Solid Earth, 11, 419–436, 2020
https://doi.org/10.5194/se-11-419-2020

Solid Earth Open Access EGU

Towards plausible lithological classification from geophysical inversion: honouring geological principles in subsurface imaging

Jérémie Giraud[1], **Mark Lindsay**[1], **Mark Jessell**[1], **and Vitaliy Ogarko**[2,3]

[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, University of Western Australia,
7 Fairway, 6009 Crawley, Australia
[3]ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D)

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

Received: 31 October 2019 – Discussion started: 11 November 2019
Revised: 13 February 2020 – Accepted: 17 February 2020 – Published: 31 March 2020

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 (SOMs) 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-regularization". 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 with 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 needs 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, 1920; 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 fundamental complementarity between geology and geophysics in modelling the same object (the Earth) has been recognized as one of the prerequisites to exploration success as early as the 1940s 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; Giraud et al., 2017). In contrast, the recovery of geological quantities from geophysical inversion has seen much less effort. Recent studies have started to rectify this by proposing the idea of lithological differentiation of inversion results (Paasche et al., 2010; Sun and Li, 2015; Paasche and Tronicke, 2007), which consists 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 (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 by the need for quantitative interpretation methods in the geosciences and by the "renaissance" phase machine learning went through after 2006 (Goodfellow et al., 2016, chap. 5). Once recovered, spatial facies distribution can be used for geological interpretation and downstream decision making. 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., 2019; Pakyuz-Charrier et al., 2018a, b, c).

Lithologies (or facies) can be characterized by a broad range of rock properties that are the result of geological processes, such as weathering, compaction, metamorphism and deformation. These physical processes are usually nonlinear, 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 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 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 or lithologies in three dimensions give little to no consideration of geological information and rules for their classification.

To complement existing methodologies, we propose a solution that partially 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 inversion results, we perform lithological classification applying plausibility filters relying on geological principles, which we refer to as "geological post-regularization". The application of geological post-regularization is to reduce the non-geological character of models obtained through classification. After classification, we calculate the frequentist probability of the different lithologies (i.e. apparition frequency relative to all lithologies) associated with each unit of the self-organizing map (SOM) and report it in each model cell discretizing the studied area for interpretation.

The methodology we propose can serve 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 inversion workflows by providing a general, automated method to derive a non-deterministic lithological interpretation of 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 work, lithologies are identified though a classification technique relying on a simple artificial neural network. We chose SOMs (Kohonen, 1982a, b), 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 SOMs 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 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 only. We obtain, for each model cell, a suite of frequentist probabilities for each lithology observed in the area. In both cases, geological post-regularization 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 (Sect. 2.1) used to obtain the models that are used as input for classification using SOM (Sect. 2.2). Post-regularization as applied to such classified lithologies (Sect. 2.3) and the related uncertainty analysis in terms of prediction accuracy and geophysical consistency is then detailed (Sect. 2.4). Following this, Sect. 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 the post-regularization operator in the area are introduced (Sect. 3.1). The classification of results from geological and geological modelling and post-regularization is then presented alongside the related uncertainty analysis, supporting a potential re-interpretation of the geological model of the area (Sect. 3.2). The discussion and conclusion sections follow and complete this contribution.

2 Methodology for geoscientific modelling and classification

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

2.1 Geophysical and geological modelling

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

$$\theta(d,m) = \|\mathbf{W}_\mathrm{d}(d - g(m))\|_2^2 + \alpha_\mathrm{m}\left\|\mathbf{W}_\mathrm{m}\left(m - m_\mathrm{p}\right)\right\|_2^2 + \alpha_\mathrm{s}\|\mathbf{W}_\mathrm{s}\nabla m\|_2^2, \quad (1)$$

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

The matrix $\mathbf{W}_\mathrm{s}$ is calculated following Giraud et al. (2019a), who use the probabilistic geological modelling approach described in Pakyuz-Charrier et al. (2018b, c, 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 $\mathbf{G}$ with model m. Therefore, we have $g(m) = \mathbf{G}m$.

Geological uncertainty is estimated from probabilistic geological modelling. During this process, an ensemble of geological models is generated using the Monte Carlo uncertainty estimator (MCUE) of Pakyuz-Charrier et al. (2018a, b, c, 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. Shannon's entropy, which can be used as a proxy for geological uncertainty, indicates how well geological information constrains the model locally. It can be used to constrain inversion in a structural sense when integrated in $\mathbf{W}_\mathrm{s}$ as per Eq. (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, 2018a, 2019a, b).

2.2 Classification using SOM

The SOM artificial neural network relies on 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; Martin and Obermayer, 2009). In essence, it consists in the projection of the SOM's latent space onto a manifold of superior dimension (i.e. our dataset). This map, which can be 2-D or 3-D, 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 approximation is optimal.

In this study, we follow common practice by training two-dimensional (2-D) maps using a hexagonal lattice topology and applying a Gaussian-shaped neighbourhood function. We chose to use a 2-D 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.

Ideally, the SOM should be trained in a controlled environment where all the variables used are perfectly known, which

motivates the utilization of synthetic geophysical data. We calculate such data from a geological structural framework derived from real-world field geological and petrophysical field measurement data in the same fashion as for a geophysical feasibility study. 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 $\mathbf{W}_{\text{s}}$;
4. spatial gradient in the inverted model $\|\nabla m_{\text{inv}}\|_2$; and
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 $x \epsilon R^{n_{\text{v}}}$, with a number of variables $n_{\text{v}} = 5$, where $x(5)$ 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. 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 3-D 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:

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

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 inputs 1 through 4 listed above are obtained from previous modelling and lithologies are the quantity sought for. In our case, the utilization 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 (Eq. 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-regularization

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

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 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 subhorizontal 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) and 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 post-regularization, 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 (2-D) geological mapping (Tarabalka et al., 2009; Stavrakoudis et al., 2014; Cracknell and Reading, 2015).

2.3.2 Implementation

The post-regularization scheme we develop for the recovery of lithologies in 3-D relies on two hypotheses. Firstly, we assume that the presence of isolated lithologies contradicts the geological principle of continuity. Although such post-regularization has been used mostly in 2-D or shallow 3-D, there is no theoretical obstacle to the extension of

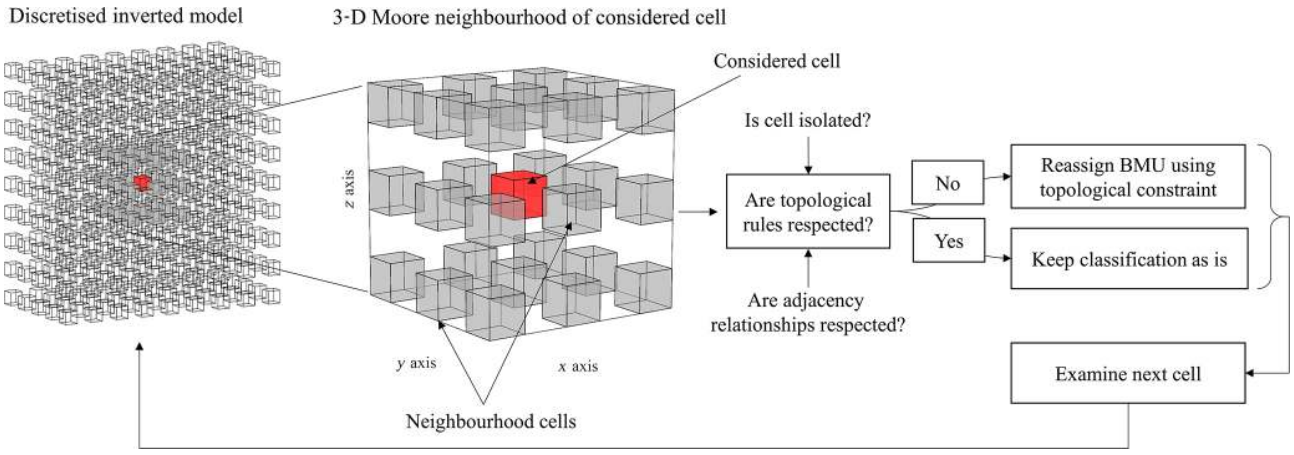

Figure 1. Summary of topological filtering used during post-regularization.

