

# Automatic processing of time domain induced polarization data using supervised artificial neural networks

Adrian S. Barfod<sup>1,2</sup>, Léa Lévy<sup>2,3</sup> and Jakob Juul Larsen<sup>1,2</sup>

<sup>1</sup>Geophysics, Instrumentation, and Signal Processing, Department of Engineering, Aarhus University, 8200 Aarhus, Denmark.

E-mail: [adrianbarfod@eng.au.dk](mailto:adrianbarfod@eng.au.dk)

<sup>2</sup>Aarhus University, Center for Water Technology (WATEC), Aarhus University, Ny Munkegade 114-166, 8000 Aarhus, Denmark

<sup>3</sup>HydroGeophysics Group, Department of Geoscience, Aarhus University, 8000 Aarhus, Denmark

Accepted 2020 September 23. Received 2020 September 18; in original form 2020 March 12

## SUMMARY

Processing of geophysical data is a time consuming task involving many different steps. One approach for accelerating and automating processing of geophysical data is to look towards machine learning (ML). ML encompasses a wide range of tools, which can be used to automate complicated and/or tedious tasks. We present strategies for automating the processing of time-domain induced polarization (IP) data using ML. An IP data set from Grindsted in Denmark is used to investigate the applicability of neural networks for processing such data. The Grindsted data set consists of eight profiles, with approximately 2000 data curves per profile, on average. Each curve needs to be processed, which, using the manual approach, can take 1–2 hr per profile. Around 20 per cent of the curves were manually processed and used to train and validate an artificial neural network. Once trained, the network could process all curves, in 6–15 s for each profile. The accuracy of the neural network, when considering the manual processing as a reference, is 90.8 per cent. At first, the network could not detect outlier curves, that is where entire chargeability curves were significantly different from their spatial neighbours. Therefore, an outlier curve detection algorithm was developed and implemented to work in tandem with the network. The automatic processing approach developed here, involving the neural network and the outlier curve detection, leads to similar inversion results as the manual processing, with the two significant advantages of reduced processing times and enhanced processing consistency.

**Key words:** Hydrogeophysics; Electrical resistivity tomography (ER6); Neural networks, fuzzy logic.

## 1 INTRODUCTION

The field of geophysics can benefit from machine learning (ML) by increasing the computational efficiency of inversion algorithms, automating data-flows and discovering new patterns, structures or relationships (Bergen *et al.* 2019). In this paper, we focus on automating data-flows. In short, ML deals with teaching a computer to autonomously complete complicated time-consuming tasks, such as cleaning and processing large geophysical data sets. ML algorithms can be divided into two classes: unsupervised and supervised learning algorithms (Bergen *et al.* 2019). Supervised learning focuses on teaching a computer to identify patterns with human supervision, that is labelled data, while unsupervised learning focuses on teaching the computer without human supervision. We focus on supervised learning algorithms. An example of a supervised learning algorithm could be binary classification of data sets with or without noise. Here, the computer is shown a set of data values, input, and corresponding labels, for example ‘good’ (0) or ‘bad’ (1). Based

on these labels, the computer is trained to identify each set of input values as either ‘good’ or ‘bad’. Provided labelled data exist, the supervised learning approach can be used to automate data processing of various types of geophysical data. The computer learns which data values to remove, based on the patterns inherent in the training data.

The focus of this paper is the direct current (DC) time-domain induced polarization (IP) method, which is an extension of the electrical resistivity tomography (ERT) method, where the time-dependent voltage signal, resulting from the injection of a square wave electrical current, is analysed. Compared to ERT, the IP method can provide additional information about the mineral composition and pore space structure of rocks and soils (e.g. Vacquier *et al.* 1957; Marshall & Madden 1959; Bodmer *et al.* 1968; Maurya *et al.* 2018; Lévy *et al.* 2019). The IP method can be applied to a wide variety of subsurface imaging problems, such as estimating permeability and water conductivity (e.g. Börner *et al.* 1996; Slater 2007; Robinson *et al.* 2018), imaging subsurface contamination (e.g. Gazoty *et al.*

2012a; Orozco *et al.* 2012; Wainwright *et al.* 2016; Maurya *et al.* 2018) and time-lapse imaging of aquifer remediation and other subsurface processes (Versteeg & Johnson 2008; Orozco *et al.* 2013).

Manual processing of IP data is common before inversion (e.g. Chongo *et al.* 2015). The IP data consists of apparent chargeability curves, which are affected by different types of noise. By applying a signal processing scheme, such as proposed by Olsson *et al.* (2016), most systematic noise sources can be identified and subtracted from the signal. For example, background drift and 50 Hz harmonics can be modelled and removed, thus increasing the signal to noise ratio. However, in order to perform spectral inversion of time-domain IP data, and more generally to ensure that the observed data can be fitted within a satisfying misfit, further processing is usually necessary before inversion. This often involves removing gates from the chargeability curves, that are influenced by noise, and in some cases the entire curves. For example, with the inversion scheme of the software AarhusInv (AarhusGeoSoftware 2020), used in this study, chargeability curves that cannot be represented by a Cole–Cole model (e.g. Pelton *et al.* 1978) will not be fitted in a satisfactory manner at the end of the inversion. Generally, the early gates show steep slopes, which we attribute to undesired electromagnetic coupling. Early gates are therefore removed (Dahlin & Leroux 2012). Currently, this processing step is carried out manually for IP field-data. A standard workflow consists of inspecting the chargeability curves one-by-one, to remove gates and curves that prevent the inversion from converging.

In order to improve and accelerate the processing step, Orozco *et al.* (2018) have presented a workflow for automating the processing of IP data, applying a rule-based approach. We explore ML methods as an alternative tool for automating the processing of IP data. Generally, ML models can be trained using different types of labelled data, for example manual or rule-based processed IP data. They will simply mimic the processing scheme embedded in the training data. We propose using supervised *multilabel binary classification* (Tsoumakas & Katakis 2007) for semi-automating the processing of IP data. We show how a supervised artificial neural network (ANN) can be trained, and used for automated processing of IP data. In particular, a field example is presented from Grindsted, Denmark, to show how large IP surveys can be processed by only manually processing a small portion of the total data.

## 2 METHODS

### 2.1 IP data acquisition and classical processing

#### 2.1.1 IP acquisition

In surface IP acquisition, the measurements are carried out between four grounded electrodes, forming a quadrupole:  $A$  and  $B$  for current injection, and  $M$  and  $N$  for voltage measurements. For each quadrupole, the apparent resistivity  $\rho_a$  is calculated:

$$\rho_a = K \frac{V_M - V_N}{I_{AB}}, \quad (1)$$

where  $I_{AB}$  is the current injected between electrodes AB,  $V_M$  and  $V_N$  are the potentials measured at electrodes  $M$  and  $N$ , respectively and  $K$  is the so-called geometrical factor (e.g. Bertin & Loeb 1976).

The IP signal can either be measured as the transient voltage builds-up upon injection of the current, during the ‘charge’  $T_{\text{on}}$ , or as the transient voltage decays when the current is turned off, during the ‘discharge’  $T_{\text{off}}$ . These measurements are usually referred to as

100 per cent duty cycle and 50 per cent duty cycle, respectively (Olsson *et al.* 2015).

The integral apparent chargeability,  $M$ , in milliseconds, represents the area enclosed by the discharge curve,  $V(t)$ , and its zero asymptote, during a given time-window  $[t_1, t_2]$ , divided by the primary voltage,  $V_p$  (Bertin & Loeb 1976; Sumner 1976):

$$M = 1000 \text{ mV V}^{-1} \int_{t_1}^{t_2} \frac{V(t)}{V_p(t_2 - t_1)} dt, \quad (2)$$

where  $t_1$  and  $t_2$  delimits the integration window.

The total chargeability corresponds to the integral chargeability described by eq. (2), when the time-window is the entire time-range recorded or, most often, the remaining time-range after manual processing. The gated chargeability corresponds to the integral chargeability defined by eq. (2), when the time-window is the time-range included in one gate, which is typically 1–100 ms wide.

Apparent resistivity is given by:

$$\rho_a = K \cdot \frac{V_p}{I}, \quad (3)$$

where  $K$  is the geometrical factor and  $I$  is injected current.

The investigation depth of the apparent resistivity,  $\rho_a$ , depends on the intensity of the current injected, the electrode configuration, but also the conductivity of the medium and the period of the electrical current injected (e.g. Telford *et al.* 1990). Each chargeability curve is integrated over a set number of gates, or time windows, that is 36 gates in the Grindsted survey, with exponentially increasing gate widths, according to the scheme presented by Gazoty *et al.* (2012b). The data processing is carried out on these gated chargeability curves.

