Delft University of Technology

A Multiscale Method For Data Assimilation

Jesus de Moraes, Rafael; Hajibeygi, Hadi; Jansen, Jan Dirk

DOI
10.3997/2214-4609.201802230

Publication date
2018

Document Version
Final published version

Published in
16th European Conference on the Mathematics of Oil Recovery, ECMOR 2018

Citation (APA)
Jesus de Moraes, R., Hajibeygi, H., & Jansen, J. D. (2018). A Multiscale Method For Data Assimilation. In D. Gunasekera (Ed.), *16th European Conference on the Mathematics of Oil Recovery, ECMOR 2018* Article Th A2 03 EAGE. https://doi.org/10.3997/2214-4609.201802230

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Green Open Access added to TU Delft Institutional Repository

'You share, we take care!' – Taverne project

https://www.openaccess.nl/en/you-share-we-take-care

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.

Th A2 03

A Multiscale Method For Data Assimilation

R. Moraes* (Delft University of Technology), H. Hajibeygi (Delft University of Technology), J.D. Jansen (Delft University of Technology)

Summary

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 down/upscaling of the data and/or the simulation model can be used, but with potential loss of important information. 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 down/upscaling of any quantity. The method relies on a recently developed mathematical framework to compute adjoint gradients via a MS strategy. The fine-scale uncertain parameters are directly updated and the MS grid is constructed in a resolution that meets the observed data resolution. The advantages of the technique are demonstrated in the 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 performance of the method is demonstrated in synthetic models and compared to down/upscaling strategies. The experiments show that the MS strategy provides advantages 1) on the computational side – expensive operations are only performed at the coarse scale; 2) with respect to accuracy – the matched uncertain parameter distribution is closer to the "truth"; and 3) in the optimization performance – faster convergence behaviour due to faster gradient computation. In conclusion, the newly developed method is capable of providing superior results when compared to strategies that rely on the up/downscaling of the response/observed data, addressing the scale dissimilarity via a robust, consistent MS strategy.

Introduction

Reservoir 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 (Emerick et al., 2007b). For instance, over the past decades, an increasing number of seismic monitoring cases has been observed (Ullmann De Brito et al., 2010, 2011). 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 in the reservoir drainage process. Moreover, it may help to characterize the formation, either via improved static geological modelling, or via dynamic assimilation (computer-assisted history matching). One hurdle in the process of assimilating spatially distributed information is the fact that, more often than 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 (Emerick et al., 2007b). The literature indicates that up/downscaling of the observed data to the simulation model scale is the most employed strategy (Gosselin et al., 2003; Gervais-Couplet et al., 2010; Le Ravalec et al., 2012).

Data assimilation and uncertainty quantification (UQ) are computationally demanding. Different techniques such as Monte Carlo (MC) methods (Liu et al., 2003), Ensemble Kalman Filter (EnKF) and derivations (Aanonsen et al., 2009; Evensen, 2009; Emerick and Reynolds, 2013a) and randomized maximum likelihood (Oliver et al., 1996) are developed to perform those studies. A comparison between the different techniques is provided by Emerick and Reynolds (2013b). 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. Upscaling (Durlofsky, 2003; Farmer, 2002), 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.

Alternatively, multiscale (MS) methods can be used (Jenny et al., 2003; Hou and Wu, 1997). These are efficient simulation strategies capable of solving the flow problem in a coarser grid, while being capable of accurately representing fine scale heterogeneities. Fu et al. (2010, 2011) developed a MS adjoint method for single-phase flow in porous media. The most efficient data assimilation methods are gradient-based, with gradients obtained with the adjoint method Oliver et al. (2008). Krogstad et al. (2011) and Moraes et al. (2017a) present multiscale gradient computation strategies for multiphase flow, the former applied to a production optimization problem, and the latter to data assimilation problems. Also, the latter is based on the general framework for MS gradient computation presented by Moraes et al. (2017b), whose algebraic nature does not rely on any assumption regarding the nature of the parameters, observations, or objective function type.

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) to seamlessly conciliate model and observed data scales. For a comprehensive review on the recent developments associated with MS methods applied to reservoir simulation, see Li et al. (2015).

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 Moraes et al. (2017b). 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 randomized maximum likelihood (RML) method. A discussion about the results and the challenges that the method can encounter are presented next. Finally, a summary of the developments and results, as well as future research perspectives, are presented.

Preliminaries

We base our developments on a Bayesian framework. According to Bayes' theorem, the posterior probability distribution function (PDF) can be computed as

$$f(\theta|\mathbf{d}_{obs}) = \frac{f(\mathbf{d}_{obs}|\theta)\, f(\theta)}{f(\mathbf{d}_{obs})}, \tag{1}$$

and, if the *a priori* PDF of the uncertain parameters, $f(\theta)$, and the measurement errors from the observations $\mathbf{d}_{obs}$ are assumed Gaussian, it can be shown that the conditional *a posteriori* distribution is given by (Tarantola, 2005)

$$f(\theta|\mathbf{d}_{obs}) \propto \exp\left(-O\left(\mathbf{y}(\mathbf{x},\theta),\theta\right)\right), \tag{2}$$

where the objective function O is given by

$$\begin{aligned} O\left(\mathbf{y}(\mathbf{x},\theta),\theta\right) = & \frac{1}{2}\left(\theta-\theta_{prior}\right)^T \mathbf{C}_\theta^{-1}\left(\theta-\theta_{prior}\right) + \\ & \frac{1}{2}\left(\mathbf{y}(\mathbf{x},\theta)-\mathbf{d}_{obs}\right)^T \mathbf{C}_D^{-1}\left(\mathbf{y}(\mathbf{x},\theta)-\mathbf{d}_{obs}\right). \end{aligned} \tag{3}$$

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

The solution of Eq. (3) can be stated as an optimization problem

$$\begin{aligned} \underset{\theta}{\text{minimize}} \quad & O\left(\mathbf{y}(\mathbf{x},\theta)\right) \\ \text{subject to} \quad & \mathbf{g}(\mathbf{x},\theta) = \mathbf{0}, \\ & \theta \in [\theta_{min}, \theta_{max}], \end{aligned} \tag{4}$$

where $\mathbf{g} : \mathbb{R}^{N_F} \times \mathbb{R}^{N_\theta} \to \mathbb{R}^{N_F}$ represent the set of forward model equations and $\theta_{min}, \theta_{max}$ are the parameter bounds. The efficient solution of Eq. (4) requires gradient-based methods (Nocedal and Wright, 2006) combined with efficient gradient computation methods. For this purpose, from the chain rule, the gradient of Eq. (3) is given by

$$\nabla_\theta O = \left(\frac{dO}{d\theta}\right)^T = \left(\frac{dO}{d\mathbf{y}}\frac{d\mathbf{y}}{d\theta}\right)^T = \mathbf{G}^T \nabla_\mathbf{y} O, \tag{5}$$

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 (Oliver et al., 2008). Rodrigues (2006) discusses how the right-multiplication of $\mathbf{G}$ by an arbitrary vector (as in Eq. (5)) can be efficiently computed via the adjoint method.

