Computational Geosciences (2006) 10:97–117
DOI: 10.1007/s10596-005-9012-4

A multiscale method for distributed parameter estimation with application to reservoir history matching

Sigurd Ivar Aanonsen and Dmitry Eydinov

Centre for Integrated Petroleum Research, University of Bergen, P.O. Box 7800, N-5020 Bergen, Norway
E-mail: Sigurd.aanonsen@cipr.uib.no

Received 29 November 2004; accepted 8 August 2005

A method for multiscale parameter estimation with application to reservoir history matching is presented. Starting from a given fine-scale model, coarser models are generated using a global upscaling technique where the coarse models are tuned to match the solution of the fine model. Conditioning to dynamic data is done by history-matching the coarse model. Using consistently the same resolution both for the forward and inverse problems, this model is successively refined using a combination of downscaling and history matching until model-matching dynamic data are obtained at the finest scale. Large-scale corrections are obtained using fast models, which, combined with a downscaling procedure, provide a better initial model for the final adjustment on the fine scale. The result is thus a series of models with different resolution, all matching history as good as possible with this grid. Numerical examples show that this method may significantly reduce the computational effort and/or improve the quality of the solution when achieving a fine-scale match as compared to history-matching directly on the fine scale.

Keywords: parameter estimation, history matching, multiscale

1. Introduction

The process of conditioning reservoir simulation models to dynamic data (history matching) is a difficult and very time-consuming process, which, for real field models in practice, is still mainly performed by tuning the model by hand. The models should honour all data, both static and dynamic, as well as being consistent with the a priori geological knowledge. Still, the amount of data will normally not be sufficient to uniquely characterize a reservoir at the degree of resolution requested. So, this inverse problem is mathematically ill conditioned and requires some kind of regularization. One approach is to perform the history matching in a stochastic framework with the aim of characterizing the uncertainty in the reservoir description and model predictions by estimating the a posteriori probability distribution based on Bayes rule

(see, e.g., Tarantola [17]). The randomized maximum likelihood (RML) method proposed by Kitanidis [12] and Oliver et al. [16] has been shown to provide a good approximation to the a posteriori probability distribution [14]. In the RML method, an approximate sampling is performed by doing repeated history matches starting from different unconditional realizations of the model parameters and data. The aim of the method presented in this paper is to improve the efficiency of history matching in this setting.

In multiscale estimation, a series of estimations is performed where the resolution of the zonation is increased for each step in the sequence [5,13,18]. This approach reduces both the computational effort as well as the overparameterization problem. The multiscale estimation was improved by Ben Ameur et al. [3] and Grimstad et al. [8], who introduced different methods for adaptive multiscale estimation where the resolution is increased only in some regions of the reservoir in each stage of the sequence.

Common for the above-mentioned multiscale techniques is that the computational grid is kept the same while the parameter resolution is varied. In this paper, a multiscale technique is suggested where also the forward problem is computed on

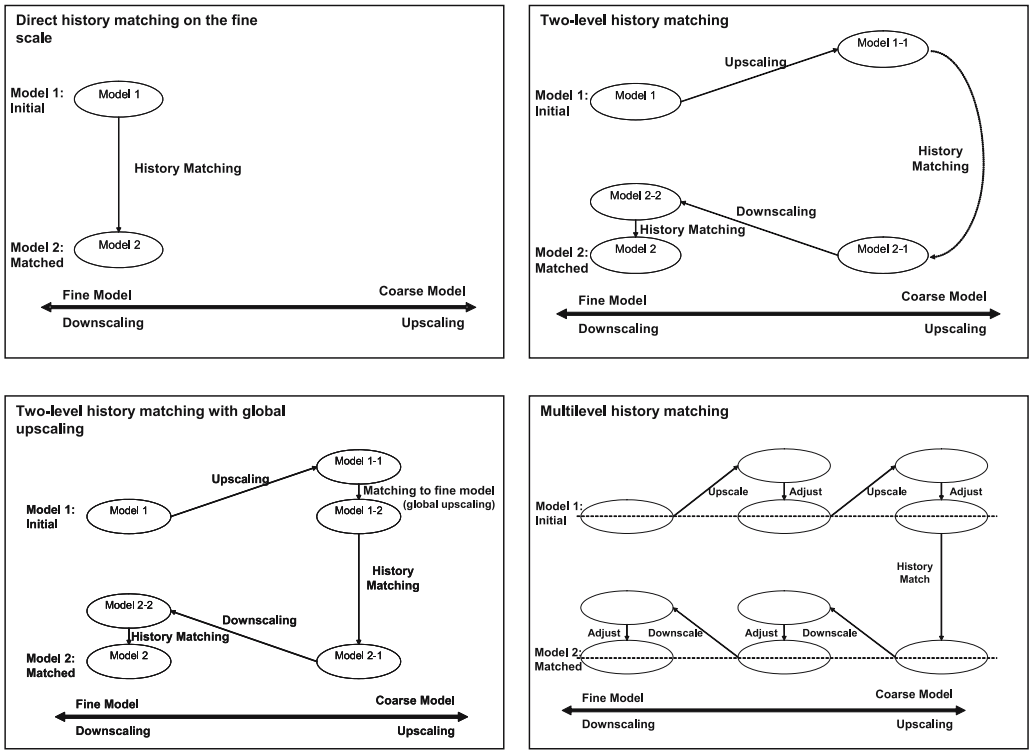

Figure 1. Multiscale history matching.

grids of varying resolution. A coarse-scale representation of the parameters is obtained by history-matching a forward model defined on the same coarse grid. Using consequently the same resolution both for the forward and inverse problems, this model is successively refined using a combination of downscaling and history matching until model-matching dynamic data are obtained at the finest scale. Large-scale corrections are obtained using fast models, which, combined with a downscaling procedure, provide a better initial guess for the final adjustment on the fine scale. The result is thus a series of models with different resolution, all matching history as good as possible with this grid. Correspondingly, a series of models that are consistent with the initial model may be generated using a global upscaling technique where the coarse models are "history-matched" with respect to the solution at the fine scale. The upscaling and downscaling techniques are illustrated in figure 1.

The advantages of this procedure are that the large-scale corrections are obtained using fast models, and that the match is retained through the downscaling steps. The coarse-scale history matching also provides a regularization of the fine-scale match.

The main focus of this paper is on the presentation of a framework for multiscale parameter estimation. No rigorous proofs of the improved efficiency or solution quality with the proposed method are presented, but the behavior is illustrated through several numerical examples. The examples show that the procedure may significantly reduce the computational effort and/or improve the quality of the solution when achieving a fine-scale match as compared to history matching directly on the fine scale.

2. Method

2.1. Problem statement

We want to estimate an unknown reservoir property $\theta(x)$ defined in a bounded domain in $\mathbb{R}^3$ (the reservoir), given the measurements y^* of a quantity $y \in \mathbb{R}^m$. We assume that the reservoir is discretized on a fine grid Ω with n grid cells, and that θ can be considered to be constant within each grid cell. A dynamic reservoir simulator defines a mapping, $f(\theta)$, between θ and y. The measurement error (data error) is denoted e_y, and the model error (on the finest grid) is denoted e_f. That is,

$$y^* = y + e_y = f(\theta) + e_f + e_y. \tag{1}$$

We further assume that the model and measurement errors are Gaussian, with covariance matrices C_f and C_{y*}, and that we have some á priori geological information about θ, which also can be represented by a Gaussian probability density function with mean $\bar{\theta}$ and covariance matrix C_θ.