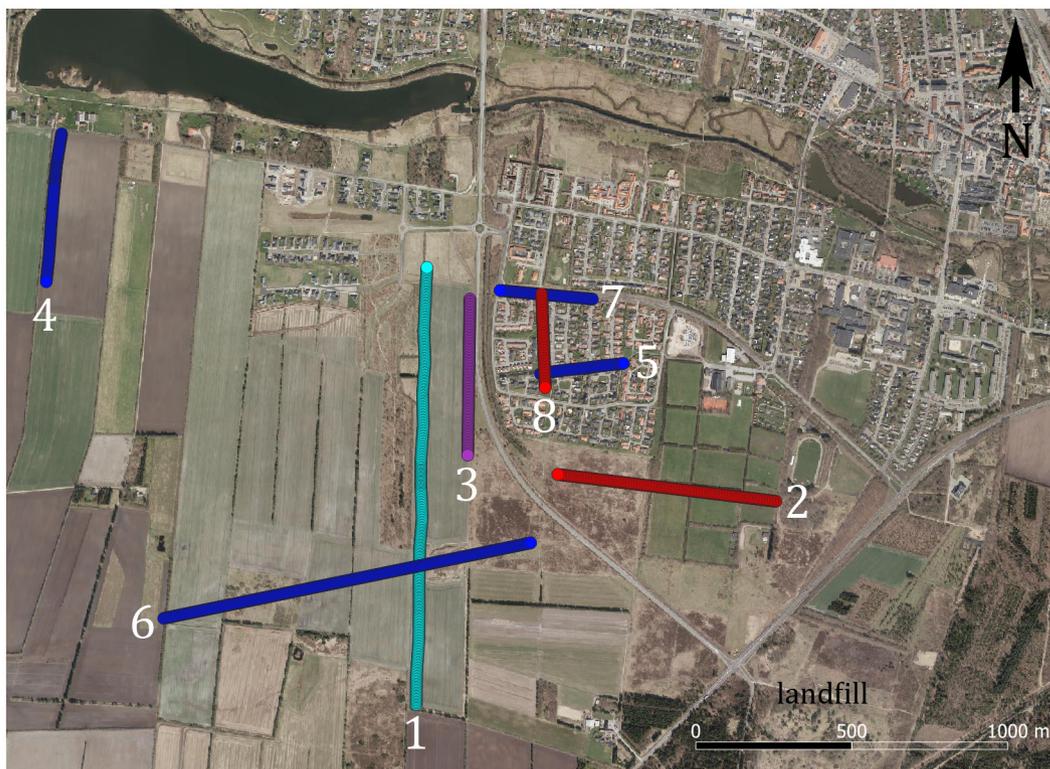
A convenient means to display results of IP acquisitions at a profile is a 2-D ‘pseudo-section’, which is obtained by placing each apparent resistivity and total chargeability data at a horizontal mid-point and a pseudo-depth (Loke & Barker 1996). The pseudo-depth is defined as the median ‘effective’ depth of investigation of the array and is calculated based on the Frechet derivatives for a homogeneous half-space (Edwards 1977; Loke 2004; Menke 2012).

#### 2.1.2 The Grindsted survey

The data originates from the Grindsted landfill site, located in the southwestern part of Jutland in Denmark (Fig. 1). Over the last years, multiple IP campaigns have been carried out in the area to investigate groundwater contamination (Maurya *et al.* 2017). The IP survey used in this study was carried out in April 2019. It contains eight profiles with a total of 16 511 chargeability curves. The eight profiles are spread out between the landfill and nearby stream situated to the northwest of the landfill (Fig. 1).

IP data were collected using the ABEM Terrameter LS instrument with a 100 per cent duty-cycle. Full waveform data were recorded at a sampling rate of 3750 Hz. A 5-m spacing set-up between neighboring electrodes was used. The measurements were carried out using a roll-along protocol (Dahlin & Bernstone 1997). The full waveform data were signal processed, using the scheme presented by Olsson *et al.* (2016). This includes:

- (i) Integrating the DC resistivity ( $\rho$ ) during the last 100 ms of the on-time.
- (ii) Gating chargeability curves, starting at 1 ms.
- (iii) Correcting for background drift related to spontaneous polarization and electrode polarization, using a drift model based on the time-domain expression of Cole–Cole model by Pelton *et al.* (1978).



**Figure 1.** An overview of the Grindsted Survey. Time domain induced polarization (IP) profiles marked in red have been manually processed and used for training and validation data (profiles 2 and 8). Profile 3, purple, was used for testing. Profile 1, cyan, was manually processed, but was inconsistent in comparison to the processing of the other profiles. The remaining profiles, marked in blue, were not manually processed and did not have labels and were not used for training, validation or testing.

- (iv) Removal of spikes at early recording times.
- (v) Removal of 50 Hz harmonic noise from surrounding infrastructure.

After signal processing, the chargeability data consist of 30–40 gated chargeability values, all of them associated to a central gate time, a gate width and a standard deviation. These gated chargeability data need to be further processed manually.

### 2.1.3 Manual processing

The signal processed data are manually processed using the software Aarhus Workbench (AarhusGeoSoftware 2020). First, profiles 1, 2, 3 and 8 undergo Direct Current (DC) processing, where only the apparent resistivity sections are considered. Resistivity outliers are identified and removed. Afterwards, each gated chargeability curve of the aforementioned profiles, are manually inspected. Parts of the curves, or entire curves, are removed when they contain non-decaying or the decay is either too slow or too steep. The goal is to prepare the data set for inversion so that it can be reasonably fitted with a Cole–Cole model. Of course, the classification of chargeability curves and gates as noise or signal is user-dependent. Therefore, a single experienced user carried out the manual processing for the results presented in this study, to make them as consistent as possible.

Chargeability curves before and after manual processing, presented in Fig. 2, illustrate the manual processing workflow. It can be observed that the early time-gates are systematically removed. The signal at early times is usually considered to be affected by electromagnetic coupling effects (Fiandaca 2018). Furthermore, entire

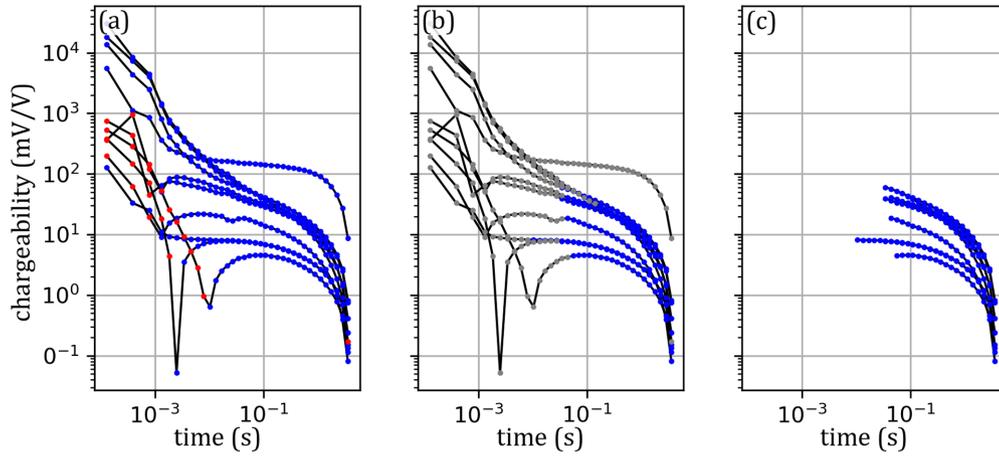
chargeability curves, considered outlier curves, are removed such as seen in Fig. 2(b), where an entire chargeability curve is removed due to it being significantly different from neighbouring data curves.

## 2.2 ANNs, set-up and training

Since some readers might not be familiar with Artificial Neural Networks (ANNs), this section is dedicated to briefly introducing ANNs, and how to setup and train them. The goal here is not to provide all the details, but simply give enough information to understand the basics.

## 2.3 A brief introduction to ML

The field of ML concerns itself with programming computers so they can learn from data (Géron 2019). A good example is an e-mail spam filter. Here, the computer is shown a feature vector,  $X$ , which contains the contents of a given e-mail. Furthermore, it is also shown a label vector,  $y$ , which in this example is a *one-by-one* vector containing a single Boolean value stating whether the given e-mail is considered spam or not. Such a vector pair  $(X, y)$ , is called an instance. After showing the computer multiple instances, it starts to identify patterns in e-mails that are considered spam (Clark *et al.* 2003). A common problem in ML is how to collect reliable labels. Again, the e-mail spam filter is a good example, since every time a user marks an e-mail as spam, he/she is essentially manually labeling the e-mail, which creates an instance pair  $(X, y)$ . However, if the user incorrectly classifies a bunch of work e-mails as spam,



**Figure 2.** The manual processing workflow of time domain induced polarization (IP) data, where (a) shows the signal processed data prior to manual processing (b) shows data after manual processing, where grey circles are the removed gates and (c) shows the final result, that is the non-flagged gates only. Blue and red circles correspond to positive and negative data, respectively. The presented chargeability curves are a selection of curves portraying the typical types of noise dealt with during processing.

the algorithm will start to get confused and identify some work e-mails as spam. Alternatively, if the people behind the spam e-mails become smarter and start composing spam e-mails that resembles work e-mails, then the computer will have a difficult time correctly identifying the spam e-mails, even though the labels are actually correct. In both cases, there is a weak link between the instance pairs. The trained ML algorithm is only as good as the training data.

#### 2.4 A brief introduction to ANNs

ANNs are a subclass of ML methods that are inspired by biological neural networks (BNNs), that is the brain's architecture. Similar to the brain, the building blocks that constitute an ANN are called neurons. In BNNs, the neurons are often connected in complicated networks. Similarly, the neurons of ANNs are also connected, but are arranged in layers, which gives them a simpler structure (Fig. 3). Each neuron mimics the behaviour of a biological neuron, which fires electrical signals when activated. This is simulated in the ANN by connection weights,  $w_{ji}$ , such as seen in Fig. 3. The activation of a neuron in the first hidden layer is a linear combination of the connection weights,  $w_{ji}$ , bias term,  $w_{j0}$  and the input variable,  $x_i$ :

$$a_j = \sum_{i=1}^N w_{ji}^{(1)} x_i + w_{j0} \quad (4)$$

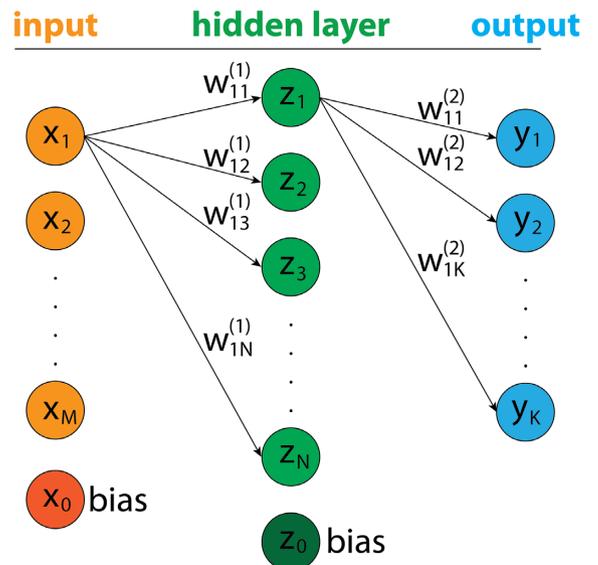
where  $j = 1, \dots, M$ , and superscript (1) refers to parameters from the first hidden layer of the ANN.

