

Geosci. Model Dev. Discuss., referee comment RC2
<https://doi.org/10.5194/gmd-2022-229-RC2>, 2023
© Author(s) 2023. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2022-229

Ute Mueller (Referee)

Referee comment on "AutoQS v1: automatic parametrization of QuickSampling based on training images analysis" by Mathieu Gravey and Grégoire Mariethoz, Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2022-229-RC2>, 2023

The paper by Gravey and Mariethoz provides an interesting approach to determining a suitable parameter set for geostatistical multiple point simulation. The method is based on simulating individual pixels and using the prediction error as the sole metric to find optimal parameters. The approach is sufficiently novel to warrant publication ultimately, but at present it lacks clarity and the quality of the scientific writing needs to be improved.

Comments on the Introduction

1. You state that the parameterisation depends on the specifics of the algorithm. This is undoubtedly true, but it would help the reader to have a quick synopsis of the two algorithms and associated parameters before going into details of selection of parameters.

2. Line 45: I would not talk about philosophies but rather about approaches here.

3. Lines 47 to 49: It would help if you explained what kind of metrics were computed and I assume you mean that the corresponding metrics computed for the simulation matched as closely as possible. But this is not stated clearly.

4. Line 60: You say "If both approaches show good results, then they are both related to optimization methods, and therefore the user has no control over the duration of the optimization process". It is not at all clear to me what you are trying to say here. Both approaches rely on optimization of some kind, irrespective of the quality of the results. So can you please be more precise?

5. Lines 70-71: I am not sure what you mean when you say "the underlying principle of our approach is that a sequence of well-simulated pixels converges to a good simulation overall". Please clarify! (what does well-simulated mean? How can a sequence of pixels converge to a good simulation? Good in what sense?)

Comments on "Challenges related to inappropriate parameters"

Most of this section is about verbatim copy, should you maybe change the title of the section to a title that highlights this? Also, it seems the issue of "verbatim copy" is one of DS and QS only? It is also not clear how constraining the conditional probability distribution of $Z(x)$ given an MPS algorithm is related to the issue of verbatim copy, (for ENESIM, SNESIM and IMAPLA this justifies the need of a large enough training image)

6. Line 90-91: it would be useful to introduce the abbreviation for the threshold here

7. Line 92 The definition of "verbatim copy" should be provided here rather than in lines 98 following.

Comments on Method

8. Line 125 : The sentence "Binary variables are a particular case of continuous and categorical variables." seems a throw-away comment whose purpose is entirely obscure.

9. In the pseudocode for the QS algorithm it would help to have a definition of the entire parameter set

10. line 138: What do you mean by "find a candidate in T those matches $N(x)$ using θ ?"

11. line 144: What do you mean by "A perfectly simulated pixel is a pixel that respects the conditional probability distribution" what probability distribution do you mean here? Presumably this is related to formula (1)?

12. line 159: please define the discretised parameter space θ (it is not really defined in line 1 of algorithm 2)

13. line 161: you need to define what "th" stands for (I do realise you mean threshold,

14. in line 162 you talk about representative stages D of the simulation, but then in line 164 you say that D represents the density of a neighbourhood. What do you mean by density of a neighborhood? Also, is D part of the parametrerisation θ ? If not, why not?

15. Should algorithm 2 not have as a first step "randomly generate a set V "

16. What do you mean when you say sample a neighborhood $N(v)$ from T respecting D ???, see also line 180

Comments on "An efficient implementation"

Section 3.1 needs to be proofread carefully. You really should adhere to principles of good scientific writing and avoid starting sentences with symbols and also check word-order.

17. Line 198 The variables θ_h and θ_S represent sets. Sets cannot be added, but you can define their union. Thus $\theta_h \cup \theta_S$ would be more appropriate than $\theta_h + \theta_S$, alternatively write $\theta = (\theta_h, \theta_S)$

18. the criterion you list in formula (4) needs further explanation and the sentence "... a given parameterisation is only further explored if the error is a range of a σ " does not make sense. To me the top line in 203 would make more sense written as

$$\epsilon(\theta, D, T) - \epsilon(\theta_{\min}, D, T) > \frac{1}{2} (\sigma(\theta, D, T) - \sigma(\theta_{\min}, D, T))$$

Comments on Results

19. Line 213 the term "uniform" has specific connotations. I would suggest the term "fixed" might better capture what you mean here. Also, does the kernel you consider here also have radial shape and why do you use ω , when later on you use w ?

In figure 3 you introduce the term "ignorance treshold" without provision of a definition.

20. Lines 227-228: "we also note that even if the parameterisation is logical it is difficult to predict" what do you mean?

21. I believe Figure 4 warrants more discussion. Also, in the discussion on line 238 you use the term "two stage process" but that seems ill chosen for the behaviour you describe

22. line 252: It is the neighbours furthest from the location considered that are allocated the negligible weights with large values of the shape parameter α , and the sensitivity of n decreases, but n doesn't become insensitive

23. in line 270 the word adaptive needs to be inserted

Figure 8: the title of the figure in row 3 column 1 needs to be corrected.

In the appendix, please provide a reference for each of the training images. It is interesting to note that the variogram reproduction for Delta Lena shows the "right" shape, but the sill are too low and connectivity is not as good as one would hope. Any thoughts on why this happens?

It would also be good to see some results for multivariable images. You allude to the

algorithms working in this case also, but being slow