this methodology to the purely 3-D classification case we present here. Secondly, we introduce the utilization of adjacency relationships between the different lithologies in post-regularization 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). Here, we extend existing post-regularization 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; Thiele et al., 2016, for the different topologies) defined by geological principles.

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

$$\begin{aligned}\mathrm{BMU} &= \mathrm{BMU}_k, \text{ where } k \\ &= \operatorname{argmin}\left(\|u_k - \mathrm{BMU}\| \,|\, \text{conditions}\right), \end{aligned} \tag{3}$$

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

$$\begin{aligned}\text{conditions} = \Big\{ &\exists\, c\ U \,|\, c = \mathrm{BMU}_k(5) \cap 1^T \mathbf{M}^k \cdot \mathbf{M}^f 1 \\ &= 0_{n_1,n_1} \cap \text{stratigraphy} \Big\}, \end{aligned} \tag{4}$$

where 1 is $n_1 \times 1$ column vector of ones, and $0_{n_1,n_1}$ is the $n_1 \times n_1$ matrix of zeros; $\mathbf{M}^f$ and $\mathbf{M}^k$ are adjacency matrices, and $\mathbf{M}^k \cdot \mathbf{M}^f$ is the Hadamard product of $\mathbf{M}^k$ and $\mathbf{M}^f$, such that $\left(\mathbf{M}^k \cdot \mathbf{M}^f\right)_{ij} = \left(\mathbf{M}^k\right)_{ij}\left(\mathbf{M}^f\right)_{ij}$, and $0_{n_1,n_1}$ is the zero matrix of dimension $n_1 \times n_1$; $\mathbf{M}^f$ encapsulates geological knowledge and principles about contacts between lithologies; $\mathbf{M}^k$ contains the adjacency relationships between the considered cell and its neighbourhood U. The derivation of $\mathbf{M}^f$ and $\mathbf{M}^k$ is detailed below.

The first part of the condition in Eq. (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 3-D 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 the 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-regularization is illustrated in Fig. 1.

We point out that in contrast to Tarabalka et al. (2009) and Stavrakoudis et al. (2014), who used the first and second Chamfer neighbourhoods in 2-D around the considered model cell, we do not follow the same approach in 3-D. Our implementation of the extension of their approach to 3-D 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 utilization of a mode filter with a $3 \times 3 \times 3$ kernel.

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 topo-

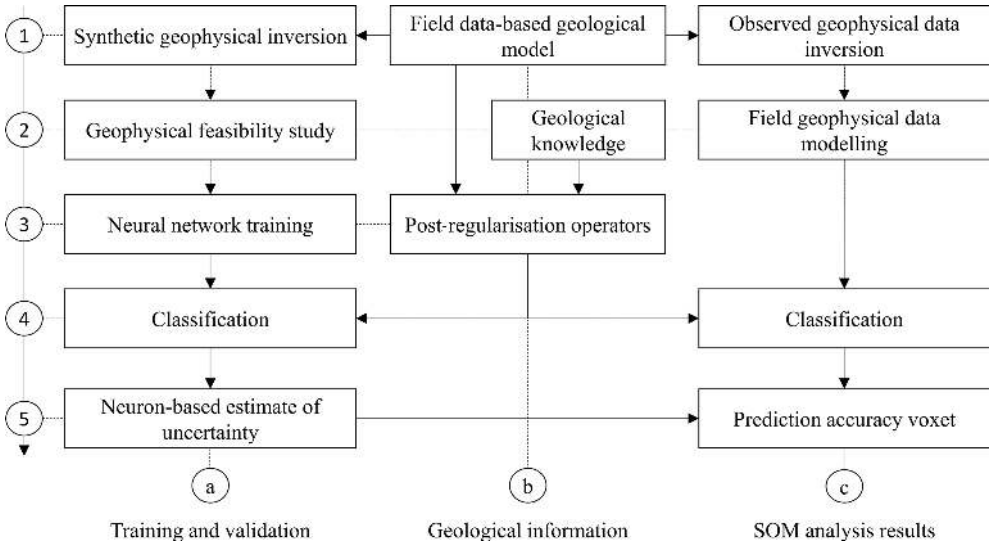

Figure 2. Schematic summary of proposed methodology.

logical 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

$$M_{i>j,j} = n_{ij}, M_{i\leq j,j} = 0, \tag{5}$$

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

$$\begin{cases} M^f_{i>j,j} = 1 & \text{if contact between } i \text{ and} \\ & j \text{ contradicts geology.} \\ M^f_{i>j,j} = 0 & \text{otherwise.} \end{cases} \tag{6}$$

From there, it is straightforward to identify occurrences of forbidden contacts by calculating $\mathbf{M}_k \cdot \mathbf{M}^f$. Therefore, $1^{\mathbf{T}}\mathbf{M}_k \cdot \mathbf{M}^f 1 = 0$ (Eq. 4) indicates that no contact violating the condition imposed by $\mathbf{M}^f$ is observed. The last condition in Eq. (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".

After identification of configurations forbidden by the conditions set in Eq. (4), its BMU is updated using the closest neuron honouring the set of conditions (Fig. 2, box 4a).

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 (Fig. 2, box 5a). 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 correctly 20 % of the time (20 % accuracy). This process is described below.

The methodology is summarized in Fig. 2.

2.4 Uncertainty analysis

2.4.1 Prediction accuracy of the recovered lithologies

The prediction accuracy τ_i of lithology $i \epsilon [1, \ldots, n_l]$ (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 $\mathbf{M}^c$. We remind that the "matching matrix" (or "confusion matrix" in supervised learning) is a matrical 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 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 Eq. (7), it appears that τ_i is equivalent to the frequentist probability of the ith lithology over the entire SOM. When considering a specific node j, it becomes the relative frequentist probability of the lithology i for cells classified as having the jth node as their BMU, noted τ_{ij}.

