Robust closed-loop reservoir management using residual analysis

M. Mohsin Siraj[a], Paul M.J. Van den Hof [1a], Jan Dirk Jansen[b]

[a] *Control Systems Group, Department of Electrical Engineering, Eindhoven University of Technology, The Netherlands (email: m.m.siraj@tue.nl, p.m.j.vandenhof@tue.nl)*
[b] *Department of Geoscience and Engineering, Delft University of Technology, The Netherlands. (e-mail: j.d.Jansen@tudelft.nl)*

Abstract

Model-based dynamic optimization of the water-flooding process suffers from high levels of uncertainty. Among different sources of uncertainties, model uncertainty has a dominant effect on this optimization. A traditional way of quantifying uncertainty in robust water-flooding optimization is by considering an ensemble of uncertain model realizations. These models are generally not validated with data and the resulting robust optimization strategies are mostly open-loop or offline, i.e., they do not take account of information revealed over time. The main focus of this work is to develop a closed-loop or online robust optimization scheme that allows the strategy to be updated whenever information (production data in this case) becomes available. The introduced scheme uses the concept of residual analysis as a major ingredient, where the models in an ensemble are confronted with data and an adapted ensemble is formed with only those models that are not invalidated. As a next step, the robust optimization is again performed (i.e., updated/adjusted) with this adapted ensemble. The steps of the introduced closed-loop (online) scheme, i.e., data collection, residual analysis and robust optimization can be repeated at multiple time step. The introduced invalidation (rejection) strategy is compared with common joint parameter/state estimation (history matching) method, Ensemble Kalman Filtering (EnKF). The adapted ensemble gives a less conservative description of uncertainty and also reduces the high computational cost involved in robust optimization. Simulation examples show that an increase in the objective function value with a reduction of uncertainty on these values is obtained with the introduced robust scheme compared to an open-loop offline robust scheme with the full ensemble and with a closed-loop online scheme with EnKF.

Keywords: Closed-loop reservoir management, water-flooding optimization, robust optimization, residual analysis, data assimilation

[1] Corresponding Author

1. Introduction

Various studies have shown that model-based dynamic optimization of the water-flooding process improves the economic life-cycle performance of oil fields, see e.g., [1–3]. One of the key challenges in this optimization is the high levels of uncertainty arising from the modeling process of water flooding and from strongly varying economic conditions. As a result, the potential advantages of dynamic optimization are not fully realized and the optimized objective value is not obtained.

Different approaches to decision making under uncertainty can be broadly divided into two categories, see e.g., [4]. In the first set of approaches, also known as open-loop or offline schemes, a decision maker selects a strategy without knowing the exact values taken by the uncertain parameters, and the exact values are assumed to belong to an uncertainty space. In the second category, also known as closed-loop or online data-based approaches, the strategy is allowed to update/adjust to information that is revealed over time. Uncertainty modeling (quantification) of the uncertainty space Θ is one of the essential steps in these robust approaches. A general practice of quantifying uncertainty in water-flooding optimization is an ensemble (scenario)-based approach where an ensemble of uncertain parameters (models), see e.g., [5], [6] is considered. These models are usually generated with geostatistical tools, see e.g., [7] or, occasionally, hand drawn, and are typically not (in)validated by the production data. Hence they may provide a (very) conservative description of uncertainty. The state of the art in the oil industry is to use a large number of realizations which are considered to be a good representation of the uncertainty space Θ.

In the petroleum engineering literature, open-loop (offline) ensemble (scenario)-based robust approaches have been studied from various perspectives. In [5], a so-called robust optimization approach has been introduced, which maximizes an average NPV over an ensemble of geological model realizations. A mean-variance optimization (MVO) approach honoring geological uncertainty, which maximizes the average NPV and minimizes the variance of the NPV distribution, has been implemented in [8] and extended to consider economic uncertainty in [9]. Similar MVO approaches, e.g., for a well-placement problem, have been described in [10–12]. Different risk averse open-loop strategies, e.g., worst-case robust optimization, CVaR optimization, have been presented, e.g., in [13–16]. The application of closed-loop (online) data-based robust approaches is limited by the complexity of the reservoir models. A closed-loop reservoir management approach has been introduced in [17], where the reservoir model variables (states and/or parameters) are frequently updated using data assimilation or Computer Assisted History Matching (CAHM) techniques such as Ensemble Kalman Filtering (EnKF), Ensemble Kalman Smoothing (EnKS), variational approaches, streamline-based approaches, etc., see e.g., [18], [19], [20]. Model-based dynamic optimization is performed with updated model(s) at each history matching time step. In the robust settings, because robust optimization uses an ensemble of model realizations, the posterior ensemble, e.g., estimated by EnKF, can be directly used in an online fashion. In [21], a closed-loop online optimization method has been introduced that combines adjoint-free ensemble-based optimization (EnOpt) with EnKF and it has been applied to the Brugge field in [22]. Various other authors have implemented closed-loop schemes with different combinations of CAHM and optimization techniques, see, e.g., [23], [24] and [6].

The purpose of this work is to devise a closed-loop (online) robust scheme that can be updated with given production data. The main focus is to address the question: how can the available information (production data) with time be used to shrink the uncertainty space? Then the key issue is to analyze uncertainty reduction and improvement in the robust optimization of the water-flooding process? Traditionally, a joint state/parameter estimation, e.g., by using EnKF, is used for uncertainty reduction in closed-loop reservoir management. We consider a different route and use a rejection mechanism, i.e., residual analysis, to find fewer representative models in an ensemble (i.e. to reduce the ensemble size). In residual analysis, models are confronted with data and are invalidated if they do not sufficiently agree with the observed data. An adapted ensemble is formed with only those models that are not invalidated and it is used in a robust optimization exercise in an closed-loop online fashion. This adapted ensemble provides a better description of uncertainty and it is also beneficial in reducing the computational complexity of robust optimization of the water-flooding process. Rejection sampling, as discussed in [25], also provides a mechanism for rejecting models based on available data. It is a probabilistic approach and requires the knowledge of the likelihood probability. Rejection sampling is not considered in this work. The parameter estimation, due to a large number of to-be-estimated parameters (overparameterized), is an ill-posed problem and therefore questions on the effectiveness of the estimation, e.g., with EnKF, can be raised. For an overparameterized model, the production data does not necessarily map to the relevant (and informative) model information. In [26], it has been stressed that the observation should only be used to falsify possible solutions (e.g., rejecting models), and not to deduce any particular solution (e.g., estimating parameters). To analyse uncertainty reduction and performance of the developed closed-loop scheme with residence analysis, simulation examples are presented and the results are compared with the closed-loop robust scheme with EnKF. This paper is based on the preliminary results of [27] but developed further with extended uncertainty reduction analysis and comparison with EnKF.

The paper is organized as follows: In the next section, a short recapitulation is given of uncertainty quantification in a waterflooding setting and closed-loop reservoir management. The concept of residual analysis with performance measures to define the invalidation test is discussed in section 3. In section 4, the closed-loop (online) robust scheme is introduced. Simulation examples with this online (closed-loop) scheme are given in section 5 and the results of offline (open-loop) and online (closed-loop) robust schemes with EnKF and residual analysis are compared followed by conclusions in section 6.

