A multidimensional AI-trained correction to the 1D approximate model for Airborne TDEM sensing

Wouter Deleersnyder[1,2], David Dudal[1,3], Thomas Hermans[2]

[1]KU Leuven Campus Kortrijk - KULAK, Department of Physics, Etienne Sabbelaan 53, 8500 Kortrijk, Belgium.
E-mail: wouter.deleersnyder@kuleuven.be
[2]Ghent University, Department of Geology, Krijgslaan 281 - S8, 9000 Gent, Belgium
[3]Ghent University, Department of Physics and Astronomy, Ghent, Krijgslaan 281 - S9, 9000 Gent, Belgium

ABSTRACT

The computational resources required to solve the full 3D inversion of time-domain electromagnetic data are immense. To overcome the time-consuming 3D simulations, we construct a surrogate model, more precisely, a data-driven statistical model to replace the 3D simulations. It is trained on 3D data and predicts the approximate output much faster. We construct a surrogate model that predicts the discrepancy between a 1D subsurface model and a deviation of the 1D assumption. The latter response is fastly computable with a semi-analytical 1D forward model. We exemplify the approach on a two-layered case. The results are encouraging even with few training samples. Given the computational cost related to the 3D simulations, there are limitations in the number of training samples that can be generated. In addition, certain applications require a wide range of parameters to be sampled, such as the electrical conductivity parameters in a saltwater intrusion case. The challenge of this work is achieving the best possible accuracy with only a few thousand samples. We propose to view the performance in terms of learning gain, representing the gain from the surrogate model whilst still acknowledging a residual discrepancy. Our works open new avenues for effectively simulating 3D TDEM data.

1 INTRODUCTION

Airborne Electromagnetic (AEM) mapping is an increasingly used method to image near-surface geological features over large areas via the bulk electrical resistivity. Common applications are, for example, in mineral exploration (Macnae and Milkereit, 2007), contamination (Pfaffhuber et al., 2017), hydrogeological mapping (Mikucki et al., 2015), and saltwater intrusion (Siemon et al., 2019; Deleersnyder et al., 2023c). Part of the AEM systems collects time-domain electromagnetic (TDEM) data, but those systems require significantly more computer power to run 3D modelling simulations, compared to frequency-domain electromagnetic (FDEM) data. 3D FDEM forward modelling is becoming applicable for smaller areas (Ansari et al., 2017; Cockett et al., 2015; Haber et al., 2004; Werthmüller et al., 2019), while 3D TDEM forward modelling remains impractical (Börner et al., 2015; Haber et al., 2007). For example, Engebretsen et al. (2022) could carry out a 2D TDEM inversion with 418 soundings, which took 30 hours, 97 GB of memory for only 16 iterations. Auken et al. (2017) argues that the computational demand for a full 3D AEM survey is still too large so that only few companies are capable of realizing such inversions.

By consequence, quasi-2D/3D inversion methods, using a much faster and approximate (1D) forward model, methods are standard for large-scale mappings. The 1D forward model assumes cylindrical

symmetry and is thus limited to horizontally layered conductivity models. Using that forward model in a quasi-2D/3D method may yield erroneous interpretations and the results may better not be interpreted quantitatively. Deleersnyder et al. (2022) provide an appraisal tool which verifies whether an erroneous interpretation may occur as a result of the approximate 1D forward model, and indicates areas that should be reinterpreted with a full 3D forward model.

To overcome the time-consuming simulations, a common data-driven approach is to construct surrogate models. It is a statistical model that can approximate the simulation output. A single evaluation of the trained statistical model is much faster than a simulation. The statistical relation is built from pairs of full simulation data and its inputs. Surrogate modelling is thus a case of supervised machine learning. Surrogate modelling and machine learning may replace 3D forward modelling on a mesh during a 3D inversion, sometimes coined Fast Forward Modelling (FFM).

The existing literature of using Machine Learning (ML) and 1D forward modelling is already significant. Bording et al. (2021) use a conventional Neural Network (NN) for ground-based TDEM data and claims a factor 13 speed up compared to the accurate 1D forward models with 98% accuracy within the 3% relative error limit. Other examples of 1D surrogate models are found in surface wave dispersion (Hou et al., 2019), seismic and marine controlled-source EM data (Puzyrev and Swidinsky, 2021) and seismic full-waveform inversion (Wu et al., 2018). For AEM modelling, Asif et al. (2022) train a NN for 1D forward modelling, including the flight altitude as additional parameter, which adds complexity to the learning problem. Additionally, two separate NNs were trained, one to predict the forward response and one for the partial derivatives. A speed-up with a factor 50 is claimed. Wu et al. (2023) have trained an alternative NN with a long short-term memory (LSTM) architecture with similar 'functionalities' and claim a factor 2700 speed-up. The step towards 3D modelling has already been made for gravity data (Zhang et al., 2021) or frequency domain magnetotellurics (Conway et al., 2019; Peng et al., 2023), which suffer less from the computational burden to generate training data.

Other work shows that NNs can be used for a direct inversion (Noh et al., 2020; Feng et al., 2020; Bai et al., 2020; Li et al., 2020; Puzyrev and Swidinsky, 2021; Laloy et al., 2021). This avoids the potentially time-consuming optimization step of the inverse problem, but one loses control over the specific tunable aspects of the inversion process. The trained NN needs to handle the non-uniqueness of the geophysical problem, as well as the noise. By training a separate forward/surrogate model, one retains control over the inversion: The type of regularization and the regularization parameter(s) can be easily modified. This allows to use the trained surrogate forward model in flexible regularization schemes where, for example, the sharpness can be tuned (Deleersnyder et al., 2023c; Klose et al., 2022).

While much work has been done in exploring the surrogate modelling and Machine Learning for 1D forward problems in TDEM problems, the step towards multidimensional forward modelling, to our knowledge, has not yet been made. While the literature above has used up to a million samples in the training dataset, this is not feasible for 3D TDEM simulations. The challenge of this work is achieving the best possible accuracy with only a few thousand (5090) samples. We, therefore, limit the parameter space to subsurface models with two layers. Rather than predicting the 3D data directly, we predict the relative error between the 3D and 1D data (Deleersnyder et al., 2023b). The computation of the 1D approximate forward model is fast and explains most of the variability in the geophysical data. The discrepancy, originating from the multidimensional nature of the subsurface, is predicted with machine learning tools. We focus on a minimal, well-defined context, namely a central-loop airborne system flying at 40 m altitude, covering an area with high conductivity contrasts, such as when imaging a fresh saltwater interface.

We have opted for more classical ML methods which are well-established in geosciences and are more insightful, rather than using (advanced) neural networks that are often data hungry for the learning step. We will use Gaussian Process Regression, typically used in forecasting (Ferkous et al., 2021; Satija and Caers, 2015), kriging (Gonçalves et al., 2022; Ching and Yoshida, 2023) or in geohazard mitigation applications (Dikshit et al., 2021; Gao et al., 2022). We work with the common Principal Component Analysis (PCA) as dimensionality reduction technique, which finds many applications (Grana et al.,

2019; Scheidt et al., 2018; Michel et al., 2020). Functional Component Analysis, e.g., used in Satija and Caers (2015), is less common in geosciences (Suhaila and Hamdan, 2023) as a dimensionality reduction technique. Demonstrating the performance of a surrogate model for a simplified case without advanced ML leaves room to use more advanced ML in more extensive contexts (with more parameters).

