Comput Geosci (2015) 19:999–1025
DOI 10.1007/s10596-015-9517-4

ORIGINAL PAPER

Ensemble-based multi-scale history-matching using second-generation wavelet transform

**Théophile Gentilhomme · Dean S. Oliver ·
Trond Mannseth · Guillaume Caumon · Rémi Moyen ·
Philippe Doyen**

Received: 20 December 2013 / Accepted: 20 October 2014 / Published online: 21 August 2015

Abstract Ensemble-based optimization methods are often efficiently applied to history-matching problems. Although satisfactory matches can be obtained, the updated realizations, affected by spurious correlations, generally fail to preserve prior information when using a small ensemble, even when localization is applied. In this work, we propose a multi-scale approach based on grid-adaptive second-generation wavelets. These wavelets can be applied on irregular reservoir grids of any dimensions containing dead or flat cells. The proposed method starts by modifying a few low frequency parameters (coarse scales) and then progressively allows more important updates on a limited number of sensitive parameters of higher resolution (fine scales). The Levenberg-Marquardt ensemble randomized maximum likelihood (LM-enRML) is used as optimization method with a new space-frequency distance-based localization of the Kalman gain, specifically designed for the multi-scale scheme. The algorithm is evaluated on two test cases. The first test is a 2D synthetic case in which several inversions are run using independent ensembles. The second test is the Brugge benchmark case with 10 years of history. The efficiency and quality of results of the multi-scale approach are compared with the grid-block-based LM-enRML with distance-based localization. We observe that the final realizations better preserve the spatial contrasts of the prior models and are less noisy than the realizations updated using a standard grid-block method, while matching the production data equally well.

Keywords Multi-scale · Second-generation wavelets transform · Grid-adaptive parameterization · Ensemble-based optimization · Multi-scale localization · Prior information preservation · History-matching · Inverse problem

1 Introduction

In standard geo-modeling workflows, integration of static and dynamic data is usually performed sequentially: Petrophysical geo-models are initially built using log data and

T. Gentilhomme (✉) · G. Caumon
GeoRessources (UMR 7359 Université de Lorraine-ENSG,
CNRS, CREGU), 2 Rue du Doyen Marcel Roubault,
54518 Vandœuvre-les-Nancy, France
e-mail: t.gentilhomme@gmail.com

G. Caumon
e-mail: guillaume.caumon@univ-lorraine.fr

T. Gentilhomme · R. Moyen
CGG - GeoConsulting, 27 Avenue Carnot,
91300 Massy, France

R. Moyen
e-mail: remi.moyen@cgg.com

D. S. Oliver · T. Mannseth
Uni CIPR, Uni Research, P.O. Box 7800,
5020 Bergen, Norway

D. S. Oliver
e-mail: dean.oliver@uni.no

T. Mannseth
e-mail: trond.mannseth@uni.no

T. Mannseth
Department of Mathematics, University of Bergen,
P.O. Box 7800, 5020 Bergen, Norway

P. Doyen
CGG - GeoConsulting, Crompton Way,
Crawley, RH10 9QN, UK
e-mail: philippe.doyen@cgg.com

geophysical data and are then perturbed to match historic production data. For this last step, ensemble-based methods of optimization [1–3] have become popular, thanks to their flexibility, computational efficiency, and ability to match dynamic data using a wide range of parameters and quantify posterior uncertainties. A quadratic approximation of the objective function is statistically estimated from an ensemble of realizations and is used to update each individual member. However, for high dimensional problems, the size of the ensemble is limited as a full fluid flow simulation needs to be run for each member. For this reason, the resulting approximation may be noisy and leads to spurious updates which will damage important geological features of the prior models and significantly reduce the variability of the ensemble. The aim of this work is to update a set of realizations, generated from a prior assimilation of data and knowledge, with production history using an efficient ensemble-based optimization method, while minimizing the perturbation of parameters and the addition of noise in order to maintain the variability of the ensemble and the consistency with all the available data used to constrain the realizations.

Low resolutions of property fields (large scales, e.g., porosity or permeability means) generally have an important impact on fluid flow in the reservoir and are well characterized by the production data [4–6]. Moreover, property contrasts, derived from facies modeling or seismic inversion, are not affected by the perturbations of low frequencies: Geological objects, such as channels, are characterized by sharp high frequency transitions of properties and seismic reflection data only give information about property contrasts and are band-limited such that they do not record low frequencies [7]. Therefore, our method postulates that the match of the production data can be improved by primarily modifying low resolutions, with a limited impact on the prior characterization and without introducing noise in higher frequencies. Under this hypothesis, a multi-scale re-parameterization of the model variables is appropriate.

Information carried by the production data is spatially heterogeneous [6], the flow response being more affected by the variations of parameters close to the wells where true correlations exist between the dynamic flow response and the high frequencies. Far from the measurements, high frequencies are generally not characterized by the data (except when they have an important impact on fluid flow, e.g., fine flow barriers). Therefore, the parameterization should be defined both in space and frequency. In this work, we decompose the property fields in space and frequency using a grid-adaptive second-generation wavelet parameterization. These wavelets are very efficient and can be applied on irregular reservoir grids containing dead or flat cells. Moreover, reservoir models typically contain hundreds to millions of unknown parameters and the available data are usually sparse and uncertain, which makes the problem under-determined. Re-parameterization then also helps stabilize the inversion by reducing the number of parameters, while preserving the main structures of the property fields, thanks to the compression properties of wavelets.

Based on this parameterization, we propose a multi-scale inverse approach: Different subsets of wavelet coefficients are successively updated with the production data using the iterative Levenberg-Marquardt ensemble-based randomized maximum likelihood (LM-enRML) [8, 9] as optimization method. The algorithm starts by optimizing an initial subset of large-scale coefficients. The parameterization is then progressively refined, and a new optimization is run each time new coefficients are added to the subset. However, the prior realizations may include uncertain fine-scale heterogeneities, which also have a significant impact on the flow response. For this reason, we propose to attenuate these high frequencies during the optimization of the large-scale coefficients in order to avoid bias in their estimation. The original values of the attenuated coefficients are then easily restored once they have been included in the optimization. Finally, the refinement process stops when all the coefficients are part of the optimization set. The mismatch being significantly reduced during the first optimizations of the large scales, the perturbation at finer scales is then limited.

Although re-parameterization helps identify sensitive parameters, it does not remove spurious correlations of the ensemble (i.e., correlations that do not exist, but appear in the ensemble because of its limited size). In ensemble-based methods, localization is used to attenuate these effects and avoid the deterioration of the ensemble. In this work, we propose a new space-frequency localization of the Kalman gain. We assume that the flow response is highly sensitive to the large-scale coefficients as they are corresponding to wavelet functions that are covering all the domain and data, whereas fine-scale coefficients (local high-frequency heterogeneities) only have an important impact close to the well locations as their corresponding wavelets have a more restricted impact on the property fields (smaller support). Thus, we adapt the localization to the current set of optimized parameters. At the beginning of the process, when only large scales are optimized, the correlations between parameters and data are assumed reliable and no regularization needs to be applied, which allows for global updates of the property fields. The localization later becomes effective when less sensitive coefficients are used in the optimization.

The algorithm is tested on two history-matching cases (Section 6). The first one is a 2D synthetic channelized reservoir with about 6 years of production history measured in nine producer and four injector wells. The second test is the Brugge benchmark case with 10 years of history. The results of the multi-scale approach are compared with the grid-block-based LM-enRML optimization with distance-based localization. While both methods perform well in matching the data, the regularization effect of the multi-scale approach avoids addition of noise and allows a better preservation of the models by minimizing the amplitude of the updates.

1.1 Features

The new multi-scale method includes the following features:

1. Grid-adaptive second generation smooth wavelets are used to re-parameterize property fields contained in complex stratigraphic reservoir grids (Section 3 and Appendix).
2. A sequence of LM-enRML optimizations is performed with an adaptive subset of optimized coefficients, initially composed of a few sensitive coarse-scale coefficients and progressively refined allowing more important updates on a limited number of sensitive coefficients of higher resolution (Section 4.2.1).
3. A reversible scheme attenuates the effects of non-included coefficients during the first optimizations in order to avoid bias in the estimation and then restores their influence on the fluid flow when they are added into the subset of optimized coefficients (Section 4.2.2 and Section 5). This allows incorporation of prior knowledge at all scales during the multi-scale inversion process.
4. A multi-scale space-frequency localization method and a specific control of the LM-enRML optimization, which depend on the subset of optimized coefficients, are used to further regularize the problem (Section 5.1 and Section 5.2).

1.2 Outline

In Section 2, we introduce the different concepts used in this work and give further references to related work. In Section 3, we present the second generation wavelets, which are used to re-parameterize the property fields. The LM-enRML optimization of a subset of the wavelets coefficients is then presented in Section 4. The multi-scale approach using space-frequency localization is described in Section 5. In Section 6, this method is applied on two history-matching test cases and results are compared with the LM-enRML method based on grid-block parameterization. Technical descriptions and implementations of the algorithm are relegated to the Appendix.

2 Previous work

2.1 Multi-scale approaches and re-parameterization

Multi-scale methods are commonly used in history-matching [10–15] in order to adapt the parameterization to the data, stabilize the inversion, and avoid over-parameterization (i.e., the use of an excessive number of parameters to explain the data). Although not formally proven, multi-scale optimization is considered to help avoid local minima [16–18] or lead to simpler optimum solution [19]. These methods typically start by optimizing a limited number of large-scale coefficients. The parameterization resolution is then increased by analyzing the results of the optimization step: Refinement indicators can be computed using the sensitivity matrix and/or the gradient in order to choose new parameters that will help reduce the value of the objective function at the next iteration without over-parameterizing the problem.

Different multi-scale parameterizations are found in the petroleum literature, as reviewed by [3, 14]. Methods based on adaptive zonation usually start with a coarse model and sequentially refine local regions depending on their impact on the flow response [10, 11]. A significant challenge with these methods is to use prior information at different scales, which is essential when other source of data are integrated in the model. Multi-scale geostatistical-based parameterization techniques [20] manage to preserve spatial variability, but cannot handle prior models that contain geological structures or seismic-derived information.

More common re-parameterizations based on linear transformation perform a change of basis. Interesting characteristics can result from these transformations, such as multi-scale decomposition, sparse representation, or parameters de-correlation. Methods based on the covariance of the model variables, such as principle components analysis (PCA) or Karhunen-Loeve transform (KLT) [21, 22], have been used to reduce the dimension of inversion problems. In petroleum reservoirs, however, the covariance is rarely known a priori and the computation of the basis can be problematic for large problems.

More recent approaches coming from the numerical image compression have the advantage to be independent of the prior knowledge and are computationally effective. Two main transformations are used in reservoir history-matching: the discrete cosine transform (DCT) [5, 23–25] and the discrete wavelet transform (DWT) [4, 6, 13]. For DCT, the transformed domain is characterized by global cosine functions, whereas the DWT uses more complicated functions called wavelets, which are also localized in space. With these parameterizations, the coarse resolutions of a signal can be optimized by only using specific coefficients whereas high resolutions characterized by the prior information (seismic, conceptual models, etc.) are not modified.

It is generally difficult in multi-scale approaches to account for prior information at all frequencies as the parameterization evolves and is not initially adapted to the frequency content of the prior information. In order to address this problem, Feng and Mannseth [26] propose a predictor-corrector strategy, which enables the downscaling of the multi-scale estimate and assimilation of prior covariance functions. Sahni and Horne [6] propose to fix some wavelet coefficients for history-matching and use the remaining coefficients to match a variogram. However, these techniques are difficult to use when detailed prior information, such as seismic derived attributes or complex geological structures, is available.

Although re-parameterization with few low-scale parameters gives satisfying results when applied to smooth synthetic cases, the optimization of the coarse scales is generally not sufficient to achieve a satisfying match of the production data. Fine-scale heterogeneities can have a large impact on the flow response where the changes of the dynamic properties are important and more frequencies need to be included in the inversion. Since these parameters are subject to spurious correlations, it is necessary to ensure that only sensitive parameters are modified during the ensemble-based optimization in order to preserve the information carried by the prior models.

2.2 Localization

In real applications, when the ensemble size is relatively small as compared to the dimension of the problem, localization helps prevent the spurious updates of variables that are not sensitive to the observation. Element-wise multiplications with a screening matrix are generally used to regularize the covariances matrices [27–29] or the Kalman gain [29, 30] directly. Distance-based localization has been used in atmospheric and petroleum applications to limit the update of the parameters within a certain distance of the observed dynamic data. Localization functions, usually chosen on the basis of the experience, are used to define the screening matrix. Their ranges are dependent on the size of the ensemble as it conditions the quality of the covariance approximations [27, 28, 31]. When distance-based functions fail to regularize correctly the ensemble update, sensitivity analysis may be used to define localization area [29], but can turn out to be fairly complex. Moreover, for parameters such as relative permeability curves or fluid contrasts, a distance interpretation is not relevant.

In this case, more general methods have been proposed for limiting the effects of the spurious correlations [4, 32, 33]. These methods can benefit from re-parameterization as it helps bring out important parameters. Chen and Oliver [4] apply the bootstrap method to wavelet coefficients and show that the estimate of the Kalman gain for the low frequencies using a small ensemble is less subject to noise. However, the automatic adaptive localization fails to estimate correctly the correlations associated with the high frequencies (fine scales).

2.3 First- and second-generation wavelets

The discrete wavelet transform is a linear transformation which decomposes a signal into different frequencies without increasing the number of coefficients. Its basis is composed of finite support functions called wavelets. Each wavelet is associated with a frequency range and a finite localization in space. This property is used in image compression where high frequencies are only included where they bring important information, for example, to characterize a sharp color variation. Since images or reservoir properties are generally spatially correlated, only a few wavelet coefficients are needed to obtain a good approximation of the original signal. For this reason, wavelet-based re-parameterization of an ill-posed inverse problem is interesting as it yields a significant reduction in the number of inverted parameters. However, unlike image compression in which coefficients are selected based on their impact on the visual aspect of an image, our history-matching method selects additional frequencies only where they have an important impact on the flow response.

Standard first-generation wavelet basis [34, 35], $\psi_{r,k}(t)$, is built from translations and dilatations (by a factor of two) of a generating function $\psi(t) \in L^2(\mathbb{R})$ called the mother wavelet:

$$\psi_{r,k}(t) = \frac{1}{\sqrt{s_0^r}}\psi\left(\frac{t - k\tau_0 s_0^r}{s_0^r}\right), \tag{1}$$

where τ_0 and s_0 correspond respectively to the translation and dilatation factor, which are usually set to $\tau_0 = 1$ and $s_0 = 2$. The indices r and k refer to the frequency and space localization of the wavelet. With the decrease of r, the frequency spectrum covered by the wavelets are compressed by a factor of two and translated from high to low resolutions (Fig. 1). For each set of wavelets, we can define a complementary set of functions, $\Phi_{r,k}(t)$, which cover the rest of the spectrum of the signal (Fig. 1) called the scaling functions [35]. Similarly to the wavelets, the scaling functions are built from a generating function (called the father wavelet) by translation and dilatation. Scaling functions have a low-pass nature and, for this reason, are sometimes called averaging functions.

Because of translation-dilatation invariance, first-generation wavelets are usually only applicable to regularly sampled infinite signals or bounded signals of dimensions equal to a power of 2.

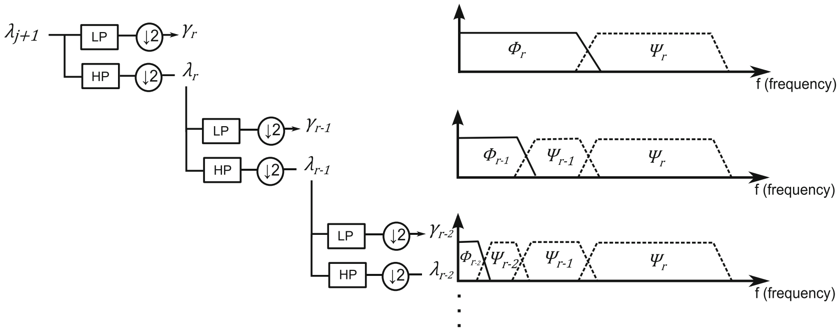


Fig. 1 Wavelet filter bank: cascade algorithm. The coefficients γ_r and λ_r correspond to the scaling and wavelet coefficients at the resolution r, respectively (*left*). Frequency coverage of the wavelets and scaling functions (*right*). The filters high-pass filter(HP) and low-pass filter (LP) can be deduced from the properties of the wavelet and scaling function [36]

For this reason, we use second generation wavelets [37] which can be applied to any grid (Appendixes 2 and 3). Second-generation wavelets lack the translation and dilatation invariance property. Instead, they are adapted to their spatial localization, but keep the same space-frequency localization property as traditional wavelets. The full wavelet transform follows the lifting scheme [37]. This process performs in-place processing of the signal which reduces memory requirements as compared to standard first-generation transform implementations. It is composed of reversible operators that are applied sequentially on a sequence of coefficients (Appendix 1). The main purpose of these operators is to improve the properties of an initial wavelet. The lifting of the Lazy wavelet [37] (Split operator, see Appendixes 1 and 2) and the Haar wavelet (Appendix 3) is presented in the Appendix. In this paper, we build grid-adaptive wavelets, which can handle dead or flat cells, by accounting for the volume of the cells during their construction.

