

Reservoir Description with Integrated Multiwell Data Using Two-Dimensional Wavelets

Abeeb A. Awotunde · Roland N. Horne

Received: 4 August 2012 / Accepted: 3 January 2013 / Published online: 12 February 2013
© International Association for Mathematical Geosciences 2013

Abstract Inferring reservoir data from dynamic production data has long been done through matching the production history. However, proper integration of available production history has always been a challenge. Different production history data such as well pressure and water cut often occur at different scales making their joint inversion difficult. Furthermore, production data obtained from the same well or even the same reservoir are often correlated making a significant portion of the dataset redundant. Thirdly, the massiveness of the data recorded from wells in a large reservoir over a long period of time makes the nonlinear inversion of such data computational demanding. In this paper, we propose the integration of multiwell production data using wavelet transform. The method involves the use of a two-dimensional wavelet transformation of the data space in order to integrate multiple production data and reduce the correlation between multiwell data. Multiple datasets from different wells, representing different production responses (pressure, water cut, etc.), were treated as a single matrix of data rather than separate vectors that assume no correlation amongst datasets. This enabled us to transform the multiwell production data into a two-dimensional wavelet domain and subsequently select the most important wavelets for history match. By minimizing the square of the Frobenius norm of the residual matrix we were able to match the calculated response to the observed response. We derived the relationship that allows us to replace a conventional minimization of the sum of squares of the l_2 norms of multi-objective functions with the minimization of the square of the Frobenius norm of the integrated data. The usefulness of the approach is demonstrated using two examples. The approach proved very effective at reducing correlation between multiwell data. In addition, the method

A.A. Awotunde (✉)
King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia
e-mail: awotunde@kfupm.edu.sa

R.N. Horne
Stanford University, Stanford, CA 94305, USA
e-mail: horne@stanford.edu

helped to reduce the cost of computing sensitivity coefficients. However, the method gave poor prediction of water cut when the datasets were not scaled before inverse modeling.

Keywords Parameter estimation · Data integration · Wavelet transform · Adjoint sensitivity · Reparameterization

1 Introduction

A common challenge in reservoir modeling is to integrate multiple reservoir and production data properly for efficient reservoir management. Reservoir and production data exist at multiple scales making such integration difficult. The presence of multiple scales in reservoir heterogeneity necessitates that the parameters be resolved into separate scales in order to identify the scale at which each parameter-component affects reservoir performance the most. Production data must also be resolved into different scales in order to identify the most relevant portion of the data. Furthermore, in evaluating large-scale reservoir performance it is necessary to determine the distributed reservoir parameter (for example, permeability) field. Estimation of reservoir parameter distributions in turn depends on the availability and efficient integration of valuable static data (for example core and well log data) and dynamic data (for example, production data). The most common method of estimating distributed reservoir parameters from dynamic production data is through inverse modeling. Integrating multiple production data during inverse modeling poses several challenges. One major challenge is the correlation in data. It is necessary to decorrelate the data in order to extract the most useful information from them. The problem of inverse analysis is further complicated when the constraining data exist at different scales of resolution. In this case, integrating the data to reduce the scale effect becomes necessary. A commonly adopted approach is to use different scaling factors to bring the different datasets to the same resolution. However, there is no unique way to select scaling factors and this introduces some bias into the inverse modeling, giving more importance to some datasets. A third difficulty is in handling a voluminous dataset within given limited computational resources.

Several attempts have been made to resolve these problems. Reservoir fluid flow theory is based on the continuum theory in which flow equations are derived based on mass conservation over a representative control volume (REV) whose properties are assumed to be averages of the properties at microscopic scales (Peaceman 1977; Aziz and Settari 1979). For several years, upscaling (Chu et al. 1998; Nakashima and Durlafsky 2010) has been used widely to integrate reservoir parameters representing multiple grid scales. The upscaled parameters are often some form of averaged grid block parameters representative of the microscopic properties over the grid blocks. In geostatistical reservoir modeling, conditional simulation (Journel 1974; Strebelle et al. 2003) has been proposed to integrate different types of static reservoir data while the integration of dynamic production data into reservoir modeling has been done mainly by history matching (Jacquard and Jain 1965; Carter et al. 1974; Chavent et al. 1975). Several history matching methods including gradual deformation (Hu 2000, 2002), probability perturbation method (Caers 2003; Johansen et al.

2007), wavelet approach (Huang and Kelkar 1996; Lu and Horne 2000; Sahni and Horne 2005, 2006), and streamline-based simulation approach (Yoon et al. 2001; He et al. 2002; Al-Harbi et al. 2005) have been proposed. All these approaches integrate information from production data into reservoir modeling. To our knowledge, none of these approaches considered the integration of production data into a single indistinguishable form before or during history matching. Awotunde and Horne (2011b, 2012) proposed the use of one-dimensional wavelet transform to integrate dynamic data from the same well. However, the focus of their work was mainly to reduce the computational overhead involved in estimating sensitivities of the integrated production data to integrated reservoir parameters.

The main theme of this paper is the integration of production data obtained from different wells producing from the same reservoir. We proceed first by presenting a framework that enables us to integrate production data using a two-dimensional wavelet transform. Here, we derived the relationship that allows us to replace a conventional minimization of the square of the norms of multiobjective functions with the minimization of the square of the Frobenius norm of the integrated data in wavelet domain. The procedure is then coupled to a wavelet decomposition of the parameter space in order to solve the inverse problem. The approach helps to integrate and decorrelate production datasets. Apart from the benefit obtained from decorrelation of data, it is also shown that the approach reduces the size of data needed for parameter estimation significantly, thereby making the process more efficient computationally. The usefulness of the approach is demonstrated using two numerical examples. Comparison is made to the conventional approach (no transformation of data) and the $wp - wk$ approach (Awotunde and Horne 2012).

2 Theory

In this section, we describe topics that are relevant to the understanding of the formulations in this paper.

2.1 Data Integration

A major challenge in reservoir flow modeling is the integration of different data obtained from the same reservoir. Because reservoir heterogeneity often exists at multiple and largely different scales, reservoir parameters at different scales must be properly integrated for accurate flow prediction. Several approaches to integrate reservoir flow parameters have been proposed. One approach is to transform the parameter space into a wavelet domain that allows reduction in parameter space (Huang and Kelkar 1996; Lu and Horne 2000; Sahni and Horne 2005). This approach reduces the correlation in the transformed parameter and has proven to be very efficient in reducing the nonuniqueness associated reservoir parameter estimation and by extension the uncertainty in estimated reservoir parameters. Liu et al. (2006) analyzed interwell heterogeneity by correlating injection and production data. Another challenge is integrating production data that exist at the same or different scales before or during inverse modeling. Panda et al. (2001) presented a method for identifying the impact

of multiscale data on reservoir modeling. History matching of pressure data from different wells is an example of an inverse modeling involving production data at the same scale of resolution. Integrating these data will help identify the most significant part of the data and reduce redundancies in the dataset. An orthogonal transformation (such as Karhunen Loeve, wavelet, etc.) of the data will help decorrelate the data, enabling better exploitation of the important components in the dataset. An example of inverse modeling involving data at different scales is the simultaneous matching of pressure and water cut data from different wells located in the same reservoir. These two data types exist at different scales and integrating both during history matching is more difficult. The method often adopted is to scale each dataset with either the standard deviation or the largest member (in absolute value) of that dataset. These two scales will lead to different estimates even when the initial guesses are the same in both cases. Awotunde and Horne (2012) introduced the use of one-dimensional wavelet decomposition ($wp - k$ and $wp - wk$) of production data in parameter estimation but applied the approach separately to one-dimensional datasets. Scale factors were used subsequently to integrate the transformed datasets during parameter estimation. In this work, we study how different datasets at different scales can be integrated using a two-dimensional wavelet transform. Our aim here is to propose a platform that allows a two-dimensional wavelet transform of multiple production datasets rather than a separate transform for each dataset.

2.2 The Two-Dimensional Wavelet Transform

Two-dimensional wavelet transforms are useful in decomposing two-dimensional datasets. In general, a two-dimensional wavelet transform is composed of four functions: a two-variable scaling function, $\phi(x, y)$ defined (Chui 1992; Percival and Walden 2000) as

$$\phi_{j,m,n}(x, y) = 2^{j/2} \phi(2^j x - m, 2^j y - n), \quad (1)$$

and three two-variable wavelet functions given by

$$\psi_{j,m,n}^i(x, y) = 2^{j/2} \psi^i(2^j x - m, 2^j y - n), \quad \forall i \in \{H, V, D\}, \quad (2)$$

where $j, m, n \in \mathbb{Z}$, $\phi(x, y)$ is a scaling function and $\psi^H(x, y)$, $\psi^V(x, y)$ and $\psi^D(x, y)$ are three different wavelet functions. If the filters are separable, then each of the two-variable scaling and wavelet functions can be written as the product of two one-variable functions

$$\phi(x, y) = \phi(x)\phi(y), \quad (3)$$

$$\psi^H(x, y) = \psi(x)\phi(y), \quad (4)$$

$$\psi^V(x, y) = \phi(x)\psi(y), \quad (5)$$

and

$$\psi^D(x, y) = \psi(x)\psi(y). \quad (6)$$

Equations (3) to (6) indicate that two-dimensional separable filters can be designed directly from their one-dimensional counterpart. Thus, in discrete form, the two-dimensional wavelet transform operates on an image (a two-dimensional dataset) by

first applying a one-dimensional discrete wavelet transform (DWT) on each column of the data and then applying the one-dimensional DWT on each row of the transformed data. This is equivalent to

$$\tilde{D}_w = W_t \tilde{D} W_{ds}^T, \tag{7}$$

where \tilde{D} is the original dataset, W_t is the wavelet matrix that transforms the columns of \tilde{D} , W_{ds} is the wavelet matrix that transforms the rows of \tilde{D} and \tilde{D}_w is the two-dimensional wavelet transform of \tilde{D} . The subscripts t and ds in Eq. (7) indicates decomposition along the time domain and across datasets, respectively. The wavelet matrix can be subdivided into two filters: the low pass filter H that computes the normalized pairwise averages of the input signal (column vector) and the high pass filter G that computes the normalized pairwise differences of the input signal. With these definitions, the linear operation in Eq. (7) can be represented by

$$\tilde{D}_w = W_t \tilde{D} W_{ds}^T = \begin{bmatrix} H_t \\ G_t \end{bmatrix} \tilde{D} \begin{bmatrix} H_{ds} \\ G_{ds} \end{bmatrix}^T. \tag{8}$$

Equation (8) can be written as

$$\tilde{D}_w = \begin{bmatrix} H_t \tilde{D} \\ G_t \tilde{D} \end{bmatrix} \begin{bmatrix} H_{ds}^T & | & G_{ds}^T \end{bmatrix}^T. \tag{9}$$

By rearranging the terms on the right-hand side of Eq. (9), we obtain

$$\tilde{D}_w = \begin{bmatrix} H_t \tilde{D} H_{ds}^T & | & H_t \tilde{D} G_{ds}^T \\ G_t \tilde{D} H_{ds}^T & | & G_t \tilde{D} G_{ds}^T \end{bmatrix} = \begin{bmatrix} B & | & V \\ H & | & D \end{bmatrix}. \tag{10}$$

In Eq. (10), the linear transform $H_t \tilde{D} H_{ds}^T$ computes the normalized pairwise averages along the columns of \tilde{D} , and then the pairwise averages along the rows of $H_t \tilde{D}$. This produces an approximation (or blur) B of \tilde{D} . $H_t \tilde{D} G_{ds}^T$ computes the normalized pairwise averages along the columns of \tilde{D} and then the normalized pairwise differences along the rows of $H_t \tilde{D}$. This will produce the vertical differences V between B and \tilde{D} . $G_t \tilde{D} H_{ds}^T$ computes the normalized pairwise differences along the columns of \tilde{D} and then the normalized pairwise averages along the rows of $G_t \tilde{D}$. This operation produces the horizontal differences H between B and \tilde{D} . The last linear operation, $G_t \tilde{D} G_{ds}^T$, computes the normalized pairwise differences along the columns of \tilde{D} and then the normalized pairwise differences along the rows of $G_t \tilde{D}$ to produce the diagonal differences D between B and \tilde{D} .

