Bayesian Inversion of Electrical Resistivity Tomography Data for Uranium Exploration Through a Localized Ensemble-Based Approach

Alessandro Vinciguerra, Guy Marquis, Jean-françois Girard

Geophysical Prospecting

Geophysical Prospecting

ORIGINAL ARTICLE OPEN ACCESS

Bayesian Inversion of Electrical Resistivity Tomography Data for Uranium Exploration Through a Localized Ensemble-Based Approach

Alessandro Vinciguerra[1] | Guy Marquis[2] | Jean-François Girard[2]

[1]CNRS, Institut Terre & Environnement de Strasbourg, UMR 7063, Université de Strasbourg, Strasbourg, France | [2]Institut Terre & Environnement de Strasbourg, UMR 7063, Université de Strasbourg, Strasbourg, France

Correspondence: Alessandro Vinciguerra (avinciguerra@unistra.fr)

Received: 23 July 2025 | **Revised:** 10 March 2026 | **Accepted:** 17 April 2026

Keywords: inverse problems | resistivity | mining exploration

ABSTRACT

The electrical resistivity tomography (ERT) is a geophysical method commonly used in mining exploration to reconstruct the subsurface resistivity. The associated inverse problem is typically addressed using deterministic approaches that minimize an objective function and guarantee fast convergence. However, due to the ill-posedness and nonlinearity of the ERT inverse problem, the solution is not unique and the deterministic algorithm can remain trapped in local minima of the objective function. Additionally, the uncertainty of the solution can be assessed using a local approximation of the inverse of the Hessian matrix. For this reason, we focus on an inversion algorithm based on the Kalman filter, the ensemble smoother multiple data assimilation algorithm, which produces an ensemble of models as a solution that approximates the posterior probability density function. The algorithm iteratively updates the models by estimating the sensitivity information using correlation matrices. When the ensemble size is finite, the correlation matrices and then the Kalman filter, can be affected by spurious correlations, introducing errors in the sampling of the posterior probability density function sampling. To mitigate this issue, we follow a strategy commonly used in other fields, such as in reservoir history matching and meteorology, which involves a distance-based function to smooth out the components of the Kalman filter affected by spurious correlations. This strategy, known as localization, ensures the attenuation of model updates driven by spurious correlations. To assess the effectiveness of the proposed approach and evaluate the impact of ensemble size on the propagation of spurious correlations, we first invert synthetic data generated through a realistic resistivity model representing a typical unconformity-related uranium setting. We then apply the method to a field dataset from the Athabasca Basin, one of the World's major uranium plays. The inversion results illustrate the propagation of spurious correlations and its attenuation when the localization strategy is employed. Moreover, the localization allows for obtaining a reasonable solution even reducing the ensemble size. Finally, the field data inversion suggests the presence of at least two low-resistivity anomalies below the unconformity zone, starting at a depth of approximately 500 m. The posterior distribution we obtain with the proposed approach allows for the evaluation of the uncertainty, enabling a more reliable interpretation of the result. Overall, this paper highlights the potential of the localized ensemble-based inversion method for estimating the subsurface resistivity and quantifying the associated uncertainties in the resistivity model.

1 | Introduction

Electrical resistivity tomography (ERT) is a direct-current geophysical method used in hydrogeophysics, civil engineering and mining exploration to reconstruct the subsurface resistivity model. Due to the ease and flexibility of data acquisition, where a dipole injects current into the ground and a pair of voltage electrodes records the potential difference between two points along the survey profile, this method is widely used in various mining contexts for different targets (Dentith and Mudge 2014).

The reconstruction of the subsurface resistivity model is obtained by solving the ERT inverse problem, which is non-linear and ill-posed, through an iterative algorithm. The standard strategy to solve the problem involves the minimization of an objective function containing a data misfit term and model regularisation, which guarantees a rapid convergence even in high-dimensional cases (Loke et al. 2022). In particular, among the most commonly used algorithms in ERT are the Gauss–Newton (Lines and Treitel 1984; De Groot-Hedlin and Constable 1990; Sasaki 1992) and the quasi-Newton technique (Loke and Barker 1996), along with their variants, such as the damped Gauss–Newton method (Farquharson and Oldenburg 2004), the smoothness-constrained method (Loke et al. 2022) and the iterative reweighted least-squares method (Zhong et al. 2021), among others.

These methods employ a local linearization of the non-linear forward operator around the current model and use the resulting Jacobian (sensitivity) matrix to iteratively update the model parameters. However, since the ERT inverse problem is non-linear (Loke and Barker 1996), the objective function approximation depends on the considered model, leading to the non-uniqueness of the solution. In the mining context, where decisions have to be taken about drilling operations, the uncertainty estimation can play a crucial role in evaluating the resistivity contrasts. Although there exist strategies to estimate the uncertainty from the local minimization approaches based on the approximation of the model covariance from the Hessian matrix (Tarantola 2005; Auken et al. 2005), the results are still dependent on the local approximation of the objective function, which can lead to uncertainty underestimation and overconfidence in results (Sambridge and Mosegaard 2002).

One of the strategies to tackle these drawbacks is to reformulate the inverse problem in a Bayesian framework, considering the solution as a posterior probability density function (*pdf*) rather than as a single resistivity model (Galetti and Curtis 2018; Peng et al. 2023; Singh et al. 2024). This is performed by Bayesian strategies such as the Markov chain Monte Carlo algorithms (MCMC), which sample the posterior *pdf* without relying on an objective function (Sambridge and Mosegaard 2002). However, the number of samples required to assess the posterior *pdf*, and consequently the computational time, can increase dramatically with the number of model parameters. Research has been conducted to reduce the computational burden of MCMC inversion of ERT involving model space compression through discrete cosine transform (DCT) as proposed by Aleardi et al. (2021) and Vinciguerra et al. (2022). Nevertheless, due to the lossy nature of the DCT compression, the strategy brings loss of spatial resolution, which is undesirable when sharp contrasts of resistivity are expected, such as in mining applications. For this reason, we propose an alternative strategy for inverting ERT data within a Bayesian framework, based on a Kalman filter and ensemble-based method. This strategy is widely employed in various fields such as oceanography, meteorology, reservoir history matching and near-surface geophysics (Vinciguerra et al. 2024). In this work, we use the ensemble smoother multiple data assimilation (ES-MDA) algorithm, which approximates the posterior *pdf* by iteratively updating an ensemble of models drawn from a prior *pdf* (Emerick and Reynolds 2013; Emerick 2016). This approach integrates features of the ensemble smoother (ES), which performs a single Gauss–Newton iteration with a full step using the sensitivity approximated from the ensemble of models (Reynolds et al. 2006) and the multiple data assimilation, which performs multiple smaller corrections to the ensemble (Emerick and Reynolds 2013).

To apply corrections, the ES-MDA algorithm does not compute sensitivities analytically (e.g., via the adjoint state method), but instead approximates them using the cross-covariance matrix between the model ensemble and the predicted data. As a result, the sensitivity information depends on the ensemble size, which can introduce spurious correlations in the cross-covariance matrix and lead to errors in posterior *pdf* sampling. In other fields where the computational cost of the forward modelling imposes a constraint on the ensemble dimension, the spurious correlation issue has been tackled through the application of a localization procedure (Morzfeld and Hodyss 2023). For instance, in meteorological applications (Houtekamer and Mitchell 1998), the sampling error is reduced using data within a certain distance from the state variables, whereas Chen and Oliver (2010), in multiphase flow data assimilation, apply a distance-based localization that attempts to attenuate covariances beyond a predetermined distance.

In this work, we follow the strategy of applying localization to the Kalman filter to reduce the spurious correlation caused by a finite ensemble size (Chen and Oliver 2010; Chen and Oliver 2017). In particular, we limit the model parameters that are updated to those located within a particular distance from the observation location (Chen and Oliver 2010). In the context of mining exploration, considering the number of current electrodes and survey profiles, the number of unknowns of the problem can increase dramatically, leading to a rise in the computational cost of the forward modelling. Consequently, the computational burden of the ES-MDA algorithm can increase significantly, as forward modelling is computed for each ensemble member and each iteration. Usually, the choice of the ensemble size is a function of the model parameters and represents a trade-off between minimization of the computational time and accurate sampling of the posterior *pdf*. To attenuate the effect of spurious correlation, thus limiting the sampling error, we propose a distance-based localization based on the exponential correlation function that smooths out the Kalman filter components beyond a certain distance.

In this paper, we first introduce the ensemble-based methods and the localization approach employed. We then validate the algorithm using synthetic data generated from a resistivity model simulating an unconformity-related uranium deposit. Finally, we

apply the inversion approach to field data from the Athabasca Basin for uranium exploration.

2 | Theory and Methods

In this section, we will first review the ensemble-based approach, then describe the adaptive ensemble-based and the localization strategy employed.

2.1 | Ensemble-Based Approach

The ERT inverse problem aims at estimating a subsurface resistivity model $\mathbf{m}$ that is able to reproduce the observed data $\mathbf{d}_{obs}$ with the given error threshold. If we denote the forward operator that solves Poisson's differential equation by G then the predicted data can be written as

$$\mathbf{d}_{obs} = G(\mathbf{m}). \tag{1}$$

Assuming that prior information of the unknown model follows a Gaussian distribution with mean $\bar{\mathbf{m}}_{prior}$ and covariance between cells of the unknown model $\mathbf{C}_m$, we express the prior *pdf* as

$$p(\mathbf{m}) \propto \exp(-\frac{1}{2}(\mathbf{m}-\bar{\mathbf{m}}_{prior})^T\mathbf{C}_m^{-1}(\mathbf{m}-\bar{\mathbf{m}}_{prior})) \tag{2}$$

and the likelihood function $p(\mathbf{d}_{obs}|\mathbf{m})$ defined as

$$p(\mathbf{d}|\mathbf{m}) \propto \exp(-\frac{1}{2}(\mathbf{d}-\mathbf{d}_{obs})^T\mathbf{C}_d^{-1}(\mathbf{d}-\mathbf{d}_{obs})), \tag{3}$$

where $\mathbf{C}_d$ is the data covariance matrix. Following the Bayes theorem, the posterior probability density function $p(\mathbf{m}|\mathbf{d}_{obs})$ is expressed by

$$p(\mathbf{m}|\mathbf{d}_{obs}) \propto p(\mathbf{m})p(\mathbf{d}_{obs}|\mathbf{m}). \tag{4}$$

In case of a non-linear forward operator, like in ERT, the goal is to find the unknown model that maximizes the *pdf* of Equation (4). It can be demonstrated (Tarantola 2005) that maximizing $p(\mathbf{m}|\mathbf{d}_{obs})$ is equivalent to minimizing the cost function $\Phi(\boldsymbol{m})$ defined as (see Equations 2 and 3):

$$\begin{aligned}\Phi(\mathbf{m}) = \frac{1}{2}&((\mathbf{m}-\bar{\mathbf{m}}_{prior})^T\mathbf{C}_m^{-1}(\mathbf{m}-\bar{\mathbf{m}}_{prior})\\ &+(\mathbf{d}-\mathbf{d}_{obs})^T\mathbf{C}_d^{-1}(\mathbf{d}-\mathbf{d}_{obs})).\end{aligned} \tag{5}$$

Since in this work, we aim to sample the posterior *pdf* rather than only find its maximum, we employ an ensemble-based approach that allows us to consider an ensemble of cost functions. Defining $\mathbf{m}_j$ as the jth ensemble member and the perturbed data $\mathbf{d}_{pert} = \mathbf{d}_{obs} + \epsilon$, where ϵ is drawn from the normal distribution $\mathcal{N}(0, \mathbf{C}_d^{1/2})$ (Emerick and Reynolds 2013), where $\mathbf{C}_d$ is the data covariance matrix. The jth cost function $\Phi(\mathbf{m}_j)$ is

$$\begin{aligned}\Phi(\mathbf{m}_j) = \frac{1}{2}&((\mathbf{m}_j-\bar{\mathbf{m}}_{prior})^T\mathbf{C}_m^{-1}(\mathbf{m}_j-\bar{\mathbf{m}}_{prior})\\ &+(\mathbf{d}_j-\mathbf{d}_{pert})^T\mathbf{C}_d^{-1}(\mathbf{d}_j-\mathbf{d}_{pert})),\end{aligned} \tag{6}$$

where j is the jth ensemble member. Employing the Gauss–Newton algorithm to minimize each cost function, we obtain (Reynolds et al. 2006; Evensen et al. 2022)

$$\mathbf{m}_j^{i+1} = \mathbf{m}_j^i + \mathbf{C}_m\mathbf{J}^T(\mathbf{J}\mathbf{C}_m\mathbf{J}^T + \mathbf{C}_d)^{-1}(\mathbf{d}_{pert}-\mathbf{d}_j), \tag{7}$$

where $\mathbf{J} = \frac{\partial \mathbf{d}}{\partial \mathbf{m}}$ is the Jacobian matrix, which expresses the sensitivity of the data $\mathbf{d}$ with respect to the model $\mathbf{m}$. Reynolds et al. (2006) demonstrated that the ensemble Kalman filter (EnKF) represents an approximation of Equation (7) expressed by

$$\mathbf{m}_j^{i+1} = \mathbf{m}_j^i + \mathbf{C}_{md}(\mathbf{C}_{dd} + \mathbf{C}_d)^{-1}(\mathbf{d}_{pert}-\mathbf{d}_j), \tag{8}$$

where $\mathbf{C}_{md}$ is the cross covariance matrix between the ensemble of models and the data (dimension $M \times N$), whereas $\mathbf{C}_{dd}$ (dimension $N \times N$) is the covariance matrix of the ensemble of data. They are defined as

$$\mathbf{C}_{md} = \frac{1}{N_e-1}\sum_{j=1}^{N_e}(\mathbf{m}_j-\bar{\mathbf{m}})(\mathbf{d}_j-\bar{\mathbf{d}})^T, \tag{9}$$

$$\mathbf{C}_{dd} = \frac{1}{N_e-1}\sum_{j=1}^{N_e}(\mathbf{d}_j-\bar{\mathbf{d}})(\mathbf{d}_j-\bar{\mathbf{d}})^T, \tag{10}$$

where N_e is the ensemble dimension, $\bar{\mathbf{m}}$ and $\bar{\mathbf{d}}$ are the mean of the model and data ensemble and $\mathbf{m}_j$ and $\mathbf{d}_j$ a member of the ensemble of models and data, respectively.

