

Bayesian approach to contaminant source characterization in water distribution systems: adaptive sampling framework

Hui Wang · Kenneth W. Harrison

Published online: 28 April 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract Bayesian analysis can yield a probabilistic contaminant source characterization conditioned on available sensor data and accounting for system stochastic processes. This paper is based on a previously proposed Markov chain Monte Carlo (MCMC) approach tailored for water distribution systems and incorporating stochastic water demands. The observations can include those from fixed sensors and, the focus of this paper, mobile sensors. Decision makers, such as utility managers, need not wait until new observations are available from an existing sparse network of fixed sensors. This paper addresses a key research question: where is the best location in the network to gather additional measurements so as to maximize the reduction in the source uncertainty? Although this has been done in groundwater management, it has not been well addressed in water distribution networks. In this study, an adaptive framework is proposed to guide the strategic placement of mobile sensors to complement the fixed sensor network. MCMC is the core component of the proposed adaptive framework, while several other pieces are indispensable: Bayesian preposterior analysis, value of information criterion and the search strategy for identifying an optimal location. Such a framework is demonstrated

with an illustrative example, where four candidate sampling locations in the small water distribution network are investigated. Use of different value-of-information criteria reveals that while each may lead to different outcomes, they share some common characteristics. The results demonstrate the potential of Bayesian analysis and the MCMC method for contaminant event management.

Keywords Uncertainty · Inverse modeling · Source identification · Adaptive sampling · Value of information · Data worth framework · Bayesian preposterior analysis

1 Introduction

Contaminant source identification in a water supply network generally has been treated as an inverse problem, has involved deterministic modeling, and therefore has aimed to identify the “best” prediction of the source (e.g., van Bloemen Waanders et al. 2003; Laird et al. 2006; Preis and Ostfeld 2006; Guan et al. 2006). The nonuniqueness characteristic of the inverse problem limits the application of such approaches. Some have adopted a probabilistic approach (e.g., Propato et al. 2010; Poulakis et al. 2003; De Sanctis et al. 2008), but with the stochastic modeling limited to sensor measurement error. Dawsey et al. (2006) use Bayesian belief networks to identify the contamination node but they do not characterize its magnitude or duration. A Bayesian approach can infer the contaminant history with limited information, which has been applied to similar fields (e.g., Liu et al. 2010; Wang and Jin 2012). With few observations, the initial inference about the contamination source may not be sufficiently strong evidence for the decision-maker. Utility managers likely would prefer stronger inference to enact countermeasures such as cutting

H. Wang (✉)
Bureau of Economic Geology, The University of Texas at
Austin, Austin, TX 78758, USA
e-mail: hui.wang@beg.utexas.edu

K. W. Harrison
Earth System Science Interdisciplinary Center, College Park,
MD 20740-3823, USA

K. W. Harrison
Hydrological Sciences Branch, NASA Goddard Space Flight
Center, Greenbelt, MD 20771, USA

linkages to possible contaminant nodes, or informing residents in possibly affected areas. Utility managers have two options: wait until more observations are available at which point a more complete picture of the source hopefully will emerge or seek out additional measurements by placing mobile sensors at strategically identified locations. The latter could help to reduce potential risk of the contaminant event including the volume of contaminated water and the population size affected. Here, we address this by asking the following research question: where should a mobile sensor be placed soon after detection so as to most strengthen the inference?

This is essentially a question posed in the context of an adaptive strategy. Methods for optimal multi-stage decision-making with uncertainty updating have been studied in the field of water resources and environmental engineering (see Harrison 2002 for discussion). At every decision stage, the uncertainty estimates of unobservable parameters (e.g., contaminant source location) are revised with available data via Bayes' theorem. The full optimal solution is generally intractable, requiring considerable simplification (e.g., limiting the number of decision variables and decision stages; see Chao and Hobbs 1997; Venkatesh and Hobbs 1999; Harrison 2007a, b).

