Probabilistic Inversion of Multiconfiguration Electromagnetic Induction Data Using Dimensionality Reduction Technique: A Numerical Study

Davood Moghadas*

Low-frequency loop–loop electromagnetic induction (EMI) offers several key advantages over many other geophysical techniques for proximal soil sensing. Yet, because of problems with the inversion of measured apparent electrical conductivity (ECa) data, application of EMI for geophysical imaging and interpretation is limited. In this study, a Bayesian inference was used to obtain electromagnetic conductivity images (EMCIs) from multiconfiguration ECa data. This approach allows analysis of highly nonlinear problems and renders an ensemble of models obtained from the posterior distribution that can be used to explore parameter uncertainty. In this respect, generalized formal likelihood function was used to more accurately describe the sensitivity of the posterior distribution to residual assumptions. Discrete cosine transform (DCT) was employed as a model compression technique to reduce the number of unknown parameters in the inversion. The DCT parameterization was performed using training image (TI)-based geostatistical simulations considering the ECa data pseudosection as a TI. The potential of the proposed approach was examined through different theoretical scenarios. The estimated subsurface EMCI shows excellent agreement with the original synthetic models subject to the appropriate choice of prior information. Moreover, DCT parameterization reduces the number of unknown parameters, increasing accuracy of the inversion with the Bayesian procedure. The proposed approach ensures accurate and high-resolution characterization of subsurface conductivity layering from measured ECa values.

Abbreviations: 1D, one-dimensional; 2D, two-dimensional; AR, autoregressive; DCT, discrete cosine transform; DREAM, differential evolution adaptive metropolis algorithm; DS, direct sampling; ECa, apparent electrical conductivity; EMCI, electromagnetic conductivity image; EMI, electromagnetic induction; ERT, electrical resistivity tomography; GLF, generalized likelihood function; HCP, horizontal coplanar loop; MAP, maximum a posteriori probability; MCMC, Markov chain Monte Carlo; MEAN, mean of the posterior distribution; MPS, multiple-point statistics; TI, training image; VCP, vertical coplanar loop.

Electromagnetic induction has shown considerable potential for characterization and monitoring of soil geophysical properties at field scale (Abdu et al., 2007; Triantafilis and Monteiro Santos, 2009, 2010, 2013; Jadoon et al., 2017; Altdorff et al., 2018; Rejiba et al., 2018; Robinet et al., 2018). Electromagnetic induction returns ECa (or σa mS m⁻¹), which is a cumulative response of the subsurface electrical conductivities. Operating in a diffusive regime (low frequency), EMI sensors allow for ECa measurements using two common source–receiver orientations namely: vertical coplanar loop (VCP) and horizontal coplanar loop (HCP). Recent advances in the calibration of the EMI data facilitate the quantitative interpretation of the measured apparent electrical conductivity values (Lavoué et al., 2010; Moghadas et al., 2012; Minsley et al., 2012). Obtaining accurate and quantitative ECa data is an essential prerequisite for establishing a reliable multilayered inversion (Jadoon et al., 2015, 2017; Moghadas et al., 2017).

The EMI antenna modes and intercoil spacings (offsets) are important factors affecting penetration depth of the system (Mester et al., 2011). Consequently, successful multilayered one-dimensional (1D) inversion of ECa data (single-site inversion) is challenging, since it involves limited subsurface information in the optimization process resulting...
to high uncertainty in parameter estimation. In this respect, the multilayered inversion of EC data may lead to an under-determined optimization problem. Moreover, inverting EC data for both electrical conductivity and thickness of the layers may present inaccuracies in estimation of layer thicknesses (Moghadas et al., 2017). Consequently, single-site inversion of EC data using 1D stratified earth model is typically ill-posed.

To enhance the efficiency of the traditional 1D multilayered conductivity inversion process, the problem can be formulated in a quasi-two-dimensional (2D) framework. This quasi-2D electromagnetic model mimics the EC measurements along a transect in pursuit of the subsurface electrical conductivity distribution. This postulates to discretize the subsurface domain of the transect in 2D grid cells (fixed thicknesses). Although this approach inverts only for the electrical conductivities taking into account all measured EC data along the transect, it requires estimation of the electrical conductivity for each discretized grid cell of subsurface structure. Consequently, this model parameterization involves the inference of a large number of unknowns and an inverse problem that is computationally expensive. These limitations hamper wider applications of EMI, motivating to develop more efficient inversion techniques.

Many researchers used the gradient-based inversion technique to infer subsurface geophysical properties from EC data (Monteiro Santos, 2004; Mester et al., 2011; Dafflon et al., 2013; Guillermot et al., 2016; Huang et al., 2016). This method allows for relatively fast inversion of measured EC data. However, the results of gradient-based inversion (as a deterministic approach) are typically interpreted as a so-called single-best solution (the best least-squares solution). Another category of the inversion routines is probabilistic methods providing posterior distributions that can be used to characterize parameter uncertainty, thus offering a richer source of information than a single estimation by deterministic inversions (Minsley, 2011; Shanahan et al., 2015; Jadoon et al., 2017; Moghadas et al., 2017). Moreover, this approach does not require linearization of the inverse problem and is particularly desirable for highly nonlinear problems (Minsley, 2011). Vrugt et al. (2009) proposed a probabilistic technique, differential evolution adaptive metropolis algorithm (DREAM), which has been shown to be robust to solve many earth and environmental problems (Steenpass et al., 2010; Laloy and Vrugt, 2012; Rosas-Carbajal et al., 2014; Moghadas et al., 2015, 2017).

