PROCEEDINGS, Thirty-Sixth Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, January 31 - February 2, 2011
SGP-TR-191

# INVESTIGATION OF THE USE OF THE ENSEMBLE KALMAN FILTER (ENKF) FOR HISTORY MATCHING PRESSURE DATA FROM GEOTHERMAL RESERVOIRS

Omer Inanc Tureyen and Mustafa Onur

Istanbul Technical University
ITU Maden Fakultesi Petrol ve Dogal Gaz Muh. Bol.
Maslak, 34469, İstanbul-TURKEY
e-mails: inanct@itu.edu.tr, onur@itu.edu.tr

## ABSTRACT

In this study we investigate the use of the Ensemble Kalman Filter (EnKF) method for estimating model parameters and quantifying uncertainty of future performance predictions of reservoir models for liquid dominated geothermal reservoirs. Specifically we concentrate on the performance and accuracy of the method. We couple the Ensemble Kalman Filter with lumped parameter models (tank models) for testing the method. The lumped parameter models used in the study are capable of modeling the average pressure behavior of liquid dominated geothermal reservoirs. This is accomplished by solving the mass balance simultaneously on all tanks that represent the various components of a geothermal reservoir (components such as the aquifer or the reservoir itself). The model parameters that are used in the inversion process are mainly recharge indices between tanks, storage capacities and initial pressures of the tanks.

Our main goal in this study is to have a clear understanding about the Ensemble Kalman Filter method and how it performs. We first present synthetic examples then use the method on real field data. The method seems to be very advantageous in terms of speed compared to other gradient-based history matching procedures (e.g., the Levenberg-Marquart) even though the problem we are dealing with in this study is composed of only a few model parameters.

## INTRODUCTION

The ultimate goal in any reservoir engineering study is to predict both the future performance predictions and to predict the uncertainty associated with the performance. This is necessary to determine the production/re-injection practices that will provide economical exploitation of the geothermal system in consideration. Uncertainty in future performance predictions is mainly caused by (i) measurement errors or noise in observed data, (ii) modeling errors.

In this paper we study parameter estimation, performance predictions and the uncertainty in performance predictions by using the Ensemble Kalman Filter (EnKF) method. The method is stochastic in nature. The petroleum engineering literature available on EnKF is now quite extensive. Its application to petroleum reservoir history matching problems have been studied by many authors; Naevdal et al. (2005); Gao et al., (2005); Zafari and Reynolds, (2005); Evensen et al. (2007); Li and Reynolds, (2007); Aanonsen et al., (2009); Gu and Oliver, (2006); Li et al., (2009).

In this work, we couple the EnKF with lumped parameter models for testing. Lumped parameter models are chosen because of their simplicity (only a few model parameters are involved) and their speed. The lumped parameter model used in this study solves the mass balance equation and hence is capable of modeling the average pressure behavior of geothermal reservoirs. These types of models have been used for modeling a number of geothermal systems in Iceland, Turkey, Philippines, China, Mexico and other countries. For instance, Axellson et al (2005), Sarak et al. (2005) and Tureyen et al. (2007) have presented several field applications of various lumped parameter models to low temperature systems. When lumped parameter models are used, model parameters are determined as a result of a history matching process to available production data. Once the model parameters are obtained, future performance predictions of average reservoir pressure (or water levels) can be made.

The paper begins with a brief review of lumped parameter models considered in this study. Then the EnKF is explained followed by an application to a synthetic example where the performance of the

method is analyzed. Finally an application to real field data is given followed by conclusions.

## LUMPED PARAMETER MODELING

The lumped parameter model used in this study is based on the conservation of mass for a single-phase liquid water and rock system. The model is assumed to be composed of an arbitrary $N_t$ number of tanks. The mass balance equations are then solved numerically for all tanks simultaneously. Fig. 1 illustrates the properties of any tank *i*.

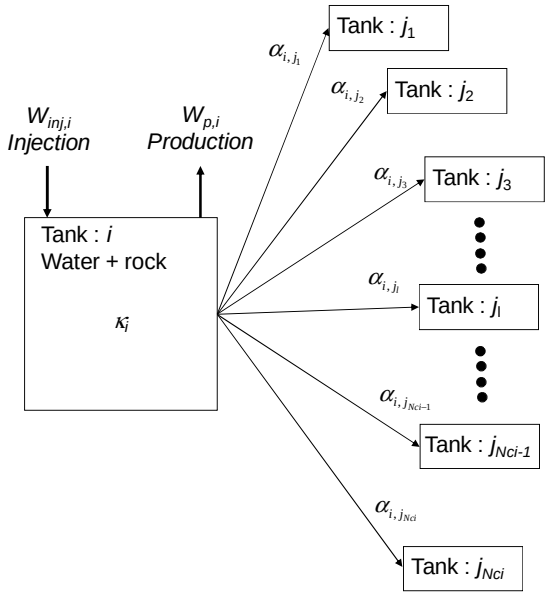



*Figure 1: Properties of a representative tank (tank i) in the model.*

The tanks are composed of two components; water and rock. Tank *i* has a storage capacity $\kappa_i$ and an initial pressure $p_i$. The storage capacity defines the amount of production obtained from a unit pressure drop and is assumed to be constant. Any tank *i* can make an arbitrary number of connections with the other tanks. The total number of connections the tank *i* makes is represented by $N_{ci}$. Note that $N_{ci}$ can vary from tank to tank because each tank in the model can make a different number of connections. Any connecting tank is represented by $j_l$, for $l$=1,2,…,$N_{ci}$. Both injection into and production from tank *i* is allowed. The injection is performed at a specified mass rate $W_{inj,i}$ and the production is at a specified mass rate of $W_{p,i}$. Our convention is that $W_{inj,i}$ is negative for injection (i.e., $W_{inj,i} < 0$) and $W_{p,i}$ is positive for production (i.e., $W_{p,i} > 0$).

The mass flow rate between any tank $j_l$ and *i* is determined using the Schilthuis (1936) relation given in Eq. 1.