Comparing Equations (7) and (8), it is evident that

$$\mathbf{C}_{md} \approx \mathbf{C}_m\mathbf{J}^T, \tag{11}$$

$$\mathbf{C}_{dd} \approx \mathbf{J}^T\mathbf{C}_m\mathbf{J}. \tag{12}$$

For non-dynamical systems, where the physical simulation is a stable function of subsurface properties, the EnKF equation coincides with ES, which is equivalent to a single Gauss–Newton iteration with full step and averaged sensitivity computed from the prior ensemble (Reynolds et al. 2006; Emerick and Reynolds 2013). For non-linear problems, instead of using the ES approach, it is convenient to employ an iterative ensemble algorithm that performs small corrections to the ensemble (multiple data assimilation; Emerick and Reynolds 2013), leading to the ES-MDA algorithm.

2.2 | Adaptive ES-MDA

ES-MDA samples the posterior *pdf* updating an ensemble of models drawn from a Gaussian prior *pdf*. Since the assumption of this strategy is the Gaussianity of the prior model parameters distribution (see Emerick and Reynolds 2013 for details), we build a prior ensemble of models drawing realizations from a prior distribution $N(\mathbf{m}_{prior}, \mathbf{C}_m)$, defining the model covariance $\mathbf{C}_m$ and through a spatial variogram.

The ensemble is updated through a relation similar to Equation (8) with an inflated data covariance:

$$\mathbf{m}_j^{i+1} = \mathbf{m}_j^i + \mathbf{C}_{md}(\mathbf{C}_{dd} + \alpha_i\mathbf{C}_d)^{-1}(\mathbf{d}_{pert}-\mathbf{d}_j), \tag{13}$$

where i is the ith update, j is the jth ensemble member and α_i is the coefficient that inflates a diagonal data covariance matrix.

Introducing the Kalman filter $\mathbf{K}$, we obtain

$$\mathbf{m}_j^{i+1} = \mathbf{m}_j^i + \mathbf{K}(\mathbf{d}_{pert} - \mathbf{d}_j), \tag{14}$$

where the matrix $\mathbf{K}$ has dimension $M \times N$, M is the number of unknowns and N is the number of data points.

The role of inflating the data covariance is to dampen the updates at early iterations (Emerick and Reynolds 2013; Emerick 2016) when we may have a large data misfit. For this reason, it is reasonable to expect, assuming convergence of the algorithm, a decreasing damping factor across the iterations that, in the case of ES-MDA, has to satisfy the condition (Emerick and Reynolds 2013):

$$\sum_{i=1}^{N_{it}} \frac{1}{\alpha_i} = 1, \tag{15}$$

where N_{it} is the number of iterations. The simplest strategy is to define α_i as a constant and is equal to the reciprocal of the iteration number. However, Le et al. (2016) and Rafiee and Reynolds (2017) suggest that this simple choice can produce too large updates at initial iterations. For this reason, Le et al. (2016) and Emerick (2016) proposed an automatic selection of the α_i coefficient according to the evolution of data misfit but satisfying Equation (15). In particular, the parameter is selected based on the progress of the objective function, which we define as

$$O(\mathbf{m}) = \frac{1}{2N} \sum_{i=1}^{N_e} (\mathbf{d}_i - \mathbf{d}_{obs})^T \mathbf{C}_d^{-1} (\mathbf{d}_i - \mathbf{d}_{obs}) \tag{16}$$

and the normalized objective function

$$\bar{O}(\mathbf{m}) = \frac{1}{N_e} \sum_{i=1}^{N_e} O(\mathbf{m}_i) \tag{17}$$

at each iteration, α_i will be equal to $\bar{O}(\mathbf{m})$. The adaptive choice of the inflation coefficient leads to the adaptive ES-MDA (Algorithm 1) (Le et al. 2016; Vinciguerra et al. 2024).

When the ES-MDA reaches its last iteration, it satisfies a convergence criterion based on the objective function value (Equation 16) or meets the condition defined by Equation (15) and the resulting ensemble of models approximates the posterior *pdf*. However, due to the finite ensemble size, the matrices $\mathbf{C}_{md}$ and $\mathbf{C}_{dd}$ may be contaminated by spurious correlations, which can propagate through the Kalman filter, introducing sampling errors.

2.3 | Localization

Localization strategies aim at reducing the effect of spurious correlations that contaminate the covariance matrices and Kalman filter due to small ensemble size approximation (Chen and Oliver 2010). To better understand the impact of spurious correlations, it is useful to express the matrix $\mathbf{C}_{md}$ (Equation 11) explicitly as

$$\mathbf{C}_{md} = \begin{bmatrix} C_{m^0 d^0} & C_{m^0 d^1} & \dots & C_{m^0 d^n} \\ C_{m^1 d^0} & C_{m^1 d^1} & \vdots & C_{m^1 d^n} \\ \vdots & \cdots & \ddots & \vdots \\ C_{m^m d^0} & C_{m^m d^1} & \cdots & C_{m^m d^n} \end{bmatrix}, \tag{18}$$

where each element $C_{m^i d^j}$ expresses the covariance between the ith cell of the model and the jth apparent resistivity value of the data. From the expression of the cross-covariance matrix (Equation 9), we expect almost zero correlation between cells and apparent resistivity data referring to distant portions of the half-space. In other words, we expect that the ith cell of the model will not have an impact on the jth data point if their distance is sufficiently large. This is consistent with Equation (11), which indicates that $\mathbf{C}_{md}$ is approximately proportional to the sensitivity matrix. However, when the ensemble dimension is not infinite, the matrix is contaminated by spurious values, which can produce sampling errors in the posterior pdf. At the same time, the matrix $\mathbf{C}_{dd}$ can be written as

$$\mathbf{C}_{dd} = \begin{bmatrix} C_{d^0 d^0} & C_{d^0 d^1} & \dots & C_{d^0 d^n} \\ C_{d^1 d^0} & C_{d^1 d^1} & \vdots & C_{d^1 d^n} \\ \vdots & \cdots & \ddots & \vdots \\ C_{d^d d^0} & C_{d^d d^1} & \cdots & C_{d^d d^n} \end{bmatrix}, \tag{19}$$

where the diagonal contains the variance of each data point whereas the off-diagonal elements represent the covariance between each data point computed on the ensemble of apparent resistivity data. Even in this case, we expect that, as the spatial distance between apparent resistivity values increases, the covariance decreases.

In the literature, localization is primarily applied in two ways: to the cross-covariance matrix $\mathbf{C}_{md}$ and $\mathbf{C}_{dd}$ to mitigate spurious correlations (Equation 13) and directly to the Kalman filter $\mathbf{K}$ (Equation 14). The former approach applies localization separately to $\mathbf{C}_{md}$ and $\mathbf{C}_{dd}$ (Emerick and Reynolds 2011), while the latter operates directly on the Kalman filter (Chen and Oliver 2017).

Since the assignment of ERT data to subsurface points is mainly done conventionally, for instance, between current and voltage dipoles at an angle of 45° with the horizontal, evaluating the spatial covariance of the data as a function of the distance becomes arbitrary. Consequently, we apply the localization directly to the Kalman filter, which shares the same dimensions as the sensitivity matrix (M × N). This allows for straightforward column-by-column evaluation. Adopting this strategy means rewriting Equation (14) as

$$\mathbf{m}_j^{i+1} = \mathbf{m}_j^i + \mathbf{\Gamma} \circ \mathbf{K}(\mathbf{d}_{pert} - \mathbf{d}_j), \tag{20}$$

where $\circ$ indicates the Schur product, a simple element-by element multiplication, which implies the localization matrix $\mathbf{\Gamma}$ has the same dimension as the Kalman filter. Each element of $\mathbf{\Gamma}$ represents the assumed correlation between each model cell and data point, as a function of the distance between them. In this work, we investigate the effect of a simple exponential correlation function, which leads to comparable results to more sophisticated

ALGORITHM 1 | Localized Adaptive ES-MDA

1: Input: N_e, N_{it}, α_1 → Set the number of ensemble members
2: Input: $\mathbf{M} \sim \mathcal{N}(\mathbf{m}_{prior}, \mathbf{C}_m)$ → Draw models from Gaussian *pdf*
3: Input: Localization matrix $\mathbf{\Gamma}$
4: Input: $\bar{O}(\mathbf{M})$ → Average objective function of $\mathbf{M}$
5: Input: Localization True or False → ES-MDA with or without localization
6: **for** $i = 1 : N_{it}$ **do** → Loop over the assimilation steps
7: $\alpha_i = \bar{O}$
8: **if** $\sum(1/\alpha_i) == 1$ **then** break
9: **else if** $\sum(1/\alpha_i) > 1$ or $i == N_{it}$ **then** $\alpha_i = (1 - \sum(1/\alpha_{i-1}))^{-1}$
10: **end if**
11: $\mathbf{C}_{dd} = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (\mathbf{d}_j - \bar{\mathbf{d}})(\mathbf{d}_j - \bar{\mathbf{d}})^T$ → Covariance matrix
12: $\mathbf{C}_{md} = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (\mathbf{m}_j - \bar{\mathbf{m}})(\mathbf{d}_j - \bar{\mathbf{d}})^T$ → Cross-covariance matrix
13: $\mathbf{K} = \mathbf{C}_{md}(\mathbf{C}_{dd} + \alpha_i \mathbf{C}_d)^{-1}$ → Kalman Filter
14: **for** $j = 1 : N_e$ **do** → Loop over the ensemble
15: $\tilde{\mathbf{d}}_j = \mathbf{d}_{obs} + \sqrt{\alpha_i \mathbf{C}_d}\mathbf{n}$ → Data perturbation $\mathbf{n} \sim N(0, \mathbf{I})$
16: **if** Localization == True **then**
17: $\mathbf{m}_j^i = \mathbf{m}_j^{i-1} + \mathbf{\Gamma} \circ \mathbf{K}(\tilde{\mathbf{d}} - G(\mathbf{m}_j^{i-1}))$ → Model updating with localization
18: **else if** Localization == False **then**
19: $\mathbf{m}_j^i = \mathbf{m}_j^{i-1} + \mathbf{K}(\tilde{\mathbf{d}} - G(\mathbf{m}_j^{i-1}))$ → Model updating without localization
20: **end if** →
21: $O(\mathbf{m}_j) = \frac{1}{2N} \sum_{j=1}^{N_e} (\mathbf{d}_j - \mathbf{d}_{obs})^T \mathbf{C}_d^{-1} (\mathbf{d}_j - \mathbf{d}_{obs})$ → Objective function
22: **end for**
23: $\bar{O}(\mathbf{m}) = \frac{1}{N_e} \sum_{i=1}^{N_e} O(\mathbf{m}_j)$ → Average objective function
24: **end for**

localization strategies (Morzfeld and Hodyss 2023). The matrix Γ is expressed as

$$\mathbf{\Gamma} = \begin{bmatrix} \Gamma_{m_0,d_0} & \Gamma_{m_0,d_1} & \cdots & \Gamma_{m_0,d_n} \\ \Gamma_{m_1,d_0} & \Gamma_{m_1,d_1} & \vdots & \Gamma_{m_1,d_n} \\ \vdots & \cdots & \ddots & \vdots \\ \Gamma_{m_m,d_0} & \Gamma_{m_m,d_1} & \cdots & \Gamma_{m_m,d_n} \end{bmatrix}. \tag{21}$$

In the case of an exponential correlation function, the first element on the diagonal is

$$\Gamma_{m_i,d_i} = \mathrm{e}^{-\left(\frac{dist}{R}\right)^k}, \tag{22}$$

where *dist* is the Euclidean distance between the cell m_i and data point d_i, R is the range expressing the correlation length and k is the order. Since this function reaches its maximum value (equal to 1) at zero distance and decreases as the distance increases (see Figure 1), it can be used as a weighting function to smooth out the undesirable values. In this work, the domain of interest is bidimensional, with range R defined along x- and z-directions. We adopt a formulation in which $\mathbf{\Gamma}$ depends on the two-dimensional (2D) distance between the model and the data, as described in previous studies (Hamill et al. 2001; Chen and Oliver 2010; Emerick and Reynolds 2011; Raniolo et al. 2013). Figure 2a illustrates an example of an exponential localization function with k equal to 3, the range along x and z, R_x and R_z, equal to 1500 m (Figure 2b), representing a column of the localization matrix $\mathbf{\Gamma}$ (Equation 21). The maximum value is close to the surface where we measure the apparent resistivity value (at the distance of 0 m along the profile), whereas the value decreases as we move away from this point. There exist various strategies to define the range, which involve prior assumptions on correlation distance (Chen and Oliver 2010), and computation of the sensitivity for linear problems (Emerick and Reynolds 2011). Since a too restrictive or too large choice of the localization region may lead to unreasonably large local updates on each ensemble member (Emerick and Reynolds 2011), we employ a simple strategy of defining an isotropic localization function with ranges R_x and R_z equal to the quadrupole dimension. This strategy, along with the choice of the order of the function ($k = 3$ in Figure 1 and Equation 22), should attenuate the risk of choosing a region of influence unreasonably smaller than the real one. On the other hand, determining the range as a function of the computed sensitivity (e.g., using finite differences) for the entire ensemble would result in a dramatic increase in computational cost. The localized adaptive ES-MDA is expressed in Algorithm 1.

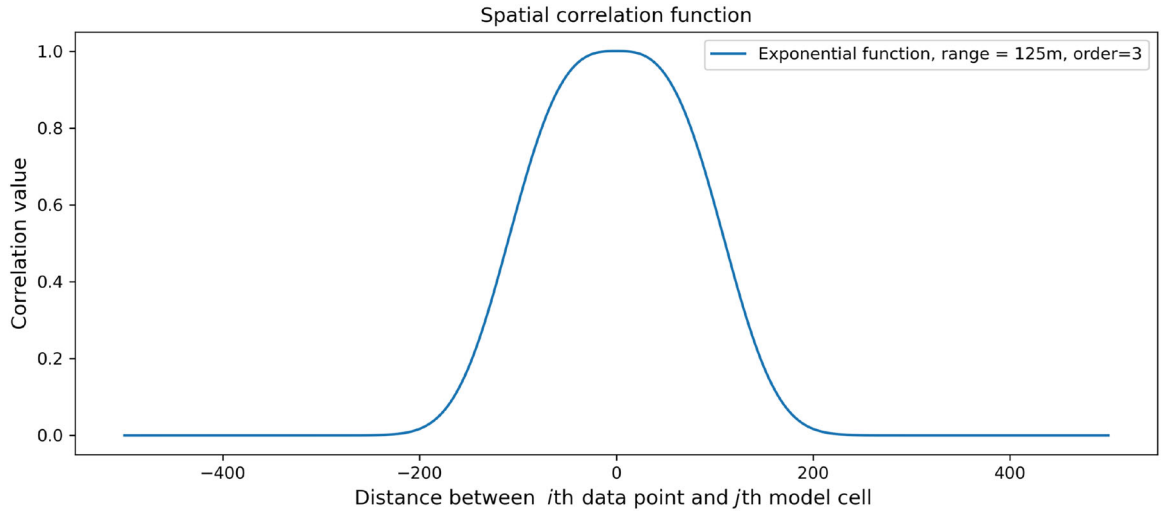

Figure 1 | Three spatial correlation functions in the one-dimensional case. Note that the x-axis represents the distance between a cell of the model and a data point (in one dimension).

