



## A Bayesian belief network approach for assessing uncertainty in conceptual site models at contaminated sites

Thomsen, Nanna Isbak; Binning, Philip John; McKnight, Ursula S.; Tuxen, Nina; Bjerg, Poul Løgstrup; Trolborg, Mads

*Published in:*  
Journal of Contaminant Hydrology

*Link to article, DOI:*  
[10.1016/j.jconhyd.2016.02.003](https://doi.org/10.1016/j.jconhyd.2016.02.003)

*Publication date:*  
2016

*Document Version*  
Peer reviewed version

[Link back to DTU Orbit](#)

*Citation (APA):*  
Thomsen, N. I., Binning, P. J., McKnight, U. S., Tuxen, N., Bjerg, P. L., & Trolborg, M. (2016). A Bayesian belief network approach for assessing uncertainty in conceptual site models at contaminated sites. *Journal of Contaminant Hydrology*, 188, 12-28. <https://doi.org/10.1016/j.jconhyd.2016.02.003>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# A Bayesian Belief Network approach for assessing uncertainty in conceptual site models at contaminated sites

Nanna I. Thomsen<sup>1</sup>, Philip J. Binning<sup>\*</sup>, Ursula S. McKnight<sup>1</sup>, Nina Tuxen<sup>3</sup>, Poul L. Bjerg<sup>1</sup>, Mads  
Troldborg<sup>2</sup>

<sup>1</sup> Department of Environmental Engineering, Miljoevej, Building 113, 2800 Kgs. Lyngby, Denmark

<sup>2</sup>The James Hutton Institute, Craigiebuckler, Aberdeen, AB15 8QH, United Kingdom

<sup>3</sup>Capital Region of Denmark, Regionsgården - Kongens Vænge 2 - 3400 Hillerød, Denmark

\*Corresponding author: Philip J. Binning, Department of Environmental Engineering Environment,  
Miljoevej, building 113, 2800 Kgs. Lyngby, Denmark; phone (+45) 45 25 21 61 fax (+45)

45932850 e-mail: [pjbi@env.dtu.dk](mailto:pjbi@env.dtu.dk)

Submitted for publication in Journal of Contaminant Hydrology 2015

Revised January 2016

## **Abstract**

A key component in risk assessment of contaminated sites is the formulation of a conceptual site model (CSM). A CSM is a simplified representation of reality and forms the basis for the mathematical modelling of contaminant fate and transport at the site. The CSM should therefore identify the most important site-specific features and processes that may affect the contaminant transport behavior at the site. However, the development of a CSM will always be associated with uncertainties due to limited data and lack of understanding of the site conditions. CSM uncertainty is often found to be a major source of model error and it should therefore be accounted for when evaluating uncertainties in risk assessments. We present a Bayesian Belief Network (BBN) approach for constructing CSMs and assessing their uncertainty at contaminated sites. BBNs are graphical probabilistic models that are effective for integrating quantitative and qualitative information, and thus can strengthen decisions when empirical data are lacking. The proposed BBN approach facilitates a systematic construction of multiple CSMs, and then determines the belief in each CSM using a variety of data types and/or expert opinion at different knowledge levels. The developed BBNs combine data from desktop studies and initial site investigations with expert opinion to assess which of the CSMs are more likely to reflect the actual site conditions. The method is demonstrated on a Danish field site, contaminated with chlorinated ethenes. Four different CSMs are developed by combining two contaminant source zone interpretations (presence or absence of a separate phase contamination) and two geological interpretations (fractured or unfractured clay till). The beliefs in each of the CSMs are assessed sequentially based on data from three investigation stages (a screening investigation, a more detailed investigation, and an expert consultation) to demonstrate that the belief can be updated as more information becomes available.

Keywords: Bayesian belief networks, contaminated sites, conceptual models, conceptual uncertainty, screening level.

# 1 Introduction

The conceptual site model (CSM) is essential to many aspects of contaminated site management including: risk assessment [Suter, 1999; Trolborg, 2010]; as the basis for groundwater models [Neuman and Wierenga, 2003]; in the modelling of contaminant transport [McMahon et al., 1999]; for sampling design [US EPA, 1996]; and/or for identifying natural attenuation [Bjerg et al., 2011]. According to the US EPA [1996] a CSM can be defined as “a three-dimensional picture of site conditions that illustrates contaminant distributions, release mechanisms, exposure pathways and migration routes, and potential receptors. The CSM documents current site conditions and is supported by maps, cross sections, and site diagrams that illustrate human and environmental exposure through contaminant release and migration to potential receptors”. A key aspect of developing a CSM is that it is an iterative process that should evolve in complexity as more data are collected [ASTM Standard E1689, 2008]. The CSM can be seen as a hypothesis for how a site operates and can be continuously tested as new data are collected. In standard statistics any hypothesis needs at least one alternative (i.e. H0 and H1), where in Bayesian hypothesis testing you must have at least one, but can have many alternatives.

The development of CSMs is challenging and will always be associated with uncertainty due to lack of data and understanding of the site conditions, but also due to the simplifications introduced to describe complex phenomena such as heterogeneous geology, hydrogeology, contaminant source distribution and transformation processes.

**Abbreviations** BBN: Bayesian belief network, BMA: Bayesian model average, CPT: Conditional probability table, CSM: Conceptual site model, DCE: Dichloroethylen, DNAPL: Dense non aqueous phase liquid, FD: Fractured dissolved, FLUTE: Flexible liner Underground technologies, FN: Fractured D(N)APL, LIF: Laser induced fluorescence, MIP: membrane interface probe, PCE: Perchloroethylene, PID: photoionization detector, TCE: Trichloroethylene, UD: Unfractured dissolved, UN: Unfractured D(N)APL, VC: Vinyl chloride.

To help overcome some of the challenges, many excellent guidelines for setting up CSMs have been constructed [ASTM Standard E1689, 2008; ASTM Standard E2531, 2009; McMahon et al., 1999; Neuman and Wierenga, 2003; Suter, 1999; US EPA, 1996; US EPA, 2002]. Regardless of the choice of guidelines and purpose of the study, modelers will eventually have to choose which features and processes (geological, chemical, hydraulic, etc.) to include, and how to represent and simplify these. In many cases, the CSM will therefore be based on the modelers' subjective belief and perception of how a specific site "operates", where the modeler relies not only on the available data, but also on past experiences from similar sites. Uncertainty is therefore an inherent part of creating a CSM. The uncertainty concerning the CSM addressed within this paper will be referred to as conceptual uncertainty.

Conceptual uncertainty is a well described phenomenon [Beven, 2009; Konikow and Bredehoeft, 1992; Refsgaard et al., 2006; Walker et al., 2003], which is often found to be a major source of uncertainty and must therefore be considered [Bredehoeft, 2005; Harrar et al., 2007; Højberg and Refsgaard, 2005; Troldborg et al., 2007]. The challenge of quantifying conceptual uncertainty has been discussed in the literature and many methods have been proposed. One of the most frequently applied methods is to use multiple CSMs to represent the uncertain settings at the site [e.g. Foglia et al., 2007; Georgakakos et al., 2004; James and Oldenburg, 1997; Li and Tsai, 2009; Neuman, 2003; Poeter and Anderson, 2005; Rojas et al., 2008; Tebaldi et al., 2005; Troldborg et al., 2010; Ye et al., 2005]. Most studies using the multi-model approach are also concerned with investigating how well the different models represent the system behavior. It is, for example, common to use Bayesian model averaging (BMA) [Hoeting et al., 1999] to aggregate the output from competing models [Li and Tsai, 2009; Neuman, 2003; Rojas et al., 2008; Troldborg et al., 2010; Ye et al., 2010]. In BMA the predictions from alternative (conceptual) models are combined using weights that reflect each model's relative ability to reproduce the system behavior. Usually, these weights

are determined by evaluating how well the different models match the available data of the predictive variable(s) (e.g. hydraulic head and/or concentration measurements) using a predefined likelihood function and Bayes' theorem [e.g. *Ye et al.*, 2010]. If such data are not available, the weights/beliefs must be assigned subjectively. Even if data are available, it is still necessary to assign (subjective) prior beliefs to the different models in order to apply Bayes' rule for determining the posterior model probabilities (although the Maximum Likelihood BMA approach proposed by *Neuman* [2003] can be applied without prior beliefs). The subjective specification of (prior) model beliefs is often based on a no preference assumption where all models are assigned equal probabilities [e.g. *Rojas et al.*, 2008; *Troldborg et al.*, 2010] but is, as noted by *Ye et al.* [2008] and *Singh et al.* [2008], ideally based on expert elicitation.

Here we explore how Bayesian Belief Networks (BBN) (also known as Belief Networks, Causal Probabilistic Networks or Knowledge Maps) can be used to facilitate the construction of multiple CSMs and determine the belief in each of them from sparse data and expert opinion. BBNs are graphical probabilistic models that represent system variables and their conditional relationships as nodes and linkages in an influence diagram. The relationships between variables are defined by conditional probability distributions, and BBNs can therefore account for uncertainty in model predictions explicitly [*Korb and Nicholson*, 2003]. The graphical representation helps to visualize and structure the relevant system components. BBNs have proven effective for aggregating data (quantitative information) and expert opinions (qualitative information), and they thus have the ability to strengthen decisions when empirical data are lacking. BBNs provide both diagnostic and predictive capabilities and allow for updating the probability distributions with new evidence when such become available.

BBNs have previously been used in the assessment of contaminated sites and groundwater quality. Examples include e.g.: (1) evaluating reductive dechlorination at TCE (trichloroethylene)

contaminated sites [*Stiber et al.*, 1999; *Stiber et al.*, 2004], (2) public participation and stakeholder engagement in integrated management of groundwater contamination [*Farmani et al.*, 2009; *Farmani et al.*, 2012; *Henriksen et al.*, 2007b; *Henriksen et al.*, 2007a; *Henriksen and Barlebo*, 2008], (3) forecasting groundwater pollution levels [*Shihab*, 2008; *Shihab and Chalabi*, 2007], (4) assessing and mapping groundwater quality [*Aguilera et al.*, 2013], and (5) detecting contaminant leakage from landfills [*Small*, 1997]. BBNs have also been applied to a wide range of other problems within the field of hydrology and water management. For example, *Chan et al.* [2010] used a BBN for assisting catchment-based water resources management, *Wang et al.* [2009] developed a BBN for assessing and managing farm irrigation systems, while *Fienen et al.* [2013] used a BBN with a numerical groundwater model to study the response of groundwater to sea level rise.

The overall aim of this paper is to examine the potential of using a BBN methodology to firstly facilitate the systematic construction of multiple CSMs, and secondly for assessing the uncertainty and assigning weights (beliefs) to each of the created CSMs. To do this, we demonstrate the proposed BBN methodology on a study site where a spill of PCE (Tetrachloroethylene) and TCE occurred in the 1970s. At this site, two specific conceptual issues, both key to risk assessment, are considered. The first is the presence of a DNAPL (Dense Non Aqueous Phase Liquid) [*ITRC*, 2013] and consequently the long term persistence of a secondary source [*Parker et al.*, 2008]. The second element concerns the presence or absence of fractures in the clay till. The long term persistence of a DNAPL source is exacerbated in low permeability geological features [*Hadley and Newell*, 2012] or dual porosity geological media [*Chambon et al.*, 2011]. Clay tills are common in northern Europe and parts of North America [*Christiansen et al.*, 2008; *Parker et al.*, 2008] and can lead to a complex system of preferential flow paths (e.g. fractures or sand lenses) in a low permeability matrix [*Damgaard et al.*, 2013a; *Damgaard et al.*, 2013b; *Gerber et al.*, 2001; *Hendry et al.*, 2004;

*Kessler et al., 2012; Kessler et al., 2013*]. Capturing a correct conceptual understanding of such sites is crucial in risk assessment and management where predictions of contaminant breakthrough, leaching time and remedial options are required. Due to the uncertainty in the conceptual understanding of the above issues at the study site, four different CSMs were formulated and implemented into a BBN. Using the BBN, the beliefs in each of the CSMs are assessed. At the study site, a screening investigation [*ASTM Standard E1739 - 95(2010)e1, 2010; Danish EPA, 2002; UK Environment Agency, 2004*] is followed by a more detailed investigation and an expert consultation, and this information is used with the BBN in three stages to demonstrate how the beliefs can be updated as more information becomes available.

