Dynamic Bayesian Networks as a Possible Alternative to the Ensemble Kalman Filter for Parameter Estimation in Reservoir Engineering

Anca Hanea[a*]**, Remus Hanea**[a,b]**, Aurelius Zilko**[a]

[a]Institute of Applied Mathematics, Delft University of Technology, The Netherlands
[b]The Netherlands Organization for Applied Scientific Research, Utrecht, The Netherlands

Abstract: The objective of reservoir engineering is to optimize hydrocarbon recovery. One of the most common and efficient recovery processes is water injection. The water is pumped into the reservoir in injection wells in order to push the oil trapped in the porous media towards the production wells. The movement of the water front depends on the reservoir characteristics. To avoid an early water breakthrough in the oil production wells, a good knowledge of the subsurface is imperative. The properties of the rock, e.g. porosity and permeability, are therefore important for oil extraction since they influence the ability of fluids to flow through the reservoir. They represent unknown parameters that need to be estimated when simulating a reservoir. In this paper the authors are concerned with estimating the permeability field of a reservoir. To characterise the fluid flow into the reservoir we use a two phase (oil-water) 2D flow model which can be represented as a system of coupled nonlinear partial differential equations which cannot be solved analytically. Consequently, we build a state-space model for the reservoir. There are many ways of representing state-space models, one of the most common being the Kalman Filter (KF) model and its variants, e.g. En(semble)KF. A more general representation is a dynamic Bayesian network. Recently, the performance of the EnKF and that of a non-parametric BN were investigated and compared in a twin experiment for permeability estimation [1]. The NPBN approach proved a promising alternative to the EnKF. Yet a fair number of open questions emerged from the comparison of the two methods. Moreover, in all investigations the assumptions of the EnKF method were used in the NPBN approach for a proper comparison. In this paper we try to answer some of the questions left open, and extend the initial work by making more realistic assumptions.

Keywords: Reservoir Engineering, Parameter Estimation, Ensemble Kalman Filter, Non Parametric Bayesian Networks

1. INTRODUCTION

Reservoir simulation is meant to provide field and research engineers with a quantitative estimate of hydrocarbon volume in a reservoir undergoing production. Both the quality of geological data for a particular reservoir, and the quality of the modelling technique contribute to the uncertainties in the estimates provided by simulation software. When simulating a reservoir one must account for the physical and chemical processes taking place in the subsurface. Rock and fluid properties are crucial when describing the flow in porous media. Large-scale numerical simulation methods for flow through heterogeneous porous media are usually used to design optimal recovery strategies for oil reservoirs. Such reservoir models consist of time and space discretised systems of nonlinear partial differential equations (i.e. state-space models) that describe the time evolution of the variables involved in these models. It is not unusual for reservoir models to contain up to millions of grid blocks with strongly varying values of parameters like the porosity and the permeability of the subsurface.

In this paper the authors are concerned with estimating the permeability field of a reservoir. The problem of estimating model parameters such as permeability is often referred to as a history matching problem in reservoir engineering. To characterise the fluid flow into the reservoir we use a two phase (oil-water), 2D flow model. From the many ways of representing state-space models, we have chosen

two to work with and compare: a Kalman Filter based model and a dynamic Bayesian network based approach.

2. STATE-SPACE MODELS

A general state-space model is in principle any model that includes an observation process Y_t and a state process X_t at the current time t. The equations may be nonlinear, or non-Gaussian. The state process describes an underlying/hidden state of the world that generates observations. This hidden state is represented by a vector of variables that we cannot measure, but whose state we would like to estimate. The state vector evolves in time and the goal of many applications is to infer the hidden state given the observations, $y_1, .., y_t$, up to the current time. Hence the belief state $P(X_t|y_1, ..y_t)$ has to be calculated and this is done recursively using Bayes' rule. A prior, $P(X_1)$, a state-transition function, $P(X_t|X_{t-1})$, and an observation function, $P(Y_t|X_t)$ are needed. We can assume the model is first-order Markov, i.e., $P(X_t|X_1, .., X_{t-1}) = P(X_t|X_{t-1})$. Similarly, we can assume that the observations are conditionally independent given the model, i.e. $P(Y_t|Y_{t-1}, X_t) = P(Y_t|X_t)$.