The  $k$ 'th output of the ANN can be computed by combining all of the activations for all layers, so considering the ANN from Fig. 3, and using eq. (4):

$$y_k(\mathbf{x}, \mathbf{w}) = f \left( \sum_{j=1}^K w_{kj}^{(2)} h \left( \sum_{i=1}^N w_{ji}^{(1)} x_i + w_{j0}^{(1)} \right) + w_{k0}^{(2)} \right), \quad (5)$$

where  $k = 1, \dots, N$ , superscript (2) refers to parameters from the second hidden layer,  $\mathbf{x}$  is a vector containing all inputs,  $x_1, \dots, x_M$ , and  $\mathbf{w}$  is the a matrix containing the weights of all layers in the ANN,  $f(\cdot)$  and  $h(\cdot)$  are the non-linear activation functions of the hidden layer and output layer, respectively.

The presented ANN has one hidden layer, which is the simplest ANN possible (Fig. 3). ANNs come in different shapes and sizes.



**Figure 3.** An overview of an artificial neural network (ANN). A subset of the connection weights are shown as arrows pointing from the first node of the input layer,  $x_1$ , to all subsequent nodes it is connected to. The same is shown for the first node of the hidden layer,  $z_1$ . Each node in each layer is connected to all nodes in the next subsequent layer.

The number of neurons in each layer can be increased or decreased, and hidden layers can be added as needed, to create quite complex neural networks. In this paper, we use a simple ANN structure with just one hidden layer, to show how well even the simplest form of ANN performs. The ANN performance is affected by so-called hyperparameters, which affect how the network is trained. However, it is beyond the scope of this paper to describe each of these hyperparameters. Readers who are interested in diving deeper into ML are referred to one of the many textbooks on ML. We recommend Géron (2019) or Bishop (2006)

##### 2.4.1 ANN setup

We used a *supervised multilabel binary classification* ANN for processing time-domain IP data. Each instance,  $(X, y)$ , consisted of

a  $N_{\text{gates}}$ -dimensional feature vector,  $X$  and a  $N_{\text{gates}}$ -dimensional label vector,  $y$ . In a *multilabel classification problem*, multiple labels exist for each instance, that is in the IP processing problem there are as many labels as gates. After systematically testing different ANN architectures, we chose a one-layer ANN with 40 neurons. Using multiple hidden layers did not improve the results, nor did increasing the number of neurons in the hidden layer(-s) beyond 40. Common activation functions were tested, such as the tangent hyperbolic function, sigmoid function, piecewise linear unit (PLU) and rectified linear unit (ReLU). The best performance was, however, achieved by using the ReLU activation function for the hidden layer, and the sigmoid activation function for the output layer, with a bias term added to both the hidden and output layers, but not the input layer. A full overview of the final ANN architecture can be found in Table 1.

We used the Python native Pytorch library for ML (Paszke *et al.* 2019). The Pytorch library provides a robust ML platform, enables training on GPUs using CUDA, was easy to debug, and was consistent with other popular scientific computing libraries, for example the *Numpy* library used for large matrix operations and high-level mathematics. The ANN was trained on a computer system with an NVIDIA GeForce RTX 2080 Titanium, and a Intel Xeon Gold 6132 CPU @2.6GHz.

The Grindsted survey data were divided into a training, validation and test set. The training and validation data was constructed by combining all chargeability curves from profiles 2 and 8, amounting to 15.2 per cent of the Grindsted data set. From this subset of data, the training data was selected by randomly drawing 85 per cent of the data and the remaining 15 per cent made up the validation data. The randomized data selection ensured consistency in the training and validation data. Finally, profile 3 was used for testing, that is the test set, since the ANN should always be tested on data that it has never seen before. An overview of the training, validation and test data can be found in Table 2. It is important to mention that the final training data set is a refined version of the raw data set. It was composed by removing all outlier curves, where entire curves are significantly different from its immediate neighbours, from the profiles used for training and validation, that is profile 2 and 8. Including the outlier curves in the training added bias to the ANN, and removing them increased the overall performance of the ANN.

#### 2.4.2 Normalizing the training data

Prior to feeding the IP data to a ANN for training, it is common practice to normalize the data as this increases the performance of ANNs (e.g. Jayalakshmi & Santhakumaran 2011). The data is normalized using a combination of MinMax-normalization (Al Shalabi & Shaaban 2006) and the  $\log_{10}$ -transform. All data, including potential outliers, are included in the normalization step. MinMax-normalization is used since it has the advantage of largely preserving all relationships in the data, and the  $\log_{10}$ -transform is used to reduce the dynamic range of chargeability curves. The first step of the normalization is the  $\log_{10}$ -transform, which due to negative values, is carried out as follows:

$$\hat{q}_{i,j} = \log_{10}(|q_{i,j}|), \quad (6)$$

where  $|q_{i,j}|$  is the absolute value of the gate chargeability for the  $i$ 'th gate of the  $j$ 'th sounding, and  $\hat{q}_{i,j}$  is the corresponding  $\log_{10}$ -transformed value.

After transforming the gate values, the MinMax-normalization is computed, for each  $\log_{10}$ -transformed gate-value,  $\hat{q}_{i,j}$ , using the

following formula:

$$z_{i,j} = \frac{\hat{q}_{i,j} - \min(\mathbf{Q})}{\max(\mathbf{Q}) - \min(\mathbf{Q})}, \quad (7)$$

where  $z_{i,j}$  is the  $i$ 'th normalized value of the  $j$ 'th sounding,  $\hat{q}_{i,j}$  is defined above in eq. (6),  $\mathbf{Q}$  is a  $(N_c, N_{\text{gates}})$ -dimensional matrix containing all  $\log_{10}$ -transformed gate values and  $N_c$  is the number of chargeability curves.

The gates are rescaled according to a particular data range, corresponding to the range of possible values that can be measured by the IP instrument, that is the MinMax-values are set manually. Based on the Grindsted data, we used a minimum value of  $\log_{10}(10^{-6})$  and a maximum value of  $\log_{10}(250\,000)$ , corresponding to a data range of:  $10^{-6}$  to 250 000 mV/V. The reason the data range is so large is due to the previously mentioned electromagnetic coupling effects, which increases the overall range of measured values. The result of the normalization strategy is presented in Fig. 4. It is seen that the overall relationships have been altered, as the negative values are mapped to positive values and the dynamic data range is significantly reduced. The transitions from negative to positive, and vice versa, are clearly seen in the normalized data as sharp inflection points in the chargeability curves. It is important to note that the stretched exponential decaying behaviour of the chargeability curves is accentuated in the normalized data (Fig. 4).

#### 2.4.3 Training the ANN

The network is trained using the AMSGrad variant of the Adam optimizer (Kingma & Ba 2014) and the back-propagation algorithm (Rumelhart *et al.* 1986) with the parameters shown in Table 3. The batch size determines how many chargeability curves are shown to the ANN at each training step. The learning rate hyperparameter determines how quickly the weights and biases are adjusted during training, and affects how fast, or slow, the network converges. The weight decay term penalizes large weights to ensure that the network does not over fit. Finally, the random seed is used for drawing the initial weights before the network is trained. The loss, which is similar to the objective function in geophysical inversion, and accuracy are tracked during training. The cross-entropy loss function was used to track the loss. To avoid over-fitting we use early-stopping (Prechelt 1998) to check that the training and validation loss decrease synchronously. The training results are seen in Fig. 5, and it is seen that the network starts overfitting after about 35 000 epochs, as indicated by the red line. Therefore, the weights after 30 000 epochs were used (Table 3). In Fig. 5 the accuracy is also being tracked. The accuracy describes the overall performance of the ANN, and is computed as:

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad (8)$$

where TP are the true positives, TN are the true negatives, FP are the false positives and the FN are false negatives. The accuracy itself is not a sufficient summary statistic. The precision and recall are considered for supplementing the accuracy. The precision, which represents the percentage of positive predictions that were correctly classified, is computed by:

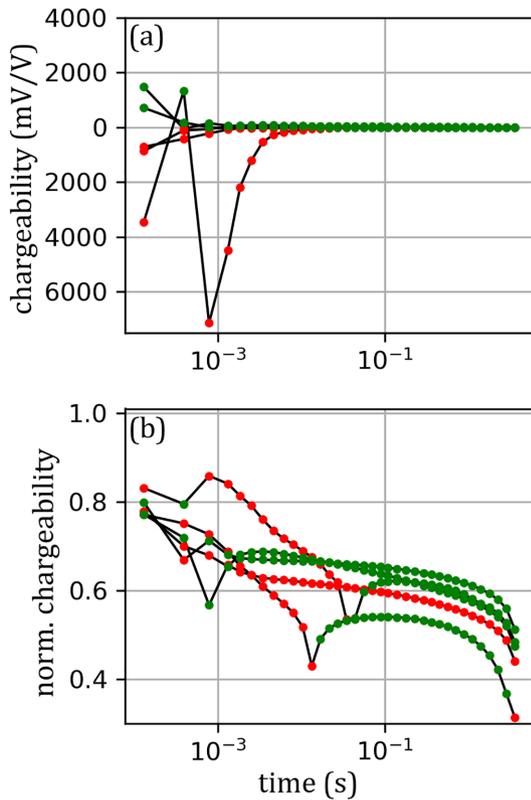
$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100. \quad (9)$$

**Table 1.** The artificial neural network architecture used for processing TDIP data. FC is short for fully connected layer.

Layer	Name	Layer type	Neurons	Activations	Bias
1	Input	Linear (FC)	36	N/A	False
2	Hidden	Linear (FC)	40	ReLU	True
3	Output	Linear (FC)	36	Sigmoid	True

**Table 2.** An overview table of the training, validation and test data. Each number represents the number of chargeability curves present for the given category. The percentages show how large a portion of the total survey the given category makes up.

