Reviewer #1: Answers / discussion (black) of specific review comments (blue)

1. *Please use "semivariogram" instead of "variogram", or state in the text that you are using "variogram" to intend "semivariogram", as I will do from now on in this review.*

   Thanks for clarifying that! We will change "variogram" to "semivariogram" throughout the entire manuscript as you suggested.

2. *Is the support used to compute the experimental variograms the whole image plane orthogonal to the computed direction? It is not completely clear to me at the moment. I find the Methods description misleading in this regard. If it's the case, I strongly advice to use for variography at least squares and not the full slices, and even better yet would be to do a full 3D variography, doing cubic supports as you did in the "classical" REV analysis of figure B1. Though, this is much more computationally intensive and may require ad-hoc coding. This may have a potentially large effect on the results, depending on the chosen "box side", especially for the sample S2, and if done properly also actually hint at the true principal axes of anisotropy for the samples, which may not be aligned with their sides.*

   Yes, that is the case, concerning your comment on the support. In the manuscript we used a "segmented image slice" which we will rephrase a little bit in order to point out that we used the "whole segmented image slice" for the analysis.
   Additionally, we understand the importance of performing the different types of a variography, even with cubic supports in 3D. In fact, this would be a very good approach in our opinion, if no information about the rocks of interest is available. For this study, we made large efforts to take the samples in the field as accurate as possible to achieve a sub-sampling procedure in line with the visible foliation of the different layers. By best of our knowledge and experience, the resulting "coordinate system" of the samples as well as of the related 3D images should be in very good alignment to the true principle axis of anisotropy. Evidence for this is given by comparing the exceptionally good results of measured and modelled permeability values and tensors. Accordingly, we have chosen the "less challenging" 2D variography path for our study, since we know all of the rock samples' features and textures very well.
   Nevertheless, we will address this topic within the discussion section of the manuscript, as a different methodical approach with its own advantages and drawbacks, inserting the additional results/charts suggested by the reviewer into the supplementary material. We believe that the application and comparison of results from the different variography methods is a big stand-alone topic. It will be of a great impact and thus deserves to be addressed within a new topical manuscript applicable to samples with various (not known) positions of the principal anisotropy axes (which we are already checking for feasibility).

3. *Fig 9-11: if you are showing the fitted variograms, based upon which you define the apparent ranges, I believe you should also state which variogram model was fitted. From the legend I may assume it's an exponential model but or spherical or something else, and the "expo. fit". If it's the case, then you need to specify if the reported "range" is the actual coefficient in the exponential model or the "practical range" of the asymptotic function.*

   Again thanks for pointing this out. In fact, we used the "practical range" of the asymptotic function as "the range" for the semivariograms. It seemed more "intuitive" to potential users which are not "deeply in this statistical topic". We will clarify that in the text.

4. *All three samples represents exemplary cases for zonal anisotropy, where the sills of the variograms are not constant following different directions. This reinforce my suggestion of making the samples available to the public.*

All data are available at the PANGÄA-repository as explained in the "Supplementary materials section". Related doi's will be added at the end of the manuscript and cited in the text of the manuscript.

5. *It is to me however striking - and this may hint to a too large support definition, cfr comment 2, or else to a graphical imprecision - that no experimental variogram displays any nugget effect. This could mean that the variable has been excessively regularised. Please state in the text how the lags for the calculation of the experimental variagrams were chosen, and if the computed pairs at each lag bin are comparable.*

Obviously, the nugget effect can be attributed to measurement errors or spatial sources of variation at distances smaller than the sampling interval or both. Measurement error occurs because of the error inherent in measuring devices. Natural phenomena can vary spatially over a range of scales. Variations at microscales smaller than the sampling distances will appear as part of the nugget effect. Nevertheless, this is more or less impossible to achieve for "mining" spatial data from µ-CT images since we have a fixed resolution limit. Hence, all variation below that "hard resolution boundary" is "invisible" for the variography analysis. This clearly is a drawback of the image analysis, and hence it is important to gain detailed understanding of the scales of spatial variation from multiple methods. Accordingly, the smallest lags are related to the resolution, i.e. the smallest segmented feature of the 3D scan. We will address this issue within the discussion section.

6. *Fig 11. Regarding the variograms of sample S2, they are clearly linear, especially the xy plane, which is a clear sign of non-stationarity, as also clear from the strong trend in subfigure (c). However, you also correctly recognised the "external drift" represented by the clay content. This is possibly a textbook example of external drift, which makes the de-trending of porosity worth. My point here is that the fact that the sample is clearly strongly anisotropic and non-stationary does not mean that it is not possible to extract a REV from it, at least for the two other directions, but with some manipulation, also in the xy plane. Moreover, a full 3D variography (if my 2. comment is valid) may give different insights and results.*

We fully agree on this comment. In fact, this is a very good example for doing exactly that what you've mentioned (Comment #2) in future work, in order to understand and extract REVs on arbitrary scales (meaning here from µm – couple of cm, i.e. the laboratory scale). As stated on your previous comment, we made quite some efforts to be sure that all samples have been derived and measured in alignment with their intrinsic foliation and textural features. For our analysis, we used the largest possible and available 3D image. By nature of the computed tomography, the "size" of the image is limited by the so-called "field of view", which is a result of the sample positioning relative to the X-ray source and the detector. The field of view is always "linked" to the achieved image resolution. Hence, changing the field of view (i.e. making it larger) would lead to a different (i.e. coarser) resolution. We will discuss this point within the manuscript as clarified within our response to the comment #2.

7. *No histogram of apparent image porosity is displayed, neither from the slices used for the variographic REV nor from the subsets of figure B1, although from that figure we get an idea of the "density" (however there is sampling involved here, I assume). Is it possible that the "cube" porosity - at a given cube size - is also lognormally distributed? Possibly then it could be worth to perform the variographic analysis on a log porosity.*

We will add and discuss the 2D "porosity histogram" as recommended in the given context. Since we discussed the 3D variography before (answers to your comments #2 and #6), we will not re-calculate the variograms with log-Phi data at this point.

8. *For sample S3 the REV is identified at 350 voxels, though only one permeability simulation is conducted. It would be nice to demonstrate that the calculated permeability is somewhat "continuous" by repeating the flow simulations on different subsets of that size of the original microCT image.*

   We agree that performing the repeated simulations on different subsets of that REV size for S3 to demonstrate the “continuity” of the permeability estimation is important. However, we do not have the (cost-intensive) Materialize (Belgium) meshing license anymore in order to transform the µ-CT images into COMSOL readable meshes that was used in this project. In order to avoid problems related to software usage, we would like to propose using the currently available toolbox GEODICT, which features a “built-in” mesh generator, for this task only. The basic numerical Navier-Stokes algorithm and the according boundary conditions will be exact the same as documented within the manuscript. Hence there should not be numerical differences (inferred from our previous experience) that would create problems or require an extensive discussion. We will first compare the results with those conducted on the current subsample of S3 with Comsol and then perform the additional simulations on the different subsets of S3 with GEODICT. We could easily do this for the S3 sample only as suggested by the reviewer, to avoid a massive re-modelling of all the derived data with GEODICT. The results will be placed within the discussion section. We would like to discuss, if this is an appropriate way for Reviewer #1, and also Reviewer #2, as well as for the topical Editor.