Building Bayesian Networks from Basin Modeling scenarios for Decision Making under geologic uncertainty

Gabriele Martinelli, Sara Tviberg, Jo Eidsvik, Richard Sinding-Larsen and Tapan Mukerji

Abstract Basin and Petroleum Systems Modeling is important for understanding the geological mechanisms that characterize reservoir units. Bayesian Networks are useful for decision making in geological prospect analysis and exploration. We unify these two methodologies in this paper. The probabilistic description of the Bayesian Network is trained by using multiple scenarios of Basin and Petroleum Systems Modeling. A range of different input parameters are used for total organic content, heat flow, porosity, and faulting, to span a full categorical design for the Basin and Petroleum Systems Modeling scenarios. Given the consistent Bayesian Network for trap, reservoir and source attributes, we demonstrate important decision making applications such as evidence propagation and the value of information.

1 Introduction

The correct integration of geological and geophysical information remains a challenge in oil and gas exploration that will increase in importance with increasing costs of new targets. Currently it is common practice among geologists to quantify information about risk through detailed exploration analysis, and then forward these

Gabriele Martinelli
NTNU, Trondheim, Norway e-mail: gabriele.martinelli@math.ntnu.no

Sara Tviberg
NTNU, Trondheim, Norway e-mail: tviberg@stud.ntnu.no

Jo Eidsvik
NTNU, Trondheim, Norway e-mail: joeid@math.ntnu.no

Richard Sinding-Larsen
NTNU, Trondheim, Norway e-mail: richard.sinding-larsen@ntnu.no

Tapan Mukerji
Stanford University, Palo Alto, USA e-mail: mukerji@stanford.edu

Ninth International Geostatistics Congress, Oslo, Norway, June 11. – 15., 2012

results to the management. In this work we propose a workflow integrating directly basin modeling scenarios and decision strategies in a Bayesian Network (BN) formalism.

The idea of modeling play element prospect dependencies with BNs was proposed in [10], where a BN model was constructed for modeling the likelihood of source presence in a part of the North Sea. One of the critical points of [10] was the substantial belief in expert opinion. In the present paper we therefore propose an alternative idea for building the BN, integrating expert opinions with hard data coming from standard geological modeling procedures. The main idea is to construct a BN model that is consistent with the results of several basin modeling outputs. We train the probabilistic structure of the BN from the multiple basin modeling outputs.

The idea of integrating statistical design of experiment (DOE) with oil and gas problems is not new: [5] and [6] propose a DOE based approach for reservoir modeling simulations; more recently [4] extends DOE and MonteCarlo (MC) methods in order to study uncertainties in geophysics, geology and reservoir engineering.

The abstract is organized as follows: In Section 2 we introduce Basin and Petroleum System Modeling (BPSM) and the case study; in Section 3 we discuss the DOE simulation design. In Section 4 we show the procedure for developing the BN model. Finally, in Section 5 we present possible applications and discuss results and conclusions.

2 A Case study for basin and petroleum systems modeling

The purpose of BPSM [9] is to simulate the geological and chemical reactions that have occurred in the basin through geological time, in order to identify the critical aspects of the HC generation, migration and accumulation. The main geological risk factors in oil and gas exploration are the trap (consisting of trap geometry, reservoir and seal), the oil and gas charge (reservoir and source factors), and the timing relationship between the charge and the potential traps.

We have decided to use as training model a synthetic basin developed in the Petroleum Geology class at NTNU, Trondheim, Norway. The controlled basin environment is called Bezurk Basin (Figure 1), and it includes three potential kinds of prospects, namely anticlinal prospects, fault prospects and a shoestring prospect. The latter is located within impermeable shale and consequently the chances of HCs migrating into this reservoir are low. The Bezurk basin mimic the behavior of a possible real basin with a main anticlinal trap on the NE sector of the basin, and a series of faults in the NS direction. A major uplift followed by a strong erosion has occurred in the western part of the basin, and this activity has caused the major faulting resulted in Faults 1 and 2.

