Dear Reviewer,

Thank you for your insights and comments. We have implemented most of the recommendations you made to the exception of a few where we thought that an alternative was possible and another one where we think that a clarification accompanied with a modification of the prose suffice. Seeing that some information that were not essential to the paper had the potential to confuse the reader, we decided to remove the incriminating bits. This removal alters neither the meaning nor the main message of the paper. Please find below our detailed answer to the different points you raised. To refer to a specific paragraph of comment, we reproduce the title you wrote in *italics*, followed by our answer.

Detailed answer:

Justifying the SOM: We have added what you suggested (page 6 line 15 on the non-revised manuscript). We added a sentence to the text of the manuscript: "In our case, the utilisation of SOM for partitioning the input models allows the recovery of lithology, which is a geological quantity reflective of all input data. It is also useful in that, as we will see later, the consistency of the recovered lithological model can be analysed from a geophysical point of view."

Lithology probabilities: We agree with the point raised here. We have adjusted the wording. We replaced 'lithology probability' with a more appropriate term. Instead of using 'lithology probability', which we calculated as the 'relative apparition frequency' for each unit of the SOM, we use 'prediction accuracy' of that same unit. In this case, the former and the latter are interchangeable as they are calculated in the same fashion (equation 7). We think that using prediction accuracy is clearer and more appropriate in the context of a paper relying on SOM.

For the sake of simplicity, and to maintain the accessibility of the manuscript to a broad audience, we plotted the box-plot of the prediction accuracy for the different lithologies as it complements the prediction accuracy for each lithology. We think that such statistics are a useful addition to Figure 10 that provide the reader with summary statistics and an estimate of the overall confidence one can place in the recovered lithologies volume without looking in the details. It also complements the results shown in the evolution of the prediction accuracy as a function of the number of elements in the SOM. Since this may be useful but is not essential to the manuscript, we added it to the appendices (Appendix B).

Probability uncertainties: The 'probability uncertainties' are not essential to the manuscript and removing it from the document does not alter the message significantly. Therefore, we have removed it from the paper. Future work in this direction would be an interesting way to infuse a little bit of real-world physics into simple machine learning algorithms like SOM; however, since this idea did not come from us we did not add it to the manuscript. Information about the positive recovery rate is provided for the reader by the boxplot mentioned above. In the revised version of the manuscript, the assessment of the prediction accuracy is shown in the appendix where the box plot of prediction accuracy is shown. We have added the following in section 3.2.3: "For completeness, assessment of the prediction accuracy of the different lithologies is shown by the corresponding boxplot in Appendix B".

Selecting the number of SOM units: We agree with your comments about the evolution of the prediction accuracy. However, in the case we present, the prediction accuracy increases until reaching approx. 750 units. Beyond this point, it stabilizes and oscillates around the maximum values. This is not obvious from the Figure. To lift this ambiguity, we have added the following statement to the legend of Figure 8: "Note that after 750 units, the quantization error for the different lithologies stabilizes and oscillates around its maximum values".

Geophysical consistency: We agree that additional information would help support the point we make about geophysical consistency and have added more information about the comparison between the updated model's response and the original inversion model's response. To compare how they differ we have added to the manuscript the misfit maps – before and after classification. We have added the corresponding maps in Appendix. Note that the data misfit maps are, overall, similar.

Figure 9: We have added the number of cells that were modified. The following was added to the manuscript: "a total of 2561 inclusions was identified" in the text accompanying the figure.

Additional modifications of the manuscript: to improve readability, we made a few minor modifications to the manuscript, dealing with wording.

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.

1. 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.

2. 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 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.

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.

3. 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 petrophysical and 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 geological 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 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.

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.

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

– Line 15: How is Wm *specified?*

We have added the following information:

"here, $\boldsymbol{W_m}$ 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 without the errors associated with well positioning or lithology interpretation errors."

– *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'? What is u(5)?*

We brough the following modifications to clarify this (underlined). "Each datum from the training and tests datasets is a vector $x \in \mathbb{R}^{n_v}$, 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 first time 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).

Towards geologically reasonable lithological classification from ~~integrated~~ geophysical ~~inverse modelling~~inversion: methodology and application case

Jérémie Giraud[1], Mark Lindsay[1], Mark Jessell[1], and Vitaliy Ogarko[2].

[1] Centre for Exploration Targeting (School of Earth Sciences), University of Western Australia, 35 Stirling Highway, 6009 Crawley, Australia.
[2] The International Centre for Radio Astronomy Research, The University of Western Australia, 7 Fairway, 6009 Crawley, Australia.

~~Correspondence~~Correspondance: Jérémie Giraud (jeremie.giraud@uwa.edu.au)

Abstract. We propose a methodology for the recovery of lithologies from geological and geophysical modelling results and apply it to field data. Our technique relies on classification using self-organizing maps (SOM) paired with geoscientific consistency checks and uncertainty analysis. In the procedure we develop, the SOM is trained using prior geological information in the form of geological uncertainty, the expected spatial distribution of petrophysical properties, and constrained geophysical inversion results. We ensure local geological plausibility in the lithological model recovered from classification by enforcing basic topological rules through a process called 'post-regularisation'. This prevents the three-dimensional recovered lithological model from violating elementary geological principles while maintaining geophysical consistency. Interpretation of the resulting lithologies is complemented by the estimation of the uncertainty associated to the different nodes of the trained SOM. The application case we investigate uses data and models from the Yerrida Basin (Western Australia). Our results generally corroborate previous models of the region but they also suggest that the structural setting in some areas need to be updated. In particular, our results suggest the thinning of one of the greenstone belts in the area may be related to a deep structure not sampled by surface geological measurements and which was absent in previous geological models.

1. Introduction

The idea of geoscientific integration is not new and has been advocated since the inception of quantitative geoscientific studies involving geophysics and during the advances of geophysics as a discipline (see for instance Wegener, 1915, Nettleton, 1949, Towles, 1952, Jupp and Vozoff, 1975, Lines et al., 1988, Li and Oldenburg, 2000,). In the natural resource sector, the exploitation of the ~~inherent duality~~fundamental complementarity between ~~of~~ geology and geophysics in modelling the same object (the earth) has been recognized as one of the pre-requisites to exploration success as early as the 1940's during the early years of the Society of Exploration Geophysicists (Eckhardt, 1940, Green, 1948). Numerous authors have since tackled the issue of integrating petrophysical and geological information to model geophysical quantities (seismic velocities, mass-density, etc.) through inversion, with an increasing trend in the past 15 years or so (see for instance references reviewed in ~~(~~Lelièvre and Farquharson 2016; Meju and Gallardo 2016; Moorkamp et al. 2016~~)~~, and ~~(~~Giraud et al., 2017~~)Lelièvre and Farquharson, 2016, Meju and Gallardo, 2016, Moorkamp et al., 2016, Giraud et al., 2017~~). In contrast, the recovery of geological quantities from geophysical ~~inverse modelling~~inversion has seen much less

effort. Recent studies have started to rectify this by proposing the idea of lithological differentiation of ~~inverse modelling~~inversion results (Paasche et al. 2010, Sun and Li 2015, Paasche and Tronicke, 2007), which consist ~~in~~ of the identification of lithologies from inversion results. While lithological differentiation is expected to hold much potential in mineral exploration it still remains underexplored (Li et al., 2019)~~,~~.

In oil and gas exploration scenarios, seismic facies analyses and classification using techniques developed for what is commonly called *machine learning* (~~e.g.,~~ neural networks, support vector machine algorithms, etc.) have become popular in recent years (Zhao et al., 2015, Chopra and Marfurt, 2018, Wrona et al., 2018, Zhang et al., 2018). This was driven, ~~on the one hand,~~ by the need for quantitative interpretation methods in the geosciences, and by the 'renaissance' phase machine learning went through after 2006 (Chap. 5, Goodfellow et al., 2016) ~~on the other hand~~. Once recovered, spatial facies distribution can be used for geological interpretation and downstream decision-making. ~~Nevertheless~~However, like all modelling results, the identification of facies or lithologies using machine learning relies upon statistical models and is affected by ambiguity and uncertainty. One reason for this is that validation datasets are usually treated as the ground 'truth' while they are fraught with uncertainty. For instance, the interpretation of borehole data or outcrops with their uncertainty can lead to significantly different models honouring geological measurements equally well (Wellmann et al., 2010, de la Varga et al., 2018, Pakyuz-Charrier et al., 2018a, b, c~~,~~).

L~~It is also evident that the modelling of l~~ithologies~~y~~ (or ~~facieses~~facies) can be ~~is~~are ~~sensitive~~ characterised ~~to~~ by a broad range of rock ~~parameters~~ properties ~~and that their characterization relies on properties~~ that are the result~~s~~ of geological processes, such as weathering, compaction, metamorphism and deformation. These~~complex, non-linear~~ physical processes are usually non-linear, especially when in combination, and produce complex representations of different lithotypes which are difficult to discriminate from geophysical data. In this context, one possible solution is to use neural networks for lithological classification as they are "universal approximators" (van der Baan and Jutten, 2000). As a consequence, however, lithological classification is affected by uncertainty from the data used to train and validate the algorithm. Such uncertainty is difficult to quantify and is rarely estimated or even considered.

To date, whether it be in oil and gas or mining exploration, uncertainty in recovered ~~rock types~~lithologies is a research avenue, which, to the best of our knowledge, only a few authors have addressed. Sun and Li (2019) assess uncertainty by varying the number of clusters in their lithological differentiation scheme, and Bauer et al. (2003) classify lithologies and estimate the resolution of their results using synthetic data. As a result of the lack of comprehensive uncertainty analyses, practitioners often lack quantitative, robust uncertainty modelling necessary to inform~~ed~~ interpretation or risk evaluation (Jessell et al., 2018). In addition, apart from ~~(~~Zhao et al., (2017) who account for seismic data-driven stratigraphy in their seismic facies classification, established workflows relying on neural networks to identify facies~~es~~ or ~~rock types~~lithologies in three-dimensions give little to no consideration of geological information and rules for their classification.

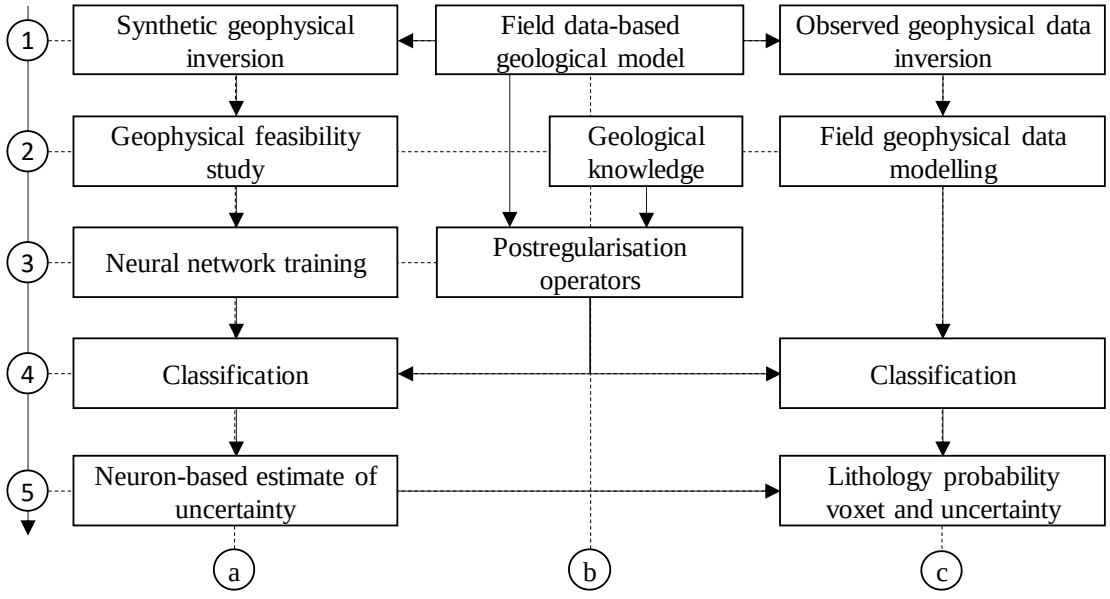
~~To mitigate this, and t~~To complement existing methodologies, we propose a solution that partially ~~addresses the issues and shortcomings highlighted above~~addresses the lack of consideration given to geological information during classification. We introduce a general post-processing (i.e., post-inversion) workflow for the recovery of lithologies from geoscientific modelling results and the estimation of the related uncertainty. For this purpose, we

complement existing classification techniques by ensuring the geological and geophysical consistency of, and estimating the confidence in, the recovered lithologies. 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~~inversion results, we perform lithological classification applying plausibility filters relying on geological principles which we refer to as 'geological post-regularisation'. The application of geological post-regularisation is to reduce the non-geological character of models obtained through classification. After classification, we calculate the ~~probability of occurrence~~ frequentist probability of the different lithologies (i.e., apparition frequency relative to all lithologies) associated with each unit of the SOM ~~and estimate the related uncertainty for~~and report it in each ~~each~~ model-cell ~~discretizing~~discretising the studied area for interpretation. ~~.~~

The methodology we propose can serve ~~multiple~~ two main objectives. Our first objective is to introduce a methodology that is made efficient by leveraging existing geoscientific inputs and prior information, and cost-effective by imposing requirements that do not exceed the computational power available on a personal computer. Secondly, our aim is to complement ~~inverse modelling~~inversion workflows by providing a general, automated method to derive a ~~probabilistic~~ non-deterministic lithological interpretation of ~~inverse modelling~~inversion results. Thirdly, we propose a real-world application based on a case study in the Yerrida Basin (Western Australia) where we build upon recent work by Giraud et al. (2019a) and Lindsay et al. (2018) who performed the geophysical inversion and geological modelling of data collected in the area, respectively.

