# Towards geologically reasonable lithological classification from integrated geophysical inverse modelling: 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:** 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 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 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 of geology and geophysics 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 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, Giraud et al., 2017). In contrast, the recovery of geological quantities from geophysical inverse modelling has seen much less effort. Recent studies have started to rectify this by proposing the idea of lithological differentiation of inverse modelling results (Paasche et al. 2010, Sun and Li 2015, Paasche and Tronicke, 2007), which consist in 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

5  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, like all modelling results, the identification of facies or lithologies

10  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,, ).

15  It is also evident that the modelling of lithology (or facies) is sensitive to a broad range of parameters and that their characterization relies on properties that are the results of complex, non-linear physical processes. 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

20  is rarely estimated or even considered.

To date, whether it be in oil and gas or mining exploration, uncertainty in recovered rock types 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

25  comprehensive uncertainty analyses, practitioners often lack quantitative, robust uncertainty modelling necessary to informed 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 faciesies or rock types in three-dimensions give little to no consideration geological information and rules for their classification.

30  To mitigate this, and to complement existing methodologies, we propose a solution that partially addresses the issues and shortcomings highlighted above. 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

35  network trained in a fully controlled environment with attributes characterizing the geophysical inverse model, 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 of the different lithologies and estimate the related uncertainty for each model-cell discretizing the studied area.

The methodology we propose can serve multiple 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 workflows by providing a general, automated method to derive a probabilistic lithological interpretation of inverse modelling 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, lithologies are identified though a classification technique relying on a simple artificial neural network. We chose 'self-organizing maps' (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. 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 of 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 methodology is summarized in Figure 1 below.
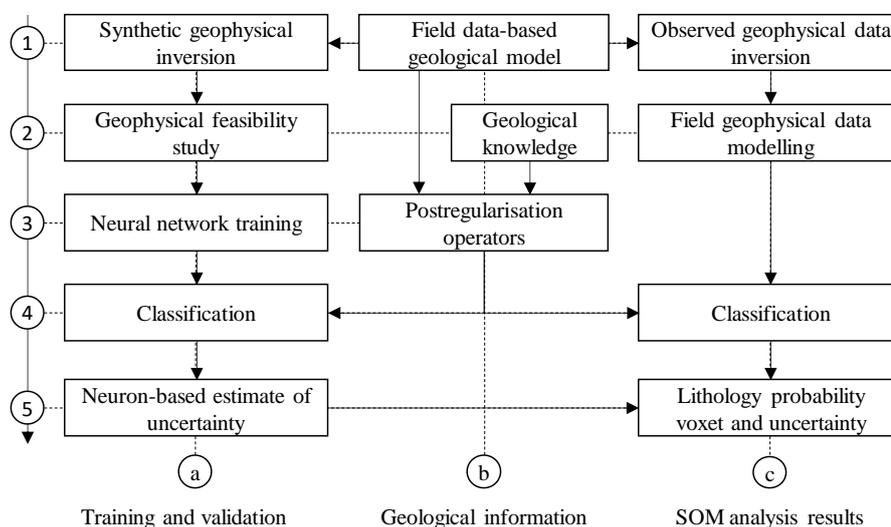


**Figure 1.** Schematic summary of proposed methodology.

## 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.

5    ### 2.1    Integrated geophysical 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 weight varies locally accordingly with geological uncertainty (Giraud et al., 2019a). The cost function to minimize is given as:

$$\theta(d, m) = \left\| W_d \left( d - g(m) \right) \right\|_2^2 + \alpha_m \left\| W_m \left( m - m_p \right) \right\|_2^2 + \alpha_H \| W_H \nabla m \|_2^2, \tag{1}$$

where $d$ represents the measurements and $m$ is the model; $g$ is the forward operator calculating the data $m$ produces; $m_p$ is the prior model. In this contribution, $W_d$ is a diagonal matrix where each element is equal to the inverse of the sum-of-squares of the geophysical measurements; $W_m$ and $W_H$ are diagonal covariance matrices; $\alpha_m$ and $\alpha_H$ are weights controlling the relative importance of the different terms in the equation; $\nabla$ is the spatial gradient operator. The last term of the equation (1) constraints the structural features of the inverted model. The values in diagonal matrix $W_H$ are determined from prior information. In this work, $W_H$ is obtained from geological modelling results and is a proxy for geological uncertainty.

The matrix $W_H$ 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 $m$ is calculated as the product of the Jacobian matrix $G$ with model $m$, i.e., we have $g(m) = Gm$.

### 2.1.2    Geological uncertainty

Geological uncertainty is estimated from probabilistic geological modelling. During this process, a series 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 through calculation of Shannon's entropy (Shannon, 1948) for the simulated geological models (Wellmann and Regenauer-Lieb, 2012). 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 $W_H$ 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 and uncertainty analysis

#### 2.2.1 Fundamentals of self-organizing maps

The 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

5   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 is made of a predefined number of interconnected neurons (also referred to as 'nodes' or 'units') that have a fixed network configuration.

10   Projection occurs during the training phase, where the locations of the neurons in the manifold are iteratively adjusted so that they approximate it optimally.

#### 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

15   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

20   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

25   Ideally, the SOM should be trained in a controlled environment where all the variables used are perfectly known. This motivates the utilisation of synthetic geophysical data. In this work, we 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 study (Figure 1, lines 1 and 2). The training and tests datasets are comprised of the following variables:

30     1.  Starting model for inversion $m_{start}$;

    2.  Inverted model $m_{inv}$;

    3.  Geological uncertainty $W_H$;

    4.  Spatial gradient in the inverted model $\|\nabla m_{inv}\|_2$;

    5.  Most likely lithology obtained from geological modelling.

35   Each datum is a vector $x \in \mathbb{R}^{n_v}$, with a number of variables $n_v = 5$, where $u(5)$ 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

5   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

10  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 $x_i$:

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

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

15  the quantity sought for.

### 2.2.3    Geological post-regularisation

#### 2.2.3.1    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

20  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

25  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.

30  The process of postregularisation, 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). s

#### 2.2.3.2    Implementation

The post-regularisation (PR) scheme we develop for the recovery of lithologies in 3D relies on two hypotheses.

35  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

5    for structural geological interpretation (Freeman et al., 2010, Godefroy et al., 2019). In this paper, we extend existing post-regularisation approaches (ie,, 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.

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

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

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 $n_l \times 1$ column vector of ones, and $\mathbf{0}_{n_l,n_l}$ is the $n_l \times n_l$ 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

15    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

20    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 this 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

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

30    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.

