



SPE 153924

Reducing the Dimensionality of Geophysical Data in Conjunction with Seismic History Matching

Javad Rezaie, SPE, Norwegian University of Science and Technology (NTNU)

Jon Sætrom, Statoil

Eivind Smørgrav, SPE, Statoil

Copyright 2012, Society of Petroleum Engineers

This paper was prepared for presentation at the EAGE Annual Conference & Exhibition incorporating SPE Europec held in Copenhagen, Denmark, 4–7 June 2012.

This paper was selected for presentation by an SPE program committee following review of information contained in an abstract submitted by the author(s). Contents of the paper have not been reviewed by the Society of Petroleum Engineers and are subject to correction by the author(s). The material does not necessarily reflect any position of the Society of Petroleum Engineers, its officers, or members. Electronic reproduction, distribution, or storage of any part of this paper without the written consent of the Society of Petroleum Engineers is prohibited. Permission to reproduce in print is restricted to an abstract of not more than 300 words; illustrations may not be copied. The abstract must contain conspicuous acknowledgment of SPE copyright.

Abstract

Seismic history matching (HM) has attracted increasing attention the last few years. As more repeated seismic surveys are acquired, the more apparent the shortcomings in modern HM tools and algorithms become. A common conception seems to be that the amount of data represented by geophysical observations and the complexity of working with 3D fields make the updating procedure hard. We investigate the nature of geophysical observations from a HM point of view by testing several data reduction techniques such as Principal Component Analysis (PCA), regression techniques such as forward stepwise, as well as state-of-the-art techniques based on neural networks. We argue that simulated geophysical fields from the prior models are prone with spatial correlations and that their information content and effective dimensionality is much smaller than the dimensionality of the observed field. The techniques are tested on a reservoir model of an anonymous North Sea oil field, using the seismic time shift, i.e. the difference travel time integrated over the reservoir between two surveys. We find that PCA is particularly promising, resulting from the versatility and robustness of the method. In practice this means that high dimensional geophysical data, e.g. 2-D seismic images or 3-D seismic cubes, can often be described using only a handful of scalars. We show how to assess the information content in the data, compress the data, and use this compressed data in a reservoir conditioning setting. The methods we present are generic; they apply equally well to all geophysical attributes regardless of representation and can be applied with any history matching algorithm, although they are mainly designed for ensemble based techniques.

Introduction

Bayesian inversion is one statistical solution to the spatiotemporal inverse problem of finding the unknown properties of a reservoir given a sequence of observed data in time, commonly known as history matching. Methods based on Bayesian inversion are attractive to use in this setting because any prior knowledge of the reservoir with associated uncertainties can be accounted for in a systematic manner. Furthermore, likelihood models are defined through the known physical equations. The unknown properties of the reservoir can then be assessed by evaluating the posterior distribution, given through Bayes rule as the product of the prior and likelihood. Monte Carlo (MC) methods are in general used to evaluate the posterior distribution, as its analytical tractability will be lost. Hence, an ensemble of posterior realizations or reservoir models describes the statistical properties of the unknown reservoir.

One major challenge in data assimilation is determining the information content in the measured data. This is particularly challenging in the reservoir conditioning setting, because the data can include both static (e.g. core samples and well logs) and dynamic (e.g. oil production rates and time-lapse seismic) measurements. There is possibly a strong correlation between the different data types, which is important to take into account prior to the conditioning to prevent a severe model overfitting. In the statistical literature, this is a well-known problem and numerous methods exist in different contexts to prevent or reduce model overfitting in the presence of uninformative or collinear data (Hastie, et al., 2009). One notable example is of course classical linear regression.

Conditioning of reservoir models to seismic data has attracted increasing attention the last few years (Skjervheim, et al., 2007). With the increasing amount 4D seismic available, it becomes imminent to find efficient and robust ways of

conditioning these data jointly with the production data. One challenge when conditioning on seismic in particular is the potentially large data dimension. This follows because of high computational demands for the assisted HM algorithms and the possibility of a dramatic model overfitting, as discussed above. Hence, reducing the dimensionality of the data, by extracting the useful information and discarding the noise prior to assimilation, is natural to consider.

