Geophysical Time Series Processing with Machine Learning

Johnathan C. Kuttai
Department of Earth, Ocean and Atmospheric Sciences
University of British Columbia University
Vancouver, BC 15213
11074580
jkutt@eoas.ubc.ca

Anran Xu
Department of Earth, Ocean and Atmospheric Sciences
University of British Columbia University
Vancouver, BC 15213
54002811
anranxu@eoas.ubc.ca

Abstract

With large scale geophysical surveys an enormous amount of time series data can be acquired. The idea is that subsurface bodies with significant property contrast (conductive, density, velocity, etc.) change the expected time series signal than that of a uniform medium. Encoded into the time series data is the property's inhomogeneity map through the subsurface. Using the data and a geophysical inversion scheme, a smoothed subsurface image is reconstructed. The reconstruction being only as good as the data going into it. Meaning quality control of the data is important. Though, it can becomes a very time consuming task when there are potentially millions of these time-series to analyse. For a single person this can be an overwhelming task. Machine learning is a relatively fresh topic in geophysics application generally but can be well suited to make an impact. Particularly in that it can be trained to pick bad chunks of time-series that corrupt processed inputs to an imaging technique. Being a new application, training data is sparse or at the very least time consuming to generate. Seismic methods are the earliest adoption of machine learning techniques with electromagnetics quickly realizing it's usefulness. This review will discuss existing supervised and unsupervised methods successfully being employed with a focus on how they can be used for quality control of time series acquisition data.

1 Introduction

Geophysics is relatively late to adopting machine learning techniques. In the past couple of years its use is growing exponentially. Many of the first applications pertain to the data that is acquired. Other applications are to the interpretation of the acquired data. We will discuss the former for this project. Geophysics is a broad discipline with many types of acquisition data. This can be in the form of making measurements of the gravity or magnetic fields of the earth, natural electromagnetic fields from the ionosphere and lightning strikes, pressure waves moving through the earth (e.g earthquakes) and artificial controlled source electromagnetic fields (Telford, 1990). Depending on the sub-surface target in need of imaging each of these measurements are sensitive to different physical properties and are applied accordingly. Electromagnetics is often applied to conductive properties, gravity is

UBC CPSC 532M Final Course Project (2022).

associated with density contrasts and magnetics is the magnetic susceptibility and remanence. This data is inherently contaminated with noise making the extraction of the desired signal difficult. This noise can be anywhere from geomagnetic noise, cultural noise or environmental. Up until now this was determined by the sensitivity of the measurement device itself. The better the signal to noise of the measurement device, the easier to record the desired signal. Though targets are becoming harder to find, mostly because they are deeper, the measurement devices are the limiting factor. This is where machine learning can get a footing into helping extract the signals from the measurements. The goal of this report will be to analyse the existing applications and the architectures used. Supervised methods are discussed in section 3.1 and the unsupervised methods are discussed in section 3.2. We will examine several papers of the current work in applying machine learning to the geophysics discipline. Each paper focuses on a different sub-genre of geophysical data acquisition. We will then compare the differences in techniques applied and the similarities of what is working across multiple applications. Finally we argue whether these techniques can be applied to our problem.

2 Data and the Problem

Generally geophysics data are interpreted via a physical property such as conductivity, porosity, density, magnetic susceptibility, velocity, etc. These measurements though are typically acquired as time-series measurement and the physical properties are extracted during the processing of the time-series. Most acquisition types process these time-series and then apply a technique of stacking the signal which provides an averaging and filtering application of the data. Electromagnetic methods stack voltage or frequency data, seismic methods stack velocity data, and magnetics and gravity are averaged over duration of measurement. For example, direct current and induced polarisation measurements derive their properties from 50% duty cycle square wave (figure 1). The apparent resistivity is extracted using the positive and negative "on time" data where the induced polarisation is extracted from the "off time" data giving two properties for the cost of a single survey (Telford, 1990). After processing, the data is then passed into interpretation tools to define sub-surface targets.

