Hydrogeophysical Parameter Estimation Using Iterative Ensemble Smoothing and Approximate Forward Solvers

Corinna Köpke [1*†], Ahmed H. Elsheikh [2] and James Irving [1]

[1] Institute of Earth Sciences, University of Lausanne, Lausanne, Switzerland, [2] School of Energy, Geoscience, Infrastructure and Society, Heriot-Watt University, Edinburgh, United Kingdom

In iterative ensemble smoother approaches and ensemble methods in general, the ensemble size governs the accuracy of the parameter estimates obtained. However, employing large ensembles may be computationally infeasible in applications with expensive forward solvers. Here, we reduce the computational cost of using large ensembles in iterative ensemble smoothing through the use of a proxy solver. To correct the proxy response for the corresponding model error, the latter of which can bias posterior parameter estimates if left untreated, we propose a local basis approach. With this approach, the discrepancy between the detailed and proxy solvers is learned for a subset of the ensemble and collected in a dictionary that grows with each iteration. For each ensemble member, the K-nearest neighbors in the dictionary are employed to build an orthonormal basis which is used to identify the model-error component of the residual by projection. The proposed methodology reduces the effects of overfitting the data with the proxy solver, but may lead to underfitting of the data in the absence of a sufficient number of dictionary entries, meaning that the number of ensemble members relative to the number of detailed-solver runs cannot be inflated arbitrarily. We present our approach in the context of the ensemble smoother with multiple data assimilations (ES-MDA) algorithm, and show its successful application to a high-dimensional synthetic example that involves crosshole ground-penetrating radar (GPR) travel-time tomography.

Keywords: ensemble methods, ES-MDA, proxy model, model error, inversion, uncertainty quantification

1. INTRODUCTION

Inverse problems commonly involve computationally expensive forward solvers and large numbers of unknown parameters that are spatially distributed. For risk assessment and effective environmental decision making, parameter uncertainties are required. These can be obtained through, for example, Bayesian stochastic inversion whereby the corresponding posterior distributions are typically sampled using Markov-chain-Monte-Carlo (MCMC) methods. The Bayesian-MCMC framework offers the advantages of providing a natural quantification of parameter uncertainties, as well as the flexibility to incorporate probabilistic information about priors and measurement errors into the inverse problem (Robert and Casella, 2004). However, depending on the forward solver and the dimensionality of the model-parameter space involved, it can be extremely computationally expensive. In many real-world applications, for example,

millions of forward model executions may be required to obtain meaningful posterior statistics with Bayesian-MCMC methods (Ruggeri et al., 2015). Although several recent modifications to the standard Metropolis-Hastings algorithm have significantly improved the computational efficiency of MCMC (e.g., Haario et al., 2001; Hansen et al., 2012; Cotter et al., 2013; Chen et al., 2016; Vrugt, 2016; Beskos et al., 2017), these modifications are often still not enough to make such methods practically feasible for many inverse problems.

One way to significantly reduce the computational cost of stochastic parameter estimation is to employ ensemble-based methods. With such methods, an initial ensemble of model parameter sets, drawn from the Bayesian prior distribution, is updated into posterior samples taking into account the available data. The most popular ensemble-based method is the ensemble Kalman filter (EnKF) (Evensen, 1994, 2007), which was developed as a robust sequential data-assimilation technique. A modification of the EnKF for solving parameter-estimation problems is the ensemble smoother (ES), whereby all available data are assimilated in one global update step rather than sequentially. The underlying equations for both the EnKF and ES may be derived from Bayesian statistics (e.g., van Leeuwen, 2001; Evensen, 2007). To deal with non-linear problems, iterative ensemble techniques have been proposed (e.g., Reynolds et al., 2006; Emerick and Reynolds, 2012b; Elsheikh et al., 2013; Stordal and Elsheikh, 2015). The ensemble smoother with multiple data assimilation (ES-MDA) is one of such techniques, in which the single update step of ES is replaced with a number of smaller updates (Emerick and Reynolds, 2012a). The large advantage of ES methods over MCMC for stochastic parameter estimation is that the executions of the forward solver can be parallelized in a straightforward manner.

Despite the computational advantages of ensemble methods over Bayesian-MCMC approaches, it is well known that large ensembles are required for the most accurate parameter estimates and predictions (e.g., Buizza and Palmer, 1998; Chen and Zhang, 2006; Evensen, 2007). As a result, we still have with ensemble methods the possibility that, for high-dimensional inverse problems involving expensive forward solvers, accurately sampling from the posterior distribution will remain computationally prohibitive. In such cases, the only solution is to employ an approximate forward solver or proxy. Generating such a proxy can be achieved by simplifying the physics of the problem (e.g., Scholer et al., 2012; Josset et al., 2015a,b), by coarsening the forward model discretization (e.g., Arridge et al., 2006; Calvetti et al., 2014), or by constructing a surrogate model based on, for example, polynomial chaos expansion, Gaussian processes, or neural network techniques (e.g., Khu and Werner, 2003; Rasmussen and Williams, 2006; Marzouk and Xiu, 2009; Goh et al., 2013). However, using a proxy forward solver in the inversion introduces model error, which has the potential to strongly bias posterior statistics (Laloy et al., 2013) and can lead to highly overconfident estimates of the wrong parameters (i.e., posterior collapse) if not accounted for.

To address the issue of model error arising from the use of proxy models in stochastic inversion, researchers have typically focused on two general approaches, both of which rely upon pairs of detailed and proxy solver runs corresponding to different sets of model parameters. In the first approach, these pairs are used to construct a global error model, whose statistics are incorporated into the estimation procedure through, for example, the Bayesian likelihood function (e.g., Kaipio and Somersalo, 2007; Lehikoinen et al., 2010; Schoups and Vrugt, 2010; Evin et al., 2014; Hansen et al., 2014; Smith et al., 2015; Piccolo and Cullen, 2016; Oliver and Alfonzo, 2018). Although this can be highly effective in some cases, we have found that the model errors for many inverse problems exhibit complex behavior that cannot be described in the same way over the entire parameter space. With the second approach, the aim is to construct a local error model, which is generally accomplished through some kind of interpolation between known model-error realizations (e.g., Kennedy and O'Hagan, 2001; Xu et al., 2014; Josset et al., 2015a; Cui et al., 2018). Although doing this effectively addresses the non-global nature of the model errors, it is implicitly assumed that the model-response surface is smooth enough for interpolation to be effective, and problems may arise in regions of the model parameter space that are not well-sampled by the model-error realizations.

Recently, Köpke et al. (2018) presented a new approach to account for model error arising from the use of proxy forward solvers in Bayesian-MCMC inversion, whereby information about the error is gathered during the inversion procedure through occasional runs of the proxy and detailed solvers together, the results of which are stored in a dictionary. In contrast to the existing methods mentioned above, the approach of Köpke et al. (2018) focuses on identifying by projection the model-error component of the residual through the construction of a local orthogonal model-error basis, rather than on the development of a global or local error model. In this paper, we adapt this methodology for use with ES parameter-estimation methods. In particular, we incorporate the related ideas into the ES-MDA algorithm, where for each ensemble member, the local basis is created using the K-nearest-neighbor (KNN) entries in the model-error dictionary. Doing this enables us to accurately solve the parameter-estimation problem using large ensembles, while at same time reduce computational costs through the use of a proxy solver.

The paper is organized as follows: In section 2 we begin with a short review of ensemble methods followed by the presentation of our approach to account for model error. In section 3, we then show the results of applying this methodology to the geophysical inverse problem of estimating spatially distributed radar-wave slowness from synthetic crosshole ground-penetrating radar (GPR) travel-time data. In this regard, results are compared with inversions based on the standard ES-MDA procedure for reference. Based on these findings, we discuss in section 4 how our results compare with standard MCMC sampling, the choice of parameters in our algorithm needed to provide an optimal balance between computational efficiency and accuracy, as well as how the inversion results progress as a function of ES-MDA iteration. Finally, in section 5, we conclude with some general comments on the methodology and provide directions for future research in this domain.

2. METHODOLOGY

2.1. Ensemble Methods

In a generic formulation of the ensemble Kalman filter (EnKF), the model state vector $\mathbf{y}^n$ at data assimilation time step n is updated after the state forecast step. The update from forecast f to analysis a is carried out using the following equation (Emerick and Reynolds, 2012a):

$$\mathbf{y}_j^{n,a} = \mathbf{y}_j^{n,f} + \mathbf{K}^n(\mathbf{d}_{pert}^n - H(\mathbf{y}_j^{n,f})),\qquad(1)$$

with Kalman matrix

$$\mathbf{K}^n = \mathbf{C}_{YD}^{n,f}(\mathbf{C}_{DD}^{n,f} + \mathbf{C}_D^n)^{-1}.\qquad(2)$$

Here, $j = 1, 2, ..., n_e$, where n_e is the number of ensemble members; $\mathbf{C}_{YD}^{n,f}$ is the calculated cross-covariance matrix between the forecast state vector $\mathbf{y}_j^{n,f}$ and the predicted data $\mathbf{d}_j^n = H(\mathbf{y}_j^{n,f})$ obtained through the observation operator $H(\cdot)$; $\mathbf{C}_{DD}^{n,f}$ is the calculated auto-covariance matrix of the predicted data; $\mathbf{C}_D^n$ is the covariance matrix of the observed-data measurement errors; and $\mathbf{d}_{pert}^n$ is the vector of perturbed observations. The latter is obtained using $\mathbf{d}_{pert}^n \sim \mathcal{N}(\mathbf{d}_{obs}^n, \mathbf{C}_D^n)$, where $\mathbf{d}_{obs}^n$ denotes the observed data.

