ORIGINAL PAPER

A multiscale method for data assimilation

Rafael J. de Moraes[1,2] · Hadi Hajibeygi[1] · Jan Dirk Jansen[1]

© The Author(s) 2019

Abstract

In data assimilation problems, various types of data are naturally linked to different spatial resolutions (e.g., seismic and electromagnetic data), and these scales are usually not coincident to the subsurface simulation model scale. Alternatives like upscaling/downscaling of the data and/or the simulation model can be used, but with potential loss of important information. Such alternatives introduce additional uncertainties which are not in the nature of the problem description, but the result of the post processing of the data or the geo-model. To address this issue, a novel multiscale (MS) data assimilation method is introduced. The overall idea of the method is to keep uncertain parameters and observed data at their original representation scale, avoiding upscaling/downscaling of any quantity. The method relies on a recently developed mathematical framework to compute adjoint gradients via a MS strategy in an algebraic framework. The fine-scale uncertain parameters are directly updated and the MS grid is constructed in a resolution that meets the observed data resolution. This formulation therefore enables a consistent assimilation of data represented at a coarser scale than the simulation model. The misfit objective function is constructed to keep the MS nature of the problem. The regularization term is represented at the simulation model (fine) scale, whereas the data misfit term is represented at the observed data (coarse) scale. The computational aspects of the method are investigated in a simple synthetic model, including an elaborate uncertainty quantification step, and compared to upscaling/downscaling strategies. The experiment shows that the MS strategy provides several potential advantages compared to more traditional scale conciliation strategies: (1) expensive operations are only performed at the coarse scale; (2) the matched uncertain parameter distribution is closer to the "truth"; (3) faster convergence behavior occurs due to faster gradient computation; and (4) better uncertainty quantification results are obtained. The proof-of-concept example considered in this paper sheds new lights on how one can reduce uncertainty within fine-scale geo-model parameters with coarse-scale data, without the necessity of upscaling/downscaling the data nor the geo-model. The developments demonstrate how to consistently formulate such a gradient-based MS data assimilation strategy in an algebraic framework which allows for implementation in available computational platforms.

Keywords Data assimilation · Multiscale inversion · Gradient-based optimization · Adjoint method · Spatial observations · Uncertainty quantification

1 Introduction

Subsurface simulation models should be conditioned to field data, whenever possible, in order to reduce uncertainty in the model parameters and hence increase forecasting reliability. Well production data (time series of fluid rates and pressures), seismic surveys, and well testing pressure data are some instances of field data that can be assimilated in

order to better estimate the uncertain parameters. In addition, spatially distributed observations (e.g., seismic, electromagnetics) provide valuable information that can considerably improve the assimilation process [10]. For instance, over the past decades, an increasing number of seismic monitoring cases has been observed [13, 27, 48, 52, 53, 58]. One of the main advantages of time-lapse seismic data is its ability to approximate the pressure/fluid distribution inside the reservoir. This may considerably help to gain insight about the subsurface fluid displacement process. Moreover, it may help to characterize the formation, either via improved static geological modeling, or via dynamic assimilation (inverse modeling). One hurdle in the process of assimilating spatially distributed information is the fact that, more often than

✉ Rafael J. de Moraes
r.moraes@tudelft.nl; rafael_moraes@petrobras.com.br

Extended author information available on the last page of the article.

not, the observed data and the forward model are described at different spatial scales. In fact, it is an open question at which scale data should best be assimilated: the simulation model scale or the observed data scale [10].

The relevancy of addressing the multiscale nature of inverse problems is observed in the recent literature on the topic. A collection of articles about multiscale forward modeling strategies and multiscale challenges associated to inverse problems can be found in [25]. In [36], the authors propose a multiscale data assimilation scheme based on the decomposition of the objective function so that the error covariance can be estimated for distinct spatial scales. A multiscale parameter field discretization designed to reduce the dimensionality of the inverse problem via an adaptive grid refinement is presented in [55]. The impact of the scale dissimilarity in terms of observation information content and the parameter space size on the ensemble collapse in ensemble-based methods [1] has been addressed by [16, 17, 40] via upscaling/homogenization techniques [7, 15]. The authors also benefit from coarse-scale simulations to improve the inverse problem computational efficiency. In [18], a multiscale method is proposed which accounts for microscale features by assuming that they can be represented by a low-dimensional parameterization. Nonetheless, the aforementioned works are based on the assumption that the fine-scale uncertain parameters can be homogenized and represented at a coarser scale.

On the other hand, data assimilation strategies based on multiscale (MS) simulation [28, 32] have also been developed. MS methods are efficient simulation strategies capable of solving the flow problem at a coarser grid, while being capable of accurately representing fine-scale heterogeneities. An adjoint-based multiscale finite volume method for computation of sensitivities has been presented in [20] and later extended to time-dependent [19] single-phase flow in porous media. More recently, a general framework for the computation of multiscale gradients has been introduced in [42], with an extension to multiphase flows [41]. The latter two are based on a general framework for derivative computation, whose algebraic nature does not rely on any assumption regarding the nature of the parameters, observations, or objective function type. Also, in [3], a multiscale inversion technique is presented based on the Markov-chain Monte Carlo method that also relies on the generalized multiscale finite element method [8].

Despite this body of work found in the inverse modeling literature, when one is interested in assimilating spatially distributed data, there is an implicit assumption that the observed data is described at the same scale of the parameters is usually made. Actually, assuming one is not interested in changing the scale of the model parameter description, some treatment must be employed in the change of the observed data or the forward model response

scale. The literature indicates that upscaling/downscaling of the observed data to the forward model scale, as a pre-processing step with respect to the data assimilation process, is the most employed strategy in practice [23, 24, 35]. In the present work, we are particularly interested in addressing the spatial scale dissimilarity between observations and the discretized forward model. In many applications, there is no observability of the spatially distributed data (due to limitations in the acquisition process; e.g., in terms of resolution) at the parameter resolution which is necessary to accurately describe important physical phenomena. Therefore, here we consider observed data that is acquired at a resolution that is coarser than the parameter resolution, and hence represented in a grid that is coarser than the one employed in the forward simulation. Note that sub-grid observations, at a resolution lower than the forward model discretization, is outside the scope of this work.

Another important aspect to be considered in data assimilation and uncertainty quantification (UQ) studies is the fact that those rely on computationally demanding algorithms. Different techniques such as Monte Carlo (MC) methods [37], Ensemble Kalman Filter (EnKF) and derivations [1, 11, 14], and randomized maximum likelihood (RML) [45] are developed to perform those studies. A comparison between the different techniques is provided by [12]. Regardless of the technique, a common feature they share is the necessity of performing many forward model runs in order to reasonably sample the posterior probability distribution of the reservoir uncertain parameters. As already mentioned, upscaling [7, 15] can build faster, reasonably accurate forward models that can speed up the sampling process. However, to accurately represent some physical phenomena, e.g., mixing, diffusion, fluid fronts, or compositional capillary effects, fine-scale resolution is of utmost importance. Hence, the ability of keeping the high fidelity description of geological parameters is fundamental for an adequate reservoir characterization. Partial-differential-equation-(PDE)-constrained optimization techniques can be employed in the solution of the inverse problem. In this case, it is well known that gradient-based algorithms are the most efficient ones, mainly if combined with efficient gradient computation. And it is also well known that gradients obtained with the adjoint method [2, 31, 33, 47] are the most efficient and accurate ones.

The objective of this work is to develop and demonstrate an inverse modelling method that, at the same time, (1) is computationally efficient, (2) addresses the scale dissimilarity issue, with minimum loss of information, and (3) is capable of updating the highest fidelity model description. To this end, we exploit multiscale (MS) simulation strategies in order to (1) speed up the forward simulation, while preserving fine-scale geological features, (2) efficiently compute gradient information, and (3) seamlessly

conciliate model and observed data scales. For a comprehensive review on the recent developments associated with MS methods applied to reservoir simulation, see [36].

The remainder of this paper is organized as follows. Firstly, a brief overview about how data assimilation is approached from a Bayesian perspective is presented. Next, we state our target forward model, consisting of incompressible single-phase flow in heterogeneous porous media. Also, we revisit the MS solution of the flow equation in a purely algebraic presentation. Thereafter, we discuss the data assimilation problem setup, focusing on the challenges of assimilating spatially distributed data. More specifically, we discuss alternatives on how to conciliate data and model scales. Then, we introduce our multiscale data assimilation strategy, consisting of, basically, a MS objective function and a MS gradient computation strategy. The method here employed is largely based on the MS gradient computation strategy discussed by [42]. We compare the different data conciliation methods and our newly introduced method based on a synthetic 2D case. We focus our experiments on both the maximum a-posteriori (MAP) estimate and UQ via the RML method. A discussion about the results and the challenges that the method can encounter is presented next. Finally, a summary of the developments and results, as well as future research perspectives, is presented.

2 Preliminaries

2.1 Problem statement

Let N_d denote the number of space dimensions. Let $\Omega \subset \mathbb{R}^{N_d}$ be the problem domain with boundary $\partial\Omega$. Let $\mathbf{s} \in \mathbb{R}^{N_d}$ be an arbitrary space position. Let $\mathbf{n} \in \mathbb{R}^{N_d}$ be an unit normal vector to $\partial\Omega$. Our analysis focuses on phenomena governed by an elliptic PDE equation, denoted by g, in the form

$$(g) \begin{cases} \nabla \cdot (\theta(\mathbf{s}) \nabla x(\mathbf{s})) = q(\mathbf{s}), & \mathbf{s} \in \Omega \\ \nabla x(\mathbf{s}) \mathbf{n} = \Gamma, & \mathbf{s} \in \partial\Omega_\Gamma \\ x(\mathbf{s}) = \Upsilon, & \mathbf{s} \in \partial\Omega_\Upsilon \end{cases} \quad (1)$$

where $\partial\Omega = \partial\Omega_\Upsilon \cup \partial\Omega_\Gamma$, $\partial\Omega_\Upsilon \cap \partial\Omega_\Gamma = \emptyset$, $x = x(\theta)$ is the variable of interest, $q = q(\theta)$ is the sink/source term, and θ is the heterogeneous uncertain coefficient which we aim to estimate via the assimilation of real system observations. We assume θ has no separation of scales; hence, homogenization techniques would lead to unavoidable approximations to the effective property.

Let

$$y = h(x, \theta) \quad (2)$$

be the observable model responses, then the inverse problem

$$d = h(x, \theta) + \epsilon \quad (3)$$

can provide an estimate for θ given the description of the real observation errors ϵ. We assume that d, the real system data, can only be observed at a resolution that is coarser or equal to the resolution at which θ is described.

2.2 Inverse problem as a PDE-constrained optimization

We base our developments on a Bayesian framework. Let N_y be number of observable responses, N_θ be the number of model parameters, and N_x the number of primary (state) variables. According to Bayes' theorem, the posterior probability distribution function (PDF) can be computed as

$$f(\theta|\mathbf{d}) = \frac{f(\mathbf{d}|\theta) f(\theta)}{f(\mathbf{d})}, \quad (4)$$

where $\theta \in \mathbb{R}^{N_\theta}$ is the vector of model parameters and $\mathbf{d} \in \mathbb{R}^{N_y}$ is the vector observable responses. If the *a priori* PDF of the uncertain parameters, $f(\theta)$, and the measurement errors from the observations are assumed Gaussian, it can be shown that the conditional a posteriori distribution is given by [50]

$$f(\theta|\mathbf{d}) \propto \exp(-O(\mathbf{y}(\mathbf{x}, \theta), \theta)), \quad (5)$$

where the objective function $O \in \mathbb{R}$ is given by

$$O(\mathbf{y}(\mathbf{x}, \theta), \theta) = \frac{1}{2}(\theta - \theta_{prior})^T \mathbf{C}_\theta^{-1}(\theta - \theta_{prior})$$
$$+ \frac{1}{2}(\mathbf{y}(\mathbf{x}, \theta) - \mathbf{d})^T \mathbf{C}_D^{-1}(\mathbf{y}(\mathbf{x}, \theta) - \mathbf{d}). \quad (6)$$

In the above equations, $\mathbf{y} \in \mathbb{R}^{N_y}$ is the vector of model responses (outputs), $\mathbf{x} \in \mathbb{R}^{N_x}$ is the state vector, $\theta_{prior} \in \mathbb{R}^{N_\theta}$ is the prior mean, $\mathbf{C}_\theta \in \mathbb{R}^{N_\theta \times N_\theta}$ is the parameter covariance matrix, and $\mathbf{C}_D \in \mathbb{R}^{N_y \times N_y}$ is the covariance matrix of the measurement errors.

The solution of Eq. 6 can be stated as a PDE-constrained optimization problem as [46]

$$\underset{\theta}{\text{minimize}} \quad O(\mathbf{y}(\mathbf{x}, \theta), \theta)$$
$$\text{subject to} \quad \mathbf{g}(\mathbf{x}, \theta) = \mathbf{0}$$
$$\theta \in [\theta_{min}, \theta_{max}], \quad (7)$$

where $\mathbf{g} : \mathbb{R}^{N_x} \times \mathbb{R}^{N_\theta} \rightarrow \mathbb{R}^{N_x}$ represents the set of forward model equations and $\theta_{min} \in \mathbb{R}^{N_\theta}, \theta_{max} \in \mathbb{R}^{N_\theta}$ are, respectively, the parameter lower and upper bound vectors. The efficient solution of Eq. 7, resulting from the discretization of Eq. 1, requires gradient-based methods [44] combined with efficient gradient computation methods.