3 Grid adaptive 3D wavelets transform

The lifting also enables the construction of grid-dependent wavelets. We use a scheme which takes into account the cell volumes, as described in Appendix 2, in order to handle zero volume cells (dead cells) and obtain a weighted average of the property fields at coarse resolutions. Because the zero volume or dead cells do not participate in the reconstruction of the property field, the associated coefficients can be neglected and the number of active parameters is not increased after the transformation. In the examples presented here, we use a quadratic wavelet based on a polynomial interpolation of degree 2, which can be seen as an improved Haar wavelet ([38], Appendix 3). This wavelet is smoother than the Haar wavelet, which helps avoid the introduction of sharp artifacts after perturbations of the wavelet coefficients caused by the optimization. The quadratic wavelets have a compact support (i.e., area where the wavelet is not nil) comparable to the Haar wavelet (wavelet with the smallest support), which allows a more precise localization in space compared to higher order wavelets associated with larger supports and an easier interpretation of the contribution of each wavelet in the reconstruction of the property fields as the overlaps between these basis functions are limited. Nevertheless, depending on the property of a signal (e.g., smooth and repeated pattern) and the field of application of the transform (e.g., compression and inversion), a wavelet transform may turn out to be more efficient than another, but this analysis is not in the scope of this article.

For two- or three-dimensional (2D and 3D) signals, it is possible to use 2D or 3D wavelet functions or to perform 1D transforms along each direction. The first approach is still rarely used as 2D or 3D wavelets are more difficult to create and the transform can be computationally less efficient. The second approach, followed in this work, is widely used in image compression because of its efficiency. The original 3D (or 2D) signal is decomposed into several 1D signals aligned along a first direction of the grid. One level of the wavelet transform is applied independently of each 1D signal, which results in the computation of detail and scaling coefficients (Fig. 2b). The 1D decomposition and wavelet transforms are repeated for each direction on both the detail and scaling coefficients (Fig. 2c, d). The same process is then repeated on the remaining scaling coefficients until the coarsest resolution is reached (Fig. 2a, b).

Different wavelets can be used in the different directions depending on the spatial variations of the properties. In a 3D reservoir model, the lateral property transitions inside a given layer can be smooth whereas the property transitions between layers can be sharp. In this case, smooth wavelets can be used laterally whereas sharp wavelets (e.g., Haar wavelet) can be used vertically. Lateral anisotropy can also be handled by using different wavelets depending on the correlation. Typically, high-order wavelets will be more efficient when the property correlation is important.

In the examples presented here, independent 2D wavelet transforms are used on either 2D grids or 3D grids with geologically independent layers.

In the following, we denote by $\mathbf{m} \in \mathbb{R}^n$ the property vector of size $n = n_p \times n_{\text{grid}} + n_d$ in the grid-block space,

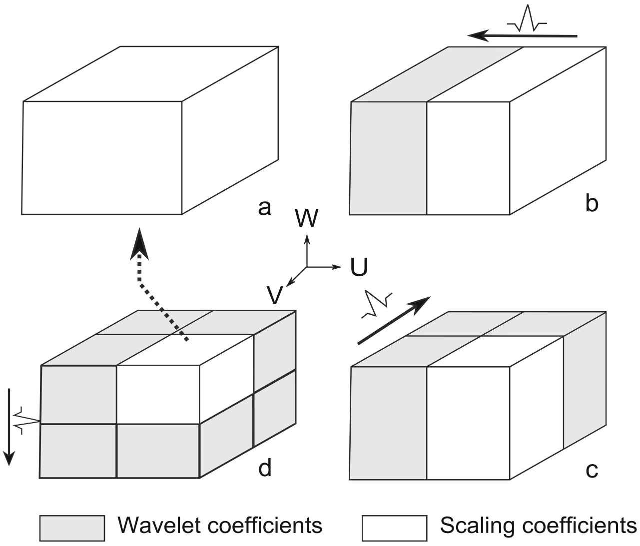


Fig. 2 3D wavelet transform. **a** Scaling coefficients at resolution r. **b** After wavelet decomposition along U direction. **c** After wavelet decomposition along V direction. **d** After wavelet decomposition along W direction. ($d \rightarrow a$) Repeat with the block of remaining scaling coefficients

where n_p, n_{grid}, and n_d are the number of spatial properties (e.g., grid-block porosity, and permeabilities), the size of the grid, and the number of flow simulator parameters (e.g., relative permeabilities); $\boldsymbol{\gamma} \in \mathbb{R}^n$ the vector of all wavelet coefficients; and $\mathbf{W}$ the direct wavelet transform such that $\boldsymbol{\gamma} = \mathbf{Wm}$ and $\mathbf{W}^{-1}$ the inverse wavelet transform.

4 Optimization with a subset of wavelet coefficients

4.1 Notations

In this section, we introduce mathematical notations used to define and access submatrices or subvectors of a given matrix or vector. Let n and k be integers with $1 \leqslant k \leqslant n$. We denote by $Q_{k,n}$ the totality of all sequences of k integers where the elements of the sequence are strictly increasing and chosen from $\{1, \ldots, n\}$. Let $\mathbf{A} = (a_{ij})$ be a $(m \times n)$ matrix with elements $a_{ij} \in \mathbb{R}$. Let r and c be positive integers with $1 \leqslant r \leqslant m$, $1 \leqslant c \leqslant n$, and $\alpha \in Q_{r,m}$, $\beta \in Q_{c,n}$ two sequences of i and j indices of matrix $\mathbf{A}$, such that $\alpha = (i_1, \ldots, i_r)$, $\beta = (j_1, \ldots, j_c)$. Then, the submatrix $\mathbf{A}[\alpha, \beta]$ has (l, t) entries equal to a_{i_l, j_t} and has r rows and c columns. We denote by "$*$" the increasing sequence $Q_{n,n}$ of all the elements of $\{1, \ldots, n\}$, such that $\mathbf{A}[*, *] = \mathbf{A}$. We denote by $\mathbf{A}(\alpha, \beta)$ the submatrix of $\mathbf{A}$ whose rows and columns are complementary to α and β, respectively.

In the following, the terms *scale* or *resolution* refer to frequency ranges carried by the wavelets (see Fig. 1) and we denote by n and n_r the total number of wavelets coefficients and the number of wavelet coefficients associated to the resolution r, with $n = \sum n_r$. We denote by $\left\{i_\gamma\right\}_r$ the set of wavelet coefficient indices of resolution r and by $q\left(\left\{i_\gamma\right\}_r\right) \in Q_{n_r,n}$ its corresponding increasing sequence. Finally, we denote by $\left\{I_\gamma\right\}_{r,r'} = \left\{i_\gamma\right\}_r \cup \ldots \left\{i_\gamma\right\}_{r'}$ the set of all coefficient indices from resolution r to r' and $q_{r,r'} \in Q_{n_{r'}+\cdots+n_r,n}$ its corresponding increasing sequence. With these notations, the universe (set of all wavelet coefficient indices) is defined as $U = \left\{I_\gamma\right\}_{0,r_{\max}}$, where $r = 0$ and $r = r_{\max}$ correspond to the coarsest (or lowest) (e.g., the mean) and finest (or highest) resolutions, respectively.

4.2 Optimization method

4.2.1 Levenberg-Marquardt ensemble-based randomized maximum likelihood

The Levenberg-Marquardt ensemble randomized maximum likelihood (LM-enRML) [9, 39] is an iterative Levenberg-Marquardt method which provides a quadratic approximation to the logarithm of the maximum a posteriori estimate using an ensemble of realizations. This approximation is then used to update simultaneously several realizations which are members of the ensemble [8]. The initial ensemble is sampled from a prior distribution (which might be conditioned by other data, such as core or seismic data) and deemed to be representative of the uncertainties before assimilation of the production data.

Each ensemble member (vector of parameters, i.e., cell porosities, permeabilities) $\mathbf{m} \in \mathbb{R}^n$ is transformed into wavelet coefficients $\boldsymbol{\gamma} \in \mathbb{R}^n$ by applying the lifting scheme. The optimization is then performed on a subvector of selected coefficients $\boldsymbol{\gamma}^* \in \mathbb{R}^{n^*}$ with $n^* \leqslant n$ such that $\boldsymbol{\gamma}^* = \boldsymbol{\gamma}[\alpha]$ and $\boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{\gamma}^* \\ \boldsymbol{\gamma}^c \end{bmatrix} = \mathbf{Wm}$, where $\alpha \in Q_{n^*,n}$ and $\boldsymbol{\gamma}^{\mathbf{c}} = \boldsymbol{\gamma}(\alpha)$ correspond to the indices of selected coefficients of γ and the complement vector of $\boldsymbol{\gamma}^*$, respectively. The corresponding property values used by the flow simulator can be reconstructed using the inverse wavelet transform, $\mathbf{W}^{-1}$, such that $\mathbf{m} = \mathbf{W}^{-1}\boldsymbol{\gamma}$.

One common method of uncertainty quantification for large history-matching problems is to condition each individual member $j \in \mathbb{N}$ to a data realization, $\mathbf{d^0}_j \in \mathbb{R}^{n^d}$, generated from a prior multivariate Gaussian distribution characterized by the data mean $\mathbf{d_{obs}} \in \mathbb{R}^{n^d}$ and the $(n^d \times n^d)$ data covariance, $\mathbf{C_D}$ (often assumed diagonal) [40, 41]. This is accomplished by minimizing an objective function of the form:

$$S_j^{rml}\left(\begin{bmatrix} \boldsymbol{\gamma}_j^* \\ \boldsymbol{\gamma}^c \end{bmatrix}\right) = \frac{1}{2}(\boldsymbol{\gamma}_j^* - \boldsymbol{\gamma}_{pr,j}^*)^T \mathbf{C}_{\gamma^*}^{-1}(\boldsymbol{\gamma}_j^* - \boldsymbol{\gamma}_{pr,j}^*) + \frac{1}{2}(g(\mathbf{m}_j) - \mathbf{d^0}_j)^T \mathbf{C_D}^{-1}(g(\mathbf{m}_j) - \mathbf{d^0}_j) \quad (2)$$

Where $\boldsymbol{\gamma}^*_{pr,j} = \boldsymbol{\gamma}_{pr,j}\,[\alpha]$, $\mathbf{C}^{-1}_{\gamma^*} = \mathbf{C}_\gamma\,[\alpha]^{-1}$, and $g(.)$ are the subvector of the jth ensemble prior member $\boldsymbol{\gamma}_{pr,j}$, the inverse of the submatrix of the covariance $\mathbf{C}_\gamma$, and the forward model (i.e., flow simulator), respectively. Since the forward model is non-linear, the solution of this problem cannot be calculated analytically and calls for an iterative resolution.

Gauss-Newton methods and, in particular, the Levenberg-Marquardt algorithm are known to be very efficient when the $(n^* \times n^d)$ sensitivity matrix (or Jacobian), $\mathbf{G} = \frac{\partial g(\mathbf{m})}{\partial \mathbf{m}}$, is cheap to compute. For each iteration, a perturbation of the parameters is computed from the gradient (first derivative, steepest descent) and the Hessian (second derivative, shape) of the objective function. For non-linear problems, the approximation of the objective function can only be trusted around the current state of the model. In the Levenberg-Marquardt method, the Hessian is modified with a multiplier λ in order to reduce the influence of data mismatch in early iterations and therefore limit the amplitude of the parameter updates. The perturbation, $\delta\boldsymbol{\gamma}^*_l$, at the lth iteration is obtained by solving:

$$[(\lambda_l + 1)\mathbf{C}^{-1}_{\gamma^*} + \mathbf{G}^T_l\mathbf{C_D}^{-1}\mathbf{G}_l]\delta\boldsymbol{\gamma}^*_l = \qquad (3)$$
$$-\ [\mathbf{C}^{-1}_{\gamma^*}(\boldsymbol{\gamma}^*_l - \boldsymbol{\gamma}^*_{pr}) + \mathbf{G}^T_l\mathbf{C_D}^{-1}(g(\mathbf{m}_l) - \mathbf{d^0})]$$

Where the right and first left term of the equality correspond to the gradient and the modified Hessian of $S^{rml}\,(\boldsymbol{\gamma})$. The $(n^d \times n^*)$ sensitivity matrix $\mathbf{G}_l$ is computed from the ensemble of n^e realizations by solving:

$$\Delta\mathbf{D}_l = \mathbf{G}_l.\Delta\boldsymbol{\Gamma}_l. \qquad (4)$$

Each column of the $(n^d \times n^e)$ matrix $\Delta\mathbf{D}_l$ and the $(n^* \times n^e)$ matrix $\Delta\boldsymbol{\Gamma}_l$ store, for each realization, the deviation of the predicted data, and deviation of the wavelets coefficients from the ensemble means, respectively [8]. Since the matrix is generally not invertible (or even square), (truncated) singular value decomposition ((T)SVD) is used to solve the system. If TSVD is applied, the parameters must be scaled in order to avoid the loss of valuable information [9]. Because of the reduction in the number of parameters $(n^* \leqslant n)$ and the small size of the ensemble $(n^e \ll n)$, the computation of $\mathbf{G}$ is relatively fast.

In this work, we use the approximate form of the Levenberg-Marquardt method [9], where the covariance matrix $\mathbf{C}_{\gamma^*}$ in the Hessian is approximated by a positive definite matrix $\mathbf{P}_l = \Delta\boldsymbol{\Gamma}_l\Delta\boldsymbol{\Gamma}_l{}^T/(n^e - 1)$, which is computed from the current ensemble in each iteration. Using the Sherman-Woodbury-Morrison matrix inversion formulas [42] and substituting $\mathbf{C}^{-1}_{\gamma^*}$ by $\mathbf{P}_l{}^{-1}$ in the Hessian, (4) is written as:

$$\delta\boldsymbol{\gamma}^*_l = -[(\lambda_l + 1)\mathbf{P}_l{}^{-1} + \mathbf{G}^T_l\mathbf{C_D}^{-1}\mathbf{G}_l]^{-}1\mathbf{C}^{-1}_{\gamma^*}(\boldsymbol{\gamma}^*_l - \boldsymbol{\gamma}^*_{pr})$$
$$-\mathbf{P}_l\mathbf{G}^T_l[(\lambda_l + 1)\mathbf{C_D} + \mathbf{G}_l\mathbf{P}_l\mathbf{G}^T_l]^{-1}(g(\mathbf{m}_l) - \mathbf{d^0}) \qquad (5)$$

Finally, the deviation from prior term is neglected (first term in Eq. 5) in which case the update is only driven by the data mismatch:

$$\delta\boldsymbol{\gamma}^*_l = -\mathbf{P}_l\mathbf{G}^T_l[(\lambda_l + 1)\mathbf{C_D} + \mathbf{G}_l\mathbf{P}_l\mathbf{G}^T_l]^{-1}(g(\mathbf{m}_l) - \mathbf{d^0}). \qquad (6)$$

It is shown [9] that this form which gives slightly better data matches is numerically more stable and is computationally more efficient than the original form (Eq. 5). From the authors' experience, this approximated form can give better matches when the initial ensemble of simulated responses is not enclosing the data, and the quality of the output realizations does not suffer from the loss of the prior deviation constraint. When the same ensemble is used to compute the sensitivity and the current covariance matrices, Eq. (6) simplifies to the ensemble smoother update and the sensitivity matrix does not need to be computed explicitly:

$$\begin{aligned}\delta\boldsymbol{\gamma}^*_l &= -\mathbf{C}^{\gamma\mathbf{d}}_l[(\lambda_l + 1)\mathbf{C_D} + \mathbf{C^{dd}}_l]^{-1}(g(\mathbf{m}_l) - \mathbf{d^0})\\ &= -\mathbf{K^e}_l(g(\mathbf{m}_l) - \mathbf{d^0}), \qquad (7)\end{aligned}$$

where $\mathbf{C}^{\gamma\mathbf{d}}_l = \Delta\boldsymbol{\Gamma}_l\Delta\mathbf{D}_l{}^T/(n^e - 1)$ is the cross-covariance between parameters included in the optimization and simulated data, $\mathbf{C^{dd}}_l = \Delta\mathbf{D}_l\Delta\mathbf{D}_l{}^T/(n^e - 1)$ is the simulated data covariance matrix and $\mathbf{K^e}_l$ is similar to the estimate of the Kalman gain.

