Response to Prof. Ute Mueller

We thank Prof. Ute Mueller for her feedback and interest in our work!

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.

We thank you for the positive feedback

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.

We appreciate your suggestion to introduce the algorithm earlier in the paper. We agree that it would improve the flow of the paper and help readers better understand our methodology. To address this, we will add a brief explanation about the algorithms in the introduction, and we will ensure that the pseudo-code for Algorithm 1 is presented clearly.

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

We agree that the term is more appropriate and we will update to "approaches".

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.

Here we distinguish between matching the training image statistics VS matching the conditioning data statistics. This was indeed not very clearly explained. To address your concern, we will add a couple of sentences to our manuscript to provide further clarifications on Dagason et al.'s approach and explain how their methodology was used.

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?

We mean that the result should be good for both approaches, at least based on the objective functions, however it is impossible to know ahead how long this optimization process can take. We will rephrase the sentence to better clarify our point.

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?)

The divide and conquer philosophy suggests that if each pixel simulation is statistically accurate, then in a sequential simulation the resulting full simulation should also be statistically accurate. This is the principle behind sequential simulation. We understand that this sentence can be difficult to understand and will attempt to clarify it.

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?

Thanks for the suggestion, we will do it.

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)

We agree that more clarifications and context are needed on the issue of verbatim copy. From a technical standpoint, if we disallow data events with less than two replicates in SNESIM (and similarly in ENSESIM or Imapla), we guarantee to avoid verbatim copy in the resulting simulation. However, in practice, users often choose to allow such unique data events because they provide texture that can lead to visually satisfying results. There are therefore examples where verbatim copy does occur with algorithms such as SNESIM, although it is much less visible with categorical variables than with continuous ones.

We partially disagree with the idea that the training image plays a significant role in pattern variability. For example, in a chessboard pattern or any other repetitive texture, the image can be infinitely large, but the patterns remain the same. If the issue is not directly related to the size of the training image, then we agree that, in practice, using a large training image is the simplest solution to increase pattern variability, but this does not guarantee the absence of verbatim copy, even if the image is stationary.

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

It will be added.

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

We will restructure the section, change the title and start with the definition of verbatim copy.

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.

Unfortunately, we have to respectfully disagree with the reviewer's comment regarding the importance of this. In our experience of teaching MPS, we have found that this property is not trivial for everyone, and hence we believe it is essential to highlight its significance. Moreover, this property removes the need to manually manage the binary cases.
That being said, we are open to improving the sentence to ensure that our meaning is clear and unambiguous.

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

We suppose that this refer to \theta, this would be added.

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

We will consider adding a sentence to the main text to clarify certain parameters like respecting the threshold or the fraction in Direct Sampling, and k (and kernel) in QuickSampling.

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)?

Yes indeed it is equation 1.

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

We agree that \theta is explicitly defined, this is because we see the method as generic, but we will specify this in Algorithm 1, and provide the specifics related to our case studies with QS.

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

It will be added.

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 theta? If not, why not?

Indeed, D represents a progression in the simulation and therefore the density of neighbors informed (assuming a random simulation path). For example, at 50% of the simulation, we have statistically 50% of neighbors informed. We realize that this is an important factor and that the complete parametrization (k or f and th) is dependent on the simulation stage. The notation will be revised to theta(D) to reflect this. Being a key part of the methodology, we will put a strong effort on clarifying this point. Thanks to highlight that this point was unclear.

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

This is not really relevant, and in practice they are generated on the fly, so the current version is more accurate. But for sure they could be generated ahead too without changing the result. However, to make this point clearer and maybe more intuitive we will describe V as random generator as a 1st step.

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

Same as comment 14, we think that clarifying the definition of the density will improve this too.

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.

This was also mentioned by reviewer 1 and we agree that the entire manuscript should be proofread carefully.

17. Line 198 The variables \theta_h and \theta_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)

Thanks to have highlighted this notation issue. In fact it is closer to a Cartesian product, we will adjust by replacing the **+** with a **x**, and an explanatory sentence.

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 \sigma" 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))

Not with a difference of sigmas, but with a sum, yes. This is maybe a taste question. But we will rewrite the equation to isolate sigmas and epsilon, and extra description to clarify the equation.

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 \omega, when later on you use w?

Here we believe that the term uniform is appropriate. It means that all neighbors have the same weight, which is the definition of uniform. However we agree that it may be worth explaining clearly in the manuscript. Since we use a distance to neighbors (or a radius), it is indeed a radial shape.

We will uniformize the \omega and w notations.

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

We will add the definition.

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

The algorithms begin by simulating a large structure, followed by simulating smaller patterns. The simulation is done in a hierarchical manner, where each smaller structure is a part of the larger one, making the simulation logically consistent. We will revise this section to provide a clearer explanation of its intended meaning, which may not be immediately apparent in its current form.

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

We will expand the description of figure 4 to better convey our interpretations and to guide the reader.

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 \alpha, and the sensitivity of n decreases, but n doesn’t become insensitive

With large \alpha values, increasing n does not change the results as the weight of the additional neighbors becomes negligible. But indeed, one could indeed argue that there is still a sensitivity to n for small n values, where all neighbors are close to the center of the kernel. This will be mentioned and discussed.

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

Will be done.

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

Will be done.

In the appendix, please provide a reference for each of the training images.

Well be added

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?

The variance reduction in MPS is a generally under-discussed point that seem to be universal to all MPS solution, especially for continuous variables. We could not find a study that looks at the question thoroughly, and we believe that it should be the topic of further studies.

Interestingly, our results provide an intuition of why this happens. It seems due to the absence of compatible large-scale patterns in the training image. Essentially, if one wants to reproduce the large-scale connectivity observed in the training image, the only solution is to do a verbatim copy of the large structural elements found in the training image. The reason is that such large connected bodies are too few in the training image. Since our approach tends to avoid verbatim copy, these large-scale elements are not reproduced.

If taking a binary version of this training image (water/no water) and doing a simulation SENESIM/Impala, we will see that this difference really appear when we change the setting of minimum number of replicated from 1 to 2. As suggest earlier maybe a larger training image could certainly will help to have more pattern variability.

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

In fact, we tested it for multiple multivariable scenarios, and it is rapidly becoming slow, due to high dimensional parameter space. It is also hard to visually analyze or interpret the results. We didn't see the added value of just dumping the results at the end. The current solution is close to brute force, and we expect in the future to come up with an improved (mainly faster) strategy that takes into account the relative order in parameters (in particular in kernel weights) using a strategy closer to gradient descent.