Geological Classification from Geophysical models within a PGI Framework

EOAS 555B: Course Project

Johnathan Kuttai
11074580
jkutt@eoas.ubc.ca

Abstract

The current work of applying image segmentation techniques to geological classification within the PGI (Petrophysical and Geologically Guided Inversion) framework ([1]) presents an opportunity to create more geologically meaningful models given geophysical and petrophysical information. These geologically classified models are then fed back into the inversion in the form of ran updated reference model in order to influence the inversion. Meaning, the better the classification, the better constrained the inversion will be. Currently the geological classification is done in a naive implementation of closest pixel to pixel relation. This was expanded to using a neighbourhood for encoded geological rules in the Implementing geological rules within geophysical inversion: A PGI perspective abstract ([2]). These results provide a stepping stone to further improve the mechanism for classification. From the advent of the 2017 paper: Attention is all you need ([9]), Attention or further yet the Transformer architecture has evolved from natural language processing to image segmentation applications. These architectures could greatly improve classification in the PGI framework contained in the opensource software simpeg ([4]).

1 Introduction

Imaging of the subsurface with geophysical methods is a widely used technique in industry use. It is used in variety of disciplines from mineral exploration, hydrocarbon exploration, environmental, to unexploded ordnance detection and void detection for civil engineering. Most methods are very low impact on the environment and deployable almost anywhere. The draw back is the resolution of the image. Geophysical models are often smooth and not a direct representation of the subsurface. Though useful, they are not directly related to the actual subsurface nature. This report will focus on how to improve this in exploration which attempts to image geological settings. Geophysical models are interpreted and geology is often inferred. Typically geophysical methods define targets while loosely representing structure. Recently methods to improve on the geological meaning extracted from geophysical models have been developed. For this study the Petrophysically and Geologically guided inversion developed by Astic et al. abbreviated PGI is considered. The open source software package simpeg containing the work of Astic et al. will be built upon for improved classification ([4]). PGI allows prior information such as petrophysical measurements to be included to influence the geophysical end model. Within this, a geological classification is accomplished as part of the 3 steps in a PGI inversion. This is done via the smallness term in a typical Tikhonov inversion objective function ([1]).

Provided this conduit, the focus the study will be on the geological classification component. Here there are avenues to improve how the classification is done. Currently classification is completed on neighbouring cells via information containing in Gaussian mixture models representing the petrophysical information. At this stage of the PGI inversion the problem can be considered an image segmentation problem. Another family of mixture models known as Gaussian mixture Markov random fields has been a popular choice in this field. This method is implemented by modifying the Gaussian mixture model as described by Nguyen et al, 2013 ([6]). Furthering the avenue's of tools for classifications are ever popular neural nets. The use and unarguable performance of the neural nets have gained much notoriety in image segmentation as of late. Specifically, the use of transformers which at its core uses a mechanism called Attention ([9]). Attention and transformers first got traction in natural language processing and then applied to vision problems. Attention is considered to give context in order to make a classification. For example in natural language processing, the prediction of a proceeding word is influenced by information given at the beginning of sentences or paragraphs for that matter. This representing contextual influence. For this feature alone, attention could prove useful to geological classification within the PGI framework.

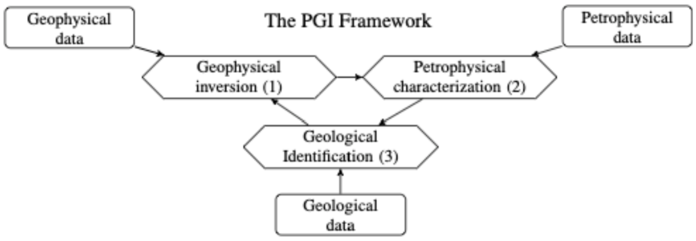

Figure 1: PGI three step framework

2 Petrophysically and Geologically Guided Inversion

To get geologically meaningful models from geophysical models there needs to be a link between the two. To get a quasi-geology model ([5]), petrophysical information is required for this connection. This is accomplished in a 3 step process withing a standard geophysical inversion **(Figure 1)**. We can represent petrophysical information in the form of a probability distribution or combination of many. This is commonly formed as a Gaussian Mixture Model ([1]). These model our petrophysical information and is interjected via the smallness term in our regularization of a typical Tikhonov geophysical inversion ([7]) described as:

$$\underset{m}{\text{minimize}}\Phi(m) = \Phi_d(m) + \beta\left(\alpha_s\Phi_s(m) + \sum_{v\in x,y,z}\alpha_v\Phi_v(m)\right) \tag{1}$$

provided $\Phi_d(m) \leq \Phi_d^*$.