Curves per profile (avg)	2000 (12.1 per cent)
Training data	2507 (15.2 per cent)
Validation data	439 (2.7 per cent)
Profile 1 (test)	5357 (32.4 per cent)
Profile 3 (test)	1512 (9.2 per cent)
Survey total	16511

**Figure 4.** An example of the data normalization using the  $\log_{10}$ -transform and MinMax-normalization. (a) the chargeability curves before normalization (b) the normalized chargeability curves. Taking the absolute value prior to the  $\log_{10}$ -transform forces the negative values to become positive, so to visually keep track, positive gate values are marked in green and negative values, which are mapped to positive values, are marked in red.**Table 3.** The final parameters used for training the artificial neural network.

Name	Value
Num. epochs	30 000
Batch size	100
Learning rate	1.00E-03
Weight decay	1.00E-06
Random seed	40
AMSgrad	True
Loss function	Cross entropy

The recall represents the percentage of actual positive cases that were correctly identified:

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100. \quad (10)$$

#### 2.4.4 Outlier curve detection

ANN processing is carried out on a curve-by-curve basis and to detect outlier curves several chargeability curves need to be considered. Since the ANN only sees one curve at a time, it cannot recognize outlier chargeability curves. The ANN processed IP data contains such outlier curves, and a strategy for outlier curve detection was developed. A given chargeability curve is grouped together with its immediate neighbouring curves, that were measured using the same quadrupole configuration, but shifted along the profile by one electrode. Since we use an electrode spacing of 5 m, we shift the quadrupole by 5 m. Such curves, which have been measured using the same electrode configuration, should yield relatively similar chargeability curves since the current and potential distributions are assumed to be homogeneous in the calculations of the apparent resistivity values. Chargeability curves are sorted into groups of three, the curve itself and its two immediate neighbours. If the curve is too dissimilar from both its neighbours, then the entire curve is considered an outlier and is removed.

The outlier analysis used, uses the following distance metric for comparing the  $j$ 'th and  $k$ 'th chargeability curves:

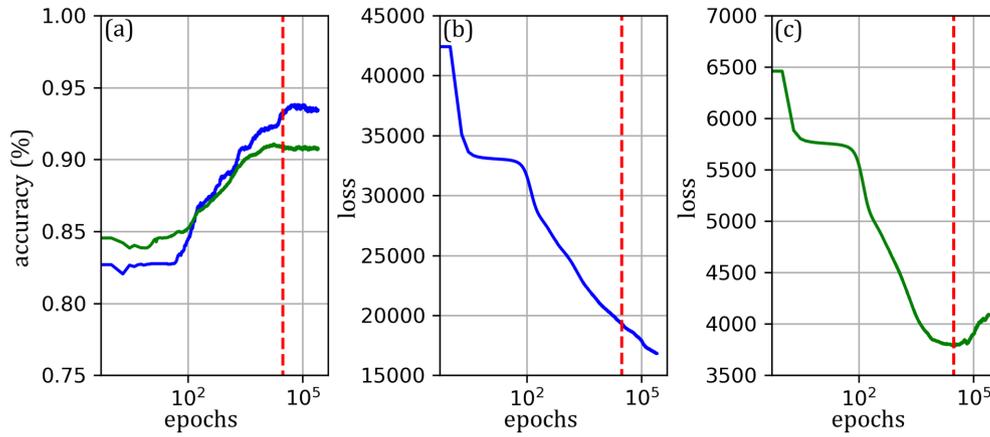
$$d_{j,k} = \sqrt{\frac{\sum_{i=N_{EM}}^{N_{\text{gates}}} (q_{i,j} - q_{i,k})^2}{N_{\text{gates}} - N_{EM}}}, \quad (11)$$

where  $N_{EM}$  is the number of gates that are affected by the EM-effect and is chosen manually, we used a  $N_{EM}$  value of 5 for the entirety of the Grindsted survey,  $N_{\text{gates}}$  is the number of gates,  $q_{i,j}$  is the  $i$ 'th gate of the  $j$ 'th chargeability curve, and  $q_{i,k}$  is the  $i$ 'th gate of the  $k$ 'th curve, and where  $j \neq k$ . The reason we remove the EM affected gates from the distance computations is to avoid them dominating the resulting distance values. Finally, note that the distances are computed for the data before normalization.

The distance metric described in eq. (11) describes the similarity between two chargeability curves, if they are similar then the distance is small, and *vice versa*. A given chargeability curve is compared to its two neighbours, yielding two distance values,  $d_{i,k-1}$  and  $d_{i,k+1}$ . A chargeability curve is considered an outlier and is removed, if both its similarity distances are above a threshold value. The threshold value is set manually based on the distribution of all similarity distances between all chargeability curves and their immediate neighbours.

## 2.5 Inversion

Inversion of IP data is carried out using the software AarhusInv (Auken *et al.* 2015) and the re-parametrized 'maximum phase angle' (MPA) Cole–Cole model suggested by Fiandaca (2018). The phase angle represents polarization as a momentarily 'frozen' impedance



**Figure 5.** The artificial neural network (ANN) training statistics (a) shows the accuracy for the training data (blue) and validation data (green) (b) shows the training loss (c) shows the validation loss. The ANN starts overfitting after about 35 000 epochs, as indicated by the red lines. Note the choice of  $\log_{10}$  for the x-scale, which emphasizes the improvements of the training statistics at late epochs.

vector (Sumner 1976). In the frequency domain, where AC current is injected at varying frequencies, the phase-angle of the impedance corresponds to the delay of the measured voltage relative to the injected current. The maximum phase angle corresponds to the maximum value of the phase spectrum in the relevant frequency range, typically 0.1–100 Hz for geophysical applications (Lévy *et al.* 2019). Moreover, for a given Cole–Cole impedance model, the phase angle can be calculated, using the arctan of the ratio between the ‘Cole–Cole predicted’ imaginary and real conductivity, at a given frequency (Cole & Cole 1941). The maximum-phase-angle, used in the MPA parametrization, corresponds to the value of this Cole–Cole-computed phase angle, taken at the frequency where the phase angle peaks/reaches a maximum. This peak frequency, or the corresponding relaxation time, is calculated from the chargeability, Cole–Cole exponent and relaxation time of the Cole–Cole model in its complex resistivity form (Pelton *et al.* 1978; Fiandaca *et al.* 2018). With the MPA parametrization, the model space consists of DC resistivity,  $\rho_0$ , maximum phase angle,  $\phi_{\max}$ , time constant,  $\tau_\phi$  (inverse of the angular frequency at which the phase angle reaches a maximum) and frequency exponent,  $C$ .

The depth of investigation (DOI) computation in AarhusInv is based on approximate covariance analysis, following Fiandaca *et al.* (2015). The analysis takes the noise in the data and the final derived model into account. In the results section, the DOI is shown as conservative (shallow) and less conservative (deeper). For each inversion, a dimensionless total data misfit  $\chi$  is calculated using:

$$\chi = \sqrt{\frac{1}{N} \sum_{i=1}^N \frac{(\log_{10}(d_{\text{obs},i}) - \log_{10}(d_{\text{fwd},i}))^2}{\delta_{d_i}^2}}, \quad (12)$$

where  $d_{\text{obs}}$  is the observed (measured) data,  $d_{\text{fwd}}$  is the forward prediction,  $\delta_{d_i}$  is the standard deviation of the observed data and  $N$  is the number of data points.

The relative standard deviation of resistivity is estimated to 1 per cent. The relative standard deviation of IP parameters is calculated as:

$$\delta_{d_{\text{IP}}} = \Delta A + \frac{V_{\text{threshold}}}{V_{\text{IP}}} \frac{1}{\sqrt{n_{\text{stack}}}} \frac{1}{\sqrt{\frac{g_{\text{width}}}{10}}}, \quad (13)$$

where  $\Delta A$  is a minimum constant standard deviation,  $\delta_{d_{\text{IP}}}$  is the standard deviation of IP parameters in a given gate,  $V_{\text{threshold}}$  is the nominal noise floor for 10 ms integration time, gate width, and one

stack in mV,  $V_{\text{IP}}$  the voltage value for a given gate in mV,  $n_{\text{stack}}$  is the stack size (dimensionless) and  $g_{\text{width}}$  is the gate width in ms normalized to 10 ms, that is  $g_{\text{width}}/10$  ms is dimensionless. The parameters  $V_{\text{threshold}}$  and  $\Delta A$  are adjusted manually until a suitable noise model is obtained. Values of 10 mV and 5 per cent are used in the present study, respectively. The objective function, which is minimized using the L2 norm, also contains vertical and lateral roughness constraints, where the relative difference (in log space) between the model parameters of two adjacent cells and the constraint value are compared (Auken & Christiansen 2004; Fiandaca *et al.* 2013, 2018). The following vertical and lateral constraint values are used here for resistivity,  $\rho$ ,  $\phi_{\max}$ ,  $\tau$  and  $C$ : [0.5,0.15], [0.5,0.15], [0.15,0.45] and [0.5,0.15], respectively. No constraint on the *a priori* value is used.

### 3 RESULTS

The manual- and ANN-based processing is compared in two different ways: (i) visual and statistical comparison of data remaining after ANN and manual processing (Fig. 6 and Table 4) and (ii) visual comparison of inversion results from the ANN and manually processed data (Fig. 7).

