AARHUS UNIVERSITY

Coversheet

This is the publisher's PDF (Version of Record) of the article.

This is the final published version of the article.

How to cite this publication:

T M Hansen, B J Minsley, Inversion of airborne EM data with an explicit choice of prior model, *Geophysical Journal International*, Volume 218, Issue 2, 1 August 2019, Pages 1348–1366, <https://doi.org/10.1093/gji/ggz230>

Publication metadata

Title:	Inversion of airborne EM data with an explicit choice of prior model
Author(s):	T M Hansen, B J Minsley
Journal:	<i>Geophysical Journal International</i>
DOI/Link:	https://doi.org/10.1093/gji/ggz230
Document version:	Publisher's PDF (Version of Record)

This article has been accepted for publication in *Geophysical Journal International* ©: The Author(s) 2019. Published by Oxford University Press on behalf of the Royal Astronomical Society. All rights reserved.

General Rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

If the document is published under a Creative Commons license, this applies instead of the general rights.

Inversion of airborne EM data with an explicit choice of prior model

T.M. Hansen¹ and B.J. Minsley²

¹Department of Geoscience, Aarhus University, 8000 Aarhus C, Denmark. E-mail: tmeha@geo.au.dk

²U.S. Geological Survey, Geology, Geophysics, and Geochemistry Science Center, Denver, CO 80225, USA

Accepted 2019 May 17. Received 2019 April 30; in original form 2018 September 7

SUMMARY

Inversion of airborne electromagnetic (AEM) data is an under-determined inverse problem, in that infinitely many resistivity models exist that will be able to explain the observed data, within measurement errors. Therefore, additional information or constraints must be taken into account to solve the inverse problem. In deterministic approaches, the goal is to locate one optimal model that can be obtained by using some form of smoothness constraints implied through a number of regularization choices. This model, however, will not necessarily represent realistic geological features. Probabilistic methods offer an alternative in which the solution is not one model, but a collection of models, whose variability represents the uncertainty. The probabilistic approach can also rely on implicit model assumptions, representing prior information (a type of regularization information) that may or may not be consistent with the actual available information. Here, we present an approach for AEM inversion in which the prior model is explicitly chosen by a user, preferably selected based on actual prior information available and then integrated with AEM data using a general Monte Carlo based sampling approach. This approach leads to a new workflow to AEM inversion in which geological prior information is independently and explicitly chosen before inversion is carried out. The main benefit of this approach is that each model obtained will, by construction, be consistent with prior (geological) information as well as geophysical data. Through examples based on synthetic and real AEM data, we will demonstrate the methodology, not least that the choice of prior information cannot be avoided: Either it is done explicitly, or it will be chosen implicitly by the choice of method used to invert the AEM data.

Key words: Inverse theory; Statistical methods.

1 INTRODUCTION

Airborne electromagnetic (AEM) surveys have in the last decades been developed and applied to map large areas of near-surface electrical resistivity variations. As for geophysical measurements in general, such measurements are ultimately carried out in order to increase knowledge about the subsurface geological structure that cannot be directly assessed over large areas and with high spatial resolution. Ideally, this knowledge is used to help inform decisions related to, for example, reservoir exploration, characterization and management or mineral exploration.

Multiple AEM methods exist that can be used to measure an EM response, sensitive to near-surface resistivity (down to hundreds of meters depending on the method) in both frequency-domain, for example, RESOLVE (CGG) and time-domain, for example, TEM-PEST (Lane *et al.* 2000; Worrall *et al.* 2001), SkyTEM (Sørensen & Auken 2004) and VTEM (Wetherly *et al.* 2004). For a more detailed overview of AEM methods and the difference between time- and frequency-domain systems see, for example, McCracken

et al. (1986), Siemon *et al.* (2009), Legault (2015) and Auken *et al.* (2017).

Inverse methods can be used to infer information about the distribution of subsurface resistivity (and properties related to resistivity) from observed EM data. Many different approaches have been considered to solve this inverse problem that can roughly be regarded as either a ‘deterministic’ or ‘probabilistic’ method. The goal of deterministic methods is to locate one, in some sense optimal, resistivity model along with some uncertainty estimate that represents observed data, see for example, Constable *et al.* (1987) and Menke (2012). The goal of probabilistic methods is to locate a (large) collection of models (realizations from a posterior probability distribution) which ideally represent all available information, see for example, Tarantola (2005).

In any case, inversion of AEM data is a non-unique inverse problem. Infinitely many models will be able to fit the observed data within its noise. This means that additional information must be assumed in order to solve the inverse problem. For deterministic inversion methods, this is provided through ‘regularization’ that

Figure 1. A sequential workflow. AEM data are inverted to obtain one optimal model (m^1); a geological expert uses that model to set up a conceptual hydrogeological model (m^2).

can, for example, control the degree of simplicity/smoothness of the solution model. For probabilistic inverse problems, additional information is quantified as ‘prior information,’ typically through a probability function. As will be discussed later, the choice of regularization parameters is one specific, sometimes non-intuitive, approach to quantify prior information.

1.1 Deterministic AEM inversion

A number of methods have been proposed for deterministic inversion of AEM data. Christensen (2002) proposed a 1-D approach for very fast imaging (inversion) of AEM data. Pseudo 2-D and 3-D inversion using spatially constrained inversion, using a 1-D forward model, is described by Auken & Christiansen (2004) and Viezzoli *et al.* (2008). Christensen & Dodds (2007) describe 1-D EM inversion in a marine setting. Tølbøll & Christensen (2006) describe 1-D inversion of helicopter AEM data. Vignoli *et al.* (2015, 2017) make use of a type of regularization that favours retrieval of sharp layer boundaries. These methods are all based on using a 1-D forward model, through a linearized least-squares type inversion (Menke 2012; Auken *et al.* 2014). Therefore, these methods are also computationally very efficient. Full 3-D inversion (using a 3-D forward model), also based on least-squares type Tikhonov regularization, is proposed by, for example, Cox *et al.* (2010, 2012), Yang & Oldenburg (2012) and Grayver *et al.* (2013). Common to all these methods is that regularization parameters are related to properties of a prior model covariance operator. In other words, using any of these deterministic methods described above involves an implicit Gaussian prior assumption. Note though that the Gaussian choice does not necessarily lead to smooth inversion results (Vignoli *et al.* 2015). Brodie & Sambridge (2006) propose an alternative approach in which layer resistivities and thicknesses are described by spline surfaces. This is another implicit prior assumption on the resistivity field.

A common problem using deterministic methods is that it is typically part of a sequential workflow, see Fig. 1. Initially, AEM data are inverted to obtain an optimal model of subsurface resistivity. Then this model is passed to a geologist, who makes a geological interpretation. The process may continue to a hydrological or other process-based model. In this deterministic sequential workflow, it is not trivial (in practice impossible) to propagate all the uncertainty from the geophysical (and other types of) data through to the final model used to make decisions.

1.2 Stochastic/probabilistic AEM inversion

The goal of probabilistic inversion is to locate a large number of models that are all consistent with both data and prior assumptions. The variability of the models reflects the combined information. Ideally, the probability of occurrence of any event can be obtained simply by scanning through all the models while computing the frequency with which that specific feature occurs.

Minsley (2011) presents an application of a probabilistic approach for inversion of frequency-domain AEM data, based on a

Figure 2. A probabilistic workflow. The information related to AEM data and geological/structural information are quantified through independent probability densities, which are then combined to one probability distribution representing the combined information.

parsimonious/trans-dimensional method (Malinverno 2002; Sambridge *et al.* 2006). Brodie & Sambridge (2012) and Christensen *et al.* (2017) apply a similar approach to inversion of time-domain AEM data. In these studies, a Monte Carlo sampling approach is considered, in which the model space is explored using a reversible jump strategy (Green 1995). This leads to an implicit prior assumption that the subsurface can be described as layered 1-D models of resistivity, with an unknown number of layers (which is to be inferred as part of the inversion). The natural parsimony of this method favours simpler (fewer layers) models. In both cases, prior constraints on resistivity values typically involve normal or uniform distributions. Hawkins *et al.* (2018) propose to invert time-domain AEM data using a 2-D trans-dimensional tree approach with a wavelet parametrization. In this approach, a choice of wavelet base and wavelet coefficients in effect determines a prior model (Hawkins & Sambridge 2015).

Hauser *et al.* (2015) propose to use the Bayesian parametric bootstrap method to invert AEM data. This is a probabilistic approach that implicitly assumes that a Gaussian probability distribution can be used to describe subsurface layer conductivities and thicknesses of a 1-D layered model. In practice, the inverse problem is solved iteratively by re-centring the prior distribution to the location of the current posterior Gaussian distribution. Thus, the final posterior distribution is a local Gaussian probability distribution, linearized around the model with maximum posterior probability. This work is generalized by Hauser *et al.* (2016) who propose a parametrization that allows for multiple thin plates.

Using any of the methods referenced above involves an implicit choice of prior information. A fundamental question is then whether the prior assumption is consistent with the actual knowledge available such as geologic structure. We will argue that, in general, this is not the case.

We suggest changing the deterministic sequential workflow, Fig. 1, into a parallel probabilistic workflow, Fig. 2, to propose a fundamentally different approach for inverting AEM data in which prior/structural information is explicitly quantified through a probability distribution prior to inversion, independent of the available AEM data. Then, an appropriate method for solving the problem of integrating all the available information should be chosen that can take into account the available information. We argue that one should not try to force specific implicit assumptions using a specific inversion method, but rather choose an inversion method that can handle the available information.

As discussed, the inversion methods discussed above involves implicit prior model assumptions (through e.g. regularization and choice of parametrization). If the information available is such that a specific type of probabilistic inversion can represent exactly that, then that inversion method can be safely chosen. However, in general, this may not be the case for the methods described above.