The smallness term is the important term here as this is where petrophyscial information can be interjected. The smallness is defined:

$$\Phi_s(\mathbf{m}) = \frac{1}{2}||W_s(\mathbf{m} - \mathbf{m}_{ref})||_2^2 \tag{2}$$

Where W_s is the confidences in the reference model $\mathbf{m}_{ref}$ and $\mathbf{m}$ is the model vector. This penalized large differences between the the recovered model and the reference model. This reference model is where we can influence the geophysical inversion by using the geological classification to update the $\mathbf{m}_{ref}$. Objective function (1) can equivalently be expressed in a Bayesian formulation ([8]) as:

$$\mathcal{P}(\mathbf{m}|d_{obs} = \frac{\mathcal{P}(d_{obs}|\mathbf{m})\mathcal{P}(\mathbf{m})}{\mathcal{P}(d_{obs}} \propto \mathcal{P}(d_{obs}|\mathbf{m})\mathcal{P}(\mathbf{m}) \tag{3}$$

$$\mathcal{P}(\mathbf{m}) = \mathcal{P}_s(\mathbf{m})\mathcal{P}_{x,y,z}(\mathbf{m}) \tag{4}$$

This results in a model $\mathbf{m}$ that maximizes this posterior distribution is a map estimate which is the same model that minimizes the objective function in (1) ([1]). The data misfit term can be referred to as the maximum likelihood estimate. This now allows us to form a Gaussian distribution with parameters means μ and covariances Σ giving the equation [rfe actic]:

$$\mathcal{P}_s(\mathbf{m}) = \mathcal{N}(\mathbf{m}|\mathbf{m}_{ref}, (\beta\alpha_s W_s^T W_s)^{-1}) \tag{5}$$

By using the negative natural log of the distributions we get the equivalent least-squares terms in our original objective function (1), but specifically for $\mathcal{P}_s(\mathbf{m})$ we get somethings similar to (2). The PGI smallness term is then represented as:

$$\Phi_s(\mathbf{m}) = \frac{1}{2}||W_s(\Theta, \mathbf{z}^*)(\mathbf{m} - \mathbf{m}_{ref}(\Theta, \mathbf{z}^*))||_2^2 \tag{6}$$

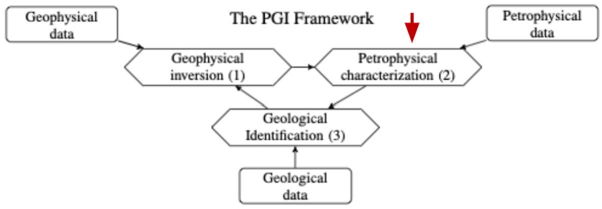

Figure 2: PGI second step: Petrophysical characterization

2.1 Petrophysical Characterization

In the previous section it was shown that petrophysical information can be interjected into the inversion using a MAP estimate (4) which is the equivalent of (2). Stepping into the second part of the PGI inversion farmework **(figure 2)**, by using a Gaussian mixture model we can represent each geological unit's physical property by fitting a Gaussian distribution ([1]). If given petrophysical data where $j = 1 \ldots c$ represents c number of geological units. For each sample a MLE of unit j will have mean μ_j, variance σ_j^2 and proportions π_j. The full probability function for some physical property sample S_i written as a GMM is defined as:

$$\mathcal{P}(S_i|\Theta) = \sum_{j=1}^{c} \pi_j \mathcal{N}(S_i|\mu_j.\sigma_j^2) \tag{7}$$

This can now be used to influence our geophysical inversion using our petrophysical information. Upon the end of each iteration, these parameters can be updated from the new "learned" physical property distributions Θ by averaging the prior information and the current iterations geophysical model parameters.

2.2 Geological classification

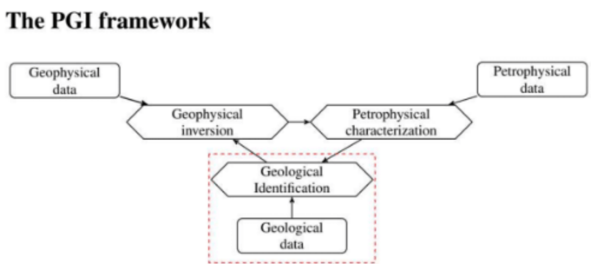

Figure 3: PGI step three: geological classification

Now, in step 3 of the PGI framework, the geological classification can be made using the prior information $\mathcal{P}(\mathbf{z})$ and the learned petrophysical distribution Θ. Using this information the geological classification influences the geophysical inversion via the reference model $\mathbf{m}_{ref}$ and weights W_s. For every pixel (or mesh cell) a geological unit is assigned. This is done by taking the value of unit with the highest probability:

$$\mathbf{z}^* = \underset{\mathbf{z}}{\text{argmax}} \mathcal{P}(\mathbf{m}|z)\mathcal{P}(z) \tag{8}$$

Then we can update the $\mathbf{m}_ref$ as function of Θ and z^*:

$$\mathbf{m}_{ref}(\Theta, \mathbf{z}^*) = \mu^* \tag{9}$$

and then W_s in a similar way:

$$W_s(\Theta, \mathbf{z}^*) = diag\left(\mathbf{w}\ o\ \frac{1}{\sigma_{\mathbf{z}*}}\right) \tag{10}$$

Both variables $W_s(\Theta, \mathbf{z}^*)$ and $\mathbf{m}_{ref}(\Theta, \mathbf{z}^*)$ give us the required elements for (6).

2.3 Simple Example with nonlinear PDE Natural Source simulation

With all the ingredients in place a simulation using PGI for a nonlinear partial differential equation problem such as natural source electromagnetics will be tested. Many examples with linear problems like potential fields exist and prove quite useful at extracting much needed geological meaning. The nonlinear problem is notorious for being difficult to stabilize but given the success of PGI a simple test case of a buried conductive sphere in a more resistive half space is considered.

First a standard Tikhonov inversion is performed as a comparison for the PGI implementation. A sphere of radius $r = 100m$ at depth of approximately 150 m and resistivity $= 1\ m$ and a background conductivity $\rho_{background} = 1$ is considered for the toy model **(figure 4)**. Topography is added to simulate a more accurate representation of acquisition terrain.