One of the most common ways of representing a state-space model is the Kalman Filter (KF) model [2]. The KF is based on a system model consisting of a state equation and an output (measurement) equation which are both linear. To model unpredictable disturbances, noise is added to both equations. The state process noise and the measurement noise are assumed to be independent, white, and normally distributed. The KF will recursively calculate the state vector X_t along with its covariance matrix, conditioned on the available measurements up to time t, under the criterion that the estimated error covariance is minimum. Conditioning on the measurements is referred to as the assimilation step of the procedure. The KF method becomes computationally expensive for large scale systems and it is not suitable for nonlinear systems. Several algorithms have been developed in order to overcome these limitations. One such algorithm is the Ensemble Kalman filter (EnKF) [3]. EnKF represents the distribution of the system state using a collection of state vectors, called an *ensemble*, and replaces the covariance matrix by the sample covariance computed from the ensemble. Advancing the probability distribution function in time is achieved by simply advancing each member of the ensemble. The main advantage of the EnKF is that it approximates the covariance matrix from a finite number of ensemble members, thus becoming suitable for large nonlinear problems. Nevertheless, very often the number of variables to be estimated is much larger than the number of ensemble members. In these situations the ensemble covariance is rank deficient, therefore inaccurate. Moreover, EnKF relies on the normality assumption although it is often used in practice for nonlinear problems, where this assumption may not be satisfied. Nevertheless, EnKF is currently one of the most widely used methodologies which address the history matching problem in reservoir engineering [4, 5]. We would like to compare the performance of this state-of-the-art methodology with a new approach that uses another representation of a state-space model, namely a dynamic Bayesian network [6, 7].

Dynamic Bayesian networks provide a more expressive language for representing state-space models. They can be interpreted as instances of a static Bayesian networks (BNs) [8] connected in discrete slices of time. A static BN is a directed acyclic graph (DAG) whose nodes represent univariate random variables, and the arcs represent direct influences. The BN stipulates that each variable is conditionally independent of all predecessors in an ordering of the variables, given its direct predecessors. The direct predecessors of a node i, corresponding to variable X_i are called parents and the set of all i's parents is denoted $Pa(i)$, or $Pa(X_i)$. In most applications, uncertainty distributions do not conform to any parametric form, therefore we shall use non-parametric Bayesian networks (NPBNs) [9] to describe the dynamical system. NPBNs associate nodes with variables for which no marginal distribution is assumed, and arcs with conditional copulae [10, 11]. These conditional copulae, together with the one-dimensional marginal distributions and the conditional independence statements implied by the graph uniquely determine the joint distribution [12]. The (conditional) copulae used in this method are parametrised by (conditional) rank correlations that can be calculated from data or elicited from experts [13]. A dynamic NPBN is a way to extend a static NPBN to model probability distributions

of collections of random variables, $Z_1, Z_2, .., Z_T$. The variables can be partitioned in $Z_t = (X_t, Y_t)$ to represent the hidden and output variables of a state-space model. A dynamic NPBN is defined to be the pair $(B_1, B_\rightarrow)$. B_1 is a NPBN which defines the prior $P(Z_1)$, and $B_\rightarrow$ is a two-slice temporal NPBN which defines $P(Z_t|Z_{t-1}) = \prod_i P(Z_t^i|Pa(Z_t^i))$, with Z_t^i the ith node at time t, which could be a component of X_t, or of Y_t.

The difference between a dynamic NPBN and a KF model is that the latter requires joint normality, whereas a dynamic NPBN allows arbitrary marginal distributions. In addition, a dynamic NPBN allows for a much more general graph structure. Figure 1 presents a general KF model as a dynamic BN.
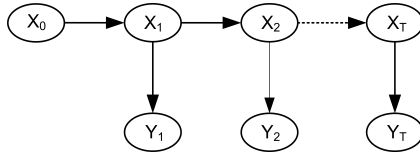

Figure 1: The KF Model as a Dynamic BN

3. EXPERIMENTAL SETUP

This exercise follows a classic *twin-experiment* setup, in which the *true model* is known and used to generate *true synthetic measurements*. These measurements are further used in the assimilation/conditioning step in order to try and recover the aforementioned *truth*. We use a five-spot injection-production strategy. That is to say the reservoir has an injector in the middle of the field (where water is injected) and four producers, one in each corner (where oil is pumped out). The reservoir model considered here is a 2D square with a uniform Cartesian grid consisting of 21 grid cells in each direction. The reservoir is considered a closed space where liquid gets in and out only through the drilled wells. A well model is available for the injection and extraction of fluids through the drilled wells. The flow is specified at the wells by bottom hole pressure (bhp) and fluid flow rates (q) [14]. It is imposed by the well model that either the bottom hole pressure or the fluid flow rates must be prescribed. Let us take the injection well to be constrained by prescribed flow rates and production wells to be constrained by bottom hole pressure. The two phase flow model is combined with the well model and implemented in a simple in-house simulator.

The state vector contains pressures (p) and saturations (S) corresponding to each grid cell. Since we want to perform a parameter estimation, the state vector is augmented with the parameter of interest, i.e. the natural logarithm of the permeability[1] ($\log(k)$). Given the well model constraints, we measure bottom hole pressure at the injector and total flow rates at the producers. The final form of the vector Z_t is:

$$Z_t = \begin{pmatrix} p(t) \\ S(t) \\ \log k(t) \\ bhp(t) \\ q(t) \end{pmatrix}$$

The reservoir is initialized with pressure equal to $3 \cdot 10^7 [Pa]$ and saturation equal to 0.2, in every grid cell. We perform simulations for 480 days, considering measurements every 60 days.