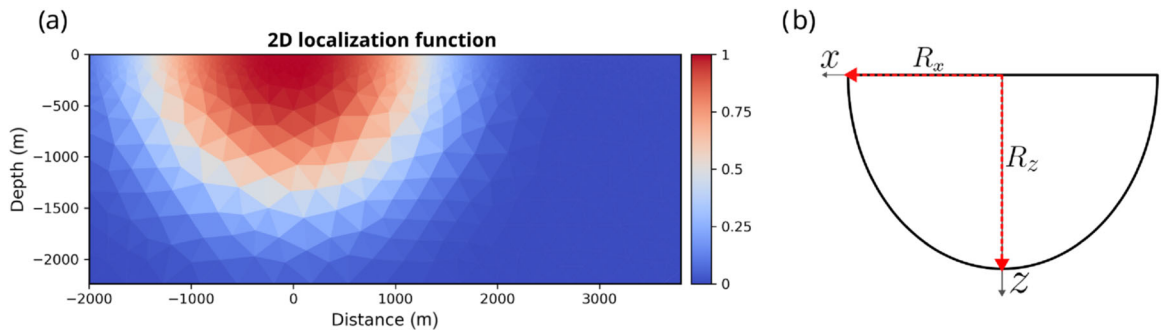

Figure 2 | (a) 2D localization function computed through the exponential correlation of (22) with order $k = 3$. Note that in this case, the range along x (R_x) and z (R_z) is coincident and equal to 1500 m. (b) Scheme of localization function. R_x represents the range along x and R_z the range along z.

3 | Synthetic Test

We evaluate the proposed strategy by inverting a synthetic dataset generated using the Python library pyGIMLi (Rücker et al. 2017). We begin by describing the data generation process, followed by the construction of the prior ensemble of models. We then present examples of Kalman filter contamination due to spurious correlations and conclude with the inversion of the synthetic dataset.

3.1 | Generation of Synthetic Data

Since we aim to apply the proposed approach for uranium exploration, we build a realistic geological scenario to generate synthetic data. Indeed, taking advantage of the vast amount of knowledge about one of the world's major uranium plays (Mir et al. 2022), the Athabasca Basin (Canada), we build a representative resistivity model.

The high-grade uranium deposits occur close to the unconformity between the basin (sandstone) and basement rocks, both characterized by high-resistivity values (Mir et al. 2022). Specifically, we model the late Paleoproterozoic to Mesoproterozoic sandstone with a resistivity of 3000 $\Omega \cdot$ m and the Archean and Paleoproterozoic basement with a resistivity of 5000 $\Omega \cdot$ m, following the work of Mir et al. (2022). Uranium mineralization is typically associated with altered sandstone and shear zones reactivated under brittle conditions that acted as fluid conduits during mineralization and alteration events (Jefferson et al. 2007; Mir et al. 2022). These reactivated shear zones typically include graphitic metapelitic characterized by extremely low resistivity (few $\Omega \cdot$ m), becoming a conductive target for electromagnetic and ERT exploration. To build the resistivity model, we construct a domain with a fine, irregular mesh comprising 4623 triangular cells. Specifically, we generate an initial high-resistivity layer (3000 $\Omega \cdot$ m) approximately 500 m thick located above a basement (5000 $\Omega \cdot$ m) that contains three graphitic bodies of resistivity 1 $\Omega \cdot$ m. In addition, we model a small alteration halo with a resistivity of about 100 $\Omega \cdot$ m (see Figure 3a).

To make the synthetic data as much realistic as possible, we simulate a 2D pole–dipole acquisition inspired by a real survey performed in the Athabasca Basin. In particular, we employ voltage dipoles of constant length (equal to 100 m) and progressively increase the distance between electrode B and the

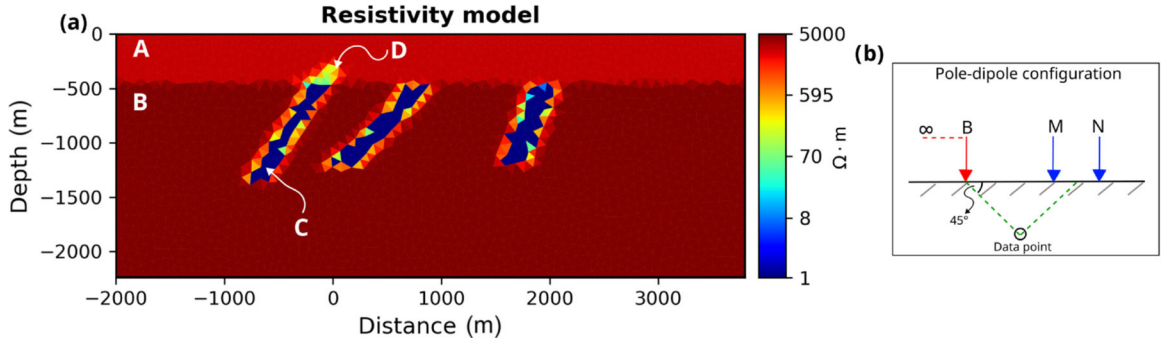

Figure 3 | (a) Resistivity model composed of sandstone (A), basement (B), three conductive bodies (C) and alteration halo (D). (b) Scheme of pole–dipole configuration.

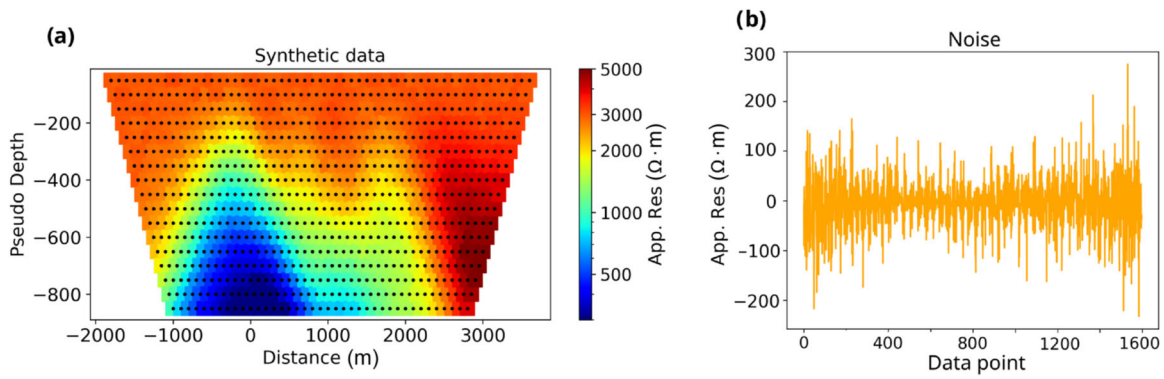

Figure 4 | (a) Synthetic data generated from the resistivity model in Figure 3. (b) White noise contaminating the synthetic data.

dipole centre up to 1650 m (Figure 3b). The apparent resistivity measurement is assigned to the point midway between the MN dipole (Figure 3b) and the electrode B. This point lies at the intersection of two lines subtending an angle of 45 degrees from the centre of MN and from the current electrode (Figure 3b). Figure 4c illustrates the simulated data as a pseudosection. The model clearly produces low apparent resistivity values in the bottom-left portion of the pseudosection, while high apparent resistivity values are observed around the low-resistivity anomaly. In order to make the simulated data more realistic and to do not penalize low apparent resistivity values too much, we add uncorrelated Gaussian noise with a standard deviation equal to 2% of each data point (Figure 4b).

3.2 | Prior Ensembles

One of the inputs of the ES-MDA algorithm is represented by an ensemble of models that incorporates prior knowledge expressed in terms of resistivity. We generate the ensemble through a random multivariate simulator, assuming a model covariance matrix that contains the variance of each model cell and a spatial correlation. The non-zero elements represent the covariance between parameters that we build through an exponential variogram defined as

$$\mathbf{C} = \exp((-\tilde{\mathbf{r}}/R)^k), \tag{23}$$

where k is the order of the exponential that we fix equal to 2, $\tilde{\mathbf{r}}$ is the Euclidean distance between each cell and R is the range that we fix equal to 250 m for both x and z directions. The choice of an isotropic variogram is motivated by the assumption, reasonable in the field case, that no prior information is available about the orientation of the anomalies, especially in the case of alteration halos. Moreover, the short correlation range, compared with the extension of the model in depth, allows for the generation of models with sharp resistivity contrasts. Since the resistivity values span several orders of magnitude, we apply a two-step transformation in the model space, preventing unreasonably high-resistivity values. First, we compute the logarithm of the resistivity model and then we constrain the values with predefined minimum and maximum bounds, consistent with the expected resistivity range of the true model, following the approach described by (Günther and Rücker 2012)

$$\mathbf{m}_T = \log(\mathbf{m} - lb) - \log(ub - \mathbf{m}), \tag{24}$$

$$\mathbf{m} = \frac{(lb + ub \cdot \exp(\mathbf{m}_T))}{(1 + \exp(\mathbf{m}_T))}, \tag{25}$$

where $\mathbf{m}_T$ is the transformed model, lb and ub are the lower and upper bounds and $\mathbf{m}$ is the anti-transformed value. Considering the expected resistivity range, we set the lower bound to 1 $\Omega \cdot$ m and the upper bound to 10^4 $\Omega \cdot$ m, which constrains the values while still allowing for large variability.

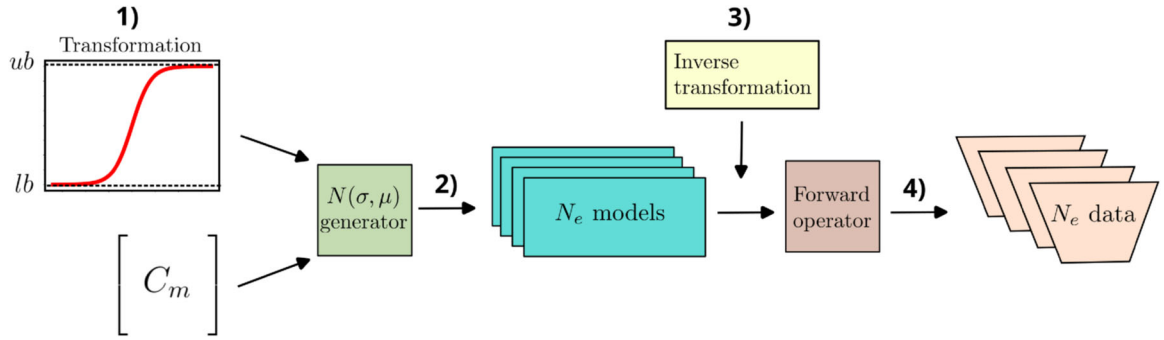

Figure 5 | Scheme of the ensemble generation. (1) Definition of the bounds for the transform and model covariance matrices. (2) Generation of N_e models from a multivariate Gaussian distribution. (3) The N_e models are anti-transformed. (4) Calculation of the ensemble of predicted data through the forward operator.

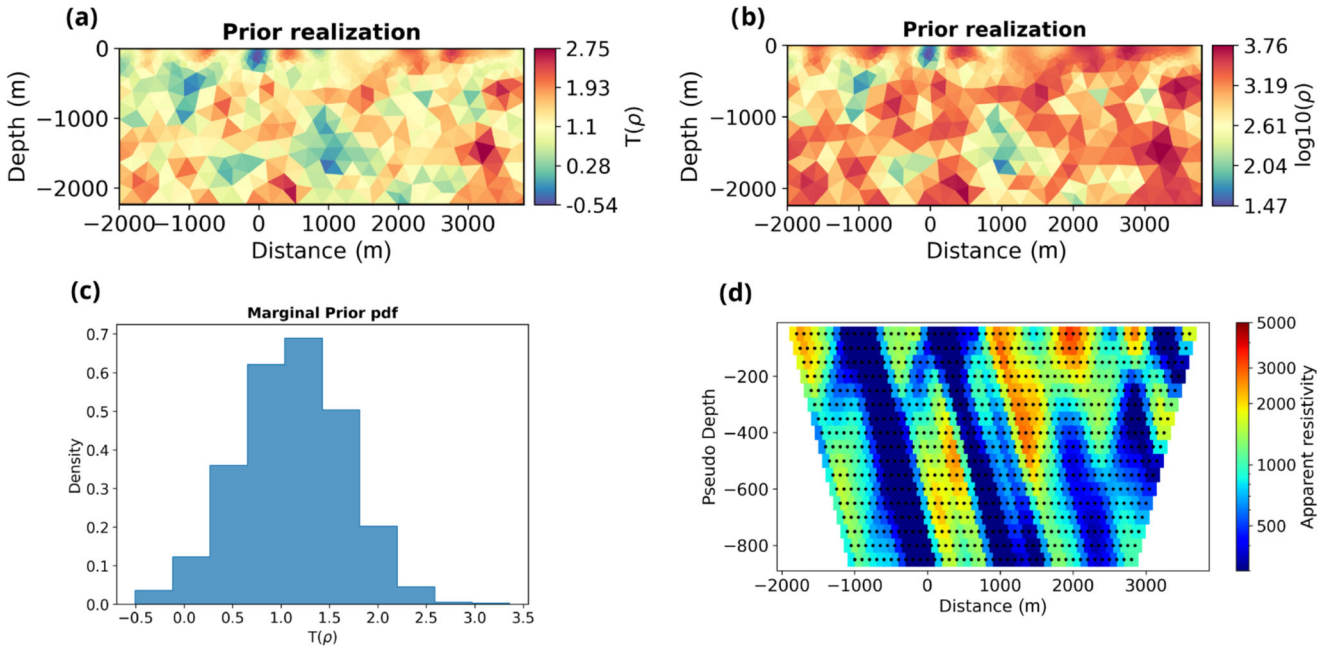

Figure 6 | (a) Model drawn from prior distribution (transformed domain). Note that the model is smoother close to the surface where the cells are smaller. (b) The same model in log10 domain. (c) Marginal pdf associated with a cell of the model. (d) Data associated with the model in Figure (a). Comparing it with the data in Figure 4a, we note how distant the data are from the synthetic observed pseudosection.

