Solid Earth

Discussions

Open Access

EGU

[SED]

Interactive comment

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

Anonymous Referee #2

1   General Comments

The authors present a new method to estimate likely subsurface lithologies and associated uncertainties by applying self-organizing maps to geophysical (a 3D density model), geological (a 3D lithology model), and derived model products (density model gradient, etc.). With the aim of quantifying uncertainty in the potential lithology, the method presented is novel and a welcome contribution to the existing literature on this topic as it relates to resource exploration. My background is geophysics and many of

Printer-friendly version

Discussion paper

my comments are in relation to those aspects, whereas aspects related to the machine learning are far afield for me.

I think the title containing 'integrated geophysical inverse modelling' is misleading, as only a single geophysical model is used in the work. Further to that, geophysical inverse modelling is not the main focus of this work that simply uses a density model obtained from geophysical inverse modelling. I suggest a title along the lines of 'Towards geologically reasonable lithological classification from geologic and geophysical models: methodology and application'.

Throughout the text, there are multiple phrases that refer to the same thing. For ease of reading, I would suggest selecting one phrase and using it throughout so there is less opportunity for confusion on what exactly is being referred to. For example, 'inversion results (Page 2, line 1)', 'inverse modelling results (Page 3, line 7), 'geophysical inverse models', 'geophysical models', and 'model' all seem to refer specifically to the density model from geophysical inversion of gravity data. Selecting a single phrase to refer to this association would be helpful.

The text was somewhat difficult to follow, and I found myself flipping through multiple pages either forward or backward as sections were referenced throughout the text. It seems that the structure of the text could be streamlined, and the complex outline could be simplified. Additionally, some section headings were not needed while others suggest the section will deliver more than it does. There are multiple instances of a section heading followed by a subsection heading, with no text or discussion introducing the reader to the flow of the text and sections to follow. I find this approach to sections and structure confusing and a deterrent to continued reading. Specifically, '2.1 Integrated geophysical modelling' is broken into two sections that could be combined into one section on 'Geophysical and geological modelling' since most of the details within the few paragraphs that compose the two sections area left as references to other works.

As I read the content within '2.2 Classification using SOM and uncertainty analysis',

I found that I was jumping from section to section (often ahead) for needed context. This section seems to be overly fragmented and content should be shifted to reduce the number of forward/backward section references. For example, the mean quantization error (Q) is stated as the metric used in '2.2.2.1 SOM specifications and training procedure' but the technical introduction to the metric is left to section '2.2.2.2 Training and testing datasets'. The content within these two sub-sub sections could easily be rearranged into one section under the single section of 'SOM training and validation'. Further, the paragraph under '2.2.1 Fundamentals of SOM' could readily be moved under '2.2 Classification using SOM and uncertainty analysis' (which currently is a lone section header with no text).

2   Specific Comments

- Paragraphs in the introduction seem piecemeal and could use some editing to link the topics together. Perhaps consider including a summary paragraph that will serve two purposes: tying all the topics included in the introduction together and providing a roadmap for what is to come in sequence within the rest of the manuscript.

- Would Figure 1 showing the method be better placed in the methods section? Placed in the introduction with a single sentence reference the figure with no discussion only causes me confusion as to what I'm supposed to learn from the diagram in the context of the introduction. The diagram itself is also a bit confusing. Is there a purpose for listing the numbers 1-5 and a-c? Data, concepts, and methods are all mixed in the diagram - might be clearer if the boxes for each were distinguishable from each other (e.g. inputs, outputs, algorithms).

- Would it be worthwhile explicitly stating that there are two separate examples, a synthetic and field case?

- Part of the method to estimate uncertainty is predicated on the fact that the post-regularisation step results in a density model that is within the null space in a geophysical sense. This is a good metric to show that changes in lithology are within the uncertainty tolerance of the individual density model being used, however this does not address the broader aspect of uncertainty where the actual density model is concerned. The density model is, as pointed out, a deterministic model. Given the non-unique nature, if the specific density model is not correct than any perturbations in the null space based on the model are also in error. How do you account for the fact that the input data (density model) may be a major source of uncertainty (it could be altogether incorrect)?

- The only geophysical data used in this method and example is gravity data and the density model resulting from geophysical inversion. This does not become clear until well into the text. Consider explicitly stating this earlier on in the text, perhaps in the methods section where the geophysical inversion is introduced.

- How are lithology changes mapped back into density changes for the forward model comparison? Is it a range for the density or a distribution?

- The phrase 'integrated inversion' is used throughout the text. It seems to me that the type of geophysical inversion used for this work is typical single data set inversion. I do not think the term 'integrated' is applicable here.

- In text citations do not need parenthesis around the citation.