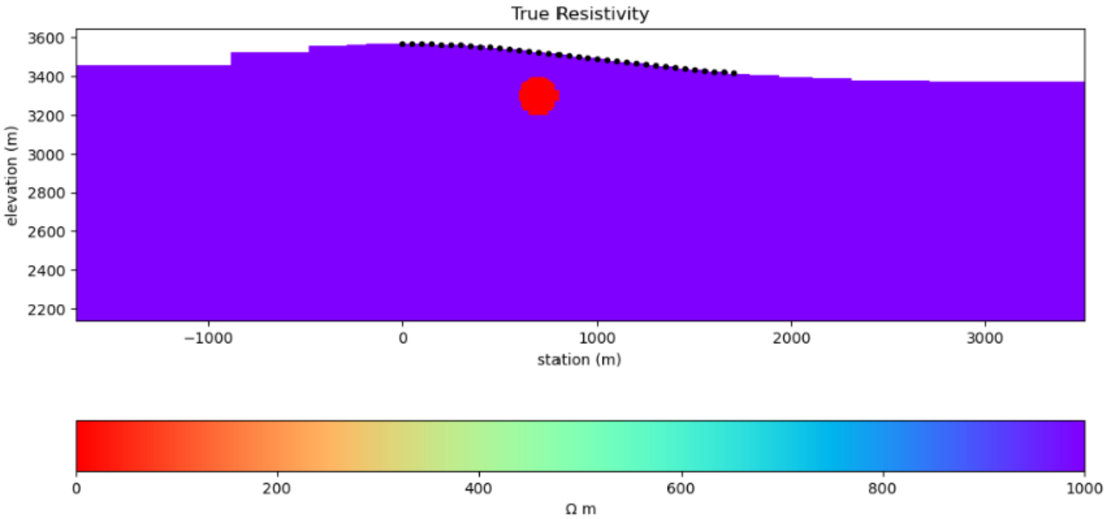

Figure 4: Simple Geologic model - Buried conductive sphere in a resistive medium

The standard Tikhonov inversion produces a very smooth model as expected. The conductive features are smeared to depth and the model looks more complex than needed **(figure 5)**. The inversion converged fully so that we have $\phi_d \leq \phi_d^*$.

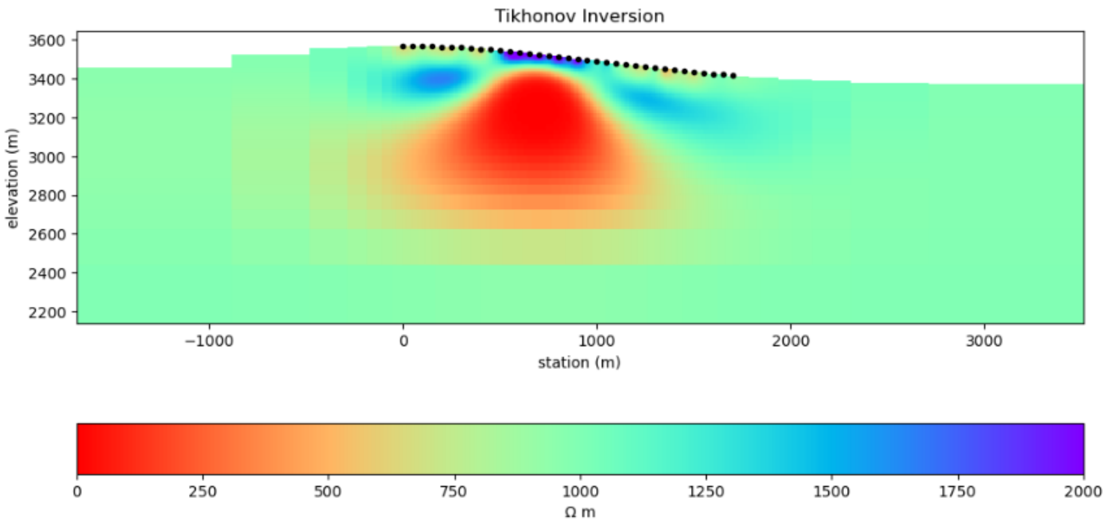

Figure 5: Plain inversion geophysical model

Next, the PGI framework is employed and we set up a GMM with two geological units with means of the value of each unit and variances 0.01. With these parameters the inversion runs runs for more iterations but the results are much more geologically meaningful **(figure 6)**. The classification under estimates the size of the sphere but not to the degree it gets exaggerated with a standard inversion.