After applying the transformations, the models are drawn from a multivariate normal generator in order to preserve the Gaussianity of the prior ensemble of models, which is one of the assumptions for ES-MDA (Emerick and Reynolds 2013). The prior mean is fixed at $10^3\ \Omega \cdot$ m for each cell, and the standard deviation is chosen to ensure that the resistivity values remain within the desired bounds.

To compute the prior data ensemble, we first apply the inverse transformation to the models and subsequently apply the forward operator. Figure 5 exhibits the workflow for model and data generation. The result of this process is an ensemble of prior models and the corresponding ensemble of predicted data, which together serve as inputs to the ES-MDA algorithm. Figure 6a and 6b shows two realizations in the transformed and in the logarithmic domain, respectively. The models exhibit smoother variations in the upper portion, where the triangular cells are small, and become less smooth at greater depths due to the increasing cell size. Each cell of the model is associated with a marginal prior distribution as shown in Figure 6c. For each model in the prior ensemble, we compute the associated forward response (Figure 6d), thus building the ensemble of predicted data. The choice of the ensemble size and its influence on spurious correlations are examined in the next subsection.

3.3 | Spurious Correlations and Choice of the Localization Function

This section investigates the effects of ensemble size on the Kalman filter $\mathbf{K}$ and the cross-covariance matrices $\mathbf{C}_{md}$, both of which can introduce errors in sampling the posterior *pdf*.

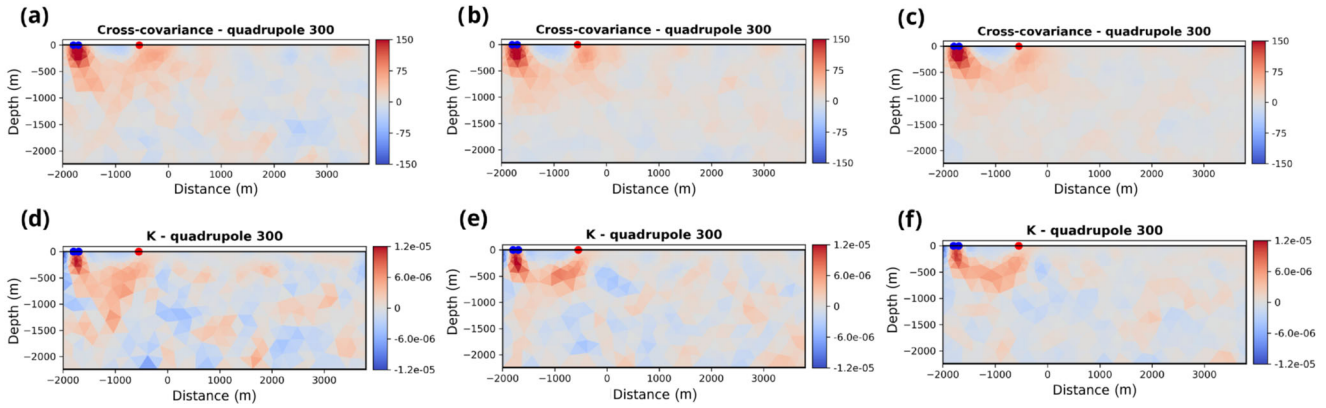

Figure 7 | (a) Column 300 of the $\mathbf{C}_{md}$ matrix at the first iteration, ensemble size 500. (b) Column 300 of the Kalman filter at the first iteration, ensemble size 1000. (c) Column 300 of the C_{md} matrix at the first iteration, ensemble size 2000. (d) Column 300 of the Kalman filter at the first iteration, ensemble size 500. (e) Column 300 of the Kalman filter at the first iteration, ensemble size 1000. (f) Column 300 of the Kalman filter at the first iteration, ensemble size 2000. The blue dots represent the voltage electrodes, whereas the red dot illustrates the current electrode.

Although this aspect of ES-MDA has not been explicitly addressed in the literature, it becomes relevant when attempting to reduce the computational costs by decreasing the ensemble size.

The cross-covariance matrices $\mathbf{C}_{md}$, $\mathbf{C}_{dd}$ and then the Kalman filter $\mathbf{K}$ are functions of the ensemble, as defined by Equations (9) and (10). Therefore, the number of models and dataset pairs forming the ensemble affects the approximation of sensitivity information (Equations 11 and 12) and may lead to the introduction of spurious correlations (Chen and Oliver 2017) that contaminate the matrices and the Kalman filter. In the absence of a well-defined rule, the ensemble size is typically chosen proportionally to the number of unknown model parameters (cells). We define a coarser mesh, consisting of larger triangular cells (1932) than those used in the data simulation.

To investigate the effects of ensemble size on the covariance matrices, we generate three ensembles of 500, 1000 and 2000 models, all drawn from the same prior distribution. We then compute the covariance matrices at the first iteration. To analyse their characteristics, we extract a representative column from the cross-covariance $\mathbf{C}_{md}$ and Kalman filter $\mathbf{K}$ (Figure 7). Since both $\mathbf{C}_{md}$ and $\mathbf{K}$ have the same dimension $M \times N$, each column can be represented in the model space with a specific data point measured by a corresponding quadrupole. For every ensemble size, column 300 of the cross-covariance shows maximum and minimum values between the potential dipole (blue dots) and current (red dots) pole, which is consistent with the fact that the matrix approximately depends on the sensitivity (Equation 11). However, at greater distances from the current and potential dipoles, the cross-covariance values do not decay as expected. Instead, they appear scattered, with non-negligible values persisting even at considerable distances, indicating the presence of spurious correlations. As evident from the comparison between Figure 7a–c, although these effects are more pronounced with an ensemble size of 500, they remain noticeable even with 1000 and 2000 members. We interpret these features as spurious correlations that affect the covariance matrices when a small ensemble size is used. These spurious correlations, along with those that may be present in the covariance matrix $\mathbf{C}_{dd}$, further affect the Kalman filter, as illustrated in Figure 7d–f. Similar to the cross-covariance, column 300 of the Kalman filter is affected by spurious values at large distances from the corresponding quadrupole, for all ensemble sizes (500, 1000 and 2000). As suggested by Equation (14), the effect on the model updates becomes clear: a contaminated Kalman gain, when applied to the difference between observed and predicted data, leads to large and spatially scattered updates even in regions of low sensitivity. This, in turn, introduces errors in the sampling of the *pdf*.

The localization strategy we employ allows us to tackle the effects of spurious correlations far away from the considered quadrupole. For this purpose, we model a localization function according to the dimension of each quadrupole (without considering the remote electrode). In particular, we employ an exponential variogram of order $k = 3$ having the range R_x and R_y is equal to the distance between electrode B and N (Figure 3b). In this way, we obtain an isotropic localization function that changes its range depending on the quadrupole dimension. Figure 8 illustrates two examples of localization applied to two different quadrupoles for an ensemble of size 500. The function corresponding to quadrupole 300 shows maximum values, up to 1, close to the electrodes and zero far away from them (Figure 8a). Consequently, the effect on the Kalman filter (Figure 8b) is to attenuate the effects of the spurious correlation, while maintaining the filter values close to the quadrupole (Figure 8c). Figure 8d illustrates the localization function associated with quadrupole number 1200, which is able to attenuate the undesired correlations (Figure 8e) and allows for obtaining the localized Kalman filter represented in Figure 8f. In the following section, we discuss the effects of localization on synthetic data inversion.

3.4 | Synthetic Inversion Results

In order to test the approach, we invert the synthetic data shown in Figure 4a using three different ensemble dimensions (500, 1000, 2000) by applying the localization function described in the previous section. The prior realizations are drawn from the prior distribution described in Section 3.2, and parameters are reported in Table 1. In all the tests performed, the algorithm converges in nine iterations, achieving a root mean square error (RMSE) of 4%

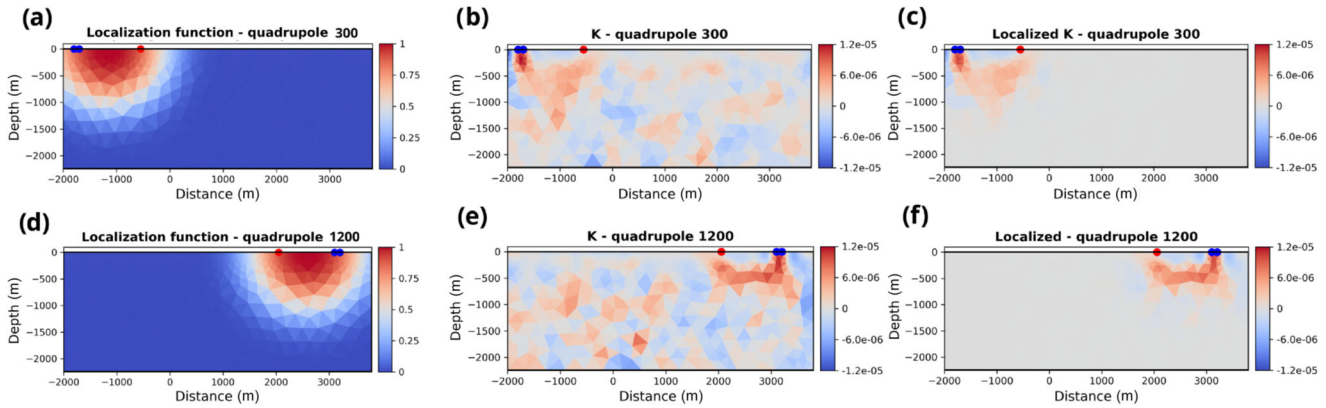

Figure 8 | (a) Localization function for the quadrupole number 300. (b) Column 300 of the Kalman filter. (c) Localized Kalman filter, column 300. (d) Localization function for the quadrupole number 1200. (e) Column 1200 of the Kalman filter. (f) Localized Kalman filter, column 1200.

Table 1 | Parameters concerning the synthetic data inversion. μ indicates the mean, σ the standard deviation of the prior distribution and R the range of the spatial correlation.

Localization	Prior ensemble				Inversion		
	Ensemble size	μ^a	σ^{2a}	R (m)	Iterations	Mean rmse (%)	Computational time (h)
Yes	500	1.09	0.3	250	9	4.02	1.6
Yes	1000	1.09	0.3	250	9	3.97	3.23
Yes	2000	1.09	0.3	250	9	4.35	6.48
No	500	1.09	0.3	250	6	2.55	1.06
No	1000	1.09	0.3	250	6	2.49	2.12
No	2000	1.09	0.3	250	6	2.49	2.12

[a] Transformed domain.

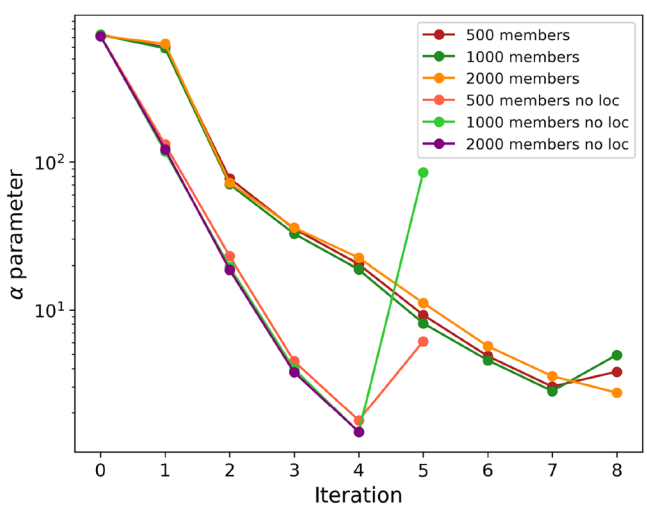

Figure 9 | Parameter α for different ensemble sizes, with and without localization. Note that the increase of α at the last iteration follows from satisfying Equation (15).

across the ensemble sizes (details in Table 1). Additionally, the parameter α, representing the average normalized objective function (Line 6 in Algorithm 1), exhibits a similar decreasing trend, further suggesting the convergence of the algorithm (Figure 9).

Although from the data misfit point of view, the three tests return similar results. The computational cost is much higher when using 2000 ensemble members (6.48 h) than when using 1000 (3.23 h) or 500 (1.6 h) for a laptop equipped with Intel Core i7-1165G7 @ 2.80 GHz.

The ensemble of models obtained at the end of the inversion approximates the posterior (*pdf*), from which the mean model and associated uncertainty are estimated. The mean models corresponding to the three different ensemble sizes suggest that in all the cases the localized ES-MDA is able to recover two high-resistivity domains (sandstone and basement) and the low resistive bodies (Figure 10a–c). In particular, the first resistivity layer is about 500 m thick with lower resistivity than the underlying domain representing the basement, which contains the graphitic bodies. Although, as expected, the dip of the graphitic bodies is not fully reconstructed, their position along the profile is well recovered with good horizontal and spatial resolution. The substantial similarity among the mean models suggests that, due to the localization of the Kalman filter, the ensemble size does not significantly affect the mean of the posterior *pdf*, either in terms of the anomaly locations or their spatial resolution.

The application of the ES-MDA algorithm without localization results in heavily scattered mean models, especially around and below the position of the conductive bodies (Figure 10d–f).