For this purpose, by applying the chain rule to Eq. 6 it follows that

$$
\begin{aligned}
\nabla_\theta O\left(\theta, \mathbf{y}\left(\mathbf{x}, \theta\right)\right) &= \left(\frac{dO}{d\theta}\right)^T \\
&= \left(\frac{\partial O}{\partial \theta}\right)^T + \mathbf{G}^T \nabla_\mathbf{y} O,
\end{aligned}
\tag{8}
$$

where $\mathbf{G} \in \mathbb{R}^{N_y \times N_\theta}$ is the so-called sensitivity matrix, representing the sensitivity of the responses w.r.t. the parameters. Efficient gradient methods are analytical, and more specifically in inverse problems where the number of parameters is greater than the number of output functionals, the adjoint method is the most accurate, efficient method [46]. The efficient computation of the right multiplication of $\mathbf{G}$ by an arbitrary vector (as in Eq. 8) via the adjoint method is discussed in [47].

2.3 Randomized Maximum Likelihood (RML)

RML [45] is an approximated sampling method for UQ, which obtains the jth sample of the posterior PDF distribution by solving Eq. 7 for a given sample $\theta_{uc,j}$ from a normal distribution $\mathcal{N}\left(\theta_{prior}, \mathbf{C}_\theta\right)$ and a given sample $\mathbf{d}_{uc,j}$ from $\mathcal{N}\left(\mathbf{d}, \mathbf{C}_D\right)$. Therefore, Eq. 6 can be re-written for $O_j\left(\mathbf{y}\left(\mathbf{x}, \theta\right), \theta\right)$ as

$$
\begin{aligned}
O_j\left(\mathbf{y}\left(\mathbf{x}, \theta\right), \theta\right) = &\frac{1}{2}\left(\theta - \theta_{uc,j}\right)^T \mathbf{C}_\theta^{-1}\left(\theta - \theta_{uc,j}\right) \\
&+ \frac{1}{2}\left(\mathbf{y}\left(\mathbf{x}, \theta\right) - \mathbf{d}_{uc,j}\right)^T \mathbf{C}_D^{-1}\left(\mathbf{y}\left(\mathbf{x}, \theta\right) - \mathbf{d}_{uc,j}\right).
\end{aligned}
\tag{9}
$$

Hence, a minimization problem has to be solved for every jth posterior PDF sample one wants to estimate. This is only feasible with efficient gradient computation methods, as described in the previous section.

3 The forward model

The set of discretized equations that describes the forward simulation at the fine scale can be algebraically expressed as [42]

$$
\mathbf{g}_F\left(\mathbf{x}, \theta\right) = \mathbf{0},
\tag{10}
$$

where $\mathbf{g}_F : \mathbb{R}^{N_F} \times \mathbb{R}^{N_\theta} \rightarrow \mathbb{R}^{N_F}$ represents the set of algebraic forward model equations resulting from the numerical discretization of Eq. 1 over a fine grid $\mathbb{G}_F \in \mathbb{R}^{N_F}$, $\mathbf{x} \in \mathbb{R}^{N_F}$ is the state vector and the subscript F refers to "fine scale." There are N_F fine-scale cells. Equation 10 implicitly assumes a dependency of the state vector $\mathbf{x}$ on the parameters θ, i.e.,

$$
\mathbf{x} = \mathbf{x}\left(\theta\right).
\tag{11}
$$

Once the model state is determined, the observable responses of the forward model are computed. The forward model responses may not only depend on the model state, but also on the parameters themselves, and can be expressed as

$$
\mathbf{y}_F = \mathbf{h}_F\left(\mathbf{x}, \theta\right),
\tag{12}
$$

where $\mathbf{h}_F : \mathbb{R}^{N_F} \times \mathbb{R}^{N_\theta} \rightarrow \mathbb{R}^{N_y}$ represents the output equations [30]. It is assumed that $\mathbf{g}_F$ can be described as

$$
\mathbf{g}_F\left(\mathbf{x}, \theta\right) = \mathbf{A}\left(\theta\right)\mathbf{x} - \mathbf{q}\left(\theta\right),
\tag{13}
$$

where $\mathbf{A} = \mathbf{A}\left(\theta\right) \in \mathbb{R}^{N_F \times N_F}$ represents the elliptic discrete operator and $\mathbf{q} = \mathbf{q}\left(\theta\right) \in \mathbb{R}^{N_F}$ is a vector of source terms and boundary conditions.

3.1 Multiscale simulation

A multiscale (MS) solution strategy can be algebraically devised [56, 59] by firstly computing a coarse-scale solution

$$
\breve{\mathbf{g}}\left(\breve{\mathbf{x}}, \theta\right) = \left(\mathbf{RAP}\right)\breve{\mathbf{x}} - \left(\mathbf{Rq}\right) = \breve{\mathbf{A}}\breve{\mathbf{x}} - \breve{\mathbf{q}} = \breve{\mathbf{0}},
\tag{14}
$$

where after an approximate fine-scale solution is formed as

$$
\mathbf{g}'\left(\mathbf{x}', \breve{\mathbf{x}}, \theta\right) = \mathbf{x}' - \mathbf{P}\breve{\mathbf{x}} = \mathbf{0}.
\tag{15}
$$

Let $\breve{\mathbf{x}} \in \mathbb{R}^{N_C}$ be the coarse-scale solution ($N_C \ll N_F$), and $\mathbf{x}' \in \mathbb{R}^{N_F}$ the approximated fine-scale solution. The prolongation operator $\mathbf{P} = \mathbf{P}\left(\theta\right)$ is an $N_F \times N_C$ matrix that maps (interpolates) the coarse-scale solution to the fine scale, where N_C is the number of coarse grid blocks. The restriction operator $\mathbf{R} = \mathbf{R}\left(\theta\right)$ is defined as an $N_C \times N_F$ matrix which maps the fine scale to the coarse scale.

In multiscale methods, the scaling operators are constructed based on locally supported basis functions. Different strategies to build MS basis functions are available in the literature [8, 28, 32, 43]. In this work, we employ the multiscale finite volume (MSFV) method [32]. However, we emphasize, as will be clear from the formulation, that the framework allows the employment of different MS methods, as long as they can be expressed in terms of $\mathbf{R}$ and $\mathbf{P}$. Next, we discuss the MSFV basis function construction.

3.1.1 Construction of scaling operators via the MSFV method

The MSFV discretization relies on two overlapping coarse grids, namely the primal and dual coarse grids, which are superimposed on a given fine grid. The grids are illustrated in Fig. 1. The primal coarse grid contains N_C control volumes $\breve{\Omega}_i$, $i \in \{1, \ldots, N_C\}$, and the dual coarse grid contains N_D local domains $\tilde{\Omega}_j$, $j \in \{1, \ldots, N_D\}$.

The MSFV basis functions are constructed based on local solutions of the elliptic governing Eq. 1 for every
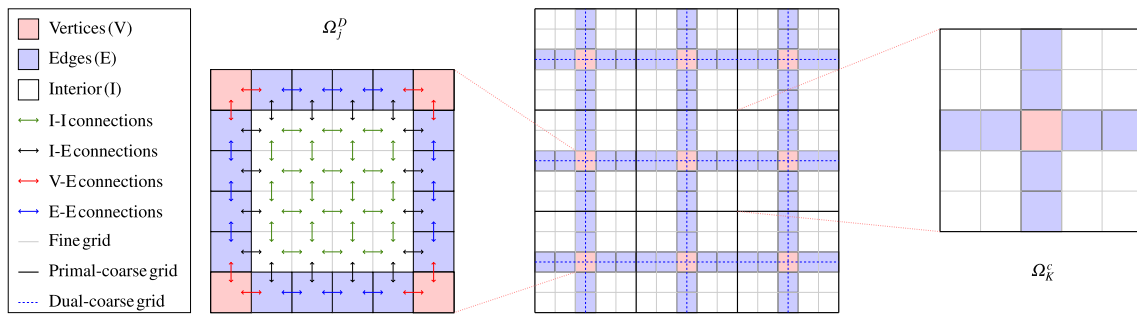
Fig. 1 Multiscale finite volume grids and illustration of interfacial connections between cells used in the wirebasket ordering

$\tilde{\Omega}_j$, with no right-hand-side and subject to special boundary conditions [32, 56]

$$\begin{cases} \nabla \cdot \left(\theta \left(\mathbf{s} \right) \nabla \varphi_j^i \left(\mathbf{s} \right) \right) = 0, & \mathbf{s} \in \tilde{\Omega}_j \\ \nabla_{\|} \cdot \left(\theta \left(\mathbf{s} \right) \nabla \varphi_j^i \left(\mathbf{s} \right) \right)_{\|} = 0, & \mathbf{s} \in \partial \tilde{\Omega}_j \\ \varphi_j^i \left(\mathbf{s}_k \right) = \delta_{ik}, & \forall \mathbf{s}_k \in \{1, ..., N_C\}, \end{cases} \quad (16)$$

where φ_j^i is the basis function associated with the vertex i in $\tilde{\Omega}_j$, the subscript $\|$ represents the projection along $\partial \tilde{\Omega}_j$, and δ_{ik} is the Kronecker delta, and $k \in \{1, \ldots, 2^{N_d}\}$ denotes the vertices in $\tilde{\Omega}_j$ (N_d is the spatial dimensionality of the problem – 1, 2, or 3).

Assuming a finite volume discretization of Eq. 1, the basis functions, and hence the prolongation operator, can be constructed directly from the given fine-grid linear system matrix as [60]

$$\mathbf{P} = \mathscr{P} \begin{bmatrix} \mathbf{A}_{II}^{-1} \mathbf{A}_{IE} \tilde{\mathbf{A}}_{EE}^{-1} \mathbf{A}_{EV} \\ -\tilde{\mathbf{A}}_{EE}^{-1} \mathbf{A}_{EV} \\ -\mathbf{I}_{VV} \end{bmatrix}, \quad (17)$$

after $\mathbf{A}$ in Eq. 13 is re-ordered in a wirebasket ordering [54] as

$$\begin{aligned} \mathbf{g} &= \mathscr{P} \begin{bmatrix} \mathbf{g}_I \\ \mathbf{g}_E \\ \mathbf{g}_V \end{bmatrix} \\ &= \left(\mathscr{P} \begin{bmatrix} \mathbf{A}_{II} & \mathbf{A}_{IE} & \mathbf{0} \\ \mathbf{A}_{EI} & \mathbf{A}_{EE} & \mathbf{A}_{EV} \\ \mathbf{0} & \mathbf{A}_{VE} & \mathbf{A}_{VV} \end{bmatrix} \mathscr{P}^T \right) \mathscr{P} \begin{bmatrix} \mathbf{x}_I \\ \mathbf{x}_E \\ \mathbf{x}_V \end{bmatrix} - \mathscr{P} \begin{bmatrix} \mathbf{q}_I \\ \mathbf{q}_E \\ \mathbf{q}_V \end{bmatrix}, \end{aligned}$$

(18)

where $\mathscr{P} \in \mathbb{R}^{N_F \times N_F}$ is a permutation matrix that re-orders from wirebasket to natural ordering, $\mathbf{A}_{II} \in \mathbb{R}^{N_I \times N_I}$, $\mathbf{A}_{IE} \in \mathbb{R}^{N_I \times N_E}$, $\mathbf{A}_{EI} \in \mathbb{R}^{N_E \times N_I}$, $\mathbf{A}_{EE} \in \mathbb{R}^{N_E \times N_E}$, $\mathbf{A}_{EV} \in \mathbb{R}^{N_E \times N_V}$, $\mathbf{A}_{VE} \in \mathbb{R}^{N_V \times N_E}$, $\mathbf{A}_{VV} \in \mathbb{R}^{N_V \times N_V}$ are, respectively, the sub-matrices of $\mathbf{A}$ corresponding to the interior-interior, interior-edge, edge-interior, edge-edge, edge-vertex, vertex-edge, and vertex-vertex cell connections and N_I, N_E, and N_V are, respectively, the total number of interior, edge, and vertex cells in the fine grid. The interfacial connections are illustrated in Fig. 1. Also, $\mathbf{g}_I \in \mathbb{R}^{N_I}$, $\mathbf{g}_E \in \mathbb{R}^{N_E}$, $\mathbf{g}_V \in \mathbb{R}^{N_V}$, $\mathbf{x}_I \in \mathbb{R}^{N_I}$, $\mathbf{x}_E \in \mathbb{R}^{N_E}$, $\mathbf{x}_V \in \mathbb{R}^{N_V}$, and $\mathbf{q}_I \in \mathbb{R}^{N_I}$, $\mathbf{q}_E \in \mathbb{R}^{N_E}$, $\mathbf{q}_V \in \mathbb{R}^{N_V}$ are,

respectively, the model equations, and the state and source term sub-vectors corresponding to the interior, edge, and vertex cells.

Note that the construction of $\mathbf{P}$ requires setting $\mathbf{A}_{VE} = \mathbf{0}$, $\mathbf{A}_{EI} = \mathbf{0}$, and $\mathbf{A}_{VV} = \mathbf{I}_V V$ and likewise the corresponding entries $\mathbf{A}_{EE}$, resulting in $\tilde{\mathbf{A}}_{EE}$, which is equivalent to the localization assumptions required to build the basis functions as stated in Eq. 16 [60].

