

IMPERIAL COLLEGE LONDON

Department of Earth Science and Engineering

Centre for Petroleum Studies

History Matching Using Principal Component Analysis

By

Akshay Sharma

**A report submitted in partial fulfillment of the requirements for
the MSc and/or the DIC.**

September 2011

DECLARATION OF OWN WORK

I declare that this thesis “**History Matching Using Principal Component Analysis**” is entirely my own work and that where any material could be construed as the work of others, it is fully cited and referenced, and/or with appropriate acknowledgement given.

Signature:.....

Name of student: Akshay Sharma

Name of supervisor: Prof. Martin Blunt

Name of the company supervisor: Thomas Dombrowsky, Schlumberger

Acknowledgements

I would like to express my sincere gratitude and acknowledgment to my project supervisors Thomas Dombrowsky (Schlumberger) and Prof. Martin Blunt (Imperial College) for their invaluable support, supervision, encouragement and helpful advice throughout the research work. I gratefully acknowledge Thomas Dombrowsky for his mentoring, exceptional support and advice during the course of research.

Special thanks to Abingdon Technology Centre, Schlumberger Oilfield UK Plc, for providing an excellent environment for my research and for generously providing all the necessary resources.

I wish to thank my parents for their constant encouragement, prayers and support throughout the whole MSc program which made it all possible. I would also like to thank my friends, especially Syed Abdul Samad Ali for his constant support and help throughout my project.

I also wish to thank the amazing people at the Abingdon Technology Centre for their valuable knowledge sharing, support and insightful discussions during the course of this research project. I am grateful to Thomas Dombrowsky for spending his valuable time with me and answering all my questions.

Table of Contents

DECLARATION OF OWN WORK	ii
Acknowledgements.....	iii
Table of Contents.....	iv
List of Figures.....	iv
List of Tables	iv
Abstract.....	1
Introduction	1
Model Description	2
Linear Principal Component Analysis – Dimension Reduction	3
Basis.....	4
Covariance & Eigen-Decomposition	4
Choosing Components	5
Reparameterization	6
Search the feature space.....	6
History Matching	7
Summary.....	8
Conclusions & Recommendations	9
Nomenclature.....	10
References	10
APPENDIX A.....	11
APPENDIX B	17
APPENDIX C	19

List of Figures

Figure 1 Aerial view of the synthetic test model.	3
Figure 2 Dimension reduction as shown by change in the co-ordinate system from <i>model space</i> to <i>feature space</i>	4
Figure 3 Loss of information. On the left is for Brugge and on the right is the curve for the synthetic test model.	5
Figure 4 History match for individual wells. On the left is the best history matched well (P302), on the right is the one of the bad history matched well (P301). The green line denotes the history matched results, while the black points denotes the observed well data.	8
Figure 5 Methodology employed in PCA for history matching.	9
Figure 6 Petrophysical Modelling Process	17
Figure 7 Workflow in Petrel to create prior.	18
Figure 8 Aerial view of the porosity distribution; on the left is one of the prior model and on the right is the history matched model.	19
Figure 9 Aerial view of the permeability (in k direction) distribution; on the left is one of the prior model and on the right is the history matched model.	19
Figure 10 Match for the field oil production rate. Black points denotes the observed data, green line is the simulated result.	20
Figure 11 Match for Water Production Rate. Black points denotes the observed data, green line is the simulated result.	20
Figure 12 Match for field oil production total. Black points denotes the observed data, green line is the simulated result.	21
Figure 13 Match for field water production total. Black points denotes the observed data, green line is the simulated result.	21

List of Tables

Table 1 Overview of the reservoir simulation parameters.	2
Table 2 Milestones in History Matching using PCA	11

History Matching with Principal Component Analysis

Akshay Sharma

Imperial College supervisor – Prof. Martin Blunt

Company supervisor – Thomas Dombrowsky, Schlumberger Abingdon Technology Center

Abstract

In the oil and gas industry, investment decisions are made based on predictions of reservoir performance. The accuracy of such predictions depends on the model being able to simulate past production behaviour. This is known as the history matching process, wherein the tuning of geostatistical parameters are undertaken so as to arrive at a model which can simulate the observed dynamic data available from the wells. Traditionally this process is done by trial and error, where the uncertain parameters are varied until a satisfactory match is obtained. But this process is highly time consuming and tedious and since the history matched model is non-unique the resultant match may give poor predictions of future production. Therefore, in recent years studies have been conducted to test the applicability of stochastic algorithms, gradient-based methods, streamline based methods and Kalman filter approaches to the history matching process. In this study one such approach, Linear Principal Component Analysis, has been discussed.

Linear Principal Component Analysis (PCA) is a stochastic approach which has been used in face recognition and image compression, and is a common technique for finding patterns in data of high dimension. The central idea of PCA is to extract the most important information from the data-set and compress the size of the data-set, without much loss of information (Jolliffe). This is achieved by using linear transformations that preserve the degree of freedom that describe the largest variation and eliminates those that describe small variations. The retained variables, known as principal components, provide a simple and efficient means for imposing the geological constraints. The resulting history matched model can then be expected to provide more accurate forecasts about future predictions.

The PCA approach is applied to a geologically complex reservoir, with porosity, permeability, net-to-gross and fault transmissibility multipliers as the uncertain parameters. The report discusses how PCA is employed to create a model representation in the reduced dimensional parameter space; this re-parameterized model is history matched using an optimization algorithm.

This study has provided insights into the workings of PCA for the history matching process. PCA holds much promise to become a standalone tool for the history matching process, but this requires future investigation into the bottlenecks. Primary of these is to better understand the relationship between the choice of priors and loss in information during dimension reduction, and to update the framework of the optimization algorithm to perform robustly and efficiently.

Introduction

The history matching process is an inverse problem to find a model that closely matches the historical reservoir performance. It is a key component of the closed-loop reservoir management, with the idea of maximizing the reservoir performance. In this process real-time surface and downhole production data provide continuous input to the history matching algorithm; and this calls for automated history matching techniques. Although several approaches have been investigated actively - stochastic algorithms (Hu et al. 2001; Caers 2003), gradient-based methods (Chen et al. 1974; Chavent et al. 1975), streamline based methods (Agarwal and Blunt 2003; Cheng et al. 2004) and Kalman filter approaches (Geir et al. 2003; Gao et al. 2005), in practice the history matching is still often done manually. This suggests that new developments are still required to better understand the need and provide a robust, efficient approach for this problem. One approach that has received a lot of attention is Ensemble Kalman filtering. A related method, Linear Principal Component Analysis (PCA) (Gavalas et al. 1976; Sarma et al. 2006), also uses an ensemble representation and is used in this study for the purpose of history matching.

Often in the history matching process, the geological constraints are usually ignored due to lack of an efficient mechanism for imposing them. This results in models that does not have a good estimation of these parameters, and as such

will lead to discrepancy in the forecast for the performance of the reservoir. Linear Principal Component Analysis (PCA) approach provides an efficient way of imposing these constraints, under a linear approximation, when the uncertainty in prior model is expressed as a collection of possible geological realizations. The result is a geologically constrained history matched model, on which more reliable predictions of the future performance can be made.

The PCA methodology had already been applied to the Brugge test case (Peters et al. 2010), and the results achieved with this had the lowest estimation error when run against the truth model (Dombrowsky et al. 2011). In the current state the algorithm takes porosity, permeability and net-to-gross as input parameters. This work deals with expanding the current state of the algorithm to incorporate fault transmissibility multiplier into the workflow, which is subsequently applied to a test model. The test model is highly faulted with 25 year production history, so presents a good complex reservoir for the methodology to be tested.

The report firstly explains the test model; the geostatistical properties are stated with a brief description of the workflow employed to generate the priors. The PCA approach as applied to the reservoir simulation is then explained with detailed description on how the dimension reduction takes place; and the information lost due to this dimension reduction. The optimization algorithm is then run on this re-parameterized model to arrive at the history matched results. The results are analysed in detail with the problem areas in the algorithm identified.

The work was conducted using Petrel 2011.1; the PCA algorithm was prototyped using Ocean and Matlab. Eclipse 2011.1 was used for reservoir simulation.

Model Description

The test model is a synthetic in-house reservoir model with a 25 year of production data. The model is highly faulted with 23 faults in total. The model is a rhombus shaped reservoir with aquifers at the western, eastern and southern edges of the reservoir. The reservoir is an undersaturated oil reservoir, so a two-phase simulation was sufficient. The original high-resolution model consists of 2.5 million cells (325 x 261 x 29). The upscaled model has 55104 cells (48 x 41 x 28). The upscaling was performed arithmetically for porosity and with directional averaging for a diagonal tensor for permeability. The reservoir has 62 wells (56 producers and 6 injectors); some producers were turned into injectors during the life of the reservoir. The fluid and rock properties are described in the Table 1 and the model with the well location is shown in Figure 1.

Overview of the reservoir parameters		
Initialization	4278 psi at 3000 m depth	
Pore Compressibility	$4.5 \times 10^{-6} \text{ psi}^{-1}$	
Density (lbm/ft ³)	Water: 63.7	Oil: 50
Compressibility (psi ⁻¹)	Water: 2.8×10^{-6}	Oil: 9.3×10^{-6}
Viscosity (cp)	Water: 0.4	Oil: 0.6
Well Constraints	Producers Minimum BHP = 1160 psi	
	Injectors Maximum BHP = 5365 psi	

Table 1 Overview of the reservoir simulation parameters.

For the purpose of using PCA on the model for dimension reduction, realizations had to be generated with different values of geostatistical parameters (porosity, permeability and net-to-gross). These realizations describe the uncertainty in the geological model as wells as define the constraints which are to be applied with respect to these parameters. To generate these realizations two-point geostatistical (Sequential Gaussian Simulation) approach, which generates parameters with a Gaussian distribution; was employed. Various approaches had to be tried to generate these realizations as they directly affect the retained information after dimension reduction. Prior models (realizations) were created with different values for the geostatistical parameters, which were randomly generated using a Monte Carlo Simulator. A brief summary of the methodology of how these realizations are created in a Petrel workflow is explained in the Appendix B. 100 realizations were generated for this particular model. These realizations, together with the production history for the wells (BHP, oil and water production rates, and water injection rates), form the input for the history matching process.