4.2.2 Initial smoothing of the property fields

Although we optimize only a subset of n^* coefficients, all coefficients are used to reconstruct the properties needed by the flow simulator. However, for non-linear problems, it is generally useful to attenuate the impact of non-selected coefficients (mostly high frequencies) on the flow response in order to avoid bias in the estimation of the large scales and avoid local minima. At well locations, high resolutions may have an important impact on the flow response and can bias the characterization of the low resolutions if not removed. For example, fine-scale heterogeneities may introduce high or low permeability values (local drain or barriers) in the cells of the model that intersect or are located close to the well perforations. As these fine scales have an important impact on the flow response, but are not included in the optimization, it could result in an unrealistic deformation of the large scales (e.g., large increase or decrease of the permeability average) that attempt to correct a local state. In other words, we are trying to limit the influence of sensitive parameters that are not yet included in the optimization. Accordingly, the effect of wavelet coefficients that are not included in the optimization is attenuated before starting the optimization by multiplying them with an attenuation vector $\mathbf{s}_0 \in \mathbb{R}^n$ such that,

$$\boldsymbol{\gamma}^\mathbf{s} = \mathbf{s}_0 \circ \boldsymbol{\gamma} = \begin{bmatrix}\boldsymbol{\gamma}^*\\ \boldsymbol{\gamma}^c_s\end{bmatrix}, \qquad (8)$$

where $\circ$ corresponds to the Schur or Hadamard element-wise product, $\boldsymbol{\gamma}^*$ is the subvector of optimized coefficients, and $\boldsymbol{\gamma}_s^c$ is the attenuated version of $\boldsymbol{\gamma}^{\mathbf{c}}$ (coefficients not used in the optimization). The application of the inverse wavelet transform on the vector $\boldsymbol{\gamma}^{\mathbf{s}}$ results in the reconstruction of smoothed property fields, as displayed in Fig. 3. Generally, $\mathbf{s}_0$ is not constant, but its elements vary with the resolution and location of the corresponding wavelet coefficients. In this work, the values of the elements of $\mathbf{s}_0$ decrease linearly with the resolution of the associated parameters from 1 (no attenuation) to a minimum value (strongest attenuation), $s_{\min}$, set to 0.1 in the applications presented here, such that:

$$\forall r \in \{0, \ldots, r_{\max}\}, \forall j_r \in \{i_\gamma\}_r$$
$$\mathbf{s}_0(j_r) = \min\left(1, \frac{(1 - s_{\min}) \times r + s_{\min} \times r_0 - r_{\max}}{(r_0 - r_{\max})}\right), \quad (9)$$

where r is the resolution associated with the wavelet coefficients from the coarsest ($r = 0$) to the finest ($r = r_{\max}$), r_0 is the starting resolution for the optimization (see Section 5), and $\{i_\gamma\}_r$ is a set of wavelet coefficient indices, j_r, for resolution r.

As developed later, the attenuation values must be greater than 0 in order to compute the sensitivity or the cross-covariance (Eqs. (4) or (7)) of the wavelet coefficients after refinement of the parameterization and to restore the original value of the coefficients. Thanks to the compression property of the wavelet basis, all important features are still preserved, as illustrated in Fig. 3. The attenuated frequencies are easily restored during the process by multiplying the wavelets coefficients by $\mathbf{s}_0^{-1}$, such that $\boldsymbol{\gamma} = \mathbf{s}_0^{-1} \circ \boldsymbol{\gamma}^{\mathbf{s}}$ (see Section 5 and Algorithm 2 in Appendix 4), where $\mathbf{s}_0^{-1}$ is the element-wise inverse of $\mathbf{s}_0$.

In general, the minimum attenuation factor $s_{\min}$ associated with high frequencies must be small in order to avoid an incorrect estimation of the lower scale coefficients that impact the entire domain. The strategy proposed in Eq. 9 does not claim to be optimal or generic, but we observed that an increase of $s_{\min}$ results in a smaller decrease of the objective function when optimizing the low frequencies and a too important attenuation (i.e., decrease of $s_{\min}$) can result in an increase of the objective function when restoring the original value of the coefficients in the last iterations (see Section 5). Note that all the elements of $\mathbf{s}_0$ corresponding to the coefficients $\boldsymbol{\gamma}^*$ are equal to one since they are included in the optimization (9).

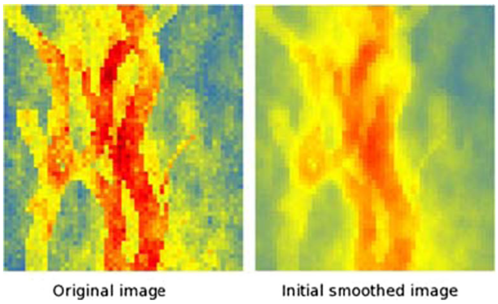

Original image
Initial smoothed image

Fig. 3 Original porosity realization (*left*) and its smoothed version (*right*) used at the beginning of the optimization after attenuation of the high frequencies

5 Multi-scale approach

We propose a multi-scale approach based on the LM-enRML optimization using wavelet parameterization, which includes highly detailed prior knowledge. An overview of this method is given in Fig. 4. The main central loop corresponds to iterations through the different resolutions. Before entering the loop (step 0), some low resolutions (from 0 to r_0) wavelet coefficients, $\boldsymbol{\gamma}_{r_0}^*$, are selected, such that $\boldsymbol{\gamma}_{r_0}^* = \boldsymbol{\gamma}\left[q_{0,r_0}\right]$ and a reversible attenuation (Section 4.2.2) is applied to the remaining coefficients, $\boldsymbol{\gamma}_{r_0}^c = \boldsymbol{\gamma}\left(q_{0,r_0}\right)$. The initial selection should only include coefficients associated with wavelets, Ψ^*, that are covering at least one element of the observed data locations set, P_d, such that,

$$\forall \psi_{r,k} \in \Psi^*, \; \exists x_d \in P_d \mid \left(\psi_{r,k} * \delta_{x_d}\right)(x) > \epsilon, \quad (10)$$

where $(*)$ is a convolution operator, $\psi_{r,k}$ denotes wavelets of resolution r and position k, δ_{x_d} is a Dirac function centered on the location x_d, and ϵ is a minimum impact value.

Moreover, the initial parameterization should include a minimum number of parameters so that the main features of the prior realizations remain discernible after attenuation of the high frequencies (Fig. 3) in order to avoid convergence at large scales that is not compatible with the prior information at finer scales. Visual qualitative controls of the preservation of the main features of property fields after smoothing tests have been used in this work to select the initial coefficients (starting resolution) for the applications examples of the Section 6.

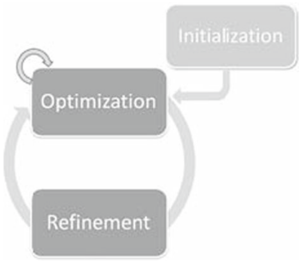


Fig. 4 Overview of the multi-scale method. Step 0: initialization of the parameterization and attenuation of the high frequencies (smoothing). Step 1: ensemble-based closed loop optimization with the current parameterization. Step 2: refinement of the parameterization and restoration of the attenuated wavelets coefficients with their original values

In the first step (step 1), the parameters $\boldsymbol{\gamma}_r^*$ are optimized using the LM-enRML algorithm as described in Section 4. Although the ensemble is generally not large enough to obtain a perfect estimate of the Kalman gain for the full parameterization, it usually gives a good approximation for the large-scale coefficients. We emphasize here that the wavelet transform by itself does not remove the spurious effects related to the use of a small ensemble, but only helps re-parameterize the problem with few sensitive parameters. In geological or atmospheric applications, only a limited number of large-scale scenarios of variation of the spatial properties are consistent with the physics, the depositional processes, and the prior knowledge, which enable a better approximation of the parameters distribution with a small ensemble. The simulated response is also very sensitive to these coefficients as their associated basis functions are covering a large part of the model. However, during the first resolutions, because all the coefficients needed to match the data that are not included, it is not recommended to seek convergence as it can lead to unrealistic updates. Moreover, the optimization of the low frequencies is only performed after attenuating the effect of the fine-scale parameters (8) in order to avoid bias in the estimation (see Section 4.2.2).

Then (step 2), new parameters corresponding to the next finer resolution $r+1$ are added with their original values and the attenuation vector $\mathbf{s}_{r+1}$ is updated, such that,

$$\boldsymbol{\gamma}_{r+1}^* = \mathbf{s}_r\left[q_{0,r+1}\right]^{-1} \circ \boldsymbol{\gamma}^{\mathbf{s}}\left[q_{0,r+1}\right] \tag{11}$$

$$\forall j \in \left\{I_\gamma\right\}_{0,r+1}, \ \mathbf{s}_{r+1}(j) = 1. \tag{12}$$

The influence of the finer scale are also partially re-introduced by multiplying the non-included coefficients by a factor ϵ_s set to 1.15 in the applications presented here, such that the original values of the coefficients are progressively restored (see Algorithm 2 in Appendix 4):

$$\forall j \in \left\{I_\gamma\right\}_{0,r+1}^c,$$

$$\boldsymbol{\gamma}_{s,r+1}^{\mathrm{c}}(j) = \boldsymbol{\gamma}_{s,r}^{\mathrm{c}}(j) \times \epsilon_s \tag{13}$$

$$\mathbf{s}_{r+1}(j) = \mathbf{s}_r(j) \times \epsilon_s, \tag{14}$$

where $\left\{I_\gamma\right\}_{0,r+1}^c$ is the complementary set of non-included wavelets coefficients indices (indices of $\boldsymbol{\gamma}_{s,r+1}^{\mathrm{c}}$). Finally, a new optimization is run with the new parameterization. In order to avoid running specific flow simulations for the unsmoothed realizations, the results of the last evaluation of the objective function before refinement are used to compute the first update with the new parameterization. For this reason, the initial attenuation factors must be greater than 0.

This process is repeated until the finest resolution is reached and a stopping criterion (Section 5.2) is fulfilled at the optimization step. Because of the initial attenuation of the coefficients, the process should not end before entering the finest resolution even if the data are sufficiently well matched since some prior knowledge could be lost.

5.1 Multi-scale localization

The objective of the localization in ensemble-based optimization is to minimize the difference between the true Kalman gain, $\mathbf{K}$, and the estimate of the Kalman gain, $\mathbf{K^e}$, affected by spurious correlations considering that it is computed from an insufficiently large ensemble. This can be achieved by multiplying element-wise the estimate of the Kalman gain by a screening matrix $\mathbf{A}$ that minimizes the function:

$$f(\mathbf{A}) = \left\|\mathbf{A} \circ \mathbf{K^e} - \mathbf{K}\right\|_f^2, \tag{15}$$

where $\|.\|_f^2$ is the Frobenius norm.

We define two screening matrices, $\mathbf{A_{gb}}$ and $\mathbf{A_{wt}}$, applied sequentially in the grid-block and wavelet spaces in order to regularize the estimate of the Kalman gain. The first regularization is based on distance functions defined in the grid-block space that vary with the resolution. The second regularization is derived from a user-defined pseudo-sensitivity map which is translated into the space-frequency domain. Although they are both related to the position of the coefficients in space and frequency, they are serving different purposes. The grid-block localization only limits the update of the parameters inside pre-defined areas without distinguishing between frequencies, whereas the wavelet localization only restricts the updates of some frequencies at some locations, but allows global modifications of the property fields.

5.1.1 Multi-scale grid-block localization

Distance-based localization aims at regularizing the estimate by limiting the effects of the data to sensitive areas. Equivalently, distance-based localization can be seen as the decomposition of one large problem into several smaller semi-independent problems that are less affected by the spurious correlations as only a reduced number of parameters with significant correlations to the data are included.

We propose here to adapt the localization to the current parameterization used in the optimization. A localization function, $\mathbf{A_{gb}}$, similar to the one proposed by Furrer and Bengtsson [31] is used to screen the estimate of the Kalman gain in the grid-block space:

$$\mathbf{A_{gb}}(h, r) = \frac{1}{1 + [f(0,r)^2/f(h,r)^2]/n^e}, \tag{16}$$

where $f(h,r)$ is a distance function which depends both on the separation distance h and the resolution r of the multi-scale loop. In the first iterations of the multi-scale loop, the dimension of the problem is highly reduced as only few large-scale parameters are optimized. As discussed before, the distribution of these parameters is generally better sampled by the ensemble and these parameters are generally sensitive, which give a better estimation of the Kalman gain. We also recall that the initial smoothing prevents introduction of a bias induced by the effect of the high resolutions. In this case, no regularization needs to be applied and the localization functions are constant functions equal to one, allowing global updating of the realizations without introducing noise in finer resolutions. With successive refinements, the number of parameters increases and the estimate of the Kalman gain becomes more affected by the spurious correlations. As a consequence, the range of the distance function decreases and the updates are applied in smaller sensitive areas (see Fig. 5, top). At the finest resolution, the localization function simplifies to the one proposed by Furrer and Bengtsson [31].

As localization is defined in the grid-block and because of the use of the Hadamard product in Eq. 15, it is not straightforward to apply localization directly in the wavelet space ($\mathbf{W}(\mathbf{B}\circ\mathbf{C}) \neq \mathbf{W}\cdot\mathbf{B}\circ\mathbf{W}\cdot\mathbf{C}$, where $\mathbf{B}$, $\mathbf{C}$ represent any matrices with consistent dimensions). Instead, the Kalman gain is transformed back into the grid-block space where the localization is applied and is then re-transformed into the wavelet space in order to update the reduced set of wavelet coefficients (see Algorithm 3 in Appendix 4), such that:

$$\begin{aligned}\mathbf{K}^{\mathbf{tmp}} &= \mathbf{W}\left(\mathbf{A_{gb}} \circ \left(\mathbf{W}^{-1}\cdot \mathbf{K}^{\gamma^*}_{n,n^d}\right)\right)\\ \mathbf{K}^{\gamma^*} &= \mathbf{K}^{\mathbf{tmp}}\left[q_{0,r},*\right],\end{aligned} \tag{17}$$

where $\mathbf{K}^{\mathbf{tmp}}$ is a $\left(n \times n^d\right)$ temporary matrix, $\mathbf{K}^{\gamma^*}_{n,n^d}$ is a $\left(n \times n^d\right)$ matrix defined such that, $\mathbf{K}^{\gamma^*}_{n,n^d}\left[q_{0,r},*\right] = \mathbf{K}^e$, and $\mathbf{K}^{\gamma^*}_{n,n^d}\left(q_{0,r},*\right) = \mathbf{0}_{n-n^*,n^d}$, where $\mathbf{K}^e$ is the $\left(n^* \times n^d\right)$ estimate of the Kalman gain for the selected wavelet coefficients and $\mathbf{0}_{n-n^*,n^d}$ is a $\left((n-n^*)\times n^d\right)$ zero matrix corresponding the non-selected coefficients, and $\mathbf{K}^{\gamma^*}$ is the resulting localized $\left(n^* \times n^d\right)$ Kalman gain in the wavelet space. Thus, even the update of global parameters, such as the mean, only have impact inside localizations areas at the current resolution. Unless a strict stationarity of the property field can be assumed, it is important to limit the effect of the global parameters when the parameterization is refined. Indeed, with the inclusion of high frequencies, the estimation of the large-scale parameters, far from the data, is biased by the local state (except in strictly stationary state). In other words, we avoid incorrect extrapolations of local observations. The wavelet transform being fast and easily parallelizable, the transformation and reconstruction of the Kalman gain take a reasonable amount of time compared to the ensemble of the fluid flow simulations.

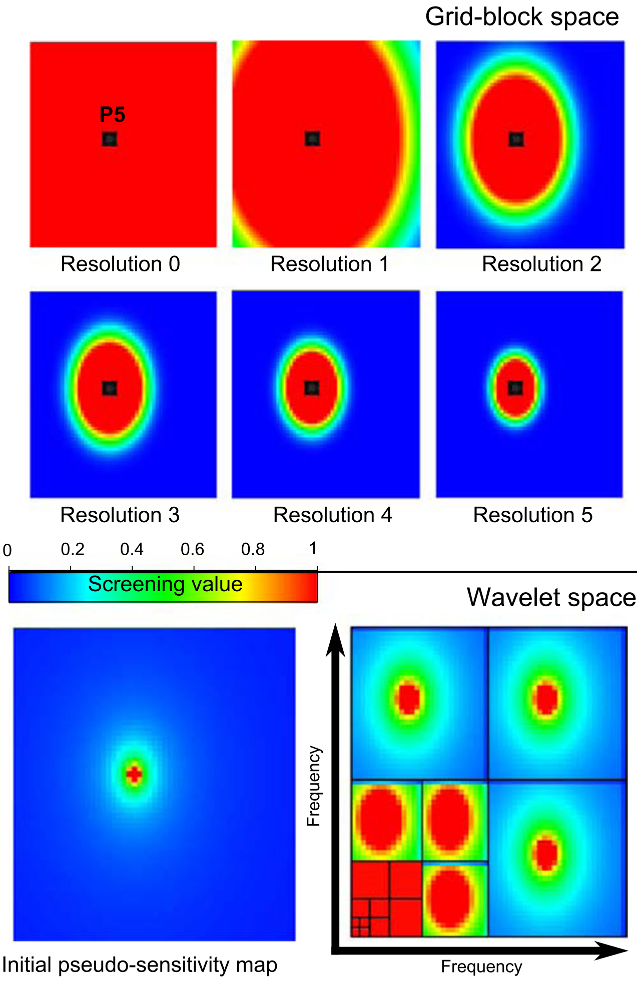


Fig. 5 Multi-scale localization for pressure data at well P5. Applied on the Kalman gain in the grid-block space at the different resolutions of the multi-scale loop (*top*). Applied on the Kalman gain in the wavelet space (*bottom*) (see Algorithm 3 in Appendix 3)

5.1.2 Regularization of the Kalman gain in the wavelet space

For large problems, even by taking into account spatial dependencies, it is generally not possible to ensure total suppression of the spurious correlations. Localization functions generally depends on the size n^e of the ensemble [31] as in Eq. 16, meaning that when the ensemble is small, the localization may become very restrictive and only allows updates of a limited numbers of parameters, which conversely may remove true correlations and lead to bad data match or create artifacts. Generally, the localization function should be larger than a single pattern and include nearby wells [29]. In this work, another kind of regularization, inspired from the analysis based on the adaptive localization presented in [4], is used in addition to the space localization described above.