If the FV method is used in the fine-scale system discretization, the restriction operator can be defined as the sum of the equations of all the fine cells contained in the coarse cell, i.e., [60]

$$\mathbf{R}_{i,j} = \begin{cases} 1, \text{if} & \Omega_i \subset \breve{\Omega}_c \\ 0, \text{otherwise.} \end{cases} (c=1, \ldots, N_C; f=1, \ldots, N_F), \quad (19)$$

hence, in combination with Eq. 17, establishing the multiscale finite volume (MSFV) method. Also, a Galerkin restriction operator could be used by making

$$\mathbf{R} = \mathbf{P}^T,$$

and hence, in combination with Eq. 17, establishing the MS finite element (MSFE) method. While the MSFV is conservative by construction, the MSFE provides monotone solutions.

3.1.2 Computational efficiency

An analytical computational efficiency estimate of the MSFV method for the solution of one phase flow is discussed in [32] and briefly revisited here. For CPU studies based on the MSFV pressure equation solution of 3D heterogeneous domains, see [56]. Let N_R be the coarsening ratio employed to construct the MS coarse grid, N_L be the number of local problems that must be solved per coarse grid vertex (4 in 2D and 8 in 3D problems), and b a constant associated specifically to the linear solver employed. Assuming that the solver employed to the MS system is equally efficient to the one employed to solve the fine-scale system, it can be shown, via an asymptotic

analysis of the operations, that the cost ratio between MS and FS elliptic PDE solutions is expressed as [42]

$$\frac{\mathscr{O}_{MS}}{\mathscr{O}_{FS}} = \frac{1}{N_R^b} + N_L \frac{N_R^{b-1}}{N_F^{b-1}}. \tag{20}$$

In order to illustrate an estimate of the MS simulation efficiency potential, let us assume $b = 1.3$, $N_R = 10$, $N_L = 4$, and $N_F = 10^8$. That gives $\frac{\mathscr{O}_{MS}}{\mathscr{O}_{FS}} = 0.082$, hence the MS solution is proportional to only 8% the cost of a fine-scale solution in this scenario.

4 Data assimilation problem setup

4.1 Adjoint gradient computation

The maximum a posteriori probability (MAP, [46]) of the uncertain parameters is obtained by solving the optimization problem stated in Eq. 7, with the objective function (OF) given by Eq. 6 (and using Eq. 12), the gradient which is given by

$$\nabla_\theta O = \mathbf{C}_\theta^{-1} \left(\theta - \theta_{prior} \right) + \left(\frac{d\mathbf{h}}{d\theta} \right)^T \mathbf{C}_D^{-1} \left(\mathbf{h}\left(\mathbf{x}, \theta\right) - \mathbf{d} \right)$$

$$= \mathbf{p} + \mathbf{G}^T \mathbf{m}, \tag{21}$$

where

$$\mathbf{p} = \mathbf{C}_\theta^{-1} \left(\theta - \theta_{prior} \right), \tag{22}$$

and

$$\mathbf{m} = \mathbf{C}_D^{-1} \left(\mathbf{h}\left(\mathbf{x}, \theta\right) - \mathbf{d} \right), \tag{23}$$

$\mathbf{m} \in \mathbb{R}^{N_Y}$. Following an implicit differentiation strategy [33, 47], the sensitivity matrix $\mathbf{G}$ can be obtained from the total derivative of Eq. 12 with respect to θ as follows [42]:

$$\mathbf{G} = \frac{d\mathbf{h}}{d\theta} = \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \frac{d\mathbf{x}}{d\theta} + \frac{\partial \mathbf{h}}{\partial \theta} = -\frac{\partial \mathbf{h}}{\partial \mathbf{x}} \left(\frac{\partial \mathbf{g}}{\partial \mathbf{x}} \right)^{-1} \frac{\partial \mathbf{g}}{\partial \theta} + \frac{\partial \mathbf{h}}{\partial \theta}, \tag{24}$$

where $\frac{\partial \mathbf{g}}{\partial \mathbf{x}} \in \mathbb{R}^{N_F} \times \mathbb{R}^{N_F}$, $\frac{\partial \mathbf{g}}{\partial \theta} \in \mathbb{R}^{N_F} \times \mathbb{R}^{N_\theta}$, $\frac{\partial \mathbf{h}}{\partial \mathbf{x}} \in \mathbb{R}^{N_y} \times \mathbb{R}^{N_F}$, $\frac{\partial \mathbf{h}}{\partial \theta} \in \mathbb{R}^{N_y} \times \mathbb{R}^{N_\theta}$ are, respectively, partial derivative matrices obtained from the derivation of Eqs. 10 and 12 with respect to $\mathbf{x}$ and θ.

The product $\mathbf{G}^T \mathbf{m} = \left(\mathbf{m}^T \mathbf{G} \right)^T$ can be solved at costs proportional to one backward simulation, regardless of the number of model parameters, via the adjoint method, by pre-multiplying Eq. 24 by the transpose of Eq. 23, as discussed in [47]. By defining

$$\mathbf{z} = \left(\frac{\partial \mathbf{g}}{\partial \mathbf{x}} \right)^{-T} \left(\frac{\partial \mathbf{h}}{\partial \mathbf{x}} \right)^T \mathbf{m}, \tag{25}$$

it follows that

$$\mathbf{m}^T \mathbf{G} = -\mathbf{z}^T \frac{\partial \mathbf{g}}{\partial \theta} + \mathbf{m}^T \frac{\partial \mathbf{h}}{\partial \theta}. \tag{26}$$

4.2 Conciliation of spatially distributed data and forward model scales

In the data assimilation of spatially distributed observations, Eq. 6 assumes that the observations $\mathbf{d}$ and the model responses $\mathbf{h}$ are described at the same scale. This is often not the case. Due to resolution issues and acquisition limitations, observations are often not available at the scale of the model responses. Therefore, if no MS simulation is available, either $\mathbf{d}$ must be downscaled to the simulation scale or $\mathbf{h}$ must be upscaled to the observation scale.

The downscaling of observed data can be expressed as

$$\mathbf{d}' = \mathbf{D}\breve{\mathbf{d}}, \tag{27}$$

where $\mathbf{D}$ is an $N_F \times N_C$ downscaling operator, $\breve{\mathbf{d}} \in \mathbb{R}^{N_C}$ is the coarse-scale observation and $\mathbf{d}' \in \mathbb{R}^{N_F}$ is the interpolated observation at the fine scale.

Additionally, one must be able to describe the data covariance matrix $\mathbf{C}_D$, originally at (coarse) observation scale, at the fine scale. This can be achieved by setting

$$\mathbf{C}_D' = \mathbf{D}\breve{\mathbf{C}}_D \mathbf{D}^T, \tag{28}$$

where $\mathbf{C}_D'$ is the covariance matrix represented at the fine scale. It is simple to show that Eq. 28 holds because of the linearity of the expectation operator given the Gaussian assumptions. From Eq. 27, the expectation of $\mathbf{d}'$ is given by

$$E[\mathbf{d}'] = \mathbf{D}E[\breve{\mathbf{d}}]. \tag{29}$$

The covariance of $\mathbf{d}'$ can then be computed as (Emerick, A. A., personal communication, March 23, 2018)

$$\begin{aligned} Cov[\mathbf{d}'] &= E[(\mathbf{d}' - E[\mathbf{d}'])(\mathbf{d}' - E[\mathbf{d}'])^T] \\ &= E[\mathbf{D}(\breve{\mathbf{d}} - E[\breve{\mathbf{d}}])(\breve{\mathbf{d}} - E[\breve{\mathbf{d}}])^T \mathbf{D}^T] \\ &= \mathbf{D}E[(\breve{\mathbf{d}} - E[\breve{\mathbf{d}}])(\breve{\mathbf{d}} - E[\breve{\mathbf{d}}])^T]\mathbf{D}^T \\ &= \mathbf{D}Cov[\breve{\mathbf{d}}]\mathbf{D}^T \end{aligned} \tag{30}$$

Alternatively, one could upscale the model responses as

$$\breve{\mathbf{d}} = \mathbf{U}\mathbf{h}, \tag{31}$$

where $\mathbf{U}$ is an $N_C \times N_F$ upscaling operator, and solve Eq. 6 by setting $\mathbf{d}_{obs} = \breve{\mathbf{d}}_{obs}$. One advantage over the dowscaling strategy is that $\mathbf{C}_D$ is kept at its original scale.

We highlight that we only consider strategies that change observed data / response scale and do not consider strategies that change the original uncertain parameter description scale. This is because we aim to update the most accurate description of the model parameters, so that important fine-scale features (crucial to describe the physical phenomena) are not lost.

5 Multiscale data assimilation

An MS solution strategy provides a coarse-scale solution that can, theoretically, be represented at any resolution coarser than the fine-scale resolution. In data assimilation studies, where the spatially distributed data resolution is known and is coarser than the model resolution, the MS grid can be chosen to be at the same resolution as the assimilation grid. This allows spatially distributed model responses to be computed at the same scale as the observed data. Next, we devise a multiscale data assimilation procedure based on this feature. This allows us, instead of manipulating the data and/or the uncertain parameters, to accurately compute responses at the observed data scale.

Therefore, a multiscale objective function is introduced by re-writing Eq. 6 as

$$
O_{MS}\left(\check{\mathbf{h}}, \theta\right) = \frac{1}{2}\left(\theta - \theta_{prior}\right)^T \mathbf{C}_\theta^{-1}\left(\theta - \theta_{prior}\right) \\
+ \frac{1}{2}\left(\check{\mathbf{h}}\left(\check{\mathbf{x}}, \theta\right) - \check{\mathbf{d}}\right)^T \mathbf{C}_D^{-1}\left(\check{\mathbf{h}}\left(\check{\mathbf{x}}, \theta\right) - \check{\mathbf{d}}\right),
\tag{32}
$$

where $\check{\mathbf{h}}$ is the response at the (coarse) observation scale and $\check{\mathbf{x}}$ is the coarse state variable, computed by Eq. 14. Hence, the misfit term is computed at the coarse scale—the scale where data is assimilated—and the regularization term is described at the fine scale—the scale at the model parameters are described.

5.1 Multiscale gradient computation

As discussed in [42], the state vector can be described as a combination of both sets of primary variables at the fine and coarse scales, i.e.,

$$
\mathbf{x} = \begin{bmatrix} \check{\mathbf{x}} \\ \mathbf{x}' \end{bmatrix},
\tag{33}
$$

and, similarly, the model equations can be represented as a combination of the equations at both scales, i.e.,

$$
\mathbf{g}\left(\mathbf{x}, \theta\right) = \begin{bmatrix} \check{\mathbf{g}} \\ \mathbf{g}' \end{bmatrix} = \mathbf{0}.
\tag{34}
$$

The definition of the state vector as in Eq. 33 is a key aspect of this development. It allows the description of the state not only at the fine scale, but also at the coarse scale. The simulator responses $\mathbf{y}$ obtained from the multiscale method are represented as

$$
\check{\mathbf{y}} = \check{\mathbf{h}}\left(\mathbf{x}, \theta\right),
\tag{35}
$$

the sensitivity matrix $\mathbf{G}$ can be computed in a multiscale fashion as

$$
\mathbf{G}' = \left(\frac{\partial \check{\mathbf{h}}}{\partial \check{\mathbf{x}}} + \frac{\partial \check{\mathbf{h}}}{\partial \mathbf{x}'}\mathbf{P}\right)(\mathbf{RAP})^{-1}\mathbf{R}\left(\frac{\partial \mathbf{A}}{\partial \theta}\mathbf{P} + \mathbf{A}\frac{\partial \mathbf{P}}{\partial \theta}\right)\check{\mathbf{x}} \\
- \frac{\partial \check{\mathbf{h}}}{\partial \mathbf{x}'}\frac{\partial \mathbf{P}}{\partial \theta}\check{\mathbf{x}} + \frac{\partial \check{\mathbf{h}}}{\partial \theta}.
\tag{36}
$$

Equation 36 allows capturing derivative information of the coarse response, namely $\dfrac{\partial \check{\mathbf{h}}}{\partial \check{\mathbf{x}}}$ and $\dfrac{\partial \check{\mathbf{h}}}{\partial \theta}$.

In a similar way that MS methods represent (interpolate) the coarse-scale solution at the fine scale, the partial derivative of the prolongation operator w.r.t. the model parameters, $\dfrac{\partial \mathbf{P}}{\partial \theta}$ in Eq. 36, allows the representation of the model parameters at the fine scale, even though the primary variables are not solved at the model scale.