## **2 Bayesian belief networks**

BBNs are graphical probabilistic models with the strength that causal relations are formulated as conditional probabilities. BBNs consist of a qualitative and a quantitative component. The qualitative component is a directed acyclic graph in which nodes and directed links represent system variables and their conditional relationships. If there is a directed link from node A to node B, node A is called the parent and B is the child (Figure 1). The quantitative component is a set of probability distributions that quantifies the strength of the conditional relations between variables. BBNs usually operate with discrete probabilities. Each node or variable in a Bayesian network is therefore assigned a finite set of state values, where each state is associated with a probability. For any child node (B) with parents ( $A_1, \dots, A_n$ ), there exists a conditional probability table (CPT)  $P(B|A_1, \dots, A_n)$ , as illustrated in Figure 1 [*Nielsen and Jensen, 2007*].

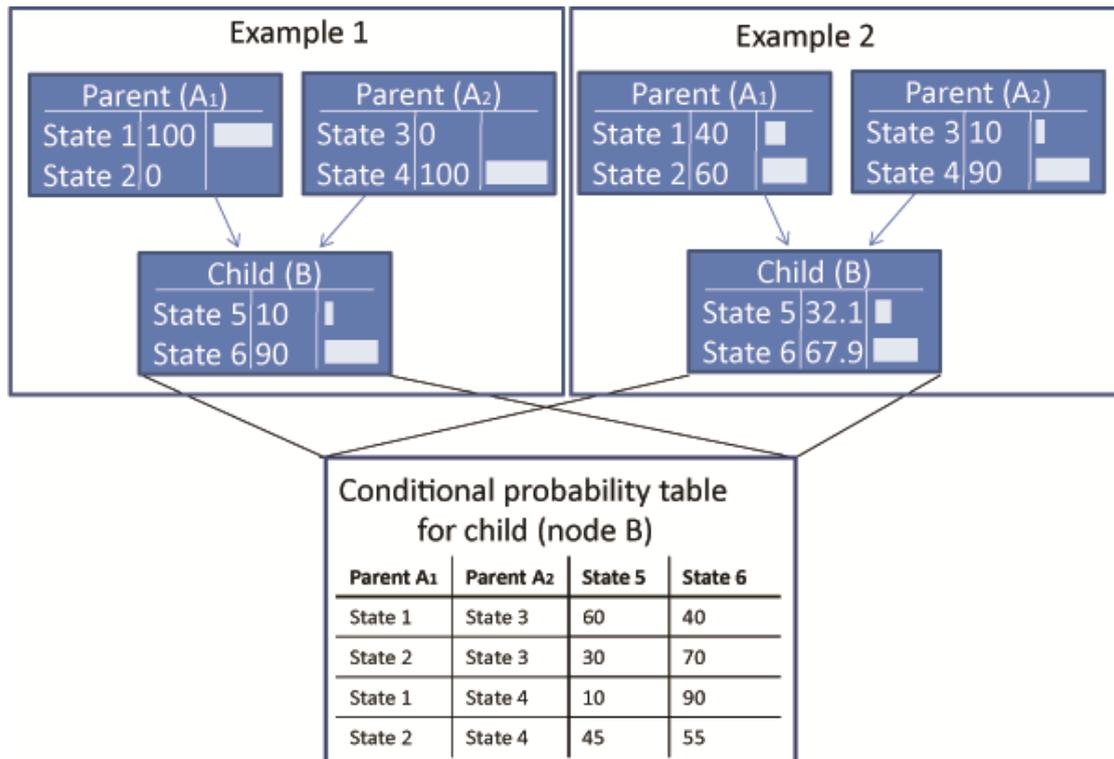
A major advantage of BBNs is that they allow inference based on observations. Probabilistic inference, also called belief updating, is simply the task of computing the posterior probability distribution of the BBN given the evidence, i.e. given the observed value of some variable(s) in the

network. The information flow in a BBN is both in the causative direction (from A to B, i.e. reasoning from new observations of causes, to new beliefs about the effects) and the diagnostic direction (from B to A, i.e. reasoning from observed effects to updated beliefs about causes). This property is useful because when we have evidence about the state of a given node, this evidence can be used to make inference about the states of connected nodes. The state of a node is determined either by providing evidence, or by making an inference about the node based on any evidence provided for the connected nodes [Nielsen and Jensen, 2007].

To illustrate how the BBN method works, two example calculations are presented, see Figure 1. For both examples, the overall network construction is the same, i.e. the CPT for the child node (B) is the same (see Figure 1), but different states are specified for the parent nodes. In example 1, node  $A_1$  and node  $A_2$  are known (i.e. 100 % probability) to be in state 1 and state 4, respectively. Propagating this through to the child node (B), the resulting probability of B being in state 5 and state 6 becomes 10 % and 90 %, respectively, as shown directly in the CPT. Example 2 shows the same calculation when the states of the parent nodes are uncertain, in which case the resulting probability of B is found by marginalizing over nodes  $A_1$  and  $A_2$ . Thus, the probability of B being in state 5 is then:

$$\begin{aligned}
 P(B = state5) &= P(B = state5|A_1 = state1, A_2 = state3)P(A_1 = state1)P(A_2 = state3) \\
 &\quad + P(B = state5|A_1 = state2, A_2 = state3)P(A_1 = state2)P(A_2 = state3) \\
 &\quad + P(B = state5|A_1 = state1, A_2 = state4)P(A_1 = state1)P(A_2 = state4) \\
 &\quad + P(B = state5|A_1 = state2, A_2 = state4)P(A_1 = state2)P(A_2 = state4) \\
 &= 0.6 * 0.4 * 0.1 + 0.3 * 0.6 * 0.1 + 0.1 * 0.4 * 0.9 + 0.45 * 0.6 * 0.9 = 0.321
 \end{aligned}$$

In the present work, the BBNs are built using the Netica<sup>TM</sup> software [Norsys Software Corp., 2014].



**Figure 1** Two example calculations using a Bayesian belief network. The probabilities are in %.

### 3 Study site

A study site located on Zealand in Denmark is used to demonstrate how the BBN methodology can be applied to address conceptual model uncertainties. During the 1970s the site was used to store chemicals and a spill occurred at the terrain surface and migrated into the underlying soil. The spill consisted of chlorinated solvents, mainly PCE and TCE. The site is located in an area where the geology is dominated by clay till settings.

#### 3.1 Available data

We have deliberately not chosen a research study site, because we want to demonstrate that our method can be applied in practice. The study site is managed by the regional authority (The Capital Region of Denmark) and the data is typical for a site in that context. This means that the method

must be robust enough to deal with cases where there is relatively poor data such as in a screening investigation. Also, since the aim of this paper is the presentation of the BBN method for constructing conceptual models, we only present an outline of some of the data available and not the details of sampling methods and chemical analysis. We first present the data collected during the screening investigation, and then proceed to describe additional data obtained in a detailed investigation and from an expert consultation. We do this because the available information will later be used sequentially to demonstrate how the beliefs in the different CSMs can be updated when new data becomes available.

### **3.1.1 Data from the screening investigation**

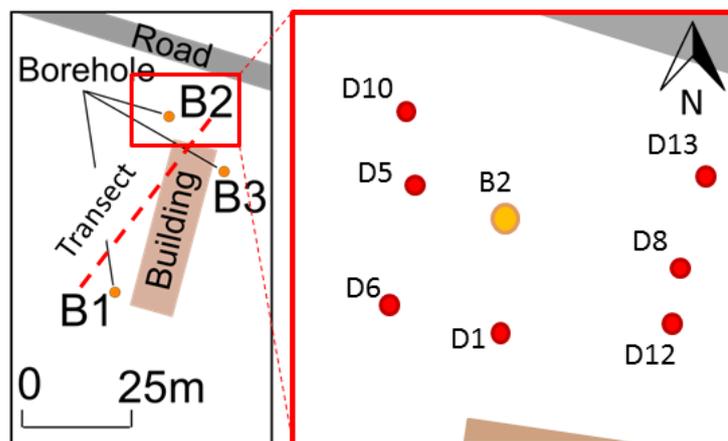
The screening investigation included a desktop investigation and a field campaign. The details of the inventory are specified in Table 1 and the results of the sampling and chemical analysis are shown in

Table 2.

**Table 1** Screening investigation inventory [*Hedeselskabet, 2005*]

<b>Screening Investigation: Desktop data</b>			
Item		Information/Reference	Comment
Consultant's report and personal communication		Spill history [ <i>Hedeselskabet, 2005</i> ]	The consultant's report documents the screening investigation
Geological maps and maps of the groundwater potential		Geological stratum [ <i>Hovedstadsrådet, 1980</i> ] Groundwater potential [ <i>Rambøll, 1999</i> ]	The hydrogeological maps are regional, i.e. not site specific
<b>Screening Investigation: Field data</b>			
Item	Number	Information	Comment
Shallow boreholes (B1, B2, and B3)	3	Geological profiles (cross sections) [ <i>Hedeselskabet, 2005</i> ]	Shallow: Boreholes that do not reach the sandy layer. Locations see Figure 2.
		Redox boundary [ <i>Hedeselskabet, 2005</i> ]	Indicated on the geological profiles Figure 2.
Soil samples	20	PID <sup>1</sup> measurements [ <i>Hedeselskabet, 2005</i> ]	PID in a soil sample every 0.5 m. Locations: see Figure 2.
Water samples	3	Aqueous concentration [ <i>Hedeselskabet, 2005</i> ]	One sample per borehole

<sup>1</sup> PID: Photoionization detector for measuring volatile organic compounds, such as chlorinated hydrocarbons or aromatics.



**Figure 2** Left: Map of the screening investigation area. The transect indicates the location of the geological cross section (presented in Figure 4). Right: Zoom on the MIP (membrane interface probe) soundings that were conducted as part of the detailed investigations described in Section 3.1.2 (note that not all MIP soundings are shown here, e.g. sounding D4 was located between D5 and D6 west of the frame) [*Orbicon A/S, 2007*].

**Table 2** Results from the chemical analysis of soil and water samples from the two field campaigns. The numbers in bold are used in Section 5.1 to update the BBNs [Hedeselskabet, 2005; Orbicon A/S 2007].

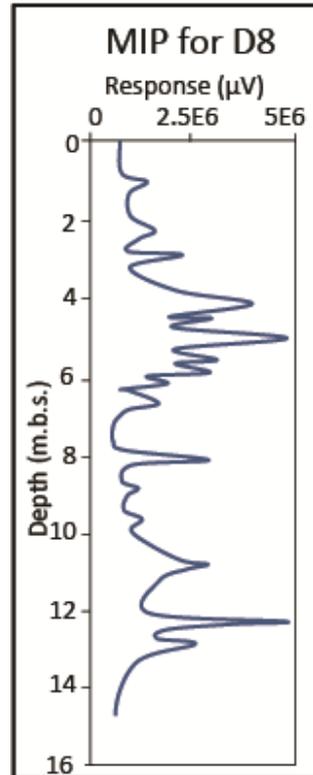
Stage	Sample ID	Sample depth (m.b.s. <sup>5</sup> )	Sample type	PCE <sup>1</sup>	TCE <sup>2</sup>	DCE <sup>3</sup>	VC <sup>4</sup>
Screening investigation	B1	12-14	Aqueous concentration (µg/l)	0.51	14	0.3	<0.2
	B2	11-13	Aqueous concentration (µg/l)	<0.05	1200	30000	31
	B3	8-14	Aqueous concentration (µg/l)	1.7	7.7	8.9	4.4
Detailed investigation	D4	8.3	Soil concentration (mg/kg)	<0.005	<b>56</b>	-	-
	D6	2.3	Soil concentration (mg/kg)	0.98	2.6	-	-
		8	Soil concentration (mg/kg)	<0.005	0.14	-	-
		13.3	Soil concentration (mg/kg)	<0.005	7	-	-
	D12	2	Soil concentration (mg/kg)	12	0.2	-	-
		5.7	Soil concentration (mg/kg)	<b>89</b>	4.8	-	-
12		Soil concentration (mg/kg)	2.4	0.09	-	-	

<sup>1</sup>PCE: Tetrachloroethylene <sup>2</sup>TCE: Trichloroethylene, <sup>3</sup>cDCE: cis-dichloroethylene, <sup>4</sup>VC; Vinyl chloride.  
<sup>5</sup>m.b.s.: meters below surface.

