Criteria for Evaluating Distribution Network Early Warning Systems

Dan Kroll, Chief Scientist
Hach Homeland Security Technologies

Hach Homeland Security Technologies
5600 Lindbergh Drive • Loveland, Colorado 80539
970-663-1377 • dkroll@hach.com

Homeland Security Technologies

Introduction

The ability to detect and act upon changes in water quality is a critical component in the drive to protect our drinking water supplies from intentional or accidental contamination.

A number of studies conducted since 9/11 have shown that bulk monitoring of basic water quality parameters has the potential to indicate the presence of many harmful agents in water at the levels of interest.(EPA, 2005; Kroll, 2006; Hall et. al., 2007).

Since 9/11, numerous communities have installed multi-parameter monitoring stations in various locations through out their distribution network as early warning systems to detect potential water security threats as well as providing operational data.

These data streams are quite complex and it becomes a Herculean task to differentiate what is normal background noise and fluctuations due to normal everyday events from changes that are indicative of a deviation in water quality deserving further attention. Intelligent algorithms are a necessary adjunct to bulk parameter monitoring if useful decision trees are to be built from this type of monitoring.

Over the past several years a number of such algorithms designed for this use have been under development by private industry, government programs, universities and national labs. The question then becomes how to evaluate the effectiveness of these potential early warning system (EWS) solutions. While a number of research studies (McKenna et al, 2006; Umberg, 2008) and programs such as the EPA's ETV program (ETV2005) have attempted to set criteria and means for evaluating such systems, many important functional criteria for such systems have been overlooked in past studies. A number of factors that are mandatory for the success of any such system are outlined here.

#1 Dual Use Value

The first priority for any EWS is that it detects contaminants so that people and infrastructure can be protected in some effective way. While the ability to detect contamination is critical, if a system is to be capable of offering a likely return on investment (ROI) and achieve widespread deployment, it would also be useful if the EWS can inherently provide information that would be useful on a day-to-day basis. It is presumed that an EWS spends very little of its total time in an actual alarm condition. What is it doing for the rest of the time? Justification of the cost can include:

1. Optimization of daily operations within the system monitored,
2. Providing alarms for operational events not related to contamination threats,
3. Replace grab sampling for compliance purposes with continuous monitoring,
4. Document system operation anomalies to assist with planning maintenance activities or planning and justifying major system upgrades (line replacement)
5. Building consumer confidence by continuously documenting system water quality

Such capabilities can reduce the cost of operations for the system being monitored, and possibly provide information that can be useful to those people who use the monitored system.

#2 Detection Class Requirements

When discussing early warning systems, there are a number of different classification levels relating to their effectiveness commonly utilized in categorizing systems for military use. The following Detection Classes are possible:

1. **Detect to treat:** very high confidence level, nearly 100%, which allows treatment to proceed on those exposed. Response time slow, cost high.
2. **Detect to protect:** high confidence level, > 99%, that allows for protection by limiting exposure without the need for confirmatory testing. Response time fast, cost medium.
3. **Detect to warn:** Presumptive confidence level, <99%, which would allow for protection by limiting exposure while confirmatory tests are run. Response time fast, cost low.

Confidence levels are usually a function of analysis time and cost, with higher confidence levels increasing both. An EWS will need to respond rapidly to be effective, and it will need to be relatively low in cost to provide wide coverage, so this paper will assume that the first deployed defense against contamination will be systems for Detect to Warn.

#3 Coverage Characteristics

1. **Cost:** Cost may not be significantly limiting when selecting an EWS for an icon or high-profile facility, but if the coverage is large, such as a major metropolitan area, budgets may be constrained, and the degree of coverage becomes a function of the cost per point monitored. Therefore, cost becomes an issue.

2. **Area of Protection:** Note that effective coverage may be a function of the hydraulics of the distribution system within the geographical setting. Not all ideal deployment sites from a protection standpoint will be capable of being utilized due to system requirements, site "noise functions" and other logistics. While "protection of all" may be a laudable goal, reality may constrain the degree of implementation, forcing tradeoffs.

3. **Communication:** Monitoring of multiple points in a geographical area immediately raises the need for communication at least from the remote points to some centralized facility where the data can be interpreted, and actions taken. The EWS can accomplish highly sophisticated interpretation of the local data, but it cannot take actions in a complex situation affecting possibly millions of people. Human interaction and analysis will be required. The information from the multiple points must be communicated to an analysis and command center.

Thus, an EWS must be structured for secure communication. Both the instrument and the network must have tools in place to affect a high level of security so that information cannot be blocked, or false information transmitted on the network.

#4 Operational Characteristics

There are a number of operational characteristics that need to be incorporated into the design of such a system.