The maximum á posteriori estimate of θ given y^* is given by the minimum of the following objective function (see, e.g., [17]):

$$J(\theta) = \frac{1}{2}\left[\left(y - f(\theta)\right)^T C_M^{-1}\left(y - f(\theta)\right) + \left(\theta - \bar{\theta}\right)^T C_\theta^{-1}\left(\theta - \bar{\theta}\right)\right], \tag{2}$$

where $C_M = C_{y*} + C_f$ is the sum of the measurement and model error covariance matrices.

Let θ^* be the á posteriori estimate that minimizes J. An estimate of the uncertainty in θ^* can be found by linearizing f around θ^*:

$$f(\theta) \approx f(\theta^*) + A(\theta^*)(\theta - \theta^*). \tag{3}$$

Here $A = \mathrm{d}f/\mathrm{d}\theta$ is the sensitivity matrix. The á posteriori covariance matrix becomes equal to the inverse of the linearized Hessian and is given by:

$$C_{\theta^*} = H^{-1} = \left(C_\theta^{-1} + A^{\mathrm{T}} C_M^{-1} A\right)^{-1}. \tag{4}$$

Here m may be less than n, since the prior covariance matrix ensures a non-singular Hessian. Without any prior information, $C_\theta^{-1} = 0$, and provided $m > n$, the solution reduces to the traditional least-squares estimate. Often, however, if a rock property is to be estimated on the fine scale, the number of measurements will be smaller than the number of unknowns, and the prior covariance matrix is necessary, providing regularization and a preconditioner for the optimization routine. In this derivation, we consider only one unknown property, i.e., the dimension of θ is equal to the number of grid cells on the finest grid. However, this may easily be extended to several properties.

2.2. Upscaling

Let $\Omega^{(j)} = \Omega, \Omega^{(j-1)}, \ldots, \Omega^{(0)}$ be a coarsening sequence of overlapping grids. The corresponding number of cells is $n^{(j)} = n > n^{(j-1)} > \cdots > n^{(0)}$. Further, let the number of cells in grid $\Omega^{(i)}$ within cell $\hat{i}$ in grid $\Omega^{(i-1)}$ be $(n^{(i)})_{\hat{i}}$. To each coarse cell, $\hat{i}$, we associate a basis vector $(\Psi^{(i-1)})_{\hat{i}} \in \mathbb{R}^{n^{(i)}}$ with components

$$\left((\Psi^{(i-1)})_{\hat{i}}\right)_k = \begin{cases} 1, & \forall k \in (M^{(i-1)})_{\hat{i}}, \\ 0, & \text{otherwise}, \end{cases} \tag{5}$$

where $(M^{(i-1)})_{\hat{i}}$ is the index set defining which of the cells in grid $\Omega^{(i)}$ are in the cell $\hat{i}$ in grid $\Omega^{(i-1)}$.

A simple arithmetic averaging from grid $\Omega^{(i)}$ to grid $\Omega^{(i-1)}$ is given by

$$\theta^{(i-1)} = R^{(i-1)}\theta^{(i)}, \tag{6}$$

where $R^{(i-1)} \in \mathbb{R}^{n^{(i-1)} \times n^{(i)}}$ is the restriction operator defined by:

$$R^{(i-1)} = \begin{Bmatrix} \frac{1}{(n^{(i)})_{\hat{i}=1}} \left[(\Psi^{(i-1)})_{\hat{i}=1} \right]^T \\ \frac{1}{(n^{(i)})_{\hat{i}=2}} \left[(\Psi^{(i-1)})_{\hat{i}=2} \right]^T \\ \vdots \\ \frac{1}{(n^{(i)})_{\hat{i}=n^{(i-1)}}} \left[(\Psi^{(i-1)})_{\hat{i}=n^{(i-1)}} \right]^T \end{Bmatrix}. \tag{7}$$

The covariance matrix for $\theta^{(i-1)}$ is given by

$$C_{\theta^{(i-1)}} = R^{(i-1)} C_{\theta^{(i)}} (R^{(i-1)})^T \tag{8}$$

Upscaling from an arbitrary grid $\Omega^{(j)}$ to another grid $\Omega^{(i)}$, $i < j$, is done by applying the restriction operator several times. That is,

$$\theta^{(i)} = R^{(i,j)} \theta^{(j)}, \tag{9}$$

and

$$C_{\theta^{(i)}} = R^{(i,j)} C_{\theta^{(j)}} (R^{(i,j)})^T, \tag{10}$$

where

$$R^{(i,j)} = R^{(i)} R^{(i+1)} \cdots R^{(j)}. \tag{11}$$

The simulated data values are given at the various grids by $f^{(j)}(\theta^{(j)})$, $f^{(j-1)}(\theta^{(j-1)}), \ldots, f^{(0)}(\theta^{(0)})$. The difference between the data simulated at the two grids defines the upscaling error going from grid i to grid $i-1$, $\epsilon^{(i-1)}$:

$$f^{(i)}(\theta^{(i)}) = f^{(i-1)}(\theta^{(i-1)}) + \epsilon^{(i-1)}. \tag{12}$$

An improved (global) upscaling may be obtained by taking $\theta^{(i-1)}$ as the minimum variance estimate based on equation (12), i.e., "history-matching" the coarse model with respect to the solution at the finer level. An updated, linear estimate for $C_{\theta^{(i-1)}}$ is then given by:

$$C_{\theta^{(i-1)}} = \left[\left(R^{(i-1)} C_{\theta^{(i)}} (R^{(i-1)})^T \right)^{(-1)} + (A^{(i-1)})^T \left(C_{f^{(i)}} \right)^{-1} A^{(i-1)} \right]^{-1}, \tag{13}$$

where $A^{(i-1)} = \mathrm{d}f^{(i-1)} / \mathrm{d}\theta^{(i-1)}$.

In practice, one would probably want to match the solution at all the grids to the solution on the finest grid, i.e., minimizing the "total" upscaling error at grid $\Omega^{(i-1)}$. Denoting this error as $\tilde{\epsilon}^{(i-1)}$, we thus have

$$f(\theta) = f^{(i-1)}(\theta^{(i-1)}) + \tilde{\epsilon}^{(i-1)}. \tag{14}$$

The parameter covariance matrix may be calculated from equation (10), or alternatively updated at every upscaling step, i.e., calculated from equation (13). This requires an estimate of the modelling error on the finest grid.

2.3. History matching

At the coarsest level, corresponding to grid $\Omega^{(0)}$, the model is history-matched to the measured data to incorporate large-scale modifications warranted by the data. Normally, the model error is not taken into account in history matching, but if we assume that the model error at the coarsest scale is dominated by the upscaling error, the model error covariance matrix is given by the covariance matrix of $\tilde{\epsilon}^{(0)}$.