When considering MC based techniques such as the ensemble Kalman filter (EnKF) (Evensen, 2009), or the response surface Bayesian approach described in (Smørgrav & Slotte, 2008), the information regarding the prior and posterior distribution can be extracted from the ensemble of realizations. The same is true for the ensemble of forecasted data, found by evaluating the likelihood model. Although the dimension of the data (production and/or seismic) can be high, the rank of the ensemble of forecasted data is equal to the minimum of the ensemble size and the data dimension. In addition, there is possibly a high spatial and temporal correlation in this forecasted data. This implies that the true rank of the ensemble of forecasted data can be significantly smaller than the data dimension. Hence, the goal of this paper is to present and evaluate methods that can be used to extract the true information seen in the ensemble of forecasted data.

This paper is organized as following; we introduce the history matching/inversion problem in a general Bayesian setting and then discuss potential solutions based on a MC approach. Then, we look at various techniques for reducing the dimensionality of the data. Finally, we evaluate the properties of some of the proposed methods on a synthetic model inspired by a North Sea reservoir.

Notation

Throughout this paper, the notation $a \in \mathbb{R}^{n_a \times 1}$, will be used to denote that a is an n_a dimensional column vector and $a^T \in \mathbb{R}^{1 \times n_a}$ its transpose. Further we will use the notation $A \in \mathbb{R}^{c \times d}$ for matrices with c rows and d columns.

In a reservoir conditioning setting, the aim is to recover the unknown reservoir state variables, x , given the observed data $d \in \mathbb{R}^{n_d \times 1}$. This is known as an inverse problem, which in the HM setting is defined as a spatio-temporal inverse problem. The unknown state variables can include the grid block permeability, net-to-gross, fault multipliers, position of the top reservoir surface, or any other grid block property or hyper-parameter used in the construction of the reservoir geo and simulation model. The data vector can contain traditional reservoir monitoring data, such as production and pressure data, but in general we can use any data type, such as geophysical reservoir monitoring (GRM) data. As of today, however, the quantitative use of GRM data in reservoir conditioning is sparse and research based. Regardless of what the data vector contains, we will assume that it is connected to the reservoir state vector through a function, $g(\cdot): \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_d}$ (possibly non-linear), so that $d = g(x, \varepsilon_d)$. Here, ε_d is a random error term, representing the observation error. Note that this error term is not necessarily additive and Gaussian in a general setting.

The dimension of the reservoir state vector, $x \in \mathbb{R}^{n_x \times 1}$, can potentially be in the order of $10^6 - 10^8$ for realistic reservoir model if all unknown properties are taken into account. This will of course limit the number of algorithms that can be applied, because of potentially high computational demands. For this reason, the unknown reservoir state vector, n_x typically only consists of a limited number of parameters ($n_x \in O(10) - O(100)$) in a traditional reservoir conditioning setting.

Assume that we have generated an ensemble of independent, identically distributed (i.i.d.) realizations of the reservoir state vector from a known probability density function (pdf) $\pi(\cdot): x^{(1)}, \dots, x^{(n_e)} \stackrel{i.i.d.}{\sim} \pi(x)$. Further let the forecasted data observations be defined as $d^{(i)} = g(x^{(i)}, \varepsilon_d^{(i)})$, for $i = 1, \dots, n_e$. For notational convenience we will let $X = \begin{bmatrix} x^{(1)} & x^{(2)} & \dots & x^{(n_e)} \end{bmatrix} \in \mathbb{R}^{n_x \times n_e}$ and $D = \begin{bmatrix} d^{(1)} & d^{(2)} & \dots & d^{(n_e)} \end{bmatrix} \in \mathbb{R}^{n_d \times n_e}$ denote the state and data ensemble matrices.

Problem Formulation

Consider that we include all prior information about the state variables, x , in a known pdf $\pi(x)$. The likelihood, describing the physical relationship between the reservoir state vector and measured data is represented by a general distribution as $\pi(d|x)$. Through Bayes' rule, we can assess the unknown properties of the reservoir state vector by evaluating the posterior distribution:

$$\pi(x|d) = \text{const} \cdot \pi(d|x) \cdot \pi(x)$$

Note, that in general this posterior distribution is not analytically tractable, because of the unknown normalizing constant. Hence, MC based techniques are required to generate realizations from the posterior distribution of interest. It should be noted, however, that these methods can potentially suffer from high computational demands for the general HM problem. This follows because a single evaluation of the likelihood model can take hours or days on high dimensional reservoir