The ensemble smoother (ES) is a variation of the EnKF update formula presented in Equations (1) and (2) that is specifically formulated for parameter estimation problems. The general forward problem

$$\mathbf{d}_{obs} = F(\mathbf{m}_{true}) + \epsilon_d\qquad(3)$$

links a set of observed data $\mathbf{d}_{obs}$ to a set of "true" model parameters $\mathbf{m}_{true}$ through the forward operator $F(\cdot)$ with measurement errors $\epsilon_d \sim \mathcal{N}(0, \mathbf{C}_D)$. The corresponding ES update equation is given by (Emerick and Reynolds, 2012a)

$$\mathbf{m}_j^a = \mathbf{m}_j^f + \mathbf{K}(\mathbf{d}_{pert} - F(\mathbf{m}_j^f)),\qquad(4)$$

with

$$\mathbf{K} = \mathbf{C}_{MD}^f(\mathbf{C}_{DD}^f + \mathbf{C}_D)^{-1}.\qquad(5)$$

Here, $\mathbf{m}_j^f$ and $\mathbf{m}_j^a$ denote the forecast and analyzed model-parameter vectors, respectively, which correspond to an update from prior to posterior; $\mathbf{C}_{MD}^f$ is the cross-covariance matrix between $\mathbf{m}_j^f$ and the predicted data $\mathbf{d}_j = F(\mathbf{m}_j^f)$; $\mathbf{C}_{DD}^f$ is the auto-covariance matrix of the predicted data; and $\mathbf{d}_{pert} \sim \mathcal{N}(\mathbf{d}_{obs}, \mathbf{C}_D)$ is again a vector of perturbed observations. The idea with equations (4) and (5) is that, after defining an initial parameter ensemble by drawing from the Bayesian prior distribution, the ensemble members are updated to represent samples from the posterior distribution in a single analysis step that incorporates all of the available data.

2.2. The ES-MDA Algorithm

ES offers an efficient tool to solve parameter-estimation problems under the assumptions that the prior parameter distribution is Gaussian and the forward operator $F(\cdot)$ is linear. If these conditions are not satisfied, then ES can lead to unacceptable data matches and unphysical results (Aanonsen et al., 2009). To deal with this issue, we focus in this paper on a recent development by Emerick and Reynolds (2012a), namely the ensemble smoother with multiple data assimilation (ES-MDA). With this approach, one standard ES step, which is comparable to a single Gauss-Newton iteration when maximizing the posterior probability of the model parameters (Tarantola, 2005), is replaced by a number of smaller update steps (or assimilation iterations) based on a Kalman matrix and perturbed data vector that are recalculated at each iteration. In order to correctly sample from the posterior distribution, the measurement-error covariance matrix $\mathbf{C}_D$ must be inflated in this procedure. Typically this is done by scaling $\mathbf{C}_D$ by the number of assimilation iterations; however, more generalized inflation coefficients may be used (Emerick and Reynolds, 2012a). For linear forward solvers, the ES-MDA algorithm is theoretically equivalent to standard ES (Emerick and Reynolds, 2012b). For non-linear problems, it can be shown that the methodology has links to annealed importance sampling (Stordal and Elsheikh, 2015).

Algorithm 1 outlines the steps involved in the ES-MDA procedure where, for simplicity, the measurement-error covariance matrix is assumed diagonal with entries σ^2 and the corresponding inflation coefficient α is set to equal the number of assimilation iterations n_{iter}. To estimate the inverse of matrix $(\mathbf{C}_{DD} + \alpha \cdot \mathbf{C}_D)$ we use the truncated singular value decomposition (TSVD) and retain 99% of the total energy of the singular values (Emerick and Reynolds, 2012a).

Algorithm 1: Standard ES-MDA

---

1   **set** $\mathbf{C}_D = \sigma^2\,\mathbf{I}$; $\alpha = n_{iter}$
2   **draw** prior ensemble $\mathbf{m}$ containing n_e sets of model parameters
3   **for** $i = 1, ..., n_{iter}$ **do**
4   **perturb** observations $\mathbf{d}_{pert} \sim \mathcal{N}(\mathbf{d}_{obs}, \alpha \cdot \mathbf{C}_D)$
5   **for** $j = 1, ..., n_e$ **do**
6   **compute** predicted data $\mathbf{d}_j = F(\mathbf{m}_j)$
7   **end**
8   **compute** Kalman matrix $\mathbf{K} = \mathbf{C}_{MD} \cdot (\mathbf{C}_{DD} + \alpha \cdot \mathbf{C}_D)^{-1}$
9   **for** $j = 1, ..., n_e$ **do**
10   **compute** residual $\mathbf{r}_j = \mathbf{d}_{pert} - \mathbf{d}_j$
11   **update** ensemble $\mathbf{m}_j = \mathbf{m}_j + \mathbf{K}\,\mathbf{r}_j$
12   **end**
13 **end**

2.3. Model Error

When working with a perfectly known forward solver $F(\cdot)$ in the ES-MDA procedure outlined above, the residual $\mathbf{r}_j$ corresponding to the jth ensemble member $\mathbf{m}_j$, which quantifies

the misfit between the perturbed observations and the predicted (forward-calculated) data, is given by

$$
\begin{aligned}
\mathbf{r}_j &= \mathbf{d}_{pert} - F(\mathbf{m}_j) \\
&= \underbrace{F(\mathbf{m}_{true}) - F(\mathbf{m}_j)}_{\text{parameter-error component}} + \tilde{\epsilon}_d,
\end{aligned}
\tag{6}
$$

where $\tilde{\epsilon}_d$ denotes the sum of the measurement errors and perturbation noise. In the case where $\mathbf{m}_j = \mathbf{m}_{true}$, we see from Equation (6) that the parameter-error term, which represents the component of the residual related to being at the wrong set of model-parameter values, will be zero and that the residual energy will tend to be minimized. In the case where a proxy forward solver $\hat{F}(\cdot)$ is used in the ES-MDA algorithm, however, the latter does not generally hold true because

$$
\begin{aligned}
\mathbf{r}_j &= \mathbf{d}_{pert} - \hat{F}(\mathbf{m}_j) \\
&= F(\mathbf{m}_{true}) - \hat{F}(\mathbf{m}_j) + \tilde{\epsilon}_d \\
&= \underbrace{F(\mathbf{m}_{true}) - F(\mathbf{m}_j)}_{\text{parameter-error component}} + \underbrace{F(\mathbf{m}_j) - \hat{F}(\mathbf{m}_j)}_{\text{model-error component}} + \tilde{\epsilon}_d.
\end{aligned}
\tag{7}
$$

Indeed, the presence of an additional model-error component in Equation (7) compared to Equation (6) means that the residual energy may be minimized for model parameter vectors $\mathbf{m}_j$ that are substantially different from $\mathbf{m}_{true}$, as such parameter sets well tend to compensate for the model errors. As mentioned previously, this can lead to strongly biased and overconfident posterior statistics.

In order to deal with model error in the ES-MDA procedure arising from use of a proxy solver, we build on the methodology presented in Köpke et al. (2018) for Bayesian-MCMC inversion, which focuses on identifying the model-error component of the residual using a projection-based method. We refer the reader to that paper for details beyond those given here. Algorithm 2 outlines the steps involved in our modified ES-MDA methodology, again assuming that $\alpha = n_{iter}$ and $\mathbf{C}_D = \sigma^2 \mathbf{I}$ for simplicity, where $\mathbf{I}$ is the identity matrix. In addition we introduce n_d, defined as the number of detailed solver runs used to learn about the model error, and set it to a value to less than or equal to the number of ensemble members n_e.

In the modified ES-MDA algorithm, initial ensemble members $\mathbf{m}_j$ are drawn from the prior parameter distribution and the corresponding predicted data $\hat{\mathbf{d}}_j = \hat{F}(\mathbf{m}_j)$ are computed using the proxy solver. In each assimilation iteration, a subset of the ensemble members having size n_d is randomly chosen, for which the detailed forward responses $\mathbf{d}_j = F(\mathbf{m}_j)$ are also calculated. The resulting n_d model-error vectors (i.e., $\mathbf{d}_j - \hat{\mathbf{d}}_j$) and corresponding parameter sets $\mathbf{m}_j$ are stored in the dictionaries $\mathbf{D}^E$ and $\mathbf{D}^M$, respectively. As $\mathbf{D}^M$ and $\mathbf{D}^E$ are further enriched with n_d entries in each ES-MDA iteration, more detailed information about the model error around the posterior solution is gathered.

For each ensemble member $\mathbf{m}_j$, the model-error component of the residual is identified and used to correct the proxy response in order to mimic the detailed forward solver. To this end,

the current model-parameter dictionary $\mathbf{D}^M$ is searched for the K-nearest-neighbor (KNN) parameter sets to $\mathbf{m}_j$ using a standard Euclidean distance measure (e.g., Hastie et al., 2009). An orthonormal basis $\mathbf{B}_j$ for the model error at $\mathbf{m}_j$ is then constructed from these parameter sets using the Gram-Schmidt technique (e.g., Strang et al., 1993). We assume in our work that the data-measurement-error and parameter-error components of the residual are orthogonal to the model-error component, and therefore cannot be well represented by the basis. An estimate of the model error $\tilde{\mathbf{e}}_j$ is thus obtained by projecting the residual onto $\mathbf{B}_j$

$$
\tilde{\mathbf{e}}_j = \mathbf{B}_j \cdot \mathbf{B}_j^T \cdot \mathbf{r}_j.
\tag{8}
$$

This result, which represents the details missing in the proxy solution, is then added to the proxy response to obtain a corrected forward response

$$
\tilde{\mathbf{d}}_j = \hat{\mathbf{d}}_j + \tilde{\mathbf{e}}_j
\tag{9}
$$

with corresponding corrected residual

$$
\tilde{\mathbf{r}}_j = \mathbf{d}_{pert} - \tilde{\mathbf{d}}_j.
\tag{10}
$$

The corrected forward responses for all of the ensemble members are used to compute the corrected Kalman matrix

$$
\tilde{\mathbf{K}} = \mathbf{C}_{M\tilde{D}} \cdot (\mathbf{C}_{\tilde{D}\tilde{D}} + \alpha \cdot \mathbf{C}_D)^{-1}
\tag{11}
$$

which is used with the corrected residuals to update the ensemble.