2.4.2 Geophysical consistency

The consistency of the classification performed using SOM after application of post-regularization with field geophysical measurements might be altered by both the classification

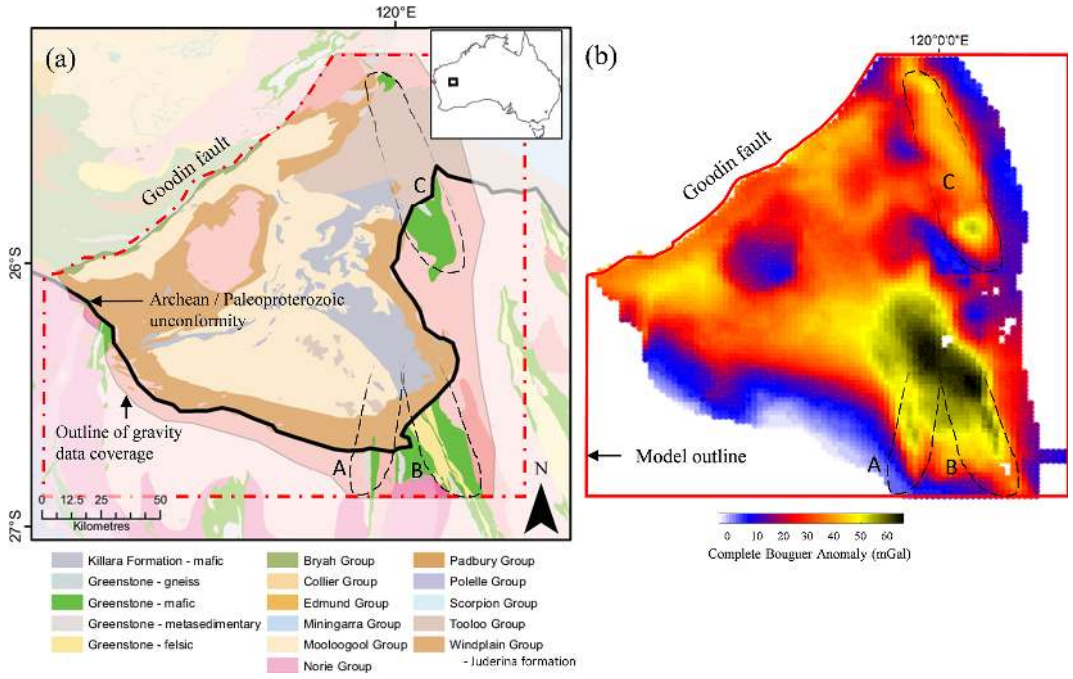

Figure 3. (a) Geological map of the area and **(b)** complete Bouguer anomaly (reproduced from Giraud et al., 2019a). The dashed red line outlines the modelled area. Capital letters "A", "B" and "C" symbolize the possible outlines for the greenstone belts in the area.

itself and by post-regularization. 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 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}, $\Delta m = m_{\mathrm{inv}} - m_{\mathrm{SOM}}$, satisfies the following condition:

$$\begin{aligned}&\|\mathbf{W}_{\mathrm{d}}\left((d - g\left(m_{\mathrm{inv}}\right)) - \left(d - g\left(m_{\mathrm{SOM}}\right)\right)\right)\|_2^2 \\ &\quad = \|\mathbf{W}_{\mathrm{d}}\mathbf{G}\Delta m\|_2^2 \leq \mathrm{tol}, \end{aligned} \tag{9}$$

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 Eq. (9) is that the difference Δm 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 Eq. (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 utilization of Eq. (9) with the visual comparison of data misfit maps corresponding to the model corresponding to m_{inv} and m_{SOM}.

3 Application case: Yerrida Basin

3.1 Survey setting

This subsection introduces and summarizes the geological and geophysical context of the application case presented here. More details about the geology of the area and the initial geophysical inversion 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 150 km N–S and 180 km E–W (Pirajno and Adamides, 2000) (Fig. 3a). The structures of interest in this work are Archean greenstone belts (Fig. 3), as they are prospective for Au and Ni and underlie the younger basin rocks. The 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 approximately 2200 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.

The density contrast of the lithologies observed in the area ranges between 0 and 330 kg m^{-3}, making it appropriate for

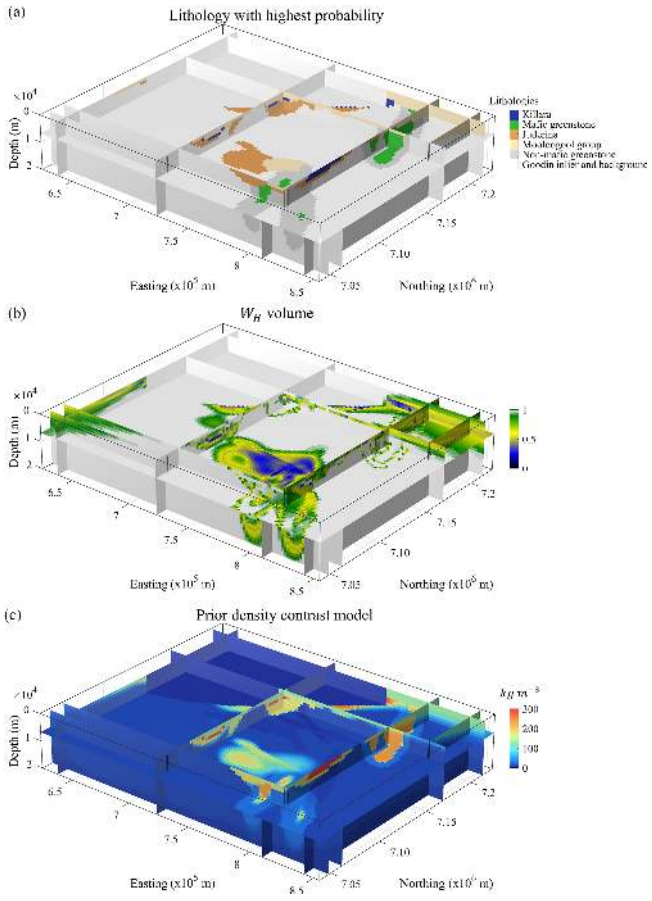

Figure 4. Prior geological modelling. Preferred lithology volume **(a)**, geological uncertainty volume **(b)** and starting density contrast model **(c)**. Modified from Giraud et al. (2019a).

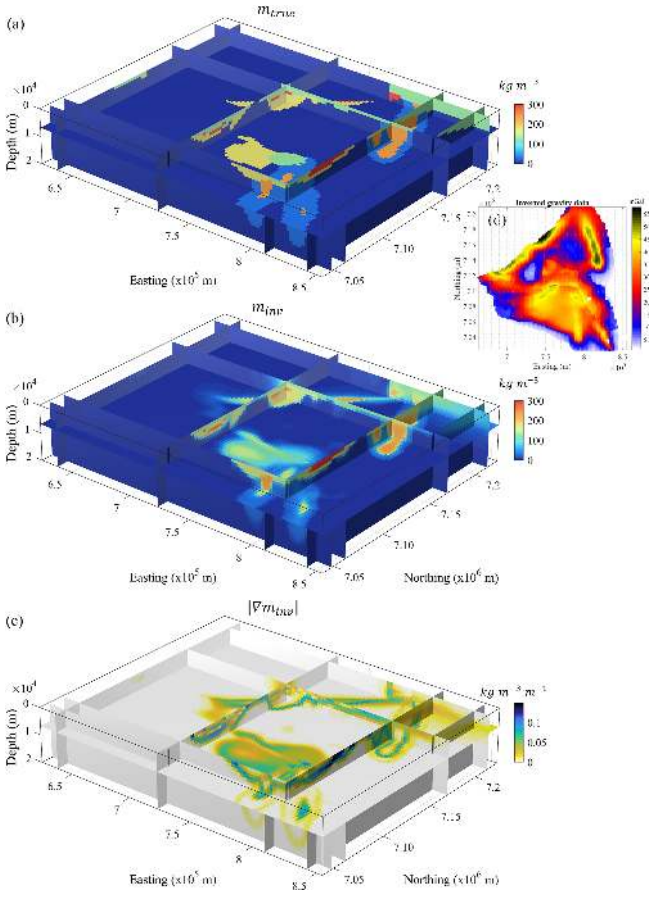

Figure 5. True density contrast model for calculation of synthetic geophysical data **(a)**, inverted model obtained from geophysical inversion of synthetic geophysical data **(b)**, corresponding spatial gradient of density contrast **(c)** and synthetic geophysical data **(d)**.

