REAL TIME ASSESSMENT OF DRINKING WATER SYSTEMS USING A DYNAMIC BAYESIAN NETWORK

W. J. Dawsey[1], B. S. Minsker[2] and E. Amir[3]

[1]Graduate Research Asst., Dept of Civil & Environmental Engineering, University of Illinois-Urbana, 205. N. Mathews Ave, Urbana, IL 61801; PH (217) 333-6979; *dawsey@uiuc.edu*

[2] Professor, Dept of Civil & Environmental Engineering, University of Illinois-Urbana, 205. N. Mathews Ave, Urbana, IL 61801; PH (217) 333-9017; *minsker@uiuc.edu*

[3]Assistant Professor, Dept of Computer Science, University of Illinois-Urbana, 201 N Goodwin Ave, Urbana, IL 61801; PH (217) 333-8756; *eyal@uiuc.edu*

Abstract

This paper presents a methodology for real-time estimation of water distribution system state parameters using a dynamic Bayesian network to combine current observations with knowledge of past system behavior. The dynamic Bayesian network presented here allows the flexibility to model both discrete and continuous variables and represent causal relationships that exist within the distribution system. The posterior belief state can be inferred using a compact approximation algorithm that has been shown to contain inference errors. Simulations over stochastic variables are proposed to define the transition and observation models for the dynamic Bayesian network.

Introduction

There is significant uncertainty involved in real time monitoring of water distribution systems (Xu and Goulter, 1998; Bargiela and Hainsworth, 1989). This uncertainty generally falls into two categories: i) internal uncertainty due to a sensor technology's relative ability to quantify a target parameter and ii) external uncertainty due to the hydraulic and chemical processes that may disguise the presence of a contaminant. Current sensor technology relies on surrogate measures such as total organic carbon, oxidation reduction potential,, turbidity, and others to indicate the presence of a target contaminant using classification algorithms such as artificial neural networks (ASCE, 2004; National Research Council, 2006). A sensor's false positive rate is reduced with time as the algorithm is trained following installation. However, the probability of a false positive would likely always exceed the probability of a true contamination event for contaminants that occur very infrequently.

The second category of uncertainty in online monitoring refers to those processes that occur outside of a sensor that obscure its ability to indicate or measure a contaminant. For example, suppose that a mass of contaminant is injected into a distribution system pipe. As the contaminant is transported downstream, its concentration profile is degraded by diffusion, reactions with pipe walls, reactions with other constituents in the water, and mixing with uncontaminated water at junctions or tanks. These factors combine to increase the uncertainty of downstream sensor measurements. In addition, uncertain knowledge of real-time hydraulics cause uncertainty in relating a sensor's signal with the state of upstream nodes.

Monitoring networks in water infrastructure are analogous to the array of sensors used by autonomous agents in the field of robotics to detect obstacles, self-locate, and infer other knowledge about the surrounding environment. Direct observation of an obstacle's size and position may not be possible, thus making it necessary that a probability distribution for those characteristics be generated from indirect sensor data. Dynamic Bayesian networks have been used extensively in Artificial Intelligence to model probabilistic causal relationships among many interacting variables. This paper presents an analogous approach for drinking water monitoring, whereby a dynamic Bayesian network is used to infer knowledge about the current state of a water distribution system in real time.

Bayesian networks have found widespread application in the environmental domain for such applications as groundwater remediation (Stiber et al., 1999, 2004a, 2004b), ecological planning and modeling (Marcot, 2001), and prediction of estuarine water quality (Stow et al., 2003). However, there have been few applications of dynamic Bayesian networks in the Environmental Science or drinking water research literature. Dynamic Bayesian networks differ from static networks in that they describe systems that change with time. This aspect is necessary for a monitoring network to be adaptive to changing operational conditions that may occur in a water distribution system. In environmental research, K. Shihab (2005) modeled groundwater quality with time using a dynamic Bayesian network, and Murphy and Weiss (2001) demonstrated a novel algorithm for network inference for a wastewater treatment plant application.