The gradient of Eq. 32 w.r.t. the model parameters at fine scale can be computed as

$$
\nabla_\theta O_{MS} = \mathbf{C}_\theta^{-1}\left(\theta - \theta_{prior}\right) + \left(\frac{d\check{\mathbf{h}}}{d\theta}\right)^T \mathbf{C}_D^{-1}\left(\check{\mathbf{h}}\left(\check{\mathbf{x}}, \theta\right) - \check{\mathbf{d}}\right) \\
= \mathbf{C}_\theta^{-1}\left(\theta - \theta_{prior}\right) + \mathbf{G}'^T\check{\mathbf{m}},
\tag{37}
$$

where

$$
\check{\mathbf{m}} = \mathbf{C}_D^{-1}\left(\check{\mathbf{h}}\left(\check{\mathbf{x}}, \theta\right) - \check{\mathbf{d}}\right).
\tag{38}
$$

The product $\mathbf{G}'^T\check{\mathbf{m}} = \left(\check{\mathbf{m}}^T\mathbf{G}'\right)^T$ can be solved at costs proportional to one coarse-scale backward simulation, regardless of the number of model parameters, via the MS adjoint method presented in [42], by pre-multiplying Eq. 36 by the transpose of Eq. 38, defining

$$
\check{\mathbf{z}} = (\mathbf{RAP})^{-T}\left(\frac{\partial \check{\mathbf{h}}}{\partial \check{\mathbf{x}}} + \frac{\partial \check{\mathbf{h}}}{\partial \mathbf{x}'}\mathbf{P}\right)^T\check{\mathbf{m}},
\tag{39}
$$

and rearranging the terms, it follows that

$$
\check{\mathbf{m}}^T\mathbf{G}' = \check{\mathbf{z}}^T\alpha + \beta\frac{\partial \mathbf{P}}{\partial \theta}\check{\mathbf{x}} + \check{\mathbf{m}}^T\frac{\partial \check{\mathbf{h}}}{\partial \theta},
\tag{40}
$$

where

$$
\alpha = \mathbf{R}\frac{\partial \mathbf{A}}{\partial \theta}\mathbf{P}\check{\mathbf{x}},
\tag{41}
$$

and

$$
\beta = \mathbf{z}^T\mathbf{RA} - \check{\mathbf{m}}^T\frac{\partial \check{\mathbf{h}}}{\partial \mathbf{x}'}.
\tag{42}
$$

5.1.1 Scaling operators partial derivative computation

The partial derivative computation of MSFV basis functions was originally discussed in [20] and recast in an algebraic, general mathematical framework expressed in terms of $\mathbf{P}$ in [42]. An efficient algorithm that computes the product $\beta\dfrac{\partial \mathbf{P}}{\partial \theta}\mathbf{x}$ in a backward-fashion was originally introduced in

[42]. Here, we recast this computation in terms of the wirebasket [60] submatrices of $\mathbf{A}$ and $\frac{\partial \mathbf{g}}{\partial \theta}$.

Proposition 1 (Partial derivative of MSFV prolongation operator w.r.t. θ) *Let* $\mathbf{A} \in \mathbb{R}^{N_F \times N_F}$ *be the elliptic discrete operator* (13), $\mathscr{P} \in \mathbb{R}^{N_F \times N_F}$ *the permutation matrix that re-orders from wirebasket to natural ordering, and* $\frac{\partial \mathbf{g}^\varphi}{\partial \theta} \in$ $\mathbb{R}^{N_F \times N_\theta}$ *the partial derivative of* Eq. 13 *w.r.t.* θ, *then* $\frac{\partial \mathbf{P}}{\partial \theta}$, *the partial derivative of the MSFV prolongation operator w.r.t.* θ *can be computed as*

$$\frac{\partial \mathbf{P}}{\partial \theta} = \mathscr{P} \begin{bmatrix} \mathbf{A}_{II}^{-1} \left(\mathbf{A}_{IE} \breve{\mathbf{A}}_{EE}^{-1} \dfrac{\partial \mathbf{g}_E^\varphi}{\partial \theta} - \dfrac{\partial \mathbf{g}_I^\varphi}{\partial \theta} \right) \\ -\breve{\mathbf{A}}_{EE}^{-1} \dfrac{\partial \mathbf{g}_E^\varphi}{\partial \theta} \\ \mathbf{0} \end{bmatrix}, \quad (43)$$

where $\frac{\partial \mathbf{g}_I^\varphi}{\partial \theta} \in \mathbb{R}^{N_I \times N_\theta}$ *and* $\frac{\partial \mathbf{g}_E^\varphi}{\partial \theta} \in \mathbb{R}^{N_E \times N_\theta}$ *are the partial derivative submatrices from* $\frac{\partial \mathbf{g}^\varphi}{\partial \theta} \in \mathbb{R}^{N_F \times N_\theta}$ *after the wirebasket re-ordering of* Eq. 13 *and the application of the MSFV localization assumptions and boundary conditions.*

Proof The basis function can be computed from Eq. 18, after applying the MSFV localization assumptions and applying the appropriate boundary conditions

$$\mathscr{P}^T \mathbf{g}^\varphi (\mathbf{x} = \boldsymbol{\varphi}, \theta) = \begin{bmatrix} \mathbf{g}_I^\varphi \\ \mathbf{g}_E^\varphi \\ \mathbf{g}_V^\varphi \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{II} & \mathbf{A}_{IE} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{A}}_{EE} & \mathbf{A}_{EV} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{VV} \end{bmatrix} \begin{bmatrix} \boldsymbol{\varphi}_I \\ \boldsymbol{\varphi}_E \\ \boldsymbol{\varphi}_V \end{bmatrix} - \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \quad (44)$$

whose partial derivative w.r.t. θ, reads

$$\begin{bmatrix} \dfrac{\partial \mathbf{g}_I^\varphi}{\partial \theta} \\ \dfrac{\partial \mathbf{g}_E^\varphi}{\partial \theta} \\ \dfrac{\partial \mathbf{g}_V^\varphi}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial \mathbf{A}_{II}}{\partial \theta} \boldsymbol{\varphi}_I + \dfrac{\partial \mathbf{A}_{IE}}{\partial \theta} \boldsymbol{\varphi}_E \\ \dfrac{\partial \tilde{\mathbf{A}}_{EE}}{\partial \theta} \boldsymbol{\varphi}_E + \dfrac{\partial \mathbf{A}_{EV}}{\partial \theta} \boldsymbol{\varphi}_V \\ \mathbf{0} \end{bmatrix}. \quad (45)$$

The partial derivative of Eq. 17 w.r.t. θ is

$$\frac{\partial \mathbf{P}}{\partial \theta} = \mathscr{P} \begin{bmatrix} \mathbf{A}_{II}^{-1} \left(\mathbf{A}_{IE} \left(\breve{\mathbf{A}}_{EE}^{-1} \left(-\dfrac{\partial \tilde{\mathbf{A}}_{EE}}{\partial \theta} \boldsymbol{\varphi}_E - \dfrac{\partial \mathbf{A}_{EV}}{\partial \theta} \boldsymbol{\varphi}_V \right) \right) - \dfrac{\partial \mathbf{A}_{IE}}{\partial \theta} \boldsymbol{\varphi}_E - \dfrac{\partial \mathbf{A}_{II}}{\partial \theta} \boldsymbol{\varphi}_I \right) \\ \tilde{\mathbf{A}}_{EE}^{-1} \left(-\dfrac{\partial \tilde{\mathbf{A}}_{EE}}{\partial \theta} \boldsymbol{\varphi}_E - \dfrac{\partial \mathbf{A}_{EV}}{\partial \theta} \boldsymbol{\varphi}_V \right) \\ \mathbf{0} \end{bmatrix}. \quad (46)$$

Substituting Eq. 45 in Eq. 46, it follows that

$$\frac{\partial \mathbf{P}}{\partial \theta} = \mathscr{P} \begin{bmatrix} \mathbf{A}_{II}^{-1} \left(\mathbf{A}_{IE} \tilde{\mathbf{A}}_{EE}^{-1} \dfrac{\partial \mathbf{g}_E^\varphi}{\partial \theta} - \dfrac{\partial \mathbf{g}_I^\varphi}{\partial \theta} \right) \\ -\tilde{\mathbf{A}}_{EE}^{-1} \dfrac{\partial \mathbf{g}_E^\varphi}{\partial \theta} \\ \mathbf{0} \end{bmatrix}.$$

$\square$

Hence, likewise, the MSFV prolongation operator can be fully determined directly from the fine-scale system matrix, while the partial derivative of the prolongation operator w.r.t. the model parameters can fully determined from both $\mathbf{A}$ and the partial derivative matrix $\frac{\partial \mathbf{g}}{\partial \theta}$. This allows a more straightforward implementation of the method in existing computational frameworks. Again, in the same way, the MS method can be integrated to existing simulation platforms if access to the system matrix and the grid topology is available, our method can be implemented in existing data assimilation frameworks if access to the grid topology, system matrix and partial derivative matrices are provided. Note that, however, as in any adjoint derivative computation framework, the computation of the partial derivative matrices for different parameters is usually a challenge. Derivative computation techniques like automatic differentiation [5] can be employed as a flexible solution to this problem..

Note that, even though Proposition 1 indicates the important ability of determining $\frac{\partial \mathbf{P}}{\partial \theta}$ from $\frac{\partial \mathbf{g}}{\partial \theta}$, it does not provide enough information about how to efficiently compute this partial derivative. It is discussed in [42] how to efficiently compute the left/right multiplication of $\frac{\partial \mathbf{P}}{\partial \theta}$ in the context of, respectively, the direct and adjoint methods. Algorithm 4 in that paper presents an efficient way to compute the product $\beta \frac{\partial \mathbf{P}}{\partial \theta} \breve{\mathbf{x}}$, at costs proportional to the number of coarse cells and independent of the number of parameters, suitable to be used in combination with Eq. 40 for its efficient solution.

Because we aim to compute the gradient of a scalar function (see Eq. 6), the computational cost is proportional to solving one so-called backward simulation. In our case, this is proportional to the solution of the transposed linear system of equations in Eq. 39, whose system matrix $\mathbf{RAP}$ has size proportional to the number of coarse grid cells, plus the solution of $\beta \frac{\partial \mathbf{P}}{\partial \theta} \breve{\mathbf{x}}$, the cost of which is equal to the solution of the basis functions (43). See [42] for details. Hence, the computational cost ratio of the proposed method is also given by Eq. 20.
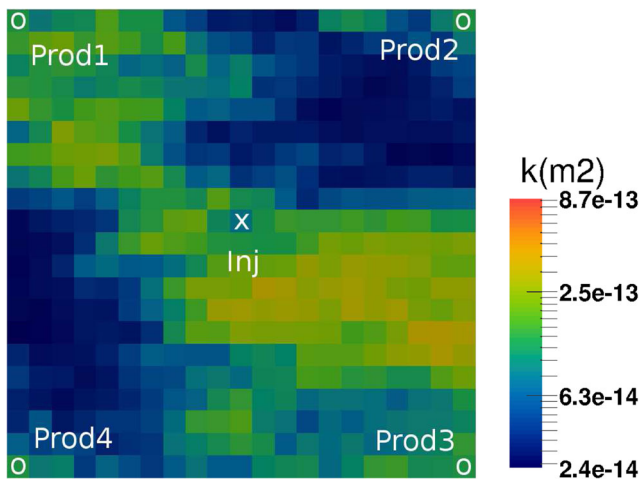
Fig. 2 The synthetic inverted five-spot model used in the numerical experiments. One of the 1000 permeability realizations is shown

6 Numerical experiments

We focus our analysis on incompressible, single-phase flow. In our experiments, a simple synthetic model is considered as proof of concept (see Fig. 2). It is a 2D inverted five-spot model, consisting of a 21×21 equidistant Cartesian mesh with grid block sizes of $33.3 \times 33.3 \times 2$ m. The reservoir porosity is constant and equal to 0.3. A fluid dynamic viscosity of 0.5×10^{-3} Pa.s is considered. The wells are controlled by bottom-hole pressure (BHP). The injection pressure is 35 MPa and the production wells' BHP is 25 MPa.

The uncertainty around the absolute permeability distribution is represented by an ensemble of different permeability realizations. The ensemble is generated via the decomposition of a reference permeability "image" using principal component analysis (PCA) parameterization [29]. Figure 3 illustrates four different permeability realizations from the ensemble of 1000 equiprobable permeability realizations. In order to focus on the MS aspect of the data assimilation process, we assume that pressures can be approximately extracted from a time-lapse seismic survey [4, 34, 38, 39, 51]. However, it is important to note that this is not a limitation. If one is interested to perform the data assimilation in different domains, say, in the impedances domain [10], the additional complexity involved is the appropriate incorporation of seismic forward model equations in the forward model set of equations [9] and, consequentially, the computation of the appropriate partial derivative information necessary to compute Eq. 21.