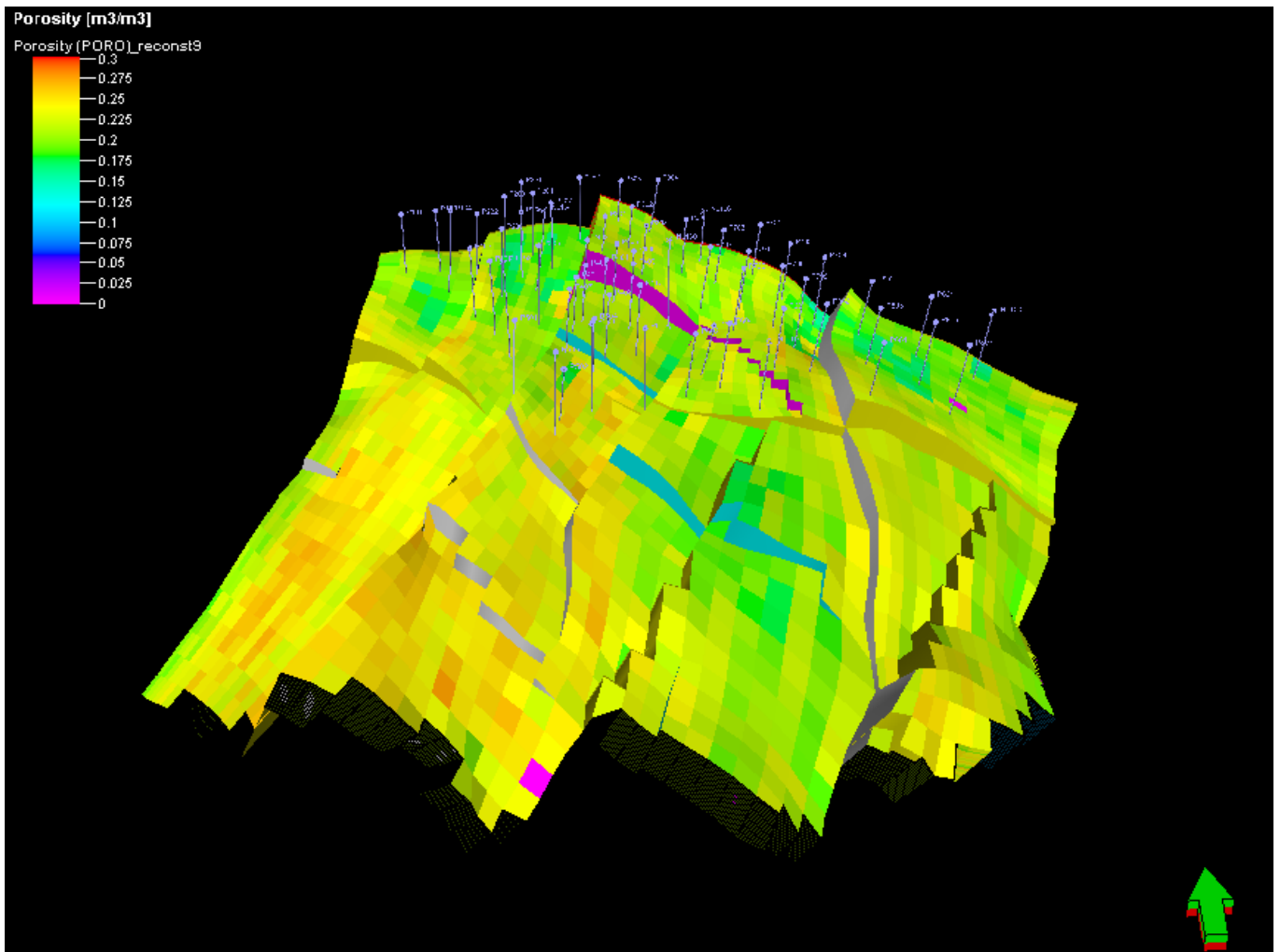


Figure 1 Aerial view of the synthetic test model.

Linear Principal Component Analysis – Dimension Reduction

Linear Principal Component Analysis (PCA) is a statistical technique that allows us to analyze interdependencies in data sets with a large number of variables and describe these variances using a smaller number of ‘latent variables’ known as Principal Components. The uncertainty in the 3D geological model arises from the different values the geostatistical parameters can take consistent with the data (logs, well test). These uncertainties are described by generating different realizations of the model known as *priors*. The latent variables constructed are uncorrelated with unit-normal distributions. The space of latent variables is known as *feature space*, while the original variables lie in the *model space*. New models are generated by mapping these latent variables back into the model space. This section will mainly deal with the dimension reduction as applied to the reservoir model, through the linear mapping between the model and feature space.

To start with, the reservoir model can be thought of in terms of geometrical concepts. A model is input to the simulator, which can be expressed as a vector of properties (porosity, permeability, net-to-gross). Every cell of the model will take a unique value of the parameter; and with different realizations generated to represent the uncertainty, this vector can be considered as point in an extremely-high dimensional space (usually of the order of thousands or millions). The space in which this point resides consists of all possible inputs and is known as the *model space*. The priors, when mapped on to this space, form a cloud of points which in the case of Gaussian distribution is in the shape of a multi-dimensional ellipsoid. PCA expresses the uncertainty in terms of a new coordinate system whose axes are aligned to the major axis of the ellipsoid, while the direction associated with the minor axis are eliminated, thereby reducing the dimensionality. All of the models in the original space thus lie closer in the reduced space. The degree of closeness is controlled by the variance threshold that defines which dimensions are eliminated. Figure 2 below shows the dimension reduction after the eigenvectors with lesser significance have been ignored; this process has been explained later on.

The basic methodology of PCA as explained below is representing the features of the input data set which is of higher dimensionality into a lower dimensionality space, on which the optimization algorithms can be run robustly to look for optimum values of the uncertain parameters. When these values are found they are mapped back on to the original model space and simulation is then run, resulting in the history matched case. PCA as applied to the history matching process only deals with finding correlations among the input data set such that the outputs that we get are bound by the constraints imposed on by the input parameters.

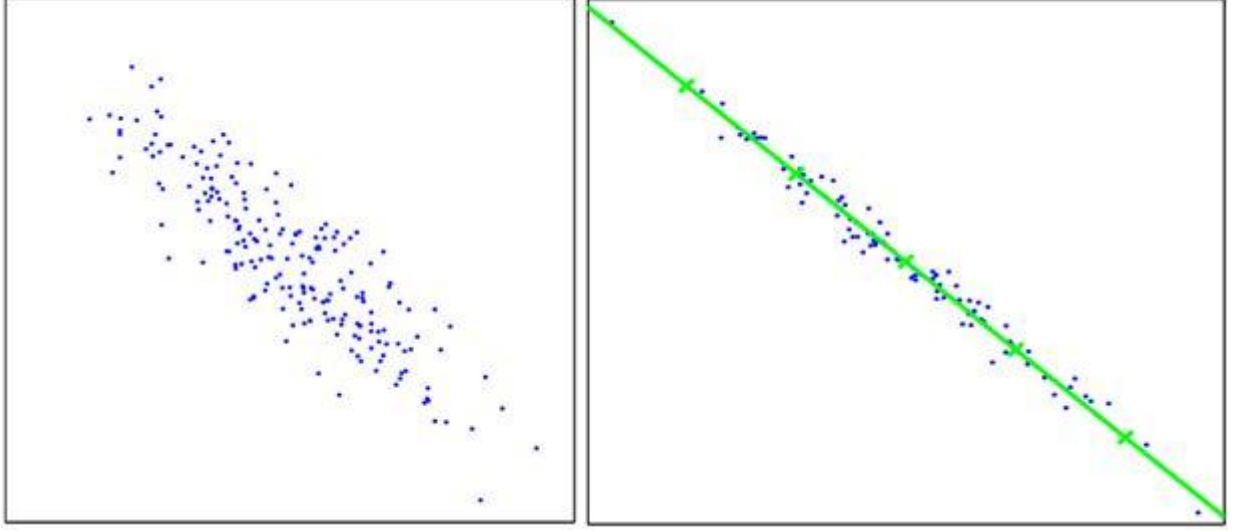


Figure 2 Dimension reduction as shown by change in the co-ordinate system from *model space* to *feature space*.

There are several methods to calculate the principal components, explained below is one of them which has been employed in this study.

- I. **Basis** - The goal of PCA is to create the most meaningful basis to re-express a noisy data set. The model has 55104 active cells, and taking 5 parameters (porosity ϕ ; permeability k_x , k_y , k_z ; net-to-gross) gives 275520 dimensions (M); the vector Y denotes a set of parameters for one cell. This represents a vector in a 5 dimensional space. So to represent 1 model we will need 55104 such vectors.

$$P = \begin{pmatrix} k_x \\ k_y \\ k_z \\ \phi \\ ntg \end{pmatrix} ; \quad Z = \begin{pmatrix} P_1 \\ \vdots \\ P_{55104} \end{pmatrix}$$

This represents one model which is a vector in 275520 dimensions. A total of 100 (N) prior realizations have been generated. So finally our data set becomes,

$$X = [Z_1 \quad \cdots \quad Z_{100}]$$

These are represented in matrix form X (M x N), where each column represent one realization of the model. Mean-adjusted data set (B) is calculated by subtracting the mean vector μ from X. Basically μ is the column matrix of the mean of all rows (thus the mean of parameter of respective cells across all 100 priors). Where h is 1 x N row vector of all 1's.

$$\mu = \frac{1}{N} \sum_{i=1}^N Z_i ; \quad B = X - \mu h$$

- II. **Covariance & Eigen-Decomposition** - The covariance measures the degree of linear relationship between two variables. A large (small) value indicates high (low) redundancy. The covariance matrix is calculated as,

$$C = \frac{1}{N} \sum B \cdot B^T$$

The covariance matrix (C) generated is a square matrix (M x M), which is also symmetric. This matrix characterises the scatter of the data. The eigen decomposition performed on the covariance matrix results in orthonormal eigenvectors (of length 1) and non-negative eigenvalues,

$$CV = DV \quad ; \quad C = VDV^T$$

where V is a matrix, with each column representing an eigenvector v_i and D is a diagonal matrix of eigenvalues λ_i (where the diagonal elements are the eigenvalues and all other elements are 0). The vectors v_i are the principal components of X. Besides being perpendicular to each other, the eigenvector also provide us with information about the patterns in the data. There is a direct relationship between eigenvalues and eigenvectors; the eigenvector associated with the highest eigenvalue denotes the largest variance in the dataset and so on.