gravity modelling and inversion (Giraud et al., 2019a; Lindsay et al., 2018). While basin rocks exhibit some density contrast, the greenstone is conspicuous in gravity data with a density contrast expected to lie between 190 and 270 kg m^{-3}, making it 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 (Fig. 3b) is comprised of a total of 4882 measurement points. The model is discretized into $100 \times 100 \times 42$ cells of dimensions $2.335\,\text{km} \times 1.875\,\text{km} \times 1.0475\,\text{km}$ down to approximately 44 km 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 subsection introduces the semi-synthetic survey we performed for the training of SOM.

The volumes of most probable lithology, W_S and the starting model are shown in Fig. 4. Volumes shown in Fig. 4a, b and c are used for the training and validation dataset for SOM training as explained in Sect. 2.2.

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

3.1.3 Field geophysical data inversion

The density contrast model obtained from inversion of field geophysical data and its gradients are shown in Fig. 6a and b. Visual comparison of inverted models shown in Figs. 6a

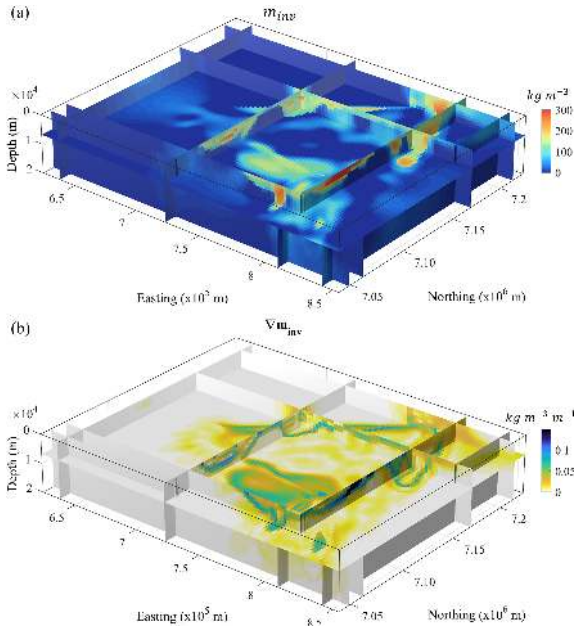

Figure 6. Inverted model obtained from geophysical inversion of field geophysical data **(a)** and corresponding spatial gradient of density contrast **(b)**.

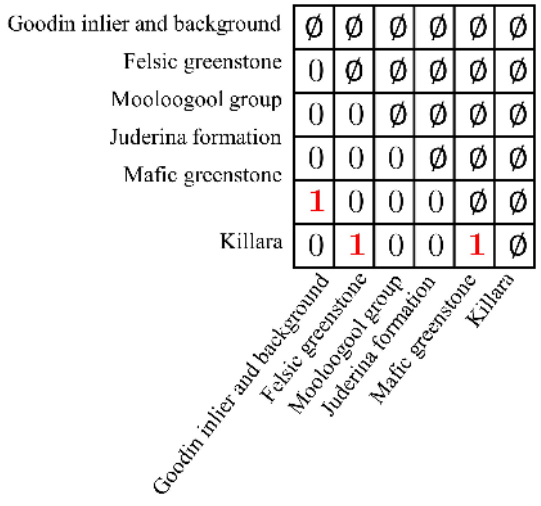

Figure 7. Matrix defining forbidden contact between lithologies in the Yerrida Basin. Here, 1 means that two units may be in contact with each other, 0 means that they may not, and φ represents symmetric relationships or when the same unit is adjacent to itself (which geologically may occur across a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

and 5b reveals that mesoscale structures are similar with the exception of large structures presenting low-density contrasts at depth (darker shades of blue in Fig. 6a). This is reflected in Fig. 6b, which exhibits low gradient values in these areas.

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

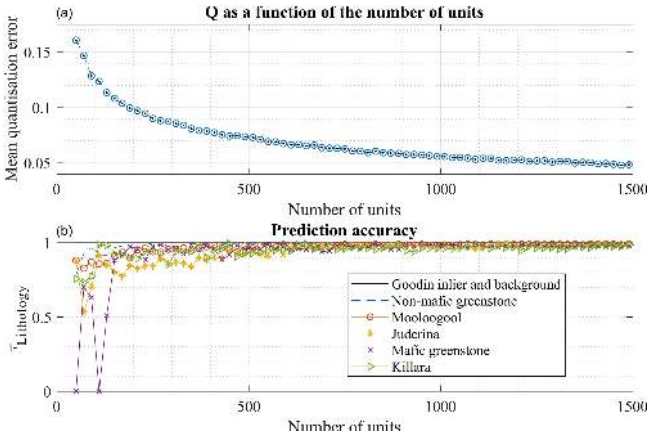

Figure 8. (a) Elbow curve of the quantization error for the determination of the optimum number of neurons (or units) in SOM and **(b)** prediction accuracy for the different lithologies present in the training and validation datasets. Note that after 750 units, the quantization error for the different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

3.1.4 Geological rules for post-regularization

As mentioned above, for clarity in this demonstration, only adjacency relationships are considered. The utilization of simple relationships such as adjacency is also chosen 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 suggest that the Killara Formation is a volcanic unit that is restricted to the Yerrida Basin and thus not in contact with the mafic greenstone. In addition, we assume that the mafic greenstone cannot be in contact with the Goodin Inlier and background (or basement), as the mafic greenstone is modelled to be enveloped by the felsic component of the greenstone.

The matrix $\mathbf{M}^f$ defining the contacts forbidden by geology as described above is given in Fig. 7.