The "true" observed data, $\mathbf{p}_{true}$, is obtained from a twin experiment, where a MS simulation is run using a permeability field randomly chosen from an ensemble of equiprobable model realizations. The coarse observed data is the coarse-scale pressure $\check{\mathbf{x}}$ computed with Eq. 14, while
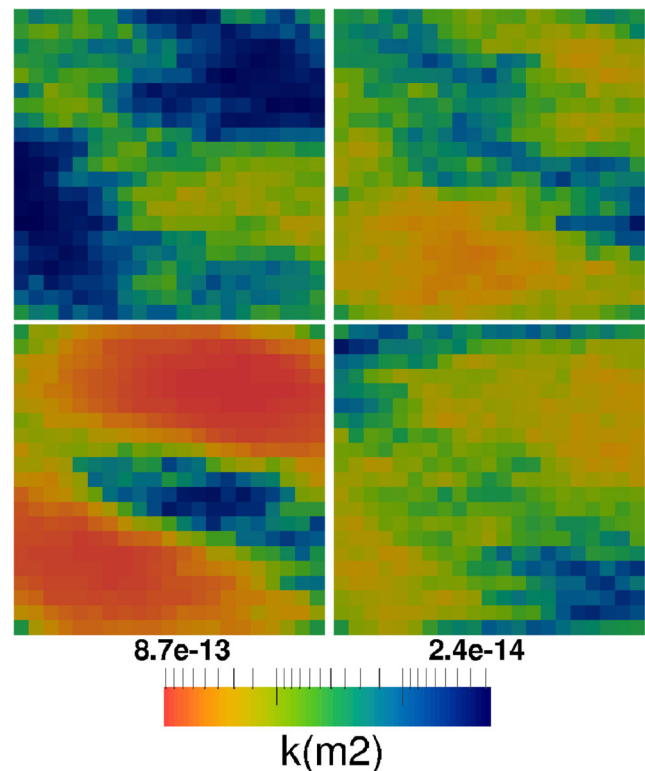

Fig. 3 Four different permeability realizations from the ensemble of 1,000 members of the toy model numerical experiment

the (hypothetical, for comparison purposes) fine observed data is computed with Eq. 13. The pressure measurement errors are considered to follow a spherical covariance model with a 1% standard deviation in all experiments and correlation lengths equal to 1 grid-block size. Noise is added to the data by setting

$$\mathbf{p}_{obs} = \mathbf{p}_{true} + \sqrt{\mathbf{C}_D}\mathbf{z}, \qquad (47)$$

where $\mathbf{z}$ is sampled from $\mathcal{N}(\mu = 0, \sigma^2 = 1)$ and $\sqrt{\mathbf{C}_D}$ is computed from a Cholesky decomposition. More details on the procedure can be found in [46]. The resulting noisy observed coarse and fine pressure fields are illustrated, respectively, in Fig. 4 d and c.

We consider the observation grid (Fig. 4b) where the observed data is represented to be three times coarser than the model grid (Fig. 4a) in the x and y directions. Hence, it has 7×7 grid blocks with grid block size $99.9 \times 99.9 \times 2$ m.

The covariance matrix $\mathbf{C}_\theta$ is computed from the ensemble of realizations as

$$\mathbf{C}_\theta = \frac{1}{N_e - 1} \left(\Theta - \mu \mathbf{e}^T \right) \left(\Theta - \mu \mathbf{e}^T \right)^T, \qquad (48)$$

Fig. 4 Schematic representation of the model (fine) grid (**a**) and observation (coarse) grid (**b**). Also, the (noisy) pressure data distribution observed at the fine-grid (hypothetical complete observations) (**c**) and at the actual observation grid resolution (**d**)

(a) Model grid

(b) Observation grid

(c) Fine-scale observation

(d) Coarse-scale observation
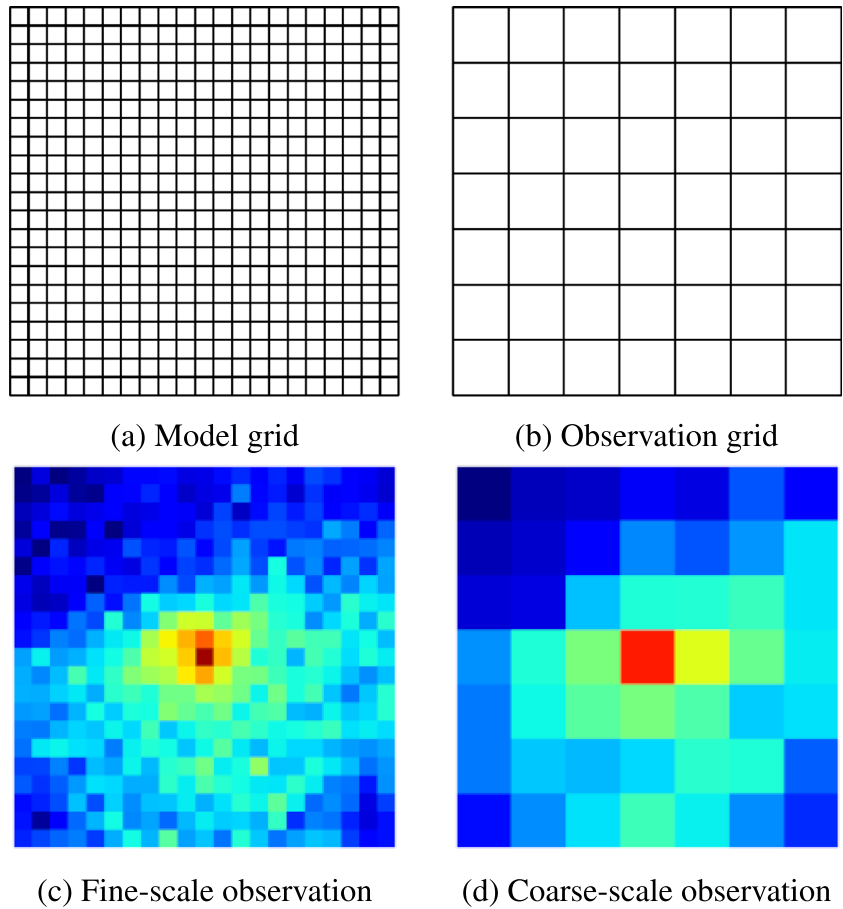
where Θ is the $N_F \times N_e$ matrix whose jth column is given by the member of the ensemble θ_j, $j \in \{1, ..., N_e\}$,

$$\mu = \frac{1}{N_e} \sum_{j=1}^{N_e} \theta_j, \tag{49}$$

is the ensemble mean, and $\mathbf{e} = [\mathbf{1}, ..., \mathbf{1}]^T$ is a vector of ones of size $N_e \times 1$. The prior is taken to be the ensemble mean,

$$\theta_{prior} = \mu. \tag{50}$$

In the fine-scale data assimilation strategy, an adjoint model [47] is used to calculate the OF gradient given by Eq. 6. In the multiscale strategy, we employ the MS adjoint gradient computation depicted in Algorithm 4 in [42]. Because the spatially distributed observed data at the coarse scale is the primary variable itself, in Eq. 32, we have

$$\frac{\partial \breve{\mathbf{h}}}{\partial \breve{\mathbf{x}}} = \breve{\mathbf{I}}, \tag{51}$$

where $\breve{\mathbf{I}}$ is the $N_C \times N_C$ identity matrix, and

$$\frac{\partial \breve{\mathbf{h}}}{\partial \mathbf{x}'} = \mathbf{0}. \tag{52}$$

when pressure is observed at the coarse scale, and

$$\frac{\partial \mathbf{h}}{\partial \mathbf{x}} = \mathbf{I}, \tag{53}$$

when pressure is observed at the fine scale.

Also, because in this case the relationship between the primary variables and the outputs is not a function of the parameter, it follows that

$$\frac{\partial \breve{\mathbf{h}}}{\partial \theta} = \breve{\mathbf{0}}, \tag{54}$$

and

$$\frac{\partial \mathbf{h}}{\partial \theta} = \mathbf{0}. \tag{55}$$

We utilize the limited memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithm as presented in [44], as it is the most efficient algorithm to deliver optimization results for the solution of Eq. 7 [22].

6.1 Construction of scaling operators

6.1.1 Observed data downscaling

Two different approaches on how to deal with the scale dissimilarity via downscaling are considered here. In the first one, we downscale the response measured at the coarse scale by setting

$$\mathbf{D} = \mathbf{R}^T. \tag{56}$$

where $\mathbf{R}$ is the MSFV restriction operator. This strategy can be viewed as a constant interpolation of the coarse-scale observations at the fine-scale model scale.

In the second strategy, we build a multiscale prolongation operator $\mathbf{P}_{prior} = \mathbf{P}\left(\theta_{prior}\right)$, whose columns are comprised of local multiscale basis functions [32], and prolong (interpolate) the coarse-scale information by setting

$$\mathbf{D} = \mathbf{P}_{prior}. \tag{57}$$

Note that $\mathbf{P}_{prior}$ is static and can be viewed as a MS downscaling operator.

In the aforementioned strategies, Eq. 6 can be used by making $\mathbf{d} = \mathbf{d}'$ and a conventional gradient–based optimization to solve Eq. 7 is run at the model (fine) scale.

6.1.2 Model response upscaling

Two upscaling strategies are considered. In the first one, a simple arithmetic average is applied by setting

$$\mathbf{U} = \mathbf{M}, \tag{58}$$

where,

$$\mathbf{M}(c, f) = \begin{cases} \dfrac{1}{N_F^C}, & \text{if } f \in \Omega_c \\ 0, & \text{otherwise} \end{cases}. \tag{59}$$

In Eq. 59, N_F^C is the number of fine-grid cells within a given coarse cell C, Ω_c is the cth primal coarse-grid domain and f is the fine-grid cell index.

In the second upscaling strategy, we again build a prolongation operator $\mathbf{P}_{prior}$ based on θ_{prior} and upscale the observed response by setting

$$\mathbf{U} = \mathbf{P}^\dagger, \tag{60}$$

where the † symbol denotes the Moore-Penrose pseudo-inverse. Here, we construct $\mathbf{P}^\dagger$ from its truncated singular value decomposition (TSVD) [49]

$$\mathbf{P}^\dagger = \boldsymbol{\Sigma}\boldsymbol{\Lambda}\boldsymbol{\Delta}^T = \begin{bmatrix} \boldsymbol{\Sigma}_p & \boldsymbol{\Sigma}_0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Delta}_p^T \\ \boldsymbol{\Delta}_0^T \end{bmatrix} = \boldsymbol{\Sigma}_p \boldsymbol{\Lambda}_p \boldsymbol{\Delta}_p^T \tag{61}$$

where $\boldsymbol{\Sigma} \in \mathbb{R}^{N_F \times N_F}$ and $\boldsymbol{\Delta} \in \mathbb{R}^{N_C \times N_C}$ are orthonormal matrices, $\boldsymbol{\Lambda} \in \mathbb{R}^{N_F \times N_C}$ is a diagonal matrix containing the singular values of $\mathbf{P}$, and the subscript p indicates the first p columns of the matrices corresponding the p non-zero singular values.

6.2 Maximum a posteriori probability estimate

In this section, we assess the performance of our newly introduced method via the estimation of the maximym a posteriori probability (MAP) in comparison to the upscaling/downscaling strategies discussed in Section 6.1. Therefore, six different data assimilation strategies are considered, namely:

1. fine-scale data assimilation with constant prolongation downscaling of observed pressure (56);
2. fine-scale data assimilation with prior MS prolongation downscaling of observed pressure (57);
3. fine-scale data assimilation with arithmetic average to upscale the simulated pressure (59);
4. fine-scale data assimilation with pseudo-inverse of MS prolongation to upscale the simulated pressure (60);
5. multiscale data assimilation strategy introduced in this work;
6. fine-scale data assimilation with complete observations available at the model (fine) scale.

The latter, a hypothetical situation, is considered as the reference case, as if enough resolution was available to resolve the observed property at the (fine) model scale. Also, note that MS operators are used in strategies 1, 2, and 4. For comparison purposes, we consider the objective function normalized by the number of data points N_d. Furthermore, according to [46, 50], an acceptable data match is achieved when $\dfrac{O}{N_d} \approx 0.5$.

It is important to highlight that strategies 1–4 have similar computation cost of strategy 6, the hypothetical situation considering complete observations. This is due to the fact that, regardless of the upscaling/downscaling of the quantities (i.e., observations or responses), which also adds extra computations, the gradient computation is given by Eq. 26. This means the solution of a transposed linear system with size proportional to the number of fine-grid cells.