models, as it involves fluid flow simulation. One solution to this problem to define a proxy flow simulator, as suggested in (Smørgrav & Slotte, 2008), which greatly speeds up the evaluation of the likelihood model. This will ensure that realizations are generated from the correct posterior distribution. However, the construction and use of this proxy model might restrict the dimensionality of the state vector. An alternative solution is to instead replace the true posterior distribution with an approximate distribution. This corresponds to the EnKF solution to the problem, which can be defined under Gaussian model assumptions (Evensen, 2009) (Sætrum & Omre, 2011).

When considering inversion methods based on Bayesian inversion, the mismatch between observed and forecasted data is central. The usual norm for describing this mismatch is based on the Mahalanobis distance:

$$\|d^{(o)} - d^{(i)}\|_M = (d^{(o)} - d^{(i)})^T \Sigma_{\varepsilon_d}^{-1} (d^{(o)} - d^{(i)}).$$

Here Σ_{ε_d} is the covariance matrix for the observation errors. As discussed above, using all available data in the conditioning will potentially lead to a severe model overfitting. This is particularly true if the majority of the data carries little or no information regarding the state vector of interest, or is highly collinear.

Dimension Reduction Techniques

To prevent the problem of model overfitting we will consider methods that transforms the data into a low dimensional subspace: $\varphi: d \rightarrow z$, so that any comparison between forecasted and measured data is performed in this reduced order subspace:

$$\|z^{(o)} - z^{(i)}\|_M = (z^{(o)} - z^{(i)})^T \Sigma_{\varepsilon_z}^{-1} (z^{(o)} - z^{(i)}).$$

The intention of this transformation is twofold: First we effectively eliminate uninformative and collinear data in the full data space. Secondly, we greatly speed up the run-time of the algorithm when conditioning on high dimensional data as we only compare data variables in the reduced order subspace. In the following we will consider both linear and non-linear transformation functions. For simplicity, we will only present one of these dimension reduction techniques in detail, and rather give a brief summary of the remaining methods.

Principal Component Analysis (PCA);

Principal Component Analysis (PCA) is one of the most frequently used dimension reduction techniques. This is mainly due to the straightforward implementation using Singular Value Decomposition (SVD), the orthonormal properties, and that PCA minimizes the mean squared reconstruction error (Pearson, 1901).

The aim of PCA is to explain the structure of the data ensemble matrix, D through a small number of vectors, termed Principal Components (PC) in a p dimensional subspace:

$$z_1 = (u_1^T D)^T, \dots, z_p = (u_p^T D)^T \in \mathbb{R}^{n_d \times 1}.$$

The sample PCs are selected based on the following criteria:

$$z_i = (u_i D)^T - \begin{cases} \max_{u_i} \{u_i^T \hat{\Sigma}_d u_i\} \\ \|u_i\|_2 = 1 \\ z_i^T z_j = 0, \text{ for all } j < i, i = 1, \dots, p \end{cases}$$

where $\|\bullet\|_2$ denotes the Euclidean norm and $\hat{\Sigma}_d$ is the empirically estimated data covariance matrix. It can be shown (Hastie, et al., 2009) that the i^{th} sample PC direction u_i , or basis vector, is given as the i^{th} eigenvector of the covariance matrix $\hat{\Sigma}_d$. Hence, the matrix of basis vectors $[u_1, \dots, u_p]$ defines the transformation function φ . Further, it can be shown that the variance explained by the i^{th} PC is given by the i^{th} eigenvalue λ_i of $\hat{\Sigma}_d$, where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$. A popular criterion for selecting the subspace dimension is therefore to choose p so that the explained variance is larger than some tolerance level, α (e.g. 95%). It should be noted, however, that in settings where $n_d \gg n_e$, this criterion is known to be severely biased, and should be used with caution (Ledoit & Wolf, 2004).