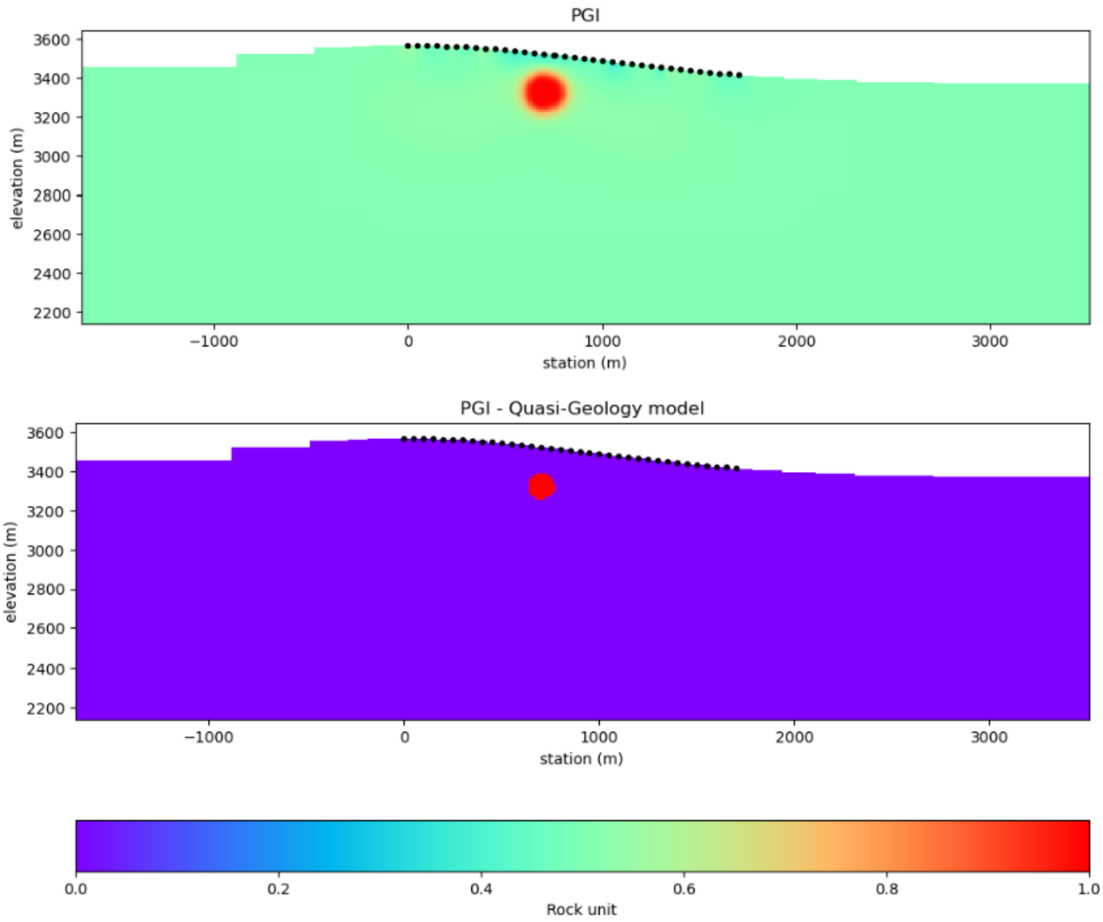

Figure 6: (top) PGI geophysical model. (bottom) PGI MT geological classification

2.4 A complex example - Direct current resistivity

In the last section the use of the geological classification greatly improves the final model created from the modified geophysical inversion. The result has real geological meaning. But what happens when we add complexity. In this test a dipole-dipole DC resistivity inversion is used to image a layered model with faulting and a dyke intrusion **(figure 7)**. Again, a standard inversion is completed resulting in the typical smoothed model. However, this time, the PGI model is not entirely geologically as meaningful as seen in the simple case. The results do discriminate a bit on the layering but appear to pull everything to surface where the highest sensitivity is **(figure 8)**.