The Forward Model

The set of discretized equations that describes the forward simulation at the fine scale can be algebraically expressed as (Moraes et al., 2017b)

$$\mathbf{g}_F(\mathbf{x},\theta) = \mathbf{0}, \tag{6}$$

where $\mathbf{g}_F : \mathbb{R}^{N_F} \times \mathbb{R}^{N_\theta} \to \mathbb{R}^{N_F}$ represents the set of algebraic forward model equations, $\mathbf{x} \in \mathbb{R}^{N_F}$ is the state vector (which, for single-phase flow, contains the grid block pressures) and the subscript F refers to 'fine scale'. There are N_F fine-scale cells and N_θ parameters. Eq. (6) implicitly assumes a dependency of the state vector $\mathbf{x}$ on the parameters θ, i.e.

$$\mathbf{x} = \mathbf{x}(\theta). \tag{7}$$

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

$$\mathbf{y}_F = \mathbf{h}_F(\mathbf{x},\theta), \tag{8}$$

where $\mathbf{h}_F : \mathbb{R}^{N_F} \times \mathbb{R}^{N_\theta} \to \mathbb{R}^{N_y}$ represents the output equations Jansen (2013b). It is assumed that $\mathbf{g}_F$ can be described as

$$\mathbf{g}_F(\mathbf{x},\theta) = \mathbf{A}(\theta)\mathbf{x} - \mathbf{q}(\theta), \tag{9}$$

where $\mathbf{A} = \mathbf{A}(\theta) \in \mathbb{R}^{N_F} \times \mathbb{R}^{N_F}$ represents the system equations and $\mathbf{q} = \mathbf{q}(\theta) \in \mathbb{R}^{N_F}$ is a vector of source terms.

Multiscale Simulation

A two-stage multiscale (MS) solution strategy Zhou and Tchelepi (2008); Wang et al. (2014) can be devised by firstly computing a coarse scale solution

$$\check{\mathbf{g}}(\check{\mathbf{x}},\theta) = (\mathbf{RAP})\check{\mathbf{x}} - (\mathbf{Rq}) = \check{\mathbf{A}}\check{\mathbf{x}} - \check{\mathbf{q}} = \check{\mathbf{0}}, \tag{10}$$

where-after an approximate fine-scale solution is formed as

$$\mathbf{g}'\left(\mathbf{x}',\check{\mathbf{x}},\theta\right) = \mathbf{x}' - \mathbf{P}\check{\mathbf{x}} = \mathbf{0}. \tag{11}$$

Let $\check{\mathbf{x}} \in \mathbb{R}^{N_C}$ 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 $\mathbf{P} = \mathbf{P}(\theta)$ is an $N_F \times N_C$ matrix that maps (interpolates) the coarse-scale solution to the fine-scale resolution, where N_C is the number of coarse grid-blocks. The restriction operator $\mathbf{R} = \mathbf{R}(\theta)$ is defined as an $N_C \times N_F$ matrix which maps the fine scale to the coarse scale. More information about how these operators are constructed for the Multiscale Finite Volume (MSFV) method can be found in Zhou and Tchelepi (2008) and Wang et al. (2014).

Data Assimilation Problem Setup

The maximum *a posteriori* probability (MAP, (Oliver et al., 2008)) of the uncertain parameters is obtained by solving the optimization problem stated in Eq. (4), with the objective function (OF) given by Eq. (3) (and using Eq. (8)), the gradient of which is given by

$$\begin{aligned}\nabla_\theta O = \mathbf{C}_\theta^{-1}(\theta - \theta_{prior}) + \left(\frac{d\mathbf{h}}{d\theta}\right)^T \mathbf{C}_D^{-1}(\mathbf{h}(\mathbf{x},\theta) - \mathbf{d}_{obs}) = \\ \mathbf{C}_\theta^{-1}(\theta - \theta_{prior}) + \mathbf{G}^T\mathbf{m}.\end{aligned} \tag{12}$$

Following an implicit differentiation strategy (Rodrigues, 2006; Kraaijevanger et al., 2007; Moraes et al., 2017b), the and sensitivity matrix $\mathbf{G}$ can be computed as

$$\mathbf{G} = \frac{\partial\mathbf{h}}{\partial\mathbf{x}}\frac{d\mathbf{x}}{d\theta} + \frac{\partial\mathbf{h}}{\partial\theta} = -\frac{\partial\mathbf{h}}{\partial\mathbf{x}}\left(\frac{\partial\mathbf{g}}{\partial\mathbf{x}}\right)^{-1}\frac{\partial\mathbf{g}}{\partial\theta} + \frac{\partial\mathbf{h}}{\partial\theta}, \tag{13}$$

where the partial derivative matrices are obtained from the forward model and output equations, and

$$\mathbf{m} = \mathbf{C}_D^{-1}\left(\mathbf{h}\left(\mathbf{x},\theta\right) - \mathbf{d}_{obs}\right). \tag{14}$$

The product $\mathbf{G}^T\mathbf{m}$ is solved with cost of one backward simulation, regardless the number of model parameters, via the Adjoint method.

Randomized Maximum Likelihood (RML) RML (Oliver et al., 1996) is an approximated sampling method for UQ, which obtains the j-th sample of the posterior PDF distribution by solving Eq. (4) for a given sample $\theta_{uc,j}$ from $\mathcal{N}\left(\theta_{prior}, \mathbf{C}_\theta\right)$ and a given sample $\mathbf{d}_{obs,uc,j}$ from $\mathcal{N}\left(\mathbf{d}_{obs}, \mathbf{C}_D\right)$. Therefore, Eq. (3) can be re-written for $O_j\left(\mathbf{h},\theta\right)$ as

$$\begin{aligned} O_j\left(\mathbf{h},\theta\right) = & \frac{1}{2}\left(\theta-\theta_{uc,j}\right)^T \mathbf{C}_\theta^{-1}\left(\theta-\theta_{uc,j}\right) + \\ & \frac{1}{2}\left(\mathbf{h}\left(\mathbf{x},\theta\right) - \mathbf{d}_{obs,uc,j}\right)^T \mathbf{C}_D^{-1}\left(\mathbf{h}\left(\mathbf{x},\theta\right) - \mathbf{d}_{obs,uc,j}\right). \end{aligned} \tag{15}$$

Hence, a minimization problem has to be solved for every j-th posterior PDF sample one wants to estimate.

Conciliation of Spatially Distributed Data and Forward Model Scales

In the data assimilation of spatially distributed observations, Eq. (12) assumes that the observations $\mathbf{d}_{obs}$ 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 avaialble at the fine scale description of the model parameters. Therefore, if no MS simulation is available, either $\mathbf{d}_{obs}$ 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

$$\mathbf{d}'_{obs} = \mathbf{D}\check{\mathbf{d}}_{obs}, \tag{16}$$

where $\mathbf{D}$ is an $N_F \times N_C$ downscaling operator, $\check{\mathbf{d}}_{obs}$ is the coarse scale observation and $\mathbf{d}'$ is the approximated observation at the fine scale.