The inverse problem can be formulated in a probabilistic setting following Tarantola & Valette (1982b), which in effect corresponds to integration of information described by probability densities (Hansen *et al.* 2016a), where the solution is the posterior probability distribution. We consider using sampling methods as a general approach to sample the posterior probability distribution, which consists of a collection of resistivity models, all consistent with all the available data and prior information (Mosegaard & Tarantola 1995). We demonstrate how such an algorithm can be applied as long as the available prior information can be described by a (possibly black box) sampling algorithm that samples the prior probability distribution. We will pay particular interest to the problem of how to quantify available spatial prior information (i.e. from geological knowledge), and how it affects the solution to the inverse problem.

First, we lay out the theory for probabilistic inversion of AEM data. Then, we demonstrate using a synthetic 1-D case that in practice the choice of the prior model cannot be avoided. If the prior model is not chosen explicitly, it is most always chosen implicitly by the choice of inversion methods used. Then we demonstrate, on data from Morrill, Nebraska (Abraham *et al.* 2012), how very different prior models can be explicitly defined and modelled, and demonstrate how such prior model choices affect the inversion results obtained from probabilistic inversion. Finally, we discuss the practical use of the methodology for decision makers and risk assessment.

2 THEORY

The main challenge in an inverse problem is to infer information about some model parameters $\mathbf{m}$ given some available information. Tarantola & Valette (1982b) developed a probabilistic approach to solving inverse problems, in which all information has to be quantified probabilistically through a probability distribution or a likelihood. Typically, as is considered here, available information consists of so-called prior information and information from geophysical data. In the following, the model parameters $\mathbf{m}$ represent subsurface log-resistivity.

$\rho(\mathbf{m})$ is the probability distribution representing prior knowledge. This can come from nearby outcrops, geological expert knowledge, previous surveys and similar sources. Information from geophysical data is described by $f_{\text{data}}(\mathbf{m})$, quantified by the likelihood function, which quantifies how well the forward response of a given model matches observed data, $L(\mathbf{d}_{\text{obs}} - g(\mathbf{m}))$. $\mathbf{d}_{\text{obs}}$ represents the observed data, and g is a forward mapping operator that maps a set of model parameters $\mathbf{m}$ into a set of noise-free data $\mathbf{d} = g(\mathbf{m})$. The likelihood function is chosen as a statistical model (often Gaussian) that describes the expected data residual ($\mathbf{d}_{\text{obs}} - \mathbf{d}$).

Once the available information is quantified, the combined states of information (the prior distribution and the likelihood) can be obtained using the concept of *conjunction of states of information*, which results in the posterior probability distribution:

$$\sigma(\mathbf{m}) \propto \rho(\mathbf{m}) f_{\text{data}}(\mathbf{m}) \quad (1)$$

$$= \rho(\mathbf{m}) L(\mathbf{d}_{\text{obs}} - g(\mathbf{m})). \quad (2)$$

In the specific case where the forward problem is linear and both the prior distribution and the likelihood can be described by Gaussian probability distributions, an analytical expression for a corresponding Gaussian distributed posterior distribution, $\sigma(\mathbf{m})$, can be obtained (Tarantola & Valette 1982a). The forward problem of

computing the AEM data response can though, in general, not be described by a linear forward model.

Instead, in the general nonlinear and non-Gaussian case, a number of sampling methods exist that allow sampling the posterior probability distribution, to generate a collection of realizations (models) of the posterior probability distribution that occur with a frequency proportional to $\sigma(\mathbf{m})$ (Hansen *et al.* 2016a). This allows detailed uncertainty and risk analysis to be carried out.

2.1 The extended Metropolis algorithm

One such sampling method is the extended Metropolis algorithm as proposed by Mosegaard & Tarantola (1995). It is specifically designed to sample from the product of two probability densities, as given in eq. (2). An implementation of the extended Metropolis algorithm is given by the following Markov chain, where the current state (the current model, $\mathbf{m}_{\text{cur}}$) is iteratively updated:

(1) Generate an initial model $\mathbf{m}_{\text{cur}}$ from $\rho(\mathbf{m})$.

(2) Start Loop.

(i) Generate a new realization $\mathbf{m}_{\text{pro}}$ from $\rho(\mathbf{m})$, in the vicinity of $\mathbf{m}_{\text{cur}}$.

(ii) Accept the move to the new model ($\mathbf{m}_{\text{pro}} \rightarrow \mathbf{m}_{\text{cur}}$), with probability

$$P_{\text{acc}} = \min \left[1, \frac{L(\mathbf{d}_{\text{obs}} - g(\mathbf{m}_{\text{pro}}))}{L(\mathbf{d}_{\text{obs}} - g(\mathbf{m}_{\text{cur}}))} \right]. \quad (3)$$

(iii) Store current state, $\mathbf{m}_{\text{cur}}$. Goto A.

(3) End Loop.

The only requirements for using the method is that one must be able to (1) sample the prior distribution $\rho(\mathbf{m})$, (2) solve the forward problem $\mathbf{d} = g(\mathbf{m})$ and (3) evaluate the likelihood model $L(\mathbf{d}_{\text{obs}} - g(\mathbf{m}))$.

The extended Metropolis algorithm is closely related to the Metropolis and Metropolis-Hastings algorithms (Metropolis *et al.* 1953; Hastings 1970). They are all guaranteed to sample the posterior distribution in finite time and differ only in computational efficiency.

The key difference between the algorithms is that the Metropolis and Metropolis-Hastings algorithms require evaluation of the posterior distribution at each iteration, which typically involves evaluating both the prior probability distribution and the likelihood. The extended Metropolis algorithm only requires evaluation of the likelihood, and does not need evaluation of the prior probability. Instead, it needs an algorithm that can sample the prior distribution through a random walk, such that forcing $P_{\text{acc}} = 1$ (ignoring the likelihood) leads to an algorithm that samples the prior distribution.

This latter constraint was utilized by Hansen *et al.* (2008, 2012) that demonstrated how most geostatistically based models that can be sampled using sequential simulation can also be efficiently sampled through a random walk, and hence be used to quantify prior information as part of a probabilistic inverse problem (2). Even in a case using a prior model that can be evaluated, as when using a Gaussian type prior probability distribution, it may be orders of magnitudes faster to use a combination of computationally efficient geostatistical sampling methods to sample the prior distribution and the extended Metropolis algorithm than to use the Metropolis algorithm. A large number of different geostatistical simulation algorithms exist, based on both 2-point and multiple point statistics, than can be used as part of the extended Metropolis algorithm,

Figure 3. EM1DFM inversion result using data from the West–East profile described in Minsley (2011).

and are capable of representing quite complex spatial variability, see, for example, Deutsch & Journel (1992), Strebelle (2002), Mariethoz *et al.* (2010), Zahner *et al.* (2015) and Hansen *et al.* (2016b).

For a more detailed introduction to the use of geostatistical prior models with probabilistically formulated inverse problems, see, for example, Hansen *et al.* (2016a).

3 AIRBORNE EM DATA FROM MORRILL, NEBRASKA

To exemplify the method, we consider the frequency-domain AEM data set acquired in 2009 in western Nebraska using the Resolve system, as described in Smith *et al.* (2010). Here, only one 22.5 km north–south trending flight line is considered, whose exact location can be found in fig. 12 in Minsley (2011). Fig. 3 shows the 2-D resistivity profile obtained by using the EM1DFM algorithm (a linearized least-squares based approach) (Farquharson *et al.* 2003).

Along the flight line, soundings were measured approximately every 3 m, and were averaged to an output interval of 50 m (451 soundings in all) measuring both the real and imaginary part of the signal for six frequencies, that is, 12 observations per sounding. For more details on the data, see Minsley (2011).

3.1 Setting up the inverse problem

In order to invert such AEM data in a probabilistic formulation one must select (1) a parametrization, (2) a prior model and (3) a probability distribution to represent the expected noise (through the likelihood).

3.1.1 Parametrization

A fundamental challenge is the choice of parametrization of the subsurface. Parametrization refers to the choice of a finite set model parameters used to describe the real Earth, such that it can be represented in a computer. Here, we follow the convention discussed in Mosegaard & Hansen (2016) to use a parametrization that is believed to represent the Earth with a ‘minimum required accuracy’. This choice is naturally related to the available *a priori* information.

Some parametrizations related to inversion of EM data use voxels whose size increase with depth; see for example Auken & Christiansen (2004). This choice is related to the physics of EM methods, where responses are increasingly insensitive to objects of a specific size with increasing depth. However, such increasingly coarse parametrization does not allow small-scale features at depth, even

if they are actually there, which is in conflict with conventions discussed in Mosegaard & Hansen (2016). The choice of parametrization we have made here is done specifically to allow smaller scale features both in the top and bottom of the model.

We choose to parametrize the subsurface into a Cartesian coordinate system of size 125×451 ($M = 56375$) voxels, represented by the model parameters $\mathbf{m} = [m_1, m_2, \dots, m_M]$. Each voxel is 1 m thick and 50 m wide, such that the model parameters span a 22 500 m by 150 m deep model below the surface. No feature smaller than $1\text{ m} \times 50\text{ m}$ can be represented by this choice of parametrization! Initially, a simple 1-D case will be considered for which $M = 125$.

3.1.2 Prior information

Three different types of 1-D prior models are initially considered differing in how much information is present in the specific choice of prior information. M_{free} refers to the ‘free number of model parameters’ for each type of prior model, that is, the minimum number of parameters needed to completely describe a realization of a prior distribution. In some cases, this number can be trivially obtained, and in other cases only an approximate estimate can be given, as will be discussed below. As the number of free parameters increases, so does the complexity of solving the corresponding inverse problem, as discussed by Hansen *et al.* (2009).