We have identified 2 main plays, corresponding to the two potential reservoir rocks (see Figure 1):

- The reservoir of the Mmd play in the Bezurk Basin is made up of sandstone, deposited in a regressive shallow marine environment during the time interval

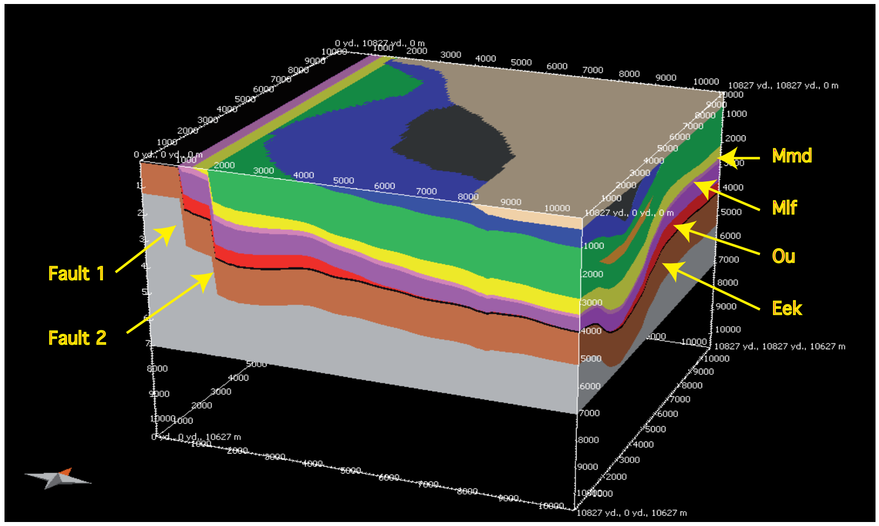

Fig. 1 Bezurk basin; we see the 100 km^2 area and the different thicknesses of the layers; in the west part of the basin we identify the two faults that characterize the system.

20Ma to 15Ma. The sandstone reservoir has porosity ranging from 12% to 30%, which is considered to be a good porosity. The reservoir covers the whole area on the east side of the faults, and has a thickness ranging from about 300-900m. We have distinguished two possible accumulations for this play, one in the eastern part of the basin, under the anticlinal trap, and one in the western part, against the fault trap. We name the accumulations TE (Top East) and TW (Top West).

- The reservoir of Ou play is deposited from 34Ma to 23Ma in a transgressive shallow marine environment with the overlying Mlf shale acting as a seal. HCs are expected to generate in the underlying Eek-coal that is deposited on a coastal plain in the same transgressive system as the reservoir. The potential traps are the western faults and the northeastern anticlinal, which are the same potential traps as the younger and priory outlined Mmd-play. The porosity of the Ou reservoir ranges from 7% to about 20%, which overall is lower than compared to the Mmd reservoir. Both reservoirs have the same kind of sandstone, but due to compaction the lower reservoir (Ou) has a lower porosity than the upper reservoir (Mmd). Also for this play we have identified two possible accumulations in the anticlinal and in the fault trap, and we have named them BE (Bottom East) and BW (Bottom West).

3 Basin modeling scenarios

During a preliminary analysis of the basin we have been able to identify four critical elements that constitute possible sources of uncertainty in our model:

1. The assigned **Total Organic Carbon** (TOC) for the source rocks was vaguely specified in the original model. The TOC is a measure of the concentration of organic material in source rocks [1], and is reported by the weight percentage of organic carbon.
2. The same argument hold for **porosity** to the reservoir rocks, whose original data are rather uncertain.
3. The boundary conditions, an in particular the **Heat Flow** (HF), could represent an exogenous source of uncertainty in our model.
4. There is a zone in the western part of the basin characterized by a prominent faulting activity; for this reason we can hypothetize a possible structural uncertainty, by adding or removing a **fault element** referred to as fault 3.

