

## ***Interactive comment on “Towards geologically reasonable lithological classification from integrated geophysical inverse modelling: methodology and application case” by Jérémie Giraud et al.***

**Tom Horrocks (Referee)**

tom.horrocks@uwa.edu.au

Received and published: 12 December 2019

### **1 General comments**

In the manuscript entitled “Towards geologically reasonable lithological classification from integrated geophysical inverse modelling: methodology and application case”, Giraud et al. present a method for integrating a variety of pre-existing subsurface models. The pre-existing models that were integrated in the case study are impressively

C1

diverse, being geophysically-derived (i.e. the inversion model and its spatial gradient), structurally-derived (i.e. the geological uncertainty matrix), or otherwise produced by a traditional workflow (i.e. the inversion start model and the most likely lithology model). The estimated lithology model, which is based on all five input models, is then refined during a ‘post-regularisation’ step which enforces some degree of geological plausibility.

This manuscript addresses an important area of research, since it is rarely the case that a single unified lithology model is computed by joint inversion of all available information, particularly in industry due to time constraints, different datasets becoming available at different times, etc. The application of the method to refine a real pre-existing lithology model is also pleasing. As my background is primarily in machine learning, my comments below mostly address the statistical aspects of the manuscript.

### **2 Specific comments**

The first step of model integration relies on a self-organising map (SOM). Ignoring the post-regularisation, the SOM is effectively (1) trained on all input models (including most likely lithology), and then (2) used to produce a new lithology model by replacing each original lithology cell with the most popular lithology within its best matching unit (BMU). The lithology population ratios within each BMU is interpreted as the probability of the cell being that lithology given all input models. The following points address this use of the SOM.

*Justifying the SOM:* Upon first reading, it was unclear to me why a clustering/partitioning algorithm (SOM) was being used for a classification problem (technically re-classifying lithology). I later realised that this approach actually allows one to view updated versions of *all* input models, which is later necessary for the geophysical consistency check. Please emphasise this earlier, perhaps where you state that

C2

“lithologies are the quantity sought for” (Page 6, Line 15).

*Lithology probabilities:* I am unaware of any research showing that probabilities taken from a SOM in this way are well-calibrated. Note that this is a separate issue to the uncertainty of these probabilities which is instead related to the variance of the SOM's probability estimates and the samples size—the probabilities could have low ‘uncertainty’ but still be entirely incorrect (i.e. poorly calibrated). A new figure showing a probability calibration curve would reveal if the probabilities are reliable; the “Probability Calibration curves” page in the Scikit-learn documentation provides a good overview if the authors are unaware of this technique.

*Probability uncertainties (Appendix A):* It seems unnecessary to use a Gaussian approximation here, especially since you then have to check that each SOM cell contains enough samples. As far as I can tell, a beta distribution would be the perfect fit this problem, since it is defined over  $x \in [0, 1]$  and will work regardless of sample size.

*Selecting the number of SOM units:* Generally speaking, the degree of regularisation (here the number of SOM units) chosen is that which maximises a performance metric (here accuracy) on the validation set. However, I note that your accuracy continually increases as the number of SOM units increases. The causes of this could be: (1) you have not reached the point of overfitting and need to increase SOM unit count further; (2) your training set accidentally includes your validation set; or (3) the statistical problem / the validation set is too ‘easy’, e.g. most samples in the validation set have near-copies in the training set. Point (3) seems most likely, and may have come about if you generated your validation set completely spatially randomly, in which case most voxels in the validation set will have spatial neighbours in the training set. In this case, a validation set containing a few strategically chosen contiguous chunks of voxels might be more suitable.

Alternatively, you may want to additionally plot the geophysical misfit (equation 9) as a function of SOM cell count. This way, you could directly select the SOM size that

C3

results in the most integrated model that is still geophysically permissible.

The following points address other aspects of the paper.

*Geophysical consistency:* The geophysical consistency is checked by comparing the updated model's response with the original inversion model's response and checking that they differ by at most the noise threshold. However, the threshold check is an model-wide average one (based on the L2-norm), and so it's possible that the response at a few stations is significantly different (beyond the noise threshold). Can you provide some summary statistics on these (if they exist)?

*Figure 9:* What I think is missing from this figure, or perhaps the accompanying prose, is an illustration of how many voxels were ‘inclusions’ (where all neighbouring voxels were different lithologies), and how many voxels were changed due to the other adjacency constraints. Currently, it's impossible to know how much of an effect the interface constraints are actually having.

### 3 Technical corrections

- In-text citations for erroneous trailing commas (e.g. page 1, line 27) and spacing errors (e.g. page 2, line 14).
- When a reference is referred to as a noun, a textual citation is preferable. For example, “In addition, apart from (Zhao et al., 2017). . .” is more correctly written as “In addition, apart from Zhao et al. (2017). . .” or similar.
- Page 1, line 11: “apply it to field data”
- Page 2, line 28: facieses → facies
- Page 4: The reason for the  $H$  subscript in  $W_H$  is never explained; a short explanatory note would be nice (is it related to entropy)?

C4