$\rho_1(\mathbf{m})$: 1-D maximum entropy. An example of a prior model that contributes little-to-no information (that results in a prior model with maximum entropy), is a spatially uncorrelated uniform probability distribution, such that the prior distribution of log-resistivity for each model parameter is $f(m_i) \sim \mathcal{U}[\log(5), \log(3000)]$. Fig. 4(a) shows 100 independent realizations from $\rho_1(\mathbf{m})$, and Fig. 5 the corresponding expected uniform 1-D marginal distribution. In this uniform 1-D case, with no *a priori* correlation between model parameters, the number of free parameters is the same as the actual number of model parameters, $M_{\text{free}} = M = 125$.

$\rho_2(\mathbf{m})$: 1-D minimum entropy. An example of a prior model that contributes a lot of information (that results in a prior model with little entropy) is a three-layer model with constant resistivity within each layer. This can be described completely using the log-resistivity value of three layers (r_1, r_2 and r_3), the depth to the first layer interface (z_1) and the thickness of layer 2 (t_2). r_3 describes the log-resistivity below $z_1 + t_2$. r_2 describes the log-resistivity in the range z_1 to $z_1 + t_2$. r_1 describes the log-resistivity above z_1 . $\rho_2(\mathbf{m})$ can therefore be described completely by five parameters, $\rho_2(r_1, r_2, r_3, z_1, t_2)$.

Table 1 lists the mean and variance of five assumed independent 1-D prior Gaussian distributions, for the five parameters. Note, the 1-D prior distribution for the resistivity in the first and third layer is the same, and *a priori* the resistivity can be higher in layer 2

Figure 4. 100 independent realizations from prior distributions (a) $\rho_1(\mathbf{m})$, (b) $\rho_2(\mathbf{m})$ and (c) $\rho_3(\mathbf{m})$.

Figure 5. 1-D marginal probability distribution for $f_1(\mathbf{m})$, $f_2(\mathbf{m})$ and $f_3(\mathbf{m})$.

than above and below. Also note that $\rho_2(z_1)$ is a truncated Gaussian distribution, such that $\rho_2(z_1) = 0 \forall z_1 < 0$.

A realization of $\rho_2(\mathbf{m})$, $\mathbf{m}^*$, is constructed by generating a realization from $\rho_2(r_1, r_2, r_3, z_1, t_2)$, from which $\mathbf{m}^*$ can be constructed. Fig. 4(b) shows 100 independent realizations from $\rho_2(\mathbf{m})$. Note how the prior assumptions results in both 2- and 3-layer models, and that the base of the second layer typically is located relatively shallow.

Fig. 5 shows the corresponding uniform 1-D marginal distribution, which has a peak around $\log(12.5)$, and a less pronounced peak around $\log(100)$, stemming from the *a priori* mean choices for

Table 1. Mean and standard deviation of five 1-D normal distributions. The ‘min’ column refers to the value of a lower truncation of the normal distribution if used.

ρ	Mean	Std	min
$\rho_2(r_1)$	$\log(12.5)$ [ohm-m]	0.2	–
$\rho_2(r_2)$	$\log(100)$ [ohm-m]	0.5	–
$\rho_2(r_3)$	$\log(12.5)$ [ohm-m]	0.2	–
$\rho_2(z_1)$	5 [m]	10	0
$\rho_2(t_2)$	20 [m]	14.1	–

$\rho_2(r_1)$, $\rho_2(r_2)$ and $\rho_2(r_3)$. In this case, the effective number model of parameters is exactly $M_{\text{free}} = 5$.

$\rho_3(\mathbf{m})$: 1-D ‘realistic’ prior model. From analysis of well logs and other data in the area (Abraham *et al.* 2012) it is expected that three different lithologies can be used to describe the subsurface. Each of these lithologies are expected to be associated with specific log-resistivity value. The 1-D marginal distribution for each of these lithologies are assumed to follow 1-D log-normal distributions with means $[\log(12.5), \log(100), \log(562)]$ and standard deviations $[0.1, 0.15, 0.175]$, respectively.

A spatially correlated Gaussian model (with zero mean, and unit variance) is used to describe spatial correlation, $\mathcal{N}(0, \mathbf{C}_m)$. $\mathbf{C}_m$ is constructed using a Gaussian type covariance model with a range of 30 m to mimic the expected thicknesses of the three lithologies. The specific choice of $\rho_3(\mathbf{m})$ given information from Abraham *et al.* (2012) can naturally be debated, but as discussed, an actual explicit

choice has to be made. One can, in principle, leave the range as a free parameter that can be inferred as part of the inversion, but this will make the sampling problem computationally more challenging (Hansen *et al.* 2013b). One can also rely on hypothesis testing due to using different ranges, prior to inversion, but this again will increase computational demands (Sambridge *et al.* 2006). The problem of choosing a prior model representing actual knowledge is not trivial, and can be dealt with in more detail, as for example, Høyer *et al.* (2017).

An actual realization from $\rho_3(\mathbf{m})$ is obtained by generating a sample from $\mathcal{N}(0, \mathbf{Cm})$, which is then transformed to follow the desired 1-D marginal distribution using normal score transformation, see for example, Goovaerts (1997). Fig. 4(c) shows 100 independent realizations from $\rho_3(\mathbf{m})$. The corresponding trimodal 1-D marginal distribution is illustrated in Fig. 5. Due to the assumption of correlation between model parameters, the number of free parameters is significantly below the number of model parameters, $M_{\text{free}} < 125$, but no exact value of M_{free} can be obtained. See for example, Hansen *et al.* (2009) for a discussion about the number of free parameters in a correlated Gaussian model.

Illustrating independent realizations, such as in Fig. 4 is an example of following the ‘movie strategy’ as discussed by Tarantola (2005). While it may be difficult to appreciate the implications of a specific choice of prior model from, for example, Table 1 (which provides all information about $\rho_2(\mathbf{m})$), actual examples of independent realizations as in Fig. 4 allow easy communication, and analysis of, what prior assumptions in effect mean. Such a figure, or movie of prior realizations should always be presented when probabilistic inversion methods are used.

3.1.3 The likelihood

The noise of the data is initially assumed to be independent uncorrelated zero mean Gaussian noise, with a standard deviation of 5 ppm (parts per million) plus 5 per cent of the data value, such that $\sigma_{d_i} = \sqrt{(0.05 * d_i)^2 + 5^2}$. Hence, the corresponding likelihood becomes

$$L(\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m})) \propto \prod_{i=1}^N \exp \left(-0.5 \frac{(d_{i_{\text{obs}}} - d_i(\mathbf{m}))^2}{\sigma_{d_i}^2} \right) \quad (4)$$

$$= \prod_{i=1}^N \exp \left(-0.5 \frac{(d_{i_{\text{obs}}} - d_i(\mathbf{m}))^2}{(0.05 * d_{i_{\text{obs}}})^2 + 5^2} \right). \quad (5)$$

3.2 Synthetic 1-D case

Initially, a synthetic 1-D case is considered. A reference model, $\mathbf{m}_{\text{ref}}$, is generated as a realization of $\rho_3(\mathbf{m})$. A reference data set, $\mathbf{d}_{\text{ref}}$, is generated by adding a realization of the noise, eq. (5), to the noise-free forward response, $\mathbf{g}(\mathbf{m}_{\text{ref}})$. The forward model $\mathbf{g}$ is the same as used in Minsley (2011). The reference model represents here the ‘true’ model about which we wish to infer information.

Then the reference data set is inverted, by sampling the posterior distribution, eq. (2), for three cases, considering $\rho_1(\mathbf{m})$, $\rho_2(\mathbf{m})$ and $\rho_3(\mathbf{m})$ as the *a priori* model. A sample of the posterior is generated by running the extended Metropolis algorithm for 200 000 iterations using the SIPPI Matlab toolbox (Hansen *et al.* 2013a).¹

When the sampling is initiated, a simulated annealing type sampling is used where each proposed model is accepted with probability $P_{\text{acc}}^{(T_a^{-1})}$, where T_a is an annealing temperature. This is done to avoid potentially sampling secondary areas with high posterior probability. Initially T_a is high ($T_a = 30$), leading to a large exploration of the model space). T_a is gradually decreased to $T_a = 1$ and fixed after 10 per cent of the iterations; hence the acceptance criteria becomes identical to eq. (3).

The current model is perturbed by re-simulating a random set of model parameters from the prior distribution. The number of perturbed model parameters are chosen such that on average around 30 per cent of the proposed models are accepted. After the first 15 per cent of iterations, the perturbation is fixed for the remainder of the simulation. The combination of the annealing schedule and the automatic adjustment of perturbation size provides a robust approach to run the extended Metropolis algorithm unsupervised (without manual interaction) for the cases demonstrated below. The first 15 per cent of the visited models are discarded, and the rest is used to describe the posterior distribution. For more details on the implementation see Hansen *et al.* (2013a).

The reference data, with 95 per cent confidence intervals for the associated noise, are shown together with the forward response of 100 realizations of the three considered prior models in Figs 6(a)–(c). Naturally, the forward response of realizations from the prior distribution does not fit within the uncertainty for most of the any considered prior models. However, using any of the three different prior models, the forward response of the corresponding realizations of the posterior fits within the noise model, as shown in Figs 7(a)–(c).

Figs 8(a)–(c) show 100 independent realizations of the three considered posterior distributions, which should be compared to the 100 independent realizations of the prior distributions in Figs 4(a)–(c). The first model (thick black lines) in Fig. 8 refers to the reference model, $\mathbf{m}_{\text{ref}}$. Corresponding 1-D marginal statistics (mean, mode and 95 per cent confidence interval from both prior and posterior) are computed for every 125 model parameters and shown in Figs 9(a)–(c).