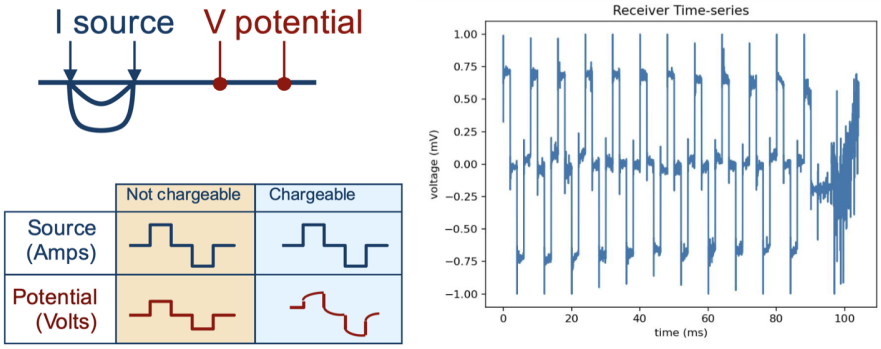

Figure 1: Left: example of a transmit-receiver pair (L.Heagy, Transform 2020). Right: example of corrupted receiver time series.

It can be observed that the interpretation will only be as good as the data going in. Acquired time-series data can quickly degrade depending on the environment they are collected in. Often this can be in urban areas, near mines, power lines, or any other power generating structures. We can even get degradation from natural disturbances in the magnetic, or gravity fields. As sensitivity in our instruments is increased these noisier signals become more and more of a bigger problem. More and more data are being collected creating an enormous task for people to monitor the quality of this collected data. For example, geophysical data acquired by aircraft can cover 100's, 1000's to tens of thousands of kilometres. This then leads into a second problem; capabilities of quality control of large amounts of data.

Going forward, geophysics data acquisition could greatly benefit from machine learning techniques. Applications of denoising to quality control of large amounts of data could easily be tasked to a computer. This makes more time for interpretation of results and quicker targeting turn around.

3 Paper review

3.1 Supervised Methods

3.1.1 Integrating neural networks in least-squares inversion of airborne time-domain electromagnetic data

In this paper, Asif et al. derive two separate neural networks ForwardNet and DeriveNet. They use artificial neural nets to predict data that represent the forward operator. With this they can get the data and the derivatives at a reduced cost than typical inversion. With the current architecture they incorporate the estimate for the flight height in airborne electromagnetic (AEM) surveys figure 2.

ForwardNet (fNet):

The input layer consists of 31 parameters corresponding to layer resistivities (Asif, 2019) for a model vector m. The output of fNet is an impulse response (db/dt) corresponding to the data collected. The interesting point here is how they incorporation estimation for flight height into the analysis. Shown in the paper that data are extremely sensitive to this parameter and having an improved representation of this leads to better forward modelling of the data (Asif, 2022).