1. **Ease of use:** A system that may be absolutely critical in a crisis cannot be difficult to use. User interface operation must be intuitive enough that minimally skilled operators can obtain necessary information without resorting to an operator's manual.
2. **Automated:** The system must normally operate without requiring the presence of a human. Human intervention should only be needed for service or maintenance.
3. **Continuous:** The system must run without any long gaps in analysis that enable contamination of significant duration to slip by undiscovered. The maximum time for non-observability should be relatively short—on the order of 3 minutes. Longer response times unduly endanger a larger percent of the population by exposing them to the hazard before remedial actions can be taken.
4. **Reliable:** A non-working system is an opportunity for exploitation.
5. **Cost Effective:** Amortized cost per day comparable to or less than the labor, travel, equipment and reagents for an existing grab sample program. When evaluating this criterion, it is important to factor in the overall better picture of operations afforded by continuous versus discrete measurements

#5 Performance Characteristics

1. Detection of a broad spectrum of contaminant classes

There are many contaminants that could cause serious harm if introduced into a drinking water distribution system. It should be noted that chemicals within a given category could have fundamentally different chemical characteristics. The commonality that one finds within a category does not preclude different responses of analytical sensors used in an EWS. The ability to detect with significant specificity is dependent upon seeing a chemical in many dimensions. Thus, single parameter systems have inherent limitations. There are a number of important categories of contaminant that such a system should detect including; Organo-phosphates, Carbamates, Other pesticides, Herbicides, Nematicides, Animal poisons, Petroleum products, Heavy metals, Infectious agents, Warfare agents (tested against real threat agents), Toxic Industrial Chemicals, Poisons, Cyanides, Toxins, Drugs and pharmaceuticals both legal and illicit

2. Rapid

The response time of any detection method that is used on a flowing stream is really a measure of the delivery rate of harmful contaminants. A simple way to view this is given by the equation:

$$\textbf{\# of lethal doses delivered} = \frac{\textbf{stream flow rate X response time X conc.}}{\textbf{toxic dose}}$$

With this in mind, it becomes readily apparent that response times of a few minutes are required. Considering the flow rates of the pipes involved for most deployment scenarios, response times of hours or even 10's of minutes are unacceptable.

3. Specific

Specificity of contaminant identification can be obtained in two ways:

a) *Analyze for a specific molecule or organism.*

b) *Analyze generally across multiple orthogonal dimensions via mathematical analysis of the data from multiple sensors.*

Detectors of both types exist, so the value of comparing the two approaches is evident. Given the immense number of possible contaminants that could be put into a water distribution system, looking for specific molecules would require an immense number of sensors. There are so many problems associated this approach that it quickly become untenable. The approach of using a manageable set of orthogonal sensors faces only the difficulty of obtaining sufficient information to apply pattern recognition methods that can differentiate between contaminants or classes of contaminants. One advantage of using an orthogonal set of different sensors is that it becomes nearly impossible to find some contaminant that goes by all sensors unnoticed.

4. Reproducible

An EWS must be reproducible to be trustworthy. Reproducibility can be demonstrated via testing with actual contaminants (for example: Aldicarb, Anthrax culture, Cyanide, Fluoroacetate, Nicotine, Ricin, Sarin, VX)

5. Low false negatives, low false positives

If a system can be blind to certain classes of chemicals (for example, not visible in the Ultraviolet 254 spectrum) then it represents an opportunity to contaminate the system without an alarm. Systems with multiple types of sensors are difficult if not impossible to fool, so that kind of system will normally produce some kind of alarm upon contamination, and false negatives will be very unlikely.

False positives can come from two sources: random system noise, and insufficient information during the analysis of event data. Given that systems are not noise-free, and there will always be some degree of insufficient information, false positives are bound to happen. The ultimate question is the determination of the number per time in a given installation (since noise is site-specific) and the acceptable frequency. False positives caused by noise can be reduced by the proper choice of alarm threshold according to ROC curves.

False positives caused by imperfect information will happen in an EWS using an inferential method. Given that a multi-parameter system implies an inferential method that can give false positives, and that an actual contamination event is a very rare occurrence, it should be recognized that there will be more false positives then actual positive events. Thus, any detection must be considered presumptive until follow-up testing verifies or denies the detection.

6. Qualitative

Contaminants can be named if the EWS is specific, but people using the information may not have sufficient training to recognize the nature of the contaminant found from its name. The system should assign a general classification category for simple clarification.

7. Quantitative

Ideally, an EWS would provide enough information to give some quantitative information about any contaminant presumably found in the water system. Such information could be useful in determining the threat level during a given episode, and also might be useful if treatment of people or infrastructure become necessary.

8. Sensitive

Clearly an EWS must be sensitive to contaminants present in harmful amounts, but quoting a simple minimum detection limit can be misleading. Contaminants change water chemistry and those changes can be analyzed as part of an alarm process that generates a Trigger signal. Changes are seen against a background of noise, or natural fluctuations in measured parameters. Accordingly, one cannot address sensitivity unless it is relative to a site's noise properties.

Fortunately, military needs have produced a useful tool for simply stating a detector's capability: the Receiver Operating Characteristic (ROC) curve.

Such curves allow an operator to select a detection alarm threshold according to local noise characteristics. Note that this allows an operator to increase sensitivity, accepting the higher probability of a false alarm when it is suspected that contamination is more likely. Figure 1 shows a representation of ROC space.