2. Uncertainty in water-flooding optimization

Typically in model-based optimization of the water-flooding process, a financial measure, i.e., Net Present Value (NPV), is maximized. NPV can be mathematically represented in the usual fashion as:

$$J = \sum_{k=1}^{K} \left[\frac{r_o \cdot q_{o,k} - r_w \cdot q_{w,k} - r_{inj} \cdot q_{inj,k}}{(1+b)^{\frac{t_k}{\tau_t}}} \cdot \Delta t_k \right], \tag{1}$$

where r_o, r_w and r_{inj} are the oil price, the water production cost and the water injection cost in $\frac{\$}{m^3}$ respectively. K represents the production life-cycle i.e., the total number of time steps k and Δt_k the time interval of time step k in days. The term b is the discount rate for a certain reference time τ_t. The terms $q_{o,k}, q_{w,k}$ and $q_{inj,k}$ represent the total flow rate of produced oil, produced water and injected water at time step k in $\frac{m^3}{day}$.

Model uncertainty is the prime source of uncertainty in the model-based optimization of the water-flooding process. Traditionally in water-flooding optimization, an ensemble of uncertain model realizations is considered to quantify the uncertainty space Θ. It is equivalent to descretizing the uncertainty space, i.e., $\{\mathcal{M}(\theta_1), \mathcal{M}(\theta_2), \cdots, \mathcal{M}(\theta_{N_{geo}})\}$, where $\mathcal{M}$ is a model with $\theta_i \in \Theta, i = 1, 2, \cdots, N_{geo}$ a realization of a vector of uncertain parameters. This ensemble-based uncertainty set can be used with various robust schemes. One of the simplest scenario-based robust approaches with adjoint-based well control optimization is to maximize the average of the NPV objective over the model uncertainty ensemble, as introduced in [5]. Robust optimization (or mean optimization (MO)) can be formulated as:

$$J_{MO} = \frac{1}{N_{geo}} \sum_{i=1}^{N_{geo}} J_i(\mathbf{u}, \boldsymbol{\theta}_i), \tag{2}$$

where J_i is the NPV objective and $\mathbf{u}$ is the input decision variable. Other robust approaches in water-flooding optimization, e.g., mean-variance and mean-CVaR have been discussed, e.g., in [8, 10, 13–15]. In these open-loop (offline) approaches, the optimal solution is devised for the complete production life of the reservoir and the uncertainty set and the optimization results are not updated/adapted to the information, e.g., production data, which becomes available over time. These approaches aim only at minimizing the negative effect of uncertainty on the achieved NPV. In other words, the sensitivity of the optimized strategy to uncertainty is reduced. The offline approaches do not focus on the problem of reducing uncertainty, i.e., shrinking the uncertainty space Θ and on minimizing the mismatch between the model and the true system.

The uncertainty reduction can be achieved with the help of available information, e.g., with production data. In the oil industry, the use of model-based optimization with the available sensors and control valves is often referred to as smart fields, intelligent fields, real-time reservoir management, or closed-loop reservoir management as discussed in the next subsection, see e.g., [28, 29].

2.1. (Robust-)closed-loop reservoir management

The concept of closed-loop reservoir management is shown in Fig. 1. The key elements of this approach are dynamic model-based optimization under physical constraints and/or geological uncertainties as indicated by the blue loop, and parameter estimation/data assimilation aimed at continuous updating of the system model(s) as indicated by the red loop.

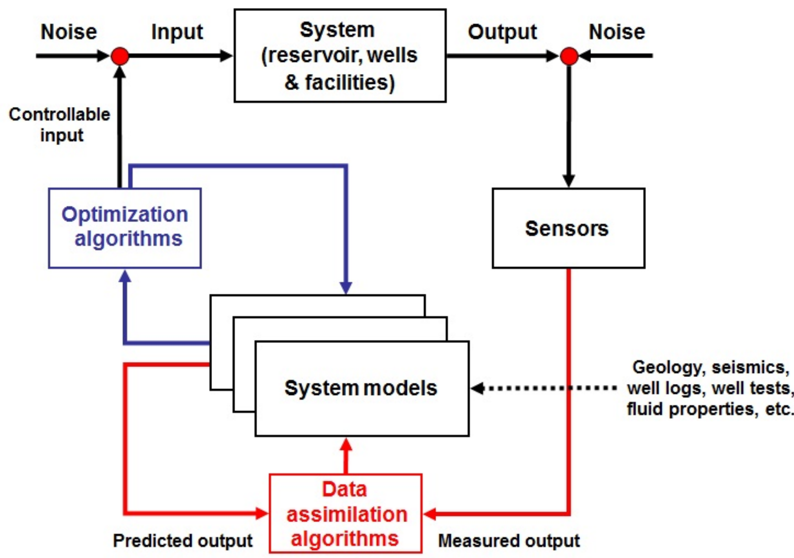


Figure 1: Closed-loop reservoir management (Source: [17])

In this closed-loop reservoir management approach, the estimation of physical parameters with the available production data is one of the ways to reduce uncertainty. Data assimilation or CAHM algorithms such EnKF, EnKS, variational methods, etc., are typically used in reservoir simulation offering either joint state and parameter estimates, or, usually, just parameter estimates. The variational methods minimize the mismatch between the model output and data by using a gradient-based approach where the gradients are obtained by solving adjoint equations, see e.g. [30], [18]. This parameter estimation problem, due to a large number of to-be-estimated parameters, is ill-posed, i.e., many combinations of parameter values will result in the same minimum value of the cost function. Therefore, CAHM typically uses a Bayesian framework with a prior distribution of the parameters reflected by a prior ensemble. Hence the estimation of physical parameters is highly influenced by the selection of this prior ensemble of parameters. The purpose of CAHM is to use production data for uncertainty reduction, but due to the ill-posed nature or overparameterization of the estimation problem the effectiveness of estimation, e.g., with EnKF, in terms of mapping production data into relevant and informative model information can be seriously reduced. These estimated parameters and the resulting adapted ensemble with CAHM can be used in a closed-loop (online) setting to update the robust strategies, resulting in a robust closed-loop reservoir management strategy as presented in [6, 21, 31, 32]. The application of CAHM with nominal optimization, i.e., with a single reservoir model, has been presented e.g., in [33], [17], [23], [34], [24], [29].

Robust optimization is computationally very expensive and as the number of realizations in the adapted ensemble with CAHM are not reduced, the computational complexity of the closed-loop (online) robust optimization steps is not decreased. Another way to adapt the ensemble is by using clustering techniques. In that case, the number of realizations in an ensemble is reduced by clustering the models with similar (static or dynamic) behavior and only a few representative models are used, see e.g., [35, 36]. However, the clustering techniques are not data driven and the ensemble size reduction will normally not result in uncertainty reduction, hence it will only minimize the computational complexity of the robust optimization problem.

3. Residual analysis

Model validation is usually performed by confronting the model with available information, e.g., production data, time-lapse seismic, etc. Validation is important to assess the quality of a model. It is a common practice in regression analysis, where tests are typically defined by computing the model residuals and giving statistics about it, see e.g., [37]. For a given ensemble of models, residual analysis follows an 'exclusion approach' to uncertainty, which focuses on starting from all possibilities (models) and then excluding those possibilities (models) that can be 'rejected' by any information available to us. Therefore, it does not only reduce the size of the ensemble but it also aims at minimizing the uncertainty space Θ. A principle difference between residual analysis and other data assimilation techniques is that residual analysis is performed on the model space which is smaller in size compared to the large parameter space used in data assimilation methods. Hence residual analysis does not suffer from the problem of ill-posedness and the subsequent effect of the selection of a poor prior ensemble. **Fig. 2**, also shown in [27], illustrates a detailed overview of techniques which can be used to update the ensemble of realizations either by using production data or by using clustering techniques. Variational methods, EnKF and clustering techniques have been discussed in the previous section. Rejection sampling, as discussed in [25], is a probabilistic approach for rejecting models based on available data and it requires the knowledge of the likelihood probability, which is the conditional probability of the data given the model. These techniques can be used in a closed-loop online robust optimization setting where the robust optimization is defined over the posterior (adapted) ensemble. For a linear regression problem, first order statistics provide a complete characterization of the validation problem, e.g., in the correlation analysis, the residual should be asymptotically uncorrelated with past input samples. For a nonlinear regression problem, the first order moments are not sufficient to draw

