Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC3
https://doi.org/10.5194/gmd-2021-297-RC3, 2021


# Comment on gmd-2021-297

Gregoire Mariethoz (Referee)

---

Referee comment on "Mapping high-resolution basal topography of West Antarctica from radar data using non-stationary multiple-point geostatistics (MPS-BedMappingV1)" by Zhen Yin et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-297-RC3, 2021

---

Review of « Mapping high-resolution basal topography of West Antarctica from radar data using non-stationary multiple-point geostatistics (MPSBedMappingV1) » by Z. Yin et al.

This manuscript addresses an important question: the interpolation of sparse data in a non-stationary context. It develops a framework that is seen as generic and demonstrates it in the specific context of interpolating subglacial bedrock topography. The novelty is the use of a large number of training images coming from a variety of deglaciated areas, and to devise a scheme to select a subset of them that is specific to each simulated region.

My overall assessment is that the manuscript is scientifically sound and will have an impact in the glaciology community, and also possibly more generally in other applications of MPS to large-scale problems, including space-time simulation. The tests in section 4 clearly show the value of the approach. Importantly, this study improves our understanding of the usage of MPS in the case of very large models that are fed by lots of data and training images. It demonstrates the need to select a specific subset of training images for each region rather than using the entire training image database for all areas. This strategy provides gains in terms of quality and computation, which is very valuable. Also, it could fuel the discussion on using machine learning approaches to address such problems (e.g. GANs) which would require a new training for each sub-area, making them inefficient.

This said, I do have remarks on some technical aspects of the proposed approach that I outline below. Despite these comments, I find the method convincing and I believe the manuscript deserves publication in GMD:

- While the approach is novel, some of the existing literature was missed. For example,

on l.63-54, it is mentioned that this is the first application of MPS to subglacial topography. A recent paper doing just this appeared in The Cryosphere: https://tc.copernicus.org/preprints/tc-2021-161/

- The proposed method involves a number of modeling steps and choices. This is fine when justified, but here I did not see a clear justification for many of the choices made. I detail some instances of this below:
  - I do not see a clear motivation for using a modified Hausfdorff distance in eq. 1. Why using a distance between patterns rather than some statistical similarity between patterns, e.g. in terms of integral scales?
  - There exists a literature on methods to select one or more TIs based on conditioning data (e.g. Pérez, C., G. Mariethoz, and J. M. Ortiz (2014), Verifying the high-order consistency of training images with data for multiple-point geostatistics, Computers and Geosciences, 70, 190-205, 10.1016/j.cageo.2014.06.001, or Abdollahifard, M. J., M. Baharvand, and G. Mariéthoz (2019), Efficient training image selection for multiple-point geostatistics via analysis of contours, Comp. & Geosci., 128, 41-50, https://doi.org/10.1016/j.cageo.2019.04.004). This is the same problem that is addressed here with the probability aggregation approach. Such methods are not considered or mentioned. It is fine that the authors propose a new methodology, but the reason for not using existing approaches should be given.
  - The general approach proposed is rather sophisticated (probability aggregation - PSO optimization – kernel density estimation), which complexifies the implementation. From a user point of view, these steps should be justified. Basically, it is pretty clear how things are done but the why is not always explicit.
  - 303-304: this is a strong assumption because these distances are considered in a high-dimensional space. Can it be justified?
- The procedure for the choice of the TIs is rather complex. This is justified by the dependence between neighboring TIs which is modeled by probability aggregation. However, since there are local data everywhere in the domain, simply selecting a set of TIs statistically similar to the data might also perform well. Has this been tested? For instance, looking at figure 10, I am wondering whether a similar ranking could be obtained with a simpler approach, for instance considering similarity in terms of histogram and variogram similarity.
- It is likely that the training images need to be optimally oriented prior to simulation. Or possibly, one could use a rotation-invariant distance. Has this been tested?
- Section 3.1.3 and the legend of figure 4 are quite unclear to the uninitiated reader and should be improved.
- 227: This statement about boundaries seems incorrect. DS is affected by the boundaries of the TI, as the edges of the TI cannot be sampled, especially for large data events.

Minor comments/typos:

l.22: sentence grammatically incorrect

l.23: provides

l.24: flight lines data

l.86-87: here the work by G. Pirot could be mentioned as it does exactly what is mentioned in this sentence (https://doi.org/10.1016/j.geomorph.2014.01.022, and also doi:10.1002/2015WR017078).

l.96-97: Awkward sentence, rephrase

l.115: remove extra parenthesis

l.149: section 0 is a mistake

l.182: verbatim copy should be explained

l.204: remove word expressed

l.286.287: this reference seems out of place. This is the TI selection, not DS simulation

l.287: erroneous reference: there is no section 3.13

l.292: parameterization … convergence

l.293: make PSO

l.320: “is to reflect” -> should be rephrased

l.370: section 0 doesn’t exist

l.397: rewrite sentence (diving ->dividing)

Figure 18: problem of color scale where white areas appear in the simulation domain

l.424: density of water is 1000 kg/m3

Check the citations: several are duplicated.