Under the stated assumptions and with appropriate choices of n_e, n_d, and K, Algorithm 2 allows us to effectively reduce the computational cost of ES-MDA when considering large ensembles through the use of a proxy solver. The dimensionality of the parameter-estimation problem and the difference in computational cost between the proxy and detailed forward solutions determine how much computational benefit is derived from this methodology. We refer the reader to Köpke et al. (2018) for a detailed discussion of the orthogonality assumption between the model-error and other components of the residual.

3. APPLICATION TO CROSSHOLE GPR TOMOGRAPHY

3.1. Experimental Design and Forward Models

As an example, we now apply our modified ES-MDA algorithm with model-error correction to the crosshole GPR travel-time tomography inverse problem. A transmitter and receiver antenna, located in two adjacent boreholes, are used to obtain the travel times of radar energy between the holes for different antenna positions. These times are linked to the spatial distribution of subsurface radar-wave velocity, the estimation of which is the goal of the inverse problem. Crosshole GPR travel-time tomography represents an excellent test problem for our purposes because (i) it has been extremely well-studied, most notably from a stochastic inverse standpoint (e.g., Giroux et al.,

---

Algorithm 2: ES-MDA with model-error correction

1 **set** $\mathbf{C}_D = \sigma^2 \mathbf{I}$; $\alpha = n_{iter}$; $n_d \leq n_e$; $l = 0$
2 **draw** prior ensemble $\mathbf{m}$ containing n_e sets of model
    parameters
3 **for** $i = 1, ..., n_{iter}$ **do**
4 $\quad$ **perturb** observations $\mathbf{d}_{pert} \sim \mathcal{N}(\mathbf{d}_{obs}, \alpha \cdot \mathbf{C}_D)$
5 $\quad$ **select** random subset $\mathbf{m}_r$ of $\mathbf{m}$ with $r = 1, ..., n_d$
6 $\quad$ **for** $j = 1, ..., n_e$ **do**
7 $\quad\quad$ **compute** predicted data using proxy solver
        $\hat{\mathbf{d}}_j = \hat{F}(\mathbf{m}_j)$
8 $\quad\quad$ **if** $\mathbf{m}_j \in \mathbf{m}_r$ **then**
9 $\quad\quad\quad$ $l = l + 1$
10 $\quad\quad\quad$ **compute** predicted data using detailed solver
          $\mathbf{d}_l = F(\mathbf{m}_j)$
11 $\quad\quad\quad$ **enrich** model-error dictionary $\mathbf{D}_l^E = \mathbf{d}_l - \hat{\mathbf{d}}_j$
12 $\quad\quad\quad$ **enrich** corresponding model-parameter dictionary
          $\mathbf{D}_l^M = \mathbf{m}_j$
13 $\quad\quad$ **end**
14 $\quad$ **end**
15 $\quad$ **for** $j = 1, ..., n_e$ **do**
16 $\quad\quad$ **search** dictionary $\mathbf{D}^M$ for K-nearest neighbors to $\mathbf{m}_j$
17 $\quad\quad$ **take** corresponding entries from $\mathbf{D}^E$ and place in set
        $\mathbf{D}^K(\mathbf{m}_j)$
18 $\quad\quad$ **build** orthonormal basis $\mathbf{B}$ having span$\{\mathbf{D}^K(\mathbf{m}_j)\}$
19 $\quad\quad$ **compute** residual $\mathbf{r}_j = \mathbf{d}_{pert} - \hat{\mathbf{d}}_j$
20 $\quad\quad$ **project** $\mathbf{r}_j$ onto $\mathbf{B}$ to estimate model error
        $\tilde{\mathbf{e}}_j = \mathbf{B} \cdot \mathbf{B}^T \cdot \mathbf{r}_j$
21 $\quad\quad$ **correct** proxy response $\tilde{\mathbf{d}}_j = \hat{\mathbf{d}}_j + \tilde{\mathbf{e}}_j$
22 $\quad$ **end**
23 $\quad$ **compute** Kalman matrix $\tilde{\mathbf{K}} = \mathbf{C}_{M\tilde{D}} \cdot (\mathbf{C}_{\tilde{D}\tilde{D}} + \alpha \cdot \mathbf{C}_D)^{-1}$
24 $\quad$ **for** $j = 1, ..., n_e$ **do**
25 $\quad\quad$ **compute** residual $\tilde{\mathbf{r}}_j = \mathbf{d}_{pert} - \tilde{\mathbf{d}}_j$
26 $\quad\quad$ **update** ensemble $\mathbf{m}_j = \mathbf{m}_j + \tilde{\mathbf{K}}\tilde{\mathbf{r}}_j$
27 $\quad$ **end**
28 **end**

2007; Looms et al., 2008; Scholer et al., 2012; Hansen et al., 2013; Linde and Vrugt, 2013); (ii) it involves a high-dimensional and spatially distributed set of model parameters that must be estimated; and (iii) the forward problem can be solved in a variety of different ways using different physical approximations.
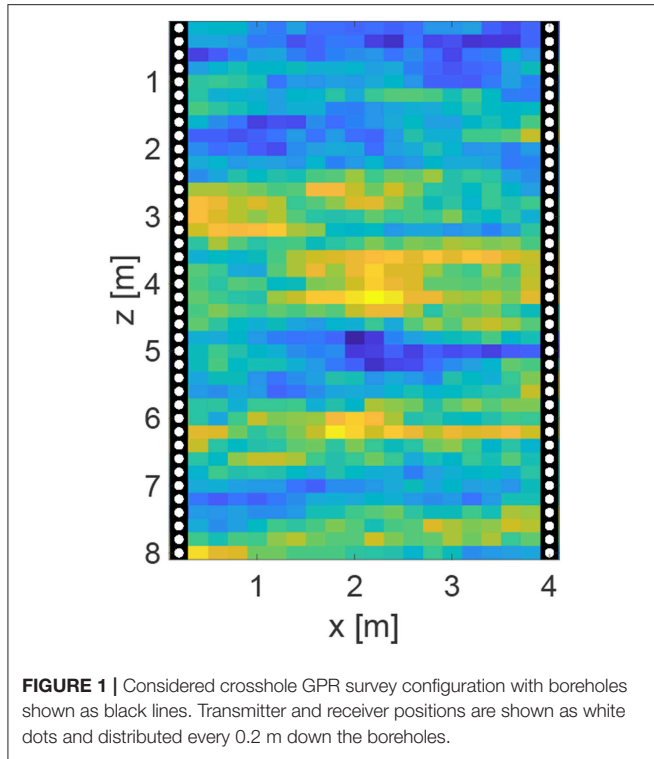
GPR travel times are linked to the spatial distribution of electrical properties between the two boreholes, predominantly the dielectric permittivity, through Maxwell's equations. Numerical solution of these equations represents the most accurate means of calculating the travel times, but at the same time it is highly computationally expensive. To reduce the computational cost, the physics of the electromagnetic wave propagation can be approximated using ray theory, whereby the effects of frequency are ignored and we solve the eikonal equation (e.g., Nowack, 1992). To decrease the computational cost even further, the straight-ray approximation may also

be considered, which means that the ray paths that connect transmitter and receiver locations are assumed to be straight lines (e.g., Cordua et al., 2008). The latter approximation is typically applied in cases where contrasts in velocity do not exceed 10%; however it is only truly valid when the subsurface is homogeneous. Here, we consider the eikonal equation to be our detailed forward solver $F(\cdot)$ and the straight-ray approach to be our proxy solver $\hat{F}(\cdot)$. This choice was made for demonstration purposes, as it allows us to compare the results of ES-MDA inversions obtained using our approach to those obtained using standard ES-MDA, as well as MCMC, based on the detailed solver alone. That is, the eikonal solution is fast enough to allow it to be used in the standard ES-MDA algorithm with a large number of ensemble members, as well as for MCMC posterior sampling. Note that, instead of estimating directly velocity from GPR travel times in our analysis, we focus on the estimation of subsurface slowness (the reciprocal of velocity) which makes the straight-ray forward problem linear.

The survey configuration for our synthetic experiments consists of two boreholes that are 8-m deep and 4-m apart (**Figure 1**). Transmitter and receiver antenna positions are distributed equally in depth every 0.2 m down the left and right boreholes, respectively. Sending a radar pulse from all transmitter positions to all receiver positions yields 1,600 travel-time data. We consider a pixel-based parameterization of the subsurface whereby the region between the boreholes by discretized into 20×40 square cells of constant-slowness and side length 0.2 m. The synthetic "true" subsurface and initial prior ensemble members are generated by sequential Gaussian simulation using the GSLIB software package (Deutsch and Journel, 1992). The mean slowness is set to 10 ns/m and an exponential auto-covariance kernel having a standard deviation of 1.7 ns/m is assumed, with horizontal and vertical correlation lengths of 6 m and 1.5 m, respectively. The corresponding synthetic observed data are generated by solving the eikonal equation and adding measurement errors, the latter of which are simulated as Gaussian random noise having covariance matrix $\mathbf{C}_D = \sigma^2 \mathbf{I}$ with standard deviation $\sigma = 0.2$ ns.

3.2. ES-MDA Results

Our goals in this analysis are to (i) study the effects of model error on ES-MDA inversions; (ii) investigate the influence of the ensemble size on the accuracy of the results obtained; and (iii) explore how the parameters of our modified ES-MDA procedure with model-error correction can be chosen to provide an optimal balance between computational efficiency and accuracy. To this end, we compare parameter-estimation results for different numbers of ensemble members when (i) there is no model error, meaning that the detailed (eikonal-equation) forward solver is used within the standard ES-MDA procedure (Algorithm 1); (ii) model error is present but not accounted for, meaning that the proxy (straight-ray) forward solver is used within the standard ES-MDA procedure; and (iii) model error is present and accounted for through the use of Algorithm 2. In each case, we examine the combined results from 10 ES-MDA inversions obtained using different initial ensembles and $n_{iter} =$

FIGURE 1 | Considered crosshole GPR survey configuration with boreholes shown as black lines. Transmitter and receiver positions are shown as white dots and distributed every 0.2 m down the boreholes.