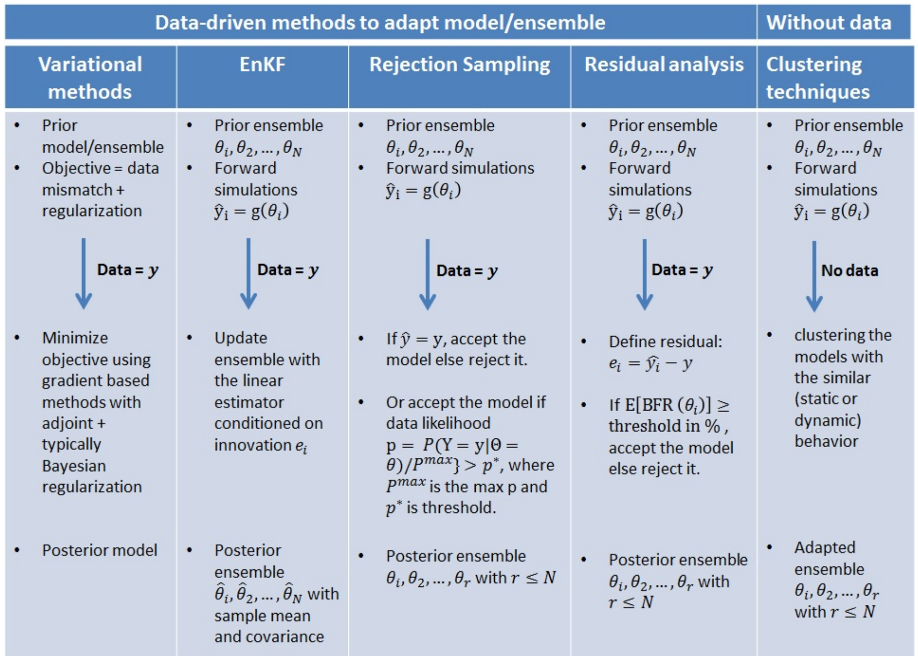


Figure 2: A comparison of methods for updating an ensemble of models to be used with robust optimization ([27])

any conclusions about the validity of the models. Because the reservoir models are strongly nonlinear in nature, in this work a deterministic metric, i.e., the Best-Fit-Ratio (BFR), is used to define an invalidation test. The available production data is used for invalidation.

The residual $\boldsymbol{\epsilon}$ is defined as the difference between the observed (measured) output $\mathbf{y}$ and the simulation output $\hat{\mathbf{y}}$.

$$\text{Residual} = \text{measured output} - \text{simulation output},$$
$$\boldsymbol{\epsilon} = \mathbf{y} - \hat{\mathbf{y}}.$$

The BFR (R_{BF}) or the fit ratio is defined as:

$$R_{BF} = 100\% \times \max\left(1 - \frac{||\mathbf{y} - \hat{\mathbf{y}}||_2}{||\mathbf{y} - \bar{y} \times \mathbf{1}||_2}, 0\right),$$

where $\bar{y}$ is the mean value of the measured output y, i.e., $\bar{y} = \mathbb{E}[\mathbf{y}]$. In our case, it is the average of the measured output $\mathbf{y}$ and $\mathbf{1}$ is a vector of ones. The BFR percentage is a relative measure often used in system identification and a low value of BFR indicates a poor fit to data, see [38]. The BFR is a dimension less quantity and gives an indication of fit in a percentage. Generally, as the reservoir models contain multiple outputs, an average BFR over each individual output channel is considered. The selection of the BFR is not unique and other metrics, e.g., Mean Squared Error (MSE) or Variance Accounted For (VAF), see [38] can also be used for defining the invalidation test. VAF measures how much variation in data (variance) is captured by the model output and disregards the mismatch (bias) of data with model output; therefore it is not considered in this work. The MSE measure is dependent on the units of the physical quantity being measured; therefore BFR will be a used for residual analysis.

The test for invalidating models is given as:

$$\mathcal{M}(\boldsymbol{\theta}_i, \mathbf{u}) \text{ is not invalidated if } \mathbb{E}[R_{BF}(\mathcal{M}(\boldsymbol{\theta}_i, \mathbf{u}))] > 30\%, \quad \text{for a given } \mathbf{u},$$

where $\mathbb{E}[.]$ is the expected value operator and, in our case, it is the average of BFR values of each output channel. It implies that all those models with an average BFR of above 30% threshold are retained in the adapted ensemble. The selection of the threshold is important and a systematic selection of this threshold is considered as future research; here 30% is chosen in an ad-hoc way. One of the risks with this selection criterion is that all the models in an ensemble can be rejected. An alternative choice for the invalidation test is by considering, e.g., 10% of the models who score the highest average BFR within the ensemble. This criterion is not used in this work. A flow chart explaining the concept of rejecting models by residual analysis and adapting the model ensemble to be used with robust optimization in a closed-loop online fashion is shown in **Fig. 3**.

4. A robust closed-loop (online) reservoir management approach

The key elements of the closed-loop online robust strategy using residual analysis are displayed in **Fig. 4**. The top of the figure represents the physical system consisting of reservoirs, wells and facilities with inputs and outputs. The center of the figure displays the residual analysis step, which starts from considering a prior ensemble. The sensors on the right side of the figure are used for measurements which are used to invalidate models with residual analysis resulting in an adapted ensemble. A robust optimization workflow is defined using the adapted ensemble as shown at the left side of the figure. Throughout this work, the MO approach of [5] is used for robust optimization. Other robust measures such as mean-variance, mean-CVaR can also be used. An implementation of the closed-loop (online) robust strategy is presented in the next section to investigate if model invalidation by residual analysis can be an appropriate tool for improving robust optimization of economic performance of oil reservoirs in an online setting. A detailed analysis and comparison is drawn with EnKF in robust closed-loop (online) settings which is a common data assimilation algorithm.

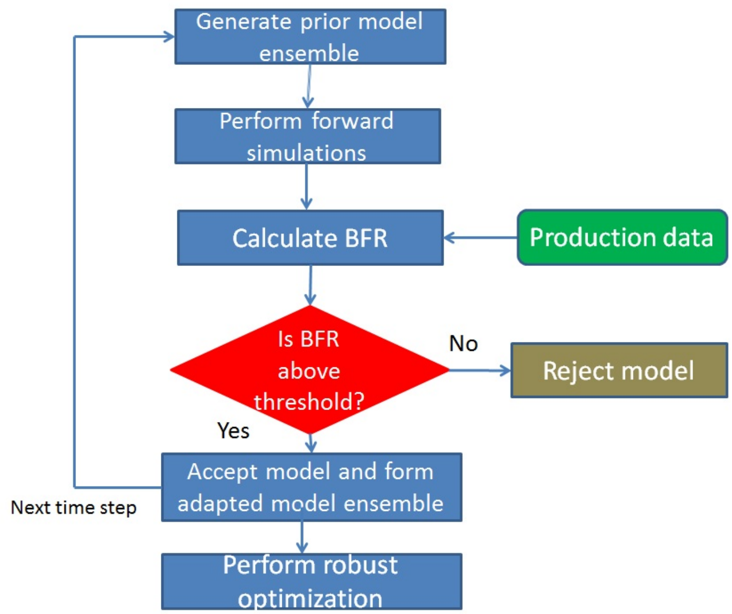


Figure 3: Adapting the model ensemble using residual analysis and robust optimization based on the adapted ensemble ([27])

5. Simulation examples

5.1. Ensemble of reservoir models

We use an ensemble of $N_{geo} = 100$ geological realizations of the standard egg model, see [39]. Each model is a three-dimensional realization of a channelized reservoir produced under water flooding conditions with eight water injectors and four producers based on the original egg model proposed in [5]. The true permeability field is considered to be the unknown parameter and the number of 100 realizations is assumed to be large enough to be a good representation of this parametric uncertainty space. The life cycle of each reservoir model is 3600 days. The absolute-permeability field of the first realization in the set is shown in **Fig. 5**. **Fig. 6** shows the permeability fields of six randomly chosen realizations of the standard egg model in an ensemble of 100 realizations. Each realization in the set is considered equiprobable.

