



1-D inversion of frequency-domain airborne electromagnetic data using the open-source aempy toolbox

Duygu Kiyani^{a,b,*}, Volker Rath^{a,b,*}, Mark R. Muller^c, Mohammednur D. Ture^c, James Hodgson^c

^a Geophysics Section, Dublin Institute for Advanced Studies, 5 Merrion Square, Dublin 2, Ireland

^b iCRAG centre, University College Dublin, Dublin 4, Ireland

^c Geological Survey Ireland, Block 1, Booterstown Hall, Booterstown, Blackrock, Co Dublin, Ireland



ARTICLE INFO

Keywords:

Applied geophysics
Airborne electromagnetics
Inversion
Stochastic
Deterministic
Ireland

ABSTRACT

Over the past decade, Geological Survey Ireland has acquired, and continues to acquire, frequency and time-domain airborne electromagnetic data as part of the Tellus project. While the project advances towards national coverage, we have developed an open-source and freely available pre-inversion processing and one-dimensional inverse modelling toolbox, aempy, for numerical modelling of the existing and future electromagnetic data sets. aempy is a flexible tool facilitating the quantitative interpretation and optimal use of the data sets by the broader scientific community, supporting the ethos and intent in the free public release of the Tellus airborne data sets. Coded in Python, the toolbox is freely and readily adaptable for specific educational and research objectives and for application to data acquired using other electromagnetic systems. The inversion algorithms in the toolbox are currently implemented in 1-D, assuming that the Earth is horizontally stratified and has laterally-constant layer model parameters, e.g., conductivity and thickness. In its current state, modules and functions of aempy work on processed frequency-domain airborne electromagnetic data. In this paper, we introduce the concept, structure and capacities of aempy and present model results derived from the use of different filtering and inversion techniques implemented in the toolbox, as well as an analysis of the influence of the flight altitude parameter on resulting subsurface inversion models.

1. Introduction

The Tellus airborne geophysical programme in Ireland started with the first airborne geophysical survey of Northern Ireland, which was flown in 2005 and 2006. It soon was complemented with the Tellus Border project covering the six northern most counties of the Republic of Ireland in 2011 to 2012. In the following years, eight more surveys were completed, so that, to date, more than 70% of the island of Ireland has been covered. The airborne geophysical surveys aim to provide spatially continuous information about the subsurface electrical conductivity, magnetic properties, and natural gamma-radiation variability. Data from these surveys have already proven valuable for geological mapping, environmental monitoring, and mineral exploration (Young, 2016). This study concentrates on the airborne electromagnetic (AEM) measurements.

Most of the Tellus surveys include acquisition of frequency-domain EM (FDEM) data carried out by Sander Geophysics Ltd. using the GTK airborne system (Finnish Geological Survey) AEM-95 and the Joint

Airborne Geoscience Capability, which is a partnership between the Finnish and British Geological Surveys, airborne system (AEM-05). While the first AEM survey in Northern Ireland was flown with the GTK AEM-95 system with only two frequencies (3125 Hz and 14,368 Hz), the system was soon redesigned with more frequencies, leading to the AEM-05 system, which operates at four frequencies, (912 Hz, 3005 Hz, 11,962 Hz, and 24,510 Hz). The transmitter (TX) and receiver (RX) are vertical, co-planar coils (VCP or CpX configuration, (Tolboll and Christensen, 2007; Hodges et al., 2016)) mounted at the tips of the wings with a fixed separation of ≈ 21.4 m, Fig. 1. A detailed description of the system can be found in (Levaniemi et al., 2009). The survey design parameters for the FDEM data acquisition in Ireland are: a nominal flight altitude of 60 m (56 m for the Tellus Border Survey) above ground level and flight lines with a fixed orientation of 345° (NNW-SSE, perpendicular to the dominant and most common lithological and structural trends in Ireland) and a 200 m lateral spacing. The spatial sampling interval depends on the speed of the plane and is generally around 6 m for a nominal flight speed of 60 m/s [e.g., (SGL, 2016)]. The VCP

* Corresponding authors at: Dublin Institute for Advanced Studies, 5 Merrion Square, Dublin 2, Ireland.

E-mail addresses: duygu@cp.dias.ie (D. Kiyani), vrrath@cp.dias.ie (V. Rath).



Fig. 1. AEM-05 four-frequency wing-tip Tellus survey aircraft. (Photo by courtesy of Geological Survey Ireland).

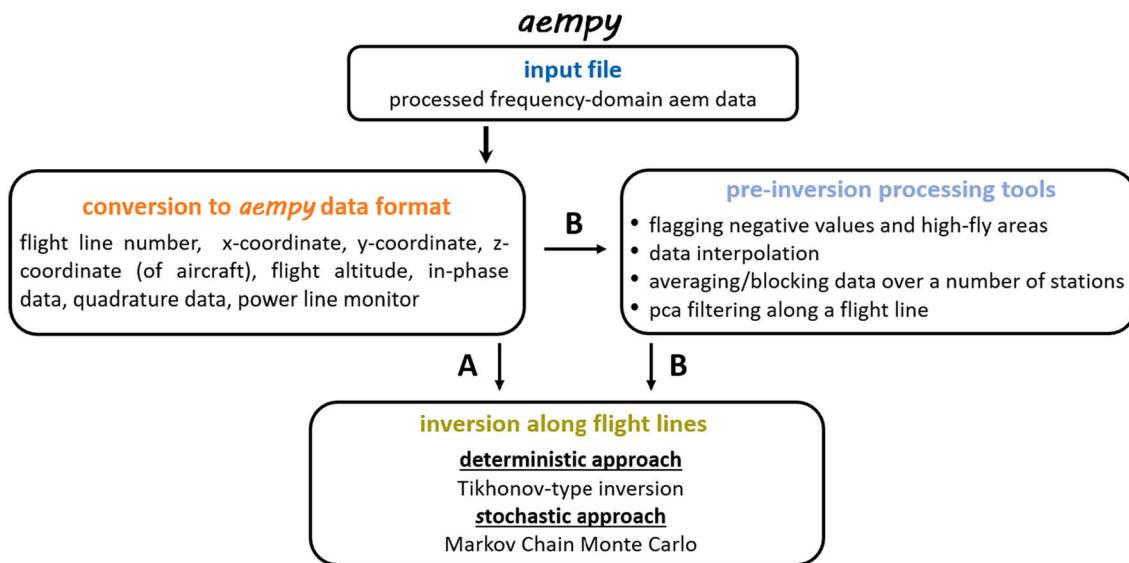


Fig. 2. Structural flowchart showing aempy features.

configuration sensitivity footprint has two times greater extension orthogonal to the flight line than in the flight line direction, which would provide a certain overlap of sensitivities between the flight lines at the lowest frequency (912 Hz) at moderate resistivities, but not at the highest ones (Beamish, 2003). Along the flight lines, under most conditions the footprint will cover a larger number of measuring locations, leading to an oversampling in this direction even at the high frequencies (24,510 Hz, 11,962 Hz), opening possibilities to take advantage of this data redundancy, for example, through decimation or averaging along flight lines before inversion.

The quantitative interpretation of the island-wide Tellus AEM data requires approaches that are able to deal with a number of challenges that do not arise, or are less prevalent, in the case of smaller, focused surveys with designs adapted to particular targets (e.g., groundwater or mineral deposits). These challenges include the very large volume of data and number of sites to be modelled and the very strong spatial variability in the signal-to-noise characteristics of the data due to variability in cultural noise and in flight altitude, variability in the strength of the geological signal itself, as well as variability resulting from temporal changes in weather (e.g., wind speed, direction and turbulence)

and ground (e.g., moisture) conditions that arise during long-duration (multi-month and multi-year) surveys.

For instance, the recently completed Block A1 in the eastern part of Ireland includes 5.5 million measurement sites with about 44 million data points (Hodgson and Ture, 2016). Any modelling or inversion of such a number of data is a considerable computational task. The Tellus VCP loop configuration and transmitted frequency range already pose limits on the modelling of the observations beyond the use of 2-D inversion, as the sensitivity footprint of the system will only overlap between neighbouring flight lines at the lowest frequency. A further challenge is related to dense population and intensive agriculture in Ireland. The presence of towns, agricultural activity and industrial infrastructure may result not only in noisy data sets, but also require data acquisition at significantly higher than nominal flight altitudes (> 200 m). A large component of Irish agriculture consists of livestock breeding, posing limitations on the allowable minimum flight altitude and requiring intensive public communications and coordination programmes prior to flying. The result of these cultural challenges is the presence of high-fly zones of varying altitude and sizes all over the areas of airborne coverage. In these high-fly zones the resolution of the AEM

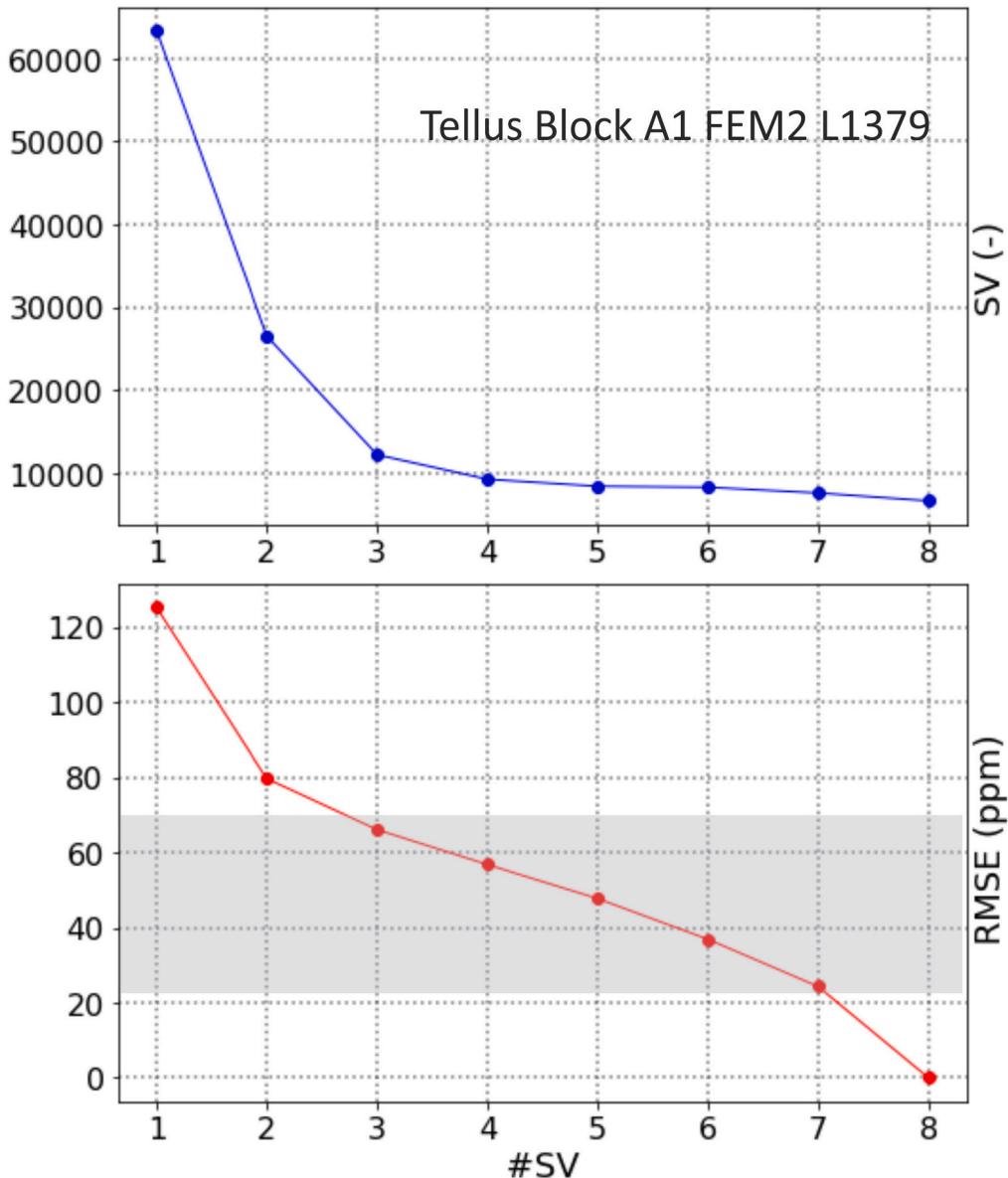


Fig. 3. SVD decomposition of data from flight line L1379 from Tellus Block A1 FEM2. Top: Singular values. Bottom: RMSE for the corresponding approximations (see Eq. (B.3) in Appendix B). The grey shaded zone marks the limits of errors estimated in this study (see Table 1).

measurements is substantially reduced or even negligible. However, dedicated Tellus high-fly test lines present an opportunity to estimate and assess error levels in the measured data. Any inversion depends on the correct estimation of the random and systematic errors, which will enter the inversion process and determine how much constraint is to be imposed on the models.

The measurement systems, modelling and inversion of AEM data have progressed considerably in the last few decades, as pointed out in recent reviews [e.g., (Legault, 2015; Auken et al., 2017)]. Most of the methodological developments are proprietary, though a few open-access implementations exist, such as the SimPEG (Heagy et al., 2017) framework, which provides a collection of geophysical simulation and inversion tools and the stand-alone code GA-AEM (Brodie, 2016) for modelling and inversion of AEM data. Most of the recent improvements took place for the different variants of time-domain electromagnetic (TDEM) systems, which now dominate this research area due to their higher flexibility, potentially better resolution, deeper penetration and, last but not least, availability of the systems. Much less development is

visible in the area of FDEM systems since the codes developed by the AMIRA consortium and CSIRO were made publicly available in 2007 (Raiche et al., 2007). These inversion codes are based on the damped Singular Value Decomposition methods presented in (Jupp and Vozoff, 1975; Vozoff and Jupp, 1975). While these well-tested, flexible, and reliable codes are an excellent starting point for the development of interpretation tools for the Tellus data, we felt that the underlying forward modelling codes should be combined with more developed inverse techniques, and should be complemented by data analysis procedures adapted to the particular conditions prevalent in Ireland, as mentioned above. Thus the building blocks for the inversions and simulations used in this study have been implemented in a Python toolbox, aempy, which makes the incorporation of the new inverse methods comparatively easy, and allows a fast development of workflows for particular needs.

