

Drinking Water Security and Public Health Disease Outbreak Surveillance

Steven M. Babin, Howard S. Burkom, Zaruhi R. Mnatsakanyan, Liane C. Ramac-Thomas, Michael W. Thompson, Richard A. Wojcik, Sheri Happel Lewis, and Cynthia Yund

The U.S. Environmental Protection Agency (EPA) strives to develop technologies and protocols to assist drinking water utilities in reducing the risks of potential terrorist attacks on our nation's infrastructure. Of particular interest are the development and feasibility testing of contamination warning systems that integrate existing public health surveillance and water quality measurements. In collaboration with the EPA, APL is developing a novel prototype warning system that employs sensor clustering and Bayesian Network analyses. These techniques address the challenges of synthesizing results from disparate data types with different data rates and complex environmental and operational responses into a warning system that can provide the user with a measure of the likelihood that data anomalies do or do not indicate a potential water-borne disease outbreak. These critical challenges and how this novel approach and its components address them will be described.

INTRODUCTION

The Need

Although relatively uncommon in the United States, the Centers for Disease Control and Prevention (CDC) report 13–14 drinking-water-related disease outbreaks per year, affecting an average of ~1000 people annually (see Table 1; note that it is based on biennial CDC reports, so annual averages were determined by taking the sum and dividing by 10 years).¹ Throughout his-

tory, there have been numerous instances of the deliberate poisoning of drinking water supplies or denial of drinking water service to an enemy.² Deliberate contamination remains a viable form of attack because it can easily be performed in a covert manner, and the resulting health effects, as well as potential panic, could be significant. As the threat of terrorism within this

Table 1. Data derived from biennial CDC reports of disease outbreaks resulting from contaminated drinking water.

Years	Number of States	Number of outbreaks	Number of ill individuals	Number of deaths	Outbreak size, range (median)	Pathogenic cause (no. of outbreaks)	Chemical cause (no. of outbreaks)	Cause indeterminate (no. of outbreaks)
1995–1996	13	22	2567	0	1–1449 (22)	7	7	8
1997–1998	13	17	2038	0	1–1400 (10)	10	2	5
1999–2000	25	39	2068	2	2–781 (13.5)	20	2	17
2001–2002	19	31	1020	7	2–230 (6)	19	5	7
2003–2004	19	30	2760	4	1–1450 (7)	17	8	5

country persists, it is therefore prudent to evaluate the vulnerabilities of our water infrastructure, to mitigate risks where possible, and to be prepared to respond in the event of an incident. In December 2003, the President of the United States issued Homeland Security Presidential Directive 7 (HSPD-7),³ which designated the U.S. Environmental Protection Agency (EPA) as the agency responsible for the protection of this country's drinking water. On January 30, 2004, the President issued HSPD-9,⁴ which directed the EPA to develop robust, comprehensive, and fully coordinated monitoring and surveillance systems for water quality. This EPA Water Security Initiative is piloting contamination warning systems at several U.S. water utilities to test their feasibility. These pilot programs are integrating automated online and routine water-sampling analyses, automated public health surveillance systems, consumer complaint monitoring, and enhanced physical security monitoring.

In a 2003 U.S. General Accounting Office report⁵ documenting the results of interviews with water experts, the water distribution network was identified as the most vulnerable component of the U.S. drinking water infrastructure. Recognizing this vulnerability, water utilities are increasingly measuring disinfectant levels and other parameters in their distribution systems. To improve surveillance further by providing an assessment of the likelihood that data anomalies are related to the occurrence of a waterborne disease outbreak, the EPA is collaborating with APL to test the feasibility of a warning-system prototype framework that integrates data from the following disparate sources: (i) traditional water quality parameters, which may be measured by online water quality sensors and by routinely collected grab samples (to be described subsequently), and (ii) community health data that may include early signs and symptoms of diseases of water origin but that are identified through automated public health surveillance

systems. Outcomes from this project will suggest the appropriateness of this type of integration as a component of the EPA Water Security Initiative.

The Approach

APL has been able to leverage its expertise in disease surveillance, data fusion, and signal analysis to arrive at a novel solution that should improve early detection of water contamination events and their associated health effects. Therefore, this pilot project was initiated to explore the feasibility of integrating measured drinking water parameters within an existing operational population-based public health syndromic surveillance system called the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE).⁶ The resulting addition to ESSENCE is called the Water Security (WS) module. This new application utilizes and integrates existing components and data elements with water measurement data and implements data fusion to combine water and health data to detect potential water-based health events.

One of the requirements of this prototype is that it be useful in routine monitoring as well as detection of deliberate contamination events, with the rationale being that, although deliberate attacks would be rare and catastrophic events, one cannot wait until these attacks are imminent or occurring before learning how to use the system. Instead, it is best if the system is used regularly so that users are familiar with features, limitations, etc. Such a dual-use approach thereby provides users with advanced preparation for use during actual or suspected bioterrorism events.