We propose to apply a regularization of the estimate of the Kalman gain in the wavelet space. Generally, parameters close to the well measurements will have a significant impact on the flow response and fine scales heterogeneities are more likely to be characterized, whereas far from wells, only low frequency content can be deduced from the data (except in very heterogeneous reservoirs). Accordingly, pseudo-sensitivity maps are first defined in the grid-block space and aim at identifying areas where production data are highly informative. Distance functions with small ranges are used to generate these maps, where values drop rapidly from 1 to 0 with the inverse of the distance from the location of the data (Fig. 5, bottom left). Because water cuts are sensitive to the heterogeneities between producer and injector, a larger range is used in order to include more fine-scale parameters between the wells. Then, this knowledge is translated into the wavelet space by considering that the reliability of the cross-correlation between data and wavelet coefficients is related to the sensitivity of the area covered by the basis functions. Thus, operators similar to the scaling functions are used to integrate the values of the pseudo-sensitivity maps over the support of the wavelet functions in order to estimate the importance of the different wavelet coefficients.

If the Haar transform is used to decompose a given 1D property and if we denote by $m^s_{j+1,k}$, the initial pseudo-sensitivity map, then the value $m^s_{j,k}$ associated with the wavelet coefficient $\gamma_{j,k}$ is computed, such that,

$$m^s_{j,k} = m^s_{j+1,k} + m^s_{j+1,k+1}. \tag{18}$$

This operation is performed on the remaining $m^s_{j,k}$ values until the coarsest resolution is reached in the same manner as the cascade algorithm (Fig. 1). For a 2D or 3D property, the process is successively applied along the different directions, and the resulting values correspond to the sum over a surface or a volume. Thus, the importance of the wavelet coefficients increases quickly when their frequencies decrease (see Fig. 5, bottom) and only the finest frequencies are affected by this regularization (large-scale correlations are supposed reliable as shown in [4]). The final values are then leveled to 1 in order to be used as elements of the $\left(n^* \times n^d\right)$ screening matrix $\mathbf{A_{wt}}$ of the Kalman gain in the wavelet space, such that:

$$\mathbf{K}^{\gamma^*}_{\text{loc}} = \mathbf{A_{wt}} \circ \mathbf{K}^{\gamma^*}, \tag{19}$$

where $\mathbf{K}^{\gamma^*}_{\text{loc}}$ is the resulting Kalman gain after both regularizations in the grid-block and wavelet space. We emphasize that even though the term pseudo-sensitivity is used to define the maps in the grid-block space, the matrix $\mathbf{A_{wt}}$ is not an approximation of the sensitivity in the wavelet space, but is only used to identify the reliable correlations between the parameters and data.

When higher order wavelets are used, the sum in Eq. 18 includes more elements (depending on the order of the wavelet) with different weights deduced from the associated scaling functions. Because scaling functions, in the second-generation context, are changing with the sampling of the signal and are not computed explicitly, it can be difficult to define correctly the corresponding integrating functions and approximations can be made. In this work, a wavelet based on a quadratic average interpolation is used to decompose the properties, but we approximate the computation of the screening coefficients to the Haar case (18) as the two wavelets are quite similar (Appendix 3).

5.2 Optimization and control of the Levenberg-Marquardt parameter

In our multi-scale method, an optimization is performed after each refinement of the parameterization. During the first resolutions, the parameterization only includes large-scale parameters and the problem is simplified, thanks to the initial smoothing of the realizations, which allows a fast convergence. However, because the problem is generally under-parameterized at this stage (i.e., the parameterization does not contain enough degrees of freedom to match the data), it is not suitable to reach full convergence as it would result in unrealistic solutions. Moreover, the initially smoothed components reintroduced during the refinement step might be incompatible with the current optimal state. For these reasons, only one LM-enRML iteration is performed at each resolution level while the total number of parameters is lower or equal to the total number of data. At intermediate scales, when iterations are allowed, refinement criteria based on the maximum number of iterations and the improvement of the mismatch are applied. Because all the parameters are not optimized and finer scales have been smoothed, only a few iterations are permitted. In the applications presented in this work, the maximum number of iterations at an intermediate scale is limited to 2, but the control of the iteration would benefit from automatic techniques based on the approximation of the objective function, as described in [11]. Several stopping criteria are used for the last optimization (at the finest resolution), including total number of flow simulations, maximum number of iterations, maximum increase of λ at one iteration, or whether the mismatch falls within production data uncertainty ($S_{av} < n^d$).

Initialization and control of the Levenberg-Marquardt parameter λ are based on the initial mismatch of data and the evolution of the ensemble mean, S_{av} (see (20) below), and standard deviation of the data mismatch, σ_S, as described in [9]. When both S_{av} and σ_S decrease, λ is reduced by a factor noted ω, but is bounded by a minimum value. If S_{av} is improved but σ_S is increased, reflecting a heterogeneous improvements of the realizations, the ensemble update is accepted, but the value of λ is kept unchanged. When S_{av} is increased, the update is rejected and λ is multiplied by a factor of ω.

However, after refinement and restoration of the wavelet coefficients with their original values, a reasonable increase of the global objective function is tolerated. If S_{av} after refinement is greater than the one before optimizing the parameters at the previous resolution, the update is rejected and only half of the frequency components are reintroduced at the next update. When the update is accepted, but the objective function is not improved, the value of λ is not changed (see Algorithm 1 in Appendix 4). A rejection at early stage may reflect a too strong initial attenuation of the coefficients or an incoherent prior information. In this case, incorporation of high frequency parameters inside sensitive areas (such as well neighborhood) from the start of the optimization can help stabilize the behavior of the objective function, but may affect the preservation of the prior information. The approach for selecting λ described here is fairly standard and has worked well on similar history-matching problems [9, 43], but it is clear that the method of Iglesias and coworkers [44] might be advantageous in stabilizing convergence.

In standard approaches, ω is usually constant. In our multi-scale approach, the value of ω is tuned depending on the current resolution. In the first multi-scale iterations, the value of ω is kept low in order to maintain a relatively large λ and limit the update of the large-scale parameters as it is generally not desirable to converge with a reduced parameterization. The value of ω is then increased with refinements and reaches its maximum at the finest resolution (see Algorithm 1 in Appendix 4). In our examples, the initial value of ω is set to 2 and increases linearly to 10 and λ is bounded to 0.01. When the same control of the Levenberg-Marquardt parameter is used with the standard grid-block approach, the optimization struggles to reduce the objective function in the first iterations without improvement of the quality of the output models, whereas the updates of large-scale sensitive parameters allow significant improvements even with a relatively large λ.

5.3 Validation of the method

Re-parameterization and multi-scale methods are generally efficient when the original inverse problem is underdetermined. In order to illustrate the advantages of our approach, two different medium-to-large test cases are used in this paper. The results obtained with our multi-scale method (MS-LM-enRML) are compared with the grid-block-based approximate form of the iterative Levenberg-Marquardt ensemble-based randomized maximum likelihood method (GB-LM-enRML) [9] with distance-based localization of the estimated Kalman gain, which proved to be an efficient method for history-matching problems. In both cases, the same criteria are used to stop the algorithm (see Section 5.2). Since our algorithm is using the LM-enRML to optimize the wavelet coefficients, the benefits of the multi-scale approach can be assessed from the comparison of the two methods.

The quality of the data match and prediction along with the quality of the updated realizations are used as comparison criteria. The magnitude of the data mismatch in the RML objective function (second term in Eq. 2) is evaluated to ensure that the realizations are possible samples from the posterior:

$$S_{av}(\mathbf{m}_j) = \frac{1}{n^e}\sum_{j=1}^{n^e}(g(\mathbf{m}_j) - \mathbf{d^0}_j)^T \mathbf{C_D}^{-1}(g(\mathbf{m}_j) - \mathbf{d^0}_j), \tag{20}$$

where $\mathbf{d^0}_j$ is a realization from the data distribution. If the data mismatch of the realizations is too large, then the likelihood of these being samples from the posterior is nil. When the mismatch is small, the realizations might be samples but further investigations would be necessary to ensure it (see [45]). In any case, it is generally not possible to reach $S_{av} \ll n^d$, even if the uncertainties are overestimated, because the samples $\mathbf{d^0}_j$ are physically not consistent as independence of samples is assumed ($\mathbf{C_D}$ is diagonal).

The quality of the output realizations is difficult to quantify since most realizations of geological models are generated algorithmically. Consequently, it is not possible to compute the probability that the updated realization is a sample from the posterior. However, spatial analysis of the realizations can provide insight on the behavior of the different inversion processes. For this reason, qualitative examination of the realizations (the "eyeball test") along with variogram analysis is performed in this article.

Because the prior ensemble may incorporate other forms of knowledge, such as seismic and/or geological interpretation, it is generally desirable to minimize the magnitude of changes to the models and, in particular, to avoid spurious updates that do not improve the data match. Average cell-by-cell absolute deviation from prior realizations are also computed using the entire ensemble in order to analyze the magnitude and quality of the changes: For each cell i of the reservoir grid and for each property p, a deviation value $\mathrm{dev}_p(i)$ is computed as:

$$\mathrm{dev}_p(i) = \frac{1}{n^e}\sum_{j=1}^{n^e}|p^j(i) - p^j_{pr}(i)|, \tag{21}$$

where $|x|$ is the absolute value of x, $p^j(i)$ and $p^j_{pr}(i)$ are the values of the property p of the realization j in the cell i after and before the optimization process, respectively. Deviation from the *prior* is further summarized by the average deviation ad_p over all the cells:

$$\mathrm{ad}_p = \frac{1}{n}\sum_{i=1}^{n}\mathrm{dev}_p(i), \tag{22}$$

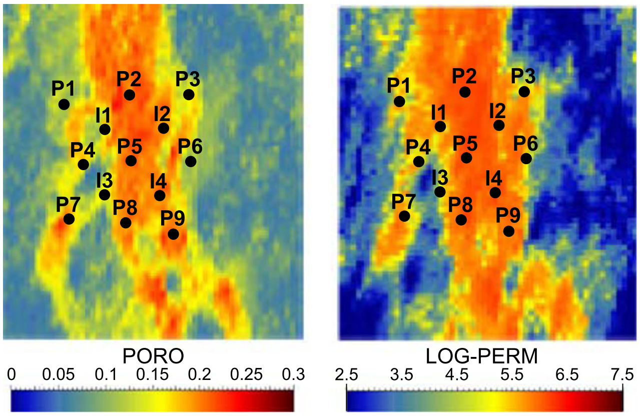


Fig. 6 Reference porosity and permeability fields for test case 1

where n is the number of cells. For a similar quality of data match, the smallest deviation to the model is preferred since in that case, the prior information is better preserved and the updates have been less subject to noise.

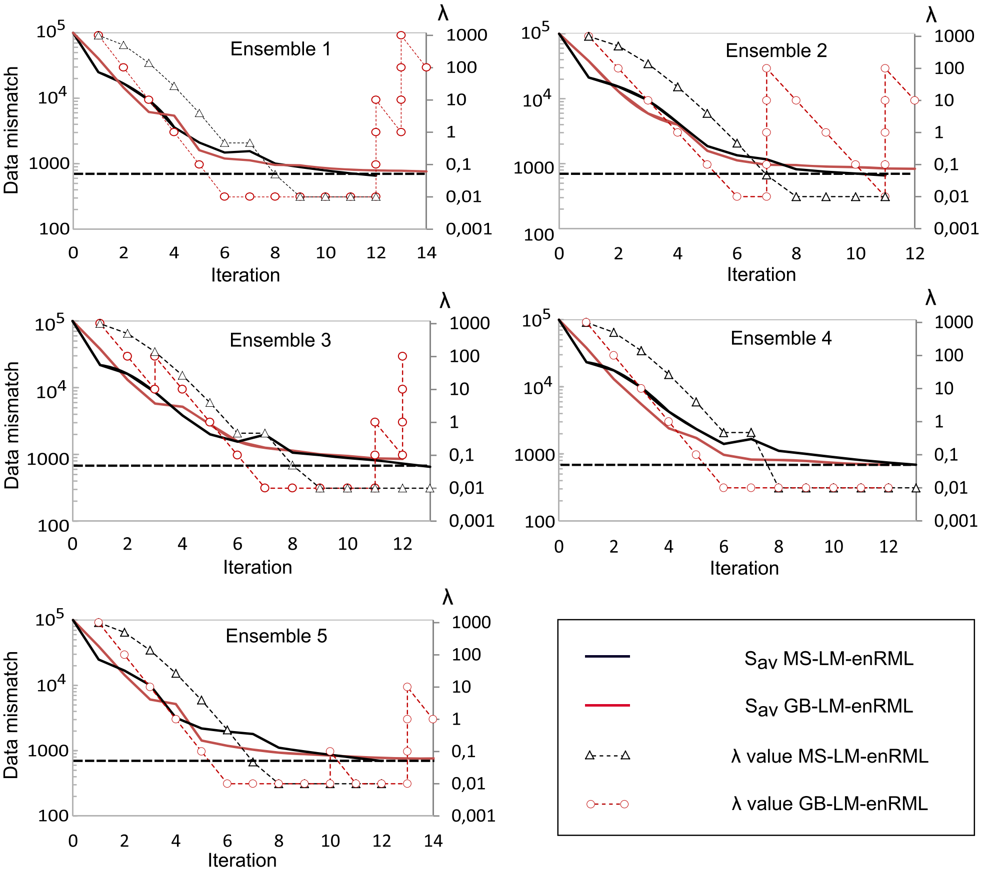
One of the major purposes of this work is to reduce the effects of spurious correlations due to small ensemble size. These spurious correlations result in addition of noise to the realizations. In order to quantify the increase of noise, we analyze the energy of the wavelet coefficients in a similar manner to the compressed sensing processing [46]. We define the energy e^j of a realization j as the $L1$-norm of its wavelet coefficient vector $\boldsymbol{\gamma}^j$, such that $e^j = \sum_{i=1}^{n} |\boldsymbol{\gamma}^j(i)|$, where $\gamma^j(i)$ is the ith wavelet coefficient of γ^j. Thus,

$$re = \frac{1}{n^e} \sum_{j=1}^{n^e} \frac{e^j}{e^j_{pr}}, \tag{23}$$

is the average ratio of energies (independent of the property type) after and before the optimization. When a small amount of noise is added to the realizations, the energy is unchanged and re takes values around 1. When noise is introduced during the inversion, the energy of the realizations increases and re becomes greater than 1.

Execution times (wall clock times) for both methods are similar as they are mainly controlled by the flow simulations of the ensemble. Wavelets transforms and reconstructions in the multi-scale method require additional computations, but the reduction of parameters helps improve the efficiency of the Levenberg-Marquardt updates. The implementations of the two methods are based on the same source code in order to be comparable, except that in the grid-block case, the wavelet transform is replaced by the identity and the optimization starts at the finest resolution.

Fig. 7 The change of the average mismatch (20) and the value λ with iteration for the five ensembles. The number of data is about 700 (*horizontal dashed lines*)



6 Numerical experiments

Both methods are applied on two history-matching test cases. The first problem is a 2D synthetic reservoir including 13 wells with about 8 years of production history. The first 6 years are used to constrain the inversion, and the remaining 2 years are used to test the prediction of the updated ensembles. The second test case is the well-known 3D Brugge benchmark [47] with 10 years of history, including 30 wells and 9 geologically independents layers. The history-matching takes about 3 h in the first case and about o1 day in the second for both MS-LM-enRML and GB-LM-enRML methods, flow simulations being simultaneously run on 25 cores.

Wavelets based on average interpolation of a quadratic polynomial (Appendix 3) are used as basis function in all MS-LM-enRML applications. Distance-based localization and control of the Levenberg-Marquardt parameter λ are applied as described in [9] for the GB-LM-enRML, whereas multi-scale localization and control of λ (Sections 5.1 and 5.2) are used for the MS-LM-enRML. In both cases, a spherical covariance function, $f(.)$, is used in the localization functions (16).

6.1 2D reservoir

6.1.1 Presentation of the problem

The first test is a 2D synthetic case composed of 3355 active cells which are populated by porosity and permeability fields generated from the blending of object-based simulation and sequential Gaussian simulation (SGS). Properties used to generate the synthetic data set are shown in Fig. 6. A probabilistic non-linear relation is used to generate the permeability field from the porosity. Final distributions of the properties are bimodal, mimicking a channelized system with floodplain deposits.

