Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2018-335-RC2, 2019
© Author(s) 2019. This work is distributed under
the Creative Commons Attribution 4.0 License.

Geoscientific Model Development Discussions
Open Access
EGU

Interactive comment on "CobWeb 1.0: Machine Learning Tool Box for Tomographic Imaging" *by* Swarup Chauhan et al.

Kirill Gerke (Referee)

kirill.gerke@gmail.com

Received and published: 7 May 2019

The paper of Chauhan et al. presents a tool developed by the leading Author through a series of papers. This particular manuscript, unlike its predecessors, focuses on presenting the toolbox for already existing methods. I am sorry to post this review late, i guess this qualifies me as being a bad referee, but it took me quite a while to absorb everything, plus i struggled with trying to run the toolbox. On the plus side i got two reviews already posted and, thus, can rely on them to put in my vision here. I believe that topic of the paper - development of a free software to process XCT tomography images of porous media, is a very relevant topic. We do not have a solution, and those available as commercial solutions are expensive and not doing their job even closely. Scientific quality of paper is good, it is in general robust except for some particular parts

Printer-friendly version

Discussion paper

picked up by Reviewer1 and now here by Reviewer2. Scientific reproducibility was a problem for me. First, i had to borrow my wife's laptop, as mine is running under Linux. She has Matlab 2016b installed - the package is not working with it. Downloading Matlab runtime took time, as it is 1.8 Gbs of an archive. After reading the license i got puzzled - totally agree with David Ham at this point. Moreover, it was not clear how much disk space on the small SSD it will occupy - guessed it will be much more than 1.8 Gbs and i gave up. Presentation quality and writing are fine with me, i did not have any problem with these (except structure, see bellow) while reading the manuscript.

Considering the potential usage and the amount of work done by the leading Author - i believe the toolbox should be published. It can be used, cited and may serve as a comparison metric/survey of techniques. In this sense, i do not agree with the Reviewer1 who mentioned that the paper lacks novel aspects - the toolbox is the novel thing presented here (+ edge enhancement correction strategy). But i do agree with other points of critique raised by Reviewer1 and, thus, do not see how this paper could be accepted without changes. To give the Authors a possibility to publish (i think the paper is appropriate for GMD), i recommend a major revision.

Major comments: 1) REV In Section 3.3 you discuss classical REV idea of property convergence with increasing subsample volume. Yet, you do not explain how do you determine REVs for all your samples. Are these REVs were chosen in terms of porosity, pore sizes, etc.? If so, what was the threshold you used to stop increasing the volume? Moreover, here you write: "In particular, while performing permeability tensor simulation using XCT data, the size of minimum REV should be assessed not only based on porosity but also on geometrical parameters such as pore size distribution, void ratio, and coordinate number (Al-Raoush and Papadopoulos, 2010; CostanzaâĂŘRobinson M.S. et al., 2011)." I did not re-read these papers to check if these authors stated exactly this, but for performing permeability tensor simulation using XCT data, the size of minimum REV should be assessed based on permeability tensor values! More generally, REVs do not necessarily exist, finding REV for one property has nothing to do

Printer-friendly version

Discussion paper

with finding it for the other. The topic is very deep and is well beyond the scope of this manuscript. The possible ways to fix the REV problem here is to either remove it altogether, or explain for which property is it analized and calculated. By the way, if the REV analysis of porosity is based on Fig.6 and 7, i.e. by calculating porosity within separate 2D slices, it is not appropriate, it should be done within increasing volumes.