If we consider well data only, m_{id} data types are measured in m_{iw} injection wells, and m_{pd} data types are measured in m_{pw} production wells at m_t time steps. Then the total number of the well data to be matched is $m = (m_{id} \cdot m_{iw} + m_{pd} \cdot m_{pw}) \cdot m_t$. The data vector may be written as

$$y = \begin{Bmatrix} y_{iw}^{id} \\ y_{pw}^{pd} \end{Bmatrix}, \tag{15}$$

where $y_{iw}^{id} \in \mathbb{R}^{m_t}, iw = 1, 2, \ldots, m_{iw}, id = 1, 2, \ldots, m_{id}$, contains the measurements of data type id in the injection well iw at each time step, and $y_{pw}^{pd} \in \mathbb{R}^{m_t}, pw = 1, 2, \ldots, m_{pw}, pd = 1, 2, \ldots, m_{pd}$, contains the measurements of data type pd in the production well pw at each time step. Letting $y_w^d \in \mathbb{R}^{m_t}$ be one of the vectors y_{iw}^{id} or y_{pw}^{pd}, we further assume that the measurement error variance $(\sigma_y^2)_w^d$ is given for each well and data type and approximates the covariance matrix of $\epsilon^{(0)}$ with a diagonal matrix having constant variance for each well and data type. This model error variance is then given by

$$\left(\sigma_f^2\right)_w^d = \frac{1}{m_t} \sum_{k=1}^{m_t} \left(\left(y_w^d\right)_k - \left(f_w^{(0)d}\right)_k \right)^2, \tag{16}$$

and the covariance matrix C_M to be used in the objective function equation (2) becomes a diagonal matrix with elements:

$$\left(\sigma_M^2\right)_w^d = \left(\sigma_y^2\right)_w^d + \left(\sigma_f^2\right)_w^d, \quad w = 1, 2, \ldots, m_w, \quad d = 1, 2, \ldots, m_d \tag{17}$$

An upscaling error for distributed measurements may be defined similarly. In principle, the model error may also be based on the locally upscaled coarse model, i.e., the model that is upscaled in the traditional way without any tuning. However, the upscaling error will then be considerably larger and sometimes too large to be useful in the coarse-scale history matching.

2.4. Downscaling

When a match is obtained at the coarsest level, the matched model is successively refined by first downscaling and then history matching to the measured data at each scale as illustrated in figure 1. At each level, updated estimates for the model and parameter error may be obtained as for the upscaling case.

Different approaches may be applied for the downscaling step. The simplest approach is a constant interpolation (sampling) given by

$$\theta^{(i)} = P_1^{(i-1)}\theta^{(i-1)}, \tag{18}$$

where the prolongation operator $P_1^{(i-1)} \in \mathbb{R}^{n^{(i)} \times n^{(i-1)}}$ is defined by:

$$P_1^{(i-1)} = \left\{ \Psi_{\hat{i}=1}^{(i-1)}, \Psi_{\hat{i}=2}^{(i-1)}, \ldots, \Psi_{\hat{i}=n^{(i-1)}}^{(i-1)} \right\}. \tag{19}$$

Alternatively, a statistical approach may be taken. Let now $\theta^{(i)}$ denote the random function representing the true values, and $\tilde{\theta}^{(i)}$ denote the estimated values on $\Omega^{(i)}$. We seek a linear, minimum variance estimate of $\theta^{(i)}$ given $\theta^{(i-1)}$ assuming intrinsic stationarity, and that the statistics are represented by a covariance matrix as before. That is,

$$\tilde{\theta}^{(i)} = P_2^{(i-1)}\theta^{(i-1)}. \tag{20}$$

The solution of this problem is given by the well-known block kriging equations (see, e.g., [11]). The equations for all the cells may be written on matrix form as:

$$\begin{pmatrix} C_{\theta^{(i-1)}} & e^{(i-1)} \\ (e^{(i-1)})^T & 0 \end{pmatrix} \begin{pmatrix} (P_2^{(i-1)})^T \\ \lambda^T \end{pmatrix} = \begin{pmatrix} R^{(i-1)} C_{\theta^{(i)}} \\ (e^{(i)})^T \end{pmatrix}, \tag{21}$$

where $\lambda^T = \{\lambda_1, \lambda_2, \ldots, \lambda_{n^{(i)}}\}$ is the vector of Lagrange multipliers, and $(e^{(i)})^T = \{1, 1, \ldots, 1\} \in \mathbb{R}^{n^{(i)}}$.

By eliminating λ^T as shown by Cressie [6], the kriging weight matrix may be expressed as

$$(P_2^{(i-1)})^T = (C_{\theta^{(i-1)}})^{-1} R^{(i-1)} C_{\theta^{(i)}} + \frac{(C_{\theta^{(i-1)}})^{-1} e^{(i-1)}}{\left((e^{(i-1)})^T (C_{\theta^{(i-1)}})^{-1} e^{(i-1)}\right)} \cdot \left[(e^{(i)})^T - (e^{(i-1)})^T (C_{\theta^{(i-1)}})^{-1} R^{(i-1)} C_{\theta^{(i)}}\right]. \tag{22}$$

If the kriging is performed on a normalized variable with zero mean, unbias is automatically achieved, and the kriging equations simplify to only the upper left part of equation (21). This requires an estimate of the mean value. However, if a minimum variance estimate (see e.g., [15]) is used for the mean, it can easily be shown that the two estimates become identical. The minimum variance estimate of the mean may be expressed as:

$$\bar{\theta} = \frac{(e^{(i-1)})^T (C_\theta^{(i-1)})^{-1} \theta^{(i-1)}}{(e^{(i-1)})^T (C_\theta^{(i-1)})^{-1} e^{(i-1)}}. \tag{23}$$

If we instead use an arithmetic average, i.e.,

$$\bar{\theta} = \frac{(e^{(i-1)})^T \theta^{(i-1)}}{n^{(i-1)}} = \frac{(e^{(i-1)})^T \theta^{(i-1)}}{(e^{(i-1)})^T e^{(i-1)}} \approx \frac{(e^{(i-1)})^T (C_\theta^{(i-1)})^{-1} \theta^{(i-1)}}{(e^{(i-1)})^T (C_\theta^{(i-1)})^{-1} e^{(i-1)}}, \tag{24}$$

and use the same approximation in equation (22), an approximative solution for $P_2^{(i-1)}$ is obtained, which requires solving only a single linear system on the coarse grid. Denote this approximation $P_3^{(i-1)}$. We get:

$$\begin{aligned}(P_3^{(i-1)})^T &= (C_{\theta^{(i-1)}})^{-1} R^{(i-1)} C_{\theta^{(i)}} + \frac{1}{n} e^{(i-1)} (e^{(i)})^T - \frac{1}{n} e^{(i-1)} (e^{(i-1)})^T \\ &\quad \times (C_{\theta^{(i-1)}})^{-1} R^{(i-1)} C_{\theta^{(i)}} \\ &= \frac{1}{n} e^{(i-1)} (e^{(i)})^T + \left(I - \frac{1}{n} e^{(i-1)} (e^{(i-1)})^T\right) (C_{\theta^{(i-1)}})^{-1} R^{(i-1)} C_{\theta^{(i)}}. \end{aligned} \tag{25}$$

For $\theta^{(i)}$, we get:

$$\begin{aligned}\theta^{(i)} &= P_3^{(i-1)} \theta^{(i-1)} \\ &= e^{(i)} \bar{\theta} + C_{\theta^{(i)}} (R^{(i-1)})^T (C_{\theta^{(i-1)}})^{-1} \left(\theta^{(i-1)} - e^{(i-1)} \bar{\theta}\right) \\ &= e^{(i)} \bar{\theta} + C_{\theta^{(i)}} (R^{(i-1)})^T \hat{\theta}^{(i-1)}, \end{aligned} \tag{26}$$

where $\hat{\theta}^{(i-1)}$ is the solution of

$$C_\theta^{(i-1)} \hat{\theta}^{(i-1)} = \theta^{(i-1)} - e^{(i-1)} \bar{\theta}. \tag{27}$$

For simplicity, we have based all the kriging calculations on P_3, i.e., equations (26) and (27).