The 100 independent realizations (the sample) of the posterior probability distribution represent ‘the solution’ to the inversion. One benefit of using sampling methods is that resolution of any feature or property of the posterior sample can easily be quantified through simple integration over the posterior distribution. A feature can be, for example, ‘pixel A and pixel B are connected by pixels with a resistivity below 10 ohm-m. One can also analyse the resolution of a single model parameter’. In practice, this amounts to simple counting operations on the sample from the posterior, as discussed in detail by Mosegaard (1998). Thus, if the resolution of a specific feature is high (e.g. the depth to a high resistivity layer), it simply means this feature is known with high certainty, and hence well resolved.

3.2.1 The effect of a maximum entropy prior model, $\rho_1(\mathbf{m})$

A completely un-informed prior model, such as $\rho_1(\mathbf{m})$, leads to a case where it is easy to fit data within its noise (Fig. 7a). At the same time, very little information seems to be contained in the posterior distribution. Fig. 4(a) reflects, as expected, the *a priori* assumption of no correlation between the model parameters, and no spatial features appear well resolved. Fig. 8(a) reflects the information using both the prior model and the observed data. Well-resolved features should appear many times in a sample; however, it is difficult to visually identify any resolved (recurring) features from Fig. 8(a).

¹Code and data are available at <http://github.com/cultpenguin/sippi/examples/>

Figure 6. Observed data, $\mathbf{d}_{\text{ref}}$ with confidence interval, and the data response from 100 realizations of (a) $\rho_1(\mathbf{m})$, (b) $\rho_2(\mathbf{m})$ and (c) $\rho_2(\mathbf{m})$. The black and red lines refer to the real and imaginary part of the signal, respectively.

Figure 7. Observed data, $\mathbf{d}_{\text{ref}}$ with confidence interval, and the data response from 100 realizations of (a) $\sigma_1(\mathbf{m})$, (b) $\sigma_2(\mathbf{m})$ and (c) $\sigma_3(\mathbf{m})$. The black and red lines refer to the real and imaginary part of the signal, respectively. Compare to Fig. 6.

In fact, it is difficult to see significant changes between the sample from the prior distribution $\rho_1(\mathbf{m})$ and the posterior distribution $\sigma_1(\mathbf{m})$. This suggests that very little information can be deemed to stem from the data themselves.

After computing posterior 1-D statistics, Fig. 9(a), it is possible to identify that some information has been obtained in the depth range from 5 to 30 m. Note though that the confidence intervals are very large everywhere in the model.

Figs 7(a), 8(a) and 9(a) show the result of what one could describe as a purely data-driven approach, in the sense that, for the specific choice of parametrization, the posterior becomes

$$\sigma(\mathbf{m}) \propto \rho(\mathbf{m}) L(\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m})) \quad (6)$$

$$\propto L(\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m})), \quad \forall \log(5) < m_i < \log(3000). \quad (7)$$

In other words, the posterior becomes proportional to the likelihood if the prior distribution is uniform, within a certain range.

However, it is important to remember that a uniform prior distribution does not restrict smooth models from being realized. In fact, a model with constant log-resistivity of 10 will have the same prior probability distribution, $\rho_1(\mathbf{m})$, value as any of the generated realizations of the prior distribution, as in Fig. 4(a), or the reference model. In that sense any of these individual models are equally likely to occur. However, the probability of realizing a model with significant spatial continuity is extremely small. If one generates 1000 000 realizations from $\rho_1(\mathbf{m})$, and computes the probability that neighbouring resistivity values do not change more than a factor of 20 (which is very mild smoothness constrain) one gets $P(\Delta m_i = |m_i - m_{i-1}| < \log_{10}(3) \forall i | \rho_1(\mathbf{m})) = 0.000004$. In practice, it will be close to impossible to realize any model with smooth characteristics. The choice of uniform prior distribution favours disorder.

If AEM data are very well informed, which is not the case here, then the use of a uniform prior distribution could be a viable approach that can be described as a data-driven approach. Instead the examples above suggest that one should, in general, avoid the use of a uniform prior distribution. It gives a false impression of assuming nothing, while the uniform prior distribution is in fact very informed with regard to the expectation of no spatial smoothness. At the same time, it leads, in this case, to an inverse problem with little resolution of the posterior probability distribution. AEM data are in this case not strong enough in themselves to infer much information about the subsurface resistivity distribution.

3.2.2 The effect of a minimum entropy prior model, $\rho_2(\mathbf{m})$

Another extreme assumption is given by $\rho_2(\mathbf{m})$ that represents an oversimplified prior model. This model is in fact inconsistent with the reference model, in the sense that the reference model is not a possible realization of the prior distribution, as $\rho_2(\mathbf{m}_{\text{ref}}) = 0$. In a typical Monte Carlo sampling approach, one might discover such inconsistencies through the sampling process, if, for example, it is found difficult to fit the observed data. Here, nothing in the sampling approach suggests that a problematic prior distribution might be used, and specifically it is possible to match the observed data within the uncertainty as shown in Fig. 7(b).

Where the use of $\rho_1(\mathbf{m})$ provided very little resolution of the posterior probability $\sigma_1(\mathbf{m})$, the use of $\rho_2(\mathbf{m})$ provides almost perfect resolution of $\sigma_2(\mathbf{m})$. All the realizations from the posterior in Fig. 8(b), show a very similar 1-D model consisting of two regions of high resistivity in the upper and lower layers, with low resistivity in between. The interface between the two layers also seems well resolved. This visual inspection corresponds well to

Figure 8. 100 independent realizations from posterior distributions (a) $\sigma_1(\mathbf{m})$, (b) $\sigma_2(\mathbf{m})$ and (c) $\sigma_3(\mathbf{m})$.

Figure 9. Point-wise statistics (mean, mode, standard deviation) (a) $\rho_1(\mathbf{m})$ and $\sigma_1(\mathbf{m})$, (b) $\rho_2(\mathbf{m})$ and $\sigma_2(\mathbf{m})$ and (c) $\rho_3(\mathbf{m})$ and $\sigma_3(\mathbf{m})$. 1-D marginal distribution (grey scales), 1-D marginal mode model (red), 1-D marginal mean model (blue), 1-D marginal 95 percent confidence interval (dashed blue) and reference model (dashed black).

the 1-D statistics computed from $\rho_2(\mathbf{m})$ and $\sigma_2(\mathbf{m})$ and presented in Fig. 9(b). These results indicate almost perfect resolution all through the model (down to 125 m depth).

This is problematic, as the actual reference model is associated with zero posterior probability. Specifically, the high resistivity layer (that exists) below 100 m depth, is associated with zero probability, which can lead to wrong decisions.

The problem here is that the prior distribution is far too narrow and restrictive, and in fact in conflict with the real reference model. While the inverse problem has been solved correctly, the prior information therefore leads to biased results. This, combined with the fact that the data themselves are not strong enough to uncover this problem, underlines the importance of not assuming too much *a priori* knowledge.

3.2.3 The effect of a 'realistic' prior model, $\rho_3(\mathbf{m})$

$\rho_3(\mathbf{m})$ represents a more realistic prior model. It is more informed than $\rho_1(\mathbf{m})$, and we know that the reference model is consistent with the prior distribution, as it has been generated from this prior model. In this respect, this prior model is the optimal choice for this particular data set.

It is then no surprise that data can fit within the noise (Fig. 7c). Fig. 8(c) reveals that the top of the model is very well resolved down to the interface at around 40 m depth. Fig. 9(c) further suggests that some resolution exists down to about 60–80 m depth, highlighting the possibility of a resistive layer from 40 to 60 m depth underlain by a layer with resistivity around 10 ohm-m, as in the reference model. There is no resolution below 65 m depth, but still, the reference model is consistent with the posterior probability, in that the reference model is a realistic realization, also below 65 m depth, as opposed to when using $\rho_2(\mathbf{m})$.

In practice, it may be difficult to define a prior model that is as well informed as $\rho_3(\mathbf{m})$ that is also broad enough to be consistent with the actual reference model. If an overly informed prior model is assumed, it may lead to good resolution but may provide biased results. The choice of prior model cannot be avoided by choosing a uniform prior distribution, as this is also a specific choice of maximum disorder that in the present case leads to little resolution of the posterior probability distribution.

3.3 Inversion of 2-D Morrill data

We now consider the data from Morrill, Nebraska, using the same noise model, and hence likelihood, as considered above. The number of model parameters in the 2-D model is $M = 125 \times 451 = 56\,375$.

3.3.1 1-D prior model

Initially, no horizontal correlation between model parameters is assumed, which means exactly the same prior models considered for the synthetic case, $\rho_1(\mathbf{m})$, $\rho_2(\mathbf{m})$ and $\rho_3(\mathbf{m})$, will be considered here. In this 2-D case, with no horizontal correlation between model parameters, this leads to $M_{\text{free}} = M = 56\,375$ for $\rho_1(\mathbf{m})$ and $M_{\text{free}} = 451 \times 5 = 2255$ for $\rho_2(\mathbf{m})$. For $\rho_3(\mathbf{m})$, M_{free} is somewhere in between the two values. Some results are visualized in the form of five-panel figures for each of the three cases (Figs 10–12).

1-D $\rho_1(\mathbf{m})$: Maximum entropy. Fig. 10 shows results related to $\sigma_1(\mathbf{m})$, that is, when using $\rho_1(\mathbf{m})$ as prior information. Fig. 10(a) shows one realization from the prior distribution $\rho_1(\mathbf{m})$, illustrating the information available before data are incorporated into the inversion. For this prior model, no coherent spatial features are apparent. The solution to the inverse problem is a number of realizations (i.e. a sample) of the posterior probability distribution $\sigma_1(\mathbf{m})$, of which Fig. 10(b) shows one. It shows some evidence of coherent resistivity values in the near surface. However, neither the realization from the prior nor the posterior distribution seems to reflect a geologically realistic scenario.