3.1 A full factorial design

In order to study the interactions among these different factors, we have designed a full factorial study [7], where each factor is represented by two to three levels. We have chosen three levels for the HF (HF): cool (50 mW/m^2), normal (60 mW/m^2) or hot (70 mW/m^2); it is expected that a cool basin mainly will stay in the oil window, consequently generating mostly oil, while a warm basin will reach the gas window at an earlier stage, and therefore generate more gas. We have further chosen two levels for the porosity of the reservoir rock, with porosity/depth functions starting from different values . We use two levels for the TOC content of both source rocks, with TOC ranging from 8% to 4% for the Mlf black shale and from 20% to 10% for the Eek coal. Finally, we select two levels, open or close, for a presence of a new fault (Fault 3) located west of Fault 2. From the first simulation, we observe that the HCs which accumulated in The Fault Trap were lost during the time period of 1.77Ma to 1.55Ma. The reason for adding the Fault 3 is to see if this fault could trap HCs and potentially create a prospect.

3.2 Simulation outcomes

From the different runs in the BPSM software Petromod, we measure the size and kind of HC accumulations. We further measure which source rock has generated them and we observe the migration path. We also gain insights of HC production, the expulsion from the source rock and accumulation in the reservoirs (see Table 1). As a result, the amounts of HC that have leaked is available, and we can try to explain this leakage phenomenon through the observation of the complete evolution of the basin.

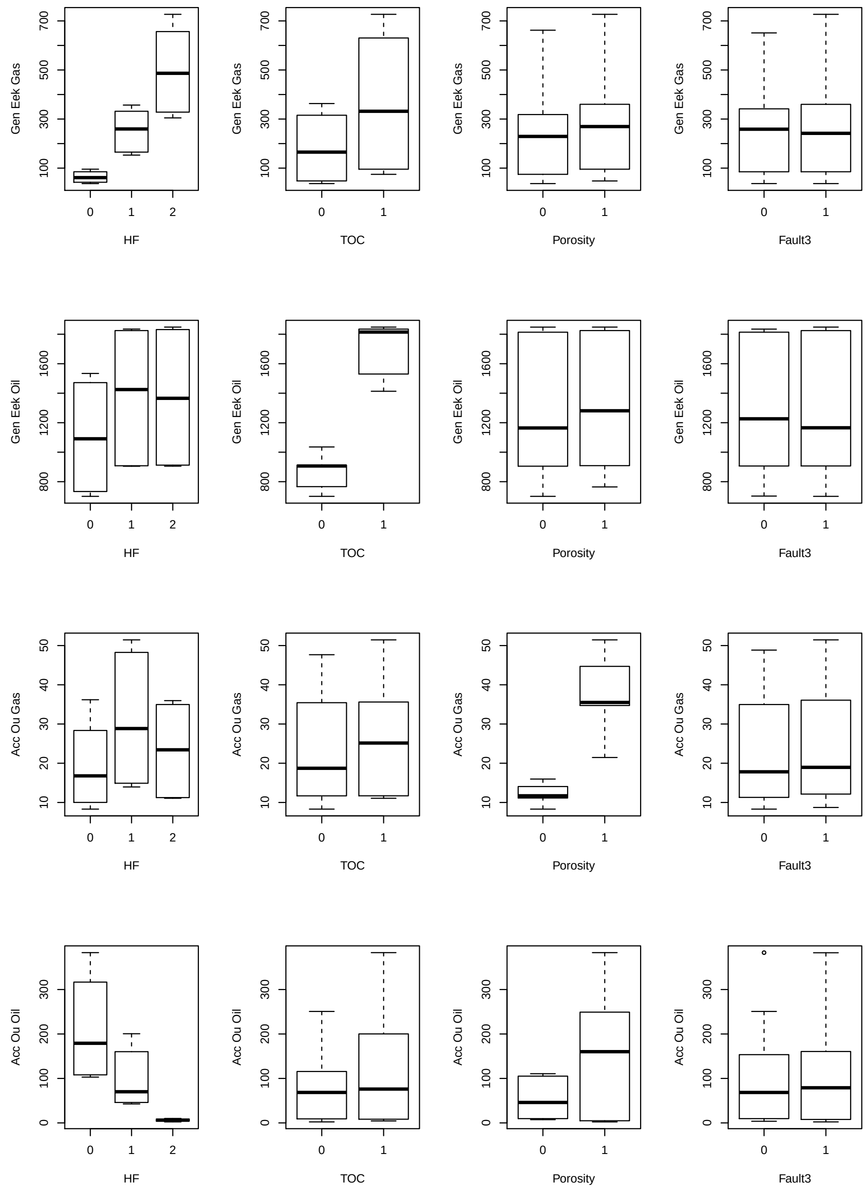

Fig. 2 Boxplot, oil and gas generation and accumulation data for play Ou (and corresponding source rock Eek)