### 3.1.2 Data from the detailed site investigation

The initial site investigation suggested that the site could pose a risk to groundwater resources, so to improve the risk assessment, additional data were collected in a detailed site investigation [Orbicon A/S, 2007].

During this investigation, membrane interface probe (MIP) soundings were conducted and extra soil samples were collected. The MIP measures volatile organic carbon components and provides its response in µV. The results are semi-quantitative, meaning that the method does not provide absolute proof of the presence of DNAPL, but if MIP data is compared with concentration measurements it can give a strong indication of the presence or absence of DNAPL. MIP soundings have a high resolution and the method is relatively fast and cheap, and so it is commonly used to delineate DNAPL sources. An example of a response graph from a MIP sounding is presented in Figure 3. The location of the soundings can be seen in Figure 2, and the results of the sampling are shown in Table 2. The soil samples were collected at the depths where peaks in the MIP soundings were observed.



**Figure 3** An example of a MIP sounding, reproduced from D8, see location on Figure 2 (m.b.s.: meters below surface).

### 3.1.3 Data from expert consultation

BBNs can also include expert opinion (on processes, etc.) as a supplement to available data. The use of this type of expert knowledge differs from that used in the CPTs by the fact that it is case-specific, i.e. an expert is consulted on the matter of the case, and not on an indicator of a more general process. Here, an expert on the geology of the area was consulted to help characterize the type of clay till. Based on an exchange of the data from Section 3.1.1 and 3.1.2, and personal communication, the expert found it most likely that the clay till at the site was a basal till type B, but suggested the till could also be a glacitectorite. This was reflected in the belief probabilities assigned by the expert to basal till type B and glacitectorite of 80% and 20%, respectively (see Section 4.3 for more information). The expert assessment was uncertain about the structure and fabric of the till.

## **3.2 Formulation of conceptual site models**

Considerable uncertainty was found to exist regarding both the conceptual understanding of the source zone and the geological settings. It was decided to reflect this uncertainty by formulating multiple CSMs, all of which were considered to be plausible representations of the site settings. Four simple CSMs have been developed for the site by combining two conceptual representations of the source zone with two representations of the clay till, as illustrated in Figure 6. The formulation of these is described in the following.

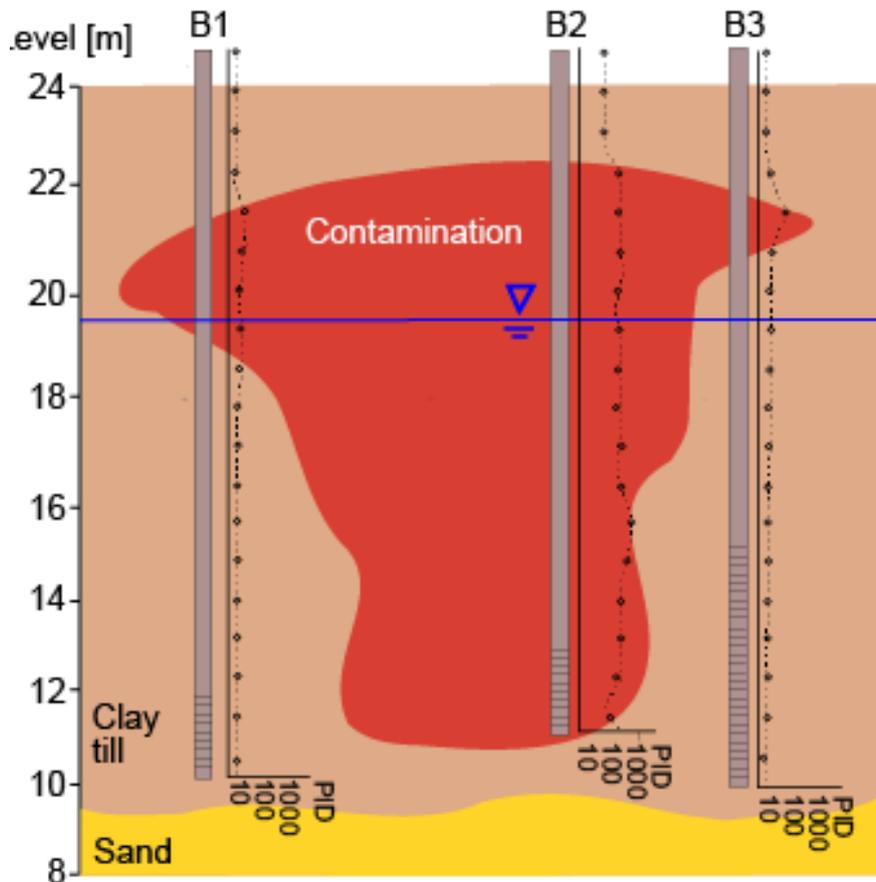
### **3.2.1 Conceptual representations of the source zone**

It is believed that the contaminant spill occurred in the 1970s at the ground surface due to leakage of above ground storage tanks and consisted of PCE and TCE. The results from the aqueous samples (Table 2) show that borehole B2 is contaminated and the available PID measurements suggest that the vertical extent of the contaminant spreading extends over the full depth of the clay till. The chlorinated solvents have thus migrated from the surface into the clay till and are a secondary source threatening the underlying aquifer (Figure 4), similar to the conditions described in *Chambon et al.* [2011].

For this particular study site, an important consideration for the conceptualization of the source zone is whether a DNAPL phase could be present, because the presence of DNAPL means that the source strength and life time are significantly increased compared to a dissolved source. Based on the chemical analyses (Table 2) it is not possible to say whether the source consists of DNAPL or dissolved phase contaminant. It is likely that the spill initially migrated into the till as DNAPL, but this too is uncertain. To reflect this uncertainty, two conceptual source zone models have been formulated: one which assumes that DNAPL is present in the source zone (N) and one which assumes that the contamination only exists in a dissolved state in the source zone (D).

### 3.2.2 Conceptual representations of the clay till

The maps of the hydrogeology suggest that the site is located in an area where 15 m of clay till is underlain by a thin sandy aquifer (approx. 0.5 m). The hydrogeology of clay tills is often difficult to characterize due to the potential presence of fractures and/or sand lenses, which can create a complex network of preferential flow paths through the clay till [Gerber *et al.*, 2001; Hendry *et al.*, 2004; Kessler *et al.*, 2013]. Migration of DNAPL into clay tills is only possible if the till is fractured [O'Hara *et al.*, 2000]. The clay till can therefore act either as a barrier and prevent migration of the DNAPL into the sandy aquifer, or if the clay till is fractured and these fractures extend through the entire thickness of the clay, DNAPL could migrate through the till and into the sandy aquifer [O'Hara *et al.*, 2000; Slough *et al.*, 1999]. The second element of conceptual uncertainty included here is therefore to determine the presence of fully penetrating fractures in the clay till defined through two conceptual geological models: one fractured (F) and one unfractured (U).

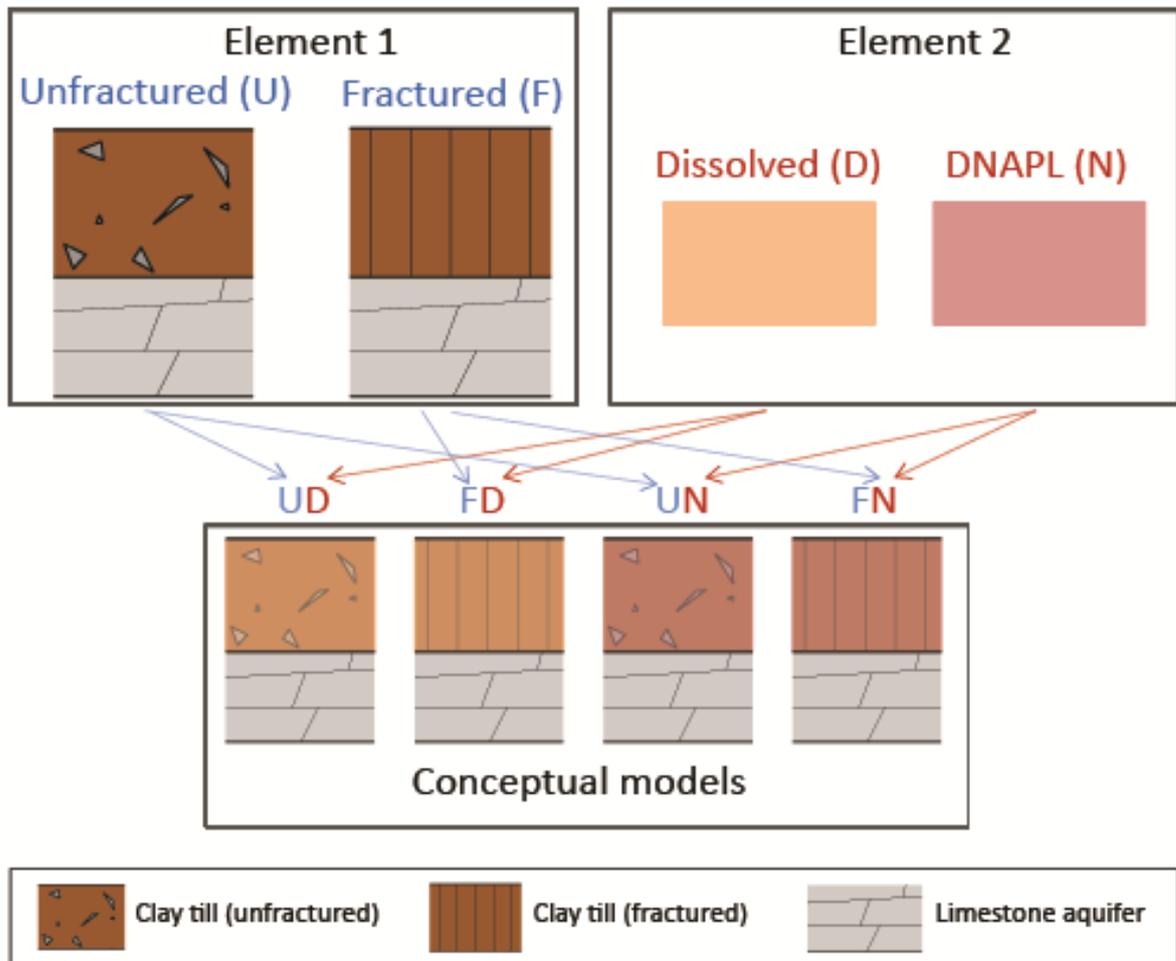


**Figure 4** A conceptual site model of the contamination. The PID measurements are shown for the 3 borehole profiles (ppm).

### 3.2.3 Combining the two sources of conceptual uncertainty

Both of the uncertain conceptual elements can be in one of two states, and by combining these, four different CSMs can be created as illustrated in Figure 5. The four different CSMs then become: (1) Unfractured and Dissolved (UD); (2) Fractured and Dissolved (FD); (3) Unfractured and DNAPL (UN); and (4) Fractured and DNAPL (FN). This is a modular approach where the uncertainty is broken into manageable quantities much like in a fault tree analysis; see e.g. *de Barros et al.* [2011] and the work by *Freeze et al.* [1992]. A similar combinatorial design approach for developing candidate models has previously been presented (see e.g. *Li and Tsai* [2009], *Rojas et al.* [2008], *Sohn et al.* [2000], *Trolldborg et al.* [2010]) and also forms the basis in hierarchical Bayesian model averaging (see e.g. *Chitsazan and Tsai* [2015], *Elshall and Tsai* [2014], *Li and Tsai* [2009], *Rojas*

*et al.* [2008], *Sohn et al.* [2000], *Troldborg et al.* [2010]). However, unlike these studies, our BBN approach requires neither numerical simulations nor the specification of formal likelihood functions, only the formulation of CPTs.



**Figure 5** The two elements of conceptual uncertainty and their four combinations. F fractured, U unfractured, DNAPL and D dissolved.