Instead, a value-of-information (VOI) decision rule approach often is adopted for multi-stage decision-making with uncertainty updating. This strategy, which may also be referred to as a data worth or optimal stopping (of sampling) rule, is a step-wise approach. At each step, the maximally beneficial next sample is sought. In the groundwater and water quality field, this framework has been explored with some sophistication (e.g., Freeze et al. 1992; James and Freeze 1993; James and Gorelick 1994; Dakins et al. 1996; Serre et al. 2003). For example, James and Gorelick (1994) investigated a groundwater remediation problem with uncertainty in the hydraulic conductivity field, in the location of the source of the contamination, and in the time that the source has been active. The worth of an additional measurement was calculated as the expected reduction in remediation costs. Such a worth measure is referred to as the expected value of sample information (EVSI). Here, we develop measures for uncertainty reduction that proxy for EVSI.

For computational efficiency, we apply the Markov chain Monte Carlo (MCMC) method. The sampling approach of the previously cited studies scales poorly in the number of uncertain parameters as it relies on a fixed sample with adjustments to the probabilities of each as observations accumulate. MCMC sampling, in contrast, samples in proportion to probability density. The MCMC implementation described in Harrison and Wang (2009) is applied as it addresses particular challenges in applying MCMC to water distribution networks that relate to the discrete nature of water distribution systems.

Observations available from a sparse network of fixed sensors alone are likely insufficient to achieve strong inference as to the source of contamination in a timely manner (Wang and Harrison 2013). The stochasticity in the water distribution network (e.g., random water demands) and modeling errors have the effect of greatly inhibiting uncertainty resolution such that it may not even be possible to identify a localized part of the network containing the source. Recognition of the limitations of a fixed sensor network is the motivation of this research into the strategic placement of mobile sensors at locations of maximum worth. This paper tries to answer the following questions: what will be the possible inferences made if a mobile sensor is installed at a particular location? Of a set of potential locations, which is the best location to next gather a sample?

The organization of the paper is as follows. In Sect. 2, the MCMC implementation of the adaptive framework is developed. In Sect. 3, an illustrative example with stochastic water demands is described and results are presented. In Sect. 4, the results are discussed in more detail. Finally, conclusions are offered (Sect. 5).

2 Methods

The Bayesian approach to adaptive sampling involves: (1) developing an initial uncertainty assessment, (2) updating uncertainty in the contaminant source characterization given available sensor measurements, (3) assessing the reduction in uncertainty for each possible simulated outcome from a potential new sample, (4) predicting the probability of each outcome, (5) evaluating a “worth” measure that combines the results of 3 and 4, and finally (6) searching for the location of maximum worth. This is an iterative process. Steps 2–6 are repeated until there is sufficient confidence in the identification of the source. In this section, these components are discussed more fully. In the discussion, for convenience we assume the sensor measurements to be categorical as opposed to continuous so that we can refer to probability and not probability density.

The conceptual framework is provided in Fig. 1. The corresponding section of the paper is indicated in the text box for each step. With regard to the first step, it is important to notice that available sensor measurements will accumulate in time, initially consisting of the fixed sensor measurements at the time of detection but later consisting of both fixed and mobile sensor measurements. In the second step, the probability of each possible measurement at each candidate sampling location is calculated. This requires runs of forward Monte Carlo simulations incorporating system uncertainties based on the existing inference of the contamination history. A “preposterior” analysis step (Step