A total of nine producers and four injectors, arranged in a five-spot pattern, are used to generate 8 years of production history. The primary constraints of wells are oil production rates and water injection rates. Bottom hole pressure (BHP) limits of 120 bars are set to all wells. The production scheme includes shut-ins of 1 year at wells P1, P3, P5, and P9 and several buildup periods for all the wells. Producing water cut (WCT), BHP, and gas-oil ratio (GOR) are used to constrain the inversion of the permeability and porosity fields, with a total number of about 700 data points. The standard

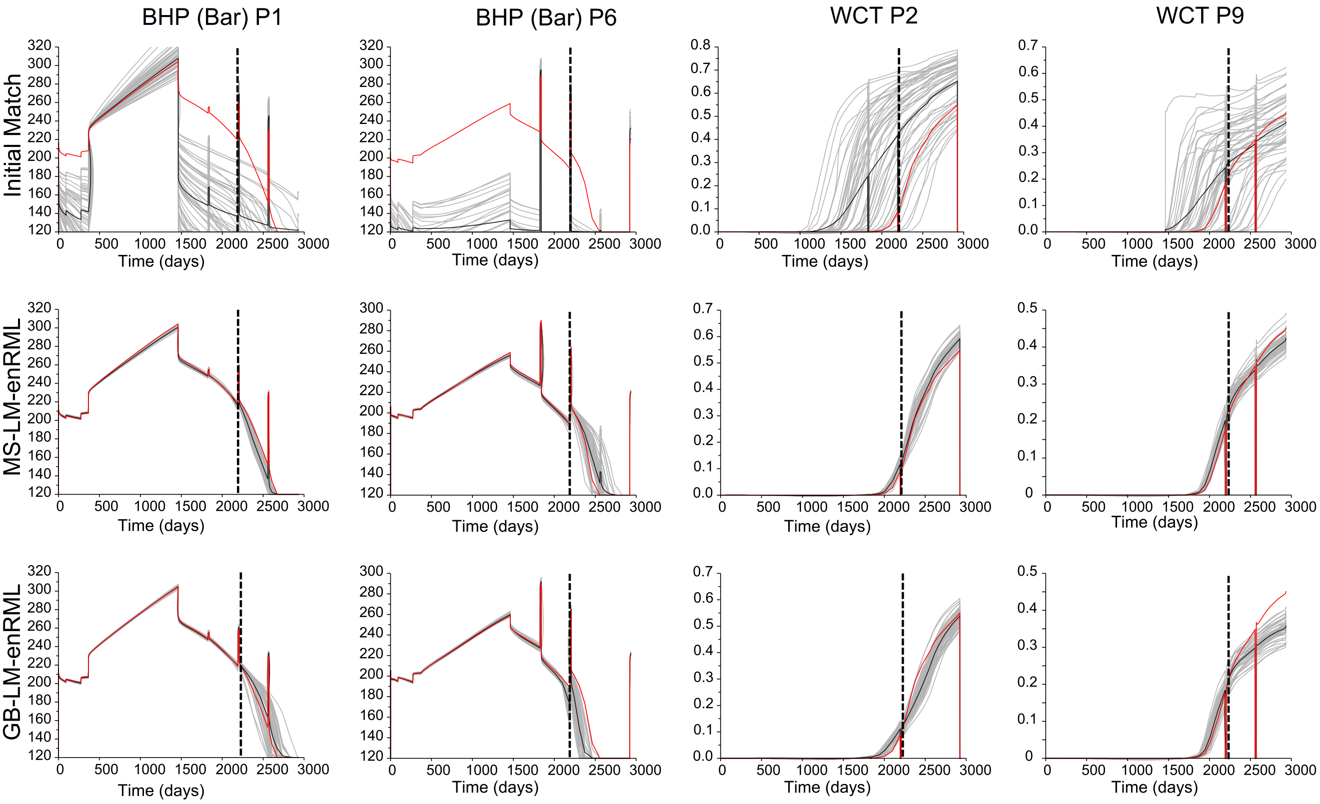


Fig. 8 The match (0 to 2200 days) and prediction (2200 to 3000 days) to water cut and pressure at four wells. The *red curve* represents the observed data. The *gray curves* show the ensemble prediction. The *black curve* is the ensemble average

deviation of noise used for the different data type are 2 bars for BHP, 2.5 % for WCT, and 3 sm^3/sm^3 for GOR.

In both GB- and MS-LM-enRML, a maximum number of 20 updates are permitted using an ensemble of 50 realizations conditioned to noisy log data and generated from object-based simulation and SGS. The orientation of the channels is known a priori with an uncertainty of 30°, such that prior models are already constrained by this information. The initial value of the Levenberg-Marquart damping parameter λ is set to 1000 for the two methods. Due to the relatively small size of the problem (compared to real cases), ensembles of only 50 realizations are used during the optimizations.

6.1.2 Application and comparison of the multi-scale approach

Final results generally depend on the initial ensemble especially when it is small. In order to avoid bias in the analysis of the results, five different ensembles are independently used to assimilate the production data using both GB and MS-LM-enRML methods. Note, however, that all the realizations in the different ensembles have been generated using the same process.

Figure 7 shows the evolution of the data mismatch S_{av} for both methods with the five ensembles. In the first iterations, the decrease of the mismatch for the two methods is quite similar even if the MS-LM-enRML uses a reduced parameterization. Only one or two iterations per resolution are performed before reaching the full parameterization (1 iteration from resolutions 1 to 4 and 2 iterations at resolution 5). The re-introduction of smoothed frequency components at the final refinement step sometimes causes a small increase of S_{av} (ensembles 1, 3, and 4), which is rapidly corrected at the next iteration. After reaching the finest resolution, all the groups using the multi-scale approach stop because the mismatch falls into data uncertainties ($S_{av} < n^d$), whereas most of the grid-block cases converge slightly above this value because the mismatch does not improve anymore.

Figure 8 shows initial and final data matches and predictions for wells P1, P2, P6, and P9. Note that it is compulsory to use an iterative scheme in this case as the non-linearity is

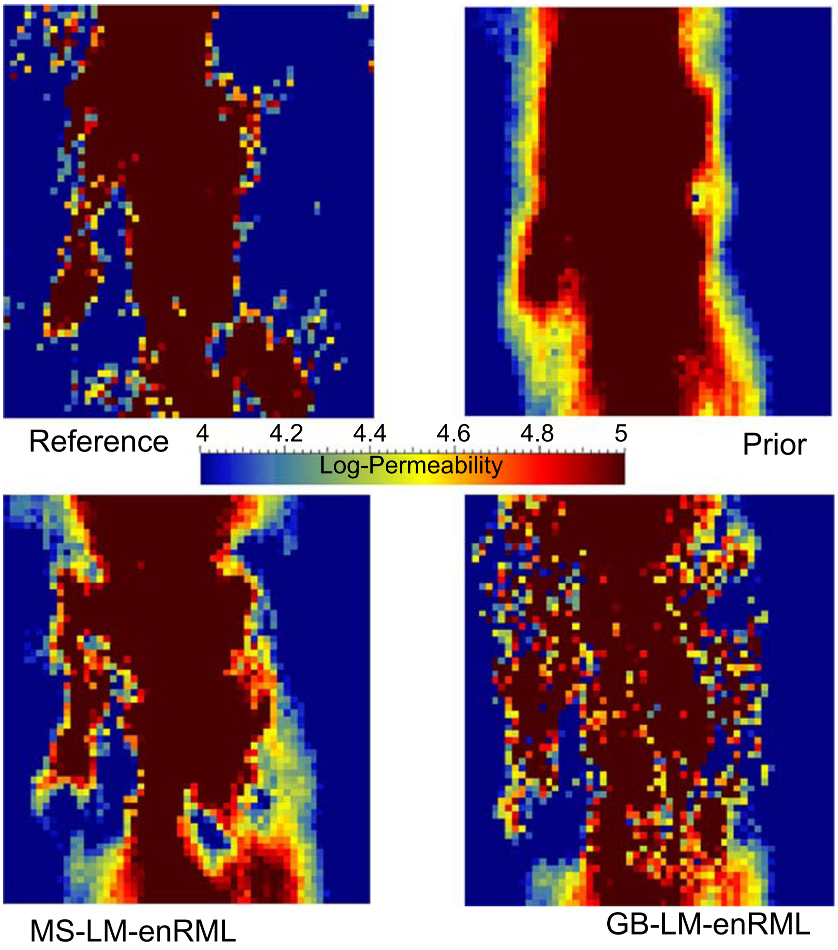


Fig. 9 Comparison of ensemble updated log-permeability averages using MS-LM-enRML (*bottom left*) and GB-LM-enRML method (*bottom right*) with the truth (*top left*) and initial mean (*top right*)

large and the initial realizations are not spanning the data. The results of ensemble number 4 are used in this comparison as that final data match is the most similar between the two methods. A good match of the data is achieved in both cases. Predictions are also similar, but ensembles sometimes fail to correctly predict the evolution of the production. The performance of prediction can slightly change depending on the method, but the analysis of the results with all the ensembles does not bring out major differences between GB- and MS-LM-enRML.

Comparison of output log-permeability averages with the truth is shown in Fig. 9. The color scale is set such that the main flow structures and property contrast, reflecting the two lithologies present in the field, can be easily identified. In both GB- and MS-enRML cases, the shape of the important flow structures are characterized after assimilation of the production data (e.g., permeable branch on the left). However, we can observe that the property contrasts are better defined and less noisy with the multi-scale approach. Figure 10 shows a set of individual realizations sampled from the five ensembles before and after assimilation of production data, and Fig. 11 summarizes the average deviation from prior maps for the porosity and log-permeability including all the realizations of the five ensembles. The final realizations generated using the multi-scale approach better preserve the spatial structure of the prior models and are less noisy while matching the data equally well. Moreover, the important deviations from the prior models are mainly localized where modifications are needed to match the data (i.e., a flow barrier between wells P7 and P8).

Figure 12 shows the variograms, oriented north 0° (orientation of the channels) with a tolerance of 20° and a bandwidth of 20 % of the model extension, corresponding to the realizations of Fig. 10. Posterior MS-LM-enRML variograms remain similar to the prior variograms, whereas GB-LM-enRML variograms are highly affected by the assimilation of the data. Preservation of the properties is only possible because prior models somehow contain prior knowledge (the existence of a channelized system, channel orientation, and log data) which is consistent with the truth. When the prior ensemble is poorly constrained, meaning that no information is used to constrain the realizations, more important perturbations are needed, but the multi-scale approach still helps stabilizing the inversion (see Brugge case, Section 6.2).

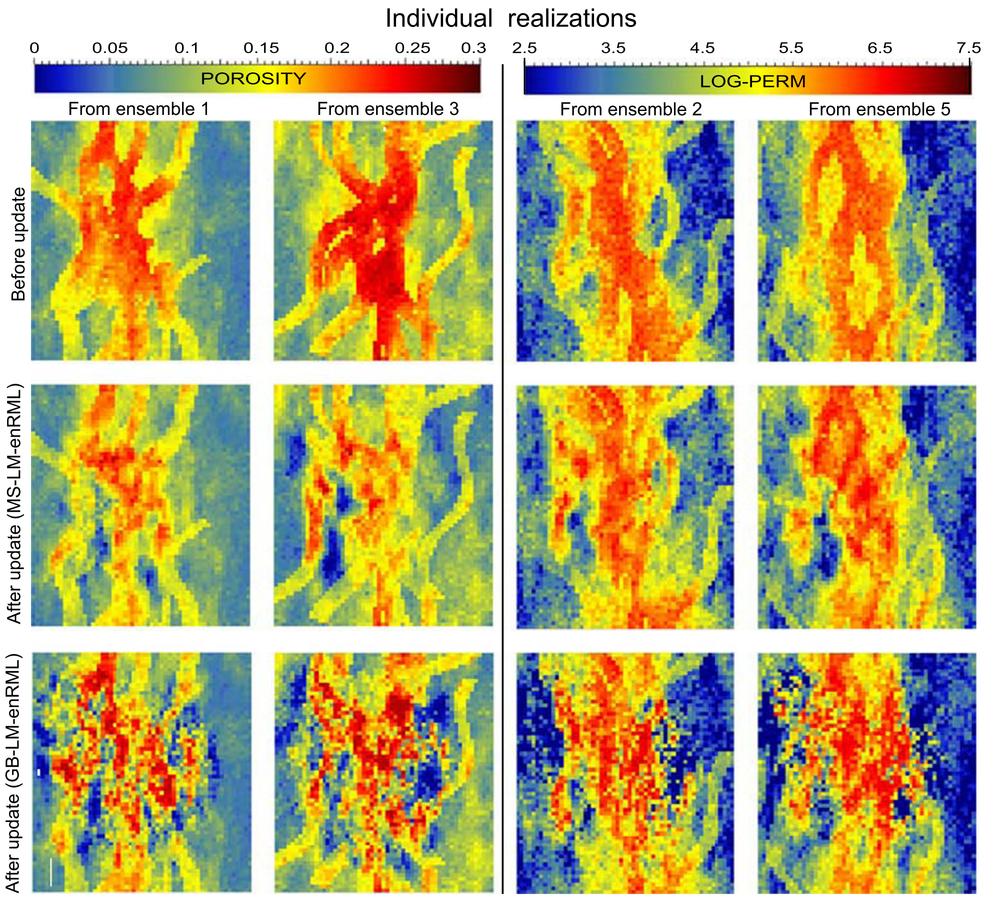


Fig. 10 Individual porosity/log-permeability realization before and after production data assimilation

Fig. 11 Average deviation (cell-by-cell the difference between initial and final model) from the prior computed from all the realizations (five independent ensembles). MS-LM-enRML (*left*) and GB-LM-enRML (*right*)

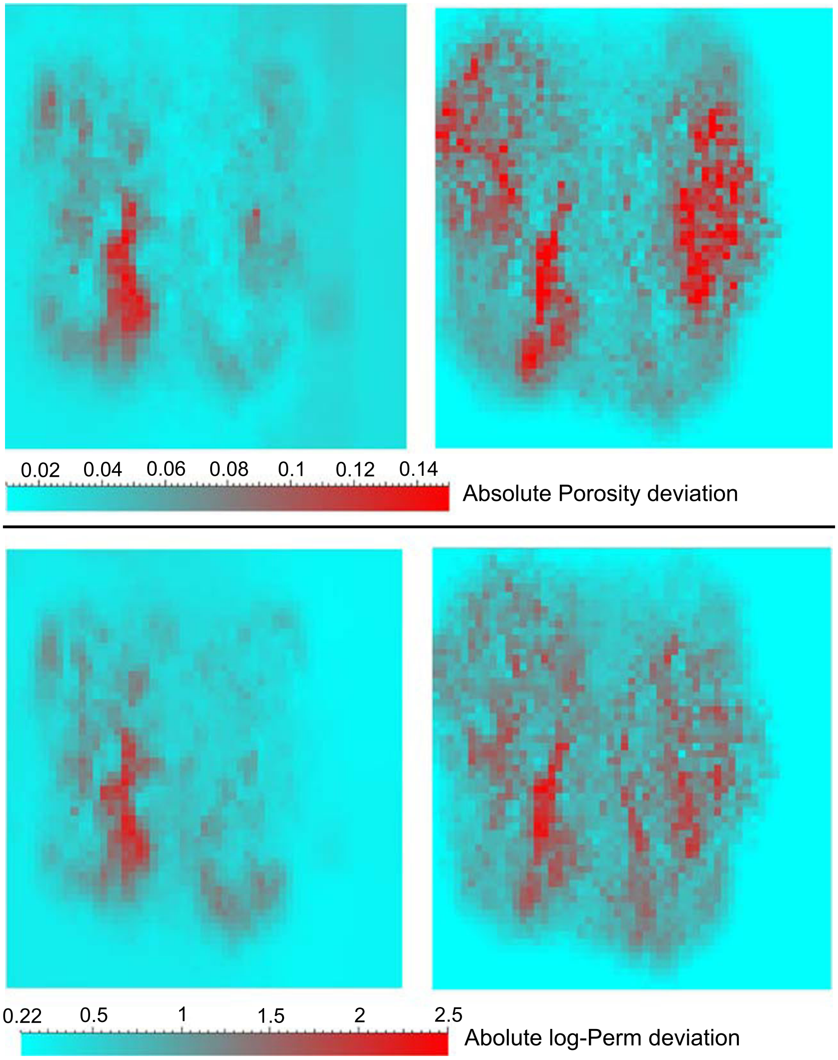
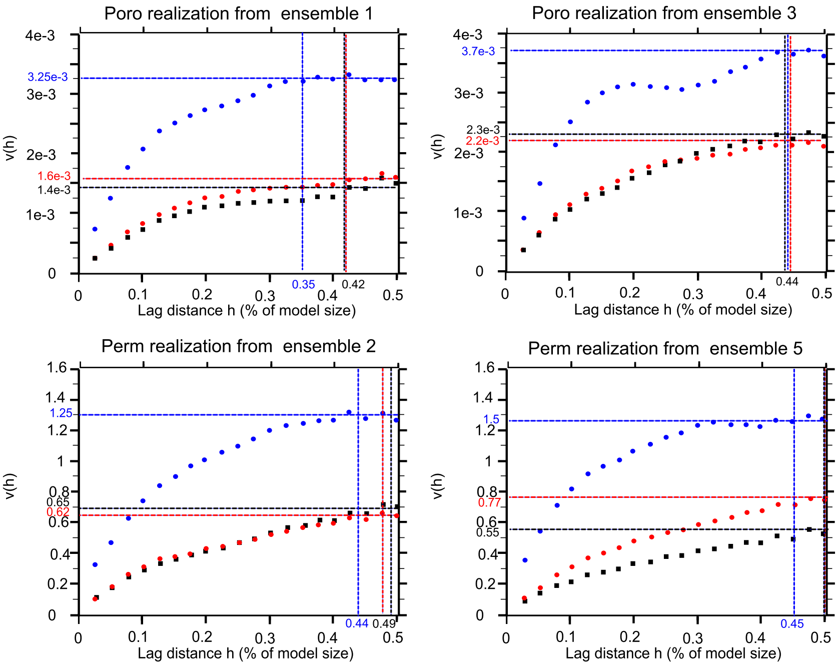


Fig. 12 Variogram analysis (north 0) of prior and posterior porosity and permeability realizations displayed in Fig. 10. *Black squares* represent the variogram of the prior realizations. *Red and blue dots* represent the variogram of the posterior realizations with MS-LM-enRML and GB-LM-enRML methods, respectively. *Vertical and horizontal lines* correspond to the interpreted sills and ranges, respectively



6.1.3 Analysis of the multi-scale algorithm

In this section, we provide examples that illustrate the effects of the adapted truncated wavelet parameterization and the localization processes. Three additional cases are run with the same ensemble of 50 realizations used in the previous section: (case 1) the standard GB-LM-enRML without localization, (case 2) the MS-LM-enRML without any localization, and (case 3) the MS-LM-enRML with only localization in the wavelet space (Section 5.1.2). For each of these cases, the control of the Levenberg-Marquart parameter λ is applied as described in [9]. Comparison is made with two previous examples: (case 4) GB-LM-enRML with localization and (case 5) MS-LM-enRML with space-frequency localization. Details on the localization and optimization controls for cases 4 and 5 are provided in the previous section. The effect of the specific control of λ is not presented here as it appears to be negligible for this test case.