8 assimilation iterations. Ensemble sizes of $n_e = 20, 40, 80, 160, 320$, and 640 are considered in our analysis.

To assess the quality of the inversion results, we consider two metrics. The average root-mean-square (RMS) travel-time misfit, which quantifies globally the ability of the posterior ensemble to represent the observed data, is defined for ES-MDA run i ($i = 1, 2, ..., 10$) as follows:

$$M_i^T = \frac{1}{n_e} \sum_{j=1}^{n_e} \frac{1}{\sqrt{n_T}} \left\| \mathbf{d}_{obs} - \mathbf{d}_{i,j} \right\|_2 . \tag{12}$$

where n_T is the number of travel-time data. For the case where model error is absent and data errors are zero-mean and Gaussian distributed with covariance matrix $\mathbf{C}_D = \sigma^2 \mathbf{I}$, the expected value of M_i^T will be σ. Note that, in the case where model error is present but not accounted for, the detailed forward solver $\mathbf{d}_{i,j}$ in equation (12) is replaced with the proxy solver $\hat{\mathbf{d}}_{i,j}$.

We also consider in our analysis the average RMS slowness misfit, defined by

$$M_i^S = \frac{1}{n_e} \sum_{j=1}^{n_e} \frac{1}{\sqrt{n_T}} \left\| \mathbf{m}_{true} - \mathbf{m}_{i,j} \right\|_2 , \tag{13}$$
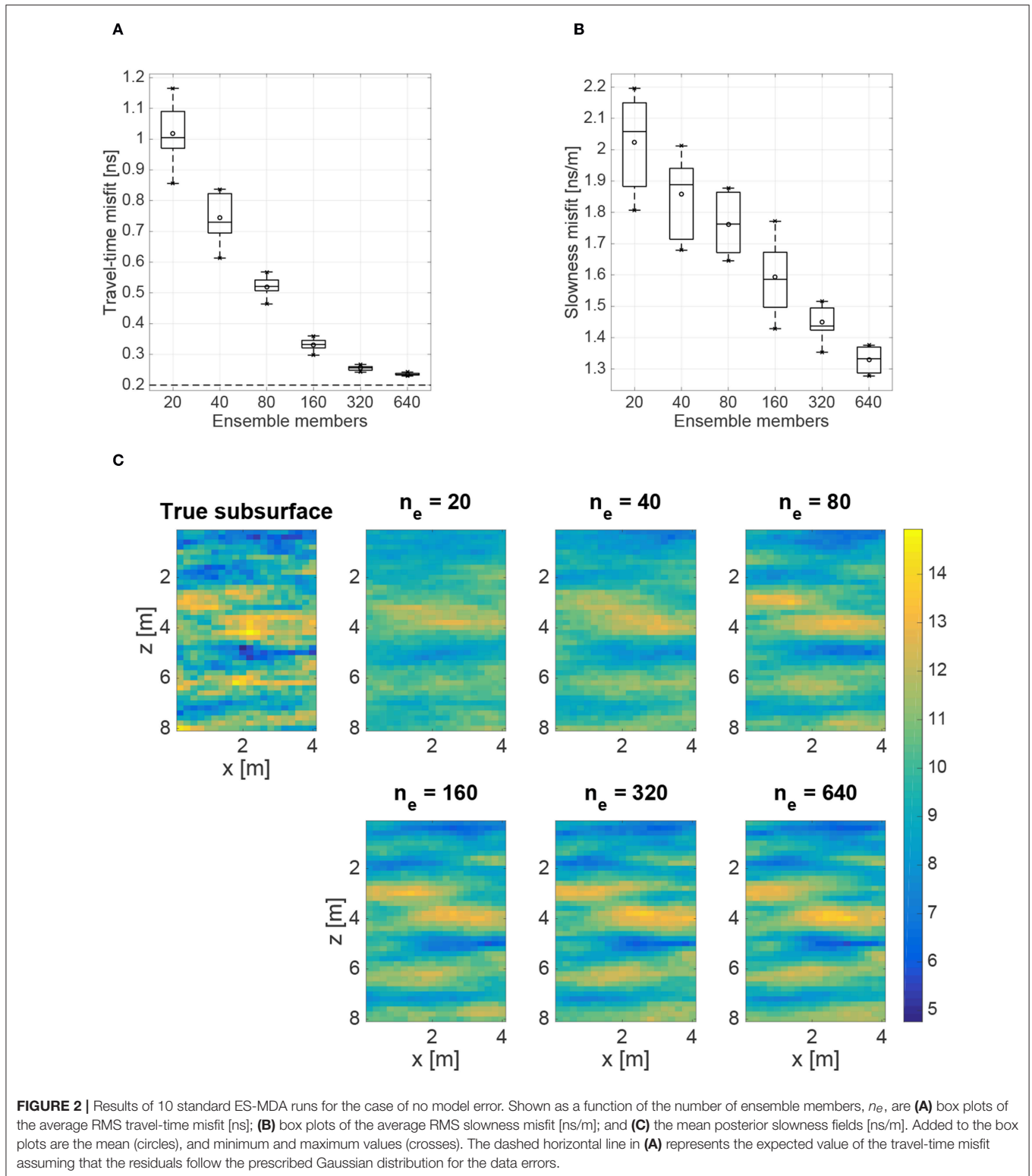
where n_S is the number of slowness cells. This metric quantifies how well the posterior ensemble captures the true underlying model parameters, and can only be employed in the case of synthetic data where the true subsurface slowness distribution is known. In addition to the two metrics in Equations (12) and (13), we plot the mean slowness fields over all ensemble members and

all ES-MDA runs in order to visually compare them with the true slowness distribution.

Figure 2 summarizes the parameter-estimation results obtained for the case where there is no model error. In **Figure 2A** we observe that the average RMS travel-time misfit decreases consistently with larger numbers of ensemble members toward the expected value of $\sigma = 0.2$ ns which reflects the prescribed data errors. After around 320 ensemble members, adding more members is seen to only slightly further improve the results. **Figure 2B** shows that the slowness misfit also decreases consistently as a function of ensemble size. This is supported by **Figure 2C**, which shows that the mean slowness fields become increasingly detailed and similar to the true subsurface distribution as the number of ensemble members increases. The increasing overall accuracy of the ES-MDA results with larger ensemble size is based on the reduction of sampling errors following the central limit theorem (Evensen, 2007). We can conclude that larger ensembles combined with the detailed forward solver enable us to obtain more reliable posterior parameter estimates, but at the cost of significantly greater computational effort when the detailed solver is computationally expensive.

Figure 3 summarizes the parameter-estimation results obtained for the case where model error is present but not accounted for. In **Figure 3A** we observe that, in accordance with **Figure 2A**, the travel-time misfit consistently decreases with larger ensemble size. However, it approaches a stable value that is well above the target value of $\sigma = 0.2$ ns, because the presence of model error does not allow data fitting to a level that is in accordance with the prescribed data errors. In addition, unlike in **Figure 2B**, the slowness misfit now decreases only until $n_e = 40$, after which it increases again (**Figure 3B**). For $n_e \leq 40$, we do not have enough ensemble members to resolve the details of the posterior distribution and therefore only the posterior mean can be represented in the parameter estimation results (Chen and Zhang, 2006). Conversely, when $n_e > 40$, the solution moves toward a biased posterior distribution. That is, with more ensemble members the parameters have the ability to compensate for the model error and the data become over-fitted, meaning that a better data match is achieved but the parameters do not represent the true subsurface model. The mean slowness fields in **Figure 3C** allow us to see how the model error introduces bias in the parameter-estimation results as n_e increases; strong artifacts are clearly observed in these fields for ensemble sizes larger than 160.

Finally, we examine the inversion results obtained for the case where model error is present and accounted for through our modified ES-MDA approach. We first consider inversions where $n_d = 20$ detailed solver runs per iteration are used to build the model-error dictionary and $K = 20$ KNN from this dictionary are used to construct the model-error basis for each ensemble member. The corresponding results are shown in **Figure 4**. Note that, in the case with $n_e = 20$, the detailed solver is executed for each ensemble member, which corresponds to the standard ES-MDA procedure in Algorithm 1. In **Figures 4A,B** we see that both the travel-time and slowness misfit decrease

FIGURE 2 | Results of 10 standard ES-MDA runs for the case of no model error. Shown as a function of the number of ensemble members, n_e, are **(A)** box plots of the average RMS travel-time misfit [ns]; **(B)** box plots of the average RMS slowness misfit [ns/m]; and **(C)** the mean posterior slowness fields [ns/m]. Added to the box plots are the mean (circles), and minimum and maximum values (crosses). The dashed horizontal line in **(A)** represents the expected value of the travel-time misfit assuming that the residuals follow the prescribed Gaussian distribution for the data errors.