4. COMPARISONS & QUESTIONS

The total number of uncertain variables involved in this toy model is 1328. At each time step the joint distribution of 1328 variables has to be described and updated with measurements. Whereas for

---

[1]We consider the $\log(k)$ instead of k because the values of the permeability are of order $10^{-13} [m^2]$.

the EnKF methodology this number of dimensions does not constitute an impediment, for the NPBN based approach only building a DAG with 1328 nodes can be somewhat cumbersome.

Given the incipient stage of modelling a petroleum engineering problem with a NPBN, a simplification of the graphical model is in order. Previous comparisons were performed in [1]; there, based on expert's opinions, the variables representing saturations were completely excluded from the NPBN analysis. Also, only parts of the grid were considered, rather than the entire reservoir. This implies that in the NPBN based approach, the distribution of only a reduced number of random variables is updated every time step, in contrast with the ENKF method, where all variables are updated in all comparisons.

We start by running the reservoir simulator for the first 60 days, and obtain the multi-dimensional joint distribution (in form of a synthetic data set). We represent the joint distribution using a static NPBN. Using a saturated NPBN translates into representing all possible dependencies present in the data set, including the noisy ones. Another choice is to learn a validated NPBN from data. A learning algorithm is introduced in [15]. The only assumption (to be validated) of the algorithm is that of a joint normal copula. A joint distribution that approximates the distribution given by the simulator is therefore obtained. The static NPBN can be further used to perform the conditionalization/assimilation step. We calculate the joint conditional distribution, given the measurement. After conditioning, we stipulate the conditional distribution by sampling it. Further, we introduce the updated distribution in the simulator, and we run it for another 60 days. In this way we obtain the distribution of the variables after 120 days (with 1 assimilation step after 60 days). The new joint distribution will be modelled with another static NPBN. The 2 NPBNs connected through the simulator are basically a dynamic NPBN. We repeat the above steps for a period of 480 days and thus we build a dynamic NPBN for 8 discrete time steps.

4.1. Previous Results

The results of estimating the permeability values for different locations using a saturated NPBN, a learned NPBN, and the EnKF method are presented in [1]. The estimated parameter mean is compared with the true value of the parameter, and the quality of the estimation is measured using the root mean square error (RMSE).

A 7×7 and a 13×13 grid (that includes the 7×7 grid) were considered in [1]. Figures 2(a) and 3(a) show the true permeabilities, the initial ensemble mean, and the estimated means of permeabilities after 480 days using the EnKF and using the saturated NPBN, for the two grids. Figures 2(b) and 3(b) provide the comparative time behaviour of the RMSE for both methods, in both cases. Both methods use 900 samples/ensemble members in the simulations.

A visual comparison would suggest that the saturated NPBN gives a better estimate than the EnKF for the 7×7 grid. This conclusion is supported by comparing the RMSE values as well (see Figure 2(b)). In the case of a 13×13 grid, Figure 3(a) only indicates that the 2 methods are comparable. However, the RMSE from Figure 3(b) contradicts the visual analysis. The RMSE for the 13×13 field is clearly smaller and more stable for the EnKF than for the saturated NPBN.

4.2. Investigation, Answers & New Questions

Even though the results for the smaller grid look very promising, when we increase the number of variables that are updated in the NPBN based procedure, the quality of the estimation decreases dramatically, at least as measured by the RMSE. Intuitively this was unexpected, since we are actually making more information available for every new time step. The only question remaining is how accurate the extra information is.

It is worth stressing that the RMSE shows only an average behavior of an estimated filed, and it can often be misleading. Keeping this in mind, the first possible explanation for the observed behaviour is
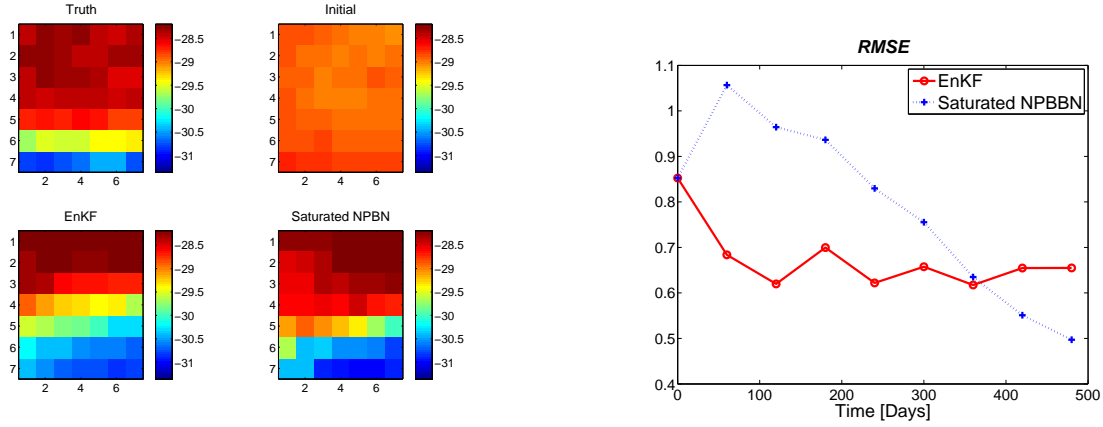
(a) The True (top-left), the Initial (top-right), the EnKF Estimated (bottom-left), and the NPBN Estimated (bottom-right) Permeability Field of a 7 × 7 Grid.

