Electromagnetic resolution—a CSEM study based on the Wisting oil field

Vemund Stenbekk Thorkildsen [†] and Leiv-J Gelius

Department of Geosciences, University of Oslo, Sem Sælands vei 1, 0371 Oslo, Norway. E-mail: vemund.s.thorkildsen@gmail.com

SUMMARY
We consider marine controlled source electromagnetic (CSEM) data and demonstrate that a typical CSEM survey is associated with significant data redundancy. Thus, it should be possible to obtain a high-quality inversion result by using only a subset of the original data. Moreover, in survey design, effort should be made to optimize the placement of the receivers. This study therefore investigates the challenges of data decimation and survey design in the case of repeated surveys by use of the *data resolution matrix* and *model resolution matrix*. A framework of analysis has been developed to efficiently use these quantities. The feasibility of the proposed approach is demonstrated using both synthetic data computed from an interpreted model of the Wisting oil field in the Barents Sea, as well as actual field data from the same oil field.

Key words: Arctic region; Controlled source electromagnetics (CSEM); Marine electromagnetics; Inverse theory.

1 INTRODUCTION

The marine controlled source electromagnetic (CSEM) technique is used to map subsurface resistivity from surface measurements of magnetic and electric fields induced by a controlled source. For an exhaustive review of the development of marine CSEM, the reader is referred to (Constable 2010) and (Zhdanov 2010). Current studies process CSEM data by using inversion techniques that can handle complex and anisotropic earth models in 3-D (Brown *et al.* 2012; Jakobsen & Tveit 2018; Wang *et al.* 2018). Ideally, every inversion result should be accompanied by a proper description of the uncertainty and resolution of the inverted model. Menke (2012) describes how to quantify the resolution of an inversion using two resolution matrices: the *data resolution matrix* and the *model resolution matrix*. The data resolution matrix describes how well the data prediction matches the observed data, while the model resolution matrix describes how well each parameter in a discrete model is resolved. This study assesses both resolution matrices to investigate their applicability within CSEM inversion.

The literature includes several examples of the use of resolution matrices to analyse various inversion problems. For example, Friedel (2003) introduced a low-contrast inversion algorithm for electrical resistivity tomography data, which also provided an estimate of uncertainty, data resolution, and model resolution. Kalscheuer *et al.* (2010) used similar techniques to evaluate the resolution and

variance properties for single and joint inversions of magnetotelluric (MT) and direct current data. Other studies discuss simple applications of the model resolution matrix within CSEM inversion (Grayver *et al.* 2014; Mattsson 2015; McKay *et al.* 2015). However a rigorous analysis of marine CSEM inversion with an emphasis on data redundancy and the resolution matrices has not yet been published.

CSEM acquisition systems can largely be divided into towed streamer and nodal acquisition approaches. In the early 2010s, Petroleum Geo-Services (PGS) developed a towed streamer acquisition system, which was later abandoned (Engelmark *et al.* 2014). Another example of towed streamer CSEM is the Scripps Institution of Oceanography's Vulcan acquisition system, which is often combined with seabed nodes (Constable *et al.* 2016). However, a significant amount of data is still acquired using seabed nodes. Because deploying the nodal receivers comprises a significant portion of the costs associated with acquiring CSEM data, it is of interest to try to minimize the number of receivers used (especially in 3-D). This is especially important in the case of repeated surveys, which can be essential in a production setting or in a CO_2 storage project. In this study, we will investigate the feasibility of using resolution matrices for survey design purposes.

Romdhane & Eliasson (2018) quantified the importance of different datapoints by evaluating the approximate Hessian as part of the CSEM inversion. This information can then be used for survey design. By omitting the least important data (quantified by the Hessian), it is possible to remove a larger part of the data set while still obtaining a good inversion result. The approximate Hessian method allows for efficient subsampling of the data set, and thus decreasing

[†]Department of Geosciences, P.O Box 1047 Blindern, NO-0316 Oslo, Norway.

the computational demand. However, because the approximate Hessian is constructed from the Jacobian matrix, which only contains first-order derivatives with respect to model parameters, the important effects of regularization of the inverse problem are not properly considered (Menke 2012; Ren & Kalscheuer 2020). We therefore present an alternative approach representing a more systematic emphasis on survey design based on the full set of resolution matrices and derived quantities.

As a starting point, this study uses the open-source inversion package MARE2DEM (Modelling with Adaptively Refining Elements 2D for Electromagnetics). This package integrates a model builder (Mamba2D), forward modelling, inversion and data display capabilities for CSEM and other electromagnetic (EM) inversion problems (Key 2016). However, MARE2DEM provides the user with a very simple measure of sensitivity based only on the Jacobian matrix. To remedy this, we have developed an extensive toolbox for post-processing of the CSEM inversion result, which includes resolution matrices and derived quantities. We demonstrate how access to such measures provides additional useful insight into the quality of the inversion results. Thus, both resolution and robustness are addressed in the context of CSEM inversion using a synthetic earth model.

Many of the synthetic models used for CSEM studies are either 1-D (Key 2009; Roux & García 2014) or very simple (Weitemeyer *et al.* 2010). Several high-quality synthetic seismic models (e.g. Marmousi, Sigsbee2a) are publicly available, allowing for the testing of algorithms on known, but complex models. This is in stark contrast to CSEM, where such models are very difficult to find. Two notable exceptions are the SEG Advanced Modelling (SEAM) initiative (Stefani *et al.* 2010), which provides modelled CSEM data that can be licensed for a fee, and the Marlim R3D model (Carvalho & Menezes 2017; Correa & Menezes 2019), which is publicly available but limited to low frequencies with focus on deep target exploration. Moreover, no field data from the same area is publicly available as part of MarlimR3D. This motivated us to create a high-quality synthetic model based on the Wisting oil field, an offshore oil field in the Barents Sea where CSEM is proven to add significant value.

This paper is organized as follows. First, the Wisting field is introduced, along with a general description of the local geology. The next section presents CSEM field data acquired across Wisting. This is followed by a description of the model-building workflow. Next, the theoretical framework of forward modelling, inversion and construction of the resolution matrices is discussed. Taking this proposed analysis framework, we demonstrate its practical use on both synthetic and field data. Finally, a discussion and conclusion section ends the paper with an eye toward future applications of the model for more efficient CSEM data collection.

2 THE WISTING OIL FIELD

To study the sensitivity of EM methods for exploration purposes, it is essential to use high-quality synthetic models of the subsurface. As mentioned, such resistivity models are not easily available. Accordingly, such a model must be constructed from the ground up. In this study, we have chosen to build a synthetic model based on the Wisting oil field. The proposed model-building workflow is described in the subsequent section. However, it is first necessary to provide a general introduction to the Wisting oil field.

The Wisting oil field is located in the Hoop Fault Complex (*cf.* Fig. 1a), a northern region of the southwestern Barents Sea. With

an estimated 500 million barrels of oil equivalents, the field is a prime candidate for further development. However, in addition to the remote location of the oil field, several geological issues pose challenges for development. Senger *et al.* (2021) describe four major tectonic phases that have shaped the southwestern Barents Sea. The first phase was governed by the Palaeozoic Caledonian orogeny, followed by erosion of the Caledonian mountain chain. The second phase was defined by a Carboniferous to Permian extension, while the third phase was dominated by a sag basin formation. Erosion of the Uraldine mountain chain and the Northern Fennoscandian shield, along with continued subsidence, created a prograding shelf delta. This tectonic phase saw the deposition of thick sandstone-dominated formations, like the Middle Jurassic Stø formation, which serves as the main oil-bearing reservoir in the Wisting field. The fourth tectonic phase was dominated by uplift and erosion and can be seen as the most important process for the preservation of the current oil accumulations. The crest of the oil-bearing structure currently lies approximately 200 m below the seafloor after being uplifted roughly 1350 m from the late Cretaceous to early Palaeogene (Senger *et al.* 2021). Such a deep burial depth led to mechanical and chemical compaction, resulting in reduced porosity and permeability in the oil-bearing formations. Moreover, Fig. 1(b) highlights the faulted nature of the oil reservoir, caused by the major uplift. The average seafloor depth is about 400 m.

The oil-bearing section of the reservoir consists of three main formations. The aforementioned Stø formation is the primary hydrocarbon bearing unit, while the Nordmela and Fruholmen formations comprise the remaining parts of the reservoir (Granli *et al.* 2017). The main geological formations are outlined in the zoomed subsection of Fig. 1(b).

2.1 Electric properties of the Wisting field

The Wisting field is of particular interest due to its high resistivity values (Fig. 2), with the Stø formation regularly exceeding the maximum limit of the resistivity logger (100 000 Ωm) locally in the borehole. Such local extremities might be explained by an oil-wet reservoir, causing exceptional resistance to electric flow. However, it is believed that there are inaccuracies in the well log when encountering such high resistivity values. Supporting this claim, forward models with the original resistivity values lead to unrealistically high field strengths. As previously mentioned, the top reservoir lies only 200 m below the seabed, and the combination of shallow burial depth and high resistivity makes it an ideal candidate for constructing a high-quality resistivity model. To guide and calibrate the building of a synthetic model of the Wisting field, it is vitally important to access both resistivity logging data, high-quality CSEM, and seismic field data from the same area. The shallow burial depth means that a large frequency band (<12 Hz) is sensitive to the reservoir, which allows for easier tuning of the electric model by use of CSEM field data. Due to the low resolution of CSEM data compared to seismic data, it is crucial to use the latter to establish the reservoir geometry.

2.2 CSEM field data

In this study, we have access to CSEM field data from a 2-D line extracted from the BSMC08W 3-D survey, which was conducted in the summer of 2008 and belongs to the multiclient library of Electromagnetic Geoservices (EMGS). The corresponding selected
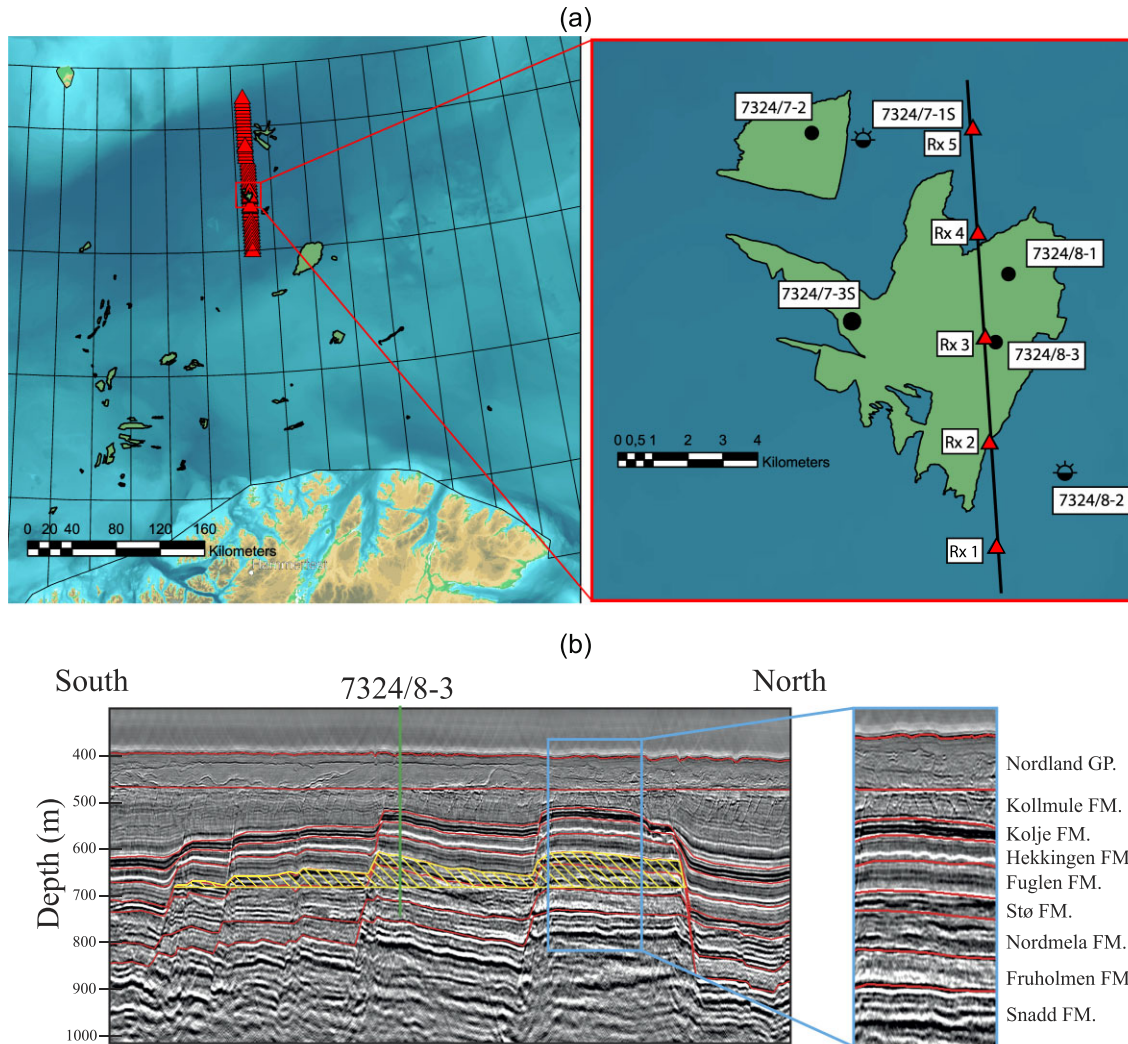
Figure 1. (a) Map of the southwestern Barents Sea along with a zoomed section of the Wisting oil field. The selected receivers of the extracted 2-D CSEM line are highlighted with red triangles, while the 2-D seismic depth line is shown in black. Nearby exploration wells are also highlighted. The data was acquired with a tow direction from south to north. Datapoints where the source is south of its corresponding receiver are therefore denoted in-tow, while the out-tow will have the source north of the corresponding receiver. (b) Seismic depth section showing the Wisting field along with the most important formations (annotated in the zoomed subsection). The approximate reservoir is highlighted in yellow.

receiver locations are highlighted in Fig. 1(a). Fig. 3 displays the source signature in the frequency domain, with the black bars representing the magnitude of the Fourier coefficients of the processed field data input to the inversion. The corresponding grey bars represent the theoretical discrete Fourier spectrum of the raw source signature. As expected, only minor differences exist. It is evident that most of the energy of this vintage data set is concentrated towards the lower frequencies (<4 Hz).