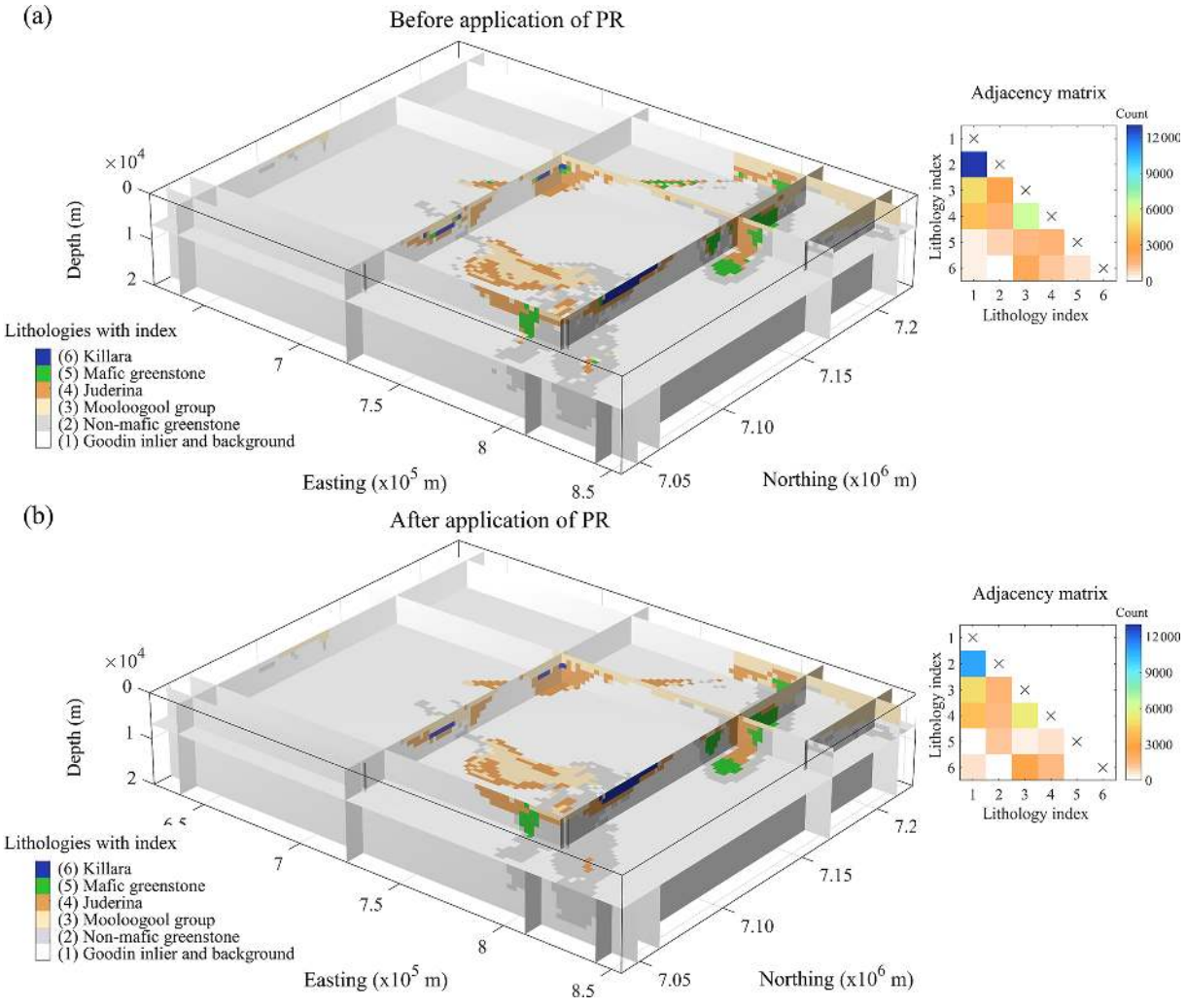

Figure 9. Recovered lithologies before post-regularization **(a)** and after **(b)** next to the corresponding adjacency matrix (right-hand side). The indices used to designate lithologies are indicated in the legend of the volume (left-hand side).

3.2 Geologically constrained SOM classification and uncertainty analysis

3.2.1 Training the neural network

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 Fig. 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 chosen number of units is corroborated by the lithology prediction accuracy (Fig. 8) as approximating the point of diminishing returns, i.e. the number of nodes beyond which additional nodes are becoming nearly redundant.

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 Fig. 8, where all lithologies are recovered with a prediction accuracy superior to 90 %.

3.2.2 Classification and post-regularization

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

As can be inferred from the adjacency matrices plotted in Fig. 9, the number of contacts between units with indices 1 and 2, respectively, is reduced by the application of post-regularization. One reason for this is the presence of a number of inclusions of lithology 1 in lithology 2, and vice versa; a total of 2561 such inclusions was identified. Overall, the number of contacts between lithologies 5 and 2 increased

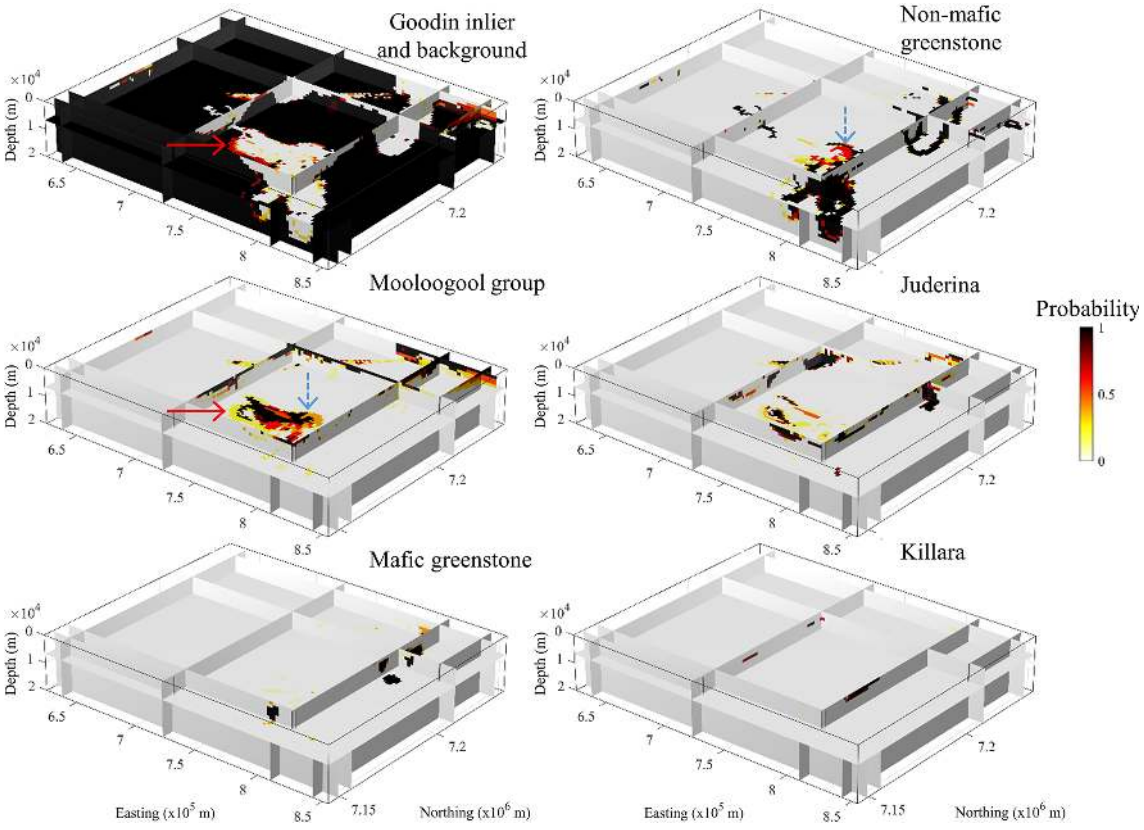

Figure 10. Frequentist probability volumes of recovered lithologies calculated as per Eq. (3). The arrows are drawn to support interpretation.

slightly due to post-regularization because a large percentage of contacts between lithologies 5 and 1 has been reassigned as contacts between lithologies 5 and 2. The elimination of contacts between lithologies 5 and 6 is also visible in Fig. 9.