We first discuss the data generation in Section 2.1 in terms of low and high fidelity data, the 1D forward model and 3D simulations respectively. In Section 2.1.2, we discuss the data transformations used in this work. Inspired by the characteristics of the transformed data (discussed in Section 3.1) and preliminary tests, a Gaussian Process Regression (GPR) is picked as surrogate model (Section 2.2). In Section 3.2, the performance of the trained surrogate models is described, suggesting successful routes for future work.

2 METHODS

2.1 Data generation

The objective of this work is to construct an ML-based correction model on an incomplete, but fast evaluable low-fidelity forward model. For TDEM data we use a 1D semi-analytical model (Hunziker et al., 2015), which assumes horizontal layers, to generate Low Fidelity (LF) data. A Python, open-source implementation is used (Werthmüller, 2017) . The correction model is trained using High Fidelity (HF) data, which must be generated by a forward model that can accommodate multidimensional structures. The only viable method is through numerical simulations, described in Section 2.1.1. Data transformations, before training, are described in Section 2.1.2.

The surrogate correction model predicts the target with the following listed features:

- The depth d (in meter) of the interface between the two layers, $d \in [10, 50]$,

- the angle θ (in $°$) of the dipping layer ($0°$ being purely horizontal), $\theta \in [-75, 75]$ and,

- the Electrical Conductivity of the first ($EC1$) and second ($EC2$) layer (in Sm^{-1}), $EC1, EC2 \in [0, 0.75]$..

To test the generalizability of this approach, the prior (or the training data) is not specifically adapted to a specific application. The data points are sampled uniformly within the feature space within the specified boundaries.

2.1.1 Finite Volume Method simulations to generate High Fidelity data

The finite volume method (FVM) is a numerical discretization technique for representing and evaluating partial differential equations by conversion to algebraic equations. The FVM evaluates exact expressions for the average value of the quantities over some (finite) volume and uses this discretization to construct an approximate solution. In this work, we resort to the SimPEG package (Cockett et al., 2015; Heagy et al., 2017), which is an open-source project in Python which implements the Maxwell's equations and the most common geophysical electromagnetic sources, and the parallel, direct PARADISO-solver (Schenk and Gärtner, 2004).

The High Fidelity mesh in this work is a 3D regular rectangular mesh with 288 000 cells. The core ranges from -50 to 50 m in the lateral orientations and from -140 to 60 m in the vertical orientation with cell sizes of 5-by-5 m. Outside the core region, cells are progressively increased with a rate of 20% up to a cell size of 200 m. The total mesh extent is 2,35 km in the lateral orientation and 2,80 km in the vertical orientation. The accuracy of the HF data is illustrated on a two horizontally layered case with 0.1 and 0.4 Sm^{-1} and an interface at 50 m depth. The HF and LF data are shown in Figure 1. It is unavoidable to have a small discrepancy. In this case the relative discrepancy averages 0.19% per time gate and a maximum of 0.8% at late time gates; see Figure 1.

2.1.2 Dataset and transformations

The training and test dataset consists of 5090 and 566 samples, respectively. The target variable $\frac{\partial \mathbf{B}}{\partial t}$ is sampled logarithmically from 10^{-5} to 10^{-2} seconds with 100 time steps. The target output is thus a vector of 100 data points, calling for a dimensionality reduction, which is described in Section 2.2.1. Before the
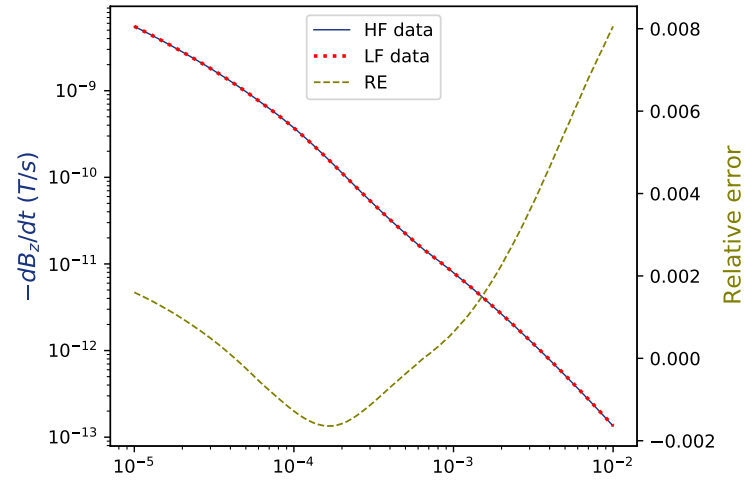
Figure 1. Illustrative comparison of Low Fidelity (*LF*) and High Fidelity (*HF*) data with horizontal layers for 0.1 and 0.4 Sm^{-1} and an interface at 50 m depth illustrate the accuracy of the *HF* mesh. The relative discrepancy between the *HF* and *LF* data is also shown; it indicates larger errors at later times.

dimensionality reduction, the magnetic data is transformed to a relative error *RE* between the *LF* and *HF* data.

$$RE = \frac{HF - LF}{LF} \tag{1}$$

By predicting the error, solely the residual multidimensional variability has to be learned. Transforming the target to a relative error, rather than an absolute error, ensures that the target outputs are within the same range, as the original magnetic data (and consequently the absolute noise) spans multiple orders of magnitude.

The sensitivity of the forward response does not vary linearly with changes in the conductivity. Bording et al. (2021) suggests the following normalization for a better optimization:

$$\tilde{EC} = a + \frac{(b-a)\left(\log_{10} EC - \log_{10}(\min EC)\right)}{\log_{10} \max(EC) - \log_{10} \min(EC)}, \tag{2}$$

where $\tilde{EC}$ is the normalized electrical conductivity. The minimum and maximum correspond to the minimum and maximum electrical conductivity in the training dataset. The values are transformed within the interval $[a, b]$, which is $[-1, 1]$ in our case.

Before training and after the dimensionality reduction of the target, all features and target outputs are standardized (by substracting the mean and dividing by the standard deviation). This is a common requirement for many machine learning estimators.

2.2 Surrogate modelling

2.2.1 Dimensionality Reduction

The target consists of 100 variables. However, the smoothness of the underlying physics implies that neighbouring variables are correlated, i.e., the *HF* data varies smoothly throughout time, which calls for a dimensionality reduction technique. Dimensionality reduction techniques transform high-dimensional data into a lower dimensional space, retaining the most descriptive features, depending on the technique considered. Dimensionality reduction is often used in machine learning, as it combats the computational cost and is able to control overfitting (Lopez-Alvis et al., 2021).