To reduce the number of unknown parameters and consequently obtain more accurate inversion results (higher certainty), dimensionality reduction approaches have been proposed (Jafarpour et al., 2009; Laloy et al., 2015; Wright et al., 2017). In this respect, DCT has proven to be of great promise for successful inversion of geophysical data (Jafarpour et al., 2009, 2010; Linde and Vrugt, 2013; Lochbühler et al., 2014; Qin et al., 2016). This approach presents high efficiency for image compression to considerably reduce the number of unknown model parameters. However, appropriate determination of the dominant DCT coefficients for model compression in the inversion is not straightforward. To overcome this issue, Lochbühler et al. (2015) suggested DCT parameterization based on multiple-point statistics (MPS). In this approach, a conceptual geological model is considered as a so-called TI and a large set of realizations from that is created using a MPS simulator. The dominant DCT coefficients and their prior distributions are obtained from the MPS realizations.

Since the early 90’s, the MPS algorithms have received great attention in the Earth sciences, as they permit to generate highly heterogeneous random fields using a TI and overcome some limitations of classical variogram-based simulation methods (Mariethoz et al., 2010; Straubhaar et al., 2016). This makes MPS potentially pertinent for the estimation of complex spatial geological features (González et al., 2008; Malone et al., 2016). This is because a TI, which is a graphical representation of a conceptual model, such as geological features, lithology, and structural patterns, can be included. Among available MPS techniques, the direct sampling (DS) method offers a great potential for modeling spatial features (Mariethoz et al., 2010). This approach is computationally fast and straightforward to implement. The DCT-based inversion of geophysical data using the TI-based parameterization is computationally efficient and substantially improves the accuracy of the posterior parameter estimations. However, complementary information that is not always readily accessible is required to build a TI.

This study tackled the issue of applying DCT-based probabilistic inversion on the multireceiver EC data using TI-based parameterization via MPS. In this respect, the EC pseudosection inferred from data was considered as a TI for the DS statistical simulations. After creating a large set of MPS realizations, the dominant DCT coefficients were determined. The DS realizations were used to formulate a TI-based parameterization for the inversion. The Bayesian sampling approach was employed using formal likelihood function to derive posterior distributions. Two strategies were also examined to determine lower and upper bounds of the parameter spaces (prior information). The first approach determined the upper and lower bounds of the DCT coefficients (in the frequency domain) in such a way that all possible models can be sampled within the specified minimum and maximum conductivity ranges, while the second approach inferred the parameter ranges from MPS realizations of the EC data pseudosection. The robustness and pertinency of the proposed methodology for inversion of multiconfiguration EC data were fully examined through different synthetic scenarios.

Materials and Methods

Electromagnetic Induction Forward Modeling

Apparent electrical conductivity forward modeling is usually performed under low induction number condition (EC < 100 mS m⁻¹) using McNeill (1980) model, which is based on some simplified assumptions. Under high induction number condition where EC > 100 mS m⁻¹ (Hendrickx et al., 2002), a more complex model (full-EM forward model) is warranted, which numerically solves the Maxwell equations (Wait, 1954; Ward and
Hohmann, 1987). Given a horizontally layered medium consisting of \( n + 1 \) layers with horizontal and vertical dipole source–receiver combination, the full-EM model is formulated as (antennas are on the surface of the earth):

\[
\sigma_{\text{HCP}}(\omega, \rho) = -\frac{4\rho}{\omega \mu_0} \text{Im} \left[ \int_0^\infty R_0 J_0(\rho \lambda) \lambda^2 d\lambda \right] \\
\sigma_{\text{VCP}}(\omega, \rho) = -\frac{4}{\omega \mu_0} \text{Im} \left[ \int_0^\infty R_0 J_1(\rho \lambda) \lambda d\lambda \right]
\]

where \( \sigma_{\text{HCP}}(\text{mS m}^{-1}) \) and \( \sigma_{\text{VCP}}(\text{mS m}^{-1}) \) are \( E_C \) for HCP and VCP configurations, respectively, \( \rho (\text{m}) \) signifies offset, \( J_0 (\text{dimensionless}) \) and \( J_1 (\text{dimensionless}) \) are Bessel functions of the zeroth and first order, respectively, \( \mu_0 (\text{m}^{-2}) \) is the free-space permeability, \( \lambda (\text{m}) \) is the radial wave number, and \( \omega (\text{rad s}^{-1}) \) represents the angular frequency. The reflection factor, \( R_0 \), is calculated from the following recursion formula:

\[
R_n(h_n, \sigma_n) = \frac{r_n + R_{n+1} \exp(-2\Gamma_n h_{n+1})}{1 + r_n R_{n+1} \exp(-2\Gamma_n h_{n+1})}
\]

where \( r_n = (\Gamma_n - \Gamma_{n+1})/\Gamma_n \), \( \Gamma_n = (\lambda^2 + \omega \mu_0 \sigma_n)^{1/2} \), \( j = (-1)^{n+1} \), and \( h_n(\text{m}) \) and \( \sigma_n(\text{mS m}^{-1}) \) denote the thickness and electrical conductivity of the \( n \)th layer, respectively. The recursion starts at the bottom layer, \( n + 1 \), for which \( R_{n+1} = 0 \), and continues to the top of the domain, where \( \sigma_0 = 0 \), to give \( R_0 \). In this study, the full-EM model was used, which is valid for both low and high induction number conditions (van der Kruik et al., 2000).

