DELFT UNIVERSITY OF TECHNOLOGY

REPORT 08-18

Variational estimation of permeability and model errors from 3D and 4D seismic data using model-driven regularization

J.R. Rommelse, J.D. Jansen and A.W. Heemink

ISSN 1389-6520

Reports of the Department of Applied Mathematical Analysis

Delft 2008

Variational estimation of permeability and model errors from 3D and 4D seismic data using model-driven regularization

J.R.Rommelse * J.D. Jansen † A.W. Heemink ‡

August 3, 2008

Abstract

Automated history matching methods, or data assimilation algorithms, can be used to support decision-making tools in closed-loop reservoir management. In variational data assimilation, the discrepancy between observed measurements and their model predicted antitheses is minimized with respect to parameters that underlie the reservoir model. Assuming that there are no uncertainties in addition to the unknown parameters, methods exist that can efficiently calculate the gradient of the discrepancy to changes in the parameters. Usually many different parameter sets exist that locally minimize the discrepancy, so the gradient must be regularized before it can be used by gradient-based minimization algorithms. In the presence of model errors, more advanced methods must be employed, like the Representer Method (RM).

This article proposes a variational method to estimate the permeability of a reservoir rock from static and time-lapse seismic data, that simultaneously estimates the model errors in the reservoir simulator. Unlike in the (classic or modified) RM, the regularization and the computation of the gradient are decoupled, which can save a lot of computation time. An analytical gradient is available, whereas the RM uses an approximated one.

First the Variational Parameter Estimation method with Model-driven Regularization (VPERM) is introduced and the relation to the RM is explained. Then experiments are shown, using three different data sets. The first data set contains synthetic pressure measurements in the injection and production wells, the second and third data sets contain P-wave impedance data in addition to pressure data. The second data set contains the baseline survey and the monitor, the third one contains the baseline and the difference.

*Delft University of Technology & Alten Nederland, E-mail: J.R.Rommelse@tudelft.nl

†Delft University of Technology & Shell International Exploration & Production, E-mail: J.D.Jansen@TUDelft.nl

‡Delft University of Technology, E-mail: A.W.Heemink@TUDelft.nl

VPERM produces results that are similar to the results of the RM, but in less computation time. Both methods produce good results when only pressure data obtained from the wells are available. Sometimes seismic data removes outliers from the "history-future" crossplot; better predictions are obtained from models that are not necessarily better history matched. In other cases, seismic data gives better history-matched models and better predictions.

1 Introduction

This article investigates the applicability of a weak constraint variational method to estimate reservoir permeability from 3D or 4D seismic data with a reservoir simulator that contains model errors. Sections 1.1 and 1.2 introduce the variational parameter estimation and weak constraints approaches. [Skjervheim et al., 2005] used the Ensemble Kalman Filter (EnKF) [Evensen, 2003] to estimate permeability from seismic data in the presence of model errors. In order to do so, a non-linear rock physics model was augmented to the non-linear reservoir model to predict changes in seismic response from changes in fluid pressure and mixture. A weak constraint variational method is promising, since [Rommelse et al., 2008] showed that for a linear measurement operator the Representer Method (RM) [Bennett and McIntosh, 1982; Eknes and Evensen, 1997; Bennett, 2002; Valstar et al., 2004; Baird and Dawson, 2005; Janssen et al., 2006; Przybysz et al., 2007] outperformed the EnKF in terms of two quality measures; root mean square difference between estimated and synthetic true permeability, and the ability to predict water breakthrough in production wells.

1.1 Variational parameter estimation

Data assimilation methods aim to improve numerical models by comparing actual measurements of a physical system with the numerical model predictions of these measurements. As the parameters of the numerical model are changed, so do the predicted state variables and the predicted measurements. The discrepancy between the "measured measurements" and the "predicted measurements" can be used to update the parameters in order to decrease this discrepancy. In variational methods, the discrepancy is formulated as an objective function that has to be minimized, often using the Euclidean norm. Usually, the dynamic model that relates parameters to state variables, is adjoined to the objective function with the aid of Lagrange multipliers [Bennett, 2006]. In reservoir management, the dynamic model is implemented in the form of a reservoir simulator, the parameters of interest can for example be porosity or permeability, and the state of the system is typically described by pressure and saturation values in all grid blocks. The first or higher order variations are iteratively used to improve the parameter estimate. Gradients can be approximated or efficiently calculated by an adjoint reservoir simulator [Rommelse et al., 2007]. Usually reservoir simulators use time-implicit numerical schemes and first order derivatives, Jacobians, are used to speed up the non-linear solvers. Even when the Jacobians are available

in a reservoir simulator, it is quite a task to construct the adjoint simulator. Higher order derivatives are usually not available. In the LBFGS [Gao and Reynolds, 2006; Ulbrich, 2002] minimization algorithm used in this article, second order derivatives are approximated by monitoring the first order derivatives over successive iterations and used in a Gauss-Newton scheme.

1.2 Model errors; strong and weak constraints

Often, the discrepancy between measured measurements and predicted measurements is formulated using the Euclidean norm. The objective of the data assimilation is then to minimize the square of this norm with respect to the model parameters while the numerical model is used as a (strong) constraint.

However, there is an additional phenomenon that may cause the discrepancy; the model is an approximation, so even if the parameters were known, the model might still produce incorrect output. These errors can be modelled as extra parameters, which are also added to the objective function using the 2-norm. The model is then used as a weak constraint in the minimization problem.

1.3 Regularization

Variational methods often get trapped in local minima. Regularization methods must be employed to reduce the number of local minima. When model errors are taken into account, these can be modelled as additional model parameters to turn the weak constraint minimization problem into a strong constraint problem of much higher order. Regularization then becomes increasingly important.

The RM [Bennett and McIntosh, 1982; Bennett, 2002; Valstar et al., 2004] turns the weak constraint problem into a strong constraint problem of the same order as the number of measurements. The regularization is part of the minimization scheme and the basis functions are recalculated at every iteration. [Rommelse et al., 2007] proposes an approximated gradient and an additional reduction of the dimension of the parameter space.

This article proposes a Variational Parameter Estimation algorithm that uses Model-driven Regularization (VPERM, section 2), similar to the RM. The regularization is decoupled from the minimization scheme. Basis functions can be updated at any iteration, or not. The weak constraint problem is reduced to a strong constraint problem of an order that is much lower than the number of measurements. Assimilating seismic data would not have been feasible otherwise. The gradient is analytical if the basis functions are kept fixed, unlike the approximated gradient of the RM.

2 The VPERM method

This section introduces the VPERM method; Variational Parameter Estimation Regularized by Model-driven basis functions.

2.1 High dimensional gradient of the weak constraint problem

The state variables, pressures and saturations in all grid blocks, at time t_i are denoted by $\mathbf{x}_i, i \in \{0, \cdots, n\}$. Running the reservoir simulator and predicting measurements is denoted by

$$\mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \theta, \varepsilon_i\right) = \mathbf{0} \quad , \quad \mathbf{y} = \mathbf{h}\left(\mathbf{x}_{\{0,\cdots,n\}}\right), \tag{1}$$

where the model parameters are collected in the vector θ and the model errors on interval $[t_{i-1}, t_i]$ are contained in the vector ε_i. The initial states may be part of the parameter estimation process, so $\mathbf{x}_0 = \mathbf{x}_0(\theta)$. The minimization relies on the availability of the first (mean) and second (covariance) order statistics of the model parameters and model errors. These are denoted by θ^{prior}, $\mathbf{P}_\theta$, $\varepsilon_i^{prior} = \mathbf{0}$ and $\mathbf{P}_{\varepsilon_i}$.

The objective function that has to be minimized is

$$\begin{aligned} J = & \frac{1}{2}\left(\mathbf{h}\left(\mathbf{x}_{\{0,\cdots,n\}}\right) - \mathbf{m}\right)^T \mathbf{P}_\mathbf{y}^{-1}\left(\mathbf{h}\left(\mathbf{x}_{\{0,\cdots,n\}}\right) - \mathbf{m}\right) + \\ & + \frac{1}{2}\left(\theta - \theta^{prior}\right)^T \mathbf{P}_\theta^{-1}\left(\theta - \theta^{prior}\right) + \frac{1}{2}\sum_{i=1}^{n} \varepsilon_i^T \mathbf{P}_{\varepsilon_i}^{-1} \varepsilon_i + \\ & + \sum_{i=1}^{n} \lambda_i^T \mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \theta, \varepsilon_i\right), \end{aligned} \tag{2}$$

where $\mathbf{m}$ contains the actual physical measurements, possibly taken at different times, and $\mathbf{h}$ is the measurement operator that operates on all state variables at all time steps. $\mathbf{P}_\mathbf{y}$ represents the uncertainty in the measurements in the form of an error covariance matrix. The last term of Eq. (2) represents the system equations $\mathbf{g}$ that have been adjoined to the objective function with the aid of Lagrange multipliers λ_i in the usual fashion [Bennett, 2006].