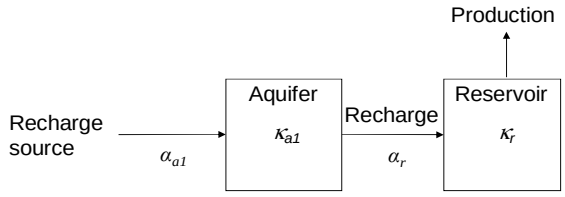
$$W_{i,j_l} = \alpha_{i,j_l}\left(p_{j_l} - p_i\right) \quad (1)$$

where $\alpha_{i,jl}$ is the recharge index which represents the amount of mass flow rate per unit pressure drop and is assumed to be constant, $W_{i,jl}$ is the mass flow rate between tank *i* and tank $j_l$, and $p_{jl}$ is the pressure in tank $j_l$. Under these assumptions the conservation of mass for a single-phase liquid water and rock system can be expressed as follows:

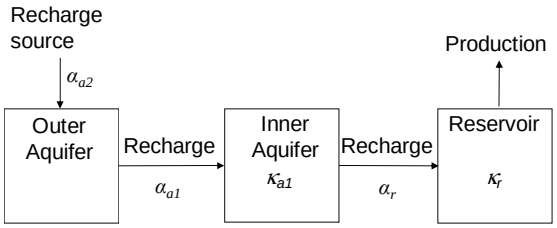
$$\kappa_i \frac{dp_i(t)}{dt} - \sum_{l=1}^{N_{ci}} \alpha_{i,j_l}\left(p_{j_l}(t) - p_i(t)\right) + W_{p,i}(t) + \quad (2)$$

For $i$=1,…,$N_t$. Here *t* represents time. The first term on the LHS of Eq. 2 represents the mass accumulation, the second term represents the mass rate from connecting tanks and the third and fourth terms represent the production and injection mass rates respectively.

The above formulation allows the modeling of average reservoir pressure for any configuration of tanks. Fig. 2 illustrates two alternative configurations of the.




(a) two-tank open lumped parameter model




(b) three-tank open lumped parameter model

*Figure 2: Two different configurations of .tanks in the lumped parameter models.*

The model shown in Fig. 1(a) represents a two-tank open lumped model, where the first tank, in which production/injection occurs, represents the innermost (or central) part of the geothermal system. The changes in pressure in this part are monitored and production/injection rates are recorded. In the second tank, representing the outer part of the reservoir that is connected to the recharge source, there is neither production nor injection and it recharges the central reservoir. Fluid production causes the pressure in the

reservoir to decline, which results in water influx from the outer to the central part of the reservoir. The recharge source represents the outermost part of the geothermal system.

When using the lumped-parameter models considered in this work (Fig. 1), the simulated model (output) response represents pressure or water level changes for an observation well for a given net production history (input). The number of model parameters increases as the number of tanks or the complexity of the lumped model increases. Eq. 2 has been used to solve for the pressures in this study, however analytical solutions to the above given configurations are also available in the literature, see Sarak et al. (2005)

## THE ENSEMBLE KALMAN FILTER (ENKF)

Our implementation of the EnKF follows that of Li et al. (2009). We have applied their algorithmic approach to the lumped parameter models. Traditionally the Kalman filter has been used in the literature for estimating the state of any system which evolves with time. The system is represented with a state vector that is forecasted forward in time to the point where observed data are available. Then the state of the system is corrected to honor the observed data. When considering lumped parameter models, the state of the system are composed of all variables that are needed to run the model and those that are uncertain. We will denote the state vector as $\boldsymbol{y}$, and it will mainly be composed of two components. The first component of the state vector contains the (unknown or uncertain) model parameters (the recharge indices and the storage capacities of the tanks). It is important to note that, in this study, the model parameters are treated independent of time. The second component of the state vector is composed of the dynamic variables (pressures of the tanks). The state vector for the lumped parameter models considered in this paper are given in Eq. 3.

$$\boldsymbol{y} = [\boldsymbol{m}^T \boldsymbol{p}^T]^T \qquad 3$$

Here $T$ represents the transpose, $\boldsymbol{m}$ represents the vector of model parameters, and $\boldsymbol{p}$ represents the dynamic variables. The state vector has a dimension of $N_y=M+N_p$, where $M$ represents the number of model parameters that are unknown or uncertain and $N_p$ represents the number of dynamic variables (for the lumped parameter applications it will be equal to the total number of tanks in the system). We should note that in many of the studies present in the literature, the state vector is composed of mainly three parts; model parameters, dynamic variables and theoretical data. In the application of the EnKF to the lumped parameter models, the theoretical data are the same as the dynamic variables since the data used for history matching are average pressure data. Hence we do not have the extra part (theoretical data) since it is the same as the dynamic variables.

For a two-tank open system (as shown in Fig. 2a), if all model parameters were treated as unknown, the state vector would have the following form:

$$\boldsymbol{y} = [\alpha_r, \kappa_r, \alpha_{a1}, \kappa_{a1}, p_r, p_{a1}]^T \qquad 4$$

It is important to note that there exists the following relationship between the model parameters and the dynamic variables:

$$\boldsymbol{p} = f(\boldsymbol{m}) \qquad 5$$

Here the function $f$ corresponds to the numerical solution of Eq. 2. It is however also possible to write an expression for $\boldsymbol{p}$ as follows (Gu and Oliver, 2006):

$$\boldsymbol{p} = \boldsymbol{H}\boldsymbol{y} \qquad 6$$

$\boldsymbol{H}$ is a matrix with only 0 and 1 as its components, $\boldsymbol{H}=[\boldsymbol{0}|\boldsymbol{I}]$ is used for extracting the dynamic variables from the state vector. Here $\boldsymbol{0}$ is a $N_p \times M$ matrix with 0 entries and $\boldsymbol{I}$ is an identity matrix with dimensions of $N_p \times N_p$.

Initially, at $t$=0, before any data are assimilated (history matched), a large number of state vectors are generated, $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{Ne}$, from a prior model. Here $N_e$ is the total number of state vectors which is also referred to as the number of ensembles. At this point only the model parameters in the state vectors are generated from a chosen prior distribution. Using these model parameters the model is advanced to the first point in time where data are observed (we will denote this time as $t^k$). This is the forecast step where the dynamic variables are predicted at the time point where data are available. Once the forecast step is completed, all ensembles of the model parameters and the dynamic variables (in other words all ensembles of the state vector) are updated based on the difference between the observed data and the forecasted dynamic variables using the following relation (Gu and Oliver, 2006):

$$\boldsymbol{y}_j^u = \boldsymbol{y}_j^f + \boldsymbol{K}_e\left(\boldsymbol{d}_j - \boldsymbol{H}\boldsymbol{y}_j^f\right) \quad (j = 1,2,\ldots,N_e) \qquad 7$$

Here $\boldsymbol{y}_j^u$ represents the updated state vector, $\boldsymbol{y}_j^f$ represents forecasted state vector and $\boldsymbol{H}\boldsymbol{y}_j^f$ represents the forecasted dynamic variables. $\boldsymbol{d}_j$ is an unconditional realization of the observed data and can be generated using the following equation:

$$\boldsymbol{d}_j = \boldsymbol{d} + \boldsymbol{C}_D^{1/2}\boldsymbol{z}_u \qquad 8$$

where $\boldsymbol{d}$ represents the observed data at $t^k$, $\boldsymbol{C}_D$ is the $N_d \times N_d$ (where $N_d$ represents the total number of available data at $t^k$) covariance matrix of all available data at $t^k$ and $\boldsymbol{z}_u$ is an $N_d$-dimensional vector of independent random normal deviates. In this work, we assume a diagonal data covariance matrix, and hence assume that noise in observed comes from an uncorrelated Gaussian distribution.