Firstly, we present a qualitative discussion based on the MAP conditioned permeability fields and final matched pressure fields in comparison to the respective "true" permeability and pressure fields. The results for the fine-scale, complete observation data assimilation exercise is illustrated in Fig. 5, followed by the results from the
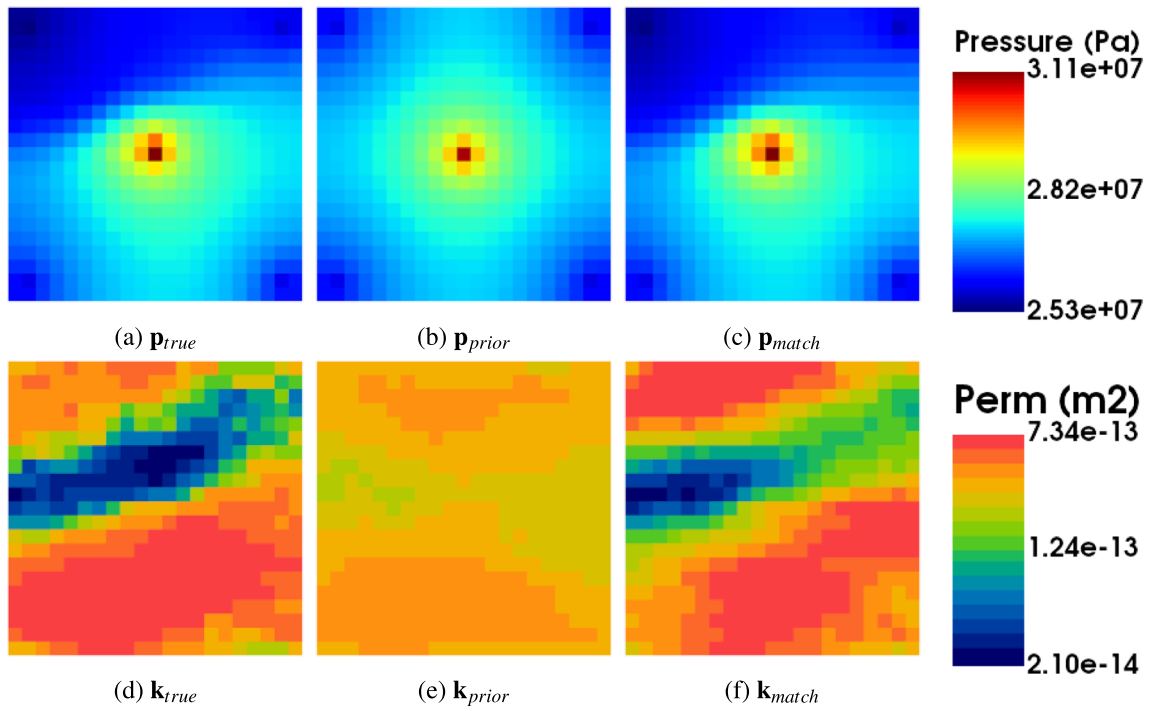
(a) $\mathbf{p}_{true}$ (b) $\mathbf{p}_{prior}$ (c) $\mathbf{p}_{match}$

(d) $\mathbf{k}_{true}$ (e) $\mathbf{k}_{prior}$ (f) $\mathbf{k}_{match}$
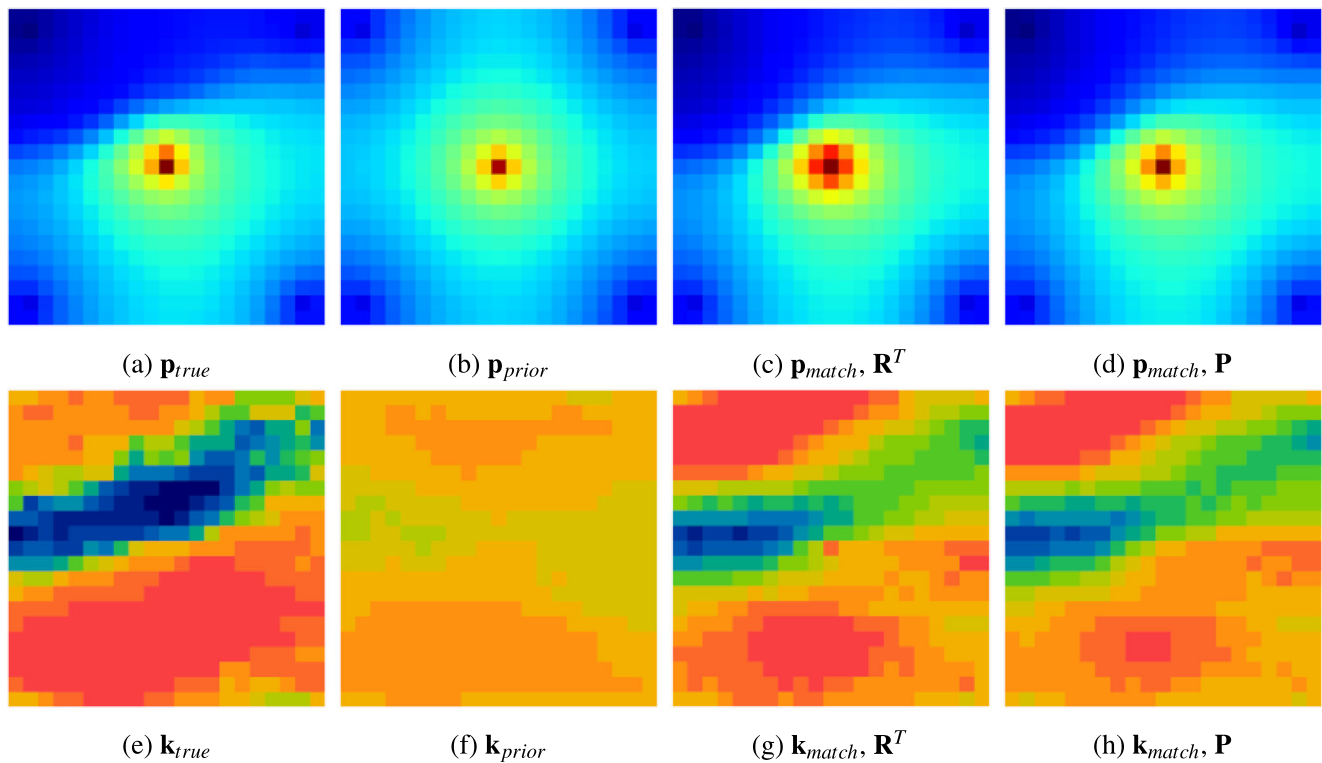
Fig. 5 Data assimilation results, fine-scale data assimilation, complete, fine-scale observations. In the first row, the **a** true, **b** initial, and **c** matched pressure fields are shown. In the second row, **d** the " true," **e** the prior, and **f** the conditioned permeability fields are shown. The color scales follow the color bars shown in Fig. 5

(a) $\mathbf{p}_{true}$ (b) $\mathbf{p}_{prior}$ (c) $\mathbf{p}_{match}$, $\mathbf{R}^T$ (d) $\mathbf{p}_{match}$, $\mathbf{P}$

(e) $\mathbf{k}_{true}$ (f) $\mathbf{k}_{prior}$ (g) $\mathbf{k}_{match}$, $\mathbf{R}^T$ (h) $\mathbf{k}_{match}$, $\mathbf{P}$

Fig. 6 Data assimilation results, fine-scale data assimilation, downscaling of data observations. In the first row, the **a** true, **b** initial, **c** matched using the constant interpolation ($\mathbf{R}^T$), and **d** matched using the MS prolongation operator (**P**) pressure fields are shown. In the second row, **d** the "true," **e** the prior, **g** the conditioned using $\mathbf{R}^T$, and **g** the conditioned using **P** permeability fields are shown. The color maps follow the color bar found in Fig. 5

downscaling and upscaling data conciliation strategies, represented, respectively, by Figs. 6 and 7. Lastly, the results of the data assimilation using our MS data assimilation strategy are illustrated in Fig. 8.

From a qualitative point of view, the matched responses from all data assimilation strategies, except the responses obtained by the constant interpolation downscaling (Fig. 6c) and arithmetic average upscaling (Fig. 7c), are fairly similar to the observed data. The pressure matches are both in accordance with the fine-scale pressure match (Fig. 5c) and with the "true" pressure field. However, the simpler upscaling/downscaling strategies result in somewhat poorer matches around the injection well. It can be noted that higher pressure responses are computed around the injection well. This is due to the simpler interpolation employed in the pressure upscaling/downscaling, which results in a constant pressure for the fine-grid cells within the coarse grid block where the injection well is located.

Also, it is possible to observe that, from the point of view of the conditioned permeability fields, all assimilation strategies were capable of recovering the main features. Furthermore, not much difference is noted in the results when comparing the upscaling to the downscaling matched permeability fields. In order to better assess the quality of

the parameter matches, we investigate the permeability distribution from the different matching exercises. Therefore, the density functions of the matched permeability fields are plotted in Fig. 9. It is possible to note that, even though the initial permeability distribution is considerably far from the true model (due to the rather homogeneous prior used in the MAP), the complete observation, fine-scale strategy is capable of reproduction of the reference permeability density function. But, more importantly, the MS data assimilation, with coarse scale only observations, can also provide a permeability field whose density function is consistent with the "true" permeability density function. Also, it can be noted that the permeability fields obtained by the other strategies are also consistent with the " true" permeability distribution.

We also analyze the optimization convergence behavior shown in Fig. 10 for quantitative assessment of the match. The fine-scale reference data assimilation reaches a normalized OF value very close to the ideal value of 0.5, while all other data assimilation strategies reach values relatively higher, with the simpler constant interpolation and arithmetic average scaling strategies reaching slightly higher OF values. It is important to note that the optimization behavior is remarkably similar for all upscaling/downscaling strategies, as well as for the MS data assimilation strategy here
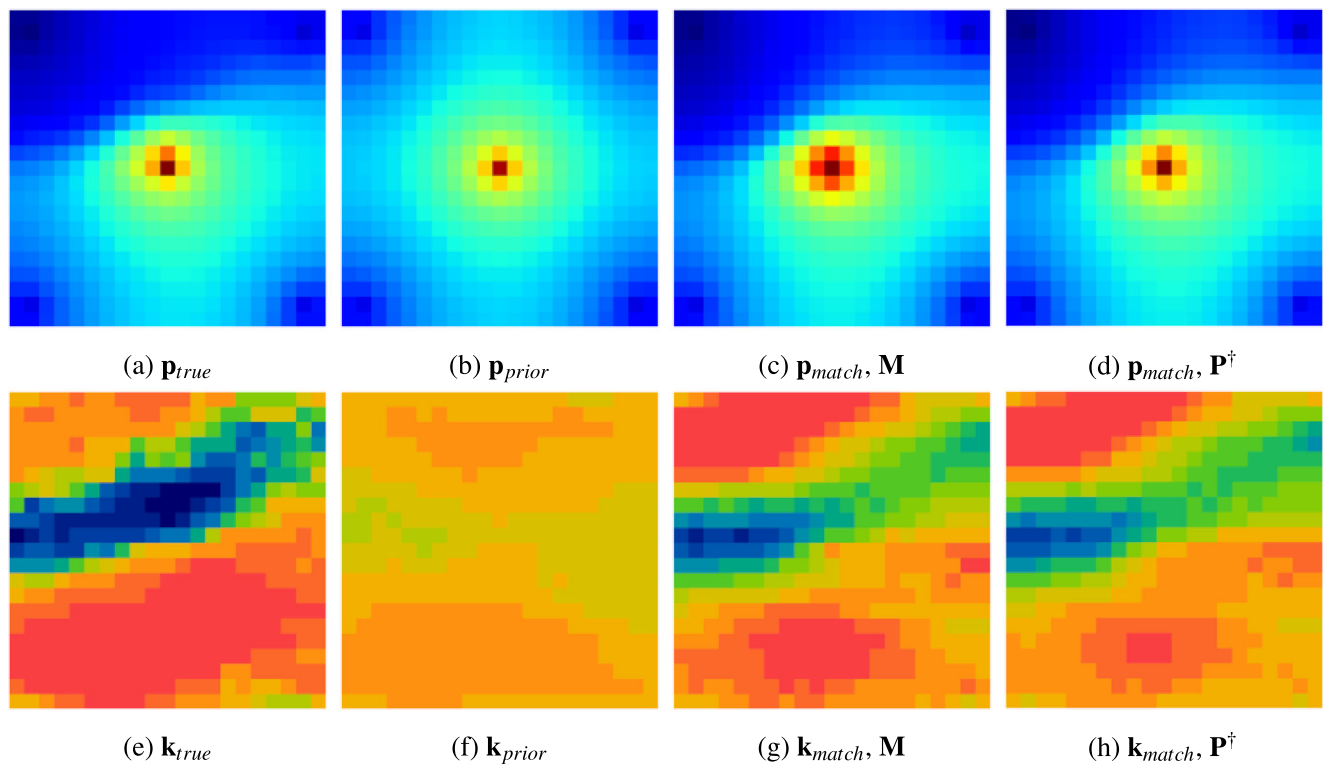

(a) $\mathbf{p}_{true}$ (b) $\mathbf{p}_{prior}$ (c) $\mathbf{p}_{match}$, $\mathbf{M}$ (d) $\mathbf{p}_{match}$, $\mathbf{P}^{\dagger}$

(e) $\mathbf{k}_{true}$ (f) $\mathbf{k}_{prior}$ (g) $\mathbf{k}_{match}$, $\mathbf{M}$ (h) $\mathbf{k}_{match}$, $\mathbf{P}^{\dagger}$

Fig. 7 Data assimilation results, fine-scale data assimilation, upscaling of model responses. In the first row, the **a** true, **b** initial, **c** matched using arithmetic average (**M**), and **d** matched using the MS prolongation operator pseudo-inverse (**P**†) pressure fields are shown. In the second row, **d** the " true," **e** the prior, **g** the conditioned using **M**, and **g** the conditioned using **P**† permeability fields are shown. The color maps follow the color bars found in Fig. 5