The main factor driving the HC generation is the level of maturation of the source rock itself, that ultimately depends on the HF and on the TOC. The analysis shows that higher HF allows a faster maturation and therefore a more abundant generation (see Figure 2, 1st and 2nd rows) of gas in both the source rocks, while for what concerns the oil generation there are no significant differences in the impact of HF for the Eek source rock. This means that the oil generation has reached the maximum potential already when HF is on the medium level, and this is consistent with our hypothesis.

For what concerns the accumulations (Figure 2, 3rd and 4th rows), we can see that the main factor is the porosity, followed by the HF again, especially for what concerns the Ou accumulations. It is quite natural that the porosity is relevant, since a sandstone with good porosity can trap much more HC than a bad reservoir. It is interesting to notice how the effects of HF and TOC tend to disappear, showing that the surplus of HC generated has almost totally been lost before the seal rock was deposited.

Finally, Fault 3 in the western part has a strong effect only when it is leaking. When the fault is not present, there is no leaking through the fault's wall. On the contrary, when the fault is present, there is some leaking against the fault, especially when there is an early maturation (high HF). The fault has clearest effect when measuring the outflow from the side. In contrast, the outflow from the top and the total outflow is governed by the HF and TOC, since the scenarios with early maturation leak most of the HC before the seal is formed.

Table 1 Simulation data for the first 12 scenarios, concerning the generation, expulsion, accumulation and leaking of the HC in the whole basin. Values are in MMBOE (millions barrels oil equivalents).

Model	1	2	3	4	5	6	7	8	9	10	11	12
Gen Mlf	2422	2057	5783	5193	9828	9239	2421	2054.8	5774	5182	8903	8477
Gen Mlf Gas	69	55	259	215	902	762	69	55	258	214	634	554
Gen Mlf Oil	2353	2002	5524	4978	8926	8477	2352	1999	5516	4968	8269	7923
Gen Eek	1621	1488	2191	2121	2575	2510	1629	1491	2191	2121	2466	2423
Gen Eek Gas	95	74	356	307	727	662	95	74	357	306	651	610
Gen Eek Oil	1526	1413	1834	1814	1848	1848	1534	1417.2	1834	1814	1815	1813
Gen Tot	4044	3545	7975	7314	12403	11749	4051	3546	7966	7304	11369	10900
Exp Mlf	1468	1152	4830	4202	9334	8656	1465	1148	4811	4184	8294	7800
Exp Eek	1479	1332	2131	2051	2552	2483	1487	1336	2132	2052	2443	2396
Exp Tot	2947	2484	6961	6252	11886	11139	2952	2485	6944	6236	10737	10196
Acc Mmd	694	282	951	268	869	252	579	263	878	245	809	233
Acc Ou	418	121	252	65	39	18	417	114	248	56	40	21
Acc Tot	1113	403	1203	333	909	270	997	377	1126	302	850	253
Outflow Top	0	0	977	1260	3837	4297	300	238	2427	2355	4666	4834
Outflow Side	1825	2076	4736	4634	7075	6527	1618	1858	2929	3313	3293	3480
HC losses	1834	2080	5757	5920	10977	10869	1955	2107	5817	5933	9887	9943

4 Building the network

The experimental design setup gives better insight of the key factors responsible for the main geological processes in the basin. We will next use the multiple-scenario information to build a dependency structure for the segments. This takes the form of a BN that will be useful for decision making.

A BN is characterized by a set of nodes and edges. The nodes are random variables, that may be discrete or continuous. As an example, we will define nodes for trap presence (on/off), which is a binary random variable. Edges define the conditional probability structure of the variables, connecting parents to children. For instance, we will define a parent node for 'Trap Anticline' that can be on/off. This node has two children: 'TrapTopEast' and 'TrapBottomEast', which are also on/off, and they have conditional probability distributions depending on the outcome of the parent node. Let V be the set of all nodes, x_v the variable at node $v \in V$, and $\mathbf{x}$ the vector of all node variables, the joint probability model is defined by

$$p(\mathbf{x}|\theta) = \prod_{v \in V} p(x_v | x_{\mathrm{pa}(v)}, \theta^v).$$

Here, $\mathrm{pa}(v)$ denotes the parents of node v. Further, θ denotes the set of model parameters required for the conditional probabilities tables (CPT), where θ^v is the local parameter for node x_v. We show below how we can train or learn these parameter values from the multiple-scenario BPSM outputs.