The derivatives of Eq. (2) with respect to λ_i, $\mathbf{x}_i$, θ and ε_i are

$$\left(\frac{\partial J}{\partial \lambda_i}\right)^T = \mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \theta, \varepsilon_i\right), \tag{3}$$

$$\begin{aligned} \left(\frac{\partial J}{\partial \mathbf{x}_i}\right)^T = & \left(\frac{\partial \mathbf{h}\left(\mathbf{x}_{\{0,\cdots,n\}}\right)}{\partial \mathbf{x}_i}\right)^T \mathbf{P}_\mathbf{y}^{-1}\left(\mathbf{h}\left(\mathbf{x}_{\{0,\cdots,n\}}\right) - \mathbf{m}\right) + \\ & + \left(\frac{\partial \mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \theta, \varepsilon_i\right)}{\partial \mathbf{x}_i}\right)^T \lambda_i + \left(\frac{\partial \mathbf{g}\left(\mathbf{x}_{i+1}, \mathbf{x}_i, \theta, \varepsilon_{i+1}\right)}{\partial \mathbf{x}_i}\right)^T \lambda_{i+1}, \end{aligned} \tag{4}$$

$$\left(\frac{\partial J}{\partial \theta}\right)^T = \mathbf{P}_\theta^{-1}\left(\theta - \theta^{prior}\right) + \sum_{i=1}^{n}\left(\frac{\partial \mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \theta, \varepsilon_i\right)}{\partial \theta}\right)^T \lambda_i, \tag{5}$$

$$\left(\frac{\partial J}{\partial \varepsilon_i}\right)^T = \mathbf{P}_{\varepsilon_i}^{-1}\varepsilon_i + \left(\frac{\partial \mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \theta, \varepsilon_i\right)}{\partial \varepsilon_i}\right)^T \lambda_i. \tag{6}$$

Setting the first two derivatives equal to zero gives the reservoir simulator, Eq. (1) and the adjoint reservoir simulator

$$\left(\frac{\partial \mathbf{h}\left(\mathbf{x}_{\{0,\cdots,n\}}\right)}{\partial \mathbf{x}_i}\right)^T \mathbf{P}_{\mathbf{y}}^{-1}\left(\mathbf{m}-\mathbf{h}\left(\mathbf{x}_{\{0,\cdots,n\}}\right)\right) \tag{7}$$
$$= \left(\frac{\partial \mathbf{g}\left(\mathbf{x}_i,\mathbf{x}_{i-1},\theta,\mathbf{0}\right)}{\partial \mathbf{x}_i}\right)^T \lambda_i + \left(\frac{\partial \mathbf{g}\left(\mathbf{x}_{i+1},\mathbf{x}_i,\theta,\mathbf{0}\right)}{\partial \mathbf{x}_i}\right)^T \lambda_{i+1}.$$

So, for the weak constraint case, the gradient $\left[\frac{\partial J}{\partial \theta} \quad \frac{\partial J}{\partial \varepsilon_1} \quad \cdots \quad \frac{\partial J}{\partial \varepsilon_n}\right]^T$ is calculated by running the simulator and the adjoint simulator, using the current estimates of the parameters and model errors as input, and substituting the results in Eq. (5) and Eq. (6). The gradient can be used in any gradient-based search algorithm to find a local minimum of the objective function.

2.2 Low dimensional gradient of a strong constraint problem

The low dimensional problem is parameterized by a vector $\mathbf{b}$. In this low dimensional space, there is no model and hence no model errors; the model is only defined in the high order space. From the low order parameters $\mathbf{b}$, the high order parameters θ are constructed by

$$\theta = \widetilde{\theta} + \mathbf{R}_\theta \mathbf{b}, \tag{8}$$

and the high order model errors are constructed by

$$\varepsilon_i = \widetilde{\varepsilon}_i + \mathbf{R}_{\varepsilon_i}\mathbf{b}. \tag{9}$$

The columns of the matrices $\mathbf{R}_\theta$ and $\mathbf{E}_i$ contain the parameter and model error basis functions. Whenever these basis functions are calculated, they are based on the results of the last iteration $\widetilde{\theta}$ and $\widetilde{\varepsilon}_i$ and the low order parameters are reset to $\mathbf{b}=\mathbf{0}$. Before the first iteration, $\widetilde{\theta}$ is initialized to $\widetilde{\theta} = \theta^{prior}$ and $\widetilde{\varepsilon}_i$ is initialized to $\widetilde{\varepsilon}_i = \mathbf{0}$. The RM also introduces state representers $\mathbf{R}_{\mathbf{x}_i}$ as

$$\mathbf{x}_i = \widetilde{\mathbf{x}}_i + \mathbf{R}_{\mathbf{x}_i}\mathbf{b}, \tag{10}$$

and adjoint representers $\mathbf{R}_{\lambda_i}$ as

$$\lambda_i = \widetilde{\lambda}_i + \mathbf{R}_{\lambda_i}\mathbf{b}. \tag{11}$$

The RM does not update $\widetilde{\theta}$, $\widetilde{\varepsilon}_i$ and $\widetilde{\lambda}_i = \mathbf{0}$, meaning that the quality of the parameterization decreases as parameters move further away from the prior. $\widetilde{\mathbf{x}}_i$ is formed from the results of the last iteration plus a correction term that is updated at every iteration. Effectively $\widetilde{\mathbf{x}}_i$ is never updated, and it is obtained by running the reservoir simulator on the prior parameters θ^{prior} without model errors. [Rommelse et al., 2007] also introduced measurement representers

$$\mathbf{h}\left(\mathbf{x}_{\{0,\cdots,n\}}\right) = \widetilde{\mathbf{h}} + \mathbf{R}_{\mathbf{y}}\mathbf{b}. \tag{12}$$

Using the basis functions, Eq. (8) and Eq. (9), the low order gradient $\left(\frac{\partial J}{\partial \mathbf{b}}\right)^T$ is calculated as

$$\left(\frac{\partial J}{\partial \mathbf{b}}\right)^T = \left(\frac{\partial \theta}{\partial \mathbf{b}}\right)^T \left(\frac{\partial J}{\partial \theta}\right)^T + \sum_{i=1}^{n} \left(\frac{\partial \varepsilon_i}{\partial \mathbf{b}}\right)^T \left(\frac{\partial J}{\partial \varepsilon_i}\right)^T = \tag{13}$$
$$= \mathbf{R}_\theta^T \left(\frac{\partial J}{\partial \theta}\right)^T + \sum_{i=1}^{n} \mathbf{R}_{\varepsilon_i}^T \left(\frac{\partial J}{\partial \varepsilon_i}\right)^T,$$

with $\left(\frac{\partial J}{\partial \theta}\right)^T$ and $\left(\frac{\partial J}{\partial \varepsilon_i}\right)^T$ from Eq. (5) and Eq. (6).

2.3 Choosing the basis functions

To find new basis functions $\mathbf{R}_\theta$ and $\mathbf{R}_{\varepsilon_i}$, equations Eq. (5) and Eq. (6) are set to zero and linearized around $\left(\widetilde{\theta}, \widetilde{\mathbf{x}}_i, \widetilde{\lambda}_i, \widetilde{\varepsilon}_i\right)$. The results are:

$$\mathbf{R}_\theta = -\mathbf{P}_\theta \sum_{i=1}^{n} \left(\frac{\partial \mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \theta, \varepsilon_i\right)}{\partial \theta}\right)^T \mathbf{R}_{\lambda_i} \tag{14}$$

and

$$\mathbf{R}_{\varepsilon_i} = -\mathbf{P}_{\varepsilon_i} \left(\frac{\partial \mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \theta, \varepsilon_i\right)}{\partial \varepsilon_i}\right)^T \mathbf{R}_{\lambda_i}. \tag{15}$$

Obviously, first basis functions $\mathbf{R}_{\lambda_i}$ for the adjoint variables must be chosen. Similar to Eq. (14) and Eq. (15), equations for the adjoint basis functions can be found

$$-\left(\frac{\partial \mathbf{h}\left(\mathbf{x}_{\{0,\cdots,n\}}\right)}{\partial \mathbf{x}_i}\right)^T \mathbf{P}_{\mathbf{y}}^{-1} \frac{\partial \mathbf{h}\left(\mathbf{x}_{\{0,\cdots,n\}}\right)}{\partial \mathbf{x}_i} \mathbf{R}_{\mathbf{x}_i} = \left(\frac{\partial \mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \theta, \varepsilon_i\right)}{\partial \mathbf{x}_i}\right)^T \mathbf{R}_{\lambda_i} + \tag{16}$$
$$+ \left(\frac{\partial \mathbf{g}\left(\mathbf{x}_{i+1}, \mathbf{x}_i, \theta, \varepsilon_{i+1}\right)}{\partial \mathbf{x}_i}\right)^T \mathbf{R}_{\lambda_{i+1}},$$

which depend on the state variable basis functions $\mathbf{R}_{\mathbf{x}_i}$, that can be calculated from

$$\frac{\partial \mathbf{g}}{\partial \mathbf{x}_i} \mathbf{R}_{\mathbf{x}_i} + \frac{\partial \mathbf{g}}{\partial \mathbf{x}_{i-1}} \mathbf{R}_{\mathbf{x}_{i-1}} + \frac{\partial \mathbf{g}}{\partial \theta} \mathbf{R}_\theta = \mathbf{O}. \tag{17}$$

In the RM, Eq. (16) and Eq. (17) can be solved sequentially, while here they are coupled through Eq. (14). If an estimate of the basis functions $\mathbf{R}_\theta$ for the parameters θ is available, for example from a previous iteration, then new basis functions can be calculated by sequentially solving Eq. (17), Eq. (16) and Eq. (14). This can be used in a Picard iteration scheme. Here a different approach is taken; $\mathbf{R}_{\mathbf{x}_i}$ in Eq. (16) is obtained by a

singular value decomposition:

$$\left(\frac{\partial \mathbf{h}\left(\mathbf{x}_{\{0,\cdots,n\}}\right)}{\partial \mathbf{x}_i}\right)^T \mathbf{P}_{\mathbf{y}}^{-1} \frac{\partial \mathbf{h}\left(\mathbf{x}_{\{0,\cdots,n\}}\right)}{\partial \mathbf{x}_i} \mathbf{R}_{\mathbf{x}_i} \tag{18}$$
$$= \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T\mathbf{X}_i = \mathbf{U}\boldsymbol{\Sigma}\begin{bmatrix}\mathbf{V}_1^T \\ \mathbf{V}_2^T\end{bmatrix}\mathbf{V}_1 = \mathbf{U}\boldsymbol{\Sigma}\begin{bmatrix}\mathbf{I} \\ \mathbf{O}\end{bmatrix}.$$

When the left-most columns of $\mathbf{U}\boldsymbol{\Sigma}$ are substituted in Eq. (16) to replace the left-hand-side, the resulting equation can be used to find approximations for $\mathbf{R}_{\lambda_i}$. Thereafter, these can be used in Eq. (14) and Eq. (15).