All the cases stop after ten iterations and show good matches with the production data. Figure 13 shows one updated log-permeability realization and the deviation maps for cases 1–3. Table 1 gives the values of $\mathrm{ad}_p, p \in \{\text{log-perm,poro}\}$ and re for all five cases. Although the adaptive truncated multi-scale parametrization (case 2) helps limit the increase of energy (noise) as compared to the standard GB-LM-enRML approach (case 1), it does not suppress the effect of the spurious correlations and does not help preserve the prior realizations by itself. After the last refinements, the parameterization includes coefficients that

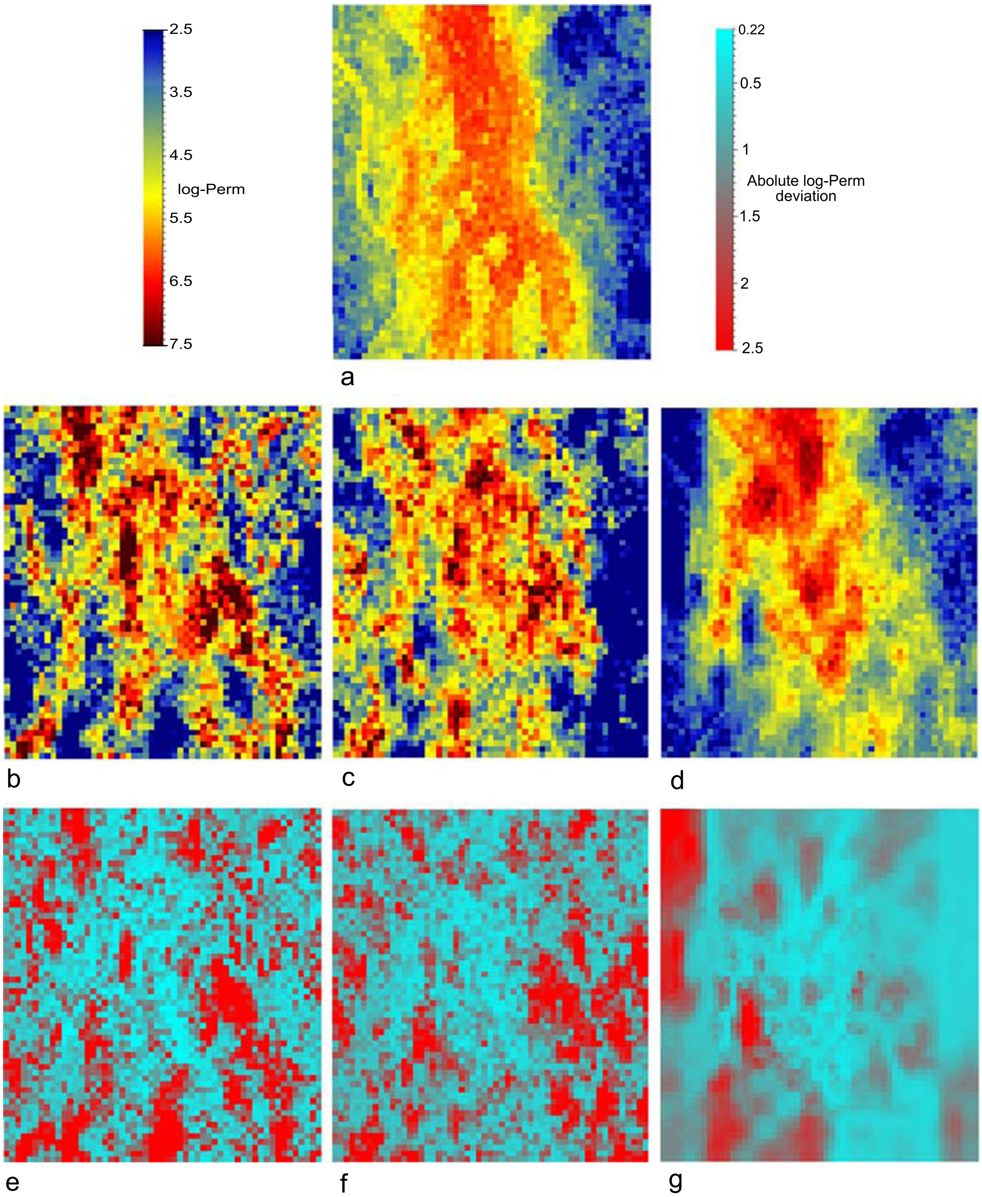


Fig. 13 Analysis of the resulting log-permeability realizations. **a** One *prior* realization. Corresponding updated realization/average cell-by-cell deviation map: **b**,**e** GB-LM-enRML without localization, **c**,**f** MS-LM-enRML without localization, **d**,/**g** MS-LM-enRML with localization in the wavelet space only

Table 1 Average cell-by-cell deviation values and energy ratios (see Section 5.3)

	ad_{Poro}	$ad_{\text{log-perm}}$	re
Case 1	0.060	1.30	2.92
Case 2	0.060	1.16	2.30
Case 3	0.038	0.67	1.12
Case 4	0.041	0.73	1.75
Case 5	0.029	0.47	1.15

Case 1: GB-LM-enRML without localization. Case 2: MS-LM-enRML without localization. Case 3: MS-LM-enRML with wavelet localization only. Case 4: GB-enRML with localization. Case 5: MS-enRML with localization in both the wavelet and grid-block space. Note that for cases 4 and 5, the values are computed using the five ensembles

can be incorrectly correlated to the production data. Moreover, the variance of these coefficients is relatively high as they have not been updated during the previous iterations. Thus, the impact of the spurious correlations is important even though the mismatch reduction at the coarse iterations helps limit the amplitude of the perturbations.

The localization in wavelet space helps limit these spurious updates (case 3), and the increase of the energy is more efficient than the standard grid-block localization (case 4). However, comparable preservations of the prior properties are obtained in cases 3 and 4. As discussed before in Section 5.1.1, we think this is caused by an inappropriate modification of the low resolutions (large scales) induced by the conditioning of well production data measured in regions that do not present the same distribution of properties (non-stationary case). This effect can be observed in the Fig. 13g (top left) where important large-scale perturbations are applied away from the wells. By limiting the regions of updates (case 5 with grid-block localization), it is possible to limit the perturbation of the properties, while matching the data.

6.2 Brugge benchmark case

The Brugge field case is a (139 × 48 × 9) 3D simulation model designed by the Dutch Organization for Applied Scientific Research (TNO). The original geological model is built in a fine resolution grid composed of 20 million grid-blocks. A first upscaling of the petro-physical properties is performed in order to generate production data from 30 vertical wells (20 producers and 10 injectors) with a half million grid-block flow model. Only oil and water phases are present in the reservoir. These data are used to test different optimization algorithms on a further upscaled models that contains 44,550 active cells. More information and optimization results can be found in the study of Peters et al. [47]. The primary constraint of the wells are the production rates (FPR) of 2000 bbl/day for the producers and the water injection rate (WIR) of 4000 bbl/day for the injector. Bottom hole pressure limits of 725 psi are set for all wells and is the active constraint of well P9 throughout the 10 years. The data used in this work include bottom-hole pressure for the 30 wells, water cuts (WCT), and fluid injection/production rates corresponding to 10 years of history for a total number of about 1330 data points. The standard deviation of noise used for different types of data are 30 psi for BHP, 3.5 % for WCT, 2 bbl/day for FPR and WIR if they are active constraint, and 20 bbl/day for FPR of P9. The update parameters are the porosity, lateral and vertical permeabilities, net-to-gross (NTG) ratio, three end points of Corey-type realtive permeability curves, and one initial oil-water contact. For the MS-LM-enRML, the wavelets coefficients of the transformed property fields are included in the optimization, whereas grid-blocks values are used in the GB-LM-enRML. Note that we do not include mean properties of individual layers in the GB-LM-enRML as in [4, 9] as it can be considered as a part of a multi-scale parameterization. It should, however, improve the results of the GB-LM-enRML. The total number of model parameters for the full parameterization is about 178,200. An ensemble of 104 realizations generated from object-based simulation process and SGS (provided by TNO [47]) is used for all the methods. The initial realizations are poorly constrained and therefore very disparate, making the preservation of the models a more challenging task. Initial Corey end points and initial oil-water contact are generated from uniform distributions defined in [30].

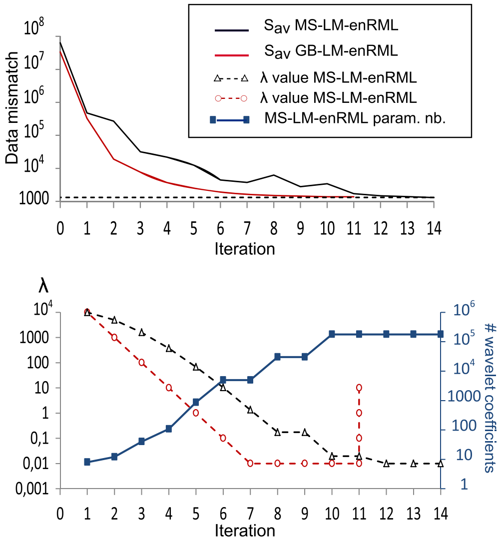


Fig. 14 *Top*: evolution of the data mismatch with iteration for MS-LM-enRML (*black curve*) and GB-LM-enRML (*red curve*). *Bottom*: evolution of the value of λ for both methods and number of parameters included in the MS-LM-enRML optimization (*blue curve*) with iteration

Figure 14 shows the evolution of the objective function (top) and the value of the λ parameter (bottom) with iterations for both methods. The GB-LM-enRML performs better at early iterations and stops with an average data mismatch of 1389 before the MS-LM-enRML because of three unsuccessful updates, whereas the MS-LM-enRML stops after 14 iterations as it falls below data uncertainties with an average data mismatch of 1300. In both cases, the initial value of λ is 10,000 and the lower bound is 0.01. At early iterations, only a few parameters are included in the MS-LM-enRML, which explains the difference of efficiency, but a significant improvement of the match is obtained. We can see that after the two last refinements, the re-introduction of the smoothed components causes a slight increase of the objective function. The dissimilarity of the prior realizations (generated from different methods without secondary data conditioning) may explain the need of individual gradient after re-introduction of the smoothed components. With the full grid-block parameterization, this effect is attenuated by the linear projection of all the parameters during the first updates, leading to a reduction of the variability at all frequencies, but damaging the prior model at the same time.

Figure 15 shows the initial and final match to bottom-hole pressure, fluid production rate, and water cut at three producers for the MS-LM-enRML and GB-LM-enRML. In both cases, very good matches are obtained and it is not possible to discriminate one method from another.

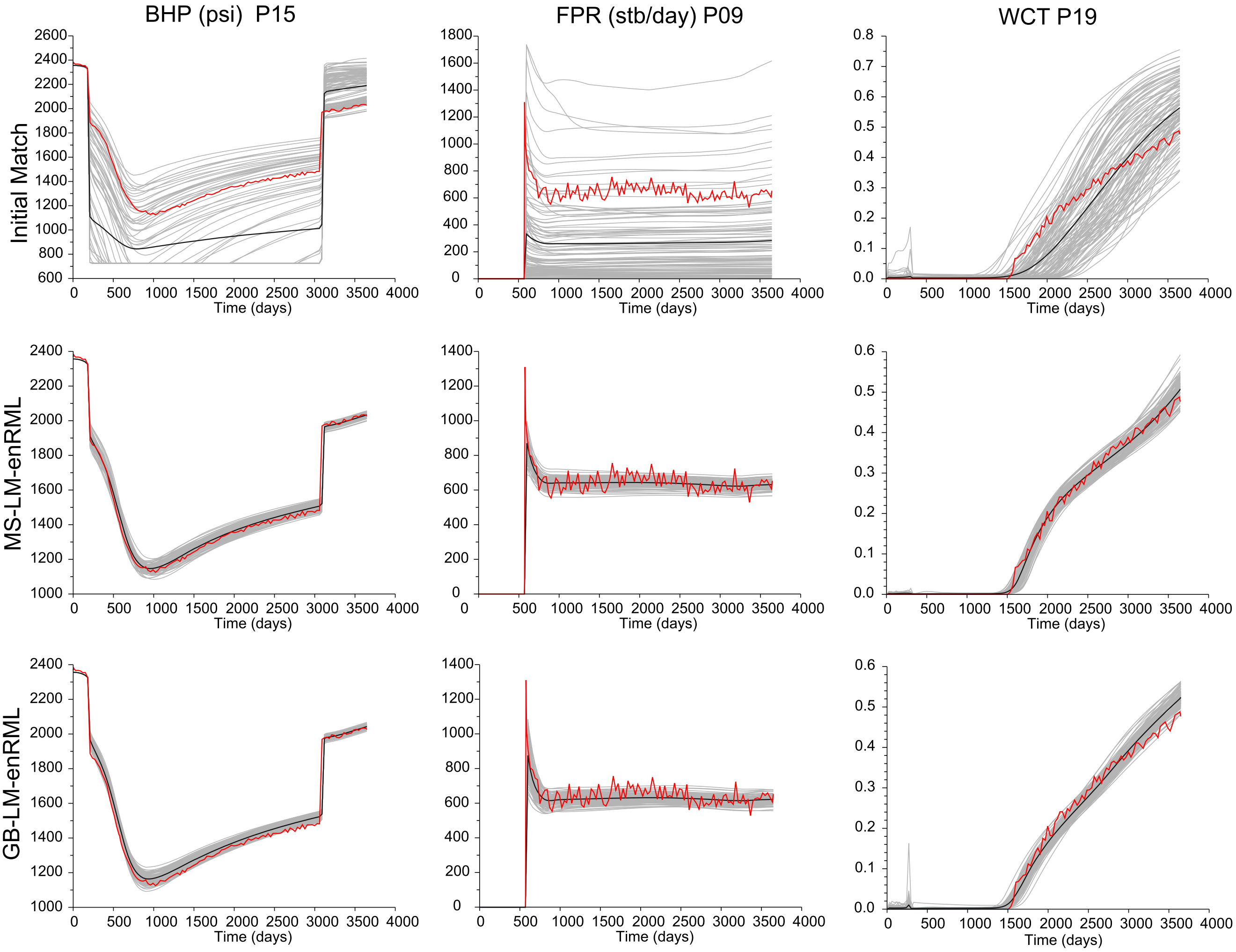


Fig. 15 The match to bottom-hole pressure, fluid production rate, and water cut at three wells. The *red curves* show the observed data. The *black and gray curves* show the average and individual predictions of the ensemble, respectively

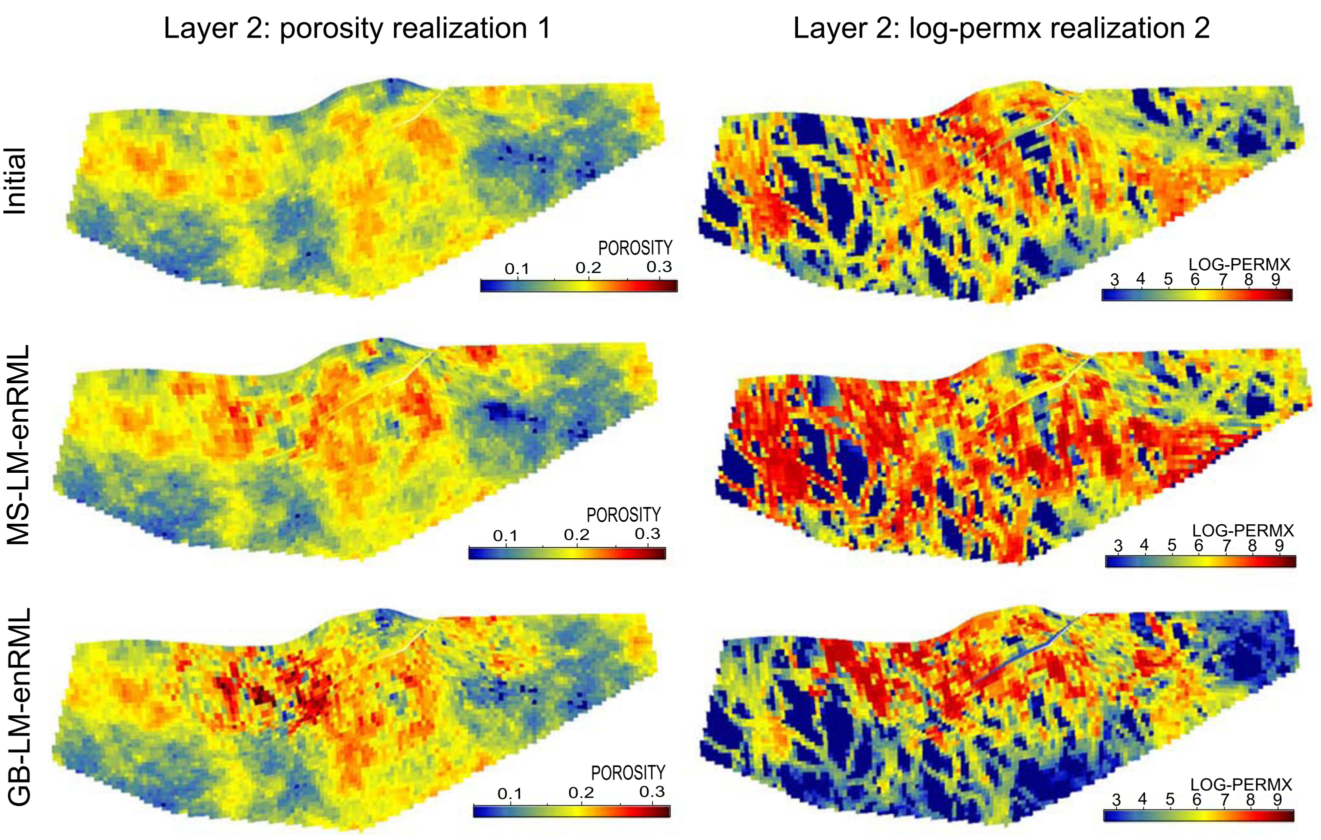


Fig. 16 One porosity realization (*left*) and one log-permeability realization (*right*) from the initial ensemble (*top row*) and the corresponding updated realizations using MS-LM-enRML (*middle row*) and GB-LM-enRML (*bottom row*). Only properties of the second layer are shown