Solid Earth
Discussions
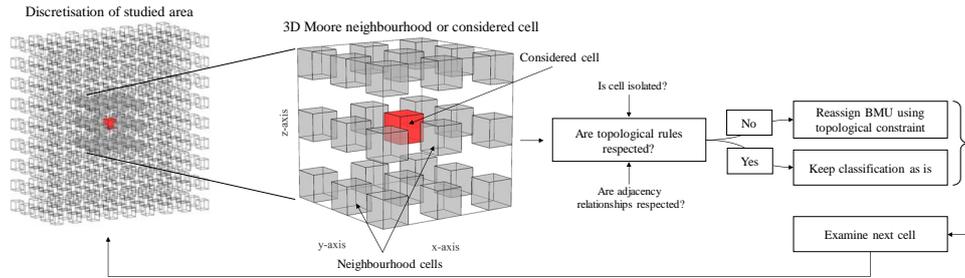


**Figure 2.** Summary of topological filtering used during post-regularisation.

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.

5      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:

$$M_{i>j,j} = n_{ij}, M_{i\leq j,j} = 0,$$
(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:

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

10     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, for instance, "lithology B must always be between lithologies A and C in a conformable sequence".

15     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, 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 predict lithology A using the validation dataset with
20     80% accuracy with index B with 20% accuracy. This process is described below.

### 2.2.4     Estimation the confidence in recovered lithologies

The prediction accuracy $\tau_i$ 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
25     representation of the number of occurrences of true/false positives/negatives.

We use the prediction accuracy $\tau_i$ as a metric measuring the capability of the nodes of the trained SOM to recover lithologies. Let $\tau_i$ be expressed as:

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

which is a particular case of the overall accuracy $\tau$:

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

From equation (7), it appears that $\tau_i$ is equivalent to the relative apparition frequency of the $i^{\text{th}}$ lithology over the entire SOM. When considering a specific node $j$, it becomes the relative apparition frequency of the lithology $i$ for cells classified as having the $j^{\text{th}}$ node as their BMU, noted $\tau_{ij}$. 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 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 the units of SOM is consistent with geophysical measurements. To this end, we verify that the geophysical response of the density contrast model $m_{SOM}$ corresponding to the BMUs of each cell in the studied area fits the field measurement $d$ within a certain tolerance assumed to approximate noise level. Consequently, we ensure that the difference between $m_{inv}$ and $m_{SOM}$, $\Delta m = m_{inv} - m_{SOM}$, satisfies the following condition:

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

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

The implication of equation (9) is that the difference $\Delta m$ 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).

### 3.     Application case: Yerrida basin

### 3.1     Survey setting

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

### 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 3a). 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.
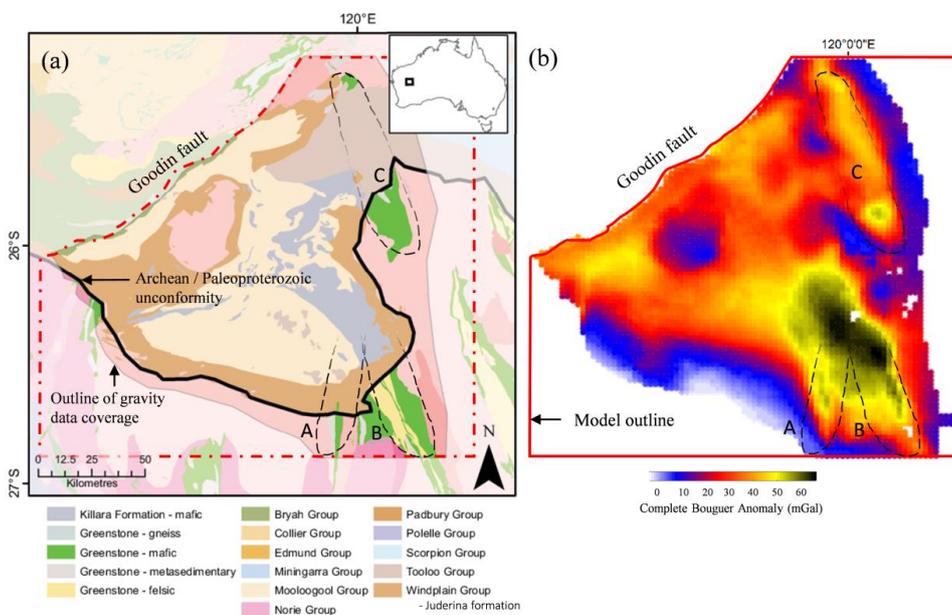


**Figure 3.** (a, to the left) Geological map of the area and (b, to the right) complete Bouguer anomaly (reproduced from Giraud et al., 2019a). The dashed lines delineate the possible sub-basin extent of the mafic greenstone belts. Capital letters 'A', 'B' and 'C' symbolize the possible outlines for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

The density contrast of the rock types observed in the area range between 0 and 330 kg/m³, making it 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

10

(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 3b) is comprised of a total of 4882 measurement points. The

5    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 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.

10    The volume of most probable lithology, $W_H$ and the starting model are shown in Figure 4. Volumes shown in Figure 4a, b, and c are used for the training and validation dataset for SOM training as explained in 2.2.2.2.
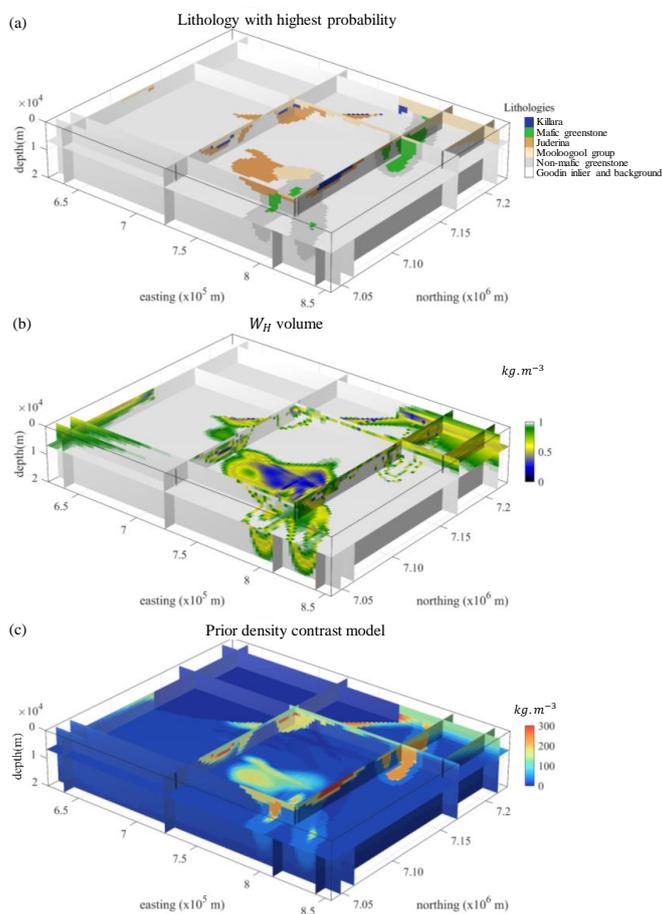


**Figure 4.** Prior geological modelling. Preferred lithology volume (a), geological uncertainty volume (b) and starting density contrast model (c). Modified from (Giraud et al., 2019a).