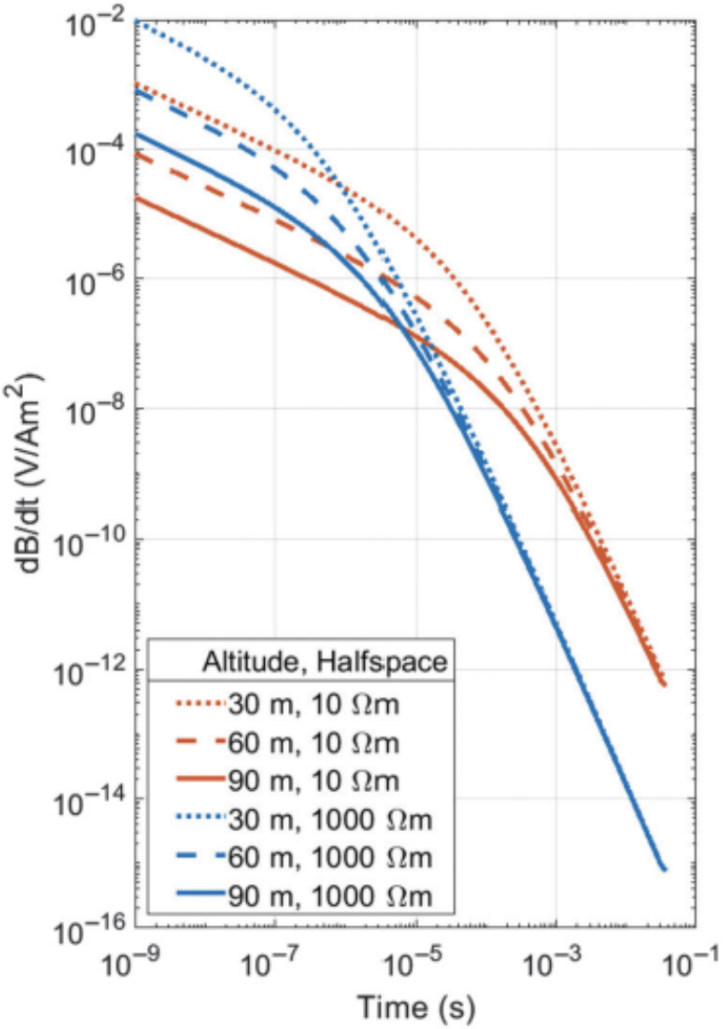

Figure 2: Effects of flight height on AEM data (Asif, 2022)

The neural net consists of 4 layers. The first is the input layer, the next two are hidden layers with 384 neurons each which is based on the grid search method of Liashchynski, 2019 (Asif, 2019). The last layer, the output layer, is the data representing the impulse response typical of an AEM survey with a time-series transmitted signal. The activation functions used are hyperbolic tangents since values can take on both positive and negative values. The loss function is a sum of squared loss with a regularizer to fight over fitting. The loss function is described as:

$$C = (1-\gamma)\sum_{i=1}^{N}(Y_i - \hat{Y})^2 + \gamma(\frac{1}{n}\sum_{j=1}^{n} w_j^2 + \frac{1}{n}\sum_{j=1}^{n} b_j^2) \tag{1}$$

Where $\mathbf{Y}$ is the numerical forward data and predicted data $\hat{\mathbf{Y}}$. The variables w and b are the weights and biases. γ is the performance ration factor.

DeriveNet (dNet)

Similar to fNet, dNet's input is layer resistivities but this time the output is the partial derivatives of the data that are fed into the inversion. There are also 5 layers in total, where they add a third hidden layer to estimate the parameters better.

Similar Techniques

A similar technique has been applied to airborne time-domain acquisition that uses a LSTM network to map the data to the model directly (Li, 2020). This is accomplished by stitching 1D geophysical models together produced from the LSTM neural net to form a multidimensional model. This method is not ideal as it does not consider the quality of the time series.

Training

For the training, synthetic data are produced from a suite of realistic subsurface patterns that ideally cover the entire model space expected from data.This suite is called the Karman models of resistive variation (Asif, 2006). To train, they use 700,000 1D models. 650,000 of which are used for training and 50,000 used for validation. To account for height parameters, simulated models are flown from 10 m to 120 m.

Results

Evaluating the effectiveness of fNet, it was used to invert field data which out of 801 inverted models 100% of them match within a 3% error threshold. However, they found that the model is sensitive to the nature of the data collected, meaning that the more resistive the subsurface the higher the error in fNet. In the end using both fNet and dNet to perform the data inversion reduces time and can be comparable in conductive environments. This is thought to be because there is much more variation in the signal itself at low resistivity as this represents high conductivity which is more evident feature in AEM data.

The advantage of method is that it relates the data processed from the time series to the inversion itself. The network models the physics and produces what it would expect the data to be. The disadvantage to this approach is that depending on the training data used it likely does not represent the all expected data derived from the field time series data. This method also does not predict the time series directly but rather the processed output from the time series. This could be used to compared the processed time series data from the field to that of the predicted from the network but not ideal for quality control.