In this work, the target variables are consecutively transformed to the functional and principal component space. This is called *Functional Component Analysis* (FCA) (Silverman, 1996; Satija and Caers, 2015). First, the target is represented in a functional data space via Functional Data Analysis (FDA) (Ramsay et al., 2006), which represents the original data as a linear combination of differentiable mathematical

functions:

$$y(t) \simeq \sum_{k}^{K} c_k \phi_k(t), \tag{3}$$

where $\phi_k(t)$ is a set of basis functions. The true, underlying basis of our target is unknown. Therefore, we view each functional datum (each sample), using a non-parametric representation, meaning that each value of the 100 target variables are stored in a finite grid of points. The underlying physics is smooth and thus the functional data are smooth and one may use interpolation to evaluate between the grid. With FDA, we represent that data in a parametric fashion, using the linear combination of Eq. (3). Common systems rely on the Fourier basis, monomial basis, finite element basis etc. We work with the B-spline basis. It can represent smooth functions in a sparse fashion and it is computationally efficient.

We can now reduce the dimensionality, by representing the non-parametric (original) data in a parametric fashion, i.e. coefficients of the corresponding spline basis, where the number of coefficients $K < 100$ can still relatively accurately represent the original data.

On those functional data coefficients, Principal Component Analysis (PCA) is performed, providing a representation in terms of the principal vector basis. It consists of a linear mapping which maximizes the variance of the data in that principal vector basis or principal components. Those principal components are linearly uncorrelated and are ordered by decreasing variance. By excluding the principal components which explain little variance, the dimensionality can be further reduced. We denote the number of retained components with L.

Note that FCA is different than the method coined Functional Principal Component Analysis (FPCA), which retains modes in functional variation of longitudinal data (Hall et al., 2006).

2.2.2 Gaussian Process Regression

In Gaussian Process Regression (GPR), the concept of a Gaussian Process (GP) is used for making predictions. Williams and Rasmussen (2006) defines a GP as a collection of random variables, any finite number of which have a joint Gaussian distribution and is fully specified by second order statistics: a mean and a covariance function. In a GPR context, the latter is referred to as the *kernel*. The kernel encodes the type of structure that is found in samples drawn from the GP prior by defining the "similarity" of two datapoints combined with the assumption that similar datapoints should have similar target values. When making a prediction, we evaluate the samples at a given point, but we must first incorporate the knowledge from the training data (Scheidt et al., 2018, p. 77).

Our proposed model is a GPR and relies on the Radial Basis Function (RBF) in combination with a White Noise kernel. The RBF is a squared-exponential kernel

$$k_{\text{RBF}}(x_i, x_j) = \exp\left(-\frac{d(x_i, x_j)^2}{2\ell^2}\right), \tag{4}$$

where ℓ is the length-scale of the kernel and $d(\cdot, \cdot)$ the Euclidean distance. It is a default kernel with appealing properties (it is stationary, i.e., it does not directly depend on the value of x, but the distance between two points, and it has infinitely many derivatives) (Duvenaud, 2014, p. 9). The RBF generates smooth samples in the GP prior and is generally a good choice if no discontinuities are expected, as in our case. Increasing the length scale leads to more slowly varying GPs.

The White Noise kernel is included,

$$k_{\text{WN}}(x_i, x_j) = \begin{cases} \sigma^2 & \text{if } x_i = x_j \\ 0 & \text{else} \end{cases} \tag{5}$$

meaning that the GPR will explicitly learn the variance σ^2 of the noise, rather than to solely rely on the Tikhonov regularization (smoothing) which is assumed on the training point's covariance matrix, as the covariance matrix may be ill-conditioned, controlled by parameter α within the GPR, it works similarly as the White Noise Kernel, but the variance is not a free parameter). The main origin of the noise in the dataset is due to the discretization of the subsurface model for the *HF* fidelity data. For example, when the interface at depth d does not coincide with the edges of a cell, the depth is rounded to coincide with the edge of that cell, causing an apparent random noise. We are not interested in the noise and want to prevent overfitting, such that the fitted length scales of the RBF kernel follow the slowly varying, general

trend, rather than the short-scale variations due to the noise. The latter would limit the interpolating power of the GPR.

We train a separate GPR for each functional component, without using a multiple-output GPR. The latter could exploit output correlations, however, we train in the functional component space, where PCA was used to create uncorrelated functional components. In Section 4, we discuss the simplicity of using single-output GPRs over one multiple-output GPR which has the advantage of considering correlation between responses. This means that we will optimize for each functional component separately. The length scales and the noise-level will potentially be different for each component.

3 RESULTS

3.1 Insights from transformed data

We can easily apply the interpretation of the diffusing currents to the obtained relative error (RE) curves in Figure 2. The peaks in the RE can be explained by the bulk of the current distribution reaching the interface, where the subsurface deviates from the horizontally layered models. If we perform a sweep over the angle, we find that larger angles yield larger maxima in the RE. A sweep over the depth shows, not unexpectedly, that a deeper interface yields a maximum at later times, because the diffusing currents travel longer to reach the discrepancy at the interface. The EC of the first layer has a large impact on the 'arrival time' of these diffusing currents. We know that electrical currents diffuse faster in lower conductive media, which is confirmed in the RE curves. The EC in the second layer does not impact the time of the peaks in RE, but the figure nicely illustrates that the contrasts between two electrical conductivities are more significant to explain the value of the RE peaks, rather than the actual electrical conductivities itself. We need to check to what extent the target can represent the original relative error after dimensionality reduction. With $L < 100$, the number of functional components will no longer be able to represent all relative errors RE accurately. For this analysis, the dimensionality reduction was trained on the training dataset and its performance, expressed in Mean Absolute Error (MAE), tested on the test dataset. The MAE is computed between the original data (RE) and the back transform of the reduced data $\tilde{RE}$,

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|RE_i - \tilde{RE}_i\right|, \tag{6}$$

where n is the number of samples in the training dataset. The minimum MAE in this analysis will be the infimum for the surrogate model performance.

With the FDA, the dimensionality is reduced to $K = 10, 20$ and 30 dimensions, meaning that we use K B-spline basis functions to represent our original RE. Then, L functional components are retained after PCA on the coefficients of the FDA. The MAE is shown in Figure 3 for each $L \leq K$. Also, the MAE for all K in $[1, 30]$, while retaining *all* functional components, is shown (black dotted line). This MAE is identical to the MAE as if you would not perform PCA on the FDA coefficients.

From Figure 3, we observe that an increase in the number of B-spline functions K can better represent the original relative error RE, as expected. However, this effect becomes insignificant for $K > 20$. Note that the minimum number of B-splines is 4. This does not hold for PCA. Secondly, we learn that many functional components are required to represent the original data in a reduced dimensional space accurately. If K is low, the results suggests to retain almost all L functional components. For larger K, one can gain quite a bit of dimensionality reduction by using PCA. For example, for $K = 30$, one can easily further halve the dimensions without much loss of accuracy.

We do not decide on the optimal number of B-spline basis functions K and the number of functional components L using this analysis. It will serve as a guide to reduce the optimization for these parameters during the training of the GPR. The MAE of the *predicted* relative error in our original space will be the only criterion to determine K and L.

