We thank the referee for the time and effort to review our manuscript. Below, we give point-by-point response to your concern.

Sincerely,

Swarup Chauhan, Kathleen Sell, Frieder Enzmann, Wolfram Rühaak, Thorsten Wille, Ingo Sass, Michael Kersten.

Anonymous Referee #1

The reviewer comments are formatted in italics and the authors response to the comments are formatted in bold.

Notation *RC1.P#* represents ReviewersComment.ParagraphNumber

RC1.P1 Manuscript GMD-2018-335 presents a new collection of image processing and analyses tools for tomographic 3-D images. As the title suggests, the implementation of machine learning tools for image segmentation is in the focus of the manuscript. It is also claimed that the presented imaging software would be particularly suited for identifying representativ elementary volumes.

Yes. Apart from that, the manuscript also highlights a new procedure termed as, dual filtering and dual segmentation to remove edge enhancement artefact in synchrotron based images using machine learning approach. To our knowledge this hasn't been published before.

The scientific community will benefit from the novelty of this approach. The code has been made available. We can certainly elaborate this in the introduction of the manuscript.

RC1.P2 The connection of this paper to models is only weak, in form of the REV detection. However, the manuscript does not elaborate on how REVs are identified. It simply states that "voxel sizes around 480ˆ3 suited best for... ", without explaining how this conclusion was reached.

OK. In the revised manuscript we will extend the section 3.3 and elaborate on the identification of REV for all the three samples.

Basically, it was a combination of visual inspection and consecutively segmenting and plotting tends in relative porosity, pore size distribution and volume fraction. This was done by loading the complete stack in the CobWeb software, during the loading process a 2D movie of the tomogram is displayed in the display window and saved in the root folder. Carefully monitoring the movie gives an objective evaluation of the heterogeneity of the respective XCT sample. Thereafter, based on this subjective information different ROIs are selected, cropped, segmented and their respective geometrical parameter are intercompared.

In the case of Berea sandstone, four different ROIs were investigated, whereas Grosmout carbonate rock seven different ROIs where need to identify the best REVs. Through our previous scientific studies on the GH sediments (Sell et al.,

2016; Sell et al., 2018) we were aware or best-suited REVs. The identification of best REV for Grosmout was relatively tedious compared to Berea sandstone and GH sediment; due to the low resolution and microporosity present in the Grosmount tomograms.

The intention of showing particularly only two REV trends of relative porosity in Figure 7. is due to a very good agreement in porosity values to the benchmark publication of Andrä et al., (2013a, 2013b).

RC1.P3. The presented software appears to be a promising piece of work, but as the authors write themselves, it still has limited capabilities. The maybe most innovative part is implementation of machine learning routines for segmentation, albeit this in itself is not a scientific novelty.

The software is built on scientific studies which have been peer-reviewed and accepted in the scientific community Chauhan et al., 2016a,b. The spinoff for these studies was not the lack of accuracy provided by manual segmentation schemes, but the subjective assessment and non-comparability caused by the individual human assessments. Therefore, the automated segmentation schemes offer speed, accuracy and possibility to intercompare results, enhancing traceability and reproducibility in the evaluation process. To our knowledge none of the XCT software used in rock science community relies on machine learning for segmentation explicitly, which makes the software unique if not novel.

Despite many review articles and scientific publication highlight potential of machine learning and deep learning (Iassonov et al., 2009; Cnudde and Boone, 2013; Schlüter Steffen et al., 2014), software libraries or toolbox are seldom made available. Thus, with CobWeb we started for the first time to fill this gap, and despite its limited volume rendering capabilities— it is a useful tool and current version of the software can be applied in scientific and industrial studies. Certainly a conscious decision need to be taken on our side if to dedicate CobWeb as a segmentation tool or expand it towards simulation software like MATH2MARKET, GeoDict or Volume Graphics. On the other side CobWeb provides an appropriate test platform, where new segmentation and filtration schemes can be tested and used as a complementary tool to the simulation software GeoDict and Volume Graphics. The simulation softwares (GeoDict and Volume Graphics) have benchmarked solvers for performing flow, diffusion, dispersion, advection type simulation, but their accuracy relies heavily on the finely segmented datasets.

RC1.P3 The manuscript is moreover vague when it comes to describing how the different machine learning options are implemented (with exception of the K-means clustering). It is neither explained how the cross-validation option function.

We acknowledge reviewers concern. But, in section 2.4.2 (page 5) we have cited Chauhan et al., (2016a); Chauhan et al., (2016b) which covers the details about the algorithms and cross-validation schemes. Also, Figure 3 gives a visual overview how the ML techniques fit into the framework. Thereafter, the

user manual published on the zendo repository (https://dx.doi.org/10.5281/zendo.2390943) explains the implementation and how to use the segmentation algorithms. Within the scope of the manuscript we find the description sufficient.

But if required, we will expand further on the implementation of the ML techniques and the cross-validation options.

RC1.P4 The manuscript is written in adequate English. Its structure could be improved (e.g. the description of the workings of the filters do not belong in the materials and methods).

Ok. The description and working of filters is written under the sub section 3.1 image processing not under materials and methods.

RC1.P4 At several points I found the manuscript rather inconcise. It is e.g. not explained whether the filters and segmentation approaches work in 3-D or only in 2-D.