The kriging operators P_2 and P_3 yield smooth functions, with more extreme values than the sampling operator. The kriged θ fields may be conditioned to other hard measurements, e.g., in wells. To get a solution at the finest scale with similar properties as the initial θ, the kriging may be combined with sequential Gaussian simulation as in Behrens et al. [2]. Similar methods for downscaling have also been used by, for instance, Grimstad and Mannseth [9]. Normally, the initial field has been obtained by a combination of kriging and stochastic simulation, so the necessary parameters and procedures will already be available.

3. Examples

3.1. *Description of model cases*

The efficiency of the multiscale approach has been tested on three different synthetic test models. The first (model 1) is a two-dimensional version of the so-called PUNQ-S3 model – a synthetic model based on a real West African oil field. The field is bound by faults to the east and south and by a strong aquifer to the north and west. Details of the model can be found in Floris et al. [7]. A two-dimensional version of the "true" case was generated by upscaling in the vertical direction using arithmetic averaging of layer permeabilities. The reservoir with initial saturations and true permeability field are shown in figure 2. Note that the North-South axis is inverted in all the plots. The quantity to be estimated is the permeability in all (active) cells, i.e.,

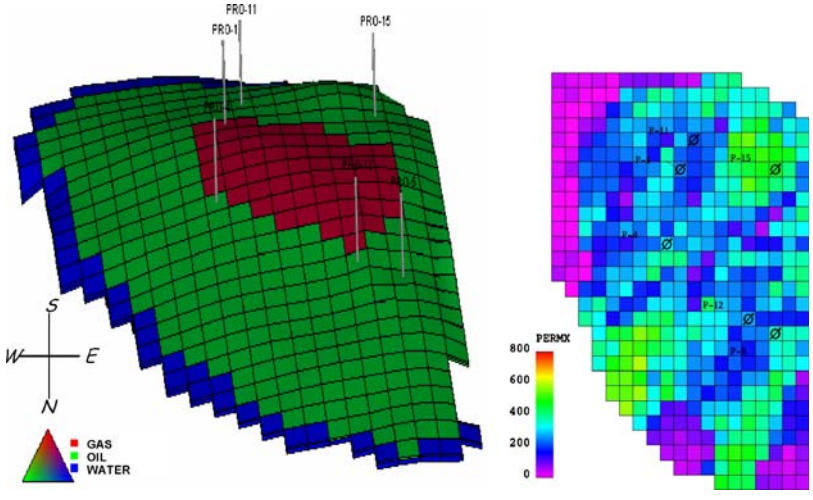

Figure 2. 2D version of PUNQ-S3 model. Overview with initial saturations (*left*) and true permeability (*right*). Note that the North-South axis is inverted.

454 parameters. Data to be matched are well bottom-hole pressure (BHP), gas-oil ratio [ratio between produced gas and oil rate (GOR)] and water cut [ratio between produced water and total liquid rate (WCT)] in the six production wells. The history simulated with the true permeability field was used as the true history without adding noise. Thus, it is theoretically possible to reduce the objective function to zero. Uncertainties in the measurements were included as a diagonal covariance matrix in the objective function. This ensures consistent weighting between the different data types.

Three cases have been investigated. In all cases, the "true" model was kept the same, whereas the initial permeability field was changed. For case 1, the initial permeability field was generated by increasing the variability and also increasing the average slightly. Case 2 was generated by first increasing all the permeability values by 500 mD and then adding a small noise. Finally, case 3 was generated by introducing an incorrect trend in the permeability values. The permeabilities corresponding to the three cases are plotted in figure 3. Note that because of the large variation in the maximum values, different colour scales are used in the permeability plots.

Two levels of coarsening were applied:

(i) Coarse (7 × 7 cell) model with 3 × 4 fine cells in each coarse cell (total number of active cells: 43).

(ii) Very coarse (4 × 4 cell) model with 6 × 8 fine cells in each coarse cell (total number of active cells: 13).

For simplicity (keeping the upscaling as uniform as possible), one column of cells was not coarsened. The grids are shown in figure 4. The aquifer influx parameters were kept the same for all the grids.

The second model (model 2) is a regular, three-dimensional model with five wells: a central water injector and four producers. The dimensions are 500 × 500 × 40 m. The fine-grid model has 12 × 12 × 3 cells. Like in model 1, the quantity to be

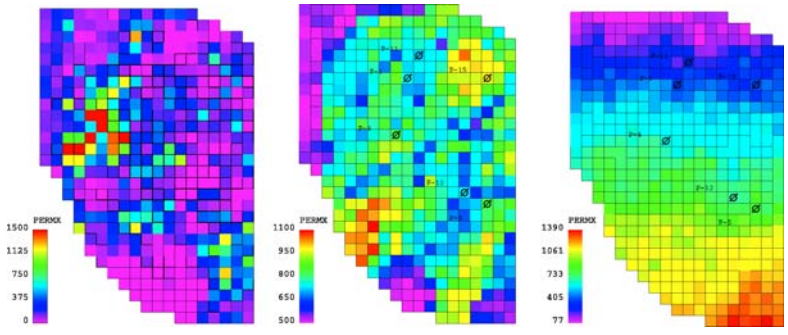

Figure 3. Initial permeabilities case 1 (*left*), case 2 (*middle*) and case 3 (*right*).

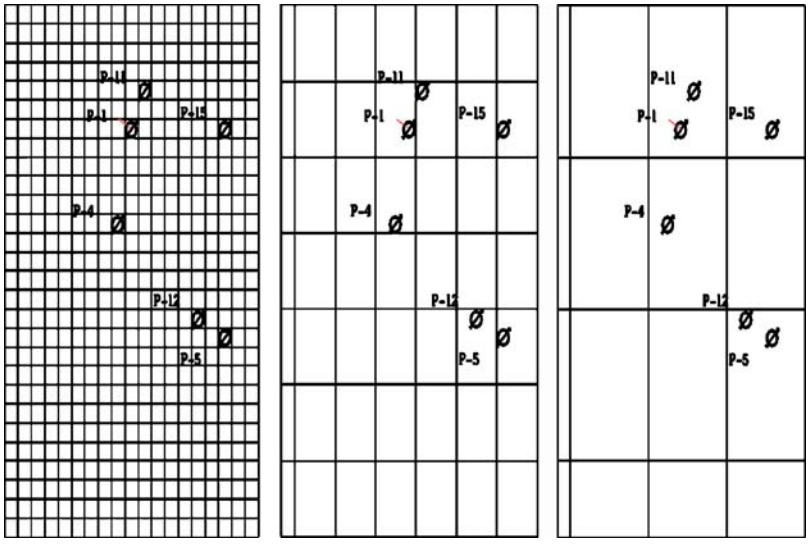

Figure 4. Original fine grid and two levels of coarsening.

estimated is the permeability in all the (432) grid cells. For simplicity, a homogeneous porosity equal to 0.2 was used. The "true" and base case (initial) permeability fields were generated as realizations from stochastic models with input parameters as listed in table 1. Vertical permeability is for all runs set to 0.5, 0.05, and 0.1 times the horizontal permeability in the top, middle, and bottom layers, respectively. The coarsening was uniform using $6 \times 6 \times 3$ and $6 \times 6 \times 1$ cell grids. The production wells in this model were controlled by a constant BHP, and the injection well rate was controlled to replace the produced volume (100% voidage injection). The history-matching data are oil production rate and water cut in the four producers. More details of this model may be found in Aanonsen [1], who presents additional cases of multiscale history matching including cases with a prior term in the objective function.