Processing of the raw data was carried out by EMGS, including the extraction of the 2-D CSEM line used in this study. Further processing was conducted by the authors to reduce the computational time of the inversion process. This included a resampling of the transmitter interval to 200 m. Fig. 4 shows the layout of the transmitter and receiver positions that were ultimately used. Note that, in case of receiver position 4, the transmitter interval is shifted 100 m.

Plots of the magnitude and phase of the inline horizontal electric field component at 2 Hz are shown in Fig. 5 (Receiver 2 in Fig. 1a). Fig. 5 also includes a normalized Magnitude Versus Offset (MVO)

plot computed for the same receiver, which shows that the maximum field strength of the out-tow direction is about three times larger than the corresponding in-tow direction. This receiver gather only shows the response from the use of a single frequency. However, by using the full available band of frequencies (*cf.* Fig. 3) and multiple receiver locations, it should be possible to fine-tune an earth model with a representative resistivity distribution of the true subsurface.

However, to achieve a geologically constrained subsurface model of the electric properties, the use of additional seismic data is needed. This is discussed in greater detail in the next section, which describes the main steps of our suggested approach to build a synthetic model.

3 MODEL BUILDING

Building a high-quality resistivity model is not a trivial task. In order to construct a realistic resistivity model, depth-migrated seismic field data is used as a structural constraint and the structural model is populated with resistivity values from well log data. The
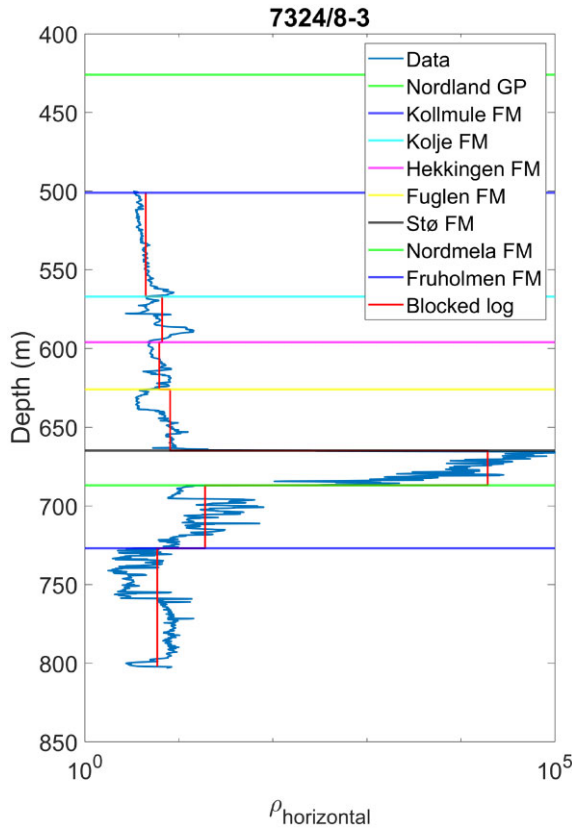
Figure 2. Deep resistivity log for well 7324/8-3 (see Fig. 1a) with labelled formation tops. Note also the discrete vertical red lines, which represent the upscaled (average) value in each formation and serve as a starting point for synthetic model building.
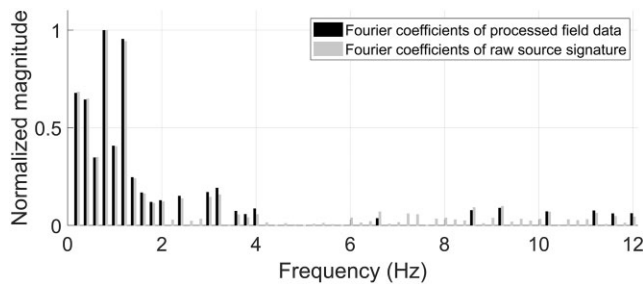

Figure 3. Source signature in the frequency domain. The black coefficients correspond to the processed data, while the grey bars show the raw source spectrum. Both sets of coefficients are normalized with the maximum amplitude of their respective data sets.
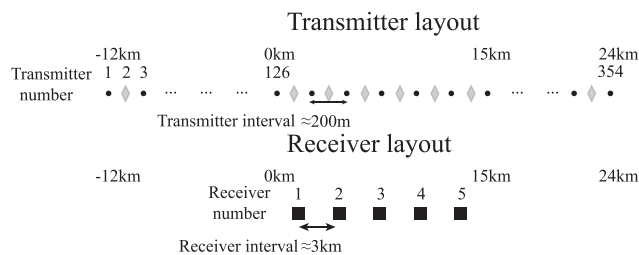

Figure 4. Survey layout after resampling to 200 m transmitter interval. The receiver interval is about 3 km, while the transmitter interval is around 200 m. Note that the transmitter positions of Receiver 4 (represented by grey diamonds in the transmitter layout) is shifted by 100 m compared to that of the other receivers.

freely available Marlim R3D uses a similar strategy for structural constraint, but takes a more sophisticated approach to populating the resistivity model (Carvalho & Menezes 2017). Building a 2-D resistivity model entails two primary problems: (i) extracting a 2-D line from a 3-D CSEM survey and (ii) identifying a 2-D seismic line in close proximity (Fig. 1a). The field data used in this paper represents the measured response of a 3-D structure. When interpreting the main (or target) horizons of the model, it is therefore essential to verify that only minor geological variations exist in the crossline direction of the selected 2-D line.

In addition, it must be recognized that the CSEM method is not sensitive to small-scale heterogeneities. Therefore, smaller faults should not invalidate the assumption of the 2-D experiment. Depth slices of a 3-D seismic data cube indicate that the selected data line used in this work should not be compromised by major structural faults (Granli *et al.* 2017), although some distortions may arise due to more rapid resistivity variations caused by changes in oil saturation. If such distortions are prominent, they would be expected to manifest in the southern part of the line (Fig. 1a). After selecting the most optimal combination of seismic and EM data, structural interpretation can be carried out. For this purpose, Petrel was used (Schlumberger 2018). After interpretation, the horizons are heavily downsampled before being imported to the model builder. For this purpose, we used the model builder Mamba2D, which is part of the MARE2DEM package. Mamba2D creates a mesh node at every point in the imported horizons. For computational efficiency, it is therefore ideal to begin with a sparse starting mesh that MARE2DEM can adaptively refine during the forward modelling process (Key 2016).

For this study, we have chosen to populate the model with resistivity values based on simple averaging of the deep resistivity logging measurements within blocks bounded by geologic formations (Fig. 2). This approach carries some limitations; well logs only provide a very local measure of the resistivity. Thus, the resistivity might vary greatly only a few metres away from the well path. In comparison, CSEM is a low-frequency technique and will only be sensitive to large volumes in our model. Moreover, vertical well logs only provide a measure of horizontal resistivity. By only using the horizontal resistivity, we are assuming the resistivity is isotropic. In the case of CSEM, this assumption can lead to poor inversion results. This can intuitively be understood by considering a layered earth. Assuming that the earth is isotropic leads to the conclusion that the electric current flows equally well in all directions. In other words, the current would be assumed to flow across lithological boundaries just as it flows along a uniform layer. Earlier studies have shown that a moderate ratio of vertical to horizontal resistivity of 2:3 may have a significant effect on the inversion result (Lu & Xia 2007; Newman *et al.* 2010; Brown *et al.* 2012).

Moreover, because CSEM data are generally more sensitive to vertical resistivity, using an isotropic inversion scheme to an anisotropic earth biases the inversion towards higher resistivity values (Hoversten *et al.* 2006). We did not have access to vertical resistivity measurements for this project. However, the operator of the Wisting oil field (Equinor) provided representative values of anisotropy. Fig. 6 shows the synthetic model colour-coded with the vertical component of the resistivity. The final vertical resistivity values in all formations are listed in Table 1, along with the ratio of vertical to horizontal resistivity $(\frac{\rho_z}{\rho_{xy}})$ and a brief lithology description. All anisotropy factors fall within the typical range found in the literature. To account for inaccuracies in the well logging tool when encountering such extreme resistivity values as seen in the Stø formation, we had to tune the model in order to achieve
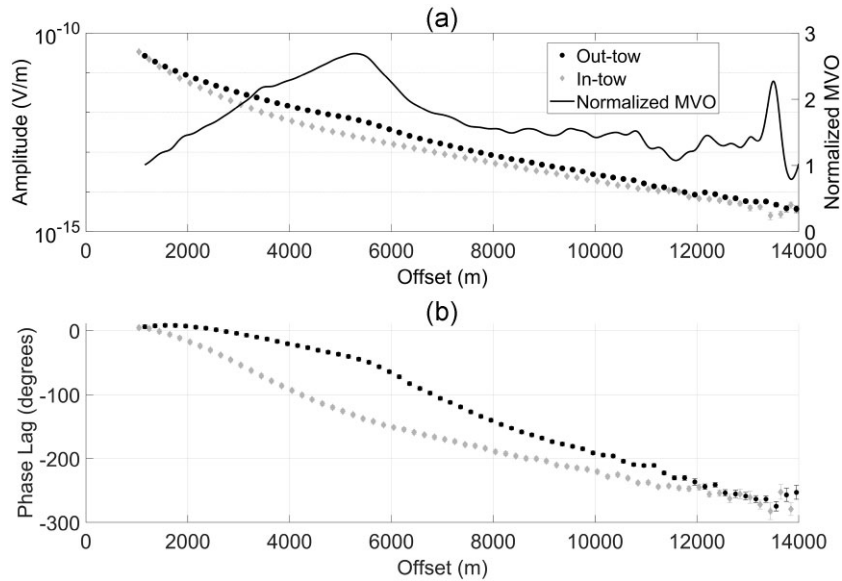
Figure 5. Field data example of the inline horizontal electric field component for Receiver 2. (a) MVO for the in-tow direction (represented by grey diamonds) and the out-tow direction (represented by black dots) along with normalized MVO, which is calculated as the ratio of the data of interest (i.e. out-tow) and an assumed background response (i.e. in-tow). (b) Phase lag for the same receiver.
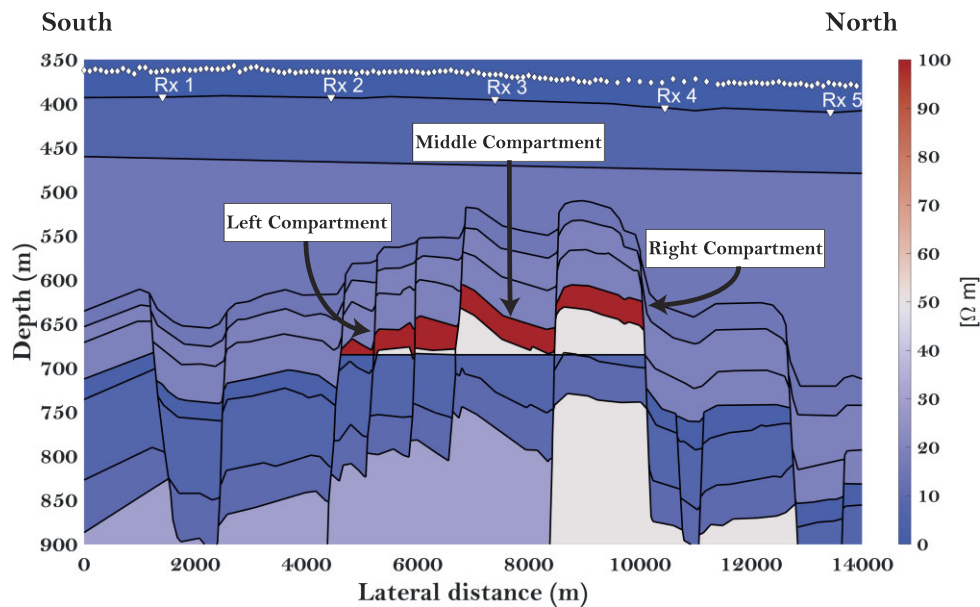

Figure 6. Synthetic model after tuning (colour coded with vertical resistivity). Note that the colour bar only shows values up to 100 Ωm to emphasize the resistivity variations in the different geological formations (see Table 1 for resistivity values). The reservoir can be divided into three compartments, defined by three main fault blocks. The left compartment is further subdivided into three fault blocks. However, when addressed in the text we refer to all three fault blocks.

a satisfactory data fit (Fig. 7). In general, the resistivity values in the inverted models from both the field data and synthetic data fit well. It should be noted that the tuning was done with the aim of creating a reasonable resistivity model for testing our method, and such tuning should not be done with interpretation in mind.