### Multireceiver Electromagnetic Induction Sensor

A measure of an \( E_C \) distribution of sensitivity in the subsurface (spatial sensitivity) is called cumulative response (McNeill 1980). In this respect, sensitivity curves introduced by McNeill (1980) are used to model the cumulative response of the subsurface to the electromagnetic field generated by EMI instruments considering HCP and VCP modes. Given a low induction number condition, the effective penetration depths in HCP and VCP modes are calculated considering intercoil spacing (Callegary et al., 2007). Here, a multiconfiguration EMI sensor called CMD Mini-Explorer (GF Instruments) was considered to simulate \( E_C \) measurements from both HCP and VCP coil orientations with three offsets. Consequently, for each measurement point, combining three offsets with two antenna modes lead to six measurement depths. The CMD Mini-Explorer uses 30 kHz frequency with 0.32, 0.71, and 1.8 m offsets. For this system, the effective penetration depths of HCP mode are 0.5, 1.0, and 1.8 m for 0.32, 0.71, and 1.8 m, respectively. These values are reduced to 0.25, 0.5, and 0.9 m using VCP mode. For each measurement point along the transect, the generated synthetic \( E_C \) data from each offset-coil configurations are plotted vs. their corresponding effective penetration depths. Then, these depth-profile \( E_C \) data are stitched together providing an \( E_C \) pseudosection of the transect down to the maximum effective penetration depth of the EMI system.

This \( E_C \) data pseudosection can be further interpolated in both vertical and horizontal directions to improve the perception about the spatial variations of the \( E_C \) data.

### Model Parameterization

To perform \( E_C \) probabilistic inversion, the subsurface was discretized in 2D grid cells considering \( N = 12 \) subsurface layers in the vertical direction and \( M = 80 \) cells in the lateral position. In this study, the \( x-z \) coordinate system is such that the \( x \) axis is horizontal with positive direction from left to right and the \( z \) axis is vertical with positive direction from top to bottom. The maximum depth of the domain was fixed to 1.8 m (the maximum effective penetration depth of the EMI sensor). For a uniformly discretized domain of \( N \times M \) cells, the DCT-II representation in two dimensions is given by (Ahmed et al., 1974)

\[
G(i, j) = \alpha_i \alpha_j \sum_{x=1}^{N} \sum_{z=1}^{M} S(x, z) \cos \left\{ \frac{\pi(2x-1)(i-1)}{2N} \right\} \cos \left\{ \frac{\pi(2z-1)(j-1)}{2M} \right\}
\]

where

\[
\alpha_i = \begin{cases} 
\frac{1}{\sqrt{N}}, & i = 1 \\
\frac{2}{\sqrt{N}}, & 2 \leq i \leq N 
\end{cases}, \quad \alpha_j = \begin{cases} 
\frac{1}{\sqrt{M}}, & j = 1 \\
\frac{2}{\sqrt{M}}, & 2 \leq j \leq M
\end{cases}
\]

In this expression, DCT coefficients stored in \( G \) render transforming from the space domain (matrix \( S \)) to the frequency domain. The low-frequency DCT coefficients (dominant coefficients) contain the main information about this transformation. By discarding the high-frequency components, one can retrieve the original spatial image without losing significant information. Incorporating the DCT with the inversion scheme (DCT-based inversion) facilitates more accurate parameter estimations, since only the dominant coefficients require estimation. After inversion, the retrieved coefficients (in the frequency domain) are transformed back to the space domain using the inverse of the DCT to construct the subsurface structures.

An important issue in DCT-based inversion is appropriate definition of the low-frequency cells in the grid. A common procedure for truncation is to consider a box or a triangle in the low-frequency corner of the transform space (an ad hoc selection of the dominant DCT coefficients). However, this approach may lead to unrealistic inversion results. In this respect, Lochbühler et al. (2015) introduced a TI-based parameterization for more accurate definition of the dominant DCT coefficients. Here, \( E_C \) data pseudosections were considered as TIs to determine sparse model parameterization. An \( E_C \) pseudosection represents the variation of the measured \( E_C \) with position and effective penetration depth (depending on the offset and coil orientation). A pseudosection is not representative of a real subsurface conductivity structure, since the true depths are unknown and the shapes (subsurface...
In this study, DeeSse software was used as a MPS simulator, which was determined by making a Pareto optimal matching between realization) of the TI, resulting in a complexity. In this respect, after inference of all low-frequency DCT providing a general impression of the subsurface features. In other pseudosection as a TI, each realization was transformed to the frequency domain using DCT (Eq. [3]). Then, the arithmetic related to different offsets were grouped together to build a vector of ECa values with \( K = 3K_m \) length (3 is the number of offsets). The grouped data sets from both HCP and VCP modes were then integrated to construct a 2K vector of ECa data, resulting to a multiconfiguration inversion.

Schoups and Vrugt (2010) introduced a generalized likelihood function (GLF), which extends the applicability of likelihood functions to situations where residual errors are correlated, heteroscedastic, and non-Gaussian with varying degrees of kurtosis and skewness. To evaluate the sensitivity of the posterior parameter distribution to residual assumptions, the GLF was used, which is formulated as

\[
L(\theta|\sigma_a\text{ meas}, \sigma, \beta, \xi, \Phi_p) = \prod_{k=1}^{2K} \left( \frac{2\sigma_k \omega_k}{\xi + \xi^{-1}} \right) \exp \left[ -\epsilon_k \frac{\tau_k(\theta, \Phi_p)}{\xi \text{ sgn} \tau_k(\theta, \Phi_p)} \right]^{2/(1+\delta)}
\]

where \( \sigma(\theta) \) and \( \sigma(\theta|\sigma_a\text{ meas}) \) denote the prior and posterior parameter distribution, respectively, and \( L(\theta|\sigma_a\text{ meas}) \equiv p(\sigma_a\text{ meas}|\theta) \) signifies the likelihood function. The model evidence, \( p(\sigma_a\text{ meas}) \) acts as a normalization factor and can be eliminated. Given \( K_m \) the number of measurement points along a transect, the \( \sigma \) values related to different offsets were grouped together to build a vector of ECa values with \( K = 3K_m \) length (3 is the number of offsets). The grouped data sets from both HCP and VCP modes were then integrated to construct a 2K vector of ECa data, resulting to a multiconfiguration inversion.