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

$$\mathbf{D} = \mathbf{R}^T. \tag{17}$$

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 (Jenny et al., 2003), and prolong (interpolate) the coarse scale information by setting

$$\mathbf{D} = \mathbf{P}_{prior}. \tag{18}$$

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

In the aforementioned strategies, Eq. (3) can be used by making $\mathbf{d}_{obs} = \mathbf{d}'_{obs}$ and a conventional gradient-based optimization to solve Eq. (4) is run at the model (fine) scale. However, this is only true if one can describe the data covariance matrix $\mathbf{C}_D$, originally at (coarse) observation scale, at the fine scale. This can be achieved by making

$$\mathbf{C}'_D = \mathbf{D}\check{\mathbf{C}}_D\mathbf{D}^T, \tag{19}$$

where $\mathbf{C}'_D$ is the covariance matrix represented at the fine scale. It is simple to show that Eq. (19) holds because of the linearity of the expectation operator. From Eq. (16), the expectation of $\mathbf{d}'_{obs}$ is given by

$$E[\mathbf{d}'_{obs}] = \mathbf{D}E[\check{\mathbf{d}}_{obs}]. \tag{20}$$

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

$$\begin{aligned} Cov[\mathbf{d}'_{obs}] &= E[(\mathbf{d}'_{obs} - E[\mathbf{d}'_{obs}])(\mathbf{d}'_{obs} - E[\mathbf{d}'_{obs}])^T] \\ &= E[\mathbf{D}(\breve{\mathbf{d}}_{obs} - E[\breve{\mathbf{d}}_{obs}])(\breve{\mathbf{d}}_{obs} - E[\breve{\mathbf{d}}_{obs}])^T \mathbf{D}^T] \\ &= \mathbf{D}E[(\breve{\mathbf{d}}_{obs} - E[\breve{\mathbf{d}}_{obs}])(\breve{\mathbf{d}}_{obs} - E[\breve{\mathbf{d}}_{obs}])^T]\mathbf{D}^T \\ &= \mathbf{D}Cov[\breve{\mathbf{d}}_{obs}]\mathbf{D}^T \end{aligned} \tag{21}$$

Alternatively, one could upscale the model responses as

$$\breve{\mathbf{d}}_{obs} = \mathbf{U}\mathbf{h}, \tag{22}$$

where $\mathbf{U}$ is an $N_C \times N_F$ upscaling operator, and solve Eq. (3) 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. Two upscaling strategies are considered. In the first one, a simple arithmetic average is applied by setting

$$\mathbf{U} = \mathbf{M}, \tag{23}$$

where,

$$\mathbf{M}(c,f) = \begin{cases} \dfrac{1}{N_F^C}, & \text{if } f \in \Omega_c \\ 0, & \text{otherwise} \end{cases}. \tag{24}$$

In Eq. (24), N_F^C is the number of fine grid cells within a given coarse cell C, Ω_c is the c-th primal coarse grid domain and f is the fine-grid cell index.

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

$$\mathbf{U} = \mathbf{P}^{\dagger}, \tag{25}$$

where the † symbol denotes for the Moore-Penrose inverse. Here, we construct $\mathbf{P}^{\dagger}$ from its truncated singualar value decomposition (TSVD) (Strang, 1993)

$$\mathbf{P} = \Sigma\Lambda\Delta^{\mathbf{T}} = \begin{bmatrix}\Sigma_{\mathbf{p}} & \Sigma_{\mathbf{0}}\end{bmatrix} \begin{bmatrix}\Lambda_{\mathbf{p}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} \begin{bmatrix}\Delta_{\mathbf{p}}^{\mathbf{T}} \\ \Delta_{\mathbf{0}}^{\mathbf{T}}\end{bmatrix} = \Sigma_{\mathbf{p}}\Lambda_{\mathbf{p}}\Delta_{\mathbf{p}}^{\mathbf{T}} \tag{26}$$

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

We highlight that we only consider strategies that change observed data / responses scale and do not consider strategies that change the original uncertain parameters 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.

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-writting Eq. (3) as

$$O_{MS}\left(\breve{\mathbf{h}},\theta\right) = \frac{1}{2}(\theta-\theta_{prior})^T\mathbf{C}_\theta^{-1}(\theta-\theta_{prior}) + \frac{1}{2}\left(\breve{\mathbf{h}}(\breve{\mathbf{x}},\theta)-\breve{\mathbf{d}}_{obs}\right)^T\mathbf{C}_D^{-1}\left(\breve{\mathbf{h}}(\breve{\mathbf{x}},\theta)-\breve{\mathbf{d}}_{obs}\right), \tag{27}$$

where $\breve{\mathbf{h}}$ is the response at the (coarse) observation scale and $\breve{\mathbf{x}}$ is the coarse state variable, computed by Eq. (10). Hence, the misfit term is computed at the coarse scale – the scale where data is assimilated – and the regularization term is described in the fine scale – the scale where the model parameters are described.

Multiscale Gradient Computation

As discussed in Moraes et al. (2017b), the state vector can be described as a combination of both sets of primary variables at the fine and coarse scales, i.e.,

$$\mathbf{x} = \begin{bmatrix} \breve{\mathbf{x}} \\ \mathbf{x}' \end{bmatrix}, \tag{28}$$

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

$$\mathbf{g}(\mathbf{x},\theta) = \begin{bmatrix} \breve{\mathbf{g}} \\ \mathbf{g}' \end{bmatrix} = \mathbf{0}. \tag{29}$$

The definition of the state vector as in Eq. (28) 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

$$\breve{\mathbf{y}} = \breve{\mathbf{h}}(\breve{\mathbf{x}},\theta), \tag{30}$$

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

$$\mathbf{G}' = \left(\frac{\partial\breve{\mathbf{h}}}{\partial\breve{\mathbf{x}}} + \frac{\partial\breve{\mathbf{h}}}{\partial\mathbf{x}'}\mathbf{P}\right)(\mathbf{RAP})^{-1}\mathbf{R}\left(\frac{\partial\mathbf{A}}{\partial\theta}\mathbf{P} + \mathbf{A}\frac{\partial\mathbf{P}}{\partial\theta}\right)\breve{\mathbf{x}} - \frac{\partial\breve{\mathbf{h}}}{\partial\mathbf{x}'}\frac{\partial\mathbf{P}}{\partial\theta}\breve{\mathbf{x}} + \frac{\partial\breve{\mathbf{h}}}{\partial\theta}. \tag{31}$$

Eq. (31) allows capturing derivative informations of the coarse response, namely $\frac{\partial\breve{\mathbf{h}}}{\partial\breve{\mathbf{x}}}$ and $\frac{\partial\breve{\mathbf{h}}}{\partial\theta}$.