3.2.3 Estimated confidence in recovered lithologies

For each node of the SOM, we calculate the prediction accuracy τ_i (Eq. 3) for the different lithologies observed in the area using the cross-validation dataset. After application of the trained SOM for the classification of inversion results obtained from the inversion of field geophysical data, we obtain a frequentist probability volume for each lithology. Figure 10 shows the resulting frequentist probability volumes for the six lithologies present in the Yerrida Basin.

Figure 10 exhibits probabilities between 0.3 and 0.6 in the area marked by the two arrows. This suggests that these zones are the least well constrained. Note that from Fig. 10, we can interpret the presence of mafic greenstone with confidence, as it shows high frequentist probability nearly everywhere classification suggests its presence. For completeness, assessment of the prediction accuracy of the different lithologies is shown by the corresponding box plot in Appendix B.

3.2.4 Geophysical null space validation and implications for geological interpretation

Applying Eq. (9), we obtain $\| W_d \mathbf{G} \Delta m \|_2^2 = 6 \times 10^{-4}$, indicating that the model can be considered geophysically equivalent overall. This is illustrated by the map of the corresponding data misfits (see Fig. B1 in Appendix B), which indicates that we can consider the recovered lithological model after application of post-regularization as reflective of both geophysical and geological information. Focusing on the mafic greenstone belts of interest in the area (Fig. 11), the classification results allow us to propose the following geological interpretations.

Figure 11 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 the thinning of mafic greenstone belt C could be attributed to faulting, folding or the topography of the palaeoenviron-

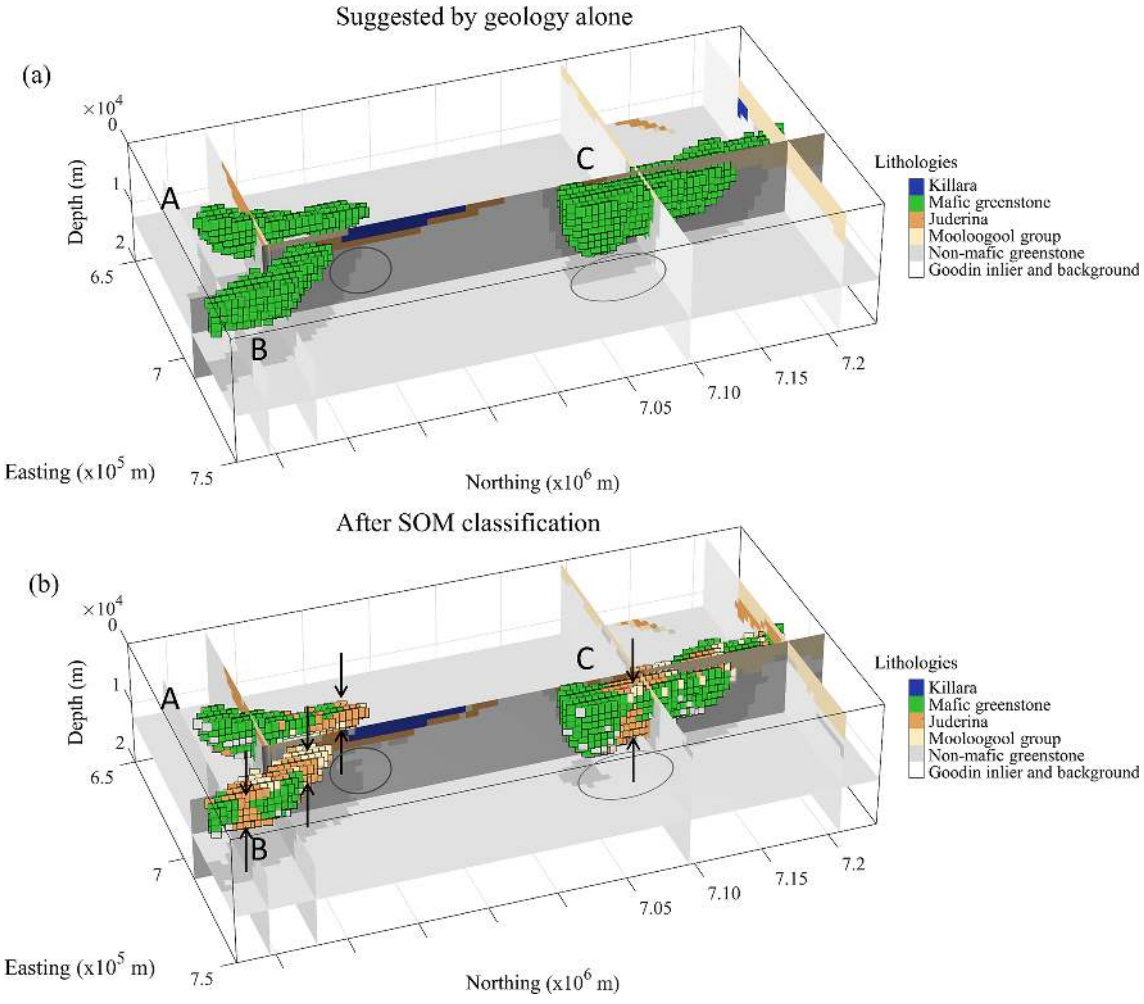

Figure 11. Mafic greenstone belts and their surroundings following geological modelling only **(a)** and after SOM classification **(b)**. The cells shows in panels **(a)** and **(b)** have the same geographical location and are coloured according to lithology. The vertical arrows show areas where mafic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

ment 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 plausible reason is 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 Fig. 11b by the arrows around greenstone C, hints at a possible false assumption that the Merrie Greenstone Belt is a single and coherent geological body.

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 stand-alone 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 (Paasche and Tronicke, 2007; Carter-McAuslan et al., 2015; Sun and Li, 2015, 2016; Maag and Li, 2018; Ward et al., 2014; 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 utilization of post-regularization can be effective for increasing 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-regularization operator apply to lithological topology and focus on the adjacency relation-

ship between cells. Ideally, post-regularization should also consider the surface area of contacts and their topology. This could be followed by, for instance, a 3-D extension of the geological model-editing approach of Anquez et al. (2019) to produce genuine geological models honouring age relationships, stratigraphic principles, etc. Provided that the resulting models honour Eq. (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-regularization can be successfully applied to other clustering techniques. In addition, the implementation of post-regularization 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 W_s (Eq. 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 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 (Fig. 11). While this issue remains unresolved, the capability of 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 frequentist probability volume associated with it. By selecting models belonging to the null space of the geophysical data (i.e. satisfying Eq. 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 frequentist probability voxet thereby remediating to some of the limitations of deterministic inversion. The proposed technique utilizes a post-regularization 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. Results allowed us to confirm previous results and bring new insights into possible reinterpretation of the geometry of prospective greenstone belts.

Appendix A: Data misfit generated by SOM classification and post-regularization

In Fig. A1, data misfit and the absolute data misfit difference (Fig. A1b and c, respectively) show values which, in places, are relatively high but which are in line with the gravity data inversions. Figure A1a and b show similar features to the exception of a patch in the central part of the model (northing $\sim 7.09 \times 10^6$ m – easting $\sim 7.7 \times 10^5$ m), where Fig. A1b shows values that are approximately 1.5 mGal higher. Figure A1c shows misfit differences generally on the order of, or lower than, 2 mGal. Also note that there are places where Fig. A1b shows lower misfit than Fig. A1a.