3.1.2 The application of artificial neural networks to magnetotelluric time-series analysis

Magnetotellurics is a natural source geophysical survey that uses lightning and solar activity as sources to image the earth's subsurface. It has the advantage of carrying information imaging shallow to extreme depths depending on the recording time (Manoj, 2003). It can often be used to image the mantle crust transition. This is done by recording time-series data of the fluctuations in the earth's electromagnetic field. From the measurements, transfer functions are derived which represent the conductivity of the subsurface. These transfer functions are estimations of the ratios between the electric and magnetic fields (Telford, 1990). From the transfer functions we can calculate apparent resistivity and phase which are then used for interpretation.

The reliability of these measurements depend greatly on the amount of noise contained in the signal. This noise is from man-made and natural sources that do not fit the normal distribution of the data. Another drawback is that noise can be highly correlated as well. The noise in magnetotelluric time series are both active and passive and to get the best estimate of our conductivity changes of the subsurface the characterization of the noise and the classification of the signal is very important. For every station there are up to 5 time series components recorded: Hx, Hy, Hz, Ex, Ey. Processing is primarily manual at this point and can be a large task for a single person to do. This is primarily a pattern recognition problem suitable for a computer to do. To automate this an artificial neural network is trained to provide acceptable processed time-series that is used to create reliable transfer functions (Manoj, 2003).

Feed-forward artificial neural network (FANN)

A most commonly used network for a problem like the MT processing problem is the feed-forward artificial neural network. Here a simple FANN architecture consisting of 3 layers is employed:

1. Input layer.
2. Hidden layer.
3. Output layer.

In the hidden and the output layers the activations are sigmoid functions.

Training

The training is then completed using backpropagation. The network feeds the 5 channels of time-series each with 256 points into a single network. The signal detection scheme was divided into two parts for training: Detection of the patterns of individual channels. From a single stack five values corresponding to the five channels are output. Detection of interchannel parameters.The inputs for this neural network are the amplitude ratios, correlation coefficients and the quality classification from the previous network. Finally, the overall quality factor is the output of this final network.

A total of 1200 MT long period time series stacks represent the training dataset. This totals 6000 total traces where 70% are training and the remaining 30% are for testing.

Results

The results of this simple neural network show the robustness of this application. In low noise environments, the results are nearly exactly the manual edits done by humans. Even in the mid to high noise regime the network is similar to manual edits however, this means the same error is made in the estimates.

Given the acceptable performance of the network, it is deemed that the automation of MT time-series processing is reliable and efficient for regular use.

This method is closely related to the task of quality control similar to our outlined goal. The advantage of this network is that it uses direct time series inputs and learns proper features to be used in the end processing of the data. A disadvantage though, is the manual labour it would take to create training data and labeling desired features. However, this method is very effective and would be worth the time.

3.1.3 Automated First Arrival Picking: A Neural Network Approach

One of the most cited papers in neural networks for geophysics, the automated picking of first arrivals in seismic data processing dates back to 1992 (Murat, 1992). This is one of the earliest adoptions of neural nets. It is interesting to see that this was considered many years before becoming popular in geophysics. The use of this back propagation neural network was also highly successful too. It was also done including noisy field seismogram converted from time series data. The network was designed to take in attributes of input data and output a classification whether each half cycle was a first pick or not. Where a first pick is the designation of a first arrival signal from a physical source input.

Network Design

In a noise free seismogram, first picks are easily observed by an interpreter. As noise is added, it can become very difficult pick. It is also difficult to form the pattern recognition in mathematical constructs (Murat, 1992). This is why the neural network approach was considered and was successful. The approach was written by McClelland and Rumelhart in 1988 for a variety of pattern recognition problems (Murat, 1992). For current network discussed, a three layer neural net was constructed to take 4 inputs and output a classification of first pick or not. The three layers consist of an input layer followed by a fully connected 10 neuron hidden layer and then finally the output layer. The input of this network is 4 attributes derived from time series data. These are:

1. Peak amplitude of a half cycle.
2. Peak to lobe difference.
3. RMS amplitude ratio.
4. RMS amplitude ratio on adjacent traces.

These attributes were selected by a clustering technique to ensure that the attributes are separable and the neural net will be able to discriminate the features. This is done in 3D plots and is the typical procedure for an interpreter to aid in their decision (Murat, 1992).