The complete set of 24 scenarios, and associated observations, are shown in Figure 3 (generation) and in Figure 4 (accumulation). We have used a standard *k-means* algorithm with $k = 2$ (accumulation) or $k = 3$ (generation) for assessing the threshold for categorizing the data. Note that the data in this way become proxy for the knowledge of geological elements, that could potentially be observed at segment level. From the categorical data we learn the BN branches for trap, reservoir, and source separately. The resulting BN model (Figure 5) is derived in an explicit way from the data, while the CPT of the last level (from Trap, Reservoir and Source to the accumulations) have been assigned with external subjective considerations.

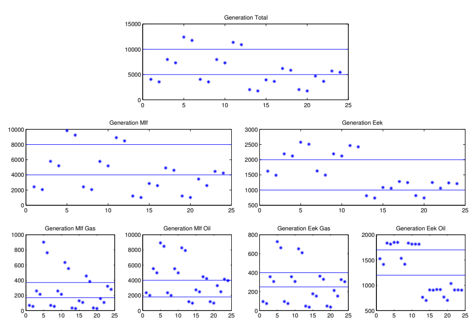

Fig. 3 Data for learning the source network. Top: values for the HC generation. Middle: values for Eek and Mlf generation. Bottom: values for oil and gas generation in each of the Eek and Mlf source rock. Values in MMBOE.

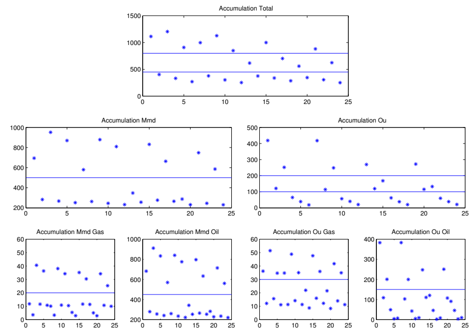

Fig. 4 Data for learning reservoir network. Top: values for the HC accumulation. Middle: values for Mmd and Ou generation. Bottom: values for oil and gas generation in each of the Mmd and Ou source rock. Values in MMBOE.

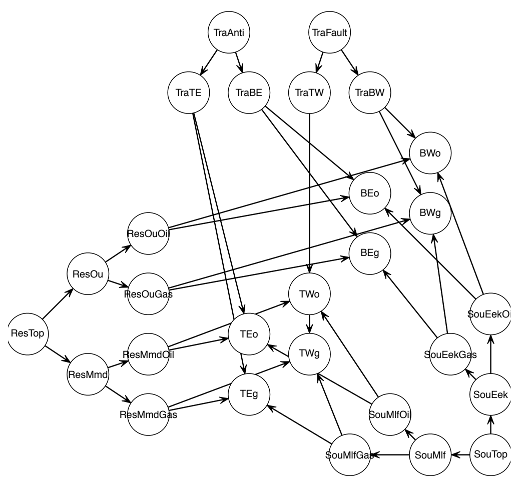

Fig. 5 BN with trap(top), reservoir(left) and source(lower right) branches. Top nodes are discrete, bottom nodes are Gaussian.

While it seems reasonable to have discrete nodes in the top parts of the network, since attributes such as source, reservoir and trap are on/off or multi-level features, it may be more realistic to have continuous bottom nodes that mimic the actual behavior of the simulated scenarios. We therefore split each of the bottom nodes TE, BE, TW and BW in two nodes, one for gas volume and the other for oil volume, and state that they represent accumulation distributions whose mean and (possibly) variance depend on the states of their parents. The simultaneous use of discrete and continuous variables in BN has been explored in [3] and [8]; an inference algorithm is presented in [11]. The related CPTs have to be assessed, for example the conditional probability density of BEg is:

$$p_{BEg}(x|Tra_{BE}, Res_{OuGas}, Sou_{EekGas}) \sim N(\mu_{BEg}, \sigma^2_{BEg}),$$

where μ_{BEg} is the conditional mean value and σ_{BEg} is the conditional standard deviation of this Gaussian distribution.