$\boldsymbol{K}_e$ in Eq. 7 is called the Kalman gain matrix and can be determined using the following equation (Gu and Oliver 2006):

$$\boldsymbol{K}_e = \boldsymbol{C}_Y^k \boldsymbol{H}^T \left(\boldsymbol{H}\boldsymbol{C}_Y^k \boldsymbol{H}^T + \boldsymbol{C}_D\right)^{-1} \qquad 9$$

$\boldsymbol{C}_Y^k$ is the covariance matrix of the state vector. The individual elements of $\boldsymbol{C}_Y^k$ is determined from the ensemble of state vector. Specifically any element of $\boldsymbol{C}_Y^k$ can be computed using the following relationship:

$$c_{m,l} = \frac{1}{N_e - 1}\sum_{j=1}^{N_e}\left(x_{m,j} - \overline{x}_m\right)\left(x_{l,j} - \overline{x}_l\right) \qquad 10$$

for $m,l=1,\ldots,N_y$. The subscripts $m$ and $l$ refer to the $m^{\text{th}}$ and $l^{\text{th}}$ entries in the covariance matrix. $x_{m,j}$ and $x_{l,j}$ correspond to the $m^{\text{th}}$ and $l^{\text{th}}$ variables for the $j^{\text{th}}$ ensemble. $\overline{x}_m$ and $\overline{x}_l$ are the means that are calculated across the ensembles.

It is important to note that the update is performed on the entire state vector. In other words both the model parameters and the dynamic variables are updated at the same time. The reason behind updating the dynamic variables is to mimic the response of the lumped parameter model if the model was run with the updated model parameters from time zero. Hence by updating the dynamic variables, we avoid re-running the model from time zero which may be time consuming. The procedure does not introduce any errors if the relationship between the model parameters and the dynamic variables ($f$ in Eq. 5) is linear. In the case of a non-linear relationship, the results are only approximate.

Once the entire state vector is updated, we again perform a forecast step to the point where the next time level ($t^{k+1}$) where a new set of observed data is available. Then we repeat the above steps and perform another updating. This procedure is repeated until all observed data are assimilated.

## APPLICATIONS OF ENKF

In this section we present synthetic and real field applications for better understanding of the EnKF method and its capabilities. We will first look at a simple linear problem. Then we will consider a synthetic example using the lumped parameter model, which constitutes an example of a nonlinear problem. Finally a real field application is performed on data obtained from the Balcova Narlidere field in Izmir, Turkey. The real field data is the same data used by Tureyen et al. (2007).

### Application to a Linear Model

In this application of the EnKF we use a very simple linear model given by the following equation:

$$h(a,b,t) = at + b \qquad 11$$

Here $a$ and $b$ are the model parameters and $t$ is the independent variable. Note that $h$ is linearly related to the model parameters.

We first generate synthetic data to be used in the assimilation process. For this purpose the true values of the model parameters are taken as $a$=5.75 and $b$=11.23. By varying $t$ from 1 to 100 with a uniform step size of 1, we generated 100 data points. Then a Gaussian noise with mean equal to zero and variance equal to 16 was added to the generated data points to mimic actual observed data (all are illustrated in Fig. 3).

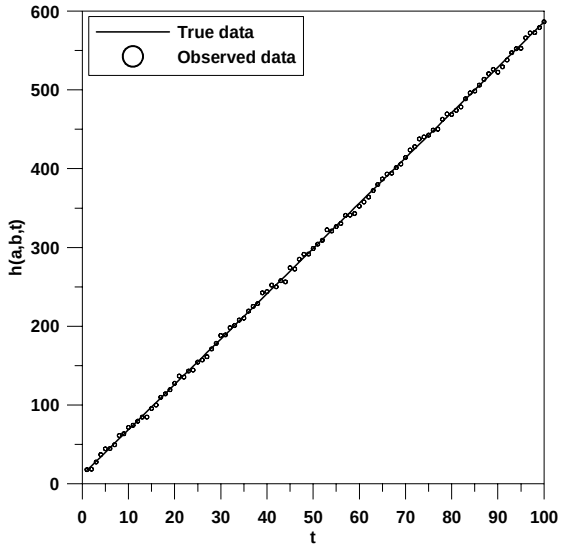



*Figure 3: True data and the observed data for the linear problem.*

We now apply the EnKF to assimilate the observed data given in Fig. 3 using 1000 ensembles for the model parameters. The initial ensembles of the model parameters have been sampled from normal

distributions. The model parameter *a* has been sampled with a mean of 5.75 (the correct mean) and a variance of 100. The model parameter *b* has been sampled with a mean of 11.23 (the correct mean) and a variance of 100. The results obtained from the initial distribution of model parameters and the results after the implementation of the EnKF are given in Fig. 4.

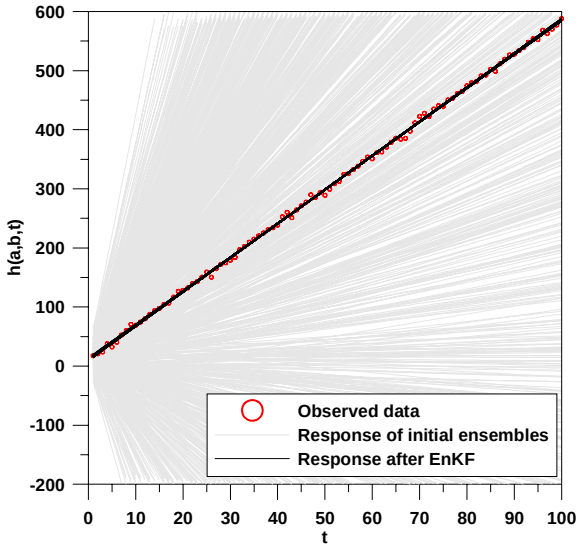



*Figure 4: Results before and after the application of the EnKF for the linear problem.*

The gray lines in Fig. 4 represent the results obtained using the initial distribution of the model parameters. The black lines on the other hand represent the results obtained with the model parameters after the application of the EnKF. It is clear that the EnKF performs very well in matching to the observed data.

