

Response to Reviewer #1

We are thankful to Reviewer #1 for his/her valuable comments and suggestions, which will certainly improve the manuscript. The response to the individual comments is given below. The original review is quoted in *italics*, whereas our response is given in **bold font**.

5

This manuscript is a well-written and clear case study on the application of MPS to a very large domain. As such, it will be valuable for a range of researchers. While I recommend eventual publication, I also have reservations that should be addressed.

10 *Regarding the content of the study, I appreciate the overall methodology and the emphasis, throughout the discussions, on the fact that the training image and the simulation algorithm are all elements structuring the final models, and as such the evaluation should take place on unconditional realizations.*

However, I also found that the conclusions would be much better supported by adding a few elements:

15 *1) Currently only a single realization is used for each setting. This is clearly insufficient. On top of p. 12 it is argued that the simulation is considered representative, however I don't agree with this statement. Multiple realizations are needed to quantify uncertainty. It is possible that the single realization is representative, but the only way to find out is to compare with a set of other realizations and decide whether the inter-realization variability is small enough, according to a given criterion (e.g. flow, transport, etc). On p.12 (l.11-12) it is also argued that the methods to use multiple realizations do not exist, which is clearly not the case.*

20

Concerning the representativeness of the realization discussed in our manuscript, it is possible to state that, by definition, each individual realization is representative. In fact, every realization is by construction compatible with all the input information (i.e., the statistics formalized by the training image, the hard data, and the soft conditioning).

25

Concerning the uncertainty assessment, even if it would be definitely very interesting in principle, we feel it would be out of the scope of the present research that is solely dealing with the development of the optimal strategies for conditioning the simulation and for preparing effective training images.

30 *2) The assessment of the results is mostly qualitative, both regarding the patterns produced in the model, and to assess the proportions variability (top of p.8, top of p.9). The tools to do exist and should be used. Also, quantitative comparison of the modeled patterns and the patterns in the conditioning data would be a good validation.*

With respect to a quantitative analysis of the proportions variability, we think that our point (i.e.: the kriged sand probability is effective in enforcing the proper spatial trend) is clearly shown by the comparison between Fig 14b and Fig. 8. The two figures allow a voxel-by-voxel comparison between the soft conditioning distribution and the final corresponding realization. So, it is not clear to us what kind of more quantitative argument we should use. On the other hand, we agree that we should make this comparison more evident in the text and, in the revised version of the manuscript, we will add few lines on this respect.

Regarding the quantitative assessment of the produced patterns, we believe that the comparison of Figs 10a and 10b, showing the unconstrained realizations associated with the training images TI1 and TI2 (in Figs 9a and 9b respectively), demonstrates quite well the large effects on the final realizations caused by relatively small perturbations in the used training image. However, even if the difference between the patterns in Figs 10a and 10b is quite evident, if requested, we could specify, in the revised version of manuscript, the value of the mean volume of the sand bodies in the realization generated with TI1 (Fig. 10a) and the corresponding value for the realization associated with TI2 (Fig. 10b).

3) The literature review part of the introduction is quite incomplete, missing a number of studies that have looked into 3D MPS models. On p.2 l. 25 it is said that not many studies have looked at 3D TI-based models. I disagree, with for example Ronayne et al (2008), Jha et al (2014), Perez et al (2014) to name a few, and a lot of other studies in reservoir engineering as well. For non-stationarity also, there are Cuhgunova et al (2008), Straubhaar et al (2011), and possibly others, who made important contributions.

We acknowledge the relevance of some of the suggested studies. In the revised version of the manuscript, we will include additional references.

Regarding the structure of the document, I also have 2 remarks:

1) There is an imbalance between the description of the data and methods, which is quite short (6 pages), and the discussion/conclusion, which is 5 pages. There is clearly too much material in the discussion, including elements that could be removed or moved to other sections. Here are some suggestions:

- P.7, l.15-20: this could go in section 5.4.

The rationale behind our original choice is that, in the initial part, we wish to simply present the different inputs and how we used/prepare them.

In the second part (from paragraph “6. Results”), we show the effects of the choices we made and the reasons for these choices. We do this by means of a detailed discussion of the corresponding results.

So, even if it might seem unbalanced in terms of length of the different parts, we believe that, in this way, the paper is more effectively organized from a logical point of view, with a clear separation between the inputs (and their preparation) and the outputs (and the associated assumptions/choices).

5 - P.9, from l.19: *this could go in the introduction*

We definitely see the Reviewer's point.

Even if we feel that a few words spent on briefly (re)introducing the concepts can make the paper more easily readable, in the new version of the manuscript, we will follow the suggestion.

10

- P.10, l.22 to p.11, l. 2: *This is not related to the purpose of the paper and could be removed.*

Also in this case, we understand the Reviewer's point.

15 **In the first place, we decided to add this part to reinforce the discussion on the possible use of seismics via the comparison between the characteristics of the seismic lines collected in this area with those acquired somewhere else for hydrostratigraphic studies.**

- P.11, l.3-10: *The method for the conversion of boreholes to probabilities should be described in the methodology section, not here.*

20

The conversion of the borehole into probability is indeed discussed in the "5.3 Borehole data" section, where we describe the methodology we use to prepare the inputs. At page 11, in the section "7 Discussion", we simply recall that strategy to mention possible, straightforward extensions of the presented approach, for example, to the case where boreholes have varying quality.

25

- P.11, l.11-19: *There could be a separate section on non-stationarity because it is mentioned often.*

30 **Non-stationarity is tackled by kriging the probability derived from the boreholes. For this reason, we think it is more logical to keep this aspect tightly connected (across the entire paper) with the discussion about the borehole data and the sand probability spatial trend.**

- P.11, l.20-34: *This is a long paragraph on something that is not done. It could be removed.*

Actually, we believe that a discussion on why we made the choice of not to do/show something could be relevant for the community and can contribute to the overall clarity and usefulness of the manuscript.

2) Sections 2, 3 and 4 could be grouped together as they all relate to the description of the study site.

5

We see the Reviewer's point.

In the original manuscript, we decided to keep these sections separate in order to maintain:

i) the prior overall geological understanding of the entire area,

ii) the observed and utilized data (seismics and boreholes), and

10 iii) the description of the specific geological unit to be investigated (the Miocene)

well distinct for expositive clarity and logical sequentiality.

Other remark:

15 *Figures 7 and 8: the green-purple color scale is very subjective and seems to highlight values around 0.4. It creates artificial discontinuities. A usual continuous color scale (rainbow or grayscale) would be better.*

This specific color scale has been selected because 40% is the target marginal distribution value (as it has been derived from the boreholes and as it is consistently formalized in the training images). The details for this choice are described in the methodological section "5.3 Borehole data" where we discuss our approach for dealing with borehole uncertainty and for translating the lithological information into probability. For these reasons, we believe that the adopted color scale is not subjective and the value 0.4 has a specific meaning as it corresponds to "no information" about the occurrence of sand or clay.

20