III. Choosing Components - The next step is to rearrange the matrix V and D in decreasing order of eigenvalue; resulting in components in order of significance. So the first principal component will have the largest possible variance; the second component besides being orthogonal to the first also has the largest possible inertia. Here arises the matter of dimension reduction, in that components with lesser significance can be completely ignored. Assuming L eigenvalues are retained (since $L \leq \min(M,N)$ there is a drastic dimension reduction), this results in size of matrices D and V being L x L and M x L respectively. The error introduced by this reduction in dimensions can be calculated as,

$$q = \sum_{i=1}^L D_i / \sum_{i=1}^M D_i$$

This error; also known as ‘*loss of information*’; very much depends on the model and the choice of priors. As shown in the Figure 3 below, for the Brugge Model, with N=104 priors and M=300240 dimensions, 70% of the information was retained by only 40 parameters (Dombrowsky et al. 2011). But for the model in this study only 50% of information was retained with 40 parameters and 70% with 60 parameters, shown on the right. This is one area which warrants further research to better understand the relationship between the choice of prior and loss in information during dimension reduction.

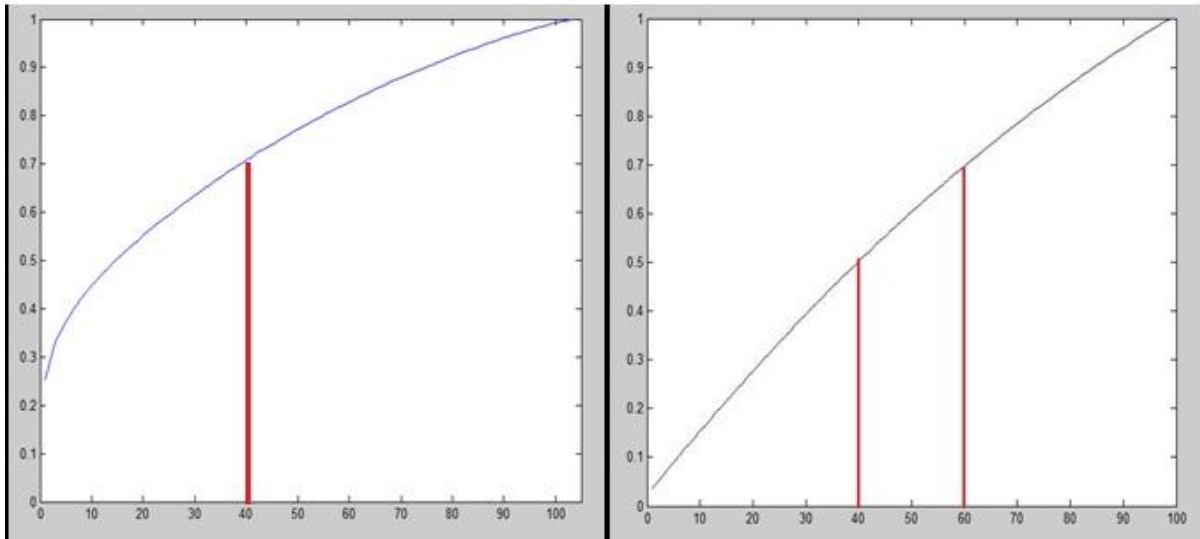


Figure 3 Loss of information. On the left is for Brugge and on the right is the curve for the synthetic test model.

The number of parameters to be retained represents a trade-off between reducing the dimensionality and minimizing the error introduced by this reduction. As this number increases, the dimensionality of the feature space increases. This will result in a significant increase in the number of iteration runs required for the history matched model to be arrived at; which will subsequently result in an increase in optimization time. Though it is expected that a more refined model will result in a better history match, this is not necessarily the case in practice.

- IV. **Reparameterization** - This is the final step in PCA. Once the dimension reduction has been established, the points are transformed to a feature space,

$$Y = V^T B$$

where Y contains the feature space co-ordinate.

To normalize the data-set the eigenvalues are used as weighting factors; this is known as a change of basis and y can be thought of as a co-ordinate vector in the lower dimensional feature space.

$$Y = \frac{1}{\sqrt{D}} V^T B$$

This will give us the original data solely in terms of the vectors retained. By doing this the data is transformed so that it is expressed in terms of the patterns between them.

- V. **Search the feature space** – An optimization algorithm for example NeuralNet, Nelder-Mead is now used in this reduced feature space to search for best fit of parameters which also adheres to the production data. It is at the optimization stage that the fault transmissibility multipliers are added as parameters to be optimized. The PCA workflow is applied to the model with 275520 dimensions for dimension reduction to 40 principal components. These principal components in addition to the 24 fault transmissibility multipliers (FTM) forms the input to the optimizer algorithm, which performs a search of the feature space for possible values of porosity, permeability, net-to-gross and FTM. This was done so that more accessibility can be given to the user with the choice of parameters that he would want to incorporate. Besides fault transmissibility the size of aquifers and throw of faults could be optimized this way. This area merits further investigation.

To map the points from the feature to the model space and hence back to a simulation model which can be simulated forward in time, we use the transform,

$$x = \sqrt{D} V y + \mu$$

where y is a point in feature space and x is a point in model space.

Through this process the goal of extracting the important information from the data set and then compressing the size of the data set by keeping only the important information is achieved. One important feature of PCA analysis is the method used in generating the priors. Modelling of property values throughout the grid cells is done by Sequential Gaussian simulation, which is a stochastic method of interpolation based on Kriging. During simulation (here we mean geostatistics simulation not flow simulation), local highs and lows will be generated between input data locations (between cells with well log values). Each of the nodes will be visited in random order. At each node the data will be kriged to determine the variance at that node, and then a value will be picked from the input distribution to match the variance at that node. As subsequent cells are visited, the previously defined cell values are used in the kriging. As a result of this method, the last cells to be defined are heavily constrained by the distribution of the defined cells. Therefore, it is important that cells are visited in truly random order and that localized areas of cells are not visited together and that upscaled well logs are used as input data. Also the position of the local highs and lows are determined by a random value, it is because of this that a high number of priors are generated to gain a better understanding of uncertainty.

PCA provides an efficient algorithm for dimension reduction of variables with multivariate Gaussian distribution. For other distributions, such as those of the multi-point geostatistical method, PCA results in models with feature that appear smoother than those from the true distribution. This is because PCA generates new models from linear combination of models from the original sample. The standard two-point geostatistical approach (Sequential Gaussian Simulation) are used while generating priors as they give rise to samples that have Gaussian distribution, allowing PCA to be employed on the model. However, these distributions have a limitation as in they cannot model complex features such as fractures and channels (Sarma et al. 2008), which are characterized by multipoint geostatistical algorithms, which expresses the joint variability at many more than two locations at a time. This limitation can be addressed by using Kernel-PCA which is a non-linear form of PCA. The applicability of this non-linear PCA approach is more suited for describing the complex geological features, and thus can be investigated further in the future.

History Matching

In the previous sections, the concept of PCA has been explained in the context of reservoir simulation. Linear PCA is explained and employed on the model for dimension reduction, defining a mapping to and from a low-dimensional feature space. It is easy to generate new points in this feature space which give rise to models in or near the region of interest and are consistent with the geological constraints. The test model was used to demonstrate how a problem with 100 prior models and more than 250,000 variables can be reduced to only 40 parameters without significant loss of information. In this section the history matching of the model is discussed. This involves an automated search of the feature space. It has been mentioned that the feature space is continuous, which means that gradient based optimization methods can be employed (even if prior models are non-continuous) (Sarma et al. 2006). Amoeba and NeuralNet which are gradient-free algorithms were used in this study (Bailey and Couët 2005).

The 40 coefficients of the feature space vector and the 24 FTM are the control parameters that are altered by the optimizer. These alterations affect the porosity, permeability, net-to-gross parameters as well as the fault transmissibility multiplier of the simulation model. With the reduced dimensions the optimization algorithm is employed in the feature space to search for models which are geologically constrained as well as fit the observed dynamic data from the wells. And as such the objective function should be able to encapsulate these features. The objective function used in the optimization is the difference between simulated and observed data (in a root mean square sense), which can be interpreted as proportional to the log of the likelihood probability density function of the model parameters. It has the form,

$$obj = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{sim_i - obs_i}{\sigma_i} \right)^2}$$

where obs_i denotes the observed rate or pressure of a well at a given time and sim_i denotes the equivalent simulation result (these are functionally dependent on the input configuration \mathbf{y}). By minimizing the objective function, we find the set of model parameters such that the discrepancy between observations and simulated results is minimal. This constitutes the maximum-likelihood (MLE) solution for the model parameters. The observed data consists of Bottom Hole Pressure for 62 wells as well as oil and water production rate for 56 wells, observed for a 25 year period on a roughly monthly basis ($n=61944$). The values of σ_i were set as constant for each measurement, with the constant chosen to normalize the measurements to have the same range of values.

Amoeba (Nelder-Mead) works on the concept of simplex, which is a special polytope of $N+1$ vertices in N dimensions. So in this case, the polytope is of 64 vertices (40 Principal Components + 24 FTM). Each iteration generates a new vertex for the simplex. The algorithm then chooses to replace one of these vertices with the new vertex; these co-ordinates are then mapped into the model space, where the resulting simulation model can be run; and so the technique progresses. This way the diameter of the simplex gets smaller and the algorithm stops when the optimum value of the objective function has reached. This derivative free optimization algorithm was derived from Aurum Optimization Library (Bailey and Couët 2005). Since the 40 principal components in this case is a representation of the input data in lower dimensionality, it encapsulates the constraints to the geological data; this mapping limits the search space to geologically constrained models, while the objective function encompasses the observed production behaviour. The optimization is a bottleneck of this process as multiple simulations cannot be run in parallel.

The above history matching process can be stated in two broad categories. The first is where the linear PCA is applied to identify the subspace spanned by the geologically feasible model; creating a mapping to the feature space which allows to discard irrelevant dimensions. The second is the optimization which will search for the model parameters that minimizes the differences between the observed and simulated data.

The history matched results for selected wells are shown in Figure 4 (Appendix C). It is evident from the results that the algorithm performs robustly. The history match results are able to simulate the observed production behaviour (as evident from the low mismatch of the objective function), at the same time the geological realism of the model is maintained.