The effects of this parametrization on the HC distributions together with the estimated recoverable resources (recovery factor 0.45 for oil and 0.75 for gas) can be seen in Figure 6.

The resulting distributions are multimodal, and the different modes reflect the likelihood of being in each of the 24 configurations taken into account.

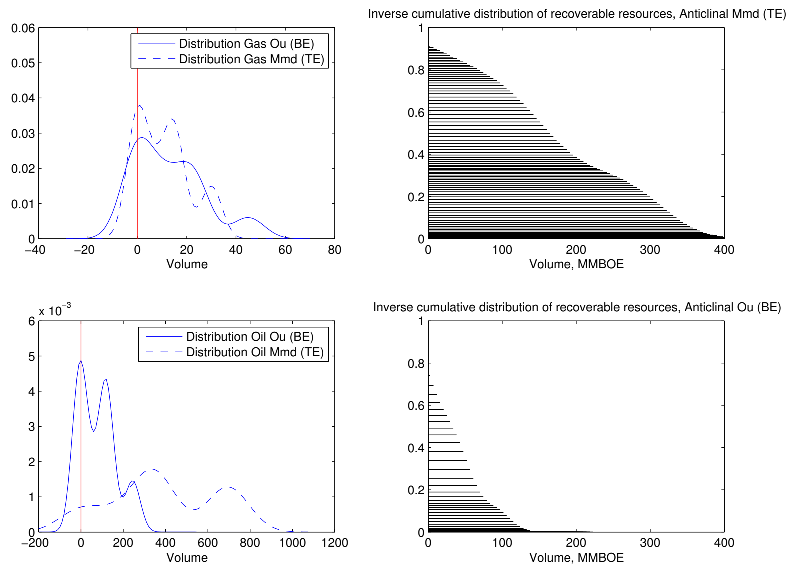

Fig. 6 Oil and gas volume distributions for segments BE and TE and expected associated total recoverable resources for TE (top right) and BE (bottom right). Volume distributions are actually truncated in 0 (red line), resulting in mixture discrete/continuous distributions.

5 Applications and conclusions

We next illustrate how BNs are useful for decision making purposes.

The distributions are immediately updated as soon as more information are available. Let us focus our attention on the gas accumulation relative to prospect TE. In this case we may receive an information that confirms our likelihood about the presence/absence of the reservoir or the trap in that prospect. The network is immediately updated: in this case the effect of a positive reservoir layer is much stronger than that of a positive trap, since the prior likelihood for the anticlinal trap is already equal to 0.9, while the uncertainty about the goodness of the reservoir layer (porosity) is much higher.

Similar measurements can lead us to the evaluation of so-called *what-if* scenarios: we are interested in the behavior of the network in case of evidence coming directly from the observation of a HC column in another prospect. We use as example the prospect TE, given possible observations from prospects BE (Figure 7) and TW (Figure 8). In the first figure we see that even a rich observation in BE is not sufficient to solve the multimodality of the original distribution, since the possible uncertainty about the quality of the reservoir remains (TE and BE belong to 2 different reservoirs). In the second figure we see that both an extremely poor and a rich observation in the fault prospect TW substantially change the shape of the posterior oil TE distribution. As we have already pointed out, a positive HC column observation in a difficult prospect such as TW confirms the quality of the reservoir and the charge, and this has a high impact on TE.

A full evaluation of such dynamics leads us to the computation of the Value of collecting perfect Information [2] in the four different prospects.