from $n_e = 20$ until they reach a minimum at around 80–160 ensemble members. This demonstrates that the consideration of larger ensembles through use of a proxy solver combined with our model-error correction can lead to more accurate

results compared to standard ES-MDA based on small ensembles and the detailed forward solver. These results are confirmed in **Figure 4C**, where we observe that the bias is largely removed from the mean slowness fields for $n_e \leq 160$ in comparison to
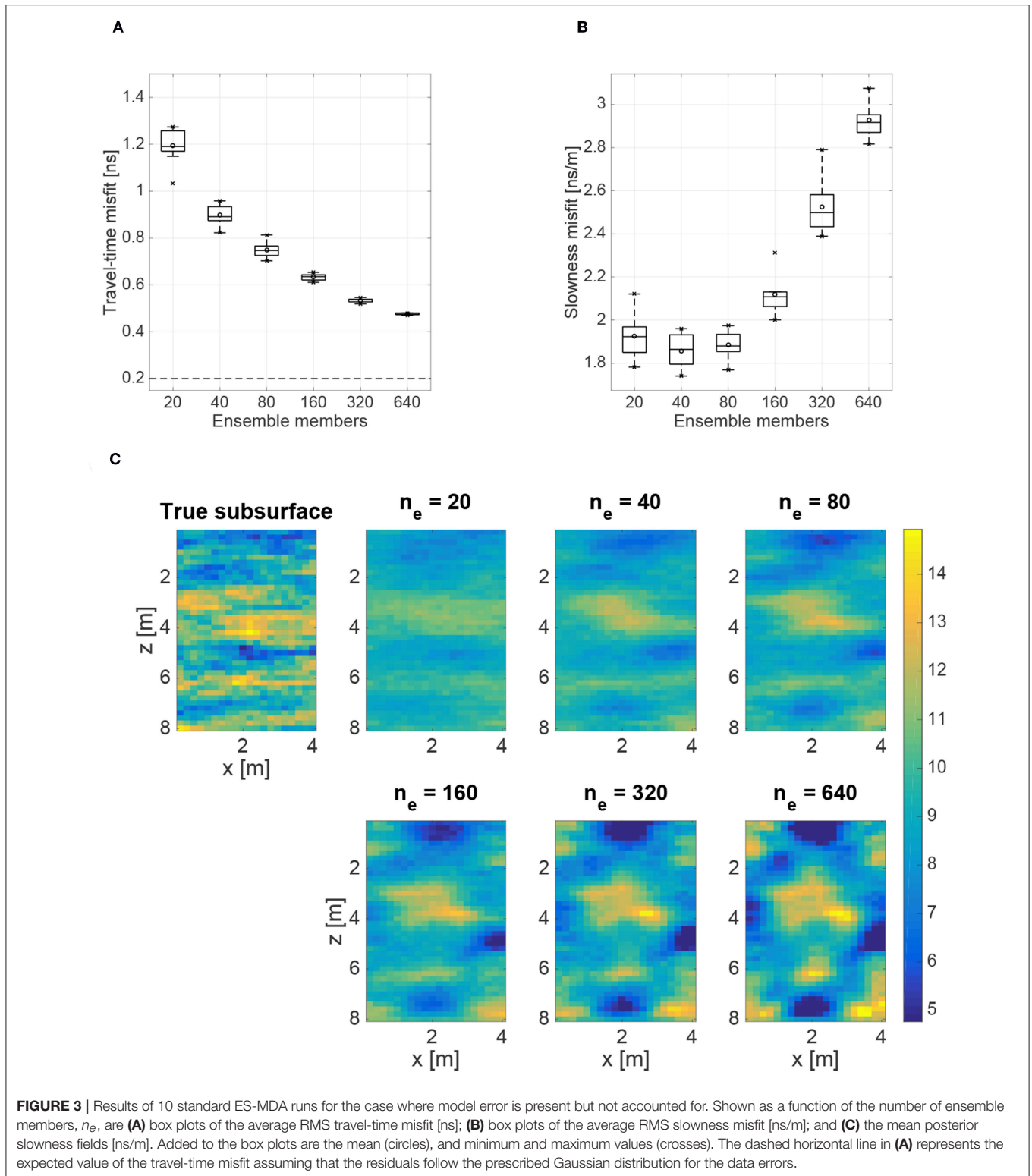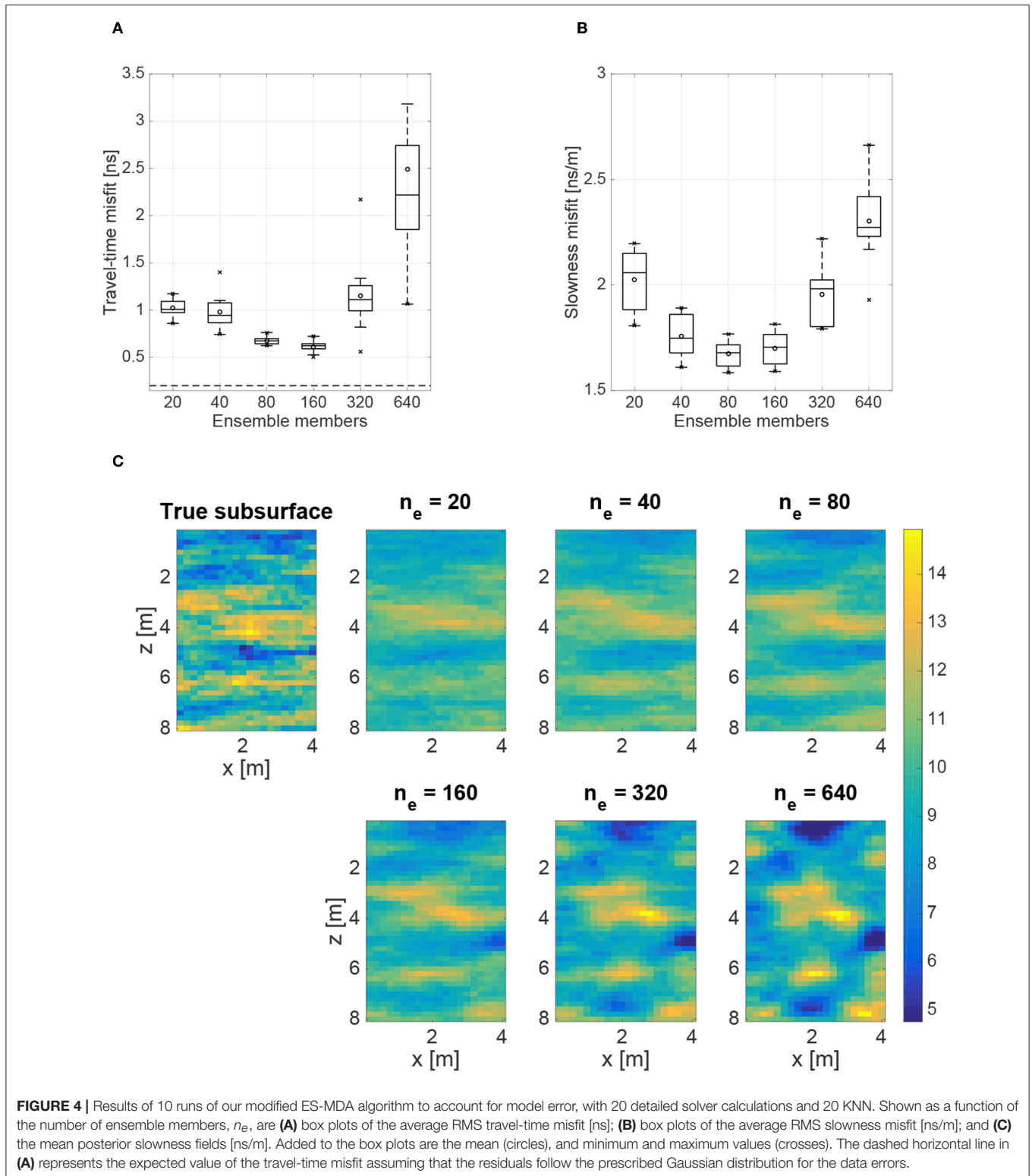
FIGURE 3 | Results of 10 standard ES-MDA runs for the case where model error is present but not accounted for. Shown as a function of the number of ensemble members, n_e, are **(A)** box plots of the average RMS travel-time misfit [ns]; **(B)** box plots of the average RMS slowness misfit [ns/m]; and **(C)** the mean posterior slowness fields [ns/m]. Added to the box plots are the mean (circles), and minimum and maximum values (crosses). The dashed horizontal line in **(A)** represents the expected value of the travel-time misfit assuming that the residuals follow the prescribed Gaussian distribution for the data errors.

Figure 4C. However, for ensemble sizes larger than around 160, the travel-time and slowness misfit are seen to again increase, meaning that the data become under-fitted. That is, the ensemble size becomes too large compa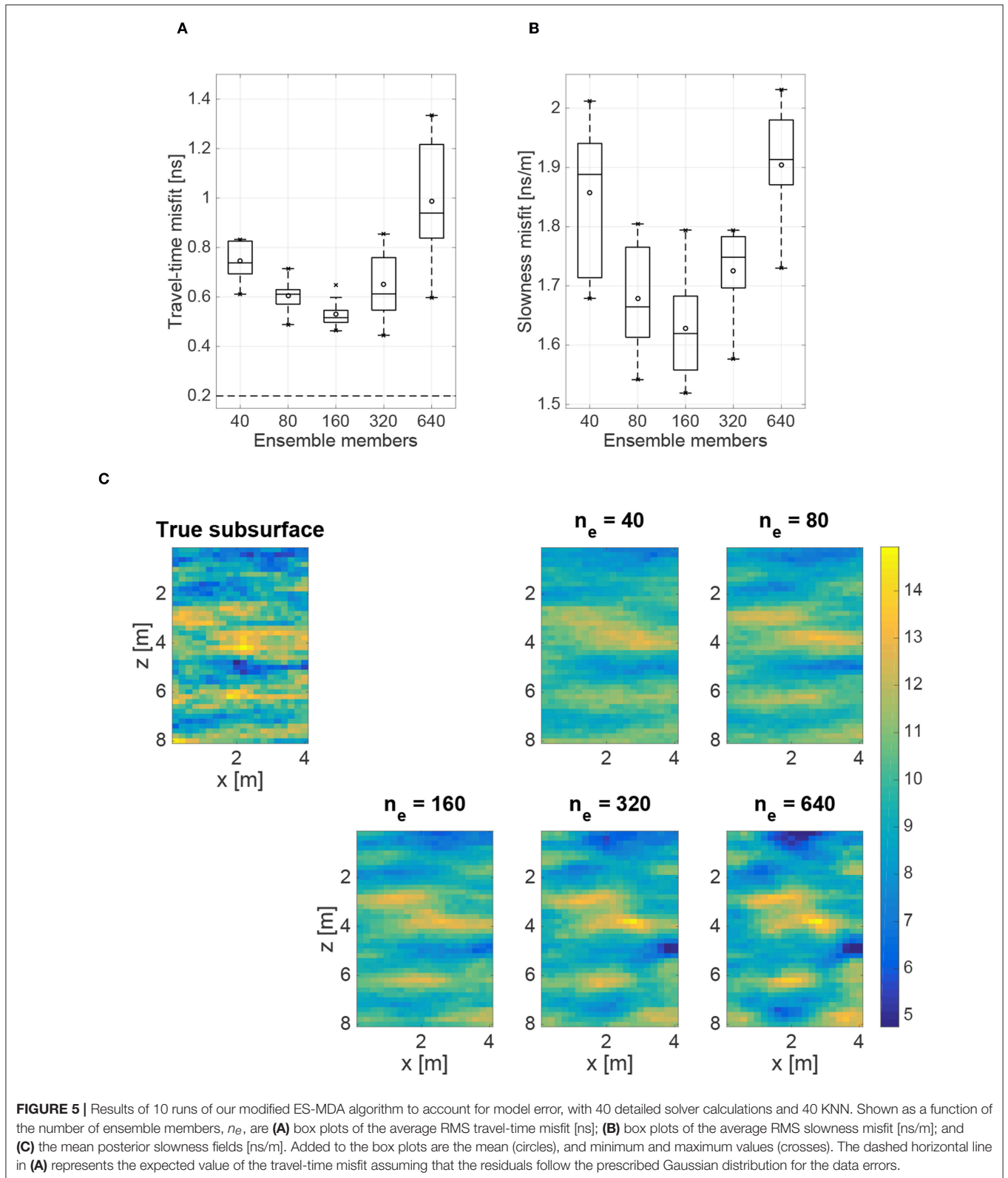red to the number of detailed solver calculations for the model error to be well represented in the dictionary, meaning that projection onto the model-error basis will not properly identify the model-error component of the residual. This has the effect of introducing bias into the inversion