Figs 10(c)–(e) show three statistics from the posterior distribution, obtained from the samples of $\sigma_1(\mathbf{m})$. Fig. 10(c) shows the point-wise 'mode' model, that is, the log-resistivity value that maximizes $\sigma_1(m_i)$ for any m_i . Fig. 10(d) shows the standard deviation of the 1-D marginal posterior distribution, $\sigma_1(m_i)$. Note the colourbar has a maximum of 0.6, even though the maximum standard deviation for this case is around 1. This is chosen for easier comparison between subsequent considered cases, for which the maximum of 0.6 is kept constant. Fig. 10(e) shows the Kullback–Leibler (KL)

distance between $\sigma_1(m_i)$ and $\rho_1(m_i)$, as $D_{\text{KL}}(\rho_1(m_i), \sigma_1(m_i))$, which is a measure of how much information is gained using the posterior $\sigma_1(m_i)$ as opposed to the prior distribution $\rho_1(m_i)$, see e.g. Kullback & Leibler (1951) and Cover & Thomas (2012). A small KL distance then suggests that little extra information is available through the posterior and vice versa.

It can be tempting to use the mode model, Fig. 10(c), or similar statistics such as the point-wise mean and median, as the 'solution' to the inverse problem. Recall that these models are 'statistics' of the posterior, which, in general, does not represent a model consistent with available information. The mode model may not be consistent with either the prior distribution or the data. The mode model averages out the pixels that vary across realizations and highlights the pixels that tend to be consistent between realizations. At the base of the model, which mostly reflect the prior information, it is clear that there is no preferred mode model, which can be seen in that no coherent areas of mode values can be found. This is also discussed in Christensen *et al.* (2017). It is extra problematic using the model as the 'solution' if there are indeed several modes in the posterior. Sampling methods, as considered here, would identify such different modes (which typically reflect different geological scenarios), but using only the most probable mode as a solution would ignore the existence of other modes, which would lead to an underestimation of uncertainty. The mode model can assist in an interpretation of the results, especially in the top part of the model, where a significant mode in the 1-D marginal is present.

The standard deviation, Fig. 10(d), reveals that most of the model parameters are not very well constrained for this prior model. Only in the top-left part of the model around $x = 2\text{--}4$ km do the model parameters seem to be resolved.

The KL distance provides an alternative approach to illustrate the information content of the data in relation to reducing the posterior parameter uncertainty. For example, the resistivity within the apparently high-resistive channel-like area in the shallow part of the model around $x = 5\text{--}8$ km (Fig. 10b) is in fact not that well resolved, as the KL distance is close to zero. Within the channel-like structure, each individual model parameter can have both high and low resistivity, as seen on the posterior realization (Fig. 10b). The reason that the channel-like area stands out (Fig. 10b) is because the surrounding low-resistivity areas are quite well resolved and less because the model parameters within the channel feature itself are particularly well resolved.

1-D $\rho_2(\mathbf{m})$: Minimum entropy. Fig. 11 shows similar figures to Fig. 10, except for the case considering $\rho_2(\mathbf{m})$ as the prior model.

The three-layer model is clearly visible from the realization of the prior model $\rho_2(\mathbf{m})$ (Fig. 11a) but lacks any horizontal correlation, suggesting little resemblance with any type of realistic geology. There is some noise/scattering between the 1-D resistivity profiles from the realization from the posterior $\sigma_2(\mathbf{m})$ (Fig. 11b). However, there is much more geologically realistic features compared with the maximum entropy model. As the prior models contain no information about horizontal correlation, any such observed features in the posterior sample, come strictly from the data.

The mode model (Fig. 11c) does not differ much from the realization from the posterior (Fig. 11b) except that it has slightly less small-scale variation.

The standard deviation, Fig. 11(d), reveals that most of the uncertainty about the model parameters is associated with the location of the boundary at the base of the resistive layer. The KL distance, Fig. 11(e), shows that significant information is gained from the posterior distribution relative to the prior distribution, almost everywhere in the model.

Figure 10. Results for the maximum entropy model. (a) Realization from $\rho_1(\mathbf{m})$ (the maximum entropy model). (b) Realization from $\sigma_1(\mathbf{m})$. (c) Point-wise mode model from $\sigma_1(\mathbf{m})$. (d) Point-wise standard deviation from $\sigma_1(\mathbf{m})$. (e) Point-wise Kullback-Leibler distance between $\sigma_1(\mathbf{m})$ and $\rho_1(\mathbf{m})$.

Thus, the posterior variability is very small, which suggests that the model parameters are well resolved, and the standard deviation and KL distance suggest an almost perfectly resolved model, except for the layer boundary, also at the very deep part of the model. Recall from the synthetic case above, however that such a well-informed posterior distribution is only valid if the chosen prior model is valid. In this case, the prior model contains too strong of a constraint, which we suggest is in fact inconsistent with the real subsurface, and results in biased estimates of posterior uncertainty. In the present context, we would not suggest using these results in practice.

1-D $\rho_3(\mathbf{m})$: ‘Realistic’ entropy. Fig. 12 shows the result using a more geologically realistic prior model $\rho_3(\mathbf{m})$. The realization from the prior distribution, Fig. 12(a), again indicates that the assumption of no horizontal correlation is probably not geologically sound. But, as when considering $\sigma_2(\mathbf{m})$, realizations from $\sigma_3(\mathbf{m})$ seem to provide a horizontally constrained upper conductive layer on top of a resistive layer, down to about 60 m depth, Fig. 12(b). Below that

depth, the effect of the horizontally unconstrained prior distribution is obvious.

The mode model, Fig. 12(c), looks remarkably sharp, with clear spatial correlation everywhere in the model. However, since the 1-D marginal prior distribution is trimodal with highest probability related to the low resistivity, then one should expect this value as the point-wise mode, anywhere data have no effect.

The standard deviation and KL distance both reveal that the main uncertainty is related to the location of the boundary between layers at depths less than 50–60 m, while there is diminishing resolution below that depth, effectively providing a depth of investigation (DOI) metric (Figs 12d–e).

3.3.2 2-D prior model

The realizations from the prior models in Figs 10(a), 11(a) and 12(a) suggest that the assumption of no spatial correlation is not consistent

Figure 11. Results for the minimum entropy model. As in Fig. 10 but for (a) $\sigma_2(\mathbf{m})$ and (b–e) $\rho_2(\mathbf{m})$.

with realistic geology. Realizations of the corresponding posterior probability densities do show some horizontal correlations, but only where model parameters are sensitive to the data, not deep in the model. Hence, two 2-D prior models are considered, based on $\rho_2(\mathbf{m})$ and $\rho_3(\mathbf{m})$, with additional horizontal constraints.

The same algorithm is used to sample the posterior distributions as for the 1-D case, except that one sample from a 2-D prior distribution at each iteration, and the extended Metropolis sampler is run for 10^6 iterations. Also to limit the numerical cost, we consider only every fourth of the total of 451 available soundings.

$\rho_{2D}(\mathbf{m})$: 2-D Minimum entropy. Recall that five parameters perfectly describe any 1-D set of model parameters through $\rho_2(\mathbf{m})$. A 2-D prior model is constructed, $\rho_{2D}(\mathbf{m})$, assuming that both $\rho_2(z_1)$ and $\rho_2(t_1)$ are correlated along the X-axis, following a Gaussian type covariance model with range of 30 m, and a variance of 10 and 14 m, respectively. These values are chosen to specifically design a ‘simple’ low-entropy prior model.

The log-resistivity r_1 , r_2 and r_3 is assumed constant along the profile with the same prior distribution as considered above. This

means that only three parameters are needed to describe the resistivity values. In addition, z_1 and t_1 are assumed to be correlated along the x -axis, suggesting that the number of free model parameters is much lower than when considering no horizontal correlation, hence $M_{\text{free}} < 2255$, that is, a small fraction of the total number of model parameters (56 375).

Fig. 13 shows the results obtained by using data from every fourth sounding, that is, 108 of the 451 available soundings. Fig. 13(a) shows one realization from the prior distribution $\rho_{2D}(\mathbf{m})$, where it is explicitly clear that horizontal correlation is now a part of the prior assumption. The realization of the posterior $\sigma_{2D}(\mathbf{m})$ (Fig. 13b) shows different resistivity values and layer boundaries than the realization from the prior distribution, but looks similar with respect to the spatial patterns and structures. The mode model, Fig. 13(c), looks very similar to the realizations from the posterior, and the point-wise standard deviation, Fig. 13(d), is very low except at the boundary between the layers.

However, the posterior distribution is not yet sampled, even after 10^6 iterations, as the fit to data is ever increasing, and never reaches

Figure 12. Results for the ‘realistic’ entropy model. As in Fig. 10 but for (a) $\sigma_3(\mathbf{m})$ and (b–e) $\rho_3(\mathbf{m})$.

a plateau of constant values. Hence, the Metropolis sampler has not yet ‘burned in’. Instead, it constantly struggles to marginally perturb a realization of the prior distribution to find an ever better fit to the data. The reason for this is typically that some inconsistency is present between the choice of prior model and the likelihood (describing measurement and modelling errors), in which case the sampling problem can become intractable.

One possibility is that the prior model may be too simple to allow fitting the data within the noise. In this case, the sampling problem turns into an optimization problem of finding the one realization of the prior distribution that fits data the best (but still in this case does not fit data according to the noise). Another possibility is that the noise model, eq. (5), may not be correct. While widely used, the assumption of uncorrelated noise may be too simple, as in practice data from one sounding may be contaminated by correlated measurement and modelling errors, see for example, Madsen & Hansen (2018) for investigation on modelling errors for seismic data. In addition, the noise may be correlated from sounding to sounding.

For the present prior model, $\rho_{2D}(\mathbf{m})$, the prior model may have been chosen to be too simplistic as it has only very few free parameters. If the prior model was in fact consistent with available data, this should render an ‘easy’ sampling problem, as the prior model only has very few free parameters which should be easy to determine.

