SAND2008-3881C

On-line Identification of Adverse Water Quality Events from Monitoring of Surrogate Data: CANARY Software

Sean A. McKenna and David B. Hart
Sandia National Laboratories
Albuquerque, New Mexico USA

June 26th, 2008

Sandia National Laboratories

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94-AL-85000

Overview

- Water Security: Both physical protection of infrastructure (RAM-W) and enhanced monitoring of distribution networks (focus here)
  - Where to place sensors (SPOT)
  - How to detect water quality events (CANARY)
  - How to rapidly determine the location of a contaminant source (PONI)
- Focus on general approach, but CANARY is the Event Detection Software (EDS) that we have developed provides examples in this presentation

WQ Monitoring: Future

- "Chem-lab on a microchip" technology promises to revolutionize in-situ monitoring of water quality
  - The Goal: Inexpensive, robust, networked, compound specific, in-situ capability
  - The Reality: Significant engineering challenges remain to go from laboratory prototypes to field deployments

WQ Monitoring: Present

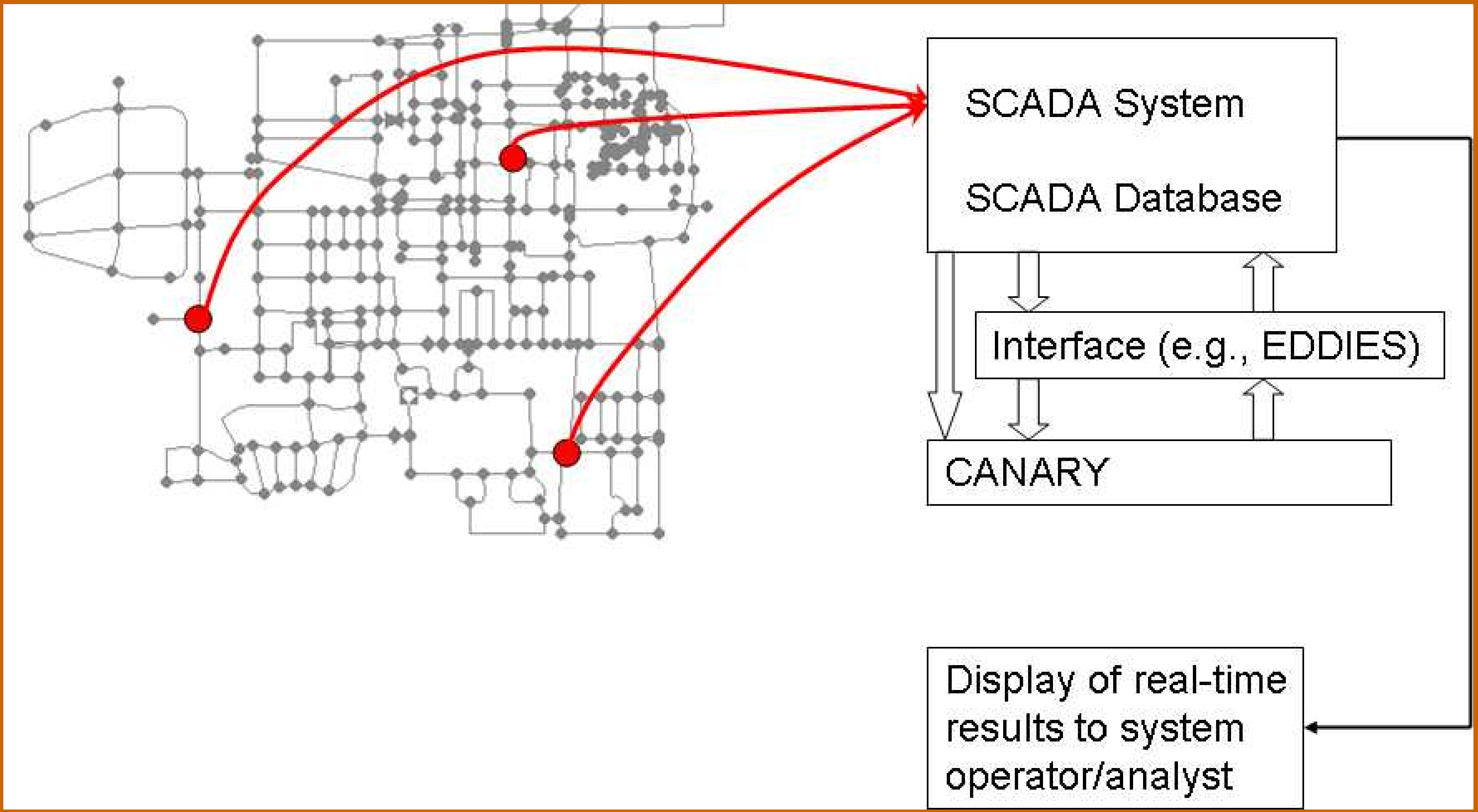
- In-situ monitoring of more basic water quality parameters is happening now and increasing all the time
- Can these indirect, or surrogate, parameters provide indication of adverse water quality?