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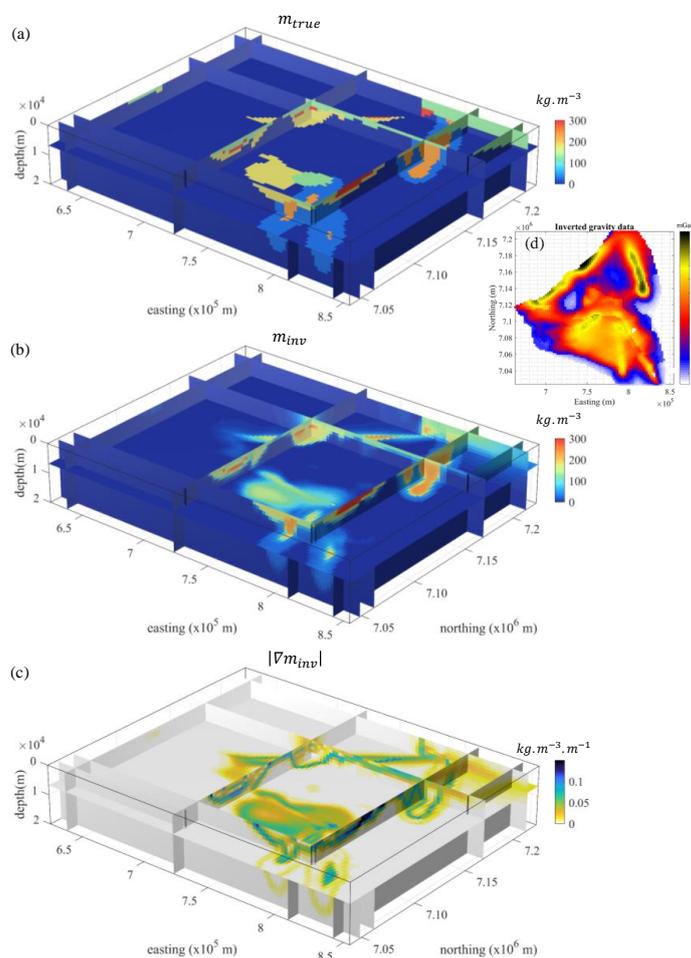


**Figure 5.** True density contrast model for calculation of synthetic geophysical data (a); inverted model obtained from integrated inversion of synthetic geophysical data (b), corresponding spatial gradient of density contrast (c) and synthetic geophysical data (d).

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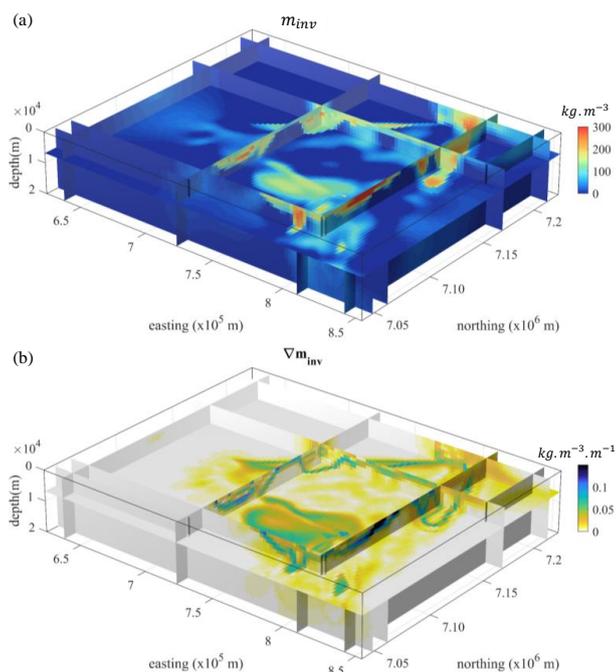
**Figure 6.** Inverted model obtained from integrated inversion of field geophysical data (a), corresponding spatial gradient of density contrast (b).

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

Solid Earth
Discussions
Open Access
EGU



**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 $\phi$ 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).

5    **3.2    Geologically-constrained SOM classification and uncertainty analysis**

**3.2.1    Training the neural network**

It 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) and approximately matches the proposed value of $5\sqrt{N}$ (with $N$ the total number of observation) proposed by (Vesanto and Alhoniemi, 2000)

10    and commonly used since (Shalaginov and Franke, 2015). The use of this value is corroborated by the lithology prediction accuracy (Figure 8) as approximating the point of diminishing returns, i.e., the number of nodes beyond which additional nodes are becoming nearly redundant.
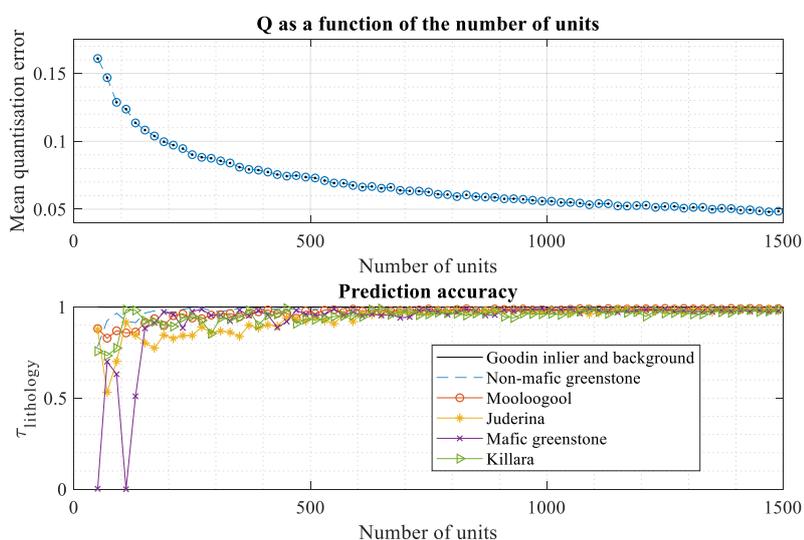


**Figure 8.** (a) Elbow curve of the quantization error for the determination of the optimum number of neurons (or units) in SOM
15    and (b) Prediction accuracy for the different lithologies present in the training and validation datasets.