(b) RMSE for the Estimated Permeability Field of a 7 × 7 Grid Using EnKF vs. NPBN.

Figure 2: A 7 × 7 Grid.
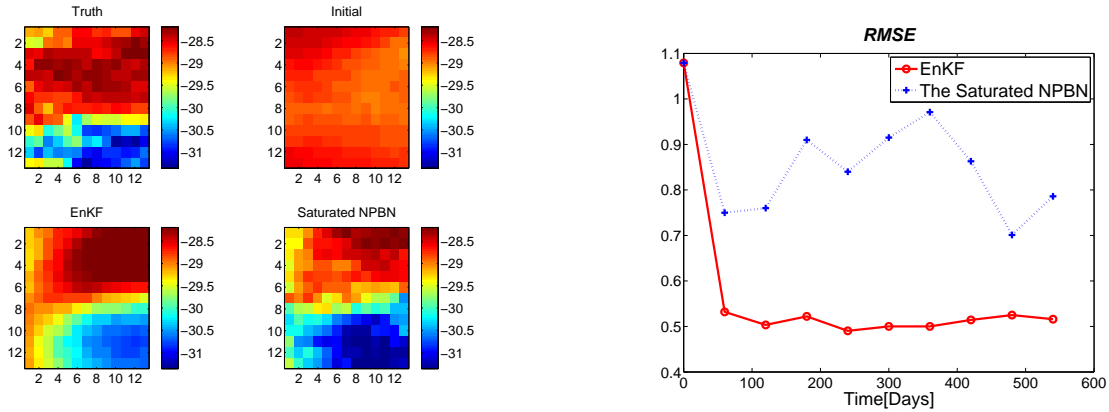
(a) The true (top-left), the Initial (top-right), the EnKF Estimated (bottom-left), and the NPBN Estimated (bottom-right) Permeability Field of a 13 × 13 Grid.

(b) RMSE for the Estimated Permeability Field of a 13 × 13 Grid Using EnKF vs. NPBN.

Figure 3: A 13 × 13 Grid.

the influence of the estimated boundary of the grid. Given the lack of information in the neighboring cells outside the boundary, one could blame the high RMSE on a presumptive bad estimation of the boundary itself. Figure 4(a) shows two RMSEs calculated for the NPBN built for the 13 × 13 grid, on 343 variables. The solid line is the same as the one in Figure 3(b), whereas the dotted line represents the time evolution of the RMSE calculated for the permeabilities of the same grid when the bordering cells are excluded. Not only do the two RMSEs exhibit the same behaviour, but the original RMSE also stays lower than the new one at each time step. Consequently we conclude that the estimated permeabilities on the boundary are not the driving values for the RMSE of the 13 × 13 grid.

It is now interesting to see what we gain or lose in the estimation of a 7 × 7 grid when we update the variables corresponding to the 13 × 13 grid that includes it. Figure 4(b) speaks for itself. The dotted line corresponds to the RMSE of the above described estimation, and the solid line corresponds to the RMSE calculated for the 7 × 7 grid when we update only the variables corresponding to the same 7 × 7 grid. The unstable and bizarre evolution in time of the dotted RMSE gives a strong indication towards not trusting the estimation made using 343 variables. The NPBN based approach, just like the EnKF, is a Monte Carlo method and its performance depends on the number of samples/ensemble members used. Unlike the EnKF, the NPBN approach estimates more than just the first two moments of the distribution, hence it is more prone to statistical errors due to a reduced number of samples.
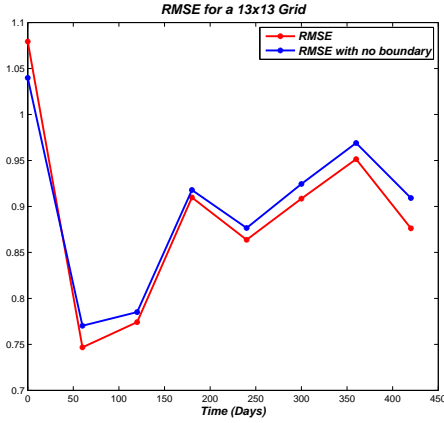
(a) RMSE for 169 vs. 132 Grid Cells.
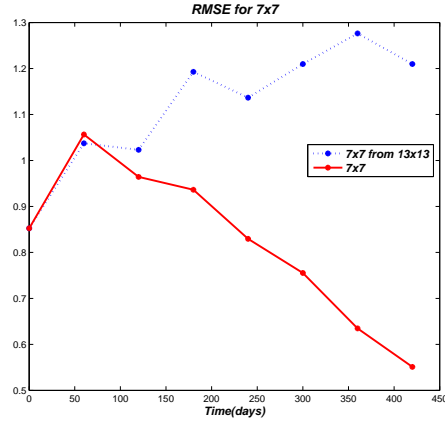
(b) RMSE for the 7×7 Grid when Using 103 vs. 343 Variables in the NPBN.