Figure 16 shows the porosity and log-permeability of two different realizations (generated from different processes) from the initial ensemble and the corresponding update realizations using MS-LM-enRML and GB-LM-enRML, and the average absolute deviation from initial porosity maps are shown in Fig. 17. Only the second layer is displayed as it has an important control on the flow response and important updates are performed. Average deviations from prior models and energy ratios indicators are also given in Table 2. For an equivalent data match, the multi-scale

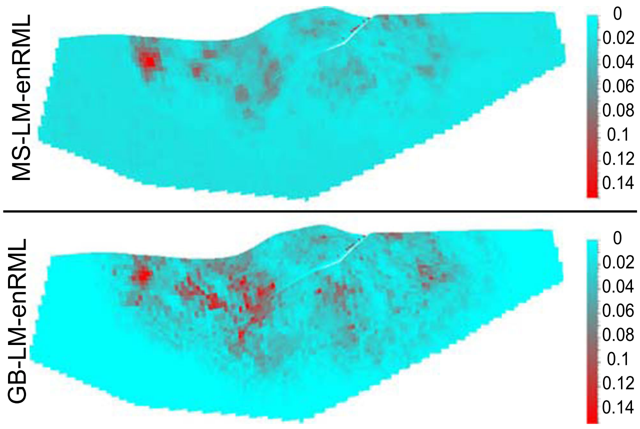


Fig. 17 Average porosity deviation (cell-by-cell the difference between initial and final model) from the prior computed from all the realizations. MS-LM-enRML (*top*) and GB-LM-enRML (*bottom*)

Table 2 Average cell-by-cell deviation values and energy ratios for the Brugge case (see Section 5.3)

	$\mathrm{ad}_{\mathrm{Poro}}$	$\mathrm{ad}_{\mathrm{log\text{-}perm}}$	re
GB-LM-enRML	0.010	0.710	1.180
MS-LM-enRML	0.008	0.450	0.940

GB-LM-enRML grid-block approach with distance based localization, *MS-LM-enRML* multi-scale approach with space-frequency localization approach clearly performed more realistic updates by avoiding the appearance of unnecessary high or low property values or addition of noise and minimizing the amplitude of the changes. We can observe that the energy is decreasing in the MS-LM-enRML case. This might be explained by a smoothing of some of the object-based realizations caused by the Gauss-Newton updates.

By making low frequency (large scale) modifications of the property fields at early iterations, the adaptive multi-scale localization approach ensures that subsequent modifications at the fine scale are relatively small so that the magnitude of spurious high-frequency updates are reduced and features of the initial property field are better preserved (e.g., log-permeability of the channels in Fig. 16, realization 2).

7 Discussions and conclusions

The adaptive multi-scale method based on second-generation wavelet parameterization appears to be useful for stabilizing the inversion and avoiding spurious effects related to ensemble-based optimization methods. In the examples presented here, a better preservation of the prior is obtained with the multi-scale approach compared to the grid-block-LM-enRML, while matching the production data equally well. However, our multi-scale approach is based on the assumption that the flow is more affected by the low than the high frequencies. In reservoirs, where the fluid flow is mainly controlled by fine-scale structures (e.g., fractured or laminated reservoirs), the multi-scale approach might fail to recover the correct structure of the reservoir. In this case, it might be useful to include high frequencies from the beginning of the optimization (e.g., for laminated reservoirs, individual 2D transforms can be used for each layers).

The multi-scale approach and the smoothing of the realizations tend to make the problem more linear by attenuating the effects of the high frequencies and reducing the number of parameters at early stage [48]. Global updates of the property fields are then possible without the influence of small-scale spurious correlations. Significant mismatch reduction is obtained by optimizing only few large-scale parameters at the first iterations, resulting in smaller perturbations of the finer scales after refinements. However, inter-scales re-introduction of the a priori smoothed components may slightly affect the convergence as compared with the standard approach. In the current implementation of the algorithm, the smoothing simply depends on the resolution of the parameters. However, a more advanced initial smoothing, based on the localization of the parameters and/or prior uncertainties, might be more efficient. Indeed, smoothing should first be applied on sensitive high frequencies (e.g., close to wells) and becomes less essential in insensitive areas. Moreover, when the uncertainty of the model has already been reduced by integration of other sources of data, the smoothing should be less important since the different realizations are more similar. The flow impact of certain features of the prior model should also be taken into account during the first updates, assuming that the different sources of information are consistent.

The regularization of the Kalman gain using localization can be a fairly complex task and a wrong localization can lead to inconsistent updates and insufficient data match. With the multi-scale approach, the effect of the localization at early iterations is limited because only sensitive parameters are included in the optimization. However, when the prior model contains information at all frequencies, it is generally required to refine the parameterization to the finest level and the need for regularization becomes essential. For this reason, a localization of the wavelet coefficients, depending on their location and associated frequency, is applied in this work in order to stabilize the inversion and preserve prior information. Nevertheless, similarly to the grid-block localization, the implementation of the regularization in the wavelet space is not always trivial. Further analysis based on cross-covariance matrices [29] or bootstrap sampling [33] might help define more appropriate multi-scale localization functions.

Preserving the prior information is one of the major purposes of this work, and the multi-scale approach helps restrict the updates to sensitive areas. Qualitative comparisons and quantitative spatial analysis are used in this article to check this preservation. When the prior ensemble integrates quantitative information, such as seismic data, it might be suitable to perform posterior quality controls using these data. This process, however, might not be straightforward as the different data might be expressed in different domains (time/depth) and sampling intervals.

If little information is used to generate the prior ensemble, the successive re-introduction of the initial smoothed frequencies may affect the efficiency of the optimization, especially when the realizations are disparate, as the local quadratic approximation of the objective function might not be appropriate for each individual realization of the ensemble. In this case, methods using full parameterization might be more efficient in term of convergence speed, but wavelet re-parameterization might still be useful in identifying sensitive parameters for regularization or to avoid over-parameterization and addition of noise.

Acknowledgments The authors are thankful for the funding of CGG and the support of the IRIS/CIPR cooperative research project ”Integrated Workflow and Realistic Geology” which is funded by industry partners ConocoPhillips, Eni, Petrobras, Statoil, and Total, as well as the Research Council of Norway (PETROMAKS). The authors would like to thank Schlumberger and Paradigm for providing Eclipse and Gocad academic licenses, respectively. The first author would like to thank Jeanne Pellerin for her valuable comments and corrections.

Appendix 1 Lifting scheme

In the following, we consider an original signal $\lambda_{r,k}$, defined at a finer scale r in all spatial locations k. The construction of the second generation wavelets is based on the lifting scheme [37], which is composed of three main reversible operators: split, predict, and update (Fig. 18). It should be made clear that these operators are not related to the enKF update and prediction steps.

First, the splitting operator, S, subdivides the initial set $\lambda_{r,k}$ into two different subsets. The correlation between the two subsets should be maximized. Usually, the subdivision is simply done between odd and even samples as neighbors have more chances to be correlated. A new sequence $\lambda_{r-1,k}$ at resolution $r-1$ is obtained by taking the even samples:

$$\lambda^*_{r-1,k} = \lambda_{r,2k}. \tag{24}$$

The loss of information is conserved in the odd samples, which correspond to the detail coefficients:

$$\gamma^*_{r-1,k} = \lambda_{r,2k+1}. \tag{25}$$

At this step, the algorithm performs a poor decorrelation of the signal and leads to aliasing effects. Proper decorrelation and subsampling are handled by the predict and update operators. The predict step, $\mathcal{P}$, uses the correlation between the odd and even subsets. The basic idea is to try to predict the odd samples by only using the even samples. Generally, the predict step must be independent of the data values in order to remain reversible (Fig. 18). The predicted value is then subtracted from the original odd sample, such as:

$$\gamma_{r-1,k} = \gamma^*_{r-1,k} - \mathcal{P}(\lambda^*_{r-1,k}). \tag{26}$$

The detail coefficients $\gamma_{r-1,k}$ reflect how good the prediction is. The aim of the predict step is to produce detail coefficients as small as possible. In this case, the original signal would be well characterized only with the subsampled coefficients. The predict step conditions the quality of the decomposition. Any prediction operators can be used and specific predictors can be built for specific applications, for example, taking into account the sampling of the data (Appendix 2) or external weights [38]. The predict step produces a set of filter coefficients, which are adapted to the localization of the data.

At this stage, the subsamples $\lambda^*_{r-1,k}$ have not been modified and are not necessarily representative of the original signal and strong aliasing problems could appear. It is often desirable to preserve some properties of the original data, such as the mean, at different scales. This is done by the update operator $\mathcal{U}$. The update operator can be compared to the action of the scaling functions in the case of first generation wavelets since it produces a coarser version of an original signal. The detail coefficients $\gamma_{r-1,k}$ are used to update the values of the sub-sampled signal:

$$\lambda_{r-1,k} = \lambda^*_{r-1,k} + \mathcal{U}(\gamma_{r-1,k}). \tag{27}$$

The update step maintains the moments (mean, first moment, …) of the sequences λ_r up to an order equal to the number of samples used in the prediction step. A set of coefficients, known as lifting coefficients, are deduced from the preservation of the moments. An efficient method of computation can be found in [49] in the case of polynomial interpolating scaling functions. The complete wavelet transform is performed by recursively applying the lifting scheme to the samples λ_r (similarly to Fig. 1) until the number of samples at one scale becomes smaller than the number of samples needed in the prediction step.

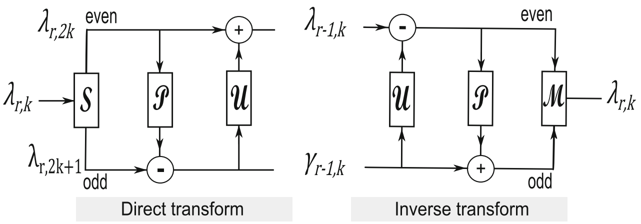
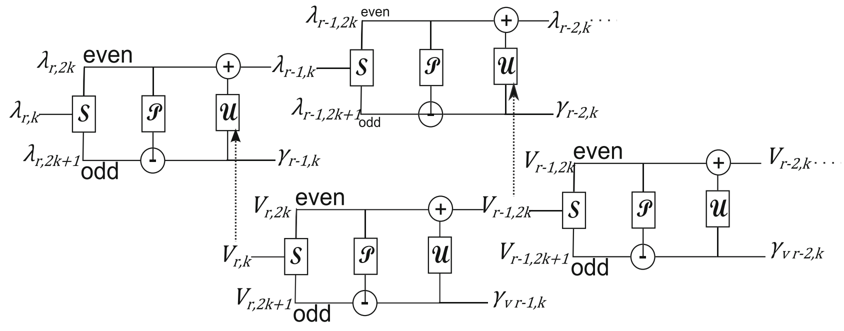


Fig. 18 The lifting scheme: decomposition of the signal. Split (S), predict ($\mathcal{P}$), update ($\mathcal{U}$), and merge ($\mathcal{M}$)

Appendix 2 Grid-adaptive Haar transform

Grid-adapted Haar transform integrates the grid cell volumes in the prediction and update steps in order to compute a weighted average sequence. This transform, also called the unbalanced Haar transform, is the generalization of the Haar wavelet to the second-generation setting [50]. It belongs to the class of wavelets based on average interpolating subdivision (Appendix 3). Values $\lambda_{r,k}$ of every

Fig. 19 Grid-adaptive Haar transform. *Top*: unbalanced Haar transform. The update step at resolution r needs volumes which are given by the sequences V_r. *Bottom*: sequences V_r are available during the Haar transform of the cell volume property



sequence in the decomposition represent an average over an interval $[x_{r,k}, x_{r,k+1}]$. In the 3D discrete wavelet transform, it can be useful to associate each $\lambda_{r,k}$ as average of a grid volume $V_{r,k}$. Before wavelet transform, the weights $V_{r,k}$ are the cell volumes of the grid and it directly follows that,

$$V_{r-1,k} = V_{r,2k} + V_{r,2k+1}. \tag{28}$$

The aim of the unbalanced Haar wavelet is to preserve the weighted average:

$$\lambda_{r-1,k} = \frac{V_{r,2k} \cdot \lambda_{r,2k} + V_{r,2k+1} \cdot \lambda_{r,2k+1}}{V_{r-1,k}}. \tag{29}$$

The lifting scheme version of the unbalanced Haar wavelet starts exactly like the classical Haar wavelet. The splitting operator divides the odd and even samples. The prediction step uses only one even sample to predict the odd. The only polynomial which can be constructed is the constant polynomial. Therefore, the prediction operator returns the even sample value and the detail coefficient is given by,

$$\gamma_{r-1,k} = \lambda_{r,2k+1} - \lambda_{r,2k}. \tag{30}$$

The update step preserves the weighted average by the such as

$$\begin{aligned} \lambda_{r-1,k} &= \lambda_{r,2k} + \frac{V_{r,2k+1} \times \gamma_{r-1,k}}{V_{r-1,k}} = \\ &\frac{(V_{r,2k} + V_{r,2k+1})\lambda_{r,2k} + V_{r,2k+1}(\lambda_{r,2k+1} - \lambda_{r,2k})}{V_{r-1,k}} = \\ &\frac{V_{r,2k} \cdot \lambda_{r,2k} + V_{r,2k+1} \cdot \lambda_{r,2k+1}}{V_{r-1,k}}. \end{aligned} \tag{31}$$

Instead of computing each weight $V_{r,k}$ directly from the grid during the decomposition and reconstruction, a usual Haar transform of the cell volume property can be performed in parallel of the unbalanced Haar transform (Fig. 19).

The unbalanced Haar wavelet is useful when the sampling is irregular which is typically the case in realistic geomodels where the cell volumes can vary a lot. Moreover, this transform can handle the presence of dead cells (inactive or zero volume cells) without increasing the number of active parameters.

Appendix 3 Wavelets based on average-interpolating subdivision

Average-interpolating subdivision is based on the idea that samples of a given sequence $\lambda_{r,k}$ are not subsamples of an original signal but are local averages over intervals.

Therefore, for a given sequence to reconstruct (Fig. 20), using the coarse sequence of $\lambda_{r-1,k}$ a prediction operator models the underlying integrated function in order to interpolate the samples at a finer scale. If the underlying function is a polynomial, its primitive will also be a polynomial. In a first time, the prediction step will find this primitive. Using this primitive, it is then possible to find, for a given coarse sample, the two samples of the signals at the finer scale, which follow the underlying polynomial. Then, the difference of these two samples is given to the Haar reconstruction. If the detail coefficient is null, the Haar reconstruction will reform the modeled polynomial, which will produce a smooth transition (see [38] for more details). The Haar transform ensures that local averages of the property are preserved at coarse scales. If the unbalanced Haar wavelet is used, the weighted average is preserved which leads to a better upscaled property.