~~In this method,~~ In this work, lithologies are identified though a classification technique relying on a simple artificial neural network. We chose 'self-organizing maps'~~-~~ Kohonen (1982a), (1982b)~~(Kohonen, 1982)~~, a well-established algorithm that has been successfully applied to seismic facies classification and geological mapping purposes (Chang et al., 2002, Klose, 2006, Köhler et al., 2010, Bauer et al., 2012, Carneiro et al., 2012, Du et al., 2015, Roden et al., 2015). We first train and test the self-organizing maps (SOM) using data extracted from a semi-synthetic dataset (i.e., a geophysical inversion feasibility study based on geological and petrophysical field data) assumed to represent the geophysical characteristics of the studied area. We utilize this controlled environment to estimate the accuracy ~~and uncertainty of~~ our predictions for each class identified in the studied volume without the errors associated with well positioning or lithology interpretation errors. We then use the trained network to perform classification using field data ~~inversion and modelling~~ only. We obtain, for each model cell, a suite of ~~probabilities equal to the relative frequency probability of occurrence~~frequentist probabilities ~~of~~ for each lithology observed in the area. In both cases, geological post-regularisation is applied before the calculation of uncertainty metrics to ensure the geological consistency of the results.

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, ~~The methodology is summarized in Figure 1 below.~~



Section 3 presents an application case using data from the Yerrida Basin (Western Australia), which was investigated using gravity data, petrophysical and geological information. Geological and geophysical modelling results are first summarized and the rules defining post-reguralisation operator in the area are introduced (subsection 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.~~ ~~Schematic summary of proposed methodology.~~

2. Methodology for geoscientific modelling and classification

In this section, we first introduce essential information about the geophysical and geological modelling used as pre-requisite to this study. We then introduce the utilisation of SOM and the tools we developed in sufficient detail to allow the reproducibility of the procedure.

2.1 ~~Integrated geophysical modelling~~Geophysical and geological modelling

~~2.1.1 Geophysical inversion scheme~~

Inverse geophysical modelling was performed using the least-square inversion platform Tomofast-x. This inversion platform enables the use of a series of constraints as detailed in Martin et al. (2018), Giraud et al. (2019a), (2019b). Constraints are enforced through a minimum-structure gradient regularisation approach ~~which~~ where weights vary~~ies~~ locally accordingly with geological uncertainty (Giraud et al., 2019a). The cost function to minimize is given as:

$$\boldsymbol{\theta}(\boldsymbol{d},\boldsymbol{m}) = \left\|\boldsymbol{W_d}\left(\boldsymbol{d}-\boldsymbol{g}(\boldsymbol{m})\right)\right\|_2^2 + \boldsymbol{\alpha_m}\left\|\boldsymbol{W_m}\left(\boldsymbol{m}-\boldsymbol{m_p}\right)\right\|_2^2 + \boldsymbol{\alpha_{sH}}\left\|\boldsymbol{W_{Hs}}\nabla\boldsymbol{m}\right\|_2^2, \quad (1)$$

where $\boldsymbol{d}$ represents ~~the measurements~~observed data ~~-~~and $\boldsymbol{m}$ is the model; $\boldsymbol{g}$ is the forward operator calculating the predicted data $\boldsymbol{m}$ produces; $\boldsymbol{m_p}$ is the prior model. In equation (1), subscripts d, m and s refer to data, model and smoothness, respectively. In this contribution, $\boldsymbol{W_d}$ is a diagonal matrix where each element is equal to the inverse of the sum-of-squares of the geophysical measurements; $\boldsymbol{W_m}$ and $\boldsymbol{W_{sH}}$ are diagonal covariance matrices; here, $\boldsymbol{W_m}$ is the identity matrix. The scalars $\boldsymbol{\alpha_m}$ and $\boldsymbol{\alpha_{sH}}$ are weights controlling the relative importance of the different terms in the equation; $\boldsymbol{\nabla}$ is the spatial gradient operator. The last term of the equation (1), the smoothness term, constraints the structural features of the inverted model. The values in diagonal matrix $\boldsymbol{W_{sH}}$ are determined from prior information. In ~~this work~~the presented work, $\boldsymbol{W_{sH}}$ is obtained from geological modelling results and is a proxy for geological uncertainty~~.~~

The matrix $\boldsymbol{W_{sH}}$ is calculated following Giraud et al. (2019a), who use the probabilistic geological modelling approach described in Pakyuz-Charrier et al. (2018c, 2018b, 2019). In the case of gravity inversion as presented here, the complete Bouguer anomaly of density contrast model $\boldsymbol{m}$ is calculated as the product of the Jacobian matrix $\boldsymbol{G}$ with model $\boldsymbol{m}$~~, i.e.,~~. Therefore, we have $\boldsymbol{g}(\boldsymbol{m}) = \boldsymbol{G}\boldsymbol{m}$.

~~2.1.2 Geological uncertainty~~

Geological uncertainty is estimated from probabilistic geological modelling. During this process~~, a series~~, and ensemble of geological models is generated using the Monte-Carlo Uncertainty Estimator (MCUE) of Pakyuz-Charrier et al. (2018a, 2018b, 2018c, 2019). MCUE relies on the perturbation of orientation measurements (interfaces and foliations) defining structures of a reference ~~-~~geological model accordingly with their uncertainty. From this series of models, geological uncertainty can be estimated (Wellmann and Regenauer-Lieb, 2012)through calculation of Shannon's entropy (Shannon, 1948) for the simulated geological models ~~(Wellmann and Regenauer-Lieb, 2012)~~. Shannon's entropy~~This metric~~, which can be used as a proxy for geological uncertainty, indicates how well geological information constraints the model locally. It can be used to constrain inversion in a structural sense when integrated in $\boldsymbol{W_{sH}}$ as per equation (1) (Giraud et al., 2019a).

More detailed information about the usage of MCUE results in geophysical inversion can be found in Giraud et al. (2017, 2018, 2019a, 2019b).

2.2 Classification using SOM~~M and uncertainty analysis~~

~~2.2.1 Fundamentals of self-organizing maps~~

The SOM artificial neural network ~~called 'self-organizing maps' (Kohonen, 1982a, 1982b) (SOM)~~ relies on ~~soft~~ competitive, non-supervised learning. The relative simplicity and the efficiency of the SOM algorithm has made it a popular tool for classification, data imputation, visualization and dimensionality reduction (Vesanto and Alhoniemi, 2000, Kalteh et al., 2008, Miljkovic, 2017, Klose, 2006, Kohonen, 1998, 2013, Roden et al., 2015, and Martin and Obermayer, 2009). In essence, it consists in the projection of the SOM's ~~(usually two-dimensional)~~ latent space onto a manifold of superior dimension (i.e., our dataset). 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~~ approximation is~~e it~~ optimal~~ly~~.

2.2.2 Training and validation process

2.2.2.1 SOM specifications and training procedure

In this study, we follow common practice by training two-dimensional (2D) maps using a hexagonal lattice topology and applying a Gaussian-shaped neighbourhood function. We chose to use a 2D map for the sake of simplicity after our testing revealed that other configurations did not improve results significantly. The hexagonal lattice topology seemed to provide better results than square lattice topology using the dataset we present here.

In the approach we follow, the optimum number of neurons (or units) is determined using the elbow curve of the mean quantization error Q (see equation 2 below in the next subsection) of the trained SOM. Note that this approach follows the same principle as the well known L curve principle (Hansen and O'Leary, 1993, Hansen and Johnston, 2001, Santos and Bassrei, 2007) for the determination of optimum weights in least square geophysical inversion. Here, we train the SOM using functions from the SOM Matlab toolbox implemented by (Vatanen et al., 2015).

2.2.2.2 Training and testing datasets

Ideally, the SOM should be trained in a controlled environment where all the variables used are perfectly known. This, which motivates the utilisation of synthetic geophysical data. In this work, wWe calculate such data from a geological structural framework derived from real-world field geological and petrophysical field measurements data in the same fashion as for a geophysical feasibility studyy (Figure 1, lines 1 and 2). The training and tests datasets are comprised of the following variables:

1. Starting model for inversion m_{start};
2. Inverted model m_{inv};
3. Geological uncertainty W_{sH};
4. Spatial gradient in the inverted model $\|\nabla m_{inv}\|_2$;
5. Most likely lithology obtained from geological modelling (training lithological model).

The starting model is obtained from prior information. Here, it is the expected petrophysical property model from geological modelling. Each datum from the training and tests datasets is a vector $x \epsilon \mathbb{R}^{n_v}$, with a number of variables $n_v = 5$, where $ux(5)$ is the lithology assigned to this unit. This choice of training variables was motivated by the necessity to account for the available information in terms of geological modelling and measurements, uncertainty and structural setting, while using interpretable datasets. The starting model for inversion encapsulates the pre-inversion state of knowledge. The inverted model comprises the update of this model using information extracted from geophysical measurements and translated into a 3D model. The lithological model refers directly to the interpreted geological observations of the area. The spatial gradients of the inverted models provide structural information about the location of the geological units that can be recovered by interpretation from the inverted density contrast model.

During training, we examine SOM quality using quantization error Q and lithology prediction accuracy. Quantization error "measures the average distance between each data vector and its best matching unit [BMU]" (Uriarte and Martín, 2005), thereby indicating how well the different BMUs approximate the dataset. It can be

interpreted as analogous to a misfit between calculated and observed data. The mean quantization error Q of the SOM is expressed as follows for n data vectors x_i:

$$Q(\boldsymbol{x}, \boldsymbol{BMU}) = \frac{1}{n}\sum_{i=1}^{n} \|BMU_i - x_i\|_2^2, \quad (2)$$

where BMU_i is the BMU of x_i. In the application of SOM to field data, the trained map is used for the classification of inversion results where input 1 through 4 listed above are obtained from previous modelling and lithologies are the quantity sought for. In our case, the utilisation of SOM for partitioning the input models allows the recovery of lithology, which is a geological quantity reflective of all input data. It is also useful in that, as we will see later, the consistency of the recovered lithological model can be analysed from a geophysical point of view.

In the approach we follow, the optimum number of neurons (or units) is determined using the elbow curve of the mean quantization error Q (equation 2) of the trained SOM. Note that we apply the same principle as the well-known L-curve principle (Hansen and O'Leary, 1993, Hansen and Johnston, 2001, Santos and Bassrei, 2007) for the determination of optimum weights in least-square geophysical inversion. Here, we train the SOM using functions from the SOM Matlab toolbox implemented by Vatanen et al. (2015).

2.3 Geological post-regularisation

~~2.2.3~~ This subsection introduces the post-regularization scheme used in this work and details its implementation and usage in the workflow introduced here.

~~2.2.3.1~~2.3.1 Motivations

Geological rules have the potential to provide an important constraint on the classification of lithologies recovered from inversion. Such rules, like adjacency (Egenhofer and Herring 1990) define which rock bodies can be in contact with each other, and which ~~are~~ cannot. These rules are typically expressed in geological terms as stratigraphy, where the relative age and event classification of geological units are stated. For example, a sedimentary depositional event of five separate units may define a simple sub-horizontal layer cake configuration where the oldest unit is never adjacent (or in contact) with the youngest unit. A magmatic event that follows may result in a vertical dyke that intrudes all sedimentary layers adjacent to all other rock units. Using geological rules as a constraint relies on finding those that are restrictive (such as the youngest unit never being in contact with the oldest), rather than permissive (such as the intruding dyke). Thiele et al. (2016), Pellerin et al. (2017), Anquez et al. (2019) show how these can constrain parametric geological modelling. It is therefore important to honour geological rules if known, and include them in classification schemes such as those to ensure that geological plausibility is not compromised in pursuit of an otherwise petrophysically and geophysically consistent model.

The process of postregul~~l~~ar~~i~~isation, which consists in the application of spatial-contextual filters to the classification results to eliminate geologically unrealistic features, has been shown to increase prediction accuracy in surface (2D) geological mapping (Tarabalka et al., 2009, Stavrakoudis et al., 2014, Cracknell and Reading, 2015).

~~2.2.3.2~~2.3.2 Implementation

The post-regularisation ~~(PR)~~ scheme we develop for the recovery of lithologies in 3D relies on two hypotheses. Firstly, we assume that the presence of isolated lithologies contradicts the geological principle of continuity~~.~~. Although such post-regularisation has been used mostly in 2D or shallow 3D, there is no theoretical obstacle to the extension of this methodology to the purely 3D classification case we present here. Secondly, we introduce the utilisation of adjacency relationships between the different lithologies in post-regularisation to ensure that base topological rules are respected across the entirety of the three-dimensional volume. This is particularly important for structural geological interpretation (Freeman et al., 2010, Godefroy et al., 2019). ~~In this paper~~Here, we extend existing post-regularisation approaches (i.e.~~,~~, Tarabalka et al., 2009, Stavrakoudis et al., 2014, Cracknell and Reading, 2015) by integrating geological information in the classification analysis in the form of topological relationships (see Egenhofer and Herring 1990, Zlatanova, 2000, and Thiele et al., 2016, for the different topologies) defined by geological principles.

The general formulation of post-regularisation is as follows, for a given model-cell:

$$BMU \coloneqq BMU_k, \text{where } k = \arg\min(\|u_k - BMU\| \mid conditions), \tag{3}$$

where arg min returns the argument k satisfying the conditions it precedes. Here, we set:

$$conditions = \{\exists\, c \subset U \mid c = BMU_k(5) \cap \mathbf{1}^T M^k \circ M^f \mathbf{1} = \mathbf{0}_{n_l,n_l} \cap stratigraphy\}, \tag{4}$$

where $\mathbf{1}$ is $\boldsymbol{n_l} \times 1$ column vector of ones, and $\mathbf{0}_{n_l,n_l}$ is the $\boldsymbol{n_l} \times \boldsymbol{n_l}$ matrix of zeros; M^f and M^k are adjacency matrices, and $M^k \circ M^f$ is the Hadamard product of M^k and M^f, such that $(M^k \circ M^f)_{ij} = (M^k)_{ij}(M^f)_{ij}$, and $\mathbf{0}_{n_l,n_l}$ is the zero matrix of dimension $n_l \times n_l$; M^f encapsulates geological knowledge and principles about contacts between lithologies; M^k contains the adjacency relationships between the considered cell and its neighbourhood U. The derivation of M^f and M^k is detailed below.