Methodology

Bayesian networks are directed acyclic graphs that represent the conditional independence relationships for a joint probability distribution over a set of variables. Nodes represent variables and arcs represent both causal relationships and conditional independence relationships. A node is independent of other non-descendant nodes, given the values of its parent nodes. In Figure 1, **d** is independent of **c**, given **a**.

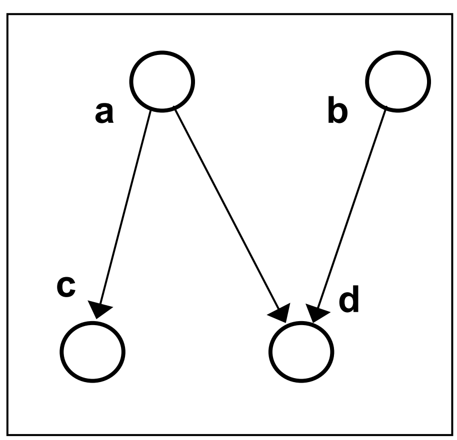

Figure 1. Bayesian network

The full joint distribution for a Bayesian network is given by the product:

$$P(X_1,...,X_n) = \prod_{i=1}^{n} P(X_i \mid Parents(X_i))$$

where $X_1,\ldots,X_n$ is the set of n variables that comprise the network. Several exact and approximate algorithms exist to infer the posterior distribution that reflects one or more observed variables (Russell and Norvig, 2003).

Dynamic Bayesian networks extend this concept for models that change with time. The dynamic Bayesian network shown in Figure 2 is updated with time steps, **t.** The definition of the conditional independence relationships is the same for both dynamic and static networks. For example, $\mathbf{a_2}$ is independent of $\mathbf{b_2}$ given $\mathbf{a_1}$ and $\mathbf{b_1}$. The conditional probability relationships between time steps, such as $P(\mathbf{a_2}|\mathbf{a_1},\mathbf{b_1})$, are referred to as the transition model. The probability relationships between observable and non-observable variables are referred to as the observation model. In Figure 2, **a** and **b** are non-observable, and **c** and **d** are observable. The task of a dynamic Bayesian network is to make inferences about the posterior distribution, or belief state, of non-observable variables from previous time steps and from current observations.