The objective of this paper is to provide an overview of the APL Water Security project and to illustrate the techniques developed for combining water quality and population-based health data to monitor for waterborne disease outbreaks.

METHODS

Water Quality Parameters

Public drinking water may come from reservoirs, surface streams, or underground aquifers, where each source has its own water characteristics (e.g., pH, minerals, etc.). Water treatment plants then use a variety of different technologies to process this source water into drinking water. In general, this involves sedimentation, filtration, and disinfection. Once the water leaves the treatment plant, it may remain in the distribution system for varying periods of time (e.g., hours to days) depending on usage and storage. Therefore, disinfectants such as chlorine or chloramines are typically added to maintain disinfection as the water travels and is stored throughout the distribution system. If certain levels of disinfectants can be maintained within the distribution system, then microorganisms may be kept to a minimum.

The water utilities routinely make measurements of certain parameters within the distribution system in order to monitor drinking water quality. These parameters may include water temperature, pressure, disinfectant level (e.g., chlorine, chloramines, etc.), pH, conductivity, *Escherichia coli*, total coliform, turbidity, and total organic carbon. Not all of these parameters are measured by all utilities in their distribution systems, and the sample measurement rates vary widely. Total coliform and *E. coli* are binary measurements, quantified only as present or absent. Other parameters are measured quantitatively on an appropriate scale. Potential drinking water contamination is most often indicated by a drop in the disinfectant level as that disinfectant reacts with the contaminant, regardless of whether the contaminant is chemical or biological. The only measurements of specific contaminants are the *E. coli* and total coliform parameters. Total coliforms include multiple species of bacteria that are common in the soil and in animal digestive tracts. One of the specific species included in total coliforms is *E. coli*, and this species consists of different strains; some of these strains can cause significant human disease, whereas other strains are innocuous.

Most water utilities rely on grab samples for these data. Grab samples are water samples collected from specific sites within the distribution system, typically at sample frequencies of weekly or less. In addition to the grab samples, some utilities have established a limited number of automated measuring systems at certain locations within the pipelines. These automated systems provide near real-time measurements as often as every few minutes. Contamination detection within a distribution system is therefore challenging because water data of different types are measured, the data are noisy, most of these data are not specific for particular contaminants, these data are measured asynchronously and at different rates, and there may be many missing data points.

Water Quality Data Preprocessing: Anomaly Detection

Previously, the EPA collaborated with Sandia National Laboratories to develop generic water-contaminant detection algorithms that employed pattern recognition.⁷ These algorithms were designed to provide early alerts based on automated online water sensor data. These ideal data were assumed to be continuously available, periodic, synchronous sensor measurements with no missing data. The pattern recognition approach does not incorporate known environmental and operational characteristics of the water distribution system and does not take into account knowledge of the specific chemical and physical responses of the measured parameters to different types of contamination. Operational and environmental responses of the water distribution system include cyclical behavior at diurnal, seasonal, and other time scales; effects of known aperiodic events (heavy rains, large ambient temperature changes, water-main breaks) on background levels; effects of human communal water usage on background levels (beyond seasonal effects); and variations among distribution system pressure zones, pump cycles, storage tank releases, etc. Knowledge of these operational and environmental responses is necessary to obtain accurate interpretation of real-world data when determining the quality of the water within the distribution system. Therefore, APL adapted the Sandia algorithms to accommodate the way water utilities currently operate (as well as future automated measurements) and in a way that takes into account the important water parameter behaviors mentioned above.

Continuously sampled water data are very noisy, so hourly averages are used to smooth the data. The mean and standard deviation (SD) are computed for samples in the previous 3-day window and are used to convert the data into normalized values. Next, a Yule–Walker autoregressive model is used to compute a normalized estimate for this value. Finally, the difference between the normalized data and the Yule–Walker estimate is compared to the threshold for the desired false-alarm rate.

Grab sample water data are processed by using a somewhat different technique than for the continuously sampled water data. A 7-day window is used, and the estimate is computed by subtracting the baseline mean from the measured value and dividing by the baseline SD. This estimate is then compared to a threshold for the desired false-alarm rate. The baseline for comparison must be found by using a different approach that will now be described. Missing water data and relatively infrequent grab sample collection mean that using a single grab sample site's historical data for its baseline might not be efficient or representative because one would have to use several weeks or months of historical data for an adequate baseline. One of the challenges was the determination of appropriate background baseline data on which a detection algorithm could be used

for comparison for the identification of an anomaly in each grab sample site's data. Figure 1 illustrates the finding by APL that certain sets of water grab sample sensors may have similar properties and therefore may be used together to obtain a more complete baseline when there are missing data from single sites. Based on technology previously used in oceanographic analysis,⁸ APL developed a divisive hierarchical clustering algorithm to identify grab sample sites that are located in waters with similar characteristics (Fig. 1). A modified version of the oceanographic algorithm was found to work well when it took into account the geographic location of the grab sample sites, the distances of these sites from the water treatment plants, and the historical disinfectant levels (see *Results*). Therefore, this algorithm was used to identify clusters of grab sample sites that were then used to determine the appropriate baseline against which each grab sample measurement in that cluster could be compared.