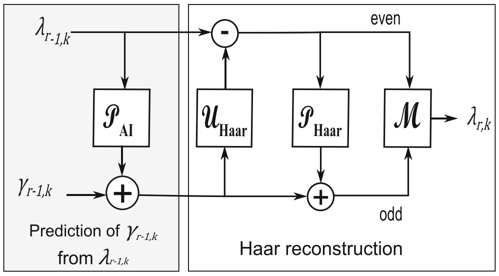


Fig. 20 Wavelet based on average-interpolating subdivision (improved Haar transform), reconstruction step (modified from [38])

Appendix 4 Algorithm

Algorithm 1: The Multi-Scale Lm-enRML algorithm. For sake of simplicity, only spatial variables are considered here, which simplifies the transformation of the ensemble of parameters

```
Input: $\mathbf{M^{pr}}, d_{obs}, r_0, \mathbf{A^{gb}}, \mathbf{A^{wt}}, \omega_0, \varepsilon_s$
Output: $\mathbf{M^{up}}$
begin initialization
    $r \leftarrow r_0$;
    $l = 1$; $SFact = 1$;
    $\mathbf{D}^0 \leftarrow d_{obs} + \varepsilon$, with $\varepsilon \sim \mathcal{N}(0, \mathbf{C_D})$;
    $\boldsymbol{\Gamma}_0 \leftarrow W \cdot \mathbf{M^{pr}}$;
    $I_r \leftarrow \{I_\gamma\}_{0,r_0}$;       // Initial set of coef. indices
    $n^* \leftarrow |I_r|$;                   // Cardinality of the set
    $\mathbf{S}_r \leftarrow \mathbf{S}_0$ ;                      // Initial smoothing
    $\boldsymbol{\Gamma}^s \leftarrow \mathbf{S_r} \circ \boldsymbol{\Gamma}_0$;
    $\boldsymbol{\Gamma}^* \leftarrow \boldsymbol{\Gamma}^s[q(I_r), *]$ ;               // Sub-matrix of $\boldsymbol{\Gamma}^s$
    $\mathbf{D}_0 \leftarrow g(\mathbf{W}^{-1} \cdot \boldsymbol{\Gamma}^s)$; $Sav_0 \leftarrow (\mathbf{D}_0 - \mathbf{D}^0)^T \mathbf{C_D} (\mathbf{D}_0 - \mathbf{D}^0)$;
    $Sav_{r_0} \leftarrow Sav_0$;
    $\lambda \leftarrow 10^{Floor(log_{10}(\frac{Sav_0}{2N^d}))}$; $\omega_{r_0} \leftarrow \omega_0$;
end
while $r \leq N^r$ do
    while $l \leq l_r^{max}$ do
        if $l = 1$ and $r \neq r_0$ then Algorithm 2
            Refinement($\boldsymbol{\Gamma}^s_{r-1}, I_{r-1}, \mathbf{S}_{r-1}, SFact, \varepsilon_s$);
        end
        $\mathbf{K} \leftarrow$ KalmanGain($\boldsymbol{\Gamma}^*_l, \mathbf{D}_l, \lambda_l$);          // see [9]
        $\mathbf{K} \leftarrow$ MSLocalization($\mathbf{K}, \mathbf{A}_r^{gb}, \mathbf{A}^{wt}, I_r$);
        // Algorithm 3
        $\delta\boldsymbol{\Gamma}^* \leftarrow -\mathbf{K}(\mathbf{D}_l - \mathbf{D}^0)$;
        $\boldsymbol{\Gamma}^*_l \leftarrow \boldsymbol{\Gamma}^*_{l-1} + \delta\boldsymbol{\Gamma}^*$; $\mathbf{D}_l \leftarrow g(\boldsymbol{\Gamma}^s)$;
        $Sav_l = (\mathbf{D}_l - \mathbf{D}^0)^T \mathbf{C_D} (\mathbf{D}_l - \mathbf{D}^0)$;
        if $Sav_l \leq Sav_{l-1}$ then
            $SFact = 1$; $l = l + 1$;
            if StopCriteria($r, Sav_l$) then
                Break;
            else
                $\lambda_{l+1} \leftarrow \lambda_l / \omega_r$;
            end
        else
            if $l = 1$ and $r - 2 \leq r_0$ then post refinement
            special case
                if $Sav_l \leq Sav_{r-2}$ then
                    $l = l + 1$; $SFact = 1$;
                else
                    $\lambda_l \leftarrow \lambda_l \times \omega_r$;
                    $\mathbf{D}_l \leftarrow \mathbf{D}_{l-1}$;
                    $SFact = SFact/2$;
                end
            else
                $\lambda_l \leftarrow \lambda_l \times \omega_r$;
                $\boldsymbol{\Gamma}^*_l \leftarrow \boldsymbol{\Gamma}^*_{l-1}$;
                $\mathbf{D}_l \leftarrow \mathbf{D}_{l-1}$;
            end
        end
    end
    $Sav_r \leftarrow Sav_{l-1}$; $Sav_0 \leftarrow Sav_{l-1}$;
    $\mathbf{D}_0 \leftarrow \mathbf{D}_{l-1}$;
    $\omega_{r+1} \leftarrow \omega_r + \frac{10 - \omega_{r_0}}{r_{max} - r_0}$;
    $r = r + 1$; $l = 1$;
end
```

Algorithm 2: Refinement function

```
Input: $\boldsymbol{\Gamma}^s_{r-1}, I_{r-1}, \mathbf{S}_{r-1}, SFact, \varepsilon_s$
Output: $\boldsymbol{\Gamma}^s_r, I_r, \mathbf{S}_r$
$I_r \leftarrow I_{r-1} \cup \{i_\gamma\}_r$; $n^* \leftarrow |I_r|$;
$\mathbf{S}_{tmp}[n^*, n^e]$ ;                          // Temporary matrix
for $e = 1$ to $e = n^e$ do
    for $p = 1$ to $p = n$ do partial re-introduction for all
    frequencies
        $\mathbf{S}_{tmp}(p, e) = \min(\mathbf{S}_{r-1}(p, e)^{-1}, \varepsilon_s)/SFact$;
    end
    forall the $p \in I_r$ do
        $\mathbf{S}_{tmp}(p, e) = \mathbf{S}_{r-1}(p, e)^{-1}/SFact$;
    end
end
$\boldsymbol{\Gamma}^s_r \leftarrow \mathbf{S}_{tmp} \circ \boldsymbol{\Gamma}^s_{r-1}$;
$\mathbf{S}_r \leftarrow \mathbf{S}_{tmp} \circ \mathbf{S}_{r-1}$;
```

Algorithm 3: MSLocalization: multi-scale Kalman Gain Localization function

```
Input: $\mathbf{K}, \mathbf{A}_r^{gb}, \mathbf{A}^{wt}, I_r$
Output: $\mathbf{K}$
$\mathbf{K} = \mathbf{A}_r^{wt} \circ \mathbf{K}$ ;      // Localization in wavelet space
$\mathbf{K}^{gb} \leftarrow zeros(n, n^e)$;          // Temporary Kalman gain
for $d = 1$ to $d = n^d$ do
    forall the $p \in I_r$ do
        $\mathbf{K^{gb}}(p, d) = \mathbf{K}(p, d)$;
    end
end
$\mathbf{K^{gb}} = \mathbf{W}^{-1} \cdot \mathbf{K^{gb}}$;
$\mathbf{K^{gb}} = \mathbf{W} \cdot (\mathbf{A}_r^{gb} \circ \mathbf{K^{gb}})$;
$\mathbf{K} \leftarrow \mathbf{K^{gb}}[q(I_r), *]$ ;                 // extract sub-matix
```

λ	Levenberg-Marquardt regularization parameter
n^e	number of realizations
n^d	number of data
n^*	current number of parameters to optimize
r	maximum resolution of the optimized parameters γ^*
l_r	iteration index at resolution r
l_r^{max}	maximum number of iteration for terminating iteration at resolution r
ω_r	factor for changing λ with the resoltuion r
SFact	control factor of unsmoothing
M^{pr}	ensemble of prior model variables
M^{up}	ensemble of updated model variables
Γ^{pr}	ensemble of prior wavelet coefficients
Γ^s	ensemble of smoothed wavelet coefficients
Γ^*	ensemble of wavelet coefficients used in the optimization
$\mathbf{W}$	direct wavelet transform
$\mathbf{W}^{-1}$	inverse wavelet transform

I_r	Set of wavelet coefficient indices included in the optimization after reaching resolution r
Floor(.)	return the largest previous integer
g(.)	function relationship between model variables and predicted data
$\mathbf{D}_l = g(W^{-1} \cdot \Gamma_l^s)$	
$\text{zeros}(n_r, n_c)$	generate a $(n_r \times n_c)$ zero matrix
$\mathbf{S}$	Smoothing of wavelet coefficients
Sav_l	Average objective function at the l^{th} iteration
Sav_r	Best value of the objective function at the r^{th} resolution
$\mathbf{d}_{obs}$	Observed data
$\mathbf{D}^0$	ensemble of perturbed data
$\mathbf{A}_r^{gb}$	Grid-block localization matrix at resolution r
$\mathbf{A}_r^{wt}$	Wavelet localization matrix at resolution r
$\mathbf{C}_D$	covariance of data noise
$\mathbf{D}_l$	ensemble of simulated data

References

1. Evensen, G.: The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dyn. **53**(4), 343–367 (2003)
2. Aanonsen, S.I., Nævdal, G., Oliver, D.S., Reynolds, A.C., Vallès, B.: The ensemble Kalman filter in reservoir engineering—a review. SPE J. **14**(3), 393–412 (2009)
3. Oliver, D.S., Chen, Y.: Recent progress on reservoir history matching: a review. Comput. Geosci. **15**(1), 185–221 (2011)
4. Chen, Y., Oliver, D.S.: Multiscale parameterization with adaptive regularization for improved assimilation of nonlocal observation. Water Resour. Res. **48**(4) (2012)
5. Bhark, E., Jafarpour, B., Datta-Gupta, A.: An adaptively scaled frequency-domain parameterization for history matching. J. Pet. Sci. Eng. **75**(3–4), 289–303 (2011)
6. Sahni, I., Horne, R.: Multiresolution wavelet analysis for improved reservoir description. SPE Reserv. Eval. Eng. **8**(1), 53–69 (2005)
7. Soubaras, R., Dowle, R.: Variable-depth streamer—a broadband marine solution. first break **28**(12) (2010)
8. Gu, Y., Oliver, D.S.: An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. SPE J. **12**(4), 438–446 (2007)
9. Chen, Y., Oliver, D.S.: Levenberg-marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. Comput. Geosci. **17**(4), 689–703 (2013)
10. Ben Ameur, H., Chavent, G., Jaffré, J.: Refinement and coarsening indicators for adaptive parametrization: application to the estimation of hydraulic transmissivities. Inverse Probl. **18**(3), 775 (2002)
11. Grimstad, A.A., Mannseth, T., Nævdal, G., Urkedal, H.A.: Adaptive multiscale permeability estimation. Comput. Geosci. **7**(1), 1–25 (2003)
12. Ben Ameur, H., Clément, F., Weis, P., Chavent, G.: The multidimensional refinement indicators algorithm for optimal parameterization. J. Inverse Ill-Posed Probl. **16**(2), 107–126 (2008)
13. Lu, P., Horne, R.: A multiresolution approach to reservoir parameter estimation using wavelet analysis. In: SPE annual technical conference and exhibition (2000)
14. Bhark, E., Rey, A., Datta-Gupta, A., Jafarpour, B.: A multiscale workflow for history matching in structured and unstructured grid geometries. SPE J. **17**(3), 828–848 (2012)
15. Bhark, E., Datta-Gupta, A., Jafarpour, B.: Subsurface flow model calibration with a spectral-domain parameterization adaptive to grid connectivity and prior model information. Math. Geosci. **44**(6), 673–710 (2012)
16. Liu, J.: A multiresolution method for distributed parameter estimation. Sci. Comput. **14**(2), 389–405 (1993)
17. Mannseth, T.: Permeability identification from pressure observations: some foundations for multiscale regularization. Multiscale Model. Simul. **5**(1), 21–44 (2006)
18. Chavent, G.: Nonlinear least squares for inverse problems: theoretical foundations and step-by-step guide for applications. Springer Verlag (2009)
19. Lien, M., Brouwer, D., Mannseth, T., Jansen, J.-D.: Multiscale regularization of flooding optimization for smart field management. SPE J. **13**(2), 195–204 (2008)
20. Gardet, C., Ravalec, M., Gloaguen, E.: Multiscale parameterization of petrophysical properties for efficient history-matching. Math. Geosci., 1–22 (2013)
21. Reynolds, A.C., He, N., Chu, L., Oliver, D.S.: Reparameterization techniques for generating reservoir descriptions conditioned to variograms and well-test pressure data. SPE J. **1**(4), 413–426 (1996)
22. Efendiev, Y., Hou, T., Luo, W.: Preconditioning Markov chain Monte Carlo simulations using coarse-scale models. SIAM J. Sci. Comput. **28**(2), 776–803 (2006)
23. Jafarpour, B., McLaughlin, D.: Reservoir characterization with the discrete cosine transform. SPE J. **14**(1), 182–201 (2009)
24. Li, L., Jafarpour, B.: A sparse Bayesian framework for conditioning uncertain geologic models to nonlinear flow measurements. Adv. Water Resour. **33**(9), 1024–1042 (2010)
25. Bhark, E.W., Jafarpour, B., Datta-Gupta, A.: A generalized grid connectivity-based parameterization for subsurface flow model calibration. Water Resour. Res. **47**(6) (2011)
26. Feng, T., Mannseth, T.: Improvements on a predictor-corrector strategy for parameter estimation with several data types. Inverse Probl. **25**(10), 105012–105032 (2009)
27. Hamill, T.M., Whitaker, J.S., Snyder, C.: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. Mon. Weather Rev. **129**(11), 2776–2790 (2001)
28. Houtekamer, P.L., Mitchell, H.L.: A sequential ensemble Kalman filter for atmospheric data assimilation. Mon. Weather Rev. **129**(1), 123–137 (2001)
29. Chen, Y., Oliver, D.S.: Cross-covariances and localization for EnKF in multiphase flow data assimilation. Comput. Geosci. **14**, 579–601 (2010)
30. Chen, Y., Oliver, D.S.: Ensemble-based closed-loop optimization applied to Brugge field. SPE Reserv. Eval. Eng. **13**(1), 56–71 (2010)
31. Furrer, R., Bengtsson, T.: Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. J. Multivar. Anal. **98**(2), 227–255 (2007)
32. Anderson, J.L.: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. Phys. D: Nonlinear Phenom. **230**(1–2), 99–111 (2007)
33. Zhang, Y., Oliver, D.S.: Improving the ensemble estimate of the Kalman gain by bootstrap sampling. Math. Geosci. **42**(3), 327–345 (2010)

34. Grossmann, A., Morlet, J.: Decomposition of hardy functions into square integrable wavelets of constant shape. SIAM J. Math. Anal. **15**(4), 723–736 (1984)
35. Mallat, S.G.: A theory for multiresolution signal decomposition: the wavelet representation. Pattern Anal. Mach. Intell., IEEE Trans. **11**(7), 674–693 (1989)
36. Burrus, C.S., Gopinath, R.A., Guo, H., Odegard, J.E., Selesnick, I.W.: Introduction to wavelets and wavelet transforms: a primer, vol. 23. Prentice hall Upper Saddle River (1998)
37. Sweldens, W.: The lifting scheme: a construction of second generation wavelets. SIAM J. Math. Anal. **29**(2), 511–546 (1998)
38. Sweldens, W., Schröder, P.: Building your own wavelets at home. In: Klees, R., Haagmans, R. (eds.) Wavelets in the Geosciences of Lecture Notes in Earth Sciences, vol. 90, pp. 72–107. Springer, Berlin, Heidelberg (2000)
39. Chen, Y., Oliver, D.S.: Ensemble randomized maximum likelihood method as an iterative ensemble smoother. Math. Geosci. **44**(1), 1–26 (2012)
40. Liu, N., Oliver, D.S.: Evaluation of Monte Carlo methods for assessing uncertainty. SPE J. **8**(2), 188–195 (2003)
41. Gao, G., Zafari, M., Reynolds, A.C.: Quantifying uncertainty for the PUNQ-S3 problem in a Bayesian setting with RML and EnKF. SPE J. **11**(4), 506–515 (2006)
42. Golub, G.H., Van Loan, C.F.: Matrix Computations, 2nd edn. Johns Hopkins University Press, Baltimore (1989)
43. Li, R., Reynolds, A.C., Oliver, D.S.: History matching of three-phase flow production data. SPE J. **8**(4), 328–340 (2003)
44. Iglesias, M.A., Dawson, C.: The regularizing Levenberg-Marquardt scheme for history matching of petroleum reservoirs. Comput. Geosci. **17**(6), 1033–1053 (2013)
45. Oliver, D.S., Reynolds, A.C., Liu, N.: Inverse theory for petroleum reservoir characterization and history matching. Cambridge University Press (2008)
46. Donoho, D.: Compressed sensing. Inf. Theory, IEEE Trans. **52**, 1289–1306 (2006)
47. Peters, L., Arts, R., Brouwer, G., Geel, C., Cullick, S., Lorentzen, R., Chen, Y., Dunlop, N., Vossepoel, F., Xu, R., et al.: Results of the Brugge benchmark study for flooding optimization and history matching. SPE SPE Reserv. Eval. Eng. **13**(3), 391–405 (2010)
48. Grimstad, A.-A., Mannseth, T.: Nonlinearity, scale, and sensitivity for parameter estimation problems. SIAM J. Sci. Comput. **21**(6), 2096–2113 (2000)
49. Fernandez, G., Periaswamy, S., Sweldens, W.: LIFTPACK: A software package for wavelet transforms using lifting. Proc. SPIE **2825**, 396 (1996)
50. Girardi, M., Sweldens, W.: A new class of unbalanced Haar wavelets that form an unconditional basis for Lp on general measure spaces. J. Fourier Anal. Appl. **3**(4), 457–474 (1997)