4 ELECTROMAGNETIC SENSITIVITY—INVERSION THEORY AND THE RESOLUTION MATRIX

In CSEM inversion, the largest computational bottleneck is the repeated forward modelling of the EM fields, due to the many source points and frequencies. However, methods to decrease the computational load do exist. Electromagnetic reciprocity can be exploited to decrease the computational burden in the case of nodal marine CSEM, where source points typically greatly outnumber receivers. Electromagnetic reciprocity states that, for a linear medium, the position and orientation of a receiver and transmitter can be swapped, and still measure the same response. The medium is defined as linear if the magnetic permeability μ, dielectric permittivity ϵ, and electric conductivity σ do not depend on the magnetic $\mathbf{H}$ or electric $\mathbf{E}$ field intensity. In practice, this means that we swap the position and orientation of the receivers and source points (Parasnis 1988). In the case of a towed streamer CSEM survey, reciprocity will not

Table 1. Vertical and horizontal resistivity values chosen for the final model. Note that Stø and Nordmela are listed twice, as these two formations form the oil reservoir. The resistivity of the oil-filled Stø formation is also listed with a range, since an optimal data fit could only be achieved if different resistivity values were assigned for each compartment of the reservoir. The lithology description is taken from Senger *et al.* (2021).

Formation	ρ_z [Ohm-m]	$\frac{\rho_z}{\rho_{xy}}$	Lithology
Nordland Gp	7	2.3	Marine shale
Kollmule Fm	15	3.4	Marine shale
Kolje Fm	15	2.7	Marine shale
Hekkingen Fm	19.5	3.2	Marine/organic rich shale
Fuglen Fm	19.5	2.4	Marine shale
Stø Fm (oil-filled)	1500–2500	1	Sandstone
Stø Fm (brine-filled)	3	2	Sandstone
Nordmela Fm (oil-filled)	50	1	Marine shale/sandstone
Nordmela Fm (brine-filled)	7	2	Marine shale/sandstone
Fruholmen Fm	10	2	Alluvial shale/sandstone
Snadd Fm	30	2	Marine shale

yield a significant decrease in computational time, as the number of source and receiver points are approximately equal.

However, even after exploiting electromagnetic reciprocity, the forward problem remains computationally demanding. This emphasizes the question of how much of the data we really need to use, and especially how many frequencies are needed. We propose the combined use of the model resolution matrix and data resolution matrix to quantify the importance of each data parameter. By examining the resolution matrices, we can remove datapoints which do not contribute significantly to our target area.

4.1 MARE2DEM

MARE2DEM is an open-source forward modelling and inversion software developed by the Scripps Seafloor Electromagnetic Consortium. The package is described in detail by Key (2016), so we will only give a brief introduction to the main concepts for completeness.

All electromagnetic induction methods aim to map the subsurface in terms of resistivity (or conductivity) by using either a natural or an artificial (i.e. active) source. Based on the measured electric field responses of the subsurface, a quantitative image of the earth model can be recovered by inversion. In CSEM inversion, we need to solve for the electric fields using Maxwell's equations on a discrete grid. This can effectively be done by using finite difference or finite element solvers. When such discrete solvers are used, the accuracy of the forward modelling critically depends on the modelling mesh. By introducing a fine grid, the forward modelling will likely be very accurate, but the cost of this accuracy is high computational demand. MARE2DEM uses an adaptively refining finite element forward modelling scheme. This means that if the base is a sparse model with few mesh nodes, the program will refine the different grid cells based on a stability criterion (Key 2016).

When field data are acquired, the earth model extends infinitely, and the measurements approach noise only when the separation between the source and receivers increases (e.g. below the detection threshold of the acquisition equipment). However, when modelling the data, it is necessary to limit the size of the model to decrease the computational burden. This can be implemented by using an absorbing boundary condition such as a perfectly matched layer (Li *et al.* 2018). However, because MARE2DEM does not have these

absorbing boundary conditions implemented, it is necessary to use a large model to avoid edge effects.

4.2 Inversion

In an iterative inversion process, the forward modelling response is calculated in the current model, and a misfit or cost function representing the error between the calculated response and the actual response is constructed. By minimizing this misfit, the model is updated in an iterative manner. Following Ren & Kalscheuer (2020), this cost function can formally be written as:

$$U[\mathbf{m}, \alpha] = Q_d[\mathbf{m}] + \alpha Q_m[\mathbf{m}], \tag{1}$$

where $\mathbf{m}$ is the model vector, $Q_d[\mathbf{m}]$ is the model dependent data misfit and $Q_m[\mathbf{m}]$ is the regularization term that simplifies the solution space. The Lagrangian multiplier α acts as a weight factor between the data misfit term and the regularization term, thereby balancing resolution and stability. MARE2DEM uses a variation of Occam inversion denoted 'fast Occam' (Key 2016). This is an implementation of the Gauss–Newton optimization scheme, in which the model update is done by building the Jacobian matrix of sensitivities. Calculating the Jacobian matrix involves simulating the response of all the sources (forward fields) and receivers (adjoint fields). Thus, reciprocity would not yield any decrease in computation time. However, each Occam iteration includes a grid search for the Lagrangian multiplier by calculating the model update and forward response for a range of α-values. The forward modelling in this grid search does not involve calculating the Jacobian matrix, and will therefore be faster by using reciprocity in the case of marine nodal acquisition. For more details regarding Occam inversion and its specific implementation in MARE2DEM, the reader is referred to, respectively, Constable *et al.* (1987) and Key (2016). The MARE2DEM package optionally includes a reference model. However, because this is not included in our analysis, the inversion relies solely on a roughness penalty for regularization. Eq. (1) can be expanded as follows:

$$U[\mathbf{m}, \alpha] = \left[(\mathbf{d} - \mathbf{F}[\mathbf{m}])^\dagger \mathbf{W}_\mathbf{d}^\dagger \mathbf{W}_\mathbf{d} (\mathbf{d} - \mathbf{F}[\mathbf{m}]) \right] + \alpha \mathbf{m}^\dagger \mathbf{W}_\mathbf{m}^\dagger \mathbf{W}_\mathbf{m} \mathbf{m}, \tag{2}$$

where $\mathbf{d}$ denotes the measured complex field data and $\mathbf{F}[\mathbf{m}]$ denotes the model response. Working with complex fields requires us to use the Hermitian † (i.e. matrix transpose + complex conjugation) for the matrices involved. The data misfit is also weighted by $\mathbf{W_d}$, a diagonal matrix consisting of the inverse of the standard error for each sample. The regularization term includes the weighting matrix $\mathbf{W_m}$ to enforce model smoothness. In MARE2DEM, this is obtained using a gradient roughness operator. For anisotropic models, the roughness is augmented by splitting the model vector into anisotropic subsets (Key 2016).

Since our problem is non-linear, the forward operator is linearized in the vicinity of the current model $\mathbf{m}_k$ by use of a Taylor series expansion:

$$\mathbf{F}[\mathbf{m}_{k+1}] \approx \mathbf{F}[\mathbf{m}_k] + \mathbf{J}(\mathbf{m}_{k+1} - \mathbf{m}_k). \tag{3}$$

The Jacobian or sensitivity matrix $\mathbf{J}$ (with entries $\frac{\partial F_i(\mathbf{m}_k)}{\partial \log(\rho_j)}$, where ρ_j is the resistivity in cell j), includes the first-order partial derivatives with respect to model parameters (log resistivity). Note also that MARE2DEM outputs the data and Jacobian matrix in the same format as the input. Wheelock *et al.* (2015) found that, in electromagnetic inversion, the most computationally efficient and robust approach is to use phase lag and logarithmically scaled amplitude as input. Thus, we adapted the same input format in this study. However, because we are working with complex fields, it is necessary
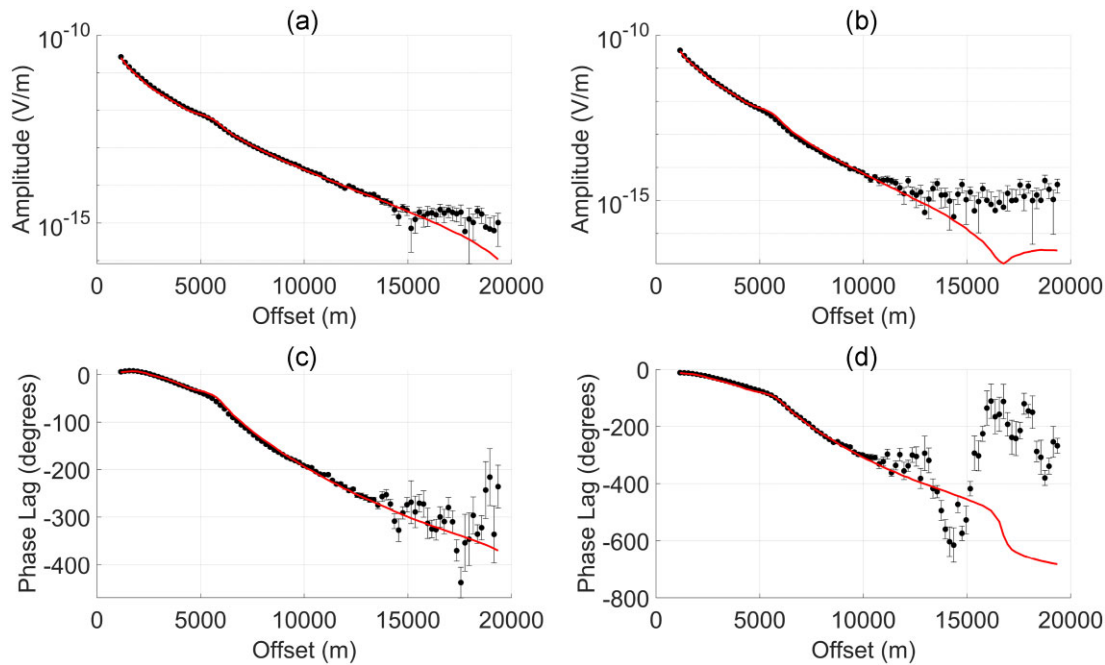
Figure 7. Field data (dots) and synthetic response (line) for Receiver 2. On the left-hand side, MVO (a) and PVO (c) are shown for a 2 Hz response, while on the right-hand side the MVO (b) and PVO (d) are shown for 4 Hz. For both frequencies, the synthetic data fit well with the field data (until the noise floor is reached at an amplitude of around 10^{-15} V m^{-1}). The vertical bars associated with the field data represent data uncertainty.

to transform the Jacobian matrix into its complex field equivalent. For more details regarding the calculation of the Jacobian and the transformation to its complex field equivalent, the reader is referred to Appendix A. Combining eqs (2) and (3) yields the following:

$$U^{lin}[\mathbf{m}_{k+1}, \alpha]$$
$$= \left[\left(\mathbf{d} - \mathbf{F}[\mathbf{m}_k] - \mathbf{J}(\mathbf{m}_{k+1} - \mathbf{m}_k)\right)^\dagger \mathbf{W}_\mathbf{d}^\dagger \mathbf{W}_\mathbf{d}\left(\mathbf{d} - \mathbf{F}[\mathbf{m}_k] - \mathbf{J}(\mathbf{m}_{k+1} - \mathbf{m}_k)\right)\right]$$
$$+ \alpha\mathbf{m}_{k+1}^\dagger \mathbf{W}_\mathbf{m}^\dagger \mathbf{W}_\mathbf{m}\mathbf{m}_{k+1}. \tag{4}$$

A least squares solution is obtained by setting $\frac{\partial U^{lin}[\mathbf{m}_{k+1}, \alpha]}{\partial \mathbf{m}_{k+1}}$ to zero and solving for $\mathbf{m}_{k+1}$:

$$\mathbf{m}_{k+1} = \mathbf{J}_w^{-g}\mathbf{W}_\mathbf{d}\mathbf{d}_k, \tag{5}$$

where $\mathbf{d}_k = [\mathbf{d} - \mathbf{F}[\mathbf{m}_k] + \mathbf{J}\mathbf{m}_k]$ and $\mathbf{J}_w^{-g}$ is the generalized inverse $[\mathbf{J}^\dagger\mathbf{W}_\mathbf{d}^\dagger\mathbf{W}_\mathbf{d}\mathbf{J} + \alpha\mathbf{W}_\mathbf{m}^\dagger\mathbf{W}_\mathbf{m}]^{-1}\mathbf{J}^\dagger\mathbf{W}_\mathbf{d}^\dagger$. It is possible to include a total of six different data components relating to the three different directions of the magnetic and electric field in the inversion. However, this study only makes use of the inline horizontal electric field (E_y).

