Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-211-RC2, 2019

Geoscientific
Model Development
Discussions
Open Access
EGU

Interactive comment on "Quantile Sampling: a robust and simplified pixel-based multiple-point simulation approach" *by* Mathieu Gravey and Grégoire Mariethoz

Thomas Mejer Hansen (Referee)

tmeha@geo.au.dk

Received and published: 27 December 2019

The authors present a novel multiple point statistical simulation algorithm that works for both discrete and continuous data, that scales well on parallel computing architectures, and that is available as open-source C++ code (G2S) with interfaces in Matlab, Python and R.

At the core of the method is the use of convolution to very efficiently compute to compute a mismatch, between a conditional event (consisting of the 'N' closest hard/simulated data) centered at all locations in the TI (except near the boundaries) (2.3) Then the authors suggest to simulate the current pixel based on a random selection between the 'k' centered pixel values associated with the smallest mismatch (2.4)

This leads to an algorithm with only two main 'tuning parameters'. The algorithm is in-itself novel and has obvious potential for used instead of some of the currently widely used MPS methods. The examples in the manuscript nicely describe the potential uses. In addition, the way the algorithm has been implemented should be applauded, as it is available as Open Source code that can be used with ease ranging from a case of "running on a single thread on a laptop in python/Matlab", to "running remote on a large cluster". This makes the code very versatile.

Therefore I find the manuscript highly appropriate for publication.

I have one major comment, that relate to the name of the algorithm and the way a pixel value is chosen based on 'k' smallest values of E/mismatch. The authors refer to these 'k' smallest values of E as a "quantile" and call the algorithm, for quantile sampling. This I do not understand and find a bit misleading. How can this represent a quantile? I think the term 'threshold' would be more fitting than 'quantile'.

The use of the term "quantile" suggests that the selection of the new pixel value is based of a probabilistic measure. Also, say 'k=18', and for a discrete case only 9 pixel values are associated with a mismatch of '0'. Why would one want to use the same probability (P=9/18) to select one of these, as opposed to one of the pixel values associated with a non-perfect match (P=9/10)? Or more extreme, say that pixel associated with the 18th best mismatch has a mismatch of 10 pixels. Why would one want to assign the same probability (1/18) to this, as to the pixels with a mismatch of 0? The use of the 'k'-'threshold' is convenient, but to me it makes the method less clear to describe in terms of the implied statistical assumptions. Some discussion on 'quantile' vs 'threshold' would be good.

Some comments to the text:

Line 150: Here ‘a’ and ‘b’ are referred to as “univariate pixel values”. It seems ‘a’ and ‘b’ has a different meaning in line 174 (eqn 3)? Here they seem to represent vectors?

Line 185, Eqn 5: Please elaborate a bit on how this allows mixing discrete and continuous variables calculating the mismatch? It seems nontrivial to compute the mismatch between for example a velocity of 2.1 km/s and a “lithology of type A” to a velocity of 2.13 km/s and “lithology of type C”?

Figure 1: What do the red dots in the middle small figure?

Line 287: Please explain clearly what is meant by “verbatim copy”. The term is used several places without a proper definition.

Line 338: Please explain “NUMA-aware” or provide a reference.

Line 392: What is meant by “..enables adaption of the parameterization…”?

Figure 5: Please help the reader here: is Qs with a kernel better than QS with no kernel? I am not sure what the figure tells us?

Line 399, Figure 6:Perhaps you could elaborate a little bit on “Euler characteristic” and whether it is a problem what Figure 6 shows?

Figure 8: I need some help appreciating how Figure 8 suggests that the use of alpha is useful?

Figure 9: Please show the ‘dots’ (the actual CPU time measurements) in the figures. Is it fair to say that the main limitation of the using QS is the size of the training image?

Lines 466-472. It is nice that one can choose to use many conditional point with not extra CPU costs. one could though argue that sometimes it is convenient in other MPS methods (SNESIM/IMPALA/DS) that the simulation becomes MUCH faster if one uses few conditioning data. If you would want to simulate with fewer conditioning data, QS would not lead to faster CPU time.. Just to say that the advantage you describe, could in a specific context, be seen as the opposite.

Some of the figures and tables in Appendix A should be excluded unless they are discussed and references in the text.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-211, 2019.