2.4 Quadratic line search

After the objective $J_0 = J(\theta)$ and the gradient $\mathbf{g}_0 = \frac{dJ}{d\theta}$ have been evaluated, a step must be taken in the direction of the gradient or a modified direction $\mathbf{d}$. If the step size is parameterized by s, then $\widetilde{J}(s) = J(\theta - s\mathbf{d})$ must be minimized. The gradient is not only used to find the direction of the line search, but it also helps in finding the step size, as is the case with Wolfe conditions [Nocedal and Wright, 1999]. However, here a different approach is taken:

1. Choose some potential step size s_p.
2. Evaluate the objective $\widetilde{J}(s_p) = J(\theta - s_p\mathbf{d})$.
3. If $\widetilde{J}(s_p) \leq \widetilde{J}(0) - \mathbf{g}_0^T\mathbf{d}$, then accept s_p and end line search.
4. Calculate a new potential step size
$$\widetilde{s}_p := \frac{1}{2} s_p \frac{\mathbf{g}_0^T\mathbf{d}s_p}{\mathbf{g}_0^T\mathbf{d}s_p + J_{s_p} - J_0}. \tag{19}$$
5. Goto 2 if $\widetilde{J}(\widetilde{s}_p) > \widetilde{J}(0)$.

Comments:

- Eq. (19) minimizes the parabola that is defined by $\widetilde{J}(0) = J_0$, $\widetilde{J}(s_p) = J_{s_p}$, and $\frac{d\widetilde{J}}{ds}(0) = -\mathbf{g}_0^T\mathbf{d}$.
- If $J_{s_p} = J_0$, then $\widetilde{s}_p := \frac{s_p}{2}$, unless $\|\mathbf{g}_0\| = 0$ (but if that were the case, then $J(\theta)$ is a local minimum of J).
- In step 2, the objective function is evaluated. In step 5 it is evaluated again. However, in numerical experiments [section 4], the step from 5 to 2 is never made.

3 The models

3.1 Weak constraint reservoir simulator

A 2-phase (water/oil) reservoir simulator can be represented symbolically as

$$\frac{d}{dt}\left(\mathbf{f}_1\left(\mathbf{x}\right)\right) = \mathbf{f}_2\left(\mathbf{x}, \theta\right), \tag{20}$$

where $\mathbf{x}$ contains the water saturation and water pressure (equal to oil pressure) for every grid block and θ contains the permeabilities of all grid blocks. $\mathbf{f}_1$ describes the presence of water and oil mass in the grid blocks and $\mathbf{f}_2$ models the flow through the grid block interfaces. Injection/production is modelled as sources/sinks, which are included in $\mathbf{f}_2$. Eq. (20) is formed by mass balance and Darcy equations [Aziz and Settari, 1979].

A fully implicit time discretization results in

$$\widetilde{\mathbf{g}}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \theta\right) = \mathbf{f}_1\left(\mathbf{x}_i\right) - \left(t_i - t_{i-1}\right)\mathbf{f}_2\left(\mathbf{x}_i, \theta\right) - \mathbf{f}_1\left(\mathbf{x}_{i-1}\right) = \mathbf{0}. \tag{21}$$

The model errors are introduced as additional sources/sinks in all grid blocks. In other words; after $\mathbf{x}_i$ has been solved from Eq. (21), the water and oil masses in the grid blocks have not correctly been predicted and must still be modified. The prediction gets worse as the time step $(t_i - t_{i-1})$ gets larger. Therefore the correction is modelled proportional to $(t_i - t_{i-1})$. If the additional sources become too strong, then unrealistically high pressures will be observed. If the additional sinks become too strong, then saturations outside $[0, 1]$ will occur. In this article, the additional sinks are non-linearly constrained by $\mathbf{f}_1$. The stochastic reservoir simulator has the form

$$\begin{aligned} &\mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \theta, \varepsilon_i\right) \\ &= \mathbf{f}_1\left(\mathbf{x}_i\right) - \left(t_i - t_{i-1}\right)\mathbf{f}_2\left(\mathbf{x}_i, \theta\right) - \mathbf{f}_1\left(\mathbf{x}_{i-1}\right) + \min\left\{\mathbf{f}_1\left(\mathbf{x}_i\right), \left(t_i - t_{i-1}\right)\varepsilon_i\right\} = \mathbf{0}. \end{aligned} \tag{22}$$

3.2 Rock-physics model

In addition to production data, seismic data can be used to estimate parameters in a reservoir simulation model, even though acoustic wave propagation is a totally different physical phenomenon than multiphase fluid flow. Seismic data can consist of travel time data and amplitude data, recorded by geophones or hydrophones. Travel time data is available as the difference between the time at which human-made sources initiate acoustic waves that travel into the subsurface and the time at which the reflected waves are recorded; usually at the surface, but sometimes inside wellbores. In "passive seismic", the travel time data is the difference between recordings of the same acoustic wave by different sets of geophones in different well bores. In this case "mother nature" is used instead of a human-made source.

Like reservoir engineers use reservoir simulators to predict the flow of fluids through the subsurface, geophysicists use wave propagation simulators to predict how acoustic waves travel through the subsurface. And like history matching in reservoir engineering,

in geophysics inverse methods are applied that bring the predictions of wave simulations closer to the actual recorded waves by changing the parameters of the wave simulator. The parameter sets of both simulators have some overlap. For example, porosity is an important parameter in both simulators. Theoretically, seismic data and production data could be merged into one big dataset and the parameter sets of both simulators could be merged to create one big inversion or parameter estimation problem. This has never been done; usually one inversion problem is solved partially, and the results are used as pseudo-data in the other inversion problem.

In this article, it is assumed that the seismic data has been inverted partially, until the level of seismic impedance. The P-wave impedance is augmented to the production data to help the history matching of the reservoir simulation model. Besides a reservoir simulator, also an impedance simulator is needed so seismic impedance can be predicted from fluid pressures and saturations. This dependence is shown in Fig (1) and described in the rest of this section.

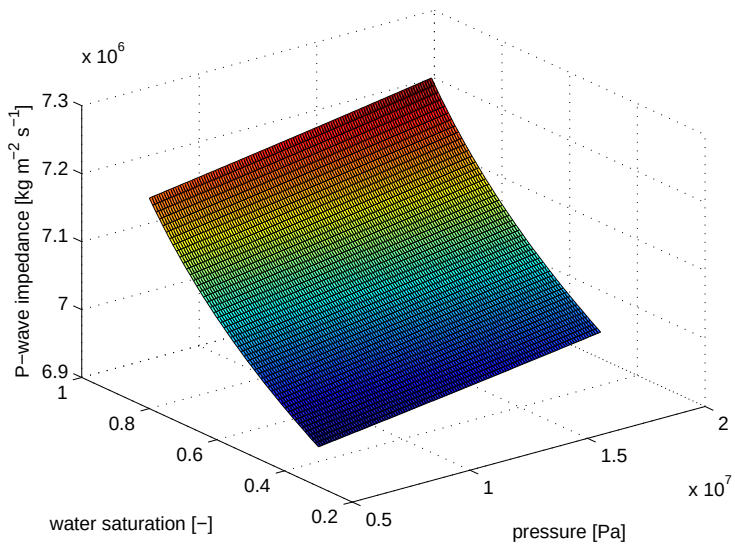

Figure 1: The dependence of the P-wave impedance on the fluid pressure and water saturation.

The P-wave impedance Z_p $[kg\ m^{-2}\ s^{-1}]$and S-wave impedance Z_s $[kg\ m^{-2}\ s^{-1}]$are defined as the velocities of the P and S-waves through the subsurface, V_p $[m\ s^{-1}]$and V_S $[m\ s^{-1}]$, multiplied by the density of the (fluid-filled reservoir) rock ρ_b $[kg\ m^{-3}]$, which is a weighted average of the densities of the sandstone, ρ_s $[kg\ m^{-3}]$, and the fluids

$$\rho_b = \phi\left(S_w\rho_w + S_o\rho_o\right) + \left(1-\phi\right)\rho_s. \tag{23}$$

Including the dependence of the velocities on the fluid filled bulk modulus K_{ff} $[Pa]$and shear modulus μ_m $[Pa]$, the impedances can be written as

$$Z_p = \sqrt{\left(K_{ff} + \frac{4}{3}\mu_m\right)\rho_b} \qquad , \qquad Z_s = \sqrt{\mu_m \rho_b}. \tag{24}$$

The fluid filled bulk modulus K_{ff} is calculated according to Gassmann [Mavko et al., 1998]

$$K_{ff} = K_s \left(\frac{1 - \frac{K_m}{K_s} + K_m \phi \left(\frac{1}{K_f} - \frac{1}{K_s} \right)}{1 - \frac{K_m}{K_s} + K_s \phi \left(\frac{1}{K_f} - \frac{1}{K_s} \right)} \right), \tag{25}$$

or with $\alpha = \phi \left(\frac{1}{K_f} - \frac{1}{K_s} \right)$ and $\beta = 1 - \frac{K_m}{K_s}$:

$$K_{ff} = K_s \left(\frac{K_m \alpha + \beta}{K_s \alpha + \beta} \right), \tag{26}$$

which combines the bulk moduli of the sandstone, K_s $[Pa]$, the dry-frame K_m $[Pa]$and the fluid K_f $[Pa]$. This last modulus is calculated from the bulk modulus of oil, K_o $[Pa]$and bulk modulus of water, K_w $[Pa]$, by

$$K_f = \frac{1}{\frac{S_w}{K_w} + \frac{S_o}{K_o}} = \frac{K_w K_o}{S_w K_o + S_o K_w}. \tag{27}$$

The elastic moduli, the porosity and the densities are assumed uncertain. However, deterministic values are used and noise is added to the synthetic P-wave impedance, rather than to the underlying parameters separately, section [3.3].