The first part of the condition in equation (4) is enforced using morphological closing where isolated lithologies are replaced by the most prevalent one in their neighbourhood (see Benavent et al., 2012, Ackora-Prah et al., 2015). In such cases, the BMU is updated as follows. Isolated cells (in terms of their lithology) are identified through examination of the 26-cell 3D cubic Moore neighbourhood U of every model-cell. A cell is considered isolated if, and only if, at least 25 cells of U have a lithology that differs from it. Once such a cell is identified, its BMU is updated using the closest neuron ensuring continuity between adjacent cells in th~~is~~e neighbourhood U, where the lithology to be assigned is determined by a majority vote in U subject to adjacency conditions. These conditions are determined by geological knowledge (as explained below). This process is repeated for all locations

until the lithological model stops changing. The general principle of post-regularisation is illustrated in

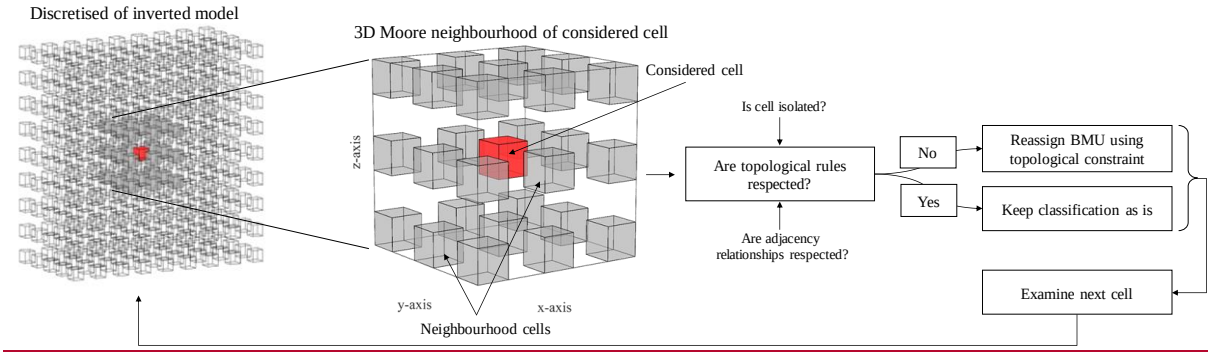


Figure 1~~Figure 2~~.

We point out that in contrast to Tarabalka et al. (2009) and Stavrakoudis et al. (2014) who use the first and second Chamfer neighbourhoods in 2D around the considered model cell, we do not follow the same approach in 3D. Our implementation of the extension of their approach to 3D showed that, in our application case study, the adjustments of the recovered lithological model it imposes are detrimental to the consistency of the classification with geophysical measurements. That is, the perturbation of the corresponding geophysical response of the model it generates exceeds noise level and compromises the geophysical validity of the recovered model (see geophysical validation subsection below for more details). The same remark applies to the utilisation of a mode filter with a 3×3×3 kernel.

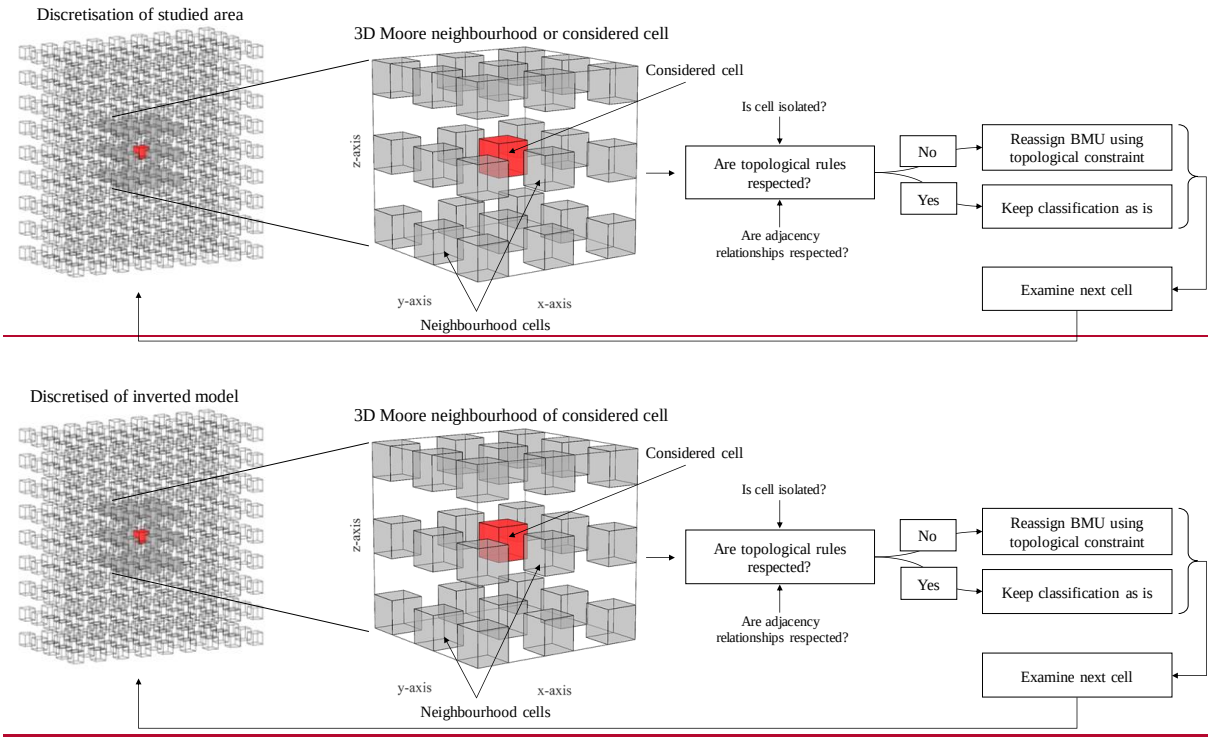


Figure 1~~2~~. Summary of topological filtering used during post-regularisation.

The conditions relating to adjacency relationships forces the model to honour adjacency relationships extracted from surface geology (Burns, 1988, Thiele et al., 2016) in the recovered lithological model.

We determine lithological topology by identifying the contacts between adjacent model-cells and represent the topological signature of lithological models using the adjacency representation of Godsil and Royle (2001). Let the adjacency matrix M of a given cell be defined as:

$$M_{i>j,j} = n_{ij}, M_{i\leq j,j} = 0, \quad (5)$$

where n_{ij} is the number of contacts between lithologies i and j. Similarly, geological laws and knowledge allow the derivation of a matrix M^f defined as follows:

$$\begin{cases} M^f_{i>j,j} = 1 \; if \; contact \; between \; i \; and \; j \; contradicts \; geology. \\ M^f_{i>j,j} = 0 \; otherwise \end{cases} \quad (6)$$

From there, it is straightforward to identify occurrences of forbidden contacts by calculating $M_k \circ M^f$. Therefore, $\mathbf{1}^T M_k \circ M^f \mathbf{1} = 0$ (equation 4) indicates that no contact violating the condition imposed by M^f is observed. The last condition in equation (4) can be used to prevent the local stratigraphy (in the Moore neighbourhood U of the considered cell) from violating geological rules such as "lithology B must be in conformable sequence between lithologies A and C".

~~from violating geological rules such as, for instance, "lithology B must always be between lithologies A and C in a conformable sequence".~~

After identification of configurations forbidden by the conditions set in equation (4), its BMU is updated using the closest neuron honouring the set of conditions (~~Figure 1~~ below, box 4-a).

The next stage of the methodology we introduce is the calculation of the apportionment of each neuron in terms of the lithologies of the testing data vectors (from the synthetic survey) they predict (~~Figure 1~~, box 5-a). For instance, in a two-lithology scenario, a given node may be found to predict lithology A using the validation dataset ~~with~~ correctly 80% of the time (80% accuracy) ~~with~~ and lithology~~index~~ B with correctly 20% of the time (20% accuracy). This process is described below.

The methodology is summarized in below.

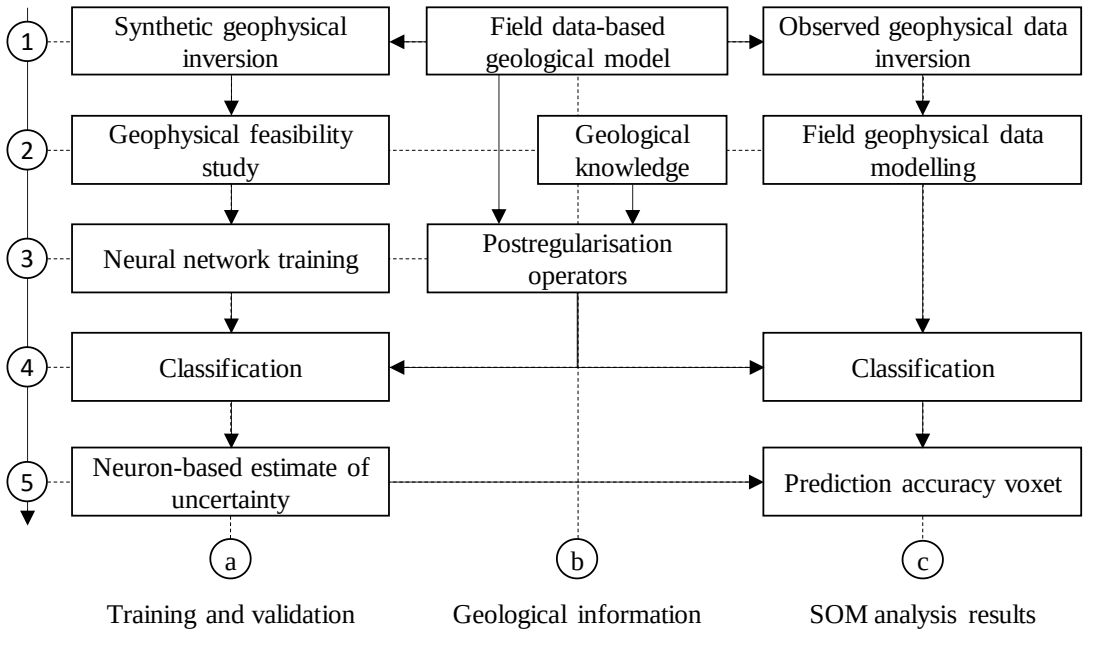


Figure 2. Schematic summary of proposed methodology.

2.4 Uncertainty analysis

~~2.2.4~~2.4.1 ~~Estimation the confidence in~~Prediction accuracy of the recovered lithologies

The prediction accuracy τ_i of lithology $i \in [1, \dots, n_l]$ (with n_l the total number of lithologies) is the ratio of correct predictions to the total number of predictions. It is obtained from the 'matching matrix' of the recovered lithologies M^c. We remind that the 'matching matrix' (or 'confusion matrix' in supervised learning) is a matricial representation of the number of occurrences of true/false positives/negatives.

We use the prediction accuracy τ_i as a metric measuring the capability of the node~~s~~ i of the trained SOM to recover lithologies. Let τ_i be expressed as:

$$\tau_i = \frac{M^c_{ii}}{M^c_{ii} + \sum_{\substack{j=1 \\ j \neq i}}^{n_l} M^c_{ij}}, \tag{7}$$

which is a particular case of the overall accuracy τ:

$$\tau = \frac{\sum_{i=1}^{n_l} M^c_{ii}}{\sum_{i=1}^{n_l} \sum_{j=1}^{n_l} M^c_{ij}}, \tag{8}$$

From equation (7), it appears that τ_i is equivalent to the frequentist probability ~~relative apparition frequency~~ of the i^{th} lithology over the entire SOM. When considering a specific node j, it becomes the relative frequentist probability ~~apparition frequency~~ of the lithology i for cells classified as having the j^{th} node as their BMU, noted τ_{ij}. ~~These percentages, which we address as probabilities, constitute apparition frequencies. Moreover, the nodes of the trained SOM each approximate a subset of the dataset that vary in size. This issue is addressed below in the next subsection.~~

~~We complement the estimation of the probability of recovering the modelled lithologies by calculating uncertainty on these probabilities (Figure 1, box 5c). We calculate the Wald type confidence interval (Wilson, 1927, DasGupta et al., 2001) of the recovered apparition frequencies, which we address here as probabilities. Details about the calculation of the confidence interval are given in Appendix A.~~

~~2.2.5~~2.4.2 Geophysical consistency

The consistency of the classification performed using SOM after application of post-regularisation with field geophysical measurements might be altered by both the classification itself and by post-regularisation. It is therefore necessary to ensure that the approximation of the dataset by values from the units of SOM is consistent with geophysical measurements. To this end, we verify that the geophysical response of the density contrast model $\boldsymbol{m_{SOM}}$ corresponding to the BMUs of each cell in the studied area fits the field measurement $\boldsymbol{d}$ within a certain tolerance assumed to approximate noise level. Consequently, we ensure that the difference between $\boldsymbol{m_{inv}}$ and $\boldsymbol{m_{SOM}}$, $\Delta \boldsymbol{m} = \boldsymbol{m_{inv}} - \boldsymbol{m_{SOM}}$, satisfies the following condition:

$$\left\| \boldsymbol{W_d} \left(\left(\boldsymbol{d} - \boldsymbol{g}(\boldsymbol{m_{inv}}) \right) - \left(\boldsymbol{d} - \boldsymbol{g}(\boldsymbol{m_{SOM}}) \right) \right) \right\|_2^2 = \| \boldsymbol{W_d G} \Delta \boldsymbol{m} \|_2^2 \leq tol, \tag{9}$$

where $\boldsymbol{tol}$ is the threshold depending on noise levels in the data above which $\boldsymbol{m_{inv}}$ and $\boldsymbol{m_{SOM}}$ are not considered geophysically equivalent.