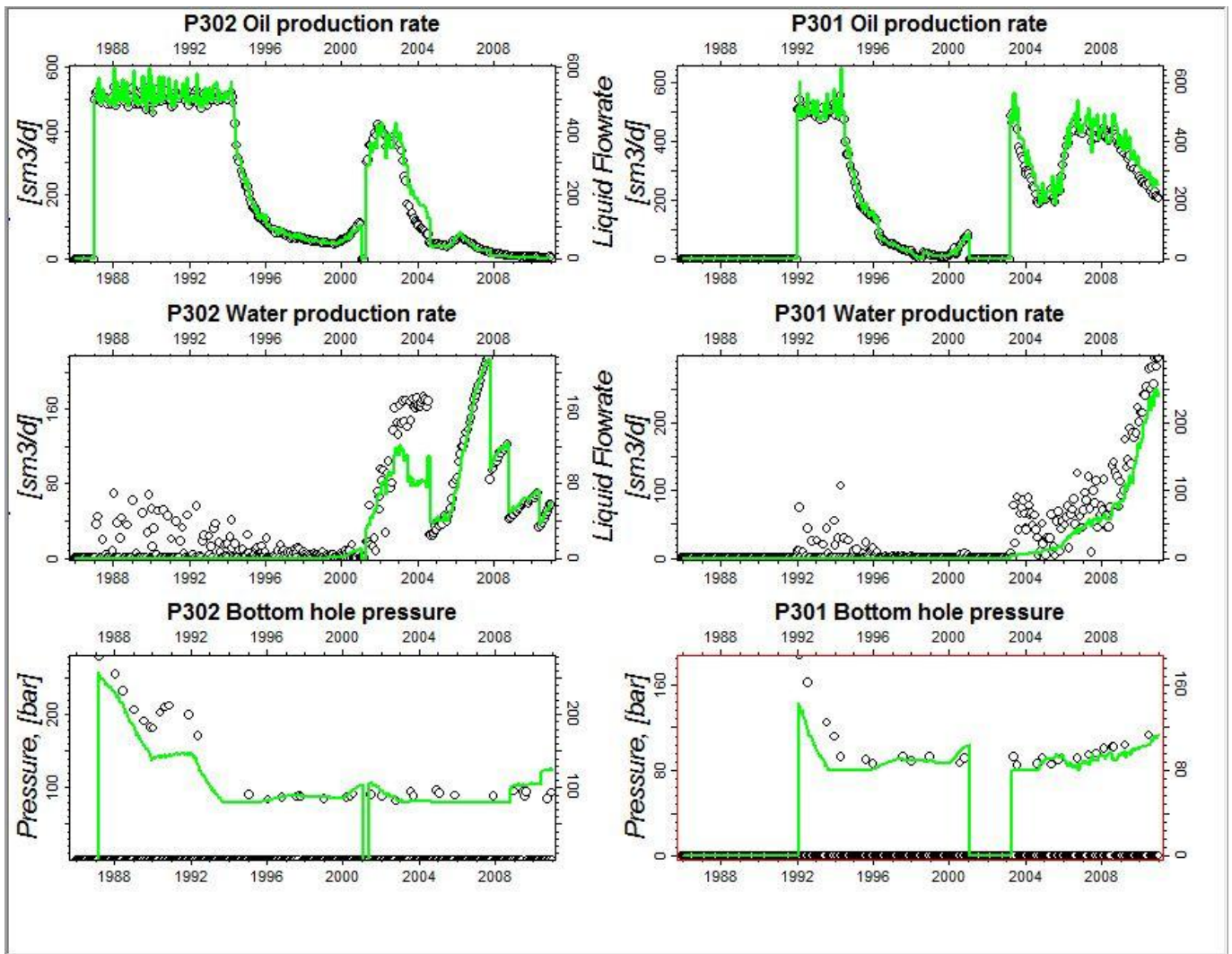


Figure 4 History match for individual wells. On the left is the best history matched well (P302), on the right is the one of the bad history matched wells (P301). The green line denotes the history matched results, while the black points denotes the observed well data.

Summary

In this article, linear PCA has been explained as applied to the history matching process. The algorithm has been successfully tested on a synthetic test model with 25 year production history. Since the prior models were generated using Sequential Gaussian Simulation (Gaussian distribution of the parameters); linear PCA could be applied to reduce the dimensionality of the problem. The analysis took into account uncertain parameters porosity, permeability, net-to-gross and fault transmissibility multipliers and generated model with values for these parameters which are consistent with the geological data. The PCA was able to reproduce the model in the reduced feature space, but the history match was not perfect. The prime reason for this is the objective function in use which was only optimizing BHP and all other mismatches get worse. This was due to the large size of the BHP component as compared to others (when the unit conversion takes place for the objective function calculation). In this example, the BHP mismatch was of the order X, oil and water rate was of the order Y, which means that improvements in BHP dominated the overall objective. This is an artefact of the chosen unit system (SI) and can be rectified by using the Petrel internal define objective function (DOF) process. Though this in itself had problems of its own as the internal coding of this process is flawed; this is being looked into further. Also the optimization runs have not converged, which means that a potentially sub-optimal solution has been found. It might be possible that these runs would have improved given further simulation runs. But since the time between the simulation runs were increasing as the iteration number was increasing, there was no way of discerning when it would have converged.

PCA is an algorithm that explains a data-set (geostatistical parameters of the simulation model in this case) in a high dimensional space reduced to a low-dimensional feature space, this makes it feasible for automated optimization algorithms to be used in the new model parameterization. The linear mapping from model to feature space limits the search space to geologically reasonable models and together with the objective functions result in models which satisfy both the observed behaviour of the field as well as the geological data. This makes it possible to generate more reliable forecasts of future

production. Since model updating is an essential feature of “real-time” reservoir management, this study has shown that PCA approach has the potential to be able to fulfil this role.

But this requires further investigation, especially the extension of linear-PCA to Kernel-PCA. Since linear-PCA works on linear mapping for model reduction, it can only be applicable to parameters which have Gaussian distribution (achieved by Sequential Gaussian Simulation, Kriging). Non-Gaussian distribution which are used to model complex features like channels can be used with linear-PCA, but this will result in features being smoothed out. This is where Kernel-PCA can be employed since it works on the non-linear mapping into the feature space. However, the efficiency and robustness of Kernel-PCA need to be investigated first.

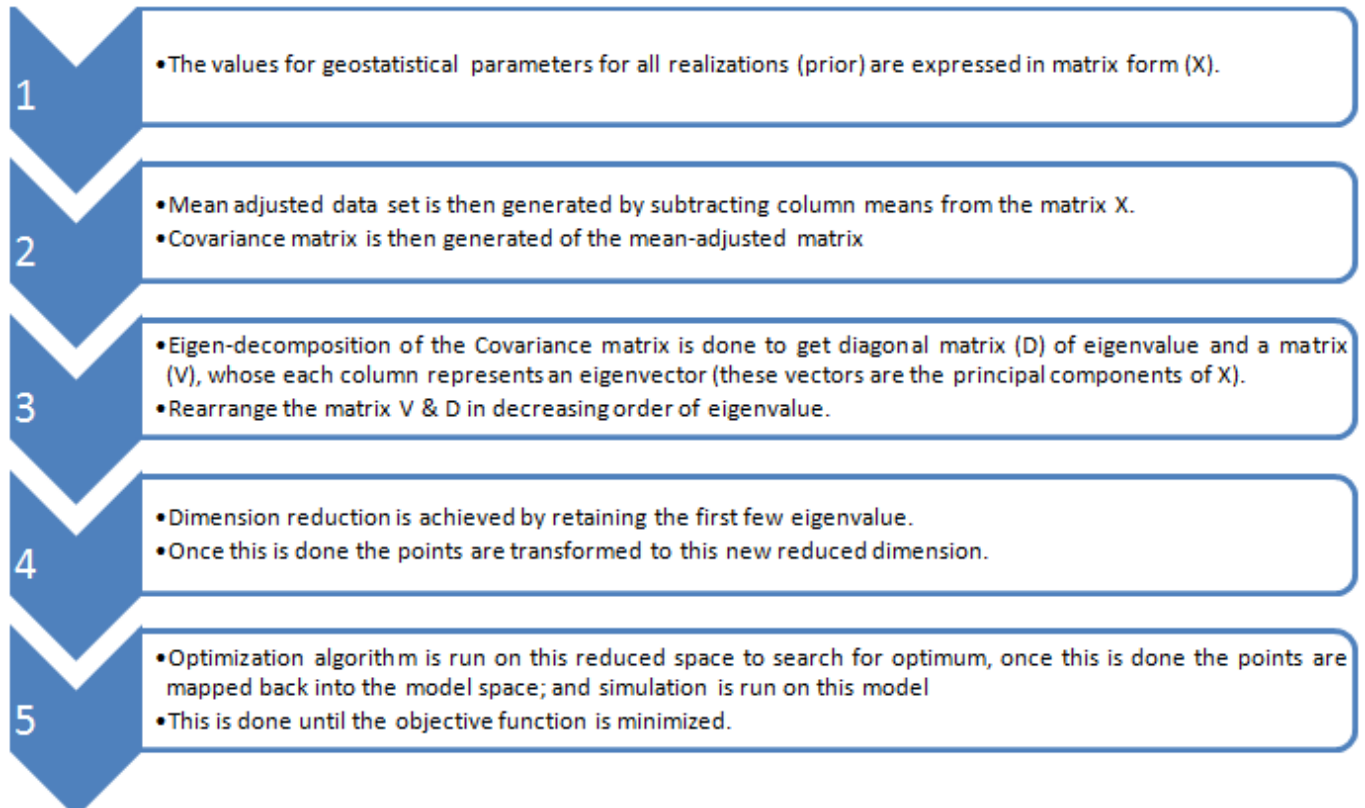


Figure 5 Methodology employed in PCA for history matching.

