Journal of Contaminant Hydrology 188 (2016) 12–28

Contents lists available at ScienceDirect

Journal of Contaminant Hydrology

journal homepage: www.elsevier.com/locate/jconhyd

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

Nanna I. Thomsen [a], Philip J. Binning [a,*], Ursula S. McKnight [a], Nina Tuxen [c], Poul L. Bjerg [a], Mads Troldborg [b]

[a] Department of Environmental Engineering, Miljoevej, Building 113, 2800 Kongens Lyngby, Denmark
[b] The James Hutton Institute, Craigiebuckler, Aberdeen, AB15 8QH, United Kingdom
[c] Capital Region of Denmark, Regionsgården - Kongens Vænge 2, - 3400 Hillerød, Denmark

ARTICLE INFO

Article history:
Received 5 September 2015
Received in revised form 10 February 2016
Accepted 14 February 2016
Available online 19 February 2016

Keywords:
Bayesian belief networks
Contaminated sites
Conceptual models
Conceptual uncertainty
Screening level

ABSTRACT

A key component in risk assessment of contaminated sites is in the formulation of a conceptual site model (CSM). A CSM is a simplified representation of reality and forms the basis for the mathematical modeling 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.

1. Introduction

The conceptual site model (CSM) is essential to many aspects of contaminated site management including: risk assessment (Suter, 1999; Troldborg, 2010); as the basis for groundwater models (Neuman and Wierenga, 2003); in the modeling 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,

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.

* Corresponding author at: Department of Environmental Engineering Environment, Miljoevej, Building 113, 2800 Kongens Lyngby, Denmark.
E-mail address: pjbi@env.dtu.dk (P.J. Binning).

http://dx.doi.org/10.1016/j.jconhyd.2016.02.003
0169-7722/

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), whereas 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.

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 (BBNs) (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