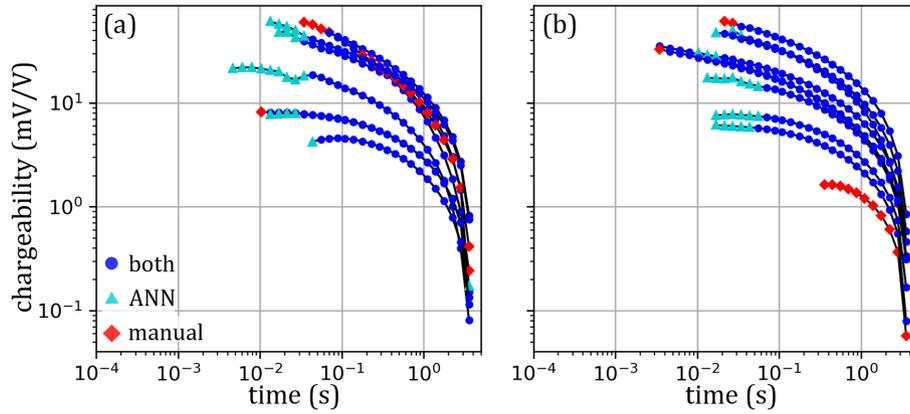
The statistical comparison of the data after ANN and manual processing are found in the confusion matrix (Table 4). The table has been compiled by considering the manual processing flags as the ‘true’ values, and the ANN processing flags as the comparison values. Using the numbers presented in the table, precision and recall can be computed.

$$\text{accuracy} = \frac{19978+29461}{19978+29461+1309+3684} = 90.8 \text{ per cent.}$$

$$\text{precision} = \frac{19978}{19978+1309} \times 100 = 93.9 \text{ per cent.}$$

$$\text{recall} = \frac{19978}{19978+3684} \times 100 = 84.4 \text{ per cent.}$$

The ANN-processing performs well, with an accuracy of 90.8 per cent, precision of 93.9 per cent and a recall of 84.4 per cent, respectively. Since the main objective of the processing is to keep only apparent chargeability curves that can be predicted by the forward response of an MPA model, a visual comparison between ANN and manually processed data is also shown (Figs 6a and b). The combination of ANN processing and outlier curve detection performs



**Figure 6.** Comparison of the manual and ANN processing for the test data (profile 3), where two sets of 10 representative chargeability curves have been hand-picked, the first set presented in (a) and the second set in (b). The same chargeability curves as presented in Fig. 2 are presented in (a). Note that the outlier from Fig. 2 is also removed by the ANN processing routine. The blue circles denote gates that were kept by both the manual and ANN processing routines. The red diamonds denote gates kept by manual processing, but removed by ANN processing. The cyan triangles denote gates kept by ANN processing, and removed by manual processing.

**Table 4.** The confusion matrix of the ANN-processing result for the test data (profile 3). The table is compiled by computing the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) relative to the manual processing result.

	Actual positive	Actual negative
Positive prediction	TP: 19978 (36.7 per cent)	FP: 1309 (2.4 per cent)
Negative prediction	TN: 29461 (54.1 per cent)	FN: 3684 (6.8 per cent)

well, and yields similar results, with only small differences, such as the ANN keeping more data at earlier times. In both plots, the outlier curve from Fig. 2 has been removed, showing that the outlier curve detection routine can successfully identify outlier curves and removed them. Furthermore, the ANN processing result is less conservative, that is less data is removed. This results in some irregular features remaining at early times for a few chargeability curves (Fig. 6b).

The outlier curve detection algorithm identifies and removes outlier curves that are significantly different from their neighbours. The results for profile 3 are affected by the removal of outlier curves, which can be seen in the summary statistics, Table 5. Removing outlier curves both reduces the accuracy and precision, but increases the recall. The outlier curve detection algorithm identified 95 (6.3 per cent) outlier curves for profile 3, which contained a total of 1512 chargeability curves. During manual processing 83 (5.5 per cent) outlier curves were identified and removed from profile 3.

Inversion results, based on manually and ANN processed data for profile 3, are presented in Fig. 7. Overall, the four inverted parameters shows similar distribution, although clear differences appear for the inverted time constant  $\tau$ . For example, the inversion results from manual processing shows patches of low time constant in the northern part of the profile, which are absent in ANN processed result.

The DC processing was identical for the manually and ANN processed IP data, and since the ANN processing has no influence on whether or not DC data are kept in the inversion, the resistivity pseudo sections are identical for manual and ANN processing. The ANN processing only deals with apparent chargeability curves. As a consequence, a few differences can be observed between the two integral chargeability,  $\langle M \rangle$ , pseudo sections. For example, a few more quadrupoles are absent at the centre of the ANN processing

pseudo section, meaning that they have been considered as outliers by the ANN processing but not by the manual processing. The total data misfit are almost identical for the manually and ANN processed data sets (2.7 and 2.8, respectively).

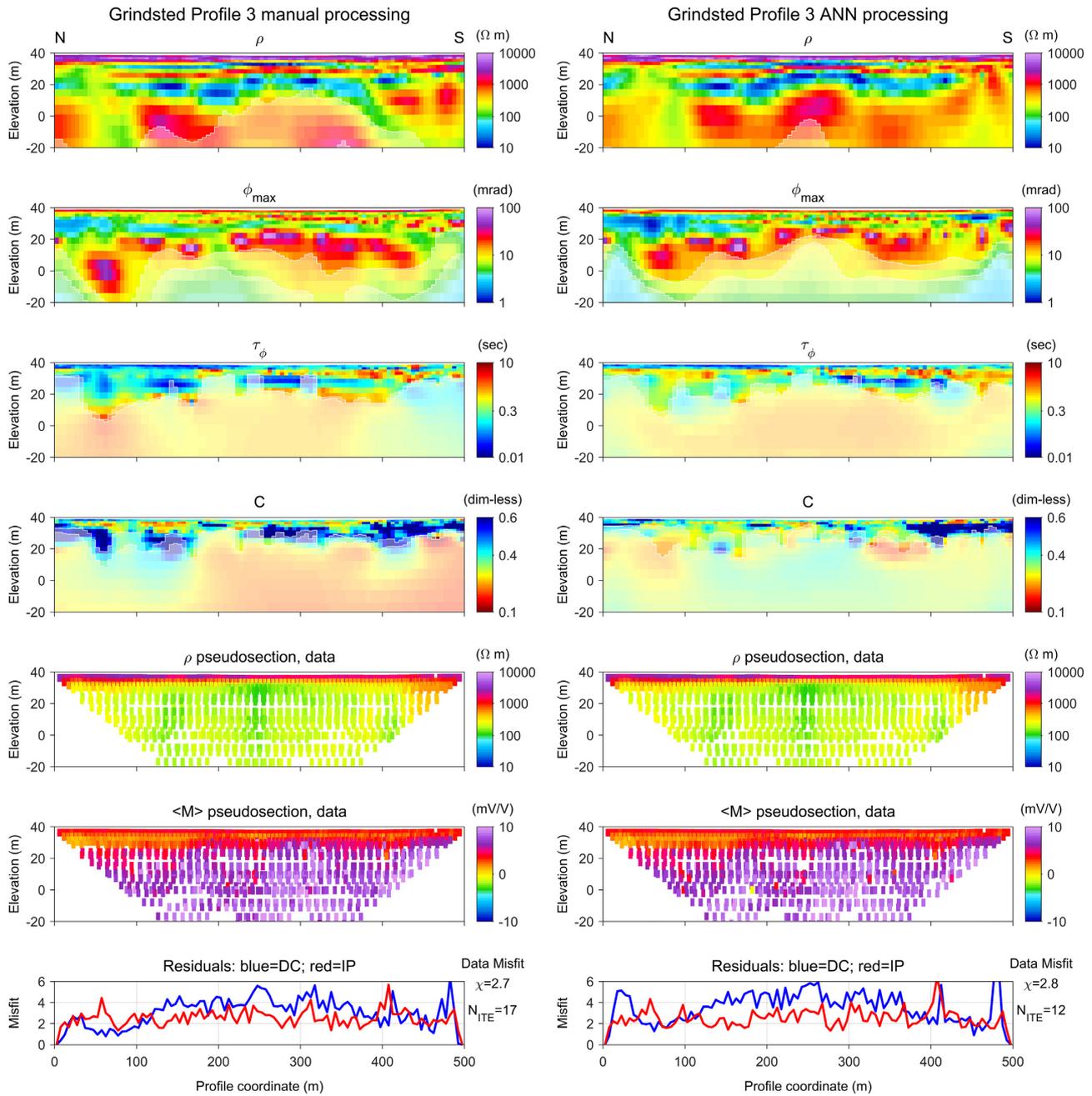
Finally, the ANN processing routine was applied to profile 1, where the manual processing had been carried out in a more conservative way, as compared to profile 2, 3 and 8 (Fig. 8). The data remaining after manual and ANN processing for profile 1 are presented in Fig. 9. With manual processing, early time gates before  $10^{-1}$  s were generally removed (Figs 8c and 9a). With ANN processing, gates as early as  $10^{-2}$  s were kept (Fig. 9b), according to the manual processing of profiles 2, 3 and 8 used for the training. Two outlier curves were identified during manual processing, a fully negative and fully positive chargeability curve, represented by the two grey chargeability curves in Fig. 8(b). These two outlier curves were successfully identified by the outlier curve detection algorithm, used by the ANN processing scheme, and were removed (Fig. 9b). The precision and recall for profile 1:

$$\text{accuracy} = \frac{93489+67921}{93489+67921+9402+22040} \cdot 100 = 83.7 \text{ per cent}$$

$$\text{precision} = \frac{93489}{93489+9402} \frac{100536}{100536+22546} \times 100 = 90.9 \text{ per cent.}$$

$$\text{recall} = \frac{93489}{93489+22040} \frac{100536}{100536+14993} \times 100 = 80.9 \text{ per cent.}$$

The performance of the ANN, when using the manual processing as a reference, is more satisfying for profile 3 than for profile 1. Especially when looking at the summary statistics: the accuracy was 83.7 per cent for profile 1 and 90.8 per cent for profile 3, the precision was 90.9 per cent for profile 1 versus 93.9 per cent for profile 3, and finally the recall was 80.9 per cent for profile 1 and 84.4 per cent for profile 3. At first glance, this might indicate an overall drop in performance from profile 3 to profile 1. Ideally, both the precision and recall would reach 100 per cent, meaning there are no false negatives nor false positives. However, it is important to remember that the summary statistics are computed relative to the manually processed data, and we do not expect the ANN to behave exactly the same as a human, especially if the reference manually processed data are different from the training data set. For profile 3, the precision was 93.9 per cent, and the recall was 84.4 per cent. For profile 1, the precision was 90.9 per cent of the positive predictions

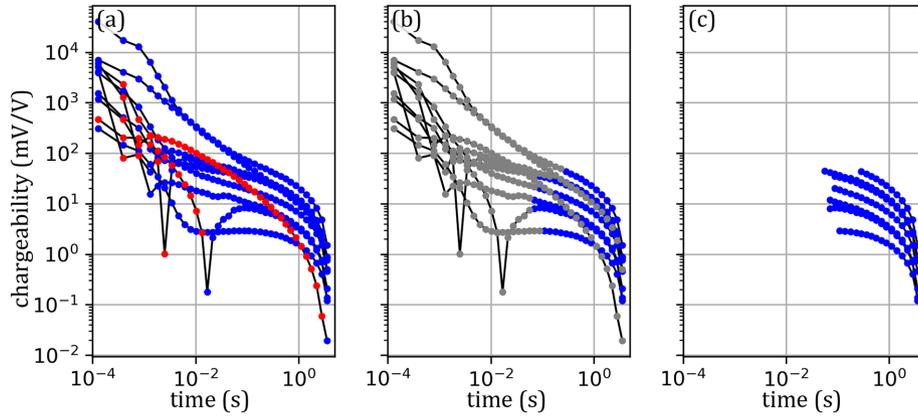


**Figure 7.** Data and inversion results for profile 3, using manually and ANN processed IP data (left- and right-hand columns, respectively). Manual and ANN processed data sets are inverted separately. Seven panels are shown for each inversion, from top to bottom: four model parameters corresponding to the MPA parametrization (resistivity  $\rho$ , maximum phase angle  $\Phi_{\max}$ , time constant  $\tau$ , frequency exponent  $c$ ), then resistivity and IP pseudo sections, and finally misfit evolution along the profile, with each value on the graph corresponding to the average misfit over the vertical column. The white lines in the lower part of the figure represent the shallow and deep DOI for each inverted parameter.