We thank the reviewer for pointing it out. The CobWeb 1.0 uses a slice-by-slice 2D approach. It was observed that the ML techniques tend to underestimate porosity values compared to manually segmented analysis at an REV scale size > 500^3. This substantial degree of uncertainty is caused due to 2D slice-by-slice processing rather than the ML techniques. The 2D slice-by-slice approach, passes only, the spatial information (X, Y coordinate direction) to the ML algorithms, which ends up sorting the intensity variation in the spatial domain (local maxima). Therefore, the lack of temporal information (Z coordinate direction) restricts the degree of freedom to find at a global spatial-temporal optimum. In other words, as the temporal changes arise, due to bedding (sedimentary rock) or micro porosity (carbonate rocks) in the rock texture, they are represented as sudden spike or dip in porosity values; which to an inexperienced eye appear as artefact or anomalies– and often-then-not discarded.

This correction will be implemented in the next software version; in the current workflow it has not been accounted for (CobWeb 1.0). Since, it requires refactoring the loop-based scalar-oriented framework to matrix and vector operation approach called *vectorization*. The 2D slice-by-slice processing scheme is much faster compared to the 3D approach. So, the choice of 2D processing for this research study was made to make it affordable to compute on desktop, laptop for near real-time and onsite evaluation.

RC1.P4 Or like at p9L19: what choice of the cluster centers influence the perfomrance of the K-means algorithm? Its initial location? The number of clusters?

Thanks again, for raising the question, we can certainly elaborate on this in the discussion section. In general, performance in terms of accuracy and speed is directly proportional to starting point (initial location) in the segmentation process. Meaning, the closer the starting point (initial location) is to the global

minima— faster will the algorithm converge and even so better is the performance (accuracy & speed).

However, in unsupervised technique by default the choice of the starting point is through random seed unless explicitly specified. So, in the case of the dual segmentation approach, the intuition was to capture all the material phases, including the edge enhancement artefact, speck and noise etc. in the first step and thereafter in the second step to rescale them to the plausible phases.

Hence, in the first step the 20 clusters where initialized using random seed. Since, the priority was to capture all phases in GH tomograms not the performance. And, after the rescaling processes, we were aware of the initial locations which we used as starting point (initial location) to assist the algorithm to move towards identifying correct phases.

RC1.P5 In summary, this manuscript presents a promising software tool for tomographic image analyses.
We thank the reviewer for the acknowledgement.

RC1.P5 But I do not think the manuscript fits within the scope of GMD, nor do I think that the manuscript is developed enough to reward revisions with another round of reviews.

We disagree with the reviewer on the above comment. The uniqueness of this journal is that, it gives the possibility of accepting six different types of manuscripts, and we were careful in placing the work in the *model description papers* category as it fulfils most of its norms if not all. This has been clearly highlighted in the cover letter to the topical editor.

References

Andrä, H., Combaret, N., Dvorkin, J., Glatt, E., Han, J., Kabel, M., Keehm, Y., Krzikalla, F., Lee, M., Madonna, C., Marsh, M., Mukerji, T., Saenger, E. H., Sain, R., Saxena, N., Ricker, S., Wiegmann, A., and Zhan, X.: Digital rock physics benchmarks—Part I: Imaging and segmentation, Benchmark problems, datasets and methodologies for the computational geosciences, 50, 25–32, doi:10.1016/j.cageo.2012.09.005, 2013a.

Andrä, H., Combaret, N., Dvorkin, J., Glatt, E., Han, J., Kabel, M., Keehm, Y., Krzikalla, F., Lee, M., Madonna, C., Marsh, M., Mukerji, T., Saenger, E. H., Sain, R., Saxena, N., Ricker, S., Wiegmann, A., and Zhan, X.: Digital rock physics benchmarks—part II: Computing effective properties, Benchmark problems, datasets and methodologies for the computational geosciences, 50, 33–43, doi:10.1016/j.cageo.2012.09.008, 2013b.

Chauhan, S., Rühaak, W., Anbergen, H., Kabdenov, A., Freise, M., Wille, T., and Sass, I.: Phase segmentation of X-ray computer tomography rock images using machine learning techniques: an accuracy and performance study, Solid Earth, 7, 1125–1139, doi:10.5194/se-7-1125-2016, 2016a.

Chauhan, S., Rühaak, W., Khan, F., Enzmann, F., Mielke, P., Kersten, M., and Sass, I.: Processing of rock core microtomography images: Using seven different machine learning algorithms, Computers & Geosciences, 86, 120–128, doi:10.1016/j.cageo.2015.10.013, 2016b.

Cnudde, V. and Boone, M. N.: High-resolution X-ray computed tomography in geosciences: A review of the current technology and applications, Earth-Science Reviews, 123, 1–17, doi:10.1016/j.earscirev.2013.04.003, 2013.

Iassonov, P., Gebrenegus, T., and Tuller, M.: Segmentation of X-ray computed tomography images of porous materials: A crucial step for characterization and quantitative analysis of pore structures, Water Resour. Res., 45, doi:10.1029/2009WR008087, 2009.

Schlüter Steffen, Sheppard Adrian, Brown Kendra, and Wildenschild Dorthe: Image processing of multiphase images obtained via X-ray microtomography: A review, Water Resour. Res., 50, 3615–3639, doi:10.1002/2014WR015256, 2014.