Objective-function minimization for models 1 and 2 was performed with a commercial history-matching software tool using the Levenberg–Marquardt optimization routine. Because of a hard-coded limitation in this software, the number of parameters had to be less than the number of measurements, and thus, the data part of the Hessian, $A^T C_M^{-1} A$, is always non-singular. That is, a prior term in the objective function is not necessary to avoid a singular Hessian and is not included in these examples. In model 1, $m = 84(\text{time steps}) * 6(\text{wells}) * 3(\text{data types}) = 1512$. In model 2, the number of data points is 328 (41 time steps $\times$ 4 wells $\times$ 2 data types), which is less than the number of parameters (423), and thus well GOR was included as data, although GOR is constant in all wells at all times. The number of data points then increases to 492.

Model 3 is similar to model 2 with a central water injector and four producers, two-dimensional, but with a larger number of grid cells (40×40). This model was run

Table 1
Input parameters for property simulations.

	Layer average			Standard deviation, all layers	Variogram range (m)			Trend
	1	2	3		x-direction	y-direction	z-direction	
Model 2								
True permeability	2.7	3.1	2.5	0.1	500	1000	10	linear
Initial permeability	2.7	2.7	2.7	0.3	500	1000	10	constant
Model 3								
True permeability	3.1	–	–	0.15	1000	1000	10	linear
True porosity	0.21	–	–	0.03	1000	1000	10	linear
Initial permeability	3.1	–	–	0.15	1000	1000	10	constant
Initial porosity	0.17	–	–	0.03	1000	1000	10	constant
Correlation porosity – log permeability				0.85				

For permeability, the parameters apply to the (base 10) logarithm of the permeability in millidarcys.

using an in-house simulator. This simulator is a compositional simulator based on an implicit pressure and saturations (IMPSAT) formulation, with various alternative discretization schemes, including the multipoint flux approximation [10]. In this case, however, a simple system with only water and a single oil component was applied. The simulator has history-matching capabilities through a coupling to an open-code LBFGS optimizer [4, 19]. Gradients with respect to porosities and tensor permeabilities (k_x, k_y, k_z and main axis directions) are efficiently calculated from the solution of the adjoint system of equations.

The parameters in this case are porosity and two permeabilities (k_x and k_y) in each grid cell, totalling 4800. True and initial models are taken as realizations from stochastic models with parameters as listed in table 1. The main difference between the initial and the true model is a linear trend in the properties. A prior term as defined in equation (2) with a diagonal covariance matrix was included. Prior standard deviations were 0.025 for porosity and 1.0 for log permeability. These values do not coincide with the values used to generate the initial fields (table 1). However, this is not believed to have any large influence on the efficiency of the multiscale method, which is the main focus of this paper.

History-matching data are oil rates in the four producers and BHP in the injector. Total simulation time is 3 years, corresponding to 0.14 pore volumes injected. The number of time steps is 60, and well data are measured at every time step. Thus, the total number of well measurements is 300. No water breakthrough was observed, but since the production is controlled by a given pressure, the distribution of the oil production between the wells will provide information about the permeability and porosity distribution.

For this model, also a case with distributed measurements, i.e., pressure and oil saturation in each grid block, was included. Such data may sometimes be obtained

from time-lapse (4D) seismic measurements. Here three sets of data were applied: after 1, 2, and 3 years of production, respectively.

In all the examples presented, the data root mean square (RMS) index was used as a measure of the quality of the match. Data RMS is defined by

$$\mathrm{RMS} = \sqrt{2\hat{J}/m}, \tag{28}$$

where $\hat{J}$ is the objective function without the prior term, and m is the total number of measurements. RMS = 1 was used as a stopping criterion for the minimization algorithm. In a practical situation, it will not be feasible to fine-tune the optimizer for each history-matching run, and therefore, the default values for the tuning parameters (e.g., trust region for the Levenberg–Marquardt optimizer) were used in all the examples presented.

3.2. Results

3.2.1. Efficiency of the multiscale approach

The main measure of the efficiency of the multiscale method is the number of iterations needed for the final fine-scale tuning. Figure 5 shows the RMS development for a direct history matching of the fine-scale model compared to two different multiscale methods for the three cases using model 1. The direct history matching of the fine-scale model is performed by adjusting all cell permeabilities individually to match the true history. The multiscale methods are denoted "n-level_i," where n denotes the number of levels, and $i = 1$ defines the method: $i = 1$ without global upscaling, and $i = 2$ with global upscaling. That is, for the "2-level_1" method, the procedure illustrated in the upper right plot of figure 1 is applied: a traditional, local upscaling and history matching followed by a downscaling. The local upscaling is in all cases based on arithmetic averaging [equations (6) and (8)]. The downscaled permeability field is then used as a new initial guess for the fine-scale tuning. The "2-level_2" method corresponds to the method illustrated in the lower left plot of figure 1. Here the upscaled permeabilities are adjusted using the global upscaling

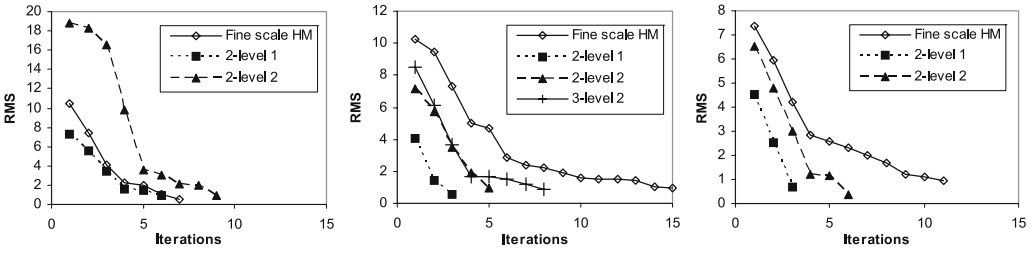

Figure 5. RMS deviation from true history *vs.* the number of fine-scale iterations model 1. Case 1 (*left*); case 2 (*middle*); case 3 (*right*).