In the following we will first introduce the aempy toolbox in Section 2 followed by the demonstration of the use of aempy for filtering and 1-D inversion of Tellus airborne EM data in Section 3. We will then demonstrate our assessment of observational error, and the numerical

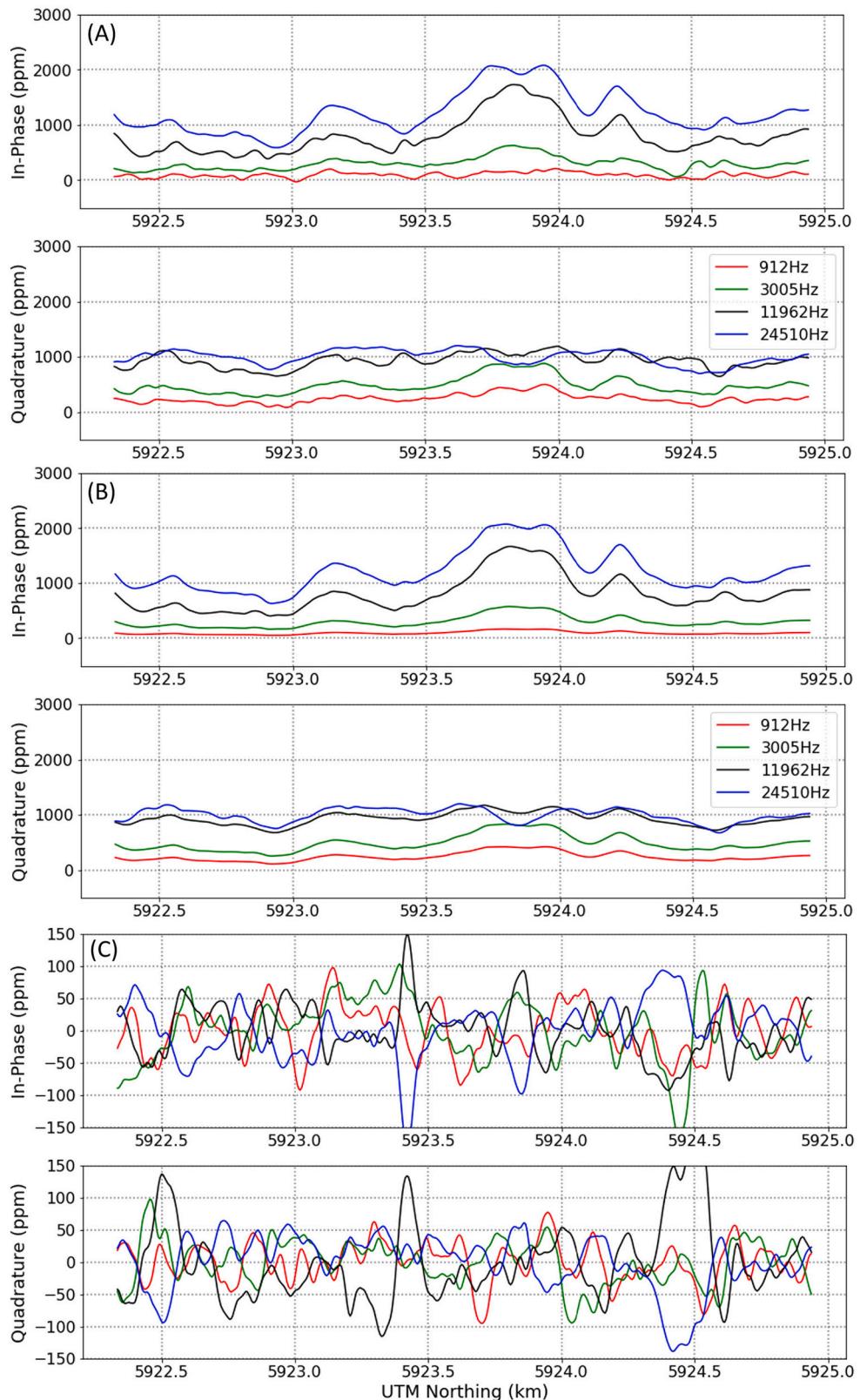


Fig. 4. Example subset of data from line L1379 from Tellus Block A1 illustrating the results of the SVD-based reconstruction. (A) Original data at four Tellus frequencies (912 Hz, 3005 Hz, 11,962 Hz, 24,510 Hz). (B) Data reconstructed using only the first two singular values and vectors. (C) Residuals between original data and data reconstructed by SVD truncation.

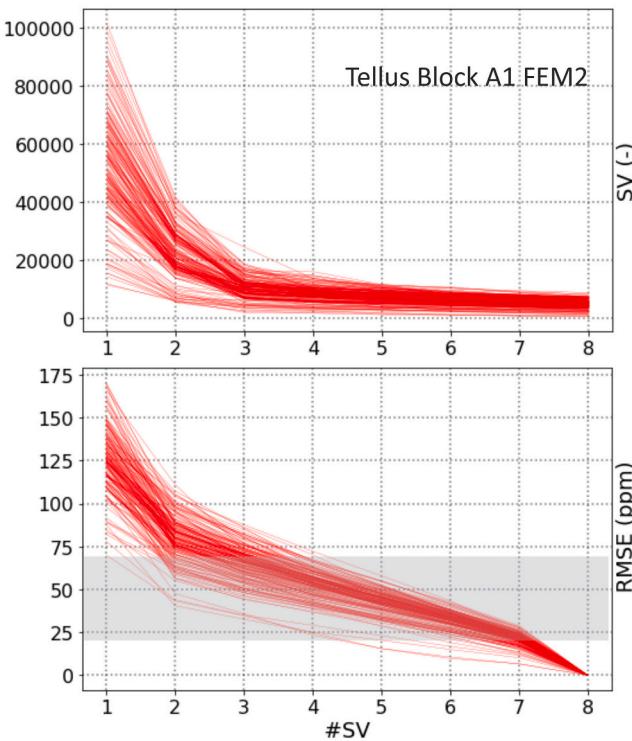


Fig. 5. SVD decomposition of data from Tellus Block A1 FEM2 (lines L1266 — L1447). Top: Singular values. Bottom: RMSE (see Eq. (B.3)) for the corresponding approximations. The grey shaded zone marks the limits of errors estimated in this study (see Table 1).

studies on the effects of high-fly areas on the inverted subsurface models in Section 4. The theory for airborne electromagnetic methods in the 1-D case and for the implemented inversion techniques is described in Appendix A, and the theory for principal component analysis approach implemented in aempy is described in Appendix B.

2. The aempy toolbox

Written in Python, aempy is open-source and contains four main subdirectories: core, doc, modules and scripts. The core subdirectory provides adapted numerical forward subroutines derived from the well-tested, open-source code Airbeo, written in Fortran 90. Instead of re-programming the forward modelling subroutines in Python, we have made use of the excellent capabilities of Python for semi-automatic interfacing with other programming languages, in this case with Fortran 90. The modules subdirectory holds the necessary toolbox libraries (currently only core1d.so), and two python modules, namely, aemprocs.py and invprocs.py. While the former one contains functions for handling data management, extraction and reformatting processed Tellus data to aempy data format, the latter one provides functions for data filtering and visualisation and contains functions necessary for data inversion. The scripts subdirectory provides scripts that contain 1-D inverse algorithms and additional scripts for graphical representation of inversion results. All plots presented in this paper have been generated using these additional scripts. The manual for the toolbox (Kiyani and Rath, 2017) is found in doc subdirectory. Fig. 2 presents the workflow implemented in the toolbox, which starts with the conversion to aempy data format. Once the data subset is defined, quality control can be achieved by visualisation using pre-inversion processing tools, for example, negative data values and high-fly areas can be flagged. The data in these flagged areas can be deleted and imputed by average or interpolated values. To improve the spatial consistency of the inversion

results, several filtering approaches including (moving) average, median window and principal component analysis are implemented in the toolbox. The inversion techniques developed within the toolbox are (i) Tikhonov-type inversion including optimal regularisation methods, and (ii) fully Bayesian inversion by Markov Chain Monte Carlo techniques. Bayesian inversion currently has only been applied to synthetic examples, aiming at better understanding of the uncertainty of the computed solutions. The details of the inversion techniques are described in Appendix A.

3. Practical AEM data inversion using aempy

As mentioned in the previous section, a procedure of particular interest for the analysis of AEM data is the Singular Value Decomposition (SVD) of the observation matrix for a particular flight line prior to inversion. The successful applications of SVD for coherency enhancement on airborne TDEM and FDEM data sets has previously been reported by (Reninger et al., 2011; Minsley et al., 2012). The principal component analysis (PCA) approach to filtering frequency-domain AEM data, which is based on the SVD of the data matrix (Hansen et al., 2006), was first proposed to reduce random noise in the data (Minsley et al., 2012) and we have implemented this filtering strategy (Appendix B) in aempy. An important strength of this approach is that it is based on the fact that the data at different frequencies are strongly correlated.

Figs. 3 and 4 examine an example of data from the Tellus Block A1 FEM2 survey and suggest that the first two singular vectors and values represent the spatially coherent part of the observations. Note that since the Tellus surveys operate at four frequencies for in-phase and quadrature components, the data signal has eight singular values and corresponding vectors. The effect of applying SVD-based filter to the EM data can be assessed by examining the differences between the input and filtered data. In the lower panel of Fig. 3, a comparison is made using the root-mean-square-error between the observed and filtered data, as a function of the number of singular values retained in the SVD-based data reconstruction. In this lower panel, for example, the data plotted at #SV = 2 correspond with the RMS error between the observed data and the data reconstructed using singular values 1 and 2 only. Retaining all eight singular values (#SV = 8) reconstructs the observed data exactly, returning an RMS error of zero. It is apparent in Fig. 3 (lower panel) that the inclusion of the fourth and higher singular components bear the risk of contaminating data signal if we assume that the assigned error estimates (between 20 and 70 ppm) are reasonable. Fig. 4 compares the observed data and reconstructed data using only the two singular values and vectors. The most obvious impact observed in the data reconstructed by SVD truncation is a general smoothing of the data for eight data components. The amplitude of residuals between the observed data and the filtered data are mostly within the assumed error limits (± 70 ppm).

In order to prove that this observation is valid, we applied SVD analysis to 185 flight lines of data from the Tellus A1 Block FEM2 survey. The analysis results in Fig. 5 demonstrate that the observed data can be represented with only the first two or three singular components. The SVD decomposition procedure can be applied to larger survey areas available in the Tellus database as a routine analysis tool as well as a pre-inversion filtering tool if appropriate.

Figs. 6 and 7 show the inversion results from line L1379 together with the apparent resistivity maps of the survey area and the flight altitude data. The location of line L1379 marked with a solid black line on the apparent resistivity maps, which are created using levelled apparent resistivity, half-space model, data provided by Sander Geophysics. A look-up procedure (Fraser, 2022) employing the in-phase and quadrature data components at each frequency was used to calculate resistivity and an associated apparent height of the sensor over an assumed conductive homogeneous half-space (SGL, 2016). The multi-frequency AEM data were inverted using 1-D smooth, multi-layer Tikhonov-type inversion as described in Appendix A. The 1-D