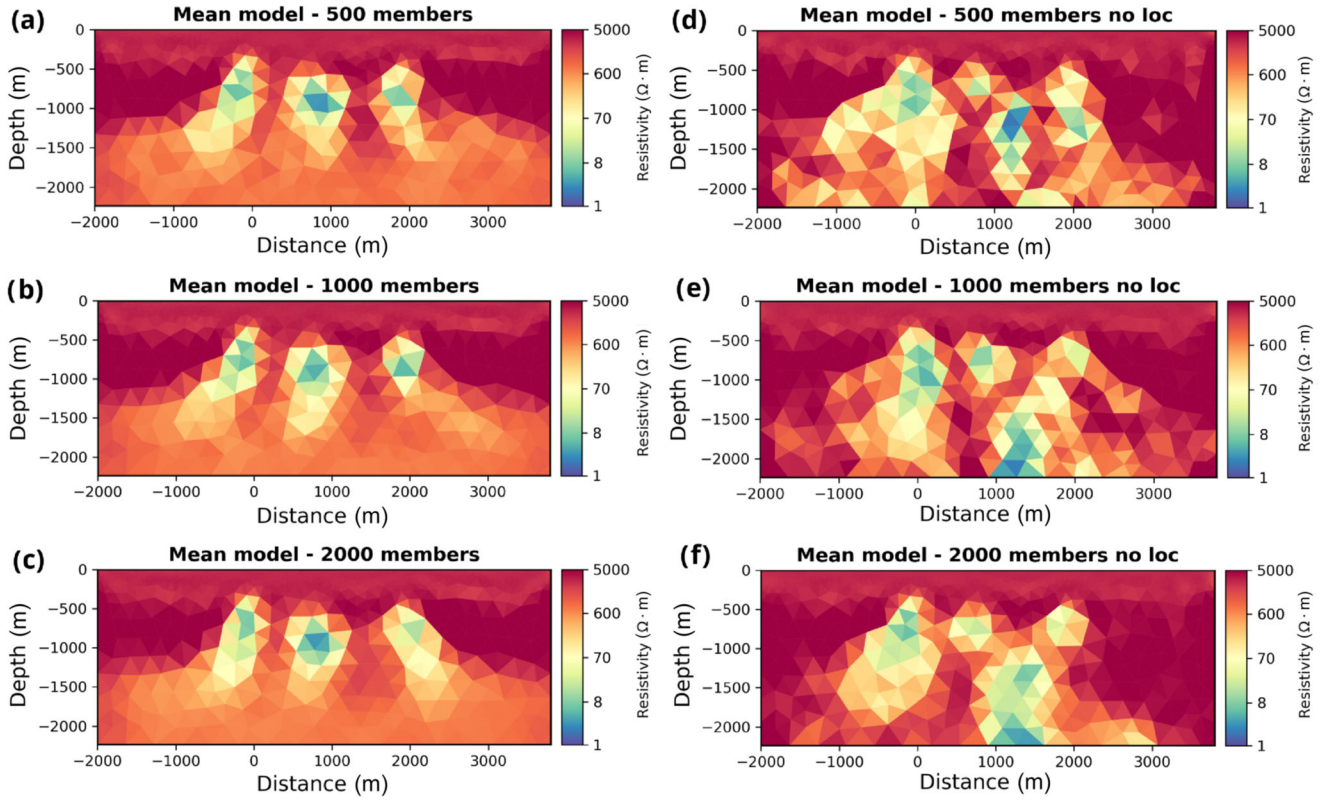

Figure 10 | (a), (b) and (c) Mean models obtained through localization of the Kalman filter and ensemble sizes of 500, 1000 and 2000, respectively. (d), (e) and (f) Mean models obtained without localization and using ensemble sizes of 500, 1000 and 2000, respectively.

Moreover, not all low-resistivity values correspond to the true locations of the graphitic bodies. Notably, such values appear at a depth of approximately 1000 m and a horizontal distance of around 1000 m, as shown in Figure 10d, as well as near the bottom of the model (see Figure 10e). These artefacts are probably caused by spurious correlations that affect the Kalman filter, thus contributing to the algorithm convergence to a local solution. However, the ES-MDA without localization converges in few iterations and with a slightly lower mean RMSE (Table 1). In accordance with the considerations made about the spurious correlations (Figure 7), the mean models tend to be smoother, increasing the ensemble size (Figure 10d–f). The main drawback of using a small ensemble size is the uncertainty underestimation (Evensen et al. 2022; Aleardi et al. 2021) that can lead to errors in the interpretation phase. The standard deviations calculated for each test (Figure 11a–c) exhibit similar features with low uncertainty in the first layer and corresponding to the graphitic bodies, while higher uncertainty laterally and in depth. It is important to note that in the portion of the model most relevant for uranium exploration, the zone between the graphitic zone and the overlying sandstone, the three tests return a comparable marginal distribution (Figure 11d–i). The uncertainty beneath the graphitic bodies does not increase significantly towards the bottom of the model, but instead appears as smooth as the mean model. This is caused by the localization that smooths and filters out the updates beyond the range of the tapering function. As suggested by Figure 12, the marginal posterior *pdf*s of the selected cells (orange, blue and green) are comparable to the marginal prior distributions in both standard deviation and mean value, indicating that these cells are not considerably updated throughout the iterations.

4 | Field Data Inversion

4.1 | Data Acquisition

The study area is located in the eastern part of the Athabasca Basin (Saskatchewan, Canada), location of some of the world's largest high-grade uranium resources (Jefferson et al. 2007). The basin is composed of Proterozoic flat-lying, quartz-rich sandstone and conglomerates, which are interpreted as deposits from river systems and nearshore to shallow-shelf environments (Hiatt et al. 2005), reaching the maximum thickness of about 2 km at the central part and thinning towards the edges (Jefferson et al. 2007; Alexandre 2020). The sequence overlies a crystalline basement mostly composed of Archean Gneisses and Paleoproterozoic metasedimentary rocks through an unconformity (Jefferson et al. 2007). The unconformity-type uranium deposits characterize the Athabasca Basin, and they are thought to result from the reaction between oxidizing brines of the basins and reducing fluids from the basement (Kotzer and Kyser 1995). Uranium precipitation forms pods, veins and semimassive replacements in the proximity of the unconformity zone and is usually associated with reactivated faults (Jefferson et al. 2007). Since the current exploration paradigm is that the mineralization occurs preferentially close to the intersection between reactivated graphitic fault zones and the basin unconformity (Lu et al. 2021), the geophysical methods

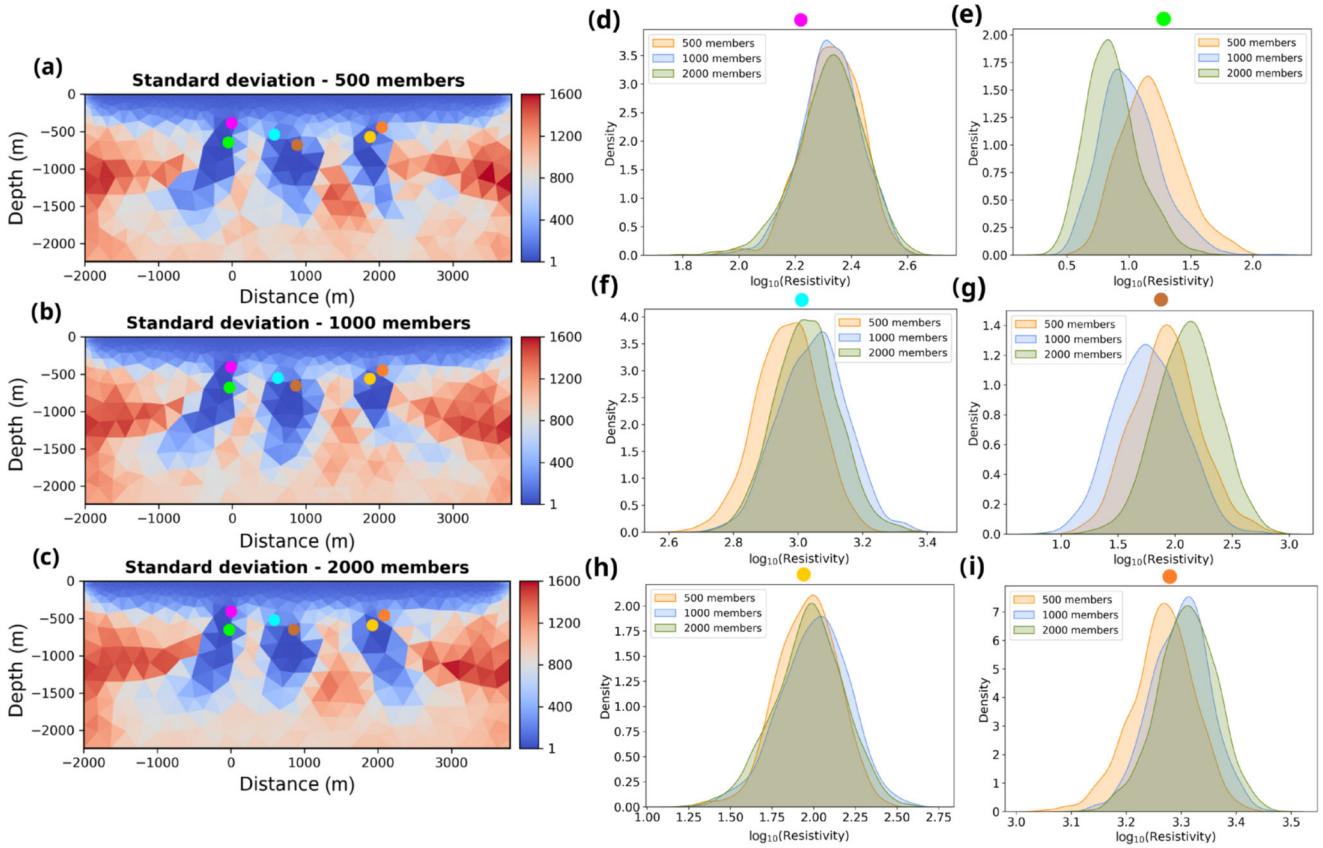

Figure 11 | (a), (b) and (c) Standard deviation obtained through an ensemble of 500, 1000 and 2000 members, respectively. (d)–(i) Marginal posterior *pdf* of the model cells is indicated by the coloured dots.

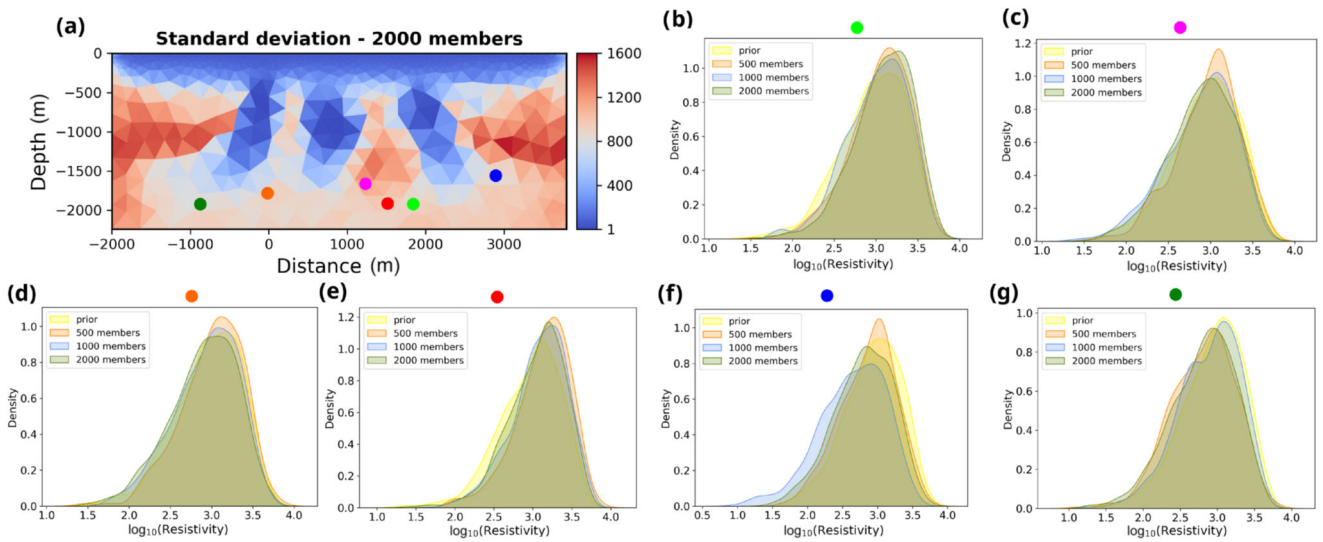

Figure 12 | (a) Standard deviation. The coloured dots represent the cells from which the marginal prior and posterior *pdf*s are extracted. (b)–(g) Marginal posterior and prior *pdf*s.

Electromagnetics and ERT focus on the graphitic bodies as the initial target.

The field data we invert in this section were acquired in the eastern part of the Athabasca Basin by Quantec Geoscience Ltd on behalf of Orano Canada Inc (formerly Areva Resources Canada Inc.) as part of a large data acquisition campaign involving EM and ERT surveys. All the 2D profiles are acquired perpendicular to the expected orientation of the graphitic bodies, which were intersected by boreholes. Figure 13 exhibits the survey area, where the black lines represent the time-domain EM profiles and the dashed red lines represent the ERT profiles. The considered

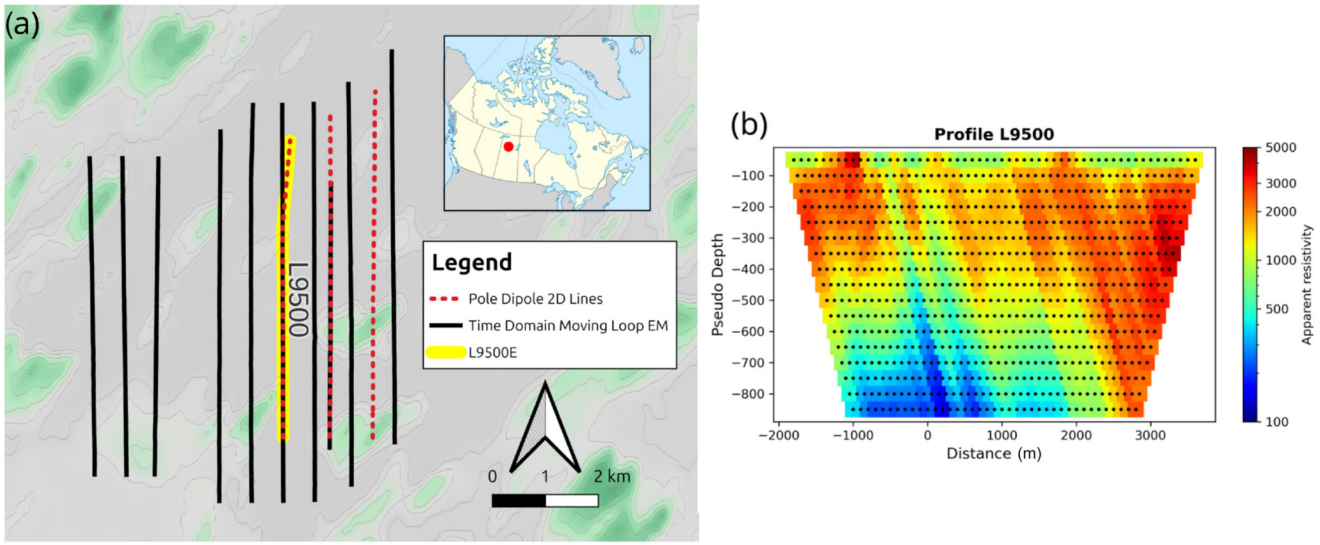

Figure 13 | (a) Location of the survey area. The black lines illustrate the EM profiles and the dashed red lines the ERT profiles. Note that the selected profile is highlighted in yellow. (b) Observed data associated with the selected profile.