Sandia National Laboratories

Monitoring Data: Dual-Use

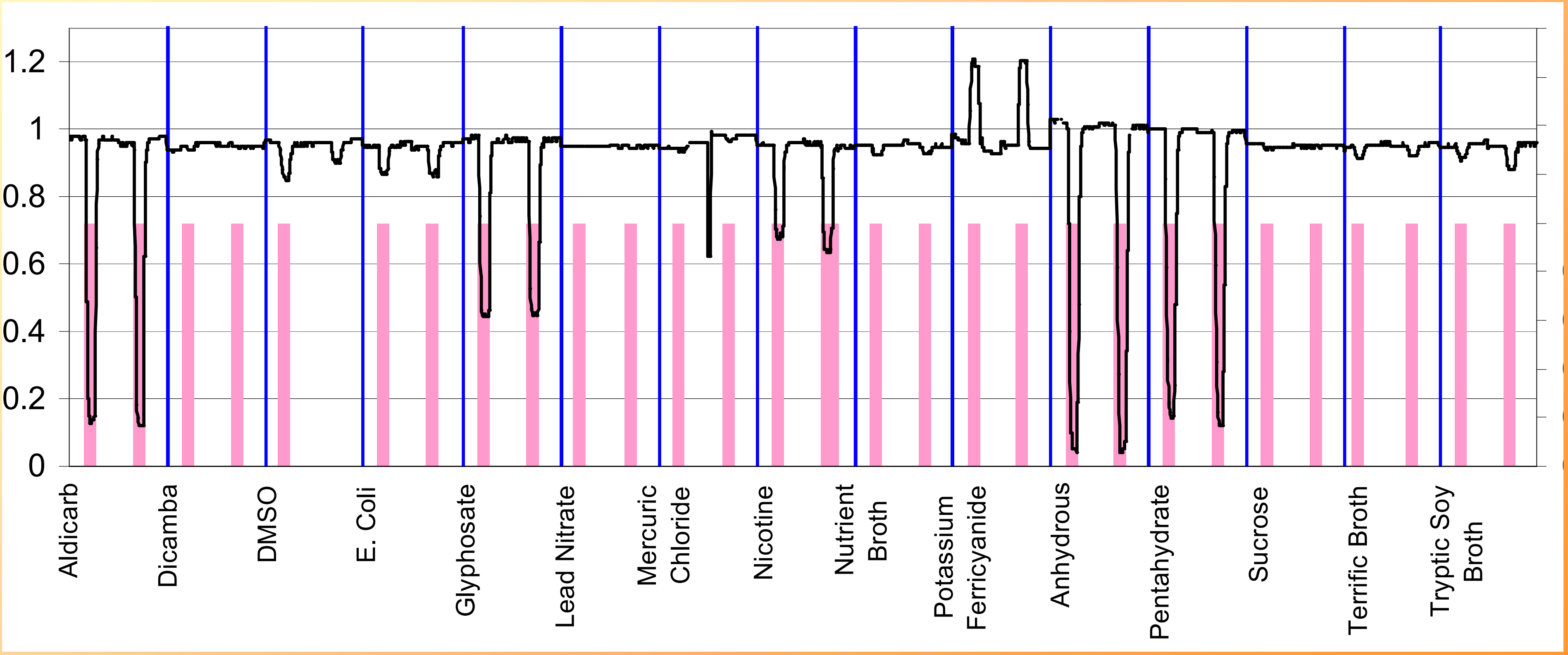
- Increasing amounts of on-line monitoring data are available through SCADA systems
  - Hydraulics (pressure, flow) and water quality
- Dual-Use (Security and Operations) benefits are achievable from these data
  - *A well-managed distribution network is a secure distribution network*