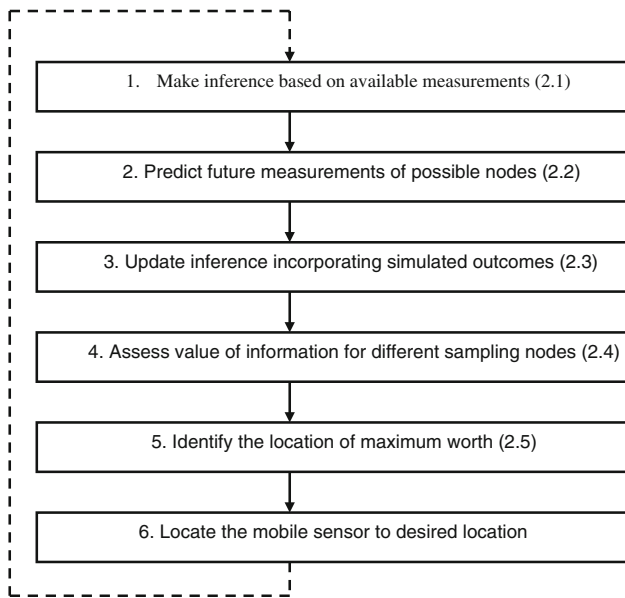


Fig. 1 Conceptual adaptive framework

3), utilizing existing data, is conducted for each and every candidate node and possible measurement. Each preposterior analysis requires a run of the MCMC algorithm and demands extensive computation. While the latest updated inference is likely stronger than the previous/initial inference, in this framework it is necessary to evaluate the degree of uncertainty reduction to be expected from new sampling. Two criteria are offered and utilized to differentiate candidate nodes (Step 4). Once the most preferred candidate node is identified (Step 5), a mobile sensor would be located to gather real data at a later time step (Step 6). The dashed arrow between the last step and the first step indicates that once new data is obtained from the mobile sensor, the process begins anew. The loop continues until the decision-maker is sufficiently confident as to the source characterization and is ready to take further action.

2.1 Updating uncertainty with sensor measurements

Let θ represent a vector of uncertain, unobserved parameters. For example, θ may refer to the contaminant source location, start time, and injection start time. The initial uncertainty in the source characterization is described with the assignment of the prior probability, $p(\theta)$. The prior probability may incorporate expert knowledge regarding, for example, aging parts of the water distribution network that are vulnerable to a contamination incident. The problem of Bayesian analysis is to incorporate data y so that $p(\theta)$ can be revised to posterior probability, $p(\theta|y)$. In the source characterization problem, the data would originate from fixed or mobile sensors in the network. The updating

of $p(\theta)$ to $p(\theta|y)$ is governed by Bayes’ theorem, a rule of probability:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int_{\theta} p(y|\theta)p(\theta)d\theta} \tag{1}$$

A standard numerical evaluation of Eq. (1) is not computationally tractable if there are many components of θ (Gilks et al. 1996). Instead, a Markov chain Monte Carlo (MCMC) approach is typically required. MCMC for Bayesian analysis involves forming a carefully constructed chain of θ . For an infinitely long chain, the frequency with which a particular state θ is visited is its posterior probability (Gelman et al. 1995). Even with MCMC, evaluation of Eq. (1) is computationally expensive.

An MCMC approach was tailored for water distribution systems (Wang and Harrison 2013). The likelihood term $p(y|\theta)$, the most computationally expensive part of an MCMC iteration, is evaluated with a forward Monte Carlo simulation: for a given θ , the water distribution model is run N times, once for each of N realizations of the likelihood terms (e.g., stochastic water demands); the likelihood of the observed y is estimated by observing the frequency of its observation. For computational efficiency, the N realizations are farmed to different processors, with control of the MCMC algorithm by one processor. This MCMC procedure is called repeatedly within the adaptive framework presented here (Steps 1 and 3 of Fig. 1).

2.2 Predictive probabilities

Next, we need to be able to evaluate the probability of each potential outcome \tilde{y} . This is given by the posterior predictive probability $p(\tilde{y}|y)$, which is defined as: (Gelman et al. 1995)

$$p(\tilde{y}|y) = \int_{\theta} p(\tilde{y}|y, \theta)p(\theta|y)d\theta. \tag{2}$$

The term $p(\tilde{y}|y)$ is the probability of potential outcome \tilde{y} given previous observations y ; $p(\tilde{y}|y, \theta)$ is the probability of potential outcome \tilde{y} given the joint of previous observations y and a certain contamination scenario θ . The right hand side of Eq. (2) integrates out the uncertainty in the contamination inference. We applied Monte Carlo to sample from the posterior predictive distribution as follows:

- (a) the MCMC chain is randomly sampled to obtain samples θ_k from $p(\theta|y)$
- (b) for each θ_k , the Monte Carlo likelihood simulation model is run; if y is reproduced, then the sample is retained; otherwise it is rejected.
- (c) $p(\tilde{y}|y)$ is estimated from the resulting sample

2.3 Bayesian preposterior analysis

If sampling a new node i , the outcome \tilde{y}_i is subject to uncertainty as described in the previous Sect. 2.2. For a speculative value of \tilde{y}_i we can evaluate how the uncertainty will change with its observation. This is a key aspect of Bayesian preposterior analysis (Berger 1993). In other words, we are interested in evaluating:

$$p(\theta|y, \tilde{y}_i) = \frac{p(y, \tilde{y}_i|\theta)p(\theta)}{\int_{\theta} p(y, \tilde{y}_i|\theta)p(\theta)d\theta} \quad (3)$$

This equation is different from Eq. (1) only in that the set of observations now includes the speculative observation. The same MCMC procedure is used as initially except now including the speculative observation. It must be performed for each potential outcome at every candidate location. Fortunately, however, this work can be done in parallel.