Recall that some water utilities use both grab sample and continuous-monitoring sensors to obtain water quality data. Because of the differences in data-collection frequency, as well as common issues such as noisy and missing data, further processing of these data was performed by using a Bayesian Network (BN) analysis to

be described later. In the next section, the public health data will be described.

Automated Syndromic Surveillance with ESSENCE

An important technology developed for earlier detection of disease outbreaks in the community is the automated syndromic surveillance system.⁹ By looking for symptoms of illness, these public health surveillance systems attempt to identify potential outbreaks or illness earlier than through clinical or laboratory diagnosis or traditional public health surveillance systems.¹⁰ Early warning of a possible incident can reduce the potential morbidity and mortality of an intentional chemical, biological, or radiological attack on a community. Specifically, early detection of water contamination incidents has the potential to reduce the extent of contamination through the distribution system, thereby partially mitigating what would undoubtedly impart massive economic impacts on the utility, its residential and commercial customers, and the community at large.

Various data sources have been investigated for syndromic surveillance systems,¹¹ including, but not limited to, chief complaint data from hospital emergency department (ED) admission records. At APL, the ESSENCE project has successfully incorporated data from these

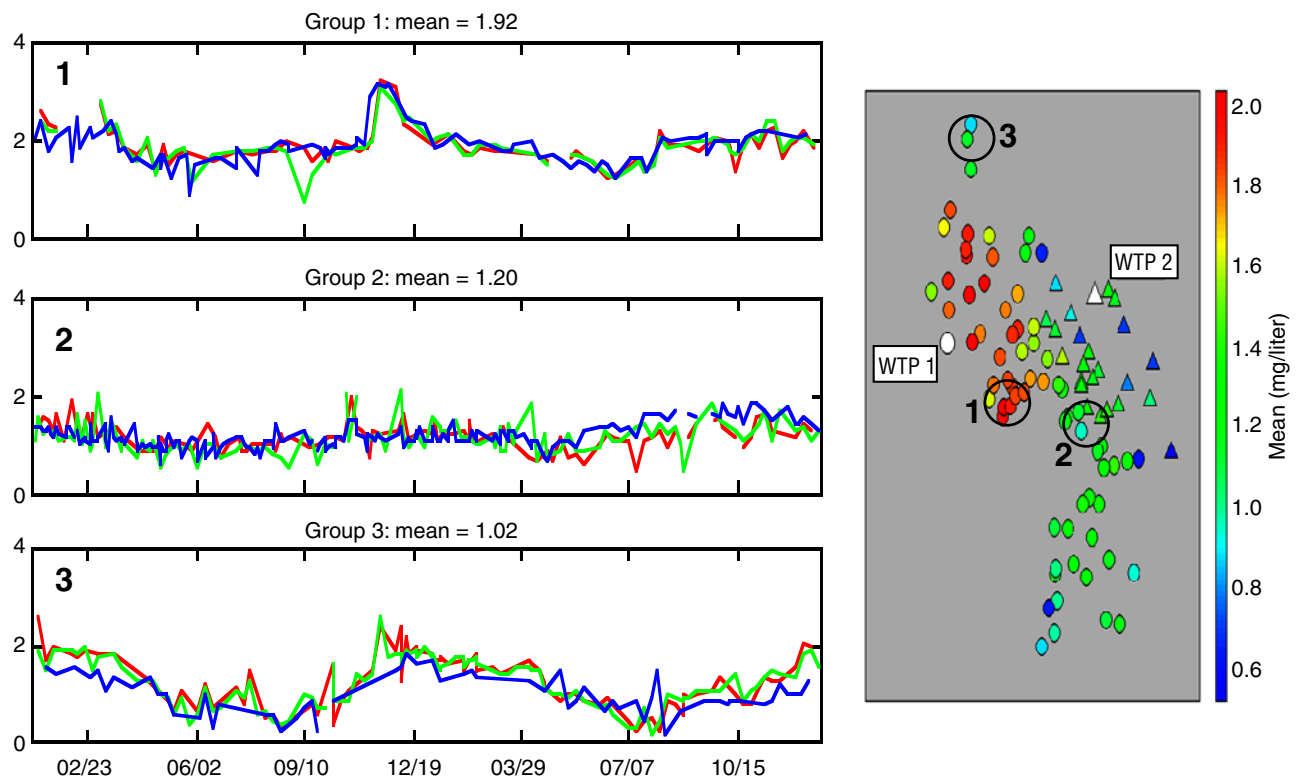


Figure 1. Regional differences in weekly disinfectant levels within a distribution system. The geographic information system (GIS) map on the right shows the locations of water sample sites. The time-series plots on the left show the disinfectant levels (in milligrams per liter) for water sample sites in regions 1–3, respectively. Note that there are larger differences among different regions than within each region. WTP, water treatment plant.