Schoups and Vrugt (2010) introduced a generalized likelihood function (GLF), which extends the applicability of likelihood functions to situations where residual errors are correlated, heteroscedastic, and non-Gaussian with varying degrees of kurtosis and skewness. To evaluate the sensitivity of the posterior parameter distribution to residual assumptions, the GLF was used, which is formulated as

\[
L(\theta|\sigma_a\text{ meas}, \sigma, \beta, \xi, \Phi_p) = \prod_{k=1}^{2K} \left( \frac{2\sigma_k \omega_k}{\xi + \xi^{-1}} \right) \exp \left[ -\epsilon_k \frac{\tau_k(\theta, \Phi_p)}{\xi \text{ sgn} \tau_k(\theta, \Phi_p)} \right]^{2/(1+\delta)}
\]

where \( \sigma(\theta) \) and \( \sigma(\theta|\sigma_a\text{ meas}) \) denote the prior and posterior parameter distribution, respectively, and \( L(\theta|\sigma_a\text{ meas}) \equiv p(\sigma_a\text{ meas}|\theta) \) signifies the likelihood function. The model evidence, \( p(\sigma_a\text{ meas}) \) acts as a normalization factor and can be eliminated. Given \( K_m \) the number of measurement points along a transect, the \( \sigma \) values related to different offsets were grouped together to build a vector of ECa values with \( K = 3K_m \) length (3 is the number of offsets). The grouped data sets from both HCP and VCP modes were then integrated to construct a 2K vector of ECa data, resulting to a multiconfiguration inversion.

Schoups and Vrugt (2010) introduced a generalized likelihood function (GLF), which extends the applicability of likelihood functions to situations where residual errors are correlated, heteroscedastic, and non-Gaussian with varying degrees of kurtosis and skewness. To evaluate the sensitivity of the posterior parameter distribution to residual assumptions, the GLF was used, which is formulated as

\[
L(\theta|\sigma_a\text{ meas}, \sigma, \beta, \xi, \Phi_p) = \prod_{k=1}^{2K} \left( \frac{2\sigma_k \omega_k}{\xi + \xi^{-1}} \right) \exp \left[ -\epsilon_k \frac{\tau_k(\theta, \Phi_p)}{\xi \text{ sgn} \tau_k(\theta, \Phi_p)} \right]^{2/(1+\delta)}
\]

where \( \sigma(\theta) \) and \( \sigma(\theta|\sigma_a\text{ meas}) \) denote the prior and posterior parameter distribution, respectively, and \( L(\theta|\sigma_a\text{ meas}) \equiv p(\sigma_a\text{ meas}|\theta) \) signifies the likelihood function. The model evidence, \( p(\sigma_a\text{ meas}) \) acts as a normalization factor and can be eliminated. Given \( K_m \) the number of measurement points along a transect, the \( \sigma \) values related to different offsets were grouped together to build a vector of ECa values with \( K = 3K_m \) length (3 is the number of offsets). The grouped data sets from both HCP and VCP modes were then integrated to construct a 2K vector of ECa data, resulting to a multiconfiguration inversion.

Schoups and Vrugt (2010) introduced a generalized likelihood function (GLF), which extends the applicability of likelihood functions to situations where residual errors are correlated, heteroscedastic, and non-Gaussian with varying degrees of kurtosis and skewness. To evaluate the sensitivity of the posterior parameter distribution to residual assumptions, the GLF was used, which is formulated as

\[
L(\theta|\sigma_a\text{ meas}, \sigma, \beta, \xi, \Phi_p) = \prod_{k=1}^{2K} \left( \frac{2\sigma_k \omega_k}{\xi + \xi^{-1}} \right) \exp \left[ -\epsilon_k \frac{\tau_k(\theta, \Phi_p)}{\xi \text{ sgn} \tau_k(\theta, \Phi_p)} \right]^{2/(1+\delta)}
\]

where \( \sigma(\theta) \) and \( \sigma(\theta|\sigma_a\text{ meas}) \) denote the prior and posterior parameter distribution, respectively, and \( L(\theta|\sigma_a\text{ meas}) \equiv p(\sigma_a\text{ meas}|\theta) \) signifies the likelihood function. The model evidence, \( p(\sigma_a\text{ meas}) \) acts as a normalization factor and can be eliminated. Given \( K_m \) the number of measurement points along a transect, the \( \sigma \) values related to different offsets were grouped together to build a vector of ECa values with \( K = 3K_m \) length (3 is the number of offsets). The grouped data sets from both HCP and VCP modes were then integrated to construct a 2K vector of ECa data, resulting to a multiconfiguration inversion.
Fig. 1. Schematic progress of the discrete cosine transform (DCT)-based probabilistic inversion strategy. The apparent electrical conductivity (ECa) pseudosection is considered as a training image (TI) for the multiple-point statistical (MPS) simulations. The TI realizations are transformed to the frequency domain using DCT. Afterward, all dominant DCT coefficients are determined and optimal number of coefficients are derived using Pareto optimality. The parameter ranges are also inferred from the TI realizations. The model space is then sampled with the MT-DREAM (ZS) algorithm to derive posterior model realizations.
subject to inference along with the model parameters, \( \theta \), assuming uniform prior distributions for the error model parameters.