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 where all lithologies are recovered with a prediction accuracy superior to 90%.

### 3.2.2    Classification and post-regularisation

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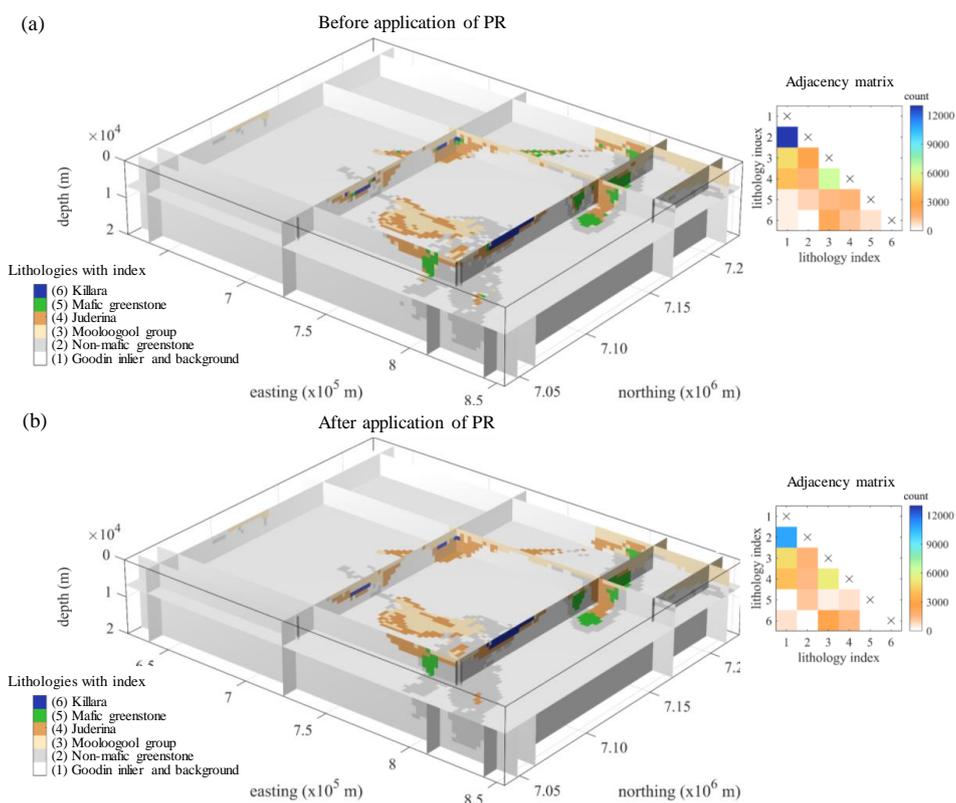


**Figure 9.** Recovered lithologies before post-regularisation (a) and after (b) next to the corresponding adjacency matrix (right hand side). The indices used to designate lithologies are indicated in the legend of the volume (left hand side).

As can be inferred from the adjacency matrices plotted in Figure 9, 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. The number of contacts between lithologies 5 and 2 increased slightly due to post-regularisation because a large percentage of contacts between lithologies 5

15    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.

### 3.2.3    Estimated confidence in recovered lithologies

For each node of the SOM we calculate the prediction accuracy $\tau_i$ (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 results obtained from the inversion of field geophysical data, we obtain a 3D probability volume for each lithology. Figure 10 shows the resulting probability volumes for the 6 lithologies present in the Yerrida basin.
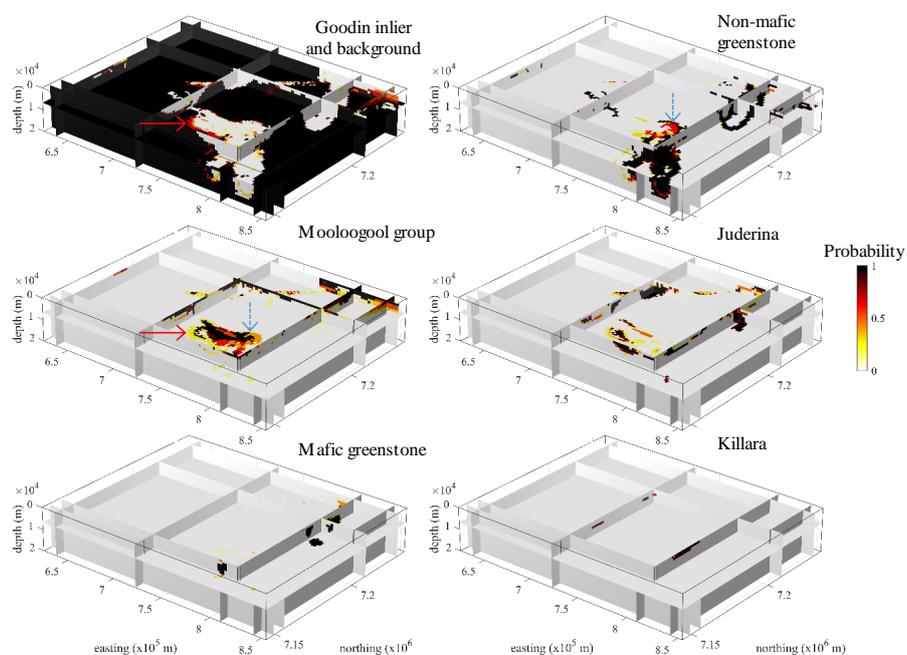


**Figure 10.** Probability volumes of recovered lithologies calculated as per equation (3). The arrows are drawn to support interpretation.