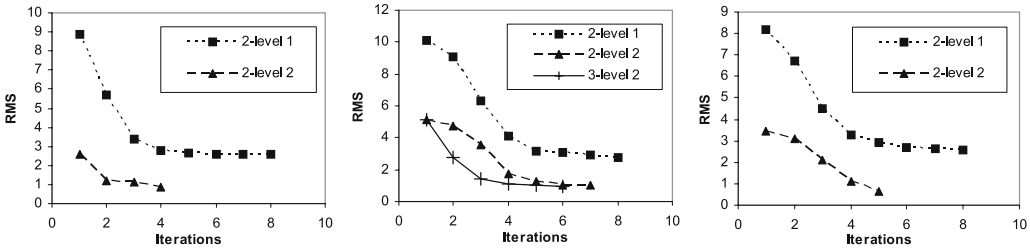

Figure 6. RMS deviation from true history *vs.* the number of coarse-scale iterations model 1. Case 1 (*left*); case 2 (*middle*); case 3 (*right*). Note that a model error is included in the error variance used in the definition of the RMS value for method 2, but not for method 1.

technique as described in section 2.2, and these are used as initial values for the coarse-scale history match. A model error covariance matrix was added to the data error covariance matrix in this regression step [cf. equation (2)]. The model error was based on the final RMS deviation between the fine- and coarse-scale simulation results as described in section 2.3: one value for each well and data type (BHP, GOR and WCT; i.e., the sum is taken over all measurements of a given data type in a given well). When this model error is added to the data error, and RMS = 1 is used as a stopping criterion for the coarse-scale history match, one ensures that large errors in the coarse model are taken into account in the coarse-scale history match. Since a larger data mismatch is accepted, the stopping criteria will normally be reached faster, as is clearly seen from figure 6. In principle, a model error could have been based on the difference between the fine-scale and coarse-scale results before global upscaling and used also in the cases where global upscaling was not applied. However, in many cases, the model error then became too large to be useful, and we have not added model error in any of the *n*-level_1 cases. A three-level method was tested on case 2, and it turned out to be impossible to perform a history match on the very coarse grid without adding a model error decreasing the weight of the GOR and water-cut data. Thus, the 3-level_1 method was not applicable in this case.

The data and model errors used are listed in table 2. Note that most of the model errors are considerably larger for the 4×4 cell model than for the 7×7 cell model. Some wells did not have free gas or water production in any of the models, and thus, the corresponding model error is small.

It is seen that in terms of the number of iterations to obtain a fine-scale match, the two-level methods are more efficient than the direct fine-scale history matching in all cases, except for case 1, method 2-level_2.

Note also that the large-scale modifications resulting from the coarse-scale history matching yields a fast convergence of the final fine-scale tuning, even though incorrect permeabilities in individual well blocks have a large effect on simulated well data. This is clearly demonstrated in case 1, method 2-level_2, where the final fine-scale history match converges fast despite a very high initial RMS value.

Table 2
Data and model errors (standard deviations) for model 1.

		Model error, 7 × 7 model			4 × 4 model
Well	Data Error	Case 1	Case 2	Case 3	Case 2
Pressure (bar)					
PRO-1	1.0	1.8	1.4	2.3	1.6
PRO-4	1.0	1.4	0.9	0.7	1.3
PRO-5	1.0	1.1	0.5	0.5	0.7
PRO-11	1.0	2.3	1.0	1.5	1.9
PRO-12	1.0	1.1	0.8	0.6	0.9
PRO-15	1.0	1.4	1.0	1.3	1.2
Gas/oil ratio (Sm^3/Sm^3)					
PRO-1	5.0	4.6	12.3	14.9	12.2
PRO-4	5.0	1.5	0.8	0.4	1.6
PRO-5	5.0	2.1	0.9	0.8	1.8
PRO-11	5.0	1.9	0.9	1.6	2.1
PRO-12	5.0	2.3	1.6	1.5	2.1
PRO-15	5.0	1.3	0.9	1.1	15.8
Water cut (fraction)					
PRO-1	0.02	0.031	0.034	0.032	0.060
PRO-4	0.02	0.022	0.007	0.008	0.021
PRO-5	0.02	0.001	0.001	0.004	0.010
PRO-11	0.02	0.022	0.029	0.024	0.056
PRO-12	0.02	0.001	0.002	0.003	0.010
PRO-15	0.02	0.004	0.030	0.018	0.127

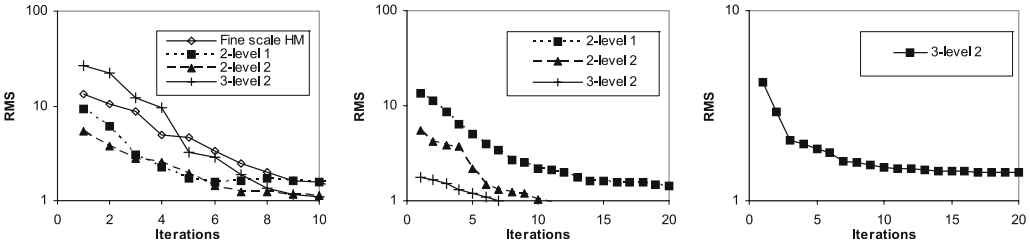

Figure 7. Development of RMS values model 2. Fine scale (*left*); coarse scale (*middle*); very coarse scale (*right*).

By comparing the final RMS values from the coarse-scale tuning 2-level_1 method (figure 6) with the corresponding initial RMS value for the final fine-scale tuning (figure 5), it is also seen that the match obtained at the coarse scale is not retained through a standard downscaling step.

Comparisons between the multiscale methods and a direct fine-scale history matching for the other two models are shown in figures 7 and 8. Again, it is seen that the multiscale methods are considerably faster than the single-level method. The only

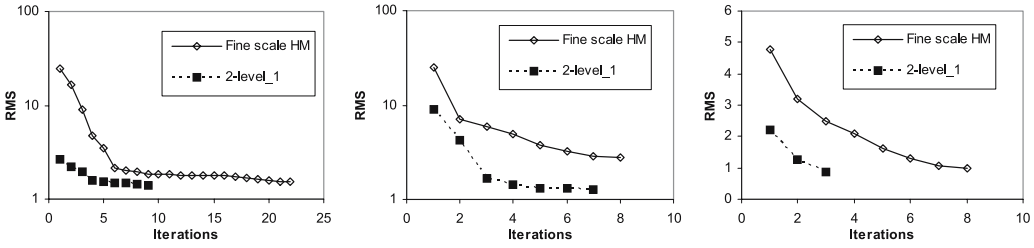

Figure 8. Development of RMS values model 3. *Left plot* shows well data RMS for case 1 (without seismic). *Middle* and *right plots* show well data RMS and well + seismic data RMS, respectively, for case 2 (with seismic).

Table 3
Data and model errors (standard deviations) for model 2.

Well	Data error	Model error	
		6 × 6 × 3 grid	6 × 6 × 1 grid
Oil rate (Sm^3/day)			
P-1	20.0	33.2	32.1
P-2	20.0	31.5	34.7
P-3	20.0	28.1	38.8
P-4	20.0	29.6	35.2
Water cut (fraction)			
P-1	0.02	0.023	0.029
P-2	0.02	0.017	0.022
P-3	0.02	0.018	0.028
P-4	0.02	0.025	0.028
Gas/oil ratio (Sm^3/Sm^3)			
P-1	5.0	3.1	3.6
P-2	5.0	2.9	3.3
P-3	5.0	2.5	3.3
P-4	5.0	2.6	2.7

exception now is the 2-level_1 method for model 2. Model errors applied for method 2 (with global upscaling) are listed in table 3 for model 2. For model 3, only method 1, without global upscaling, has been tested. Data errors (standard deviations) used for model 3 were 3% relative error for all the well data, 10% relative error for distributed pressure data and an absolute value of 0.05 for distributed saturation measurements. Arithmetic averaging based on equations (6) and (8) was used to upscale the distributed data.