3.3 Measuring impedance with synthetic noise

After the true impedance has been synthesized, noise is added. However, this is done in such a way that the difference of impedance measured at different times contains less noise than the separate measurements. So

$$\widetilde{Z}_{t_1} = Z_{t_1} + e_d + e_{t_1}, \tag{28}$$

$$\widetilde{Z}_{t_2} = Z_{t_2} + e_d + e_{t_2}, \tag{29}$$

$$\widetilde{Z}_{t_1} - \widetilde{Z}_{t_2} = Z_{t_1} - Z_{t_2} + e_{t_1} - e_{t_2}. \tag{30}$$

.

If the error in the impedance measurements is modelled by a Gaussian distribution

$$e_d + e_{t_1} \backsim N\left(0, \sigma_Z^2\right), \tag{31}$$

and so is the error in the impedance difference

$$e_{t_1} - e_{t_2} \backsim N\left(0, \sigma_{dZ}^2\right), \tag{32}$$

then e_{t_1}, e_{t_2} and e_d are also Gaussian

$$e_{t_1}, e_{t_2} \backsim N\left(0, \frac{\sigma_{dZ}^2}{2}\right), \tag{33}$$

$$e_d \backsim N\left(0, \sigma_Z^2 - \frac{\sigma_{dZ}^2}{2}\right). \tag{34}$$

The uncertainty in the baseline survey σ_Z^2, the monitor σ_Z^2 and the difference σ_{dZ}^2 are parameters of interest in the numerical experiments. The uncertainties in the synthetic noise, $\frac{\sigma_{dZ}^2}{2}$ and $\sigma_Z^2 - \frac{\sigma_{dZ}^2}{2}$, are derived from these.

4 Numerical experiments

Twin experiments were done on a horizontal 2D 2phase waterflooding application with two different permeability and well configurations; a 5-spot with vertical wells (5SPOT) and 2 horizontal wells perpendicular to two high-permeable streaks (2STREAKS). Synthetic pressure and impedance data was generated by running a reservoir simulator on a "synthetic truth" permeability field. The synthetic truths for 5SPOT and 2STRAKS are shown in Fig. (2) and Fig. (3). The well configurations for both experiments are shown in Fig. (4).

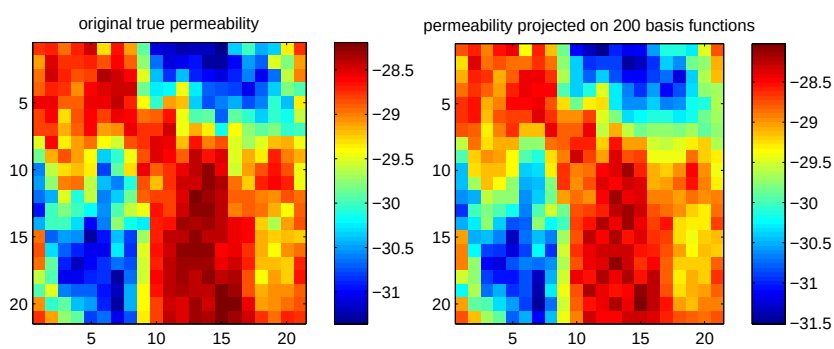

Figure 2: Synthetic truth permeability field for 5SPOT.

Water is injected at a rate of one pore volume per year and the production wells are constrained to 0.25 (5SPOT) or 1 (2STREAKS) pore volumes per year. The state of the reservoir is described by pressure and water saturation in all 21x21x1 grid blocks of

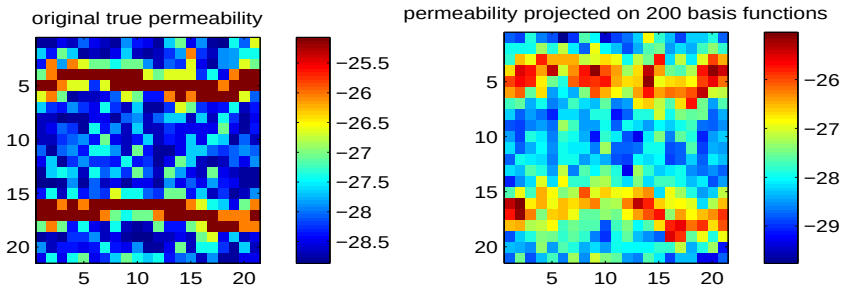

Figure 3: Synthetic truth permeability field for 2STREAKS.

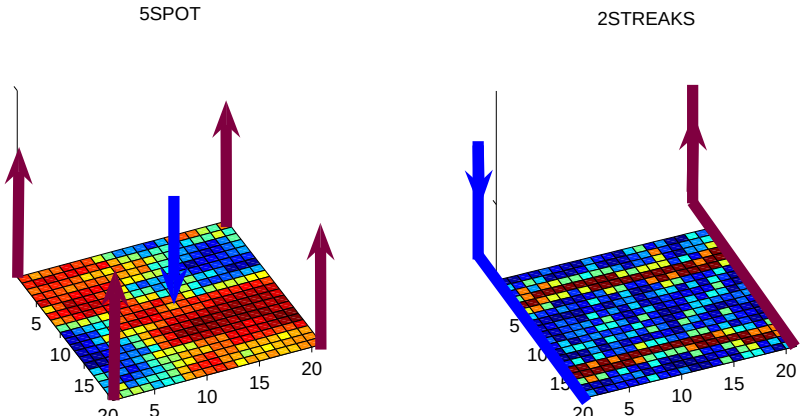

Figure 4: Well configurations. One vertical injector and four vertical producers for 5SPOT and two horizontal wells for 2STREAKS

10x10x20 m each. Capillary pressure is ignored, as well as gravity effects. For 5SPOT, the true permeability is picked as one realization out of a database of 1000 realizations. The others realizations are used to construct a covariance matrix that is used in the objective function that has to be minimized. For 2STREAKS, the truth is an academic caricature and realizations were sampled from a covariogram [Vossepoel and Douma, 2008]. The synthetic true model errors were sampled as white noise, Fig. (5).

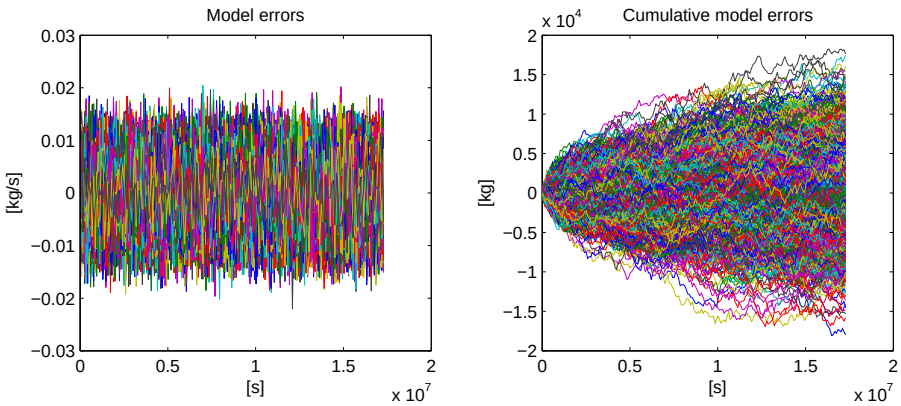

Figure 5: Synthetic model errors.

The state variables of the synthetic truth, located at the well positions, are shown in Fig. (6) for 5SPOT and Fig. (7) for 2STREAKS as functions of time. The synthetic pressure measurements are the reservoir pressures in the grid blocks that are penetrated by wells at 100 and 200 days of the simulation, resulting in 10 measurements for 5SPOT and 84 for 2STREAKS. Measurement errors are sampled and added to the synthetic measurements. Fig. (8) and Fig. (9) show the state variables after 200 days of simulation. The 882 impedance measurements, including measurement noise, are shown in Fig. (10) and Fig. (11).