A ROC curve is a graphical representation of the trade off between the false negative and false positive rates for every possible cut off. Equivalently, the ROC curve is the representation of the tradeoffs between sensitivity (Sn) and specificity (Sp). By tradition, the plot shows the false positive rate on the X axis and 1 - the false negative rate on the Y axis. You could also describe this as a plot with 1-Sp on the X axis and Sn on the Y axis. A good diagnostic test is one that has small false positive and false negative rates across a reasonable range of cut off values. A bad diagnostic test is one where the only cut offs that make the false positive rate low have a high false negative rate and vice versa.

An evaluator is usually satisfied when the ROC curve climbs rapidly towards upper left hand corner of the graph. This means that 1- the false negative rate is high and the false positive rate is low. We are less happy when the ROC curve follows a diagonal path from the lower left hand corner to the upper right hand corner. This means that every improvement in false positive rate is matched by a corresponding decline in the false negative rate. You can quantify how quickly the ROC curve rises to the upper left hand corner by measuring the area under the curve. The larger the area, the better the diagnostic test. If the area is 1.0, you have an ideal test, because it achieves both 100% sensitivity and 100% specificity. If the area is 0.5, then you have a test, which has effectively 50% sensitivity and 50% specificity. This is a test that is no better than flipping a coin. In practice, a diagnostic test is going to have an area somewhere between these two extremes. The closer the area is to 1.0, the better the test is, and the closer the area is to 0.5, the worse the test is.

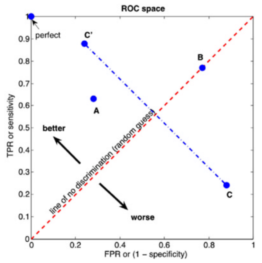

Figure 1. Classic ROC Curve Space

#6 Determining ROC Curves for Data from Sites on Military Bases

The US Army Corps of Engineers Research and Engineering Laboratory (CERL) under a Cooperative Research and Development Agreement (CRADA) with Hach Homeland Security Technologies have worked for the past several years to evaluate early warning systems in real world deployment scenarios on military bases. As well as helping to develop theories of operation and response, these deployments have resulted in a large body of data from a variety of sites for which ROC curves have been calculated.

In the system evaluated in this study, once each minute, the EDS algorithm processes five water quality parameter readings: pH, conductivity, chlorine, turbidity, and Total Organic Carbon. The algorithm calculates the deviations of those parameters from the process baseline, numerically scales those five deviations and then calculates a single Trigger Signal via a distance measure. The Trigger Signal is compared to a user selectable Threshold level. If the Trigger Signal is greater than the Threshold, the system is said to be in alarm, indicating significant deviation from baseline conditions in the water. U.S. Patent 6,999,898

EDS systems with the above described algorithm were tested at two military installations (a BASE in the southeast and a FORT on the east coast) and at a municipal site in a large Midwestern city where data were collected for many months. In addition to providing the on-line monitoring function, the EDS at each site provided data that could be used for two parallel analyses: 1) Receiver Operating Characteristic (ROC) curve analysis via a Monte Carlo simulation in a spreadsheet, and 2) simulation of agent additions to the water at each site via superposition of agent deviation data on the site data, followed by playing the resultant data files through the EDS algorithm. The results from the Monte Carlo simulation were compared to the simulation with actual site data over time.

Conclusion

A number of projects have been initiated to develop and/or evaluate EWS. After these early studies verified the efficacy of bulk parameter monitoring, later studies focused on the development and testing of event detection algorithms and their utilization in interpreting the generated data streams. While these studies utilized some of the criteria for evaluation listed above, including ROC curves, none have concentrated on all of the criteria or held these criteria to meaningful goals that would indicate that these systems are ready for real world deployment.

The systems that were deployed in the CERL study were found to perform acceptably under all of the evaluation criteria set forth in this paper. They demonstrated dual use by detecting a number of operational and non-security events including, dead ends, contamination by aviation fuel, pumping schedule problems and others. These results show that the EDS under examination can detect low levels of threat agents in a few minutes at 100% probability of alarm. These systems were detect to warn type systems and lived up to that definition in speed (time to respond was a little as 3 minutes) and cost criteria (When compared to grab sampling protocols). Coverage characteristics were found to be adequate and communications were determined to be simple and secure.

All operational criteria including continuous operation and reliability were verified. The systems were under a service contract with the manufacturer whom preformed routine maintenance and calibrations. Any required service beyond what was covered in the service contracts was minimal. The operational characteristics of the systems were deemed to meet the criteria. It was also demonstrated that the Monte Carlo analysis for the determination of ROC curves and Minimum Detection Limits at a site, closely matches the results of simulations with agent-superimposed data sets. Water distribution systems early warning systems are shown to be an effective means of enhancing the security of base water supplies and the ROC curve method is shown to be an effective way of validating these systems with a tool that can be useful to operators in running and tuning these systems.