Conclusions & Recommendations

The history matching of the model with the updated algorithm was successfully done. It is evident from the results that the PCA algorithm performs robustly. Furthermore, the geological formations form statistically independent subsets of parameters, so the geological realism is preserved. However, conducting a history match does not, in itself, provide any real indication about the *predictive* quality of the model; it only minimizes the difference between the observed and the simulated data. Several bottlenecks have been encountered in the study which needs to be understood in more detail. The recommendations for further research based on the study are summarized below:

1. The relationship between the choice of prior and loss in information during dimension reduction needs to be investigated in more detail.
2. The uncertainty estimation of the history matched model has to be looked. As this will directly tie in with the forecast estimation, the uncertainty needs to be addressed and explored.
3. Forecast optimization, as to how accurate the prediction from the history matched model are, need to be addressed.
4. The optimization algorithm run needs to be more efficient and robust. Since the time only permitted for the testing of two algorithms (Nelder-Mead and NeuralNet); more algorithms can be tried and tested. The framework for these also needs to be updated based on the time results from this study (this is being already undertaken at Schlumberger USA).
5. An extension of Linear-PCA; Kernel PCA can be studied in more detail with focus on how to tie up with linear-PCA.
6. Advanced PCA methods like using PCA in conjunction with Ensemble Kalman filter can be explored as this will address the problem posed by non-Gaussian distributions.
7. More parameters like fault throw; size of aquifer can be incorporated into the analysis and applied to test models.

Nomenclature

M	No. of dimensions ($55104 * 5 = 277520$)
N	No. of Priors (100)
X	M * N Matrix of the data set (Geostatistical Parameters)
μ	M * 1 Matrix of means
B	M * N Matrix of the mean adjusted data set
C	M * M Covariance Matrix
λ_i, D	Eigenvalue; Diagonal Matrix of Eigenvalues
v_i, V	Eigenvector; Matrix with each column representing 1 eigenvector
Y	L * N Matrix of the points in the reduced dimension (feature space)
\emptyset	Porosity
k_x, k_y, k_z	Permeability

References

- Agarwal, B. and M. Blunt. "Streamline-based method with full-physics forward simulation for history-matching performance data of a North Sea field." *SPE Journal* **8**(2) (2003): 171.
- Bailey, W. and B. Couët. "Field Optimization Tool for Maximizing Asset Value." *SPE Reservoir Evaluation & Engineering* **8**(1) (2005): 7.
- Caers, J. "History matching under training-image-based geological model constraints." *SPE Journal* **8**(3) (2003): 218.
- Chavent, G., Dupuy, M., Lemonnier, P. "History matching by use of optimal theory." *SPE Journal* **15**(1) (1975): 74.
- Chen, W., Gavalas, G.R., Seinfeld, J.H., Wasserman, M.L. "A new algorithm for automatic history matching." *SPE Journal* **14**(6) (1974): 593.
- Cheng, H., Xian-Huan Wen, William J. Milliken, Akhil Datta-Gupta. "Field experiences with assisted and automatic history matching using streamline models". SPE paper 89857 presented at the *SPE Annual Technical Conference and Exhibition*, Houston, TX, (2004).
- Dombrowsky, T. D., Michael D. Prange, William J. Bailey. "Geologically Constrained History Matching with Principal Component Analysis." Submitted for Publication (2011).
- ECLIPSE Reference Manual and ECLIPSE Technical Description, Copyright 2011, Schlumberger
- Gao, G., Zafari, M., Reynolds, A. C. "Quantifying uncertainty for the PUNQ-S3 problem in a Bayesian setting with RML and EnKF". SPE paper 93324 presented in the *SPE Reservoir Simulation Symposium*, Houston, TX, (2005).
- Gavalas, G., Shah, P.C., Seinfeld, J.H. "Reservoir history matching by Bayesian estimation." *SPE Journal* **16**(6) (1976): 337.
- Geir, N., Liv Merethe Johnsen, Sigurd Ivar Aanonsen, Erlend H. Vefring. "Reservoir Monitoring and Continuous Model Updating Using Ensemble Kalman Filter". SPE paper 84372 presented at the *SPE Annual Technical Conference and Exhibitions*, Denver, CO, (2003).
- Hu, L. Y., Blanc, G., Noetinger, B. "Gradual deformation and iterative calibration of sequential stochastic simulations." *Mathematical geology* **33**(4) (2001): 475.
- Jolliffe, I. Principal component analysis. 1986. Second Edition, Springer, New York. 1.
- Peters, L., Arts, R.J., Brouwer, G.K., Geel, C.R.Cullick, S., Lorentzen, R.J., Chen, Y., Dunlop, K.N.B., Vossepoel, F.C., Xu, R., Sarma, P., Alhuthali, A.H., Reynolds, A.C. "Results of the Brugge benchmark study for flooding optimization and history matching." *SPE Reservoir Evaluation & Engineering* **13**(3) (2010): 391.
- Petrel RE 2011.1, Copyright Schlumberger, 2011.
- Sarma, P., Durlofsky, L., Aziz, K., Chen, W. "Kernel principal component analysis for efficient, differentiable parameterization of multipoint geostatistics." *Mathematical Geosciences* **40**(1) (2008): 3.
- Sarma, P., Durlofsky, L., Aziz, K. "Efficient real-time reservoir management using adjoint-based optimal control and model updating." *Computational Geosciences* **10**(1) (2006): 3.

APPENDIX A

Critical Literature Review

Table 2 Milestones in History Matching using PCA

Paper No.	Year	Title	Authors	Contribution
SPE 119094	2010	Results of the Brugge Benchmark Study for Flooding Optimization and History Matching	E.Peters, R.J.Arts, G.K.Brouwer, C.R.Geel, S.Cullick, R.J.Lorentzen, Y.Chen, K.N.B.Dunlop, F.C.Vossepoel, R.Xu, P.Sarma, A.H.Alhutali, A.C. Reynolds	Published results of all competing companies with their respective history matching methods.
SPE 106176	2007	A New Approach to Automatic History Matching Using Kernel PCA	Pallav Sarma, Louis J. Durlofsky, Khalid Aziz, Wen H. Chen	First to apply Kernel PCA to history matching with results that were superior where data was scarce.
Paper submitted for publication	2011	Geologically constrained History Matching with PCA	Thomas P Dombrowsky, Michael D Prange, William J Bailey	Introduced the concept of Linear-PCA, its application to Brugge with NPV that was highest compared to SPE 119094.
Computational Geoscience 10:3-36	2006	Efficient real-time reservoir management using adjoint based optimal control and model updating	Pallav Sarma, Louis J. Durlofsky, Khalid Aziz, Wen H. Chen	PCA approach applied to dynamic waterflooding resulting in increase in NPV & sweep efficiency.
Math Geosci 40:3-32	2008	Kernel PCA for Efficient, Differentiable Parameterization of Multipoint Geostatistics	Pallav Sarma, Louis J. Durlofsky, Khalid Aziz	First to use higher order Kernel PCA.

SPE 119094 (2010)**Results of the Brugge Benchmark Study for Flooding Optimization and History Matching**

Author: L. Peters; R.J. Arts; G.K. Brouwer and C.R. Geel; S. Cullick; R.J. Lorentzen; Y. Chen; K.N.B. Dunlop, Roxar; F.C. Vossepoel; R. Xu; P. Sarma; A.H. Alhuthali; and A.C. Reynolds

Contribution to the understanding of history matching:

Summarized all the available mathematical tools for the history matching process with a comparison against the truth model.

Objective of the paper:

To summarize the results of the benchmark project to test different mathematical tools employed by the participating companies to the history matching process and optimal waterflooding strategy.

Methodology used:

The results of the history matched models of participants, with a 20 year forecast optimization were compared with the truth case (TNO results).

Conclusion reached:

1. By estimating the petrophysical properties, ensemble-based methods (Ensemble Kalman Filter and Ensemble randomized maximum likelihood) resulted in reliable history matched models.
2. The results also showed that NPV achieved with the flooding strategy that was updated after additional production data became available was consistently higher than before the data became available.
3. Increase in NPV by having three control intervals per well instead of one was considerable.

Comments:

The paper presented a concise overview of the methodology employed for history matching. Since the model on which this was done was same for all participants, the results could be compared against the TNO.

SPE 106176 (2007)

A new approach to Automatic History Matching Using Kernel PCA

Author: Pallav Sarma; Louis J. Durlowsky; Khalid Aziz; Wen H. Chen

Contribution to the understanding of history matching:

This paper presented a detailed description of the Kernel-PCA approach to the history matching process; and this methodology applied to synthetic models.

Objective of the paper:

To apply Kernel-PCA parameterization to model permeability fields characterized by the multipoint geostatistics; for the history matching of models.

Methodology used:

A non-linear mapping from model space (R) to feature space (F) results in linear realizations in F; and thus PCA can be applied on this space.

Conclusion reached:

1. Linear PCA may be adequate in cases where there is a large amount of data; higher order kernels give better representation of geology where data is scarce.
2. Kernel PCA approach successfully applied to the history matching of synthetic fields with permeability as the uncertain parameter.

Comments:

The paper presented an extension of linear PCA for the history matching process by changing permeability characterized by the multipoint geostatistics. Though its applicability to other parameters porosity, net-to-gross, FTM and aquifer size were not considered in this case, this methodology can be expanded to incorporate these as uncertain parameters.

Paper submitted for publication (2011)

Geologically constrained History Matching with Principal Component Analysis

Author: Thomas P. Dombrowsky; Michael D. Prange; William J. Bailey

Contribution to the understanding of history matching:

PCA approach applied to the Brugge Model with Permeability, porosity and net-to-gross as the uncertain parameters.

Objective of the paper:

To test the applicability of the PCA approach to geologically constrain the model to simulate the observed dynamic behaviour, thus resulting in a reliable history matched model.

Methodology used:

PCA methodology employed to reduce the dimensionality of the model, on which optimization algorithm can be run, which optimizes the objective function until it gives local minima.

Conclusion reached:

1. The history matched model from this analysis outperformed all other methodology employed in the challenge as indicated by small discrepancy between predicted and achieved non-optimized 20 year forecast NPV.
2. Also the optimized 20 year production forecast strategy created with this PCA history match achieves the highest of all published overall NPV.

Comments:

The paper presents the basis of PCA as employed to the history matching process on which our work is based. The main methodology employed here is taken and expanded to include more parameters and the updated workflow has been tested on the synthetic model.

Math Geosci (2008) 40: 3 - 32

Kernel Principal Component Analysis for Efficient, Differentiable Parameterization of Multipoint Geostatistics

Author: Pallav Sarma; Louis J. Durlofsky; Khalid Aziz

Contribution to the understanding of history matching:

This paper presented a detailed description of the Kernel-PCA approach to the history matching process; and this methodology applied to synthetic models.