In the reservoir conditioning setting, the ultimate goal is to recover the unknown state vector using the forecasted and observed data. One potential problem with PCA in this setting is that any information contained in the state vector ensemble matrix is not used when constructing the reduced order PC. Hence, we can potentially discard all the information important for the unknown state vector if PCA is applied (Jolliffe, 2002). Model validation is therefore extremely important when PCA

is used in this inverse problem setting (Sætrum & Omre, 2011), (Sætrum, et al., 2011). Note also that PCA is not scale invariant. That is, multiplying each grid block with some constants, where at least two of the constants are different, will lead to a different solution (Mardia, et al., 1979). An implication of this can be that the grid block where there is low or high variability in e.g. the forecasted seismic data, compared to the majority of the field, can dominate the principal components. To circumvent this issue, the approach often used is to compute eigenvectors based on the correlation matrix instead of the usual covariance matrix. Note that this is equal to standardizing the data ensemble matrix prior to computing the SVD. Finally, one important property of the PCA is that it does not rely on any model assumptions between the state and data variables. This is not true for the dimension reduction techniques discussed in the remaining part of this paper.

Regression Based Dimension Reduction Techniques;

It is well known that the EnKF updating scheme can be equally defined as a multivariate linear regression problem [e.g.] (Sætrum & Omre, 2011), where the classical least squares solution defines the unknown matrix of regression coefficients:

$$\hat{K} = \arg \min_K \left\{ (X - KD)(X - KD)^T \right\}.$$

In the statistical literature, it is well known that this classical least squares solution is far from optimal in the presence of collinear data. Shrinkage estimators are therefore used to prevent this problem which locates and down-weight/eliminate uninformative data variables from the regression model (Hastie, et al., 2009). Possible solutions are partial least squares regression (PLS), forward stepwise regression, least angle regression, and elastic-net regression. PLS behaves similar to the PCA scheme defined above, although both the state and data variables are applied while constructing the transformation function φ . In addition there is an implicit assumption of a linear relationship between the state and data variables (Sætrum & Omre, 2011). Hence, similar to PCA, linear combinations of all data variables, $d_i, i = 1, \dots, n_d$ defines the reduced order variables z_i . The remaining three regression approaches considered above, on the other hand performs the dimension reduction by completely eliminating data variables classified as uninformative. The main difference between these three methods is the use of the L_1 or L_2 norm when solving the multivariate linear regression problem. Details can be found in (Hastie, et al., 2009). Hence, for the dimension reduction techniques based on sparse linear regression the transformation function, φ , is in general a non-linear function taking a subset of the data variables as input.

Simulation

We will now consider a synthetic 3-D reservoir case that has been previously used as a test case in the EnKF literature (Seiler, et al., 2009), (Sætrum, et al., 2011), (Skjervheim, et al., 2011). In this paper, the unknown state vector of interest is the grid block permeability and porosity. The reservoir domain is of size (10000×6500×40) meters, discretised into (40×64×14) grid blocks, with 27755 active grid cells and top reservoir at depth 1558 meters. Figure 1 displays the reference horizontal permeability (PERMX). The reservoir has five vertical production-, and three vertical water injection wells, turned on at the start of the production. The five production wells are operated by a constant bottom-hole pressure of 30 MPa, and we assume that the initial reservoir pressure is 33 MPa. Production of the reservoir is continued for three years. Following the approach in (Skjervheim, et al., 2011), we represent seismic data by both an accumulated 4-D time-shift surface, calculated from the porosity, pressure and saturation in each grid cell using a petro-elastic model in (Skjervheim, et al., 2011). The measured seismic time-shift is shown in Figure 2.

In this study, the unknown reservoir state vector consists of one multiplier for the pore volume in the grid blocks of the first y, z-vertical cross section, m_{AQ1} , the connate water saturation (SWL), the critical gas saturation (SGCR), the critical oil-in-water saturation (SOWCR), and fault multipliers m_{F1}, \dots, m_{F6} , modifying the grid block transmissibility, in the six faults defined in Figure 3. This implies that the unknown reservoir state vector consists of $n_x = 10$ parameters.

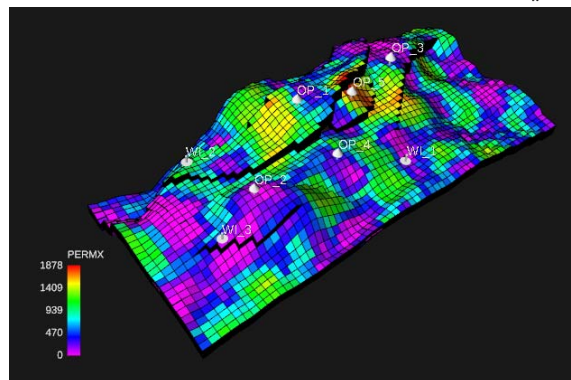


Figure 1: Reference horizontal permeability (PERMX).