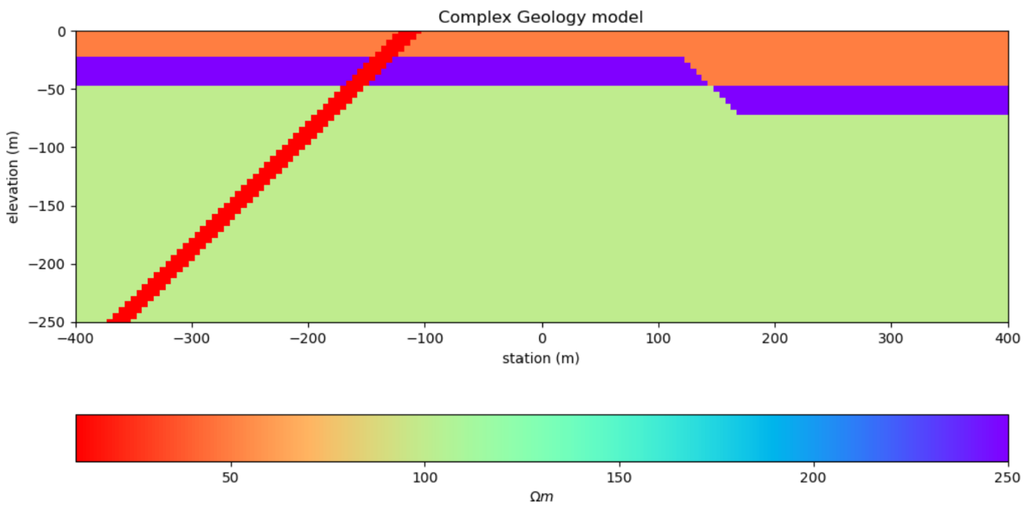

Figure 7: Complex Geology model

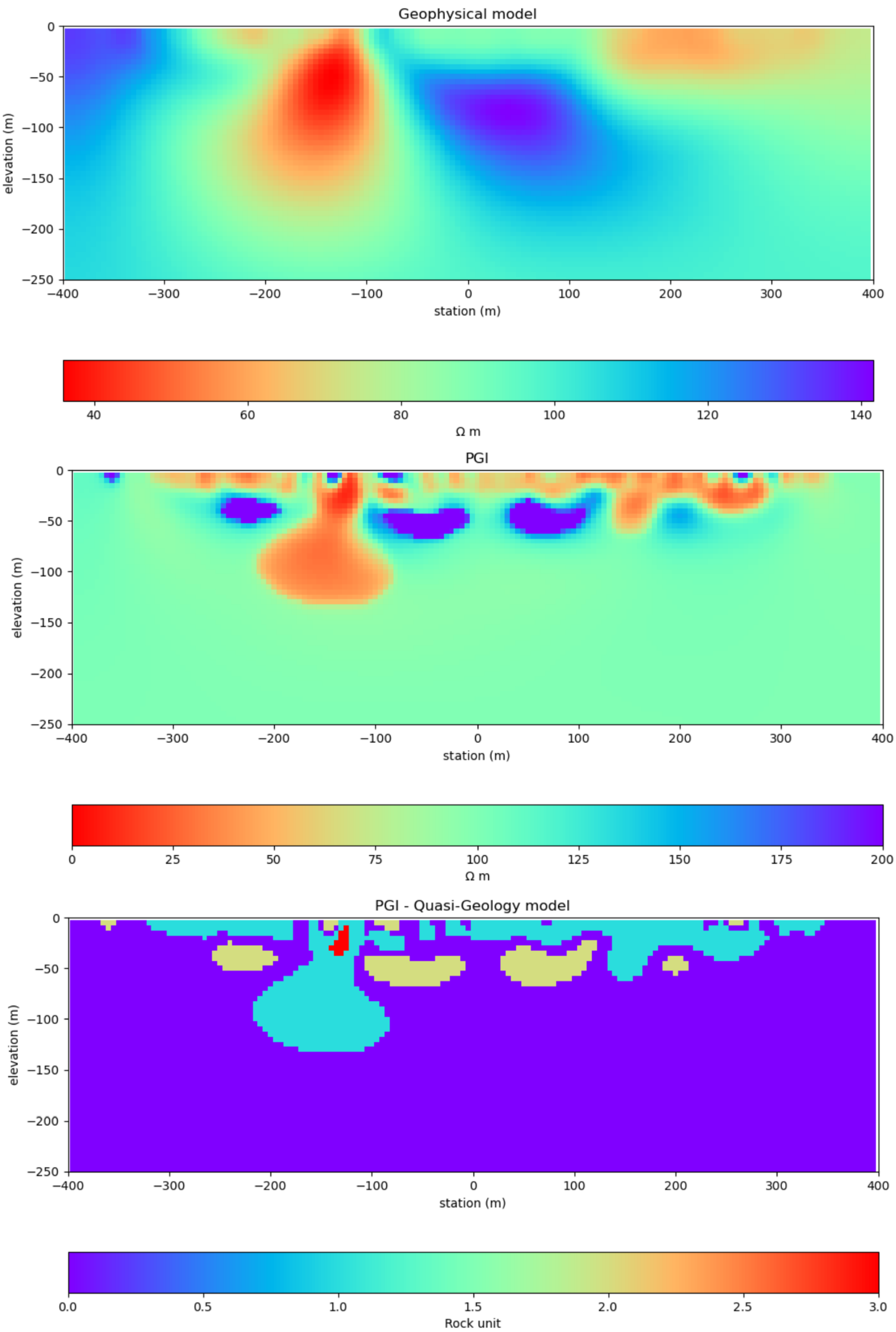

Figure 8: Plain inversion comparison with a PGI inversion

It is clear that more work is needed to stabilize the inversion in PGI's extra steps. Mostly though, a better method of classification could also help to improve our interpretation.

3 Gaussian Mixture Markov Random fields for classification

If units are not overly separated in there physical property measurements, meaning the GMM has too much of overlap, there is no way for the petrophysical characterization to discriminate between the two. Even if the properties are not overlapping, complexity is hard to capture like in our toy model with multiple layers, faulting and intrusive dyke. Gaussian mixture models lack spatial information and can be computationally expensive the more complicated they get. Here, a way to incorporate local information is based on a Markov random field ([6]). The principle here is to have the GMMRF effect the classification by iteratively learning proportions π_{ij} for our newly defined classification $\mathbf{z}$ (11) ([2]). replacing (4).

$$\mathbf{z} = \text{argmax} \prod_i \mathcal{N}(\mathbf{m}_i|\tilde{\mathbf{z}})\mathcal{P}(\Pi)_i \tag{11}$$

$$\mathcal{P}(\Pi)_i = \mathcal{P}(\tilde{\mathbf{z}}) \left(\prod_{j \in \partial_{i,\tilde{\mathbf{z}}_i}} \psi(\tilde{\mathbf{z}}_i, \tilde{\mathbf{z}}_j) \right) \tag{12}$$

where $\partial_{i,\tilde{\mathbf{z}}_i}$ defines the neighbourhood and pairwise weights of a classification cell and its neighbour $\tilde{\mathbf{z}}_j$.