The implication of equation (9) is that the difference $\boldsymbol{\Delta m}$ belongs to the null-space of the inverse problem considered. The null-space is characteristic of geophysical inversion's non-uniqueness. It is defined as the ensemble of models that reproduce geophysical data with a comparable misfit. The models honouring equation (9) can therefore be considered equivalent from a geophysical data point of view (Muñoz and Rath, 2006, Chen et al., 2007, Deal and Nolet, 1996). For qualitative assessment of geophysical consistency, we complement the utilisation of equation (9) with the visual comparison of data misfit maps corresponding to the model corresponding to $\boldsymbol{m_{inv}}$ and $\boldsymbol{m_{SOM}}$.

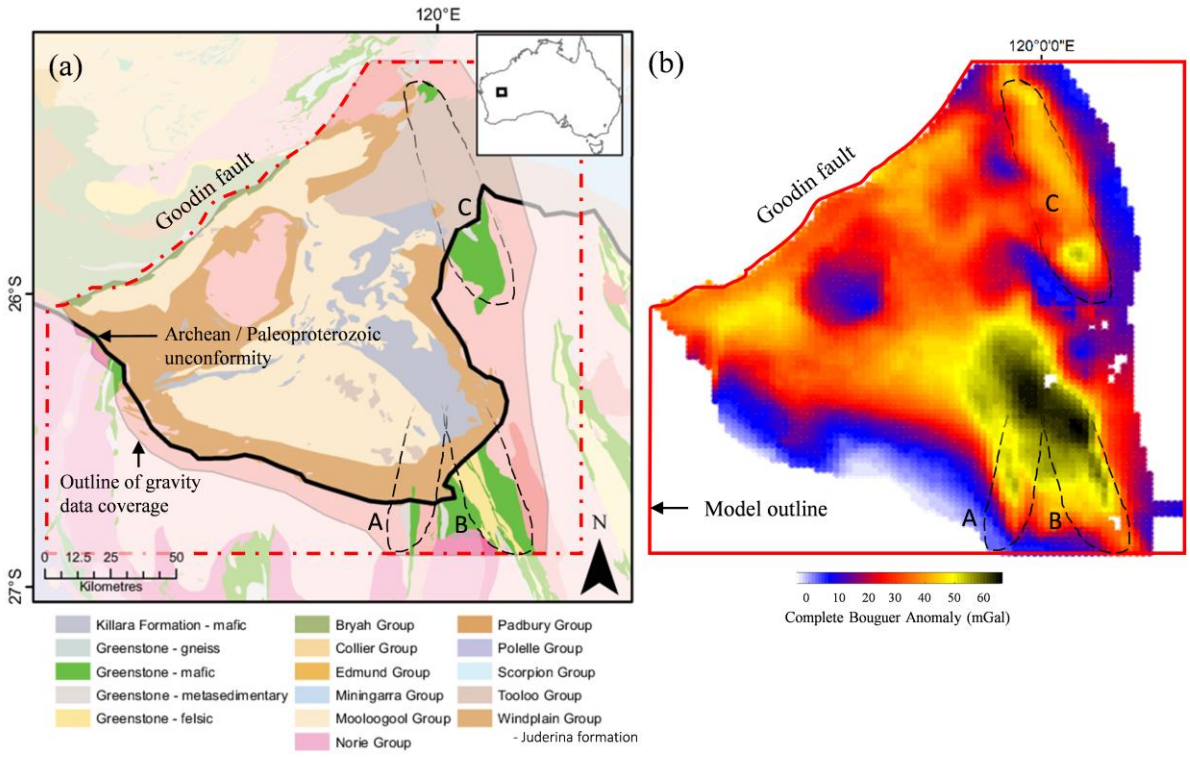
3. Application case: Yerrida basin

3.1 Survey setting

This subsection introduces and summarises the geological and geophysical context of ~~this~~ the application case presented here~~study~~. More details about the geology of the area and the initial geophysical inversion~~field data geophysical modelling~~ can be found in Giraud et al. (2019a) and Lindsay et al. (2018), (2020).

3.1.1 Geological and geophysical setting

The Paleoproterozoic Yerrida Basin is located in the southern part of the Capricorn Orogen (WA), and covers approximately 150km N-S and 180 km E-W (Pirajno and Adamides (2000)) (Figure 3~~Figure 3~~a). The structures of interest in this work are Archean greenstone belts (Figure 3) as they are prospective for Au and Ni and underlie the younger basin rocks. Basement to the Yerrida Basin is considered to be Archean granite-gneiss or greenstone rocks of the Yilgarn Craton. Lithospheric extension initiated the formation of the Yerrida Basin at c. 2200 Ma to



1990 Ma with deposition of the Windplain Group ~~(~~(Occhipinti et al., 2017~~),;~~ ~~(~~Pirajno and Adamides, 2000~~)~~). The Goodin Inlier remains exposed in the central part of the basin and is in unconformable contact with the Windplain Group. A hiatus ensued, followed by deposition of the younger Mooloogool Group, which was then overlain in the east by the Tooloo Group of the Earaheedy Basin.

Figure 3. (a, to the left) Geological map of the area ~~(modified from Lindsay et al., 2018)~~ and (b, to the right) complete Bouguer anomaly (reproduced from Giraud et al., 2019a)~~(modified from Giraud et al., 2018b)~~. The **red** dashed line~~s delineate the possible sub-basin extent of the mafic greenstone belts~~**outline the modelled area**. Capital letters 'A', 'B' and 'C' symbolize the possible outlines for the greenstone belts in the area.

The density contrast of the ~~rock types~~lithologies observed in the area range between 0 and 330 kg/m³, making it ~~suitable~~ appropriate for gravity modelling and inversion (Giraud et al., 2019a, Lindsay et al., 2018~~).~~). While basin rocks exhibit some density contrast, the greenstones are conspicuous in gravity data with a density contrast expected to lie between 190 and 270 kg/m³ making them an attractive subject for gravity inversion. Field geological measurements (orientation data in the form of interfaces and foliations) and petrophysical data were used to build the reference geological model. Airborne geophysical data, Landsat and Aster 8 satellite data were also used to support the interpretation of geological measurements.

The gravity anomaly dataset we consider (Figure 3~~Figure 3~~b) is comprised of a total of 4882 measurement points. The model is discretized into 100×100×42 cells of dimensions 2.335 km ×1.875 km ×1.0475 km down to approximately 44km depth. Weights and parameters used for the inversion of synthetic data follow the ~~work~~ settings of Giraud et al. (2019a) on field data.

3.1.2 Geological modelling and synthetic geophysical survey

This subsections introduces the semi-synthetic survey we performed for the training of SOM.

The volume of most probable lithology, W_{HS} and the starting model are shown in Figure 4~~Figure 4~~. Volumes shown in Figure 4~~Figure 4~~a, b, and c are used for the training and validation dataset for SOM training as explained in 2.2~~2.2.2.2~~.

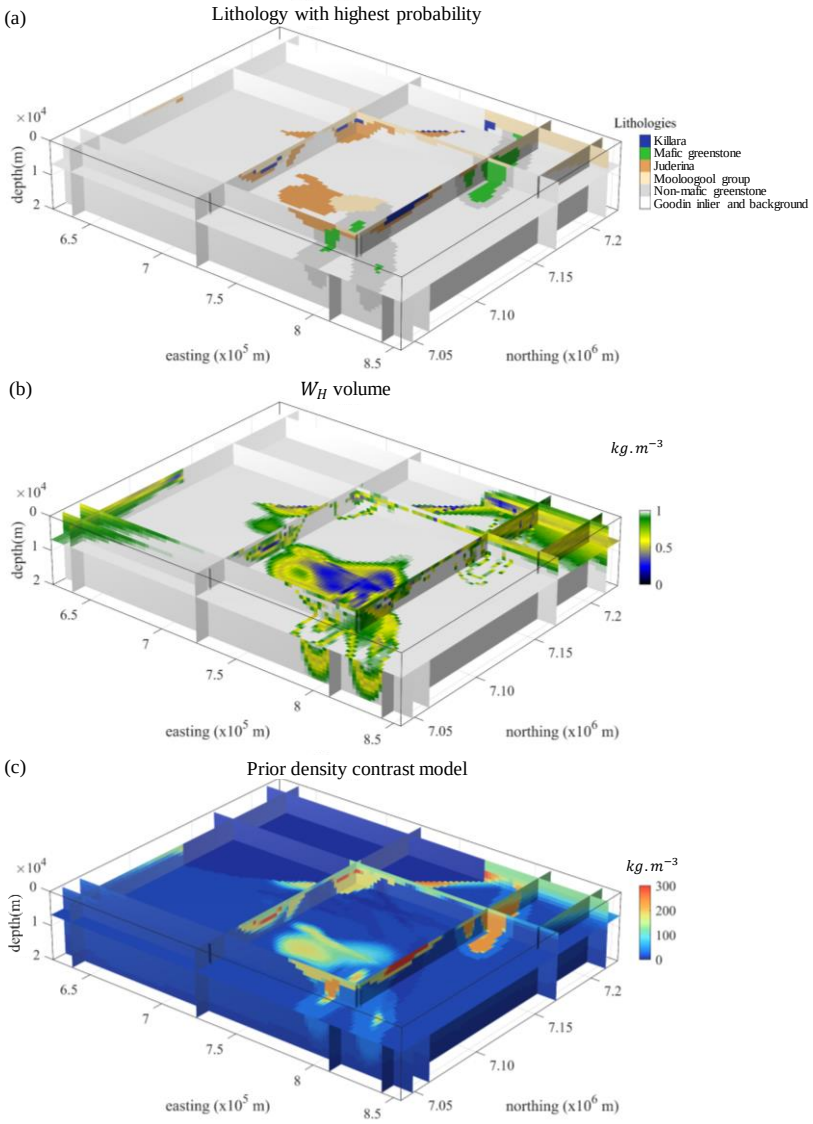


Figure 4. Prior geological modelling. Preferred lithology volume (a), geological uncertainty volume (b) and starting density contrast model (c). Modified from Giraud et al. (2019a)**.**

We use the modelling results shown in Figure 4~~Figure 4~~ to calculate a synthetic geophysical dataset (Figure 5~~Figure 5~~d). The model used to generate the synthetic geophysical ~~measurements, the inverted model and its spatial gradients are~~measurements is shown in Figure 5~~Figure 5~~a. The corresponding inverted model and its gradients are shown in Figure 5~~Figure 5~~b and Figure 5~~Figure 5~~c. Volumes shown in Figure 5~~Figure 5~~b, and c are used for training and validation ~~dataset for SOM training as detailed in 2.2.2.2~~.

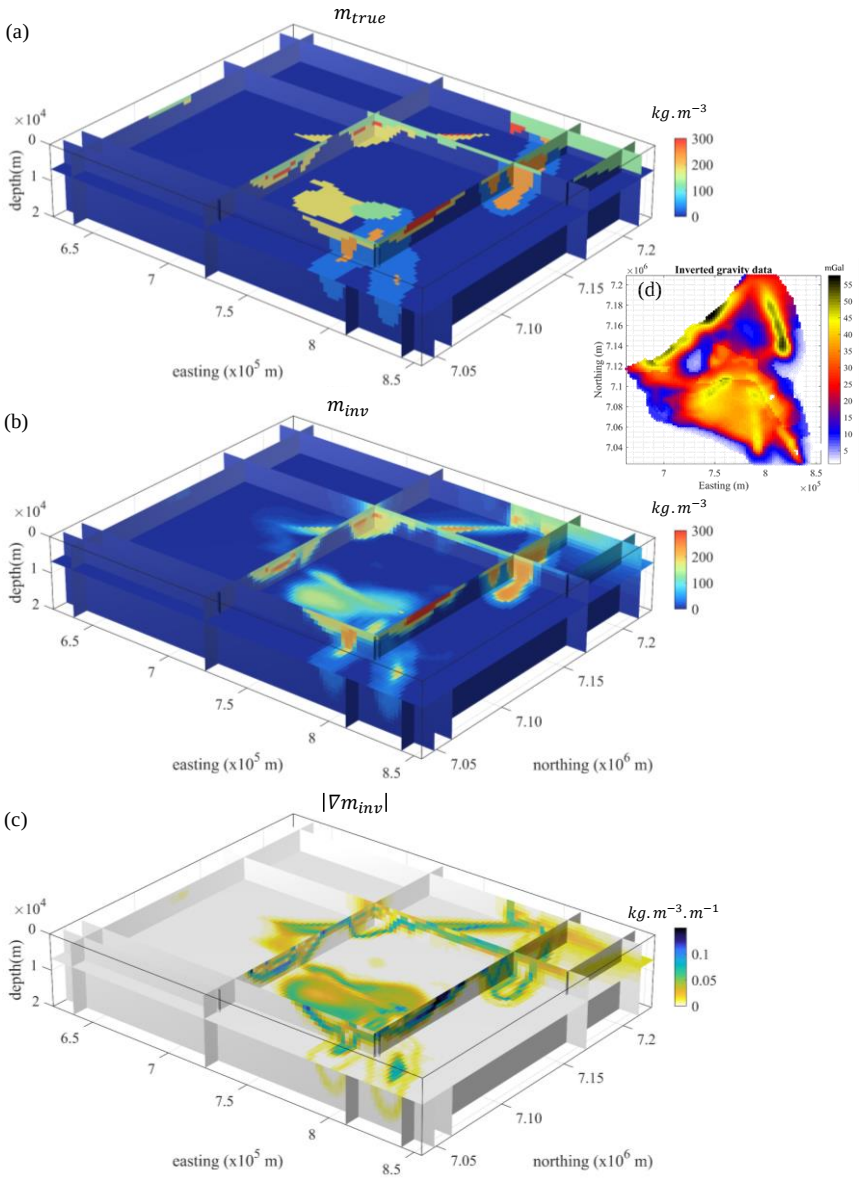


Figure 5. True density contrast model for calculation of synthetic geophysical data (a); inverted model obtained from geophysical inversion ~~integrated inversion~~ of synthetic geophysical data (b), corresponding spatial gradient of density contrast (c) and synthetic geophysical data (d).