The lower and upper bounds for the nuisance parameters were selected based on the procedure described in Schoup and Vrugt (2010). The skewed exponential power density function is symmetric when \( \xi = 1 \), positively skewed when \( \xi > 1 \), and negatively skewed when \( \xi < 1 \). A value of \( \beta = -1 \) contributes to a uniform distribution, \( \beta = 0 \) generates a normal distribution, and \( \beta = 1 \) depicts a double-exponential distribution. As a consequence, the prior uncertainty ranges of these parameters were set to \( -1 < \beta < 1 \), and \( 0.1 < \xi < 10 \). The value of the autocorrelation coefficient was assigned to \( 0 < \phi_1 < 1 \) to account for the correlations of error residuals. The ranges of heteroscedasticity intercept are inferred from the data that were set to \( 0 < \epsilon_0 < 0.1 \text{ S m}^{-1} \). Since \( \epsilon_x \) represents the heteroscedasticity slope, the lower and upper bounds for this parameter were selected between 0 and 1, respectively. During the Bayesian inversion, some proposed models had electrical conductivity values outside the specified parameter ranges. Such proposals were discarded by assigning very low likelihood values.

To approximate the posterior distribution, a Markov chain Monte Carlo (MCMC) simulation was used. The MCMC algorithm used here is called MT-DREAM(\( ZS \)) (Laloy and Vrugt, 2012), which is a modified version of the original DREAM sampling routine (Vrugt et al., 2009). This algorithm renders generating a candidate point in each chain using sampling information from the past states. The MT-DREAM(\( ZS \)) permits accelerated convergence for high-dimensional problems using only three chains. The \( \hat{R} \) statistic of Gelman and Rubin (1992) is used to assure the convergence of the algorithm. The threshold of \( \hat{R} = 1.2 \) is applied for convergence diagnosis of this algorithm. (For further information regarding the DREAM software package, see Vrugt, 2016).

Figure 1 demonstrates different steps of the proposed methodology for DCT-based probabilistic inversion of multiconfiguration EC\(_a\) data. The RMS error between the observed (\( \sigma_a^{\text{meas}} \)) and the modeled (\( \sigma_a^{\text{mod}} \)) EC\(_a\) data was formulated as

\[
\text{RMSE} = \frac{1}{2K} \sum_{k=1}^{2K} \left( \sigma_a^{\text{meas}}(k) - \sigma_a^{\text{mod}}(k) \right)^2
\]

As mentioned above, the dominant DCT coefficients in the frequency domain were considered as unknown model parameters in the inversion scheme. In this respect, the uniform prior distribution was considered for these parameters. To set the search space, it is necessary to first define the minimum (\( \sigma_{\text{min}} \)) and maximum (\( \sigma_{\text{max}} \)) electrical conductivity values in the space domain. These values were adjusted to \( \sigma_{\text{min}} = 1 \text{ mS m}^{-1} \) and \( \sigma_{\text{max}} = 80 \text{ mS m}^{-1} \) to cover a relatively large conductivity range. To adjust upper and lower bounds for the low-frequency DCT coefficients in the inversion, two approaches were examined. As a first approach (Method I), the strategy of Linde and Vrugt (2013) was employed. In this respect, the values of \( G(1,1) \) were scaled by adjusting the values of the other cells \( [G(i,j)] \) to zero to find the bounds for which the corresponding inverse of DCT models fell within \( \sigma_{\text{min}} \) and \( \sigma_{\text{max}} \). This element of matrix \( G \) contains the maximum information regarding the DCT transform. The other cells were scaled in a similar way such that after the inverse transform each individual coefficient had a corresponding value of \( (\sigma_{\text{min}} - \sigma_{\text{max}})/2 \).

As a second approach (Method II), first the minimum and maximum values of dominant DCT coefficients from the MPS realizations were determined. Since the EC\(_a\) data are not representative of the real subsurface conductivity values, these ranges cannot be directly used for the MCMC inference. However, they provide valuable insights about the variations of the lower and upper bounds of the search space. Consequently, the MPS-derived ranges were then rescaled (normalized) by adjusting the minimum and maximum ranges of the first cell to the \( G_{\text{min}}(1,1) \) and \( G_{\text{max}}(1,1) \), respectively. The resulting parameter space could be further broadened by multiplying by a scale factor, allowing for wider search domain for the problem. However, special care should be taken to select the scaling factor, since a very wide search space resulted to produce a large set of out of range proposals and a computationally expensive simulation.

**Synthetic Models**

Three different synthetic models were considered to examine the applicability of the proposed methodology. The first model corresponds to a synthetic EMCI (Model 1) consisting of three layers representing different lateral depth variations (Fig. 2a). The first, second, and third layers have \( \sigma \) of 15, 40, and 60 mS m\(^{-1} \), respectively. The second synthetic model (Model 2) consists of two layers with different lateral extensions. The top layer is characterized by \( \sigma \) of 15 mS m\(^{-1} \), while the lower layer has 45 mS m\(^{-1} \) conductivity (Fig. 2c). Both Model 1 and Model 2 assume a maximum depth of 1.8 m for the domain with 80 m horizontal length. These models mimic the depth to clay-rich and more conductive subsoil or wetter subsoil conditions.