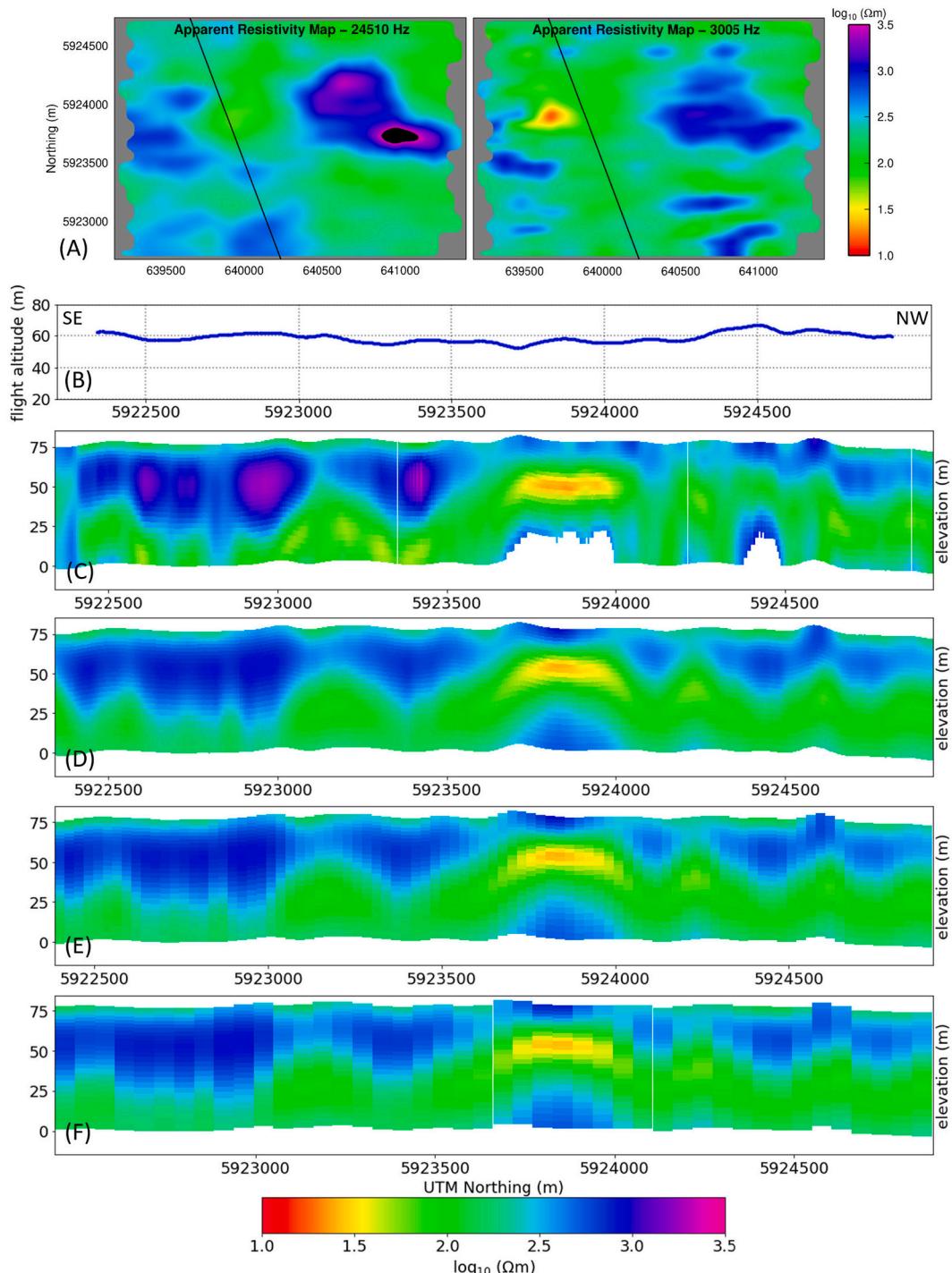


Fig. 6. Inversion results of a data subset from line L1379. (A) Location of line L1379 with the apparent resistivity maps at two frequencies 24,510 Hz and 3005 Hz. Apparent resistivity data, which were provided by Sander Geophysics, were smoothed using the nearest neighbour algorithm implemented in the Generic Mapping Tool (GMT) package. Black regions on the map of 24,510 Hz data represent data values greater than 5000 Ωm . (B) Flight altitude data. (C–F) 1-D smooth, multi-layer Tikhonov type inversion results. From top to bottom the inversions use the original data (station spacing is roughly 6 m), and PCA-filtered data using the first two SVs, with different block sizes of 1 (6 m), 5 (30 m, the average of 5 data points), and 10 (60 m, the average of 10 data points), respectively. Blocking, i.e., decimation is achieved by averaging over a number of stations. Regularisation parameters, $\tau_0 = 0.001$ and $\tau_1 = 5.3$ were chosen by trial-and-error.

inversions were carried out using both in-phase and quadrature components from four frequencies (24,510 Hz, 11,962 Hz, 3005 Hz, and 912 Hz). The model consists of 30 layers, with the thickness of the top layer starting at 2 m and subsequent layers increasing logarithmically in thickness with depth till 98 m depth. The layer thicknesses were preserved during the inversion process. A prior resistivity of 100 Ωm was chosen and each site was modelled using the previous site's inversion

model as the initial model for the next site. Fig. 6C and D compare inversion model results for L1379 derived using the unfiltered EM data and the PCA-filtered data. Except for the application of the PCA-filtered using the first two SVs, the two inversions are identical in their parameters. In comparing the two models, the PCA-filtered result shows generally improved model continuity. There is no evidence of loss of geological details with respect to the unfiltered result, indicating that the

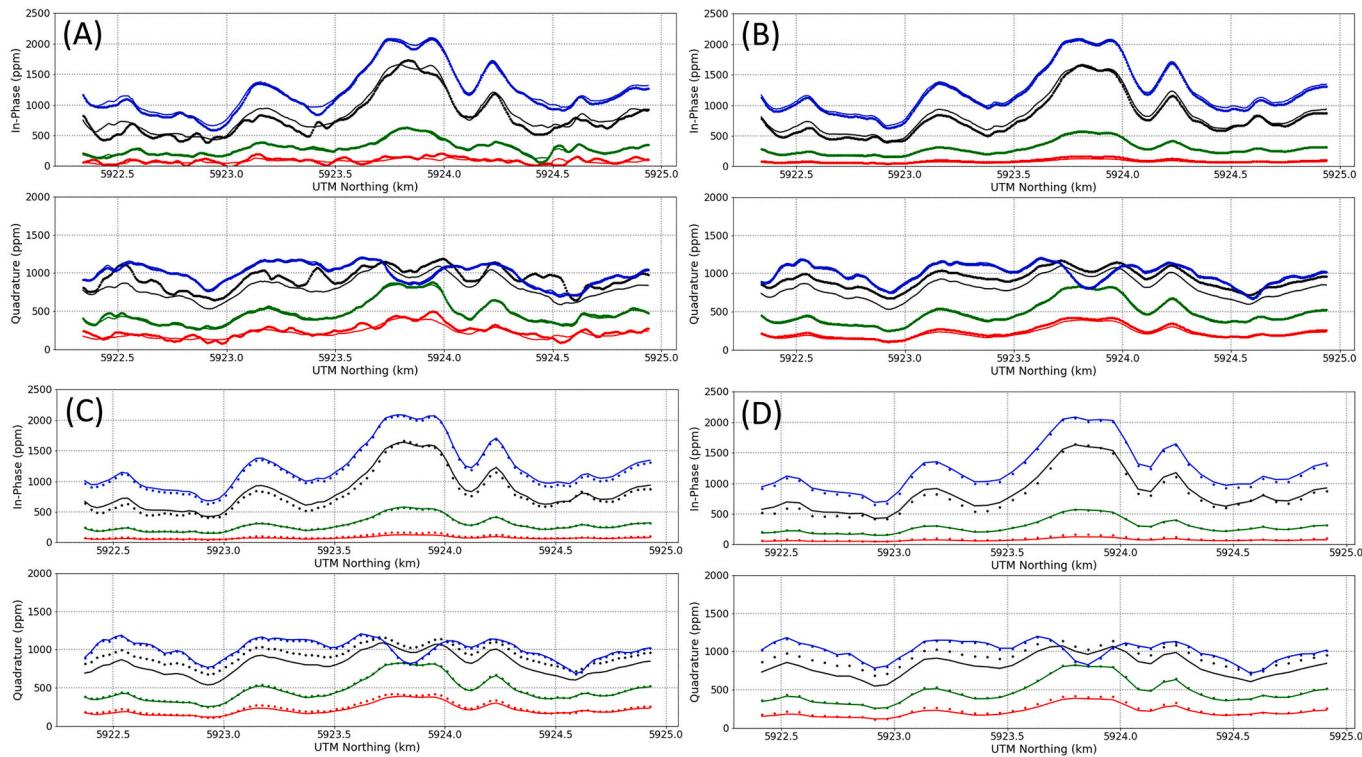


Fig. 7. Observed (solid circles) and predicted data (thin solid lines) at the four survey frequencies 24,510 Hz (blue), 11,962 Hz (black), 3005 Hz (green) and 912 Hz (red) for the inverted models presented in panels (C), (D), (E), and (F) in Fig. 6. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

PCA-filter has not removed coherent signal from the EM data. Fig. 6D and E show inversion results for PCA-filtered data with blocks of 5 (the average of 5 data points) and 10 (the average of 10 data points) stations, respectively. Blocking is achieved by summarising the data by their per-frequency arithmetic mean over a number of stations, which takes advantage of the implicit redundancy in flight line sampling (Annett and Hauser, 2016). Alternative blocking methods available in the aempy are a moving average and median window, which are not used for L1379 data inversion. It can be clearly seen that the main features in Fig. 6C are still preserved in the resulting subsurface models in Fig. 6D and E. Practically, there is some advantage in inversion time in running, as data from less number of stations are used.

4. Flight altitude and model uncertainty

4.1. The Tellus high-fly experiment

As part of the Tellus airborne geophysical surveys, a 6-km-long Test Line was established close to the town of Bundoran, Donegal. The test line was flown at different times and multiple altitudes to help assess both the repeatability of the data and the effect of altitude on the data quality. The first data from the test line were acquired in October 2011 with further data collected during 2014, 2015 and 2016 over multiple flights. Standard processing of the Bundoran Test Line FDEM data was undertaken by Sander Geophysics (SGL, 2016).

The Bundoran test line data analysed here were acquired along the same flight line flown at seven different flight altitudes (60 m, 80 m, 90

m, 120 m, 150 m, 180 m, and 240 m) in 2015.¹ Fig. 8 illustrates the in-phase (top, A) and quadrature (bottom, B) data for the altitudes of 60 m, 90 m, 150 m, and 240 m. As one can clearly see, the lower the altitudes, the higher the signal strength received. The 150 m and 240 m data sets show little to no separation and have negative (i.e., non-physical) minimum values between -100 and -300 ppm, meaning that little (geologically) informative data may be extracted from these altitudes in practice, particularly at the lowest frequency of 912 Hz. It should be noted that in-phase at 912 Hz is generally very low, even at low altitudes.

As the contribution of the data acquired at the highest flight altitude (240 m) to the subsurface model is negligible, we have estimated data errors by calculating the variance and corresponding standard deviation of the 240 m altitude data. The estimated data errors, summarised in Table 1, are between 20 ppm and 70 ppm, with no dependence on frequency or data component (in-phase and quadrature). Furthermore, the data plots presented in Fig. 8 show a spatial autocorrelation. A rough estimation of the correlation length is about 200 m, corresponding to a correlation time of a few seconds. Note that the variances and corresponding errors calculated from the high-fly experiment should be considered as rough estimates of the Tellus data uncertainty as they are derived from only one 240 m flight height line acquired in 2015.

We have made use of the test line data set to exploit the SVD methodology with an ultimate aim of understanding the effect of the survey altitude on the observed data. Fig. 9 provides a summary of the results obtained from the SVD. The top panel clearly illustrates that the data collected at 60 m survey altitude have the largest singular values

¹ The Bundoran test line data used are as follows, taken from Sander Geophysics Ltd.'s Delivery DLV2062 to Geological Survey Ireland: line L2060120.15 (60 m), L2080120.15 (80 m), L2090120.15 (90 m), L2120120.15 (120 m), L2150120.15 (150 m), L2180120.15 (180 m), and L2240120.15 (240 m)

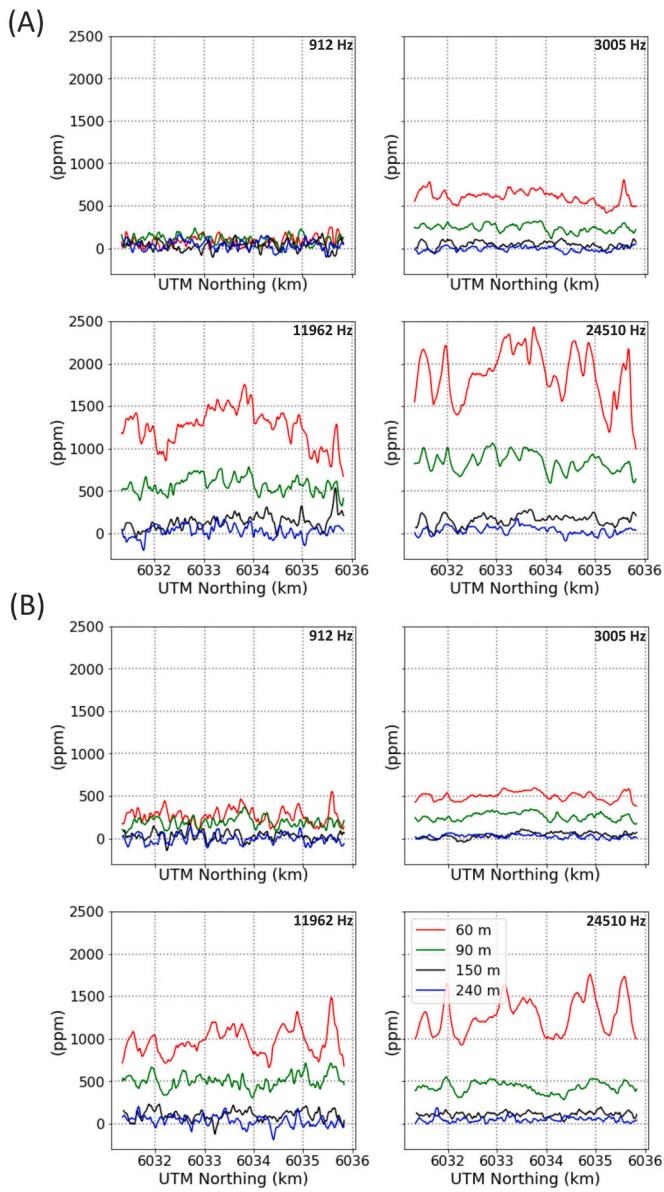


Fig. 8. Bundoran Test Line. Examples of observed In-Phase (A) and Quadrature (B) data acquired at the four Tellus frequencies (912 Hz, 3005 Hz, 11,962 Hz, 24,510 Hz) at survey altitudes of 60 m, 90 m, 150 m, and 240 m above ground level.

and vectors and the first two singular vectors and values can reconstruct the coherent portion of the data, further supported by the residual plots presented in Fig. 10. The values of the residuals in Fig. 10 fall within the

data errors (Table 1) estimated for the Tellus surveys when more than the first two singular vectors and values are used in the data reconstruction. As anticipated, the data acquired at 240 m survey altitude have the smallest singular values and vectors and one possibly cannot extract any geologically meaningful components of the data from the noise. The bottom panel of Fig. 9 presents the RMSE, indicating the deviations between the observed and the decomposed data, reconstructed using an increasing number of singular components.