Fig. (12) and Fig. (13) show some permeability estimates and the objective function during 38 iterations of VPERM on 5SPOT. Fig. (14) shows the estimated well responses. Fig. (15), Fig. (16) and Fig. (17) show the same for 2STREAKS. All iterations use the pressure data from the grid blocks that are penetrated by wells. Baseline and time-lapse data is used starting from iteration 20. If impedance data is used from the first iteration, then a different local minimum of the objective function is found, usually one that gives a worse history match. Fig. (18) shows the objective function of RM for the cases where seismic data is used in all iterations or only after the twentieth iteration. The first case converges much slower than the second. Moreover, the root-mean-square-error (RMSE) between the true and the estimated parameters is better for the second case; 0.5660 compared to 0.7536. If a history matching workflow without seismic data gives reasonable results, then it is not yet a trivial exercise to add seismic data. Additional data does potentially give more information, but it also increases the dimensionality of the non-linear estimation problem. Using impedance data from the twentieth iteration instead of

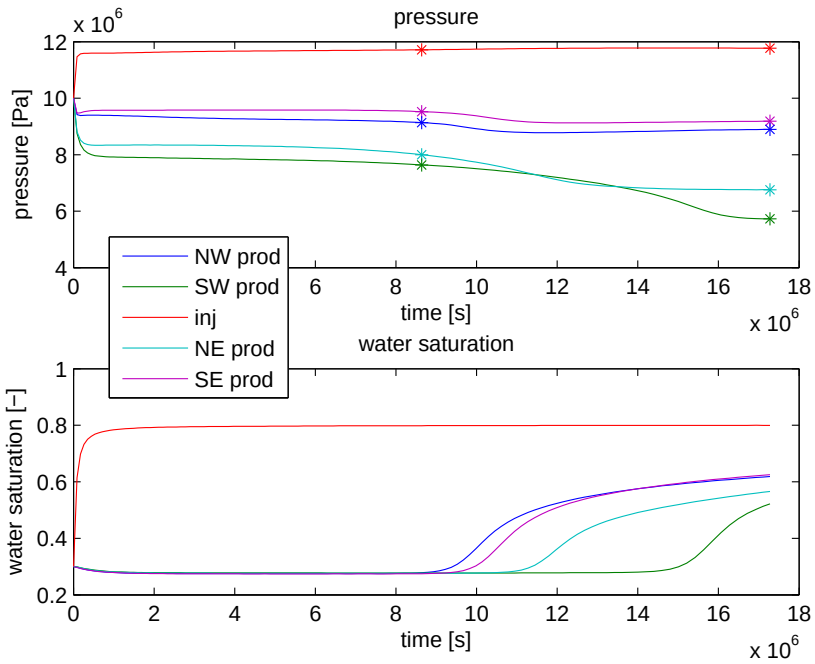

Figure 6: State variables of the synthetic truth for 5SPOT. Only the 5 grid blocks that are penetrated by the injection well and the north west (NW), SW, NE and SE production wells are shown. The top plot also shows synthetic pressure measurements prior to adding measurement noise, indicated by the asterisks.

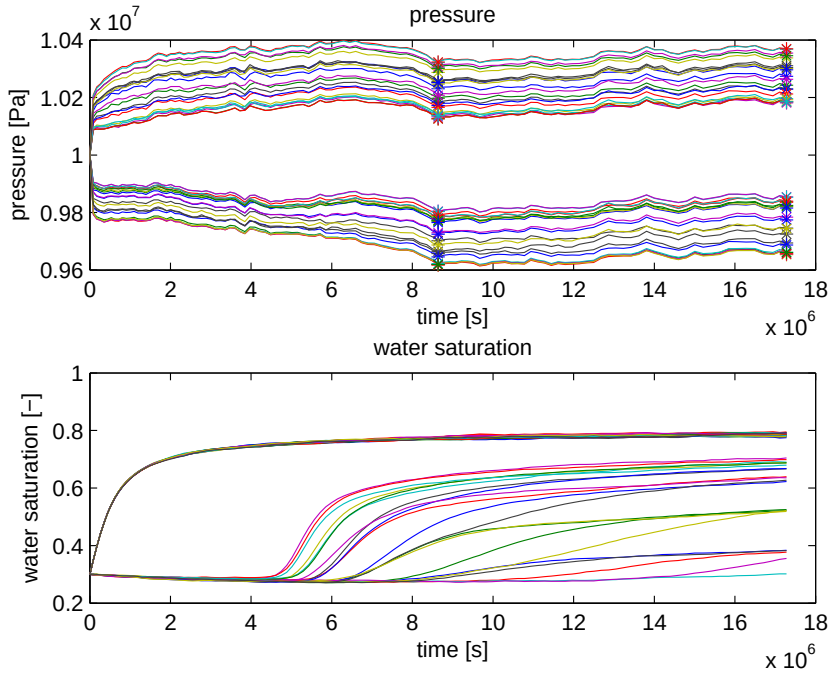

Figure 7: State variables of the synthetic truth for 2STREAKS. Only the 42 grid blocks that are penetrated by the horizontal injector and producer are shown. The top plot also shows synthetic pressure measurements prior to adding measurement noise, indicated by the asterisks.

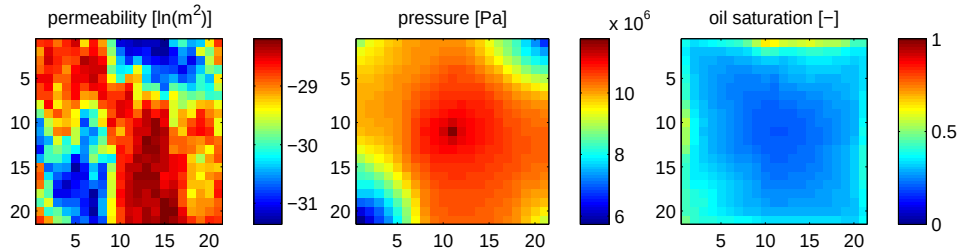

Figure 8: State variables after 200 days of simulation for 5SPOT.

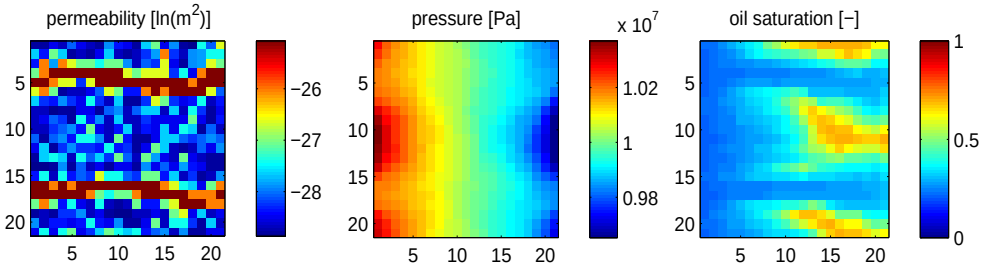

Figure 9: State variables after 200 days of simulation for 2STREAKS.

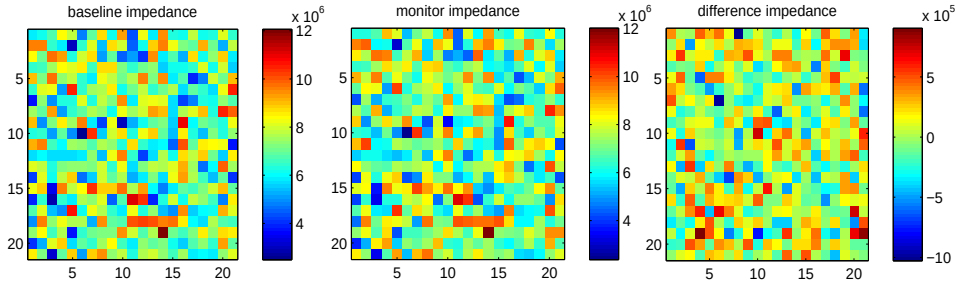

Figure 10: Impedance measurements for 5SPOT.

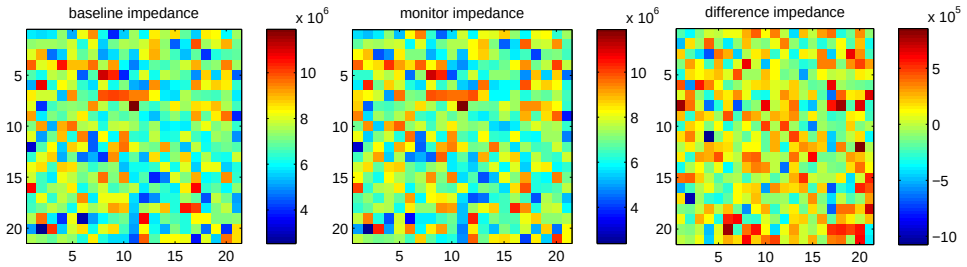

Figure 11: Impedance measurements for 2STREAKS.

the first one is a practical trick to get a variational history matching workflow to work with impedance data. The extra seismic data introduces new local minima in the neighborhood of the initial/prior estimate. Modifications to get an Ensemble Kalman Filter to work with saturation data for 2STREAKS were investigated by [Vossepoel and Douma, 2008].

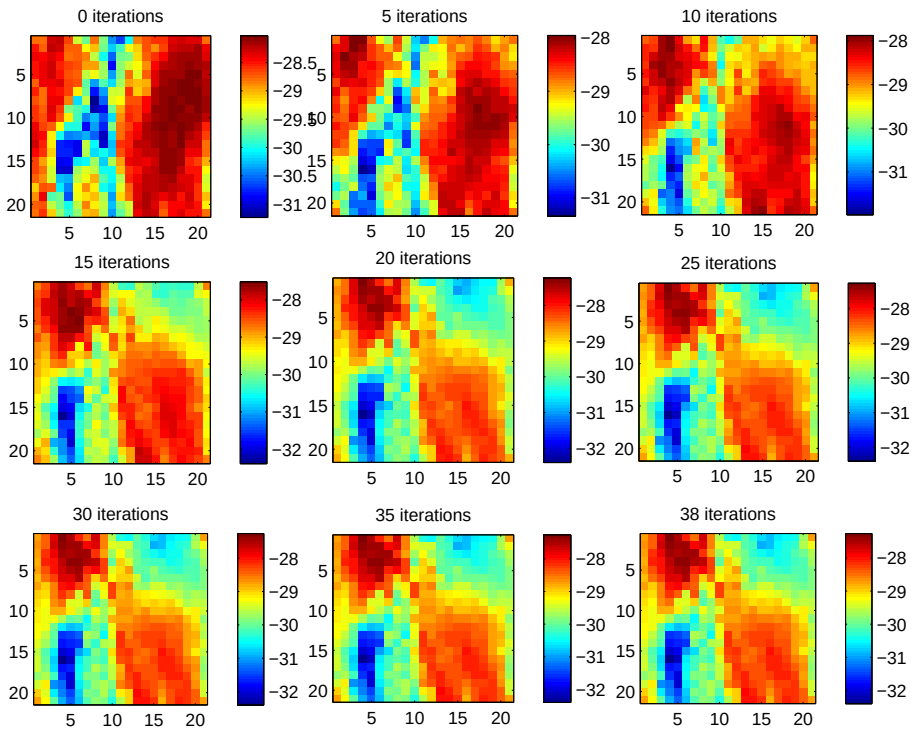

Figure 12: Permeability estimate during 38 iterations with VPERM for 5SPOT. Pressure data in the wells is used for all iterations, baseline and time-lapse impedance data is used in iterations 21 and up.