2.3 Frobenius Norm

The Frobenius norm of an $N \times L$ matrix, \tilde{D} , is defined as the square root of the sum of the absolute squares of its elements (Golub and Van Loan 1996)

$$\|\tilde{D}\|_F = \sqrt{\sum_{n=1}^N \sum_{l=1}^L |d_{n,l}|^2} = \sqrt{\text{trace}(\tilde{D}^* \tilde{D})}. \tag{11}$$

In Eq. (11), $d_{n,l}$ is the element in row n and column l of \tilde{D} , \tilde{D}^* is the conjugate transpose of \tilde{D} , and the trace function is used. The square of the Frobenius norm of \tilde{D} is equivalent to the sum of squares of the l_2 -norms of the column vectors in \tilde{D} . Thus, the sum of squares of l_2 -norms of several vectors of equal length can be replaced by the Frobenius norm of a single matrix whose columns are composed of those vectors.

3 Analysis of Multiwell Data

In practical field scenarios, it is often the case that different types of data from the same reservoir are available for history-matching. In such cases, it becomes necessary to constrain the model to match all available data simultaneously. Thus, this becomes a multiobjective minimization problem. The objective here is to minimize simultaneously the errors in all the datasets. Such minimization presents several challenges ranging from correlation of datasets to existence of data at different scales of resolution. Measured data often have different scales of resolution. For example, water cut often exists at a scale vastly different from the scale at which bottomhole pressures occur. However, residuals of these two measurements must be minimized simultaneously. The approach often adopted is to use different scaling factors for different datasets. The choice of scaling factors will no doubt introduce some level of bias in the minimization algorithm. In this work, we investigated the use of wavelet transform to integrate multiple production data. Furthermore, history data from multiple wells producing from the same reservoir are often correlated (well-to-well). Such correlation may adversely affect the capability of history matching algorithms to solve the inverse problem because correlation creates redundancies in history data. However, the discrete wavelet transform is known to decorrelate certain time series (Percival and Walden 2000). In this regard, we explored a means to decorrelate the multiple production histories using a two-dimensional wavelet transform.

3.1 Two-Dimensional Wavelet Transform of Data Space

Production data can be integrated by taking a two-dimensional wavelet transform of the data. If we consider a production history composed of L datasets $\{\mathbf{d}_{1,meas}, \mathbf{d}_{2,meas}, \dots, \mathbf{d}_{L,meas}\}$, each of length N , then the equation for multiobjective function minimization can be expressed as

$$\min_{\alpha} \{\Phi(\alpha)\} = \min_{\alpha} \left\{ \frac{1}{2} [\|\mathbf{d}_{1,cal} - \mathbf{d}_{1,meas}\|_2^2 + \|\mathbf{d}_{2,cal} - \mathbf{d}_{2,meas}\|_2^2 + \dots + \|\mathbf{d}_{L,cal} - \mathbf{d}_{L,meas}\|_2^2] \right\}. \quad (12)$$

In Eq. (12), $\mathbf{d}_{l,cal}$ is the l th calculated dataset, $\mathbf{d}_{l,meas}$ is the l th measured dataset, and $\mathbf{d}_{l,cal}$ is a function of the model parameter vector α . The sum of squares of l_2 -norms of the residuals appearing in Eq. (12) can be replaced by the square of the Frobenius norm of a single residual matrix. Thus, the measured datasets and calculated datasets can be grouped into two separate matrices with the square of the Frobenius norm

of their residual replacing the sum of squares of the several l_2 -norms appearing in Eq. (12). This relationship is given by

$$\begin{aligned} & \|\tilde{D}_{cal} - \tilde{D}_{meas}\|_F^2 \\ &= \|\mathbf{d}_{1,cal} - \mathbf{d}_{1,meas}\|_2^2 + \|\mathbf{d}_{2,cal} - \mathbf{d}_{2,meas}\|_2^2 + \dots + \|\mathbf{d}_{L,cal} - \mathbf{d}_{L,meas}\|_2^2, \end{aligned} \tag{13}$$

where

$$\tilde{D}_{cal} = [\mathbf{d}_{1,cal} \mathbf{d}_{2,cal} \dots \mathbf{d}_{L,cal}], \tag{14}$$

and

$$\tilde{D}_{meas} = [\mathbf{d}_{1,meas} \mathbf{d}_{2,meas} \dots \mathbf{d}_{L,meas}]. \tag{15}$$

However, a two-dimensional orthogonal wavelet transform of a matrix does not change the Frobenius norm of the matrix (Appendix A). Consequently, the following relationship holds

$$\|\tilde{D}_{wcal} - \tilde{D}_{wmeas}\|_F^2 = \|\tilde{D}_{cal} - \tilde{D}_{meas}\|_F^2, \tag{16}$$

where

$$\tilde{D}_{wmeas} = W_t \tilde{D}_{meas} W_{ds}^T, \tag{17}$$

and

$$\tilde{D}_{wcal} = W_t \tilde{D}_{cal} W_{ds}^T. \tag{18}$$

In Eqs. (17) and (18), W_t is an $N \times N$ wavelet matrix that transforms the columns of \tilde{D} and W_{ds} is an $L \times L$ wavelet matrix that transforms the rows of \tilde{D} . The equivalence of the Frobenius norms in the two domains (time-space and wavelet) as shown in Eq. (16) implies that we can replace the summation of several l_2 -norms in Eq. (12) by the Frobenius norm of a two-dimensional wavelet transform of the residual matrix. Therefore, Eq. (12) becomes

$$\min_{\alpha} \{\Phi(\alpha)\} = \min_{\alpha} \left\{ \frac{1}{2} \|\tilde{D}_{wcal}(\alpha) - \tilde{D}_{wmeas}\|_F^2 \right\}. \tag{19}$$

If the parameter space is transformed equally into a wavelet domain then the parameter α in Eq. (19) should be replaced by \mathbf{c}_{α} . Although the formulations in Eqs. (12) through (19) help us to move from the real space to the wavelet space, it does not provide any advantage unless a reduction in the transformed domain is made. Thus, in practice, W_t is an $N_r \times N$ matrix that reduces the number of rows in \tilde{D} from N to N_r and W_{ds} is an $L_r \times L$ matrix that reduces the number of columns in \tilde{D} from L to L_r . The bases in W_t and W_{ds} are selected such that most of the information in \tilde{D} are retained after the transformation. Even though a matrix representation is used to illustrate the process of transformation, we do not form the matrices W_t and W_{ds} explicitly due to high cost of matrix storage and multiplication. Rather, the Mallat pyramidal algorithm (Mallat 1989) is used first to compute all the wavelet coefficients and the domain is subsequently thresholded to yield the reduced set of coefficients. The approach described here shows that we can perform multiobjective

Table 1 Description of approaches and their inputs

| Approaches | $p - k$ | $p - wk$ | $wp - k$ | $wp - wk$ | $2Dwp - k$ | $2Dwp - wk$ |
|-------------------|---------------------|---------------------|-------------------------|-------------------------|---------------------------------|---------------------------------|
| System parameters | α | α | α | α | α | α |
| Model parameters | α | \mathbf{c}_α | α | \mathbf{c}_α | α | \mathbf{c}_α |
| Measured data | \mathbf{d}_{meas} | \mathbf{d}_{meas} | \mathbf{d}_{meas} | \mathbf{d}_{meas} | $\tilde{\mathbf{D}}_{meas}$ | $\tilde{\mathbf{D}}_{meas}$ |
| Observation data | \mathbf{d}_{meas} | \mathbf{d}_{meas} | $\mathbf{c}_{d_{meas}}$ | $\mathbf{c}_{d_{meas}}$ | $\tilde{\mathbf{D}}_{w_{meas}}$ | $\tilde{\mathbf{D}}_{w_{meas}}$ |
| Transformation | none | par | meas | meas and par | meas | meas and par |

function minimization by matching a subset of the wavelets derived from the two-dimensional decomposition of the data space. The subset selected for history match comprises the most important coefficients to constrain the model. This dimension reduction in data space improves the efficiency of the algorithm by (1) decorrelating the dataset and (2) reducing cost associated with the computation of sensitivities (Awotunde and Horne 2012).

3.2 The $2Dwp - wk$ Approach

In this work, six different approaches are considered and compared (Table 1). All the approaches, except the $2Dwp - k$ and $2Dwp - wk$ have been presented previously (Lu and Horne 2000; Awotunde and Horne 2011a, 2011b, 2012). The $p - k$ approach is the conventional nonlinear regression in which the parameter space is reconstructed by matching all available data without any transformation (Levenberg 1944; Marquardt 1963). The $p - wk$ approach (Lu and Horne 2000) transforms the parameter space into wavelet, and subsequently reduces the transformed space by thresholding. The $wp - k$ and $2Dwp - k$ approaches transform only the measurement space into a reduced wavelet space while the $wp - wk$ and $2Dwp - wk$ approaches transform both the measurement and parameter spaces into reduced wavelet spaces. However, while $wp - k$ and $wp - wk$ perform one-dimensional wavelet transform and reduction on the measurement space, $2Dwp - k$ and $2Dwp - wk$ perform a two-dimensional wavelet transform and reduction of the data space. The two approaches become equivalent to the $p - k$ approach if all wavelets (no data space or model space reduction) are used in the nonlinear regression estimation (Appendix A). The last row in Table 1 describes the different approaches by indicating whether the parameter space, the measurement space or both spaces were transformed. In particular, the $2Dwp - wk$ approach involves wavelet transformation and thresholding of both the data space and the parameter space. This makes it possible to integrate data from different wells. The transformation and reduction of the data space is carried out in the following way:

1. Group all measured datasets into a single $N \times L$ matrix $\tilde{\mathbf{D}}_{meas}$, with each column of the matrix representing a separate dataset.
2. Perform a two-dimensional wavelet decomposition of the matrix using the pyramidal algorithm. There is no reduction at this stage.
3. Set a threshold value for the columns and another threshold value for the rows.
4. In each column, find the entry with the largest absolute value and compare this absolute value with the preset threshold values for columns. Retain all columns

that have their largest absolute entry greater than the preset column threshold and discard the other columns. This yields a column-reduced matrix of dimension $N \times L_r$.

5. In each row of the new matrix find the entry with the largest absolute value and compare this absolute value with the preset threshold values for rows. Retain all rows that have their largest absolute entry greater than the preset row threshold and discard the other rows. This yields a row-and-column reduced matrix of dimension $N_r \times L_r$.