2.4 Development of an uncertainty reduction measure conditioned on \tilde{y}

We next require a measure of the uncertainty reduction $W(\tilde{y}_i)$ that comes with the observation of \tilde{y}_i . This is needed input to distinguish the worth of the candidate sampling locations. A more formal decision analytic loss measure could be formulated, incorporating actual costs and building in risk aversion. Here, we offer two simpler measures for uncertainty reduction that operate off of the marginal posterior probability of the nodes, $P(x_i|y, \tilde{y}_i)$. The first is the maximum nodal probability across the network:

$$W_{\max}(\tilde{y}_i) = \max_i P(x_i|y, \tilde{y}_i)$$

The measure of maximum nodal probability simply suggests, for the most probable node of contamination, the confidence that it is the true source location. However, it ignores the whole distribution of nodal probabilities. There may be more worth in having, for example, five nodes of moderate probability, with all other nodes having zero probability, than one highly probable node with many nodes of low probability.

In contrast, the second measure, based on information entropy, incorporates the whole distribution of nodal probabilities. Information entropy is defined as:

$$\text{entropy} = - \sum_i P(x_i|y, \tilde{y}_i) \ln P(x_i|y, \tilde{y}_i)$$

The entropy of random variables was first developed by Shannon (1948) and is an indispensable concept in information theory and decision science. In the context of contaminant inference, it simply provides an approach to quantify the degree of uncertainty in the inference. Newly observed data should in general reduce the inference

uncertainty to some degree, depending on the value of the measurement:

$$W_{\text{ent}}(\tilde{y}_i) = \text{entropy}$$

These two criteria are offered for their easy understandability and simple calculation. No exhaustive list of evaluation criteria is offered in this paper mainly because it is somewhat subjective. The decision-maker can choose other criteria for differentiating the worth of inferences while keeping within this overall framework, which is further discussed further in Sect. 4.

2.5 Expected worth of a new sample

In developing the aggregate worth measure, we evaluate the expectation of the uncertainty reduction measure $W(\tilde{y}_i)$ using the predictive probability:

$$E[W(\tilde{y}_i)] = \int_{\tilde{y}_i} W(\tilde{y}_i)p(\tilde{y}_i|y)d\tilde{y}_i \quad (4)$$

2.6 Optimization problem

Next, a search strategy is needed to identify the location of maximum worth. The optimization problem is:

$$\max_i E[W(\tilde{y}_i)] \quad (5)$$

The search strategy could involve an exhaustive search through all nodes or a more formal search method. It should be noted that the evaluation of expected worth of one sampling location is independent of and therefore can be done in parallel with another location, a property that can be exploited by the search algorithm. In this study, we consider a finite number of potential sampling locations to demonstrate the methodology.

3 Results

Example Network 3 in EPANET 2.0 is used as the illustration in this study, as it has been studied extensively (e.g., Preis and Ostfeld 2006), including Wang and Harrison (2013). The physical network, depicted in Fig. 2, is left unchanged, with 97 nodes, two of which are water supply sources and three of which are tanks.

The scenario is as following. The total simulation period for the illustrative example is 28 h (168 time steps). The hydraulic simulation step is 10 min; the water quality time step is 10 min. Categorical measurements are taken at the sensor location shown in Fig. 2. U, L and H, represent three categories: Undetected, Low and High. The threshold between U and L is 0.05 mg/L and it is 0.1 mg/L between L and H. The prior on the parameters of the contaminant