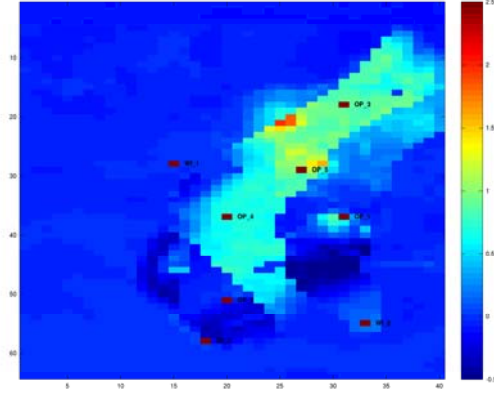


Figure 2: Measured Seismic time-shift map.

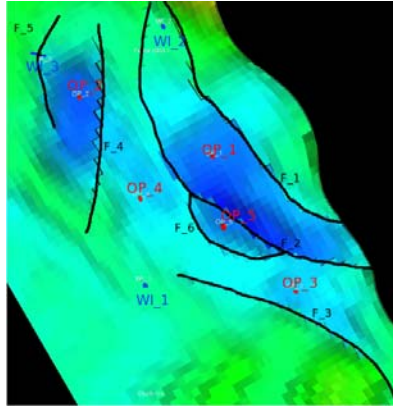


Figure 3: Location of faults F_1, \dots, F_6 .

An ensemble of $n_e = 40$ realizations have been generated from uniform prior distributions, where $m_{AQ1} \sim Unif(8, 100)$, $\log(m_{F_1}), \dots, \log(m_{F_6}) \sim Unif(-6, 0)$, $SGCT, SOWCR \sim Unif(0.01, 0.4)$ and $SWL \sim Unif(0.05, 0.3)$. Note that we use Latin Hypercube sampling, instead of the standard random sampling, to better span the prior sample space. The corresponding data ensemble matrix $D \in \mathbb{R}^{2560 \times 40}$, consists of the forecasted time-shift seismic surfaces. Figure 4 displays the forecasted seismic data from four different realizations. The purpose of this study is to investigate how the the dimension reduction techniques described above performs on this forecasted seismic data.

Figure 5 displays the first four basis vectors, u_1, \dots, u_4 , obtained when we apply PCA to the data ensemble matrix of forecasted 4-D time shift surfaces. Worth noting is the apparent spatial structure of these basis vectors, which appears to be aligned with the faults of the reservoir model. The first basis vector apparently describes the contrast between areas with small and large differences in the forecasted seismic data, while the interpretation is more complicated for the remaining $n_e - 1$ basis vectors. Notice, however, that especially the compartment between faults F_1 and F_2 is clearly visible in the first four basis vectors. Further notice that, the grid blocks with zero sample variability in the data ensemble, will necessarily give a value of zero in the corresponding grid block of the basis vectors. Figure 6 displays the last four basis vectors, u_{36}, \dots, u_{40} . Compared with the basis vectors in Figure 5, these basis vectors appears as less structured with isolated grid blocks having large or small values. This behavior is expected, because the underlying model assumption of PCA is that the data signal can be contained in a low dimensional subspace, implying that the remaining basis vectors should represent the data noise. It should be noted that for this particular example, the first five principal components account for more than 90% of the total variation in the data. Hence, the data dimension of $n_d = 2560$ can be dramatically reduced to $n_z = 5$, without a large information loss which greatly simplifies the use of this seismic data in a HM context.

Figure 7 displays the leading four basis vectors obtained by applying the PLS dimension reduction scheme on the forecasted data ensemble. Compared with the basis vectors obtained with the PCA scheme in Figure 5, there are only subtle differences

for the first three basis vectors, other than a difference in the sign and scaling. In the fourth one, however, grid blocks in the neighborhood of the three injection wells are given a large positive weight contrary to the fourth PCA basis vector where the same grid blocks are given weights close to zero. Note also that the PLS scheme takes the predictive power of the underlying regression model into account, hence we typically require a smaller subspace dimension to capture the information compared with the unsupervised PCA scheme (Sætrum & Omre, 2011).