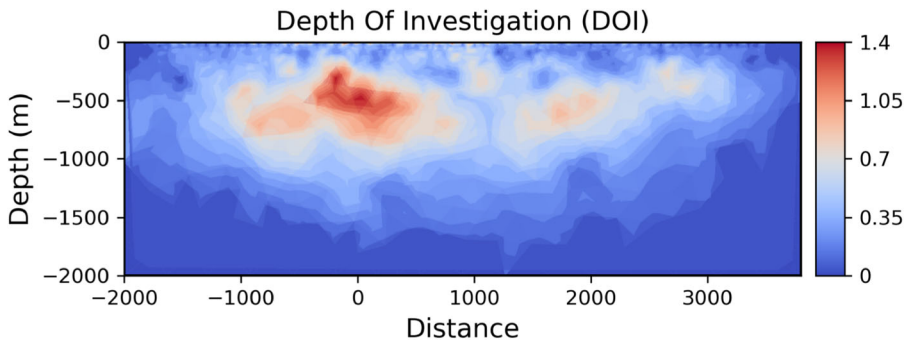

Figure 14 | Depth of investigation computed using cumulative sensitivity on a homogeneous model of resistivity 4000 Ω · m.

ERT profile (highlighted in Figure 13a) consists of a pole–dipole configuration (Figure 3b) with a 100-m voltage dipole, which is shifted by 100 m along the profile, increasing its distance from the current pole and reaching a maximum distance of 1650 m for a total profile length of 5600 m. Figure 13b illustrates the acquired data along the profile L9500, which is characterized by the contrast between the predominant high apparent resistivity and the low values at the bottom of the pseudosection. Even though the graphical representation of apparent resistivity values does not represent exactly the spatial position of the anomaly in the half-space, it seems to suggest the presence of a conductive target at depth. To assess the depth of investigation (DOI) associated with the array configuration, we follow the strategy from Christiansen and Auken (2012), which is based on the cumulated sensitivity (Figure 14). Since in the real case we do not know the exact subsurface resistivity, we compute the DOI on a homogeneous model with resistivity equal to 4000 Ω · m, which is the mean between the resistivity of the sandstone and the basement. The DOI shows maximum values around 500 m and seems to suggest a sensitivity until a depth of 1500 m. However, the interpretation of the DOI should be done carefully due to the empirical choice of the threshold value (Christiansen and Auken 2012).

4.2 | Prior Ensembles and Localization Functions

In order to test the proposed scheme and discuss the effect of ensemble size on Kalman filter contamination, we generate three ensembles of 500, 1000 and 2000 models following the strategy illustrated in Figure 5. As for the synthetic test, according to the expected resistivity to recover, we built a multivariate Gaussian distribution having a mean value equal to 1 (equal to 1000 Ω · m in resistivity domain) and a standard deviation of about 0.3 (in the transformed domain, Equations 24 and 25). The spatial correlation employed for generating the ensemble of models is defined by an exponential variogram, as in Equation (23), which we assume to be isotropic, with a range of 300 m in both the x and the z directions. This spatial correlation not only allows the generation of smooth models but also acts as a form of spatial regularization, reducing the degrees of freedom in the model space. The choice of spatial correlation is typically made in relation to the cell size and the expected resolution of the survey. Figure 15a shows a model drawn from a prior ensemble in the transformed domain, while Figure 15b presents the same model in the logarithmic domain. Since the ensembles are generated by defining the mean and standard deviation, each cell shares the same prior marginal probability density function

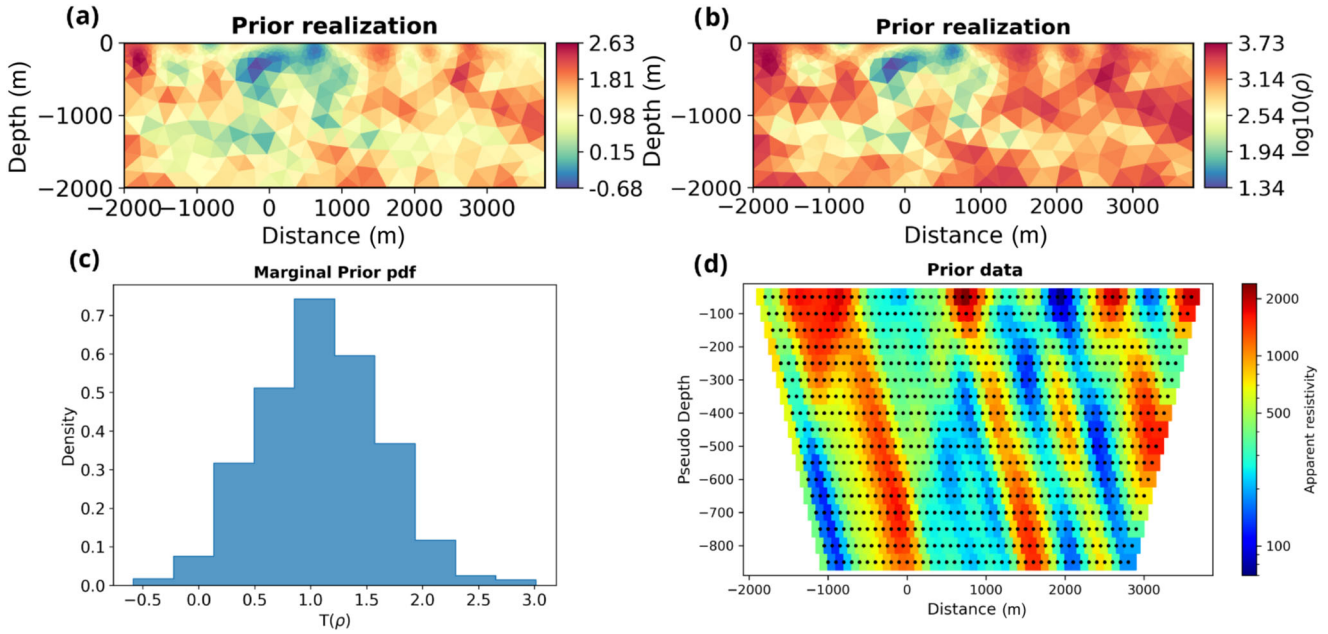

Figure 15 | (a) Model drawn from the prior ensemble (transformed domain). (b) Model in log10 resistivity domain. (c) Prior probability density function associated with a cell of the model. (d) Data predicted employing the model in (b).

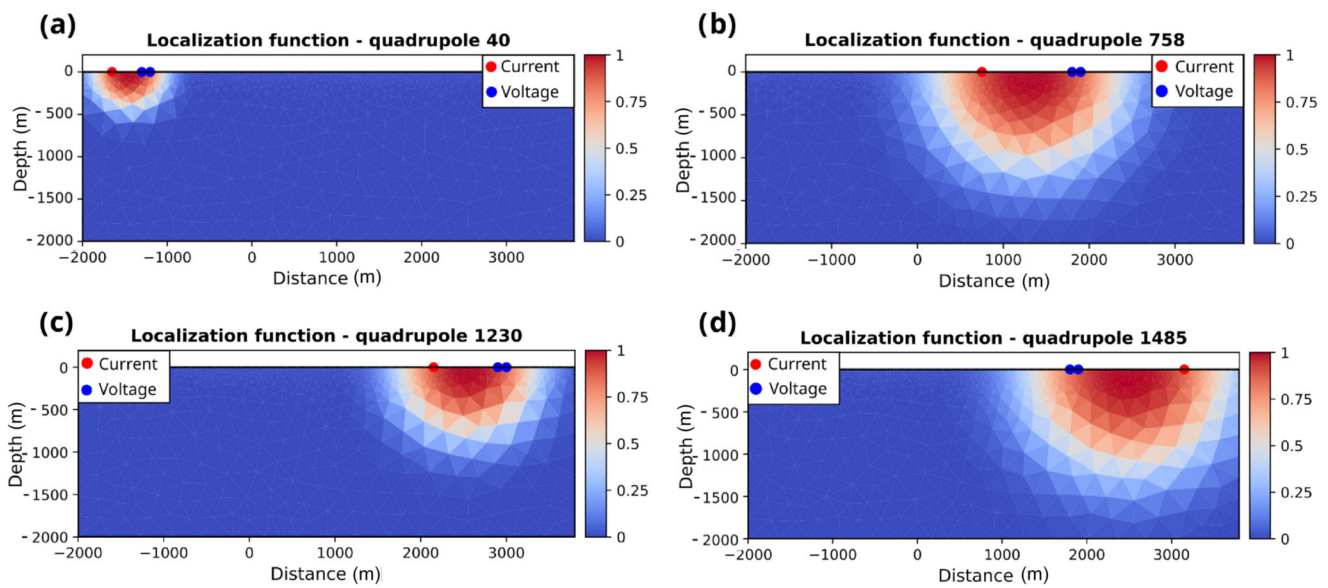

Figure 16 | Localization functions associated with four dipoles used for the survey.

pdf as illustrated in Figure 15c. To obtain the prior ensemble of data, the forward operator is applied to compute the predicted data corresponding to each model (Figure 15d). The localization function used in the field data inversion is built following the strategy developed for the synthetic test, specifically a spatial correlation-like function defined by the following relation:

$$\boldsymbol{L}(x,z) = \exp\left(-\left(\sqrt{\left(\frac{\tilde{\mathbf{d}}_x}{R_x}\right)^2 + \left(\frac{\tilde{\mathbf{d}}_z}{R_z}\right)^2}\right)^3\right), \tag{26}$$

where $\tilde{\mathbf{d}}_x$ and $\tilde{\mathbf{d}}_z$ are the distances between the considered data point and the cells of the model, R_x and R_z are the ranges along x and z that we choose equal to the distance between the injection pole and the farthest voltage electrode. Figure 16 shows four localization functions associated with four quadrupoles characterized by different ranges. As shown in Figure 16a,b, the cut-off region becomes larger as the quadrupole size decreases.

4.3 | Results

We performed the inversion using three different ensemble sizes and repeated the tests without applying localization. The localized adaptive ES-MDA converges in nine iterations, returning a mean RMSE of about 5% for all the tests with varying the ensemble size (Table 2). The objective function decreases rapidly

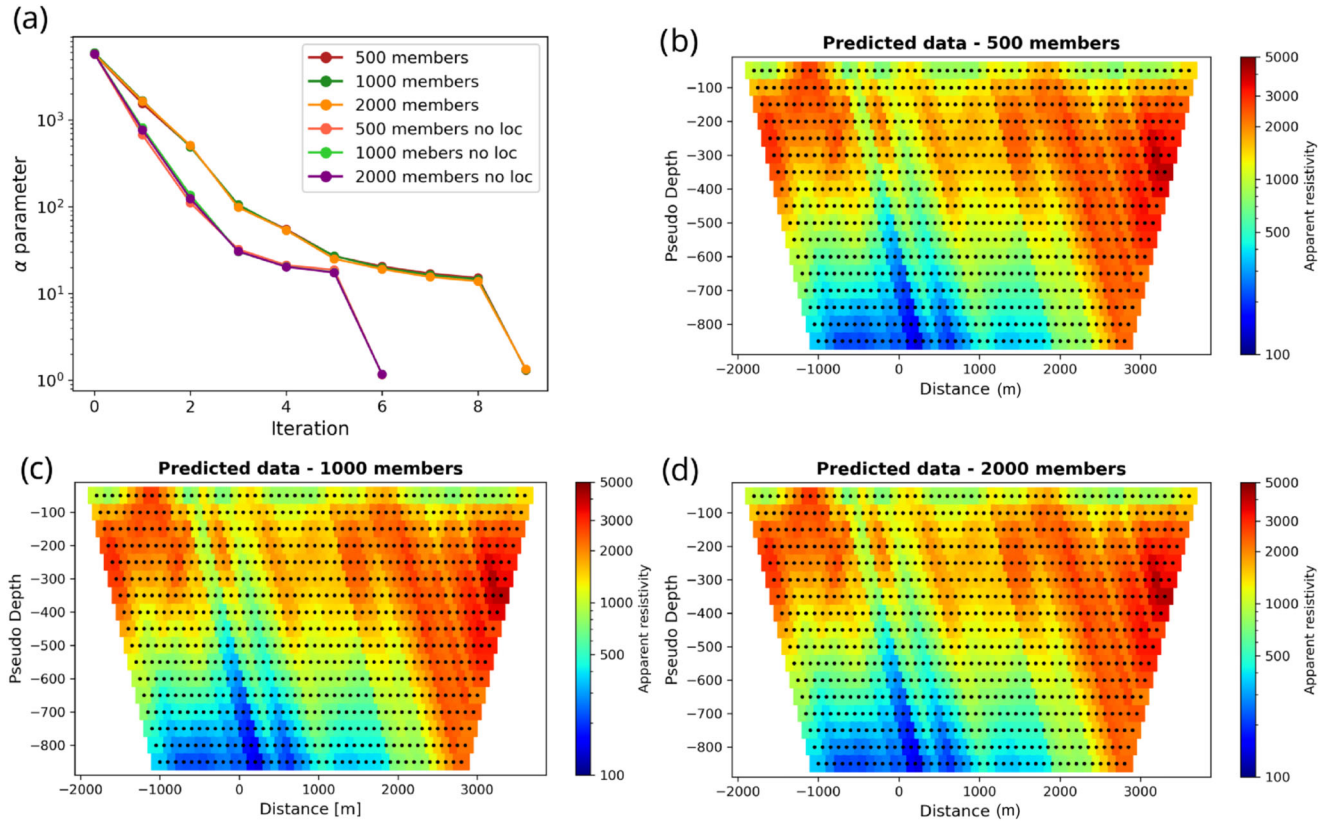

Figure 17 | (a) α parameter of Equation (15), which represents the averaged objective function value. Note the almost identical trend of the values among the tests performed by varying the ensemble sizes. (b), (c) and (d) Data predicted using an ensemble size of 500, 1000 and 2000, respectively.

Table 2 | Parameters concerning the field data inversion. μ indicates the mean, σ the standard deviation of the prior distribution and R the range of the spatial correlation.

Localization	Prior ensemble				Inversion		
	Ensemble size	μ^a	σ^{2a}	R (m)	Iterations	Mean rmse (%)	Computational time (h)
Yes	500	1.09	0.3	300	10	5.0	1.49
Yes	1000	1.09	0.3	300	10	4.9	2.97
Yes	2000	1.09	0.3	300	10	4.9	5.95
No	500	1.09	0.3	300	7	6.43	1.04
No	1000	1.09	0.3	300	7	6.15	2.08
No	2000	1.09	0.3	300	7	6.26	4.16

[a] Transformed domain.

until iteration 5, and it continues to steadily decrease until the last iteration (Figure 17a). At the ninth iteration, the α parameter decreases significantly to satisfy Equation (15); therefore, this change is not directly related to a drop in the objective function value. Similar to the synthetic test, the algorithm appears to converge earlier without localization (after seven iterations compared to ten), although with a slightly higher value of the normalized objective function (Figure 17a).