Once the approximate data were generated by SVD reconstruction using the first two singular components, they were inverted for a 1-D subsurface resistivity model using aempy. The initial model consists of

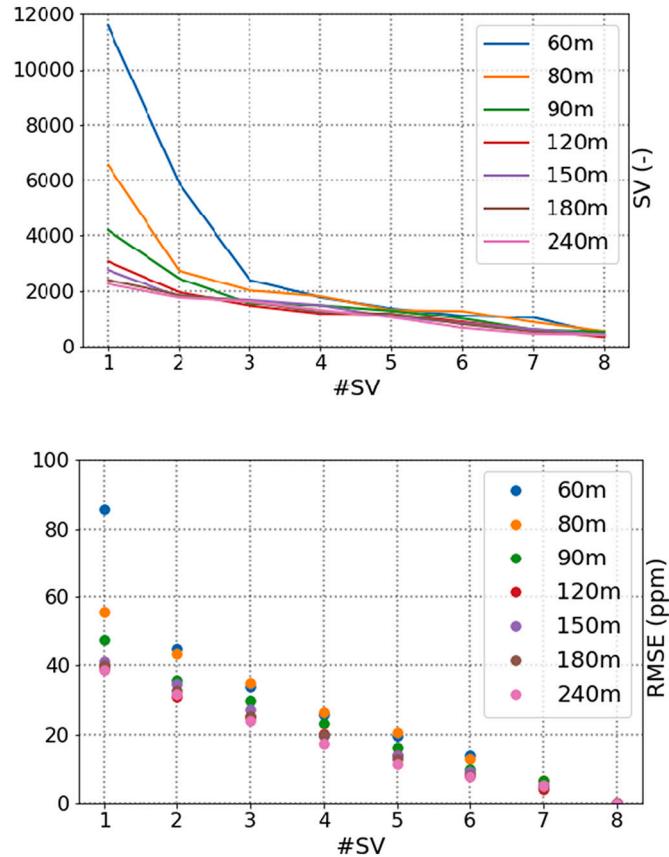


Fig. 9. Summary of the SVD analysis results for the Bundoran Test Line data for seven flight altitudes. Upper panel shows singular value amplitude (SV) for each singular value (#SV). Lower panel shows the root-mean-square-error (RMSE) between the observed data and the reconstructed data versus singular value number (#SV). In the lower panel, for example, the data plotted at #SV = 3 correspond to the RMSE between observed data and reconstructed data using singular values 1, 2 and 3 only (see Eq. (B.3)).

Table 1

Data errors for In-Phase (IP) and Quadrature (Q) at each frequency. Errors are estimated as the standard deviation of the data acquired along the Bundoran test line in 2015 for each component and frequency, for the highest flight altitude (240 m). The data errors highlighted in bold were used as input for Tikhonov-type 1-D inversion of the Bundoran test line data. L2240020.15: Line 2, 240 m flight altitude, Northbound, 2nd attempt, version 0, year 2015; L2240120.15: Southbound, 2nd attempt; L2240030.15: Northbound, 3rd attempt; L2240130.15: Southbound, 3rd attempt.

	L2240020.15	L2240120.15	L2240030.15	L2240130.15	mean	median
f (Hz)	IP Q (ppm)	IP Q (ppm)	IP Q (ppm)	IP Q (ppm)	(ppm)	(ppm)
912	58 45	54 54	54 41	63 47	57 47 (52)	54 54 (54)
3005	39 25	27 21	54 14	67 17	47 19 (33)	47 19 (33)
11,962	54 53	68 67	63 46	75 39	65 51 (58)	66 50 (58)
24,510	55 25	46 28	50 41	47 45	50 35 (42)	49 35 (42)
mean	52 37 (45)	49 43 (46)	55 36 (46)	63 37 (50)	55 38 (47)	–
median	55 35 (45)	50 41 (46)	54 41 (48)	65 42 (54)	–	52 41 (47)

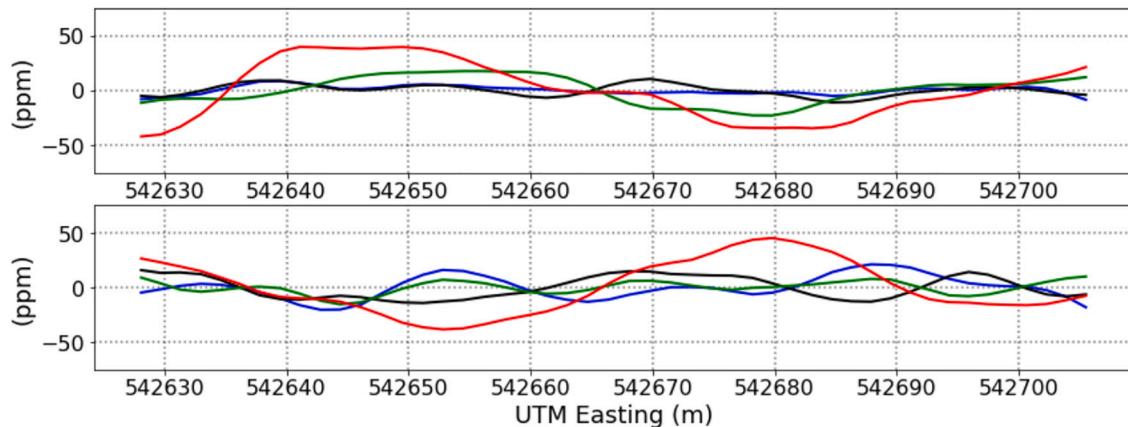


Fig. 10. Residuals between original data and data reconstructed by SVD truncation for the Bundoran Test Line for data acquired at a 60 m flight altitude. The data were reconstructed using the first two singular values (i.e., 1 and 2) only.

30 layers, with the thickness of the top layer starting at 2 m and subsequent layers increasing logarithmically with depth till 98 m depth. The layer thicknesses were preserved during the inversion process. The prior/initial model was a uniform half-space with a resistivity of $100 \Omega\text{m}$ for the first site on the profile and each site was modelled using the previous site's inversion model as the initial model for the next site. Errors assigned for in-phase and quadrature data are provided in Table 1, which are the estimates based on the standard deviation of the highest altitude (240 m) data.

Fig. 11 summarises the inversion results together with a coincident Electrical Resistivity Tomography (ERT) model. The ERT model is obtained from inversion of data acquired in 2012 along an approximately 300 m-long portion of the airborne test line. The ERT data were collected by GSI using 64 electrode Wenner array at 5 m electrode spacing and then inverted using the RES2DINV software (Loke, 2019). The results suggest that both methods are sensitive to a common subsurface model. The model sensitivity defined in Appendix A indicates that the AEM data have the highest sensitivity to the highly conductive middle layer. Note that inversion results were obtained without including in-phase data at 912 Hz and quadrature data at 11,962 Hz for the 60 m altitude inversion and in-phase and quadrature data at 912 Hz and 11,962 Hz for the 90 m altitude inversion due to poor data quality and corresponding bad data fits.

4.2. Numerical studies

In order to understand the effects of the high-fly zones on the inverted subsurface models, we have performed several numerical experiments. First of all, an AEM survey was simulated over a simple subsurface model which consists of three conductive targets embedded in a uniform $100 \Omega\text{m}$ half-space (Fig. 12). The targets have a resistivity of $5 \Omega\text{m}$ and are embedded at depths of 5 m, 15 m and 25 m respectively. The 2-D forward calculations were performed using ArjunAir (a product of the CSIRO and the Amira Consortium, project P223F) code. Note that these forward calculations were carried out separately since ArjunAir is not integrated into aempy. Measurements were simulated every 10 m along the flight line at flight altitudes of 60 m (Fig. 12A and C) and 100 m (Fig. 12B and D). The in-phase and quadrature components of the synthetic data were computed at four frequencies and were then randomly perturbed by noise of 30 ppm (Fig. 12A and B) and 50 ppm (Fig. 12C and D) for all frequencies, with a correlation length of 200 m, comparable to the values given in Table 1.

These data were inverted using 1-D smooth, multi-layer Tikhonov-type inversion. The model consists of 30 layers, with the thickness of the top layer starting at 2 m and subsequent layers increasing logarithmically in thickness with depth till 98 m depth. The layer thicknesses

were preserved during the inversion process. A prior resistivity of $50 \Omega\text{m}$ was chosen and each site was modelled using the previous site's inversion model as the initial model for the next site. Although not used here, aempy does provide other options, performing inversions fully independent for each site.

Figs. 13 and 14 present the inversion results and shows that the conductive bodies in the subsurface model in Fig. 12 have been successfully recovered. The resistivity of the surrounding host rock ($100 \Omega\text{m}$) is generally not well recovered, particularly near the conductors. The down-dropping structures at the boundaries of the conductors are a well-known effect of the 1-D inversion of 2-D "observations". The sensitivity of the airborne data to resistors in the subsurface is generally low and is significantly reduced at depths below the conductive bodies, as clearly demonstrated in the sensitivity plots. As expected, inversions of 60 m flight altitude data do a much better job recovering the conductive bodies in particular for the deep conductor. The inversion result from 100 m data with 50 ppm noise added is not able to recover the true structures well.

Secondly, to further analyse the uncertainty introduced by the flight altitude we applied a Bayesian Markov Chain Monte Carlo (MCMC) technique (Sambridge and Mosegaard, 2002; Tarantola, 2005) as described in Appendix A. We used a simple three-layer resistivity model to demonstrate the effectiveness of our algorithm and to understand the flight altitude effect on the data and thus the recovered subsurface models. The resistivity model comprises a 15 m thick first layer with a resistivity of $100 \Omega\text{m}$, a 25 m thick middle layer with a resistivity of $5 \Omega\text{m}$, and a lower half-space with resistivity of $100 \Omega\text{m}$. The forward response of this model is computed using the method described in Appendix A. The simulated airborne system is the one used for the Tellus surveys: two pairs of co-planar coils operating at four frequencies with TX–RX separation of ~ 21.4 m. An assumed data error was set to 30 ppm and 30 ppm random noise was applied to the data.

The results of single-site Bayesian simulations using the Metropolis-Hastings (M-H) algorithm are summarised in Fig. 15A for the flight altitudes of 60 m, 90 m and 150 m. The M-H algorithm was run for 1,000,000 model samples. The prior/initial model was a uniform half-space with a resistivity of $100 \Omega\text{m}$. The results obtained with 60 m flight altitude appear to be accurate with respect to the resistivity and the thickness of the first two layers, which are recovered successfully. However, the resistivity of the bottom layer has not been accurately resolved. Inclusion of the highly conductive (or its inverse less resistive, $5 \Omega\text{m}$) middle layer decreases the resolution of the bottom layer. On a positive note, the results confirm that the characteristics of the Tellus FDEM system provide a useful tool for the investigation of shallow subsurface structures, perhaps to depth of around 50 m, resistivity dependant. A summary of the results for all altitudes is shown in

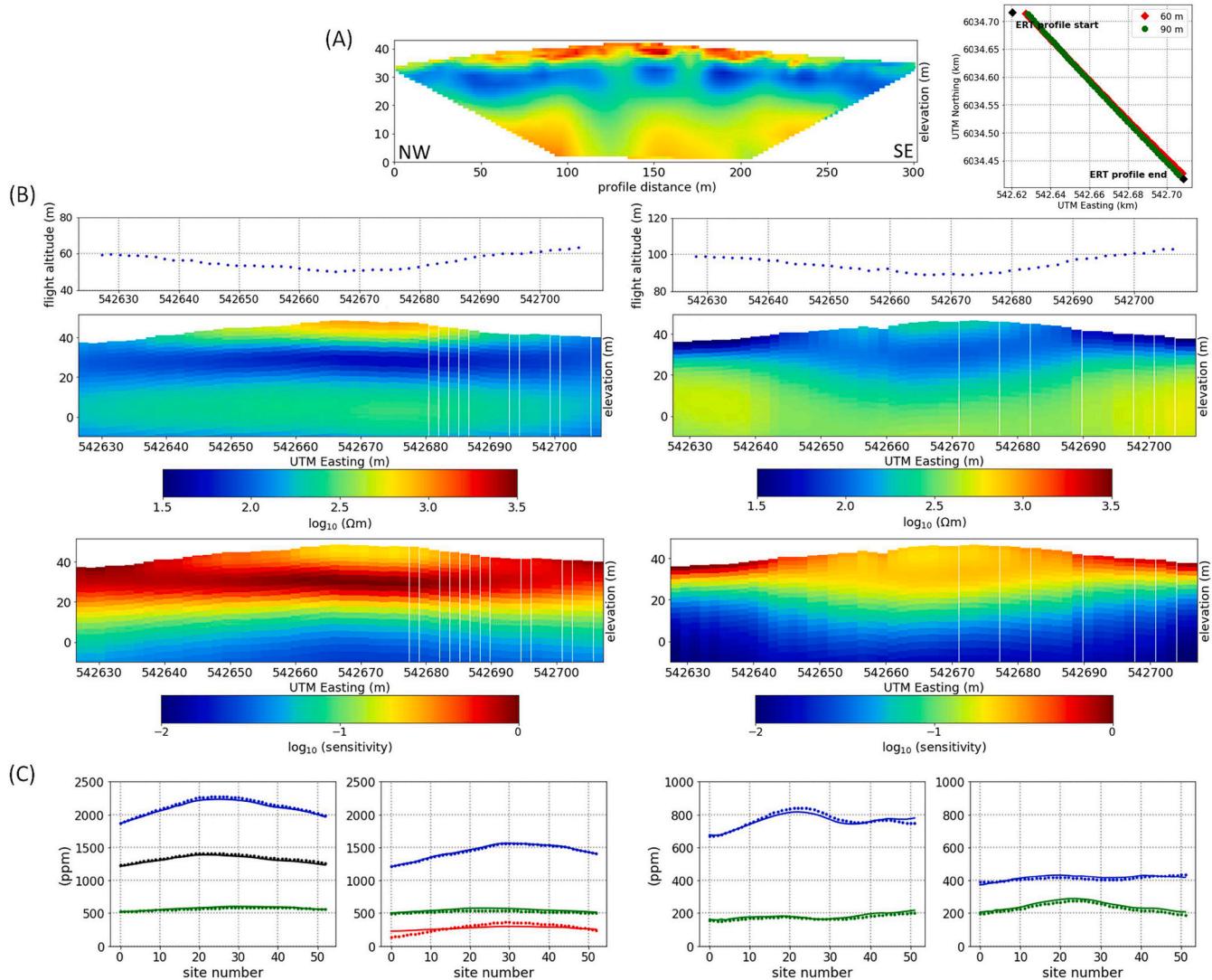


Fig. 11. Bundoran Test Line 1-D inversion results. (A) Electrical resistivity model from the inversion of Electrical Resistivity Tomography data together with an overview map showing the locations of ERT profile and Bundoran Test Line data acquired at a flight altitude of 60 m and 90 m. ERT data were acquired with a 5 m spacing between electrodes distributed along the profile length. (B) Coincident electrical resistivity model and model sensitivity from the inversion of the PCA-filtered AEM data acquired at 60 m (left) and 90 m flight altitude (right). Regularisation parameters, $\tau_0 = 0.1$ and $\tau_1 = 6.6$ were chosen by trial-and-error. (C) Observed data (solid circles) for In-Phase (left) and Quadrature (right) and calculated data (lines) at the four survey frequencies 24,510 Hz, 11,962 Hz, 3005 Hz and 912 Hz, shown in blue, black, green and red colours respectively. In-Phase data at 912 Hz and 11,962 Hz were deactivated for the 60 m altitude inversion and In-Phase and Quadrature data at 912 Hz and 11,962 Hz were deactivated for the 90 m altitude inversion. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 15B, separated into the low altitudes (left panel) and the high altitudes (right panel). The range of flight altitudes used for the simulations - 60 m, 70 m, 80 m, 90 m, 100 m, 120 m, 150 m, 180 m and 240 m - was selected as it corresponds with the range of altitudes typically flown during the Tellus survey. The median of accepted samples for each altitude is shown together with the true model (dashed line). It can be concluded that AEM data acquired at altitudes over 90 m should be inverted with caution.