Table 1 presents the elements of the computed covariance (and the correlation coefficients which are given with the numbers in parentheses) matrix of the model parameters prior to applying the EnKF. As mentioned previously, both initial distributions of the model parameters have been sampled from a normal distribution with a variance of 100. The diagonal terms of the covariance matrix give a variance of almost 100 which is consistent with the input variance.

*Table 1: Covariance matrix for the model parameters prior to applying the EnKF (the numbers in parenthesis give the correlation coefficients).*

| Parameters | a | b |
|---|---|---|
| **a** | 99.99 (1.0) | 0.7 ($7.14\times10^{-3}$) |
| **b** | 0.7 ($7.14\times10^{-3}$) | 94.77 (1.0) |

The off diagonal elements of Table 1 give the covariance and the correlation coefficient between the model parameters *a* and *b*. The low correlation coefficient between the model parameters represent very low linear correlation. Table 2 gives the posterior covariance matrix from the model parameters, computed analytically (from linear parameter estimation assuming Gaussian statistics for observed data and the model parameters) using the following relationship (Oliver et al., 2008):

$$\boldsymbol{C}_{mp} = \boldsymbol{C}_m - \boldsymbol{C}_m \boldsymbol{X}^T \left(\boldsymbol{C}_{D'} + \boldsymbol{X}\boldsymbol{C}_m\boldsymbol{G}^T\right)^{-1} \boldsymbol{X}\boldsymbol{C}_m \qquad 12$$

Here $\boldsymbol{C}_{mp}$ is the posterior covariance matrix, $\boldsymbol{C}_m$ is the prior covariance matrix of the model parameters (for this problem this is given in Table 1), $\boldsymbol{C}_{D'}$ is the data covariance matrix (but in this case it is the covariance matrix of all data), and $\boldsymbol{X}$ is the transformation (design or sensitivity) matrix and, for the specific example considered here is given by:

$$\boldsymbol{X} = \begin{bmatrix} 1 & 1 \\ \vdots & \vdots \\ 100 & 1 \end{bmatrix} \qquad 13$$

The results of Table 2 show that when the observed data are matched, the variances of the model parameters, as expected, decrease significantly (compare diagonal elements of the matrices given in Tables 1 and 2). Note that there exists nearly no correlation in the initial sample distribution of the model parameters. However, once data are history matched, a correlation is introduced between the model parameters (compare off-diagonal elements of matrices given in Tables 1 and 2).

*Table 2: Analytically determined posterior covariance matrix for the model parameters (the numbers in parenthesis give the correlation coefficients).*

| Parameters | a | b |
|---|---|---|
| **a** | $1.91\times10^{-4}$ (1.0) | $-9.63\times10^{-3}$ (-0.87) |
| **b** | $-9.63\times10^{-3}$ (-0.87) | 0.65 (1.0) |

*Table 3: Covariance matrix for the model parameters obtained from the distribution of model parameters after EnKF (the numbers in parenthesis give the correlation coefficients).*

| Parameters | a | b |
|---|---|---|
| **a** | $1.91\times10^{-4}$ (1.0) | $-9.81\times10^{-3}$ (-0.87) |
| **b** | $-9.81\times10^{-3}$ (-0.87) | 0.66 (1.0) |

Table 3 illustrates the covariance matrix obtained from the 1000 ensembles of model parameters after the EnKF. It is clear from the results of Table 3 that the EnKF manages to reproduce the statistics given by Table 2, which was computed by the use of the analytical formula given by Eq. 12.

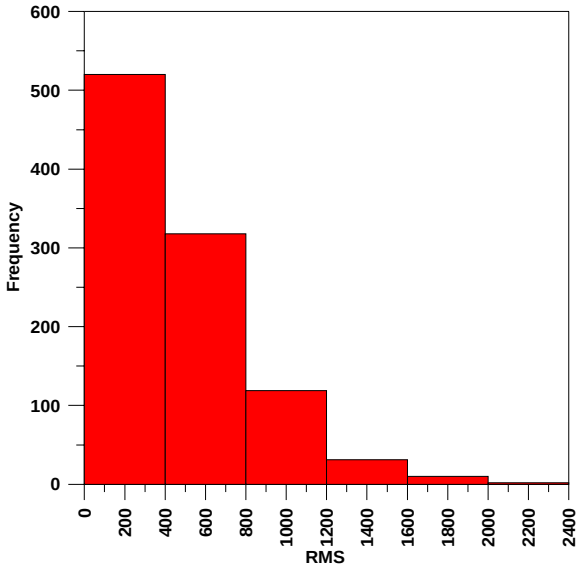



*Figure 5: Histogram of RMS based on observed data for the results of the prior model parameters.*

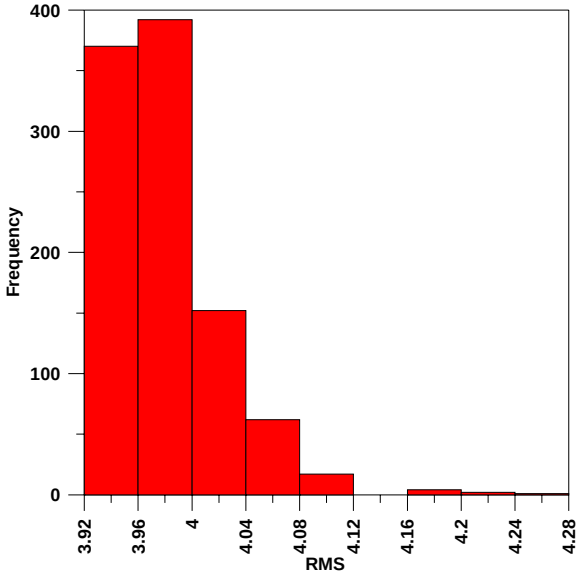



*Figure 6: Histogram of RMS based on observed data for the results of the posterior model parameters.*

Figs 5 and 6 give the histograms of the RMS (root mean squared error) for the results of the prior (gray lines in Fig 4) and posterior (black solid lines) distributions of the model parameters respectively.