Other steps in the $2Dwp - wk$ approach are similar to Steps 4 to 10 of the $wp - wk$ approach presented in Awotunde and Horne (2012). The benefits of the $2Dwp - wk$ approach include:

1. The approach helps to integrate different types of production data from the same reservoir.
2. It helps to decorrelate multiwell production data.
3. It helps to reduce the cost of storage and computation of sensitivity coefficients.

Care must be taken to avoid overcompressing the data space. Too large compression of the data space can lead to loss of vital information in the data resulting in a poor match to production data. This is particularly important in large fields where the amount of production history data is enormous. In such cases, the number of wavelets retained after compression can be large enough to make the computation of sensitivity matrix very cumbersome and the Levenberg–Marquardt approach unattractive.

3.3 Decorrelation of Datasets

Usually measured data are correlated. Correlation in data means that such data carry the same or almost the same information about the model they describe. This situation is observed in many practical cases where the number of measured data is much larger than the number of parameters of the system. However, these data cannot resolve the parameters uniquely. The adverse effect of correlation in data can be illustrated with a simple example. Consider matching 20 data points in which 12 are uncorrelated and the remaining eight are correlated and carrying almost the same information. This is nearly equivalent to matching 13 distinct and uncorrelated data points with one of the data points assigned a weight that is eight times that of each of the other data points. A history match of these data will lead to better assimilation of the information in the data with the bigger weight and not-so-good assimilation of the information in the other data points leading to poor history match results. In essence, when data are correlated, many data points will carry the information that could be carried by just one data point thus concentrating the history matching effort on the same information at the expense of other information contained in the other uncorrelated data points. Decorrelation refers to any process that is used to reduce autocorrelation within a signal (a single dataset) or cross-correlation between a set of signals (datasets). A wavelet decomposition and reduction of a correlated dataset removes all or part of the correlation in the data so that the wavelet coefficients retained carry the relevant information at appropriate scales. Thus, the $2Dwp - wk$ approach is able to remove correlation in a measured dataset leading to a better history match.

Care must however be taken to ensure that relevant information in the data is not removed during wavelet reduction of the data space. The $wp - wk$ approach is also able to reduce the correlation in the dataset. However, examples presented in Sect. 4 of this paper show that the amount of reduction achieved by $wp - wk$ is smaller than that achieved by $2Dwp - wk$. It is also necessary to note that while the $wp - wk$ approach decomposes and reduces each data-vector along the time domain only, the $2Dwp - wk$ approach decomposes the data along the time domain as well as across datasets. Decomposition across datasets makes it possible for the $2Dwp - wk$ approach to remove/reduce correlation in data from different wells producing from the same reservoir.

3.4 Solution Method

To obtain a solution to the inverse problem, Eq. (19) is solved iteratively using the Levenberg–Marquardt method (Levenberg 1944; Marquardt 1963). This method involves computing the gradient $\mathbf{g}_{\mathbf{c}_\alpha}^{\tilde{D}^{w_{cal}}}$ of the objective function $\Phi(\mathbf{c}_\alpha)$ from

$$\mathbf{g}_{\mathbf{c}_\alpha}^{\tilde{D}^{w_{cal}}} = (S_{\mathbf{c}_\alpha}^{\tilde{D}^{w_{cal}}})^T \text{vec}(\tilde{D}^{w_{cal}} - \tilde{D}^{w_{meas}}), \tag{20}$$

computing an approximate Hessian matrix

$$H_{\mathbf{c}_\alpha}^{\tilde{D}^{w_{cal}}} = (S_{\mathbf{c}_\alpha}^{\tilde{D}^{w_{cal}}})^T S_{\mathbf{c}_\alpha}^{\tilde{D}^{w_{cal}}} + \lambda I, \tag{21}$$

solving for an approximate Newton direction $\delta \mathbf{c}_\alpha^\kappa$ from

$$H_{\mathbf{c}_\alpha}^{\tilde{D}^{w_{cal} \cdot \kappa}} \delta \mathbf{c}_\alpha^\kappa = -\mathbf{g}_{\mathbf{c}_\alpha}^{\tilde{D}^{w_{cal} \cdot \kappa}}, \tag{22}$$

at each iteration and then computing the new estimate of the parameter field from

$$\mathbf{c}_\alpha^{\kappa+1} = \mathbf{c}_\alpha^\kappa + \nu^\kappa \delta \mathbf{c}_\alpha^{\kappa+1}. \tag{23}$$

In Eq. (21), $S_{\mathbf{c}_\alpha}^{\tilde{D}^{w_{cal}}}$ is the sensitivity of the thresholded two-dimensional wavelet transform of the measured data to the model-space wavelet coefficients, I is an identity matrix the same size as $H_{\mathbf{c}_\alpha}^{\tilde{D}^{w_{cal}}}$ and λ is the Levenberg–Marquardt parameter (Levenberg 1944; Marquardt 1963) and vec is a function that reshapes a matrix into a column vector. The function vec reshapes a matrix into a column vector containing all the elements of that matrix. $S_{\mathbf{c}_\alpha}^{\tilde{D}^{w_{cal}}}$ and $\mathbf{g}_{\mathbf{c}_\alpha}^{\tilde{D}^{w_{cal}}}$ are calculated using a wavelet adjoint sensitivity method (Appendix B), and ν is the step-length calculated within each iteration using the backtracking-Armijo-line search method (Gill et al. 1981; Nocedal and Wright 2006; Griva et al. 1996). While λ ensures that the Hessian is positive-definite, ν ensures that an optimum step is taken in the computed direction. In the main iteration loop, $\mathbf{g}_{\mathbf{c}_\alpha}^{\tilde{D}^{w_{cal}}}$ is computed from $S_{\mathbf{c}_\alpha}^{\tilde{D}^{w_{cal}}}$ using Eq. (20) because it is cheaper to compute the gradient using Eq. (20) if the sensitivity matrix is already known. However, within the line-search iteration loop where $S_{\mathbf{c}_\alpha}^{\tilde{D}^{w_{cal}}}$ is not required or available, the wavelet adjoint method (Appendix B) is used to compute $\mathbf{g}_{\mathbf{c}_\alpha}^{\tilde{D}^{w_{cal}}}$.

4 Applications

Two example applications are presented to illustrate the usefulness of the approach. Both examples used simulated data with added Gaussian noise as the measured data for history matching. In all the examples, we used six different approaches: $p - k$, $p - wk$, $wp - k$, $wp - wk$, $2Dwp - k$, and $2Dwp - wk$ to estimate the permeability distributions in the reservoir. In addition, two separate versions each of the $p - k$, $2Dwp - k$ and $2Dwp - wk$ approaches were considered. The first version involves scaling the datasets before transformation while the second version does not involve scaling the datasets. In the first version, we scaled all pressure data with one scale factor and scaled all water cut data with another scale factor. This procedure is expressed as

$$\min_{\alpha} \{ \Phi(\alpha) \} = \min_{\alpha} \left\{ \frac{1}{2} \left[\frac{1}{\beta_1} \| \tilde{D}_{wcal}^{press} - \tilde{D}_{wmeas}^{press} \|_F^2 + \frac{1}{\beta_2} \| \tilde{D}_{wcal}^{wc} - \tilde{D}_{wmeas}^{wc} \|_F^2 \right] \right\}. \tag{24}$$

In Eq. (24), all pressure data are integrated into one lump ($\tilde{D}_{wmeas}^{press}$) and all water cut data are integrated into another lump (\tilde{D}_{wmeas}^{wc}). β_1 and β_2 are the scaling factors for the pressure and water cut data respectively. In the second version, all pressure and water cut data are integrated into one single lump as given by Eq. (19). In order to avoid confusing the effects of neglecting data at small scale to the effect of data integration by wavelet decomposition, we included two versions of the $p - k$ approach in these example applications. The first version involves scaling of datasets while the second version used the data without scaling.

The data space was thresholded based on the magnitudes of the wavelets of data while the model space was thresholded based on the magnitudes of entries in the wavelet sensitivity matrix. There were multiple injectors and producers in the example reservoirs. Pressure and water cut were recorded in all production wells and injection pressure was recorded simultaneously in all injectors. In the $p - k$ approach, all the measured data were matched by varying the logarithm of reservoir permeability. We tested two versions of the $p - k$ approach, one with and the other without scaling factors. This was done to enable the study of the true impact of the unscaled $2Dwp - k$, and $2Dwp - wk$ approaches. In the $p - wk$ approach, all the measured data were matched while varying a subset of wavelets of logarithm of reservoir permeability. The $p - k$ and $p - wk$ approaches require that the sensitivities of all datasets be computed. The $wp - k$, $wp - wk$, $2Dwp - k$, and $2Dwp - wk$ approaches, however, require the computation of a data-reduced wavelet sensitivity matrix. The $wp - k$, $wp - wk$, $2Dwp - k$, and $2Dwp - wk$ models used the wavelet approach to adjoint sensitivity computation to reduce the computational time and storage requirements. The reduction achieved is approximately equal to the compression ratio defined as

$$CR = \frac{\text{No. of measured data}}{\text{No. of wavelets retained}}. \tag{25}$$

While the $wp - k$ and $wp - wk$ approaches perform a one-dimensional wavelet decomposition of the data space, the $2Dwp - k$, and $2Dwp - wk$ perform a two-dimensional (standard Haar wavelet) decomposition of the data space. A two-

Table 2 Reservoir, well and fluid properties

| Properties | Values | Units |
|---------------|--------|-------------------|
| c_o | 1.2E-5 | psi ⁻¹ |
| c_w | 5E-7 | psi ⁻¹ |
| p_{ini} | 5000 | psi |
| $S_{w_{ini}}$ | 1E-6 | none |
| ϕ_{ini} | 0.25 | none |
| $skin$ | 0 | none |

dimensional wavelet transform (standard implementation) is performed on the parameter space by the $p - wk$, $wp - wk$ and $2Dwp - wk$ approaches. Because the parameter space in a multidimensional reservoir system is often large, it is not always possible to search for an optimum set of model parameters at every iteration. Consequently, we used a multiscale procedure in which a small set of model coefficients was retained at the start of the nonlinear regression and the set was gradually expanded until a desirable number of model coefficients (near-optimal set) were included in the model space. All the examples presented here used this multiscale procedure.

The l_1 -norms of the differences between the true permeability distribution and the estimated distributions are used to measure the closeness of the estimates to the true distribution. Thus, in this work, the residual norm of \mathbf{k} is defined as

$$k_{norm} = \frac{\|\mathbf{k}_{est} - \mathbf{k}_{true}\|_1}{L_{\mathbf{k}}}, \quad (26)$$

where $L_{\mathbf{k}}$ is the length of \mathbf{k} , \mathbf{k}_{true} is the true permeability distribution and \mathbf{k}_{est} is the estimated permeability distribution. These norms, although providing some insight into the performances of the approaches, cannot be taken as an ultimate measure of performance of the approaches. The reservoir, well and fluid properties are presented in Table 2.