Once a history matching workflow is set up to work with or without seismic data, then the influence of the seismic data can be assessed. In this research seismic data never had a negative effect and usually had a positive effect. In order to assess the effect of seismic data, crossplots were made between the history match and the "future match", Fig. (19), Fig. (20) and Fig. (21), where the history matching workflows were repeated with different priors/initial estimates, and with three different datasets. All datasets contain the pressure data in the grid blocks that are penetrated by wells. The second and third dataset contain the baseline impedance data. The second dataset contains the monitor impedance data and the third dataset contains the time-lapse impedance data. The "history match" is quantified by the Root Mean Square Error (RMSE) between the estimated permeability and the synthetic truth permeability. The "future match" is quantified by looking at the accuracy of the water breakthrough curves; in fact, it is the square of the 2-norm of a

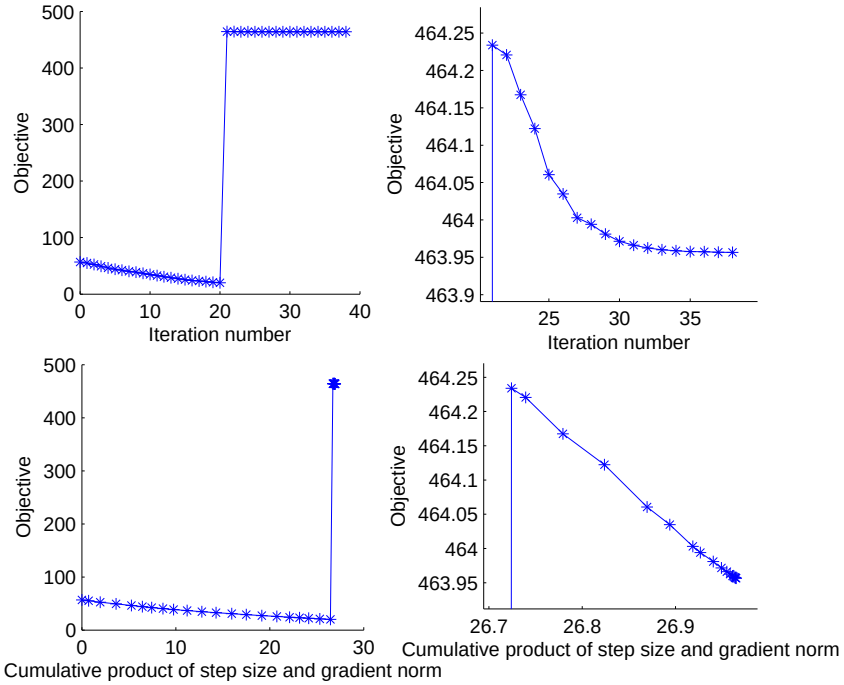

Figure 13: Objective function during 38 iterations with VPERM for 5SPOT. Pressure data in the wells is used for all iterations, baseline and time-lapse impedance data is used in iterations 21 and up. The plots on the right are close-ups of the plots on the left.

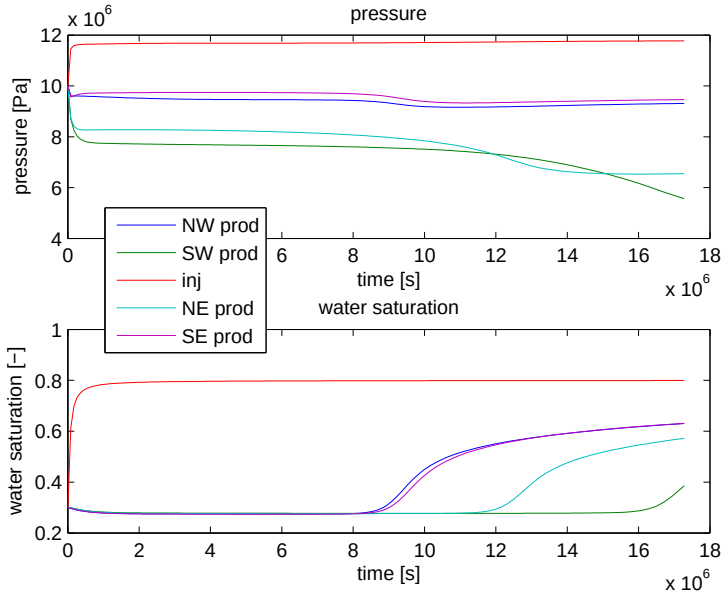

Figure 14: Estimated well response after 38 iterations with VPERM for 5SPOT. Pressure data in the wells is used for all iterations, baseline and time-lapse impedance data is used in iterations 21 and up.

vector that contains the time shifts of the estimated well responses to best match the synthetic well responses. For 5SPOT, Fig. (19), 3D seismic data does contribute to better history-matched models, although these models do not have better predictive capability than models that were history-matched without seismic data. 4D seismic data reduces the outliers in the crossplot, meaning that the models do have a better future predictive ability, without being better history matched. This is also the message of [Walker and Lane, 2007]. It must be noted that 3D seismic data is actually 2D, and 4D is actually 3D, since the reservoir model is only 2D and not 3D. For 2STREAKS, Fig. (21), 3D seismic data has hardly any effect. 4D seismic data gives better history matched models as well as better "future-matched" models. Fig. (20) is a combination of 5SPOT and 2STREAKS; 2STREAKS is used as the synthetic truth, and realizations were created by subtracting the mean from 5SPOT, adding the synthetic truth of 2STREAKS and performing some random shuffle operations on the permeability fields. Again, having seismic data available removes the outliers in the crossplot.

The VPERM method turns a high-dimensional weak constraint parameter estimation problem into a lower dimensional strong constraint parameter estimation problem by choosing basis functions for the parameters and the model errors inspired by the representer method. In the absence of model errors, it is very common to reduce the dimensionality of the strong constraint problem by choosing basis functions for the parameters, for example by calculating the principal components of the parameter covariance using SVD.

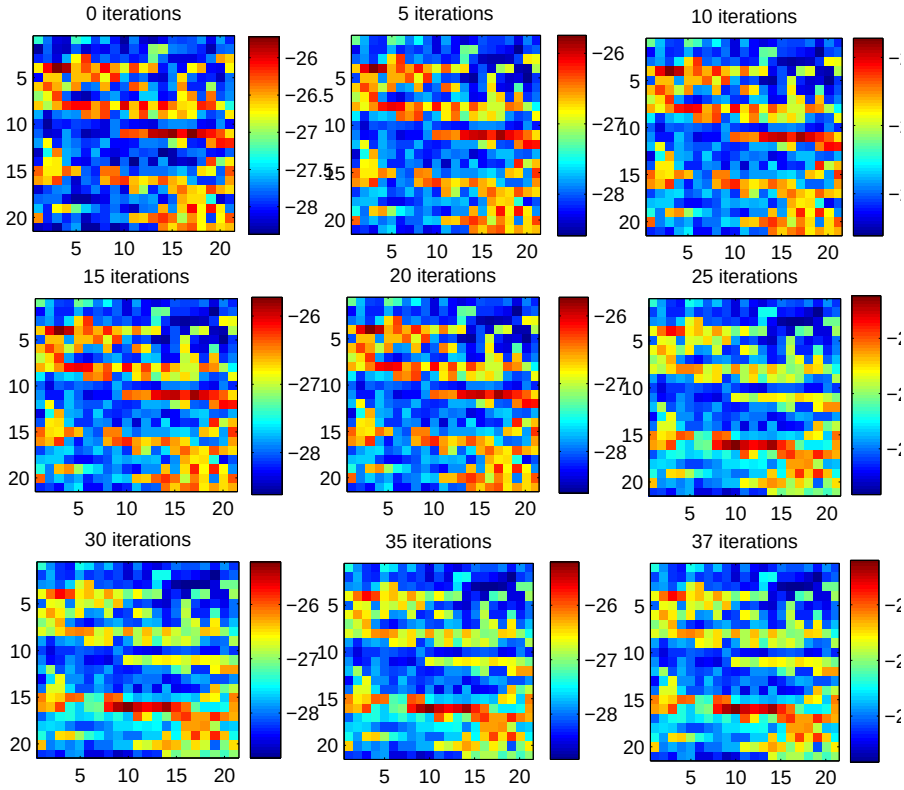

Figure 15: Permeability estimate during 37 iterations with VPERM for 2STREAKS. Pressure data in the wells is used for all iterations, baseline and time-lapse impedance data is used in iterations 21 and up.