- Citation punctuation throughout needs attention, as there are unnecessary commas and periods throughout.

- In the References section, many of the references are missing the journal or publisher (i.e. Lelievre and Farquharson (2016); Lindsay et al. (2018); Meju and Gallardo (2016); Moorkamp et al. (2016)).

- Throughout the text, 'self-organizing maps (SOM)' appears. Introducing the abbreviation SOM should only be necessary in the first instance of fully writing out 'self-organizing maps'.

- There are several instances where 'This' or 'That' is used to begin a sentence. While it is appropriate English, I find it difficult to follow when this is the norm rather than the exception. An example of this is on Page 4, Line 29; 'This metric' could be replaced with 'Shannon's entropy' with little increase in text while making the reading easier.

- I have found the use of 'lithological model', particularly in the results section, confusing as this could refer to the initial model used for training; the predicted model; or the model generated from geologic observations. I suggest identifying three different phrases to consistently use and refer to each of these models to eliminate any possible confusion.

3 Technical corrections

- Page 1

  - Line 25-30: Please expand on the 'inherent duality' of geology and geophysics
  - Line 30: What are the geophysical quantities modelled from petrophysical and geological information?
  - Line 35: citations should be in chronological order?
  - Line 35: 'which consist in' should be 'which consist of'

- Page 2

C5

- – Line 4: Both e.g. and etc. imply a subset, use one or the other
- – Line 5-10: Language 'on the one hand', 'on the other hand', 'Nevertheless, like all modelling results' make it sound to me like these are opinions
- – Line 15: Is it possible to provide examples of the 'broad range of parameters'?
- – Line 15: 'results' should be result
- – Line 26: 'informed' should be inform
- – Line 28: need 'to' or 'for' in between 'consideration geological'
- – Line 30: Are you 'mitigating' that fact that 'no consideration is given to geological information and rules' or providing a method to address the gap?
- – Line 30: 'partially addresses the issues and shortcomings': please elaborate on the aspects still be tackled by you and the geoscience community
- – Line 35: Is this a 'fully controlled environment'? Using a 'semi-synthetic' dataset (Page 3, Line 15) implies there may be some unknowns in this environment.

- Page 3

- – Line 3: The phrase 'can serve multiple objectives' I think might be a bit misleading as these are not necessarily objectives of the method; as written it sounds like the first two items stated in this paragraph are motivating goals for why you have developed the method as it is. The third is an example of the method rather than an objective of the methodology.
- – Line 12: The a or b reference for Kohonen (1982)?

- Page 4

- – Line 9: 'approach which' should be 'approach where'

– Line 12: 'd represents measurements'; 'calculating the data'; the terminology for these should be consistent. Call d 'observed data' and then use 'predicted data' or similar consistent phrasing. 'Measurements' is not a common word when referring to geophysical data and in context, I took it to mean any measurement geophysical/geological when this is actually a reference to geophysical data only.

– Line 15: How is W_m specified?

– Line 25: 'geological models is' should be 'are'

– Line 29: Move citation Wellmann and Regenauer-Lieb (2012) to line 28 just after 'can be estimated'

• Page 5

– Line 1: What is meant by soft?

– Line 14: What exactly are the 2D maps?

– Line 20-22: Provide references for the Q elbow curve rather than the L-curve.

– Line 26: Will the fact that this is using semi-synthetic data have an effect? Will it alter the results?

– Line 30: Where does the starting model for inversion come from?

– Line 35: By 'datum', do you mean 'variable'? What is u(5)?

• Page 6

– Line 1: 'interpretable datasets'; What is an example of a non-interpretable dataset?

– Line 7-15: The text on the Q metric could be moved to the SOM fundamentals or where it is referred to a page earlier.

- – Line 20: remove 'are' in which are cannot'
- – Line 30-33: 'post-regularisation, which consists of'; typo 's' at the end
- – Line 34: Use of PR to reference post-regularisation could be defined the first time this phrase is used (in the introduction).

- Page 7

    - – Equation 3: What is u_k?
    - – Equation 4: What is l in n_l?

- Page 8

    - – Line 19-20: "with index B with 20% accuracy", what does this mean?
    - – Line 21: 2.2.4 Estimation should be Estimating

- Page 10

    - – Figure 3: is the dashed line showing sub-basin extent black or red?

- Page 12

    - – Line 2: "the inverted model and its spatial gradients" is restated in the next sentence

- Page 14

    - – Line 7: "It" should be "In"

- Page 19

    - – Line 5-8: Include in the literature review in the introduction rather than here.

- Page 21

– Line 12: "are available" can be removed

Interactive comment on Solid Earth Discuss., https://doi.org/10.5194/se-2019-164, 2019.

Printer-friendly version

Discussion paper