5. Discussion

The need to process and invert large volumes of AEM data places a number of requirements on the inversion strategy used and on the inversion codes that support it. The primary requirements are: (i) automation of the process as far as possible, (ii) generation of a range of inversion-model quality-control parameters that be used collectively for automated, objective rejection of poor model solutions, and (iii)

relatively fast inversion computational speeds (Muller, 2022). One of the main advantages of the aempy toolbox lies in its Python coding, allowing the development of customised scripts for automation and for customised output of inversion models and desired model quality-control parameters. The capacities of the toolbox are implemented in a number of high-level scripts that cover a full work flow (i) loading and reformatting of raw EM data, (ii) pre-inversion processing of EM responses, (iii) inversion modelling and (iv) visualisation of outputs. Future work will involve quantitative treatment of model uncertainty which goes beyond the current version of the aempy toolbox. Post-inversion outputs provided by aempy are the model sensitivity matrix (essentially the Jacobian matrix), model posterior covariances and resolution matrices. However, the significance of the latter quantities depends on the assumptions about error statistics and the regularisation parameters chosen. These assumptions are probably too narrow. Future development will be based on improved stochastic methods [e.g., (Foks and Minsley, 2020)], which are challenging to be used on data sets as large

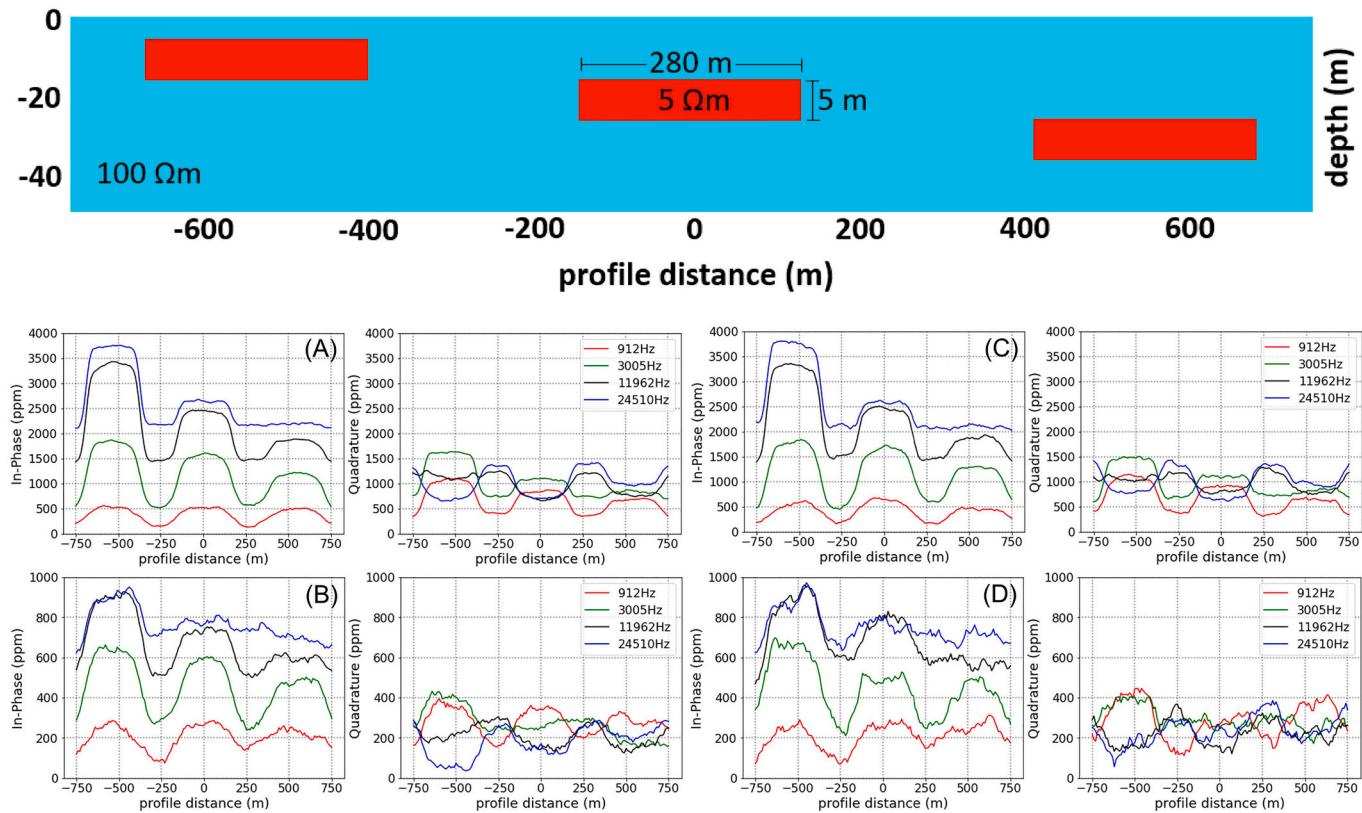


Fig. 12. Simulated model consisting of three conductive ($5 \Omega\text{m}$) targets embedded in a $100 \Omega\text{m}$ half-space and the forward response of the model calculated at four Tellus survey frequencies using 2-D ArjunAir code, at 10 m intervals along the profile for flight altitudes of 60 m (A and C) and 100 m (B and D). (A)-(B): The In-Phase and Quadrature components of the synthetic data were randomly perturbed by noise of 30 ppm for all frequencies. Note that 30 ppm is a simplified error assumption for this synthetic example. (C)-(D): The In-Phase and Quadrature components of the synthetic data were randomly perturbed by noise of 50 ppm for all frequencies.

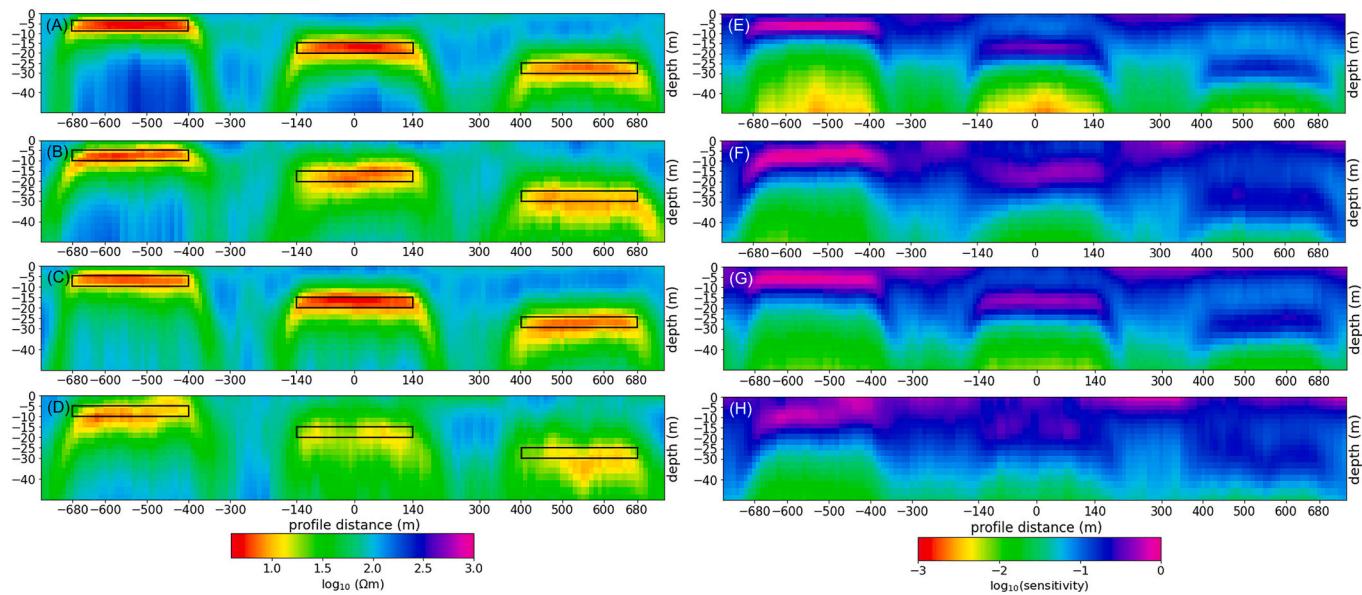


Fig. 13. 1-D smooth, multi-layer Tikhonov-type inversion results of the synthetic data of Fig. 12: (A-B) the inversion models and (E-F) the model sensitivity plots for the flight altitudes of 60 m and 100 m, respectively. Data error assumption was 30 ppm. (C-D) the inversion models and (G-H) the model sensitivity plots for the flight altitudes of 60 m and 100 m, respectively. Data error assumption was 50 ppm. Regularisation parameters, $\tau_0 = 0.3$ and $\tau_1 = 9.0$ were chosen.

as the ones generated by the Tellus programme due to the number of data points, and the requirements related to the “curse of dimensionality” (Bellman, 1957) for each of the sites.

As mentioned previously, the recorded Tellus AEM data are subject

to high cultural noise levels across many parts of Ireland and geological signal strength is strongly attenuated in high-fly zones, both of which impact on the reliability and quality of the resulting inversion models. We have investigated the estimation of data observational errors that are

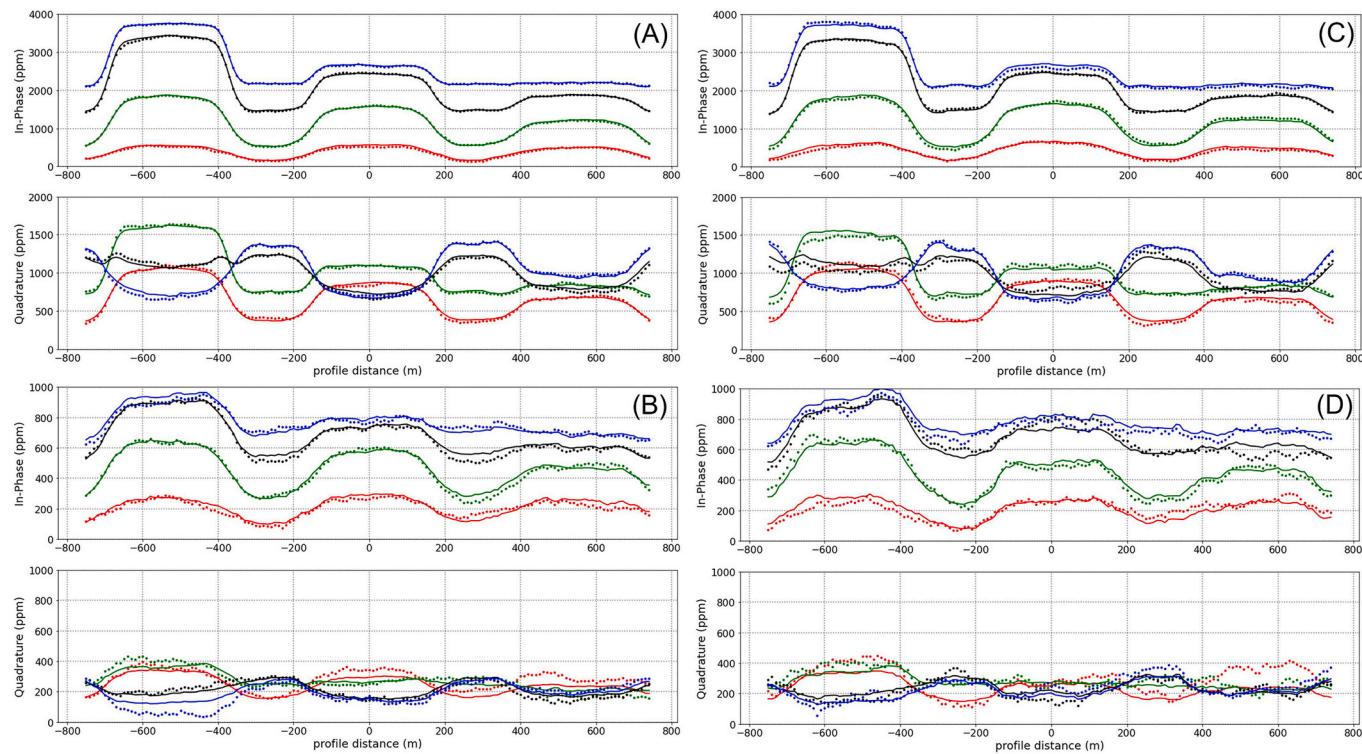


Fig. 14. 1-D smooth, multi-layer Tikhonov-type inversion results of the synthetic data of Fig. 12: the data fit curves for the inverted models in Fig. 13A - 13D, respectively. Predicted (solid lines) and observed (solid circles) for In-Phase and Quadrature components at the four frequencies 912 Hz (red), 3005 Hz (green), 11,962 Hz (black) and 24,510 Hz (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

crucial for any inversion, as well as the influence of the flight altitude parameter on the inverted subsurface models using synthetic data and field data from dedicated Tellus high-fly experiment in 2015. The SVD approach implemented in aempy can be applied prior to inversion with the objective of reducing noise and imposing regularity and consistency on the EM response data. The approach can be useful to analyse data and can be applied to the very large Tellus project data sets. We have used approximate (in other words reconstructed) data based on the first two or three singular components (e.g., determined by $RMSE > 50$ ppm) successfully for the stable inversion of flight line data sets, leading to smooth and geologically informative models. However, further improvements are possible. Clearly, the approach lends itself to the development of an iterative procedure by taking into account an increasing number of singular components, until an optimal trade-off between resolution and stability is achieved. A good prior/initial model for further inversions could be derived from an initial inversion using only the largest singular component.