3.2 Performance

The performance of the trained surrogate model is measured with the MAE in terms of (original) data (RE) and the back transform of the reduced data $\tilde{RE}$. The surrogate model is trained for the number of basis functions K in the functional data analysis and the number of functional components L to retain. The parameter α controls the Tikhonov regularization (smoothing) and during optimization, we experienced that a higher value for α was required for functional components that explain less variance (apparently,
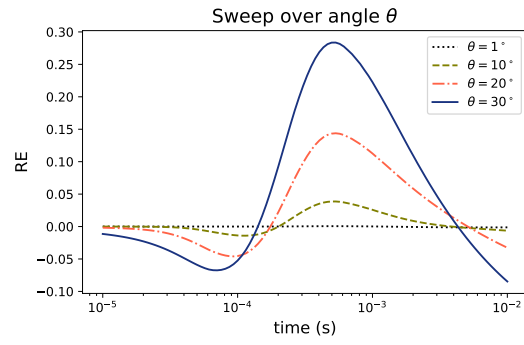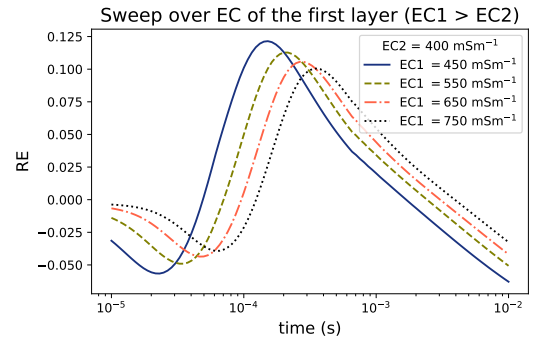
(a) Sweep over angle θ. The Relative Error (RE) increases with increasing angle θ.
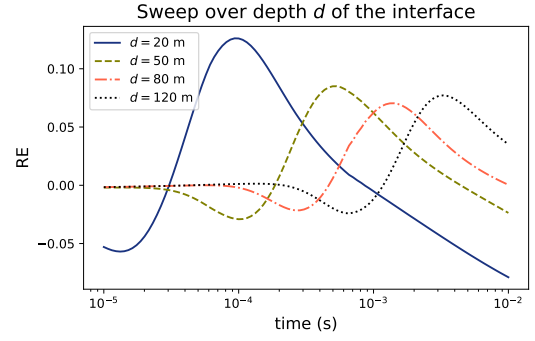
(b) Sweep over depth d of the interface. The peak of the max Relative Error (RE) shifts to later times for larger depths d.
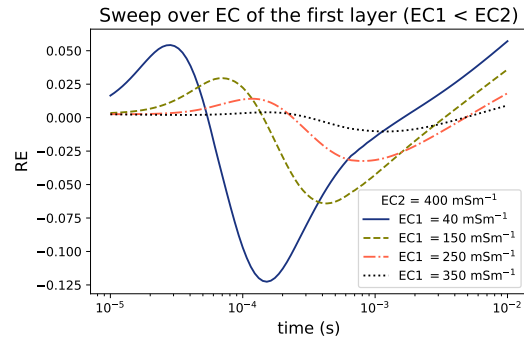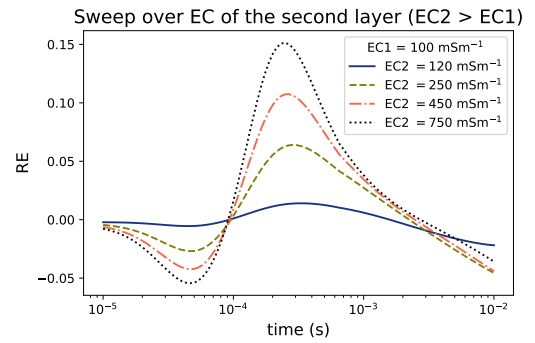
(c) Sweep over the electrical conductivity of the first layer, in the case where the conductivity of the first layer is lower than the conductivity of the second layer.

(d) Sweep over the electrical conductivity of the first layer, in the case where the conductivity of the first layer is layer than the conductivity of the second layer.
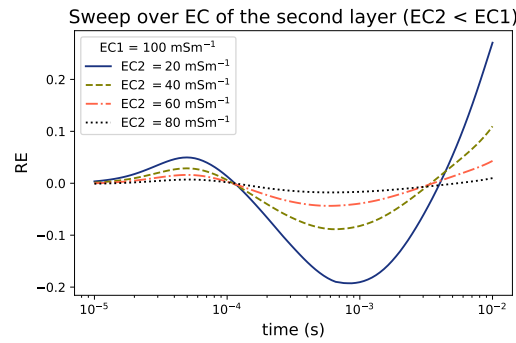
(e) Sweep over the electrical conductivity of the second layer, in the case where the conductivity of the second layer is lower than the conductivity of the first layer.

(f) Sweep over the electrical conductivity of the second layer, in the case where the conductivity of the second layer is larger than the conductivity of the first layer.

Figure 2. Relative error (RE) curves in a parameter study. In each plot, only one parameter is varied at a time. The fixed parameters were $\theta = 15°$, $d = 50$ m, EC1 = 100 mSm^{-1}, EC2 = 400 mSm^{-1}.

Table 1. The features from four samples from the test dataset shown in Figure 5.

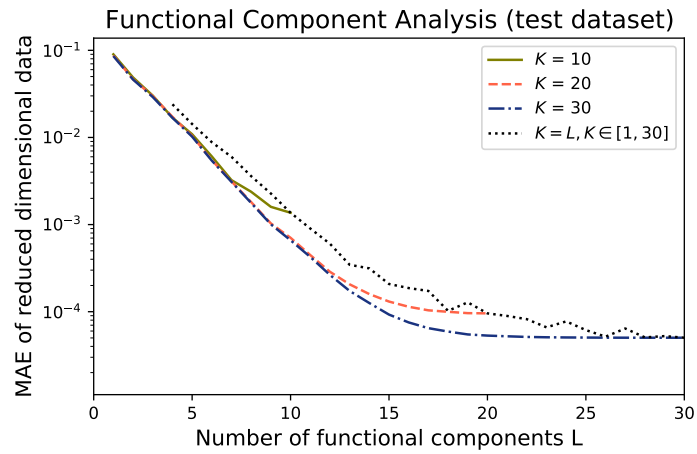|	d (m)	θ (°)	$EC1$ (S/m)	$EC2$ (S/m)
Case A	50	50.5	0.47	0.80
Case B	29	-16	0.69	0.89
Case C	11.6	-2.33	0.91	0.77
Case D	37.5	8.3	0.016	0.72