4.3 Model resolution matrix

Here, let $\mathbf{m}_k$ denote the inversion obtained from the final (*k*th) iteration of an inversion cycle. By assuming that model $\mathbf{m}_k$ is linearly close to the true earth model $\mathbf{m}_{\text{true}}$, we can write the following equation:

$$\mathbf{d} = \mathbf{F}[\mathbf{m}_{\text{true}}] + \mathbf{n} \approx \mathbf{F}[\mathbf{m}_k] + \mathbf{J}(\mathbf{m}_{\text{true}} - \mathbf{m}_k) + \mathbf{n}, \tag{6}$$

where $\mathbf{n}$ denotes noise. Consequently, the data prediction for iteration k can be approximated as follows:

$$\mathbf{d}_k = (\mathbf{d} - \mathbf{F}[\mathbf{m}_k] + \mathbf{J}\mathbf{m}_k) \approx \mathbf{J}\mathbf{m}_{\text{true}} + \mathbf{n}. \tag{7}$$

Combining eqs (5) and (7) yields:

$$\mathbf{m}_{k+1} = \mathbf{R}_\mathbf{M}\mathbf{m}_{\text{true}} + \mathbf{J}_w^{-g}\mathbf{W}_d\mathbf{n}. \tag{8}$$

In eq. (8), $\mathbf{R}_\mathbf{M}$ is the *model resolution matrix* (Menke 2012), and is explicitly given as:

$$\mathbf{R}_\mathbf{M} = \Re\left[\left[\mathbf{J}^\dagger\mathbf{W}_\mathbf{d}^\dagger\mathbf{W}_\mathbf{d}\mathbf{J} + \alpha\mathbf{W}_\mathbf{m}^\dagger\mathbf{W}_\mathbf{m}\right]^{-1}\mathbf{J}^\dagger\mathbf{W}_\mathbf{d}^\dagger\mathbf{W}_\mathbf{d}\mathbf{J}\right], \tag{9}$$

where $\Re$ implies taking the real part. If the inversion terminates at iteration k, $\mathbf{m}_{k+1}$ is considered the preferred inversion model. The model resolution matrix indicates how close the preferred inversion model is to the true model. Because the model resolution matrix depends on the Lagrangian multiplier α, letting $\alpha \to 0$ allows the model resolution matrix to approach the identity matrix. In such a case, $\mathbf{m}_{k+1}$ is said to be perfectly resolved, and the preferred inversion model only has contributions from $\mathbf{m}_{\text{true}}$ and the noise term (Ren & Kalscheuer 2020). In a real-world case, $\mathbf{m}_{\text{true}}$ is unobtainable, so it is substituted with $\mathbf{m}_k$.

The model resolution matrix can be seen as a blurring filter that describes how the unobtainable true model is reproduced by the inversion. Figs 8(a) and (b) depict a schematic representation of eq. (8) without the error term. Note that these figures show the relationship for a 1-D model. In case of a 2-D model, the model resolution matrix takes the form of a block Toeplitz matrix with Toeplitz blocks (BTTB) if the 2-D image is represented as a vector (lexicographic ordering) (Hansen *et al.* 2006). Parts (a) and (b) of Fig. 8 illustrate, respectively, the column information with an emphasis on the point spread function (PSF) and the row information with an emphasis on the Smoothing Kernel. The PSF is well known from imaging theory (Rossmann 1969) and describes how an imaging system responds to an impulse. Assigning a delta function in $\mathbf{m}_{\text{true}}$, the PSF describes how this delta function spreads across the inverted model $\mathbf{m}_{k+1}$ (Fig. 8a). The Smoothing Kernel describes the extent to which each parameter in the true model contributes to a single model parameter in the inverted model. Although it is well defined, the Smoothing Kernel is more challenging to interpret than
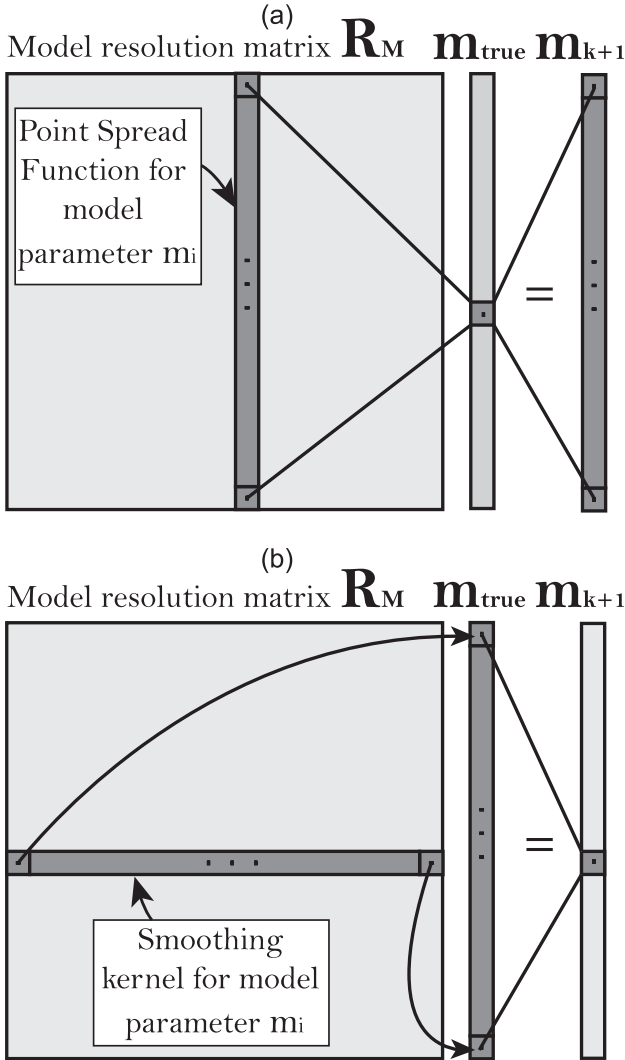
(a)
Model resolution matrix $\mathbf{R_M}$ $\mathbf{m_{true}}$ $\mathbf{m_{k+1}}$

(b)
Model resolution matrix $\mathbf{R_M}$ $\mathbf{m_{true}}$ $\mathbf{m_{k+1}}$

Figure 8. Computational relationship between the resolution matrix ($\mathbf{R_M}$), the unobtainable true model $\mathbf{m_{true}}$ and the preferred inversion model $\mathbf{m_{k+1}}$. The relationship is shown with an emphasis on the PSF (a) and the Smoothing Kernel (b).

the PSF. We have therefore only used the PSFs to quantify resolution power.

In an ideal case, where the model is perfectly resolved, the associated PSFs and Smoothing Kernels are delta functions ($\mathbf{R_M = I}$). In most cases, such a model is impossible to obtain; the PSF will vary across the model space. However, in a general inversion, it is likely that some areas will be well resolved, and others more poorly resolved. The PSF in well-resolved areas will be characterized by a small spread centred on the associated model parameter. PSFs in poorly resolved areas can be characterized by a large spread, an off-centred maximum, or a combination of the two.

Fig. 9 shows the resolution matrix of an inversion of synthetic data calculated from the model in Fig. 6. The resolution matrix is difficult to evaluate in its full form. However, it can be reorganized to form 2-D slices representing either PSFs or Smoothing Kernels. Because we are free to choose which datapoints should contribute when constructing $\mathbf{R_M}$, it is therefore possible to select different subsets of input data and evaluate the corresponding changes in the PSFs. However, it is extremely labour-intensive to evaluate all the

PSFs, so it is therefore advisable to construct a metric summarizing the quality of the selected data. Friedel (2003) tackles this challenge by introducing the *radius of resolution*:

$$r_{\mathrm{res},i} = \frac{r_0}{\sqrt{R_{M,ii}}}, \tag{10}$$

where r_0 represents an inscribed circle for a given model element i and $R_{M,ii}$ represents the corresponding value of the diagonal element of the resolution matrix for the same model parameter. This measure breaks down if the highest value is off-diagonal. In such cases, Friedel (2003) introduces a simple distortion flag to mark those cells where the highest value is not centred on the diagonal. As an alternative, we suggest a combined measure that takes into account both the width of the PSF and distortions when the PSF is off-diagonal. We denote this metric the *ratio of resolution*. It is constructed by dividing the diagonal element of $\mathbf{R_M}$ with the sum of all elements falling inside a user-defined ellipsoid. Let V_i denote the ellipsoid centred around model parameter i. The ratio of resolution for model parameter i is then defined as:

$$\mathrm{ratio}_{\mathrm{res},i} = \frac{R_{M,ii}}{\sum_{j=1}^{M} |R_{M,ij}| \in V_i}. \tag{11}$$

The size of V_i is found by trial and error. This study used an ellipsoid defined by a 150-m vertical minor axis and a 1000-m lateral major axis. The model resolution matrix can be constructed irrespective of the data input. Thus, it is possible to combine different subsets of data and quantify their resolving power by evaluating either the associated PSFs or the ratio of resolution.

4.4 Data resolution matrix

It is possible to construct the ratio of resolution for all combinations of input data. However, this study takes a different approach based on the data resolution matrix, which is constructed as follows. The predicted data for iteration $k + 1$ can be written as:

$$\mathbf{d}_{k+1} = \mathbf{F}[\mathbf{m}_{k+1}], \tag{12}$$

which can be combined with eq. (3) to give

$$\mathbf{d}_{k+1} \approx \mathbf{F}[\mathbf{m}_k] + \mathbf{J}(\mathbf{m}_{k+1} - \mathbf{m}_k). \tag{13}$$

A further combination of eqs (5) and (13) with the definition $\mathbf{d}_k = (\mathbf{d} - \mathbf{F}[\mathbf{m}_k] + \mathbf{Jm}_k)$ leads to:

$$\mathbf{d}_{k+1} \approx \mathbf{R_D}\mathbf{d} + (\mathbf{I} - \mathbf{R_D})(\mathbf{F}[\mathbf{m}_k] - \mathbf{Jm}_k), \tag{14}$$

where $\mathbf{R_D}$ is denoted the data resolution matrix and is given explicitly as:

$$\mathbf{R_D} = \Re \left[\mathbf{J} \left[\mathbf{J}^\dagger \mathbf{W}_d^\dagger \mathbf{W}_d \mathbf{J} + \alpha \mathbf{W}_m^\dagger \mathbf{W}_m \right]^{-1} \mathbf{J}^\dagger \mathbf{W}_d^\dagger \mathbf{W}_d \right]. \tag{15}$$

The diagonal of $\mathbf{R_D}$ is often called *Data Importances* (Maurer *et al.* 2000; Ren & Kalscheuer 2020), and describes how important a data point is in its own prediction.