The key to the generation of informative and stable subsurface models is to assign observational errors for the input data. For this purpose, we have made use of the Tellus test line data sets which provided a reasonable and consistent view. Further investigations are needed to study the physical origin of these errors, which is currently unknown. While numerical experiments presented in Section 4.2 show that direct influences from the surface is improbable, it is well-known that significant errors can originate from systematic errors of altitude measurements, which are related to surface conditions such as spatial changes in vegetation ("canopy effect") and are analysed in detail in (Beamish, 2002; Beamish and Levanniemi, 2010). Although we expect these effects to be smaller than the ones calculated in these publications, where a lower flight altitude (30 m) is assumed, it is possible to generate errors comparable to the ones estimated in this study. However, it is plausible to think that atmospheric processes are involved here. It is well-known that the atmospheric boundary layer, and in particular its lower parts can be highly heterogeneous in space and time [e.g., (Lee,

2018)]. While a direct influence is not easy to conceive, the inhomogeneities in atmospheric conditions (e.g., in wind speed) may produce perturbations in the movements of the plane, i.e., in pitch, roll and yaw. These parameters are currently not available in the Tellus database. If this hypothesis is correct, they might be used for a correction during the pre-inversion processing.

With respect to the widespread problem of high-fly zones in Ireland, our study has shown that the content of geological information for altitudes above ≈ 90 m is low, leading to smooth and smeared images of the subsurface. There is the danger of artefacts resulting from the low signal-noise-ratio at high altitudes. While altitude requirements can be somewhat relaxed when using the data for mapping purposes, the extraction of quantitative information on subsurface structures is only possible at lower flight altitudes. Our results are consistent with a previous repeatability study (Hegarty et al., 2022), which is based on purely statistical considerations. The authors conclude that the repeatability of the derived apparent resistivities is already strongly reduced at ≈ 90 m for all data components at four Tellus frequencies.

6. Conclusions

We have presented aempy, an open-source toolbox for pre-inversion processing and inversion in airborne electromagnetics, is particularly designed for extracting quantitative information from the Tellus AEM data. This was demonstrated by one-dimensional inversion of synthetic and Tellus data sets that allow to image the electrical conductivity distribution of the subsurface. Although these examples were primarily intended for demonstrating the performance of aempy, they also serve to improve our understanding of the sources of error in Tellus frequency-domain AEM measurements and in particular the effects of ubiquitous Tellus high-fly zones. We also propose that separating signal from noise prior to inversion using the SVD approach implemented in aempy may improve the quality of the resulting subsurface models.

Our results were obtained using the aempy toolbox which is

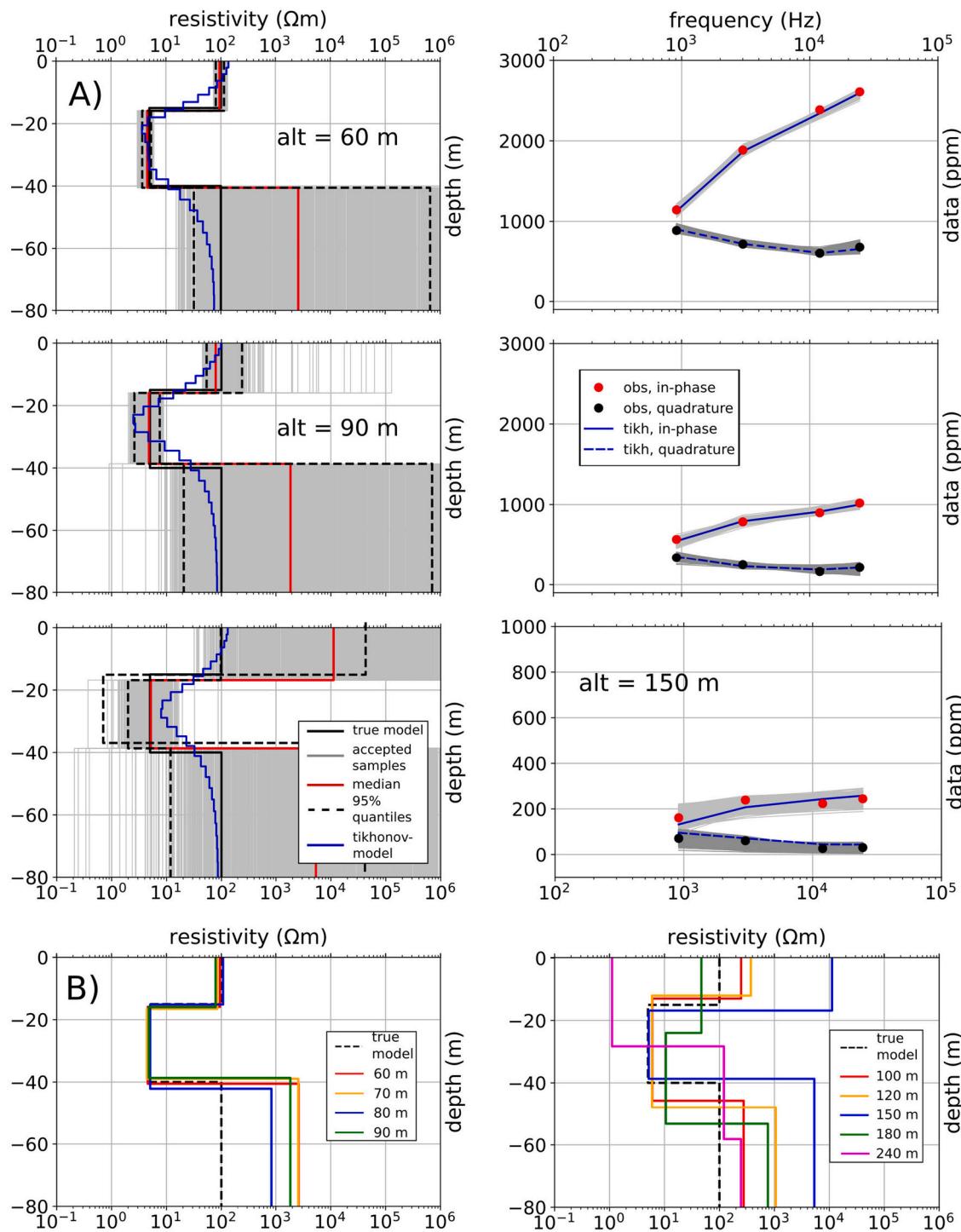


Fig. 15. Summary of the output from the Metropolis-Hastings (M-H) simulations. The M-H algorithm was run for 1,000,000 model samples. The computational core is based on an adapted frequency-domain forward modelling algorithm from the well-tested, open-source Airbeo, which was developed by the CSIRO (Australia) and the AMIRA Consortium. A) The inversion results for the altitudes of 60 m (top row), 90 m (middle row) and 150 m (bottom row). Every 500th of accepted models (light grey lines) are shown on the left panel. Superimposed are the true model (black solid line), bounds that contain 95% of the accepted models (black dashed lines), and the median resistivity values (red solid line). The right panel illustrates the data fit on which the predicted data are represented with light and dark grey solid lines, and the observed data are represented with the solid circles. The blue lines represent the corresponding deterministic Tikhonov inversion. B) The summary of all altitudes showing median value for low altitudes (left panel) and high altitudes (right panel) of the three-layer model. The dashed line on each graph represents the true model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

available from the repository https://git.dias.ie/vrath/aempy_public. The scripts and data sets for reproducing the results presented in this study can be obtained from the corresponding author. aempy goes beyond the capabilities of other packages currently available, as it

integrates not only a number of different inversion approaches, but also provides a range of pre-inversion processing tools specifically relevant to the analysis, and interpretation of AEM data, which are under continuous development. While aempy offers many options for pre-inversion

processing and inversion, we want to emphasise that the examples given in this study are not the only reasonable workflows. The necessary decisions on the methods of processing and inversion parameters must be made by the user, and very probably need to be adapted to the local situation, and the requirements of the particular target.

Author contribution statement

DK and VR developed the aempy toolbox. DK performed the computations presented in this manuscript. All the authors discussed the results and contributed to the manuscript. DK wrote the manuscript with support from VR, MRM, MDT, and JH.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

the work reported in this paper.

Acknowledgements

This work was primarily supported by the Geological Survey Ireland Research Programme (Short Call) under grant sc2015-004. DK and VR acknowledge the support by a research grant from Science Foundation Ireland (SFI) under Grant Number 13/RC/2092 and co-funded under the European Regional Development Fund and iCRAG industry partners. Tellus AEM data analysed in this study are provided by the Geological Survey Ireland, part of the Department of Environment, Climate and Communications©. DK acknowledges her student Méabh Ní Mhuilleoir's help on the Bayesian approach work, which was carried out as part of her Senior Sophister Project (October 2018 - January 2019). Australias's CSIRO and the Amira International Consortium are thanked for making their P223 modelling suite open access.

Appendix A. 1-D inversion of FDEM data

The details on the inversion applied to the synthetic and observed data in the study are given below. More information can be found in the aempy manual (Kiyani and Rath, 2017), available with the toolbox.

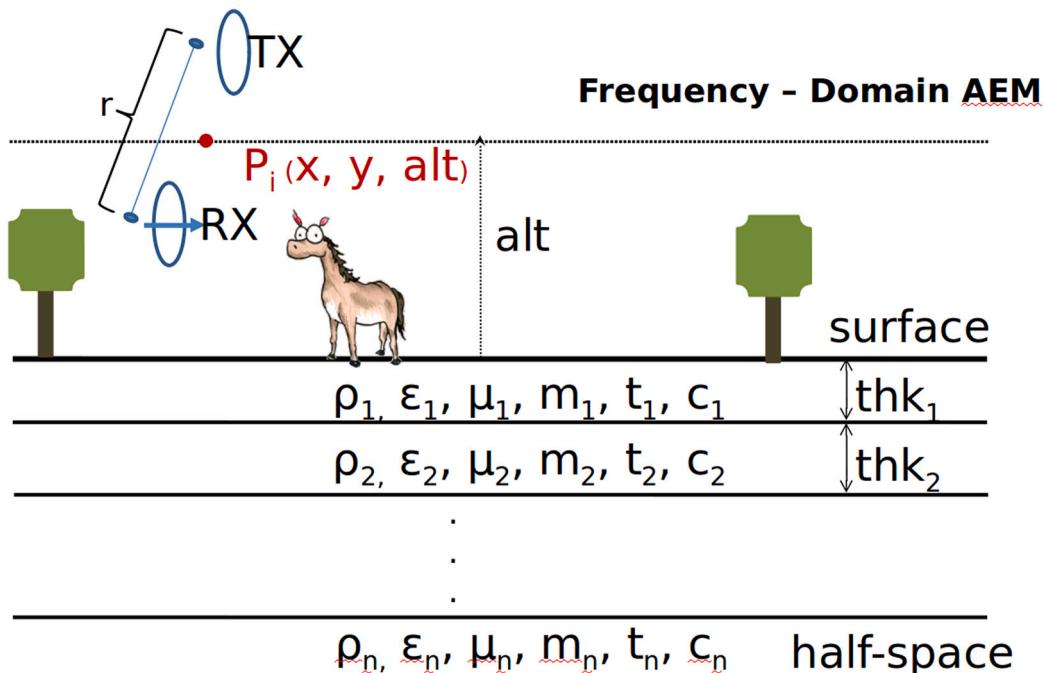


Fig. A.16. Schematic view of a VCP loop configuration over a layered-earth model. The AEM response is positioned at the centre between the Transmitter (TX) and Receiver (RX). The spacing (r) between them is kept constant. For the Tellus FDEM Surveys, $r = 21.35$ m for 912 and 3005 Hz and $r = 21.38$ m for 11,962 and 24,510 Hz.

Forward problem in 1-D. The FDEM methods considered in this study are based on the EM induction where time varying magnetic field generated by the sinusoidal current of the transmitter (TX) generates currents in the subsurface conductors. These currents (primary EM fields) generates secondary EM fields where the eddy currents induced in the conducting subsurface are included. The total magnetic field that will be measured at the receiver (RX) loop is the sum of the primary (B_p) and secondary (B_s) magnetic fields. The FDEM response is defined as the secondary magnetic field (B_s) normalised by the primary (free space) magnetic field ($B_p = B_0$). Note that due to the complex nature of the fields in the frequency domain a phase shift between B_p and B_s exists. Thus the AEM response is commonly reported as In-Phase (IP) and Quadrature (Q) components defined as

$$\begin{aligned} IP &= \left(\frac{\Re(B_{tot})}{B_0} - 1 \right) \\ Q &= \left(\frac{\Im(B_{tot})}{B_0} \right), \end{aligned} \tag{A.1}$$

respectively, and is presented in units of parts per million (ppm). The response is computed at data location $P_i(x, y, h)$, defined as the centre point between TX and RX. x , and y are the coordinates of the data location, and h is the flight altitude.