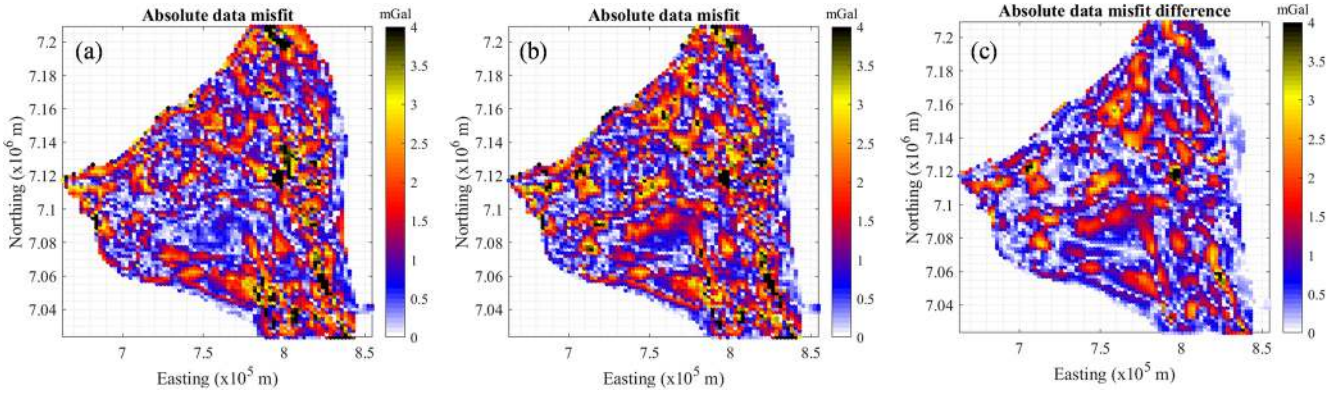

Figure A1. **(a)** Comparison of absolute data misfit maps obtained for the inverse model used here (reproduced from Giraud et al., 2019a) and **(b)** after post-regularization, with **(c)** the misfit differences between panels **(a)** and **(b)**.

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

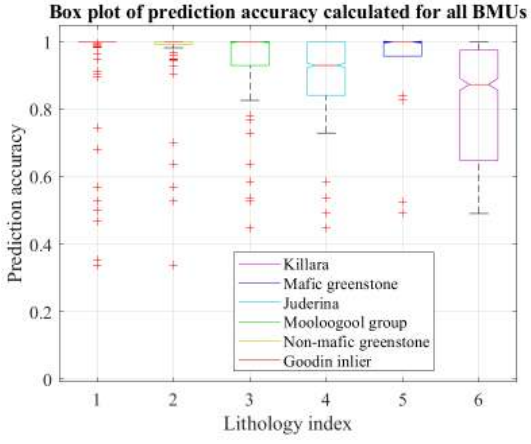

Figure B1. Box plot of prediction accuracies for the different lithologies. Red crosses mark outliers.

Data availability. The input and output of the synthetic survey are made available at https://doi.org/10.5281/zenodo.3522841 by Giraud (2019). The field data from the Yerrida Basin are made available at https://doi.org/10.5281/zenodo.3522841 (Giraud et al., 2018b).

Author contributions. JG designed the methodology, adapted the SOM algorithm and performed all modelling except geological modelling. MG is the main writer of the manuscript, which was redacted with support from the rest of the authors. ML performed geological modelling and interpretation of the recovered lithologies. MJ provided guidance and supervision while the project was being carried out. VO 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 3D (ASTRO 3D) 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 Jessell was supported by a Western Australian fellowship. Mark Lindsay was supported by the Geological Survey of Western Australia and the Exploration Incentive Scheme and the Australia Research Council (grant no. 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. The 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.

Review statement. This paper was edited by Michal Malinowski and reviewed by Tom Horrocks and one anonymous referee.

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., 5, 155–161, https://doi.org/10.4236/apm.2015.54019, 2015.

Anquez, P., Pellerin, J., Irakarama, M., Cupillard, P., Lévy, B., and Caumon, G.: Automatic correction and simplification of geological maps and cross-sections for numerical simulations, C. R. Geosci., 351, 48–58, https://doi.org/10.1016/j.crte.2018.12.001, 2019.

Bauer, K., Schulze, A., Ryberg, T., Sobolev, S. V., and Weber, M. H.: Classification of lithology from seismic tomography: A case study from the Messum igneous complex, Namibia, J. Geophys. Res.-Sol. Ea., 108, 1–15, https://doi.org/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, 984–998, https://doi.org/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, https://doi.org/10.1155/2012/678326, 2012.

Burns, K.: Lithologic topology and structural vector fields applied to subsurface prediction in geology, in: Proceedings of GIS/LIS 88, ACSM-ASPRS, San Antonio, 25–34, 1988.

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

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, W1–W15, https://doi.org/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, 223–229, https://doi.org/10.1016/S0098-3004(01)00067-X, 2002.

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, WA85–WA95, https://doi.org/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, Society of Exploration Geophysicists, 2056–2060, 2018.

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, 8, 1–14, https://doi.org/10.1109/JSTARS.2014.2382760, 2015.

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

de la Varga, M., Schaaf, A., and Wellmann, F.: GemPy 1.0: open-source stochastic geological modeling and inversion, Geosci. Model Dev., 12, 1–32, https://doi.org/10.5194/gmd-12-1-2019, 2019.

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, https://doi.org/10.1016/j.jappgeo.2014.11.007, 2015.

Eckhardt, E. A.: Partnership between geology and geophysics in prospecting for oil, Geophysics, 5, 209–214, https://doi.org/10.1190/1.1441804, 1940.

Egenhofer, M. and Herring, J.: Categorizing binary topological relations between regions, lines, and points in geographic databases, The, 1–28, available at: https://pdfs.semanticscholar.org/b303/39af3f0be6074f7e6ac0263e9ab34eb84271.pdf (last access: 18 March 2020), 1990.

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, 1668–1676, https://doi.org/10.1016/j.jsg.2009.11.001, 2010.

Giraud, J.: Synthetic geophysical survey using geological modelling from the Yerrida Basin (Western Australia), https://doi.org/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, ID19–ID34, https://doi.org/10.1190/geo2016-0615.1, 2017.

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, 1–6, https://doi.org/10.1071/ASEG2018abM1_2F, 2018a.

Giraud, J., Lindsay, M., and Ogarko, V.: Yerrida Basin Geophysical Modeling – Input data and inverted models, https://doi.org/10.5281/zenodo.1238216, 2018b.

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, 193–210, https://doi.org/10.5194/se-10-193-2019, 2019a.

Giraud, J. 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, 666–688, https://doi.org/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., 51, 1091–1107, https://doi.org/10.1007/s11004-019-09800-0, 2019.

Godsil, C. and Royle, G.: Algebraic Graph Theory, 1–18, available at: http://link.springer.com/10.1007/978-1-4613-0163-9_1 (last access: 18 March 2020), 2001.