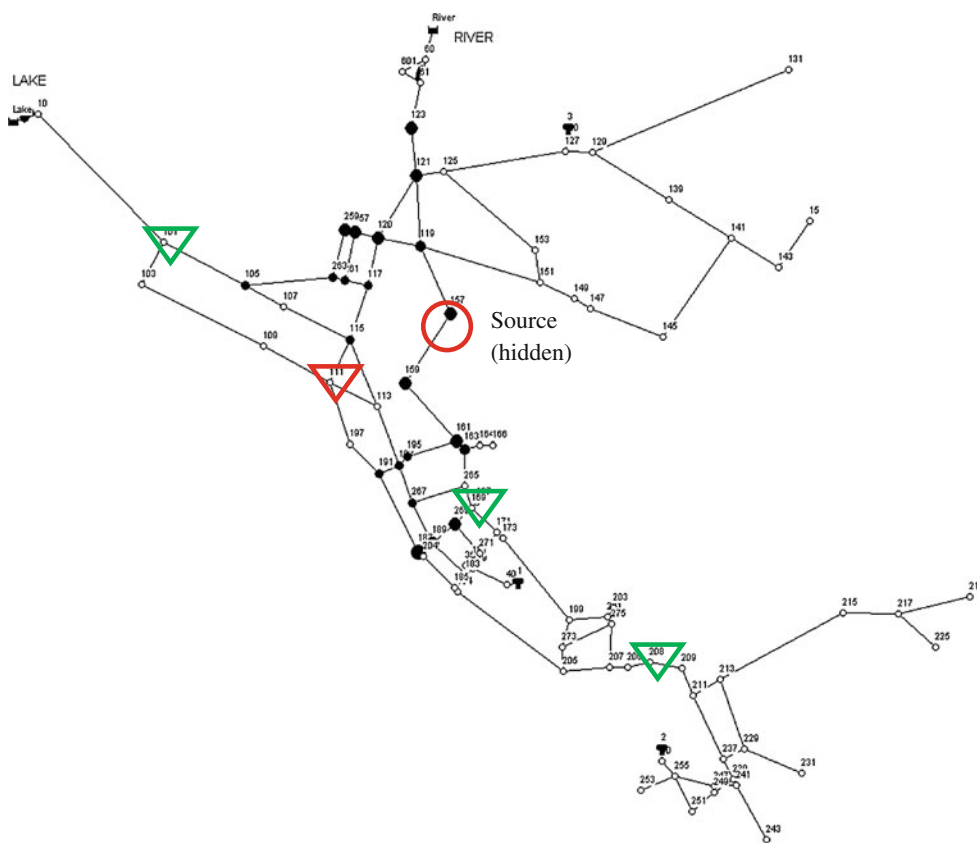


Fig. 2 Example EPANET 2.0 Network 3. The inference after initial detection is shown with the probability of each node being the source proportional to the area of the node (*circles*); nodes with estimated

zero probability are shown as open circles. Also, the candidate sensor locations (*triangles*) are shown. The fixed sensor node is node 111 (*red triangle*). (Color figure online)

event, including source node (X), magnitude (M) and start time (T) is shown in Table 1. The proposal function for each parameter is adjusted in three phases of MCMC simulation as described in detail in Harrison and Wang (2009). The only stochastic source of error considered is the stochastically varying set of water demands, described by a simple stochastic model: $D_i(t)$, the demand at node i and time t , normally and independently distributed with mean μ_i with coefficient of variation, COV , set equal to 0.40. In the study scenario, node 157 is the hidden contaminant source with magnitude 1.0 kg/min. One fixed sensor is deployed at node 111 as shown in Fig. 2. For demonstration purposes, we consider the four candidate sampling locations shown. They are chosen to represent different areas of the water distribution network.

3.1 Initial inference after contaminant detection

The uncertainty in the source location after initial detection (at the L level) is shown in Table 2. The reported probabilities are the averages using five random seeds. Only the nodes with nonzero estimated probability are shown. Prior to detection, the probability of each node being the contaminant

source is 1.03 % (=1/97). After initial detection, it is inferred that one node has probability greater than 10 % (node 187 with 12.1 %), ten nodes are between 5 and 10 % probability, two between 1 and 5 % probability, and nine with positive probability less than 1 %; the remaining 75 nodes (not shown) are estimated to have near 0 % probability. The same information is shown graphically in Fig. 2. A swath of posterior probability extends across the network.