Training & Results

Two source types producing data were used for training. One from a vibroseis source and another from signals produced from air shots. Training data from these sources included poor quality time series in order to train on data from typical difficult acquisition. Even with the difficult time series traces, the network classified at a satisfactory rate. Higher frequency data being the best while lower frequency data far from the source degraded the accuracy. Close to the source 95 of the 96 classifications were picked correctly. Further from the source the lower frequency accuracy was 67 or 96 classification correctly identified. Of these 67 classification 51 had confidences of greater than 90

It is interesting to see a technique developed decades ago become relevant. The greatest advantage and surprising for an early implementation, is the use of real data to train the neural network. It is not always adopted and synthetic is data is done because its easier to produce. Another advantage is inputting attributes derived from the time series which lightens the computation load. The disadvantage here is requiring users to label and gather these features for training. As well, significant preprocessing is required to extract the input attributes from the time series.

3.1.4 P, S arrival time picking in seismic time series using deep neural network

PhaseNet is a supervised neural network whose inputs are seismograms in three coordinates and outputs are probabilities of P-pick (P wave arrivals), S-pick (S wave arrivals), and noises. The final P-pick and S-pick are determined based on the peak in the corresponding distribution. The training data comes from the seismograms in the last 30 years in California. The training data applies the simplest pre-processing: downsampling to 100Hz. And each training sample contains 1 manually picked P arrival time point and 1 manually picked S arrival time point. The P-pick and S-pick distributions were generated by using the Gaussian distribution whose centre is the corresponding time point and standard deviation 0.1s. The distribution of the noise is set to 1- Prob(p)-Prob(s); namely, all time points which are not P or S arrivals are considered as noises (Zhu and Beroza, 2018).

U-Net was chosen in this study since it can localise the properties in an image. See the scheme in figure 4.

The cross-entropy was chosen to be the loss function. The distributions were calculated using the softmax function where i = 1, 2, 3 (P-pick, S-pick, noise) and z(x) are the values of the last layer:

$$q_i(x) = \frac{e^{z_i(x)}}{\sum_{k=1}^{3} e^{z_k(x)}} \tag{2}$$

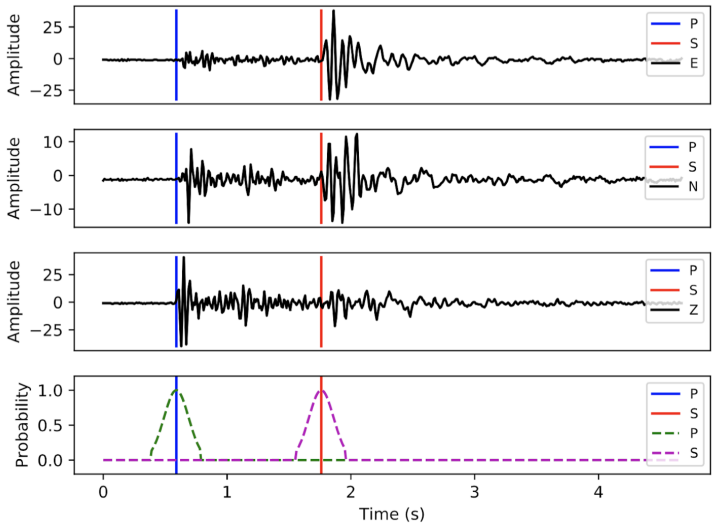

Figure 3: A sample from the data set. (a) – (c) Seismograms of the "ENZ" (East, North, Vertical) components. The blue and red vertical lines are the manually picked P and S arrival times. (d) The converted probability masks for P and S picks (Zhu and Beroza, 2018).