#### 4 Development of the Bayesian belief networks

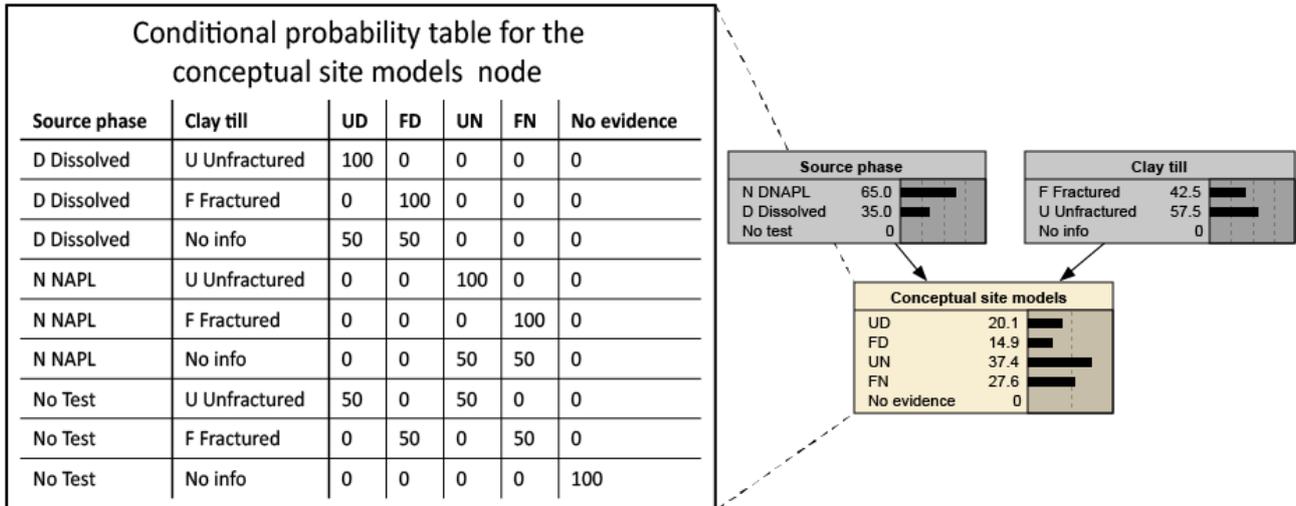
In this section we develop a BBN consisting of three smaller networks; one for each of the two elements of conceptual uncertainty and one that combines their output. We present the smaller networks individually. The CPTs for each of the child nodes can be found in the Supporting Information.

## 4.1 The combining network

Figure 6 shows the core of the developed BBN, which is where the weights (beliefs) from the different elements of conceptual uncertainty are propagated and combined to assess the belief in each of the formulated CSMs.

The central node is the “Conceptual site models” node, which in this specific case determines the weights of the four CSMs described in Section 3.2.3. The parent nodes are the two elements of conceptual uncertainty: the source phase and the fracturing of the clay till. The CPT for the “Conceptual site models” node is also provided in Figure 6 and shows the probability of the different CSMs given each possible combination of the states of the parent nodes. Other elements of conceptual uncertainty could be accounted for by adding additional parent nodes to the central “Conceptual site models” node. For example, the occurrence of reductive dechlorination could be added as another parent node (e.g. with states ‘Occurring’ and ‘Not occurring’, as suggested in *Stiber et al.* [1999], in which case the central node would then consider a total of eight CSMs and the number of rows in the resulting CPT would be twice as many.

In this paper, the specific assessments of both the source phase and the clay till are based on a number of variables and indicators, the details of which are presented in the following.



**Figure 6** Left: The conditional probability table (CPT) for the conceptual site models node. Right: The combining network.

## 4.2 The source phase network

The network for assessing the presence or absence of DNAPL in the source is shown in Figure 7 and consists of 15 nodes and 14 links. The functionality of each node is described in Table 3. The network and the CPTs are mainly developed and populated on the basis of the information contained in *Jørgensen et al. [2010]*, *Janniche et al. [2013]* and *ITRC [2013]*. *Jørgensen et al. [2010]* investigated the international literature on DNAPL sources and adapted it to Danish conditions. They also provide conceptual models and guidelines for estimating the presence and amounts of DNAPL. *Janniche et al. [2013]* extended this work and described field investigation methods and their documentation. Their work also contains information from field studies conducted in Danish clay tills. The *ITRC [2013]* discusses general scientific approaches to characterize sites contaminated with DNAPLs and can be used as a guideline if location-specific information is not available.

The “Source phase” node is the central node in this network and has three states: ‘DNAPL (N)’, ‘Dissolved (D)’ and ‘No test’. The parent nodes are indicators of the presence or absence of a separate phase, and their inputs determine the state of the “Source phase” node. The inputs are user-

specified and based on a site investigation. The indicators included are: DNAPL observation, concentration (aqueous, pore air, soil (i.e. PCE and TCE)), MIP and three different tests for the occurrence of DNAPL (Sudan IV, FLUTE<sup>TM</sup>, LIF/dye LIF; see Table 3 for an explanation of the abbreviations). Table 3 summarizes how these indicators are used to inform whether DNAPL is present or not. A detailed description of all the indicators and their CPTs are presented in the supporting information. The CPT for the “Source phase” node has generally been specified in such a way that the presence of DNAPL is considered increasingly likely as more of the indicators nodes are suggesting that NAPL is present. However, the strength of the different indicator nodes differs, which has been reflected in the CPT. For example, if presence of DNAPL is suggested by either the “NAPL observation” node (i.e. the state is 100% observed) or any of the Tests nodes (i.e. the state is 100% positive), then it is considered absolutely certain that DNAPL is actually present in the source zone (i.e. the source phase node is 100% in state NAPL) regardless of what the outcomes of the other indicators are. The presence of a separate phase can also be indicated through the measured concentration levels in samples of soil, pore water and/or pore air. To what extent the measured concentrations indicate presence of NAPL has been assessed here by comparing the concentrations to published guideline values and rules-of-thumb [Jørgensen *et al.*, 2010; Janniche *et al.*, 2013; ITRC, 2013; US EPA, 1994]. Water samples are compared to the effective aqueous solubility, which has been calculated based on Raoult’s law (*e.g.*, US EPA [1992]) and by considering the effect of the presence of all the compounds in Table 2. For water samples, a commonly employed rule-of-thumb states that if concentrations of a (D)NAPL-related compound are greater than 1% of the effective solubility, then (D)NAPL phase is probably present at the site [Bedient *et al.*, 1999; ITRC, 2013; Pankow and Cherry, 1996]. Here, we assume that contaminant concentrations in water samples between 1-10 % of the compound’s effective solubility give some indication of the presence of DNAPL (belief of 65 %), while concentrations above 10% of the

effective solubility give stronger evidence of DNAPL being present (belief of 95 %). If the concentration is less than 1 % of the effective solubility, the presence of a separate phase contaminant is still possible.

For the soil samples node we have focused on PCE and TCE, since only these two compounds are covered in *Janniche et al* [2013] and *Jørgensen et al.* [2010]. According to these two studies, soil samples with concentrations between 36.8-368 mg PCE/kg and 121-1218 mg TCE/kg give some indication of the presence of DNAPL, while concentrations above these intervals strongly indicate presence of DNAPL. These guideline values for soil samples are lower than those presented in *ITRC* [2013], where it is suggested that DNAPL chemical concentrations greater than 10,000 mg/kg indicate the presence of DNAPL. We use here the values from *Janniche et al.* [2013] because they are based on experience from a location with a geology similar to our study site.

It is assumed that concentrations in pore air in the range of 100-1000 ppm. make occurrence of DNAPL likely, while concentrations above this interval are assumed to strongly indicate DNAPL, in accordance with the guideline values presented in *US EPA* [1994] and *Janniche et al.* [2013], and have thus been used here to populate the CPT for the pore air samples node (see Table 3). However, because a large proportion of the contamination may have disappeared from the unsaturated zone due to evaporation, high concentrations may still remain in the saturated zone. Low concentrations in pore air samples are therefore assumed not to exclude the presence of a separate phase in the saturated zone.

It should be noted that because this network has been designed for sites with limited data availability, the CPTs here have been specified to account for the fact that the field sampling could have missed an area of DNAPL. For example, soil samples are typically very location-specific and may thus easily miss hotspots in the source zone. Therefore, even if all the data suggest that there is

no DNAPL at the site (e.g. measured concentration levels are low), it is still considered possible that DNAPL could be present. The BBN for the assessment of the source phase could be extended to account for the number and the quality of the available data at the site to give an assessment of the strength of the field campaign.

It should also be noted that all indicator nodes have a state that negates the presence of the indicator, e.g. no test or no sample. These states ensure that if there is no information for the specific indicator, then the conditional probabilities of that indicator will not propagate through the network, effectively meaning that unobserved indicator nodes are removed from the network. This approach was developed as part of this paper and essentially corresponds to operating with an adaptive BBN structure that uses or neglects certain nodes, depending on whether there is evidence entered or not. An alternative and more traditional way of dealing with nodes where no information exists is to assign equal weights or some other prior probability distribution to the unobserved states and carry out marginalization. However, specifying suitable and justifiable prior distributions to unobserved nodes is challenging. Also, if this approach is taken, the probabilities of the indicators that are not investigated will influence the state of the “Source phase” node, which should only be affected by the state of the indicators for which we have evidence. For example, if information is only available for the samples nodes, then assigning equal weights to all the other indicators (the state of the observations, MIP and tests nodes) will pull the probability of the “Source phase” node towards an equal distribution, rather than relying on the information obtained from the samples. In addition, specifying equal weights implies that we have an expectation of equally likely outcomes, i.e. at investigations of contaminated sites there is a 50-50 chance of encountering a separate phase vs. not finding it, and this information is not available. By introducing the ‘no evidence’ state as a dummy value, we avoid having to specify prior distributions to unobserved nodes.

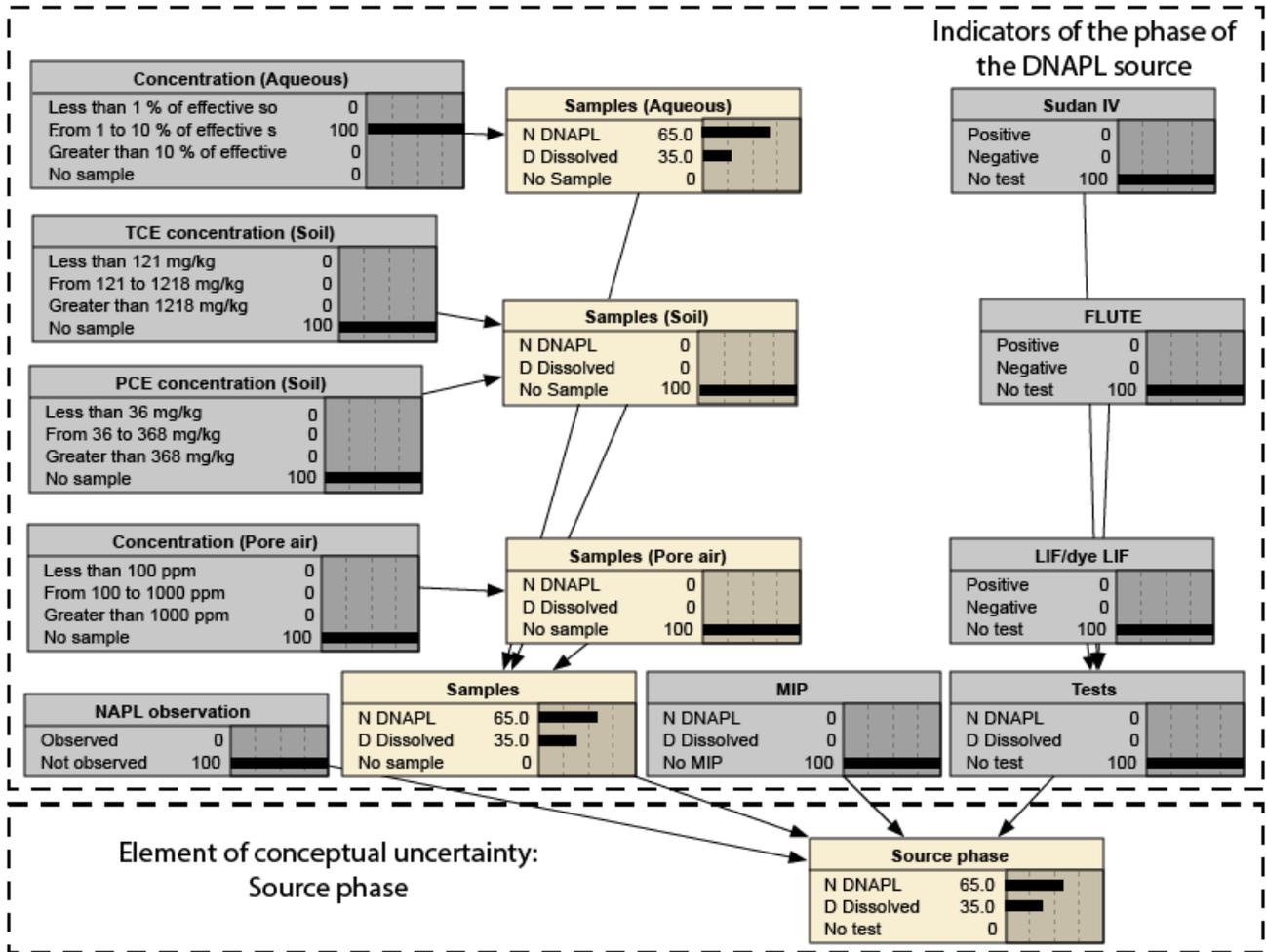


Figure 7. Bayesian belief network for assessing the phase of a TCE and PCE source.