The RMS values are computed using the following equation:

$$RMS_j = \sqrt{\frac{1}{N_{d'}}\sum_{i=1}^{N_{d'}}\left[d_{obs,i} - p_{j,i}\right]^2} \qquad (j = 1,..,N \qquad (14)$$

Here $N_{d'}$ represents the total number of data, $d_{obs,i}$ represents the i$^{th}$ observed data and $p_{j,i}$ represents the i$^{th}$ computed pressure of the j$^{th}$ ensemble.

The RMS values obtained from the model response using the prior distribution of model parameters are indeed very high. After the application of EnKF, the RMS values are considerably reduced. As mentioned earlier, the noise added to the true model response had a standard deviation of 4 (variance of 16). When Fig. 6 is inspected, it is clear that all RMS values are gathered around 4. This shows that the matches obtained by EnKF are indeed very good that they reproduce the noise initially added to the true data.

## Application to Lumped Parameter Models

We will assume that the true model for the application on lumped parameter models is a two-tank open model as shown in Fig. 2a. The true model response (in this application the average reservoir pressure) was corrupted by adding noise from a $N(0,0.49)$ to mimic actual observed data. Hence the observed pressure change to be used in the application is now the corrupted pressure. We further assume that we know the level of noise associated with the data. The corrupted pressure data represent the observed data $\boldsymbol{d}$ and contain 193 data points. The observed pressure data are shown in Fig. 7 along with the production history of the field.

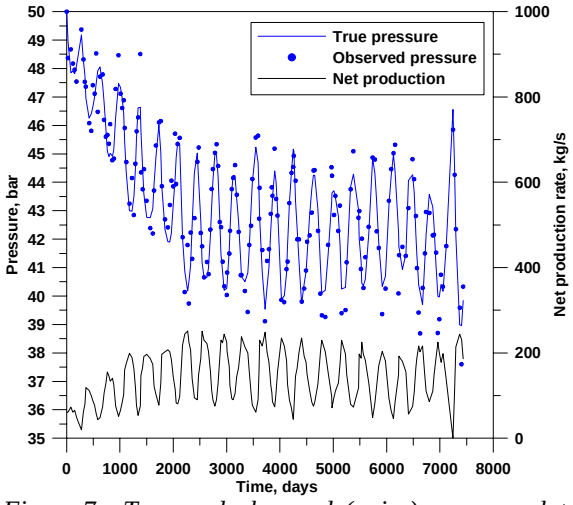



*Figure 7: True and observed (noisy) pressure data and net production rate.*

The four unknown model parameters for this case are the storage capacities of the tanks and the recharge indices. Just as in the linear problem we again use 1000 ensembles of the state vector. Table 4 summarizes the true model parameters used to generate the true pressure response given in Fig. 7, the initial ensemble of realizations and the resulting distribution of model parameters from EnKF. The top rows of the entries given in the third and fourth columns represent the mean of the distributions while $\sigma^2$ represents the variance of the distribution. In this example, the initial ensemble of both recharge indices have been sampled from a uniform distribution between 1 kg/(bar-s) and 50 kg/(bar-s). The natural logarithm of the storage capacities have been sampled from a uniform distribution between 16 and 28. The storage capacities in this case have values ranging from $8.8\times10^6$ kg/bar to $1.44\times10^{12}$ kg/bar. This covers a very wide range of both the recharge indices and storage capacities. The main idea behind using uniform initial distributions is that in actual applications, we may not have any prior knowledge about model parameters except perhaps a minimum and a maximum. Hence, all realizations of the model parameters between this minimum and maximum would be all equi-probable. A uniform distribution reflects this event. The fourth column of Table 4 gives the means and statistical properties of the ensemble of model parameters after the EnKF has been applied.

*Table 4: Summary of the EnKF process.*

| Model parameters | True parameters | Initial ensembles | Resulting ensembles |
|---|---|---|---|
| $\alpha_r$, kg/bar-s | 30 | 25.58<br>$\sigma^2$=193.2 | 30.51<br>$\sigma^2$=3.46 |
| ln $\kappa_r$, kg/bar | 18.304 | 22.03<br>$\sigma^2$=11.8 | 18.176<br>$\sigma^2$=0.011 |
| $\alpha_{a1}$, kg/bar-s | 37 | 25.09<br>$\sigma^2$=194.9 | 30.14<br>$\sigma^2$=34.8 |
| ln $\kappa_{a1}$, kg/bar | 23.121 | 22.113<br>$\sigma^2$=11.3 | 23.07<br>$\sigma^2$=0.31 |

It is clear that the EnKF has managed to decrease the uncertainty (in this case represented by the variances denoted by $\sigma^2$ in Table 4) considerably for all parameters. Furthermore the means of the ensemble of resulting model parameters are indeed very close to the true parameters.

Fig. 8 illustrates how the EnKF performs in terms of matching the history. The gray lines represent the model response when the initial ensembles are used in the model. The solid black lines represent the model responses of the resulting ensemble of model parameters. The red solid line represents the true pressure data and the black dashed line represents the production rate. As it is clear from Fig. 8 the EnKF has managed to reduce the very wide band of the initial ensembles to better match the production data.

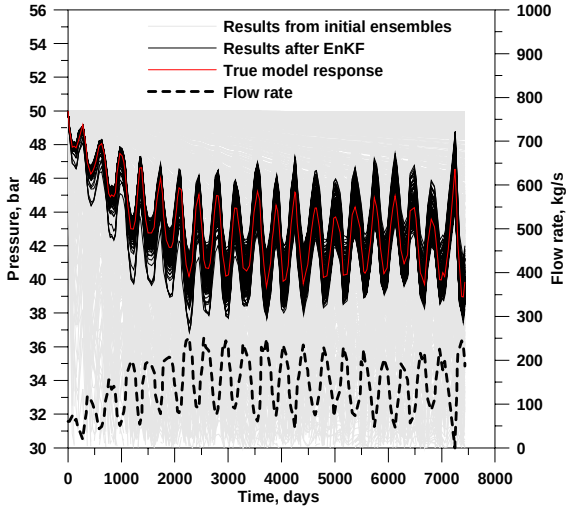



*Figure 8: Results of the EnKF process.*

To quantitatively test the match of the 1000 ensembles (the solid black lines given in Fig. 8) we compute the RMS based on the observed data. As mentioned earlier the observed data have been created by adding noise to the true pressure that has been sampled from a normal distribution that has a mean equal to 0 and a standard deviation equal to 0.7. Hence, if the matches are good, then the RMS based on the observed data should yield a value of approximately 0.7. Fig. 9 gives the histogram for the RMS values of all of the ensembles (black solid lines in Fig. 8). Not all ensembles have an RMS value of .0.7. Although the mode of the distribution is around 0.7, there still exist many ensembles with higher RMS.