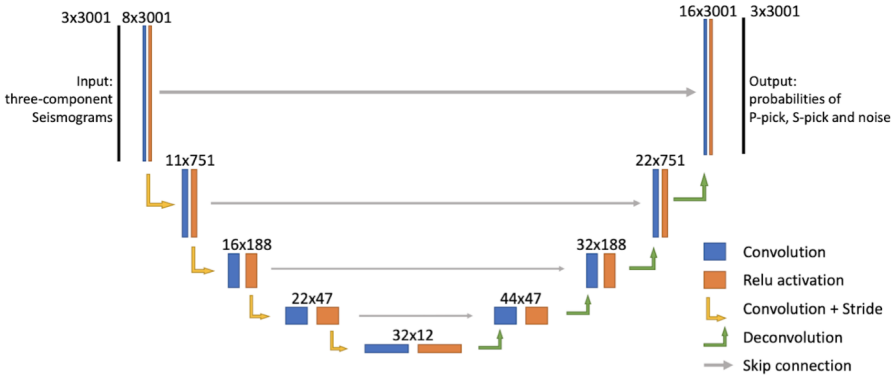

Figure 4: The U-Net scheme used in PhaseNet

The results from PhaseNet are compared with the results from AR picker by showing the corresponding precises, recalls, and F1 score. Moreover, it gives some predictions that are more consistent than the ground truth, which implies that the manually picked time points may be incorrect.

Since it was only trained on the seismograms from California with known waveform, it won't give accurate results for the regions where the earthquake types are quite different from the limited types of earthquake in California. It fails to pick some small earthquakes that were not labeled in the training set, which can be detected using certain unsupervised methods. Moreover, since the data set only contains the samples that have both P-pick and S-pick, it may return a false positive since it lacks the data that has no arrival times in it. However, the idea of generating probability instead of a single time point makes the algorithm more reasonable since it gives an uncertainty of the prediction. We can migrate this idea into our quality control problem where the groundtruth would be a Gaussian distribution centre at the start time of the corrupted data and assign a reasonable standard deviation to the distribution.

3.2 Unsupervised Methods

3.2.1 Earthquake detection in seismic time series using similarity search (3 papers to cite)

FAST is an earthquake detection method proposed by Bergen et al.(2015). The method is based on searching similarity over a long period (more than one week) of seismograms, which is a record of the ground motion, at different stations. As an earthquake occurs, we expect to see some impulsive signals on the seismograms collected from the stations near the epicentre of the earthquake. If the seismograms have a high signal-to-noise ratio and the magnitude of the earthquake is large enough, the common energy detector methods such as STA/LTA method would be good enough for detecting the earthquake. However, in more complicated cases such as detecting consequential small earthquakes after a large earthquake, it is hard to find a threshold for STA/LTA that can distinguish the small earthquake events and cultural noise; therefore, an end-to-end and unsupervised method was proposed to solve the problem. Without any information of labels of the time series or known event waveform, FAST can label the timestamps on the seismogram for potential earthquake events. FAST mainly requires the following four steps:

1. Data pre-processing
2. Generate Fingerprints
3. Similarity search
4. Network detection

The first step involves applying bandpass for noise removal since this technique is sensitive to the background noise. Namely, if we choose the wrong bandpass, high similarity would be the similar cultural noises data and earthquake events would have a relatively lower similarity score. The second step was done following these steps (Bergen et al., 2016):

1. Convert seismic time series into spectrograms which reflect the change of frequency spectrum over time.
2. Divide spectrogram into overlapping m$\times$n spectral images.
3. Apply wavelet transform to Haar basis for each spectral image to emphasis the localised features of the frequency spectrum to get a m*n matrix of wavelet coefficients (for each spectral image).
4. Keep only the sign of the coefficients, then convert coefficients to a binary sequence of 0's and 1's (negative: 01, Zero 00, Positive: 10) to get a 2m*n matrix of binary fingerprints.