If one insists that the prior model is well chosen, then one can instead try to select a more appropriate noise model. A simple, but in practice useful approach, is to increase the standard deviation of the assumed noise model. Fig. 14 shows the point-wise standard deviation obtained for two cases with $\sigma_{d_i} = 10$ and $\sigma_{d_i} = 20$, that is, two and four times the standard deviation assumed in eq. (5). In both cases, the posterior was efficiently sampled, leading to about 100 and 160 independent realizations from the posterior distribution in the 10^6 iterations, or about 10 000 and 6000 iterations between independent realizations (see Table 2). In addition, the posterior variability is very low even using two times the assumed standard deviation for the noise (Fig. 14a). Thus, increasing the magnitude of

Figure 13. Result using the minimum entropy prior model in 2-D. As in Fig. 11 but for (a) $\rho_{2D}(\mathbf{m})$ and (b–e) $\sigma_{2D}(\mathbf{m})$.

the expected noise changed a difficult (even inconsistent) sampling problem to a manageable problem.

$\rho_{32D}(\mathbf{m})$: 2-D ‘realistic’ prior model. $\rho_{32D}(\mathbf{m})$ is similar to $\rho_3(\mathbf{m})$, except that a horizontal correlation length of 1.5 km is assumed (instead of no correlation). As when considering the 2-D minimum entropy prior model above, using the original noise level of eq. (5), leads to an optimization problem that suggests inconsistency between the choice of prior and the noise model. Instead, we double the assumed standard deviation of the Gaussian noise to $\sigma_{d_i} = 10$. This leads to a sampling problem in which about 13 independent realizations are obtained in 10^6 iterations, with about 80 000 iterations between independent realizations (see Table 2).

Fig. 15 shows the results obtained by using data from every fourth sounding, that is, 108 of the 451 available soundings. Fig. 15(a) shows a realization of this prior model, and Fig. 15(b) a realization of the corresponding posterior, $\sigma_{32D}(\mathbf{m})$.

The first thing to notice is how the structures in the realizations of the prior and posterior distributions look similar. The resolution in the top and bottom of the model is different, but the same type of spatial variability is seen in both the top and the bottom of the posterior realization (Fig. 15b). Using any of the prior models with no spatial correlation, the realization from the posterior has significantly different characteristics in the top and bottom part of the model, suggesting a conflict between the data and the prior information. This is not the case here suggesting there is no conflict with the prior information and the data. The mode model, Fig. 15(c), lacks realistic variability in the deeper parts of the model, due to the lack of resolution as can be seen from the high standard deviation and small KL distance below 50–60 m depth (Figs 15d,e).

Fig. 16 is similar to Fig. 15, except we use a noise model with $\sigma_{d_i} = 20$, or four times the original standard deviation. This leads

Figure 14. Point-wise posterior standard deviation, as Fig. 13(d), but using a noise model with (a) $\sigma_{d_i} = 10$, and (b) $\sigma_{d_i} = 20$ (two and four times the initially assumed standard deviation on the noise).

Table 2. Computational demands for sampling a specific posterior distributions using an Intel Core i7-8650U CPU running at 2.11 GHz. σ_{d_i} refer to the standard deviation of the Gaussian noise on data. N_{ind} is the number of iterations between independent realizations. t_{ite} is the time in ms it takes to perform one iteration of the extended Metropolis sampler. t_{ind} is the time in minutes per independent realization of the posterior. $\log(L)$ is the average log-likelihood for realizations of the posterior (for the cases with a 2-D prior distribution).

$\sigma(\mathbf{m})$	σ_{d_i}	N_{ind}	t_{ite} (ms)	t_{ind} (m)	$\log(L)$
$\sigma_1(\mathbf{m})$ 1D	5	8.4	400	0.056	—
$\sigma_2(\mathbf{m})$ 1D	5	8.4	400	0.056	—
$\sigma_3(\mathbf{m})$ 1D	5	8.4	200	0.028	—
$\sigma_{22D}(\mathbf{m})$	5	∞	—	—	—
$\sigma_{22D}(\mathbf{m})$	10	10 000	650	108	−620
$\sigma_{22D}(\mathbf{m})$	20	6000	650	65.0	−190
$\sigma_{32D}(\mathbf{m})$	5	∞	—	—	—
$\sigma_{32D}(\mathbf{m})$	10	80 000	650	867	−280
$\sigma_{32D}(\mathbf{m})$	20	44 000	650	477	−125

to an easier sampling problem in which about 23 independent realizations are obtained in 10^6 iterations, with about 44 000 iterations between independent realizations (see Table 2). The decrease in resolution compared to using $\sigma_{d_i} = 10$, Fig. 15, may be small enough that one may choose the more efficient sampling using a higher noise model to a more demanding sampling algorithm using a smaller noise model. This choice will be dependent on the specific use of the obtained posterior sample.

Table 2 shows the average log-likelihood (from eq. 5) for realizations of the posterior in cases using the prior models $\rho_{22D}(\mathbf{m})$ and $\rho_{32D}(\mathbf{m})$ and is determined by a combination of the choices of noise and prior model. It illustrates that naturally it is easier to fit the data within the noise, when the noise level is higher (as in $\sigma_{d_i} = 20$). But it also illustrates that when using the low entropy prior model $\rho_{22D}(\mathbf{m})$, the data fit is not as good as when using $\rho_{32D}(\mathbf{m})$, which suggests the low entropy prior model is less consistent with the data (and the noise model) than $\rho_{32D}(\mathbf{m})$.

4 DISCUSSION

A key benefit of the probabilistic approach is that it allows different types of information to be integrated seamlessly, without requiring a geological expert to be an expert in geophysical data too, and vice versa. Further, analysis of the inversion result (i.e. the realizations

from the posterior distribution) can be analysed by a person that is not an expert in either geology or geophysics. It is enough if one knows what question to ask, such as ‘what is the volume of coherent blocks of high resistivity?’ One can simply run through the posterior realizations and make a histogram of the computed volume for each realization. The uncertainty (the width/shape of the histogram) will then by construction be consistent with available information (here, geology and geophysics).

Further, all realizations from the posterior can be transformed into other parameters, such as hydrological parameters, $\mathbf{n}$, while maintaining consistency to the prior model and AEM data, as long as a function exists to allow mapping from log-resistivity, $\mathbf{m}$, into the parameter $\mathbf{n}$. This can be in the form of a probabilistic, $f(\mathbf{n}|\mathbf{m})$, or deterministic $\mathbf{n} = h(\mathbf{m})$ relationship. Again, the expert setting up, for example $f(\mathbf{n}|\mathbf{m})$, need not be an expert in either geology or geophysics. It is enough that she can handle the set of realizations from the posterior.

This is the key benefit of using a probabilistic approach. Ideally, it can empower decision makers to take informed decision based on all available information, which would be extremely difficult to achieve using a sequential set of deterministic inversions as discussed in relation to Fig. 1.

The least-squares based approaches are much faster and provide a single optimal model as the least-squares solution in Fig. 3 that

Figure 15. Results using the ‘realistic’ entropy prior model in two 2-D. As in Fig. 10 but for (a) $\sigma_{32D}(\mathbf{m})$ and (b–e) $\rho_{32D}(\mathbf{m})$ with $\sigma_{d_i} = 10$.

convey similar information to the mode model in, for example, Fig. 15(c). Whether such a deterministic modelling is enough to take informed decisions depends heavily on the problem at hand and on the skills of the decision maker.

4.1 Computational demands

Solving inverse problems with a sampling approach is typically related to large computational costs. The computational costs for the examples given above is summarized in Table 2.

4.1.1 1-D prior model / 1-D forward

Generating a realization of a 1-D prior model takes about 1 ms. Solving the forward problem for one sounding takes around 6 ms.

Sampling the prior model takes about 1 ms. Evaluation of the likelihood takes around 0.06 ms. Each iteration of the extended Metropolis sampler takes on average 8.4 ms.

On average, around 200–400 iterations are needed to obtain an independent realization of the posterior distribution when sampling $\sigma_1(\mathbf{m})$, $\sigma_2(\mathbf{m})$ and $\sigma_3(\mathbf{m})$. This is computed by analysing the correlation coefficient between one posterior realization and all other posterior realizations as in Hansen *et al.* (2013b). Thus, on average it takes 1.7 to 3.4 s to generate an independent realization, and 100 independent realizations from the 1-D posterior distributions can be obtained in less than 6 min.

Obviously such computation times are much larger than what is needed by deterministic inversion methods (Tølbøll & Christensen 2006; Christensen & Dodds 2007; Auken *et al.* 2014).

Still around 6 min for 1-D probabilistic inversion of a sounding is certainly doable. Especially considering the inversion problem is

Figure 16. As in Fig. 15 but with noise level $\sigma_{d_i} = 20$.

‘embarrassingly parallel’ as each 1-D inversion can be done independently of each other.

4.1.2 2-D

Generating a realization of a 2-D prior model takes about 32 ms. Solving the forward problem for one sounding takes around 6 ms per sounding, which leads to around 2.6 s using all soundings. Evaluation of the likelihood takes around 0.13 ms. Each iteration of the extended Metropolis algorithm takes around 2.6 s considering all data, 0.65 s considering every fourth sounding.

It takes between 1 and 15 hr to generate an independent realization from the posterior distribution when using the 2-D prior models considered here, as summarized in Table 2. Thus, as is, the present sampling methodology is not currently viable for large 3-D cases or large 2-D profiles. However, in select zones where specific prior

information is available, or where one is particularly interested in exploring the full posterior uncertainty the methodology is viable.

One way to reduce the computational cost is to solve the forward problem in parallel, which should be doable, as each 1-D forward model can be evaluated independently. In addition, for each iteration only a fraction of the model is perturbed, and hence, one need only evaluate the forward model, for those data sets that are sensitive to the local perturbations. Both these suggestions will provide identical results, but with significantly less computational time. We leave this for future research.