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, $\frac{\partial\mathbf{P}}{\partial\theta}$ in Eq. (31), 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. (27) w.r.t. the model parameters at fine scale can be computed as

$$\nabla_\theta O_{MS} = \mathbf{C}_\theta^{-1}(\theta-\theta_{prior}) + \left(\frac{d\breve{\mathbf{h}}}{d\theta}\right)^T\mathbf{C}_D^{-1}\left(\breve{\mathbf{h}}(\breve{\mathbf{x}},\theta)-\breve{\mathbf{d}}_{obs}\right) = \mathbf{C}_\theta^{-1}(\theta-\theta_{prior}) + \mathbf{G}'^T\breve{\mathbf{m}}, \tag{32}$$

where

$$\breve{\mathbf{m}} = \mathbf{C}_D^{-1}\left(\breve{\mathbf{h}}(\breve{\mathbf{x}},\theta)-\breve{\mathbf{d}}_{obs}\right). \tag{33}$$

The product $\mathbf{G}'^T\breve{\mathbf{m}}$ can be accurately and efficiently computed via the MS adjoint method (Moraes et al., 2017b).

Numerical Experiments

We focus our analysis to incompressible, single phase-flow. In our experiments, a simple synthetic model is considered as proof of concept. It is a 2D inverted five-spot model, consisting of a 21×21 equidistant Cartesian mesh with grid block size dimensions 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 (Jansen, 2013a). Fig. 1 illustrates four different permeability realizations from the ensemble of 1,000 equiprobable permeability realizations. In order to focus on the MS aspect of the data assimilation process, we assume that pressures

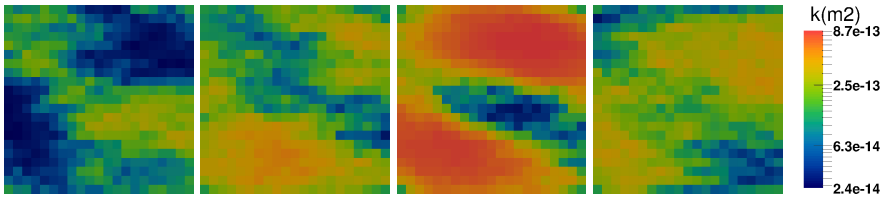


Figure 1 *Four different permeability realizations from the ensemble of 1,000 members of the toy model numerical experiment.*

can be approximately extracted from time-lapse seismic survey (Tura and Lumey, 1999; Landrø, 2001; Cole et al., 2002; Lumley et al., 2003; MacBeth et al., 2006). 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 (Emerick et al., 2007b), the additional complexity involved is the appropriate incorporation of seismic forward model equations in the forward model set of equations (Emerick et al., 2007a) and, consequentially, the computation the appropriate partial derivative information necessary to compute Eq. (12).

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. (10), while the (hypothetical, for comparison purposes) fine observed data is computed with Eq. (9). 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

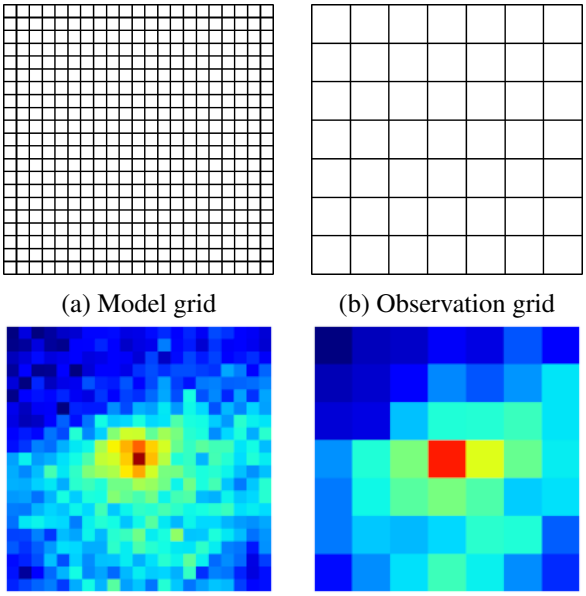
$$\mathbf{p}_{obs} = \mathbf{p}_{true} + \sqrt{\mathbf{C}_D}\mathbf{z}, \tag{34}$$

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 Oliver et al. (2008). The resulting noisy observed coarse and fine pressure fields are illustrated, respectively, in Fig. 2d and Fig. 2c.

We consider the observation grid (Fig. 2b) where the observed data is represented to be three times coarser than the model grid (Fig. 2a) 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

$$\mathbf{C}_\theta = \frac{1}{N_e - 1}\left(\Theta - \mu\mathbf{e}^T\right)\left(\Theta - \mu\mathbf{e}^T\right)^T, \tag{35}$$



(a) Model grid (b) Observation grid

(c) Fine-scale observation (d) Coarse-scale observation

Figure 2 *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).*

where Θ is the $N_F \times N_e$ matrix whose j-th column is given by the member of the ensemble $\theta_j, j \in \{1,...,N_e\}$,

$$\mu = \frac{1}{N_e}\sum_{j=1}^{N_e}\theta_j, \tag{36}$$

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,

$$\theta_{prior} = \mu. \tag{37}$$

In the fine scale data assimilation strategy, an adjoint model (Rodrigues, 2006) is used to calculate the OF gradient given by Eq. (3). In the multiscale strategy, we employ the MS adjoint gradient computation depicted in Algorithm 4 in Moraes et al. (2017b). Because the spatially distributed observed data at the coarse scale is the primary variable itself, in Eq. (27) we have

$$\frac{\partial \breve{\mathbf{h}}}{\partial \breve{\mathbf{x}}} = \breve{\mathbf{I}}, \tag{38}$$

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

$$\frac{\partial \breve{\mathbf{h}}}{\partial \mathbf{x}'} = \mathbf{0}. \tag{39}$$

when pressure is observed at the coarse scale. And when pressure is observed at the fine-scale,

$$\frac{\partial \mathbf{h}}{\partial \mathbf{x}} = \mathbf{I}, \tag{40}$$

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

$$\frac{\partial \breve{\mathbf{h}}}{\partial \theta} = \breve{\mathbf{0}}, \tag{41}$$

and

$$\frac{\partial \mathbf{h}}{\partial \theta} = \mathbf{0}. \tag{42}$$

We utilize the limited memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithm as presented in Nocedal and Wright (2006), as it is the most efficient algorithm to deliver optimization results for the solution of Eq. (4) (Zhang and Reynolds, 2002).