3.2.2. Effect of initial downscaling method

In the results presented so far, direct sampling [equations (18) and (19)] has been used for the initial downscaling step. However, with this technique, the coarse grid

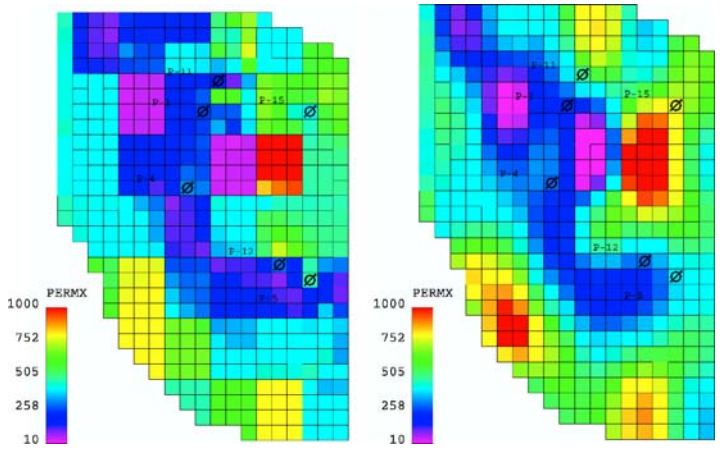

Figure 9. Permeability field for case 2 estimated with the 2-level_1 method. Initial downscaling based on sampling (*left*) and kriging (*right*).

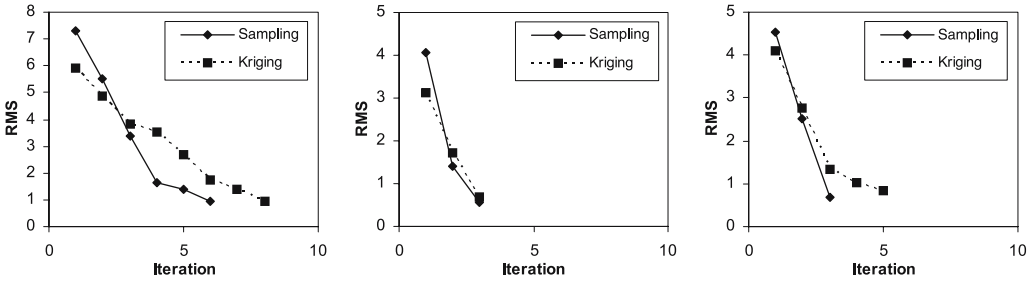

Figure 10. RMS deviation from true history *vs.* the number of fine-scale iterations. Initial downscaling based on sampling *vs.* kriging. Case 1 (*left*); case 2 (*middle*); case 3 (*right*).

structure will typically be reflected in the final permeability field. Kriging, or a combination of kriging and stochastic simulation, will yield a geologically more plausible result. A typical example is shown in figure 9, where the final permeability fields for case 2 using the 2-level_1 method based on equations (18), (19) and equations (26), (27) are compared. A comparison between different downscaling methods with respect to the convergence of the final fine-scale tuning is shown in figure 10. The differences are not very large, and other cases presented by Aanonsen [1] also indicate that as long as care is taken to avoid extreme parameter values during the initial downscaling, the efficiency of the multiscale method is relatively insensitive to the initial downscaling method. The actual solution, however, may be very different, as illustrated in figure 9 and also demonstrated in the next section.

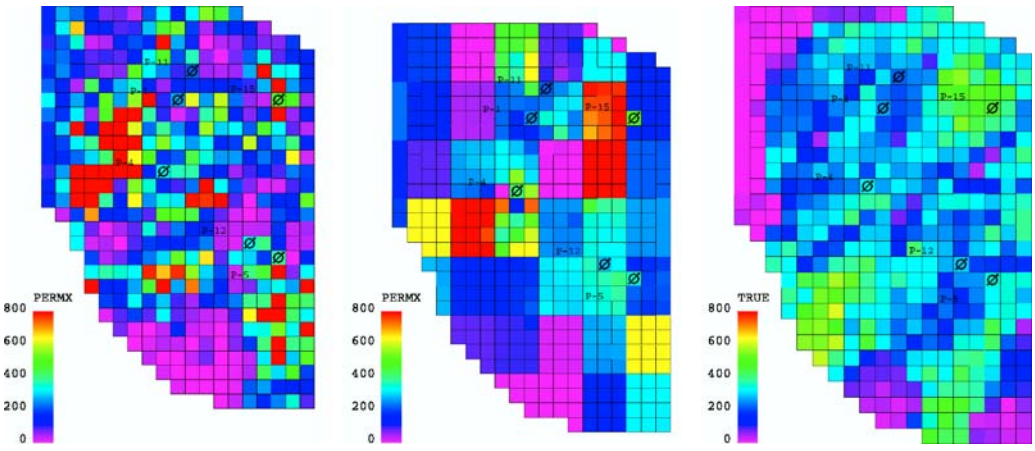

Figure 11. Estimated permeabilities case 1. Direct fine-scale history matching (*left*) and 2-level_1 method (*middle*) compared with true permeability (*right*).

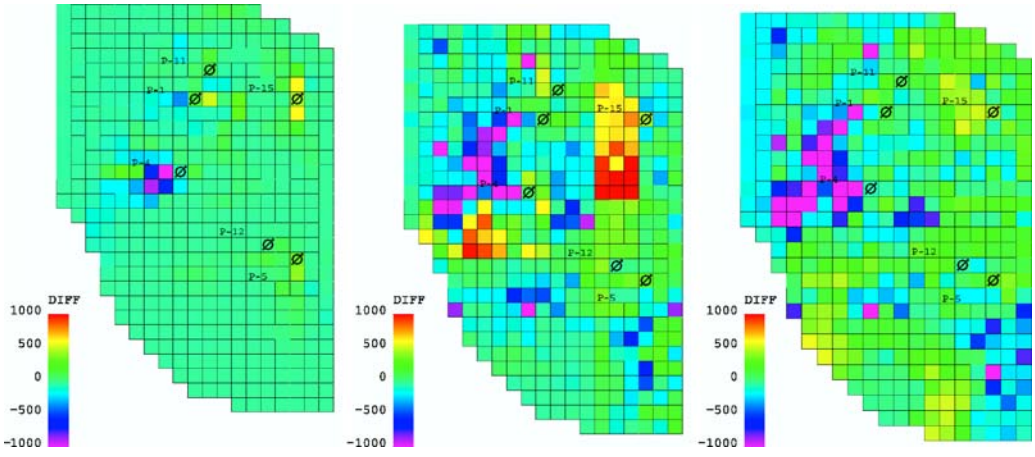

Figure 12. Difference between final and initial permeabilities case 1. Direct fine-scale history matching (*left*) and 2-level_1 method (*middle*) compared with the difference between the true and initial permeability (*right*).

3.2.3. Solution quality

The main objective of the multiscale technique is to obtain a solution that matches data on the fine scale faster than with a direct fine-scale matching. However, it is believed that this procedure in many cases also will give a better solution. Large-scale trends in the permeability field should be captured in the coarse runs, and in the fine-grid models, only small modifications, necessary to match details in well behavior, should be incorporated. The solution is also believed to be less dependent on a correct a priori model.