Figure 3. Mean Absolute Error (MAE) introduced by working on a reduced dimensional space (rather than inaccurate prediction of the estimator), for different retained functional components L of the K B-spline functions. MAE for $L > 30$ is approximately constant and therefore not shown.
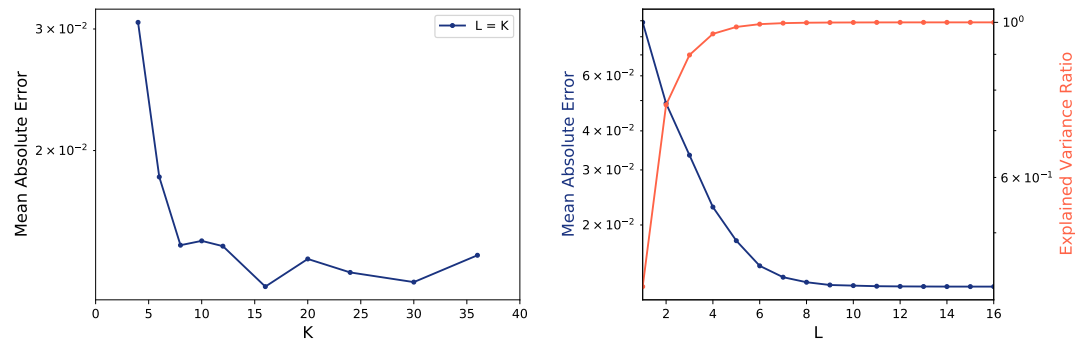

(a) The accuracy for varying the number of functional basis functions K and retaining all functional components $(K = L)$.

(b) The MAE and the (PCA) explained variance ratio for varying the number of retained functional components L for $K = 16$.
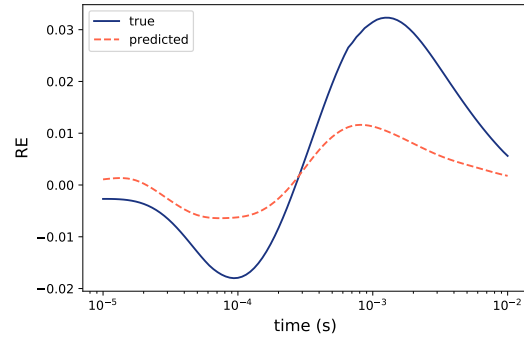
Figure 4. The accuracy (in terms of Mean Absolute Error (MAE)) of the trained surrogate model on the test dataset, varying the dimensionality reduction dimensions.
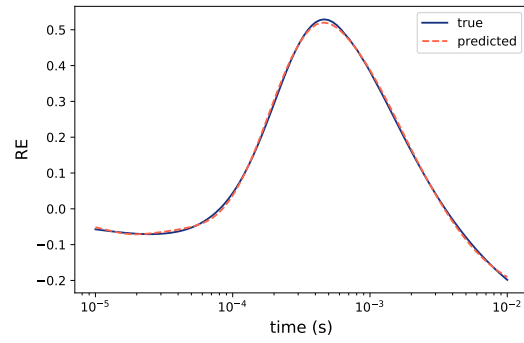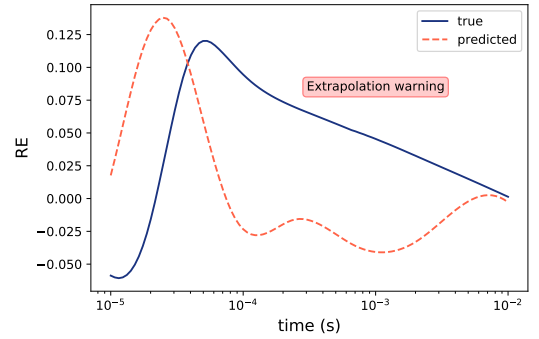
(a) Test case for $d = 50$ m, $\theta = 50.5°$, EC1 = 0.47 Sm^{-1}, EC2 = 0.80 Sm^{-1}.
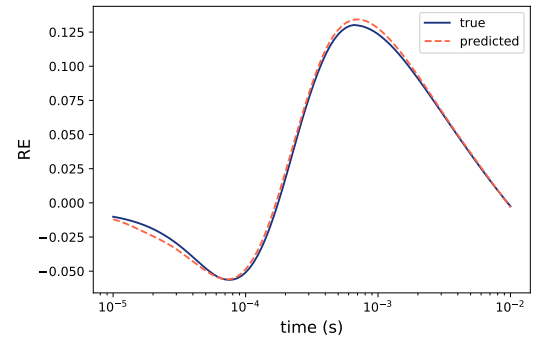
(b) Test case for $d = 29$ m, $\theta = -16°$, EC1 = 0.69 Sm^{-1}, EC2 = 0.89 Sm^{-1}.

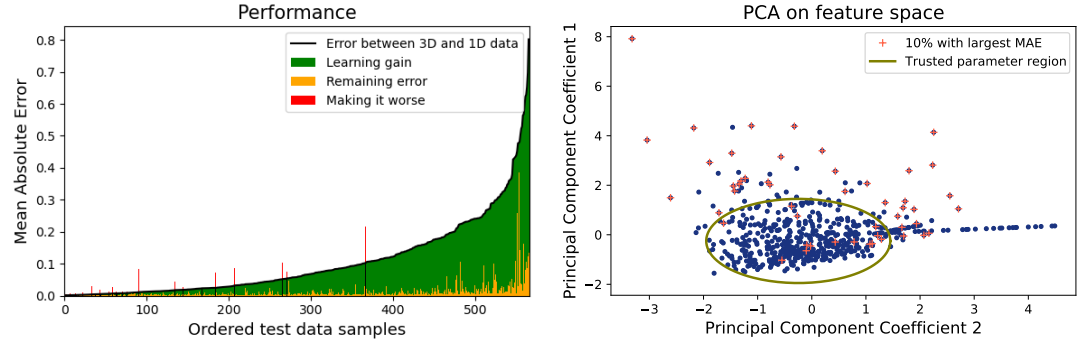**(c)** Test case for $d = 11.6$ m, $\theta = -2.33°$, EC1 = 0.91 Sm^{-1}, EC2 = 0.77 Sm^{-1}.

(d) Test case for $d = 37.5$ m, $\theta = 8.3°$, EC1 = 0.016 Sm^{-1}, EC2 = 0.72 Sm^{-1}.

Figure 5. The predicted and the true Relative Error (RE) for four samples from the test dataset (see Table 1).

(a) The learning gain using the surrogate model as correction to the *LF* model. The black line represents the ordered total MAE of the test samples. The green area shows the total reduced MAE between the 1D (*LF*) and 3D (*HF*) data and the orange area shows the residual MAE of the RE. The cases where the surrogate model makes the prediction of the *HF* data worse are shown in red.

(b) The first and second PC coefficients on the feature space. Samples in the test dataset with the 10% largest MAE are generally found further from the centre and indicate that the large MAE is a result of being at on the periphery of the trained parameter space. A warning will be printed if the parameters are outside of the trusted region, indicated in green. This is caused by the limited extrapolating capabilities of the surrogate model.

Figure 6. Performance in terms of learning gain and PCA on the feature space.

the White Noise Kernel could not handle the apparent noise properly). A value of 10^{-5} was used for the first four functional components, 10^{-3} for the next two and 10^{-2} for $l > 6$.