In Figures 8 to 10 the value and position of the non-zero regression coefficients for fault multiplier m_{F3} is shown. Note that all three schemes are keeping data variables in a neighborhood of the fault location. The forward-stepwise regression scheme is only keeping three data variables, while nine non-zero variable are kept in the least angle regression scheme. For the dimension reduction scheme based on elastic net regression, however, 25 data variables are kept.

It should be noted that although the sparse dimension reduction techniques in theory can be applied to any history matching algorithm; it is most natural to consider these in an EnKF context. This follows because each of the linear regression problems needs to be solved separately for each reservoir state variable, $x_i, i = 1, \dots, n_e$, which is implicitly done in the EnKF. Hence, it is not trivial how to combine all non-zero data variables to define the complete transformation function, φ , in the general setting. Both PCA and PLS does not have this potential problem, and therefore applies equally well for any ensemble based HM algorithm.

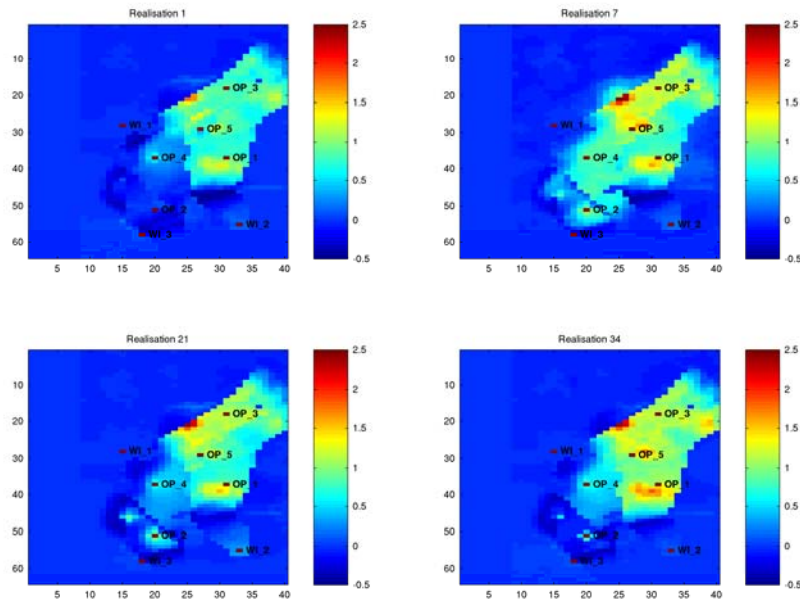


Figure 4: Forecasted seismic data from four different realizations.

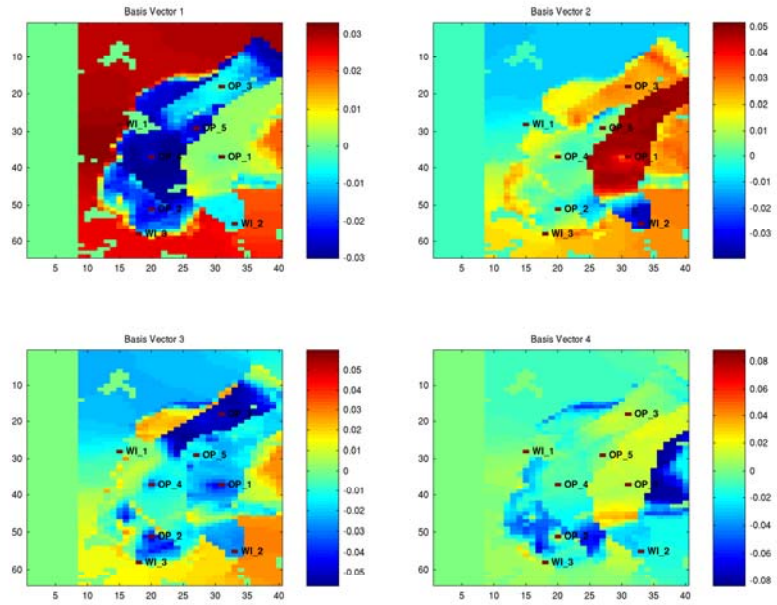


Figure 5: The first four basis obtained by applying PCA to the data ensemble matrix of forecasted 4-D time shift surfaces.