Figure 4: Boundary Influence Investigation for a 13×13 Grid.

Is the dependence structure (expressed in terms of conditional rank correlations) of a 343 - dimensional distribution well described with 900 samples? Figure 4(b) gives a brisk answer: no. For this reason, the following comparisons and investigations will only be made on 7×7 or smaller grids.

The fact that the NPBN approach needs many samples (many more than the EnKF does) is neither surprising, nor does it constitute an impediment - fast updating is one of the advantages of the NPBNs. A favorable conclusion that can be drawn from the above analysis is that truncating the reservoir domain and estimating different parts of the reservoir can be performed without the menace of losing information from the boundaries. This may suggest that interpolation methods could help in estimating the entire field.

It is obvious from Figures 2(a) and 3(a) that the two methods recover different information about the true permeability field. Extreme values of permeability are definitely of great interest. How the two methods compare in extreme situations is presented in Figures 5(a) and 5(b). We call *false extremes* the situations when the models predict extremely low (solid lines) or extremely high (dotted lines) permeabilities where, in reality, the values are not so low or so high, respectively. We calculate the relative frequencies of such situations for both the EnKF (square markers) and the NPBN (circle markers). The situations when the models fail to find extreme values of permeability are called *missed extremes*. Similar graphical distinctions are used for these in Figure 5(b).
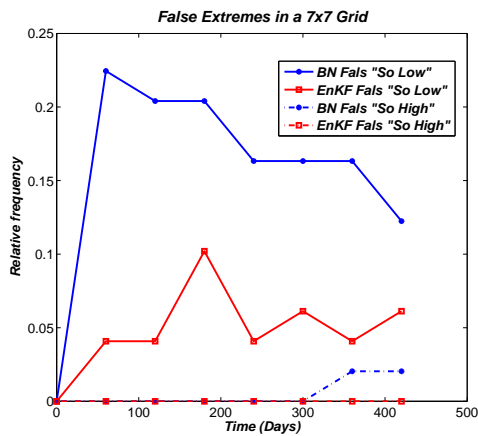

(a) False Extremes.
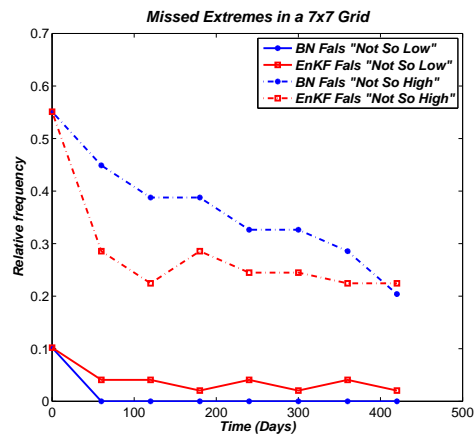
(b) Missed Extremes.

Figure 5: False and Missed Extremes for a 7×7 Grid.

The NPBN model overestimates, much more than the EnKF does, the extremely low permeability

zones, and none of the methods finds extremely high permeability zones where there are none. On the other hand, the NPBN model does not miss the very low permeability zones. It does miss some of the very high permeability ones in the first time steps, but it improves as new information is assimilated through conditioning.