4.5 Data reduction and planning of repeated surveys

As mentioned, this study proposes using a combination of the data resolution matrix and model resolution matrix to reduce the number of data points used in the inversion. To assess which data points to use, we propose a method in which the Data Importances guide the data selection. We then assess the resulting ratio of resolution to
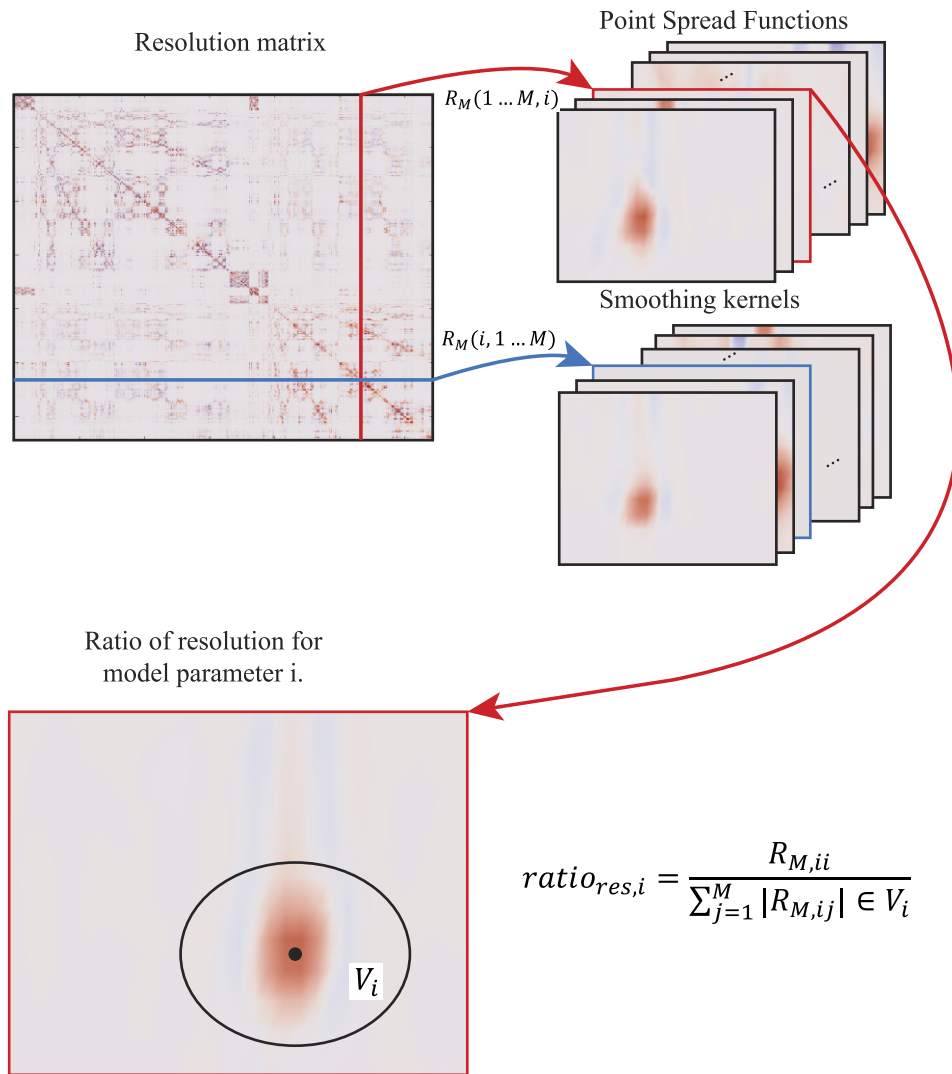
Figure 9. The full model resolution matrix is of size $M \times M$, where M is the length of the model vector. The columns in the resolution matrix define the PSFs, while the rows represent the Smoothing Kernels. The resolution matrix is difficult to evaluate in its original form. However, by reorganizing the rows or columns to fit the model space, it is possible to construct meaningful information in form of Smoothing Kernels or PSFs. To summarize the quality of the set of PSFs chosen, we propose the metric ratio of resolution. In case of model parameter i, it is calculated by dividing the corresponding diagonal element of $\mathbf{R_M}$ ($R_{M,ii}$) with the sum of the absolute value of all elements that fall inside a user-defined ellipsoid V_i.

determine whether the target area has lost any important resolving power. The workflow of our proposed method is summarized in Fig. 10.

In the proposed method for data reduction, it is also helpful to evaluate whether Data Importances can truly be used like the name suggests. Therefore, the first step involves computing the data resolution matrix and extracting the Data Importances. This study proposes calculating this quantity on a frequency-per-frequency and receiver-per-receiver basis. The Data Importances are then subsampled by first selecting the percentile value of their full range and then discarding all the values falling below this threshold. The selected data can then be used to calculate the model resolution matrix for the subsampled data set, along with its associated ratio of resolution. A direct comparison between the ratio of resolution map of the complete and decimated data sets should then reveal whether any essential resolving power has been lost. For a given model parameter m_i, a loss in resolving power will manifest as a decrease in $ratio_{res, i}$, with the note that this is an a priori indicator of the resolving power

of the subsampled data. This reduced data set can then be used as input for a new inversion.

Another method of using Data Importances relates to repeated surveys. By plotting the Data Importances for each receiver, it is possible to evaluate whether some are more important than others. If one or more receivers are characterized by a very low importance, it might be possible to remove such receivers in a repeated survey.

5 SYNTHETIC DATA

This section presents the results obtained from testing the workflow in Fig. 10, using the synthetic Wisting data set with the acquisition layout described in Section 2.2. Trials with different initial models did not significantly alter the final inversion result. Thus, we use a simple gradient model in all the inversions presented here. However, for computational efficiency, the inversion is constrained to a region around the known target area. This region reaches down to $\approx$1600 m below seabed and from -5000 to $20\,000$ m inline distance (*cf.*
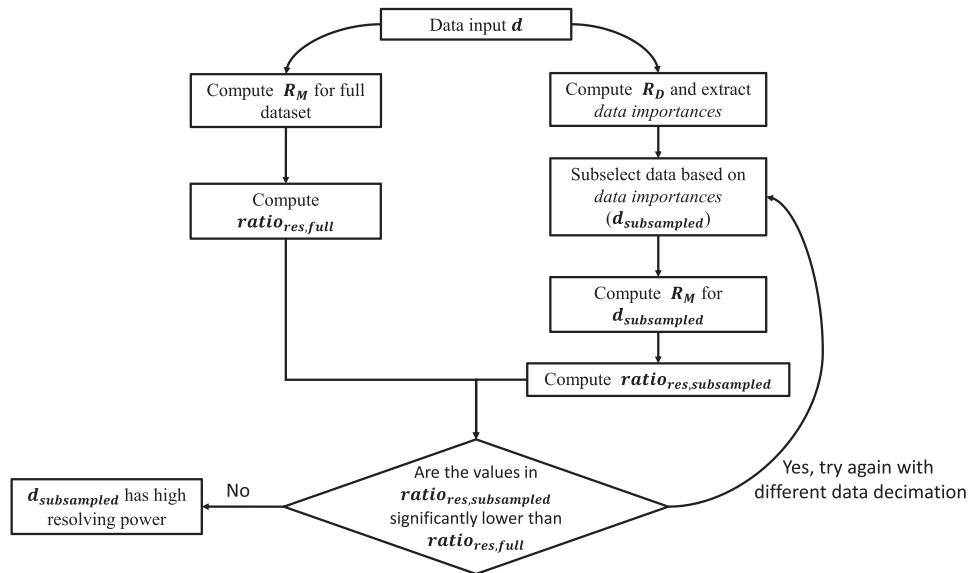
Figure 10. Workflow for data reduction.

Fig. 6). Outside this region, the model parameters are kept fixed and equal to those of the initial model. First, the results obtained from full inversion are presented in the form of PSFs and the calculated Data Importance. Next, the data are subsampled in three different ways and analysed using the proposed workflow. Finally, as a validation, a direct comparison is made between the different inversions.

5.1 Case 1—Full data set

The first experiment uses the full data set up to 12 Hz as data input. The plots shown in Fig. 11 can now be formed following the post-processing steps described in the previous section. Fig. 11(a) shows the PSF of a well-resolved model parameter and demonstrates that the PSF is well centred, exhibiting only one main lobe. Fig. 11(b) shows how the PSF for a poorly resolved model parameter contrasts with the well-resolved model parameter case. The PSF is now off-centre, smeared over a large area with several sidelobes. A direct analysis of the PSFs reveals useful information about resolving power. However, a more efficient computational approach is to use the previously introduced ratio of resolution. An example of this metric is shown in Fig. 12(a). In general, the highest values (associated with good resolution) are found inside the reservoir, while the zones above and below the reservoir are defined by lower values (with the exception of some boundary effects). The final inverted resistivity model is characterized by three main compartments, as shown in Fig. 12(b). Direct comparison with the true (i.e. synthetic) model shown in Fig. 6 demonstrates that the inversion has captured the main features, especially with regard to the lateral extension. However, the image is characteristically smeared over a larger vertical area due to the general lack of resolution of the CSEM method.

5.2 Case 2—Removing the least influential receiver

Figs 11(c) and (d) show the Data Importance for Receivers 2 and 4, respectively. These two receivers are laterally placed on opposing edges of the reservoir (*cf*. Figs 1a and 6). There are clear differences between the two plots, wherein Receiver 2 has higher Data Importance throughout. It should be noted that the Data Importance values of the remaining receivers exhibit the same character as Receiver 2, while Receiver 4 stands out with much lower values. As shown in Fig. 4, Receiver 4 is associated with a shifted line of transmitters compared to the other receivers. This may have played a role.

Based on the Data Importance panels, Receiver 4 carries less important information in the inversion. It should therefore be possible to remove this receiver without losing significant resolving power. A direct comparison of the ratio of resolution for Cases 1 and 2 (Figs 12a and b) reveals that removing Receiver 4 does not significantly change the resolving power. However, an exhaustive validation can only be found by comparing the actual inversion results. Figs 12(b) and (d) show the inverted (vertical resistivity) model for these two cases and confirms that our proposed approach to subselect the data is feasible. It should be noted that this observation is particularly valuable for design of repeated surveys, especially in a 3-D setting.

5.3 Case 3—30 per cent cut-off including lowest frequencies

The promising results of Case 2 imply that further data decimation is possible. However, because none of the remaining receivers are characterized by a low overall Data Importance, the data decimation should now be performed in a different way. As previously suggested, such a decimation can be achieved by calculating a user-provided percentile value of the full data set, then discarding values that fall below this threshold. However, this data decimation would remove all data samples of the lowest frequency (0.2 Hz) from the data set. The absence of this low frequency information in initial trials led to a vertical displacement of the reservoir in the inverted model. In Case 3, we therefore propose to discard Receiver 4 and all data points with a Data Importance falling below the 70th percentile. However, all data points of the two lowest frequencies (0.2
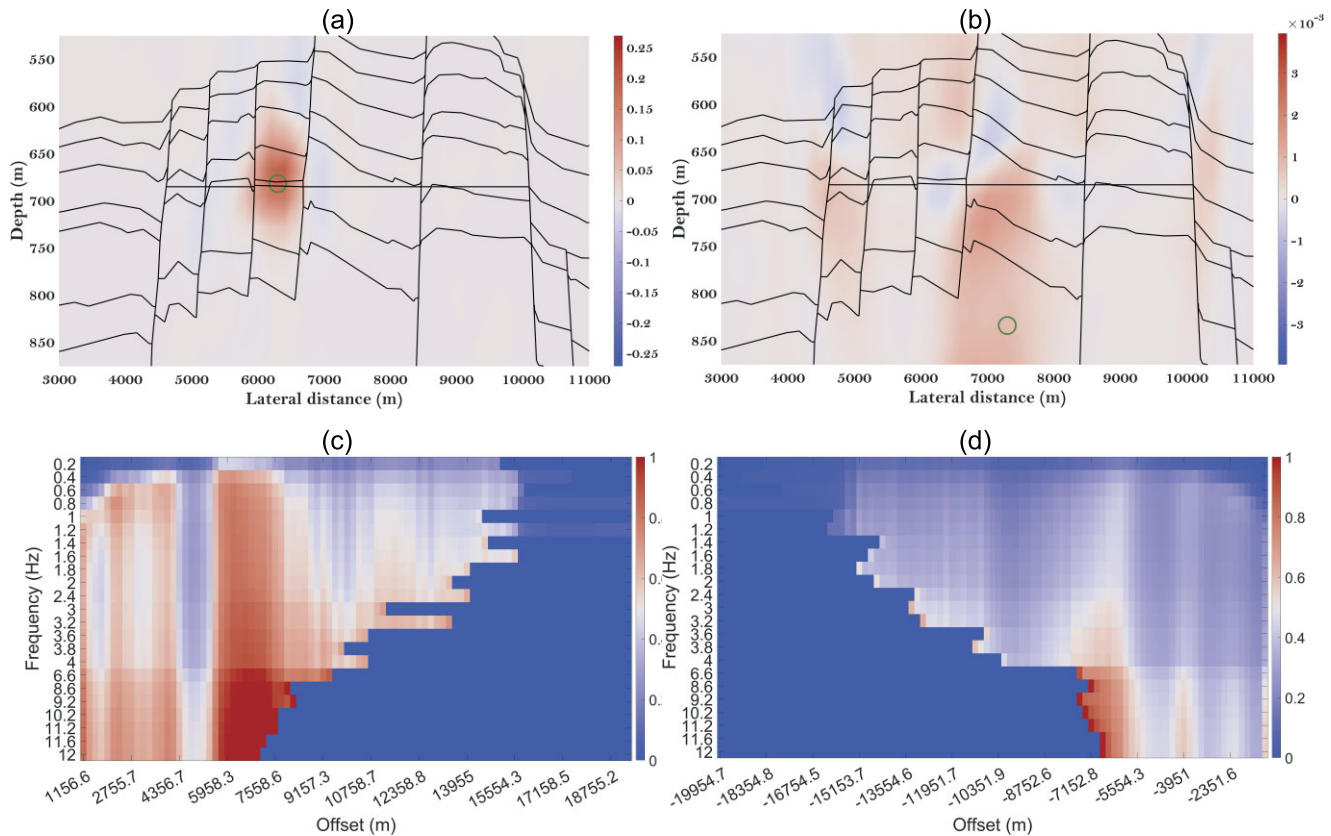
Figure 11. PSF for a well-resolved (a) and poorly resolved model parameter (b), with the circle indicating the position of the parameter. Note that there are two orders of magnitude between the maximum value of the two PSFs. Data Importance is shown as a function of frequency and offset for Receivers 2 (c) and 4 (d). Refer to Figs 1(a) and 6 for receiver locations.

and 0.4 Hz) are retained for the remaining receivers (1, 2, 3 and 5) to avoid the observed vertical displacement. The left-hand column in Fig. B1 shows the data subsampling introduced for each receiver (Appendix B). Fig. 12(e) shows the ratio of resolution for this subsampled data set. Based on the quality of this map, the resolving power of this new subsampled data set is expected to be fair. A direct comparison of the inverted models for both Cases 2 and 3 (Figs 12d and f, respectively) reveals that most of the character is retained.