3.1 GMMRF's for image segmentation

Mixture models consist of two steps when fitting and calculating parameters known as an EM algorithm; the e step and the m step ([1]). The e-step which estimated the Gaussian distribution parameters and the m-step is the maximum likelihood calculation. Incorporating the Markov random field is simple in that only the M step is modified. This is done by introducing a new factor $\mathbf{G}$ defined:

$$\mathbf{G}_{ij} = e^{\left[\frac{\beta}{2N_i} \sum_{m \in \partial_i} (z_{mj} - \pi_{mj})\right]} \tag{13}$$

Where $\mathbf{z}_{ij}$ is the posterior probability and π_{ij} are the current weights of the neighbour i. The Markov random field distribution to plug into (12) is then as follows for an iteration step t:

$$\mathcal{P}(\Pi_i) = Z^{-1} e^{\left[\frac{1}{T} \sum_{i=1} \sum_{j=1} G_{ij} log(\pi_i j))\right]} \tag{14}$$

Where T and Z are constants temperature and normalization constant. Temperature here is in regards to likelihood potential.

Using this formaulation, a comparison between the standard GMM and the newly formed GMMRF are shown in **(figure 9)** using a sample from ADE20K image set. The test is completed setting the GMM and GMMR to have 8 possible units and the GMMRF to use a neighbourhood of 4. Results between the two are comparable, though the GMMRF is less noisy and much smoother between the less distinct units.

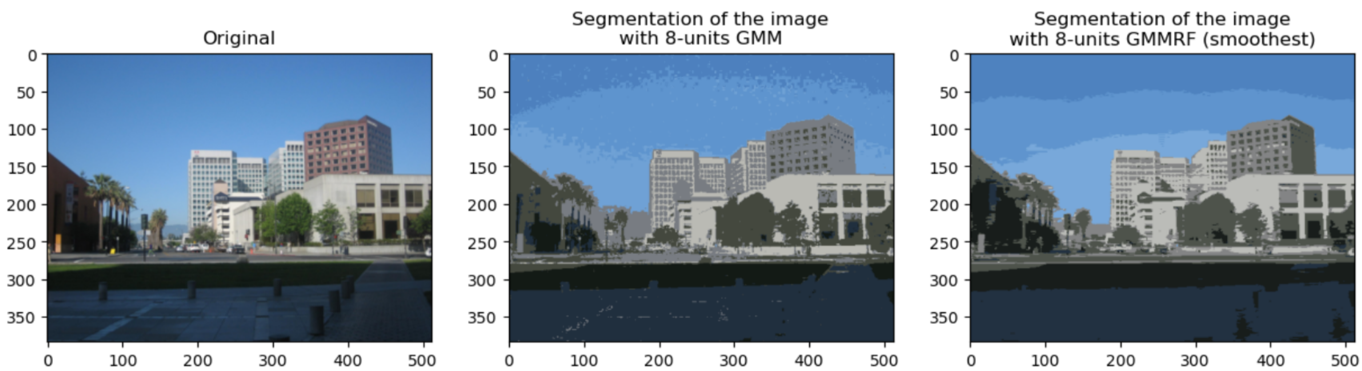

Figure 9: GMM - GMMRF comparison

4 Classification with an Attention mechanism

With the GMMRF a neighbourhood is defined giving us a way to incorporate spatial information in the maximum likelihood estimate. In a geological setting this spatial information is important, specifically in the case near contacts between separate units. Here, there is a potential for ambiguity for classification depending where the classification cell is located and its defined neighbours. Attention is at the core of many successful neural networks in natural language processing and recently in vision tasks. This mechanism allows for contextual point of view. Meaning that it can focus on important features of an input. This could be useful to capture geologic context. The attention mechanism is but a small part in a much bigger architecture know as a transformer ([9]). In this study working with attention by adding it into a simple neural network for testing its use in geological classification.

4.1 Self Attention

Commonly used is self attention([9]). It can be represented as a mapping of a query and a set of key-value pairs to an output, where the query, keys, values, and output are vectors.

For each input element $\mathbf{x}_i$:

- **Feature vectors**
  - **Query**: describes what to look for in a sequence.
  - **Key**: designed in a way to identify which elements to pay attention to given a query.
- **Query**
  - **Value**: the vector to average over.

These three matrices are created by multiplying the embedded input matrix $\mathbf{x}$ by three weight matrices created during the training process:

$$Q = W^{(Q)} \times X$$
$$K = W^{(K)} \times X$$
$$V = W^{(V)} \times X$$

$W^{(q)}, W^{(k)}$ with dimensions number of input embedding X dimension of the model and $W^{(v)}$ can vary in dimension depending on the target embedding (y) with size that is number of target embedding X dimension of the model

Lastly, the final vector is the Score function. It takes a query and a key and outputs the score & attention weight of the query-key pair.

4.2 Scaled Dot Product

The attention & score values from element **i** to **j** is based on its similarity of the query $\mathcal{Q}_i$ and key $\mathcal{K}_j$, using the dot product as the similarity metric. Can calculate the dot product attention as follows:

$$Attention_{score} = \sum_i \alpha_i * value_i = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{15}$$

Where

$$\alpha_i = Attention = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \tag{16}$$

This resulting in a matrix with score outputs for every possible pair of queries and keys.