The dependence structure of the modelled variables is estimated differently in the two approaches. Let us compare the correlation matrices of the variables corresponding to the permeability of each grid cell from the 7×7 truncation. We wish to compare the values in the correlation matrices rather than some average measure of dependence (e.g. the determinants, the norms, etc.). To do this we plotted the colour-coded matrices obtained at the last time step of the simulations with the two methods in Figures 6(a) and 6(c). Figure 6(b) shows the numbering of the permeability variables as they correspond to the grid cells. The two matrices exhibit the same patterns of correlation, but the values seem to be higher in the EnKF matrix when first examined.
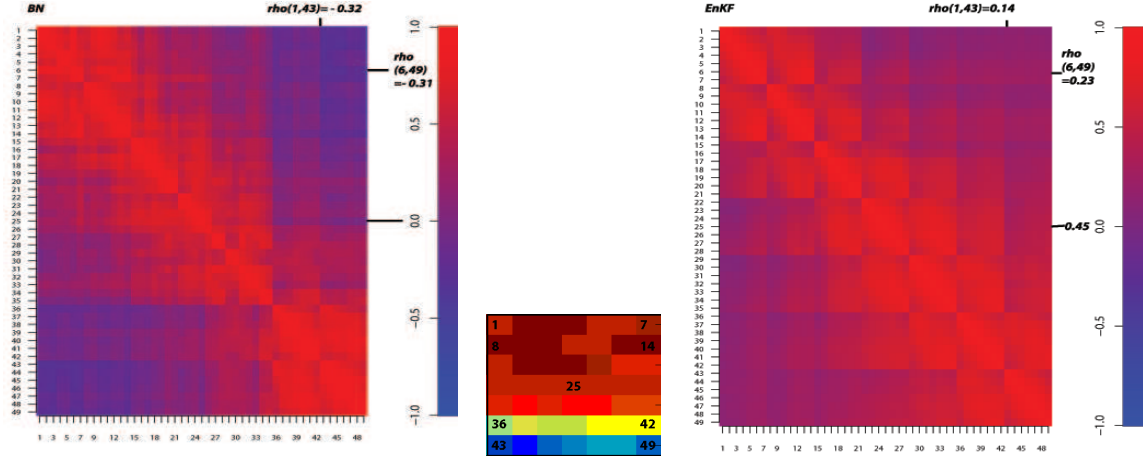
(a) The Correlation Matrix of the Uncertain Permeabilities in Each Grid Cell as Estimated by the NPBN.

(b)

(c) The Correlation Matrix of the Uncertain Permeabilities in Each Grid Cell as Estimated by the EnKF.

Figure 6: Correlation Matrices of the Permeability Variables.

After a closer inspection of the two matrices, we find that the EnKF matrix does not contain negative values, i.e all permeability variables are either positively correlated or uncorrelated. In contrast, the NPBN matrix shows negatively correlated variables as well. When the permeabilities of two far apart grid cells are from the two different ends of the spectrum of permeability values (one very low and one very high), the NPBN matrix shows a negative correlation (e.g. $\rho_{1,43} = -0.32$), whereas the EnKF matrix shows a small positive correlation ($\rho_{1,43} = 0.14$). For closer grid cells with similar values, the NPBN considers them uncorrelated ($\rho_{25,48} = 0$) and the EnKF strongly positively correlated ($\rho_{25,48} = 0.45$). Which of these two descriptions of the dependence structure is more appropriate is a question that remains open. For more insights into subsurface properties like the permeability, expert opinion should be sought.

4.3.   Different Assumptions

The synthetic measurements from the previous comparisons are obtained by adding normally distributed errors (with zero mean and 5% variance) to the production data. Running the simulator, observing the distributions of the flow rates and the bottom hole pressure variables, and assuming that the best estimators are the means of these distributions, allows us to fit different measurement error distributions. For each of the five measurement errors we fit mixtures of Gaussian distributions with different parameters. It turns out that the variance of the bottom whole pressure error is 2% and the variances for the errors of the flow rates are all above 20%. Because of the higher variance of the error, there is a higher chance to obtain *bad measurements*. By *bad measurements* we mean unreliable

and sometimes even unrealistic measurements. Since the uncertainty in the measured variable is not only a result of the uncertainty in the input variables and parameters of the model, the fitted error distribution might still be erroneous.

We consider again the 7×7 grid used previously, but now with the new measurements. Figure 7(a) shows the true permeabilities, the initial ensemble mean, and the estimated means of permeabilities after 420 days and after 480 days using the EnKF and using the saturated NPBN. Figure 7(b) provides the comparative time behaviour of the RMSE for the two methods.
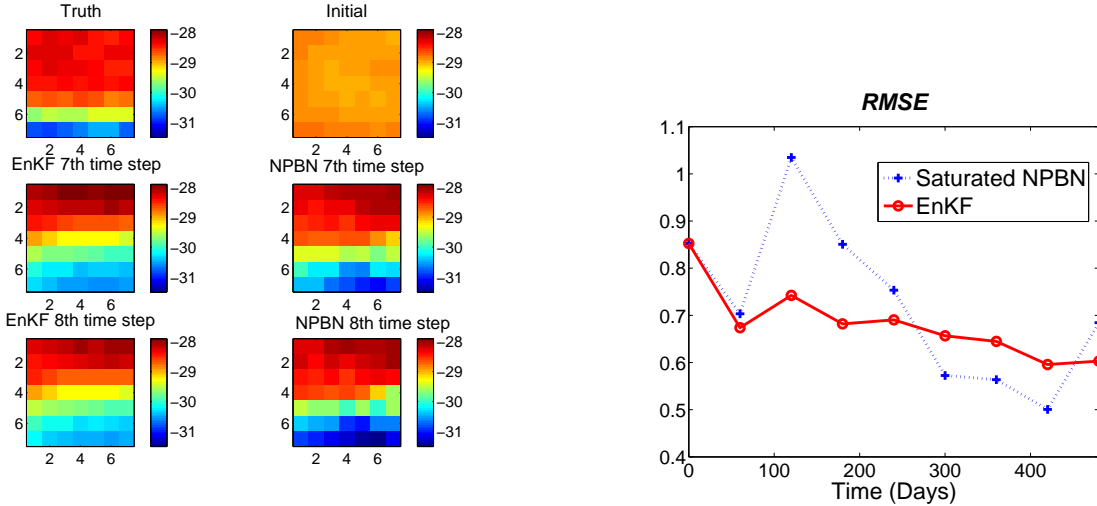
(a) The True (top-left), the Initial (top-right), the EnKF Estimated (bottom-left), and the NPBN Estimated (bottom-right) Permeability Field of a 7×7 Grid with Non-Gaussian Noise.