3.1.3 Field geophysical data inversion

The density contrast model obtained from inversion of field geophysical data and its gradients are shown in Figure 6~~Figure 6~~a and b. Visual comparison of inverted models shown in Figure 6~~Figure 6~~a and Figure 5~~Figure 5~~b reveals that mesoscale structures are similar with the exception of large structures presenting low density contrasts at depth (darker shades of blue in Figure 6~~Figure 6~~a). This is reflected in Figure 6~~Figure 6~~b, which exhibits low gradient values in these areas.

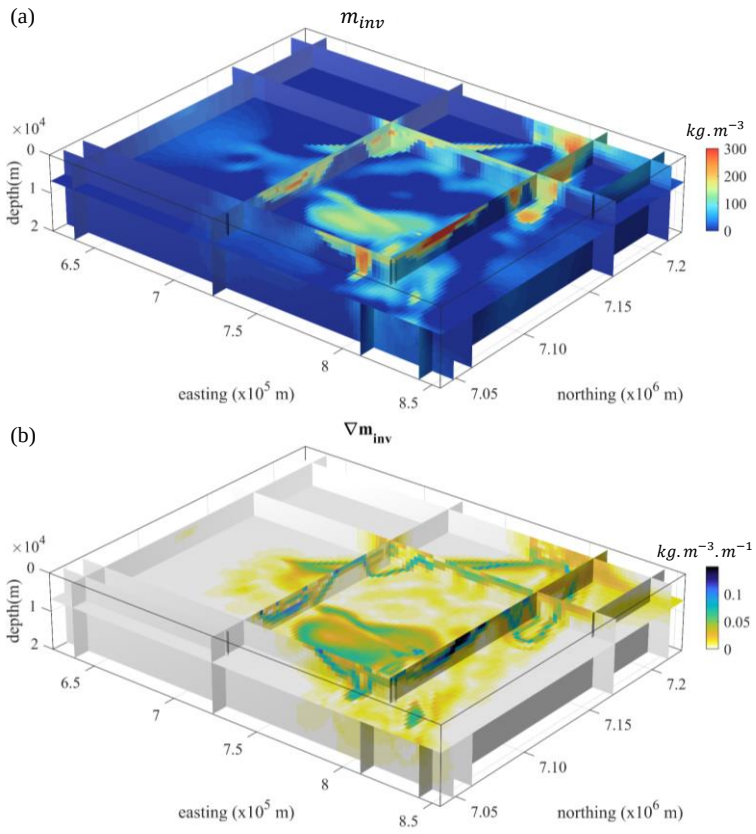


Figure 6. Inverted model obtained from geophysical inversion ~~integrated inversion~~ of field geophysical data (a), corresponding spatial gradient of density contrast (b).

The classification of lithologies using SOM is performed applying the trained network to volumes shown in Figure 4~~Figure 4~~b, c and Figure 6~~Figure 6~~a, b. The next subsections describe the geological laws used for post-regularisation (subsection 3.1.4~~3.1.3~~) and the classification process (subsection 3.2.2).

3.1.4 Geological rules for post-regularisation

As mentioned above, for clarity in this demonstration, only adjacency relationships are considered. ~~This~~ The utilisation of simple relationships such as adjacency is also chosen ~~in part~~ because in areas of sparse data, a full description of geological rules (fault relationships with fault and stratigraphy) is often not known. Given the complexity of the Yerrida Basin and its magmatic and deformation history, several base geological rules can be derived to assess the plausibility of recovered lithological models. Using fundamental geological principles (such as uniformitarianism, superposition, Walther's Law, cross-cutting relationships and original horizontality), the two most likely restrictive adjacency rules are as follows. We assume that the mafic greenstone bodies cannot be in contact with the Killara Formation (in the Mooloogool Group) since our field data suggests that the Killara Formation is a volcanic unit that is restricted to the Yerrida Basin and thus not in contact with the mafic greenstones. In addition, we assume that the mafic greenstones cannot be in contact with the Goodin Inlier and background (or basement) as the mafic greenstones are modelled to be enveloped by the felsic component of the greenstone.

The matrix M^f defining the contacts forbidden by geology as described above is given in Figure 7~~Figure 7~~.

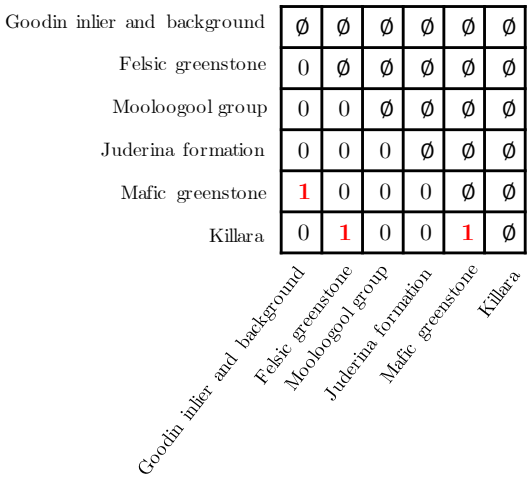


Figure 7. Matrix defining forbidding contact between lithologies in the Yerrida basin. 1 means that two units may be in contact with each other, 0 means that they may not and ϕ represents symmetric relationships or when the same unit is adjacent to itself (which geologically may occur across a fault, but cannot be resolved by the geophysical data available).

3.2 Geologically-constrained SOM classification and uncertainty analysis

3.2.1 Training the neural network

~~It~~ In this work, we use approximately 500 neurons (units) for the training of SOM. This number is inferred from the analysis of the elbow curve we calculated using the validation datasets (see Figure 8~~Figure 8~~) and approximately matches the proposed value of $5\sqrt{N}$ (with N the total number of observation) proposed by Vesanto and Alhoniemi (2000) and commonly used since (Shalaginov and Franke, 2015). The ~~use of this value~~chosen number of units is corroborated by the lithology prediction accuracy (Figure 8~~Figure 9~~) as approximating the point of diminishing returns, i.e., the number of nodes beyond which additional nodes are becoming nearly redundant.

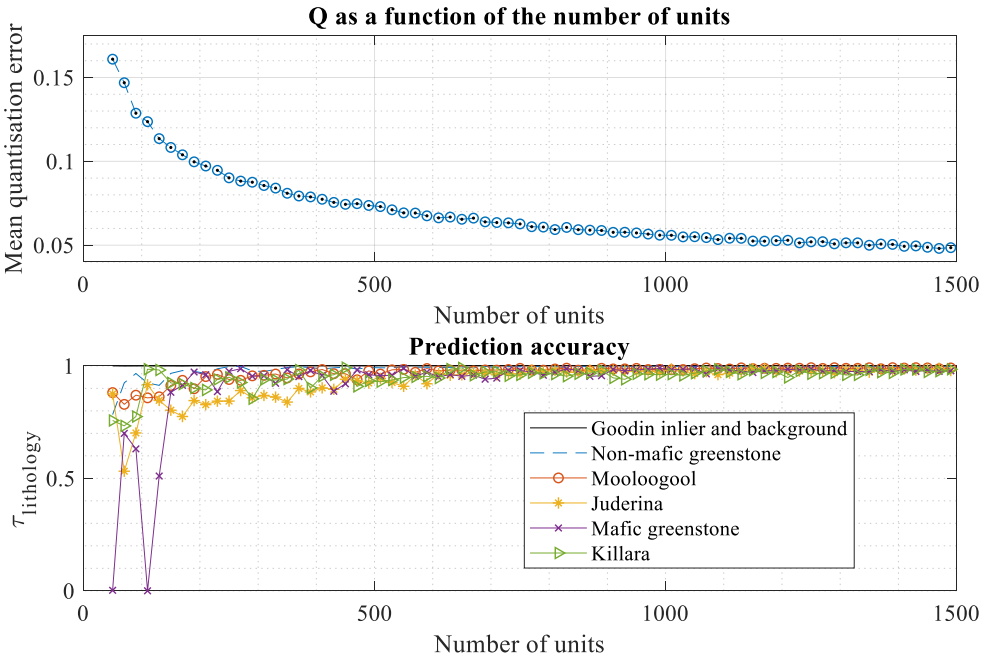
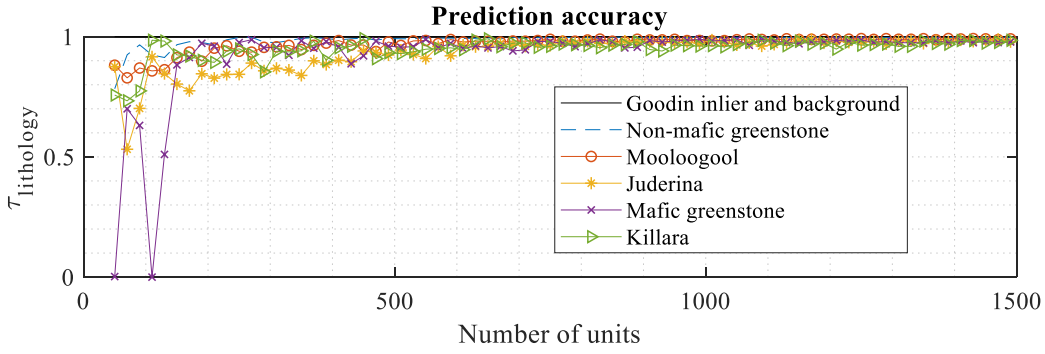


Figure 8. (a) Elbow curve of the quantization error for the determination of the optimum number of neurons (or units) in SOM and (b) Prediction accuracy for the different lithologies present in the training and validation datasets. ~~,~~Note that after 750 units, the quantization error for the different lithologies stabilizes and oscillates around its maximum values.



~~**Figure 9.** Prediction accuracy for the different lithologies present in the training and validation datasets.~~

The trained map presents a mean quantization Q equal to 0.075, which indicates relatively good approximation of the datasets by the trained SOM. This is illustrated by Figure 8~~Figure 9~~ where all lithologies are recovered with a prediction accuracy superior to 90%.

3.2.2 Classification and post-regularisation

After classification of the recovered model, we performed post-regularisation to remove geologically unrealistic features from the classified lithological volume. Figure 9~~Figure 10~~ show the classification results before and after post-regularisation, along with the associated adjacency matrix.

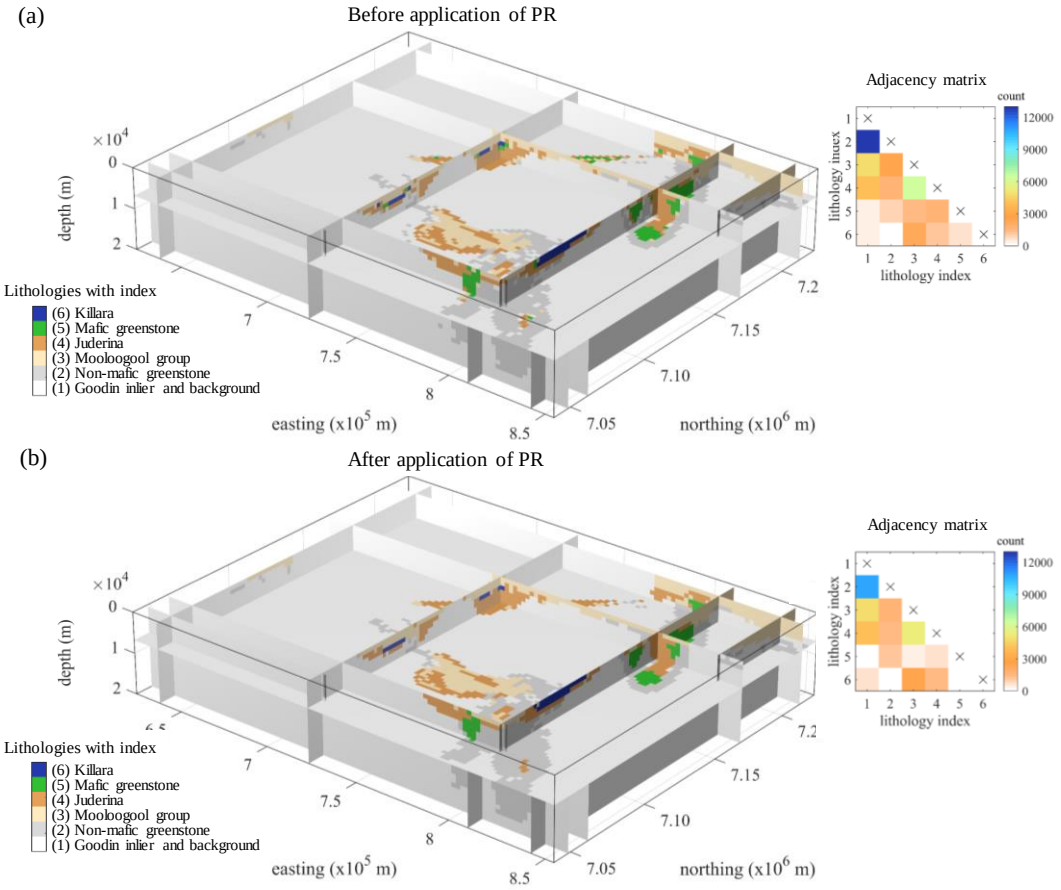


Figure 9~~10~~. Recovered lithologies before post-regularisation (a) and after (b) next to the corresponding adjacency matrix (right hand side). The indices used to designate lithologies are indicated in the legend of the volume (left hand side).