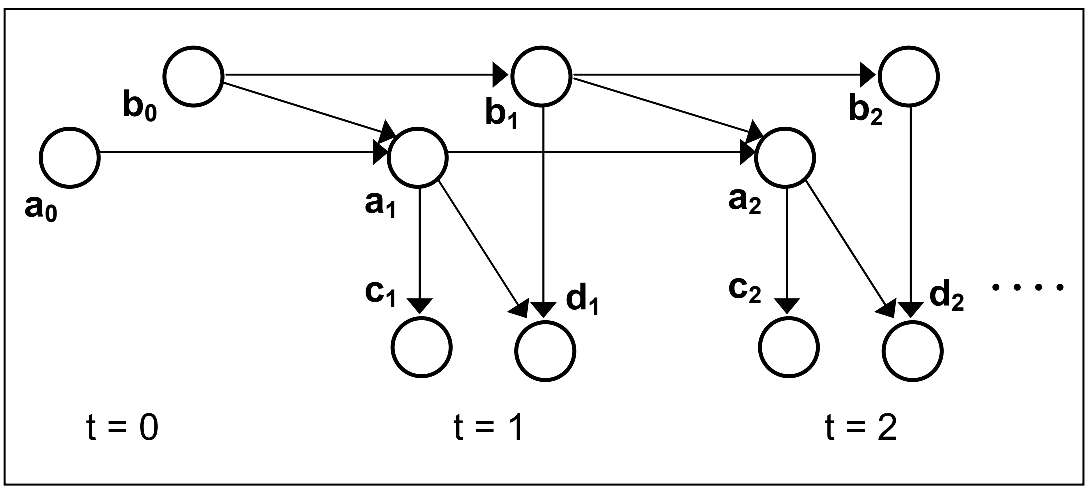

Figure 2. Dynamic Bayesian network

The formulation of a water distribution system model as a dynamic Bayesian network is relatively straightforward. Variables are defined that describe the state of the distribution system. These may be either continuous or discrete, observable or non-observable. Example variables may include the hydraulic head at a node, a pump operation status, the presence of a biological contaminant at a monitoring node, the status of an intermittent high demand node, or the concentration of a chemical contaminant at a node. Figure 3 shows an example dynamic Bayesian network with some possible observable and non-observable variables. For clarity, implied individual arcs from non-observable state variables between time periods are represented with a single bold arrow.

The conditional probability relationships among these variables can be determined using simulations of a network solver such as EPANET. A simple solution would be to assume a Gaussian distribution for input parameters such as consumer demand or pipe roughness and perform simulations using a Monte Carlo or Latin

Hypercube approach. A transition model of $P(X_{1,t+1}|X_{1,t},\ldots,X_{i,t})$ will result. A more representative input distribution could be used if such data are available. The time cost of such a computation may be significant, however, it would likely be less than a typical EPANET extended period simulation. The observation model is determined by the probability distribution of an observation given a value of its parent node. This probability would be largely determined by the internal error rates of the monitoring technology when a sensor's parent node is the parameter state at the same junction.