5.4 Case 4—30 per cent cut-off and limited frequencies

Both Cases 2 and 3 only considered data decimation based on Data Importances. Thus, this approach did not place any additional emphasis on which frequencies to transfer to the decimated data set. However, there is good coverage over a large range of frequencies. Therefore, Case 4 subsamples the data set from Case 3 along the frequency dimension. This selection was completed based on several criteria. The source signature in the frequency domain was used as a guide (Fig. 3), along with the Data Importance panels. However, the most important criterion was to evaluate the resolving power of the subsampled data set in form of its ratio of resolution. This can be done by testing different frequency combinations and then evaluating the corresponding ratios of resolution. Directly comparing the ratio of resolution maps for Cases 3 and 4 shows that they are nearly identical (Figs 12e and g). Moreover, this similarity is also reflected in the inverted model domain (*cf.* Figs 12f and h). Refer to

the right column of Fig. B1 (Appendix B) for the data subsampling introduced for each receiver.

5.5 Synthetic data—Summary

The four cases presented demonstrate that this study's proposed data decimation scheme is feasible. Table 2 gives the decimation of each case as a percentage of the full data set. It is important to note that all the inverted models exhibit the same character, especially with regard to the three compartments and the lateral extension. Case 2 shows that an entire receiver can be removed without significantly changing the inversion result. As demonstrated by Case 3, even a severe data decimation of ≈61 per cent is feasible without losing essential resolving power. Upon further inspection of the Data Importance panels (Figs 11c and d), it is clear that a large range of frequencies are covered even after Case 3's data decimation. In the final Case 4, we therefore limit the number of frequencies from the original 23 down to 11. Note that this case uses only ≈23 per cent of the original data, while still preserving the main features of the model.

6 FIELD DATA

The results presented in the synthetic data section demonstrate that it is possible to severely downsample the original data while still preserving the main features in the final inverted model. However, synthetic data are associated with an ideal model case. Field data,
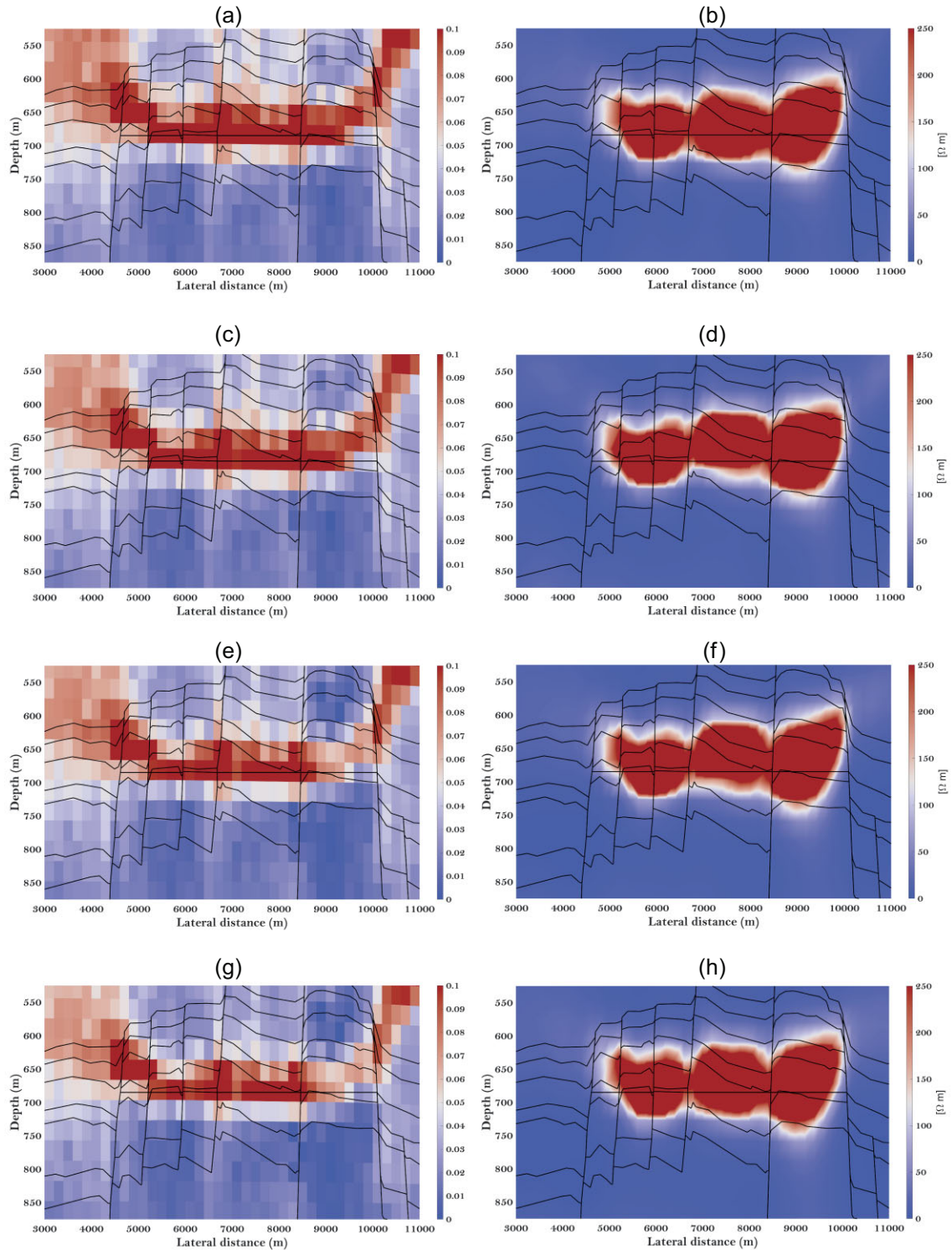
Figure 12. Ratio of resolution for Cases 1 through 4 (a, c, e, g) and corresponding inverted models (b, d, f, h).

on the other hand, represent the response from a more complicated earth model, with further complications resulting from imperfections in survey and instrumentation. Nonetheless, our proposed approach is still useful in the case of real data. To support this claim, this section provides a simple example from the

Wisting field data. In the field data inversion, we used the same starting model as described in Section 5. However, synthetic data allows for full control of the background model. Conversely, for the field data lacking this level of control, a poor inversion result was observed from using the full data set. Thus, the maximum

Table 2. Description of the four cases and the percentage of the full data set used in each inversion.

	Description	Per cent of full data set
Case 1	Full data set Frequencies [Hz]:0.2, 0.4, 0.8, 1, 1.2, 1.4, 1.6,1.8, 2, 2.4, 3, 3.2, 3.6, 3.8,4, 6.6, 8.6, 9.2, 10.2, 11.2, 11.6, 12	100 per cent
Case 2	Removed least influential receiver Frequencies [Hz]:0.2, 0.4, 0.8, 1, 1.2, 1.4, 1.6,1.8, 2, 2.4, 3, 3.2, 3.6, 3.8,4, 6.6, 8.6, 9.2, 10.2, 11.2, 11.6, 12	$\approx$ 82.5 per cent
Case 3	Removed least influential receiver Cut-off at the 70th percentile No cut-off for the two lowest frequencies (0.2 and 0.4 Hz) Frequencies [Hz]:0.2, 0.4, 0.8, 1, 1.2, 1.4, 1.6,1.8, 2, 2.4, 3, 3.2, 3.6, 3.8,4, 6.6, 8.6, 9.2, 10.2, 11.2, 11.6, 12	$\approx$ 39 per cent
Case 4	Removed least influential receiver Cut-off at the 70th percentile No cut-off for the two lowest frequencies (0.2 and 0.4 Hz) Frequencies [Hz]:0.2, 0.4, 0.8, 1.2, 2, 3, 4, 6.6, 9.2, 10.2, 12	$\approx$ 23 per cent

offset was limited to about 10 km to remove such artefacts. The inversion was not constrained by the interpreted horizons, and no constraints were introduced on the maximum resistivity to be recovered.

6.1 Field Case 1—Full data set

A direct comparison between the inversion of the field data and synthetic data reveals that the reservoir has a shallower placement in the field data (Figs 12b and 14b). This effect might arise from attempting to solve a 3-D problem by using a 2.5-D inversion algorithm. When using this 2.5-D technique, we assume no variations in the electrical properties along the strike direction. Moreover, there is also an underlying assumption that the model extends to infinity along the same direction. Thus, the inversion tries to compensate for these inconsistencies by placing the reservoir at a shallower depth. In both the field and synthetic inversions, the reservoir shows three compartments. However, these three compartments are more distinctly separated in the synthetic inversion. Moreover, the lateral extension of the reservoir is virtually the same in both. No distinct differences in resistivity between the three compartments were observed.

Another distinct feature of the Field Case 1 inversion can be seen by examining the rightmost compartment. This compartment is placed deeper in the model, which might be explained by examining the ratio of resolution plot (Fig. 14a). This map reveals that the rightmost compartment is more poorly resolved than the remaining part of the reservoir. Fig. 13 shows the field data along with the modelled response from the inverted model. As briefly mentioned, the offset range in this field data example is limited in order to remove unwanted artefacts. The recovered model accurately describes the field data response for most offsets.

6.2 Field Case 2—Limited frequencies

Some similarities are evident from direct comparison between the Data Importance panels for the field data and synthetic data. The Data Importance of Receiver 4 is generally lower for both the synthetic and field inversions. However, none of the receivers stand out to the extent as seen in the synthetic data case. Thus, it seems that the best approach for field data is to retain all receiver positions. In Field Case 2, we therefore subsample the input data along the frequency direction. We use the same frequency range as in Case 4 for the synthetic data. The subsampled data set corresponds to $\approx$47 per cent of the full field data set (Table 3). Even with such a severe downsampling, the ratio of resolution and inverted models of Field Cases 1 and 2 are virtually the same (Fig. 14).

7 DISCUSSION

The results show that the resolution matrices carry essential information in the case of CSEM inversion. Such information can be used to subsample data without losing essential resolving power. However, some challenges are also observed regarding the proposed method.

This study introduces Vertical Transverse Isotropy in the inversion, which implies access to both vertical and horizontal resolution matrices. However, we have chosen to limit our reported study to vertical resolution matrices, since CSEM is generally known to be more sensitive to vertical resistivity. Another issue relates to the field components selected for the inversion. Here, the inline horizontal field component (E_y) has been chosen, since it is known to be the most important carrier of information. However, a superior constraint of horizontal resistivity might be achieved by introducing broadside data in the inversion (Masnaghetti & Ceci 2010). While the proposed method for subsampling should also be valid for broadside data, the data analysis would be more comprehensive, as the resolution matrices need to be investigated separately for the different data inputs.

In an exploration setting, it is common to acquire a full survey with regular spatial sampling. However, Case 2 of the synthetic data study illustrates how Data Importances can be used to design a repeated survey. For example, say that the objective is to monitor the Wisting field throughout its lifespan by use of CSEM. First, a baseline survey (with regular spatial sampling) would be acquired before production starts. Thereafter, this baseline survey could be used to calculate resolution matrices and derived quantities. This information can then be used as a guide before acquiring a repeated survey. In a 3-D layout including up to 200 receivers, it is highly unlikely that all receivers contribute equally to the inversion. It should therefore be possible to remove the least influential receivers without losing essential resolving power. Another important point relates to the actual placement of the receivers. By moving away from a regular spatial sampling, it might still be possible to preserve a high resolution. Shantsev *et al.* (2020) describe how the acquisition of a monitor survey with known changes in receiver positions does not necessarily compromise the result, as long as the comparison is made in the model domain. Thus, it may be possible to replace the conventional regularly sampled survey with its sparsely optimized counterpart.

