Solid Earth
Discussions

Open Access

EGU

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

Jérémie Giraud et al.

jeremie.giraud@uwa.edu.au

Received and published: 4 February 2020

Dear Reviewer,

Thank you for your comments and remarks about the manuscript. We have addressed your comments and followed most of your suggestions. We found them useful in improving the manuscript.

In what follows, we first provide a point by point answer to your general comments,

and then to the specific comments you made. Finally, we go through the technical corrections.

General comments

Title. We have changed the title of the manuscript accordingly with your comment.

Lexicon about previous modelling. Phrasing about 'inverse modelling results', 'geophysical inversion results' , 'inverse modelling results', 'inverse modelling results', that refer to the same thing. We have replace these terms by 'inversion results'.

Text structure. We have simplified the text structure following your suggestions. Sections "2.2.1. Geophysical inversion scheme" and "2.2.2 Geological uncertainty" were merged under "2.1 Geophysical and geological modelling". The sub-subsection 2.2.3 introducing post-regularisation was move up one level to become 2.3. and we added an introductory sentence directly below the title: "This subsection introduces the post-regularization scheme used in this work and details its implementation and usage in the workflow introduced here".

We moved 2.2.4 and 2.2.5 into the under "2.4 Uncertainty analysis", and adjusted the name of the subsection to make it more reflective of the contents. Having fewer sub-subsections also addresses your comment about the lack of introductory material between section and subsection headers.

We re-arrange the content introducing SOM as you suggested: the second paragraph of the corresponding section "In the approach we follow, the optimum number of neurons..." was moved to the end of the section so it follows the logical order of the process more closely.

Specific comments

Paragraphs in the introduction. We brought minor alterations to the text in several paragraphs in the introduction and reduced the number of "this" (see comments about usage of the word "this" below), which we think helps improving the readability of the

introduction. We have also added a paragraph at the end of the introduction that summarises the paper while introducing its general structure:

"The rest of this paper develops as follows. Section 2 provides the theoretical background necessary to reproduce the work presented. It first briefly describes the geophysical and geological modelling schemes (subsection 2.1) used to obtain the models that are used as input for classification using SOM (subsection 2.2). Post-regularization as applied to such classified lithologies (subsection 2.3) and the related uncertainty analysis in terms of prediction accuracy and geophysical consistency are then detailed (subsection 2.4). Following this, Section 3 presents an application case using data from the Yerrida Basin (Western Australia). Geological and geophysical modelling results are first summarized and the rules defining post-regularisation operator in the area are introduced (3.1). The classification of results from geological and geological modelling and post-regularization are then presented alongside the related uncertainty analysis, supporting a potential re-interpretation of the geological model of the area (subsection 3.2). The discussion and conclusion sections follow and complete this contribution."

Figure 1. We have moved Figure 1 to the methodology section as you suggested. It now appears in *subsection "2.3.2 Implementation"* after it is referred to in the text.

Concern/question about impact of uncertainty in density model: "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)?" We do not address the uncertainty in the density model directly. However, in the training set, the density model we used is the one that is recovered from inversion of synthetic data using the same parameterization of the geophysical inverse problem as for the inversion of field data only. So, the uncertainty in the density model is indirectly accounted for in the training phase and the assumption is made that field data inversion results are affected in the same way. We assume that non-uniqueness and measurement uncertainty are affecting field gravity data in the same way it does affect the synthetic case, which is does to a large extent since the

C3

assumed noise in the data are the same, are is the data acquisition setup.

We have added the following in the discussion:

"While we do not address the uncertainty in the density model directly, we assume that non-uniqueness and measurement uncertainty affect both field data and synthetic data in the same manner due to the noise component and parameterization of each being the same."

We have stated in the text that we deal with the inversion of gravity only, and gravity modelling is now mentioned at the end of the introduction.

'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 lithologies are mapped back into density by assigning the corresponding BMU's value in place of the lithology. In the case of post-regularisation (PR), the value assigned corresponds to the closest unit from the SOM that does not violate the geological rules that are enforced, hence the modification of the recovered density contras generated by PR.

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. We have replaced the two occurrences of the term 'integrated inversion' by 'geophysical inversion' to maintain generality.

We followed the suggestions you made regarding citations and references.

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'. The word SOM appears 30 times while string of words "self-organizing maps (SOM)" appeared only three times. The aim was to make it clear for the read that SOM = self-organizing maps. Following your comments, we kept only the first instance. Overall, the string of characters "self-organizing maps" appears a total of 15 times in the manuscript, 2 of which are in the main body of the text.

C4

Usage of words 'this' and 'that'. We have reduced the number of occurrences of the word 'this' from 76 to 63. It did not seem to us that the word "that" was overused so we have not tried to reduce the number of times it was used.

"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." To clarify things, we have added the words 'training' and 'predicted' wherever we felt that necessary to make it clearer to the reader.