**Table 3** Descriptions of the indicator nodes in the source phase network and how evidence from the screening level investigation is employed.

<b>Node</b>	<b>Description</b>	<b>Evidence</b> (From screening investigation)
DNAPL observation	If DNAPL is observed, the Source phase node will be in the state DNAPL regardless of the states of all other indicator nodes.	Not observed
Concentration (aqueous)	The aqueous concentration is compared to the effective solubility <sup>1</sup> of the individual compounds (PCE; TCE, DCE and VC). Concentrations less than 1 % of the solubility do not indicate presence of DNAPL, and concentrations greater than 10 % of the solubility give a strong indication that DNAPL is present [ <i>Janniche et al.</i> , 2013, <i>Jørgensen et al.</i> , 2010].	The TCE concentration in B2 (1.2 mg/l) was above 1 % (0.3 mg/l) of the effective solubility (31 mg/l).
PCE and TCE concentration (Soil)	For PCE a concentration of 368 mg/kg and for TCE a concentration of 1218 mg/kg strongly indicates the presence of DNAPL [ <i>Janniche et al.</i> , 2013]. Concentrations from 36.8-368 mg/kg for PCE and from 121-1218 mg/kg for TCE suggest the presence of DNAPL. Concentrations below 36.8 mg/kg and 121 mg/kg do not indicate the presence of DNAPL, but can also not reject it.	PCE was 3.6 mg/kg and TCE was 4.8 mg/kg.
Concentration (Air)	If the concentration in the air is less than 100 ppm, there is little chance of DNAPL; if it is between 100-1000 ppm, it indicates the presence of DNAPL; if it is greater than 1000 ppm, the presence of DNAPL is highly likely [ <i>Janniche et al.</i> , 2013]	None
MIP <sup>2</sup>	MIP <sup>TM</sup> soundings provide in situ measurements of volatile organic carbon components and are ideal for rapidly screening an area for the presence of DNAPL, but do not provide actual concentration measurements. MIP is therefore best when it can be compared to actual concentration measurements [ <i>ITRC</i> , 2013; <i>Jørgensen et al.</i> , 2010].	None
Sudan IV, FLUTE <sup>TM,3</sup> , LIF/dye LIF <sup>4</sup>	Methods that prove the presence of DNAPL. If they are positive, DNAPL is present. If they are negative DNAPL could still be found elsewhere [ <i>ITRC</i> , 2013]. If there is no test information, then the indicator has no influence on the state of the TCE source phase node. <ul style="list-style-type: none"> <li>• Sudan IV: Hydrophobic dye test that shows the presence of DNAPL in soil samples [<i>ITRC</i>, 2013].</li> <li>• FLUTE<sup>2</sup> Liner: A ribbon NAPL sampler method that pinpoints DNAPL locations in boreholes where DNAPL is expected [<i>ITRC</i>, 2013].</li> <li>• LIF<sup>3</sup>/dye LIF: A direct push sounding with an energy source that activates the fluorinating DNAPL compound and measures the response.</li> </ul>	None

### 4.3 The network for determining fractures in the clay till

Figure 8 shows the BBN for assessing whether or not fractures are present in the clay till. This BBN consists of 10 nodes and 9 links. The central node of this BBN is the “Clay till” node, which has been assigned three states: fractured (F), unfractured (U) and No evidence. Here, we are interested in investigating if fractures that penetrate the entire clay till profile are present, because such fractures lead to the greatest risk of contaminating the underlying aquifer. The state of the Clay till node is inferred from the following six indicator nodes (Table 4): Drainage conditions, Redox boundary, Poly-morphological type, Thickness of till, Till types and Observed fracture. These indicators provide information about fractures in the clay till and are primarily based on the work presented in *Klint*, [2001], *Klint* [2014] and *Klint et al.* [2013]. The individual indicator nodes are briefly described in the following. More information on the different nodes as well as their associated CPTs can be found in the Supporting Information.

The “Potential development of fractures 1” node describes the potential for occurrence and distribution of fractures based on the type of clay till. The till types have here been classified based on the simple system developed by *Klint* [2001], in which tills are divided into: i) basal till type A and type B, ii) glacioteconites, iii) melt-out tills, iv) flow tills, and v) drop tills. This classification system takes into account the distribution of fractures within the tills and is mainly based on directional elements embedded in the matrix and internal deformation structures such as faults, fractures and folds [*Klint et al.*, 2013]. For each of these till types, *Klint et al.* [2013] provide a description of their physical characteristics, as well as an expert assessment of their geological

heterogeneity, and this information has been used here to inform the CPT for the “Potential development of fractures 1” node.

The node “Potential development of fractures 2” combines till thickness with the poly-morphological landform type. According to *Klint et al.* [2013], tills that are greater than 15 meters in thickness are generally not fractured, while tills of between 10-15 m thickness are occasionally fractured through the entire till cover. Clay tills of less than 10 meters in thickness have medium to high probability of fractures, but this depends greatly on the poly-morphological type [*Klint et al.*, 2013]. The classification of poly-morphology types is based on the work by *Klint et al.* [2013]. This classification was developed to describe glacial land system areas, where two or more glacial landforms are superimposed. The different poly-morphological types therefore refer to areas in glacial terrains that are potentially unique in geological heterogeneity, especially in terms of the distribution of fractures and sand lenses in tills. Ten poly-morphological types are defined and mapped for the island of Zealand, Denmark in *Klint et al.* [2013]. The different types are given in Figure 8 and summarized in Table 4. For each of these poly-morphological types, *Klint et al.* [2013] present an expert judgment of the potential degree of geological heterogeneity (primarily fracturing) in clay tills of up to 10 m in thickness, and this information has been used here to populate the CPT for the “Potential development of fractures 2” node.

The last indicator node, “Potential development of fractures 3”, summarizes the effect of the indicator nodes “Drainage conditions” and “Redox boundary”. The CPT for the “Potential development of fractures 3” node generally reflects that the size and intensity of fractures in well drained tills (tills that overlay permeable deposits) are greater than in poorly drained tills (tills that overlay deposits with a low permeability) [*Evans et al.*, 2006; *Klint et al.*, 2013] and that the density of fractures decreases in till units below the redox boundary [*Hendry, 1982; Jørgensen et al.*, 2003; *Klint et al.*, 2013]. The CPT has been designed in such way that for the case where both variables

(Drainage conditions and Redox boundary) indicate fractures, the belief of the occurrence of fractures has been set to 1, while the case where both variables indicate no fractures, a belief of the occurrence of fractures of 0.1 rather than 0 has been specified. The reason is that, while the combination of a ‘well-drained till’ and ‘till is above the redox boundary’ will lead to fractures, the opposite is not true since other factors such as depositional history, etc. can lead to the development of fractures.

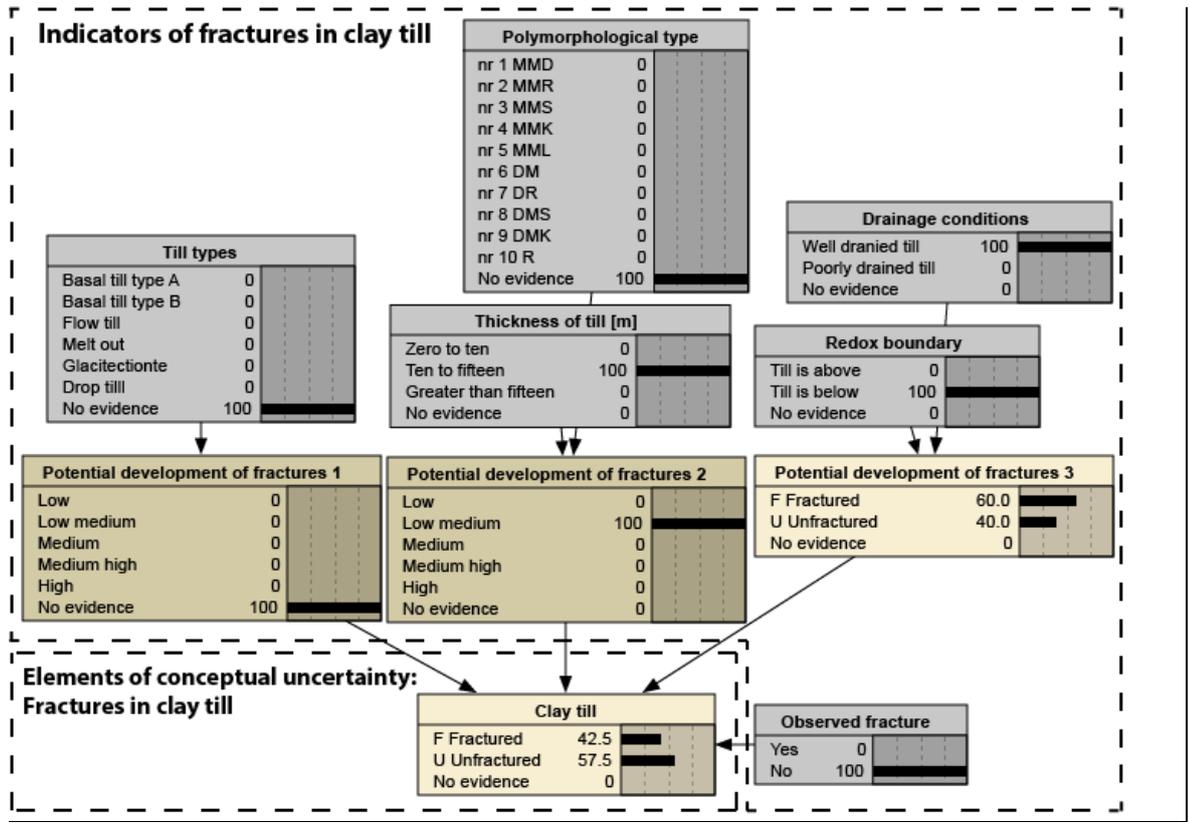


Figure 8. Bayesian belief network for the occurrence of fractures in clay till.

**Table 4.** Descriptions of the indicator nodes in the fractures network and the screening investigation data that is relevant for each node.