Maximum a posteriori Probability (MAP) Estimate

In this section we assess the performance of computing the MAP estimate via the newly introduced method in comparison the to down/upscaling strategies discussed before. Therefore, six different data assimilation strategies are considered, namely:

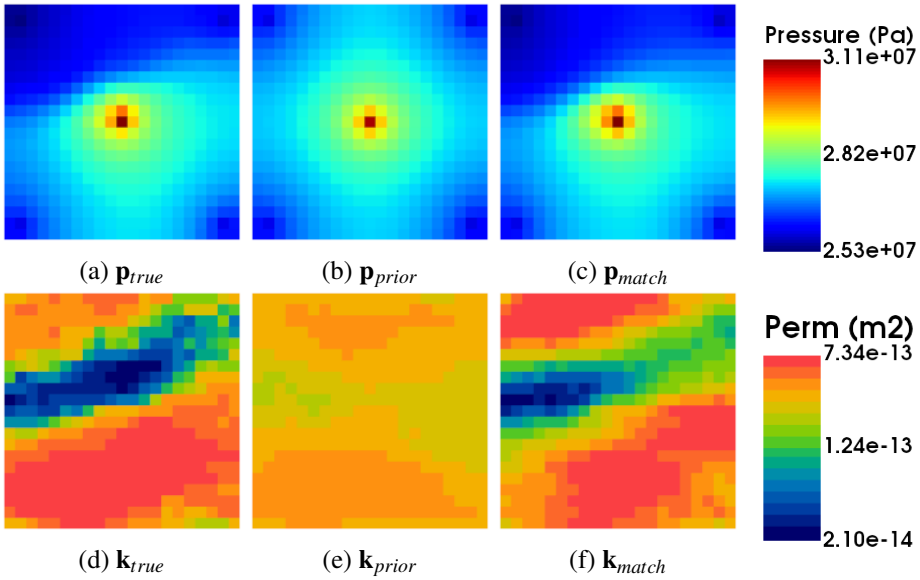
1. fine-scale data assimilation with constant prolongation downscaling of observed pressure (Eq. (17));
2. fine-scale data assimilation with prior MS prolongation downscaling of observed pressure (Eq. (18));
3. fine-scale data assimilation with arithmetic average to upscale the simulated pressure (Eq. (24));
4. fine-scale data assimilation with pseudo-inverse of MS prolongation to upscale tje simulated pressure (Eq. (25));
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 Tarantola (2005); Oliver et al. (2008), an acceptable data match is achieved when $\frac{O}{N_d} \approx 0.5$.

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. 3, followed by the results from the downscaling and upscaling data conciliation strategies, represented, respectively, by Fig. 4 and Fig. 5. Lastly, the results of the data assimilation using our MS data assimilation strategy are illustrated in Fig. 6.

From a qualitative point of view, the matched responses from all data assimilation strategies, except the responses obtained by the constant interpolation downscaling (Fig. 4g) and arithmetic average upscaling (Fig. 5g), are fairly similar to the observed data. The pressure matches are both in accordance with the fine-scale pressure match (Fig. 3f) and with the 'true'pressure field. However, the simpler up/downscaling strategies result in somewhat poorer matches around the injection well.

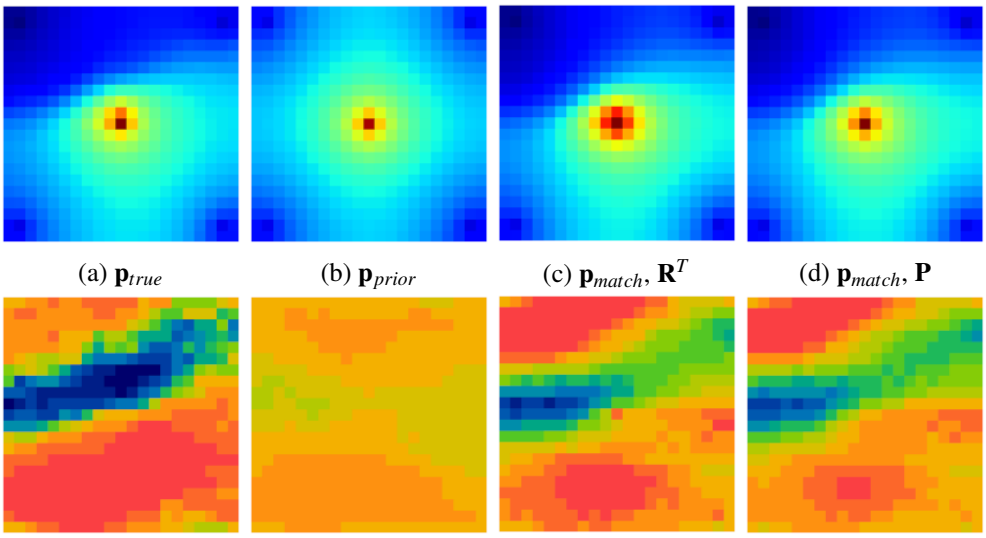
Also, it is possible to observe that, from the point of view of the conditioned permeability fields, all assimilation strategies were capable of recovering the main features. Furthermore, not much difference is noted in the results when comparing the upscaling to the downscaling matched permeability 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. 7. It is possible to note that, even though the initial permeability distribution



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

(d) $\mathbf{k}_{true}$ (e) $\mathbf{k}_{prior}$ (f) $\mathbf{k}_{match}$

Figure 3 *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.*

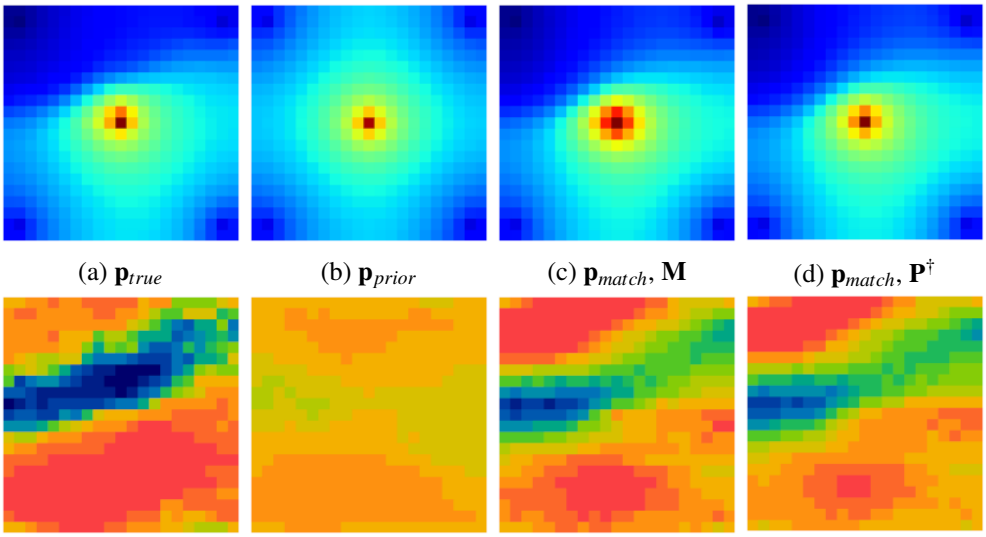


(a) $\mathbf{p}_{true}$ (b) $\mathbf{p}_{prior}$ (c) $\mathbf{p}_{match}$, $\mathbf{R}^T$ (d) $\mathbf{p}_{match}$, $\mathbf{P}$

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

Figure 4 *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 ($\boldsymbol{R}^T$) and (d) matched using the MS prolongation operator ($\boldsymbol{P}$) pressure fields are shown. In the second row, (d) the 'true', (e) the prior, (g) the conditioned using $\boldsymbol{R}^T$ and (g) the conditioned using $\boldsymbol{P}$ permeability fields are shown. The color maps follow the color bar found in Fig. 3*