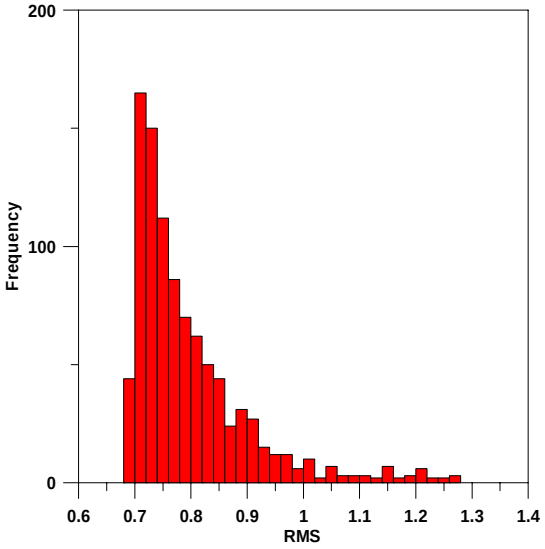



*Figure 9: Histogram of the RMS based on observed data.*

Fig. 10 illustrates the behavior of the mean of the ensembles. The RMS between the mean of the ensembles of predicted pressures and the observed data is 0.728, which is very close to the input value of standard deviation of noise added to the true pressures.

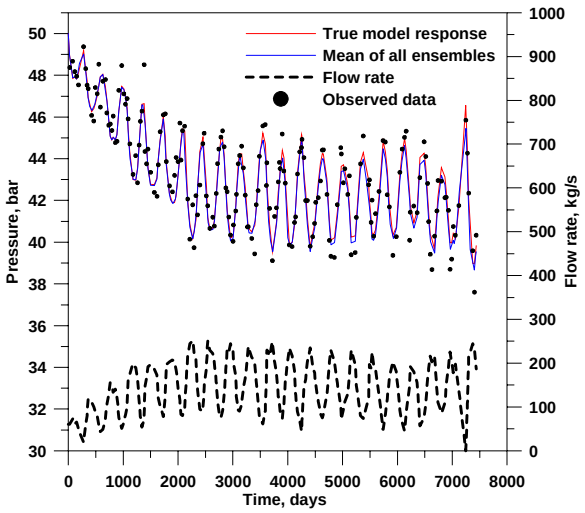



*Figure 10: Comparison of the mean of all ensembles and the true model response.*

In the linear problem given in the previous subsection, almost all RMS values of the model response based on the posterior ensemble of model parameters were very close to the noise added to the true data. In the application of EnKF on lumped parameter models an inspection of Fig. 9 shows that there are many model responses that give higher RMS values then the noise added to the true data.

To investigate why this is so, we will use a very simple 1 tank open lumped parameter model. This model will have two model parameters, one recharge index and one storage capacity. The aim of this simple toy problem will be to see how the model response behaves with respect to each model parameter. We again take two cases where in the first case the storage capacity is kept constant at ln ($\kappa$)=18 kg/bar and the recharge index is varied between 1 and 50. In the second case we fix the recharge constant at $\alpha$=10 kg/(bar-s) and vary ln ($\kappa$) between 16 and 28. In both cases we have used a net production rate of 100 kg/s and we look at how the pressure at 100 days behaves with changing recharge index and storage capacity. Both cases are given in Figs. 11 and 12 respectively.

Figs 11 and 12 give the behavior of the model response with respect to the recharge constant and the storage capacity respectively. It is clear that in both cases the models behave non-linearly. We believe that this non-linear behavior causes the result of Fig. 9 where many model responses have RMS values are higher then the noise added to the true data.

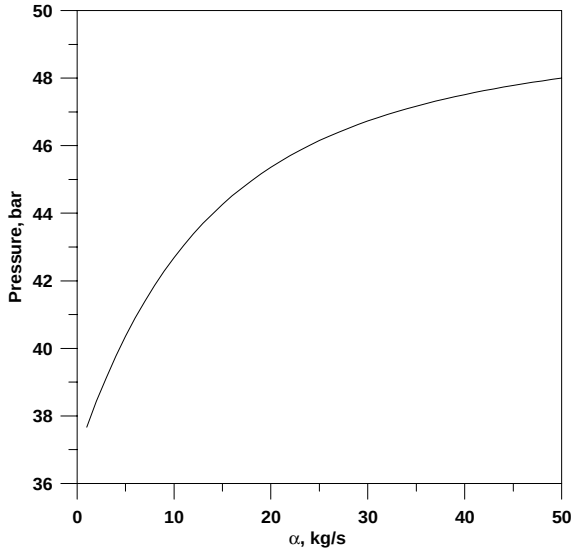



*Figure 11:The variation of pressure with varying recharge indices.*

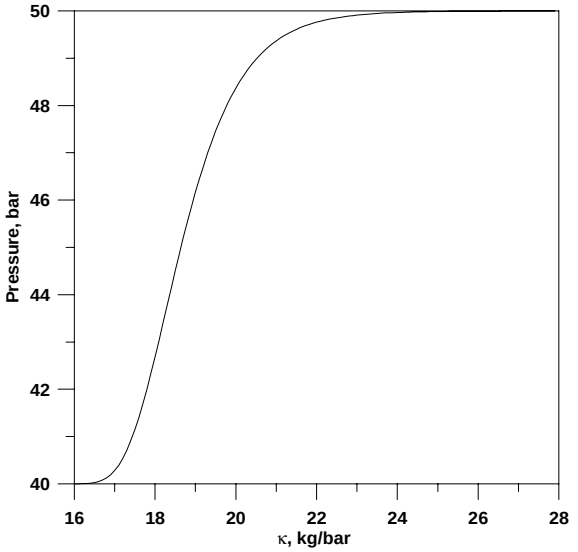



*Figure 12:The variation of pressure with varying storage capacities.*

We will now consider the prediction problem where we try to assess the uncertainty in future prediction of pressures. Fig. 13 illustrates the future predictions of all ensembles, mean of all ensembles and the true response for the two-tank open problem. For simplicity, the flow rate for the prediction period has been taken to be constant (equal to the final rate during the matching period). The true model response lies within the band of uncertainty during the prediction period. If one wants to asses the uncertainty at any time in the future, then the statistics regarding the uncertainty could be obtained simply from the ensembles.