The uncertainty associated with the confidence level (equation 10) shown in Figure 10 is shown in Figure 11.
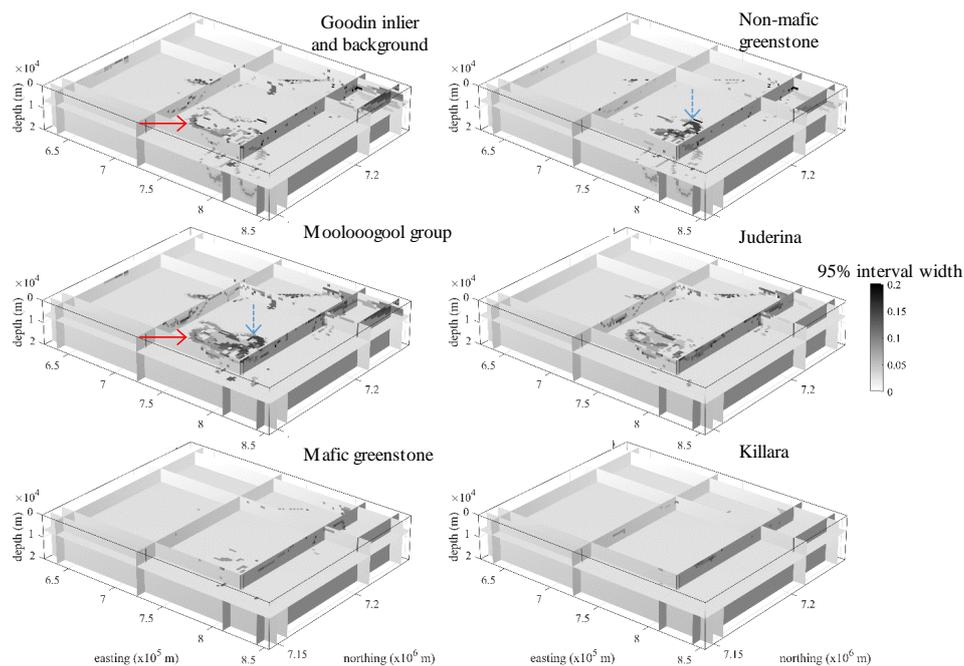
**Figure 11.** Value of 95% confidence interval of the probability of recovered lithologies, calculated as per equation (9).

A straightforward observation that can be made from Figure 11 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 9 and Figure 10. Note that the zone of the model shown by the blue arrows in Figure 10 and Figure 11 exhibit probabilities between 0.3 and 0.6 on the former and 95% confidence intervals superior to 0.1. 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 and Figure 11 we can interpret the presence of mafic greenstone with confidence as this lithology is well constrained.

### 3.2.4 Geophysical null space validation and implications for geological interpretation

Applying equation (9), we obtain $\|W_d G \Delta m\|_2^2 = 6 \times 10^{-4}$, indicating that the model can be considered geophysically equivalent overall. This indicates that we can consider the classified 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 12), the classification results allow us to propose the following geological interpretations.

Figure 12 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 thinning 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. Plausible reasons 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 12b by the arrows around greenstone C hints a possible false assumption that the Merrie Greenstone Belt is one coherent geological body.
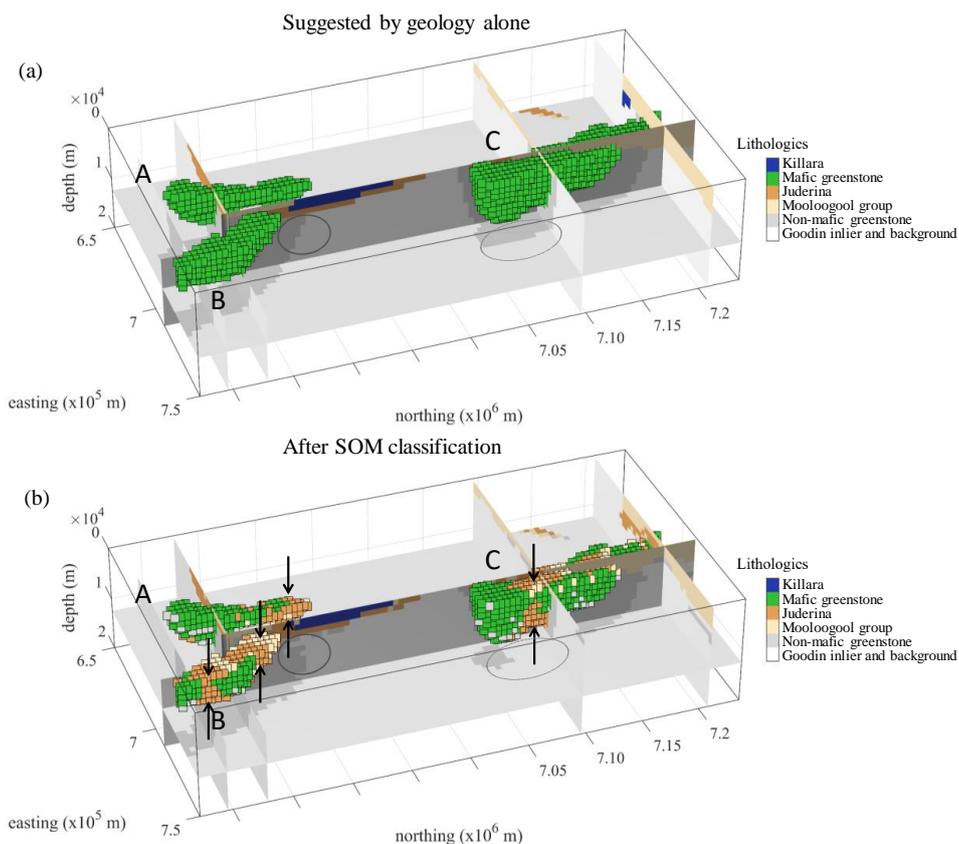


**Figure 12.** Mafic greenstone belts and their surroundings following geological modelling only (a) and after SOM classification (bottom). The cells shows in (a) and (b) have the same geographical location, and are coloured according to lithology. The vertical arrows show areas where mafic greenstone is thinner than suggested by geology only. The elliptical shapes shows zones where expected non-mafic greenstone is replaced by the background lithology.

Solid Earth
Discussions

Open Access

EGU

## 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.

5      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
10     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 honouring age relationships,
15     stratigraphic principles, etc. Provided that models obtained in this fashion honour equation (9), this 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,
20     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 $W_H$ (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
25     modelling, be it using the same or different geophysical methods.

An important result produced here involves the identification of regions which do not adequately conform to the initial model parameters (Figure 12). While this issue remains unresolved, the capability of this method to identify problematic regions is useful to drive reinterpretation of data, consideration of additional models and, eventually, increased geological knowledge of the target.

30     Future work may include the generation of multiple lithological models using the trained SOM and the 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 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 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.

Solid Earth
Discussions

Open Access

EGU

## 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

5   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:

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

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

10   becomes:

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

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

In 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 $\alpha$ is subjective. Here, we

15   arbitrarily chose a value that it is widely used in studies involving statistics as initially proposed by (Fisher, 1925).

## 7.  References

Ackora-Prah, J., Ayekple, Y. E., Acquah, R. K., Andam, P. S., Sakyi, E. A. and Gyamfi, D.: Revised Mathematical Morphological Concepts, Adv. Pure Math., 05(04), 155–161, doi:10.4236/apm.2015.54019, 2015.