sources into an operational disease surveillance system that includes techniques for handling noisy, sparse, missing, and asynchronously sampled data from disparate sources. The ESSENCE project therefore provides an existing framework upon which the WS module could be added.

Because of differences in data sources among different locations using the ESSENCE system, it was decided to develop the initial WS module with hospital ED chief complaint data as the source of health data in the prototype because it was commonly used by ESSENCE in multiple locations. A chief complaint is a brief text description of the major reason expressed by the patient for seeking health care from the ED. In addition, these ED data had to be filtered in order to reveal only those data that may be related to a waterborne illness. Although there are a wide variety of diseases that may be related to water contamination, it was decided to focus on gastrointestinal (GI) disease. The primary reason for this was that there would be GI outbreaks that could be observed in the data against which the disease detection algorithms could be tested. Examples of ED chief complaints related to GI disease include nausea and vomiting, diarrhea and vomiting, stomach cramps and diarrhea, food poisoning, bloody diarrhea, etc. The methods for selecting health data that may be related to waterborne disease will be described in the subsequent section on BN analysis.

BN Analysis

This analysis is performed after the site detection and clustering analysis described above and applies BN technology.¹² In addition to the water data, BN technology was used for the public health data and for fusing the public health and water data. BNs are directed acyclic graphs whose nodes typically represent hypotheses that could be true or false based on probability and the inputs to the nodes. These nodes are linked by conditional dependencies that can be based on expert and/or evidential reasoning. Therefore, BNs can produce probabilities as a degree of belief that an anomaly has or has not occurred. By incorporating conditional probabilities and linkages among different data types and sources, BNs are capable of using historical data and expert knowledge to determine likelihoods of significant anomalies in the data. BNs are able to include disparate types of evidence¹³ and can account for continuous versus discrete data, multiple data rates, missing or sparse data, and expert or heuristic knowledge. Because they contain graphical representations of conditional dependencies, they are transparent enough that the user can visualize the basis for

anomaly probability. Therefore, the APL strategy was to use algorithm outputs and select preprocessed water data as inputs to the BNs in order to keep computational intensity manageable and scaleable.

The WS module prototype applies BN technology using historical data and expert knowledge from public health professionals, physicians, and water engineers. Three types of BNs are used in the WS module (Fig. 2). The Water Quality BN takes into account environmental and operational interactions with the water sensors and fuses the water sensor data collected at different sample rates. The Water Quality BN examines algorithm outputs for the various water parameters described above to test for anomalies in order to provide a degree of belief that a water contamination event has or has not occurred, based only on the water data.

The examination of population health data (i.e., ED chief complaints) for anomalous levels of GI-related illness potentially caused by waterborne pathogens is performed by the GI Health BN (Fig. 2). This BN selects syndromic public health data for consistency with waterborne illnesses and takes into account expert knowledge based on reported cases of confirmed waterborne disease outbreaks that have occurred in the past. It does not distinguish the specific cause of the disease other than the degree of belief that the data are consistent with a waterborne disease of some type.

Finally, the Health/Water Quality Fusion BN (Fig. 2) examines the outputs from these two BNs, as well as ancillary preprocessed water data, to determine the degree of belief that the water quality anomalies are or are not associated with GI illnesses identified in the ED data.

RESULTS

Clustering Analysis

The divisive hierarchical water sensor site clustering technique described earlier provided a much improved

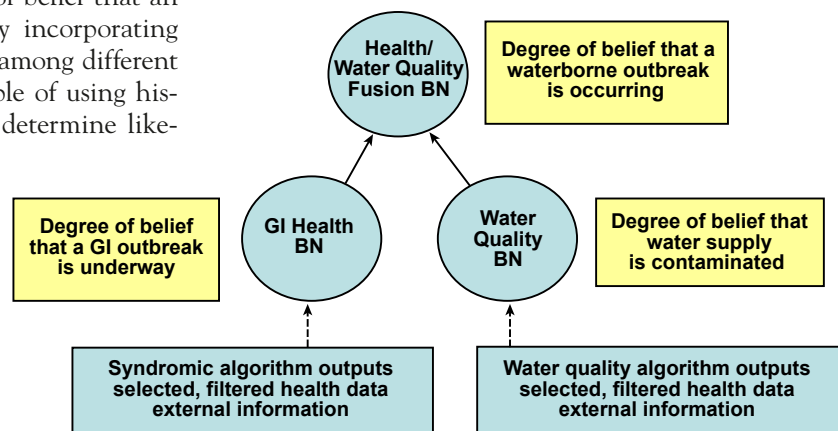


Figure 2. Conceptual scheme of data fusion for waterborne disease event detection.