4.3 Multi-headed Attention

Taking things a step further and doing many Self Attentions, multi headed attention gives multiple different query-key-value triplets over same features to capture different notions of similarity. Given a query, key, and value matrix, sub-queries, sub-keys, and sub-values are created and pass through the scaled dot product attention independently. Therefore given a query $\amalg \in R^{d_q}$ keys $\| \in R^{d_k}$ and values $\sqsubseteq \in R^{d_v}$ the $\rangle$'th head is:

$$\mathbf{h}_i = Attention(W_i^{(q)}q, \{W_i^{(k)}k, W_i^{(v)}v\}) \in R^{p_v}$$

where $W_i^{(q)}, W_i^{(k)}, W_i^{(v)}$ are projection matrices learnable parameters. Now stack or concatenate the heads together and project to R^{p_0}:

$$\mathbf{h} = MultiHeadAttn(q, \{k_j, v_j\}) = W_o \begin{bmatrix} h_1 \\ \vdots \\ h_h \end{bmatrix} \in R^{p_0} \tag{17}$$

where $W_o \in R^{p_0 \times hp_v}$

By creating sub queries and keys we can get more of an idea of context. We can see how queries score given different sub-set of keys and values. For example: The dog did not cross the road because it was too hot. Now what is "it" referring to, the dog or the road? Another perk of the multi-headed attention is that each head can be calculated in parallel.

4.4 Training & Prediction

To measure the usefulness of the attention mechanism a simple neural net with a single layer of multi-headed attention with 81 heads. The inputs are the model size and the output is the same but consist of scores for each model cell corresponding to 4 possible units that make up the original model. Since the samples are geophysical models a training set is quite costly. Only 12 models are considered for training which are variations of the complex model from DC example previously. With the small training set only the practicability of the mechanism is assessed. Merely seeing if at the very least a mapping can be accomplished.

The training used the Adam optimizer and a mean squared loss. It looped over the data 1000 times and the results, though noisy, do show some coherency and defined structure. This is then passed through the GMMRF and the defined neighbours use these values to create a geological classification. The results are observed in **(figure 10)**.

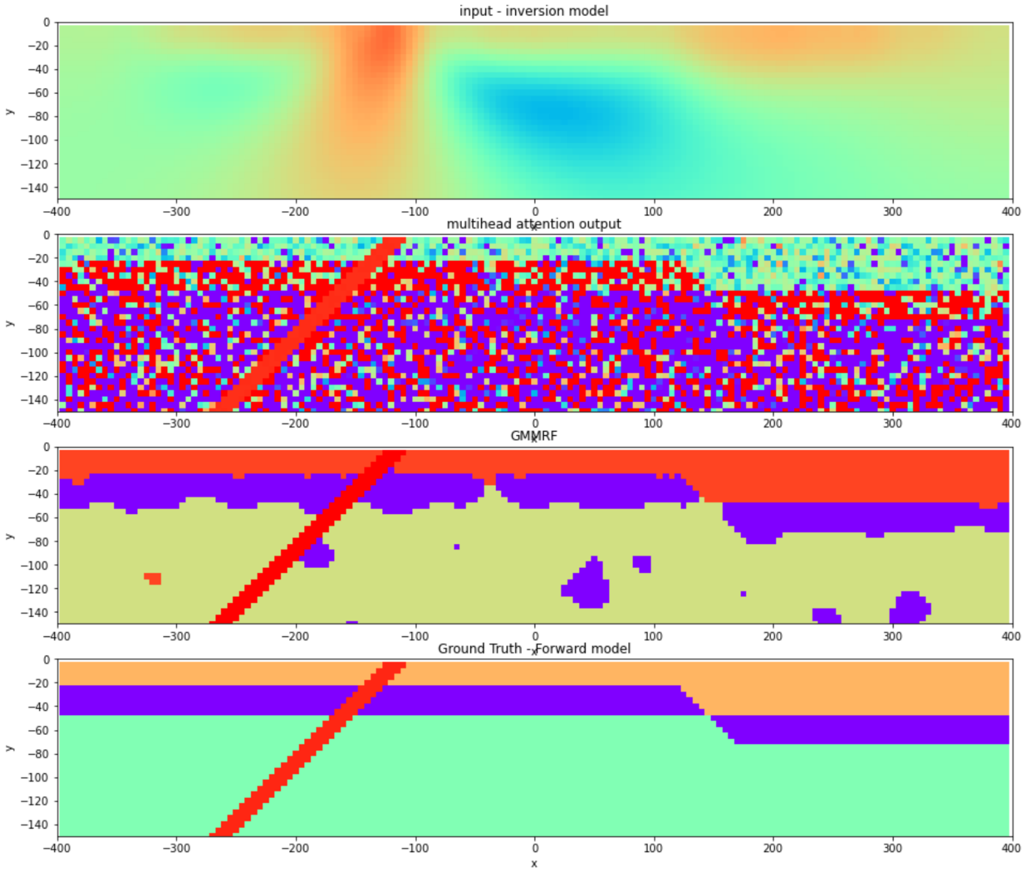

Figure 10: Classification using Multi-headed attention

Next is to work this into the PGI framework to measure the usefulness in guiding the geophysical inversion at different iterations and evolution of a model.

At this point there are some hurdles to incorporate this into the classification step. First is that not all models contain the same amount of cells. The second is the cost of creating training data and training. Lastly is at the current state a training would need to occur for every new data set that required inverting. To avoid retraining and variable amount of cells a windowing technique can be applied that has a predefined set of cells and their neighbours. For a quick brute force attempt a 9 x 9 cell neighbourhood is attempted resulting in 81 total cells for classification of a window. With this many more training samples can be extracted from a signal inversion. In this the hope is that the neural net with multi-headed attention can learn the geological context on a small scale. Unfortunately, the simple model fails at the this small scale windowing **(figure 11)**. Training for this involved 3000 samples and a learning rate of 0.001 but still the predictions are of no use. There are architectures of windowed attention that have been successful but do require the use of a full on transformer Architecture ([3]). Perhaps the full architecture is necessary to accomplish this task.