Technical corrections

Page 1

– Line 25-30: Please expand on the 'inherent duality' of geology and geophysics.

We have modified this statement, and replaced it by: 'fundamental complementarities [. . .] in modelling the same object (the earth)'

– Line 30: What are the geophysical quantities modelled from petrophysicaland geological information?

We added the following information: "(seismic velocities, mass-density, etc.)".

– Line 35: citations should be in chronological order? Yes we put them in chronological order. *– Line 35: 'which consist in' should be 'which consist of'.* Done.

Page 2

Line 4: Both e.g. and etc. imply a subset, use one or the other.
Done.

– 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. We have removed 'on

the one hand', 'on the other hand' and replaced 'nevertheless' by 'however'.

– Line 15: Is it possible to provide examples of the 'broad range of parameters'?

We have rephrased the incriminated sentence to make it clearer: 'It is also evident that lithologies (or facieses) are characterised by a broad range of properties that are the result of complex, non-linear physical processes'

– Line 15: 'results' should be result.

Done.

– Line 26: 'informed' should be inform.

Done.

– Line 28: need 'to' or 'for' in between 'consideration geological'.

Done.

– 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?

We have rephrased the sentence as "To complement existing methodologies, we propose a solution that partially addresses the lack of consideration given to geological information during classification."

– Line 30: 'partially addresses the issues and shortcomings': please elaborate on the aspects still be tackled by you and the geoscience community.

To avoid a lengthy explanation we have removed this statement (see rewritten sentence above, first comment about page 1).

– 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.

We use the term 'semi-synthetic' because the synthetic dataset is derived from a geo-

logical framework built using field geological measurements.

To make things clearer we have added the following (underlined in the original sentence):

"Using an artificial neural network trained in a fully controlled environment (all variables in the model used for training being perfectly known) with attributes characterizing the geophysical inverse model*."*

Page 3

– Line 3: The phrase 'can serve multiple objectives' I think might be a bitmisleading as these are not necessarily objectives of the method; as writtenit 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.

Correct. We have modified the text to: 'The methodology we propose can serve two main objectives'.

– Line 12: The a or b reference for Kohonen (1982)?

Both actually. Done.

Page 4

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

Done.

Line 12: 'd represents measurements'; 'calculating the data'; the terminology for these should e 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.

C7

Done. We have replaced 'measurements' by 'data'.

– Line 15: How is Wm *specified?*

We have added the following information:

"here, W_m is the identity matrix".

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

We changed 'series' to 'ensemble'. Then we keep 'geological models is'.

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

Done.

 Page 5

– Line 1: What is meant by soft?

That was a typo, legacy from a previous version of the manuscript.

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

To make the explanation clearer we have enriched the previous paragraph with the
underlined information:

"This map, which can be 2D or 3D, is made of a predefined number of interconnected
neurons (also referred to as 'nodes' or 'units') that have a fixed network configuration.
Projection occurs during the training phase, where the locations of the neurons in the
manifold are iteratively adjusted so that they approximate it optimally. "

And to make the text more readable we have deleted the section headers separating
this paragraph from the next.

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

We have rephrased the explanation as: "Note that we apply the same principle as the well-known L-curve principle".

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

Using field-data only would lead to potential bias in the classification in that some input may be affected by, for example, lithological interpretation errors. Using the semi-synthetic dataset as we do here alleviate this issue. We have adapted the text as follows:

"We utilize this controlled environment to estimate the accuracy our predictions for each class identified in the studied volume <u>without the errors associated with well positioning or lithology interpretation errors</u>."

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

The following information was added: "The starting model is obtained from prior information. Here, it is the expected petrophysical property model from geological modelling."

– Line 35: By 'datum', do you mean 'variable'?

We brought the following modifications to clarify this (underlined). "Each datum <u>from the training and tests datasets</u> is a vector $x \, \epsilon \, \mathbb{R}^{n_v}$, with a number of variables $n_v = 5$, where $x\,(5)$"

Page 6

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

We have removed the words *'interpretable datasets'*.

– 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'

Done. *– 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 firsttime this phrase is used (in the introduction).

We do not use 'PR' anymore and just use the full term 'post-regularization'.

Page 7

– Equation 3: What is uk*?*

– Equation 4: What is l *in* nl*?*

We have clarified the equation.

Page 8

– Line 19-20: "with index B with 20% accuracy", what does this mean?

*We have rephrased the sentence as: "*For instance, in a two-lithology scenario, a given node may be found to predict lithology A using the validation dataset correctly 80% of the time (80% accuracy) and lithology B with correctly 20% of the time (20% accuracy)."

– Line 21: 2.2.4 Estimation should be Estimating

We changed the title for 'Prediction accuracy of the recovered lithologies'

Page 10

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

Rewritten: "The red dashed line outline the modelled area"

Page 12

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

Addressed.

Page 14

– Line 7: "It" should be "In"

Done.

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.

The section was removed (see answer to the other reviewer).