5.2. An open-loop (offline) MO approach with complete ensemble

At first, we perform an open-loop robust (mean) optimization with an objective function as defined in eq. (2) with the complete ensemble. In this open-loop (offline) approach, the optimal solution is devised for the complete production life of the reservoir, i.e., 3600 days, and ensemble is not updated with the available information (production data). All the simulation experiments in this work are performed using MATLAB Reservoir Simulation Toolbox (MRST), see [40].

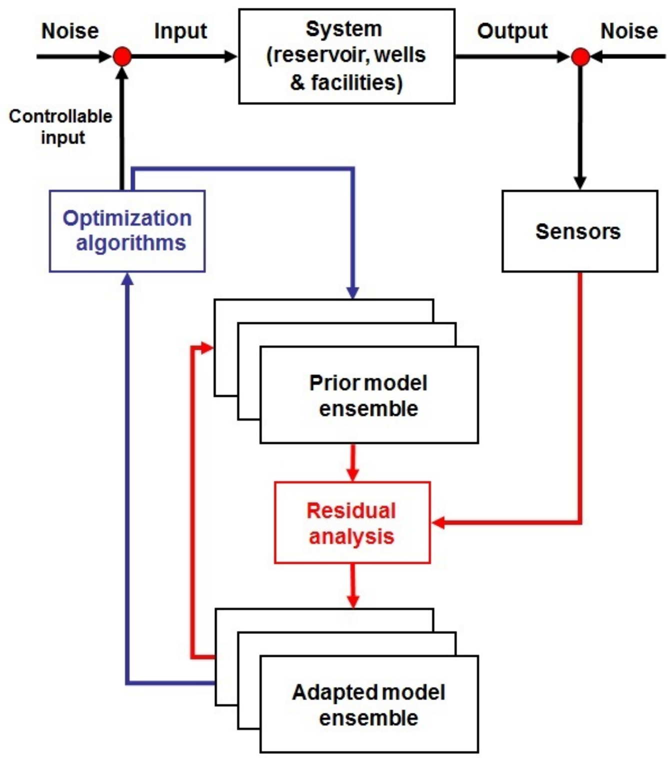


Figure 4: A closed-loop (online) robust approach by updating the ensemble with residual analysis at each time step ([27])

Economic data for NPV

An un-discounted NPV is used. Other economic parameters, i.e., oil price r_o, water injection r_{inj} and production cost r_w are chosen as $126\frac{\$}{m^3}$, $6\frac{\$}{m^3}$ and $19\frac{\$}{m^3}$ respectively.

Control input

The control input $\mathbf{u}$ involves injection flow rate trajectories for each of the eight injection wells. The minimum and the maximum rates for each injection well are set as $0.2\frac{m^3}{day}$ and $79.5\frac{m^3}{day}$ respectively. The production wells operate at a constant bottom-hole pressure of $395bar$. The control input $\mathbf{u}$ is reparameterized in control time intervals with input parameter vector $\boldsymbol{\varphi}$. For each of the eight injection wells, the control input $\mathbf{u}$ is reparameterized into twenty time periods of t_{φ} of 180 days during which the injection rate is held constant at value φ_i. Thus the input parameter vector $\boldsymbol{\varphi}$ consists of $N_u = 8 \times 20 = 160$ elements. The initial input value for the optimization is the maximum possible injection rate, i.e.,$79.5\frac{m^3}{day}$ for each injection well. The optimal input, $\mathbf{u}_{\text{off}}$, obtained by

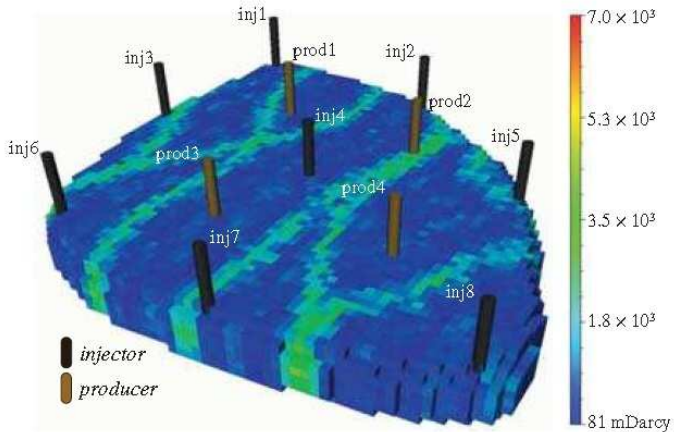


Figure 5: Permeability field of realization 1 of a set of 100 realizations.

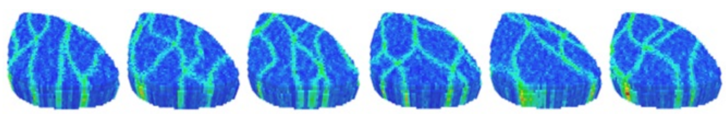


Figure 6: Permeability fields of 6 randomly chosen realizations. ([5])

maximizing average NPV as in eq. (2) with the complete ensemble is shown in **Fig. 7**.

5.3. The closed-loop (online) robust approach with adapted ensemble(s)

In [27], one of the models in the ensemble is considered as the synthetic truth to generate data **y** and the closed-loop online approach has been implemented with residual analysis and the results in terms of obtained NPV and uncertainty reduction are compared with an open-loop offline approach and also with a closed-loop online approach with EnKF. In this work, we have extended the results and provide an in-depth analysis and comparison with EnKF and subsequently consider each model in the ensemble as a plausible truth. Later, the step of robust optimization is performed for each adapted (posterior) ensemble from residual analysis and EnKF respectively. A comparison in terms of the achieved NPV value and an analysis of the uncertainty reduction resulting from residual analysis and EnKF are studied and compared with the open-loop offline approach. Though, the steps of the introduced closed-loop (online) scheme, i.e., data collection, residual analysis and robust optimization, can be repeated at each time step, in this work, these steps are performed at one time step.