FIGURE 4 | Results of 10 runs of our modified ES-MDA algorithm to account for model error, with 20 detailed solver calculations and 20 KNN. Shown as a function of the number of ensemble members, n_e, are **(A)** box plots of the average RMS travel-time misfit [ns]; **(B)** box plots of the average RMS slowness misfit [ns/m]; and **(C)** the mean posterior slowness fields [ns/m]. Added to the box plots are the mean (circles), and minimum and maximum values (crosses). The dashed horizontal line in **(A)** represents the expected value of the travel-time misfit assuming that the residuals follow the prescribed Gaussian distribution for the data errors.

results, which is clearly seen in the mean slowness fields in **Figure 4C** when $n_e > 160$.

To explore the latter findings, we consider again Algorithm 2, but this time using $n_d = 40$ detailed solver runs per iteration and

$K = 40$ KNN to build the model-error dictionary and construct the model-error basis, respectively. Here, when $n_e = 40$, the detailed solver is executed for each ensemble member, which again corresponds to the standard ES-MDA procedure. Similar

FIGURE 5 | Results of 10 runs of our modified ES-MDA algorithm to account for model error, with 40 detailed solver calculations and 40 KNN. Shown as a function of the number of ensemble members, n_e, are **(A)** box plots of the average RMS travel-time misfit [ns]; **(B)** box plots of the average RMS slowness misfit [ns/m]; and **(C)** the mean posterior slowness fields [ns/m]. Added to the box plots are the mean (circles), and minimum and maximum values (crosses). The dashed horizontal line in **(A)** represents the expected value of the travel-time misfit assuming that the residuals follow the prescribed Gaussian distribution for the data errors.

to before, we observe in **Figures 5A,B** that the travel-time and slowness misfit decrease from $n_e = 40$ until a minimum is reached. Although it is difficult to determine the exact position of this minimum due to the limited discretization, we see that it falls somewhere around 160 ensemble members. After the minimum value, the travel-time and slowness misfit are again

seen to increase and the data become under-fitted. This behavior is well reflected in the mean slowness fields in **Figure 5C**, which show good agreement with the true field for $n_e \leq 320$, but clearly contain model-error-related artifacts when $n_e = 640$.

4. DISCUSSION

We saw above that use of the modified ES-MDA approach described in Algorithm 2 can allow for a significant reduction in posterior bias when employing a proxy forward solver compared to the standard ES-MDA procedure. This offers the possibility of considering large ensemble sizes within ES-MDA, which can be computationally prohibitive in the context of an expensive detailed forward solver. One issue requiring further discussion, however, is the balance between (i) the number of ensemble members considered n_e, which in the case of no model error controls the accuracy of the results obtained; and (ii) the number of detailed solver runs n_d, which determines the success of the model-error correction. We saw in **Figure 4** that, when $n_d = 20$ detailed solver runs per iteration were considered in the modified ES-MDA procedure, use of ensemble sizes between 40 and 160 allowed for an improvement in parameter estimates compared to standard ES-MDA based on the detailed solver with $n_e = 20$. When $n_d = 40$ detailed solver runs per iteration were considered, on the other hand, a corresponding improvement was found for ensemble sizes between 80 and 320. These results suggest that, at least for the application presented in this paper, Algorithm 2 can be successfully applied only for ensembles having size less than 8 times the number of detailed solver runs per iteration. Past this number, there will not be enough entries in the model-error dictionary to allow for an accurate correction of the model error for all ensemble members, and the benefits of using an approximate solver with model-error correction will be compromised. Further exploration of these findings in the context of other inverse problems is required.

Another issue in need of some discussion is how the results of using the standard and modified ES-MDA algorithms presented in **Figures 2–5** compare with samples from the "true" posterior distribution, the latter of which we assume to be available through MCMC sampling based on the detailed forward solver. To this end, we show in **Figure 6** five randomly chosen posterior slowness realizations obtained via MCMC sampling based on the eikonal equation (**Figure 6A**); standard ES-MDA based on both the eikonal equation and straight-ray approximation (**Figures 6B,C**); and our modified ES-MDA procedure with $n_d = 20$ and $n_d = 40$ (**Figures 6D,E**). The point-wise posterior mean and standard deviation, computed over all available samples, are also shown for reference. The results in **Figure 6A** were obtained using the sequential geostatistical simulation technique (e.g., Ruggeri et al., 2015), where after burn-in, the results of 140,000 MCMC iterations were thinned to provide 140 posterior samples. For **Figures 6B–E**, the number of ensemble members considered was chosen to be the maximum investigated value ($n_e = 640$) for standard ES-MDA, whereas for our modified ES-MDA procedure it was set equal to 8 times the number of detailed solver runs, as discussed above.

In comparing the posterior realizations in **Figures 6A,B**, we see that they are highly similar, which suggests that ES-MDA based on the detailed forward solver and using a large number of ensemble members allows for adequate sampling of the Bayesian posterior distribution. The corresponding standard deviation images generally show a pattern that reflects the degree of ray coverage; regions of higher slowness contain a smaller ray density. However, the ES-MDA solution is seen to contain more variability, which may arise because the 140,000 MCMC iterations utilized were not enough to adequately explore the posterior space. In examining the stochastic realizations in **Figure 6C**, the proxy-related bias in this solution is clearly apparent. Here, the symmetric pattern of variability reflects variations in ray density that are controlled solely by the antenna locations in the straight-ray case. Finally, in comparing the results in **Figures 6D,E** with those in **Figure 6A**, we see that our modified ES-MDA algorithm largely removes the proxy-related bias and allows for the generation of posterior samples that are close in appearance to the MCMC solution, which again validates our approach. These samples do, however, show a slightly higher degree of variability with less correlation compared to **Figures 6A,B**, with the $n_d = 40$ solution providing a better match than the $n_d = 20$ solution. As discussed above, with a fixed number of detailed solver runs per iteration, there is an upper limit to the number of ensemble members that can be effectively considered in our procedure, which in turn may not be enough to characterize exactly the posterior distribution (see **Figure 2**). More accurate results would thus likely require greater numbers of detailed solver runs to allow for an increase in the ensemble size. The higher degree of variability in these results may also reflect imperfect removal of the model error, or the lesser number of samples used to compute the point-by-point mean and standard deviation.

