Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2018-256-RC2, 2018

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Locality-based 3-D multiple-point statistics reconstruction using 2-D geological cross-sections" *by* Qiyu Chen et al.

Anonymous Referee #2

Received and published: 31 July 2018

Dear Chen et al.,

I read with interest the paper entitled "Locality-based 3D multiple-point statistics reconstruction using 2D geological cross-sections". The paper describes a new methodology to build 3D realizations constrained by 2D sections that act as local training images and hard data for MPS algorithms. The method seems to give realistic results and perform better than existing algorithms. Globally, the paper and the presented results are convincing. However, I would like the following general comments to be addressed by the author:

1) The introduction is a little bit messy, giving an overview of many papers related to MPS, but missing the point of highlighting the specific issues tackled by the paper. For example, the 2nd paragraph (P2L15) makes the history of MPS. This is not the point,

you should rather insist on the importance of the TI (what you do in paragraph 3) and the description of the state of the art (relatively vague in the current version). Many sentences in the introduction are too vague such as "some assumptions have been implemented to reconstruct 3D models" (P3L14, which assumptions?) or "A promising reconstruction method . . . by adapting the DS algorithm. However, large-scale" (P3L16, the "however" refers to something that is not explained, the problem should be clearly stated).

2) I found the methodology section difficult to follow. Indeed, the proposed approach borrows some techniques from existing algorithms (mostly direct sampling), so that part of the methodology is described in other papers. Although, some parameters are common with DS, the philosophy is quite different as DS is never explicitly calculating cpdf. In that sense, the proposed methodology is closer to classical approaches such as snesim (except that the cpdf are not stored). I would therefore recommend to explicitly describe every part of the algorithm, without (too many) references to previous publications. Doing so, the methodology would be self-sufficient. Part of the methodology (dissimilarity metrics and MDS) is introduced in the result section and should be moved to the methodology.

3) In the parameter sensitivity, it is argued that some parameters are similar to DS, and thus the paper focuses on the new parameters. Nevertheless, although the proposed method borrow some ideas from DS, it is clear that the approach is different (In DS, you stop searching as soon as you find an occurrence with distance below the threshold, here you continue scanning to get the cpdf). Therefore, we cannot assume without checking that some parameters such as the threshold, the search neighborhood or the fraction of the TI that is investigated will have the same influence in DS and in the proposed approach. In addition, some interaction between the parameters is expected. For example, between the threshold and the maximum number of occurrence, or the number of available sections, some interactions can be expected. Indeed, if you increase the threshold, you will reach faster the maximum number of occurrence.

4) The application example is not a real application, but another synthetic benchmark using a real analog. Indeed, there is no verification data nor specific application of the model. I would therefore merge section 3 and 4.

Specific comments:

1) P3L6-25. You don't discuss the paper by Gueting et al. (2017) in the introduction, although it is a recent paper on the topic and you borrow some ideas later in the field example. You should reconsider this paragraph to describe with more details previous approaches and how your method is new.

2) P4L29-30. This sentence does not necessarily refer to the approach proposed by Comunian et al. (2012), but more globally, to any method assuming stationarity. It is always possible to use auxiliary variable to account for non-stationarity, as it is done by the s2Dcd approach in the application example.

3) P5L3-8. For reconstruction algorithm, there might be a confusion between training image and hard data. Here, the sections that are used are both training images and hard data. To some extent, using the whole section as TI and the sub-sections as hard data is already a good way to locally constrain the simulations. The S2Dcd algorithm is already performing well in that sense.

4) P5-L9-10. The choice of the subsections within domains is not necessarily selecting the most local information. Indeed, this depends on the location of the node within the subgrid, some other sections might be closer, for example when close to a domain boarder. Since the approach does not store any cpdf, you could center the sub-domain on the node and the TI would change for each node. This would avoid the issue of case 2 in Figure 2, where many nodes closer to the node to simulate are actually out of the TI.

5) P7L18-20. I am not sure it is intuitive, you might expect that the parallel sections are somehow correlated (except in case of strong non-stationarity), the multiplication of

probabilities could then make more sense... and otherwise for perpendicular directions (if the field is highly anisotropic, the orthogonal directions are bringing totally different information)...

6) P8L15-18. In short, you compute the cpdf using all the neighborhood whose distances are below the threshold, right?

7) P9L1-3. Multi-grid approach. The description is not clear. How do you divide the data event within several grids? I thought you were selecting all the data and previously simulated nodes within the radius. Are you only considering neighboring nodes that are on the current grid? Or you just mean that for the first grid there are less points simulated? In practice, everything depends on the radius. Please clarify.

8) P9L4-5. I understand that you take the "diagonal nodes" (figure 3) for the smaller grid as they are previously simulated in the multi-grid 2, but the other nodes should also be consider (horizontally and vertically) according to your radius, thus 8 nodes in 2D and 26 nodes in 3D, no? The remark about 3D is strange since you are only considering 3 directions in 2D, this is not a 3D neighborhood or am I missing something? How do you consider previously simulated points or hard data that are out of the 2D planes? Are they simply disregarded, or somehow projected on the plane?

9) P18. Figure 9. The red line shows the proportion in the cross-section, what about the "true" proportion?

10) P19L6-14. You could actually quantify how realizations are realistic using kernel smoothing (to average the information of the different realizations in a metrics), estimating the density distribution of the realizations around the reference. You can refer to the already cited paper of Hermans et al. (2015) for an application of MDS in the context of 3D MPS in hydrology.

11) P20L15-16. Still, in DS no cpdf is computed as the first matching sample is selected. Here, you continue scanning, so the effect of those parameter is not necessar-

ily similar. For example, an interaction between t, the number of sections and Nmax is expected.

12) P22L6. You say that the number of sections is insufficient, but you don't provide guidelines about a sufficient density to use your algorithm. What if the sections are not oriented in orthogonal directions?

13) P23L2-3. A short description of the auxiliary variable would be welcome. I guess it describes the proportion of facies in the different zones along the vertical direction. It does not have to be long.

14) P26L2-3. "partial lower dimensional data" is not true. You can only use 2D orthogonal sections in a sufficient amount. Borehole or analogs cannot be used since you argue that local information is important.

15) P26L7. But you need a lot of 2D sections, which is a clear drawback of the method.

Technical comments:

1) P12L6-7. Does the "maximum search distance" correspond to the Radius of previous sections?

2) P21L9. You previously mentioned (P17L15) that S2Dcd was using DS with 4 processors. Please check.

3) P23L11. I think it should be Figure 17 instead of 16.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-256, 2018.