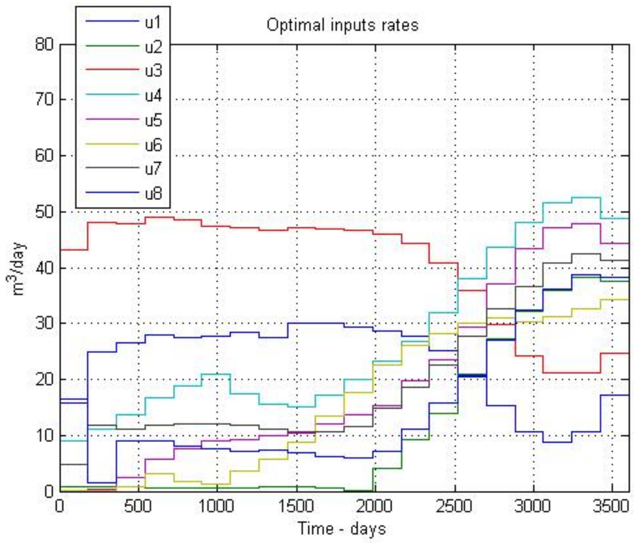


Figure 7: Optimal injection rates input $\mathbf{u}_{\text{off}}$ for robust open-loop offline MO based on the complete ensemble

Residual analysis with plausible synthetic truth models

For a synthetic truth realization, the optimal solution from the open-loop (offline) approach, i.e., $\mathbf{u}_{\text{off}}$ is applied to the respective truth realization to collect synthetic data $\mathbf{y}$. The output $\mathbf{y}$ is defined as the total production rate from each production well. The data is collected at time $t = 360$ days. The input $\mathbf{u}_{\text{off}}$ is also applied to each member of the ensemble to collect simulation data $\hat{\mathbf{y}}$. An average BFR is calculated for each model simulation output $\hat{\mathbf{y}}$ and subsequently the invalidation test is performed. This procedure of residual analysis is then repeated at time $t = 360$ days for each plausible truth realization in the ensemble. For a randomly chosen synthetic truth, i.e., model 10, the average BFR values for each model in the ensemble are shown in **Fig. 8** ([27]). The adapted ensemble for the truth realization number 10 contains 22 models. Similarly, adapted ensembles are formed for each truth realization. In this way 100 adapted ensembles are formed. The adapted ensembles contain only those models that are not invalidated with data and hence they provide a less conservative description of uncertainty. **Fig. 9** shows the number of models appearing in the ensemble adapted after day 360 for each choice of synthetic truth realization. It can be observed that the biggest adapted ensemble has 35 members for the truth realization number 76 while the smallest set has 9 members for truth realization numbers: $29, 30$ and 48.

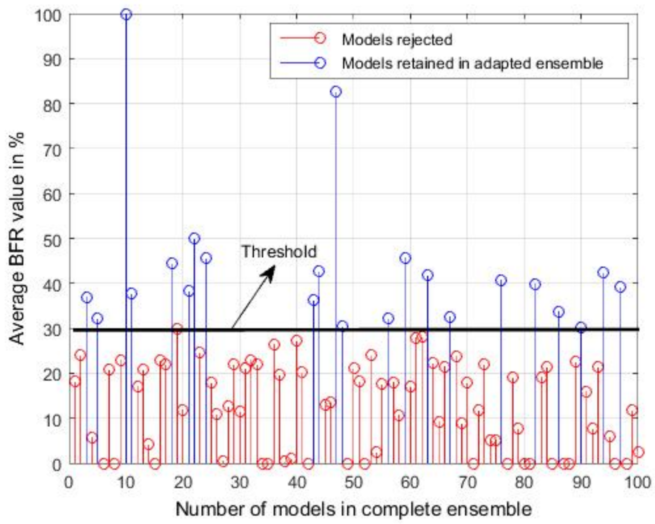


Figure 8: For truth realization number 10: Average BFR values for each model in the ensemble and the models retained in the adapted ensemble([27])

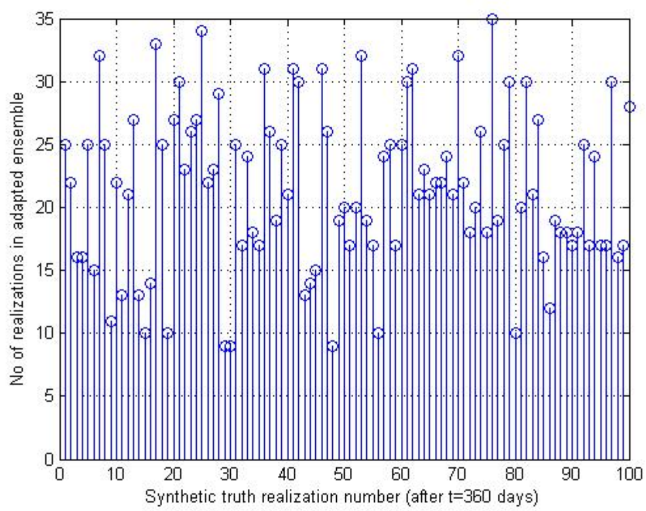


Figure 9: Number of models in the adapted ensembles after $t = 360$ day as a function of the choice of synthetic truth realization number

EnKF with plausible synthetic truth models

The EnKF is a Monte-Carlo implementation of the Kalman filter ([41]), which is suitable for non-linear complex systems and uses an ensemble of N

realizations ([19]). In EnKF, the first and second moments (means and covariances) are represented as sample mean and sample covariances. EnKF is acommonly used data assimilation algorithm in oil reservoirs. Here, EnKF is also implemented with each truth realizations to estimate the permeability field based on production data measurements. Therefore, 100 posterior ensembles at time $t = 360$ days are formed for each synthetic truth. We used a straightforward implementation using the EnKF module of MRST without localization or inflation. The output variable $\mathbf{y}$ and the time of measurement are the same as used in the residual analysis case. In EnKF, all ensemble members are updated, and so this also leads to an adapted (posterior) ensemble which can be directly used in robust optimization for each synthetic truth. However, unlike residual analysis, it has the same number of realizations as the original ensemble. Hence the complexity of robust optimization is not reduced.

MO with the adapted (posterior) ensembles

To 'close' the loop and to implement the robust closed-loop reservoir management approach, robust optimization results are updated/adjusted with information revealed, i.e., MO is again performed with the adapted ensembles generated by residual analysis and EnKF. The economic parameters are kept the same. The initial input value for the optimization is $\mathbf{u}_{\text{off}}$. As the first sample of $\mathbf{u}_{\text{off}}$, i.e., from time 0 to 360 days, is already applied to the system, the remaining part of the input from 360 days to 3600 days (end of simulation time) is used. The time horizon for MO is reduced to $3600 - 360 = 3240$ days. Optimal inputs $\mathbf{u}^i_{\text{on,RA}}$ and $\mathbf{u}^i_{\text{on,EnKF}}$ for $i = 1, \cdots, 100$ are obtained as a result of robust optimization using adapted ensembles from residual analysis and EnKF respectively. The optimal inputs are applied to the respective truth realizations and the NPV values are collected. They are then compared to the NPV obtained by applying $\mathbf{u}_{\text{off}}$ to each truth realization. A percentage change of NPV from the offline strategy for all plausible truths is shown in **Fig. 10**. For the closed-loop online approach with residual analysis, a maximum increase of 6% in NPV is observed for the truth realization number 97. Most of the realizations experience an increase of over 1% in NPV. A decrease in the NPV value is observed for a few truth realizations. A maximum decrease of 0.9% is observed for truth realization number 44. On average over all truth realizations, an increase of 1.03% is obtained. For the closed-loop online approach with EnKF, a maximum increase of 2.7% in NPV is observed for truth realization number 50. In case of EnKF, an increase in NPV is observed for almost half of the truth realizations while for the others a decrease in NPV compared to offline approach is obtained. A maximum decrease of 3.2% is observed for truth realization number 4. On average over all truth realizations, a decrease of 0.005% is obtained. There can be different reasons for this decrease. In case of EnKF, a poor prior can result in a decrease of NPV for the synthetic truth. The results for EnKF can be improved with a better choice of prior ensemble and/or by keeping the channelized structure in the posterior ensembles. The selection of MO as the robust optimization method may not be the best choice as MO does not attempt to minimize the negative effect of uncertainty on the achieved NPV. A better choice of the robust objective, e.g., worst-case optimization or mean-variance optimization may provide improved results. On average, NPV values for the closed-loop approach (online with residual analysis) are improved compared to the offline approach (open-loop with complete ensemble) and closed-loop approach (online with EnKF).