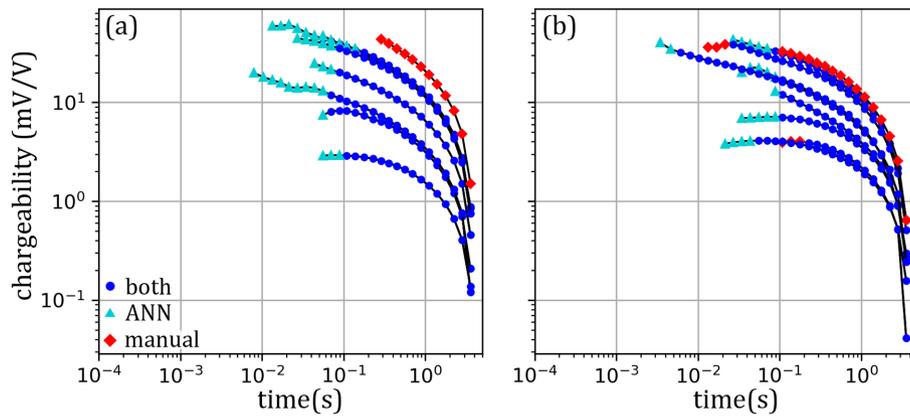
**Table 5.** The summary statistics before and after removing outliers in profile 3.

	Before	After	Difference
Accuracy (per cent)	91.2	90.8	-0.4
Precision (per cent)	95.8	93.9	-1.9
Recall (per cent)	83.5	84.4	+0.9

correctly (precision), and the recall was 80.9 per cent. Both the precision and recall were lower. This was to be expected, since the neural network is trained on data, where the manual processing is less conservative than on profile 1. For example, many false positives are observed in the ANN processing of profile 1, since the ANN keeps more early-time gates. However, the data misfit after inversion are similar for the manually and ANN processed data (1.3 and 1.4, respectively). Comparing the inversion results of the two processing methods (Fig. 7) reveals a similar resistivity,  $\rho$ , and  $\Phi_{\max}$



**Figure 8.** Manual processing for profile 1. Panel (a) shows a selection of chargeability curves before processing, with red circles showing data that were negative before normalization. Panel (b) shows the same selection of chargeability curves, with gates removed by the manual processing shown in grey. Panel (c) shows the remaining data after manual processing.



**Figure 9.** Comparison of the manual and ANN processing results for profile 1 as test data, where two sets of 10 representative chargeability curves have been hand-picked, the first set presented in (a) and the second set in (b). The same chargeability curves as presented in Fig. 8 are presented in (a). The blue circles denote gates that were kept by both the manual and ANN processing routines. The red diamonds denote gates kept by manual processing, but removed by ANN processing. The cyan triangles denote gates kept by ANN processing, and removed by manual processing.

sections for profile 3. On the other hand, the  $\tau_\phi$  and C sections are different.

In order to study the  $\tau_\phi$  discrepancy between inversion models stemming from manual and ANN processing, observed at profile coordinates 45–60 m in profile 3, all model cells in this area are analysed closely. At profile coordinate 55 m we also note a peak in IP data misfit, see bottom two frames Fig. 7. In Fig. 10, the forward prediction are shown together with the corresponding data, that is remaining chargeability gates after processing. The nine chargeability curves presented in Fig. 10 are a representative selection of the curves at profile coordinates 45–60 m. The forward predictions in Figs 10(a) and (b) show some examples where the ANN performs well, whereas Figs 10(c) and (d) show some cases where the ANN performs poorly.

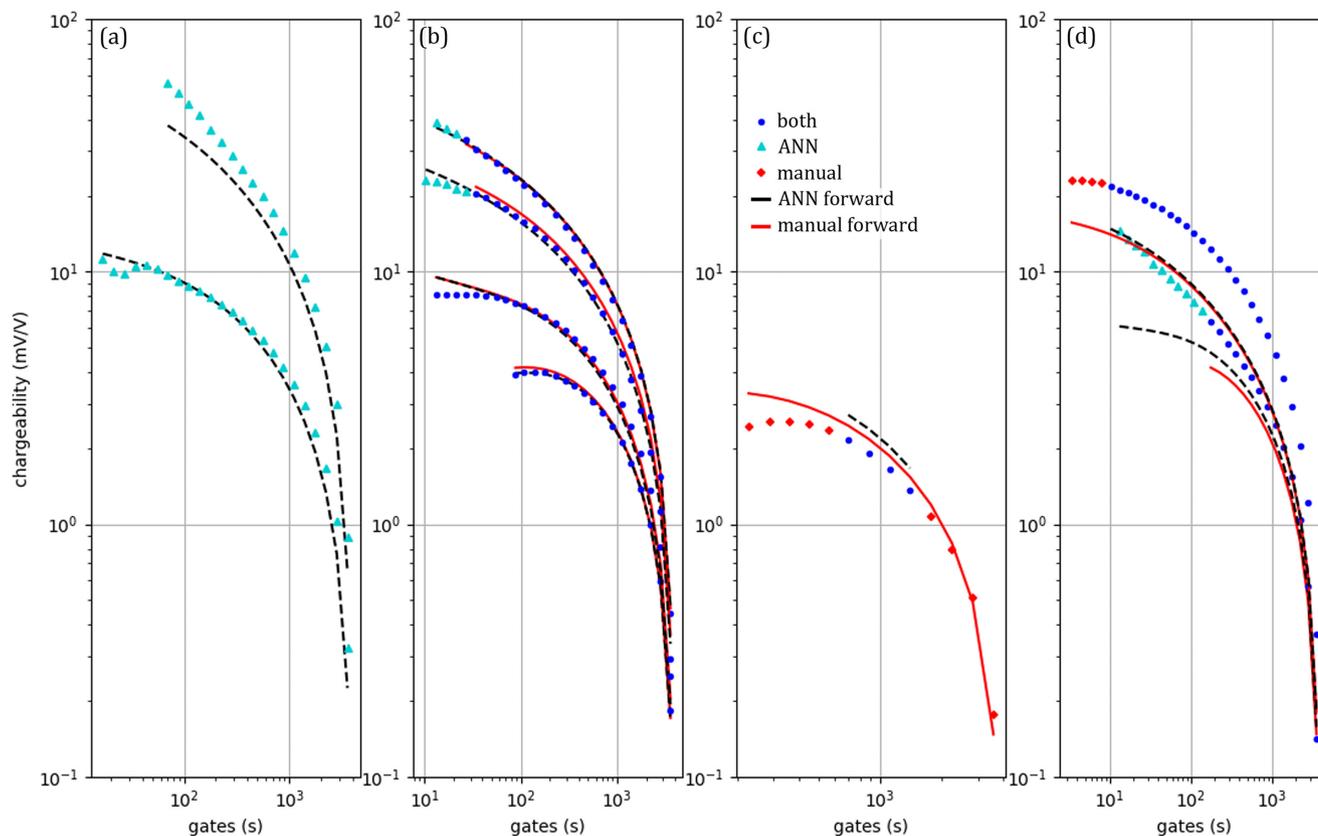
## 4 DISCUSSION

In this study, outliers are defined as data that cannot be fitted by a Cole–Cole model and are removed from the data set. A clear advantage of this model is the limited number of parameters. However, this approach can face some limitations when real IP data, corresponding to polarization within the geological units where the electrical current, does not follow a Cole–Cole behaviour. This can

be the case if, for example, polarization occurs at different scales at the same time, in which case a single Cole–Cole model may not represent the complexity of the processes in a satisfactory manner (Lévy *et al.* 2019). Given the resolution and signal level of time-domain IP data, adding parameters to the model space is not an option. Approaches including Debye decomposition, limited to 2 or 3 fixed relaxation times, could be considered in the future but this is beyond the scope of this paper.

The processing of profile 3 took 6.3 s using a trained ANN, including the sequence sorting, outlier detection algorithm, and writing the processing result to the disk. On the other hand, the manual processing of a profile of comparable size takes about 4 hr, and processing the whole survey takes several days. ANN processing significantly increases the computational efficiency, as well as automates IP processing. The training of the ANN took 3.5 hr, and might take less/more time to train depending on the size of the training data set.