Based on our insights from the dimensionality reduction in Figure 3, we have optimized the surrogate model for ten different Ks. The MAE of the predicted data for varying K is shown in Figure 4A, where the maximum number of functional components $L = K$ is retained. Starting from $K = 8$, the prediction accuracy starts to stabilize with $K = 16$ as minimum. In Figure 4B, we show the MAE in terms of number of functional components L for $K = 16$. Starting from $L = 8$, not much is gained from adding more functional components. Note that this type of optimization leads to a different choice than relying on the (cumulative) explained variance ratio, which is often done for PCA. For the remainder of this text, we proceed with $K = 16$ and $L = 8$.

To illustrate the reduction in the computational burden, the *HF* data production requires one hour and 20 minutes of computation time on one 36-core High Performance Computing node (2 x 18-core Intel Xeon Gold 6140 (Skylake @ 2.3 GHz)) and requires 150 GB of RAM. Evaluating the correction model takes two milliseconds on a laptop (MacBook Pro, 10-core M1 Pro chip, 16 GB). The initialization of a pre-trained GPR requires 30 seconds for a training dataset with 5000 samples (this initialization time grows with the number of training samples).

In Figure 5, we show some samples illustrating some typical behaviour from the test dataset. Figure 5A and B are quite successful, the MAE is 0.00455 and 0.0028 respectively, which is in the order of best MAE given the dimensionality reduction (see Figure 4). The sample from Figure 5C is an example where the angle θ is small ($-2.33°$). The trend in the predicted relative error $\tilde{RE}$ is good, however the absolute values are too small. The surrogate model thus underestimates the true relative error. The sample in Figure 5D is a curious one, as it does not really resemble the typical shape that is observed in the other samples and in the sweeps in Figure 2. Yet, it cannot be ascribed to the dimensionality reduction: the MAE of the relative error of the back transformed reduced representation is only 0.00040. The reason can be ascribed to the underlying physics, where the dissipated currents move much more rapidly in more resistive media. This calls for a more thorough sampling of this part of the parameter space.

We can also view the performance per test sample separately in Figure 6A. The test samples are ordered from small discrepancies between *LF* and *HF* data to large discrepancies. The overall green part is what one gains by using the correction/surrogate model to correct the 1D forward *LF* model. The bottom orange parts indicate the residual error, which remains after applying the surrogate model correction. In some cases, the surrogate model corrects the *LF* data such that the predicted *HF* data is worse than when only the *LF* data was used (red bars), this is in 6% of the cases. Those cases correspond to low angles, small EC contrasts between layers or a low EC of the first layer and combinations thereof. No particular
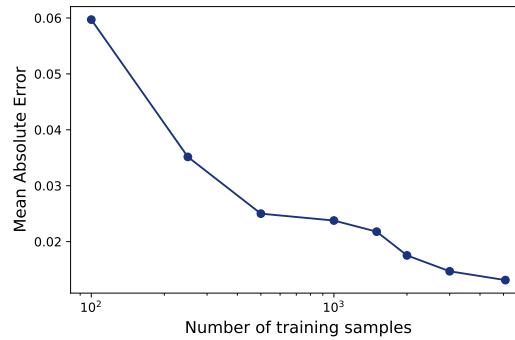
Figure 7. The learning curve suggests that more data could increase the overall accuracy.

cause could be identified. In Figure 6B we apply PCA to the four input parameters or features. By plotting the first two coefficients, we show the sampled parameter space in reduced dimensions. Samples far from the center (the origin) are extreme values in the sampled data set. The 10% predictions with the largest MAE are usually in the periphery, suggesting that the limiting extrapolating capabilities of the GPR in combination with the RBF kernel is the cause of the large MAE. Additional sampling in those parameter regions could improve the MAE. A zone of the parameter space was demarcated where the samples of the test data space performed well. We defined the sphere based on the first two PCA components of the feature space, with the median as the centre and a radius as a cut-off parameter. Points outside this sphere lie in a less densely sampled area, and a prediction in that region is less reliable. We have implemented a warning in the accompanying code that alerts when the correction model is used for parameters outside this trusted region. Furthermore, we uncovered that the surrogate correction model performs less satisfactorily at low total MAE, occurring for small angles and for low contrasts in electrical conductivity between the layers. In this case, the maximum relative error at any time is less than 1.5%, and the surrogate is less relevant. With small angles (less than 2°), slight contrasts (less than 0.025) or a combination of these (contrast^2 times the angle), the surrogate will warn the end-user that the relative error prediction is less accurate, which is not a problem as the 1D model is then sufficiently accurate. These warnings filter out 62% of the few test samples for which the MAE after prediction is larger than without using the surrogate model.

Figure 7 shows the learning curve of the training. It shows the MAE of the GPR accuracy for various training dataset sizes. The MAE was each time computed on the same test dataset of the above correction model. It is of course impossible to make predictions for the course of the curve for more training samples, but it is reasonable to assume that the accuracy can be further improved by providing more training data.

4 DISCUSSION

As expected, perfect accuracy is not achievable with these straightforward ML techniques, even in this simplified case with two layers. To our knowledge, we lack a measure to determine the overall performance of such a correction model. The effect of a modelling error, e.g., on an inversion model, is still relatively unknown. The approach to look at the learning gain, i.e. to look at what we gain with such correction model compared to only relying on a *LF* model, requires further validation. However, such a performance measure will be needed. Better accuracy could be achieved by adapting the prior or feature space to the specific application, which will require significant *HF* data for each project and may not be the most elegant and sustainable approach.

GPR provides relatively simple data fit models, meaning that it interpolates between observations. It is versatile, due to the option for different kernels. The RBF (default) kernel works in most smooth cases. It is difficult to obtain general models, however, and extrapolation is only possible ℓ (see Eq. (4)) units away from the trained prior (Duvenaud, 2014, p. 10). If one could identify or construct a more closely related underlying kernel, a more general GPR can be constructed with some extrapolating power. This would mean that sampling can be done with a lower data density in the space parameter. Note that GPR's are not sparse and thus use the whole dataset to perform one prediction. If one would have the resources to produce millions of training samples, alternatives to GPR, such as neural networks, should be considered.

In Section 2.2, we have explicitly opted to train a single-output GPR on each functional component, rather than using a multiple output GPR. The latter turns out not to be so trivial (Wang and Chen, 2015; Liu et al., 2018; Constantinescu and Anitescu, 2013). A truly multi-output GPR should not only describe the correlation between data points, but also the correlation between responses (Wang and Chen, 2015). Hence, it could leverage information from one output to provide a more accurate prediction for another output, rather than modeling them individually (Liu et al., 2018). However, multiple-output GPR is still a field of active research and existing codes are sparse. Constantinescu and Anitescu (2013) point out the difficulty of finding good covariance functions, which turn out to be much more challenging than for single-output GPs.

This correction model to the 1D approximation is relevant in contexts with significant modelling errors, such as with large electrical conductivity contrasts between layers and even for relatively small angles ($\theta = 10°$) (see Section 3.1). For example, we are thinking about imaging the freshwater lenses resting on top of salt water, which exhibit large conductivity contrasts and lateral variability. The two-layered forward model may already be sufficient. The correction model can be used with the traditional quasi- or pseudo-2D/3D inversion. The angle is an extra parameter in the inversion and can vary for each sounding. A smoothening constraint could be applied between neighbouring soundings, similar to the height parameter in AEM inversion.