3.2 Preposterior analysis

The predictive probability (Step 2 of Fig. 1) for a time 1 h after initial detection ($t = 1$ h) is shown in column 1 of Table 3 for each potential outcome and at each candidate node. The predictive probabilities vary considerably among the candidate nodes.

The worth measures are then computed (Step 3 of Fig. 1) based on the inferences made for each potential outcome (columns 2 and 3 of Table 3). As a reference point, the maximum probability after the initial inference update was 12.1 %; the information entropy measure was 2.56. For all nodes and outcomes, there is a reduction in uncertainty from acquisition of the additional measurement.

Table 1 Prior distributions assigned to the unobservable parameters θ

	Component of θ		
	Source node (X)	Source magnitude (M)	Source start time (T)
Distribution type	Discrete uniform	Uniform	Discrete uniform
Distribution parameters	Probability = 1.03 %	Low = 0.20 kg/min high = 1.8 kg/min	Low = 18.5 h prior to detection high = time of detection

Table 2 Inference after initial detection

Rank after detection	X	P(X) (%)	Rank after detection	X	P(X) (%)
1	187	12.1	12	119	4.6
2	259	9.6	13	163	2.1
3	257	9.6	14	263	0.58
4	269	9.3	15	261	0.29
5	159	9.0	16	117	0.19
6	189	8.5	17	195	0.06
7	120	8.0	18	193	0.06
8	121	7.8	19	191	0.044
9	123	6.5	20	267	0.036
10	161	5.8	21	105	0.016
11	157	5.5	22	115	0.016

Nodes with zero estimated probability are not shown

The expected worth measures, the dot product of the predictive probabilities and worth measures, are also indicated in Table 3. Using both the maximum probability and entropy measures, node 101 appears to be the best next node to sample, with node 169 as the second best. Node 101 has the highest expected maximum probability, 25.8 %, and the lowest expected entropy, 1.78.

4 Discussion

The goal of this work was to develop an adaptive framework for contaminant source identification and to demonstrate its value for an illustrative network. After the triggering of a sensor, rather than waiting for measurements from the fixed sensor network, the question was whether a mobile sensor could be strategically placed at a location that will yield information that will more quickly home in on the true source node. The results offer support for an adaptive framework for sampling.

In previous work (Harrison and Wang 2009; Wang and Harrison 2013), we demonstrated the updating of a statistical inference given measurements from a fixed sensor (node 111). The inference was updated at the time of initial

detection and at a time 1 h later ($t = 1$). Here, we consider the option to place a mobile sensor at an alternate location in time for a measurement made at $t = 1$ h. Four locations across the network were considered. For comparison purposes, the original fixed sensor (111) was included as one of the four locations.

All three of the alternate locations were evaluated as being better than the fixed sensor, suggesting the value of the adaptive framework. Use of both expected worth measures indicated that node 101 was the best of the four. Whereas a sample made at the fixed sensor at $t = 1$ h has a maximum probability worth measure of 19.8 %, a sample at node 101 has a maximum probability worth measure of 25.8 %. That is, on an expected value basis, the maximum confidence in the source node is greater by about 6 % if sampling node 101 instead of the fixed sensor. Similarly, information entropy, a measure of the spread in uncertainty across the nodes, is better with sampling 101, 1.78 as compared to 2.21.

A possible explanation for the comparatively low worth of the measurement at the fixed sensor at $t = 1$ h is that the information content at that location was in some sense already exhausted. Much had been gained with the detection (and, importantly, preceding U observations) by the fixed sensor. Moving forward, to discern further between potential source nodes requires measurements in another part of the network.

For a fuller picture, it is necessary to account for the stochastic water demands. The stochastic variation inhibits the uncertainty resolution in the source node and source characteristics. Reversals in flow directions occur. The travel time to the sensor is affected. The contaminant can reach larger portions of the network. It is not straightforward to identify meaningful locations to place sensors absent the kind of analysis presented here and the consideration of stochastic water demands.