Some challenges exist when seeking to improve survey design. The most fundamental issue is the impossible task of inferring the resolving power of a receiver location a priori. This implies that researchers need to acquire data at a proposed new location to know for certain if it constitutes an improvement. This issue might
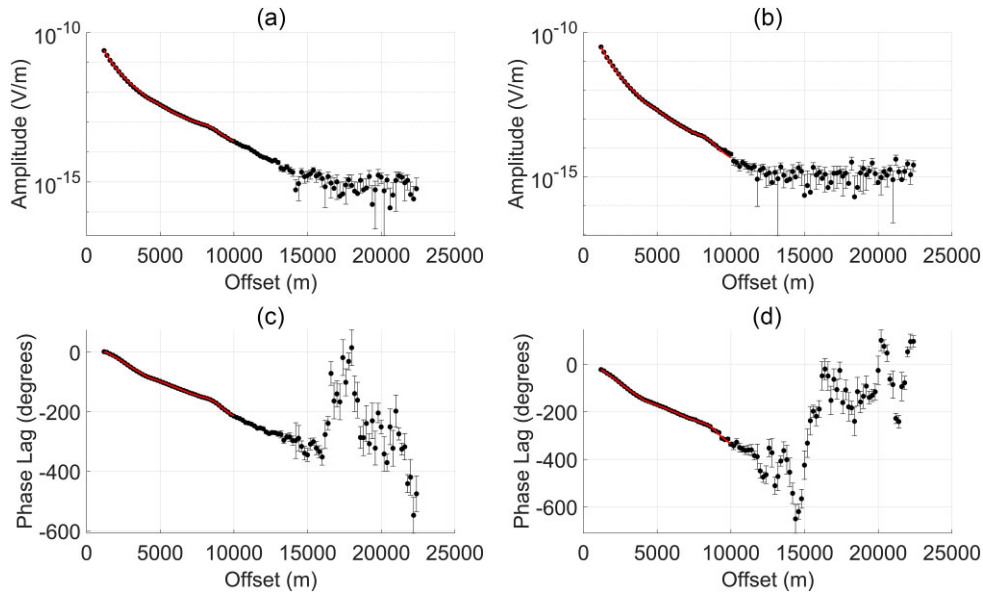
Figure 13. Field data (dots) and modelled response (line) obtained from the inverted model for Receiver 1 (*cf*. Fig. 6). On the left-hand side, MVO (a) and PVO (c) are shown for a 2 Hz response, while on the right-hand side the MVO (b) and PVO (d) are shown for 4 Hz. Note that the offset range has been limited due to the observations described above.

Table 3. Description of the two field cases and the percentage of the full data set used in each inversion.

	Description	Per cent of full data set
Field Case 1	Full data set Frequencies [Hz]:0.2, 0.4, 0.8, 1, 1.2, 1.4, 1.6,1.8, 2, 2.4, 3, 3.2, 3.6, 3.8,4, 6.6, 8.6, 9.2, 10.2, 11.2, 11.6, 12	100 per cent
Field Case 2	Frequencies: 0.2 0.4, 0.8, 1.2, 2, 3, 4, 6.6, 9.2, 10.2, 12	≈ 47 per cent

be resolved if a high-quality resistivity model is available. For a synthetic data case, where the earth model is well known, receiver sensitivity studies can be performed in advance. However, in case of field data, this cannot always be ensured. In this investigation, we have studied both field data and synthetic data associated with the Wisting oil field. Comparing the inversions of the synthetic and field data demonstrates clear similarities as well as discrepancies. For both data types, Receiver 4 stands out for its overall low Data Importance. However, it does not distinguish itself as clearly in the case of field data. Thus, the removal of this receiver might result in an unacceptable decrease in resolving power. The differences observed between the synthetic and field inversions might be due to the fact that the latter represents a 3-D earth response inverted using a 2.5-D inversion algorithm.

Another challenge regarding improvements to survey design in a production setting relates to corresponding changes in the reservoir. During production, resistivity is expected to decrease. This phenomenon might again introduce changes in the sensitivity of the different receiver locations. Shantsev *et al.* (2020) address this concern and demonstrate that time-lapse effects due to production are preserved in the inverted domain even in cases with major differences in survey layout between base and monitor data. Nonetheless, a detailed study of such effects with emphasis on the resolution matrices would be of value.

It is possible to significantly subsample the data along the frequency direction without losing essential resolving power. However, such selections should be guided by using ratio of resolution maps. Moreover, it seems that the best results are obtained by retaining the highest- and lowest-frequency components and then more sparsely filling in key frequencies between these two endpoints. Key (2009) made the same observation with regard to frequency sampling in 1-D CSEM inversion. However, while (Key 2009) emphasized that this observation might not hold in higher dimensions, our observations substantiate that this claim at least holds some merit in two dimensions.

8 CONCLUSION

The purpose of this work is to introduce and investigate the use of the resolution matrices in CSEM inversion and evaluate how such information can be used for data decimation and survey design in the case of a repeated survey. Proper testing of our proposed strategy required the construction of a high-quality resistivity model using well logs, seismic and CSEM data from the Wisting oil field in the southwestern Barents Sea. The MARE2DEM forward modelling and inversion package was used as a starting point for this study. We suggested a new metric, denoted ratio of resolution, to better evaluate the resolving power of a given data set. We also introduced a detailed framework to describe how resolution matrices can be used for both survey design and data decimation. Finally, we demonstrated the proposed approach on both synthetic and field data sets.

Our results show that the resolution matrices carry important information that can be used for more efficient data decimation and survey design. It is likely that significant data redundancy may exist in the acquisition of a full CSEM survey. Thus, utilizing the information carried by the resolution matrices allows the original data set to be downsampled without losing essential resolving power.
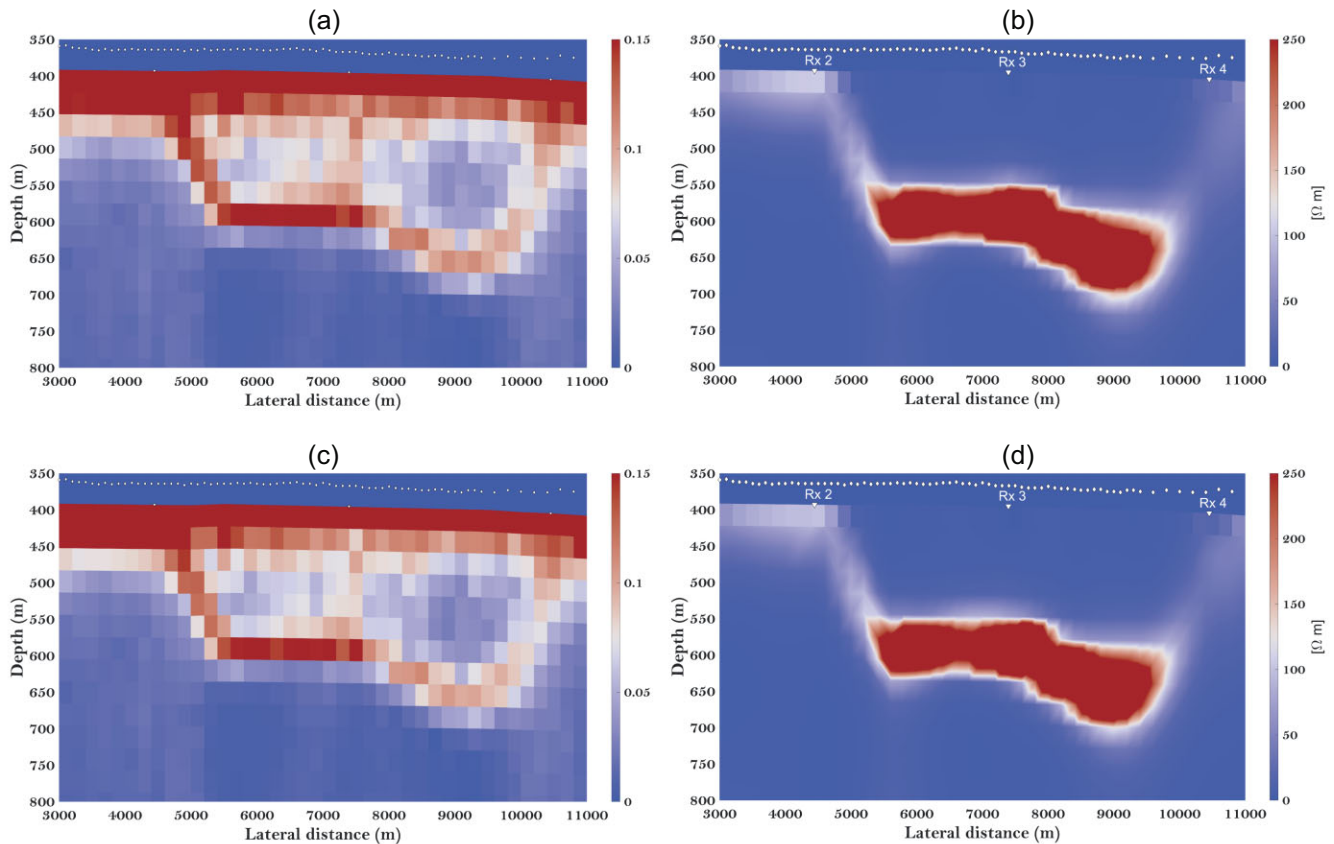
Figure 14. Ratio of resolution for Field Case 1 and 2 (a, c) and corresponding inverted models (b, d).

DATA AVAILABILITY

The data underlying this paper were provided by EMGS ASA under license. The data will be shared on a reasonable request to the corresponding author subject to permission from EMGS ASA.

REFERENCES

Brown, V., Hoversten, M., Key, K. & Chen, J., 2012. Resolution of reservoir scale electrical anisotropy from marine CSEM data, *Geophysics,* **77**(2), E147–E158.

Carvalho, B.R. & Menezes, P. T.L., 2017. Marlim R3D: a realistic model for CSEM simulations-phase I: model building, *Braz. J. Geol.,* **47,** 633–644.

Constable, S., 2010. Ten years of marine CSEM for hydrocarbon exploration, *Geophysics,* **75**(5), 75A67–75A81.

Constable, S., Parker, R.L. & Constable, C.G., 1987. Occam's inversion: a practical algorithm for generating smooth models from electromagnetic sounding data, *Geophysics,* **52**(3), 289–300.

Constable, S., Kannberg, P.K. & Weitemeyer, K., 2016. Vulcan: a deep-towed CSEM receiver, *Geochem. Geophys. Geosyst.,* **17**(3), 1042–1064.

Correa, J.L. & Menezes, P.T., 2019. Marlim R3D: a realistic model for controlled-source electromagnetic simulations—Phase 2: the controlled-source electromagnetic data set, *Geophysics,* **84**(5), E293–E299.

Engelmark, F., Mattsson, J., McKay, A. & Du, Z., 2014. Towed streamer EM comes of age, *First Break,* **32**(4), doi:10.3997/1365-2397.32.4.74381.

Friedel, S., 2003. Resolution, stability and efficiency of resistivity tomography estimated from a generalized inverse approach, *Geophys. J. Int.,* **153**(2), 305–316.

Granli, J.R., Veire, H.H., Gabrielsen, P. & Morten, J.P., 2017. Maturing broadband 3D CSEM for improved reservoir property prediction in the Realgrunnen Group at Wisting, Barents Sea, in *Proceedings of the 2017 SEG International Exposition and Annual Meeting,* OnePetro.

Grayver, A.V., Streich, R. & Ritter, O., 2014. 3D inversion and resolution analysis of land-based CSEM data from the Ketzin CO_2 storage formation, *Geophysics,* **79**(2), E101–E114.

Hansen, P.C., Nagy, J.G. & O'Leary, D.P., 2006. *Deblurring Images: Matrices, Spectra, and Filtering,* SIAM.

Hoversten, G.M., Røsten, T., Hokstad, K., Alumbaugh, D., Horne, S. & Newman, G.A., 2006. Integration of multiple electromagnetic imaging and inversion techniques for prospect evaluation, in *Proceedings of the SEG Technical Program Expanded Abstracts 2006,* pp. 719–723, Society of Exploration Geophysicists.

Jakobsen, M. & Tveit, S., 2018. Distorted Born iterative T-matrix method for inversion of CSEM data in anisotropic media, *Geophys. J. Int.,* **214**(3), 1524–1537.

Kalscheuer, T., De los Ángeles García Juanatey, M., Meqbel, N. & Pedersen, L.B., 2010. Non-linear model error and resolution properties from two-dimensional single and joint inversions of direct

current resistivity and radiomagnetotelluric data, *Geophys. J. Int.,* **182**(3), 1174–1188.

Key, K., 2009. 1D inversion of multicomponent, multifrequency marine CSEM data: methodology and synthetic studies for resolving thin resistive layers, *Geophysics,* **74**(2), F9–F20.

Key, K., 2016. MARE2DEM: a 2-D inversion code for controlled-source electromagnetic and magnetotelluric data, *Geophys. J. Int.,* **207**(1), 571–588.

Li, G., Li, Y., Han, B. & Liu, Z., 2018. Application of the perfectly matched layer in 3-D marine controlled-source electromagnetic modelling, *Geophys. J. Int.,* **212**(1), 333–344.

Lu, X. & Xia, C., 2007. Understanding anisotropy in marine CSEM data, in *Proceedings of the 2007 SEG Annual Meeting,* OnePetro.