Network Monitoring



Surrogate Parameter Response

- Example responses of a free chlorine sensor to 15 different contaminants injected 24m upstream of the sensor



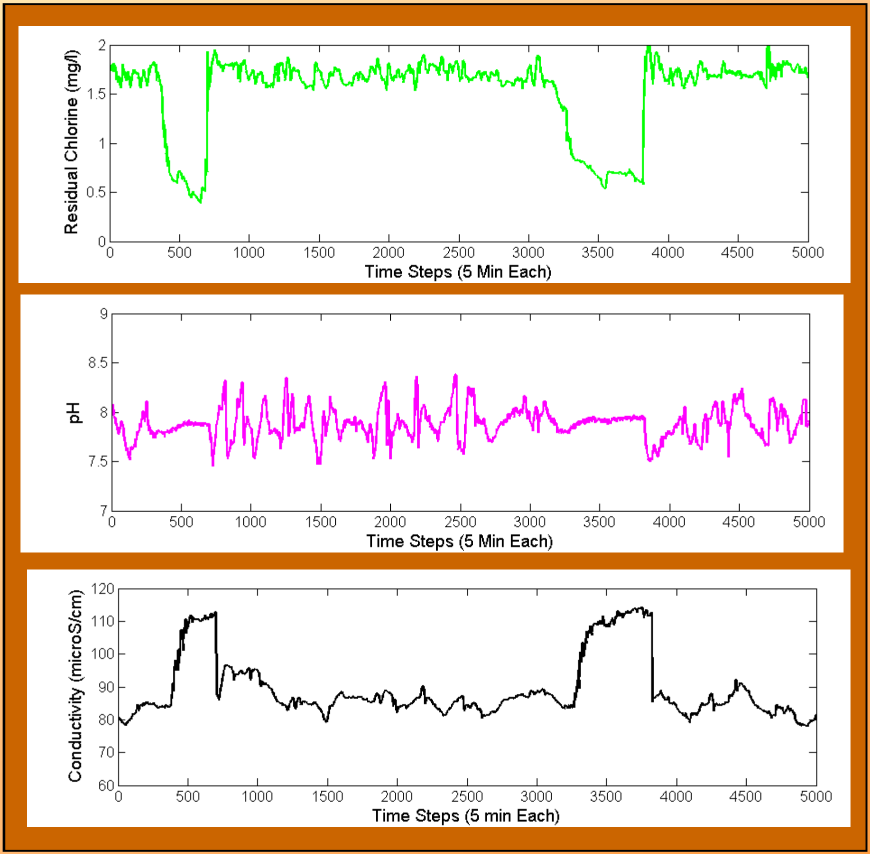
Surrogate Parameter Response

- In a study of nine different types of contaminants injected into a test loop, Hall et al. (2007) found:
  - "All of the contaminants injected caused at least one or more water quality parameters to change significantly"
  - Sensors that responded to the largest number of contaminants were: Specific conductivity, TOC, free chlorine, chloride and ORP