(a) $\mathbf{p}_{true}$ (b) $\mathbf{p}_{prior}$ (c) $\mathbf{p}_{match}$

(d) $\mathbf{k}_{true}$ (e) $\mathbf{k}_{prior}$ (f) $\mathbf{k}_{match}$
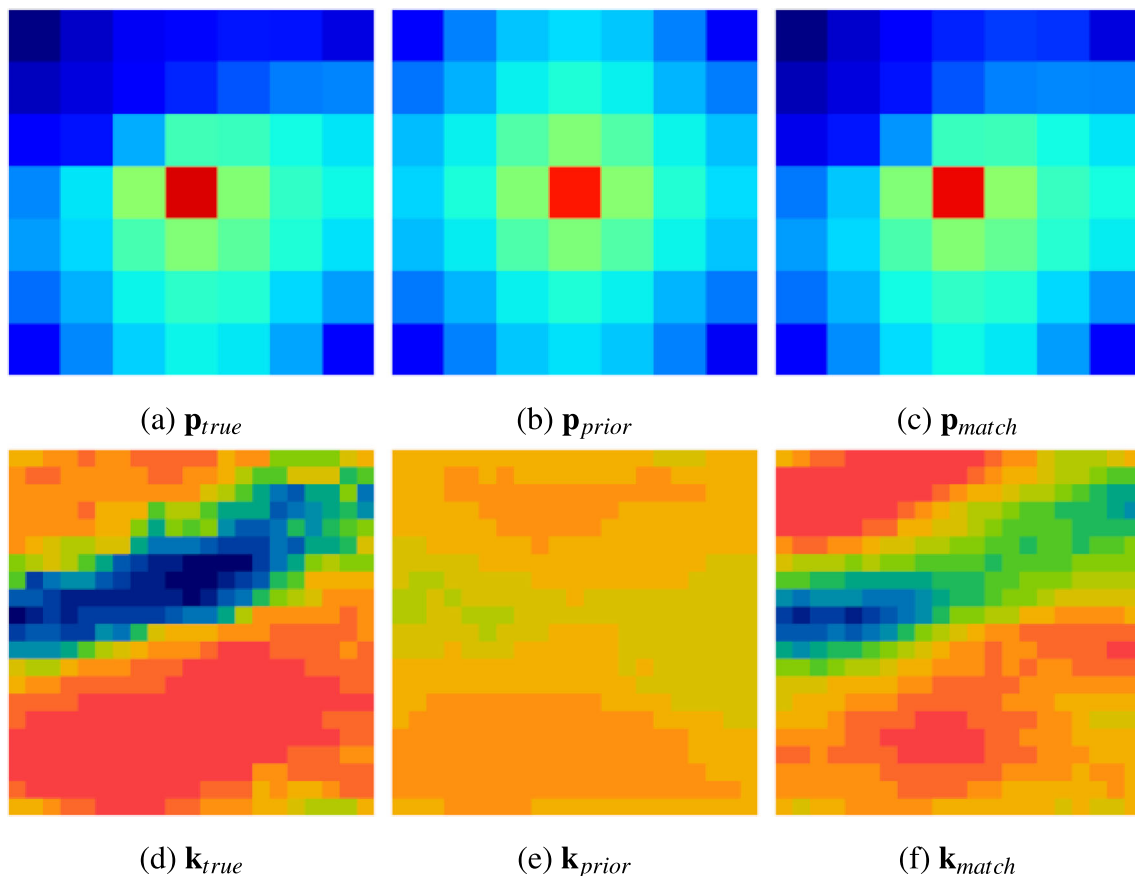
Fig. 8 Data assimilation results, multiscale data assimilation. In the first row, the **a** true, **b** initial and **c** matched pressure fields are shown. In the second row, **d** the " true," **e** the prior, and **f** the conditioned permeability fields are shown. The color maps follow the color bars found in Fig. 5. Note that MS assimilation provides a fine-scale improved permeability field **f** while the observation is at coarse scale. The color scales follow the color bars shown in Fig. 5

presented. As discussed in [42], even though in this exercise the convergence behavior of our method is similar to the others, the computational cost of the gradient computation is proportional to the coarse grid dimensions, while in all other methods the cost is proportional to fine-grid dimensions.

6.3 Uncertainty quantification

A RML is run for 100 randomly chosen permeability realization from the 1000 members ensemble (Fig. 3), for each data conciliation strategy. In order to estimate the conditioned permeability distribution for each permeability realization, a LBFGS optimization is run for each chosen member. The results for the exercise are shown in Fig. 11.

It can be observed that the permeability marginal PDFs conditioned to the pressure data obtained by the MS data assimilation here introduced (Fig. 11e) are closer to the reference fine-scale conditioned PDFs (Fig. 11f). Additionally, by observing the spread of the conditioned PDFs obtained from the RML employing the upscaling/downscaling strategies, one can note that the MS

strategy is also capable of somewhat better representing the uncertainty.

6.4 Discussion

Firstly, even though all data assimilation strategies were capable of achieving similar MAP estimates, one should note that the synthetic case used in the experiments has low permeability contrasts. Moreover the five-spot configuration is very simple and the well spacing is relatively dense compared to the characteristic size of the heterogeneities. Nonetheless, given the good results observed in the employment of our MS data assimilation strategy, we believe that the performance of the method in more challenging scenarios is worth investigating. A systematic study of the effects of the underlying geological complexity on the MS assimilation procedure is necessary. The MS ability to preserve fine-scale features is expected to allow for more detailed description of the fine-scale uncertain parameters. Additionally, we emphasize the importance of the proper representation of the measurement
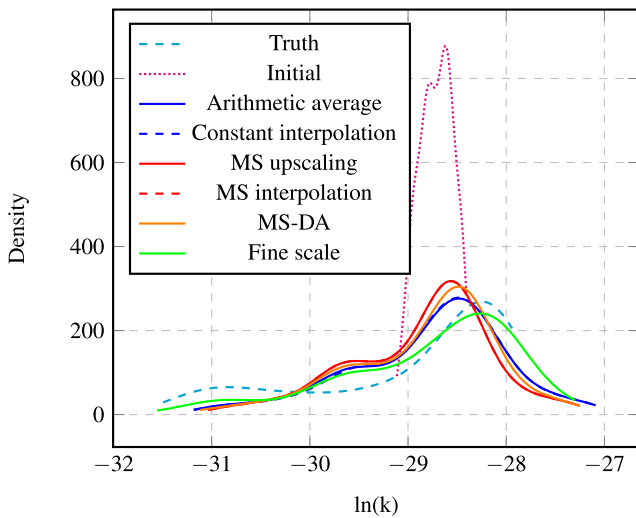
Fig. 9 Permeability conditioned marginal PDFs for the different data assimilation strategies

errors at different scales. One must take into account the data redundancy in the case of downscaling the observed data to the model scale.

One could consider a third, fine-scale only, MS-based approach, based on the reconstruction of the prolongation operator at every optimizer iteration γ, so that changes in the permeability during the optimization process are also captured by the basis functions update. Hence, one could write

$$\mathbf{d}'^{\gamma}_{obs} = \mathbf{P}^{\gamma} \check{\mathbf{d}}, \qquad (62)$$

where $\mathbf{P}^{\gamma}$ is the reconstructed prolongation operator at every optimization iteration γ. This can only be achieved at the
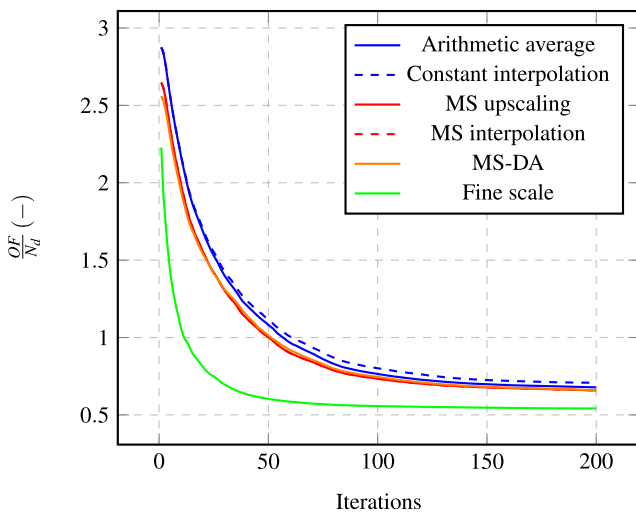
Fig. 10 Optimization performance of the data assimilation utilizing the 6 different scale conciliation strategies as presented in this section. Note that the FS represent the hypothetical case where both observed data and model parameters are at fine scale (i.e., item 6 on our list of strategies)

expense of the reconstruction of the basis function every γ. We performed studies (not reported here) where we neglect the partial derivative of $\mathbf{P}$ w.r.t θ but we did update $\mathbf{P}$. Similar results were obtained when $\mathbf{P}$ is not updated and only based on the prior (57), as reported here. Moreover, it is discussed in [42] how to efficiently compute $\partial\mathbf{P}/\partial\theta$. This can be an alternative to further take advantage of MS principles even when a MS forward simulation is not available.

In this work, we employ a two-stage MS simulation strategy, and consequently a two-stage MS gradient computation strategy. We make the primal MS coarse grid to be coincident to the observation grid resolution. However, the idea of the MS data assimilation can be extended to seamlessly address data available at multiple scales, or even consider one, or multiple MS grid resolution(s) for assimilation purposes only and different one(s) for the forward simulation. To this end, multilevel multiscale strategies [6] could be applied. Following the same multilevel multiscale strategy, data acquired at different scales (e.g., electromagnetics, high resolution close to the well, along with seismic data, low resolution in the vertical direction) could also be seamlessly and simultaneously be assimilated.

Following our studies, and also reported by [42] and [20], it can be noted that MSFV gradients can be less accurate for highly heterogeneous media. In addition, one may want to have error control on the MSFV gradient quality for practical applications. Furthermore, LBFGS proposes under/overshooting updates, mainly close to the wells [21], which also configures a challenging scenario for the MSFV gradient computation. These challenges can be addressed from the optimization point of view or from the gradient computation perspective. The former can be considered via data misfit damping or parameter constraints [21]. The latter by improved MS gradient quality, via more accurate MSFV solutions [26, 57]. An iterative MSFV gradient computation, following the solution strategy proposed by [26], could allow for additional error control over the gradient computation.

7 Final Remarks

Our numerical experiments indicate that the presented method has the potential to outperform strategies that rely on upscaling/downscaling of model responses/observed data. An important result is the ability of our MS data assimilation strategy to closely reproduce the reference fine-scale uncertainty quantification results. Applications in more complex cases, and for different types of assimilation problems, should give more insights about the computational and methodological advantages of MS data assimilation, as indicated by the results of the simple
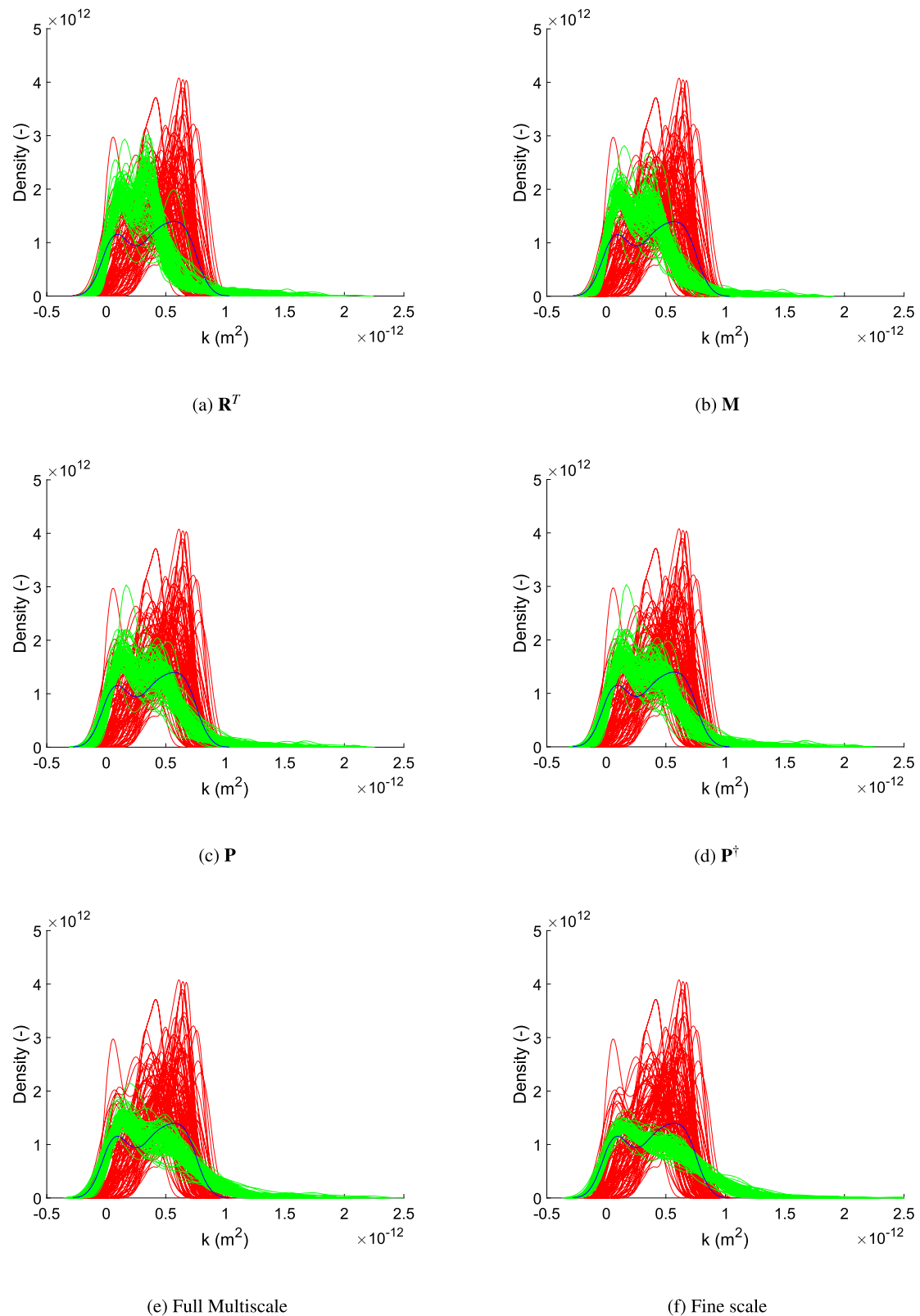
(a) $\mathbf{R}^T$

(b) $\mathbf{M}$

(c) $\mathbf{P}$

(d) $\mathbf{P}^\dagger$

(e) Full Multiscale

(f) Fine scale

Fig. 11 RML probability density functions for 100 permeability realizations randomly chosen from the original ensemble of 1000 realizations (see Fig. 3). The curves in red represent the prior permeability distributions, while the curves in green the conditioned permeability distributions and the curves in blue represent the reference ("true")
permeability distribution. **a** Constant interpolation downscaling ($\mathbf{R}^T$), **b** arithmetic average downscaling ($\mathbf{M}$), **c** MS prolongation operator downscaling, **d** MS prolongation operator upscaling $\mathbf{P}^\dagger$, **e** the multiscale data assimilation strategy, and **f** the reference fine scale

example addressed in our study. Clearly, larger-scale tests, with more complex heterogeneity, are required to further quantify these potential benefits of MS data assimilation. Our paper demonstrates how to consistently formulate such MS data assimilation strategy, in particular in combination with the use of adjoint-based techniques to efficiently obtain MS gradient information, and in an algebraic framework which allows for implementation in existing computational platforms.