2) Structure of the methods/results Again, i totally agree with Reviewer1 and believe that current structure is hard to follow. I would suggest putting the methods description first, next describe the toolbox and functionality, then the objects and specifics of their processing. So, the structure of the paper could be something like this: - Intro as it is - Methods: 1. Image processing algorithms (here you describe all filters and segmentation techniques in better detail with references) 2. Toolbox and its functionality (you current Section 2.4) 3. Objects and specs (here you describe a sample+its processing steps, next sample, ...) - Results: 1. you current section 3.1, as well as all descriptions of filters, equations and such, belongs more to methods, not results! 2. just describe what you got for each sample 3. you do not have discussion

3) Code/toolbox The way you deliver your toolbox is not suitable for many researchers. Indeed, the code would be more useful for me. But i understand that Authors wanted a way to performs computations +deliver a GUI. And doing this with Matlab is much easier than using other tools. Simply giving away the code is not that useful for many users, because the code itself is much harder to use - you need to write a script to run something. GUI provide a possibility to use a couple of buttons and slider to do the job. Making a good GUI is very challenging. This is the reason we never released our in-house C++ image processing code with Qt GUI into public - we can use it, but for anybody else it will be buggy and not user-friendly. So, both releasing code and executables have their downsides, but in this case the need for Matlab runtime is a big obstacle for potential users. It should be relatively easy to port Matlab code into Python or Julia, which would avoid Matlab - slow and expensive dev-environment. In short, i understand why Authors decided to redistribute executables and not the code,

Printer-friendly version

Discussion paper

but still see more value in free environments (python and julia) - this does not imply that Authors have to port the code as a part of the revision.

4) You write: "The NLM filter was implemented in 3D mode to attain desired spatial and temporal accuracy and was processed on an CPU device." But Manual states that "Non-local means is only implemented as a 2D filter because filtration is done slice-by-slice".

5) Multiphase segmentations Any segmentation method can be made multiphase, be it indicator kriging, converging active contours or region growing. But the machine learning algorithms used in this paper do not solve the major problem for all multiphase segmentations - the phase having intermediate grey scale values gets sandwiched between the other two phases, e.g., see numerous fugues in the paper and in the manual. I would like you to elaborate on this and why do you think clustering and k-means approaches are better than aforementioned techniques?

6) Verification In the Section 6.2 in the Manual your provide validation metrics to prove your segmentation results. I do not see how any of these are actually do the job! They are simply the metrics for performance of the machine learning algorithms, not the segmentation results. I strongly suggest remove or re-write this part. We technically do not have verification data - the ideal segmentations, which can be produced only using synthetic tomography. You could discuss these things a little bit, so that comments 5-6) will give material for small discussion subsection you lack at the moment.

Minor comments: 1) In introduction you write: "Our hypothesis is that 3D tomographic REV analysis of Berea SandstoneTM (BS), Grosmont carbonate rock (GCR), and gas hydrate (GH)-bearing sediment datasets, would benefit of this new approach." I find this statement scientifically awkward (even without considering the REV problems in this paper), moreover this hypothesis is nowhere to be tested. 2) Section 3.1.1.2, you write: "In this study, the NLM filter was set to a search window of 21, local neighbourhood of 6 and a similarity value of 0.71." As i was unable to test the toolbox, it seems

CC BY

after reading the Manual that too many things are fixed. How useful is it going to be if these NLM and other parameters are not changable within the toolbox? With the parameters you specify from my experience with NLM (our version is different from yours and follows closely Buades and Morel paper) i would expect the results to be kind of oversmoothed (again, i did not run the toolbox, so this is a guess). 3) As far as i understood, technically you do not compute pore-size distribution (e.g., extract the pore-network model), but distance-map transform and analyse its distribution. If so, please, change the description of the methodology accordingly. Moreover, i do not understand the legends on the graphs, be it fig.8 in the Manual or fig.8 in the paper (what are these numerous pore-size distributions for the same sample?). 4) You mention a number of free software solutions. I strongly suggest adding our recent effort - FDMSS software to simulate single-phase flow within 3D pore geometries into your list of tools that could be used after segmentation of 3D images using CobWeb. Moreover, you could compute permeabilities for your segmentations and compare then against available data, e.g., "Laboratory measurements of porosity and permeability reported by (Andrä et al., 2013b) are around 21 % (φ = 0.21) and κ = 150 mD âŤĂ 470 mD, respectively".

This review got too long. I appreciate the efforts of the Authors to produce this toolbox and strongly support its publication with GDM. For these reasons i was trying to provide all possible solutions for current manuscript's problems.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-335, 2019.

Printer-friendly version

Discussion paper