As can be inferred from the adjacency matrices plotted in Figure 9~~Figure 10~~, the number of contacts between units with indices 1 and 2, respectively, is reduced by the application of post-regularisation. One reason for this is the presence of a number of inclusions of lithologies 1 in 2 and vice versa; a total of 2561 such inclusions was identified.~~a.~~ Overall, ~~T~~the number of contacts between lithologies 5 and 2 increased slightly due to post-regularisation because a large percentage of contacts between lithologies 5 and 1 have been re-assigned as contacts between 5 and 2. The elimination of contacts between 5 and 6 is also visible in Figure 9~~Figure 10~~.

3.2.3 Estimated confidence in recovered lithologies

For each node of the SOM we calculate the prediction accuracy $\boldsymbol{\tau_i}$ (equation 3) for the different lithologies observed in the area using the cross-validation dataset. After application of the trained SOM for the classification of ~~inverse modelling~~inversion results obtained from the inversion of field geophysical data, we obtain a frequentist probability ~~3D probability~~ volume for each lithology. Figure 10~~Figure 11~~ shows the resulting frequentist probability ~~probability~~ volumes for the 6 lithologies present in the Yerrida basin.

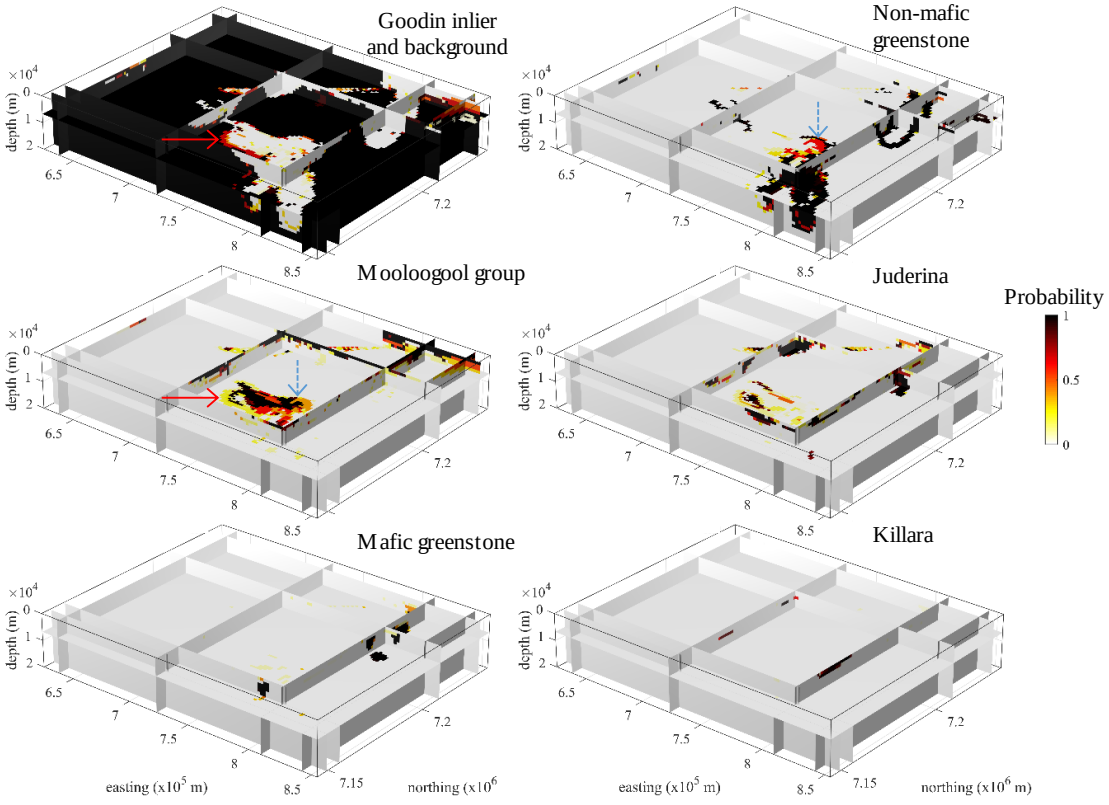
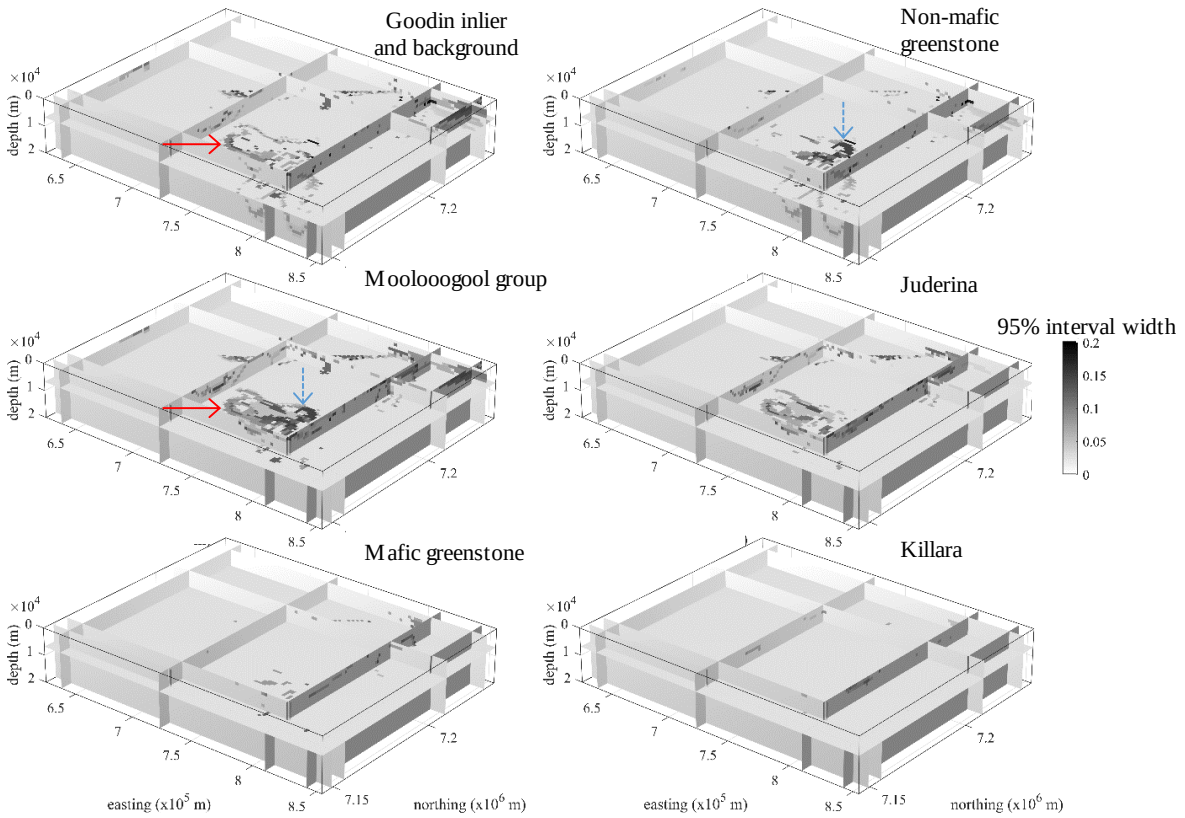


Figure 10~~11~~. Frequentist probability ~~Probability~~ volumes of recovered lithologies calculated as per equation (3). The arrows are drawn to support interpretation.

~~The uncertainty associated with the confidence level (equation 10) shown in Figure 11 is shown in Figure 12.~~



~~**Figure 12.** Value of 95% confidence interval of the probability of recovered lithologies, calculated as per equation (9).~~

~~A straightforward observation that can be made from Figure 12 is that uncertainty in recovered probabilities is very low for the majority of the model cells. It is non-negligible, however, in the central part of the model, where variations in terms of density contrast and geological uncertainty are more pronounced. However, the values of the 95% confidence interval generally confirm interpretations that can be made directly from Figure 10 and Figure 11. Note that the zone of the model shown by the blue arrows in~~ Figure 10~~Figure 11 and Figure 12 exh~~exhibit probabilities between 0.3 and 0.6 ~~on the former and 95% confidence intervals superior to 0.1~~in the area marked by the two arrows. This suggests that these ~~lithology probabilities in~~ these zones are the least-well constrained. ~~In contrast, the zone marked by the red arrow, which also exhibit probabilities between 0.3 and 0.6, presents a narrow 95% confidence interval (<0.01), indicating that recovered lithology probabilities are less uncertain.~~ Note that from Figure 10~~Figure 11, and Figure 12~~ we can interpret the presence of mafic greenstone with confidence as ~~this lithology~~it shows high frequentist probability nearly everywhere classification suggest its presence~~is well constrained~~. For completeness, assessment of the prediction accuracy of the different lithologies is shown by the corresponding boxplot in Appendix B.

3.2.4 Geophysical null space validation and implications for geological interpretation

Applying equation (~~10~~9), we obtain $\|\boldsymbol{W_d G \Delta m}\|_2^2 = 6 \times 10^{-4}$, indicating that the model can be considered geophysically equivalent overall. This is illustrated by the map of the corresponding data misfits (see **Error! Reference source not found.** in appendix B), which ~~. This~~ indicates that we can consider the ~~classified~~ recovered lithological model after application of post-regularisation as reflective of both geophysical and geological

information. Focusing on the mafic greenstone belts of interest in the area (Figure 11~~Figure 13~~), the classification results allow us to propose the following geological interpretations.

Figure 11~~Figure 13~~ shows that the northern portion of the greenstone belts A and B recovered by geophysical inversion and SOM classification is thinner in their Northern part than was proposed by the initial geological model. Likewise, greenstone belt C seems to be much thinner near its centre than expected. Given the data density and lack of understanding we have about the depth of this greenstone belt, this observation is plausible. It confirms and refines considerably the crude, preliminary lithology differentiation of Giraud et al. (2019a) that was based only on density contrast value. The cause of ~~this~~ the thinning of mafic greenstone belt C could be attributed to faulting, folding or to the topography of the palaeoenvironment where the protoliths to the belt were formed, which are not captured directly by surface geology. The portion of the southern Merrie Greenstone Belt (mafic greenstone C) is shown to be thinner than expected, prompting a review of the structure of existing models. A ~~P~~plausible reason is ~~range from~~ the presence of structure that has not been identified from the initial interpretation of geophysical data. In addition, the potential presence of deep-penetrating faults or shear zones, as shown in Figure 11~~Figure 13~~b by the arrows around greenstone C hints a possible false assumption that the Merrie Greenstone Belt is ~~one~~ single and coherent geological body.

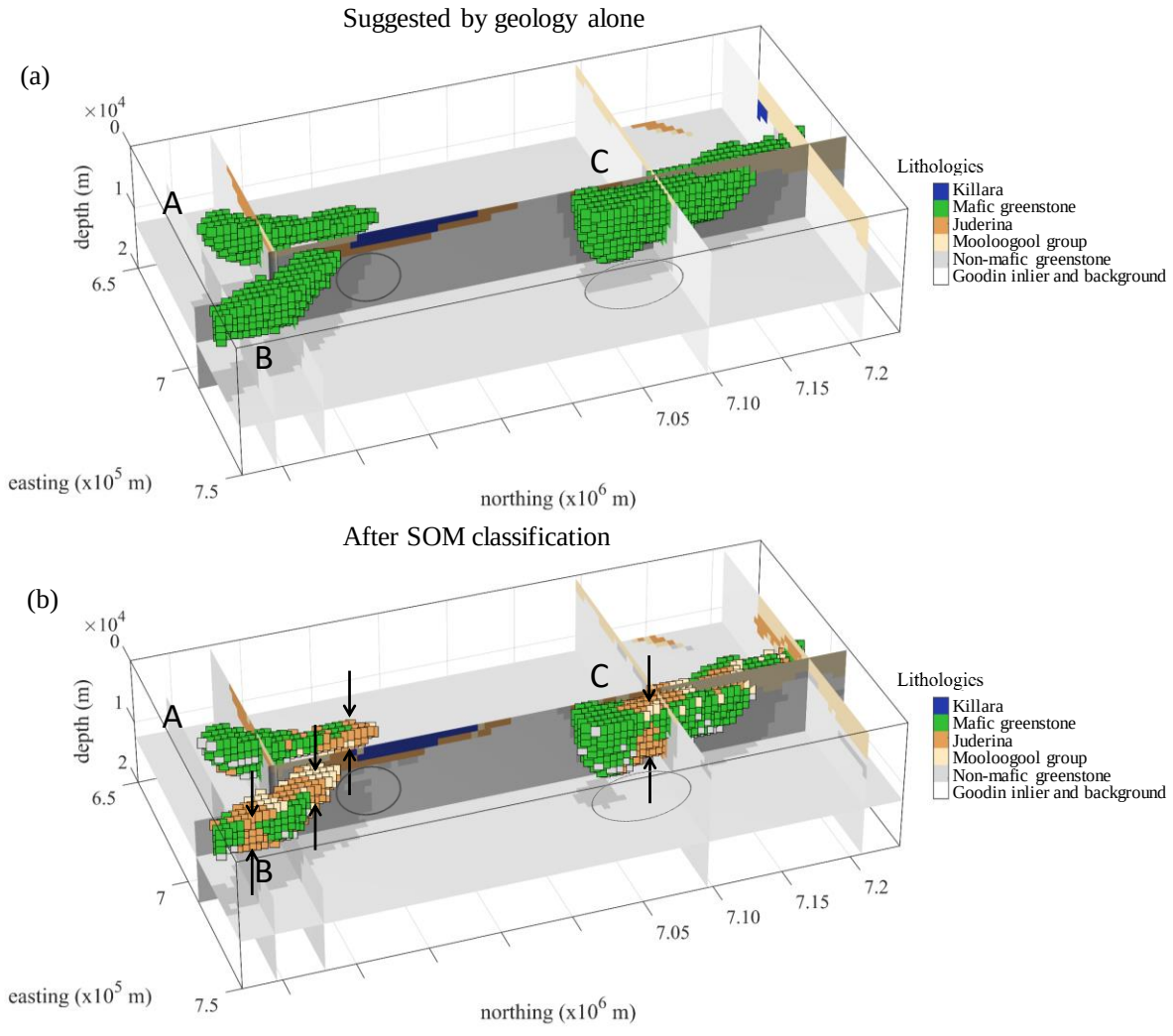


Figure 11~~13~~. Mafic greenstone belts and their surroundings following geological modelling only (a) and after SOM classification (bottom). The cells shows in (a) and (b) have the same geographical location, and are ~~colored~~coloured according to lithology. The vertical arrows show areas where mafic greenstone is thinner than suggested by geology only. The elliptical shapes shows zones where expected non-mafic greenstone is replaced by the background lithology.