References

1. Aanonsen, S.I., Nævdal, G., Oliver, D.S., Reynolds, A.C., Vallés, B.: Review of ensemble kalman filter in petroleum engineering. SPE J. **14**(3), 393–412 (2009)
2. Chavent, G., Dupuy, M., Lemmonier, P.: History matching by use of optimal theory. Soc. Pet. Eng. J. **15**(01), 74–86 (1975)
3. Chung, E.T., Efendiev, Y., Jin, B., Leung, W.T., Vasilyeva, M.: Generalized multiscale inversion for heterogeneous problems. arXiv:1707.08194 (2017)
4. Cole, S., Lumley, D., Meadows, M., Tura, A.: Pressure and saturation inversion of 4d seismic data by rock physics forward modeling. In: SEG Technical Program Expanded Abstracts 2002, pp. 2475–2478. Society of Exploration Geophysicists (2002)
5. Corliss, G., Faure, C., Griewank, A., Hascoet, L., Naumann, U.: Automatic differentiation of algorithms: from simulation to optimization, vol. 1. Springer Science & Business Media (2002)
6. Cusini, M., van Kruijsdijk, C., Hajibeygi, H.: Algebraic dynamic multilevel (adm) method for fully implicit simulations of multiphase flow in porous media. J. Comput. Phys. **314**, 60–79 (2016)
7. Durlofsky, L.J.: Upscaling of geocellular models for reservoir flow simulation: a review of recent progress. In: 7th International Forum on Reservoir Simulation Bühl/Baden-Baden, Germany, pp. 23–27 (2003)
8. Efendiev, Y., Galvis, J., Hou, T.Y.: Generalized multiscale finite element methods (gmsfem). J. Comput. Phys. **251**, 116–135 (2013)
9. Emerick, A.A., Moraes, R., Rodrigues, J., et al.: Calculating seismic attributes within a reservoir flow simulator. In: Latin American & Caribbean Petroleum Engineering Conference. Society of Petroleum Engineers (2007)
10. Emerick, A.A., Moraes, R., Rodrigues, J., et al.: History matching 4d seismic data with efficient gradient based methods. In: EUROPEC/EAGE Conference and Exhibition. Society of Petroleum Engineers (2007)
11. Emerick, A.A., Reynolds, A.C.: Ensemble smoother with multiple data assimilation. Comput. Geosci. **55**, 3–15 (2013)
12. Emerick, A.A., Reynolds, A.C.: Investigation of the sampling performance of ensemble-based methods with a simple reservoir model. Comput. Geosci. **17**(2), 325–350 (2013). https://doi.org/10.1007/s10596-012-9333-z
13. Emerick, A.A., Reynolds, A.C., et al.: History-matching production and seismic data in a real field case using the ensemble smoother with multiple data assimilation. In: SPE Reservoir Simulation Symposium. Society of Petroleum Engineers (2013)
14. Evensen, G.: Data assimilation: the ensemble Kalman filter. Springer Science & Business Media (2009)
15. Farmer, C.: Upscaling: a review. Int. J. Numer. Methods Fluids **40**(1-2), 63–78 (2002)
16. Fossum, K., Mannseth, T.: Coarse-scale data assimilation as a generic alternative to localization. Comput. Geosci. **21**(1), 167–186 (2017)
17. Fossum, K., Mannseth, T.: A novel multilevel method for assimilating spatially dense data. In: ECMOR XVI-16th European Conference on the Mathematics of Oil Recovery (2018)
18. Frederick, C., Engquist, B.: Numerical methods for multiscale inverse problems. arXiv:1401.2431 (2014)
19. Fu, J., Caers, J., Tchelepi, H.A.: A multiscale method for subsurface inverse modeling: Single-phase transient flow. Adv. Water Resour. **34**(8), 967–979 (2011)
20. Fu, J., Tchelepi, H.A., Caers, J.: A multiscale adjoint method to compute sensitivity coefficients for flow in heterogeneous porous media. Adv. Water Resour. **33**(6), 698–709 (2010)
21. Gao, G., Reynolds, A.C., et al.: An improved implementation of the lbfgs algorithm for automatic history matching. In: SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers (2004)
22. Gao, G., Reynolds, A.C., et al.: An improved implementation of the lbfgs algorithm for automatic history matching. SPEJ **11**(01), 5–17 (2006)
23. Gervais-Couplet, V., Roggero, F., Feraille, M.D., Ravalec-Dupin, L., Seiler, A., et al.: Joint history matching of production and 4d-seismic related data for a north sea field case. In: SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers (2010)
24. Gosselin, O., Aanonsen, S., Aavatsmark, I., Cominelli, A., Gonard, R., Kolasinski, M., Ferdinandi, F., Kovacic, L., Neylon, K., et al.: History matching using time-lapse seismic (huts). In: SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers (2003)
25. Graham, I.G., Hou, T.Y., Lakkis, O., Scheichl, R.: Numerical analysis of multiscale problems, vol. 83. Springer Science & Business Media (2012)
26. Hajibeygi, H., Jenny, P.: Adaptive iterative multiscale finite volume method. J. Comput. Phys. **230**(3), 628–643 (2011)
27. Haugen, V.E.J., Natvik, L.J., Evensen, G., Berg, A.M., Flornes, K.M., Naevdal, G., et al.: History matching using the ensemble kalman filter on a north sea field case. In: SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers (2006)
28. Hou, T.Y., Wu, X.H.: A multiscale finite element method for elliptic problems in composite materials and porous media. J. Comput. Phys. **134**(1), 169–189 (1997)
29. Jansen, J.D.: A simple algorithm to generate small geostatistical ensembles for subsurface flow simulation. Research note. Dept. of Geoscience and Engineering, Delft University of Technology, The Netherlands. https://doi.org/uuid:6000459e-a0cb-40d1-843b-81650053e093 (2013)
30. Jansen, J.D.: A Systems Description of Flow Through Porous Media. Springer, Berlin (2013)
31. Jansen, J.D.: Gradient-based optimization of flow through porous media: Version 3. Course notes. Delft University of Technology (2016). https://doi.org/10.4233/uuid:0010fdac-32ec-459b-bb9b-3e6327a85496

32. Jenny, P., Lee, S.H., Tchelepi, H.A.: Multi-scale finite-volume method for elliptic problems in subsurface flow simulation. J. Comput. Phys. **187**(1), 47–67 (2003)

33. Kraaijevanger, J.F.B.M., Egberts, P.J.P., Valstar, J.R., Buurman, H.W.: Optimal waterflood design using the adjoint method. In: SPE Reservoir Simulation Symposium. Society of Petroleum Engineers. https://doi.org/10.2118/105764-MS (2007)

34. Landrø, M.: Discrimination between pressure and fluid saturation changes from time-lapse seismic data. Geophysics **66**(3), 836–844 (2001)

35. Le Ravalec, M., Tillier, E., Da Veiga, S., Enchéry, G., Gervais, V.: Advanced integrated workflows for incorporating both production and 4d seismic-related data into reservoir models. Oil Gas Sci. Technol.–Revue d'IFP Energies nouvelles **67**(2), 207–220 (2012)

36. Li, Z., McWilliams, J.C., Ide, K., Farrara, J.D.: A multiscale variational data assimilation scheme: formulation and illustration. Mon. Weather. Rev. **143**(9), 3804–3822 (2015)

37. Liu, N., Oliver, D.S., et al.: Evaluation of monte carlo methods for assessing uncertainty. SPE J. **8**(02), 188–195 (2003)

38. Lumley, D., Meadows, M., Cole, S., Adams, D.: Estimation of reservoir pressure and saturations by crossplot inversion of 4d seismic attributes. In: SEG Technical Program Expanded Abstracts 2003, pp. 1513–1516. Society of Exploration Geophysicists (2003)

39. MacBeth, C., Floricich, M., Soldo, J.: Going quantitative with 4d seismic analysis. Geophys. Prospect. **54**(3), 303–317 (2006)

40. Mannseth, T., Fossum, K.: Assimilating spatially dense data for subsurface applications–balancing information and degrees of freedom. Comput. Geosci. **22**(5), 1323–1349 (2018)

41. Moraes, R., Rodrigues, J., Hajibeygi, H., Jansen, J.D.: Multiscale gradient computation for multiphase flow in porous media. In: SPE Reservoir Simulation Conference. Society of Petroleum Engineers (2017)

42. Moraes, R.J.D., Rodrigues, J.R., Hajibeygi, H., Jansen, J.D.: Multiscale gradient computation for flow in heterogeneous porous media. J. Comput. Phys. **336**, 644–663 (2017)

43. Møyner, O., Lie, K.A.: A multiscale restriction-smoothed basis method for high contrast porous media represented on unstructured grids. J. Comput. Phys. **304**, 46–71 (2016)

44. Nocedal, J., Wright, S.: Numerical optimization. Springer Science & Business Media (2006)

45. Oliver, D.S., He, N., Reynolds, A.C.: Conditioning permeability fields to pressure data. In: ECMOR V-5th European Conference on the Mathematics of Oil Recovery (1996)

46. Oliver, D.S., Reynolds, A.C., Liu, N.: Inverse Theory for Petroleum Reservoir Characterization and History Matching. Cambridge University Press, Cambridge (2008)

47. Rodrigues, J.R.P.: Calculating derivatives for automatic history matching. Comput. Geosci. **10**(1), 119–136 (2006)

48. Skjervheim, J.A., Evensen, G., Aanonsen, S.I., Ruud, B.O., Johansen, T.A., et al.: Incorporating 4d seismic data in reservoir simulation models using ensemble kalman filter. In: SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers (2005)

49. Strang, G.: Introduction to Linear Algebra. Wellesley-Cambridge Press, Wellesley (1993)

50. Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation, vol. 89. SIAM (2005)

51. Tura, A., Lumey, D.E.: Estimating pressure and saturation changes time-lapse avo data. In: SEG Technical Program Expanded Abstracts 1999, pp. 1655–1658. Society of Exploration Geophysicists (1999)

52. Ullmann De Brito, D., Caletti, L., Moraes, R., et al.: Incorporation of 4d seismic in the re-construction and history matching of marlim sul deep water field flow simulation model. In: SPE EUROPEC/EAGE Annual Conference and Exhibition. Society of Petroleum Engineers (2011)

53. Ullmann De Brito, D., Moraes, R., Emerick, A.A., et al.: The marlim field: incorporating time-lapse seismic in the assisted history matching. In: SPE Latin American and Caribbean Petroleum Engineering Conference. Society of Petroleum Engineers (2010)

54. Wallis, J., Tchelepi, H.A.: Apparatus, method and system for improved reservoir simulation using an algebraic cascading class linear solver. US Patent 7,684,967 (2010)

55. Wan, J., Zabaras, N.: A bayesian approach to multiscale inverse problems using the sequential monte carlo method. Inverse Prob. **27**(10), 105,004 (2011)

56. Wang, Y., Hajibeygi, H., Tchelepi, H.A.: Algebraic multiscale solver for flow in heterogeneous porous media. J. Comput. Phys. **259**, 284–303 (2014)

57. Wang, Y., Hajibeygi, H., Tchelepi, H.A.: Monotone multiscale finite volume method. Comput. Geosci. **20**(3), 509–524 (2016)

58. Zhang, Y., Oliver, D.S., et al.: History matching using the ensemble kalman filter with multiscale parameterization: a field case study. SPE J. **16**(02), 307–317 (2011)

59. Zhou, H., Tchelepi, H.A.: Operator-based multiscale method for compressible flow. SPE J. **13**(02), 523–539 (2008)

60. Zhou, H., Tchelepi, H.A., et al.: Two-stage algebraic multiscale linear solver for highly heterogeneous reservoir models. SPE J. **17**(02), 523–539 (2012)

Affiliations

Rafael J. de Moraes[1,2] (ID) · **Hadi Hajibeygi[1]** · **Jan Dirk Jansen[1]**

Hadi Hajibeygi
h.hajibeygi@tudelft.nl

Jan Dirk Jansen
j.d.jansen@tudelft.nl

1   Department of Geoscience and Engineering, Faculty of Civil Engineering and Geosciences, Delft University of Technology, P.O. Box 5048, 2600, Delft, The Netherlands

2   Petrobras Research and Development Center – CENPES, Av. Horácio Macedo 950, Cidade Universitária, Rio de Janeiro, RJ, 21941-915, Brazil