Alternatively to TDEM applications, the proposed approach could be applied to frequency-domain electromagnetic data. For example, to image the environmental contamination of former gypsum landfills which exhibit large conductivity contrasts with the backfill material (Deleersnyder et al., 2023a) The applicability of the proposed methodology to other geohazard mitigation applications, which tend to become increasingly important in the rising circular economy (Koley, 2022) and climate change, should be further verified.

5 CONCLUSION

We have presented a novel approach to constructing a surrogate model for 3D forward TDEM simulations with only a few thousand (5090) training samples with relatively simple data-fit models. Rather than predicting the 3D data directly, we predict the relative error between the 3D and 1D data, as the computation of the 1D approximate forward model is fast and explains most of the variability in the geophysical data. By analysing the relative error, we have uncovered that the features of relative error can easily be explained (both height and time of the peak). The contrast in electrical conductivity between the two layers and the deviation from the 1D assumption (via its angle) explain the magnitude of the relative error. In contrast, the interface's depth and the first layer's conductivity determine when the peaks occur.

The surrogate model in the form of a correction term for the 1D forward model leads to a better prediction of the 3D data. Due to its limited extrapolating capabilities, a warning was implemented when the model is evaluated outside the sufficiently sampled parameter space. In summary, we have successfully tackled the challenge alluded to in the abstract.

Possible improvements could consider (1) active learning, to more knowingly sample the parameter space which maximizes the accuracy gain per extra training sample, especially because the 3D forward simulations are so expensive, (2) generating data on a 'Medium fidelity' mesh to increase the number of training samples (while the test dataset should still be generated on an accurate high-fidelity mesh), (3) examine alternative sparse dimensionality reduction techniques (fitting functions to the RE curves did not turn out to perform better, so we did not include these in the current work.). Eventually, as a community, we should define meaningful targets to evaluate the performance of different surrogate models.

Acknowledgments

Code availability section
The Gaussian Process Regression model is trained with the openly-available Python package scikit-learn. The script that controls the above software and produces the correction model are available from

Contact: wouter.deleersnyder@kuleuven.be

Program language: Python

The datasets are available on Zenodo via `https://doi.org/10.5281/zenodo.8366262.`

REFERENCES

Ansari, S., Farquharson, C., and MacLachlan, S. (2017). A gauged finite-element potential formulation for accurate inductive and galvanic modelling of 3-D electromagnetic problems. *Geophysical Journal International*, 210(1):105–129.

Asif, M. R., Foged, N., Maurya, P. K., Grombacher, D. J., Christiansen, A. V., Auken, E., and Larsen, J. J. (2022). Integrating neural networks in least-squares inversion of airborne time-domain electromagnetic data. *Geophysics*, 87(4):E177–E187.

Auken, E., Boesen, T., and Christiansen, A. V. (2017). A review of airborne electromagnetic methods with focus on geotechnical and hydrological applications from 2007 to 2017. *Advances in Geophysics*, 58:47–93.

Bai, P., Vignoli, G., Viezzoli, A., Nevalainen, J., and Vacca, G. (2020). (quasi-) real-time inversion of airborne time-domain electromagnetic data via artificial neural network. *Remote Sensing*, 12(20):3440.

Bording, T. S., Asif, M. R., Barfod, A. S., Larsen, J. J., Zhang, B., Grombacher, D. J., Christiansen, A. V., Engebretsen, K. W., Pedersen, J. B., Maurya, P. K., et al. (2021). Machine learning based fast forward modelling of ground-based time-domain electromagnetic data. *Journal of Applied Geophysics*, 187:104290.

Börner, R.-U., Ernst, O. G., and Güttel, S. (2015). Three-dimensional transient electromagnetic modelling using rational Krylov methods. *Geophysical Journal International*, 202(3):2025–2043.

Ching, J. and Yoshida, I. (2023). Data-drive site characterization for benchmark examples: Sparse bayesian learning versus gaussian process regression. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 9(1):04022064.

Cockett, R., Kang, S., Heagy, L. J., Pidlisecky, A., and Oldenburg, D. W. (2015). SimPEG: An open source framework for simulation and gradient based parameter estimation in geophysical applications. *Computers & Geosciences*, 85:142–154.

Constantinescu, E. M. and Anitescu, M. (2013). Physics-based covariance models for Gaussian processes with multiple outputs. *International Journal for Uncertainty Quantification*, 3(1).

Conway, D., Alexander, B., King, M., Heinson, G., and Kee, Y. (2019). Inverting magnetotelluric responses in a three-dimensional earth using fast forward approximations based on artificial neural networks. *Computers & geosciences*, 127:44–52.

Deleersnyder, W., Bah, A., Blanchy, G., Hermans, T., and Van De Vijver, E. (2023a). Estimating the volume of recoverable materials from a former gypsum landfill using frequency-domain electromagnetics. In *NSG2023 29th European Meeting of Environmental and Engineering Geophysics*, volume 2023, pages 1–5. European Association of Geoscientists & Engineers.

Deleersnyder, W., Dudal, D., and Hermans, T. (2022). Novel Airborne EM Image Appraisal Tool for Imperfect Forward Modeling. *Remote Sensing*, 14(22):5757.

Deleersnyder, W., Dudal, D., and Hermans, T. (2023b). Machine learning assisted fast forward 3D modelling for time-domain electromagnetic induction data–lessons from a simplified case. In *EGU General Assembly 2023, Location: Vienna, Austria*.

Deleersnyder, W., Maveau, B., Hermans, T., and Dudal, D. (2023c). Flexible quasi-2D inversion of time-domain AEM data, using a wavelet-based complexity measure. *Geophysical Journal International*, 233(3):1847–1862.

Dikshit, A., Pradhan, B., and Alamri, A. M. (2021). Pathways and challenges of the application of artificial intelligence to geohazards modelling. *Gondwana Research*, 100:290–301.

Duvenaud, D. (2014). *Automatic model construction with Gaussian processes*. PhD thesis, University of Cambridge.

Engebretsen, K. W., Zhang, B., Fiandaca, G., Madsen, L. M., Auken, E., and Christiansen, A. V. (2022). Accelerated 2.5-D inversion of airborne transient electromagnetic data using reduced 3-D meshing. *Geophysical Journal International*, 230(1):643–653.

Feng, B., Zhang, J.-f., Li, D., and Bai, Y. (2020). Resistivity-depth imaging with the airborne tran-

sient electromagnetic method based on an artificial neural network. *Journal of Environmental and Engineering Geophysics*, 25(3):355–368.

Ferkous, K., Chellali, F., Kouzou, A., and Bekkar, B. (2021). Wavelet-gaussian process regression model for forecasting daily solar radiation in the saharan climate. *Clean Energy*, 5(2):316–328.

Gao, Z., Ding, M., Huang, T., Liu, X., Hao, Z., Hu, X., and Chuanjie, X. (2022). Landslide risk assessment of high-mountain settlements using gaussian process classification combined with improved weight-based generalized objective function. *International Journal of Disaster Risk Reduction*, 67:102662.