baseline for comparing individual grab sample site sensors. The left column of Table 2 shows how many sensor sites are available to provide a 7-day baseline when no site clustering is used and no data dropouts are allowed. The center column shows slight improvement when up to four baseline data dropouts are allowed. The right column shows the results when the divisive hierarchical clustering algorithm is applied. Note in this column that now an adequate 7-day baseline is provided for individual sensor anomaly detection algorithms to operate for 265 days during a year, thereby covering most of the weekdays during which grab sample measurements are performed. If grab sample data collection frequency were increased, this coverage could be further improved.

Depending on the water utility location, the water quality spatial cluster analysis yields on the order of five to seven water characteristic regions. These clusters enable the inclusion of sensors with intermittent data and provide a stable baseline. As an example of the algorithm detection performance achieved for Water Quality BN input with this stabilization, the measured probability of detection for transient signals was 0.69 (SD = 0.10) at a background alert rate of 1 every 190 days. Because there was no record of any actual water contamination in the historical water data provided by the utility, there was no way to know whether a signal in these data was an indication of actual contamination. Therefore, there is no bona fide measure of a false alert rate,

so we measure the alert rate when no contamination signal was added and designate it a “background alert rate.” Consequently, this background alert rate may or may not be a false-positive rate. If we empirically determine a detection threshold for a chosen background alert rate, we may then add a contamination signal to the data to determine a probability of detection for that signal at that alert rate.

Water Quality BN

After data processing by the water anomaly detection algorithms described above, BN technology is then used to take into account environmental and operational characteristics of the water distribution system, as well as the differing data rates (Fig. 3). For example, note that higher temperatures, such as those found within the distribution system during the summer months, lead to faster evaporation of chlorine gas from the water and therefore a drop in chlorine levels. The total chlorine in this figure represents the amount of disinfectant added to the water by the utility, and the free chlorine represents that fraction of disinfectant actually available for reaction to inactivate a contaminant. Therefore, a drop in both total and free chlorine may be the result of the operational environment rather than a true contaminat-

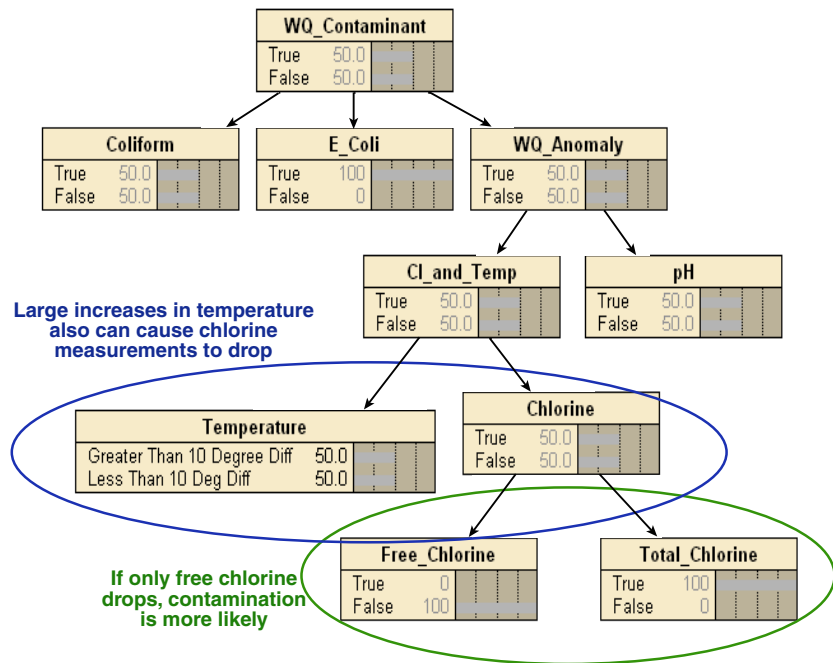


Figure 3. Example of nodes for the Water Quality BN and how environmental and utility operational knowledge can be added to the algorithm. WQ, water quality.

Table 2. Results of water grab sample site hierarchical clustering to provide baseline.