The choice of profiles to use as training data influences the ANN performance. We tried training the ANN using profiles 2 and 3 as training/validation data, and profile 8 as a test set. This yielded a different result. The final accuracy was 89.8 per cent, however, the precision and recall values were lower, with a precision of 83.1 and recall of 84.9. Generally, profile 3 contained a lot of data, and was difficult in regards to processing, that is more outlier gates,



**Figure 10.** Chargeability curves, including data and forward predictions. Panel (a) shows two curves (corresponding to two quadruples) removed during manual processing, but not during ANN processing. Panel (b) shows four curves where both the manual and ANN processing do a good job, although the ANN slightly improves the results. Panel (c) shows a case where the ANN processing removes too many gates; it should be noted that it is the only quadruple we found resulting in such a poor processing. Panel (d) shows two chargeability curves (quadruples) that were kept by both the manual and ANN processed data but were not correctly fitted in either case.

outlier curves, and more noisy chargeability curves. This means the correlation between the input data,  $X$ , and the corresponding labels,  $y$ , is lower. For the purpose of this study, profile 3 was used as a test set, since this yielded better ANN performance. However, it is important that the profiles used as training data are processed in a similar and consistent manner. We tried to include profile 1 in the training set, which decreased the overall performance of the ANN, since the training data set contained discrepant or contradictory data. This was reflected in the ANN processing result, where data was processed in a more conservative fashion. Therefore, increasing the size of the training data set does not necessarily improve the performance of the ANN. Especially, if the added data is of poor quality. The final ANN used a training data set comprised of profiles 2 and 8, amounting to 17.8 per cent of the data set. Alternatively, using profiles 2 and 3 as training data, which is a larger training data set amounting to 23.1 per cent of the data set, yielded worse performance. The precision and recall of the smaller training data set was 91.1 and 88.3 per cent, respectively. The larger training data set had a precision of 81 per cent and a recall of 85 per cent.

The performance of the outlier detection algorithm can be seen in the summary statistics before and after removing outliers. The total percentage of positive predictions that were correctly classified was slightly reduced (precision), but the percentage of actual positive cases is increased (recall, see Table 5). This means that the number of false negatives are reduced, and moved to true positives in cases where we have identified the same outliers as in the manual

processing. Outlier curves that are removed manually, but not removed by the outlier detection algorithm, decreases the accuracy and precision, since they increase the number of false positives. Overall, the outlier detection algorithm removes obvious outlier curves, but does not mimic the manual approach.

The ANN processing results are consistent and reproducible in the sense that the same ANN will always yield the same processing results. In our experience, in some cases, the ANN can also yield better results than manual processing, for example for profile 1. However, this is only the case when the ANN was trained using a good and consistent training data set. A common problem in ML is bad training data. Here, bad training data can be considered as data with a weak correlation between the patterns of the input data,  $X$ , and the corresponding labels,  $y$ . In such cases, the training data needs to be cleaned prior to training the ANN. One issue we encountered, was that detected outlier curves are removed by setting all gates to 'off', for the entire chargeability curve. Such an outlier curve, although it still might portray a decaying chargeability signal, is removed based on the fact that it has a too high, or too low, signal level, or is too different from neighbouring curves. Such outlier curves confuse the ANN. The loss function cannot be reduced beyond a certain point if it sometimes has to remove entire chargeability curves with a decaying pattern, and in other cases keep them. In our case, outliers are dealt with using an outlier detection routine, and therefore all chargeability curves, which were considered outlier curves, were removed from the training data. This improved the

overall performance of the ANN, as well as the generalization, since these outlier chargeability curves no longer confuse the network during training.

Due to the nature of human-based manual processing, the training data can be strongly affected by human error, for example when the processing used in the training has not been carried out in a consistent manner. In manual processing, as opposed to rule-based approach, similar soundings can have small fluctuations as to whether a gate is removed or not. This will confuse the ANN. If a given gate is kept in one chargeability curve, and removed in another similar curve, then the loss function cannot do better than favouring one or the other—never both. The consequence of this is that the performance can not increase beyond a certain limit, since the network cannot prioritize such contradictory data at the same time. In order to improve the training data, a rule-based approach should be used to produce a relatively consistent training data set.

Testing the ANN on profile 3, it had an accuracy of 90.8 per cent, a precision of 93.9 per cent, and recall of 84.4 per cent, which means differences between the ANN and manual processing do exist. Generally, the ANN keeps gates at much earlier times, and is less conservative, as is evident in Fig. 6. The outlier detection algorithm used in tandem with the ANN also does not remove all outliers. For profile 3 the outlier curve detection algorithm identified 95 outlier curves, while during manual processing 83 outlier curves were identified. This is the key difference we observe between the manual and ANN processing. Furthermore, the ANN processing is judged on its ability to process the gates correctly, but also on how well the outlier curve detection algorithm works. If the outlier detection does not identify the same outliers as the manual processing, or removes outlier curves that were not removed manually, this also increases the uncertainty of the ANN processing. Overall, ANN processing should never be expected to be identical to the manual processing, and with precision of 93.9 per cent and recall of 84.4 per cent, respectively, we are confident that the ANN works. This is further supported by the fact that the IP inversion results are similar in nature, except for the  $\tau_\phi$  and  $C$  section. The differences between the  $\tau_\phi$  and  $C$  sections in the inversion results based on the manual and ANN processing results, can be seen clearly in Fig. 10, where some key differences are presented in an area with a high data misfit, that is between profile coordinates 45–60 m. Generally, in areas of high data misfit, the inversion result is influenced by chargeability curves in the vicinity of the high misfit. Since there are some significant differences in the processing result in the area presented (Fig. 10), it is expected that there will also be some differences in the final inversion result. The ANN is not as conservative as manual processing, both in regards to removing outlier curves, and individual gates. This slightly alters the forward predictions, in some cases even improving the data misfit (Figs 10a and b). However, in one case the ANN did not recognize the chargeability pattern properly, and did a poor processing job (Fig. 10c). This is an indication that we either need more training instances that resembles the pattern, or that there is an instance in the training set that is poorly processed and is confusing the ANN. Also a few examples are presented where both the manual and ANN processing yielded high data misfits (Fig. 10d).

The presented ANN is trained for 36 gate chargeability curves, and hence the input vector consists of a 36-element array (Table 1). If the measurement and/or gating is different, and contains fewer, or additional gates, then the presented ANN will not work. To generalize the ANN it would be necessary to up-sample the curves to a larger number of gates. Then, the network is trained using the up-sampled curves, and when a survey containing, for example 25

gates is conducted, it is simply up-scaled, and then fed through the ANN. Such a generalization would also mean that any IP data set could be used for training and prediction.

In its current form, the presented ANN processing scheme can be applied as a tool to aid in manual IP processing. Once a subset of the data has been processed, in our case about 15–20 per cent, an ANN can be trained. The trained ANN can then process the remaining data set in less than a minute. The end goal is a generalized ANN, which can be applied on other surveys, but that requires training data from many different surveys. The ANN recognizes outlier patterns found in the chargeability curves it has been shown, and therefore does not know how to deal with new outlier patterns. Applying the ANN on a different survey than it was trained on might yield variable results, depending on how similar the surveys are. Future plans include expanding the training data set and testing the ANN on additional surveys.

## 5 CONCLUSIONS

We apply ANNs for quality assurance and processing of time-domain induced polarization data. The presented results revealed that ML works as a viable tool for processing chargeability curves, and was tested on a real-world survey from Grindsted, Denmark. The Grindsted survey consists of eight profiles, 2085 chargeability curves per profile on average, amounting to 16 511 curves in total. Two profiles, profiles 2 and 8, which contain 2946 (17.8 per cent) chargeability curves, were manually processed and used to train an ANN, which in turn is used to process the remaining data. The network can process an IP profile in 6–15 s, and took 3.5 hr to train on a modern computer system with a dedicated graphics processing unit (GPU). The manual processing of a data set can take days, depending on the experience of the person processing, how large the data set, and the amount of noise present in the IP data. The presented approach is consistent and rapid, in contrast to the manual approach, which is time consuming and relatively often inconsistent.

We showed that by processing about 18 per cent of the induced polarization data, we could train a neural network to process the remaining 82 per cent of the data set in less than a minute. Although, the performance of the neural network was good, there were also some caveats. The biggest caveat was the inability to identify outlier curves based on one chargeability curve at a time. This problem was remedied by applying a outlier curve detection scheme. After incorporating the outlier detection into the neural network, a precision of 93.9 per cent and a recall of 84.4 per cent was achieved. The inversion results based on the neural network processed data, was comparable to the inversion results based on the manually processed data. The average misfit for the manual based inversion in the test profile, that is profile 3, was 2.7, while the same data misfit for the neural network based inversion was 2.8.

Even though the ANN performed well, strategies still need to be developed and tested for improving and generalizing neural networks for quality assurance and processing of induced polarization data. Such strategies include finding a better way to obtain consistent training data, such as exploring rule-based processing as a means of making training data. Additionally, a avenue could be to explore more complex ML methods, such as convolutional neural networks, to see if they can improve the results. This is a trade-off between computation time, performance, and availability of training data. The more complex the network, the larger the computational demand, but with better overall performance, and vice versa.

## ACKNOWLEDGEMENTS

We thank Anders Kristian Kühl and Pradip Kumar Maurya for the many fruitful discussions on induced polarization. The presented research was conducted as part of Guided Injection Remediation (GIRem), and in affiliation. The funding for GIRem is provided by the Innovation Fund Denmark. The Grindsted data presented in this paper is not publicly available.