The third model (Model 3) corresponds to the electrical resistivity tomography (ERT) inversion results reported by Lavoué et al. (2010). The ERT data were measured along a 120-m transect at the Selhausen test site (North Rhine-Westphalia, Germany). These data were collected using dipole–dipole array with 0.25-m spacing between electrodes. The measured data were inverted using RES2DINV software (Loke et al., 2003). Figure 2e shows the sub-surface model derived from ERT inversion. This model presents a horizontal gradient of decreasing \( \sigma \) for increasing distance, which is due to soil water content and texture variations. Moreover, the electrical conductivity increases with depth, which can be related to an increase in soil moisture and clay content with depth (Lavoué et al., 2010).

The CMD Mini-Explorer configuration was used to estimate synthetic EC\(_a\) data assuming an 80-m transect with 1-m measurement step for Models 1 and 2 and a 120-m transect with 1.5-m measurement step for Model 3. To assess the sensitivity of the proposed approach to the presence of noise and evaluate the applicability of the method for real case scenarios, the EC\(_a\) data were contaminated with noise. In this respect, Gaussian random
noise of 2% of the average of $\sigma_a$ was introduced to the synthetic EC\textsubscript{a} data. Figures 2b, 2d, and 2f present the EC\textsubscript{a} pseudosections for Models 1, 2, and 3, respectively. As can be seen, the EC\textsubscript{a} data pseudosections present a general pattern of the subsurface conductivity changes.

**Results**

**Multiple-Point Statistical Simulations**

Given EC\textsubscript{a} data pseudosections (see Fig. 2) as TIs, MPS simulations were performed considering a total of 1000 realizations with the DS algorithmic parameters described in the Model Parametrization section. Figure 3 shows a few of these simulations for all models. Figures 4a to 4c demonstrate all dominant DCT coefficients obtained from full ensemble of TI realizations. Note that before applying the DCT, the electrical conductivity models were converted to the logarithmic scale. The trade-off between $P_t$ and $n_t$ (Pareto optimality) resulted to Pareto solution (red stars) with $P_t^\text{f}$ equal to 16, 18, and 14% for the first, second, and third

models, respectively (Fig. 4d–4f). This contributed to reduce the number of coefficients from $n_s$ equals 143, 108, and 275, respectively, to $n_t^\text{f}$ equal to 34, 26, and 79 for Models 1, 2, and 3, respectively (Fig. 4g–4i).

The MPS simulations of the EC\textsubscript{a} data pseudosections predicted the location and number of low-frequency DCT coefficients of the regular grid. These coefficients were compared with their counterparts derived from synthetic EMCIs interpolated at the regular grid points to evaluate the accuracy of the proposed approach. Figures 4j to 4l show the dominant DCT coefficients obtained from transformation of the synthetic models to the frequency domain. The number of these low-frequency components is denoted by $n_s$. As clearly seen, the predicted coefficients are almost consistent with their counterparts, but some discrepancies can be observed such that some of the low-frequency coefficients are dismissed.

To further investigate the robustness of the proposed approach to predict the location and number of the low-frequency DCT coefficients of the regular grid, the predicted dominant DCT coefficients were used to recover the synthetic EMCI transformed to the frequency domain. In this respect, the frequency components of the synthetic EMCI interpolated at the regular grid points were determined using Eq. [3]. Then, the predicted coefficients from MPS simulations were kept and the inverse of the DCT were applied on the frequency components by zeroing the other elements. Figures 5a to 5c shows the transform domain data for the first, second, and third models, respectively (in logarithmic scale). The reconstructed EMCIs are presented in Fig. 5d to 5f for Models 1, 2, and 3, respectively.

Regarding Model 1, there were relatively small inaccuracies in recovering the conductivity of the second layer around the two corners of the domain. However, the main features of the model were well recovered. Regarding the second and third models, the synthetic conductivity models were successfully recovered. These results confirm that accurate retrieval of the EMCI does not require involving all low-frequency DCT coefficients in the transformation procedure. However, the applied DCT relatively smoothed the sharp transitions between boundaries of different layers in both Model 1 and Model 2. Yet, the accuracy of the retrieved models demonstrates the pertinency and robustness of the proposed strategy and ensures appropriate selection of the low-frequency DCT coefficients for the subsequent probabilistic inversion.
Inversion Results (Method I)

The MCMC sampling was performed using model parameterization described above. Figure 6 presents posterior MCMC realizations of the ECa data using Method I. The first, second, and third columns depict maximum a posteriori probability (MAP), mean of the posterior distribution (MEAN), and 95% confidence intervals, respectively. For all scenarios, a close agreement between MAP and posterior MEAN images are observed. This manifests...
Fig. 5. Reconstruction of the synthetic subsurface electromagnetic conductivity images (EMCIs) interpolated at the regular grid points using the transform domain data. The truncation of the coefficients was performed using the predicted dominant coefficients (colored cells) determined from the training image realizations. The upper figures represent the 12 by 80 matrix of low-frequency coefficients. The lower diagrams show the reconstructed EMCIs using inverse of the discrete cosine transform. The first, second, and third columns correspond to the Model 1, Model 2, and Model 3, respectively. The transform domain data are presented in a logarithmic scale. Note that the inverse of the transform was performed by zeroing the other coefficients.