	No clustering, no baseline dropouts allowed	No clustering, ≤4 baseline dropouts allowed	Clustering, ≤2 baseline dropouts allowed
Average no. of usable sites per day	0	0.05	12.24
Maximum no. of usable sites per day	0	3	38
No. of days with at least one usable site per day	0	15	265

The number of grab sample sites available for providing a 7-day window for deriving a baseline over a 1-year period are shown.

tion event. Because a drop in disinfectant level is the primary evidence for a contamination event, it is important to take such knowledge into account.

GI Health BN

As mentioned earlier under *Methods*, the public health data were restricted to hospital ED chief complaints for this prototype. Filtering was performed in order to retain only those ED visits that were consistent with diseases that could be waterborne. Because these data are chief complaints and not diagnoses, it would be impossible to determine from the health data alone whether the chief complaints were specifically water-related. For example, because water is typically used in cooking, one may not readily distinguish a waterborne from a foodborne illness. To provide some likelihood of whether these ED visits represented waterborne illness, an additional BN was used to find spatiotemporal correlations among both health and water data. This BN is called the Health/Water Quality Fusion BN and is described below.

Health/Water Quality Fusion BN

As illustrated in Fig. 4, this BN model can be based on expert knowledge combined with detailed correlation of relevant cases. The final output (Fig. 5) is presented as a summary alert webpage showing the outputs of the three BNs over the previous several days. Webpage alert hyperlinks provide the user with maps (Fig. 5), time series, and other details to help characterize the potential outbreak. It is important to note that the WS module is a tool for examining large quantities of disparate data, and the final decision as to whether a

Region	Water/Health Fusion	GI Health	Water Quality
Area A	*****	*** ** *	*****
Area B	***** **	***** **	***** **
All	*****	*****	*****

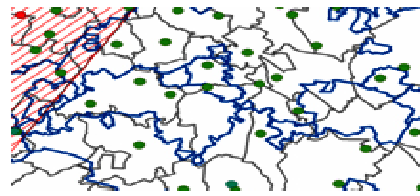


Figure 5. Examples of user screens showing alert levels and GIS maps of sensor locations.

waterborne disease outbreak is occurring is the responsibility of the public health official.

CONCLUSIONS

This article describes the development of a WS module for the ESSENCE syndromic surveillance system that will undergo its initial beta testing early in 2008. Further development of this prototype will ensue based on this initial feasibility beta testing. Although the prototype is designed for detection of waterborne disease outbreaks affecting GI health, the next phase will also involve the inclusion of other types of health effects resulting from water contamination. For example, chemical and toxin contamination may result in neurological and dermatological symptoms. The current prototype was designed to be used with water utilities that employ either chlorine or chloramines as disinfectants. While this

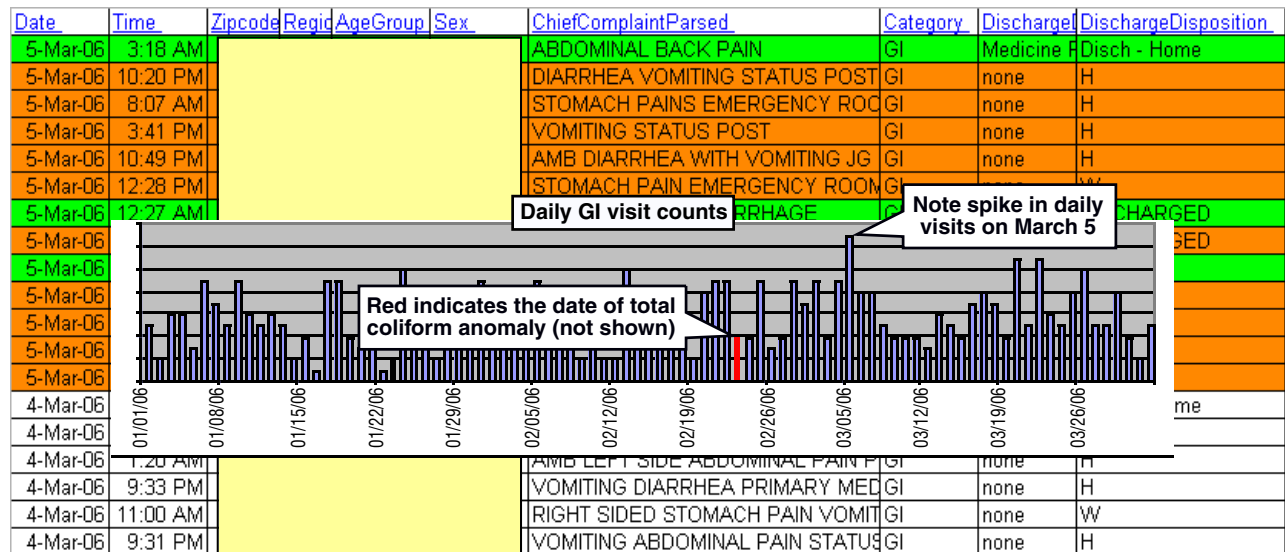


Figure 4. Example of how expert knowledge and actual case reports may be used to find correlations in the water and health data. For example, knowledge of disease incubation periods may mean that the spike in daily visits is associated with the presence of coliform in the water.

prototype will be tested with the participation of two water utilities, there are plans for the addition of water utilities at other locations with existing syndromic surveillance systems.