<b>Node</b>	<b>Description</b>	<b>Evidence</b> (From screening investigation)
Till types	The till type is an indicator of the presence and intensity of fractures . “The classification of clay tills is primarily based on directional elements embedded in the matrix such as clast fabric, scour marks on clasts, and internal deformation structures such as faults, fractures and folds” [Klint <i>et al.</i> , 2013].	None
Poly-morphological type	Poly-morphological types are areas defined in terms of fractures and sand lens distributions in glacial terrain. Ten poly-morphological types are mapped for the island of Zealand, Denmark and presented in Klint <i>et al.</i> [2013]. The poly-morphological units that form the ten poly-morphological types are: M: Undulating till plain D: Hummocky moraine S: Outwash plain R: marginal moraine K: Basement limestone L: Basement marine clay	None
Thickness of till	The thickness of the till is defined as the depth to the uppermost aquifer. Tills that are greater than 15 m are generally not fractured, tills between 10-15 m are occasionally fractured, and tills of less than 10 m have medium to high probability of fractures but depend greatly on the poly-morphological type [Klint <i>et al.</i> , 2013].	Clay thickness data is available from borehole profiles. The till is 15 m thick. In order to model conservatively, we chose 10-15 m.
Drainage conditions	The size and intensity of fractures in tills that overlay permeable deposits (well drained tills) are greater than in tills that overlay deposits with a low permeability (poorly drained tills) [Evans <i>et al.</i> , 2006; Klint <i>et al.</i> , 2013].	The till is well drained because a sandy aquifer is underlying it.
Redox boundary	Fracture density decreases across the transition between the oxidized and the underlying reduced till [Hendry, 1982; Jørgensen <i>et al.</i> , 2003; Klint <i>et al.</i> , 2013].	We know from the borehole profiles that the till is below the redox boundary
Observed fracture	Observations of fractures that cover the entire clay profile. If a fracture is observed, the clay till node will be in the state of fractured, regardless of the states of all other indicator nodes. If it is not observed, the state of the other indicators determine the outcome.	No fracture was observed

## **5 Results and discussion**

To demonstrate the BBN method and illustrate how the beliefs in the different CSMs can be updated when more information becomes available, we start by using the data only from the screening investigation (Section 3.1.1) as evidence to update the nodes. We then use the information from the detailed investigations (Section 3.1.2) as additional evidence, and lastly incorporate the expert knowledge (Section 3.1.3) in the assessment.

### **5.1 Belief updating using data from the screening investigation**

As shown in Table 1 and Table 2, the screening investigations provided evidence for the following source phase-related indicators: the aqueous concentrations of PCE and TCE. By entering this as evidence in the network and specifying that no test, sample or observation are available for all the other indicator nodes (as shown in Figure 7), it is found that the probability of DNAPL being present in the source is 65 %. If no data or information had been available, the probability would have been 100% no test.

The screening investigation also provided information on some of the clay till indicators, as described in Table 4. From the screening investigation it is known that: i) the till is well drained due to the underlying sandy aquifer; ii) the till is 15 meters thick; iii) the till is below the redox boundary; and iv) that fractures were not observed during the field investigation. By entering this information as evidence into the clay till network (as shown in Figure 8, where all the grey nodes are nodes where evidence from the screening investigation exists), beliefs of 42.5 % and 57.5 % for the presence and absence of fractures in the clay till, respectively, have been calculated. Combining the beliefs from the source phase network and the clay till network results in the weights shown in Table 5 for each of the CSMs. Thus, the available information from the screening investigations appears to favour the UN model slightly. The probabilities of the four models are fairly uniform,

suggesting that the available information at this stage, as expected, is rather weak and the degree of conceptual uncertainty is high.

## **5.2 Belief updating using data from the detailed investigations of the source phase**

The detailed site investigation provided more information on the source phase. The MIP soundings were successful at determining the depths where the concentrations were elevated, indicating high concentrations of contamination on either side of borehole B2. In D4 (west of B2, map on Figure 2) 56 mg/kg TCE were reported, while 89 mg/kg PCE were found in D12 (east of B2). This was interpreted as evidence for the presence of two sources with different compositions, one mainly comprising of TCE and one of PCE. The PCE concentration changes the state of the PCE concentration soil node from less than 36 mg/kg, to between 36 mg/kg and 368 mg/kg. In addition, the MIPs strengthen our belief in the presence of a DNAPL phase, and we thus changed the MIP node to a state of 75 % DNAPL and 25 % Dissolved. The beliefs were assigned on account of the authors' evaluation of the results.

The data from the detailed investigation (Section 3.1.2) resulted in an increase in the belief in a DNAPL source from 65 % to 70 %. By combining the updated belief in the source phase with the clay till beliefs from Section 5.1, the beliefs in the four CSMs are also updated, as shown in Table 6. After incorporating the information from the detailed investigations, the UN model is still favored and the belief in this is now more distinctive, with an increase from 37.4% to 40.2%.

## **5.3 Belief updating using expert knowledge as evidence**

In a final step, the information from the expert consultation is used as evidence in the assessment of the CSMs. Based on the expert consultation, it is believed that the clay till is either basal till type B (80 %) or a glaciectonite (20 %). This is again entered as evidence into the network by changing the states of the till type node accordingly. Applying the expert knowledge (Section 3.1.3) increased

the belief in the presence of fractures in the clay till from 42.5 % to 61.3 %, resulting in a new set of CSM weights, as shown in Table 5. Because of the increased belief in fractures being present in the clay till, the favored model is now FN and it is more distinct compared to when only the initial and detailed site investigation data were incorporated.

**Table 5.** The weights (%) for the four models after the screening investigation, the detailed investigation of the site with MIP and soil samples, and after consulting an expert. The weight of the most likely model after each site investigation is shown in bold.

Model	Screening investigation	Detailed investigation	Expert opinion
FN (Fractured and (D) NAPL)	20.1	29.8	<b>42.9</b>
UN (Unfractured and (D) NAPL)	<b>37.4</b>	<b>40.2</b>	27.1
FD (Fractured and Dissolved)	14.9	12.8	18.4
UD (Unfractured and Dissolved)	20.1	17.3	11.6

#### 5.4 Applicability of the developed BBN method

Above we have demonstrated how the BBN method can be applied to a contaminated site to systematically formulate a number of CSMs and then determine and update the beliefs in each of them using the available data and information. The method can thus explicitly assess the conceptual uncertainty at the site.

The use of the BBNs to assess the beliefs in individual conceptual models is promising, as it allows for integrating available quantitative data with qualitative information and expert opinion in a probabilistic assessment of the different models. In most cases, quantitative risk assessments of contaminated sites are based on very limited quantitative data (from screening investigations) and usually do not consider uncertainties. A method which allows for incorporating more qualitative information into the assessment, and is flexible with regards to data types and levels while accounting for uncertainties, is of practical interest and may help strengthen decision-making in situations where empirical data are lacking.

Formulating and using multiple CSMs for e.g. management of contaminated site will clearly result in more work and is more time consuming. We therefore have suggested and adopted a modular-type approach for formulating the different CSMs by systematically combining different uncertain conceptual elements. The graphical component of the BBN makes this process very transparent. Once a network has been developed for a given conceptual element (here: assessment of the presence of DNAPL and fractures), we believe it should be possible to transfer and apply such a network to other similar sites. A main challenge in developing a BBN for a given conceptual element is to formulate and agree on a suitable causative model structure, i.e. decide which nodes (variables) should be included, how many discrete state values each node should have and how the different nodes should be linked to each other. In this work, the structures of the BBNs were developed following an iterative process involving reviewing literature on DNAPL and clay tills, respectively, as well as meetings between the authors and other experts in the field. Another main challenge is the specification of the conditional probability tables for all of the child nodes in the network. The specification of CPTs can be done in various ways (e.g. from empirical data, existing models or expert consultation) and can be a very time-consuming process, especially where it involves a formal elicitation of expert opinions (e.g. as done in *Stiber et al.* [1999]). The developed BBN method allows for different predictive and/or diagnostic inference scenarios to be conducted based on observations. Predictive inference (information flows with the direction of the arrows as in the study site in this paper) can explore how changes in the beliefs/knowledge of one or more of the indicators influence the conceptual understanding at the site. With diagnostic inference scenarios (information flow against the direction of the arrows) it is possible to explore which conditions need to be met in order for a specific conceptual element to be confirmed or rejected. For example, if we wanted to investigate what data to collect to document that the clay till is fractured, then this information could be retrieved from the clay till network simply by specifying the state of the “Clay

till” node to 100% fractured. The BBN then performs a backward propagation to determine the updated beliefs of all other variables in the network. Such predictive and diagnostic inference scenarios, together with the fact that the method determines the beliefs in each CSM explicitly, can potentially be helpful for informing and directing future investigations at a site. For example, new investigations could be specifically targeted at confirming or rejecting one (or more) of the conceptual models, by use of diagnostic inference, to reduce the overall conceptual uncertainty at the site.

## 5.5 Method limitations and future developments

It is important to note that while the inclusion of multiple CSMs in the BBN method results in a more robust assessment, it does not mean that all conceptual uncertainty has been fully accounted for. In the end we can of course only include those models we are able to perceive and the results do therefore only represent a subset of all the possibilities. The number of feasible conceptual models may in principle be much larger, but an exhaustive inclusion of all potential models can quickly become impractical. This is a common problem in the multi-model and BMA related literature [e.g. *Neumann, 2003; Elshall and Tsai, 2014*]. In practice, this problem is typically overcome by considering and averaging over a manageable subset of the most parsimonious models in view of the knowledge about the system [*Neuman, 2003; Rojas et al., 2008*]. For simplicity and for demonstration purposes, only two elements of conceptual uncertainty have been considered here. Nevertheless, our four models have addressed the key questions of conceptualization that were found to be crucial at the site. Other factors that might be important to consider at the site could for instance include PCE and TCE degradation, sand lenses and/or the spreading of the contaminant to down-gradient receptors. The modular and systematic approach we use for formulating the CSMs would allow us to extend the network to include other conceptual elements. As noted previously, we could for example include the reductive dechlorination BBN presented in *Stiber et al. [1999]* as

an additional indicator network, in which case we would have three elements of conceptual uncertainty and a total of eight possible CSMs.

There is also the potential to further refine the developed BBNs to e.g. include more indicator variables and/or by consulting more experts, as well as to test and validate them against more evidence. The CPTs in the developed BBNs are currently based on a general assumption that the available data and information are rather limited. A future modification of the BBNs could be to extend the networks to explicitly account for the number and the quality of the available data at the site to give a better assessment of the strength of the field campaign. Also, for some of the nodes in the networks it might be possible to replace the use of a ‘No evidence’ state by specifying a prior belief distribution that reflects the average conditions for that node. For example, rather than using a ‘No evidence’ state for the ‘Till types’ node in the fracture network (Figure 8), it might be possible to specify a distribution of the different till types based on how typical they are for the considered region. In the absence of any site-specific information on clay till types, regional knowledge would be the best available information.

Both the development of the BBN structure and the specification of the CPTs in this study are based on a number of subjective choices, and hence the BBN largely reflects our perception of the processes occurring at the specific site. For more extensively investigated study sites, it might be possible to derive CPTs or even the structure directly from the data. However, site assessments with sparse data, like the one studied here will always be more subjective than those for well documented sites. This means that the model results will be sensitive to the subjective choices in both the CPTs and the BBN structure. While it is fairly straightforward to carry out sensitivity analyses of how strongly the different nodes in the network influence the inference results, given the structure and the CPTs (this can be done e.g. using Netica’s in-built sensitivity analysis tools), it

is more challenging to assess the influence of the subjectively chosen structure and the CPTs, and this is to our knowledge not something that is routinely done in the literature.

In order to assess the sensitivity of the structure of the BBN one option is to consult and elicit advice from multiple experts when developing the BBN and specifying the CPTs. Different modelers and experts may have different beliefs or even conflicting opinions about the same phenomenon. By eliciting and using the opinions from multiple experts, the specified probabilities in the CPTs could then be an ‘average’ reflection of the range of opinions offered or the opinions elicited from the multiple experts could be used to populate individual BBNs, each representing the opinion of an individual expert. The latter approach was for example used in the paper by *Stiber et al.* [1999], where 21 experts were asked to develop BBNs for the assessment of anaerobic dechlorination at TCE contaminated sites, resulting in 21 individual BBNs. In this study, we developed only a single model.