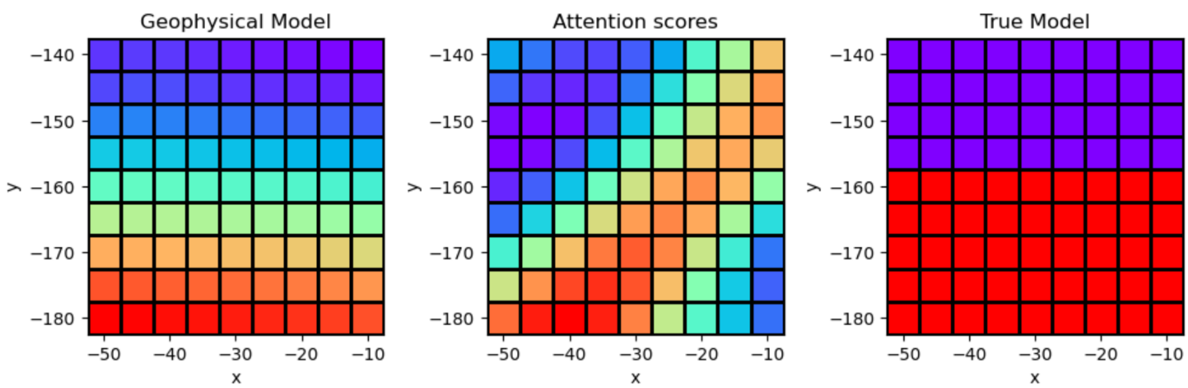

Figure 11: Classification using Multi-headed attention and smaller samples

5 Conclusion

By using the PGI framework the capabilities of making a geological classification are available and work well for simple simulations. Moving to a more complicated geological setting presents difficulties. The unsupervised approach of the mixture models integrate well into an inversion framework but require a supervised intervention with artificial neural net architecture like the transformer's attention.

Improvements can be made by using more sophisticated mixture models like the GMMRF as described in Nguyen, 2013. This allows for a defined neighbourhood to help make the classification but, again, the results from Astic et al. 2021 were only marginal. Though further study of Markov random fields will be required to better understand there capabilities. The GMMRF does provide a conduit to incorporate more advanced techniques such as attention. The attention scores can replace the weights in the neighbourhood choice defined by the GMMRF giving us some interesting preliminary results. Though as we move to a smaller windowing technique for training, the attention appears to fail.

To hopefully improve the classification, bigger windows, or even moving to a more formal transformer may be required. Also required will be to create a larger training data set as attention can require significant amount of training to be very effective. Future work on this subject will explore these ideas and investigating the architecture for neighbourhood attention which will benefit classification both in training cost for making trainable data and flexibility in geophysical model input.

References

[1] Astic, Thibaut, and Douglas W Oldenburg. 'A Framework for Petrophysically and Geologically Guided Geophysical Inversion Using a Dynamic Gaussian Mixture Model Prior'. Geophysical Journal International 219, no. 3 (1 December 2019): 1989–2012. https://doi.org/10.1093/gji/ggz389.

[2] Astic, Thibaut, Douglas W. Oldenburg, and Lindsey J. Heagy. 'Implementing Geological Rules within Geophysical Inversion: A PGI Perspective'. In First International Meeting for Applied Geoscience & Energy Expanded Abstracts, 1226–30. Denver, CO and virtual: Society of Exploration Geophysicists, 2021. https://doi.org/10.1190/segam2021-3583615.1.

[3] Hassani, Ali, Steven Walton, Jiachen Li, Shen Li, and Humphrey Shi. 'Neighborhood Attention Transformer', 2022. https://doi.org/10.48550/ARXIV.2204.07143.

[4] Heagy, Lindsey J., Rowan Cockett, Seogi Kang, Gudni K. Rosenkjaer, and Douglas W. Oldenburg. 'A Framework for Simulation and Inversion in Electromagnetics'. Computers & Geosciences 107 (October 2017): 1–19. https://doi.org/10.1016/j.cageo.2017.06.018.

[5] Li, Yaoguo, Aline Melo, Cericia Martinez, and Jiajia Sun. 'Geology Differentiation: A New Frontier in Quantitative Geophysical Interpretation in Mineral Exploration'. The Leading Edge 38, no. 1 (January 2019): 60–66. https://doi.org/10.1190/tle38010060.1.

[6] Nguyen, Thanh Minh, and Q. M. Jonathan Wu. 'Fast and Robust Spatially Constrained Gaussian Mixture Model for Image Segmentation'. IEEE Transactions on Circuits and Systems for Video Technology 23, no. 4 (April 2013): 621–35. https://doi.org/10.1109/TCSVT.2012.2211176.

[7] Oldenburg, Douglas and Yaoguo Li. "INVERSION FOR APPLIED GEOPHYSICS: A TUTORIAL." (2003).

[8] Tarantola, Albert. Inverse Problem Theory and Methods for Model Parameter Estimation. Society for Industrial and Applied Mathematics, 2005. https://doi.org/10.1137/1.9780898717921.

[9] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 'Attention Is All You Need', 2017. https://doi.org/10.48550/ARXIV.1706.03762.