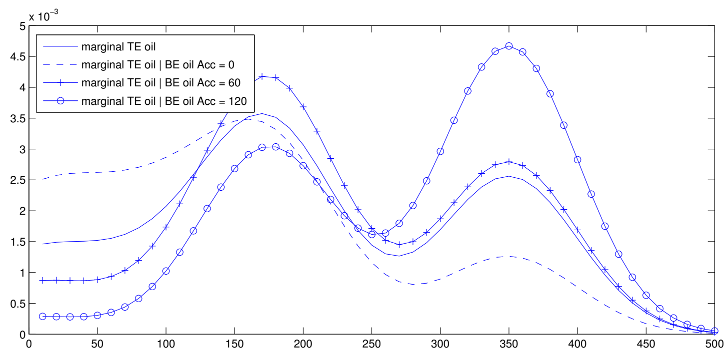

Fig. 7 Oil TE distribution before and after observation of oil column in BE

The work underlines the importance of assessing uncertainty in petroleum systems. The emphasis is less on knowing the right answer, that may never be known before drilling, but rather on determining the range of outcomes given the available data and state of understanding of the petroleum system. Problems are caused by the complex and often non linear interactions among the different parameters, that make the prediction problem extremely difficult. Currently these problems are solved running several simulations with different parameters, and studying the uncertainty in the resulting accumulation distribution as main or sole output. We believe that this process is not sufficient any longer, since there are too many parameters that remain hidden (implicit parameters) when the effect of many parameters is tested at the same time. With our framework we provide an alternative solution by making explicit all the correlated parameters, though not chosen arbitrarily, but derived from a multiple scenario evaluation. We have shown how BPSM allows partially auto-

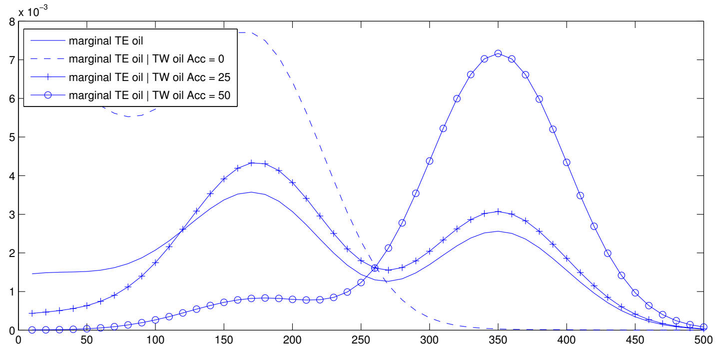

Fig. 8 Oil TE distribution before and after observation of oil column in TW

matic assessment of the BN probability structure, and our interest is in deepening this aspect by integrating even more the two frameworks.

References

1. P.A. Allen and J.R. Allen. *Basin Analysis, Principles and Applications. 2th ed.* Blackwell Publishings, 2005.
2. D. Bhattacharjya, J. Eidsvik, and T. Mukerji. The value of information in spatial decision making. *Mathematical Geosciences*, 42(2):141–163, 2010.
3. K.C. Chang and R.M. Fung. Symbolic probabilistic inference with both discrete and continuous variables. *IEEE Transactions on Systems, Man and Cybernetics*, 25(6):910–917, 1995.
4. B. Corre, P. Thore, V. deFeraudy, and G. Vincent. Integrated uncertainty assessment for project evaluation and risk analysis. *SPE European Petroleum Conference*, 2000.
5. Elvind Damsleth, Asmund Hage, and Rolf Volden. Maximum information at minimum cost: A north sea field development study with an experimental design. *Journal of Petroleum Technology*, 44(12):1350–1356, 1992.
6. J.-P. Dejean and G. Blanc. Managing uncertainties on production predictions using integrated statistical methods. *SPE Annual Technical Conference and Exhibition*, 1999.
7. R.A. Fisher. *The Design of Experiments, 9th Edition*. Macmillan, 1971.
8. N. Friedman and M. Goldszmidt. Discretizing continuous attributes while learning bn. *Machine Learning: Proceedings of the International Conference*, 1996.
9. Thomas Hantschel and Armin I. Kauerauf. *Fundamentals of Basin and Petroleum Systems Modeling*. Springer, 2009.
10. G. Martinelli, J. Eidsvik, R. Hauge, and M. Drange-Forland. Bayesian networks for prospect analysis in the north sea. *AAPG Bulletin*, 95(8):1423–1442, 2011.
11. Kevin P. Murphy. A variational approximation for bayesian networks with discrete and continuous latent variables. *Decision Analysis*, 1999.