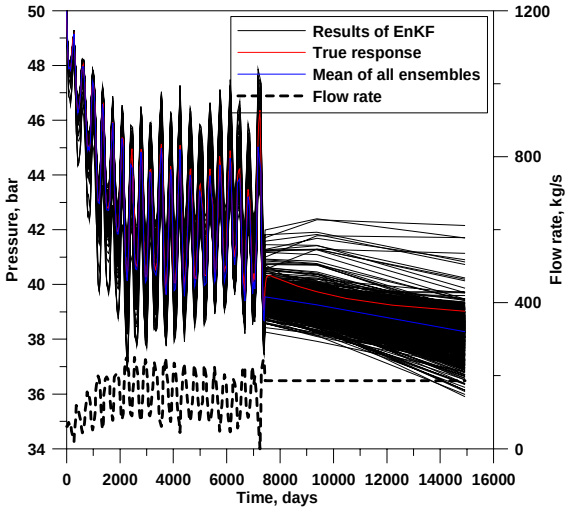



*Figure 13:Future predictions of the ensembles, the true model and the mean of all ensembles.*

Finally we look at the same example this time using a different prior distribution for studying the effects of the initial distribution of model ensembles on the pressure response. For this purpose we generate the initial ensembles from normal distributions with statistics given in the third column of Table 5. This time the means of the initial distribution is taken to be closer to the true model parameters.

*Table 5: Summary of the EnKF process with different initial distribution.*

| Model parameters | True parameters | Initial ensembles | Resulting ensembles |
|---|---|---|---|
| $\alpha_r$, kg/bar-s | 30 | 30.96<br>$\sigma^2$=38.4 | 30.33<br>$\sigma^2$=1.34 |
| ln $\kappa_r$, kg/bar | 18.304 | 18.0<br>$\sigma^2$=1.02 | 18.17<br>$\sigma^2$=0.008 |
| $\alpha_{a1}$, kg/bar-s | 37 | 35.18<br>$\sigma^2$=34.8 | 32.66<br>$\sigma^2$=8.35 |
| ln $\kappa_{a1}$, kg/bar | 23.121 | 22.05<br>$\sigma^2$=0.98 | 23.21<br>$\sigma^2$=0.02 |

The last column of Table 5 indicates the decrease in the standard deviation of the model parameters after the EnKF. More importantly we look at the impact of using such an initial distribution on the pressure matches. Fig. 14 illustrates the histogram of the RMS values. When compared with the histogram given in Fig. 9, there is considerable improvement. Now there are more ensembles that result with lower RMS. This is an expected result since the standard deviations of the initial distribution of the model parameters were lower and the means were close to the true means of the model parameters. This also causes a reduction in the uncertainty band of future predictions of pressure as shown in Fig. 15.

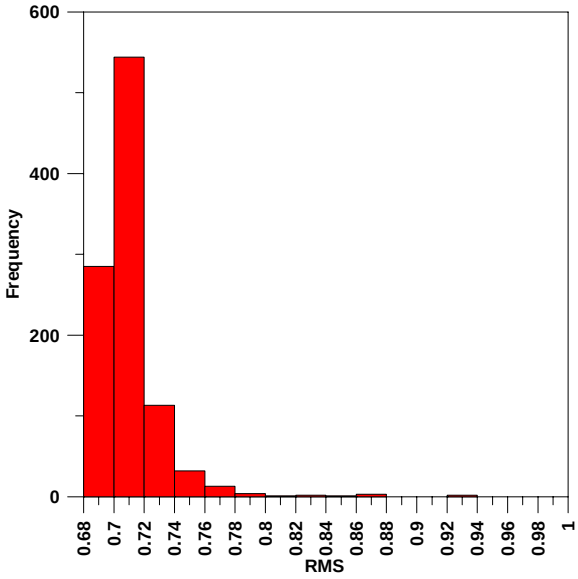



*Figure 14: Histogram of the RMS based on observed data.*

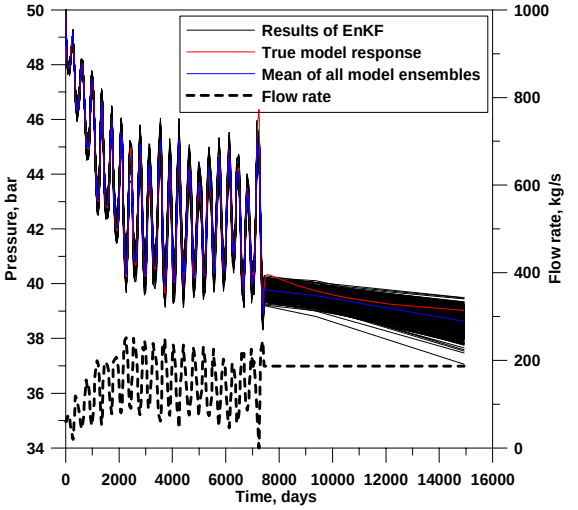



*Figure 15:Future predictions of the ensembles, the true model and the mean of all ensembles.*

## Real Field Example

Here we extend the application of the EnKF to a real field. The field at study is the Balcova Narlidere Geothermal field. This field is known as the oldest geothermal system in Turkey and is situated 10 km west of Izmir. The geothermal water with a temperature ranging from 80 °C to 140 °C is produced from the wells with depths ranging from 48.5 m to 1100m.

The application of the EnKF is performed on the data collected from one of the wells in the field. The data consists of net rate (production rate – injection rate) information starting from 01/01/2000 and corresponding water level data starting from 17/06/2001. All data had been collected until

10/11/2005. Here it is important to note that the net history is obtained from the entire field whereas the water level data is collected only from a single well. Fig. 16 illustrates the collected data to be used in the EnKF application.

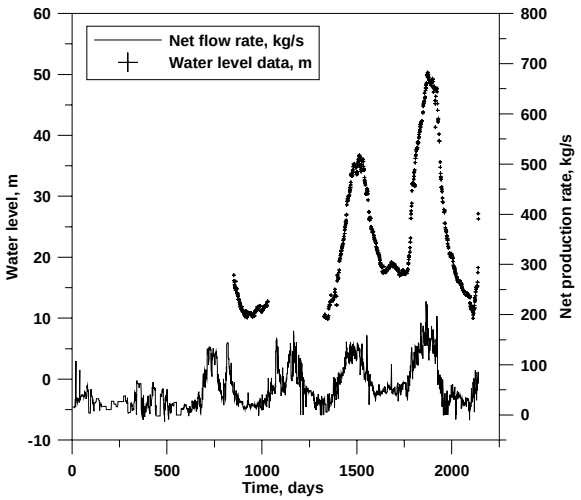



*Figure 16:Net production history and water level data.*

We considered a two-tank open model to be used with EnKF. Table 6 summarizes the statistics of the initial distribution of the ensembles and the resulting ensembles. We have used uniform distributions for the initial distribution of the model parameters with varying intervals.