A brief comparison among the pseudosections in Figure 17 computed from the mean models and the observed data (Figure 14b) suggests that all the tests provide a satisfactory prediction of the data, with an of 5%. The mean models are characterized by high-resistivity values in the upper 450 m of depth, with clear low-resistivity anomalies appearing below this depth, particularly between 500 and 2000 m along the profile (Figure 18a–c). Although these anomalies are consistently detected across all tests, the spatial resolution of the mean model obtained with an ensemble size of 500 is lower than that produced with ensemble sizes of 1000 and 2000. This reduced resolution is likely due to the influence of spurious correlations associated with the smaller ensemble. The results obtained without applying localization in the Kalman filter are presented in Figures 18d–f. As expected, for the ensemble sizes 500 and 1000, the models

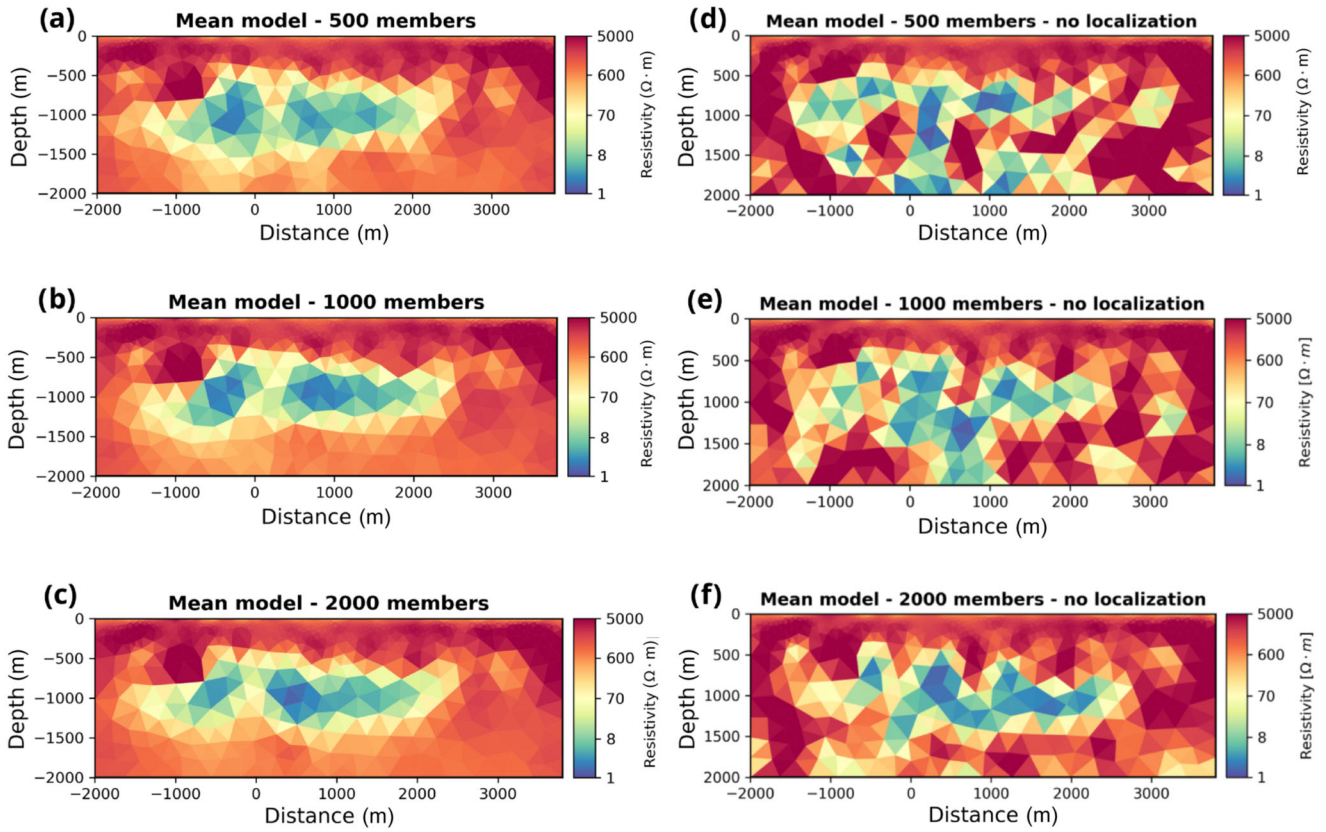

Figure 18 | (a), (b), (c) Mean model obtained through localization for an ensemble size of 500, 1000 and 2000, respectively. (d), (e), (f) Mean models obtained without localization for an ensemble size of 500, 1000 and 2000, respectively.

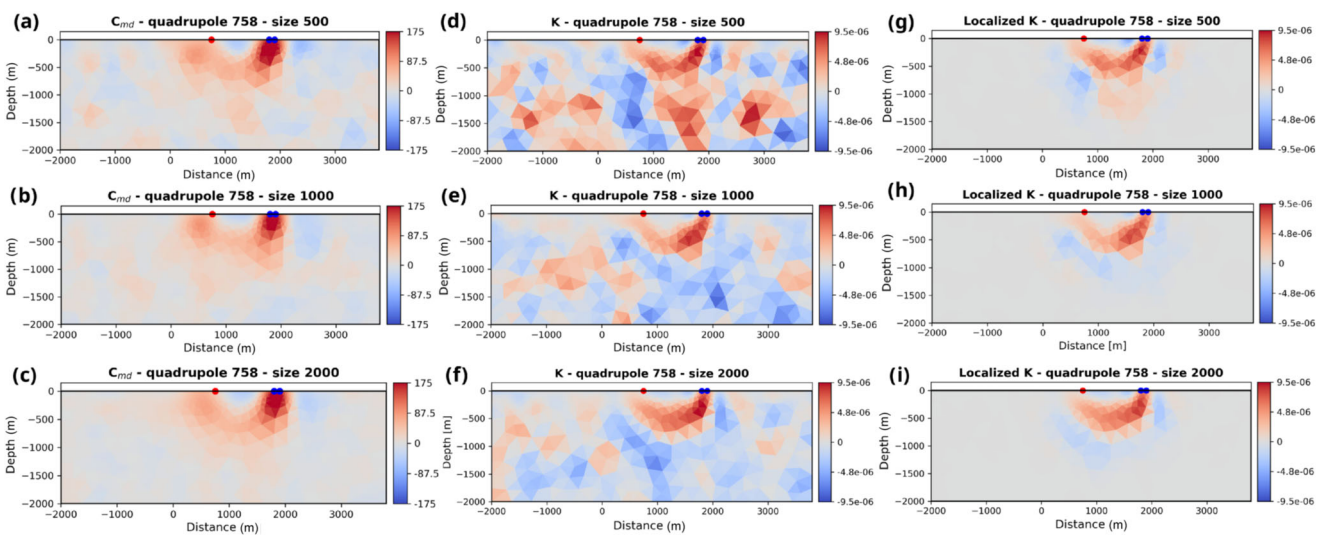

Figure 19 | (a), (b) and (c) Representation of a column of the cross-covariance matrix obtained from an ensemble of 500, 1000 and 2000 members, respectively. (d), (e) and (f) Column of the Kalman filter for an ensemble of size 500, 1000 and 2000, respectively. (g), (h) and (i) Localized Kalman filter for different ensemble sizes.

are heavily scattered below 500 m with some low-resistivity anomalies extended until 2000 m, with a poor lateral continuity (Figure 18d,e). Although the use of 2000 models helps attenuate the effect of spurious long-range correlations (Figure 18f), the mean model still appears contaminated by artefacts, which may induce interpretation errors. In particular, the model exhibits an unrealistically high resolution, showing a sub-vertical conductive body approximately 300 m wide at depths between 1000 and 1750 m (indicated by the black arrow in Figure 18f), which seems to be unreliable. To illustrate the impact of localization

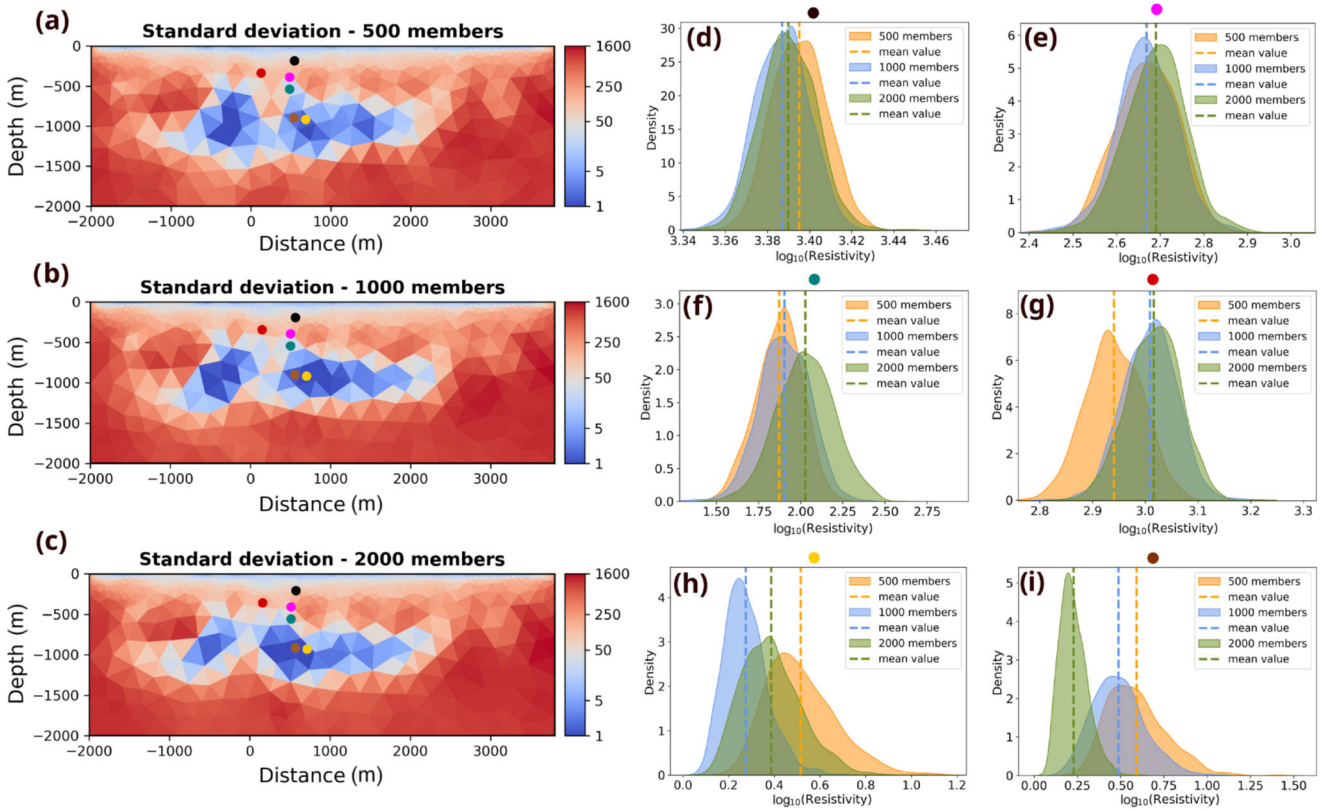

Figure 20 | (a), (b), (c) Standard deviation computed from the final ensemble of size 500, 1000 and 2000, respectively. (d)–(i) Marginal posterior distribution associated with the cells of the model indicated by the coloured dots.

functions on the Kalman filter, we extract and analyse the column corresponding to the quadrupole number 758 (Figure 19). As we have already observed in the synthetic test, increasing the ensemble size leads to a less scattered cross-covariance; however, artefacts still appear in regions far from the selected quadrupole (Figure 19a–c). Spurious correlations significantly affect the Kalman filter, particularly when the ensemble size is 500, leading to widespread scattered values (Figure 19d). While increasing the ensemble size to 1000 and 2000 reduces the influence of these artefacts, they are not completely eliminated (Figure 19e,f). The localization applied to the Kalman filter allows for smoothing out and to removing the spurious correlation far away from the quadrupole, as shown in Figure 19g–i. Nevertheless, in the case of size 500, the application of the localization does not guarantee the complete removal of spurious values that are still present down to the depth of 1000 m (Figure 19g), producing errors in the posterior pdf.

The standard deviation obtained from the three tests shows a strong correlation with the mean resistivity values; specifically, lower standard deviations are associated with low-resistivity values, and higher standard deviations are associated with high-resistivity values (Figure 18a–c). The standard deviation obtained from the three tests exhibits higher values where the model is strongly resistive, and lower values where it is more conductive (Figure 18a–c). Therefore, to further investigate the uncertainty, it is useful to examine the marginal posterior distribution of some selected model cells (Figures 20d–i and 21). All six posterior marginal pdf correspond to model cells that are potentially relevant from an interpretational perspective, as they are located either just below or just above the low-resistivity anomalies. In particular, Figure 20f shows the distributions related to the cell indicated by the green dot, located just below 500 m depth and approximately 500 m along the profile. This cell exhibits mean resistivity values between approximately 80 and 110 $\Omega \cdot$ m across the three tests. As it emerges from a brief comparison of the distributions, the larger mean value for the green distribution (ensemble of 2000 members) is due to its longer tail. Although the posterior pdf for the three tests do not completely overlap, their mean values are consistent and the associated uncertainties are of the same order of magnitude. Even for the distribution in Figure 20h,i, which shows the marginal posterior pdf at the depth of 1000 m, the differences in uncertainty remain small relative to the overall range of resistivity values. Figure 21 exhibits six posterior *pdf*s representing six cells below the low-resistivity anomalies. For all the marginal *pdf*s, both the mean values and their uncertainty are coherent.