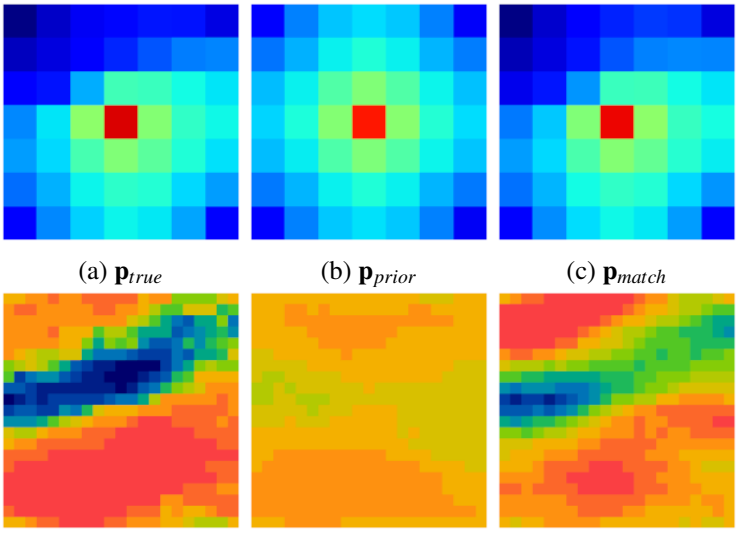
.



(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) $\mathbf{k}_{match}$, $\mathbf{P}^{\dagger}$

Figure 5 *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. 3.*



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

(d) $\mathbf{k}_{true}$ (e) $\mathbf{k}_{prior}$ (f) $\mathbf{k}_{match}$

Figure 6 *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. 3. Note that MS assimilation provides a fine-scale improved permeability field (f) while the observation is at coarse scale.*

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.

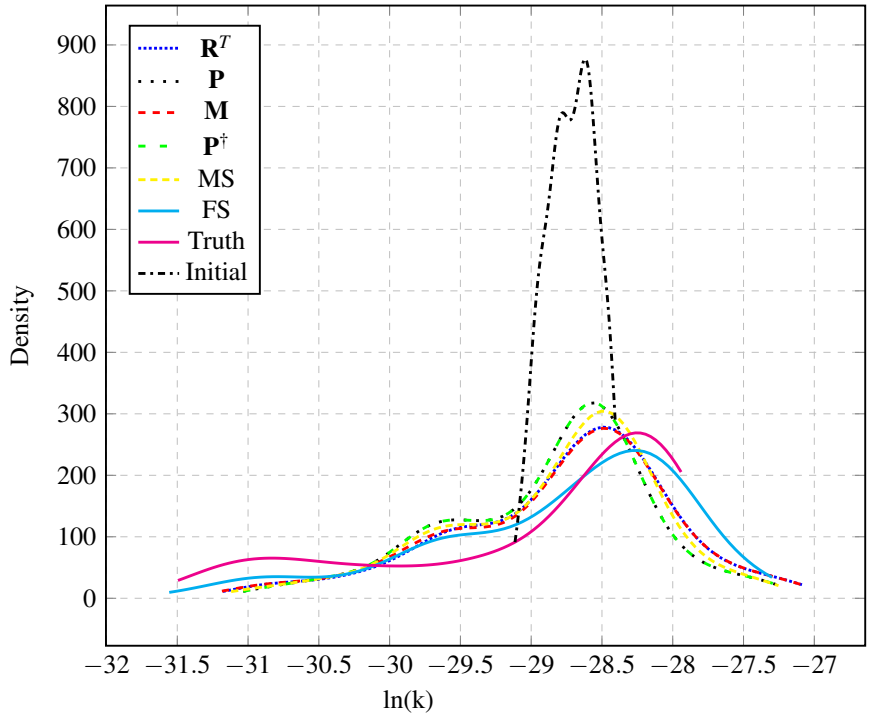


Figure 7 *Permeability conditioned marginal PDFs for the different data assimilation strategies.*

We also analyse the optimization convergence behaviour shown in Fig. 8 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 behaviour is remarkably similar for all up/downscaling strategies, as well as for the MS data assimilation strategy here presented.

Uncertainty Quantification

A RML is run for 100 randomly chosen permeability realization from the 1,000 members ensemble (Fig. 1), 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. 9.

It can be observed that the permeability marginal PDFs conditioned to the pressure data obtained by the MS data assimilation here introduced (Fig. 9e) are closer to the reference fine-scale conditioned PDFs (Fig. 9f). Additionally, by observing the spread of the conditioned PDFs obtained from the RML employing the up/downscaling strategies, one can note that the MS strategy is also capable of somewhat better representing the uncertainty.

Discussion

Firstly, even though all data assimilation strategies were capable of achieving simular MAP estimates, one should note that the synthetic case used in the experiments has low permeability contrasts. Moreover