Objective of the paper:

To apply a novel approach to create a differentiable parameterization of large scale non-Gaussian random fields that is capable of reproducing complex geological structures like channels.

Methodology used:

It expands on the methodology of PCA by working out a Kernel matrix which is able to parameterize distributions non-Gaussian in nature. A non-linear mapping from model space (R) to feature space (F) results in linear realizations in F; and thus PCA can be applied on this space.

Conclusion reached:

The methodology employed is able to overcome the two major problems of linear-PCA

1. Using high order polynomial Kernels, multipoint geostatistics are preserved thereby recreating complex geological features.
2. Furthermore, the approach requires eigen-decomposition of a small kernel matrix instead of a covariance matrix.

Comments:

The paper presented an extension of linear PCA for the history matching process by changing permeability characterized by the multipoint geostatistics. Though its applicability to other parameters porosity, net-to-gross, FTM and aquifer size were not considered in this case, this methodology can be expanded to incorporate these as uncertain parameters.

Computational Geosciences (2006) 10:3 - 36

Efficient real-time reservoir management using adjoint-based optimal control and model updating

Author: Pallav Sarma; Louis J. Durlowsky; Khalid Aziz; Wen H. Chen

Contribution to the understanding of history matching:

Karhunen-Loeve expansion used in combination with Bayesian inversion theory for efficient history matching, which is one cog of the closed-loop approach employed in this study to provide a substantial improvement in NPV.

Objective of the paper:

To implement real time reservoir management that combines efficient optimal control and model updating algorithms on synthetic models.

Methodology used:

Use of adjoint models to choose new search directions and control strategy; then K-L expansion employed for history matching and control strategy employed to this model to predict future performance.

Conclusion reached:

1. Efficient parameterization of permeability in term of K-L expansion, in combination with Bayesian inversion and adjoint models, provides a history matched model under geological constraints.
2. Since adjoint models were used for both optimization and model updating, many components of the code can be reused, resulting in added efficiency.

Comments:

The paper successfully implemented the closed loop approach which has significant future utilization. Since history matching is an essential feature of this approach, significant improvements are needed to bring this to fruition. One such approach has been discussed here with just permeability as the uncertain parameter.

APPENDIX B

Prior Model Creation

In the Petrel workflow the Uncertainty and Optimization tool allows to create multiple realizations of a model with different values for uncertain parameters. Here, the petrophysical modeling tool, constrains uncertain parameters resulting in models that differ in geological (i.e. rock) properties only. During every iteration, a new prior is created with a unique set of rock properties. The algorithm used for generating these values was **Sequential Gaussian Simulation** which honors well data, input distributions, variograms, and trends. The variogram and distribution are used to create local variations, in rock properties away from input data. As a stochastic simulation, the result is dependent on the input of a random seed. The Figure 6 below shows the tool window used in Petrel workflow. The seed number defines a random number used in simulation, each time the workflow is run a new random seed is chosen which generates a new realization. The major and minor direction ranges define the influence of the well logs away from the cells where these logs intersect. The vertical influence range defines the vertical continuity; larger the range thicker the beds will be.

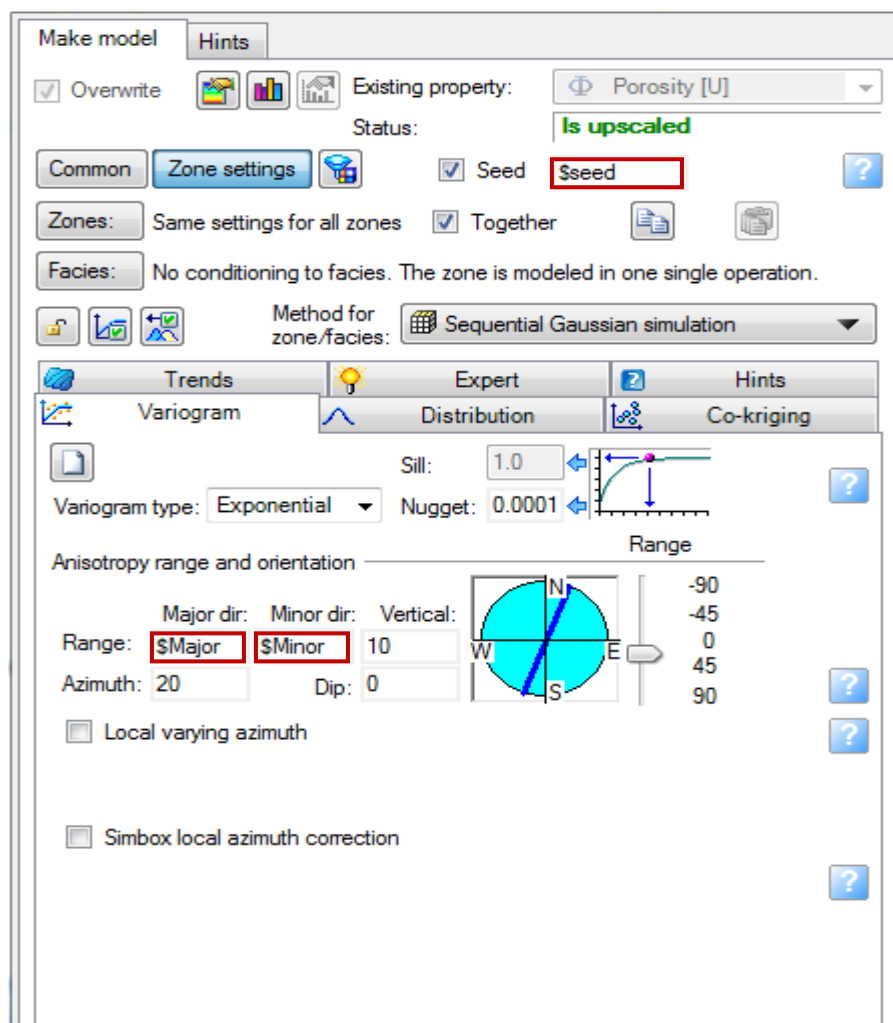


Figure 6 Petrophysical Modelling Process

Here the \$variables (seed, major and minor) are used as the control parameter for a unique value to be generated for the rock properties resulting in 1 realization. The workflow shown in Figure 7 is utilized in Petrel to create these realizations. The Monte-Carlo sampler is used in this workflow; it is a stochastic sampling algorithm which samples the uncertain variables randomly from their assigned distributions. The Latin-hypercube sampling is used as an add-on to it; this option ensures that the sampled values for each parameter are distributed over the entire range of that parameter.

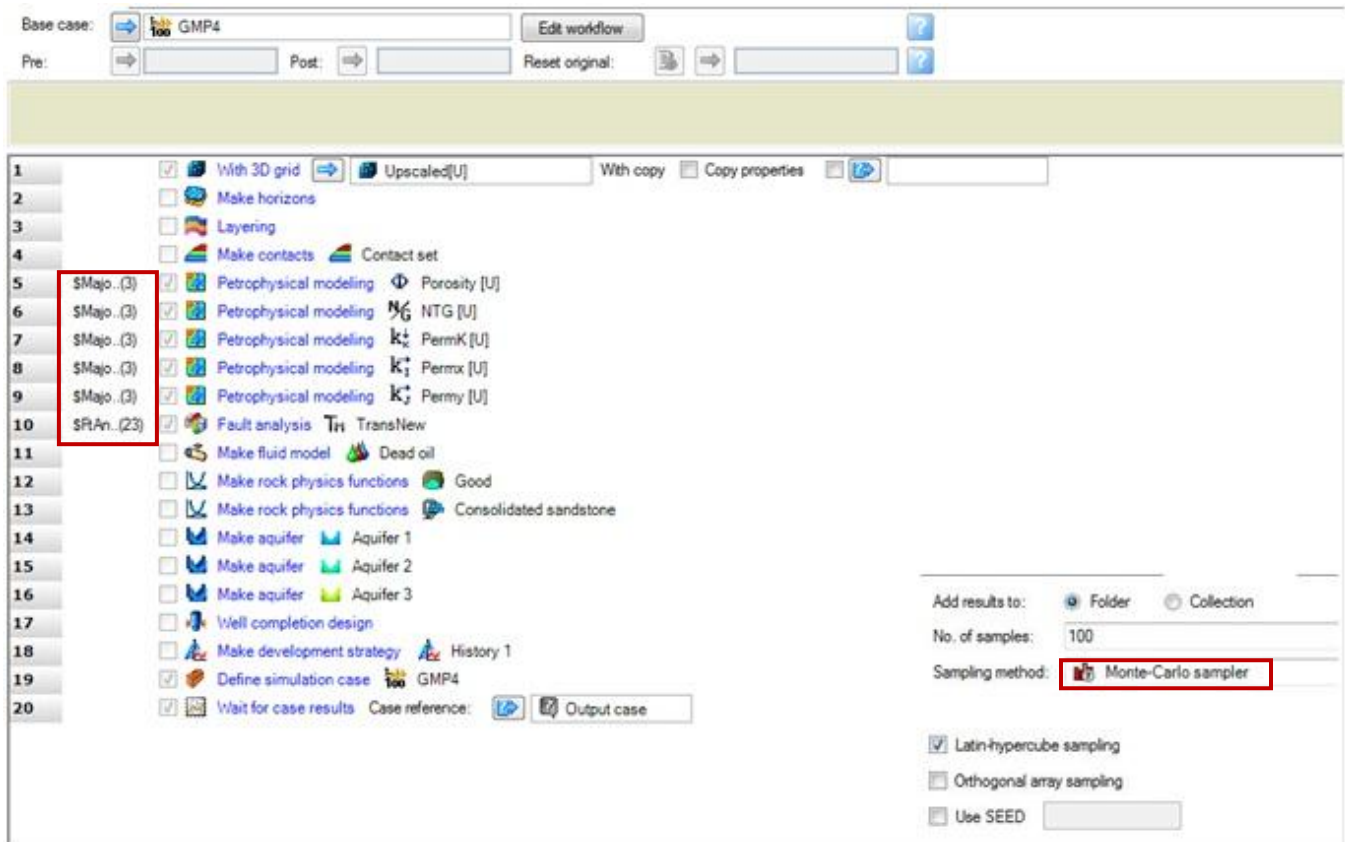


Figure 7 Workflow in Petrel to create prior.