The AEM response of a dipole source and a receiver above a layered-earth is computed using the solutions given in (Keller and Frischknecht, 1966) as

$$\frac{B_s}{B_0} = r^2 \int_0^\infty R(\lambda) \lambda e^{-2\lambda h} J_1(\lambda r) d\lambda , \quad (\text{A.2})$$

where r is the coil separation, λ is the variable of integration, h is the elevation of the TX and RX coils, and J_1 Bessel function of the first kind of order one. The term $R(\lambda)$ can be written as

$$R(\lambda) = \frac{Y_0 - \hat{Y}_1}{Y_0 + \hat{Y}_1} , \quad (\text{A.3})$$

where \hat{Y}_1 is the surface admittance and $Y_0 = u_0/(i\omega\mu_0)$ intrinsic admittance, i is the imaginary unit ($i^2 = -1$), ω is the angular frequency ($\omega=2\pi f$), and μ_0 is the magnetic permeability of free space. For an N-layered earth, \hat{Y}_1 can be obtained by the following recurrence relation (Wait, 1962):

$$\hat{Y}_n = \frac{\hat{Y}_{n+1} + Y_n \tanh(u_n t_n)}{Y_n + \hat{Y}_{n+1} \tanh(u_n t_n)} , \quad (\text{A.4})$$

with n ranging from 1, ..., N , $Y_n = u_n/(i\omega\mu_n)$, and $\hat{Y}_N = Y_N$ in the N -th layer. In the n -th layer, t_n is the layer thickness, and $u_n = (\lambda^2 + k_n^2)^{1/2}$. k_n is the wave number given by

$$k_n = \sqrt{-i\omega\mu_n\sigma_n + \omega^2\epsilon_n\mu_n} , \quad (\text{A.5})$$

where σ_n is the complex conductivity (e.g., in the Cole-Cole parametrisation) of the n -th layer, $\mu_n = \mu_0\mu_r, n$ and $\epsilon_n = \epsilon_0\epsilon_r, n$ are the magnetic permeability and the dielectric permittivity of the n -th layer, respectively.

The required Hankel transform integrals in Eq. (A.2) are realised using the filtering technique introduced by (Anderson, 1975).

In contrast to many existing codes, the 1-D (layered) forward model (Figure A.16) assigns seven physical properties to each layer n , including its thickness of $\delta z_n = th_k n$. The layer physical parameters are the electrical resistivity ρ_n , the electrical permittivity $\epsilon_n = \epsilon_r, n\epsilon_0$, the magnetic permeability $\mu_n = \mu_r, n\mu_0$, where the index 0 marks the corresponding free space value of the magnetic permeability and the dielectric permittivity, the r denotes the relative value used for the parameterisation in the respective layer. The remaining properties are the three Cole-Cole parameters related to induced polarisation, namely the chargeability m_n , the time constant t_n , and the frequency constant c_n (Cole and Cole, 1941). Induced polarisation effects may be relevant in areas with high clay content or disseminated metallic particles (e.g., ores) (Slater, 2007; Revil et al., 2017). Therefore, this general set-up makes it possible to model a layered Earth response for fairly general targets. It is used exclusively for the electrical resistivity parameter in this study. Potential IP effects within the Tellus data (e.g. related to the Moffat shales (Römhild et al., 2019)) will be investigated in the future.

Solving the inverse problem. Inverse techniques may be classified by the following different paradigms. From a more philosophical point of view, the most important division is between the regularisation-based approach [e.g., (Tikhonov, 1963; Parker, 1994; Aster et al., 2013)], and the Bayesian approach [e.g., (Rodgers, 2000; Mosegaard and Sambridge, 2002; Tarantola, 2005)]. The former seeks to find the “best” model according to a given criterion which is usually achieved by deriving an appropriate optimisation procedure which leads to one preferred model. Inversion/optimisation approaches usually (though not necessarily) require the calculation of gradients, Jacobians, or even Hessians with respect to the estimated parameters, often leading to high computational costs.

Starting from the observation that many problems in practical mathematics are ill-posed in the sense of Hadamard (Hadamard, 1923; Hansen, 2010), Tikhonov (Tikhonov, 1963; Tikhonov and Arsenin, 1977) suggested solving a “near-by” problem, which is formed by adding an additional regularising term, which renders the problem solvable, in particular ensuring the uniqueness of the outcome. In this way the solution of the inverse problem is transformed into an optimisation problem, where an objective function

$$\theta = \mathcal{D}(\mathbf{d}, \mathbf{m}) + \mathcal{R}(\mathbf{m}) , \quad (\text{A.6})$$

is minimised. \mathcal{D} measures the data fit (fidelity), while \mathcal{R} is called regulariser, and might contain one or more regularisation operators with corresponding weighting parameters τ , which are required to be determined beforehand or during the optimisation process.

This framework is very general, and the result of this inversion depends strongly on the type of the operator chosen for \mathcal{D} , and the regularisation parameters, \mathcal{R} . While the choice of \mathcal{D} is usually a weighted Euclidean norm of the residual $\mathbf{r} = \mathbf{d} - \mathbf{g}(\mathbf{m})$, i.e., $\|\mathbf{r}\|_2^2$, other options are possible which may increase the robustness of the approach (Huber, 1981; Farquharson and Oldenburg, 1998). In this study, an Euclidean semi-norm is used, leading to

$$\mathcal{D} = (\mathbf{d} - \mathbf{g}(\mathbf{m}))^T \mathbf{W}_d^T \mathbf{W}_d (\mathbf{d} - \mathbf{g}(\mathbf{m})) . \quad (\text{A.7})$$

In this equation, $\mathbf{g}(\mathbf{m})$ represents the solution of the forward problem for parameter vector \mathbf{m} at the data points \mathbf{d} . The weighting \mathbf{W}_d is commonly constructed from an estimation of the associated data error as a diagonal matrix of its inverse value.

The most common type of \mathcal{R} operators (and the ones implemented here) are a sum of simple operators, including the identity matrix \mathbf{I} , and discrete representations of low order derivatives [e.g., (Donatelli et al., 2012; Aster et al., 2013; Donatelli and Reichel, 2014)]. Each subterm is built from appropriate matrices \mathbf{W}_k , multiplied by a corresponding weighting τ_k , leading to

$$\mathcal{R} = \sum_k \tau_k (\mathbf{m} - \mathbf{m}_{prior})^T \mathbf{W}_k^T \mathbf{W}_k (\mathbf{m} - \mathbf{m}_{prior}) , \quad (\text{A.8})$$

where we have added a reference (“prior”) model \mathbf{m}_{prior} for convenience. With these settings, the common procedure of differentiation with respect to

the model parameters and setting the result to zero leads to an iteration scheme

$$\begin{aligned} & \left(\mathbf{J}_w^T \mathbf{J}_w + \sum_k \tau_k \mathbf{W}_k^T \mathbf{W}_k \right) \delta \mathbf{m}_{iter} = \\ & \mathbf{J}_w^T [\mathbf{d} - \mathbf{g}(\mathbf{m}_{iter})] - \sum_{k=0}^1 \tau_k \mathbf{W}_k^T \mathbf{W}_k \left(\mathbf{m}_{iter} - \mathbf{m}_{prior} \right) \\ \mathbf{m}_{iter+1} &= \mathbf{m}_{iter} + \mu \delta \mathbf{m}_{iter} \end{aligned} \quad (\text{A.9})$$

Here, we have used the Jacobian matrix \mathbf{J} , defined by $J_{ij} = \frac{\partial g_i(\mathbf{m})}{\partial m_j}$ scaled with the data errors, $\mathbf{J}_w = \mathbf{W}_d \mathbf{J}$. After solving the system of linear equations for $\delta \mathbf{m}_{iter}$, in the second step, an additional line search is performed if the objective function is not decreased.

Finally, the values for the weighting parameters are determined by the Generalised Cross Validation (GCV) method (Wahba, 1990; Farquharson and Oldenburg, 2004; Rath and Mottaghy, 2007). Here, for each choice of τ , the GCV function is calculated and minimised:

$$GCV(\tau) = \frac{N \|\mathbf{d} - \mathbf{g}(\mathbf{m}_\tau)\|_2^2}{\text{trace}((\mathbf{I} - \mathbf{J}_w \mathbf{J}_w^\dagger)^2)} \quad (\text{A.10})$$

Here, N is the number of data points. Finding the minimum implies not only the evaluation of many forward models, as trial values for τ are used, but also the calculation of the Generalised Inverse $\mathbf{J}_w^\dagger = \left(\mathbf{J}_w^T \mathbf{J}_w + \sum_k \tau_k \mathbf{W}_k^T \mathbf{W}_k \right)^{-1} \mathbf{J}_w$, and thus is computationally expensive for large scale problems. In practice a choice of fixed values based on numerical experiments on the data set considered, or a schedule of relaxing the regularisation with iteration number often works well. In spite of the need to calculate the Jacobian, this approach requires orders of magnitude shorter computational time than the MCMC approach. The error-weighted Jacobian is also used to calculate the model sensitivity $\text{diag}(\mathbf{J}_w^T \mathbf{J}_w)^{1/2}$. For the figures presented in this study, sensitivity values were divided by the respective layer thicknesses in order to eliminate their influence (Schwabenberg et al., 2002).

The Bayesian approach aims at characterising the full parameter space involved, starting from the well-known formula

$$p(\mathbf{m} | \mathbf{d}) = \frac{p(\mathbf{d} | \mathbf{m}) p(\mathbf{m})}{p(\mathbf{d})}, \quad (\text{A.11})$$

which relates the posterior probability $p(\mathbf{m} | \mathbf{d})$ of the model \mathbf{m} given the data \mathbf{d} to the likelihood $p(\mathbf{d} | \mathbf{m})$ and the prior $p(\mathbf{m})$. The denominator is usually disregarded, unless it is needed for some of the more advanced methods as trans-dimensional schemes (Green, 1995; Minsley, 2011). In this approach, we seek to characterise the full posterior probability distribution of the parameters, given our prior knowledge, and the additional information contributed by the data under consideration.

From the strictly Bayesian point of view, Markov Chain Monte Carlo (MCMC) methods will be the natural choice. The most common of these is the famous Metropolis-Hastings (M-H) algorithm (Metropolis et al., 1953; Hastings, 1970), which is still the “working horse” for most cases although many improvements and generalisations have been published in recent years (Green, 1995; Minsley, 2011; Haario et al., 2006; Goodman and Weare, 2010). Note that there is no necessity to calculate Jacobians or even Gradients as in traditional inversion techniques. However, the numbers of required evaluations can become excessively large, as this number grows highly non-linear with the number of parameters - the “curse of dimensionality” (Bellman, 1957; Curtis and Lomax, 2001).

We have implemented the M-H algorithm in its most simple form with a Gaussian prior. However, in order to be able to include spatially correlated priors, this approach was modified to include this type of constraints. The prior is not drawn from a Gaussian random variable, from a spatially correlated distribution. Using the assumed or estimated spatial covariance \mathbf{C}_m of the model, sampling from this distribution can be realised by first calculating the Cholesky decomposition of this matrix as $\mathbf{C}_m = \mathbf{L} \mathbf{L}^T$, where \mathbf{L} is a lower triangular matrix and \mathbf{L}^T denotes the transpose of \mathbf{L} . Then, a correlated distribution $\tilde{\mathbf{m}}$ can be constructed from \mathbf{m}_0 drawn from a Gaussian distribution $N(0, \sigma)$ as $\tilde{\mathbf{m}} = \mathbf{L}^T \mathbf{m}_0$. In this study, we have assumed 1-D Markovian (exponential) prior covariances

$$C(d_{ij})_m = \sigma^2 \exp\left(-\frac{|\mathbf{r}_i - \mathbf{r}_j|}{L}\right), \quad (\text{A.12})$$

characterised by a vertical correlation lengths $L_{(z)}$. This approach is highly flexible and can easily be adapted to general 2-D or 3-D anisotropic conditions by introducing appropriate transformed distances d . As spatial parameter correlations are rarely known beforehand, the covariances are usually chosen based on prior physical or geological considerations.

Note that the deterministic approaches developed above are structurally similar to the Bayesian Maximum A Posteriori (MAP) estimation and lead to very similar solution procedures by redefining the weighting operators. It is well-known that there are relationships between covariance matrices and the differential operators mentioned above. As shown in (Tarantola, 1987) or (Rodgers, 2000), the inverse of the exponential covariance given in Eq. (A.12) can be generated using a weighted sum of the identity \mathbf{I} and the $\mathbf{W}_1^T \mathbf{W}_1$ terms. However, no such simple representations of the inverse are known for other spatial covariances (Xu, 2005; Yaremchuk et al., 2018).

We have invested a lot of time and effort in making the inversion set-up as flexible as possible. On the data side, data can be activated for inversion by flagging them as inactive individually or in groups. The 1-D models used now can have up to seven parameters per layer for inversion, if the layer thickness is included: the electrical resistivity ρ , the relative dielectric constant ϵ , the relative magnetic permeability μ , and the three Cole-Cole parameters chargeability m_0 , time constant t , and frequency constant c . In principle all these parameters can be estimated, although it might be too optimistic at least for the frequency domain data to obtain such a large number of parameters from eight observation points. However, this capability may help understand the observations in particular when there is no reasonable distribution of resistivity consistent with the data (e.g., negative values in the responses). As in the case of the observed data, the parameters used in the inversion can easily be activated and deactivated. This is also true for the layer thicknesses, so that the inversion allows for highly resolved models with fixed layers, as well as for joint estimation of properties and thicknesses. Positivity of parameters may be enforced by working with their logarithmic forms.