Lastly, we wish to elaborate on the number of internal ES-MDA assimilation iterations considered in our approach, which was held constant at a value of $n_{iter} = 8$ for all of the results presented section 3.2. To this end, we study in **Figure 7** the travel-time and slowness misfit as a function of iteration for an ensemble size of $n_e = 160$ when (i) no model error is present; (ii) model error is present but not accounted for; and (iii) model error is present and accounted for using $n_d = 40$ detailed solver runs per iteration and $K = 40$ KNN. We observe overall that the first assimilation iteration has the largest influence on reducing the travel-time and slowness misfit from prior to posterior. This arises because of the linearity of the proxy (straight-ray) solver and the weak non-linearity of the detailed (eikonal) solver in our travel-time tomography application. Indeed, Emerick and Reynolds (2012b) proved that, for linear problems, one ES update using the measurement noise covariance matrix is equivalent to multiple ES-MDA updates using the inflated covariance matrix. When using the detailed solver in the inversion where there is no model error, for example, **Figures 7A,D** show a large decrease in travel-time and slowness misfit after one iteration and a slow decrease from iterations 2–8. In this case 4 iterations would be enough to obtain similar parameter-estimation results to those obtained using 8 iterations. When model error is present but not accounted
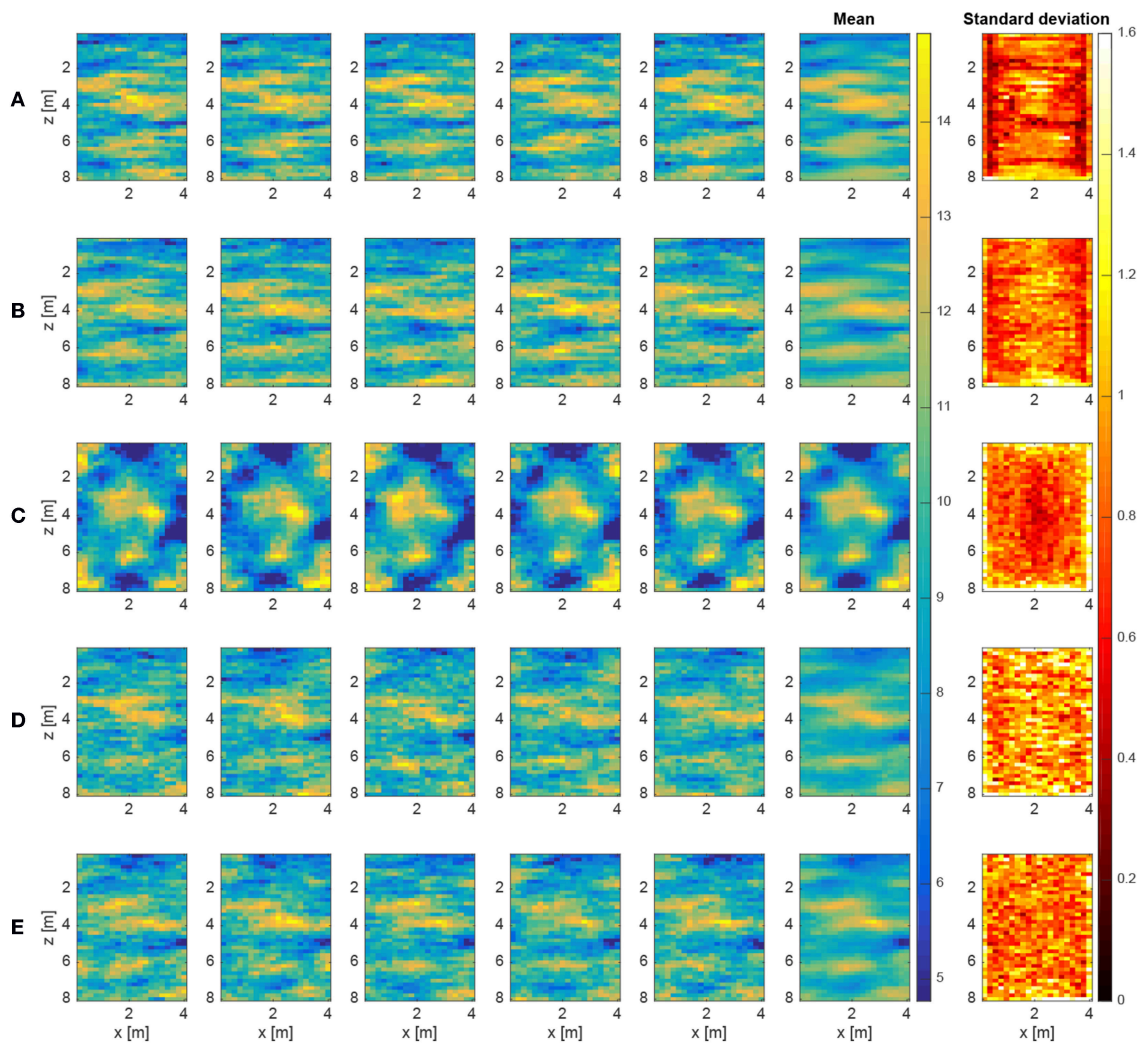
FIGURE 6 | Five randomly chosen posterior slowness realizations along with the point-wise mean and standard deviation [ns/m] obtained via **(A)** MCMC sampling based on the detailed eikonal solver (140 samples total); **(B)** standard ES-MDA based on the detailed eikonal solver with $n_e = 640$ (**Figure 2**; 6,400 samples total); **(C)** standard ES-MDA based on the proxy straight-ray solver with $n_e = 640$ (**Figure 3**; 6,400 samples total); **(D)** our modified ES-MDA algorithm with 20 KNN, $n_d = 20$, and $n_e = 160$ (**Figure 4**; 1,600 samples total); and **(E)** our modified ES-MDA procedure with 40 KNN, $n_d = 40$, and $n_e = 320$ (**Figure 5**; 3,200 samples total).

for, we see in **Figure 7B** that a good travel-time data match is achieved and no further improvement is observed after one iteration. However, **Figure 7E** shows that the corresponding slowness misfit is still large after one iteration compared to **Figure 7D**, which arises because of overfitting; that is, the inversion attempts to fit the model error. Applying our proposed method, we see in **Figure 7C** that the travel-time misfit is again primarily reduced in the first iteration and behaves similarly to the case where no model error is present (**Figure 7A**). More importantly, over-fitting is significantly reduced (**Figure 7F**) and the slowness misfit after only 4 iterations is similar to that seen in **Figure 7B**. This again confirms that employing our proposed approach can effectively remove proxy-related bias and allow the ES-MDA procedure to yield results that are comparable to inversions when no model error is present.

Although it may be possible to arrive at these results in less iterations than the 8 considered in this paper, it is difficult to know in advance how the combined approach of proxy solver and model-error correction behaves in terms of the internal ES-MDA iterations.

5. CONCLUSIONS

We have presented in this paper an approach that builds on the work of Köpke et al. (2018) in order to remove the bias associated with the use of proxy forward solvers in ES-MDA inversions. This allows for the consideration of larger ensemble sizes, which help to improve the accuracy of the parameter estimates obtained. Instead of constructing a local or global error
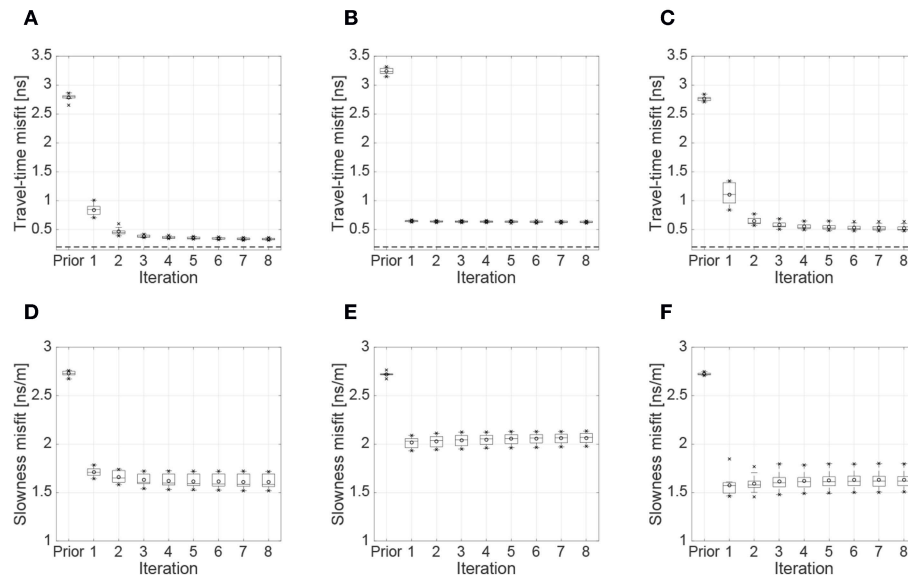
FIGURE 7 | Box plots showing the results of 10 ES-MDA runs using 160 ensemble members, as a function of the number of assimilation iterations, for the cases of **(A,D)** no model error; **(B,E)** model error present but not accounted for; and **(C,F)** model error present and accounted for using Algorithm 2 with 40 detailed solver calculations and 40 KNN. Added to the box plots are the mean (circles), and minimum and maximum values (crosses). The dashed horizontal line in **(A–C)** represents the expected value of the travel-time misfit assuming that the residuals follow the prescribed Gaussian distribution for the data errors.

model, our approach importantly aims to identify the model-error component of the residual during the ES-MDA procedure, which is used to correct the proxy forward response. This is accomplished through construction of an orthonormal model-error basis for each ensemble member and at each iteration based on a prescribed number of KNN entries selected from a model-error dictionary. The latter is created as the inversion proceeds, and thus no prior information about the model error is required before running the procedure.

With regard to the considered example problem of estimating the spatial distribution of subsurface slowness from crosshole GPR travel times, we saw that our modified ES-MDA approach allows us to obtain accurate posterior estimates characteristic of large ensembles with a computational cost comparable to a small number of runs of the detailed forward solver. The results did show, however, that the success of the approach depends on the ratio between the number of ensemble members and the number of detailed solver runs per iteration used to learn about the model error. In particular, for the crosshole GPR tomographic example considered, this ratio should not exceed a value of approximately 8.

Despite the successful application of our model-error approach, there remain a number of topics that should be investigated further. For example, in the work presented here, we set the number of KNN equal to the number of detailed solver runs per ES-MDA iteration used to learn about the model

error. In this case, the same model-error basis is constructed in the first iteration for each ensemble member. In subsequent iterations and with a growing dictionary, for each ensemble member the KNN are used to extract local information about the model error from the model-error dictionary and a local model-error basis is calculated, respectively. However, this choice could be validated and optimized to improve the identification of the model-error component of the residual. Further, we assume with our method that the latter is approximately orthogonal to the data-measurement and parameter-error components, which allows for its identification using a projection approach. Although we have found this assumption to yield acceptable results for a range of test problems examined so far, it requires further investigation.

AUTHOR CONTRIBUTIONS

CK: developing ideas, setting up the codes and writing the manuscript; AE: discussing ideas and revising the manuscript; JI: discussing ideas, revising the manuscript, and obtaining funding for the research.