APPENDIX C

History Matched Results

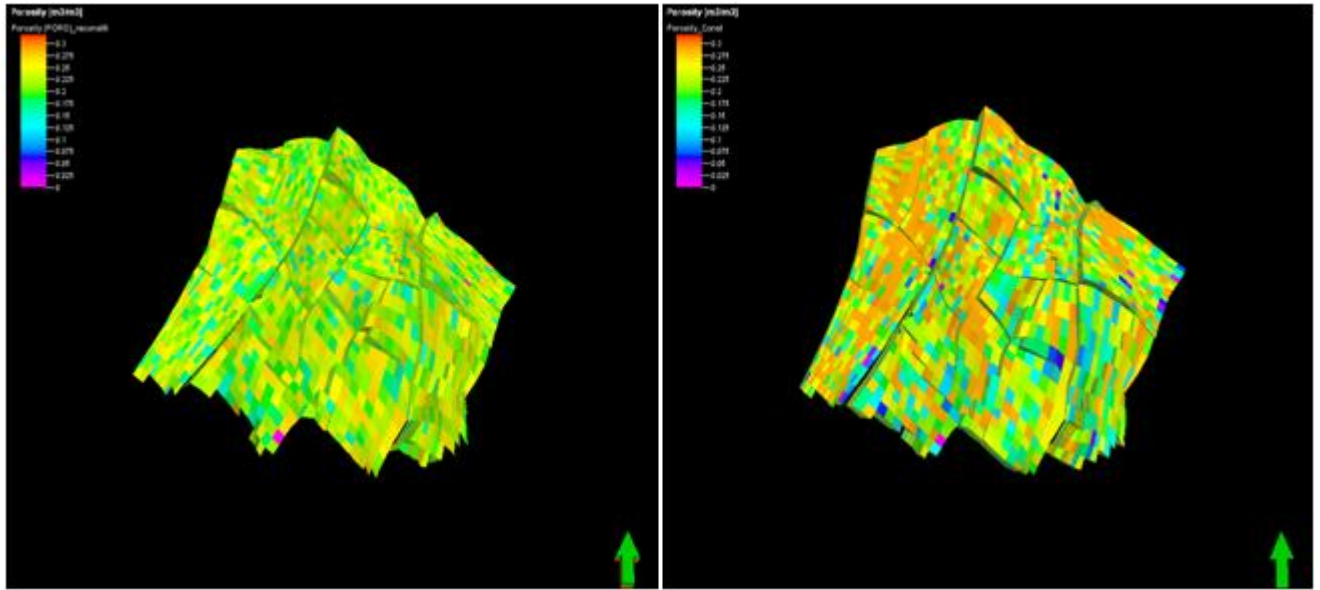


Figure 8 Aerial view of the porosity distribution; on the left is one of the prior model and on the right is the history matched model.

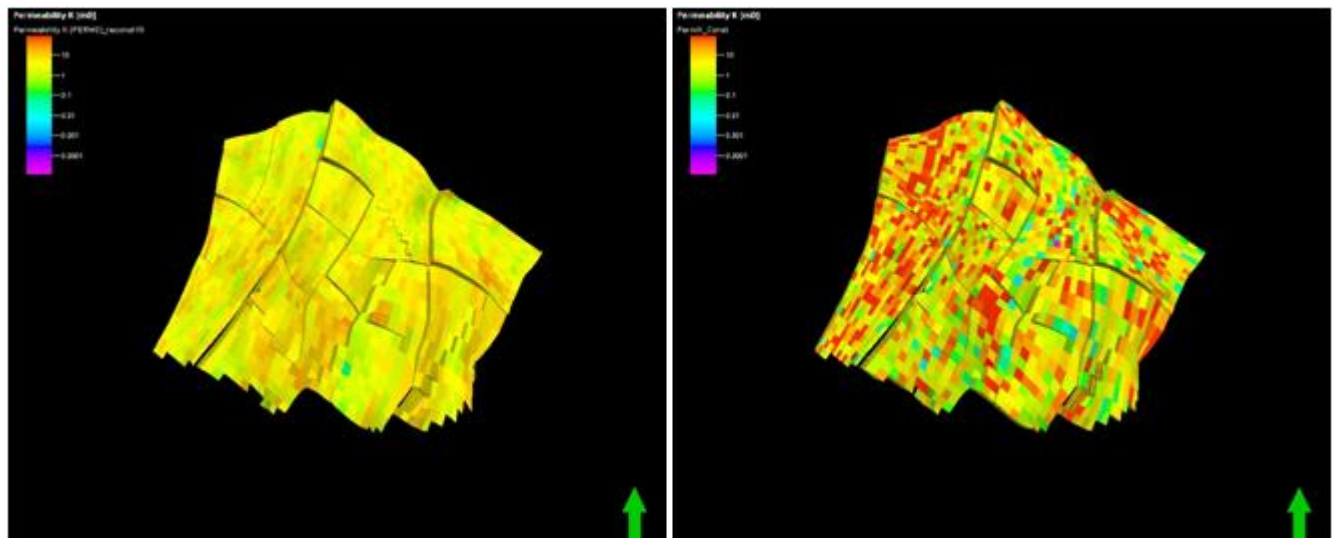


Figure 9 Aerial view of the permeability (in k direction) distribution; on the left is one of the prior model and on the right is the history matched model.

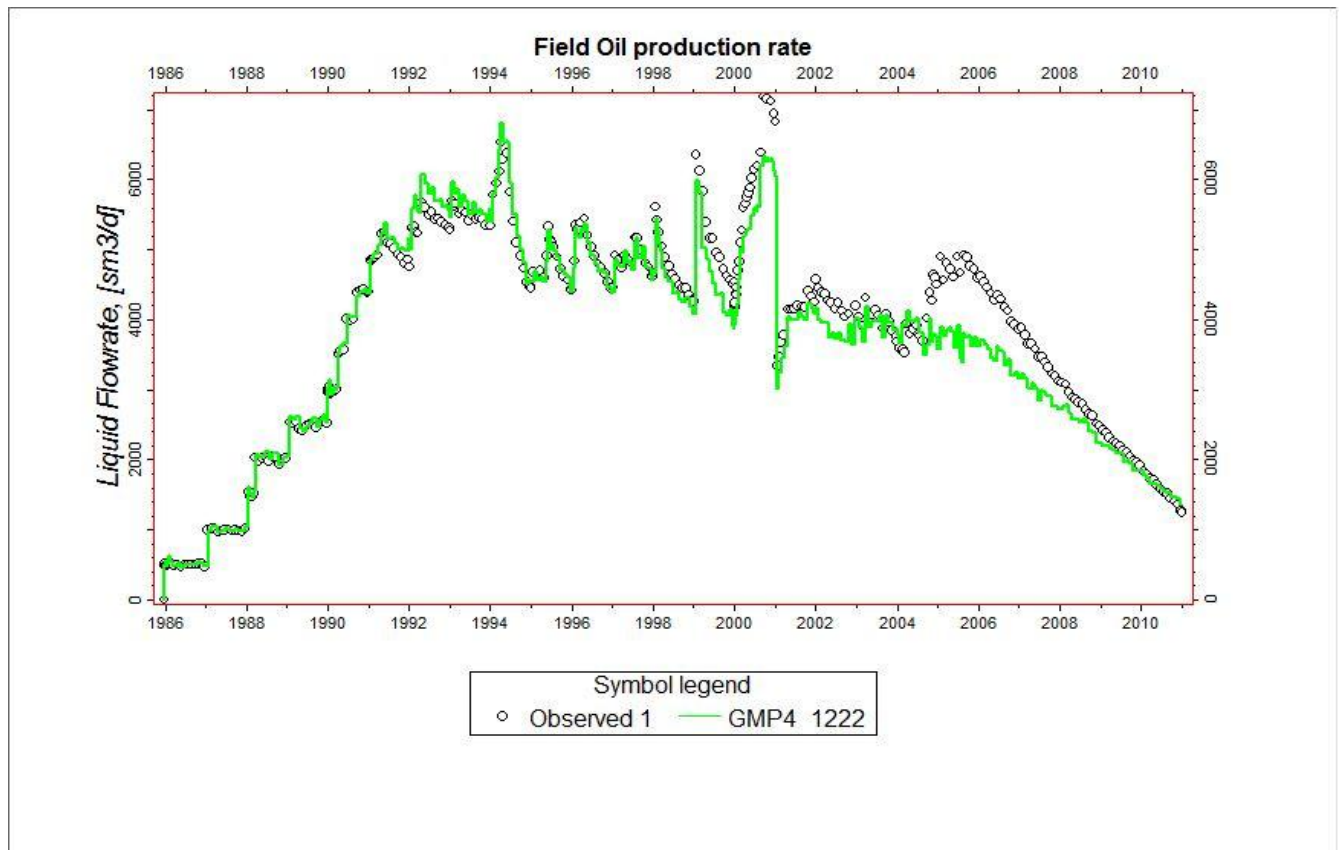


Figure 10 Match for the field oil production rate. Black points denotes the observed data, green line is the simulated result.