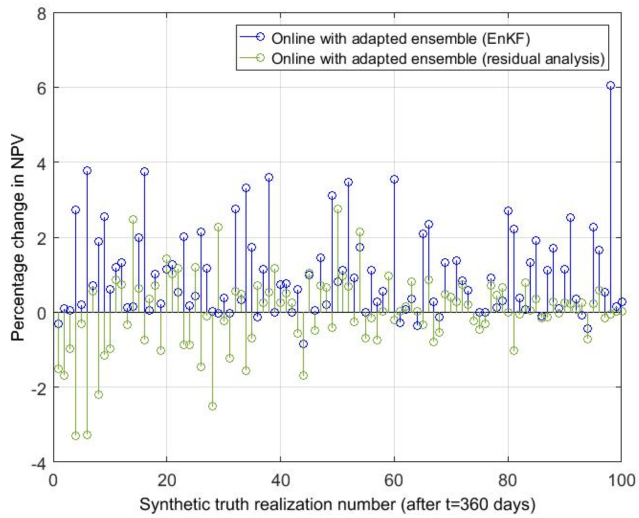


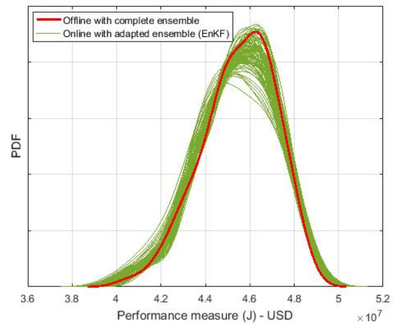
Figure 10: Percentage change in NPV w.r.t the offline strategy (complete ensemble) for online strategies (residual analysis and EnKF) as a function of the choice of synthetic truth realization number. At $t = 360$ days new optimizations have started in the online case.

5.4. Uncertainty reduction with the closed-loop (online) approach

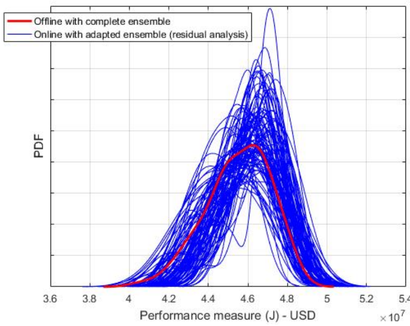
To analyze the uncertainty reduction, the optimal solutions, i.e., $\mathbf{u}_{\text{off}}, \mathbf{u}^i_{\text{on,RA}}$ and $\mathbf{u}^i_{\text{on,EnKF}}$, $i = 1, \cdots, 100$ are applied to the complete and respective adapted/posterior ensembles and NPV points are collected, leading to a distribution of NPV values of each of the truth realizations. The corresponding PDFs are obtained by approximating a non-parametric KDE with MATLAB routine '*ksdensity*' on these NPV data values as shown in **Fig. 11**.

It can be observed that robust optimization with posterior ensembles from EnKF does not deviate and stays close to the NPV distribution from the offline strategy. The NPV distribution from residual analysis has higher variations per ensemble compared to the one from EnKF. For uncertainty quantification, we have considered standard deviation and worst-case value as measures of comparison for each distribution.

Fig. 12 shows the standard deviation of NPV points for all adapted ensembles from residual analysis and EnKF. It also shows the standard deviation of the complete ensemble as a result of $\mathbf{u}_{\text{off}}$. For the case of residual analysis,



(a) PDF of long-term NPV by applying optimal inputs $\mathbf{u}^i_{\text{on,EnKF}}, i = 1, 2, \cdots, 100$ from each truth realization using online (EnKF) and offline approaches to the respective ensembles



(b) PDF of long-term NPV by applying optimal inputs $\mathbf{u}^i_{\text{on,RA}}, i = 1, 2, \cdots, 100$ from each truth realization using online (residual analysis) and offline approaches to the respective ensembles

Figure 11: Results comparison in terms of long-term NPV for offline and online robust strategies with complete and adapted ensembles by residual analysis and EnKF respectively.

standard deviations for most of the adapted ensembles are smaller compared to the standard deviation of the complete ensemble. A maximum decrease of 41.87% in standard deviation from the complete ensemble is observed for truth realization 21. With EnKF, very few adapted ensembles result in a smaller standard deviation compared to the complete ensemble. A maximum decrease of 5.83% in standard deviation from the complete ensemble is observed for truth realization 83.

Another measure for the effect of uncertainty is the worst-case NPV value. **Fig. 13** shows the worst-case values for the adapted ensembles from residual analysis and EnKF for each truth realization and for the complete ensemble. It can be observed that with residual analysis, all adapted ensembles result in higher worst-case values compared to the complete ensemble. For the case of EnKF, most of the posterior ensembles result in a better worst-case value with a few exceptions in a lower worst-case value. The results are summarized in Table 1.

Table 1 - Uncertainty reduction with closed-loop online approaches (residual analysis and EnKF)

Control strategies	Average standard deviation	Average worst-case value	Avg. st. deviation % change from offline	Avg. worst-case % change from offline
Online (residual analysis)	1.3 million USD	42.3 million USD	-17.62%	4.3%
Online (EnKF)	1.6 million USD	40.2 million USD	3.04%	-0.77 %

In summary, in terms of both obtained NPV and uncertainty reduction, residual analysis is shown to perform better compared to EnKF. The main purpose of this work is to use production data to reduce uncertainty, and it is observed that estimation with EnKF, for the case of an overparameterized model such as an oil reservoir model, may not be the best choice to turn production

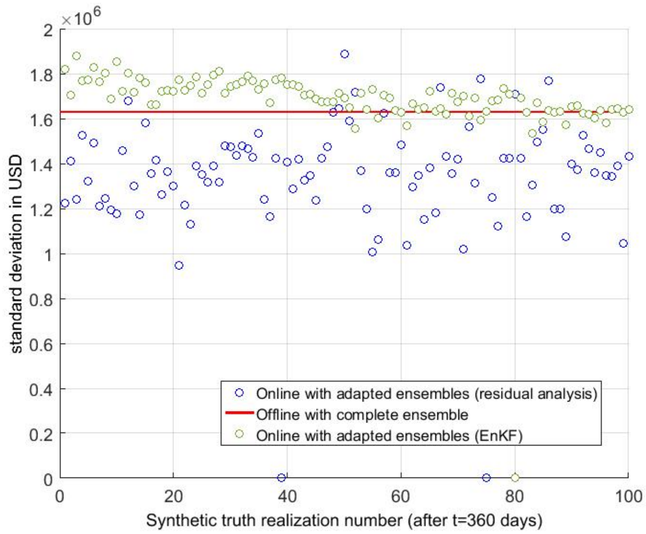


Figure 12: Standard deviation as a result of applying optimal inputs from online strategies for each truth realization ($\mathbf{u}^i_{\text{on,RA}}$ and $\mathbf{u}^i_{\text{on,EnKF}}$), and from the offline strategy ($\mathbf{u}_{\text{off}}$) to the respective adapted/posterior and complete ensembles

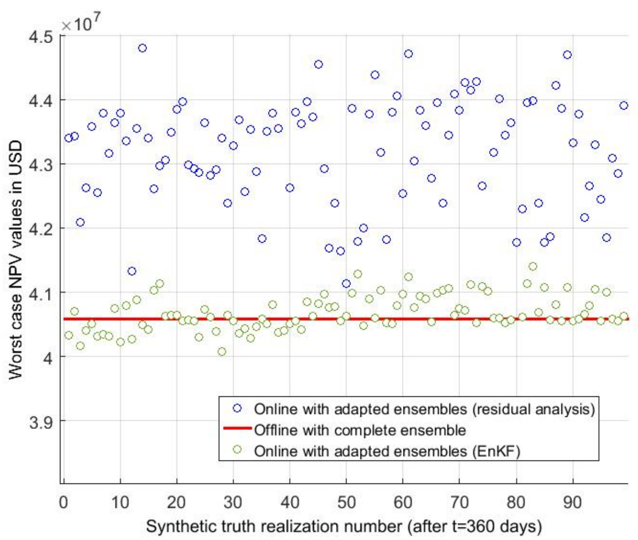


Figure 13: Worst-case values as a result of applying optimal inputs from online strategies for each truth realization ($\mathbf{u}^i_{\text{on,RA}}$ and $\mathbf{u}^i_{\text{on,EnKF}}$), and from the offline strategy ($\mathbf{u}_{\text{off}}$), to the respective adapted/posterior and complete ensembles

data into relevant model (uncertainty) information. It is shown that in this situation the information in the production data which is used to invalidate particular model realizations with residual analysis results in better uncertainty handling.