The third step is based on the Jaccard Similarity, which measures the similarity of two fingerprints A and B by calculating $\frac{|A \cap B|}{|A \cup B|}$(higher the Jaccard Similarity score, higher similarity between A and B). To accelerate the similarity search, MinHash and Locality Sensitive Hashing are applied to approximate the Jaccard Similarity (Bergen et al., 2016). However, this approximation method introduces two new hyperparameters: the number of hash tables and the number of hash functions per table. The output of the similarity search would be a list of pairs of similar fingerprints and their corresponding similarities. The last step is Network detection which is based on the sparse similarity matrix (k$\times$k, k is the number of fingerprints, and the value of each element is the similarity of the corresponding pair calculated in the last step.). It's called network detection since each network has multiple stations and each station has multiple channels and each channel has one time series, and by combining sparse similarity matrices in all channels we get the final network level sparse similarity matrix (Bergen et al., 2018). High similar events would be observed in dense regions in the sparse similarity matrix. Since the index of each fingerprint represents the start time of the corresponding spectral image, we could label the corresponding time in the origin time series as the potential occurrence of an earthquake.

One disadvantage of this method is that it cannot detect an earthquake that occurs only once and is not similar enough to any other earthquake in the continuous data. Another disadvantage is that it has many hyperparameters to choose. However, it is very efficient in detecting small or recycled earthquakes and it's much faster than the similarity search based on calculating the auto-correlation of the time domain data. And the idea of generating binary fingerprint in the frequency domain is an efficient feature engineer that can be used to extract the features of other time series data. Moreover,

we can try different clustering methods other than the Jaccard Similarity search since in the quality control task, we only have two group data–good signals and corrupted signals, and by doing so we can reduce the number of hyperparameters in the method.

3.2.2 Outlier removal pCo2 time series data using local outlier factor

Baynes Sound Mooring (BSM) is an ocean acidification monitoring system, which delivers important data to monitor changes in seawater pH and dissolved carbon dioxide concentrations that will assist the local aquaculture operators in managing their shellfish stocks. However, the pCO2 sensors at a station were experiencing several issues, and the goal is to identify and remove the outliers in the data by using an unsupervised anomaly detection method called local outlier factor (LOF). LOF is a metric that quantifies the likelihood that a certain data point is an outlier/anomaly. LOF compares the local density of a point to local density of its k-nearest neighbours and returns a score as the final output (Breunig, 2000).

Reachability distance (RD) of x and o is defined as max(k-distance(o), distance(x, o)), where k-distance(o) is the distance between points o and its kth neighbour. Local reachability density (LRD) is defined as

$$LRD_k(x) = \frac{|N_k(x)|}{\sum_{o \in N_k(x)} d_k(x, o)} \tag{3}$$

where d_k is RD.

LOF score for x is defined as

$$LOF(x) = \frac{\sum_{o \in N_k(x)} \frac{LRD_k(o)}{LRD_k(x)}}{|N_k(x)|} \tag{4}$$

As some other outlier detection methods, the results of the method highly depend on the choice of the hyperparameters K (number of nearest neighbours) and time intervals (number of points in set N_k). With higher value of K, the values of LOF are lower. Consider an outlier whose LRD is anomaly low, since we are calculating LOF by averaging over more non-anomaly neighbour points, the final average is decreased. Due to the similar reason, the threshold for outliers would decrease as K increases.

The advantage of LOF is that it filters out obvious outliers quickly, which saves a lot of time compared to filtering out them manually. It also detects outliers based on density, so it may find some outliers that may not be easily observed by eyes. The disadvantages are if an improper value of K is selected, we may not get the expected results. And if we set a very restricted threshold for LOF, we may filter out some inliers as well. It works well with a smaller number of data points when the data points are grouped based on other criteria of the neighbour. In this case, time. Considering the scenario where the feature value is a linear function of time (pCO2 = a*time +b, a>0), we separate it into two time slots, and the sampling rate in the earlier time slot is greater than the one in the later time slot. If we worked with the whole data set, we are likely to filter out points in the later time slot since their values are "sparse" even though they are inliers. However, if we work with two groups separately, we would avoid the issue by properly choosing K and threshold.

4 Discussion

In this study, we present several supervised methods. All of them require a large training set to make the neural networks have good performance. For example, the PhaseNet has 30 years of seismic data as the training set, and ForwardNet has synthetic data that ideally covers the entire model space expected from data. Especially in the scenario where we use numerical simulations to generate synthetic data since we would need to call forward modeling functions tons of thousands of times. The FANN network for magnetotellurics gives the advantage of working with time series directly and acts as a form of quality control. In the contrast, the unsupervised methods are more flexible regarding the data set size.