Fig. 6. Maximum a posteriori probability (MAP), mean of the posterior distribution (MEAN), and 95% confidence interval related to the discrete cosine transform (DCT)-based probabilistic inversion of the apparent electrical conductivity (ECa) data using Method I: (a–c) Model 1; (d–f) Model 2; (g–i) Model 3.
the well-posedness of the DCT-based probabilistic inversion problem with posterior EMCI that remains in close vicinity of the MAP image. However, the inversion approach provided totally irrelevant subsurface images in comparison with the synthetic models. Consequently, the DCT-based probabilistic inversion with the first approach failed to accurately estimate the subsurface EMCI. Figure 7 shows histograms of the sampled RMSE values of different MCMC realizations using Method I. The posterior conductivity images derived from Model 1 have RMSE values in the range of around 1 to 4 mS m\(^{-1}\) (Fig. 7a). Regarding the second scenario (Fig. 7b), the RMSE ranges from 1 to 1.5 mS m\(^{-1}\). The posterior conductivity images derived from Model 3 have RMSE values in the range of around 12.5 to 14.2 mS m\(^{-1}\) (Fig. 7c). These high RMSE values reiterate the rather inferior results of the probabilistic inversion approach using an inadequate prior distribution.

Figure 8 shows the posterior distribution of the nuisance parameters \(s_0, s_1, \xi, \beta, x,\) and \(\phi_1\) using Method I. The first, second, and third columns correspond to Model 1, Model 2, and Model 3, respectively. The blue and black stars signify the MAP and MEAN of the posterior distribution, respectively. The most important results are as follows. First, a close agreement between MAP and MEAN of the posterior distributions is observed. Second, most histograms are well described with a Gaussian distribution with the exception of \(s_0\) for Model 1 and Model 2 (Fig. 8a and 8b) and...
$s_1$ for Model 1 (Fig. 8d). Third, $s_0$ presents rather negligible values for all scenarios. Forth, the values of $b$ suggest that the shape of the residual errors are non-Gaussian (Fig. 8g–8i). Fifth, the values of $x$ propose rather small skewness for the residual errors for Models 1 and 2, and relatively sharp right skewness for Model 3 (Fig. 8j–8l). Last, the marginal distribution of $f_1$ for all models (Fig. 8m–8o) indicates a very strong serial correlation at the first lag.

To better understand how GLF characterizes the residual errors, the residuals correspond to the MAP of the MCMC realizations (Method I) are illustrated in Fig. 9. The first, second, and third columns correspond to the Model 1, Model 2, and Model 3, respectively. The top panel demonstrates the ECa residuals. The middle panel compares the histogram of the ECa residuals against the assumed distribution of the residuals (red line). The third panel presents the partial autocorrelation of the residuals. As clearly seen, the residual errors present a homoscedastic pattern for all scenarios (Fig. 9a–9c) with non-Gaussian shape for all models (Fig. 9d–9f). The bottom panel confirms that the decorrelated residuals do not exhibit serial correlation (Fig. 9g–9i).

**Inversion Results (Method II)**

Figure 10 presents posterior MCMC realizations of the ECa data using Method II. The first, second, and third columns depict MAP, MEAN, and 95% confidence intervals, respectively. These figures highlight several important findings for the synthetic scenarios. First, the MAP and MEAN of the posterior distributions are almost in close agreement. Second, the MAP and MEAN images of the electrical conductivities demonstrate an excellent agreement with their counterparts presented in Fig. 2. Third, the 95% confidence intervals of the conductivity images appear rather small close to the surface and tend to increase downward. Last but not the least, the model reduction using DCT smoothed the sharp transitions between boundaries of different layers in both Model 1 and Model 2. Figure 11 shows histograms of the sampled RMSE values of different MCMC realizations using Method II. The posterior conductivity images derived from Model 1 have RMSE values in the range of around 0.7 to 0.8 mS m$^{-1}$ (Fig. 11a). Regarding the second scenario (Fig. 11b), the RMSE ranges from 0.5 to 0.6 mS m$^{-1}$. The posterior conductivity images derived from Model 3 have RMSE values in the range of around 3 to 8 mS m$^{-1}$ (Fig. 11c). All scenarios present acceptable fitting, reasserting the robustness of the TI-based prior for DCT-based probabilistic inversion of ECa data.

Figure 12 shows the posterior distribution of the nuisance parameters ($s_0$, $s_1$, $b$, $x$, and $f_1$) using Method II. The first, second, and third columns correspond to the Model 1, Model 2, and Model 3, respectively. The blue and black stars signify the MAP and MEAN of the posterior distribution, respectively. The most
important results are as follows. First, a close agreement between MAP and MEAN of the posterior distributions is observed. Second, most histograms are well described with a Gaussian distribution with the exception of $s_0$ for Model 1 and Model 2 (Fig. 12a and 12b). Third, $s_0$ presents rather negligible values for all scenarios. Forth, the values of $\beta$ suggest that the shape of the residual errors are non-Gaussian (Fig. 12g–12i). Fifth, the values of $\xi$ propose rather small skewness for the residual errors (Fig. 12j–12l). Last, the marginal distribution of $\phi_1$ for Model 3 (Fig. 12o) demonstrates values around 0.9, indicating a very strong serial correlation at the first lag.

Figure 13 shows residuals correspond to the MAP of the MCMC realizations using Method II. The first, second, and third columns correspond to the Model 1, Model 2, and Model 3, respectively. The top panel demonstrates the $EC_{a}$ residuals. The middle panel compares the histogram of the $EC_{a}$ residuals against the assumed distribution of the residuals (red line). The third panel presents the partial autocorrelation of the residuals. As can be seen, the residual errors present a homoscedastic pattern (Fig. 13a–13c) with non-Gaussian shape for all models (Fig. 13d–13f). The bottom panel confirms that the decorrelated residuals do not exhibit serial correlation (Fig. 13g–13i).