## REFERENCES

- AarhusGeoSoftware, 2020. AarhusWorkbench, <https://www.aarhusgeosoftware.dk>.
- Al Shalabi, L. & Shaaban, Z., 2006. Normalization as a preprocessing engine for data mining and the approach of preference matrix, in *Proceedings of the 2006 International Conference on Dependability of Computer Systems*, pp. 207–214, IEEE.
- Auken, E. & Christiansen, A.V., 2004. Layered and laterally constrained 2D inversion of resistivity data, *Geophysics*, **69**(3), 752–761, doi.org/10.1190/1.1759461.
- Auken, E. *et al.*, 2015. An overview of a highly versatile forward and stable inverse algorithm for airborne, ground-based and borehole electromagnetic and electric data, *Explor. Geophys.*, **46**(3), 223–235.
- Bergen, K.J., Maarten, V.M. & Beroza, G.C., 2019. Machine learning for data-driven discovery in solid earth geoscience, *Science*, **363**(6433), doi:10.1126/science.aau0323.
- Bertin, J. & Loeb, J., 1976. *Experimental and Theoretical Aspects of Induced Polarization*, Vol. 1, Schweizerbart Science Publishers.
- Bishop, C.M., 2006. *Pattern Recognition and Machine Learning*, Springer.
- Bodmer, R., Ward, S.H. & Morrison, H.F., 1968. On induced electrical polarization and groundwater, *Geophysics*, **33**(5), 805–821, doi.org/10.1190/1.1439972.
- Börner, F., Schopper, J. & Weller, A., 1996. Evaluation of transport and storage properties in the soil and groundwater zone from induced polarization measurements, *Geophys. Prospect.*, **44**(4), 583–601.
- Chongo, M., Christiansen, A.V., Fiandaca, G., Nyambe, I.A., Larsen, F. & Bauer-Gottwein, P., 2015. Mapping localised freshwater anomalies in the brackish paleo-lakesediments of the Machile–Zambezi basin with transient electromagneticsounding, geoelectrical imaging and induced polarisation, *J. appl. Geophys.*, **123**, 81–92.
- Clark, J., Koprinska, I. & Poon, J., 2003. A neural network based approach to automated e-mail classification, in *Proceedings IEEE/WIC International Conference on Web Intelligence (WI 2003)*, pp. 702–705, IEEE, doi:10.1109/WI.2003.1241300.
- Cole, K.S. & Cole, R.H., 1941. Dispersion and absorption in dielectrics. I. Alternating current characteristics, *J. Chem. Phys.*, **9**, 341–351, doi.org/10.1063/1.1750906.
- Dahlin, T. & Bernstone, C., 1997. A roll-along technique for 3D resistivity data acquisition with multi-electrode arrays, in *Proceedings of the 10th EEGS Symposium on the Application of Geophysics to Engineering and Environmental Problems*, pp. cp–204, European Association of Geoscientists & Engineers, doi:10.3997/2214-4609-pdb.204.1997\_096.
- Dahlin, T. & Leroux, V., 2012. Improvement in time-domain induced polarization data quality with multi-electrode systems by separating current and potential cables, *Near Surf. Geophys.*, **10**(6), 545–565.
- Edwards, L.S., 1977. A modified pseudosection for resistivity and IP, *Geophysics*, **42**(5), 1020–1036, doi.org/10.1190/1.1440762.
- Fiandaca, G., 2018. Induction-free acquisition range in spectral time- and frequency-domain induced polarization at field scale, *Geophys. J. Int.*, **220**(3), e1–e66.
- Fiandaca, G., Ramm, J., Binley, A., Gazoty, A., Christiansen, A.V. & Auken, E., 2013. Resolving spectral information from time domain induced polarization data through 2-d inversion, *Geophys. J. Int.*, **192**, 631–646.
- Fiandaca, G., Christiansen, A.V. & Auken, E., 2015. Depth of investigation for multi-parameters inversions, in *Proceedings of the Near Surface Geoscience 2015-21st European Meeting of Environmental and Engineering Geophysics (Extended Abstract)*, doi:10.3997/2214-4609.201413797.
- Fiandaca, G., Madsen, L.M. & Maurya, P.K., 2018. Re-parameterisations of the Cole–Cole model for improved spectral inversion of induced polarization data, *Near Surf. Geophys.*, **16**(4), 385–399.
- Gazoty, A. *et al.*, 2012a. Application of time domain induced polarization to the mapping of lithotypes in a landfill site, *Hydrol. Earth Syst. Sci.*, **16**(6), 1793.
- Gazoty, A., Fiandaca, G., Pedersen, J., Auken, E. & Christiansen, A.V., 2012b. Mapping of landfills using time-domain spectral induced polarization data: the Eskelund case study, *Near Surf. Geophys.*, **10**(6), 575–586.
- Géron, A., 2019. *Hands-On Machine Learning with Scikit-Learn, Keras, and Tensor Flow: Concepts, Tools, and Techniques to Build Intelligent Systems*, O'Reilly Media.
- Jayalakshmi, T. & Santhakumaran, A., 2011. Statistical normalization and back propagation for classification, *Int. J. Comp. Theory Eng.*, **3**(1), 1793–8201.
- Kingma, D.P. & Ba, J., 2014. Adam: a method for stochastic optimization, arXiv preprint arXiv:1412.6980.
- Lévy, L., Gibert, B., Sigmundsson, F., Parat, F., Deldicque, D. & Hersir, G.P., 2019. Tracking magmatic hydrogen sulphur circulations using electrical impedance: complex electrical properties of core samples at the Krafla volcano, Iceland, *J. geophys. Res.*, **124**(3), 2492–2509.
- Loke, M.H., 2004. Tutorial: 2-D and 3-D electrical imaging surveys, Set of notes, available at: [https://www.researchgate.net/profile/Meng\\_Loke/publication/264739285\\_Tutorial\\_2-D\\_and\\_3-D\\_Electrical\\_Imaging\\_Surveys](https://www.researchgate.net/profile/Meng_Loke/publication/264739285_Tutorial_2-D_and_3-D_Electrical_Imaging_Surveys).
- Loke, M.H. & Barker, R., 1996. Rapid least-squares inversion of apparent resistivity pseudosections by a quasi-newton method I, *Geophys. Prospect.*, **44**(1), 131–152.
- Marshall, D.J. & Madden, T.R., 1959. Induced polarization, a study of its causes, *Geophysics*, **24**(4), 790–816, doi.org/10.1190/1.1438659.
- Maurya, P.K., Rønde, V.K., Fiandaca, G., Balbarini, N., Auken, E., Bjerg, P.L. & Christiansen, A.V., 2017. Detailed landfill leachate plume mapping using 2D and 3D electrical resistivity tomography - with correlation to ionic strength measured in screens, *J. appl. Geophys.*, **138**, 1–8.
- Maurya, P.K., Balbarini, N., Møller, I., Rønde, V., Christiansen, A.V., Bjerg, P.L., Auken, E. & Fiandaca, G., 2018. Subsurface imaging of water electrical conductivity, hydraulic permeability and lithology at contaminated sites by induced polarization, *Geophys. J. Int.*, **213**(2), 770–785.
- Menke, W., 2012. Chapter 9 - Nonlinear inverse problems, in *Geophysical Data Analysis: Discrete Inverse Theory*, 3rd ed, pp. 163–188, ed. Menke, W., Academic Press.
- Olsson, P.-I., Dahlin, T., Fiandaca, G. & Auken, E., 2015. Measuring time-domain spectral induced polarization in the on-time: decreasing acquisition time and increasing signal-to-noise ratio, *J. appl. Geophys.*, **123**, 316–321.
- Olsson, P.-I., Fiandaca, G., Larsen, J.J., Dahlin, T. & Auken, E., 2016. Doubling the spectrum of time-domain induced polarization by harmonic de-noising, drift correction, spike removal, tapered gating and data uncertainty estimation, *Geophys. J. Int.*, **207**(2), 774–784.
- Orozco, A.F., Kemna, A., Oberdörster, C., Zschornack, L., Leven, C., Dietrich, P. & Weiss, H., 2012. Delineation of subsurface hydrocarbon contamination at a former hydrogenation plant using spectral induced polarization imaging, *J. Contamin. Hydrol.*, **136**, 131–144.
- Orozco, A.F., Williams, K.H. & Kemna, A., 2013. Time-lapse spectral induced polarization imaging of stimulated uranium bioremediation, *Near Surf. Geophys.*, **11**(5), 531–544.
- Orozco, A.F., Gallistl, J., Bücker, M. & Williams, K.H., 2018. Decay curve analysis for data error quantification in time-domain induced polarization imaging, *Geophysics*, **83**(2), E75–E86.
- Paszke, A., *et al.*, 2019. PyTorch: An imperative style, high-performance deep learning library, in *Advances in Neural Information Processing Systems*, pp. 8024–8035, <http://papers.nips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library>.
- Pelton, W.H., Ward, S.H., Hallof, P.G., Sill, W.R. & Nelson, P.H., 1978. Mineral discrimination and removal of inductive coupling with multifrequency IP, *Geophysics*, **43**(3), 588–609.
- Prechelt, L., 1998. Early stopping-but when?, in *Neural Networks: Tricks of the Trade*, pp. 55–69, Springer.

- Robinson, J., Slater, L., Weller, A., Keating, K., Robinson, T., Rose, C. & Parker, B., 2018. On permeability prediction from complex conductivity measurements using polarization magnitude and relaxation time, *Water Resour. Res.*, **54**(5), 3436–3452.
- Rumelhart, D.E., Hinton, G.E. & Williams, R.J., 1986. Learning representations by back-propagating errors, *Nature*, **323**(6088), 533–536.
- Slater, L., 2007. Near surface electrical characterization of hydraulic conductivity: from petrophysical properties to aquifer geometries—a review, *Surv. Geophys.*, **28**(2–3), 169–197.
- Sumner, J.S., 1976. Principles of induced polarization for geophysical exploration, in *Developments in Economic Geology*, Vol. 5, Elsevier Scientific Publishing Company Amsterdam.
- Telford, W.M., Geldart, L.P. & Sheriff, R.E., 1990. *Applied Geophysics*, 2nd edn, Cambridge Univ. Press.
- Tsoumakas, G. & Katakis, I., 2007. Multi-label classification: an overview, *Int. J. Data Warehous. Mining (IJDWM)*, **3**(3), 1–13.
- Vacquier, V., Holmes, C.R., Kintzinger, P.R. & Lavergne, M., 1957. Prospecting for ground water by induced electrical polarization, *Geophysics*, **22**(3), 660–687.
- Versteeg, R. & Johnson, T., 2008. Using time-lapse electrical geophysics to monitor subsurface processes, *Leading Edge*, **27**(11), 1488–1497.
- Wainwright, H.M., Flores Orozco, A., Bücken, M., Dafflon, B., Chen, J., Hubbard, S.S. & Williams, K.H., 2016. Hierarchical Bayesian method for mapping biogeochemical hot spots using induced polarization imaging, *Water Resour. Res.*, **52**(1), 533–551.