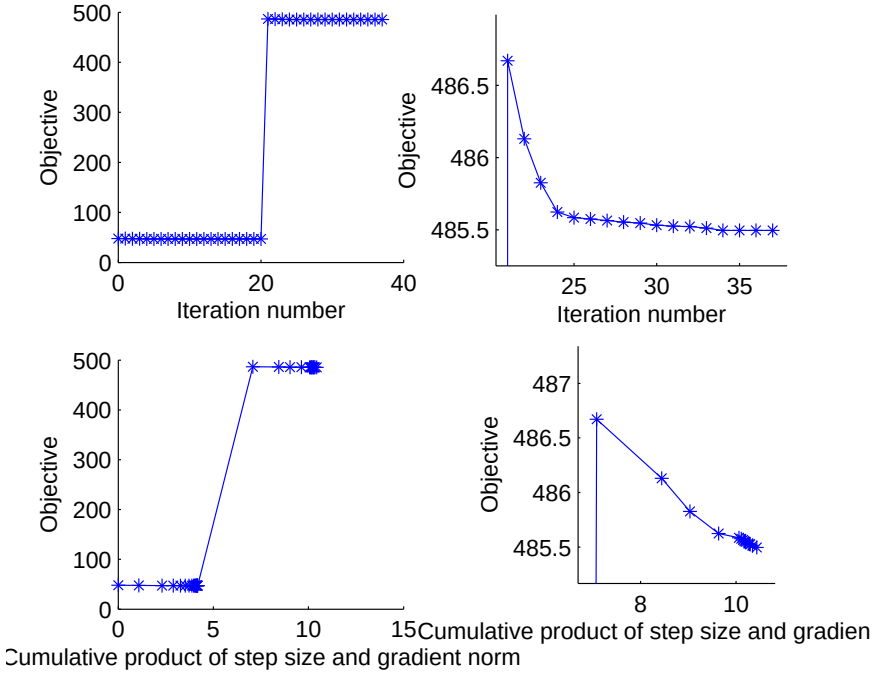

Figure 16: Objective function during 37 iterations with VPERM for 2STREAKS. Pressure data in the wells is used for all iterations, baseline and time-lapse impedance data is used in iterations 21 and up. The plots on the right are close-ups of the plots on the left.

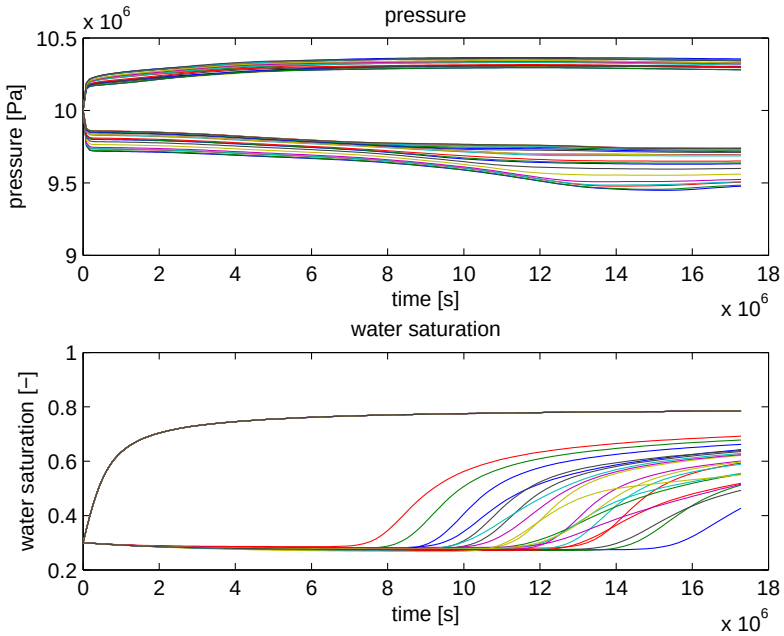

Figure 17: Estimated well response after 37 iterations with VPERM for 2STREAKS. Pressure data in the wells is used for all iterations, baseline and time-lapse impedance data is used in iterations 21 and up.

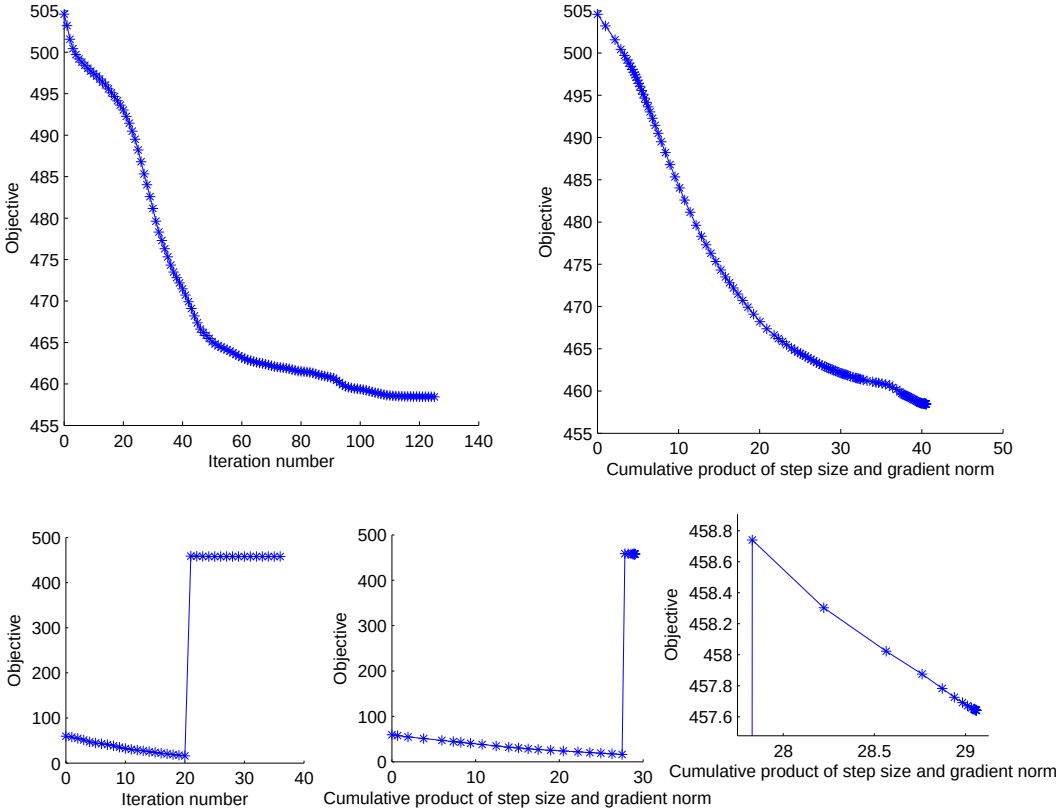

Figure 18: Objective function of 5SPOT using RM. Top: all iterations use pressure data and baseline and time-lapse impedance data. Bottom: impedance data is only used after iteration 20.

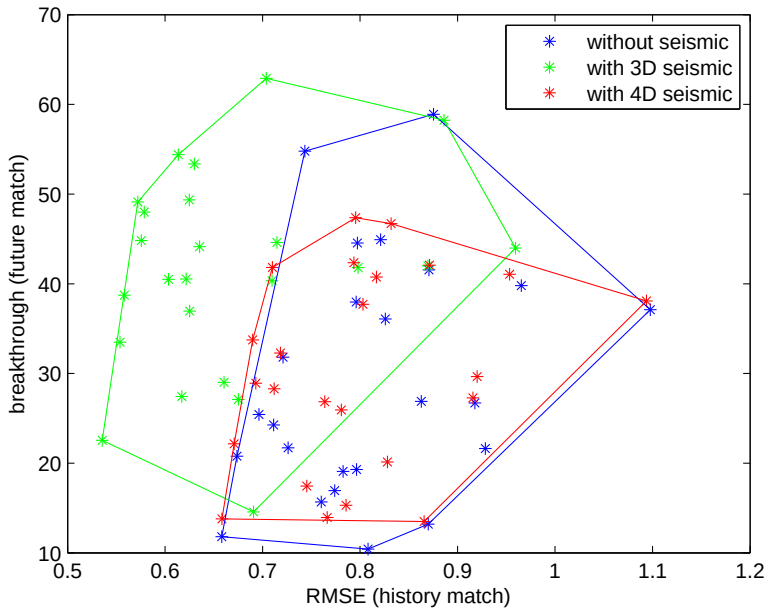

Figure 19: History-match future-match crossplot for 5SPOT using 25 different priors and 3 different datasets.

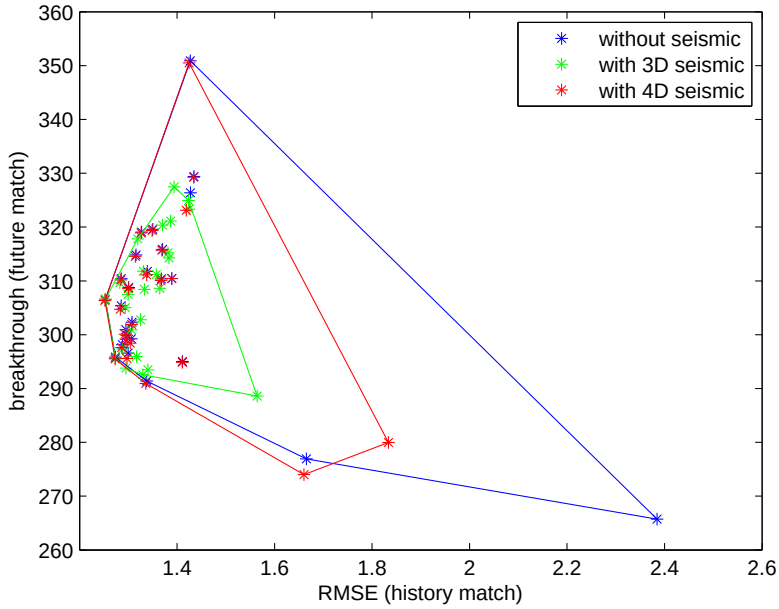

Figure 20: History-match future-match crossplot for 2STREAKS/5SPOT using 25 different priors and 3 different datasets.