Goodfellow, I., Bengio, Y., and Courville, A.: Deep Learning, MIT Press, Cambridge, MA, UWA, available at: http://www.deeplearningbook.org (last access: 18 March 2020), 2016.

Green, C. H.: Integration in exploration, Geophysics, 13, 365–370, https://doi.org/10.1190/1.1437404, 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, 119–142, available at: https://www.sintef.no/globalassets/project/evitameeting/2005/lcurve.pdf (last access: 18 March 2020), 2001.

Hansen, P. C. and O'Leary, D. P.: The Use of the L-Curve in the Regularization of Discrete Ill-Posed Problems, SIAM J. Sci. Comput., 14, 1487–1503, https://doi.org/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, chap. 4, in: Metals, Minerals, and Society, 63–74, https://doi.org/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, 835–845, https://doi.org/10.1016/j.envsoft.2007.10.001, 2008.

Klose, C. D.: Self-organizing maps for geoscientific data analysis: geological interpretation of multidimensional geophysical data, Comput. Geosci., 10, 265–277, https://doi.org/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, 1619–1630, https://doi.org/10.1111/j.1365-246X.2010.04709.x, 2010.

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

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

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

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

Lelièvre, P. G. and Farquharson, C. G.: Integrated Imaging for Mineral Exploration, in: Integrated Imaging of the Earth: Theory and Applications, American Geophysical union, 137–166, 2016.

Li, Y. and Oldenburg, D. W.: Incorporating geological dip information into geophysical inversions, Geophysics, 65, 148–157, https://doi.org/10.1190/1.1444705, 2000.

Li, Y., Melo, A., Martinez, C., and Sun, J.: Geology differentiation: A new frontier in quantitative geophysical interpretation in mineral exploration, Lead. Edge, 38, 60–66, https://doi.org/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.

Lindsay, M., Occhipinti, S., Laflamme, C., Aitken, A., and Ramos, L.: Mapping undercover: integrated geoscientific interpretation and 3D modelling of a Proterozoic basin, Solid Earth Discuss., https://doi.org/10.5194/se-2019-192, in review, 2020.

Lines, L. R., Schultz, A. K., and Treitel, S.: Cooperative inversion of geophysical data, Geophysics, 53, 8–20, https://doi.org/10.1190/1.1442403, 1988.

Maag, E. and Li, Y.: Discrete-valued gravity inversion using the guided fuzzy c – means clustering technique, Geophysics, 83, G59–G77, https://doi.org/10.1190/geo2017-0594.1, 2018.

Martin, R. and Obermayer, K.: Self-Organizing Maps, in Encyclopedia of Neuroscience, Elsevier, 551–560, 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, 428–445, https://doi.org/10.1016/j.measurement.2018.05.099, 2018.

Meju, M. A. and Gallardo, L. A.: Structural Coupling Approaches in Integrated Geophysical Imaging, American Geophysical union, 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, K1–K13, https://doi.org/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), IEEE, Opatija, Croatia, 1061–1066, 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, American Geophysical union, 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, 301–311, https://doi.org/10.1111/j.1365-246X.2005.02825.x, 2006.

Nettleton, L. L.: Geophysics, geology and oil finding, Geophysics, 14, 273–289, https://doi.org/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, 121–140, https://doi.org/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, A35–A39, https://doi.org/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, P11–P22, https://doi.org/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, https://doi.org/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, 16–39, https://doi.org/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, 385–402, https://doi.org/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, 1663–1684, https://doi.org/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, 93–100, https://doi.org/10.1016/j.cageo.2017.03.005, 2017.

Pirajno, F. and Adamides, N. G.: Geology and Mineralization of the Palaeoproterozoic Yerrida Basin, Western Australia, Perth, available at: https://catalogue.nla.gov.au/Record/524116 (last access: 18 March 2020), 2000.

Roden, R., Smith, T., and Sacrey, D.: Geologic pattern recognition from seismic attributes: Principal component analysis and self-organizing maps, Interpretation, 3, SAE59–SAE83, https://doi.org/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, 618–629, https://doi.org/10.1016/j.cageo.2006.08.013, 2007.

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), Springer, Palma de Mallorca, 549–563, https://doi.org/10.1007/978-3-319-19222-2_46, 2015.

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

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

Stavrakoudis, D., Dragozi, E., Gitas, I., and Karydas, C.: Decision Fusion Based on Hyperspectral and Multispectral Satellite Imagery for Accurate Forest Species Mapping, Remote Sens., 6, 6897–6928, https://doi.org/10.3390/rs6086897, 2014.

Sun, J. and Li, Y.: Multidomain petrophysically constrained inversion and geology differentiation using guided fuzzy c-means clustering, Geophysics, 80, ID1–ID18, https://doi.org/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 intrusion, in SEG Technical Program Expanded Abstracts 2016, Society of Exploration Geophysicists, 2175–2179, 2016.

Sun, J. and Li, Y.: Magnetization clustering inversion Part II: Assessing the uncertainty of recovered magnetization directions, Geophysics, 8, 1–86, https://doi.org/10.1190/geo2018-0480.1, 2019.

Tarabalka, Y., Benediktsson, J. A., and Chanussot, J.: Spectral–Spatial Classification of Hyperspectral Imagery Based on Partitional Clustering Techniques, IEEE Trans. Geosci. Remote, 47, 2973–2987, https://doi.org/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, https://doi.org/10.1016/j.jsg.2016.08.009, 2016.

Towles, H. C.: A study in integration of geology and geophysics, Geophysics, 17, 876–899, https://doi.org/10.1190/1.1437821, 1952.

Uriarte, E. A. and Martín, F. D.: Topology Preservation in SOM, Int. J. Appl. Math. Comput. Sci., 1, 19–22, 2005.

van der Baan, M. and Jutten, C.: Neural networks in geophysical applications, Geophysics, 65, 1032–1047, https://doi.org/10.1190/1.1444797, 2000.

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

Vesanto, J. and Alhoniemi, E.: Clustering of the self-organizing map, IEEE Trans. Neural Networks, 11, 586–600, https://doi.org/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 of electrical resistivity tomography images for interface detection, Geophys. J. Int., 197, 310–321, https://doi.org/10.1093/gji/ggu006, 2014.

Wegener, A.: Die Entstehung der Kontinente und Ozeane, Braunschweig, Braunschweig, 1920.

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, https://doi.org/10.1016/j.tecto.2011.05.001, 2012.

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

Wrona, T., Pan, I., Gawthorpe, R. L., and Fossen, H.: Seismic facies analysis using machine learning, Geophysics, 83, O83–O95, https://doi.org/10.1190/geo2017-0595.1, 2018.

Zhang, G., Wang, Z., and Chen, Y.: Deep learning for Seismic Lithology Prediction, Geophys. J. Int., 215, 1368–1387, https://doi.org/10.1093/gji/ggy344, 2018.

Zhao, T., Jayaram, V., Roy, A., and Marfurt, K. J.: A comparison of classification techniques for seismic facies recognition, Interpretation, 3, SAE29–SAE58, https://doi.org/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, T163–T171, https://doi.org/10.1190/int-2016-0132.1, 2017.

Zlatanova, S.: On 3D topological relationships, in Proceedings 11th International Workshop on Database and Expert Systems Applications, IEEE Comput. Soc., 2000, 913–919, 2000.