The use of water quality data for health monitoring must take into account the operational and environmental characteristics of the water distribution system while restricting attention to the health data characteristic of waterborne disease. The nested BN approach described above (Fig. 2) allows fusion of water information with dissimilar population health data to encapsulate both engineering and epidemiological expertise.

ACKNOWLEDGMENTS: We thank Sean McKenna and David Hart at Sandia National Laboratories for sharing an early version of the CANARY Event Detection Software. The expert knowledge of personnel at the Washington Suburban Sanitary Commission, Milwaukee Water Works, and the Milwaukee County, Montgomery County, and Prince George's County Health Departments is gratefully appreciated. We also express our appreciation to Charles Hodanics, Josh Suereth, Raj Ashar, Logan Hauenstein, Mohammed Hashemian, Wayne Loschen, and Larry Frank for their valuable assistance in this project. The EPA, through its Office of Research and Development, funded and managed the research described herein via Contract EP-C-06-074. This work has been subjected to the EPA's administrative review and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendations for use.

REFERENCES

¹Liang, J. L., Dziuban, E. J., Craun, G. F., Hill, V., Moore, M. R., et al., "Surveillance for Waterborne Disease and Outbreaks Associated with

Drinking Water and Water Not Intended for Drinking—United States, 2003–2004," *MMWR Surveill. Summ.* **55**(SS12), 31–58 (2006).

²Lombardo, J. S., and Ross, D., "Disease Surveillance, A Public Health Priority," Chap. 1, in *Disease Surveillance: A Public Health Informatics Approach*, J. S. Lombardo and D. L. Buckeridge (eds.), John Wiley and Sons, Hoboken, NJ, pp. 20–21 (2007).

³White House Office of the Press Secretary, Homeland Security Presidential Directive (HSPD), *HSPD-7: Critical Infrastructure Identification, Prioritization, and Protection* (17 Dec 2003).

⁴White House Office of the Press Secretary, Homeland Security Presidential Directive (HSPD), *HSPD-9: Defense of United States Agriculture and Food* (3 Feb 2004).

⁵U.S. General Accounting Office, "Drinking Water: Experts' Views on How Future Federal Funding Can Best Be Spent to Improve Security," Report to the Committee on Environment and Public Works, U.S. Senate, GAO-04-029, U.S. Government Printing Office (October 2003).

⁶Lombardo, J. S., Burkom, H. S., Elbert, Y., Magruder, S., Happel Lewis, S. L., et al., "A Systems Overview of the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE II)," *J. Urban Health* **80**(2), i32–i42 (2003).

⁷McKenna, S. A., Wilson, M., and Klise, K. A., "Detecting Changes in Water Quality Data," *J. Am. Water Works Assoc.* **100**(1), 74–85 (2008).

⁸Mandelberg, M. D., and Frizzell-Makowski, L. J., "Acoustic Proving of Ocean Basins," in *Proceedings of the OCEANS 2000 MTS/IEEE Conference and Exhibition*, Providence, RI, pp. 105–108 (2000).

⁹Bravata, D. M., McDonald, K. M., Smith, W. M., Rydzak, C., Szeto, H., et al., "Systematic Review: Surveillance Systems for Early Detection of Bioterrorism-Related Diseases," *Ann. Intern. Med.* **140**, 910–922 (2004).

¹⁰Buckeridge, D. L., "Outbreak Detection Through Automated Surveillance: A Review of the Determinants of Detection," *J. Biomed. Inform.* **40**, 370–379 (2007).

¹¹Babin, S., Magruder, S., Hakre, S., Coberly, J., and Lombardo, J. S., "Understanding the Data: Health Indicators in Disease Surveillance," Chap. 2, in *Disease Surveillance: A Public Health Informatics Approach*, J. S. Lombardo and D. L. Buckeridge (eds.), John Wiley and Sons, Hoboken, NJ, pp. 43–90 (2007).

¹²Pearl, J., "Fusion, Propagation, and Structuring in Belief Networks," *Artif. Intell.* **29**(3), 241–288 (1986).