Hall, J., A.D. Zaffiro, R.B. Marx, P.C. Kefauver, E.R. Krishnan, R.C. Haught and J.G. Herrmann, 2007, On-line Water Quality Parameters as Indicators of Distribution System Contamination, Journal of the American Water Works Association, 99 (1), January

Event Detection: Complications

- Detecting adverse water quality events in network monitoring data is complicated by variations in background water quality:
  - What are we looking for?
  - Suppress false events caused by changes in operations



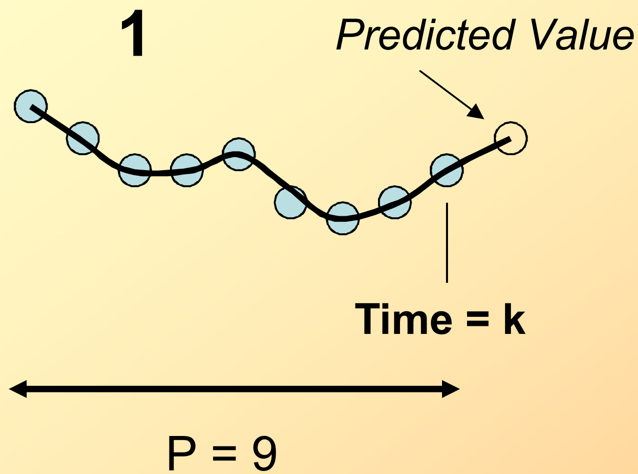
Sandia National Laboratories

Event Detection: Steps

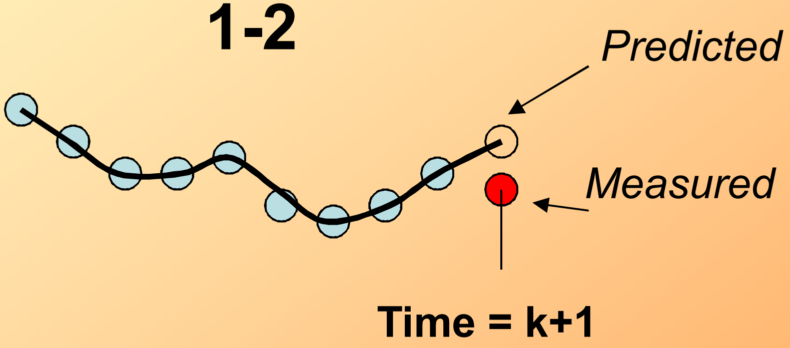
- Filter:
  - Use an adaptive filter to model background variations and predict next water quality value
- Compare:
  - Compare predicted and measured values for each time step (difference = residual)
- Combine:
  - Combine residuals across all water quality signals at a location to identify outliers in the data
- Aggregate:
  - Aggregate results across multiple time steps to determine the probability of an event occurring

