are pores. For interpretation, plots are zoomed-in on the ranges: a) 0-250 µm, b) 0-2000 µm and c) 0-6300 µm. Sill=1.12. Number of voxel pairs used is ~10^6.

Reviewer's Comment 6:

Given the flaws of this variographic analysis it's not surprising to me that the "apparent", interpreted range is ten times smaller than the classical REV determination. The authors should consider either doing it properly, possibly involving some experienced geostatistician in the analysis, or else leave it all together.

Authors' Response 6:

We agree with the reviewer, and removed all of the mentioned variography. Nevertheless, the semivariogram is a useful tool to estimate distances of correlation and lack of correlation between measurements. Determining this range of correlation is useful to estimate the scale of *heterogeneity*, and is not directly related to the *directional REV*. It is more an apparent range, influenced by cementation, clay minerals and grains. For S3, this apparent range is shorter compared to the REV determined by the classical approach. The shorter apparent range from the semivariogram is in our opinion clearly related to the grain packing, the only structural phenomenon at the size of the sample domain under investigation. As a rule of thumb, one would use at least four (4) grain diameters as REV edge size.

Reviewer's Comment 7

I salute however the new simulations of permeability within the whole sample S3. In my opinion this part should not be in appendix but discussed in more detail in the main text.

Authors' Response 7:

The simulations in sub-volumes and a full volume of S3 appear now at the main text (Figures 15, 16) and discussed in lines 858-862, 990-999, 1109-1155 in our revised manuscript.