¹³Lin, J. S., Burkom, H. S., Murphy, S. P., Elbert, Y., Hakre, S., et al., "Bayesian Fusion of Syndromic Surveillance with Sensor Data for Disease Outbreak Classification," Chap. 6, in *Life Science Data Mining*, S. Wong and C.-S. Li (eds.), *Science, Engineering, and Biology Informatics – Vol. 2*, World Scientific Publishing, Inc., Hackensack, NJ (2007).

The Authors

Steven M. Babin earned a B.S. in engineering physics (Special Distinction) from the University of Oklahoma, an M.D. at the University of Oklahoma, and an M.S.E. in electrical engineering and science at the University of Pennsylvania. He joined APL in 1983 and earned M.S. (1994) and Ph.D. (1996) degrees in meteorology from the University of Maryland. He has over 50 publications and presentations and is a member of the IEEE (Senior), the American Meteorological Society, the American Geophysical Union (life), Sigma Xi, the International Association for Urban Climate, and the International Society for Disease Surveillance. His recent activities have included environmental science and public health research and development of novel techniques for disease detection. **Howard S. Burkom** received a B.S. degree from Lehigh University and M.S. and Ph.D. degrees in mathematics from the University of Illinois at Urbana–Champaign. He has 7 years of teaching experience at the university and community college levels. Since 1979, he has worked at APL developing detection algorithms for underwater acoustics, tactical oceanography, and public health surveillance. Dr. Burkom has worked exclusively in the field of biosurveillance since 2000, primarily adapting analytical methods from epidemiology, biostatistics, signal processing, statistical process control, data mining, and other fields of applied science. He is an elected member of the Board of Directors of the International Society for Disease Surveillance. **Zaruhi R. Mnatsakanyan** is leading research activities in the field of decision support and data fusion for The Johns Hopkins University's Center of Excellence in Public Health and Disease Surveillance Program in the National Security Technology Department (NSTD) at APL. Dr. Mnatsakanyan has a Doctor of Engineering degree in computer science and machinery engineering from the State Engineering University of Armenia as well as master's degrees in computer science and in computer science and automated engineering from the same university. Her research interests include sensor data fusion, Bayesian Networks, pattern recognition, machine learning, and cognitive science. **Liane C. Ramac-Thomas** (not pictured) is a member of the Senior Professional Staff in APL's NSTD. She received a B.S. in electrical engineering from the University of Washington and an M.S. and Ph.D. (both in electrical engineering) from Syracuse University. Dr. Ramac-Thomas joined APL in 2000. She currently works on modeling, analysis, and algorithm development for various tasks related to surveillance of disease outbreaks and terrain-aided navigation. She is a member of the Engineering in Medicine and Biology and Signal Processing Societies of the IEEE. **Michael W. Thompson** earned a B.A. in music and a B.S. in applied physics from Brigham Young University in 1998. He subsequently earned an M.S. in



Steven M. Babin



Howard S. Burkom

Zaruhi R.
Mnatsakanyan

Michael W. Thompson



Richard A. Wojcik



Sheri Happel Lewis



Cynthia Yund

physics from Brigham Young in 2000 doing computer simulations of nonlinear sound production in trombones. He continued his education at the Pennsylvania State University, where he earned a Ph.D. in acoustics in 2004 doing experimental studies of nonlinear acoustic streaming in thermoacoustic devices. In 2004, he joined the Acoustics and Electromagnetics Group (STX) in NSTD at APL. He has since been working in the areas of disease surveillance and underwater acoustics. Dr. Thompson is a member of the Acoustical Society of America. **Richard A. Wojcik** is a member of the Principal Professional Staff in APL's NSTD. He received his B.S. in computer science from the Pennsylvania State University in 1980 and his M.S. in computer science from The Johns Hopkins University Whiting School of Engineering in 1985. Mr. Wojcik joined APL in 1980 and currently is part of the Information Technology (IT) development team in support of disease surveillance initiatives and system integration of ESSENCE. **Sheri Happel Lewis** is a member of the Senior Professional Staff in APL's NSTD. She received her B.S. in biology with a minor in Asian studies from Loyola College in Maryland and her master's of public health (M.P.H.) in international public health programs from the George Washington University School of Public Health and Health Services. Ms. Happel Lewis joined APL in 2001 and is a program manager within APL's disease surveillance program. She works closely with the public health community on the development and implementation of electronic disease surveillance systems. Her current work focuses on the

implementation of such systems in resource-poor settings. **Cynthia Yund** is an Environmental Health Scientist with the National Homeland Security Research Center in the Office of Research and Development at the EPA. Dr. Yund holds a B.S. in nursing from the Ohio State University and a Ph.D. in environmental health from the University of Cincinnati. Her expertise is in public health epidemiology, infectious disease investigation, and syndromic surveillance. Current activities include project management of the public health programs for the Water Security Initiative. For further information on the work reported here, contact Steven Babin. His e-mail address is steven.babin@jhupl.edu.