Event Detection: Steps

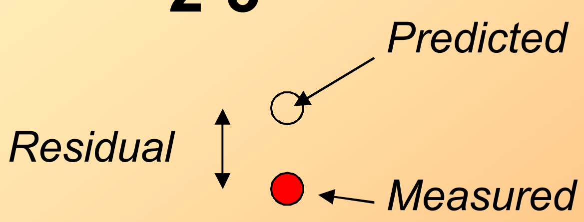
1



1-2

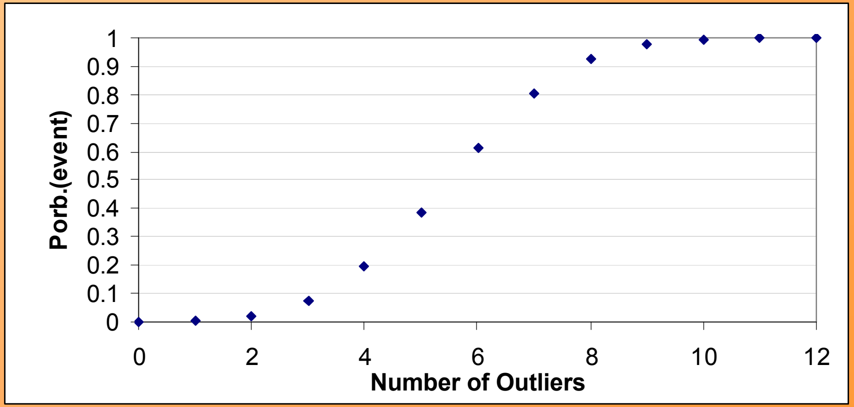


2-3



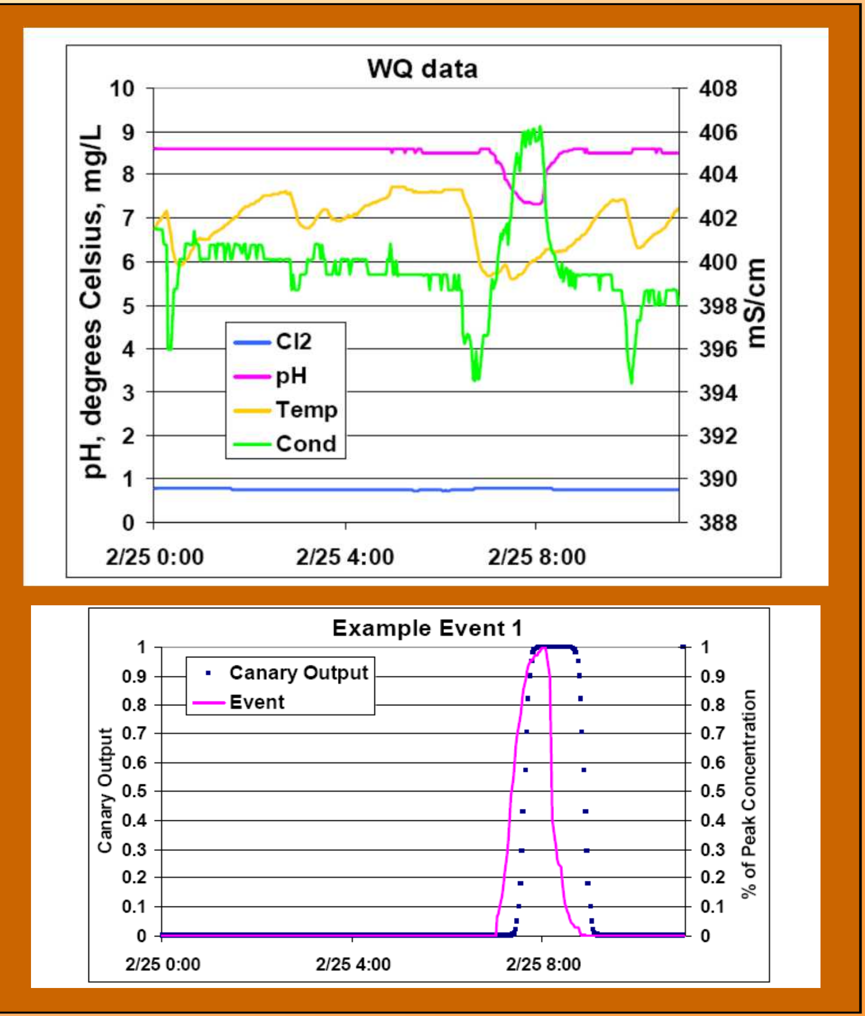
Compare the residual at each time step to a threshold. Those that exceed the threshold are “outliers”

4 Use binomial distribution to determine the probability of an event from the number of outliers over a given number of time steps