Appendix B. Singular value decomposition of the observation matrix

Assuming that the standard site-wise pre-inversion processing (flagging, deleting, or interpolating measurement sites in high-fly zones or in the vicinity of man-made installations) has been successfully applied to the observed data, the data can then be arranged in a matrix \mathbf{D} with n_{site} (number of sites) rows and n_{comp} (number of data components) columns. Prior to computing the SVD defined in Eq. (B.1), the mean value for each data component is subtracted from matrix \mathbf{D} so that the columns have zero mean.

$$\mathbf{D} = \mathbf{U}\mathbf{S}\mathbf{V}^T, \quad (\text{B.1})$$

where \mathbf{U} and \mathbf{V} are unitary matrices, and \mathbf{S} is a diagonal matrix, and contains the singular values in decreasing sequence, but only n_{comp} of these have non-zero values.

Filtering is accomplished by reconstructing \mathbf{D} from only k of the singular values and singular vectors, where $k < n_{comp}$,

$$\mathbf{D}_{\text{filtered}} = \mathbf{D}_k = \mathbf{U}_k \mathbf{S}_k \mathbf{V}_k^T. \quad (\text{B.2})$$

Choosing the k largest values, truncating the matrices correspondingly, and reintroducing the column averages, an approximate matrix \mathbf{D}_k is obtained and preserves only the correlated portion of the data \mathbf{D} .

The decomposition can be used for the characterisation of the observations with respect to noise by examining the remaining $n_{comp} - k$ vectors. The choice of an appropriate k depends strongly on the assumed or estimated error level for the data, and should be determined consistent with the assigned noise level. A convenient measure of the quality of the approximation is the Root Mean Square Error (RMSE) calculated as

$$RMSE(k) = \sqrt{\frac{1}{N} \sum_{ij} (D_{ij} - D_{k,ij})^2}, \quad (\text{B.3})$$

where $N = n_{site}n_{comp}$ is the total number of elements in the observation matrix \mathbf{D} , and $i=1,\dots,n_{site}$, $j=1,\dots,n_{comp}$.

References

- Anderson, W.L., 1975. Improved digital filters for evaluating fourier and hankel transform integrals. Tech. Rep. In: U.S. Geological Survey, Denver, CO.
- Annett, D., Hauser, J., 2016. Summarising AEM data for mapping applications. In: ASEG Extended Abstracts 2016, 1, pp. 1–5. URL: <http://www.publish.csiro.au/paper/ASEG2016ab298>.
- Aster, R.C., Borchers, B., Thurber, C.H., 2013. Parameter Estimation and Inverse Problems. Elsevier, Amsterdam, Netherlands.
- Aukun, E., Boesen, T., Christiansen, A.V., 2017. A review of airborne electromagnetic methods with focus on geotechnical and hydrological applications from 2007 to 2017. In: Advances in Geophysics, vol. 58. Elsevier, pp. 47–93. <https://doi.org/10.1016/bs.agph.2017.10.002>. Ch. 2.
- Beamish, D., 2002. The canopy effect in frequency domain airborne em. Geophysics 67 (6), 1720–1728. <https://doi.org/10.1190/1.1527073>.
- Beamish, D., 2003. Airborne EM footprints. Geophys. Prospect. 51 (1), 49–60. <https://doi.org/10.1046/j.1365-2478.2003.00353.x>.
- Beamish, D., Levanieemi, H., 2010. The canopy effect in aem revisited: investigations using laser and radar altimetry. Near Surf. Geophys. 8 (2), 103–115.
- Bellman, R., 1957. Dynamic Programming. Princeton University Press, Princeton NJ, USA.
- Brodie, R., 2016. User manual for geoscience Australia's airborne electromagnetic inversion software. Tech. Rep. In: Geoscience Australia.
- Cole, K., Cole, R., 1941. Dispersion and absorption in dielectrics - I alternating current characteristics. J. Chem. Phys. 9, 341–352. <https://doi.org/10.1063/1.1750906>.
- Curtis, A., Lomax, A., 2001. Prior information, sampling distributions, and the curse of dimensionality. Geophysics 66 (2), 372–378. <https://doi.org/10.1190/1.1444928>.
- Donatelli, M., Reichel, L., 2014. Square smoothing regularization matrices with accurate boundary conditions. J. Comput. Appl. Math. 272, 334–349. <https://doi.org/10.1016/j.cam.2013.08.015>.
- Donatelli, M., Neuman, A., Reichel, L., 2012. Square regularization matrices for large linear discrete ill-posed problems. Numer. Linear Algebra Appl. 19 (6), 896–913. <https://doi.org/10.1002/nla.1833>.
- Farquharson, C.G., Oldenburg, D.W., 1998. Non-linear inversion using general measures of data misfit and model structure. Geophys. J. Int. 134 (1), 213–227. <https://doi.org/10.1046/j.1365-246x.1998.00555.x>.
- Farquharson, C.G., Oldenburg, D.W., 2004. A comparison of automatic techniques for estimating the regularization parameter in non-linear inverse problems. Geophys. J. Int. 156 <https://doi.org/10.1111/j.1365-246x.2004.02190.x>.
- Foks, N.L., Minsley, B.J., 2020. GeoBIPy - Geophysical Bayesian Inference in Python. Tech. Rep. In: United States Geological Survey, Denver, CO.
- Fraser, D.C., 2022. Resistivity Mapping with an Airborne Multicoil Electromagnetic System, 43, pp. 144–172.
- Goodman, J., Weare, G., 2010. Ensemble samplers with affine invariance. Commun. Appl. Math. Comput. Sci. 5, 65–80. <https://doi.org/10.2140/camcos.2010.5.65>.
- Green, P.J., 1995. Reversible jump MCMC computation and Bayesian model determination. Biometrika 82 (4), 711–732. <https://doi.org/10.1093/biomet/82.4.711>.
- Haario, H., Laine, M., Mira, A., Saksman, E., 2006. DRAM: efficient adaptive MCMC. Stat. Comput. 16, 339–354. <https://doi.org/10.1007/s11222-006-9438-0>.
- Hadamard, J., 1923. Lectures on the Cauchy Problem in Linear Partial Differential Equations. Yale University Press, New Haven CT.
- Hansen, P.C., 2010. Discrete inverse problems: Insight and algorithms. In: Vol. 7 of Fundamentals of Algorithms. Society for Industrial and Applied Mathematics (SIAM).
- Hansen, P.C., Nagy, J.G., O'Leary, D.P., 2006. Deblurring Images: Matrices, Spectra, and Filtering. SIAM.
- Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109. <https://doi.org/10.1093/biomet/57.1.97>.
- Heagy, L.J., Cockett, R., Kang, S., Rosenkjaer, G.K., Oldenburg, D.W., 2017. A framework for simulation and inversion in electromagnetics. Comput. Geosci. 107, 1–19. <https://doi.org/10.1016/j.cageo.2017.06.018>.
- Hegarty, A., Stanley, G., Kashdan, E., Hodgson, J., Parnell, A.C., 2022. Repeatability analysis of airborne electromagnetic surveys, 7 (1), 6. <https://doi.org/10.1186/s40929-016-0008-1>.
- Hodges, G., Daly, E., Lafrenière, L., O'Connell, Y., 2016. Environmental and geotechnical mapping with a fixed-wing frequency-domain electromagnetic system. In: Symposium on the Application of Geophysics to Engineering and Environmental Problems, pp. 17–20. <https://doi.org/10.4133/SAGEEP.29-004>.
- Hodgson, J.A., Ture, M.D., 2016. Tellus A1 airborne geophysical survey. Logistics, processing and merging report. Tech. Rep. In: Geological Survey Ireland, Dublin, Ireland.
- Huber, P., 1981. Robust Statistics. John Wiley, New York.
- Jupp, D.L.B., Vozoff, K., 1975. Stable iterative methods for the inversion of geophysical data. Geophys. J. R. Astron. Soc. 42, 957–976.
- Keller, G.V., Frischknecht, F.C., 1966. Electrical Methods in Geophysical Prospecting. Pergamon Press.
- Kiyan, D., Rath, V., 2017. Inverse Methods for Airborne Electromagnetic Data from the Tellus Surveys. The aempy Toolbox. Dublin Institute for Advanced Studies (DIAS).
- Lee, X., 2018. Fundamentals of Boundary-Layer Meteorology. Springer, Cham, Switzerland.
- Legault, J.M., 2015. Airborne electromagnetic systems - state of the art and future directions. CSEG Rec. 40, 38–49.
- Levanieemi, H., Beamish, D., Hautaniemi, H., Kurimo, M., Suppala, I., Vironmaki, J., Cuss, R., Lahti, M., Tartaras, E., 2009. The jac airborne em system : Aem-05. J. Appl. Geophys. 67 (3), 219–233.
- Loke, M.H., 2019. RES2DINVx64. Rapid 2-D resistivity & IP inversion using the least-squares method. In: Geotomo Software SDN BHD.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. J. Chem. Phys. 21 <https://doi.org/10.1063/1.1699114>.
- Minsley, B.J., 2011. A trans-dimensional Bayesian Markov chain Monte Carlo algorithm for model assessment using frequency-domain electromagnetic data. Geophys. J. Int. 187 (1), 252–272. <https://doi.org/10.1111/j.1365-246x.2011.05165.x>.
- Minsley, B.J., Smith, B.D., Hammack, R., Sams, J.I., Veloski, G., 2012. Calibration and filtering strategies for frequency domain electromagnetic data. J. Appl. Geophys. 80, 56–66.
- Mosegaard, K., Sambridge, M., 2002. Monte Carlo analysis of inverse problems. Invers. Prob. 18, 29–54.

- Muller, M.R., 2022. Tellus A5 block: *aempy* electromagnetic inversion report. In: Tech. Rep., Geological Survey Ireland, Dublin, Ireland.
- Parker, R.L., 1994. *Geophysical Inverse Theory*. Princeton University Press, Princeton, New Jersey.
- Raiche, A., Wilson, G., Sugeng, F., 2007. Practical 3d inversion - the p223f software suite. In: *19th Geophysical Conference and Exhibition*. ASEG, pp. 68–69.
- Rath, V., Mottaghy, D., 2007. Smooth inversion for ground surface temperature histories: estimating the optimum regularization parameter by generalized cross-validation. *Geophys. J. Int.* 171 (3), 1440–1448. <https://doi.org/10.1111/j.1365-246X.2007.03587.x>.
- Reninger, P.-A., Martelet, G., Deparis, J., Perrin, J., Chen, Y., 2011. Singular value decomposition as a denoising tool for airborne time domain electromagnetic data. *J. Appl. Geophys.* 75 (2), 264–276. <https://doi.org/10.1016/j.jappgeo.2011.06.034>.
- Revil, A., Coperey, A., Shao, Z., Florsch, N., Fabricius, I.L., Deng, Y., Delsman, J.R., Pauw, P.S., Karaoulis, M., de Louw, P.G.B., van Baaren, E.S., Dabekaußen, W., Menkovic, A., Gunnink, J.L., 2017. Complex conductivity of soils. *Water Resour. Res.* 53 (8), 7121–7147. <https://doi.org/10.1002/2017WR020655>.
- Rodgers, C.D., 2000. *Inverse Methods for Atmospheric Sounding. Theory and Practice*. World Scientific, Singapore.
- Römhild, L., Sonntag, M., Kiyani, D., Rogers, R., Rath, V., Börner, J.H., 2019. Anisotropic complex electrical conductivity of black shale and mudstone from the Moffat Shale Group (Ireland). *Near Surf. Geophys.* 17, 675–690. <https://doi.org/10.1002/ng.12073>.
- Sambridge, M., Mosegaard, K., 2002. Monte Carlo methods in geophysical inverse problems. *Rev. Geophys.* 40 (3) <https://doi.org/10.1029/2000RG000089>.
- Schwalenberg, K., Rath, V., Haak, V., 2002. Sensitivity studies applied to a two-dimensional resistivity model from the Central Andes. *Geophys. J. Int.* 150 <https://doi.org/10.1046/j.1365-246X.2002.01734.x>.
- SGL, 2016. Fixed-wing high-resolution aeromagnetic, gamma-ray spectrometric and frequency-domain electromagnetic survey. Tellus A1 block, Republic of Ireland, 2015 for Geological Survey of Ireland. Tech. Rep. TR816-2015-003. Sander Geophysics Limited.
- Slater, L., 2007. Near surface electrical characterization of hydraulic conductivity: from petrophysical properties to aquifer geometries: a review. *Surv. Geophys.* 28, 169–197. <https://doi.org/10.1007/s10712-007-9022-y>.
- Tarantola, A., 1987. *Inverse Problem Theory*. Elsevier, Amsterdam, the Netherlands.
- Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, Philadelphia PA, USA.
- Tikhonov, A.N., Arsenin, V.Y., 1977. *Solutions of Ill-Posed Problems*. V.H. Winston & Sons, Washington, DC.
- Tølbøll, R.J., Christensen, N.B., 2007. Sensitivity functions of frequency-domain magnetic dipole-dipole systems. *Geophysics* 72 (2), F45–F56. <https://doi.org/10.1190/1.2409623>.
- Vozoff, K., Jupp, D.L.B., 1975. Joint inversion of geophysical data. *Geophys. J. R. Astron. Soc.* 42, 977–991.
- Wahba, G., 1990. *Spline Models for Observational Data*, Vol. 59 of CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia.
- Wait, J.R., 1962. *Electromagnetic Waves in Stratified Media*. Pergamon, Boulder CO, USA.
- Xu, Q., 2005. Representations of inverse covariances by differential operators. *Adv. Atmos. Sci.* 22 (2), 181–198. <https://doi.org/10.1007/BF02918508>.
- Yaremchuk, M., D'Addazio, J., Pantaleev, G., Jacobs, G., 2018. On the approximation of the inverse error covariances of high resolution satellite altimetry data. *Q. J. R. Meteorol. Soc.* 144 (715), 1995–2000. <https://doi.org/10.1002/qj.3336>.
- Young, M. (Ed.), 2016. *Unearthed: Impacts of the Tellus Surveys of the North of Ireland*. Royal Irish Academy, Dublin, Ireland.