Finally, in this work we have focused entirely on using the BBN approach to assess the beliefs in a set of CSMs. Future work could look into coupling this approach with suitable fate and transport models to explicitly illustrate the impact of the conceptual uncertainty on e.g. a groundwater risk assessment. The developed BBNs could potentially also be of relevance in other contexts. For example, the BBN for assessing the presence of fractures could be incorporated into a groundwater vulnerability assessment method, such as DRASTIC [*Aller et al., 1987*] which considers seven geological and hydrogeological parameters including groundwater recharge and the impact of the geological unit above the aquifer, both of which will be strongly influenced by presence of fractures.

## 6 Conclusions

A Bayesian Belief Network (BBN) approach for assessing the conceptual uncertainty at a contaminated site has been presented. The proposed approach accounts for conceptual model uncertainty by incorporating multiple conceptual site models, each representing an alternative interpretation of the site settings, into a BBN. The BBN is subsequently used to determine the beliefs in each CSM using a variety of data from desktop studies and initial site investigations together with expert opinion. The formulation of the different CSMs is done systematically by combining different uncertain conceptual elements, a process which is made transparent due to the graphical feature of the BBN.

The method was demonstrated on a Danish study site contaminated with chlorinated ethenes. Four different CSMs were developed for this site by combining two source zone interpretations (presence or absence of NAPL-phase in the source zone) with two geological interpretations (presence or absence of fractures in the clay till). The four CSMs were implemented in a BBN, which was developed based on a wide range of information and data types including information from borehole profiles, chemical data from sampling of soil water and air, different tests for the presence of DNAPL (e.g. FLUTE, SUDAN IV, LIF/dye LIF), geological information including the impact of strain/deformation on the clay till, and expert opinion.

At the study site, data was available in three stages (a screening investigation, a detailed investigation, and an expert consultation). Using the developed BBN, the beliefs in each of the CSMs were assessed sequentially based on the information from the three investigation stages to explore how the belief in each CSM changed as more information became available. An increasing belief in both the presence of a separate DNAPL phase in the source zone and the presence of fully penetrating fractures in the clay till was found with the increasing amount of available information,

thus mainly supporting the CSM accounting for both a DNAPL source and a fractured clay (model FN).

The proposed BBN method has several advantages. It explicitly accounts for uncertainty and allows for incorporating both quantitative and qualitative information in the assessment of the different CSMs. It is flexible and easy to use, and can be used for both predictive and diagnostic inference scenarios. The method also allows for updating the beliefs in the CSMs when additional data has been collected. Because of these features, the method may be of both practical and scientific interest. For practitioners, the method might be particularly useful for risk assessments and to help direct site investigations and remedial actions.

## **7 Acknowledgements**

We thank the Capital Region of Denmark for access to the data concerning the site. We gratefully acknowledge the support of the Danish Agency for Science Technology and Innovation for funding the project RiskPoint – Assessing the risks posed by point source contamination to groundwater and surface water resources and the Technical University of Denmark for funding this study. We thank our two anonymous reviewers for constructive comments and guidance for how to improve the manuscript.

## 8. References

- Aguilera, P., A. Fernández, R. Roper, and L. Molina, 2013, Groundwater quality assessment using data clustering based on hybrid Bayesian networks, *Stoch. Environ. Res. Risk Assess.*, 27(2): 435-447, doi: 10.1007/s00477-012-0676-8.
- Aller, L. A., T. Bennet, J. H. Lehr, and G. Hackett, 1987, *DRASTIC: A standardized system for evaluating ground water pollution potential using hydrogeologic settings*, Rep: EPA/600/2-87/035, US EPA, Worthington, Ohio, USA.
- ASTM Standard E1689, 2008, *Standard Guide for Developing Conceptual Site Models for Contaminated Sites*, ASTM International, USA, doi: 10.1520/E1689-95R08, [www.astm.org](http://www.astm.org).
- ASTM Standard E1739 - 95(2010)e1, 2010, *Standard Guide for Risk-Based Corrective Action Applied at Petroleum Release Sites*, ASTM International, USA, doi: 10.1520/E1739-95R10, [www.astm.org](http://www.astm.org).
- ASTM Standard E2531, 2009, *Standard Guide for Developing Conceptual Site Models and Remediation Strategies for Nonaqueous-Phase Liquids Released to the subsurface*, ASTM international, USA, doi: 10.1520/E1689-95R08, [www.astm.org](http://www.astm.org).
- Bedient, P. B., H. S. Rifai, and C. J. Newell, 1999, *Ground Water Contamination: Transport and Remediation (2nd Edition)*, Prentice Hall, USA.
- Beven, K., 2009, *Environmental modelling: an uncertain future? An Introduction to Techniques for Uncertainty Estimation in Environmental Prediction*, Routledge, UK.
- Bjerg, P. L., N. Tuxen, L. A. Reitzel, H. J. Albrechtsen, and P. Kjeldsen, 2011, Natural Attenuation Processes in Landfill Leachate Plumes at Three Danish Sites, *Ground Water*, 49(5): 688-705, doi: 10.1111/j.1745-6584.2009.00613.x.
- Bredehoeft, J., 2005, The conceptualization model problem-surprise, *Hydrogeol. J.*, 13(1): 37-46, doi: 10.1007/s10040-004-0430-5.

Chambon, J. C., P. J. Binning, P. R. Jørgensen, and P. L. Bjerg, 2011, A risk assessment tool for contaminated sites in low-permeability fractured media, *J. Contam. Hydrol.*, 124(1-4): 82-98, doi:10.1016/j.jconhyd.2011.03.001.

Chan, T., H. Ross, S. Hoverman, & B. Powell, 2010, Participatory development of a Bayesian network model for catchment based water resource management. *Water Resour. Res.*, 46(7): W07544, doi:10.1029/2009WR008848.

Chitsazan, N. and F. T. C. Tsai, 2015, A hierarchical Bayesian model averaging framework for groundwater prediction under uncertainty, *Ground Water*; 53(2): 305-16. doi: 10.1111/gwat.12207.

Christiansen, C. M., C. Riis, S. B. Christensen, M. M. Broholm, A. G. Christensen, K. E. Klint, J. S. A. Wood, P. Bauer-Gottwein, and P. L. Bjerg, 2008, Characterization and Quantification of Pneumatic Fracturing Effects at a Clay Till Site, *Environ. Sci. Technol.*, 42(2): 570-576, doi: 10.1021/es071294s.

Damgaard, I., P. L. Bjerg, J. Bælum, C. Scheutz, D. Hunkeler, C. S. Jacobsen, N. Tuxen, and M. M. Broholm, 2013a, Identification of chlorinated solvents degradation zones in clay till by high resolution chemical, microbial and compound specific isotope analysis, *J. Contam. Hydrol.*, 146(0): 37-50, doi: 10.1016/j.jconhyd.2012.11.010.

Damgaard, I., P. L. Bjerg, C. S. Jacobsen, A. Tsitonaki, H. Kern-Jespersen, and M. M. Broholm, 2013b, Performance of Full-Scale Enhanced Reductive Dechlorination in Clay Till, *Groundwater Monit. R.*, 33(1): 48-61, doi: 10.1111/j.1745-6592.2012.01405.x.

Danish EPA, 2002, Guidelines on Remediation of Contaminated Sites, Environmental Guidelines No.7 2002, Danish EPA, Denmark.

de Barros, F. P. J., Bolster, D., Sanchez-Vila and Nowak, W., 2011, A divide and conquer approach to cope with uncertainty, human health risk, and decision making in contaminant hydrology, *Water resources research*, 47 (5):W05508, doi: 10.1029/2010WR009954.

Elshall, A. S., and F. T. C. Tsai, 2014, Constructive epistemic modeling of groundwater flow with geological structure and boundary condition uncertainty under the Bayesian paradigm. *J. Hydrol.*, 517 (2014): 105-119, doi: 10.1016/j.jhydrol.2014.05.027.

Evans, D. J. A., E. R. Phillips, J. F. Hiemstra, and C. A. Auton, 2006, Subglacial till: Formation, sedimentary characteristics and classification, *Earth-Sci. Rev.*, 78(1-2): 115-176, doi:10.1016/j.earscirev.2006.04.001.

Farmani, R., H. J. Henriksen, and D. Savic, 2009, An evolutionary Bayesian belief network methodology for optimum management of groundwater contamination, *Environ. Modell. Softw.*, 24(3): 303-310, doi:10.1016/j.envsoft.2008.08.005.

Farmani, R., H. J. Henriksen, D. Savic, and D. Butler, 2012, An evolutionary Bayesian belief network methodology for participatory decision making under uncertainty: An application to groundwater management, *Integrated. Environ. Assess. Manag.*, 8(3): 456-461, doi: 10.1002/ieam.192.

Fienen, M. N., J. P. Masterson, N. G. Plant, B. T. Gutierrez, and E. R. Thieler, 2013, Bridging groundwater models and decision support with a Bayesian network. *Water Resour. Res.*, 49(10): 6459-6473, doi: 10.1002/wrcr.20496.

Foglia, L., S. W. Mehl, M. C. Hill, P. Perona, and P. Burlando, 2007, Testing Alternative Ground Water Models Using Cross-Validation and Other Methods, *Ground Water*, 45(5): 627-641, doi: 10.1111/j.1745-6584.2007.00341.x.

Freeze, R., A., B. James, J. Massmann, T. Sperling, and L. Smith, 1992, Hydrogeological Decision-Analysis .4. the Concept of Data Worth and Its Use in the Development of Site Investigation Strategies. *Ground Water*. 30 (4), 574-588.

Georgakakos, K. P., D. J. Seo, H. Gupta, J. Schaake, and M. B. Butts, 2004, Towards the characterization of streamflow simulation uncertainty through multimodel ensembles, *J. Hydrol.*, 298(1-4): 222-241, doi:10.1016/j.jhydrol.2004.03.037.

Gerber, R. E., J. I. Boyce, and K. W. F. Howard, 2001, Evaluation of heterogeneity and field-scale groundwater flow regime in a leaky till aquitard, *Hydrogeol. J.*, 9(1): 60-78, doi: 10.1007/s100400000115.

Hadley, P. W. and C. J. Newell, 2012, Groundwater Remediation: The Next 30 Years, *Ground Water*, 50(5): 669-678, doi: 10.1111/j.1745-6584.2012.00942.x.

Harrar, W. G., L. C. Murdoch, B. Nilsson, and K. E. S. Klint, 2007, Field characterization of vertical bromide transport in a fractured glacial till, *Hydrogeol. J.*, 15(8): 1473-1488, doi: 10.1007/s10040-007-0198-5.

Hedeselskabet, 2005, Indledende forureningsundersøgelse, tidligere kemikaliefordelingscentral, Vadsbyvej 16 A, English: Screening investigation, former chemical distribution central, Vadsbyvej 16A, Copenhagen County, Denmark.

Hendry, M. J., 1982, Hydraulic conductivity of a glacial till in Alberta, *Ground Water*, 20(2): 162-169, doi: 10.1111/j.1745-6584.1982.tb02744.x.

Hendry, M. J., C. J. Kelln, L. I. Wassenaar, and J. Shaw, 2004, Characterizing the hydrogeology of a complex clay-rich aquitard system using detailed vertical profiles of the stable isotopes of water, *J. Hydrol.*, 293(1-4): 47-56, doi:10.1016/j.jhydrol.2004.01.010.

Henriksen, H. J., P. Rasmussen, G. Brandt, D. von Bulow, and F. V. Jensen, 2007a, Public participation modelling using Bayesian networks in management of groundwater contamination, *Environ. Modell. Softw.*, 22(8): 1101-1113, doi:10.1016/j.envsoft.2006.01.008.

Henriksen, H. J., P. Rasmussen, J. Bromley, A. D. L. H. Portillo, and M. R. Llamas, 2007b, Bayesian belief networks as a tool for participatory integrated assessment and adaptive groundwater management: the Upper Guadiana Basin, Spain, Geological Survey of Denmark and Greenland Bulletin, (13): 69-72.

Henriksen, H. J. and H. C. Barlebo, 2008, Reflections on the use of Bayesian belief networks for adaptive management, J. Environ. Manage., 88(4): 1025-1036, doi:10.1016/j.jenvman.2007.05.009.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky, 1999, Bayesian model averaging: A tutorial, Stat. Sci., 14(4): 382-401.