Anquez, P., Pellerin, J., Irakarama, M., Cupillard, P., Lévy, B. and Caumon, G.: Automatic correction and

20   simplification of geological maps and cross-sections for numerical simulations, Comptes Rendus Geosci., 351(1), 48–58, doi:10.1016/j.crte.2018.12.001, 2019.

van der Baan, M. and Jutten, C.: Neural networks in geophysical applications, GEOPHYSICS, 65(4), 1032–1047, doi:10.1190/1.1444797, 2000.

Bauer, K., Schulze, A., Ryberg, T., Sobolev, S. V and Weber, M. H.: Classification of lithology from seismic

25   tomography: A case study from the Messum igneous complex, Namibia, J. Geophys. Res. Solid Earth, 108(B3), 1–15, doi:10.1029/2001JB001073, 2003.

Bauer, K., Muñoz, G. and Moeck, I.: Pattern recognition and lithological interpretation of collocated seismic and magnetotelluric models using self-organizing maps, Geophys. J. Int., 189(2), 984–998, doi:10.1111/j.1365-246X.2012.05402.x, 2012.

Benavent, X., Dura, E., Vegara, F. and Domingo, J.: Mathematical Morphology for Color Images: An Image-Dependent Approach, Math. Probl. Eng., 2012, 1–18, doi:10.1155/2012/678326, 2012.

Brookes, B. C., Gnedenko, B. V., Kolmogorov, A. N. and Chung, K. L.: Limit Distributions for Sums of Independent Random Variables, Math. Gaz., doi:10.2307/3608621, 2007.

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

Carneiro, C. D. C., Fraser, S. J., Crósta, A. P., Silva, A. M. and Barros, C. E. de M.: Semiautomated geologic mapping using self-organizing maps and airborne geophysics in the Brazilian Amazon, GEOPHYSICS, 77(4), K17–K24, doi:10.1190/geo2011-0302.1, 2012.

10   Carter-McAuslan, A., Lelièvre, P. G. and Farquharson, C. G.: A study of fuzzy c -means coupling for joint inversion, using seismic tomography and gravity data test scenarios, GEOPHYSICS, 80(1), W1–W15, doi:10.1190/geo2014-0056.1, 2015.

Chang, H.-C., Kopaska-Merkel, D. C. and Chen, H.-C.: Identification of lithofacies using Kohonen self-organizing maps, Comput. Geosci., 28(2), 223–229, doi:10.1016/S0098-3004(01)00067-X, 2002.

15   Chen, J., Hoversten, G. M., Vasco, D., Rubin, Y. and Hou, Z.: A Bayesian model for gas saturation estimation using marine seismic AVA and CSEM data, Geophysics, 72(2), WA85, doi:10.1190/1.2435082, 2007.

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

20   Cracknell, M. J. and Reading, A. M.: Spatial-Contextual Supervised Classifiers Explored: A Challenging Example of Lithostratigraphy Classification, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 1–14, doi:10.1109/JSTARS.2014.2382760, 2015.

DasGupta, A., Cai, T. T. and Brown, L. D.: Interval Estimation for a Binomial Proportion, Stat. Sci., doi:10.1214/ss/1009213286, 2001.

25   Deal, M. M. and Nolet, G.: Nullspace shuttles, Geophys. J. Int., 124(2), 372–380, doi:10.1111/j.1365-246X.1996.tb07027.x, 1996.

Du, H., Cao, J., Xue, Y. and Wang, X.: Seismic facies analysis based on self-organizing map and empirical mode decomposition, J. Appl. Geophys., 112, 52–61, doi:10.1016/j.jappgeo.2014.11.007, 2015.

Eckhardt, E. A.: PARTNERSHIP BETWEEN GEOLOGY AND GEOPHYSICS IN PROSPECTING FOR OIL,
30   GEOPHYSICS, 5(3), 209–214, doi:10.1190/1.1441804, 1940.

Egenhofer, M. and Herring, J.: Categorizing binary topological relations between regions, lines, and points in geographic databases, The, (July 2016), 1–28 [online] Available from: http://didattica.univaq.it/moodle/file.php/1508/9intReport_Egenhofer_1992.pdf, 1990.

Fisher, R. A.: Statistical methods for research worker, First Edit., edited by F. A. E. Crew and D. Ward Cutler,

Oliver & Boyd, Edinburg., 1925.

Freeman, B., Boult, P. J., Yielding, G. and Menpes, S.: Using empirical geological rules to reduce structural uncertainty in seismic interpretation of faults, J. Struct. Geol., 32(11), 1668–1676, doi:10.1016/j.jsg.2009.11.001, 2010.

5   Giraud, J.: Synthetic geophysical survey using geological modelling from the Yerrida Basin (Western Australia), , doi:10.5281/zenodo.3522841, 2019.

Giraud, J., Pakyuz-Charrier, E., Jessell, M., Lindsay, M., Martin, R. and Ogarko, V.: Uncertainty reduction through geologically conditioned petrophysical constraints in joint inversion, GEOPHYSICS, 82(6), ID19–ID34, doi:10.1190/geo2016-0615.1, 2017.

10  Giraud, J., Pakyuz-Charrier, E., Ogarko, V., Jessell, M., Lindsay, M. and Martin, R.: Impact of uncertain geology in constrained geophysical inversion, ASEG Ext. Abstr., 2018(1), 1, doi:10.1071/ASEG2018abM1_2F, 2018.

Giraud, J., Lindsay, M., Ogarko, V., Jessell, M., Martin, R. and Pakyuz-Charrier, E.: Integration of geoscientific uncertainty into geophysical inversion by means of local gradient regularization, Solid Earth, 10(1), 193–210, doi:10.5194/se-10-193-2019, 2019a.

15  Giraud, J., Ogarko, V., Lindsay, M., Pakyuz-Charrier, E., Jessell, M. and Martin, R.: Sensitivity of constrained joint inversions to geological and petrophysical input data uncertainties with posterior geological analysis, Geophys. J. Int., 218(1), 666–688, doi:10.1093/gji/ggz152, 2019b.

Godefroy, G., Caumon, G., Laurent, G. and Bonneau, F.: Structural Interpretation of Sparse Fault Data Using Graph Theory and Geological Rules, Math. Geosci., doi:10.1007/s11004-019-09800-0, 2019.

20  Godsil, C. and Royle, G.: Algebraic Graph Theory, pp. 1–18. [online] Available from: http://link.springer.com/10.1007/978-1-4613-0163-9_1, 2001.

Goodfellow, I., Bengio, Y. and Courville, A.: Deep Learning, MIT Press, Cambridge, MA, UWA. [online] Available from: http://www.deeplearningbook.org, 2016.