Another important aspect of robust optimization is the computational complexity. On average, with residual analysis, 21 model realizations are not invalidated per ensemble, compared to 100 model realizations in posterior ensembles with EnKF or the offline approach. Thus the introduced approach reduces the computational complexity of robust optimization by a factor of 80%.

The developed closed-loop (online) optimization workflow, i.e., data collection, residual analysis and re-optimization with the adapted ensemble, can be repeated at each time step of the simulation till the end of the simulation time. Note that in this work, the online (closed-loop) approach is performed at only one time step, i.e., at $t = 360$ days.

5.5. If a model is rejected once, is it rejected for all times?

One of the interesting questions with invalidating models in an ensemble is whether the new adapted ensembles, at later time steps, contain only those models that have been retained in the previously selected adapted ensembles, or also include realizations that have already been rejected. The data is collected from time $t_1 = 360$ days to $t_2 = 720$ days. The optimal inputs obtained from each plausible truth $\mathbf{u}_{\text{on}}$ are applied to the complete ensemble to collect data. **Fig. 14** shows the number of models retained in the adapted ensemble of each plausible truth for time duration, $t_1 = 360$ days to $t_2 = 720$ days. For this example and at this particular time instant, it is observed that all the new adapted ensembles are subsets of the previously selected adapted ensembles from time $t_0 = 0$ day to $t_1 = 360$ day. An interesting observation is that for some truth realizations the adapted ensembles only contain the synthetic truth. It clearly shows the inadequate modeling of the true system, in as much that no single realization in the ensemble is a representative model of reality. One of the remedies to avoid this situation is to generate more models with the dominant features of models that are not invalidated, e.g. by using Kernel PCA techniques, see e.g., [42],[43]. The highest number of models in an ensemble is 5 for truth realization number 27. The step of robust optimization is not repeated here.

6. Conclusions

An closed-loop (online) robust scheme has been presented where robust optimization is updated with adapted ensembles generated by residual analysis.The following conclusions can be drawn:

- The adapted ensembles consist of smaller numbers of representative model realizations ensembles and provide a less conservative description of uncertainty as shown in the simulation examples. The adapted ensembles also substantially reduce the computational complexity of robust optimization.

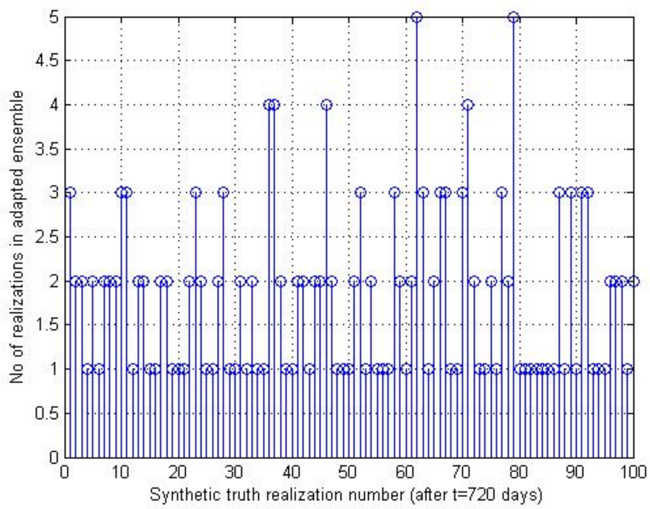


Figure 14: Number of models in the adapted ensembles after $t = 720$ day as a function of the choice of synthetic truth realization number

- The developed closed-loop (online) approach with these adapted ensembles, on average improves the economic performance of the oil reservoir and reduces the effect of uncertainty on the achieved NPV. Reduction of uncertainty is evident by a reduction in variance of the NPV distributions and an improvement in the worst-case performance compared to EnKF.

- The selection of the metric BFR and the 30% threshold is not unique and the threshold is chosen in an ad-hoc way, such that the results highly depend upon this choice.

7. Nomenclature

$$
\begin{aligned}
b &= \text{discount factor,}\\
J &= \text{NPV objective,}\\
J_{MO} &= \text{mean optimization NPV,}\\
K &= \text{total number of time steps,}\\
k &= \text{discrete instants of time,}\\
N_{geo} &= \text{number of model realizations in an ensemble,}\\
N_u &= \text{number of input elements,}\\
q_{inj,k} &= \text{total flow rate of injected water,}\\
q_{o,k} &= \text{total flow rate of produced oil,}\\
q_{w,k} &= \text{total flow rate of produced water,}\\
r_{inj} &= \text{water injection cost,}\\
r_o &= \text{oil price,}\\
r_w &= \text{water production cost,}\\
\Delta t_k &= \text{time interval of time step k,}\\
\mathbf{u} &= \text{control vector,}\\
\hat{\mathbf{y}} &= \text{Simulation output,}\\
\bar{y} &= \text{average output,}\\
\boldsymbol{\theta} &= \text{parameter vector}
\end{aligned}
$$

8. Acknowledgements

The authors acknowledge financial support from the Recovery Factory program sponsored by Shell Global Solutions International.

References

[1] D. R. Brouwer, J. D. Jansen, Dynamic optimization of waterflooding with smart wells using optimal control theory, SPE Journal 9 (04) (2004) 391–402, DOI: 10.2118/78278–PA.

[2] B. Foss, Process control in conventional oil and gas fields – Challenges and opportunities, Control Engineering Practice 20 (10) (2012) 1058–1064.

[3] P. M. J. Van den Hof, J. D. Jansen, A. Heemink, Recent developments in model-based optimization and control of subsurface flow in oil reservoirs, in: Proc. of the 2012 IFAC Workshop on Automatic Control in Offshore Oil and Gas Production, 189–200, 2012.

[4] D. Bertsimas, A. Thiele, Robust and data-driven optimization: Modern decision-making under uncertainty, INFORMS tutorials in operations research: models, methods, and applications for innovative decision making (2006) 137.

[5] G. Van Essen, M. Zandvliet, P. M. J. Van den Hof, O. Bosgra, J. D. Jansen, Robust waterflooding optimization of multiple geological scenarios, SPE Journal 14 (01) (2009) 202–210, DOI: 10.2118/102913–PA.

[6] A. Capolei, E. Suwartadi, B. Foss, J. B. Jørgensen, Waterflooding optimization in uncertain geological scenarios, Computational Geosciences 17 (6) (2013) 991–1013.

[7] G. Mariethoz, J. Caers, Multiple-point Geostatistics: Stochastic Modeling with Training Images, John Wiley & Sons, 2014.

[8] A. Capolei, E. Suwartadi, B. Foss, J. B. Jørgensen, A mean–variance objective for robust production optimization in uncertain geological scenarios, Journal of Petroleum Science and Engineering 125 (2015) 23–37.

