Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-211-RC1, 2019

Geoscientific Model Development Discussions

Open Access

GMDD

Interactive comment

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

Ute Mueller (Referee)

u.mueller@ecu.edu.au

Received and published: 26 December 2019

The paper describes a new algorithm for multiple point simulation of continuous and discrete spatial variables. To start with a short review of the various types of MPS algorithms is provided, which distinguishes patching from pixel based approaches. The algorithm described here falls into the second category. Shortcomings of the method are discussed briefly, including the need for a threshold and sensitivity of the simulation quality to this threshold, but which can also lead to very long simulation times. In this paper the authors exploit a decomposition of the distance measures to apply FFT to speed up computation of mismatch maps with the aim to more quickly identify

Printer-friendly version

Discussion paper

candidate patterns in the training image, which may be complete of incomplete. The use of the FFT to compute the mismatch map is attractive in that it is fast to compute irrespective of dimension.

The mismatch map is calculated by computing for each pair (s, t) a dissimilarity measure where t belongs to the training image and s to the conditioning set. It is this dissimilarity measure which is then identified in terms of cross correlation. The authors provide a description of the metrics applied and a rewrite of the metrics in terms of cross correlations, and while the reader gets a general idea as to what is being calculated the derivation is patchy and somewhat sloppy in that summation indices are missing and critical steps are not described satisfactorily, such as the derivation of equation 9, which introduces cross correlations. Also, is it correct to assume that "l" is a grid operator? Once the mismatch map is computed, the k best matches are identified and a sample is drawn at random from this pool. The possibility of having non-integer values for k is touched upon, and allow unequal weighting of the first ceiling(k) candidates, with the first floor(k) candidates equally likely and the final candidate less likely (probability of being chose): 1-floor(k)/k) . The main advantage appears to lie in being able to choose between 2 instead of just one candidate (case of k between 1 and 2)

Simplifications and computational implementation details for speeding up the computation are discussed reasonably thoroughly and provide other practitioners with useful suggestions on how to potentially improve the efficiency of their own MPS algorithms. The proposed algorithm is benchmarked by means of standard sample data sets and a sensitivity analysis is provided demonstrating that QS performs well subject to the choice of a suitable kernel and that the quality of QS simulations is similar to that of DS simulations. It would have been interesting to see an exploration of kernels other than one of Gaussian type. Also, the metrics being used to assess the performance would benefit from going beyond variograms and connectivity (I acknowledge that the Euler characteristic was also used, but what good is it without a definition? Reference to another paper is all fine and well, but a definition and an explanation of what it measures would have been nice.) It would be really nice to see an evaluation in terms of a multipoint statistics.

Please amend all the formulae to ensure summation indices are clear, eg: Line 149: It is not clear over what is summed in equation 1. You clarify this to some extent below in lines 150 to 183, but I find this a little unsatisfying Line 174: The description preceding equation 2 talks about vectors, but the formula seems to be univariate. If you have c categories, is “a” a vector with c entries or simply one of the values from 1 to c if you label the categories in that manner?, It looks to me that “a” is simply a category . . . so looking at the equation, it would seem that it is equal to c, if “a” and “b” are distinct and equal to c-1 if they are equal, while the sum on the right is equal to 1 if “a” and “b” are equal and 0 else. There are also brackets missing in the middle expression (you should have \sum_{j \in C} (1-\delta_{aj}\delta_{bj}) Line 200: N(t) is not just a location but a neighbourhood? Please clarify Line 230: define the cross-correlation operator. Also, T_i has not been defined. You identify “*” with convolution and then apply the convolution theorem. Provide a derivation that this is true in an appendix. There are also some typos in the figure captions

---

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

Printer-friendly version

Discussion paper