In the case under consideration, there is little consequence in the choice of expected worth measures. Both the maximum probability and information entropy measures point to node 101 as being the best node. Also, it may be noted that 101 has the best worst case outcome (maximum probability of 21.2 % and information entropy of 1.91). In general, for the best management of the water distribution

Table 3 Predictive probability (column 1) and the uncertainty reduction using two different measures for worth—maximum probability across nodes (column 2) and information entropy (column 3)

Candidate node		1 Predictive probability (%)	2 Maximum probability (%)	3 Information entropy (–)
111(fixed sensor)	Measurement @ t = 1 h			
	U	7.7	29.9	1.73
	L	22.0	22.1	1.96
	H	70.2	18.0	2.33
	Expected worth		19.8	2.21
101	Measurement @ t = 1 h			
	U	57.8	21.2	1.91
	L	6.2	88.7	0.46
	H	35.9	22.3	1.79
	Expected worth		25.8	1.78
169	Measurement @ t = 1 h			
	U	30.8	25.2	1.84
	L	0.3	44.3	1.05
	H	68.9	21.2	1.93
	Expected worth		22.5	1.90
208	Measurement @ t = 1 h			
	U	39.8	14.4	2.38
	L	7.8	20.5	2.15
	H	52.3	23.0	2.04
	Expected worth		19.4	2.19

Both expected worth measures point to node 101 as being the best among the candidates. For a point of reference, upon first detection, the maximum probability was 12.1 % and the entropy measure was 2.56. Three significant digits are shown for matching text references to the table but, given the sampling errors of MCMC, only two significant digits are warranted

system, the worth measure that is selected should be the one that most accurately captures the consequence of uncertainty reduction. Is it more important to have a great degree of confidence that the most probable source is indeed the source (the maximum probability measure) or to shrink the overall uncertainty in the source (information entropy)? Importantly, the framework can accommodate more meaningful measures, for example, reflecting remediation and exposure costs as well as attitudes towards risk (i.e., risk aversion). The use of such measures within a data worth framework has been demonstrated in the ground-water field (e.g., Freeze et al. 1992; James and Freeze 1993; James and Gorelick 1994).

As pointed out in Harrison and Wang (2009), the key limitation to an adaptive framework in an operational setting relates to the likelihood evaluation. The evaluation currently relies on Monte Carlo simulation with sampling of water demands. This evaluation is required at each MCMC iteration. Fortunately, the inference updating for each node and outcome can be done in parallel, and indeed the Monte Carlo simulation can be done in parallel, as was the case here. Nevertheless, a more efficient means of evaluating the likelihood would make the approach more practical for application in operational settings. It is more imperative when considering multiple sensors and longer observation time series as sample sizes need to be larger to

obtain sufficient sampling densities. General strategies do exist for improving the efficiencies but testing and further development are needed.

5 Summary and conclusion

An adaptive sampling framework was presented that facilitates inference of the contamination history in water distribution networks. Numerous previous works have focused on making inferences by incorporating sensor measurements collected during a lengthy simulation period. The philosophy behind the adaptive approach presented here is to infer the contaminant history as soon after detection as possible. This would aid utility managers charged with taking actions to limit exposure and re-establish service.

Previous work (Wang and Harrison 2013) described the updating of the source characterization inference using data from a fixed sensor. The adaptive framework here seeks to speed up the source identification. After the inference made at the time of contaminant detection by the fixed sensor, four candidate sampling locations, including the original sensor for comparison, were investigated to determine which would result in the greatest uncertainty reduction. The framework involves probabilistic predictions based on

the initial inference and stochastic water demand modeling, and requires worth measures for uncertainty reduction.

The results indicate the value of an adaptive framework over reliance on fixed sensors. A subsequent measurement at the original, fixed sensor made 1 h after detection yielded less uncertainty reduction than the three alternate locations. This was robust to the choice of worth measure. An adaptive framework for placing mobile sensors can avoid the costs of installing and maintaining a dense network of sensors throughout the water distribution system.

Acknowledgments This work was supported by National Science Foundation (NSF) under Award No. 0849064. Any opinions, findings and conclusion expressed in this materials are those of the authors and do not necessarily reflect the views of the NSF.