Green, C. H.: INTEGRATION IN EXPLORATION, GEOPHYSICS, 13(3), 365–370, doi:10.1190/1.1437404, 25  1948.

Hansen, P. C. and Johnston, P. R.: The L-Curve and its Use in the Numerical Treatment of Inverse Problems, in Computational Inverse Problems in Electrocardiography, pp. 119–142. [online] Available from: file://g/Publications/PDFs/hansen2001_Lcurve.pdf, 2001.

Hansen, P. C. and O'Leary, D. P.: The Use of the L-Curve in the Regularization of Discrete Ill-Posed Problems, 30  SIAM J. Sci. Comput., 14(6), 1487–1503, doi:10.1137/0914086, 1993.

Jessell, M., Pakyuz-charrier, E., Lindsay, M., Giraud, J. and de Kemp, E.: Assessing and Mitigating Uncertainty in Three-Dimensional Geologic Models in Contrasting Geologic Scenarios, , (21), 63–74, doi:10.5382/SP.21.04, 2018.

Jupp, D. L. B. and Vozoff, K.: Joint inversion of geophysical data, Geophys. J. R. Astron. Soc., 42, 977–991,

1975.

Kalteh, A. M., Hjorth, P. and Berndtsson, R.: Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application, Environ. Model. Softw., 23(7), 835–845, doi:10.1016/j.envsoft.2007.10.001, 2008.

5    Klose, C. D.: Self-organizing maps for geoscientific data analysis: geological interpretation of multidimensional geophysical data, Comput. Geosci., 10(3), 265–277, doi:10.1007/s10596-006-9022-x, 2006.

Köhler, A., Ohrnberger, M. and Scherbaum, F.: Unsupervised pattern recognition in continuous seismic wavefield records using Self-Organizing Maps, Geophys. J. Int., 182(3), 1619–1630, doi:10.1111/j.1365-246X.2010.04709.x, 2010.

10    Kohonen, T.: Analysis of a simple self-organizing process, Biol. Cybern., 44(2), 135–140, doi:10.1007/BF00317973, 1982a.

Kohonen, T.: Self-organized formation of topologically correct feature maps, Biol. Cybern., 43(1), 59–69, doi:10.1007/BF00337288, 1982b.

Kohonen, T.: The self-organizing map, Neurocomputing, 21(1–3), 1–6, doi:10.1016/S0925-2312(98)00030-7, 15    1998.

Kohonen, T.: Essentials of the self-organizing map, Neural Networks, 37, 52–65, doi:10.1016/j.neunet.2012.09.018, 2013.

de la Varga, M., Schaaf, A. and Wellmann, F.: GemPy 1.0: open-source stochastic geological modeling and inversion, Geosci. Model Dev. Discuss., (March), 1–50, doi:10.5194/gmd-2018-61, 2018.

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

Li, Y. and Oldenburg, D. W.: Incorporating geological dip information into geophysical inversions, GEOPHYSICS, 65(1), 148–157, doi:10.1190/1.1444705, 2000.

Li, Y., Melo, A., Martinez, C. and Sun, J.: Geology differentiation: A new frontier in quantitative geophysical 25    interpretation in mineral exploration, Lead. Edge, 38(1), 60–66, doi:10.1190/tle38010060.1, 2019.

Lindsay, M., Occhipinti, S., Ramos, L., Aitken, A. and Hilliard, P.: An integrated view of the Yerrida Basin with implications for its architecture and mineral prospectivity, Crawley., 2018.

Lines, L. R., Schultz, A. K. and Treitel, S.: Cooperative inversion of geophysical data, GEOPHYSICS, 53(1), 8–20, doi:10.1190/1.1442403, 1988.

30    Maag, E. and Li, Y.: Discrete-valued gravity inversion using the guided fuzzy c -means clustering technique, GEOPHYSICS, 83(4), G59–G77, doi:10.1190/geo2017-0594.1, 2018.

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

Martin, R., Ogarko, V., Komatitsch, D. and Jessell, M.: Parallel three-dimensional electrical capacitance data imaging using a nonlinear inversion algorithm and Lp norm-based model, Measurement, 128(May), 428–445, doi:10.1016/j.measurement.2018.05.099, 2018.

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

Melo, A. T., Sun, J. and Li, Y.: Geophysical inversions applied to 3D geology characterization of an iron oxide copper-gold deposit in Brazil, GEOPHYSICS, 82(5), K1–K13, doi:10.1190/geo2016-0490.1, 2017.

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

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

Muñoz, G. and Rath, V.: Beyond smooth inversion: the use of nullspace projection for the exploration of non-uniqueness in MT, Geophys. J. Int., 164(2), 301–311, doi:10.1111/j.1365-246X.2005.02825.x, 2006.

Nettleton, L. L.: GEOPHYSICS, GEOLOGY AND OIL FINDING, GEOPHYSICS, 14(3), 273–289, doi:10.1190/1.1437535, 1949.

Occhipinti, S., Hocking, R., Lindsay, M., Aitken, A., Copp, I., Jones, J., Sheppard, S., Pirajno, F. and Metelka, V.: Paleoproterozoic basin development on the northern Yilgarn Craton, Western Australia, Precambrian Res., 300, doi:10.1016/j.precamres.2017.08.003, 2017.

Paasche, H. and Tronicke, J.: Cooperative inversion of 2D geophysical data sets: A zonal approach based on fuzzy c-means cluster analysis, Geophysics, 72(3), A35–A39, doi:10.1190/1.2670341, 2007.

Paasche, H., Tronicke, J. and Dietrich, P.: Automated integration of partially colocated models: Subsurface zonation using a modified fuzzy c-means cluster analysis algorithm, GEOPHYSICS, 75(3), P11–P22, doi:10.1190/1.3374411, 2010.

Pakyuz-Charrier, E., Giraud, J., Lindsay, M. and Jessell, M.: Common Uncertainty Research Explorer Uncertainty Estimation in Geological 3D Modelling, ASEG Ext. Abstr., 2018(1), 1, doi:10.1071/ASEG2018abW10_2D, 2018a.

Pakyuz-Charrier, E., Giraud, J., Ogarko, V., Lindsay, M. and Jessell, M.: Drillhole uncertainty propagation for three-dimensional geological modeling using Monte Carlo, Tectonophysics, doi:10.1016/j.tecto.2018.09.005, 2018b.

Pakyuz-Charrier, E., Lindsay, M., Ogarko, V., Giraud, J. and Jessell, M.: 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(2), 385–402, doi:10.5194/se-9-385-2018, 2018c.