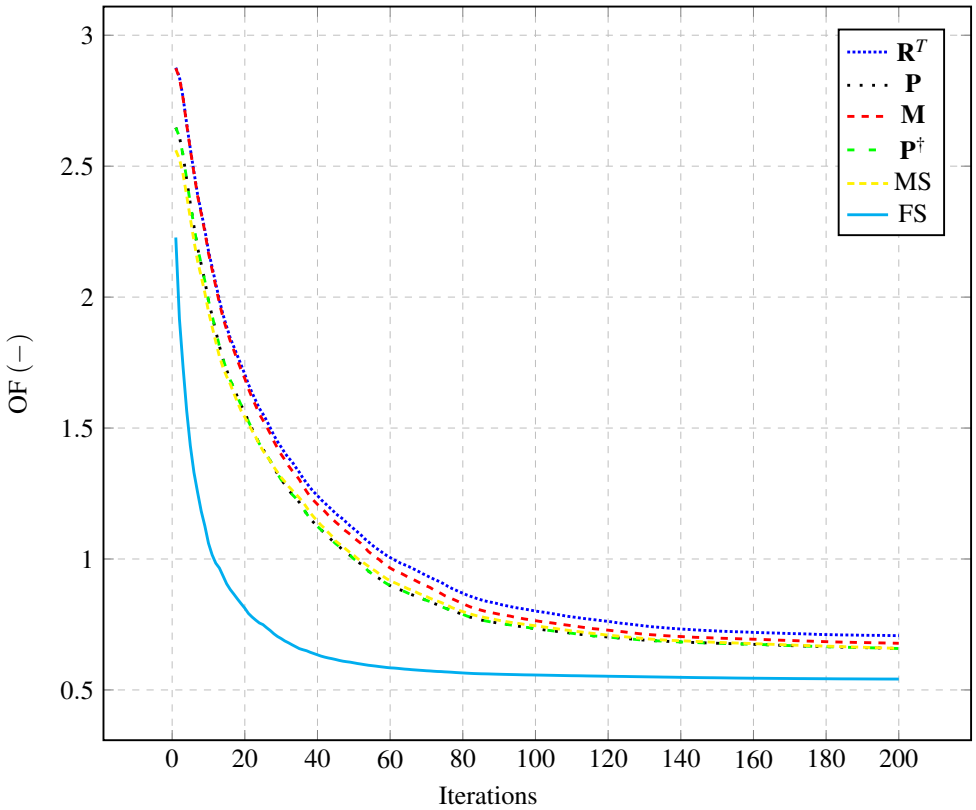
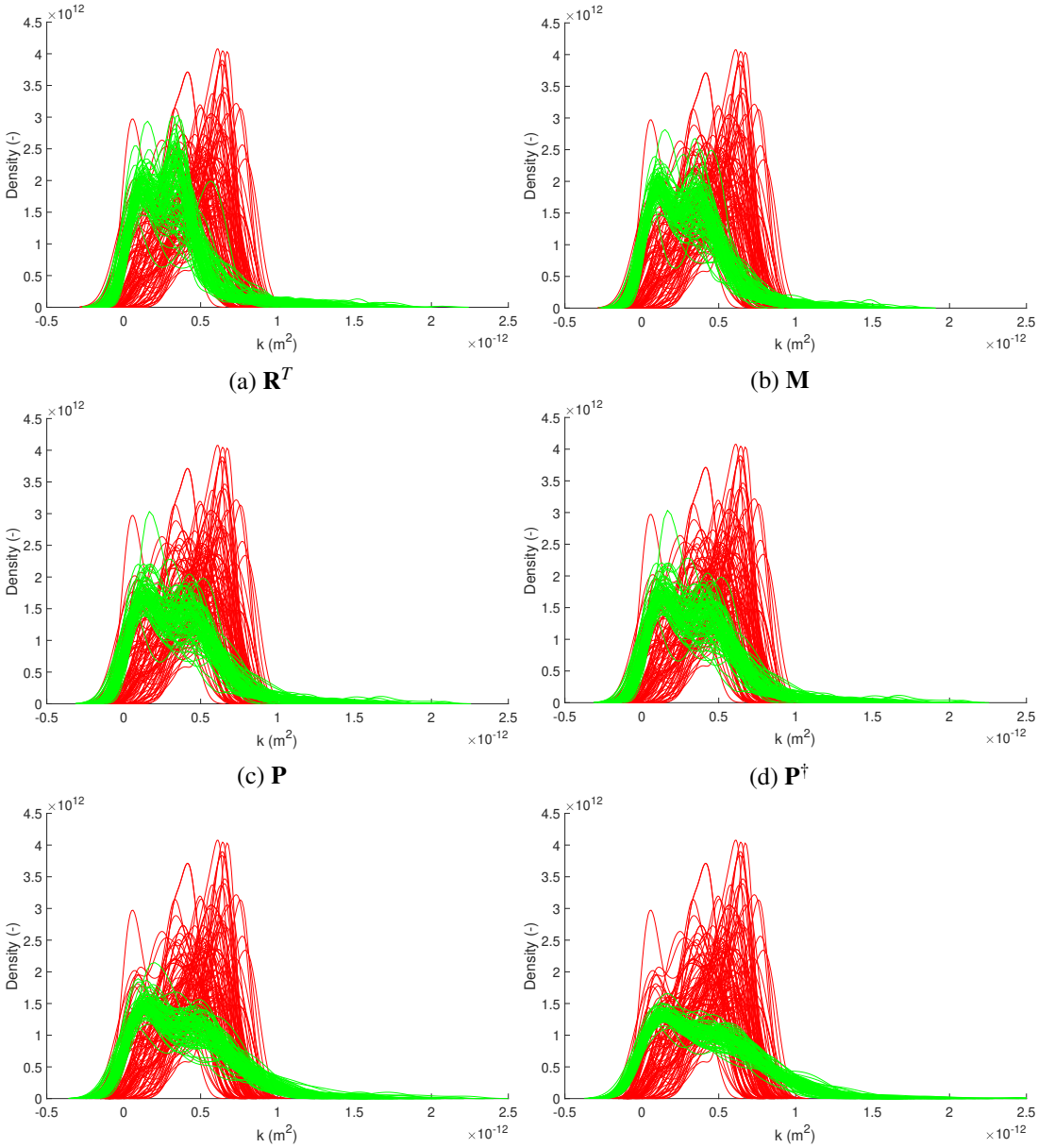


Figure 8 *Optimization performance of the data assimilation utilizing the 6 different scale conciliation strategies as presented in this section. Note that the FS represent the hypothetical case where both observed data and model parameters are at fine scale (i.e. item 6 on our list of strategies).*



(a) $\mathbf{R}^T$ (b) $\mathbf{M}$

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

(e) Full Multiscale (f) Fine scale

Figure 9 *RML probability density functions for 100 permeability realizations randomly chosen from the original ensemble of 1,000 realizations (see Fig. 1). The curves in red represent the prior permeability distributions, while the curves in green the conditioned permeability distributions. (a) Constant interpolation downscaling ($\boldsymbol{R}^T$), (b) arithmetic average downscaling ($\boldsymbol{M}$), (c) MS prolongation operator downscaling, (d) MS prolongation operator upscaling $\boldsymbol{P}^\dagger$, (e) the multiscale data assimilation strategy and (f) the reference fine-scale.*

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 errors in 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

$$\mathbf{d}'^{\gamma}_{obs} = \mathbf{P}^{\gamma}\breve{\mathbf{d}}_{obs}, \tag{43}$$

where $\mathbf{P}^{\gamma}$ is the reconstructed prolongation operator at γ. This can only be achieved at the expense of the reconstruction of the basis function every γ iteration. 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 (Eq. (18)), as reported here. Moreover, it is discussed in Moraes et al. (2017b) how to efficiently compute $\mathbf{P}/\theta$. 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 (Cusini et al., 2016) 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 Moraes et al. (2017b) and Fu et al. (2010), 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 (Gao et al., 2004), 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 (Gao et al., 2004). The latter by improved MS gradient quality, via more accurate MSFV solutions (Wang et al., 2016; Hajibeygi and Jenny, 2011). An iterative MSFV gradient computation, following the solution strategy proposed by Hajibeygi and Jenny (2011) could allow for additional error control over the gradient computation.

Final Remarks

Our simple numerical experiments indicate that the presented method outperforms the strategies that rely on the upscaling/downscaling of model responses/observed data. The examples also indicate that the method has the potential to address the problem of data dissimilarity in the data scale assimilation and is worth further investigation. Applications in more complex cases, and for different types of assimilation problems, should give more insights about the potential of the method. An important result is the ability of our MS data assimilation strategy to closely reproduce the reference fine-scale uncertainty quantification results. The experiments also show that MS simulation concepts can also provide accurate up/downscaling operators.

Acknowledgements

This work is part of the first author's PhD project, which was sponsored by Petrobras' employee training program. The insightful discussions and valuable suggestions from Dr. Alexandre A. Emerick (Petrobras Research Center) are immensely appreciated.