References

- Berger JO (1993) Statistical decision theory and Bayesian analysis. Springer, Berlin
- Chao PT, Hobbs BF (1997) Decision analysis of shoreline protection under climate change uncertainty. *Water Resour Res* 33(4): 817–829
- Dakins ME, Toll JE, Small MJ, Brand KP (1996) Risk-based environmental remediation: Bayesian Monte Carlo analysis and the expected value of sample information. *Risk Anal* 16(1):67–79
- Dawsey WJ, Minsker BS, VanBlaricum VL (2006) Bayesian belief networks to integrate monitoring evidence of water distribution system contamination. *J Water Resour Plan Manag* 132(4): 234–241
- De Sanctis AE, Boccelli DL, Shang F, Uber JG (2008) Probabilistic approach to characterize contamination sources with imperfect sensors. *Proceeding of the ASCE World and Environmental Congress*
- Freeze RA, James B, Massmann J, Sperling T, Smith L (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
- Gelman A, Carlin JB, Stern HS, Rubin DB (1995) Bayesian data analysis. Chapman & Hall, London
- Gilks WR, Richardson S, Spiegelhalter DJ (1996) Introducing Markov chain Monte Carlo. In: Gilks WR, Richardson S, Spiegelhalter DJ (eds) *Markov chain Monte Carlo in practice*. Chapman & Hall, London
- Guan J, Aral MM, Maslia ML, Grayman WM (2006) Identification of contaminant source in water distribution system using simulation-optimization method: case study. *J Water Resour Plan Manag* 132(4):252–262
- Harrison KW (2002) Environmental and water resources decision-making under uncertainty. *Doctoral Dissertation*. August, 2002
- Adviser: Ranji Ranjithan, Department of Civil Engineering, North Carolina State University
- Harrison KW (2007a) Two-stage decision-making under uncertainty and stochasticity: Bayesian programming. *Adv Water Resour* 30(3):641–664
- Harrison KW (2007b) Test application of Bayesian programming: adaptive water quality management under uncertainty. *Adv Water Resour* 30(3):606–622
- Harrison KW, Wang H (2009) A Markov chain Monte Carlo implementation of Bayesian contaminant source characterization in water distribution systems under stochastic demands. *Proceeding of the ASCE World and Environmental Congress*
- James BR, Freeze RA (1993) The worth of data in predicting aquifer continuity in hydrogeological design. *Water Resour Res* 29(7):2049–2065
- James B, Gorelick SM (1994) When enough is enough: the worth of monitoring data in aquifer remediation design. *Water Resour Res* 30(12):3499–3513
- Laird CD, Biegler LT, van Bloemen Waanders BG (2006) Mixed-integer approach for obtaining unique solutions in source inversion of water networks. *J Water Resour Plan Manag* 132(4):242–251
- Liu X, Cardiff MA, Kitanidis PK (2010) Parameter estimation in nonlinear environmental problems. *Stoch Environ Res Risk Assess* 24:1003–1022
- Poulakis Z, Valougeorgis D, Papadimitriou C (2003) Leakage detection in water pipe networks using a Bayesian probabilistic framework. *Probab Eng Mech* 18:315–327
- Preis A, Ostfeld A (2006) Contamination source identification in water systems: a hybrid model trees-linear programming scheme. *J Water Resour Plan Manag* 132(4):263–273
- Propato M, Sarrazay F, Tryby M (2010) Linear algebra and minimum relative entropy to investigate contamination events in drinking water systems. *J Water Resour Plan Manag* 136(4):483–492
- Serre ML, Christakos G, Li H, Miller CT (2003) A BME solution of the inverse problem. *Stoch Environ Res Risk Assess* 17(5–6): 354–369
- Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27:379–423
- Van Bolemen Waanders BG, Bartlett RA, Bigler LT, Laird CD (2003) Nonlinear programming strategies for source detection of municipal water networks. *Proceeding of the ASCE World and Environmental Congress*, Philadelphia, PA, June 23–26
- Venkatesh BN, Hobbs BF (1999) Analyzing investments for managing Lake Erie levels under climate change uncertainty. *Water Resour Res* 35(5):1671–1683
- Wang, H, Harrison K (2013) Bayesian update method for contaminant source characterization in water distribution systems. *J Water Resour Plan Manag* 139(1):13–22
- Wang H, Jin X (2012) Characterization of groundwater contaminant source using Bayesian method. *Stoch Environ Res Risk Assess*. doi:10.1007/s00477-012-0622-9