From an interpretational perspective, the mean resistivity models and their associated uncertainties Figures 18 and 20 allow for meaningful geophysical considerations, in particular:

- The array configuration, designed to focus on deep targets, impedes the reconstruction of the shallow low-resistivity cover that characterizes the region (Mir et al. 2022).
- The mean models consistently reveal an uppermost resistivity layer that we interpret as the Athabasca sandstone formation characterized by values between 2000 and 4000 $\Omega \cdot$ m, extending to a depth of about 500 m

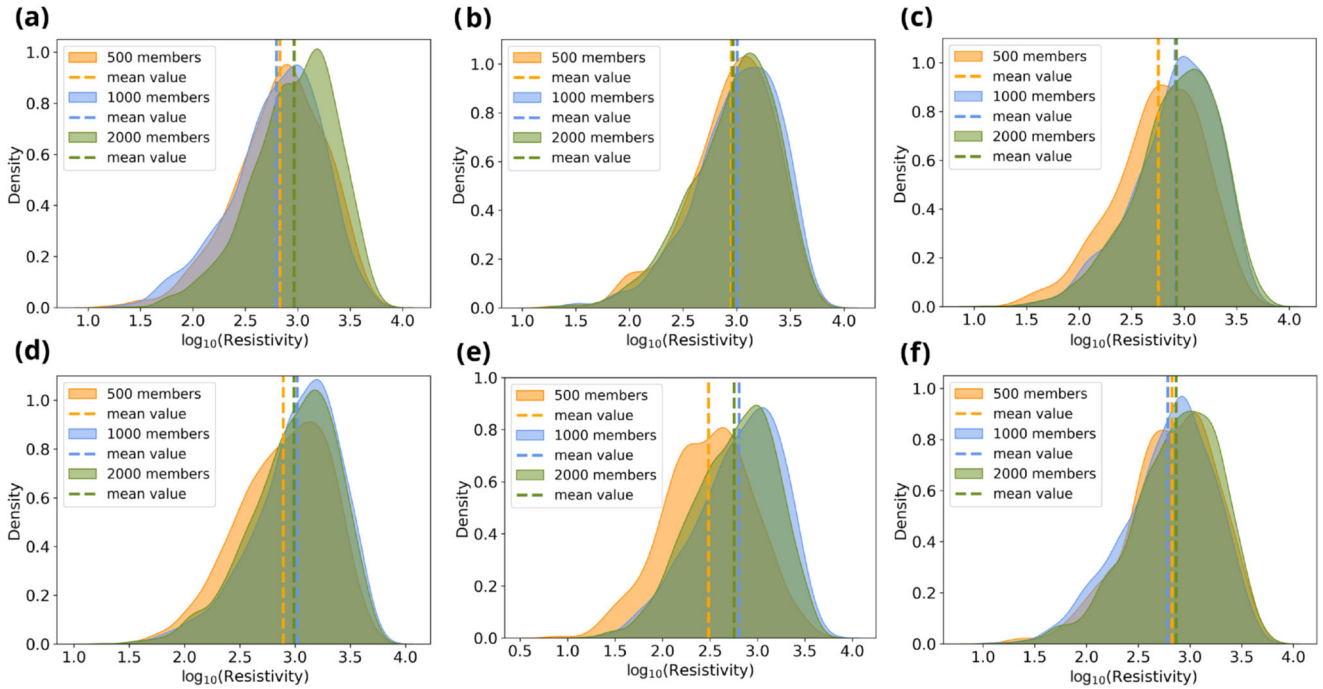

Figure 21 | Marginal posterior pdf associated with six cells of the model placed below the low-resistivity anomalies.

- At a depth of 500 m, a sub-horizontal resistivity contrast is observed, which we interpret as the unconformity delineating the top of the basement. The transition zone around this unconformity is usually the target for uranium exploration, as it may host the alteration halo associated with uranium mineralization.
- The analysis of that zone evaluating the marginal *pdf*s may enhance the interpretation, especially when integrated with boreholes, geochemical and geological information (Figure 20).
- Below and around the unconformity, the models exhibit resistivity values from a few to about hundreds of $\Omega \cdot m$. In particular, at least two distinct targets are evident, which we interpret as graphitic bodies. However, as expected, the resolution of the ERT inversion is insufficient to fully resolve the structural parameters. Therefore, integration with electromagnetic data may improve the graphitic body reconstruction.

5 | Conclusions

In this study, we apply the adaptive ES-MDA algorithm to the uranium exploration context in the Athabasca Basin, characterized by a wide range of resistivity values and sharp resistivity contrasts. We investigated the impact of the ensemble size on the sensitivity information contained in covariance matrices and then on the Kalman filter. We tested a simple localization strategy based on the quadrupole dimension to smooth out and eliminate spurious correlations as a function of the distance between data points and cells of the model.

The test on synthetic data suggests that the localization function is able to attenuate the spurious correlations iteration by iteration, preventing the model update where we expect low or no sensitivity. However, even although without localization the inversion is able to predict the synthetic observed data, the mean model is strongly affected by spurious correlations, resulting in scattered values and a poorly reconstructed target. On the contrary, the localization enables satisfactory target detection even when a smaller ensemble size is chosen. Moreover, the uncertainty estimation does not appear to be significantly affected by the ensemble size.

The inversion of the field ERT dataset employing the localized ES-MDA is able to clearly detect at least two conductive bodies. Moreover, the standard deviation, which is used, may be a relevant tool to evaluate the uncertainty in the resistivity values, but it is not significantly affected by the tested ensemble size. On the other hand, the ES-MDA without localization shows models affected by scattered values that can bring interpretation errors, where, especially as in the case of the Athabasca Basin, we expect a high contrast of resistivity. In conclusion, the inversion of the field dataset suggests that localization allows for a reduction in ensemble size without significantly affecting the estimation of the posterior probability density.

Further developments may involve the extension of the implemented strategy to different array configurations, the inclusion of sensitivity information to build localization functions and application on a three-dimensional dataset.

Acknowledgements

We thank Elodie Willard from Orano Mining and Grant Harrison from Orano Canada for their valuable suggestions. We also thank Cameco for providing and allowing us to present the field data we have used in this work.

This research was supported by the GeomIn3D Industrial Chair jointly funded by Orano and Agence Nationale de la Recherche (project ANR-21-CHIN-0006).

Open access publication funding provided by COUPERIN CY26.

Data Availability Statement

Research data are not shared.

References

Aleardi, M., A. Vinciguerra, and A. Hojat. 2021. "A Convolutional Neural Network Approach to Electrical Resistivity Tomography." *Journal of Applied Geophysics* 193: 104434.

Alexandre, P. 2020. "Geochemistry of the Athabasca Basin, Saskatchewan, Canada, and the Unconformity-Related Uranium Deposits Hosted by It." *The Canadian Mineralogist* 59, no. 5: 847–868.

Auken, E., A. V. Christiansen, B. H. Jacobsen, N. Foged, and K. I. Sørensen. 2005. "Piecewise 1D Laterally Constrained Inversion of Resistivity Data." *Geophysical Prospecting* 53: 497–506.

Chen, Y., and D. S. Oliver. 2010. "Cross-Covariances and Localization for EnKF in Multiphase Flow Data Assimilation." *Computational Geosciences* 14: 579–601.

Chen, Y., and D. S. Oliver. 2017. "Localization and Regularization for Iterative Ensemble Smoothers." *Computational Geosciences* 21: 13–30.

Christiansen, A. V., and E. Auken. 2012. "A Global Measure for Depth of Investigation." *Geophysics* 77: WB171–WB177.

De Groot-Hedlin, C., and S. Constable. 1990. "OCCAM's Inversion to Generate Smooth, Two-Dimensional Models From Magnetotelluric Data." *Geophysics* 55: 1613–1624.

Dentith, M., and S. T. Mudge. 2014. *Geophysics for the Mineral Exploration Geoscientist*. Cambridge University Press.

Emerick, A., and A. Reynolds. 2011. "Combining Sensitivities and Prior Information for Covariance Localization in the Ensemble Kalman Filter for Petroleum Reservoir Applications." *Computational Geosciences* 15: 251–269.

Emerick, A. A. 2016. "Analysis of the Performance of Ensemble-Based Assimilation of Production and Seismic Data." *Journal of Petroleum Science and Engineering* 139: 219–239.

Emerick, A. A., and A. C. Reynolds. 2013. "Ensemble Smoother With Multiple Data Assimilation." *Computers & Geosciences* 55: 3–15.

Evensen, G., F. C. Vossepoel, and P. J. van Leeuwen. 2022. *Data Assimilation Fundamentals. A Unified Formulation of the State and Parameter Estimation Problem*. Springer Cham.

Farquharson, C. G., and D. W. Oldenburg. 2004. "A Comparison of Automatic Techniques for Estimating the Regularization Parameter in Non-Linear Inverse Problems." *Geophysical Journal International* 156: 411–425.

Galetti, E., and A. Curtis. 2018. "Transdimensional Electrical Resistivity Tomography." *Journal of Geophysical Research: Solid Earth* 123: 6347–6377.

Günther, T., and C. Rücker. 2012. Electrical Resistivity Tomography (ERT) in Geophysical Applications - State of the Art and Future Challenges. https://www.researchgate.net/profile/Thomas-Guenther-9/publication/235978636_Electrical_Resistivity_Tomography_ERT_in_geophysical_applications_-_state_of_the_art_and_future_challenges/links/004635151e4096dead000000/Electrical-Resistivity-Tomography-ERT-in-geophysical-applications-state-of-the-art-and-future-challenges.pdf.

Hamill, T., J. Whitaker, and C. Snyder. 2001. "Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter." *Monthly Weather Review* 129: 2776–2790.

Hiatt, E., R. Christophe, K. Durocher, G. Holk, and K. Deckart. 2005. DIAGENETIC Diagenetic fluids in paleo - and meso- proterozoic sedimentary basins and their implications for long protracted fluid histories. 1–38. https://doi.org/10.13140/2.1.1033.1847.

Houtekamer, P. L., and H. L. Mitchell. 1998. "Data Assimilation Using an Ensemble Kalman Filter Technique." *Monthly Weather Review* 126: 796–811.

Jefferson, C. W., D. Thomas, S. Gandhi, et al. 2007. "Unconformity-Associated Uranium Deposits of the Athabasca Basin, Saskatchewan and Alberta." In *EXTECH IV: Geology and Uranium EXploration TECHnology of the Proterozoic Athabasca Basin, Saskatchewan and Alberta*, edited by C. W. Jefferson and G. & Delaney. Bulletin Geological Survey of Canada 588: 23–67. Natural Resources Canada.

Kotzer, T., and T. Kyser. 1995. "Petrogenesis of the Proterozoic Athabasca Basin, Northern Saskatchewan, Canada, and Its Relation to Diagenesis, Hydrothermal Uranium Mineralization and Paleohydrogeology." *Chemical Geology* 120: 45–89.

Le, D. H., A. A. Emerick, and A. C. Reynolds. 2016. "An Adaptive Ensemble Smoother With Multiple Data Assimilation for Assisted History Matching." *SPE Journal* 21: 2195–2207.

Lines, L., and S. Treitel. 1984. "A Review of Least-Squares Inversion and Its Application to Geophysical Problems." *Geophysical Prospecting* 32: 159–186.

Loke, M., and R. Barker. 1996. "Rapid Least-Squares Inversion of Apparent Resistivity Pseudosections by a Quasi-Newton Method." *Geophysical Prospecting* 44: 131–152.

Loke, M. H., P. B. Wilkinson, J. Gance, J.-P. Malet, C. Truffert, and O. Leite. 2022. "Measurement and Inversion Strategies for 3-D Resistivity Surveys with Vector Arrays." *Geophysical Prospecting* 70: 578–592.

Lu, X., C. G. Farquharson, J.-M. Miehé, and G. Harrison. 2021. "3D Electromagnetic Modeling of Graphitic Faults in the Athabasca Basin Using a Finite-Volume Time-Domain Approach With Unstructured Grids." *Geophysics* 86: B349–B367.

Mir, R., P. Fullagar, M. Darijani, et al. 2022. "Forward Modeling and 3D Inversion of Electromagnetic Data Collected Over the McArthur River Uranium Deposit in the Athabasca Basin, Canada." *Geophysics* 87: B129–B143.

Morzfeld, M., and D. Hodyss. 2023. "A Theory for Why Even Simple Covariance Localization Is So Useful in Ensemble Data Assimilation." *Monthly Weather Review* 151: 717–736.

Peng, R., P. Yogeshwar, Y. Liu, and X. Hu. 2023. "Quasi-2-D Bayesian Inversion of Central Loop Transient Electromagnetic Data Using an Adaptive Voronoi Parametrization." *Geophysical Journal International* 234: 650–663.

Rafiee, J., and A. Reynolds. 2017. "Theoretical and Efficient Practical Procedures for the Generation of Inflation Factors for ES-MDA." *Inverse Problems* 33: 115003.

Raniolo, S., L. Dovera, A. Cominelli, C. Callegaro, and F. Masserano. 2013. History Match and Polymer Injection Optimization in a Mature Field Using the Ensemble Kalman Filter. In IOR 2013 - 17th European Symposium on Improved Oil Recovery. European Association of Geoscientsts & Engineers. https://doi.org/10.3997/2214-4609.20142642.

Reynolds, A., M. Zafari, and G. Li. 2006. "Iterative Forms of the Ensemble Kalman Filter." In *ECMOR X - 10th European Conference on the Mathematics of Oil Recovery*, Amsterdam, the Netherlands. European Association of Geoscientists & Engineers.

Rücker, C., T. Günther, and F. M. Wagner. 2017. "pyGIMLi: An Open-Source Library for Modelling and Inversion in Geophysics." *Computers and Geosciences* 109: 106–123.

Sambridge, M., and K. Mosegaard. 2002. "Monte Carlo Methods in Geophysical Inverse Problems."*Reviews of Geophysics* 40: 3–1–3–29.

Sasaki, Y. 1992. "Resolution of Resistivity Tomography Inferred From Numerical Simulation." *Geophysical Prospecting* 40: 453–463.

Singh, A., P. Yogeshwar, M. Israil, and B. Tezkan. 2024. "2-D Transdimensional Joint Inversion of Radio Magnetotelluric and Electrical Resistivity Tomography Data." *Geophysical Journal International* 239: 1863–1878.

Tarantola, A. 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics.

Vinciguerra, A., M. Aleardi, A. Hojat, M. H. Loke, and E. Stucchi. 2022. "Discrete Cosine Transform for Parameter Space Reduction in Bayesian Electrical Resistivity Tomography." *Geophysical Prospecting* 70: 193–209.

Vinciguerra, A., M. Aleardi, L. M. Madsen, T. S. Bording, A. V. Christiansen, and E. Stucchi. 2024. "Stochastic Inversion of Time-Lapse Electrical Resistivity Tomography Data by Means of an Adaptive Ensemble Based Approach." *Geophysical Prospecting* 72: 268–284.

Zhong, S., Y. Wang, Y. Zheng, S. Wu, X. Chang, and W. Zhu. 2021. "Electrical Resistivity Tomography With Smooth Sparse Regularization." *Geophysical Prospecting* 69: 1773–1789.