Figure 11 shows estimated permeabilities for case 1 compared with the true permeability. Apparently, the 2-level result is not any better than the result from the direct fine-scale history matching. However, by comparing the differences from the initial permeabilities (figure 12), the picture is considerably different. Now, it is clearly seen that for a direct history match on the fine scale, the solution is obtained by doing large modifications in just a few cells, a solution that is normally not desirable from a geological point of view. Notice that in the difference field, the coarse cell structure seen in the middle plot of figure 11 is smeared out by the large small-scale variations in the initial permeability field.

This result is for model 1, case 1 using the 2-level_1 method, but similar results have been obtained for all the multiscale runs, indicating that the results using the multiscale technique are more robust and less dependent of a correct prior model.

4. Discussion

Through several examples, and using two different simulators and optimization algorithms, we have shown that a multiscale approach may be considerably more efficient for obtaining a model that matches the data than a direct parameter estimation on the fine scale. The main reason for the improvement at a given scale seems to be that the initial guess is based on a coarser scale solution, independent on the method used for the initial downscaling and the number of coarser scales applied. However, since this applies at any level, it may be advantageous to introduce several levels. For model 1, the coarsest model is very coarse, with more than one well in a single cell. Still, going to this extreme level of coarsening was positive for the tuning at the finer scale. The optimum degree of coarsening applied at each level will have to be further investigated.

Two multilevel methods have been presented: one method with global upscaling and one without. With respect to the number of iterations needed for the final fine-scale tuning, the 2-level_1 method (without global upscaling) is performing better than the other method in most of the cases. One problem with the 2-level_2 method (with global upscaling) is that some large, unphysical parameters may be introduced during the upscaling procedure, and if this occurs in, or close to, well blocks, the initial RMS for the fine-scale tuning may be large. However, the large-scale trend still seems to have been recovered, and the objective function in these cases decreases very rapidly (figure 5). Also, the global upscaling provides a method for a more accurate estimation of an upscaling error, which, when added to the data error, will have a positive effect on the coarse-scale runs. The coarse-scale tuning converges faster since the accuracy required for convergence is reduced, and problem data with large model errors are given a lower weight. On the other hand, the global upscaling requires an additional optimization to be performed at each level, except at the finest scale.

The main focus of this paper is on the effort required to obtain a model that matches the data, and not on the quality of the solution. However, the multiscale method will provide a regularization, which may have a positive effect on the solution, by avoiding some extreme values. Although the convergence rate at a given scale seems to be relatively insensitive to the method used for the initial downscaling, different initial downscaling techniques will result in very different solutions, and this may be used to obtain solutions with desired properties. For instance, solutions with fine-scale heterogeneities may be obtained by using an initial downscaling based on a combination of sequential Gaussian simulation and block kriging as demonstrated by Aanonsen [1].

5. Conclusions

A general framework for multiscale parameter estimation, with application to reservoir history matching, has been presented.

It is shown through numerical examples that this method may significantly reduce the computational effort and/or improve the quality of the solution when achieving a fine-scale match as compared to history-matching directly on the fine scale.

Acknowledgments

The authors thank Professor Albert C. Reynolds for many fruitful discussions on these matters, Jarle Haukås for valuable help with the IMPSAT code and Schlumberger Information Solutions for providing the software licences used to run models 1 and 2.

References

[1] S.I. Aanonsen, Efficient history matching using a multiscale technique, in: *2005 SPE Reservoir Simulation Symposium, SPE 92758* (Society of Petroleum Engineers, Feb. 2005).

[2] R.A. Behrens, M.K. MacLeod, T.T. Tran and A.O. Alimi, Incorporating seismic attribute maps in 3D reservoir models, *SPE Reserv. Evalu. Eng.* (April 1998) 122–126.

[3] H. Ben Ameur, G. Chavent and J. Jaffré, Refinement and coarsening indicators for adaptive parameterization: application to the estimation of hydraulic transmissivities, Inverse Probl. 18(3) (2002) 775–794.

[4] R.H. Byrd, P. Lu and J. Nocedal, A limited memory algorithm for bound constrained optimization, SIAM J. Sci. Statist. Comput. 16(5) (1995) 1190–1208.

[5] G. Chavent and R. Bissell, Indicator for the refinement of parameterization, in: *Proc. of the 2nd Internat. Conf. in Inverse Problems in Engineering, ISIP 98*, eds. M. Tanaka and G.S. Dulikravich (Elsevier, Amsterdam, 1998) pp. 309–314.

[6] N.A.C. Cressie, *Statistics for Spatial Data* (Wiley, New York, 1991).

[7] F.J.T. Floris, M.D. Bush, M. Cuypers, F. Roggero and A.-R. Syversveen, Methods for quantifying the uncertainty of production forecasts: a comparative study, Pet. Geosci. 7(Suppl.) (2001) 87–96.

[8] A.-A., Grimstad, T. Mannseth, G. Nævdal and H. Urkedal, Adaptive multiscale permeability estimation, Comput. Geosci. 7 (2003) 1–25.

[9] A.-A. Grimstad and T. Mannseth, Comparison of methods for downscaling of coarse scale permeability estimates, in: *Proc., 9th European Conference on the Mathematics of Oil Recovery* (Cannes, France, 30 Aug. – 2 Sept., 2004).

[10] J. Haukås, I. Aavatsmark and M. Espedal, A black-oil and compositional IMPSAT simulator with improved compositional convergence, in: *Proc., 9th European Conference on the Mathematics of Oil Recovery*, (Cannes, France, 30 Aug. – 2 Sept., 2004).

[11] E.H. Isaaks and R.M. Shrivastava, *An Introduction to Applied Geostatistics* (Oxford University Press, New York, 1989).

[12] P.-K. Kitanidis, Quasi-linear geostatistical theory for inversing, Water Resour. Res. 31(10) (1995) 2411–2420.

[13] J. Liu, A multiresolution method for distributed parameter estimation, SIAM J. Sci. Comput. 14(2) (1993) 389–405.

[14] N. Liu and D.S. Oliver, Evaluation of Monte Carlo methods for assessing uncertainty, Soc. Pet. Eng. J. 8(2) (2003) 188–195.

[15] D.G. Luenberger, *Optimization by Vector Space Methods* (Wiley, New York, 1969).

[16] D.S Oliver, N. He and A.C. Reynolds, Conditioning permeability fields to pressure data, in: *Proc., 6th European Conference on the Mathematics of Oil Recovery,* Leoben, Austria, 3–6 Sept., 1996.

[17] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation* (Society for Industrial and Applied Mathematics, Philadelphia, 2005).

[18] S. Yoon, A.H. Malallah, A. Datta-Gupta, D.W. Vasco and R.A. Behrens, A multiscale approach to production data integration using streamline models, in: *1999 SPE Annual Technical Conference and Exhibition, SPE 56653* (Society of Petroleum Engineers, Oct. 1999).

[19] C. Zhu, R.H. Byrd and J. Nocedal, L-BFGS-B: Algorithm 778: L-BFGS-B, FORTRAN routines for large scale bound constrained optimization, ACM Trans. Math. Softw. 23(4) (1997) 550–560.