[9] M. M. Siraj, P. M. J. Van den Hof, J. D. Jansen, Handling geological and economic uncertainties in balancing short–term and long–term objectives in waterflooding optimization, SPE Journal, DOI: 10.2118/185954-PA. 22 (4) (2017) 1313–1325.

[10] B. Yeten, L. J. Durlofsky, K. Aziz, Optimization of Nonconventional Well Type Location and Trajectory, SPE Journal 8 (3) (2003) 200–210, DOI:10.2118/86880–PA.

[11] W. J. Bailey, B. Couët, D. Wilkinson, Framework for field optimization to maximize asset value, SPE Reservoir Evaluation and Engineering 8 (1) (2005) 7–21, DOI: 10.2118/87026–PA.

[12] E. Yasari, M. R. Pishvaie, F. Khorasheh, K. Salahshoor, R. Kharrat, Application of multi-criterion robust optimization in water-flooding of oil reservoir, Journal of Petroleum Science and Engineering 109 (2013) 1–11, DOI: 10.1016/J.PETROL.2013.07.008.

[13] X. Liu, A. C. Reynolds, Multi Objective optimization for maximizing expectation and minimizing uncertainty or risk with application to optimal wall control, in: Proc. of 2015 SPE Reservoir Simulation Symposium, 23-25 February 2015, Houston, TX, USA., 2015.

[14] A. Capolei, B. Foss, J. B. Jorgensen, Profit and Risk Measures in Oil Production Optimization, in: Proc. of 2nd IFAC Workshop on Automatic Control in Offshore Oil and Gas Production in Florianopolis, Brazil, 2015.

[15] M. M. Siraj, P. M. J. Van den Hof, J. D. Jansen, Robust optimization of water-flooding in oil reservoirs using risk management tools, in: Proc. of 11th IFAC Symp. Dynamics and Control of Process Systems, including

Biosystems (DYCOPS-CAB 2016), 6-8 June 2016, Trondheim, Norway, 133–138, 2016.

[16] M. M. Siraj, Reducing the effect of uncertainty in robust optimization for oil recovery, Ph.D. thesis, TU Eindhoven, Eindhoven University of Technology, 2017.

[17] J. D. Jansen, D. R. Brouwer, G. Naevdal, C. P. J. W. Van Kruijsdijk, Closed-loop reservoir management, First Break 23 (1) (2005) 43–48.

[18] D. S. Oliver, A. C. Reynolds, N. Liu, Inverse theory for petroleum reservoir characterization and history matching, Cambridge University Press, 2008.

[19] G. Evensen, Data Assimilation: the Ensemble Kalman Filter, Springer Science & Business Media, 2009.

[20] S. I. Aanonsen, G. Nævdal, D. S. Oliver, A. C. Reynolds, B. Vallès, The ensemble Kalman filter in reservoir engineering–a review, SPE Journal 14 (03) (2009) 393–412.

[21] Y. Chen, D. S. Oliver, D. Zhang, Efficient ensemble-based closed-loop production optimization, SPE Journal 14 (04) (2009) 634–645.

[22] Y. Chen, D. S. Oliver, Ensemble-based closed-loop optimization applied to Brugge field, SPE Reservoir Evaluation & Engineering 13 (01) (2010) 56–71.

[23] P. Sarma, L. J. Durlofsky, K. Aziz, W. H. Chen, Efficient real-time reservoir management using adjoint-based optimal control and model updating, Computational Geosciences 10 (1) (2006) 3–36.

[24] C. Wang, G. Li, A. C. Reynolds, Production optimization in closed-loop reservoir management, SPE journal 14 (03) (2009) 506–523.

[25] J. Caers, Modeling Uncertainty in the Earth Sciences, John Wiley & Sons, 2011.

[26] A. Tarantola, Popper, Bayes and the inverse problem, Nature physics 2 (8) (2006) 492–494.

[27] M. M. Siraj, P. M. Van den Hof, J. D. Jansen, An adaptive robust optimization scheme for water–flooding optimization in oil reservoirs using residual analysis, in: Preprints 2017 IFAC World Congress, 9-14 July 2017, Toulouse, France, 11767–11771, 2017.

[28] J. D. Jansen, O. H. Bosgra, P. M. J. Van den Hof, Model-based control of multiphase flow in subsurface oil reservoirs, Journal of Process Control 18 (9) (2008) 846–855, DOI: 10.1016/J.JPROCONT.2008.06.011.

[29] J. D. Jansen, S. G. Douma, D. R. Brouwer, P. M. J. Van den Hof, O. H. Bosgra, A. W. Heemink, Closed-loop reservoir management, Paper SPE 119098, SPE Reservoir Simulation Symposium, The Woodlands, Texas, USA .

[30] T.-y. Lee, J. H. Seinfeld, Estimation of absolute and relative permeabilities in petroleum reservoirs, Inverse Problems 3 (4) (1987) 711.

[31] I. Aitokhuehi, L. J. Durlofsky, Optimizing the performance of smart wells in complex reservoirs using continuously updated geological models, Journal of Petroleum Science and Engineering 48 (3) (2005) 254–264.

[32] K. G. Hanssen, A. Codas, B. Foss, Closed-Loop Predictions in Reservoir Management Under Uncertainty, SPE Journal, DOI:10.2118/185956-PA 5 (22) (2017) 1585–1595.

[33] D. R. Brouwer, G. Naevdal, J. D. Jansen, E. H. Vefring, C. P. J. W. Van Kruijsdijk, Improved reservoir management through optimal control and continuous model updating, in: SPE Annual Technical Conference and Exhibition, Houston, USA, 26-29 September., Society of Petroleum Engineers, 2004.

[34] G. Nævdal, D. R. Brouwer, J.-D. Jansen, Waterflooding using closed-loop control, Computational Geosciences 10 (1) (2006) 37–60.

[35] P. Sarma, W. H. Chen, J. Xie, Selecting representative models from a large set of models, in: SPE Reservoir Simulation Symposium, 18-20 February, The Woodlands, Texas, USA, Society of Petroleum Engineers, 2013.

[36] T. H. Yeh, E. Jimenez, G. Van Essen, C. Chen, L. Jin, A. Girardi, P. Gelderblom, L. Horesh, A. R. Conn, Reservoir uncertainty quantification using probabilistic history matching workflow, in: SPE Annual Technical Conference and Exhibition, 27-29 October, Amsterdam, The Netherlands, Society of Petroleum Engineers, 2014.

[37] N. R. Draper, H. Smith, Applied Regression Analysis, John Wiley & Sons, 2014.

[38] L. Ljung, System Identification – Theory for the User, Prentice-Hall, 1999.

[39] J. D. Jansen, R. M. Fonseca, S. Kahrobaei, M. M. Siraj, G. M. Van Essen, P. M. J. Van den Hof, The egg model-a geological ensemble for reservoir simulation, Geoscience Data Journal 1 (2) (2014) 192–195.

[40] K. A. Lie, S. Krogstad, I. S. Ligaarden, J. R. Natvig, H. M. Nilsen, B. Skaflestad, Open-source MATLAB implementation of consistent discretisations on complex grids, Computational Geosciences 16 (2) (2012) 297–322.

[41] R. E. Kalman, A new approach to linear filtering and prediction problems, Journal of basic Engineering 82 (1) (1960) 35–45.

[42] I. Sahni, R. N. Horne, Generating multiple history-matched reservoir model realizations using wavelets, in: SPE Annual Technical Conference and Exhibition, Society of Petroleum Engineers, 2004.

[43] P. Sarma, L. J. Durlofsky, K. Aziz, Kernel principal component analysis for efficient, differentiable parameterization of multipoint geostatistics, Mathematical Geosciences 40 (1) (2008) 3–32.