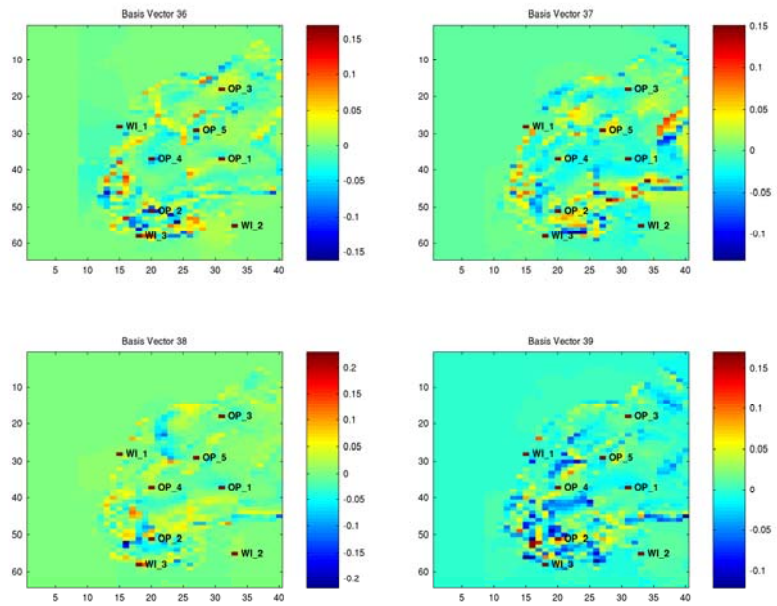


Figure 6: The last four basis vectors obtained by applying PCA to the data ensemble matrix of forecasted 4-D time shift surfaces.

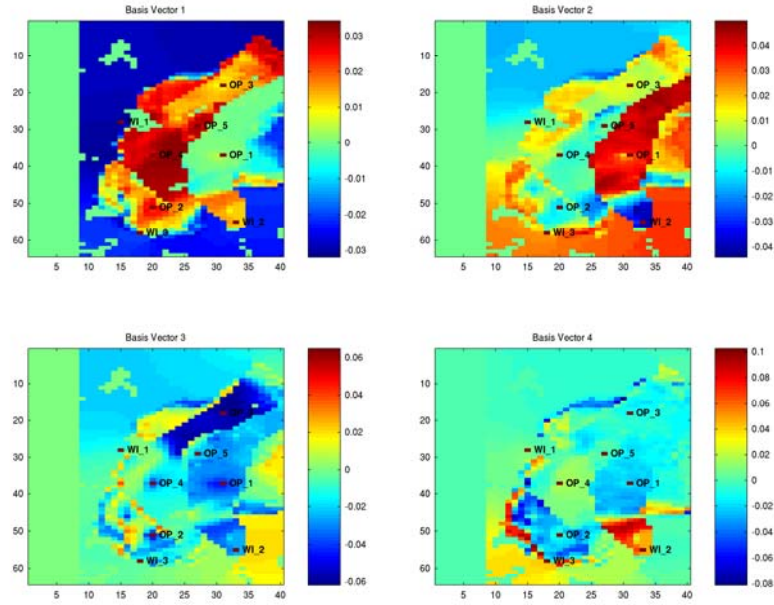


Figure 7: First four basis vectors obtained by applying PLS to the data ensemble.

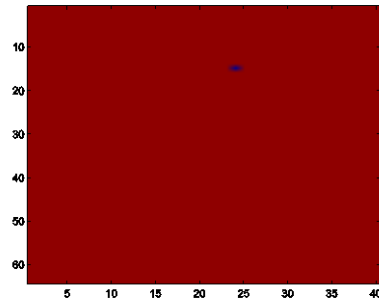


Figure 8: Position and value of non-zero regression coefficients for the multiplier of fault F_3 for the dimension reduction scheme based on forward stepwise regression.