4. Discussion

The application of the technique presented here is not restricted to usage of the particular geophysical or geological modelling schemes generating the modelling inputs to this study. The methodology we introduced is general and any different standalone geophysical and geological modelling schemes could also be used.

The work presented here relied on SOM, which can be seen as an extension of the k-means and c-means clustering algorithms used for lithological differentiation by (Paasche and Tronicke, ~~(~~2007~~)~~, Carter-McAuslan et al., ~~(~~2015~~)~~, Sun and Li, ~~(~~2015, 2016b~~)~~, Maag and Li, ~~(~~2018~~)~~, ~~(~~Ward et al., 2014~~)~~, and ~~(~~Singh and Sharma, 2018~~)~~), with which it shares a number of characteristics. We can therefore assume that our findings may hold true for these techniques.

We have shown that the utilisation of post-regularisation can be effective to increase geological realism in the recovered lithological models while preserving the geophysical validity of the corresponding model. The geological principles we used to design our post-regularisation operator apply to lithological topology and focus on adjacency relationship between cells. Ideally, post-regularisation should also consider the surface area of contacts and their topology. This could be followed by, for instance, a 3D extension of the geological model-editing approach of Anquez et al. (2019) to produce genuine geological models ~~honoring~~honouring age relationships, stratigraphic principles, etc. Provided that ~~models obtained in this~~the resulting models ~~fashion honor~~honour equation (9), this approach would ensure that while they are geophysically valid, they can be readily used for interpretation or by commercial or non-commercial geological modelling engines, reservoir simulations, etc. without further processing.

We also believe that post-regularisation can be successfully applied to other clustering techniques. In addition, the implementation of post-regularisation presented here can be readily applied to existing classification, regardless of the classification algorithm used as it only adjusts the classification using spatial-contextual features in the classified model, and could assist the geological characterization of inversion results (Melo et al., 2017).

The example we have shown uses a covariance matrix $\boldsymbol{W}_{sH}$ (equation 1) that results from geological modelling. It is used as a proxy for the uncertainty about our knowledge in terms of structural geology. Such prior information could also be derived from techniques other than geological modelling such as, for instance, prior geophysical modelling, be it using the same or different geophysical methods. 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.

An important result produced here involves the identification of regions which do not adequately conform to the initial model parameters (Figure 11~~Figure 13~~). While this issue remains unresolved, the capability of ~~this~~ our method to identify problematic regions is useful to drive reinterpretation of data, consideration of additional models and, eventually, increased geological knowledge of the target.

Future work may include the generation of multiple lithological models using the trained SOM and the ~~probability~~ frequentist probability volume associated to it. By selecting models belonging to the null-space of the geophysical data (i.e., satisfying equation 9), we expect that this would allow the identification of a series of a few archetypes that would be representative of the various datasets used in the geoscientific modelling workflow.

5. Conclusions

We have introduced a post-inversion classification technique relying on SOM that enables the recovery of lithologies, the corresponding ~~probability~~ frequentist probability voxet ~~and the associated uncertainty,~~ thereby remediating to some of the limitations of deterministic inversion. The proposed technique utilizes a post-regularisation scheme enforcing elementary geological principles to the recovered lithological model while maintaining geophysical validity. We have applied this new methodology to the Yerrida basin (Western Australia) and shown how it improves the geological plausibility of the recovered model. ~~This~~ Results allowed us to confirm previous results and to bring new insights into possible reinterpretation of the geometry of prospective greenstone belts.

Data availability. The input and output of the synthetic survey are made available by Giraud (2019).

Authors contribution. Jérémie Giraud designed the methodology, adapted the SOM algorithm and performed all modelling except geological modelling. Jeremie is the main writer of the manuscript, which was redacted with support from the rest of the authors. Mark Lindsay performed geological modelling and interpretation of the recovered lithologies. Mark Jessell provided guidance and supervision while the project was being carried out~~-~~. Vitaliy Ogarko assisted in the development of the parts of the methodology relating to geophysics and the writing of the paper on aspects relating to SOM.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. Vitaliy Ogarko acknowledges the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3-D (ASTRO 3-D) for supporting some of his research efforts. Finally, the authors thank Evren Pakyuz-Charrier and Roland Martin for interesting discussions relating to topics covered in this paper.

Financial support. Mark W. Jessell was supported by a Western Australian fellowship. Mark D. Lindsay was supported by the Geological Survey of Western Australia and the Exploration Incentive Scheme and the Australia Research Council (DE190100431). Part of this work has been supported by an Australian Government International Postgraduate Research Scholarship. The authors acknowledge partial financial support from the MinEx Cooperative Research Centre. ~~This~~ The research ~~research~~presented here has been supported, in part, by LP170100985: Loop – Enabling Stochastic 3D Geological Modelling, funded by the Australian Research Council and supported by Monash University, University of Western Australia, Geoscience Australia, the Geological Surveys of Western Australia, Northern Territory, South Australia and New South Wales as well as the Research for Integrative Numerical Geology, the Université de Lorraine, RWTH Aachen, the Geological Survey of Canada, the British Geological Survey, the Bureau de Recherches Géologiques et Minières (French Geological Survey), and AuScope.

6. Appendices

~~**Appendix A:** Uncertainty in estimated probabilities~~

~~We estimate the uncertainty attached to the association of an apparition frequency to a probability assuming that it is calculated as the average of a sufficiently large number of samples of a random variable. Under these conditions, the Central Limit Theorem holds true (Sivia and Skilling, 2006, Brookes et al., 2007), meaning that this average can be described using a normal distribution. On this premise, we can calculate the Wald type confidence interval (Wilson, 1927, DasGupta et al., 2001) for a selected target confidence value $\alpha \in]0,1[$. For the i^{th} lithology in the j^{th} node, we calculate:~~

$$\sigma = \sqrt{2}\,\mathrm{erf}^{-1}(\alpha)\left[\frac{n_c n_f}{n_c + n_f}\right]^{1/2} \quad (10)$$

~~where n_c is the number of correct predictions and n_f the number of failed predictions. In the work we present, it becomes:~~

$$\sigma_i = \sqrt{2}\,\mathrm{erf}^{-1}(\alpha)\left[\frac{\tau_i(1-\tau_i)}{\sum_{i=1}^{n_t}\tau_i}\right]^{1/2}, \quad (11)$$

~~where erf is the error function. More details about confidence intervals as calculated here are available can be found in (Simonoff, 2003).~~

~~In the remainder of this paper, we use $\alpha = 0.95$, which approximates a 95% confidence interval where the calculated probability lies in the interval defined as $[max(\tau_i - \sigma_i, 0), min(\tau_i + \sigma_i, 1)]$. This choice of α is subjective. Here, we arbitrarily chose a value that it is widely used in studies involving statistics as initially proposed by (Fisher, 1925).~~ **Appendix A:** data misfit generated by SOM classification and post-regularization

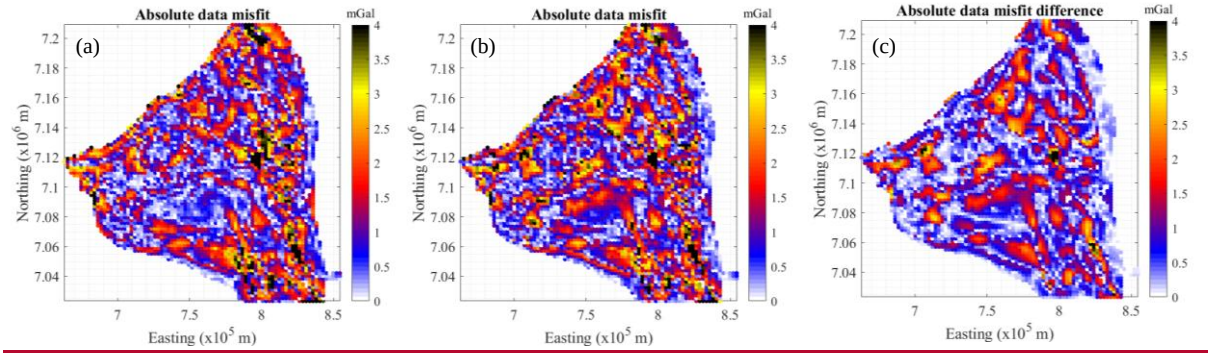


Figure A 1. (a) Comparison of absolute data misfit maps obtained for the inverse model used here (reproduced from ~~(~~Giraud et al., 2019a~~)~~) and (b) after post-regularization, with (c) the misfit differences between (a) and (b).

In Figure A 1, data misfit and the absolute data misfit difference (Figure A 1b and Figure A 1c, respectively) show values which, in places, are relatively high, but which are in line with the gravity data inversion's. Figure A 1a and Figure A 1b show similar features to the exception of a patch in the central part of the model (Northing ~7.09 10^6 m– Easting ~7.7 10^5 m) where Figure A 1b shows values that are approx. 1.5 mGal higher. Figure A 1c shows misfit differences generally in the order of, or lower, than 2 mGal. Also note that there are places where Figure A 1b shows lower misfit than Figure A 1a.

Appendix B: data misfit generated by SOM classification and post-regularization

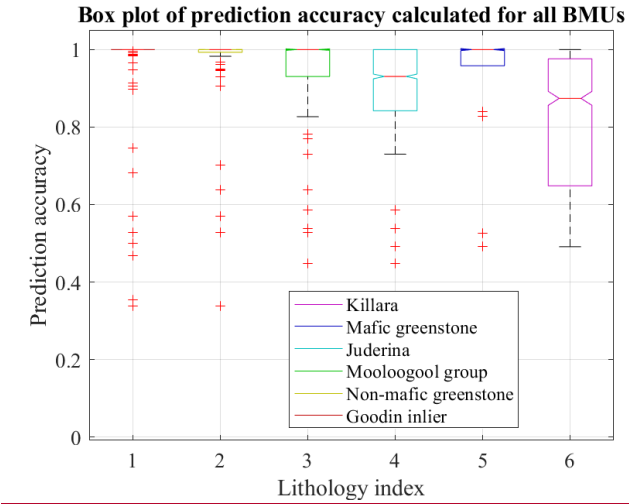


Figure B 1. Boxplot of prediction accuracies for the different lithologies. Red crosses mark outliers.

7. References

Ackora-Prah, J., Y. E. Ayekple, R. K. Acquah, P. S. Andam, E. A. Sakyi, and D. Gyamfi, 2015, Revised Mathematical Morphological Concepts: Advances in Pure Mathematics, **05**, 155–161.

Anquez, P., J. Pellerin, M. Irakarama, P. Cupillard, B. Lévy, and G. Caumon, 2019, Automatic correction and simplification of geological maps and cross-sections for numerical simulations: Comptes Rendus Geoscience, **351**, 48–58.

van der Baan, M., and C. Jutten, 2000, Neural networks in geophysical applications: GEOPHYSICS, **65**, 1032–1047.

Bauer, K., G. Muñoz, and I. Moeck, 2012, Pattern recognition and lithological interpretation of collocated seismic and magnetotelluric models using self-organizing maps: Geophysical Journal International, **189**, 984–998.

Bauer, K., A. Schulze, T. Ryberg, S. V Sobolev, and M. H. Weber, 2003, Classification of lithology from seismic tomography: A case study from the Messum igneous complex, Namibia: Journal of Geophysical Research: Solid Earth, **108**, 1–15.

Benavent, X., E. Dura, F. Vegara, and J. Domingo, 2012, Mathematical Morphology for Color Images: An Image-Dependent Approach: Mathematical Problems in Engineering, **2012**, 1–18.

Burns, K., 1988, Lithologic topology and structural vector fields applied to subsurface prediction in geology: Proceedings of GIS/LIS 88, ACSM-ASPRS.

Carneiro, C. D. C., S. J. Fraser, A. P. Crósta, A. M. Silva, and C. E. de M. Barros, 2012, Semiautomated geologic mapping using self-organizing maps and airborne geophysics in the Brazilian Amazon: GEOPHYSICS, **77**, K17–K24.

Carter-McAuslan, A., P. G. Lelièvre, and C. G. Farquharson, 2015, A study of fuzzy c -means coupling for joint inversion, using seismic tomography and gravity data test scenarios: GEOPHYSICS, **80**, W1–W15.

Chang, H.-C., D. C. Kopaska-Merkel, and H.-C. Chen, 2002, Identification of lithofacies using Kohonen self-organizing maps: Computers & Geosciences, **28**, 223–229.

Chen, J., G. M. Hoversten, D. Vasco, Y. Rubin, and Z. Hou, 2007, A Bayesian model for gas saturation estimation using marine seismic AVA and CSEM data: Geophysics, **72**, WA85.

Chopra, S., and K. J. Marfurt, 2018, Seismic facies classification using some unsupervised machine-learning methods: SEG Technical Program Expanded Abstracts 2018, 2056–2060.

Cracknell, M. J., and A. M. Reading, 2015, Spatial-Contextual Supervised Classifiers Explored: A Challenging Example of Lithostratigraphy Classification: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 1–14.