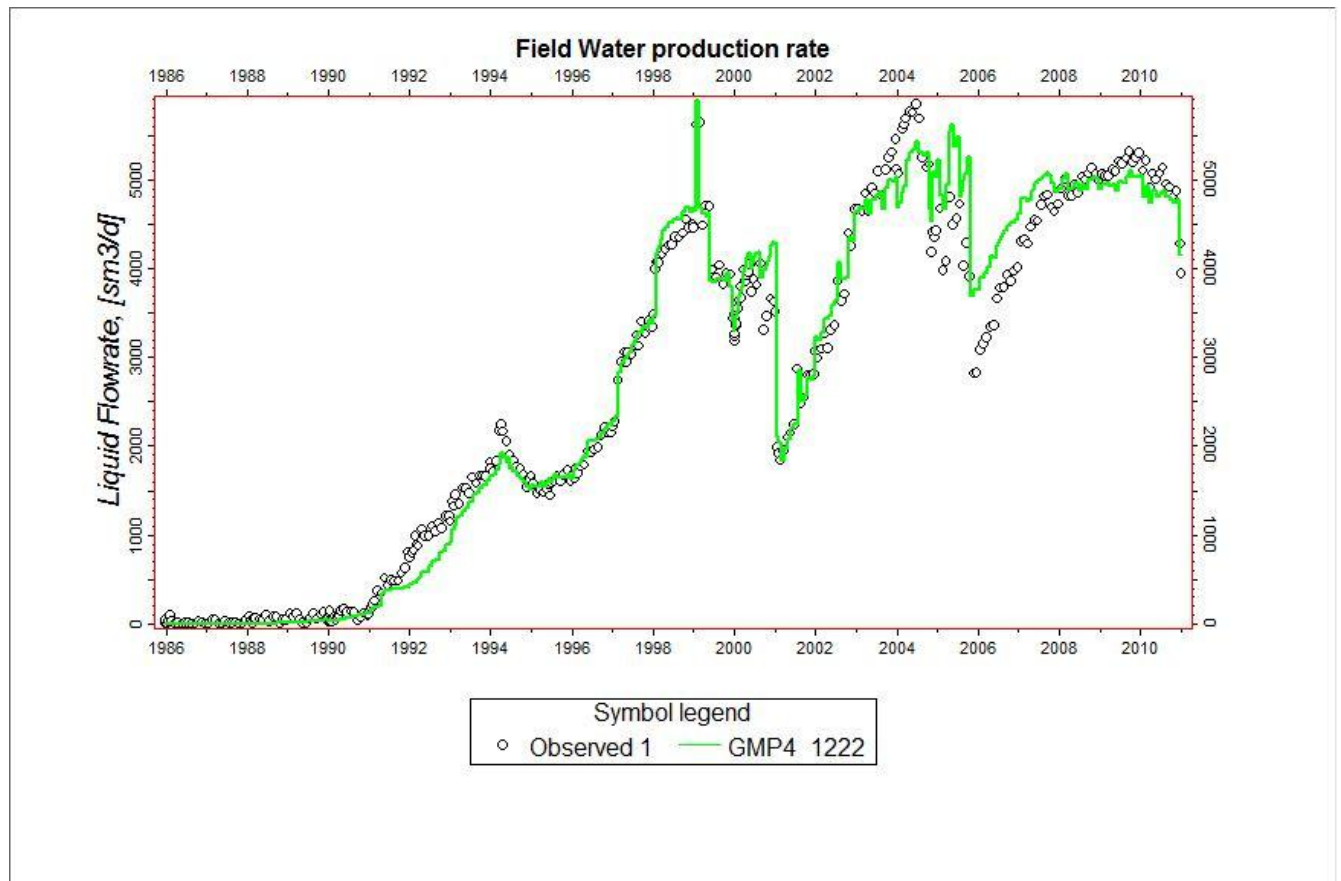


Figure 11 Match for Water Production Rate. Black points denotes the observed data, green line is the simulated result.

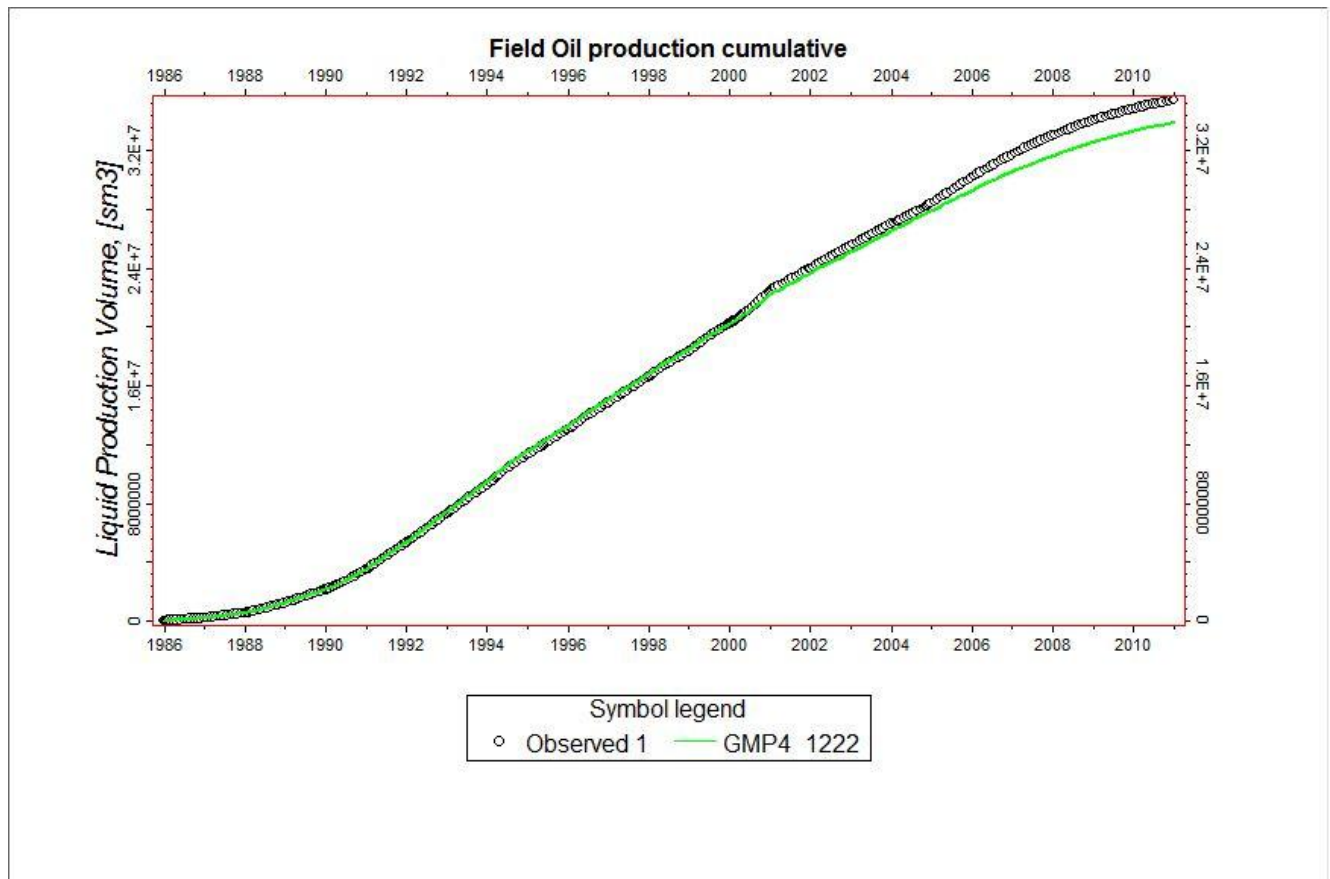


Figure 12 Match for field oil production total. Black points denotes the observed data, green line is the simulated result.

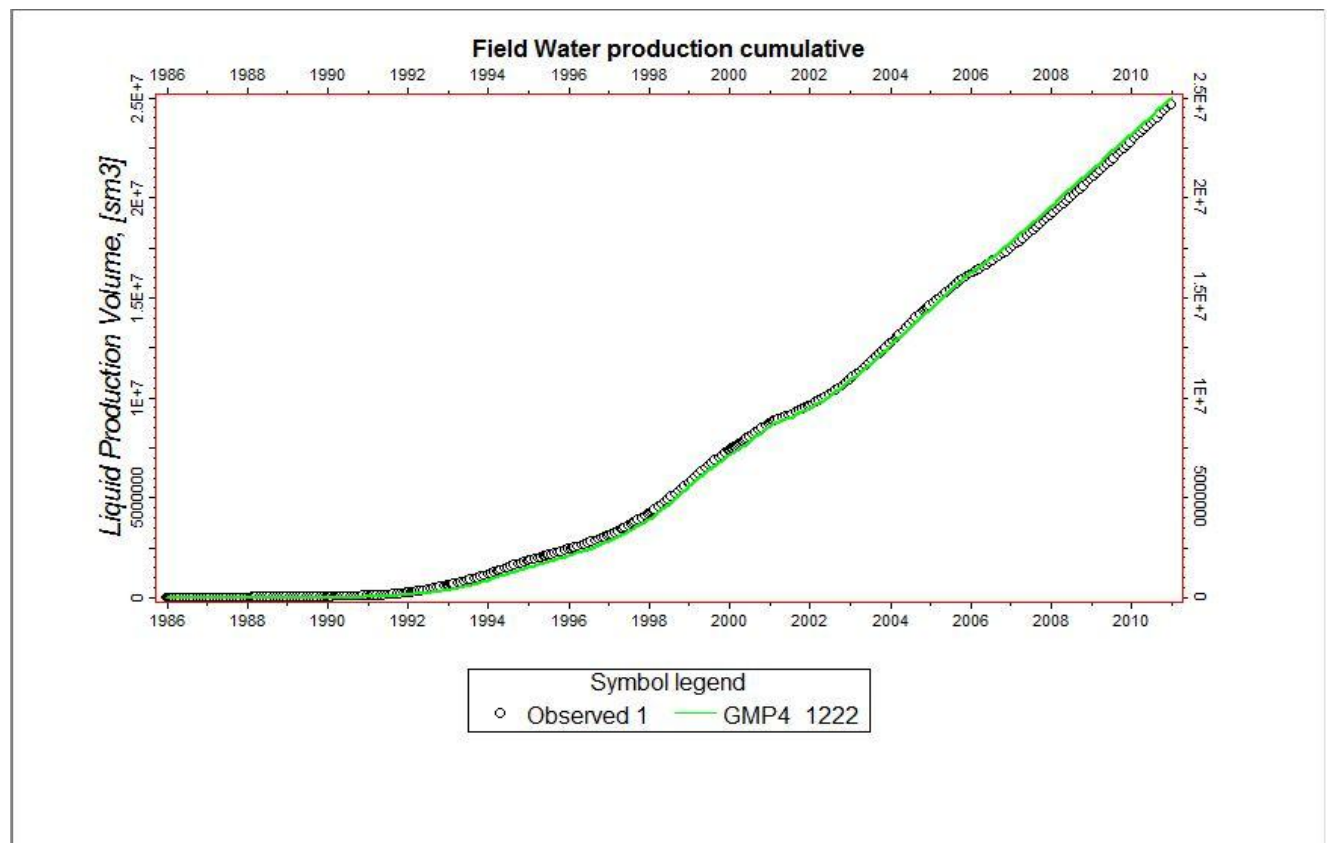


Figure 13 Match for field water production total. Black points denotes the observed data, green line is the simulated result.