Quantification of modelling errors can potentially also have significant impact on computational costs. As discussed previously, the most widely assumed noise model is that of uncorrelated Gaussian noise, as in eq. (5). However, in many cases uncorrelated noise have been filtered from geophysical data (Madsen & Hansen 2018) which means that actual error will appear as correlated noise. In addition, modelling errors due to the use of inaccurate system description

(Christiansen *et al.* 2011), approximate forward models (Arridge *et al.* 2006; Hansen *et al.* 2014; Köpke *et al.* 2018), may lead to significant correlated errors. If these are not taken into account, it may lead to fitting part of the noise as data, which leads to biased results, and which may render the sampling problem intractable as it may end up being in practice an optimization problem.

4.3 Choosing a prior model explicitly

The main challenge, compared to most other methods for AEM inversion, is that a prior model needs to be explicitly chosen, independently of the available data and choice of algorithm.

As difficult as it may be, the simple 1-D synthetic case also makes it clear that the choice of the prior model (and parametrization) is a choice that cannot be avoided.

The choice of a simple non-informative prior model (such as a uniform prior model) may be viewed as a conservative one. However, as demonstrated above, a uniform prior model will also imply that spatially uncorrelated models are much more probable than smooth models. In addition, unless the data are strongly sensitive to model parameters, the use of a uniform prior model may lead to a posterior with very little resolution. See for example, Scales & Tenorio (2001) for a discussion on the use of uninformative prior models.

Too little prior information may lead to poor posterior resolution (consistent with the actual target model), while a lot of prior information will lead to high posterior resolution, but perhaps biased results. A good guideline is to try to quantify *a priori* what is in fact known, but nothing more. Any hard constraints chosen *a priori* (such as here the use of a layered model with a maximum of three layers with constant resistivity) will also be enforced on the posterior, which can cause serious bias.

It can be argued that the use of this probabilistic approach involves an extra complexity in that a prior model has to be explicitly chosen. However, the problem with choosing the prior model is not unique to this probabilistic approach. As the inversion of AEM data is an under-determined inverse problem, some prior information is always needed when using either deterministic or probabilistic methods. In practice, the prior model is most often an implicit choice guided by the choice of inversion methods or by adjusting some regularization parameters related to a specific choice of inversion algorithm. Here, we have argued to explicitly choose the prior model, and then use an inversion methodology that allows utilizing this information. The choice of assumed prior information should in principle never be guided by the specific algorithm at hand (which typically is associated with an implicit mathematically defined prior model), but instead by available information.

4.4 Depth of investigation and the choice of the prior model

The concept of depth of investigation (DOI) is widely used to assess the validity of resistivity models inverted based on EM data. One definition of the DOI is that it defines the depth below which AEM data are insensitive to the distribution of resistivity (Oldenburg & Li 1999; Vest Christiansen & Auken 2012; Asch *et al.* 2015). This definition, the theoretical DOI, is unrelated to the type inversion algorithm and prior information used. In practice, however, the DOI is often used as the depth below which the inverted resistivity model is unresolved, which we refer to as the practical DOI. This is, for example, the case when Oldenburg & Li (1999) make use of multiple inversion to estimate the DOI, and when Vest Christiansen

& Auken (2012) define the DOI as a measure of the depth at which the model can be considered reliable. Minsley (2011) defines the DOI as the depth where the variability of the posterior distribution is under a specific level.

The DOI is often used to either blank out or fade out the colours, when visualizing inverted resistivity sections, to convey to the user which part of the model parameter space is well resolved. In this context, the practical DOI is more useful than the theoretical DOI, as the practical DOI acknowledges that prior information contributes information that affect resolution.

The results described above clearly demonstrate that the resolution and hence the practical DOI is intimately linked to the choice of a prior model. Both 1-D marginal standard deviation of the posterior and the 1-D marginal KL distance can be used to investigate the DOI directly from a sample of the posterior.

For example, using an uninformed prior model, $\rho_1(\mathbf{m})$, Figs 10(d) and (e) suggest no resolution below 20 m depth, hence DOI = 20 m. This measure of the DOI is perhaps most similar to the theoretical DOI discussed above, as the prior model is uninformative.

Using the minimum entropy prior model, $\rho_2(\mathbf{m})$ and $\rho_{2D}(\mathbf{m})$, Figs 11(d) and (e) and 14 suggest resolution down to the deepest model parameters, hence DOI = 150 m, with the uncertainty only associated to the location of a layer interface. This of course is only valid if the choice of prior information is valid, which is most certainly not the case here.

The more realistic type prior model, $\rho_3(\mathbf{m})$ and $\rho_{3D}(\mathbf{m})$, results in a practical DOI of about 50 m, Figs 12(d) and (e), 15 and 16. Clearly, using the theoretical DOI obtained using the uniform prior model $\rho_1(\mathbf{m})$, and applying it to the inversion results with the realistic prior model, $\sigma_3(\mathbf{m})$ and $\sigma_{3D}(\mathbf{m})$, will lead to ignoring well-resolved features at intermediate depths. We therefore suggest using the practical DOI when analysing inversion results.

5 CONCLUSIONS

We have presented a general approach to invert AEM data in a probabilistic formulation where a prior model must be explicitly chosen. This approach allows integration of information from different sources such as (structural) geology and AEM data simultaneously. Further, the solution to the inverse problem is a collection of models (a sample from the posterior distribution) that allows detailed uncertainty and risk assessment consistent with all used information without losing information and without underestimating uncertainty. It also allows further propagation of the uncertainty into other physical parameters of interest.

We argue that the choice of prior information should not be left to be chosen by whatever inversion methodology is available. Instead, prior information should be chosen explicitly to best represent the prior geological knowledge at hand. Then, the inversion problem should be solved using the chosen method such as, for example, the sampling approach proposed here that can consider the actual prior information and data available.

ACKNOWLEDGEMENTS

This work is funded by the Danish Free Research Council, project 717-00160B, and Innovation Fund Denmark, project 7039-00006B. Contributions from B. Minsley were supported by the USGS Mineral Resources Program through the Interdisciplinary Methods and Applications in Geophysics (IMAGe) project. Any use of trade, product or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