4.1 Example 1: Reservoir with 16×16 Grid Blocks

This example comprises a reservoir discretized into 16×16 grid blocks (Fig. 1(a)). The reservoir has two injectors and three producers as shown in Fig. 1(b). The flowing bottomhole pressure (BHP) was measured at each injector while water cut and BHP were measured at each producer. In all we had eight sets of data; two sets of measured pressure signals from the injectors, three sets of pressure signals and three sets of water cut signals from the three producers. Each dataset had 256 data points. Thus, overall, there were 2048 measured data to be matched. Figures 2(a) and 2(b) show, respectively, the sets of pressure and water cut histories prior to wavelet transformation. While Fig. 2(c) shows the data after two-dimensional wavelet transformation and thresholding of the combined datasets, Fig. 2(d) shows the wavelet coefficients after two-dimensional wavelet transformation of the pressure data and Fig. 2(e) shows the coefficients after a two-dimensional wavelet transformation and thresholding of the water cut history data. In the unscaled versions of $2Dwp - k$ and $2Dwp - wk$,

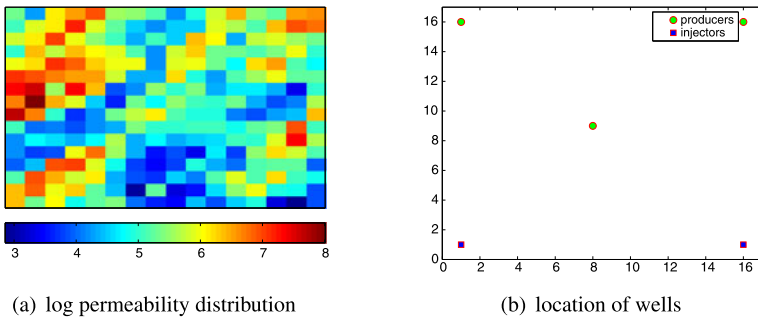


Fig. 1 Log permeability distribution and location of wells in the 16×16 reservoir model. (a) Log permeability distribution, (b) location of wells

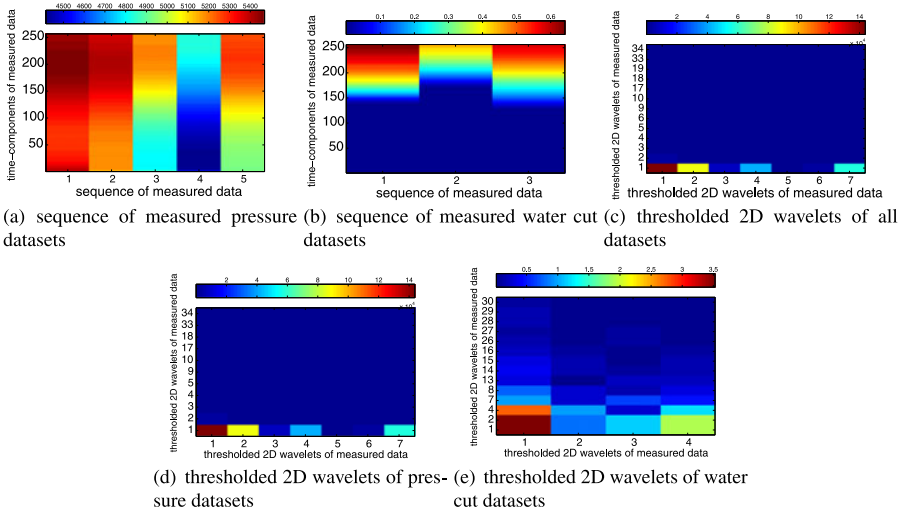


Fig. 2 Two-dimensional transformation of a sequence of datasets into thresholded wavelets (16×16 reservoir model). (a) Sequence of measured pressure datasets, (b) sequence of measured water cut datasets, (c) thresholded two-dimensional wavelets of all datasets, (d) thresholded two-dimensional wavelets of pressure datasets, (e) thresholded two-dimensional wavelets of water cut datasets

the dimension of the data space was reduced from 256×8 to 13×7 . A larger reduction is achieved along the time domain (256 to 13) than across datasets (8 to 7) indicating that the data from the same well taken over a time period is more correlated than data from different wells taken at the same time. In the scaled version, the dimension of the pressure dataset was reduced from 256×5 to 11×7 and the water cut dataset was reduced from 256×3 to 14×4 . Prior to wavelet transformation, the pressure and water cut datasets were supplemented with zero columns to make the number of columns be of order 2^j where j is an integer. Table 3 shows the number of data/coefficients matched as well as the number of model parameters used by each approach. The number of observation (data) points was reduced from 2048 to 376 by the $wp - k$ and $wp - wk$ approaches while the same num-

Table 3 Model performance data (16×16 reservoir model)

| Models | $p - k$ | $(p - k)_{w/osc}$ | $p - wk$ | $wp - k$ | $wp - wk$ | $2Dwp - k$ | $(2Dwp - k)_{w/osc}$ | $2Dwp - wk$ | $(2Dwp - wk)_{w/osc}$ |
|------------|---------|-------------------|----------|----------|-----------|------------|----------------------|-------------|-----------------------|
| N_{data} | 2048 | 2048 | 2048 | 376 | 376 | 133 | 91 | 133 | 91 |
| CR | 1 | 1 | 1 | 5.45 | 5.45 | 15.4 | 22.5 | 15.4 | 22.5 |
| N_p | 256 | 256 | 53–101 | 256 | 56–103 | 256 | 256 | 60–110 | 51–103 |
| k_{norm} | 285.6 | 287.9 | 211.3 | 283.4 | 239.3 | 340.9 | 292.7 | 238.1 | 195 |

ber was reduced by $2Dwp - k$ and $2Dwp - wk$ to 133 when a scaling factor was applied and to 91 when no scaling factor was applied, indicating that a higher compression ratio of the data space was achieved by the two-dimensional decomposition of the data space. The adjoint method was used for computing sensitivities in the $2Dwp - k$ and $2Dwp - wk$ method because the number of wavelets of data is smaller than the number of system or model parameters. Also note that the reduction in time achieved, by using a wavelet approach to adjoint formulation as opposed to the conventional adjoint sensitivity computation, is approximately proportional to the compression ratio defined in Eq. (21). In this example, the ratio was 15.4 and 22.5 for the two versions of $2Dwp - k$ and $2Dwp - wk$ considered. The residual norm of the difference between the true and estimated permeability distributions was used to measure the closeness of the modeled distributions to the true distribution (Table 3). The initial residual norm of \mathbf{k} for this example case was 309. Table 3 also shows that, for this example, the residual norm of \mathbf{k} was generally lower when the model space was reduced. This indicates that the number of system parameters used to parameterize this example model is larger than needed. Using fewer wavelets to parameterize the model space gave closer match to the true permeability distribution. Even though the number of measured data was much larger than the number of model parameters, none of the approaches (including the standard $p - k$ approach) gave an accurate estimate of the true permeability distribution. This shows that the information content in the measured data was not enough to resolve all the model parameters. Constraining the model with hard data or any accurate a priori information during inverse modeling can help improve the estimation of the model parameters.

In order to test the predictive ability of the approaches, the true reservoir permeability model, \mathbf{k}_{true} was used to predict the pressure and water cut 750 days beyond the end ($t = 1153$ days) of the measured data and the predicted data was compared with the data predicted by the six approaches. The matches to the measured data ($t \leq 1153$ days) and the predicted data ($t > 1153$ days) are shown in Figs. 3 and 4. All the approaches gave good matches to the measured pressure (Fig. 3) and to the predicted pressure data. Match to the water cut data was good in all cases except those from the unscaled versions of $2Dwp - k$ and $2Dwp - wk$ (Fig. 4). The predicted water cut data from the unscaled versions of the $p - k$, $2Dwp - k$ and $2Dwp - wk$ approach was worse than those from the other approaches. Clearly, there were mismatches due to the inability of the unscaled versions to reconcile data at vastly different scales. This effect is less conspicuous in the match to water cut data given by the unscaled version of the $p - k$ approach. The inability of the unscaled versions of the $2Dwp - k$

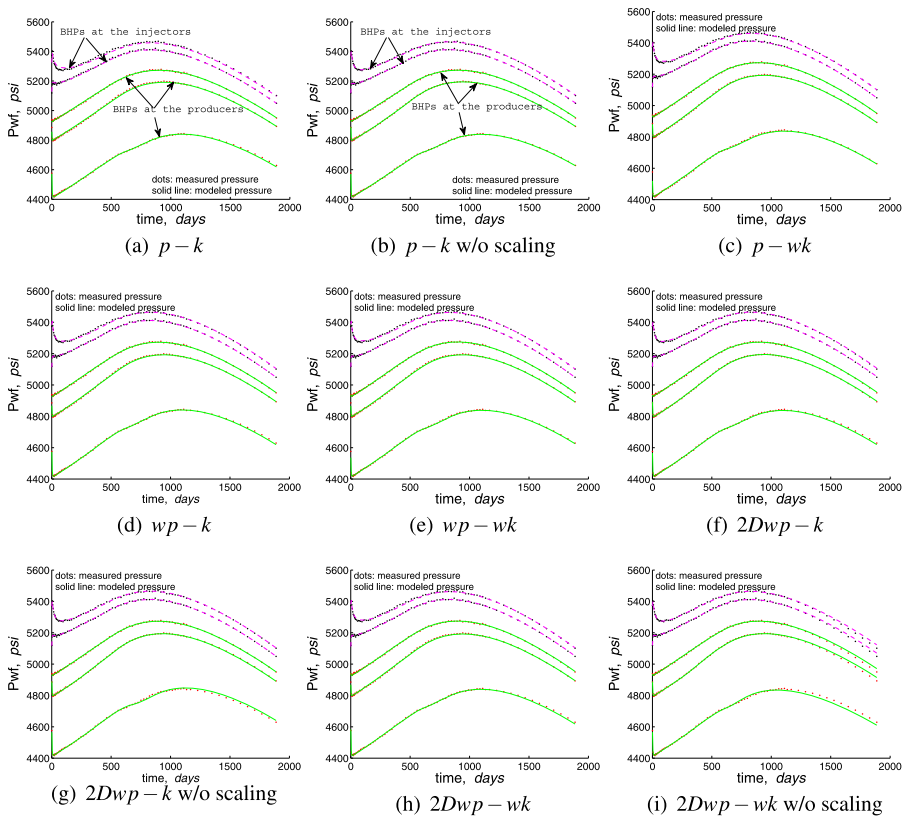


Fig. 3 Match to bottomhole pressure at the injectors and producers (16 × 16 reservoir model). (a) $p - k$, (b) $p - k$ w/o scaling, (c) $p - wk$, (d) $wp - k$, (e) $wp - wk$, (f) $2Dwp - k$, (g) $2Dwp - k$ w/o scaling, (h) $2Dwp - wk$, (i) $2Dwp - wk$ w/o scaling

and $2Dwp - wk$ approaches to reconcile data at the vastly different scales leads to the elimination of some wavelet coefficients carrying important information about the water cut in these approaches. This shows that the unscaled versions of the $p - k$, $2Dwp - k$, and $2Dwp - wk$ were unable to fully integrate data at the different scales of resolution encountered in this problem. Nevertheless, the unscaled versions gave lower overall residual values (better match to production data) than their scaled counterparts. Figure 5 shows the initial log permeability guess, the true log permeability distribution and the estimates of the log permeability given by all approaches. We observe that the unscaled versions of the $p - k$, $2Dwp - k$ and $2Dwp - wk$ approaches gave estimates (Figs. 5(d), 5(i), and 5(k), respectively) that are closer to the true log permeability distribution than the scaled versions. Since the unscaled versions match the pressure better than they match the water cut data, these results suggest that pressure data carry more important information about the reservoir permeability distribution than the water cut data. Because the overall objective of the history matching procedure is to obtain accurate estimates of reservoir parameter distribution, the unscaled versions may serve as better alternatives to the scaled versions. We also observe that the results obtained by the unscaled versions of the $2Dwp - k$ and