Gonçalves, Í. G., Guadagnin, F., and Cordova, D. P. (2022). Learning spatial patterns with variational gaussian processes: Regression. *Computers & Geosciences*, 161:105056.

Grana, D., Passos de Figueiredo, L., and Azevedo, L. (2019). Uncertainty quantification in bayesian inverse problems with model and data dimension reduction. *Geophysics*, 84(6):M15–M24.

Haber, E., Ascher, U. M., and Oldenburg, D. W. (2004). Inversion of 3D electromagnetic data in frequency and time domain using an inexact all-at-once approach. *Geophysics*, 69(5):1216–1228.

Haber, E., Oldenburg, D. W., and Shekhtman, R. (2007). Inversion of time domain three-dimensional electromagnetic data. *Geophysical Journal International*, 171(2):550–564.

Hall, P., Müller, H.-G., and Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *The annals of statistics*, 34(3):1493–1517.

Heagy, L. J., Cockett, R., Kang, S., Rosenkjaer, G. K., and Oldenburg, D. W. (2017). A framework for simulation and inversion in electromagnetics. *Computers & Geosciences*, 107:1–19.

Hou, S., Angio, S., Clowes, A., Mikhalev, I., Hoeber, H., and Hagedorn, S. (2019). Learn to invert: Surface wave inversion with deep neural network. In *81st EAGE Conference and Exhibition 2019 Workshop Programme*, pages 1–5. European Association of Geoscientists & Engineers.

Hunziker, J., Thorbecke, J., and Slob, E. (2015). The electromagnetic response in a layered vertical transverse isotropic medium: A new look at an old problem. *Geophysics*, 80(1):F1–F18.

Klose, T., Guillemoteau, J., Vignoli, G., and Tronicke, J. (2022). Laterally constrained inversion (LCI) of multi-configuration EMI data with tunable sharpness. *Journal of Applied Geophysics*, 196:104519.

Koley, S. (2022). Sustainability appraisal of arsenic mitigation policy innovations in west bengal, india. *Infrastructure Asset Management*, 10(1):17–37.

Laloy, E., Linde, N., and Jacques, D. (2021). Approaching geoscientific inverse problems with vector-to-image domain transfer networks. *Advances in Water Resources*, 152:103917.

Li, R., Zhang, H., Zhuang, Q., Li, R., and Chen, Y. (2020). BP neural network and improved differential evolution for transient electromagnetic inversion. *Computers & Geosciences*, 137:104434.

Liu, H., Cai, J., and Ong, Y.-S. (2018). Remarks on multi-output Gaussian process regression. *Knowledge-Based Systems*, 144:102–121.

Lopez-Alvis, J., Laloy, E., Nguyen, F., and Hermans, T. (2021). Deep generative models in inversion: The impact of the generator's nonlinearity and development of a new approach based on a variational autoencoder. *Computers & Geosciences*, 152:104762.

Macnae, J. and Milkereit, B. (2007). Developments in broadband airborne electromagnetics in the past decade. In *Proceedings of Exploration*, volume 7, pages 387–398.

Michel, H., Nguyen, F., Kremer, T., Elen, A., and Hermans, T. (2020). 1d geological imaging of the subsurface from geophysical data with bayesian evidential learning. *Computers & Geosciences*, 138:104456.

Mikucki, J. A., Auken, E., Tulaczyk, S., Virginia, R., Schamper, C., Sørensen, K., Doran, P., Dugan, H., and Foley, N. (2015). Deep groundwater and potential subsurface habitats beneath an Antarctic dry valley. *Nature communications*, 6(1):1–9.

Noh, K., Yoon, D., and Byun, J. (2020). Imaging subsurface resistivity structure from airborne electromagnetic induction data using deep neural network. *Exploration Geophysics*, 51(2):214–220.

Peng, Z., Yang, B., Liu, L., and Xu, Y. (2023). Rapid surrogate modeling of magnetotelluric in the frequency domain using physics-driven deep neural networks. *Computers & Geosciences*, page 105360.

Pfaffhuber, A. A., Lysdahl, A. O., Sørmo, E., Skurdal, G. H., Thomassen, T., Anschütz, H., and Scheibz, J. (2017). Delineating hazardous material without touching - AEM mapping of Norwegian alum shale. *First Break*, 35(8).

Puzyrev, V. and Swidinsky, A. (2021). Inversion of 1D frequency-and time-domain electromagnetic data with convolutional neural networks. *Computers & Geosciences*, 149:104681.

Ramsay, J. O. S., Kotz, C., Read, N., Balakrishnan, B., Vidakovic, and Johnson, N. (2006). *Functional

Data Analysis, volume 1. John Wiley & Sons, Ltd.

Satija, A. and Caers, J. (2015). Direct forecasting of subsurface flow response from non-linear dynamic data by linear least-squares in canonical functional principal component space. *Advances in Water Resources*, 77:69–81.

Scheidt, C., Li, L., and Caers, J. (2018). *Quantifying uncertainty in subsurface systems*, volume 236. John Wiley & Sons.

Schenk, O. and Gärtner, K. (2004). Solving unsymmetric sparse systems of linear equations with PARDISO. *Future Generation Computer Systems*, 20(3):475–487.

Siemon, B., van Baaren, E., Dabekaussen, W., Delsman, J., Dubelaar, W., Karaoulis, M., and Steuer, A. (2019). Automatic identification of fresh – saline groundwater interfaces from airborne electromagnetic data in Zeeland, the Netherlands. *Near Surface Geophysics*, 17(1):3–25.

Silverman, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *The Annals of Statistics*, 24(1):1–24.

Suhaila, J. and Hamdan, M. F. (2023). Research trends on functional data analysis using scopus database: A bibliometric analysis. *Malaysian Journal of Fundamental and Applied Sciences*, 19(4):494–512.

Wang, B. and Chen, T. (2015). Gaussian process regression with multiple response variables. *Chemometrics and Intelligent Laboratory Systems*, 142:159–165.

Werthmüller, D., Mulder, W. A., and Slob, E. C. (2019). emg3d: A multigrid solver for 3D electromagnetic diffusion. *Journal of Open Source Software*, 4(39):1463.

Werthmüller, D. (2017). An open-source full 3D electromagnetic modeler for 1D VTI media in Python: empymod. *GEOPHYSICS*, 82(6):WB9–WB19.

Williams, C. K. and Rasmussen, C. E. (2006). *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA.

Wu, S., Huang, Q., and Zhao, L. (2023). A deep learning-based network for the simulation of airborne electromagnetic responses. *Geophysical Journal International*, 233(1):253–263.

Wu, Y., Lin, Y., and Zhou, Z. (2018). InversionNet: Accurate and efficient seismic waveform inversion with convolutional neural networks. In *2018 SEG International Exposition and Annual Meeting*. OnePetro.

Zhang, L., Zhang, G., Liu, Y., and Fan, Z. (2021). Deep learning for 3-D inversion of gravity data. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–18.