REFERENCES

- Abraham, J.D., Cannia, J.C., Bedrosian, P.A., Johnson, M.R., Ball, L.B. & Sibray, S.S., 2012. *Airborne Electromagnetic Mapping of the Base of Aquifer In Areas of western Nebraska*, U.S. Geological Survey Scientific Investigations Report 2011-5219, p. 38.
- Arridge, S., Kaipio, J., Kolehmainen, V., Schweiger, M., Somersalo, E., Tarvainen, T. & Vauhkonen, M., 2006. Approximation errors and model reduction with an application in optical diffusion tomography, *Inverse Probl.*, **22**(1), 175, doi:10.1088/0266-5611/22/1/010.
- Asch, T., Abraham, J. & Irons, T., 2015. A discussion on depth of investigation in geophysics and aem inversion results, in *SEG Technical Program Expanded Abstracts 2015*, pp. 2072–2076, Society of Exploration Geophysicists.
- Auken, E. & Christiansen, A.V., 2004. Layered and laterally constrained 2D inversion of resistivity data, *Geophysics*, **69**(3), 752–761.
- Auken, E. *et al.*, 2014. An overview of a highly versatile forward and stable inverse algorithm for airborne, ground-based and borehole electromagnetic and electric data, *Explor. Geophys.*, **46**(3), 223–235.
- Auken, E., Boesen, T. & Christiansen, A.V., 2017. A review of airborne electromagnetic methods with focus on geotechnical and hydrological applications from 2007 to 2017, *Adv. Geophys.*, **58**, 47–93.
- Brodie, R. & Sambridge, M., 2006. A holistic approach to inversion of frequency-domain airborne EM data, *Geophysics*, **71**(6), G301–G312.
- Brodie, R.C. & Sambridge, M., 2012. Transdimensional Monte Carlo inversion of AEM data, *ASEG Extended Abstracts*, **2012**(1), 1–4.
- Christensen, N.B., 2002. A generic 1-D imaging method for transient electromagnetic data, *Geophysics*, **67**(2), 438–447.
- Christensen, N.B. & Dodds, K., 2007. 1D inversion and resolution analysis of marine CSEM data, *Geophysics*, **72**(2), WA27–WA38.
- Christensen, N.K., Minsley, B.J. & Christensen, S., 2017. Generation of 3-D hydrostratigraphic zones from dense airborne electromagnetic data to assess groundwater model prediction error, *Water Resour. Res.*, **53**(2), 1019–1038.
- Christiansen, A.V., Auken, E. & Viezzoli, A., 2011. Quantification of modeling errors in airborne tem caused by inaccurate system description, *Geophysics*, **76**(1), F43–F52.
- Constable, S.C., 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.
- Cover, T.M. & Thomas, J.A., 2012. *Elements of Information Theory*, John Wiley & Sons.
- Cox, L.H., Wilson, G.A. & Zhdanov, M.S., 2010. 3D inversion of airborne electromagnetic data using a moving footprint, *Explor. Geophys.*, **41**(4), 250–259.
- Cox, L.H., Wilson, G.A. & Zhdanov, M.S., 2012. 3D inversion of airborne electromagnetic data, *Geophysics*, **77**(4), WB59–WB69.
- Deutsch, C.V. & Journel, A.G., 1992. *GSLIB: Geostatistical Software Library and User's Guide*, Oxford University Press.
- Farquharson, C.G., Oldenburg, D.W. & Routh, P.S., 2003. Simultaneous 1D inversion of loop–loop electromagnetic data for magnetic susceptibility and electrical conductivity, *Geophysics*, **68**(6), 1857–1869.
- Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*, *Applied Geostatistics*, Oxford University Press.
- Grayver, A.V., Streich, R. & Ritter, O., 2013. Three-dimensional parallel distributed inversion of CSEM data using a direct forward solver, *Geophys. J. Int.*, **193**(3), 1432–1446.
- Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, **82**(4), 711–732.
- Hansen, T., Cordua, K., Looms, M. & Mosegaard, K., 2013a. SIPPI: a Matlab toolbox for sampling the solution to inverse problems with complex prior information: Part 1, methodology, *Comput. Geosci.*, **52**, 470–480.
- Hansen, T., Cordua, K., Looms, M. & Mosegaard, K., 2013b. SIPPI: a Matlab toolbox for sampling the solution to inverse problems with complex prior information: Part 2, Application to cross hole GPR tomography, *Comput. Geosci.*, **52**, 481–492.
- Hansen, T.M., Mosegaard, K. & Cordua, K.C., 2008. Using geostatistics to describe complex a priori information for inverse problems, in *VIII International Geostatistics Congress*, Vol. **1**, pp. 329–338, Mining Engineering Department, University of Chile.
- Hansen, T.M., Mosegaard, K. & Cordua, K.S., 2009. Reducing complexity of inverse problems using geostatistical priors, in *Proceedings of IAMG*, Vol. **9**, Stanford.
- Hansen, T.M., Cordua, K.C. & Mosegaard, K., 2012. Inverse problems with non-trivial priors - efficient solution through sequential Gibbs sampling, *Comput. Geosci.*, **16**(3), 593–611.
- Hansen, T.M., Cordua, K.S., Jacobsen, B.H. & Mosegaard, K., 2014. Accounting for imperfect forward modeling in geophysical inverse problems—exemplified for crosshole tomography, *Geophysics*, **79**(3), H1–H21.
- Hansen, T.M., Cordua, K.S., Zunino, A. & Mosegaard, K., 2016a. Probabilistic integration of geo-information, in *Integrated Imaging of the Earth: Theory and Applications*, Vol. **218**, pp. 93, eds Moorkamp M., Lelièvre P.G., Linde N., Khan A., American Geophysical Union.
- Hansen, T.M., Vu, L.T. & Bach, T., 2016b. MPSLIB: a C++ class for sequential simulation of multiple-point statistical models, *SoftwareX*, **5**, 127–133.
- Hastings, W., 1970. Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, **57**(1), 97–109.
- Hauser, J., Gunning, J. & Annetts, D., 2015. Probabilistic inversion of airborne electromagnetic data under spatial constraints, *Geophysics*, **80**(2), E135–E146.
- Hauser, J., Gunning, J. & Annetts, D., 2016. Probabilistic inversion of airborne electromagnetic data for basement conductors, *Geophysics*, **81**(5), E389–E400.
- Hawkins, R. & Sambridge, M., 2015. Geophysical imaging using trans-dimensional trees, *Geophys. J. Int.*, **203**(2), 972–1000.
- Hawkins, R., Brodie, R.C. & Sambridge, M., 2018. Trans-dimensional Bayesian inversion of airborne electromagnetic data for 2D conductivity profiles, *Explor. Geophys.*, **49**(2), 134–147.
- Høyer, A.-S., Vignoli, G., Hansen, T.M., Vu, L.T., Keefer, D.A. & Jørgensen, F., 2017. Multiple-point statistical simulation for hydrogeological models: 3-d training image development and conditioning strategies, *Hydrol. Earth Syst. Sci.*, **21**(12), 6069–6089.
- Köpke, C., Irving, J. & Elsheikh, A.H., 2018. Accounting for model error in Bayesian solutions to hydrogeophysical inverse problems using a local basis approach, *Adv. Water Resour.*, **116**, 195–207.
- Kullback, S. & Leibler, R.A., 1951. On information and sufficiency, *Ann. Math. Stat.*, **22**(1), 79–86.
- Lane, R., Green, A., Golding, C., Owers, M., Pik, P., Plunkett, C., Sattel, D. & Thorn, B., 2000. An example of 3D conductivity mapping using the tempest airborne electromagnetic system, *Explor. Geophys.*, **31**(1/2), 162–172.
- Legault, J.M., 2015. Airborne electromagnetic systems—state of the art and future directions, *CSEG Recorder*, **40**(6), 38–49.
- Madsen, R.B. & Hansen, T.M., 2018. Estimation and accounting for the modeling error in probabilistic linearized amplitude variation with offset inversion, *Geophysics*, **83**(2), N15–N30.
- Malinverno, A., 2002. Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem, *Geophys. J. Int.*, **151**(3), 675–688.
- Mariethoz, G., Renard, P. & Straubhaar, J., 2010. The direct sampling method to perform multiple-point geostatistical simulations, *Water Resour. Res.*, **46**(11), 1–14.
- McCracken, K., Oristaglio, M. & Hohmann, G., 1986. A comparison of electromagnetic exploration systems, *Geophysics*, **51**(3), 810–818.
- Menke, W., 2012. *Geophysical Data Analysis: Discrete Inverse Theory*, Vol. 45, Academic Press.
- Metropolis, N., Rosenbluth, M., Rosenbluth, A., Teller, A. & Teller, E., 1953. Equation of state calculations by fast computing machines, *J. Chem. Phys.*, **21**, 1087–1092.
- 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.
- Mosegaard, K., 1998. Resolution analysis of general inverse problems through inverse Monte Carlo sampling, *Inverse probl.*, **14**(3), 405, doi:10.1088/0266-5611/14/3/004.

- Mosegaard, K. & Hansen, T.M., 2016. Inverse methods: problem formulation and probabilistic solutions, in *Integrated Imaging of the Earth: Theory and Applications, Geophysical Monograph*, Vol. **218**, pp. 9–27, eds Moorkamp M., Lelièvre P.G., Linde N., Khan A., American Geophysical Union.
- Mosegaard, K. & Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems, *J. geophys. Res.*, **100**(B7), 12 431–12 447.
- Oldenburg, D.W. & Li, Y., 1999. Estimating depth of investigation in dc resistivity and IP surveys, *Geophysics*, **64**(2), 403–416.
- Sambridge, M., Gallagher, K., Jackson, A. & Rickwood, P., 2006. Trans-dimensional inverse problems, model comparison and the evidence, *Geophys. J. Int.*, **167**(2), 528–542.
- Scales, J.A. & Tenorio, L., 2001. Prior information and uncertainty in inverse problems, *Geophysics*, **66**(2), 389–397.
- Siemon, B., Christiansen, A.V. & Auken, E., 2009. A review of helicopter-borne electromagnetic methods for groundwater exploration, *Near Surf. Geophys.*, **7**(5-6), 629–646.
- Smith, B.D., Abraham, J.D., Cannia, J.C., Minsley, B.J., Deszcz-Pan, M. & Ball, L.B., 2010. Helicopter Electromagnetic and Magnetic Geophysical Survey Data, Portions of the North Platte and South Platte Natural Resources Districts, Western Nebraska, *U.S. Geological Survey Scientific Investigations Report 2010-1259*, p. 33.
- Sørensen, K.I. & Auken, E., 2004. SkyTEM—a new high-resolution helicopter transient electromagnetic system, *Explor. Geophys.*, **35**(3), 194–202.
- Strebelle, S., 2002. Conditional simulation of complex geological structures using multiple-point statistics, *Math. Geol.*, **34**(1), 1–21.
- Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*, Elsevier.
- Tarantola, A. & Valette, B., 1982a. Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. Geophys. Space Phys.*, **20**(2), 219–232.
- Tarantola, A. & Valette, B., 1982b. Inverse problems= quest for information, *J. Geophys.*, **50**(3), 150–170.
- Tølbøll, R.J. & Christensen, N.B., 2006. Robust 1D inversion and analysis of helicopter electromagnetic (HEM) data, *Geophysics*, **71**(2), G53–G62.
- Vest Christiansen, A. & Auken, E., 2012. A global measure for depth of investigation, *Geophysics*, **77**(4), WB171–WB177.
- Viezzoli, A., Christiansen, A.V., Auken, E. & Sørensen, K., 2008. Quasi-3D modeling of airborne TEM data by spatially constrained inversion, *Geophysics*, **73**(3), F105–F113.
- Vignoli, G., Fiandaca, G., Christiansen, A.V., Kirkegaard, C. & Auken, E., 2015. Sharp spatially constrained inversion with applications to transient electromagnetic data, *Geophys. Prospect.*, **63**(1), 243–255.
- Vignoli, G., Sapia, V., Menghini, A. & Viezzoli, A., 2017. Examples of improved inversion of different airborne electromagnetic datasets via sharp regularization, *J. Environ. Eng. Geophys.*, **22**(1), 51–61.
- Witherly, K., Irvine, R. & Morrison, E.B., 2004. The Geotech VTEM time domain helicopter em system, in *SEG Technical Program Expanded Abstracts 2004*, pp. 1217–1220, Society of Exploration Geophysicists.
- Worrall, L., Lane, R., Meyers, J. & Whitaker, A., 2001. Exploring through cover - The integrated interpretation of high resolution aeromagnetic, airborne electromagnetic and ground gravity data from the Grant's Patch Area, Eastern Goldfields Province, Archaean Yilgarn Craton Part A: Mapping geology using airborne electromagnetics (TEMPEST), *ASEG Extended Abstracts*, **2001**(1), 1–4.
- Yang, D. & Oldenburg, D.W., 2012. Three-dimensional inversion of airborne time-domain electromagnetic data with applications to a porphyry deposit, *Geophysics*, **77**(2), B23–B34.
- Zahner, T., Lochbühler, T., Mariethoz, G. & Linde, N., 2015. Image synthesis with graph cuts: a fast model proposal mechanism in probabilistic inversion, *Geophys. J. Int.*, **204**(2), 1179–1190.