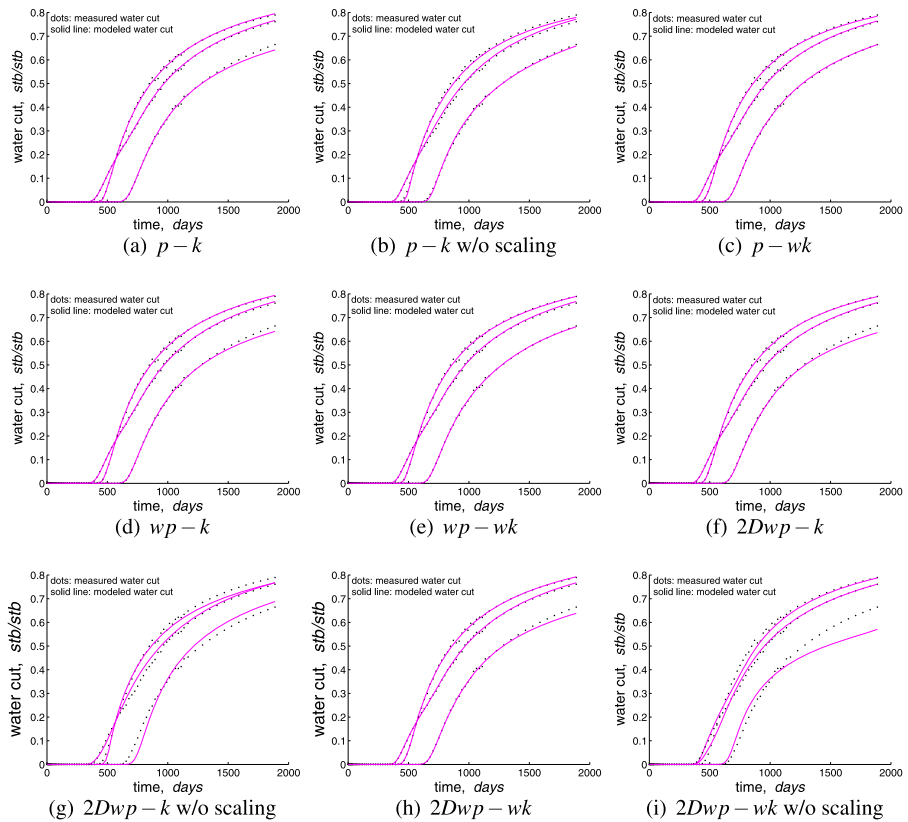


Fig. 4 Match to water cut at the producers (16×16 reservoir model). (a) $p - k$, (b) $p - k$ w/o scaling, (c) $p - wk$, (d) $wp - k$, (e) $wp - wk$, (f) $2Dwp - k$, (g) $2Dwp - k$ w/o scaling, (h) $2Dwp - wk$, (i) $2Dwp - wk$ w/o scaling

$2Dwp - wk$ approaches are better than (in terms of more accurate log permeability maps in Fig. 5 and lower data mismatch in Fig. 6) those obtained from the unscaled version of the $p - k$ approach indicating that it is better to integrate the multiwell data before solving the inverse problem. In addition, Fig. 6 shows that the unscaled versions of the $2Dwp - k$ and $2Dwp - wk$ approaches converged faster than the other approaches. The number of model parameters selected by the $p - wk$, $wp - wk$, and $2Dwp - wk$ approaches at different iteration counts are shown in Fig. 7. In these three approaches, we retained approximately 40 % of the model-space coefficients for history matching.

4.2 Example 2: Reservoir with 64×64 Grid Blocks

This example was designed to test the abilities of the six approaches to map the high permeability path in a heterogeneous reservoir (Fig. 8(a)). The system was a 100 ft thick reservoir of length 6400 ft and width 6400 ft discretized into 64×64 gridblocks. The reservoir had an underlying permeability of 100 md with some medium to high

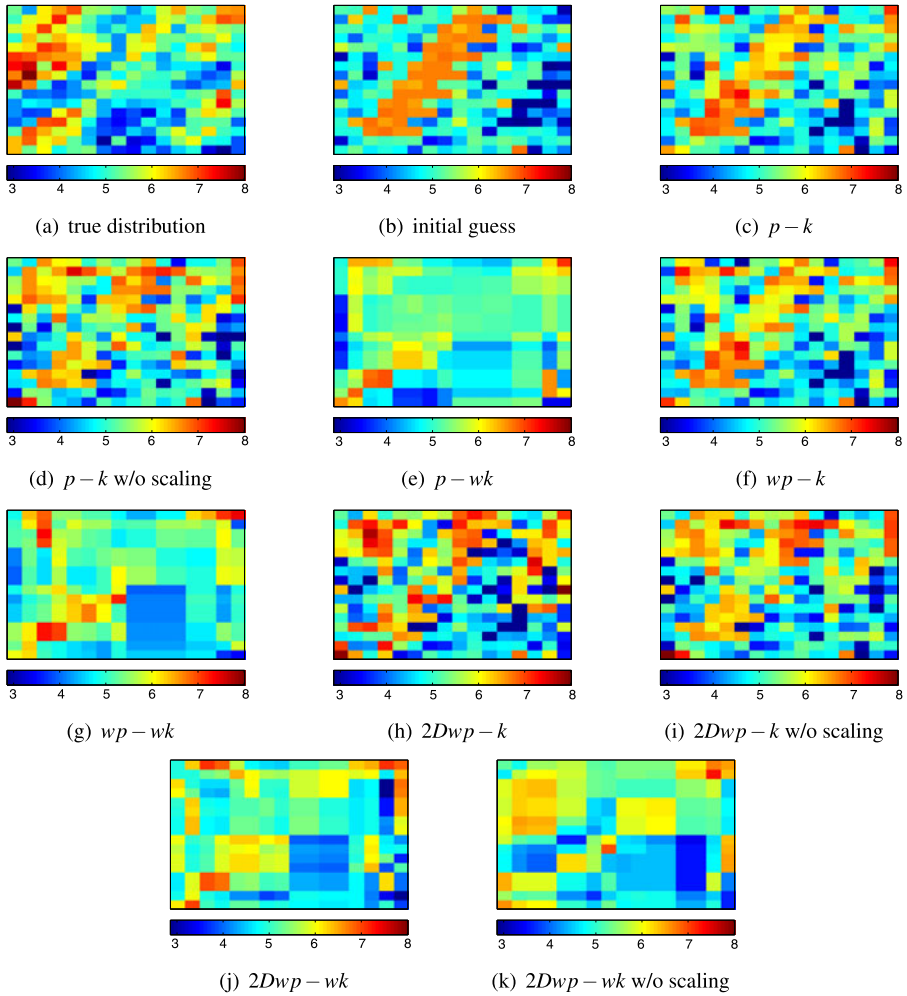


Fig. 5 log k distributions (16×16 reservoir model). (a) True distribution, (b) initial guess, (c) $p - k$, (d) $p - k$ w/o scaling, (e) $p - wk$, (f) $wp - k$, (g) $wp - wk$, (h) $2Dwp - k$, (i) $2Dwp - k$ w/o scaling, (j) $2Dwp - wk$, (k) $2Dwp - wk$ w/o scaling

permeability streaks running in a north-eastern direction. There were 16 injectors and 16 producers in the reservoir (Fig. 8(b)). Measurements of pressure in the injectors and pressure and water cut in the producers were taken concurrently so that there were 48 sets of data from all 32 wells. Each set of data contained 256 data points making up a total of 12 288 data points.

Table 4 shows that the number of observation data was reduced from 12 288 to 688 by the scaled version of $2Dwp - k$ and $2Dwp - wk$ and from 12 288 to 429 by the unscaled versions. These reductions translate to approximately 17.8 and 28.6 times saved in terms of computing sensitivity coefficients using the adjoint method. In fact, the original data matrix was reduced from 256×48 to 13×33 in the unscaled versions

Fig. 6 Decay of data mismatch (16×16 reservoir model)

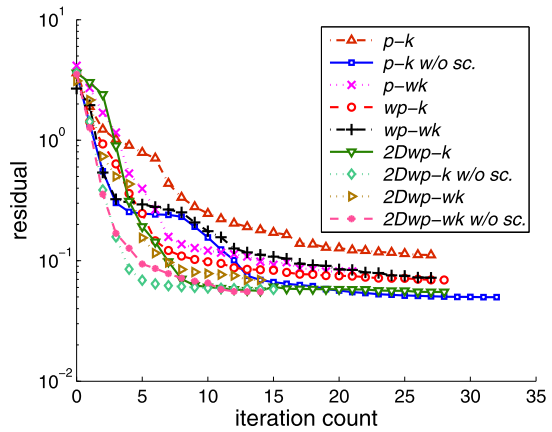
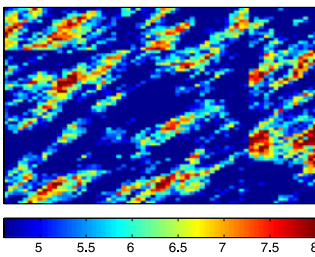
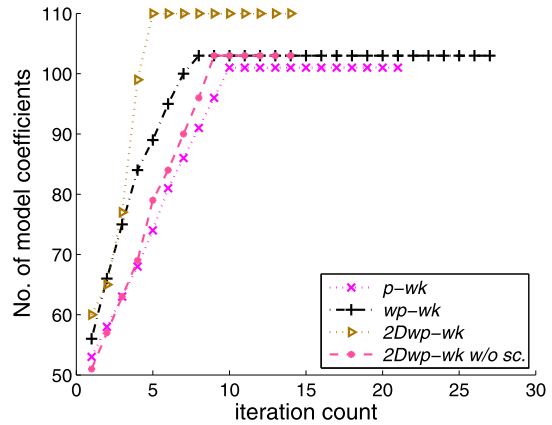
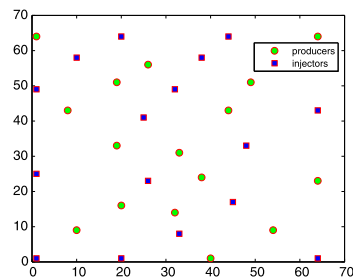


Fig. 7 Model space composition by $p-wk$, $wp-wk$ and $2Dwp-wk$ (16×16 reservoir model)



(a) log permeability distribution



(b) location of wells

Fig. 8 Log permeability distribution and location of wells in the 64×64 reservoir model. (a) Log permeability distribution, (b) location of wells

of $2Dwp-k$ and $2Dwp-wk$. The reduction achieved across datasets (from 48 to 33) is not obtainable in the $wp-wk$ approach thus showing the advantage of $2Dwp-wk$ over $wp-wk$ in integrating and decorrelating multiwell data. In the scaled versions,