Højberg, A. and J. Refsgaard, 2005, Model uncertainty parameter uncertainty versus conceptual models, Water Sci. Technol., 52(6): 177-186.

Hovedstadsrådet, 1980, Geologisk basisdatakort, English: Geological maps of the Capital region of Denmark, Hovedstadsrådet, Denmark.

ITRC, 2013, An Introduction to Characterizing Sites Contaminated with DNAPLs, The Interstate Technology and Regulatory Council, USA, <http://www.itrcweb.org/Documents/DNAPLs-4.pdf>.

James, A. L. and C. M. Oldenburg, 1997, Linear and Monte Carlo uncertainty analysis for subsurface contaminant transport simulation, Water. Resour. Res., 33(11): 2495-2508, doi: 10.1029/97WR01925.

Janniche, G. S., A. S. Fjordbøge, and M. M. Broholm, 2013, DNAPL i moræner og kalk - vurdering af undersøgelsesmetoder og konceptuel modeludvikling, Naverland 26AB, Albertslund, English: DNAPL in clay till and limestone - evaluation of investigation methods and conceptual modeling, Naverland 26AB, Albertslund, DTU Environment, Denmark, <http://www.sara.env.dtu.dk/>.

Jørgensen, I. V., M. M. Broholm, and P. L. Bjerg, 2010, DNAPL i kildeområder - konceptuelle modeller, karakterisering og estimering af forureningsmasse, English: DNAPL in source zones- conceptual models, characterization and estimation of mass, DTU Environment, Denmark, <http://www.sara.env.dtu.dk/>.

Jørgensen, P. R., K. E. S. Klint, and J. P. Kistrup, 2003, Monitoring Well Interception with Fractures in Clayey Till, *Ground Water*, 41(6): 772-779, doi: 10.1111/j.1745-6584.2003.tb02418.x.

Kessler, T. C., K. E. S. Klint, B. Nilsson, and P. L. Bjerg, 2012, Characterization of sand lenses embedded in tills, *Quaternary Science Reviews*, 53(0): 55-71, doi:10.1016/j.quascirev.2012.08.011.

Kessler, T. C., A. Comunian, F. Oriani, P. Renard, B. Nilsson, K. E. Klint, and P. L. Bjerg, 2013, Modeling Fine-Scale Geological Heterogeneity - Examples of Sand Lenses in Tills, *Ground Water*, 51(5): 692-705, doi: 10.1111/j.1745-6584.2012.01015.x.

Klint, K. E., 2001, Fractures in Glacigene Diamict Deposits; Origin and Distribution., Ph. D. thesis, Geological Survey of Denmark and Greenland-GEUS, Copenhagen, Denmark.

Klint, K. E., 2014, Hydrologisk evaluering - klassificering af jordforhold ved punktforurening, English: Hydrological evaluation - classification of soil conditions at point source contaminations, Web site: <http://geuskort.geus.dk/siteeval/>, accessed on 02-09-2014.

Klint, K. E. S., B. Nilsson, L. Troldborg, and P. R. Jakobsen, 2013, A poly morphological landform approach for hydrogeological applications in heterogeneous glacial sediments, *Hydrogeol. J.*, 21(6): 1247-1264, doi: 10.1007/s10040-013-1011-2.

Konikow, L. F. and J. D. Bredehoeft, 1992, Ground-water models cannot be validated, *Adv. Water Resour.*, 15(1): 75-83, doi: 10.1016/0309-1708(92)90033-X.

Korb, K. B. and A. E. Nicholson, 2003, Bayesian artificial intelligence, CRC Press, USA.

Li, X. B. and F. T. C. Tsai, 2009, Bayesian model averaging for groundwater head prediction and uncertainty analysis using multimodel and multimethod, *Water Resour. Res.*, 45(9), W09403, doi: 10.1029/2008WR007488.

McMahon, A., J. Heathcote, M. Carey, and A. Erskine, 1999, Guide to Good Practice for the Development of Conceptual Models and the Selection and Application of Mathematical Models of Contaminant Transport Processes in the Subsurface, Publication: NC/99/38/3, UK Environment Agency, UK, <https://www.gov.uk/government/publications>.

Neuman, S. P., 2003, Maximum likelihood Bayesian averaging of uncertain model predictions, *Stoch. Env. Res. Risk. A.*, 17(5): 291-305, doi: 10.1007/s00477-003-0151-7.

Neuman, S. P. and P. J. Wierenga, 2003, A Comprehensive Strategy of Hydrogeologic Modeling and Uncertainty Analysis for Nuclear Facilities and Sites, University of Arizona, Tucson Arizona, United States, Rep. NUREG/CR-6805.

Nielsen, T. D. and F. V. Jensen, 2007, Bayesian networks and decision graphs, Springer, USA.

Norsys Software Corp., 2014, Netica Version 4.16, Norsys Software Corp, Canada, [www.norsys.com/](http://www.norsys.com/).

O'Hara, S. K., B. L. Parker, P. R. Jorgensen, and J. A. Cherry, 2000, Trichloroethene DNAPL flow and mass distribution in naturally fractured clay: Evidence of aperture variability, *Water Resour. Res.*, 36(1): 135-147, doi: 10.1029/1999WR900212.

Orbicon A/S, 2007, Vadsbyvej 16A - supplerende undersøgelser, English: Vadsbyvej 16A - further investigations, Capital Region of Denmark, Denmark.

Pankow, J. F. and J. A. Cherry, 1996, Dense chlorinated solvents and other DNAPLs in groundwater: history, behavior, and remediation, Waterloo Press, USA.

Parker, B. L., S. W. Chapman, and M. A. Guilbeault, 2008, Plume persistence caused by back diffusion from thin clay layers in a sand aquifer following TCE source-zone hydraulic isolation, *J. Contam. Hydrol.*, 102(1-2): 86-104, doi:10.1016/j.jconhyd.2008.07.003.

Poeter, E. and D. Anderson, 2005, Multimodel Ranking and Inference in Ground Water Modeling, *Ground Water*, 43(4): 597-605, doi: 10.1111/j.1745-6584.2005.0061.x.

Rambøll, 1999, Grundvandspotentiale i Kalkmagasinet oktober 1999, English: Groundwater potential on the limestone aquifer October 1999, Copenhagen County and HOFOR, Denmark.

Refsgaard, J. C., J. P. van der Sluijs, J. Brown, and P. van der Keur, 2006, A framework for dealing with uncertainty due to model structure error, *Adv. Water Resour.*, 29(11): 1586-1597, doi:10.1016/j.advwatres.2005.11.013.

Rojas, R., L. Feyen, and A. Dassargues, 2008, Conceptual model uncertainty in groundwater modeling: Combining generalized likelihood uncertainty estimation and Bayesian model averaging, *Water Resour. Res.*, 44(12), doi: 10.1029/2008WR006908.

Shihab, K., 2008, Dynamic modeling of groundwater pollutants with Bayesian networks, *Appl. Artif. Intell.*, 22(4): 352-376, doi: 10.1080/08839510701821645.

Shihab, K. and N. Chalabi, 2007, Dynamic modeling of ground-water quality using Bayesian techniques, *J. Am. Water Resour. As.*, 43(3): 664-674, doi: 10.1111/j.1752-1688.2007.00053.x.

Singh, A, Minsker, B. S., Valocchi, A. J., 2008, An interactive multi-objective optimization framework for groundwater inverse modeling, *Adv. Water Resour.*, 31 (10): 1269–1283, doi:10.1016/j.advwatres.2008.05.005 .

Slough, K. J., E. A. Sudicky, and P. A. Forsyth, 1999, Numerical simulation of multiphase flow and phase partitioning in discretely fractured geologic media, *J. Contam. Hydrol.*, 40(2): 107-136, doi: 10.1016/S0169-7722(99)00051-0.

Small, M. J., 1997, Groundwater detection monitoring using combined information from multiple constituents, *Water Resour. Res.*, 33(5): 957-969, doi: 10.1029/96WR03977.

Stiber, N. A., M. Pantazidou, and M. J. Small, 1999, Expert system methodology for evaluating reductive dechlorination at TCE sites, *Environ. Sci. Technol.*, 33(17): 3012-3020, doi: 10.1021/es981216s.

Stiber, N. A., M. J. Small, and M. Pantazidou, 2004, Site-specific updating and aggregation of Bayesian belief network models for multiple experts, *Risk Anal.*, 24(6): 1529-1538, doi: 10.1111/j.0272-4332.2004.00547.x.

Sohn, M. D., M. J. Small, and M. Pantazidou, 2000, Reducing uncertainty in site characterization using Bayes Monte Carlo methods. *J. Environ. Eng.* 126 (10), 893-902, doi: 10.1061/(ASCE)0733-9372(2000)126:10(893).

Suter, G. W., 1999, Developing conceptual models for complex ecological risk assessments, *Hum. Ecol. Risk Assess.*, 5(2): 375-396, doi: 10.1080/10807039991289491.

Tebaldi, C., R. L. Smith, D. Nychka, and L. O. Mearns, 2005, Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multimodel ensembles, *J. Climate.*, 18(10): 1524-1540, doi: 10.1175/JCLI3363.1.

Troldborg, L., J. C. Refsgaard, K. H. Jensen, and P. Engesgaard, 2007, The importance of alternative conceptual models for simulation of concentrations in a multi-aquifer system, *Hydrogeol. J.*, 15(5): 843-860, doi: 10.1007/s10040-007-0192-y.

Troldborg, M., 2010, Risk assessment and uncertainty estimation of groundwater contamination from point sources, Ph. D. thesis, DTU Environment, Denmark.

Troldborg, M., W. Nowak, N. Tuxen, P. L. Bjerg, R. Helmig, and P. J. Binning, 2010, Uncertainty evaluation of mass discharge estimates from a contaminated site using a fully Bayesian framework, *Water Resour. Res.*, 46(12): W12552, doi: 10.1029/2010WR009227.

UK Environment Agency, 2004, Model Procedures for the Management of Land Contamination, Publication: Contaminated Land Report 11, UK Environment Agency, UK, <https://www.gov.uk/government/publications>.

US EPA, 1992, Estimating the potential for occurrence of DNAPL at Superfund sites, Publication: 9355.4-07FS, US EPA, USA <http://www.epa.gov/superfund/health/conmedia/gwdocs/>.

US EPA, 1994, DNAPL site characterization, Publication: 9355.4-16 FS, US EPA, USA <http://www.epa.gov/superfund/health/conmedia/gwdocs/>.

US EPA, 1996, Soil Screening Guidance: User's Guide, Publication: 9355.4-23, US EPA, USA <http://www.epa.gov/superfund/health/conmedia/gwdocs/>.

US EPA, 2002, Supplemental Guidance for Developing Soil Screening Levels for Superfund Sites, Publication: OSWER 9355.4-24, US EPA, USA <http://www.epa.gov/superfund/health/conmedia/gwdocs/>.

Walker, W. E., P. Harremoes, J. Rotmans, J. P. van der Sluijs, M. B. A. van Asselt, P. Janssen, and M. P. Kraye von Krauss, 2003, Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support, *Integrated Assessment*, 4(1): 5-17, doi: 10.1076/iaij.4.1.5.16466.

Wang, Q. J., D. E. Robertson, and C. L. Haines, 2009, A Bayesian network approach to knowledge integration and representation of farm irrigation: 1. Model development. *Water Resour. Res.*, 45(2), W02409, doi:10.1029/2006WR005419.

Ye, M., S. P. Neuman, P. D. Meyer, and K. Pohlmann, 2005, Sensitivity analysis and assessment of prior model probabilities in MLBMA with application to unsaturated fractured tuff, *Water Resour. Res.*, 41(12), doi: 10.1029/2005WR004260.

Ye, M., K. F. Pohlmann, and J. B. Chapman, 2008, Expert elicitation of recharge model probabilities for the Death Valley regional flow system, *J. Hydrol.*, 354(1): 102-115, doi:10.1016/j.jhydrol.2008.03.001.

Ye, M., K. F. Pohlmann, J. B. Chapman, G. M. Pohll, and D. M. Reeves, 2010, A Model-Averaging Method for Assessing Groundwater Conceptual Model Uncertainty, *Ground Water*, 48(5): 716-728, doi: 10.1111/j.1745-6584.2009.00633.x.