Fig. 10. Maximum a posteriori probability (MAP), mean of the posterior distribution (MEAN), and 95% confidence interval related to the discrete cosine transform–based probabilistic inversion of the apparent electrical conductivity data using Method II: (a–c) Model 1; (d–f) Model 2; (g–i) Model 3.

Fig. 11. Histograms of the sampled RMSE values of the different Markov chain Monte Carlo inversion runs using Method II: (a) Model 1; (b) Model 2; (c) Model 3.
**Discussion**

The proposed DCT-based probabilistic inversion successfully recovered the subsurface EMCIs with high spatial resolution using TI-based prior. Since Method I and Method II share the same low-frequency DCT components, the use of an adequate prior distribution describes the superiority of the second approach for inversion of all synthetic scenarios. This method allows for incorporating all measured ECa data along the transect with high computational efficiency as a result of model compression via DCT. This is an important achievement in comparison with the commonly used single-site multilayered inversion that involves limited subsurface information in the optimization procedure. Moreover, prior information is obtained directly from the measured data and incorporation of the external source of information (as used by Lochbühler et al. [2015]) is not necessary for model parameterization. It is important to note that an ECa pseudosection for the same site and with the same instrument can be varied because of system calibration. As a consequence, the measured ECa data should be calibrated before applying the proposed methodology presented here. The choice of likelihood function is important for accurate subsurface characterizations. Thanks to the more realistic residual error assumptions, the formal likelihood function has demonstrated superiorities than the commonly used normal likelihood (Schoups and Vrugt, 2010). In this study, a 1D earth model was assumed for the simulations. However, in the presence of multidimensional subsurface structures, a 2D or 2.5D forward modeling should be used to more accurately describe complex geological features.

The proposed methodology for inversion of loop–loop EMI data substantially relies on the ECa data pseudosection. The CMD Mini-Explorer system presented here allows different depth sensing because of the multireceiver capacity of the sensor. However, EMI sensors with less intercoil spacing return rather limited subsurface information. For instance, EM38 (Geonics Ltd.) renders conductivity measurements with 1-m offset and two coil orientations. This hampers inferring an ECa pseudosection that appropriately reflects conceptual subsurface features. Moreover, in the presence of complex subsurface geology, the dominant response from the shallower targets may mask signature of the deeper layers known as geological noise (Everett and Weiss, 2002). This causes the inferred pseudosection to present entirely different conductivity patterns for deeper parts of the domain. As a result, the inaccuracies of the ECa values introduce uncertainties in both model parameterization and inversion procedure (Monteiro Santos et al., 2010). In other words, the ambiguities in the measured ECa data are reflected in the inversion results.

Incorporation of the DCT as a model reduction in the inversion of ECa data resulted in smooth subsurface EMCIs.

---

**Fig. 12.** Posterior histograms of the nuisance parameters using generalized likelihood function and Method II. The first, second, and third columns correspond to the Model 1, Model 2, and Model 3, respectively. The blue and black stars signify the maximum a posteriori probability and mean of the posterior distribution of the posterior distribution, respectively.
Consequently, this approach may generate dipping layers in the presence of sharp conductivity variations within the layers as presented by the inversion of Model 1 and Model 2. Nevertheless, the consistency between the synthetic and inversely estimated models is promising and warrants further investigations. It is worth noting that the posterior intervals of the conductivity images appear rather small close to the surface and tend to increase downward. This is due to the fact that sensitivity of the returned EMI signal decreases with increasing depth. This suggests the need to examine the potential advantages of an irregular parameterization over a regular one, since an irregular grid allows for introducing higher resolution cells in specific parts of the domain. For instance, an irregular grid with logarithmically spaced cells in vertical direction may provide higher resolution results for shallower layers and maximizes the information inferred from the data. As a consequence, DCT-based inversion of $EC_a$ data using an irregular grid will be investigated in the future.

**Conclusion**

In this study, the probabilistic inversion approach was integrated with the dimensionality reduction technique via DCT to back out subsurface EMCIs. The proposed methodology was benefitted as well from TI-based prior. In this respect, the $EC_a$ data pseudosection was used as a TI for the DS simulations to obtain dominant DCT coefficients and their prior information for the subsequent DCT-based probabilistic inversion. The inversion strategy was successfully validated using different synthetic scenarios. This study demonstrated that although appropriate selection of the dominant DCT coefficients is important, a proper choice of the prior parameter range plays a crucial role for accurate subsurface characterizations. The proposed inversion scheme using TI-based prior allowed for obtaining high-resolution subsurface EMCIs with excellent agreements with their synthetic counterparts. Compared with the commonly used multilayered conductivity inversion techniques, this approach offers several advantages. It allows for model reduction inferring prior information directly from data and incorporating all measured values along the transect in a quasi-2D optimization framework (an over-determined inverse problem), which is particularly pertinent for EMI inversion. Moreover, the information from TI realizations renders to reduce the posterior uncertainty in probabilistic inversion of $EC_a$ data. Furthermore, the TI-based prior suppresses inversion artifacts and provides high-resolution EMCIs. This methodology is also computationally efficient since the model compression through DCT allows for a very fast convergence of the inverse problem. Future work will focus on field applications of the proposed methodology and comparing the results with the other inversion routines.
Acknowledgments

This work was supported by the Brandenburg University of Technology Cottbus-Senftenberg (BTU, Germany), which is acknowledged. The author kindly acknowledges Philippe Renard and Julien Straubhaar (University of Neuchâtel) for providing the DeCCe simulation code. Two anonymous reviewers and the associate editor, Andrew Binley, helped to substantially improve the manuscript, which is acknowledged.

References