Masnaghetti, L. & Ceci, F., 2010. Analysis of the sensitivity to anisotropy of CSEM data using 2.5 D modeling and inversion, in *Proceedings of the SEG Technical Program Expanded Abstracts 2010,* pp. 614–618, Society of Exploration Geophysicists.

Mattsson, J., 2015. Resolution and precision of resistivity models from inversion of towed streamer EM data, in *Proceedings of the 2015 SEG Annual Meeting,* OnePetro.

Maurer, H., Boerner, D.E. & Curtis, A., 2000. Design strategies for electromagnetic geophysical surveys, *Inverse Problems,* **16**(5), doi:10.1088/0266-5611/16/5/302.

McKay, A., Mattsson, J. & Du, Z., 2015. Towed streamer EM-reliable recovery of sub-surface resistivity, *First Break,* **33**(4), doi:10.3997/1365-2397.33.4.79748.

Menke, W., 2012. *GeophySical Data Analysis: Discrete Inverse Theory,* MATLAB edition, Vol. **45**, Academic Press.

Newman, G.A., Commer, M. & Carazzone, J.J., 2010. Imaging CSEM data in the presence of electrical anisotropy, *Geophysics,* **75**(2), F51–F61.

Parasnis, D., 1988. Reciprocity theorems in geoelectric and geoelectromagnetic work, *Geoexploration,* **25**(3), 177–198.

Ren, Z. & Kalscheuer, T., 2020. Uncertainty and resolution analysis of 2D and 3D inversion models computed from geophysical electromagnetic data, *Surv. Geophys.,* **41**(1), 47–112.

Romdhane, A. & Eliasson, P., 2018. Optimised geophysical survey design for CO_2 monitoring—a synthetic study, in *Proceedings of the 14th Greenhouse Gas Control Technologies Conference,* Melbourne, pp. 21–26.

Rossmann, K., 1969. Point spread-function, line spread-function, and modulation transfer function: tools for the study of imaging systems, *Radiology,* **93**(2), 257–272.

Roux, E. & García, X., 2014. Optimizing an experimental design for a CSEM experiment: methodology and synthetic tests, *Geophys. J. Int.,* **197**(1), 135–148.

Schlumberger, 2018. *Petrel 20 years book,* https://www.software.slb.com/-/media/software-media-items/software/documents/external/product-sheets/petrel-20-years-book.pdf, Accessed: 14-02-2022.

Senger, K., Birchall, T., Betlem, P., Ogata, K., Ohm, S., Olaussen, S. & Paulsen, R.S., 2021. Resistivity of reservoir sandstones and organic rich shales on the Barents Shelf: Implications for interpreting CSEM data, *Geosci. Front.,* **12**(6), doi:10.1016/j.gsf.2020.08.007.

Shantsev, D.V., Nerland, E.A. & Gelius, L.-J., 2020. Time-lapse CSEM: how important is survey repeatability?, *Geophys. J. Int.,* **223**(3), 2133–2147.

Stefani, J., Frenkel, M., Bundalo, N., Day, R. & Fehler, M., 2010. SEAM update: models for EM and gravity simulations, *Leading Edge,* **29**(2), 132–135.

Wang, F., Morten, J.P. & Spitzer, K., 2018. Anisotropic three-dimensional inversion of CSEM data using finite-element techniques on unstructured grids, *Geophys. J. Int.,* **213**(2), 1056–1072.

Weitemeyer, K., Gao, G., Constable, S. & Alumbaugh, D., 2010. The practical application of 2D inversion to marine controlled-source electromagnetic data, *Geophysics,* **75**(6), F199–F211.

Wheelock, B., Constable, S. & Key, K., 2015. The advantages of logarithmically scaled data for electromagnetic inversion, *Geophys. J. Int.,* **201**(3), 1765–1780.

Zhdanov, M.S., 2010. Electromagnetic geophysics: notes from the past and the road ahead, *Geophysics,* **75**(5), 75A49–75A66.

APPENDIX A: SENSITIVITY CALCULATIONS AND TRANSFORMATION TO COMPLEX FIELDS

In order to extract the resistivity model from observed data, iterative inversion schemes are used. In deterministic inversion, the partial derivatives with respect to model parameters form a crucial part of the inversion scheme. These partial derivatives are often denoted sensitivities and together makes up the Jacobian matrix. We start by defining sensitivity as introduced in Key (2016). Let σ_j represent an arbitrary conductivity parameter in our earth model. The sensitivity of one datapoint (i.e. one unique source, receiver and frequency combination) with respect to this model parameter can now be calculated as:

$$\frac{\partial F}{\partial \sigma_j}(x, y, z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{s}_j(k_x, y, z) e^{ik_x(x_r - x_s)} dk_x, \quad (A1)$$

where $\hat{s}_j(k_x, y, z)$ is given by

$$\hat{s}_j(k_x, y, z) = \int_{A_j} \hat{\mathbf{E}}^{\mathbf{a}}(-k_x, y, z) \left(\frac{\partial \bar{\bar{\sigma}}}{\partial \sigma_j} \hat{\mathbf{E}}(k_x, y, z) \right) dA_j. \quad (A2)$$

$\mathbf{E}$ and $\mathbf{E}^{\mathbf{a}}$ denote, respectively, the electric field and adjoint electric field in the wavenumber domain. The adjoint field is created by turning the corresponding receiver into an adjoint source. In eqs (A1) and (A2), A_j denotes the area of the cell containing conductivity parameter σ_j, while x_r and x_s describe the along strike position of the receivers and sources. The entries of the Jacobian matrix related to this unique datapoint can now be written as

$$J_j = \frac{\partial F[\mathbf{m}]}{\partial m_j} = \frac{\ln(10)}{\rho_j} \frac{\partial F[\mathbf{m}]}{\partial \sigma_j} = \frac{\partial F[\mathbf{m}]}{\partial \log(\rho_j)} = \frac{\partial \mathbf{d}}{\partial \log(\rho_j)}. \quad (A3)$$

Eq. (A3) takes this special form since the inversion in MARE2DEM is parametrized with respect to log transformed resistivities. Moreover, the actual sensitivity output from MARE2DEM is given separately for log amplitude and phase (in degrees). However, in order to carry out the analysis proposed in this paper, it is crucial that the entries in the Jacobian matrix represent the complex field. For one unique datapoint in the data vector $\mathbf{d}$, the complex data sample d can be formally written as

$$d = ae^{i\phi_{\mathrm{rad}}} = a(\cos(\phi_{\mathrm{rad}}) + i\sin(\phi_{\mathrm{rad}})), \quad (A4)$$

where a is the amplitude and ϕ_{rad} is the phase given in radians. By taking the derivative with respect to the base 10 logarithm of the model we get

$$\frac{\partial d}{\partial \log(\rho_j)} = \frac{\partial a}{\partial \log(\rho_j)} e^{i\phi_{\mathrm{rad}}} + ae^{i\phi_{\mathrm{rad}}} i \frac{\phi_{\mathrm{rad}}}{\partial \log(\rho_j)}$$

$$= d \left(\frac{1}{a} \frac{\partial a}{\partial \log(\rho_j)} + i \frac{\partial \phi_{\mathrm{rad}}}{\partial \log(\rho_j)} \right). \quad (A5)$$

The quantities inside the brackets in eq. (A5) can be computed as:

$$\frac{1}{a} \frac{\partial a}{\partial \log(\rho_j)} \approx \ln(10) \frac{\partial \log(a)}{\partial \log(\rho_j)}, \quad (A6)$$

and

$$\frac{\partial \phi_{\mathrm{rad}}}{\partial \log(\rho_j)} = \frac{\pi}{180} \frac{\partial \phi_{\mathrm{degree}}}{\partial \log(\rho_j)}. \quad (A7)$$

Thus, by combining eqs (A4), (A5), (A6) and (A7) we can construct the Jacobian matrix for the complex field as follows:

$$\frac{\partial d}{\partial \log(\rho_j)} \approx d \left(\ln(10)\frac{\partial \log(a)}{\partial \log(\rho_j)} + i\,\frac{\pi}{180}\frac{\partial \phi_{\text{degree}}}{\partial \log(\rho_j)} \right), \tag{A8}$$

where $\frac{\partial \log(a)}{\partial \log(\rho_j)}$ and $\frac{\partial \phi_{\text{degree}}}{\partial \log(\rho_j)}$ represent sensitivity outputs from MARE2DEM. The Jacobian entries as calculated from eq. (A8)

must also be accompanied with corresponding standard errors (δ) of the measurement data. These errors form the diagonal weighting matrix $\mathbf{W_d}$ used in the expressions for the resolution matrices. For one particular datapoint, δ is computed as

$$\delta = \frac{1}{\sqrt{2}a\sigma_l}, \tag{A9}$$

where σ_l denotes the user defined noise level (i.e. 1 per cent $= 0.01$)
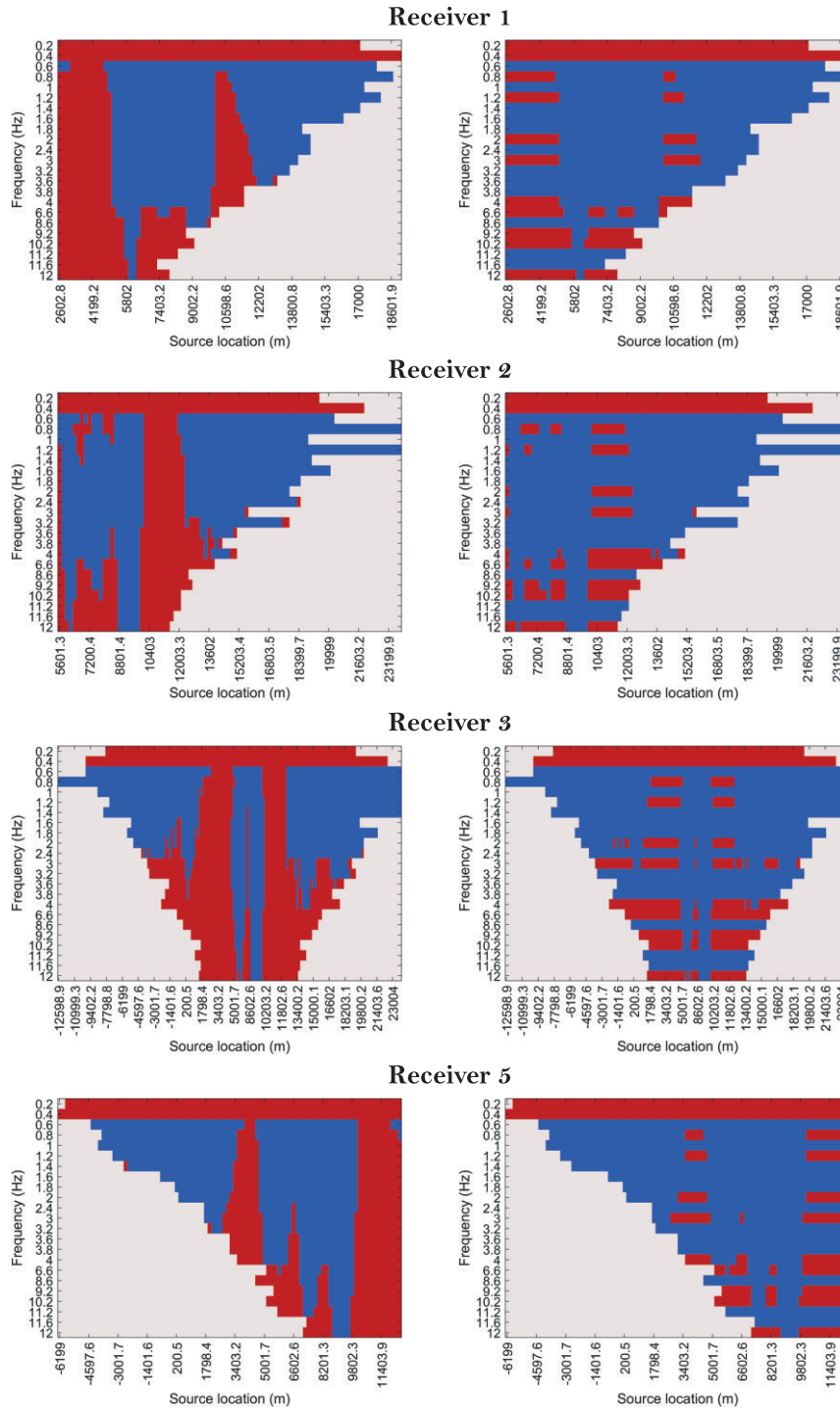
APPENDIX B: SUBSAMPLED DATA SETS

Figure B1. Data subsampling for Cases 3 (left-hand column) and 4 (right-hand column) in the main body of the text. Red indicates data input to the inversion, while blue signifies data which have been removed by the thresholding described in Section 4.5. Row 1, 2, 3 and 4 refers to receiver 1, 2, 3 and 5, respectively (Fig. 1a). Note that the data subsampling plot for Receiver 4 is omitted due to the findings in Case 2.