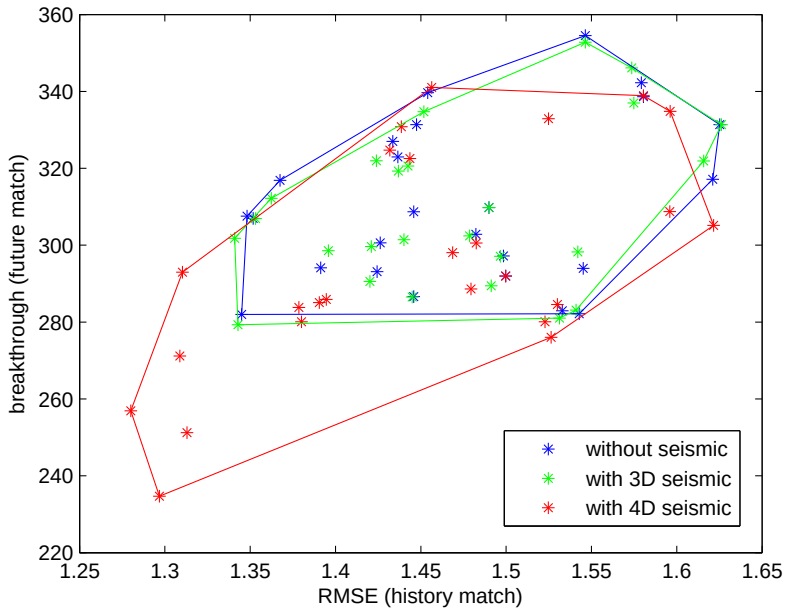

Figure 21: History-match future-match crossplot for 2STREAKS using 25 different priors and 3 different datasets.

Optionally, these basis functions may also be used for the parameters in a VPERM workflow. Sometimes this even results in faster convergence or convergence to a better local minimum, Fig. (22).

5 Conclusions

In this article, a weak constraint Variational Parameter Estimation algorithm with Model-driven Regularization (VPERM) is introduced. The regularizations turns the high order weak constraint parameter estimation problem into a lower order strong constraint problem. The basis functions that perform the regularization are model-driven, like in the Representer Method (RM). However, if the basis functions are kept fixed in the VPERM, then an analytical gradient can be obtained, whereas the gradient in the RM is merely an approximation; it is the direction of a step in an iterative scheme with Picard iteration that is not proven to converge. In the RM, basis functions, or representers, are calculated in every iteration. In VPERM, the basis functions are not updated at every iteration. This reduces computation time per iteration, but it might also increase the total number of iterations.

Experiments were done with VPERM, attempting to assimilate production data, baseline and monitor P-wave impedance data and difference P-wave impedance data. From these experiments, the following can be concluded:

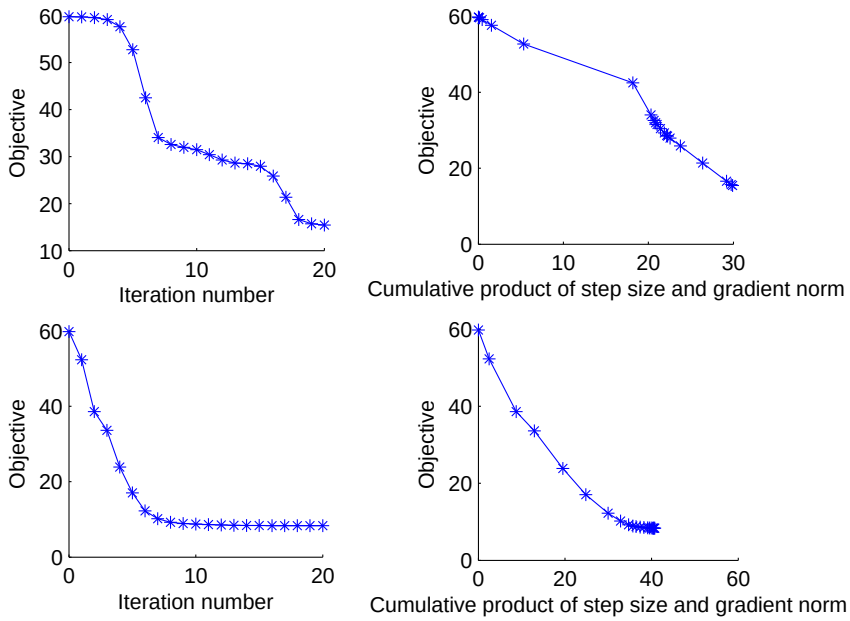

Figure 22: Top: VPERM for 5SPOT without seismic data. Bottom: parameter basis functions replaced using SVD.

- VPERM produces results that are similar to the results of the RM, but in less computation time per iteration. Most figures in this article show VPERM results and only Fig. (18) shows RM results. However, many experiments were done with both methods. In fact VPERM was developed when RM provided disappointing results (slow convergence, too high history match RMSE) with seismic data, but then performed similar to RM. Both methods produce good results when only pressure data obtained from the wells are available.

- If a data assimilation workflow gives reasonable results without seismic data, then adding seismic data is not a trivial exercise. Although additional data does potentially give more information, it also increases the dimensionality of the parameter estimation problem. The seismic data introduced new local minima of the objective function in the neighborhood of the initial estimate. Perturbing the initial estimate in a random direction will improve the seismic impedance predictions in some grid-blocks, but will also move predicted impedance values away from the true impedance in other grid blocks. In a mean-square-difference type of objective function these effects are cancelled out if there are as many improvements as deteriorations. In the direction of the objective function gradient there are only improvements, as long as the step size is small enough. In the neighborhood of the initial estimate the step size must be chosen so small, that improvement is only marginal and even below rea-

sonable thresholds. In this research, using pressure data in all iterations and using seismic data from the 20th iteration gave good results. Similar modifications must be used for other data assimilation workflows. For example, [Vossepoel and Douma, 2008] investigated modifications for the Ensemble Kalman Filter to give good results with saturation data on the 2STREAKS model that was also used in this article.

- Using seismic data with VPERM never gives worse results than not using seismic data.
- In some cases, assimilating seismic data does not result in better history-matched models, whereas the models do give better future predictions; outliers in the history-future crossplot are removed. This was earlier reported by [Walker and Lane, 2007]. In other cases, both the history match and the "future match" are improved by seismic data.
- Sometimes convergence can be improved, or a better local minimum of the objective function can be found, if the parameter basis functions are overwritten with the principal components of the prior parameter covariance.

References

K. Aziz and A. Settari. *Petroleum reservoir simulation*. Applied Science Publishers, 1979.

J. Baird and C. Dawson. The representer method for data assimilation in single-phase Darcy flow in porous media. *Computational Geosciences*, 9(4):247–271, 2005. DOI: 10.1007/s10596-005-9006-2.

A. F. Bennett. *Inverse modeling of the ocean and atmosphere*. Cambridge University Press, 2002.

A. F. Bennett. *Lagrangian fluid dynamics*. Cambridge Monographs on Mechanics. Cambridge University Press, New York, 2006.

A. F. Bennett and P. C. McIntosh. Open ocean modeling as an inverse problem; tidal theory. *Journal of Physical Oceanography*, 12(10):1004 – 1018, 1982.

M. Eknes and G. Evensen. Parameter estimation solving a weak constraint formulation for an Ekman flow model. *Journal of Geophysical Research*, 102(6C):12479 – 12491, 1997.

G. Evensen. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53, 2003.

G. Gao and A. C. Reynolds. An improved implementation of the LBFGS algorithm for automatic history matching. *SPE Journal*, 11(1):5 – 17, 2006. DOI: 10.2118/90058-PA.

G. M. C. M. Janssen, J. R. Valstar, and S. E. A. T. M. van der Zee. Inverse modeling of multimodal conductivity distributions. *Water Resources Research*, 42(3):Art. No. W03410, 2006.

G. Mavko, T. Mukerji, and J. Dvorkin. *The rock physics handbook; tools for seismic analysis in porous media*. Cambridge University Press, Cambridge, 1998.

J. Nocedal and S. J. Wright. *Numerical optimization*. Springer Verlag, New York, 1999.

J. K. Przybysz, R. G. Hanea, J. D. Jansen, and A. W. Heemink. Application of the representer method for parameter estimation in numerical reservoir models. *Computational Geosciences*, 11(1):73 – 85, 2007. DOI: 10.1007/s10596-006-9035-5.

J. R. Rommelse, J. D. Jansen, and A. W. Heemink. An efficient gradient-based parameter estimation algorithm using representer expansions. Submitted July 2008 for publication in SPEJ. Preprint available as. Technical Report 07-05, Delft University of Technology, March 2007.

J. R. Rommelse, J. D. Jansen, A. W. Heemink, F. Wilschut, and O. Kleptsova. Comparison of the ensemble Kalman filter and a modified representer method for sensitivity to prior data. Submitted July 2008 for publication in COMG. Preprint available as. Technical Report 08-16, Delft University of Technology, July 2008.

J. A. Skjervheim, G. Evensen, S. I. Aanonsen, B. O. Ruud, and T. A. Johansen. Incorporating 4D seismic data in reservoir simulation models using ensemble Kalman filter. In *SPE Annual Technical Conference and Exhibition*, Dallas, 2005.

S. Ulbrich. Limited-Memory BFGS-Verfahren mit Powell-Wolfe Schrittweitenregel, 2002. `http://www.mathematik.tu-darmstadt.de:8080/ags/ag10/Team/ulbrich/opt3/codes.html`.

J. R. Valstar, D. B. McLaughlin, C. B. M. te Stroet, and F. C. van Geer. A representer-based invserse method for groundwater flow and transport applications. *Water Resourses Research*, 40 W05116, 2004. DOI:10.1029/2003WR002922.

F. Vossepoel and S. G. Douma. Assimilating saturation data with the ensemble Kalman filter (personal communication), June 2008.

G. J. Walker and H. S. Lane. Assessing the accuracy of history-matching predictions and the impact of time-lapse seismic data; a case study for the Harding reservoir. In *SPE Reservoir simulation symposium*, Houston, 2007.