FUNDING

REFERENCES

Aanonsen, S. I., Nævdal, G., Oliver, D. S., Reynolds, A. C. and Vallès, B. (2009). The ensemble Kalman filter in reservoir engineering–a review. *Spe J.* 14, 393–412. doi: 10.2118/117274-PA

Arridge, S., Kaipio, J., Kolehmainen, V., Schweiger, M., Somersalo, E., Tarvainen, T., et al. (2006). Approximation errors and model reduction with an application in optical diffusion tomography. *Inverse Prob.* 22:175. doi: 10.1088/0266-5611/22/1/010

Beskos, A., Girolami, M., Lan, S., Farrell, P. E., and Stuart, A. M. (2017). Geometric MCMC for infinite-dimensional inverse problems. *J. Comput. Phys.* 335, 327–351. doi: 10.1016/j.jcp.2016.12.041

Buizza, R., and Palmer, T. N. (1998). Impact of ensemble size on ensemble prediction. *Month. Weather Rev.* 126, 2503–2518. doi: 10.1175/1520-0493(1998)126<2503:IOESOE>2.0.CO;2

Calvetti, D., Ernst, O., and Somersalo, E. (2014). Dynamic updating of numerical model discrepancy using sequential sampling. *Inverse Prob.* 30:114019. doi: 10.1088/0266-5611/30/11/114019

Chen, Y., Keyes, D., Law, K. J., and Ltaief, H. (2016). Accelerated dimension-independent adaptive Metropolis. *SIAM J. Sci. Comput.* 38, S539–S565. doi: 10.1137/15M1026432

Chen, Y., and Zhang, D. (2006). Data assimilation for transient flow in geologic formations via ensemble Kalman filter. *Adv. Water Resour.* 29, 1107–1122. doi: 10.1016/j.advwatres.2005.09.007

Cordua, K. S., Looms, M. C., and Nielsen, L. (2008). Accounting for correlated data errors during inversion of cross-borehole ground penetrating radar data. *Vadose Zone J.* 7, 263–271. doi: 10.2136/vzj2007.0008

Cotter, S. L., Roberts, G. O., Stuart, A. M., and White, D. (2013). MCMC methods for functions: modifying old algorithms to make them faster. *Stat. Sci.* 28, 424–446. doi: 10.1214/13-STS421

Cui, T., Fox, C., and O'Sullivan, M. J. (2018). Adaptive approximation error models for efficient uncertainty quantification with application to multiphase subsurface fluid flow. *arXiv:1809.03176.*

Deutsch, C. V., and Journel, A. G. (1992). *GSLIB: Geostatistical Software Library and User's Guide.* New York, NY: Oxford University Press.

Elsheikh, A. H., Wheeler, M. F., and Hoteit, I. (2013). An iterative stochastic ensemble method for parameter estimation of subsurface flow models. *J. Comput. Phys.* 242, 696–714. doi: 10.1016/j.jcp.2013.01.047

Emerick, A. A., and Reynolds, A. C. (2012a). Ensemble smoother with multiple data assimilation. *Comput. Geosci.* 55, 3–15. doi: 10.1016/j.cageo.2012.03.011

Emerick, A. A., and Reynolds, A. C. (2012b). History matching time-lapse seismic data using the ensemble kalman filter with multiple data assimilations. *Comput. Geosci.* 16, 639–659. doi: 10.1007/s10596-012-9275-5

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* 99, 10143–10162. doi: 10.1029/94JC00572

Evensen, G. (2007). *Data Assimilation: The Ensemble Kalman Filter.* Heidelberg: Springer Science & Business Media.

Evin, G., Thyer, M., Kavetski, D., McInerney, D., and Kuczera, G. (2014). Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. *Water Resour. Res.* 50, 2350–2375. doi: 10.1002/2013WR014185

Giroux, B., Gloaguen, E., and Chouteau, M. (2007). bh_tomoa Matlab borehole georadar 2D tomography package. *Comput. Geosci.* 33, 126–137. doi: 10.1016/j.cageo.2006.05.014

Goh, J., Bingham, D., Holloway, J. P., Grosskopf, M. J., Kuranz, C. C., and Rutter, E. (2013). Prediction and computer model calibration using outputs from multifidelity simulators. *Technometrics* 55, 501–512. doi: 10.1080/00401706.2013.838910

Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli* 7, 223–242. doi: 10.2307/3318737

Hansen, T. M., Cordua, K. S., Jacobsen, B. H., and Mosegaard, K. (2014). Accounting for imperfect forward modeling in geophysical inverse problems - Exemplified for crosshole tomography. *Geophysics* 79, H1–H21. doi: 10.1190/geo2013-0215.1

Hansen, T. M., Cordua, K. S., Looms, M. C., and Mosegaard, K. (2013). Sippi: a Matlab toolbox for sampling the solution to inverse problems with complex prior information: part 2—application to crosshole GPR tomography. *Comput. Geosci.* 52, 481–492. doi: 10.1016/j.cageo.2012.10.001

Hansen, T. M., Cordua, K. S., and Mosegaard, K. (2012). Inverse problems with non-trivial priors: efficient solution through sequential Gibbs sampling. *Comput. Geosci.* 16, 593–611. doi: 10.1007/s10596-011-9271-1

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning, Data Mining, Inference and Prediction, 2nd Edn.* New York, NY: Springer Series in statistics.

Josset, L., Demyanov, V., Elsheikh, A. H., and Lunati, I. (2015a). Accelerating Monte Carlo Markov chains with proxy and error models. *Comput. Geosci.* 85, 38–48. doi: 10.1016/j.cageo.2015.07.003

Josset, L., Ginsbourger, D., and Lunati, I. (2015b). Functional error modeling for uncertainty quantification in hydrogeology. *Water Resour. Res.* 51, 1050–1068. doi: 10.1002/2014WR016028

Kaipio, J., and Somersalo, E. (2007). Statistical inverse problems: discretization, model reduction and inverse crimes. *J. Comput. Appl. Math.* 198, 493–504. doi: 10.1016/j.cam.2005.09.027

Kennedy, M. C., and O'Hagan, A. (2001). Bayesian calibration of computer models. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* 63, 425–464. doi: 10.1111/1467-9868.00294

Khu, S.-T., and Werner, M. G. (2003). Reduction of Monte-Carlo simulation runs for uncertainty estimation in hydrological modelling. *Hydrol. Earth Syst. Sci.* 7, 680–692. doi: 10.5194/hess-7-680-2003

Köpke, C., Irving, J., and Elsheikh, A. H. (2018). Accounting for model error in Bayesian solutions to hydrogeophysical inverse problems using a local basis approach. *Adv. Water Resour.* 116, 195–207. doi: 10.1016/j.advwatres.2017.11.013

Laloy, E., Rogiers, B., Vrugt, J. A., Mallants, D., and Jacques, D. (2013). Efficient posterior exploration of a high-dimensional groundwater model from two-stage Markov chain Monte Carlo simulation and polynomial chaos expansion. *Water Resour. Res.* 49, 2664–2682. doi: 10.1002/wrcr.20226

Lehikoinen, A., Huttunen, J. M. J., Finsterle, S., Kowalsky, M. B., and Kaipio, J. P. (2010). Dynamic inversion forhydrological process monitoring with electrical resistance tomography under model uncertainties. *Water Resour. Res.* 46:W04513. doi: 10.1029/2009WR008470

Linde, N., and Vrugt, J. A. (2013). Distributed soil moisture from crosshole ground-penetrating radar travel times using stochastic inversion. *Vadose Zone J.* 12:vzj2012.0101. doi: 10.2136/vzj2012.0101

Looms, M. C., Binley, A., Jensen, K. H., and Nielsen, L. (2008). Identifying unsaturated hydraulic parameters using an integrated data fusion approach on cross-borehole geophysical data. *Vadose Zone J.* 7, 238–248. doi: 10.2136/vzj2007.0087

Marzouk, Y., and Xiu, D. (2009). A stochastic collocation approach to Bayesian inference in inverse problems. *Commun. Comput. Phys.* 6, 826–847. doi: 10.4208/cicp.2009.v6.p826

Nowack, R. L. (1992). Wavefronts and solutions of the eikonal equation. *Geophys. J. Int.* 110, 55–62. doi: 10.1111/j.1365-246X.1992.tb00712.x

Oliver, D. S., and Alfonzo, M. (2018). Calibration of imperfect models to biased observations. *Comput. Geosci.* 22, 145–161. doi: 10.1007/s10596-017-9678-4

Piccolo, C., and Cullen, M. (2016). Ensemble data assimilation using a unified representation of model error. *Month. Weather Rev.* 144, 213–224. doi: 10.1175/MWR-D-15-0270.1

Rasmussen, C. E., and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning.* Cambridge: MIT Press.

Reynolds, A. C., Zafari, M., and Li, G. (2006). "Iterative forms of the ensemble kalman filter," in *ECMOR X-10th European Conference on the Mathematics of Oil Recovery*, A030. Amsterdam.

Robert, C. P., and Casella, G. (2004). *Monte Carlo Statistical Methods.* New York, NY: Springer.

Ruggeri, P., Irving, J., and Holliger, K. (2015). Systematic evaluation of sequential geostatistical resampling within MCMC for posterior sampling of near-surface geophysical inverse problems. *Geophys. J. Int.* 202, 961–975. doi: 10.1093/gji/ggv196

Scholer, M., Irving, J., Looms, M. C., Nielsen, L., and Holliger, K. (2012). Bayesian Markov-chain-monte-carlo inversion of time-lapse crosshole GPR data to characterize the vadose zone at the arrenaes site, Denmark. *Vadose Zone J.* 11. doi: 10.2136/vzj2011.0153

Schoups, G., and Vrugt, J. A. (2010). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resour. Res.* 46:W10531. doi: 10.1029/2009WR008933

Smith, T., Marshall, L., and Sharma, A. (2015). Modeling residual hydrologic errors with Bayesian inference. *J. Hydrol.* 528, 29–37. doi: 10.1016/j.jhydrol.2015.05.051

Stordal, A. S., and Elsheikh, A. H. (2015). Iterative ensemble smoothers in the annealed importance sampling framework. *Adv. Water Resour.* 86, 231–239. doi: 10.1016/j.advwatres.2015.09.030

Strang, G., Strang, G., Strang, G., and Strang, G. (1993). *Introduction to Linear Algebra*, Vol. 3. Wellesley, MA: Wellesley-Cambridge Press.

Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia, PA: SIAM.

van Leeuwen, P. J. (2001). An ensemble smoother with error estimates. *Month. Weather Rev.* 129, 709–728. doi: 10.1175/1520-0493(2001)129<0709:AESWEE>2.0.CO;2

Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: theory, concepts, and MATLAB implementation. *Environ. Modell. Softw.* 75, 273–316. doi: 10.1016/j.envsoft.2015.08.013

Xu, T., Valocchi, A. J., Choi, J., and Amir, E. (2014). Use of machine learning methods to reduce predictive error of groundwater models. *Groundwater* 52, 448–460. doi: 10.1111/gwat.12061