(b) RMSE for the Estimated Permeability Field of a 7×7 Grid Using EnKF vs. NPBN with Non-Gaussian Noise.

Figure 7: A 7×7 Grid with Non-Gaussian Noise.

The RMSE for EnKF decreases slowly and steadily with an almost constant behaviour for about 4 time steps. The RMSE for the NPBN has a very good decrease between the second and the seventh time steps, but it increases at the second and the eighth steps. The reason for these increases are the *bad measurements* and the adjustment done to those. For the second time step one can observe a slight increase in the RMSE of the EnKF as well but not as high as in the NPBN case. The NPBN method seems to be much more sensitive to the poor measurements than the EnKF method is. On the other hand it also seems to recover fast, outperforming the EnKF until another poor measurement occurs. A visual inspection of the estimated fields for the last two steps shows two very similar pictures given by EnKF, motivating the almost constant value of the corresponding RMSE. The NPBN estimated field from the seventh time step, which approximates the truth better than EnKF does, changes with the new measurement in a field that still looks slightly better than the EnKF estimated field, even though the RMSE values disagree with this visual conclusion. As mentioned earlier, the RMSE as an average measure of fit may sometimes be deceiving.

When we consider a smaller grid of 5×5, closer to the measurement with less variation and further away from one of the bad measurements, we obtain the results from Figures 8(a) and 8(b). Even though, intuitively, both methods should benefit from the position of this field, Figure 8(b) indicates otherwise. The RMSE for the NPBN decreases steadily with the exception of the two time steps with bad measurements, only that the increase from both times is negligible when compared with the one in Figure 7(b). The RMSE of the EnKF method, on the contrary, increases almost every time step. Not only that EnKF does not seem to appreciate the proximity of a good measurement, it also performs worse than in the previous case.

A visual comparison of the fields in Figure 8(a) is somewhat disappointing. The fields estimated using EnKF are very smooth, unlike the true one, and fail to recover both the very permeable and the very impermeable areas. Yet the permeable grid cells are much better approximated than in the fields

(a) The True (top-left), the Initial (top-right), the EnKF Estimated (bottom-left), and the NPBN Estimated (bottom-right) Permeability Field of a 5×5 Grid with Non-Gaussian noise.
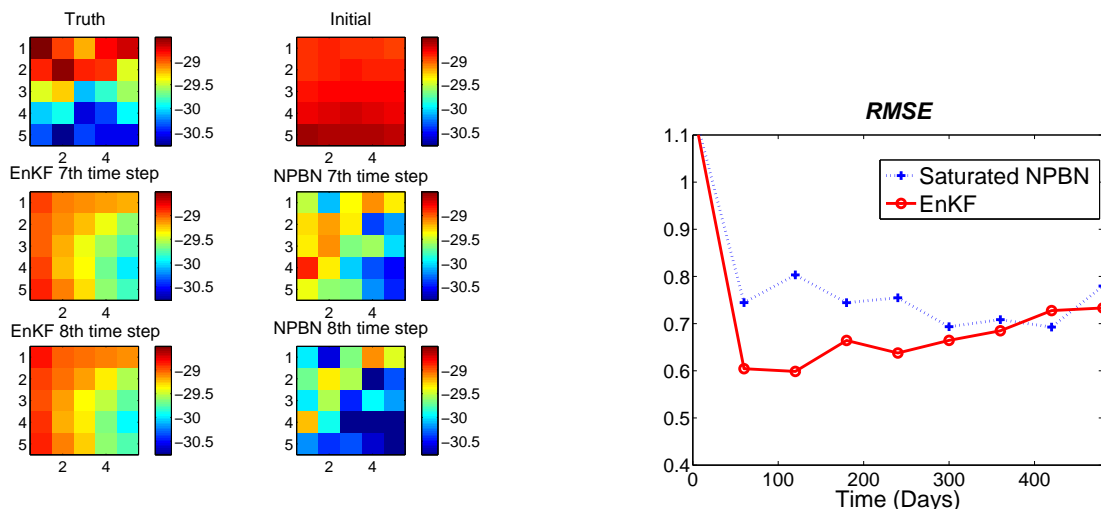
(b) RMSE for the Estimated Permeability Field of a 5 × 5 Grid Using EnKF vs. NPBN with Non-Gaussian Noise.