Table 4 Model performance data (64×64 reservoir model)

| Models | $p - k$ | $(p - k)_{w/o sc}$ | $p - wk$ | $wp - k$ | $wp - wk$ | $2Dwp - k$ | $(2Dwp - k)_{w/o sc}$ | $2Dwp - wk$ | $(2Dwp - wk)_{w/o sc}$ |
|------------|---------|--------------------|----------|----------|-----------|------------|-----------------------|-------------|------------------------|
| N_{data} | 12 288 | 12 288 | 12 288 | 716 | 716 | 688 | 429 | 688 | 429 |
| CR | 1 | 1 | 1 | 15 | 15 | 17.8 | 28.6 | 17.8 | 28.6 |
| N_p | 4096 | 4096 | 816–1679 | 4096 | 822–1647 | 4096 | 4096 | 854–1645 | 791–1687 |
| k_{norm} | 249.1 | 232.2 | 249.2 | 258.6 | 259 | 294.1 | 234.4 | 267.9 | 233.2 |

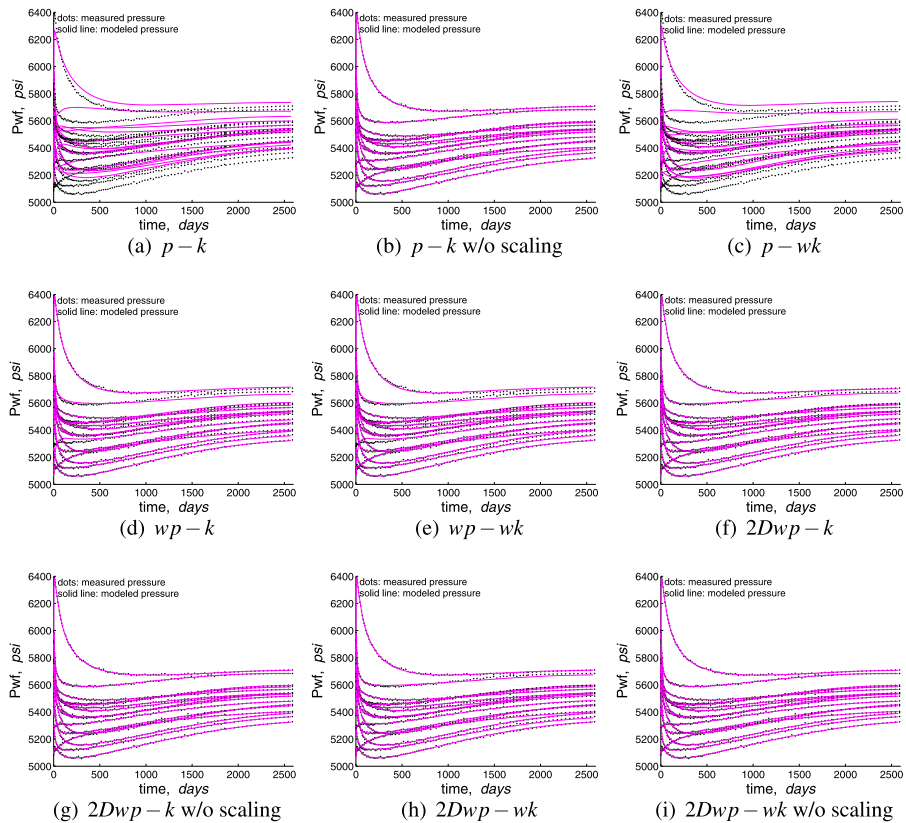


Fig. 9 Match to pressure and water cut *Column 1*: pressure at injectors; *Column 2*: pressure at producers; *Column 3*: water cut (64×64 reservoir model). (a) $p - k$, (b) $p - k$ w/o scaling, (c) $p - wk$, (d) $wp - k$, (e) $wp - wk$, (f) $2Dwp - k$, (g) $2Dwp - k$ w/o scaling, (h) $2Dwp - wk$, (i) $2Dwp - wk$ w/o scaling

the pressure datasets were reduced from 256×32 to 10×32 and the water cut datasets were reduced from 256×16 to 23×16 . In addition, reducing the observation space by a two-dimensional wavelet transformation makes the adjoint approach a cheaper method of sensitivity computation. Table 4 also shows the residual norm of \mathbf{k} given by the six approaches. The initial residual norm of \mathbf{k} for this example was 290. In this

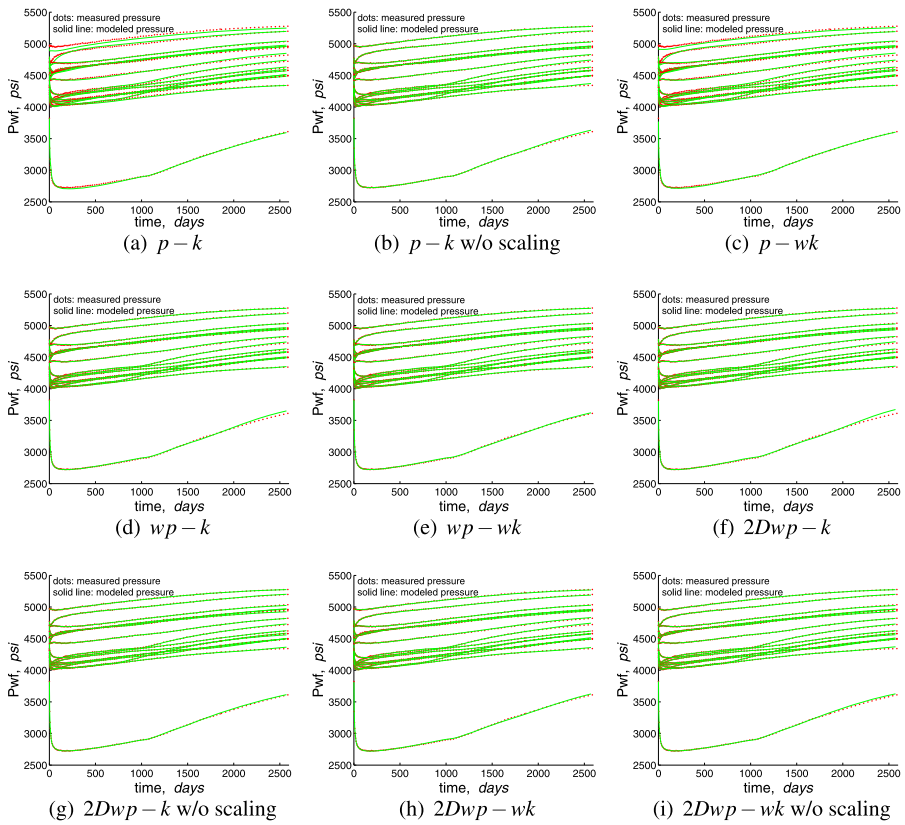


Fig. 10 Match to pressure at the producers (64×64 reservoir model). (a) $p - k$, (b) $p - k$ w/o scaling, (c) $p - wk$, (d) $wp - k$, (e) $wp - wk$, (f) $2Dwp - k$, (g) $2Dwp - k$ w/o scaling, (h) $2Dwp - wk$, (i) $2Dwp - wk$ w/o scaling

example, the approaches that do not decompose (and reduce) the model space gave lower residual norms of \mathbf{k} . The $2Dwp - k$ approaches gave the smallest norm.

The matches to the production histories ($t \leq 1852$ days) and predicted data ($t > 1852$ days) are presented in Figs. 9, 10, 11. The predicted data are produced from the true permeability distribution and compared with the predictions from the six approaches. The $p - k$ (scaled version) and the $p - wk$ approaches had difficulty matching the pressure history, particularly the injection pressure (Figs. 9(a), 9(c), 10(a) and 10(c)). The unscaled versions of the $p - k$, $2Dwp - k$ and $2Dwp - wk$ provided excellent matches to the pressure data but only acceptable matches to most of the water cut histories (Figs. 11(b), 11(g), and 11(i)). However, the predictions of future water cut profile from all the approaches are generally poor. It is important to note that the inability of the unscaled versions of the $2Dwp - k$ and $2Dwp - wk$ approaches to match all the water cut histories accurately indicate that the approaches do not satisfactorily integrate multiple datasets at different scales. However, the integration achieved is acceptable given the benefits derived from decorrelating the dataset and reducing the computational costs.

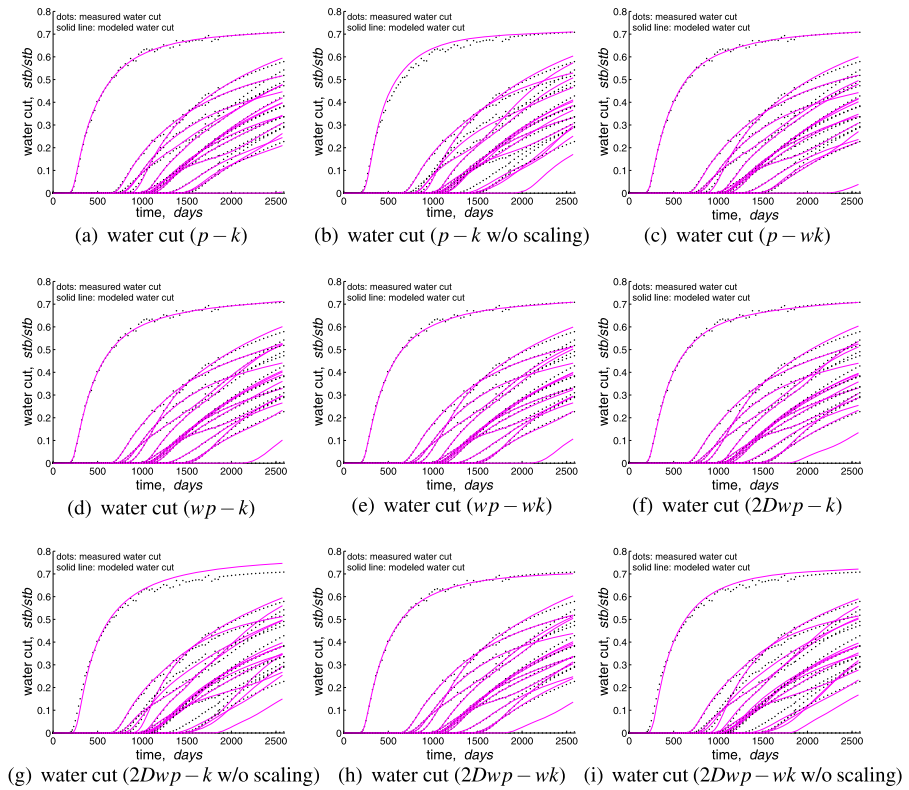


Fig. 11 Match to water cut at the producers (64×64 reservoir model)

Estimates of permeability distribution obtained from the six approaches are presented in Fig. 12. Although (as expected) none of the approaches was able to give an accurate estimate of the true log permeability distribution, the approaches showed varying ability to map out the medium-to-high permeability paths. Visual inspection shows that the unscaled versions of the $p - k$, $2Dwp - k$, and $2Dwp - wk$ approaches produced better estimates than the scaled versions. We also observe that the $2Dwp - k$ and $2Dwp - wk$ approaches (involving two-dimensional integration of the datasets) produced better estimates than their counterparts. The scaled version of $p - k$ and the $p - wk$ (also scaled) gave the worst estimates in this example. The results obtained here are in agreement with those obtained in Example 1, further indicating that (1) pressure data carry more significant information about the reservoir parameter distribution than water cut data, and (2) the two-dimensional wavelet transform is an efficient tool for integrating and decorrelating multiwell data. In Fig. 13, we observe that the $2Dwp - k$ and $2Dwp - wk$ approaches (scaled and unscaled) exhibited better convergence properties than the other approaches. Of all the approaches considered, the scaled $p - k$ and the $p - wk$ exhibited the worst convergence. Figure 14 shows the number of model parameters used at different iteration counts. We retained approximately 40 % of the total number of model coefficients for history match.