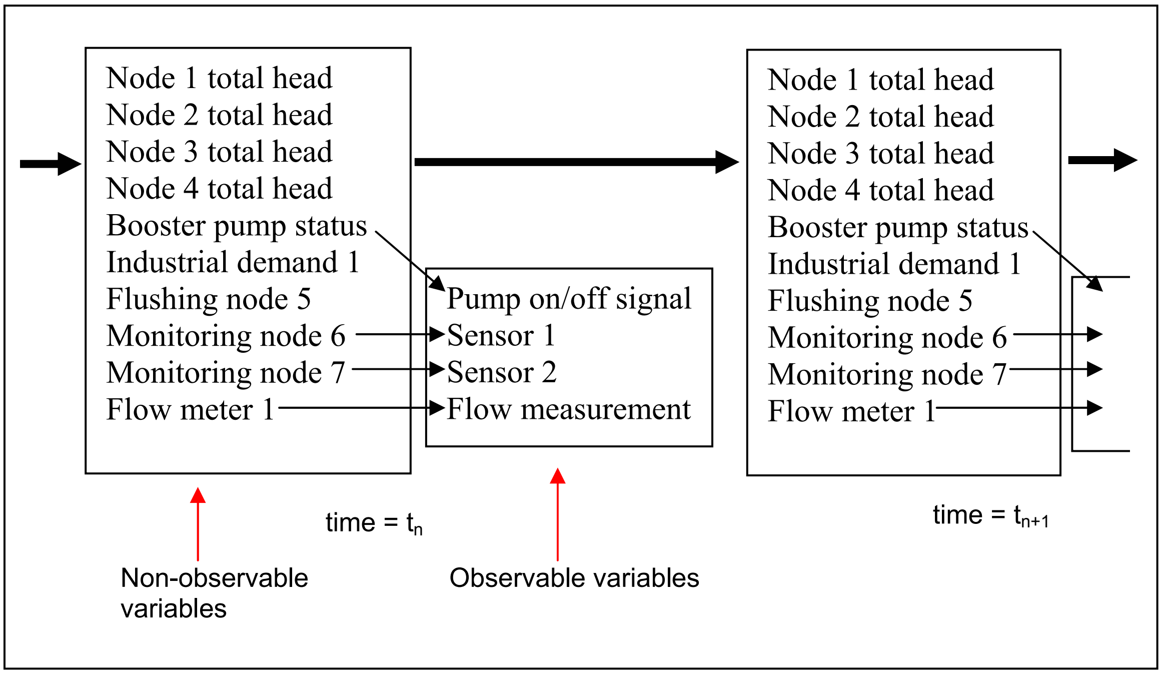

Figure 3. Example dynamic Bayesian network for water distribution system

Ideally, once the transition and observation models are defined for a given time step, a probability distribution describing the current state of non-observable variables can be inferred from current observation data and knowledge of previous states. Unfortunately, an exact solution for this inference problem quickly becomes infeasible as time progresses and the Bayesian network expands without bound (Russell and Norvig, 2003). A solution to this well-known problem is presented by Boyen and Koller (1998). These authors utilize a compact approximation algorithm to solve the inference problem and show that errors contract exponentially with time. The belief state is conditioned on observations at each time step and then represented in a compact form prior to the subsequent step. Thus, the belief state representation remains small. Further details of this algorithm are given by Boyen and Koller (1998).

Discussion

Dynamic Bayesian networks provide a great deal of flexibility in modeling the belief state of discrete and continuous variables over time. This approach is robust to problems associated with incomplete and noisy data. The compact approximation inference algorithm reduces the computing cost of networks containing large numbers of state variables, although the computing cost of simulating a water distribution system's behavior over many realizations of stochastic variables may still be an issue.

The internal structure of the dynamic Bayesian network must be fitted to a specific water distribution system. Combinations of infrastructure components would be unique to a given water system, and network topology would define many of the causal relationships among variables.

This approach has the potential to improve online contaminant monitoring by conditioning local sensor observations on knowledge of past system behavior. This conditioning occurs in the context of the larger distribution system, which is a departure from surrogate sensors that learn only at the local level without awareness of other variables.

References

American Society of Civil Engineers (ASCE), Interim Voluntary Guidelines for Designing an Online Contaminant Monitoring System, Reston VA, 2004.

Bargiela, A., and G. D. Hainsworth, Pressure and flow uncertainty in water systems, J. Water Resour. Plann. Manage., 115(2), 212–229, 1989.

Boyen, X., and Koller, D., Tractable Inference for Complex Stochastic Processes, Uncertainty in Artificial Intelligence (UAI 1998), 33-42, 1998.

Marcot, B. G., Using Bayesian belief networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement, Forest Ecology and Management, v 153 (1-3), n 1-3, pp. 29-42, 2001.

Murphy, K. P., and Weiss, Y., The Factored Frontier Algorithm for Approximate Inference in DBNs, Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence, 378 – 385, 2001.

National Research Council (NRC), Drinking Water Distribution Systems: Assessing and Reducing Risks, Washington D.C., The National Academies Press, 2006.

Russell, S., and Norvig, P., *Artificial Intelligence: A Modern Approach*, Prentice Hall, New Jersey, 2003.

Shihab, K., Modeling groundwater quality with Bayesian techniques, Proceedings. 5th International Conference on Intelligent Systems Design and Applications, 73-8, 2005.

Stiber, N. A., M. Pantazidou, and M. J. Small, Expert system methodology for evaluating reductive dechlorination at TCE sites, Environmental Science and Technology, v 33(17), pp. 3012-3020, 1999.

Stiber, N.A., M. Pantazidou and M.J. Small, Embedding expert knowledge in a decision model: Evaluating natural attenuation at TCE sites. Journal of Hazardous Materials, v 110(1-3), pp. 151-160, 2004.

Stiber, N.A., M.J. Small and M. Pantazidou. 2004. Site-specific updating and aggregation of Bayesian Belief Network models for multiple experts. Risk Analysis, v 24(6), pp. 1529-1538, 2004.

Stow, C. A., C. Roessler, M. E. Borsuk, J. D. Bowen, and K. H. Reckhow, Comparison of Estuarine water quality models for total maximum daily load development in Neuse River Estuary, Journal of Water Resources Planning and Management, v 129 (4), pp. 307-314, 2003.

Xu, C., and I. C. Goulter, Probabilistic model for water distribution reliability, J. Hydraul. Eng., 124(4), 218–228, 1998.