Figure 8: A 5 × 5 Grid with Non-Gaussian Noise.

estimated using the NPBN. The NPBN has its merits in estimating the very impermeable grid cells. In fact the NPBN overestimates the impermeability of the field and this agrees with the findings from Figure 5(a).

We were puzzled to observe that for very similar values of the RMSE, the estimated fields could look so different to each other and to the truth. Moreover there is no visual correspondence between a 0.7 RMSE for the NPBN in the 7 × 7 grid (eighth time step) and a 0.7 RMSE for the NPBN in the 5 × 5 grid (seventh time step). The NPBN estimated field from the eighth time step, in the 7 × 7 grid, approximates the truth fairly well, while the one from the seventh time step, in the 5 × 5 grid, shows a completely different field from the true one. These add to our concern about using the RMSE as a measure of performance.

5. CONCLUSIONS

The NPBN based approach for parameter estimation in reservoir engineering is still in its infancy, but there are reasons to believe it is a promising approach. The investigations described in this paper showed that even though, with 900 samples, the number of variables that a NPBN could handle in an accurate manner is around one hundred, estimating truncated parts of the reservoir domain does not affect the accuracy of the estimation. In light of these findings estimating parts of the reservoir and interpolating the resulting subfields seems a possible way of estimating the entire reservoir.

An analysis of the methods' behaviour with respect to extreme permeabilities shows that the NPBN approach is very cautious in terms of impermeability, overestimating it, but in general it exhibits a more adventurous performance, when recovering heterogeneous fields. On the contrary all EnKF estimated fields have a very smooth transition between different permeabilities.

The two methods represent different spatial dependence between the permeability variables. The NPBN analysis finds both positively and negatively correlated variables, whereas the EnKF methodology finds mostly positively correlated variables. Experts should further assert the validity of either method.

The NPBN methodology appears to be very sensitive to *bad measurements*, much more so than the EnKF, hence the former would greatly benefit from a better estimation of the measurement noise.

Another possible improvement of the NPBN method is the incorporation of saturation variables in the model. The EnKF method uses continuous variables to represent saturation. A better representation of saturation would use discrete random variables. Unlike EnKF, a NPBN can model both discrete and continuous variables.

REFERENCES

[1] A.M. Hanea and M. Gheorghe. Parameter estimation in a reservoir engineering application. In *Proceedings of ESREL 2011 - European Safety and Reliability Conference 2011*, 2011.

[2] R.E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, page 3545, 1960.

[3] G. Evensen. Sequential data assimilation with nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C6):1014310162, 1994.

[4] G. Evensen, J. Hove, H.C. Meisingset, E. Reiso, K.S. Seim, and O. Espelid. Using the EnKF for assisted history matching of a North Sea reservoir model. *SPE Reservoir Simulation Symposium*, Houston(Texas):USA, 2007.

[5] S.I. Aanonsen, G. Naedval, D.S. Oliver, A.C. Reynolds, and B. Valles. The Ensemble Kalman Filter in Reservoir Engineering. *SPE Journal*, 2009.

[6] T. Dean and K. Kanazawa. A model for reasoning about persistence and causation. *Artificial Intelligence*, 93(1-2):127, 1989.

[7] T. Dean and M. Wellman. Planning and control. *Morgan Kaufmann*, 1991.

[8] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufman Publishers, San Mateo, 1988.

[9] A.M. Hanea. *Algorithms for Non-parametric Bayesian belief nets*. PhD thesis, TU Delft, Delft, the Netherlands, 2008.

[10] H. Joe. *Multivariate Models and Dependence Concepts*. Chapman & Hall, London, 1997.

[11] R.B. Nelsen. *An Introduction to Copulas*. Lecture Notes in Statistics. Springer - Verlag, New York, 1999.

[12] A.M. Hanea, D. Kurowicka, and R.M. Cooke. Hybrid Method for Quantifying and Analyzing Bayesian Belief Nets. *Quality and Reliability Engineering International*, 22(6):613–729, 2006.

[13] O. Morales, D. Kurowicka, and A. Roelen. Eliciting conditional and unconditional rank correlations from conditional probabilities. *Reliability Engineering and System Safety*, 2007. doi: 10.1016/j.ress.2007.03.020.

[14] J.D. Jansen. The use of principal analysis (pca) to create constrained geological realizations from a training image. Technical report, TU Delft.

[15] A.M. Hanea, D. Kurowicka, R.M. Cooke, and D.A. Ababei. Mining and visualising ordinal data with non-parametric continuous BBNs. *Computational Statistics and Data Analysis*, 54(3):668–687, 2010.