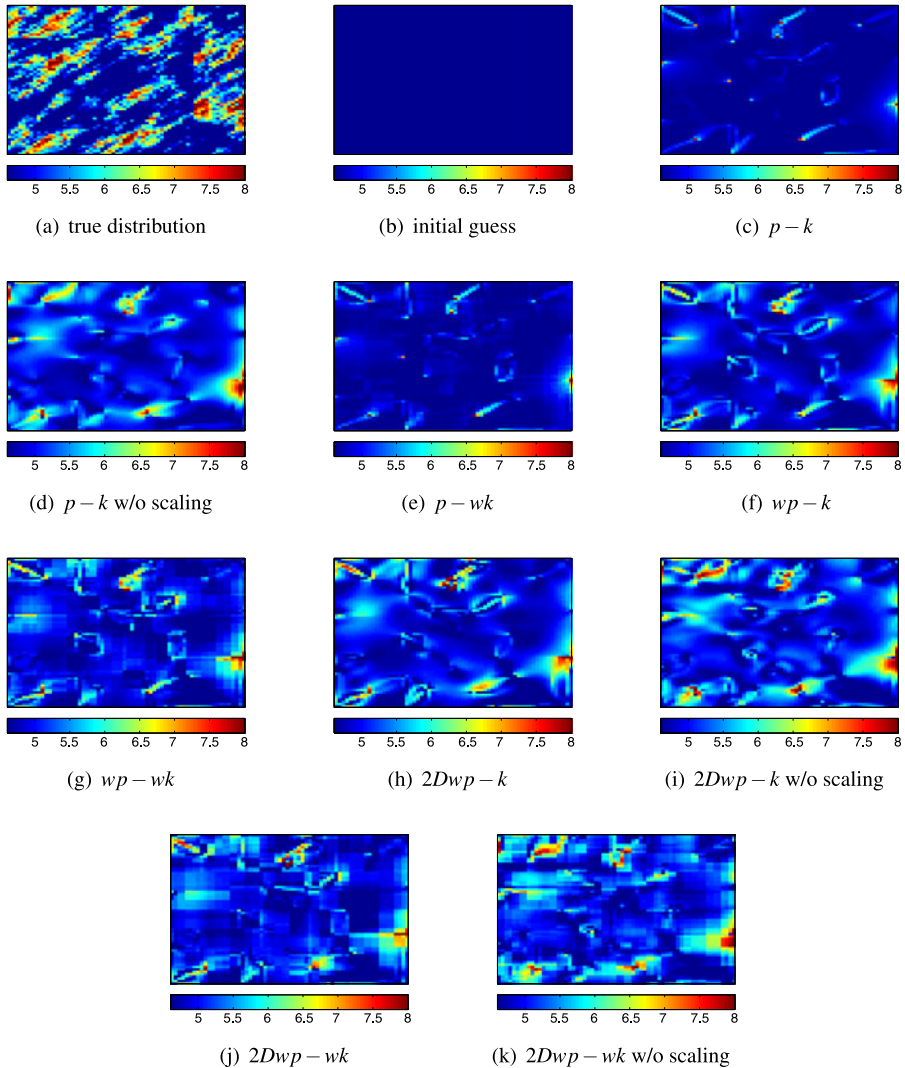


Fig. 12 $\log k$ distributions (64×64 reservoir model). (a) True distribution, (b) initial guess, (c) $p - k$, (d) $p - k$ w/o scaling, (e) $p - wk$, (f) $wp - k$, (g) $wp - wk$, (h) $2Dwp - k$, (i) $2Dwp - k$ w/o scaling, (j) $2Dwp - wk$, (k) $2Dwp - wk$ w/o scaling

5 Conclusions

A two-dimensional wavelet transform and reparameterization of the data space was used to integrate and decorrelate multiwell production data. The scaled and unscaled versions of the transformed datasets were considered and used for reconstructing the reservoir permeability field. Sample applications were used to study the effects of (a) integrating the datasets and (b) scaling the datasets on the ability to reconstruct the reservoir model. Results from sample applications show that (1) the approaches

Fig. 13 Decay of data mismatch (64×64 reservoir model)

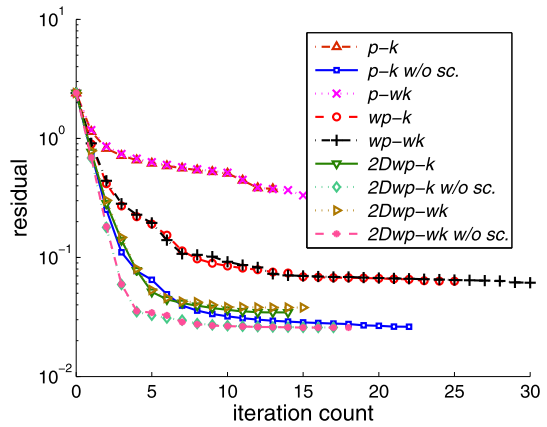
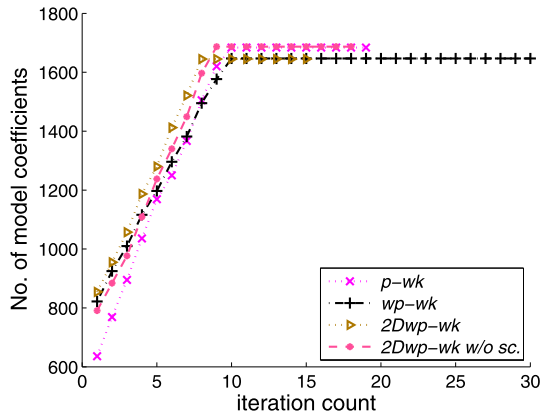


Fig. 14 Model space composition by $p-wk$, $wp-wk$ and $2Dwp-wk$ (64×64 reservoir model)



involving two-dimensional integration of the data space produced better estimates of reservoir parameter distribution than those involving one-dimensional or no integration of the data space, (2) the unscaled versions of the transformed data produced poorer but acceptable matches to the water cut data, and (3) the unscaled versions of the data ($p-k$, $2Dwp-k$ and $2Dwp-wk$) gave the lowest residual error in data but poor predictions of water cut. We conclude that the two-dimensional integration of multiwell production datasets is able to integrate and decorrelate the data thus removing redundancies in the dataset. Furthermore, a large volume of data does not necessarily lead to a better estimate of model parameters. The amount of information about the model carried by the data is more important than the volume of the data available for history match. Finally, observations show that the $2Dwp-wk$ approach gives a larger compression of data, and hence faster adjoint wavelet sensitivity computations than the $wp-wk$ approach.

Acknowledgements We acknowledge the support received from the Saudi Aramco Research Collaboration Fellowship and the Stanford University Petroleum Research Institute for Innovation in Well Testing (SUPRI-D) toward completing this work.

Appendix A: Relationship Between Regression in Real Space and Regression in Wavelet Space

In this appendix, we give two proofs. First we prove that the Frobenius norm remains unchanged when the residual matrix is transformed into an orthonormal wavelet domain. Furthermore, we show that when all observation-space wavelet coefficients and all model-space wavelet coefficients are included in the nonlinear regression estimation, the $2Dwp - wk$ approach is equivalent to the $p - k$ approach.

Lemma 1 Given two matrices $\tilde{X} = (\mathbf{x}_1 \ \mathbf{x}_2 \ \dots)$ and $\tilde{Y} = (\mathbf{y}_1 \ \mathbf{y}_2 \ \dots)$, of equal dimensions, and having orthonormal two-dimensional wavelet transforms $\mathbb{W}_{2D}(\tilde{X}) = \tilde{S} = (\mathbf{s}_1 \ \mathbf{s}_2 \ \dots)$ and $\mathbb{W}_{2D}(\tilde{Y}) = \tilde{R} = (\mathbf{r}_1 \ \mathbf{r}_2 \ \dots)$, respectively. The Frobenius norm of the difference between \tilde{X} and \tilde{Y} , is exactly the same as the Frobenius norm between \tilde{S} and \tilde{R} . That is,

$$\|\tilde{X} - \tilde{Y}\|_F^2 = \|\tilde{S} - \tilde{R}\|_F^2. \tag{27}$$

Proof Given that

$$\tilde{S} = W_t \tilde{X} W_{ds}^T, \tag{28}$$

and

$$\tilde{R} = W_t \tilde{Y} W_{ds}^T, \tag{29}$$

then the Frobenius norm of the difference between \tilde{S} and \tilde{R} can be expressed as

$$\begin{aligned} \|\tilde{S} - \tilde{R}\|_F^2 &= \|W_t \tilde{X} W_{ds}^T - W_t \tilde{Y} W_{ds}^T\|_F^2 \\ &= \|W_t (\tilde{X} - \tilde{Y}) W_{ds}^T\|_F^2 \\ &= \text{trace}\{[W_t (\tilde{X} - \tilde{Y}) W_{ds}^T]^T [W_t (\tilde{X} - \tilde{Y}) W_{ds}^T]\} \\ &= \text{trace}\{W_{ds} (\tilde{X} - \tilde{Y})^T (\tilde{X} - \tilde{Y}) W_{ds}^T\}. \end{aligned} \tag{30}$$

Because the *trace function* is invariant under cyclic permutations, we can rearrange the matrices inside the *trace* to have

$$\begin{aligned} \|\tilde{S} - \tilde{R}\|_F^2 &= \text{trace}\{(\tilde{X} - \tilde{Y})^T (\tilde{X} - \tilde{Y}) W_{ds}^T W_{ds}\} \\ &= \text{trace}\{(\tilde{X} - \tilde{Y})^T (\tilde{X} - \tilde{Y})\} \\ &= \|\tilde{X} - \tilde{Y}\|_F^2. \end{aligned} \tag{31}$$

□

Lemma 2 If all observation-space wavelets and all model-space wavelets are used in a minimization involving the $2Dwp - wk$ approach, then regression performed in the wavelet domain will follow exactly the same path and yield the same estimates as regression performed in the real space.