Deal, M. M., and G. Nolet, 1996, Nullspace shuttles: Geophysical Journal International, **124**, 372–380.

Du, H., J. Cao, Y. Xue, and X. Wang, 2015, Seismic facies analysis based on self-organizing map and empirical mode decomposition: Journal of Applied Geophysics, **112**, 52–61.

Eckhardt, E. A., 1940, PARTNERSHIP BETWEEN GEOLOGY AND GEOPHYSICS IN PROSPECTING FOR OIL: GEOPHYSICS, **5**, 209–214.

Egenhofer, M., and J. Herring, 1990, Categorizing binary topological relations between regions, lines, and points in geographic databases: The, 1–28.

Freeman, B., P. J. Boult, G. Yielding, and S. Menpes, 2010, Using empirical geological rules to reduce structural uncertainty in seismic interpretation of faults: Journal of Structural Geology, **32**, 1668–1676.

Giraud, J., 2019, Synthetic geophysical survey using geological modelling from the Yerrida Basin (Western Australia): .

Giraud, J., E. Pakyuz-Charrier, M. Jessell, M. Lindsay, R. Martin, and V. Ogarko, 2017, Uncertainty reduction through geologically conditioned petrophysical constraints in joint inversion: GEOPHYSICS, **82**, ID19–ID34.

Giraud, J., E. Pakyuz-Charrier, V. Ogarko, M. Jessell, M. Lindsay, and R. Martin, 2018, Impact of uncertain geology in constrained geophysical inversion: ASEG Extended Abstracts, **2018**, 1.

Giraud, J., M. Lindsay, V. Ogarko, M. Jessell, R. Martin, and E. Pakyuz-Charrier, 2019a, Integration of geoscientific uncertainty into geophysical inversion by means of local gradient regularization: Solid Earth, **10**, 193–210.

Giraud, J., V. Ogarko, M. Lindsay, E. Pakyuz-Charrier, M. Jessell, and R. Martin, 2019b, Sensitivity of constrained joint inversions to geological and petrophysical input data uncertainties with posterior geological analysis: Geophysical Journal International, **218**, 666–688.

Godefroy, G., G. Caumon, G. Laurent, and F. Bonneau, 2019, Structural Interpretation of Sparse Fault Data Using Graph Theory and Geological Rules: Mathematical Geosciences.

Godsil, C., and G. Royle, 2001, Algebraic Graph Theory, *in* , 1–18.

Goodfellow, I., Y. Bengio, and A. Courville, 2016, Deep Learning: MIT Press.

Green, C. H., 1948, INTEGRATION IN EXPLORATION: GEOPHYSICS, **13**, 365–370.

Hansen, P. C., and D. P. O'Leary, 1993, The Use of the L-Curve in the Regularization of Discrete Ill-Posed Problems: SIAM Journal on Scientific Computing, **14**, 1487–1503.

Hansen, P. C., and P. R. Johnston, 2001, The L-Curve and its Use in the Numerical Treatment of Inverse Problems, *in* Computational Inverse Problems in Electrocardiography, , 119–142.

Jessell, M., E. Pakyuz-charrier, M. Lindsay, J. Giraud, and E. de Kemp, 2018, Assessing and Mitigating Uncertainty in Three-Dimensional Geologic Models in Contrasting Geologic Scenarios: , 63–74.

Jupp, D. L. B., and K. Vozoff, 1975, Joint inversion of geophysical data: Geophysical Journal of the Royal Astronomical Society, **42**, 977–991.

Kalteh, A. M., P. Hjorth, and R. Berndtsson, 2008, Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application: Environmental Modelling & Software, **23**, 835–845.

Klose, C. D., 2006, Self-organizing maps for geoscientific data analysis: geological interpretation of multidimensional geophysical data: Computational Geosciences, **10**, 265–277.

Köhler, A., M. Ohrnberger, and F. Scherbaum, 2010, Unsupervised pattern recognition in continuous seismic wavefield records using Self-Organizing Maps: Geophysical Journal International, **182**, 1619–1630.

Kohonen, T., 1982a, Analysis of a simple self-organizing process: Biological Cybernetics, **44**, 135–140.

Kohonen, T., 1982b, Self-organized formation of topologically correct feature maps: Biological Cybernetics, **43**, 59–69.

Kohonen, T., 1998, The self-organizing map: Neurocomputing, **21**, 1–6.

Kohonen, T., 2013, Essentials of the self-organizing map: Neural Networks, **37**, 52–65.

de la Varga, M., A. Schaaf, and F. Wellmann, 2018, GemPy 1.0: open-source stochastic geological modeling and inversion: Geoscientific Model Development Discussions, 1–50.

Lelièvre, P. G., and C. G. Farquharson, 2016, Integrated Imaging for Mineral Exploration, *in* Integrated Imaging of the Earth: Theory and Applications, .

Li, Y., and D. W. Oldenburg, 2000, Incorporating geological dip information into geophysical inversions: GEOPHYSICS, **65**, 148–157.

Li, Y., A. Melo, C. Martinez, and J. Sun, 2019, Geology differentiation: A new frontier in quantitative geophysical interpretation in mineral exploration: The Leading Edge, **38**, 60–66.

Lindsay, M., S. Occhipinti, L. Ramos, A. Aitken, and P. Hilliard, 2018, An Integrated View of the Yerrida Basin with Implications for Its Architecture and Mineral Prospectivity:

Lindsay, M., S. Occhipinti, C. Laflamme, A. Aitken, and L. Ramos, 2020, Mapping undercover: integrated geoscientific interpretation and 3D modelling of a Proterozoic basin: Solid Earth Discussions.

Lines, L. R., A. K. Schultz, and S. Treitel, 1988, Cooperative inversion of geophysical data: GEOPHYSICS, **53**, 8–20.

Maag, E., and Y. Li, 2018, Discrete-valued gravity inversion using the guided fuzzy c -means clustering technique: GEOPHYSICS, **83**, G59–G77.

Martin, R., and K. Obermayer, 2009, Self-Organizing Maps, *in* Encyclopedia of Neuroscience, Elsevier, 551–560.

Martin, R., V. Ogarko, D. Komatitsch, and M. Jessell, 2018, Parallel three-dimensional electrical capacitance data imaging using a nonlinear inversion algorithm and Lp norm-based model: Measurement, **128**, 428–445.

Meju, M. A., and L. A. Gallardo, 2016, Structural Coupling Approaches in Integrated Geophysical Imaging, *in* , 49–67.

Melo, A. T., J. Sun, and Y. Li, 2017, Geophysical inversions applied to 3D geology characterization of an iron oxide copper-gold deposit in Brazil: GEOPHYSICS, **82**, K1–K13.

Miljkovic, D., 2017, Brief review of self-organizing maps: 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 1061–1066.

Moorkamp, M., B. Heincke, M. Jegen, R. W. Hobbs, and A. W. Roberts, 2016, Joint Inversion in Hydrocarbon Exploration, *in* Integrated Imaging of the Earth: Theory and Applications, , 167–189.

Muñoz, G., and V. Rath, 2006, Beyond smooth inversion: the use of nullspace projection for the exploration of non-uniqueness in MT: Geophysical Journal International, **164**, 301–311.

Nettleton, L. L., 1949, GEOPHYSICS, GEOLOGY AND OIL FINDING: GEOPHYSICS, **14**, 273–289.

Occhipinti, S., R. Hocking, M. Lindsay, A. Aitken, I. Copp, J. Jones, S. Sheppard, F. Pirajno, and V. Metelka, 2017, Paleoproterozoic basin development on the northern Yilgarn Craton, Western Australia: Precambrian Research, **300**.

Paasche, H., and J. Tronicke, 2007, Cooperative inversion of 2D geophysical data sets: A zonal approach based on fuzzy c -means cluster analysis: Geophysics, **72**, A35–A39.

Paasche, H., J. Tronicke, and P. Dietrich, 2010, Automated integration of partially colocated models: Subsurface zonation using a modified fuzzy c -means cluster analysis algorithm: GEOPHYSICS, **75**, P11–P22.

Pakyuz-Charrier, E., J. Giraud, M. Lindsay, and M. Jessell, 2018a, Common Uncertainty Research Explorer Uncertainty Estimation in Geological 3D Modelling: ASEG Extended Abstracts, **2018**, 1.

Pakyuz-Charrier, E., J. Giraud, V. Ogarko, M. Lindsay, and M. Jessell, 2018b, Drillhole uncertainty propagation

for three-dimensional geological modeling using Monte Carlo: Tectonophysics.

Pakyuz-Charrier, E., M. Lindsay, V. Ogarko, J. Giraud, and M. Jessell, 2018c, Monte Carlo simulation for uncertainty estimation on structural data in implicit 3-D geological modeling, a guide for disturbance distribution selection and parameterization: Solid Earth, **9**, 385–402.

Pakyuz-Charrier, E., M. Jessell, J. Giraud, M. Lindsay, and V. Ogarko, 2019, Topological analysis in Monte Carlo simulation for uncertainty propagation: Solid Earth, **10**, 1663–1684.

Pellerin, J., A. Botella, F. Bonneau, A. Mazuyer, B. Chauvin, B. Lévy, and G. Caumon, 2017, RINGMesh: A programming library for developing mesh-based geomodeling applications: Computers & Geosciences, **104**, 93–100.

Pirajno, F., and N. G. Adamides, 2000, Geology and Mineralization of the Palaeoproterozoic Yerrida Basin, Western Australia:

Roden, R., T. Smith, and D. Sacrey, 2015, Geologic pattern recognition from seismic attributes: Principal component analysis and self-organizing maps: Interpretation, **3**, SAE59–SAE83.

Santos, E. T. F., and A. Bassrei, 2007, L- and Θ-curve approaches for the selection of regularization parameter in geophysical diffraction tomography: Computers & Geosciences, **33**, 618–629.

Shalaginov, A., and K. Franke, 2015, A new method for an optimal SOM size determination in neuro-fuzzy for the digital forensics applications: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).

Shannon, C. E. E., 1948, A Mathematical Theory of Communication: Bell System Technical Journal, **27**, 379–423.

Singh, A., and S. P. Sharma, 2018, Identification of different geologic units using fuzzy constrained resistivity tomography: Journal of Applied Geophysics, **148**, 127–138.

Stavrakoudis, D., E. Dragozi, I. Gitas, and C. Karydas, 2014, Decision Fusion Based on Hyperspectral and Multispectral Satellite Imagery for Accurate Forest Species Mapping: Remote Sensing, **6**, 6897–6928.

Sun, J., and Y. Li, 2015, Multidomain petrophysically constrained inversion and geology differentiation using guided fuzzy c-means clustering: Geophysics, **80**, ID1–ID18.

Sun, J., and Y. Li, 2016, Joint-clustering inversion of gravity and magnetic data applied to the imaging of a gabbro intrusion: SEG Technical Program Expanded Abstracts 2016, 2175–2179.

Sun, J., and Y. Li, 2019, Magnetization clustering inversion Part II: Assessing the uncertainty of recovered magnetization directions: GEOPHYSICS, 1–86.

Tarabalka, Y., J. A. Benediktsson, and J. Chanussot, 2009, Spectral–Spatial Classification of Hyperspectral Imagery Based on Partitional Clustering Techniques: IEEE Transactions on Geoscience and Remote Sensing, **47**, 2973–2987.

Thiele, S. T., M. W. Jessell, M. Lindsay, V. Ogarko, J. F. Wellmann, and E. Pakyuz-Charrier, 2016, The topology of geology 1: Topological analysis: Journal of Structural Geology, **91**, 27–38.

Towles, H. C., 1952, A STUDY IN INTEGRATION OF GEOLOGY AND GEOPHYSICS: GEOPHYSICS, **17**, 876–899.

Uriarte, E. A., and F. D. Martín, 2005, Topology Preservation in SOM: International Journal of Applied Mathematics and Computer Sciences.

Vatanen, T., M. Osmala, T. Raiko, K. Lagus, M. Sysi-Aho, M. Orešič, T. Honkela, and H. Lähdesmäki, 2015, Self-organization and missing values in SOM and GTM: Neurocomputing, **147**, 60–70.

Vesanto, J., and E. Alhoniemi, 2000, Clustering of the self-organizing map: IEEE Transactions on Neural Networks, **11**, 586–600.

Ward, W. O. C., P. B. Wilkinson, J. E. Chambers, L. S. Oxby, and L. Bai, 2014, Distribution-based fuzzy clustering of electrical resistivity tomography images for interface detection: Geophysical Journal International, **197**, 310–321.

Wegener, A., 1915, Die Entstehung Der Kontinente Und Ozeane, Braunschwe.:

Wellmann, J. F., and K. Regenauer-Lieb, 2012, Uncertainties have a meaning: Information entropy as a quality measure for 3-D geological models: Tectonophysics, **526**–**529**, 207–216.

Wellmann, J. F., F. G. Horowitz, E. Schill, and K. Regenauer-Lieb, 2010, Towards incorporating uncertainty of structural data in 3D geological inversion: Tectonophysics, **490**, 141–151.

Wrona, T., I. Pan, R. L. Gawthorpe, and H. Fossen, 2018, Seismic facies analysis using machine learning: GEOPHYSICS, **83**, O83–O95.

Zhang, G., Z. Wang, and Y. Chen, 2018, Deep learning for Seismic Lithology Prediction: Geophysical Journal International, 1368–1387.

Zhao, T., F. Li, and K. J. Marfurt, 2017, Constraining self-organizing map facies analysis with stratigraphy: An approach to increase the credibility in automatic seismic facies classification: Interpretation, **5**, T163–T171.

Zhao, T., V. Jayaram, A. Roy, and K. J. Marfurt, 2015, A comparison of classification techniques for seismic facies recognition: Interpretation, **3**, SAE29–SAE58.

Zlatanova, S., 2000, On 3D topological relationships: Proceedings - International Workshop on Database and Expert Systems Applications, DEXA, **2000**-**January**, 913–919.