Two unsupervised methods are presented in processing time series data in geophysics. One lesson we learned from the FAST algorithm is to transfer time domain data into the frequency domain and make predictions on spectrograms that contain two dimensions–time and frequency. Generally speaking,

adding more features would decrease the training error, and frequency domain data can be considered as an extra feature added to the original data. We can add two features to the quality control problem: frequency and derivatives of the original time series since both features have different behaviours when an anomaly/outlier occurs. Some non-parametric unsupervised methods may not be very suitable for solving the problem since it would not be easy to find the right hyperparameters such as the K value in LOF method. Namely, the value of K would change for different projects. However, there are some parametric outlier detection methods available such as PCA. For example, we can reduce the dimension of time series by projecting them into k principal components calculated from the whole data set. Ideally, the corrupted data would have anomaly values in certain dimensions (called dimension set M) so that corrupted and good data can be distinguished on this lower dimension space. We can store the dimension set M disclosed in one project where we have the labels for corrupted and good data and reuse it in another project where we have no knowledge about the labels. And the two ideas can be combined: Instead of calculate k principal components in the time domain, we can transfer data into spectrograms and then calculating k principal components in the spectrogram space and test which cases work better.

References

[1] Asif, Muhammad Rizwan, Nikolaj Foged, Pradip Kumar Maurya, Denys James Grombacher, Anders Vest Christiansen, Esben Auken, and Jakob Juul Larsen. 'Integrating Neural Networks in Least-Squares Inversion of Airborne Time-Domain Electromagnetic Data'. GEOPHYSICS 87, no. 4 (1 July 2022): E177–87. https://doi.org/10.1190/geo2021-0335.1.

[2] Breunig, Markus Kröger, Peer, Ng, Raymond, Sander, Joerg, 'LOF: Identifying Density-Based Local Outliers.', ACM Sigmod Record. 29. 93-104(2000). 10.1145/342009.335388.

[3] C. Yoon, O. O'Reilly, K. Bergen and G. C. Beroza, 'Earthquake detection through computationally efficient similarity search', Sci. Adv. 1, e1501057(2015).

[4] K. Bergen, C. Yoon and G. C. Beroza, 'Scalable Similarity Search in Seismology: A New Approach to Large-Scale Earthquake Detection', 9th International Conference on Similarity Search and Applications, Tokyo, Japan, 1-8 (2016).

[5] K. Bergen and G.C. Beroza, 'Detecting Earthquakes over a Seismic Network using Single-Station Similarity Measures', Geophys. J. Int.(2018), doi:10.1093/gji/ggy100

[6] Li, Jinfeng, Yunhe Liu, Changchun Yin, Xiuyan Ren, and Yang Su. 'Fast Imaging of Time-Domain Airborne EM Data Using Deep Learning Technology'. GEOPHYSICS 85, no. 5 (1 September 2020): E163–70. https://doi.org/10.1190/geo2019-0015.1.

[7] Manoj, C., and Nandini Nagarajan. 'The Application of Artificial Neural Networks to Magnetotelluric Time-Series Analysis'. Geophysical Journal International 153, no. 2 (May 2003): 409–23. https://doi.org/10.1046/j.1365-246X.2003.01902.x.

[8] Murat, Michael E., and Albert J. Rudman. 'AUTOMATED FIRST ARRIVAL PICKING: A NEURAL NETWORK APPROACH1'. Geophysical Prospecting 40, no. 6 (August 1992): 587–604. https://doi.org/10.1111/j.1365-2478.1992.tb00543.x.

[9] Telford, W. M., L. P. Geldart, and Robert E. Sheriff. Applied Geophysics. 2nd ed. Cambridge [England]; New York: Cambridge University Press, 1990.

[10] Weiqiang Zhu, Gregory C Beroza, 'PhaseNet: a deep-neural-network-based seismic arrival-time picking method', Geophysical Journal International, Volume 216, Issue 1, January 2019, Pages 261–273, https://doi.org/10.1093/gji/ggy423