Proof Because all observation-space wavelets and all model-space wavelets are used, $S_{\mathbf{c}_\alpha}^{\tilde{D}^{w_{cal}}}$, $H_{\mathbf{c}_\alpha}^{\tilde{D}^{w_{cal}}}$, and $\mathbf{g}_{\mathbf{c}_\alpha}^{\tilde{D}^{w_{cal}}}$ are of full dimensions. By substituting the definitions of $H_{\mathbf{c}_\alpha}^{\tilde{D}^{w_{cal}}}$ and $\mathbf{g}_{\mathbf{c}_\alpha}^{\tilde{D}^{w_{cal}}}$ into Eq. (22) and omitting the iteration index, we obtain

$$[(S_{\mathbf{c}_\alpha}^{\tilde{D}^{w_{cal}}})^T S_{\mathbf{c}_\alpha}^{\tilde{D}^{w_{cal}}} + \lambda I] \delta \mathbf{c}_\alpha = -(S_{\mathbf{c}_\alpha}^{\tilde{D}^{w_{cal}}})^T \text{vec}(\tilde{D}_{w_{cal}} - \tilde{D}_{w_{meas}}). \tag{32}$$

Replacing $S_{\mathbf{c}_\alpha}^{\tilde{D}^{w_{cal}}}$ by $W_t S_{\mathbf{c}_\alpha}^{\tilde{D}^{cal}} W_{ds}$ and \tilde{D}_w by $W_t \tilde{D} W_{ds}^T$ in Eq. (32) and simplifying leads to

$$[S_{\mathbf{c}_\alpha}^{\tilde{D}^{cal}T} S_{\mathbf{c}_\alpha}^{\tilde{D}^{cal}} + \lambda I] \delta \mathbf{c}_\alpha = -(S_{\mathbf{c}_\alpha}^{\tilde{D}^{cal}})^T \text{vec}(\tilde{D}_{cal} - \tilde{D}_{meas}) \tag{33}$$

Recognizing that $\mathbf{c}_\alpha = W_s \alpha$ we may write

$$[(S_{\mathbf{c}_\alpha}^{\tilde{D}^{cal}} W_s^T)^T (S_{\mathbf{c}_\alpha}^{\tilde{D}^{cal}} W_s^T) + \lambda I] W_s \delta \alpha = -(S_{\mathbf{c}_\alpha}^{\tilde{D}^{cal}} W_s^T)^T \text{vec}(\tilde{D}_{cal} - \tilde{D}_{meas}). \tag{34}$$

Further simplification yields

$$[(S_{\alpha}^{\tilde{D}^{cal}})^T S_{\alpha}^{\tilde{D}^{cal}} + \lambda I] \delta \alpha = -(S_{\alpha}^{\tilde{D}^{cal}})^T \text{vec}(\tilde{D}_{cal} - \tilde{D}_{meas}). \tag{35}$$

Equation (35) may be written as

$$H \delta \alpha = -\mathbf{g}, \tag{36}$$

where H and \mathbf{g} are the Hessian and gradient of the objective function in the $p - k$ approach. This proves that the $2Dwp - wk$ approach is equivalent to the $p - k$ approach when all wavelets of the data and model spaces are retained for regression. \square

Appendix B: Adjoint-State Method for Two-Dimensional Data Integration

In this appendix, we discuss the method of directly computing the adjoint sensitivities of the two-dimensional wavelets of measured data to model parameters. Consider any vector-valued function $\xi(\alpha)$, which depends on state variables $\mathbf{u}^n(\alpha)$ and is thus represented by

$$\xi(\alpha) = \eta(\mathbf{u}^1, \mathbf{u}^2, \dots, \mathbf{u}^N, \alpha), \tag{37}$$

in which $\eta(\mathbf{u}^1, \mathbf{u}^2, \dots, \mathbf{u}^N, \alpha)$ is, in general, any vector-valued function. However, for the purpose of sensitivity computation it represents the vector of computed data. These computed data correspond to only times at which measurements are made. We may then form an augmented functional ξ_a by adjoining the constraints \mathbf{f} to η using adjoint variables $\tilde{\lambda} \in U^*$. Thus,

$$\xi_a(\mathbf{u}^{n+1}, \tilde{\lambda}^{n+1}, \alpha) = \eta + \sum_{n=0}^N [(\tilde{\lambda}^{n+1})^T \mathbf{f}^{n+1}]. \tag{38}$$

In Eq. (38), η , of length N_L , is the vector of data for which sensitivities are to be computed and $\tilde{\lambda}^{n+1}$ is the matrix of adjoint variables at time-step $n + 1$. $\tilde{\lambda}^{n+1}$ is of

dimension $N_u \times N_L$. In vectorial notation η can be written as

$$\eta = \text{vec}(\tilde{D}_{cal}) = [\mathbf{d}_{1,cal}^T \mathbf{d}_{2,cal}^T \cdots \mathbf{d}_{L,cal}^T]^T, \tag{39}$$

where the function vec reshapes a matrix into a column vector containing all the elements of that matrix. By taking a wavelet transform of Eq. (38) we obtain

$$\xi_{w_a}(\mathbf{u}^{n+1}, \tilde{\lambda}_w^{n+1}, \alpha) = \eta_w + \sum_{n=0}^N [(\tilde{\lambda}_w^{n+1})^T \mathbf{f}^{n+1}]. \tag{40}$$

In Eq. (40), we have used the definitions

$$\xi_{w_a} = \text{vec}(W_i \tilde{\xi}_a W_{ds}^T), \tag{41}$$

$$\eta_w = \text{vec}(W_i \tilde{D}_{cal} W_{ds}^T), \tag{42}$$

and $\tilde{\lambda}_w$ is computed from the adjoint system of equations. Using the following definitions,

$$J^n = \frac{\partial \mathbf{f}^n}{\partial \mathbf{u}^n}, \tag{43}$$

$$D^{n+1} = \frac{\partial \mathbf{f}^{n+1}}{\partial \mathbf{u}^n}, \tag{44}$$

$$Y^{n+1} = \frac{\partial \mathbf{f}^{n+1}}{\partial \alpha}, \tag{45}$$

and following the same approach presented in Awotunde and Horne (2012), we obtain

$$(J^n)^T \tilde{\lambda}_w^n = - \left[(D^{n+1})^T \tilde{\lambda}_w^{n+1} + \left(\frac{\partial \eta_w}{\partial \mathbf{u}^n} \right)^T \right], \tag{46}$$

and

$$\frac{\partial \xi_{w_a}}{\partial \alpha} = \frac{\partial \eta_w}{\partial \alpha} + \sum_{n=1}^N [(\tilde{\lambda}_w^n)^T Y^n]. \tag{47}$$

Equation (46) is the adjoint equation through which all the adjoint variables $\tilde{\lambda}_w^n$ are evaluated. J^n is the Jacobian matrix, D^{n+1} is the is a block-diagonal matrix containing the derivatives of the accumulation terms at current time-step with respect to the state variables at the previous time step and Y^{n+1} is a matrix containing the derivatives of the residual vector with respect to model parameters. All the three matrices are obtained from the simulator at the last step of the Newton–Raphson iteration at every time-step n . The Jacobian matrix and the matrix of derivatives of the accumulation terms are constant for all parameters. Thus, we need to compute them only once at every time step and, therefore, there is no computational overhead. To compute the reduced wavelet adjoint gradient, we follow the procedure presented in Awotunde and Horne (2012), but replace the one-dimensional wavelet decomposition with a two-dimensional wavelet transform of the two-dimensional dataset and then reshapes the matrix into a vector using vec .

References

- Al-Harbi M, Cheng H, He Z, Datta-Gupta A (2005) Streamline-based production data integration in naturally fractured reservoirs. *SPE Reserv Eval Eng* 10:426–439
- Awotunde AA, Horne RN (2011a) A multiresolution analysis of the relationship between spatial distribution of reservoir parameters and time distribution of well-test data. *SPE Reserv Eval Eng* 14:345–356
- Awotunde AA, Horne RN (2011b) A wavelet approach to adjoint state sensitivity computation for steady-state differential equations. *Water Resour Res* 47:1–21
- Awotunde AA, Horne RN (2012) An improved adjoint-sensitivity computations for multiphase flow using wavelets. *SPE J* 17:402–417
- Aziz K, Settari A (1979) *Petroleum reservoir simulation*. Applied Science Publishers, London
- Caers J (2003) Geostatistical history matching under training-image based geological constraints. *SPE J* 8:218–226
- Carter RD, Kemp LF, Piece AC, Williams DL (1974) Performance matching with constraints. *SPE J* 14:187–196, trans., AIME 257
- Chavent GM, Dupuy M, Lemonnier P (1975) History matching by use of optimal theory. *SPE J* 15:74–86
- Chu L, Schatzinger RA, Tham MK (1998) Application of wavelet analysis to upscaling of rock properties. *SPE Reserv Eval Eng* 1:75–81
- Chui CK (1992) *An introduction to wavelets*. Academic Press, San Diego.
- Gill PE, Murray W, Wright MH (1981) *Practical optimization*. Academic Press, New York
- Golub GH, Van Loan CF (1996) *Matrix computations*. John Hopkins University Press, Baltimore
- Griva I, Nash SG, Sofer A (1996) *Linear and nonlinear optimization*. SIAM Press, Philadelphia
- He Z, Yoon S, Datta-Gupta A (2002) Streamline-based production data integration with gravity and changing field conditions. *SPE J* 7:423–436
- Hu LY (2000) Gradual deformation and iterative calibration of Gaussian-related stochastic models. *Math Geol* 32:87–108
- Hu LY (2002) Combination of dependent realizations within the gradual deformation method. *Math Geol* 34:953–963
- Huang X, Kelkar M (1996) Integration of dynamic data for reservoir characterization in the frequency domain. Paper SPE 115795 presented at the SPE annual technical conference and exhibition, Denver, Colorado, USA, 6–9 October 1996
- Jacquard P, Jain C (1965) Permeability distribution from field pressure data. *SPE J* 5:281–294, trans., AIME 234
- Johansen K, Caers J, Suzuki S (2007) Hybridization of the probability perturbation method with gradient information. *Comput Geosci* 11:319–331
- Journel AJ (1974) Geostatistics for conditional simulation of ore bodies. *Econ Geol* 69:673–687
- Levenberg K (1944) A method for the solution of certain non-linear problems in least squares. *Q Appl Math* 2:164–168
- Liu M, Liu YZ, Guo T, Cai L (2006) Interwell heterogeneity analysis by dynamic correlation method. Paper SPE 104399 presented at the SPE international oil and gas conference and exhibition, Beijing, China, 5–7 December 2006
- Lu P, Horne RN (2000) A multiresolution approach to reservoir parameter estimation using wavelet analysis. Paper SPE 62985 presented at the SPE annual technical conference and exhibition, Dallas, 1–4 October 2000
- Mallat S (1989) A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans Pattern Anal Mach Intell* 11:674–693
- Marquardt D (1963) An algorithm for least-squares estimation of nonlinear parameters. *SIAM J Appl Math* 11:431–441
- Nakashima T, Durlafsky L (2010) Accurate representation of near-well effects in coarse-scale models of primary oil production. *Transp Porous Media* 83:741–770
- Nocedal J, Wright SJ (2006) *Numerical optimization*, 2nd edn. Springer, New York
- Panda MN, Mosher CC, Chopra AK (2001) Reservoir modeling using scale-dependent data. *SPE J* 6:157–170
- Peaceman DW (1977) *Fundamentals of numerical reservoir simulation*. Elsevier, New York
- Percival D, Walden AT (2000) *Wavelet methods for time series analysis*. Cambridge University Press, Cambridge, England
- Sahni I, Horne RN (2005) Multiresolution wavelet analysis for improved reservoir description. *SPE Reserv Eval Eng* 8:53–69

- Sahni I, Horne RN (2006) Generating multiple history-matched realizations using wavelets. *SPE Reserv Eval Eng* 9:217–226
- Strebelle S, Payrazyan K, Caers J (2003) Modeling of a deepwater turbidite reservoir conditional to seismic data using principal component analysis and multi-point geostatistics. *SPE J* 8:227–235
- Yoon S, Malallah AH, Datta-Gupta A, Vasco DW, Behrens RA (2001) A multiscale approach to production data integration using streamline models. *SPE J* 6:182–192