References

Aanonsen, S.I., Nævdal, G., Oliver, D.S., Reynolds, A.C. and Vallés, B. [2009] Review of ensemble Kalman filter in petroleum engineering. *SPE Journal*, **14**(3), 393–412.

Cole, S., Lumley, D., Meadows, M. and Tura, A. [2002] Pressure and saturation inversion of 4D seismic data by rock physics forward modeling. In: *SEG Technical Program Expanded Abstracts 2002*, Society of Exploration Geophysicists, 2475–2478.

Cusini, M., van Kruijsdijk, C. and Hajibeygi, H. [2016] Algebraic dynamic multilevel (ADM) method for fully implicit simulations of multiphase flow in porous media. *Journal of Computational Physics*, **314**, 60–79.

Durlofsky, L.J. [2003] Upscaling of geocellular models for reservoir flow simulation: a review of recent progress. In: *7th International Forum on Reservoir Simulation Bühl/Baden-Baden, Germany*. 23–27.

Emerick, A.A., Moraes, R., Rodrigues, J. et al. [2007a] Calculating seismic attributes within a reservoir flow simulator. In: *Latin American & Caribbean Petroleum Engineering Conference*. Society of Petroleum Engineers.

Emerick, A.A., Moraes, R., Rodrigues, J. et al. [2007b] History matching 4D seismic data with efficient gradient based methods. In: *EUROPEC/EAGE Conference and Exhibition*. Society of Petroleum Engineers.

Emerick, A.A. and Reynolds, A.C. [2013a] Ensemble smoother with multiple data assimilation. *Computers & Geosciences*, **55**, 3–15.

Emerick, A.A. and Reynolds, A.C. [2013b] Investigation of the sampling performance of ensemble-based methods with a simple reservoir model. *Computational Geosciences*, **17**(2), 325–350.

Evensen, G. [2009] *Data assimilation: the ensemble Kalman filter*. Springer Science & Business Media.

Farmer, C. [2002] Upscaling: a review. *International journal for numerical methods in fluids*, **40**(1-2), 63–78.

Fu, J., Caers, J. and Tchelepi, H.A. [2011] A multiscale method for subsurface inverse modeling: Single-phase transient flow. *Adv. Water Resour.*, **34**(8), 967–979.

Fu, J., Tchelepi, H.A. and Caers, J. [2010] A multiscale adjoint method to compute sensitivity coefficients for flow in heterogeneous porous media. *Advances in water resources*, **33**(6), 698–709.

Gao, G., Reynolds, A.C. et al. [2004] An improved implementation of the LBFGS algorithm for automatic history matching. In: *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers.

Gervais-Couplet, V., Roggero, F., Feraille, M.D., Ravalec-Dupin, L., Seiler, A. et al. [2010] 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.

Gosselin, O., Aanonsen, S., Aavatsmark, I., Cominelli, A., Gonard, R., Kolasinski, M., Ferdinandi, F., Kovacic, L., Neylon, K. et al. [2003] History matching using time-lapse seismic (HUTS). In: *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers.

Hajibeygi, H. and Jenny, P. [2011] Adaptive iterative multiscale finite volume method. *Journal of Computational Physics*, **230**(3), 628–643.

Hou, T.Y. and Wu, X.H. [1997] A multiscale finite element method for elliptic problems in composite materials and porous media. *J. Comput. Phys.*, **134**(1), 169–189.

Jansen, J.D. [2013a] *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.

Jansen, J.D. [2013b] *A systems description of flow through porous media*, 570. Springer.

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

Kraaijevanger, J.F.B.M., Egberts, P.J.P., Valstar, J.R. and Buurman, H.W. [2007] Optimal waterflood

design using the adjoint method. In: *SPE Reservoir Simulation Symposium*. Society of Petroleum Engineers.

Krogstad, S., Hauge, V.L. and Gulbransen, A. [2011] Adjoint multiscale mixed finite elements. *SPE Journal*, **16**(01), 162–171.

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

Le Ravalec, M., Tillier, E., Da Veiga, S., Enchéry, G. and Gervais, V. [2012] Advanced integrated workflows for incorporating both production and 4D seismic-related data into reservoir models. *Oil & Gas Science and Technology–Revue dâĂŹIFP Energies nouvelles*, **67**(2), 207–220.

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

Liu, N., Oliver, D.S. et al. [2003] Evaluation of Monte Carlo methods for assessing uncertainty. *SPE Journal*, **8**(02), 188–195.

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

MacBeth, C., Floricich, M. and Soldo, J. [2006] Going quantitative with 4D seismic analysis. *Geophysical Prospecting*, **54**(3), 303–317.

Moraes, R., Rodrigues, J., Hajibeygi, H. and Jansen, J.D. [2017a] Multiscale Gradient Computation for Multiphase Flow in Porous Media. In: *SPE Reservoir Simulation Conference*. Society of Petroleum Engineers.

Moraes, R.J.d., Rodrigues, J.R., Hajibeygi, H. and Jansen, J.D. [2017b] Multiscale gradient computation for flow in heterogeneous porous media. *Journal of Computational Physics*, **336**, 644–663.

Nocedal, J. and Wright, S. [2006] *Numerical optimization*. Springer Science & Business Media.

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

Oliver, D.S., Reynolds, A.C. and Liu, N. [2008] *Inverse theory for petroleum reservoir characterization and history matching*. Cambridge University Press.

Rodrigues, J.R.P. [2006] Calculating derivatives for automatic history matching. *Computational Geosciences*, **10**(1), 119–136.

Strang, G. [1993] *Introduction to linear algebra*. Wellesley-Cambridge Press Wellesley, MA.

Tarantola, A. [2005] *Inverse problem theory and methods for model parameter estimation*, 89. siam.

Tura, A. and Lumey, D.E. [1999] Estimating pressure and saturation changes time-lapse AVO data. In: *SEG Technical Program Expanded Abstracts 1999*, Society of Exploration Geophysicists, 1655–1658.

Ullmann De Brito, D., Caletti, L., Moraes, R. et al. [2011] Incorporation of 4D seismic in the reconstruction and history matching of Marlim Sul deep water field flow simulation model. In: *SPE EUROPEC/EAGE Annual Conference and Exhibition*. Society of Petroleum Engineers.

Ullmann De Brito, D., Moraes, R., Emerick, A.A. et al. [2010] 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.

Wang, Y., Hajibeygi, H. and Tchelepi, H.A. [2014] Algebraic multiscale solver for flow in heterogeneous porous media. *J. Comput. Phys.*, **259**, 284–303.

Wang, Y., Hajibeygi, H. and Tchelepi, H.A. [2016] Monotone multiscale finite volume method. *Computational Geosciences*, **20**(3), 509–524.

Zhang, F. and Reynolds, A.C. [2002] Optimization algorithms for automatic history matching of production data. In: *ECMOR VIII-8th European Conference on the Mathematics of oil Recovery*.

Zhou, H. and Tchelepi, H.A. [2008] Operator-Based Multiscale Method for Compressible Flow. *SPE J.*, **13**(02), 523–539.