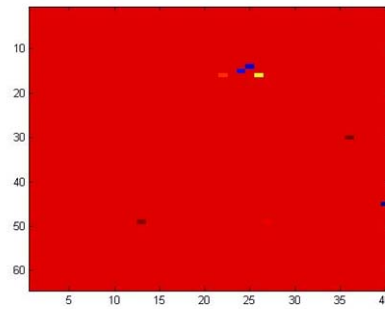


Figure 9: Position and value of non-zero regression coefficients for the multiplier of fault F_3 for the dimension reduction scheme based on least angle regression.

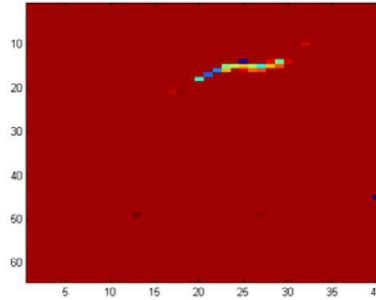


Figure 10: Position and value of non-zero regression coefficients for the multiplier of fault F_3 for the dimension reduction scheme based on elastic-net regression.

Summary

In recent years repeated seismic surveys have become more frequent and quantitative methods that can be used to integrate the information found in these surveys into reservoir models are called for. One common misconception is that the potentially large dimension of such data makes this a difficult problem to solve. In this paper we solve this problem using a Bayesian approach, where the statistical properties of the reservoir can be described through an ensemble of realizations generated from the posterior distribution of interest. When considering ensemble based methods for high dimensional data, the rank of the ensemble of forecasted data is at most equal to its size of 20-200. However, in practice there are collinearities between the ensemble members, which means that the effective rank can be dramatically smaller than the data dimension. This implies that we can essentially represent high dimensional data, such as time-lapse, in terms of a few scalars. In this paper we therefore applied dimension reduction techniques to efficiently handle problems related to the data dimensionality. We considered both basis vector-, and pixel based approaches. The proposed algorithms were applied on a reservoir model inspired by a North Sea oil field as the test benchmark. We find that the basis vector approach based on Principal Component Analysis (PCA) is particularly promising, resulting from the versatility and robustness of the method. PCA applies equally well for any ensemble based history matching algorithm, while the pixel based methods are designed mainly for the ensemble Kalman filter.

Acknowledgment

The first author acknowledges Statoil for providing a very good atmosphere for discussing and using their resources. The first author also acknowledges the sponsors of Uncertainty Reservoir Evaluation (URE) project at the Department of Mathematical Sciences, NTNU. Finally, the authors acknowledge Professor Jo Eidsvik for his useful discussions and comments.

Bibliography

- Evensen, G., 2009. *Data assimilation, The Ensemble Kalman Filter*. Berlin: Springer.
- Hastie, T., Tibshirani, R. & Friedman, J., 2009. *The Elements of Statistical Learning; Data Mining, Inference, and Prediction*. New York: Springer.
- Jolliffe, I. T., 2002. *Principal Component Analysis*. s.l.:Springer.
- Ledoit, O. & Wolf, M., 2004. well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, p. 365–411.
- Mardia, K. V., Kent, J. T. & Bibby, J. M., 1979. *Multivariate Analysis*. London: Academic Press.
- Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, Volume 2, p. 559–572.
- Seiler, A., Rivenæs, J. C., Aanonsen, S. I. & Evensen, G., 2009. *Structural uncertainty modelling and updating by production data integration*. Abu Dhabi, SPE-125352.
- Skjervheim, A. J. et al., 2007. Incorporating 4D Seismic Data in Reservoir Simulation Models Using Ensemble Kalman Filter.. *SPE Journal*, 12(3), p. 282–292.
- Skjervheim, J. A., Evensen, G., Hove, J. & Vabø, J. G., 2011. *An Ensemble Smoother for Assisted History Matching*. Woodlands, SPE-141929 .
- Smørgrav, E. & Slotte, P. A., 2008. *Response Surface Methodology Approach for History Matching and Uncertainty Assessment of Reservoir Simulation Models*. s.l., SPE-113390.
- Sætrom, J., Hove, J., Skjervheim, J. A. & Vabø, J. G., 2011. Improved Uncertainty Quantification in the Ensemble Kalman Filter Using Statistical Model Selection Techniques. *SPE Journal*.
- Sætrom, J. & Omre, H., 2011. Ensemble Kalman Filtering Using Shrinkage Regression Techniques. *Computational Geosciences*, 15(2), p. 271–292.