*Table 6: Summary of the EnKF process for the real field case.*

| Model parameters | Initial ensembles | Resulting ensembles |
|---|---|---|
| $\alpha_r$, kg/bar-s | 30.07<br>$\sigma^2$=131.6 | 34.49<br>$\sigma^2$=0.21 |
| ln $\kappa_r$, kg/bar | 22.09<br>$\sigma^2$=11.5 | 18.04<br>$\sigma^2$=0.0009 |
| $\alpha_{a1}$, kg/bar-s | 124.7<br>$\sigma^2$=210.2 | 139.8<br>$\sigma^2$=11.8 |
| ln $\kappa_{a1}$, kg/bar | 21.97<br>$\sigma^2$=12.3 | 21.37<br>$\sigma^2$=0.13 |

Fig. 17 summarizes the results. The gray lines show the initial ensemble of model parameters, the black lines show the results of the EnKF and the blue line shows the mean of all resulting ensembles. The water levels obtained by using the prior ensemble of model parameters give many realizations that do not match the history. The application of EnKF results in 1000 history matched water level responses, which fits the observed data quite well.

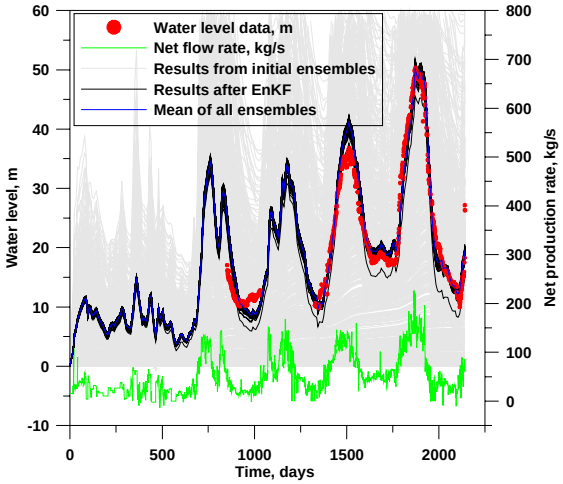



*Figure 17:Results of the EnKF process for the real field case.*

## CONCLUSIONS

The following conclusions have been obtained from this study:

- The Ensemble Kalman Filter method has been successfully used with lumped parameter models both with synthetic examples and real field data.
- The EnKF coupled with lumped parameter models provides very fast history matching for geothermal reservoirs. For the lumped parameter models considered here, history matching of 1000 ensembles takes only on the order of seconds.
- Due to the non-linear relationship between the model parameters and the model responses, when EnKF is performed with lumped parameter models, the RMS values can show a wider spread when compared with the application of EnKF with linear models.
- The results of the EnKF seem to be sensitive to the initial distribution of model parameters. As we have observed, if the means of the initial ensembles are close to the truth and the variance is decreased, this seems to provide better matches as expected.

## REFERENCES


Aanonsen, S.I., Naevdal, G., Oliver, D.S., Reynolds, A.C. and Vallès, B. (2009) "The Ensemble Kalman Filter in Reservoir Engineering – a Review", *SPE Journal*, **14**, 393-412.

Axelsson, G., Björnsson, G. and Quijano, J.E. (2005) "Reliability of Lumped Parameter Modeling of Pressure Changes in Geothermal Reservoirs", *proceedings the World Geothermal Congress*, Antalya, Turkey 24-29 April.

Evensen, G., Hove, J., Meisingset, H.C., Reiso, E., Seim, K.S., and Espelid, Ø., "Using the EnKF for Assisted History Matching of a North Sea Reservoir Model," *proceedings of the 2007 SPE Reservoir Simulation Symposium*, Woodlands, Texas, USA 26-28 February.

Gao, G., Zafari, M. and Reynolds, A.C. (2005), "Quantifying Uncertainty for the PUNQ-S3 Problem in a Bayesian Setting With RML and EnKF", *proceedings of the 2005 SPE reservoir Simulation and Symposium*, Houston, Texas, USA, 31 Jan. – 2 Feb.

Gu, Y. and Oliver, D.S. (2006) "The Ensemble Kalman Filter for Continuous Updating of Reservoir Simulation Models", *Journal of Energy Resources Technology*, **128**, 79 – 87.

Li, G., Han, M., Banerjee, R. and Reynolds, A.C. (2009) "Integration of Well Test Pressure Data Into Heterogeneous Geological Reservoir Models", *proceedings of the 2009 SPE Annual Technical Conference and Exhibition*, New Orleans, Louisiana, USA 4 – 7 Oct.

Li, G. and Reynolds, A.C. (2007) "An Iterative Ensemble Kalman Filter for Data Assimilation", *proceedings of the 2007 SPE Annual Technical Conference and Exhibition*, Anaheim, California, USA, 11 – 14 Nov.

Naevdal, G., Johnsen, L.M., Aanonsen, S.I., and Vefring, E.H. (2005) "Reservoir Monitoring and Continous Model Updating Using Ensemble Kalman Filter," SPE Journal (March), 66-74.

Oliver, D. S., Reynolds, A. C. and Liu, N. (2008) "Inverse theory for Petroleum Reservoir Characterization and History Matching", Cambridge University Press, Cambridge, UK, 380 pp.

Onur, M. and Tureyen, O.I. (2006) "Assesing Uncertainty in Future Pressure Changes Predicted by Lumped Parameter Models For Low Temperature Geothermal Systems", *31st Stanford Geothermal Workshop*, Stanford, California, USA, Jan 30 – Feb 1.

Sarak, H., Onur, M. and Satman, A. (2005) "Lumped Parameter Models for Low Temperature Geothermal Reservoirs and Their Application", *Geothermics*,**34**, 728-755.

Schilthuis, R.J., (1936) *Active Oil and Energy Trans*. AIME, **118**, 33-52.

Tureyen, O.I., Sarak, H. and Onur, M. (2007) "Assesing Uncertainty in Future Pressure Changes Predicted by Lumped Parameter Models: A Field Application", *32nd Stanford Geothermal Workshop*, Stanford, California, USA, 22 – 24 January.

Zafari, M. and Reynolds, A.C. (2005) "Assesing the Uncertainty in Reservoir Description and Performance Predictions With the Ensemble Kalman Filter" *proceedings of the 2005 SPE Annual Technical Conference and Exhibition*, Dallas, Texas, USA, 9 – 12 Oct.