Pakyuz-Charrier, E., Jessell, M., Giraud, J., Lindsay, M. and Ogarko, V.: Topological analysis in Monte Carlo simulation for uncertainty propagation, Solid Earth, 10(5), 1663–1684, doi:10.5194/se-10-1663-2019, 2019.

Pellerin, J., Botella, A., Bonneau, F., Mazuyer, A., Chauvin, B., Lévy, B. and Caumon, G.: RINGMesh: A programming library for developing mesh-based geomodeling applications, Comput. Geosci., 104(March), 93–100, doi:10.1016/j.cageo.2017.03.005, 2017.

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

5   Australia, Perth. [online] Available from: https://catalogue.nla.gov.au/Record/524116, 2000.

Roden, R., Smith, T. and Sacrey, D.: Geologic pattern recognition from seismic attributes: Principal component analysis and self-organizing maps, Interpretation, 3(4), SAE59–SAE83, doi:10.1190/INT-2015-0037.1, 2015.

Santos, E. T. F. and Bassrei, A.: L- and Θ-curve approaches for the selection of regularization parameter in geophysical diffraction tomography, Comput. Geosci., 33(5), 618–629, doi:10.1016/j.cageo.2006.08.013, 2007.

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

Shannon, C. E. E.: A Mathematical Theory of Communication, Bell Syst. Tech. J., 27(3), 379–423, doi:10.1002/j.1538-7305.1948.tb01338.x, 1948.

15   Simonoff, J. S.: Analyzing Categorical Data, Springer New York, New York, NY., 2003.

Singh, A. and Sharma, S. P.: Identification of different geologic units using fuzzy constrained resistivity tomography, J. Appl. Geophys., 148, 127–138, doi:10.1016/j.jappgeo.2017.11.014, 2018.

Sivia, D. S. and Skilling, J.: Data Analysis: A Bayesian Tutorial, 2nd Editio., Oxford University Press., 2006.

Stavrakoudis, D., Dragozi, E., Gitas, I. and Karydas, C.: Decision Fusion Based on Hyperspectral and

20   Multispectral Satellite Imagery for Accurate Forest Species Mapping, Remote Sens., 6(8), 6897–6928, doi:10.3390/rs6086897, 2014.

Sun, J. and Li, Y.: Multidomain petrophysically constrained inversion and geology differentiation using guided fuzzy c-means clustering, Geophysics, 80(4), ID1–ID18, doi:10.1190/geo2014-0049.1, 2015.

Sun, J. and Li, Y.: Joint-clustering inversion of gravity and magnetic data applied to the imaging of a gabbro

25   intrusion, in SEG Technical Program Expanded Abstracts 2016, pp. 2175–2179, Society of Exploration Geophysicists., 2016.

Sun, J. and Li, Y.: Magnetization clustering inversion Part II: Assessing the uncertainty of recovered magnetization directions, GEOPHYSICS, 1–86, doi:10.1190/geo2018-0480.1, 2019.

Tarabalka, Y., Benediktsson, J. A. and Chanussot, J.: Spectral–Spatial Classification of Hyperspectral Imagery

30   Based on Partitional Clustering Techniques, IEEE Trans. Geosci. Remote Sens., 47(8), 2973–2987, doi:10.1109/TGRS.2009.2016214, 2009.

Thiele, S. T., Jessell, M. W., Lindsay, M., Ogarko, V., Wellmann, J. F. and Pakyuz-Charrier, E.: The topology of geology 1: Topological analysis, J. Struct. Geol., 91, 27–38, doi:10.1016/j.jsg.2016.08.009, 2016.

Towles, H. C.: A STUDY IN INTEGRATION OF GEOLOGY AND GEOPHYSICS, GEOPHYSICS, 17(4), 876–899, doi:10.1190/1.1437821, 1952.

Uriarte, E. A. and Martín, F. D.: Topology Preservation in SOM, Int. J. Appl. Math. Comput. Sci., 2005.

Vatanen, T., Osmala, M., Raiko, T., Lagus, K., Sysi-Aho, M., Orešič, M., Honkela, T. and Lähdesmäki, H.: Self-
5  organization and missing values in SOM and GTM, Neurocomputing, 147(1), 60–70, doi:10.1016/j.neucom.2014.02.061, 2015.

Vesanto, J. and Alhoniemi, E.: Clustering of the self-organizing map, IEEE Trans. Neural Networks, 11(3), 586–600, doi:10.1109/72.846731, 2000.

Ward, W. O. C., Wilkinson, P. B., Chambers, J. E., Oxby, L. S. and Bai, L.: Distribution-based fuzzy clustering
10  of electrical resistivity tomography images for interface detection, Geophys. J. Int., 197(1), 310–321, doi:10.1093/gji/ggu006, 2014.

Wegener, A.: Die Entstehung der Kontinente und Ozeane, Braunschwe., 1915.

Wellmann, J. F. and Regenauer-Lieb, K.: Uncertainties have a meaning: Information entropy as a quality measure for 3-D geological models, Tectonophysics, 526–529, 207–216, doi:10.1016/j.tecto.2011.05.001, 2012.

15  Wellmann, J. F., Horowitz, F. G., Schill, E. and Regenauer-Lieb, K.: Towards incorporating uncertainty of structural data in 3D geological inversion, Tectonophysics, 490(3–4), 141–151, doi:10.1016/j.tecto.2010.04.022, 2010.

Wilson, E. B.: Probable Inference, the Law of Succession, and Statistical Inference, J. Am. Stat. Assoc., doi:10.1080/01621459.1927.10502953, 1927.

20  Wrona, T., Pan, I., Gawthorpe, R. L. and Fossen, H.: Seismic facies analysis using machine learning, GEOPHYSICS, 83(5), O83–O95, doi:10.1190/geo2017-0595.1, 2018.

Zhang, G., Wang, Z. and Chen, Y.: Deep learning for Seismic Lithology Prediction, Geophys. J. Int., 1368–1387, doi:10.1093/gji/ggy344, 2018.

Zhao, T., Jayaram, V., Roy, A. and Marfurt, K. J.: A comparison of classification techniques for seismic facies
25  recognition, Interpretation, 3(4), SAE29–SAE58, doi:10.1190/INT-2015-0044.1, 2015.

Zhao, T., Li, F. and Marfurt, K. J.: Constraining self-organizing map facies analysis with stratigraphy: An approach to increase the credibility in automatic seismic facies classification, Interpretation, 5(2), T163–T171, doi:10.1190/int-2016-0132.1, 2017.

Zlatanova, S.: On 3D topological relationships, in Proceedings - International Workshop on Database and Expert
30  Systems Applications, DEXA, vol. 2000-Janua, pp. 913–919., 2000.