Example Event Detection

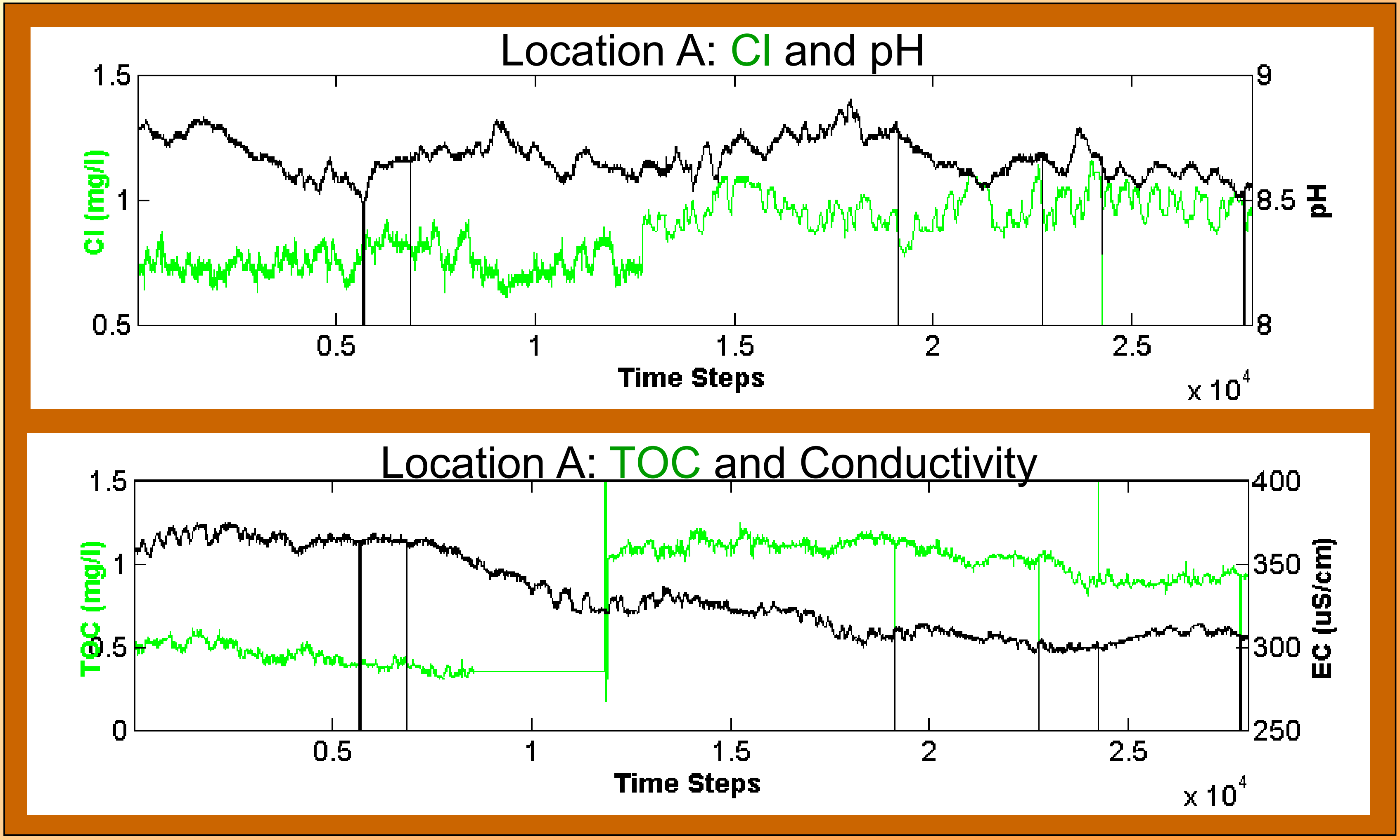
- Example event detection from a location in the US
  - Simulated event on top of measured water quality
  - Water quality signals shown at top
  - True event (magenta line) and CANARY response (blue squares) shown at bottom



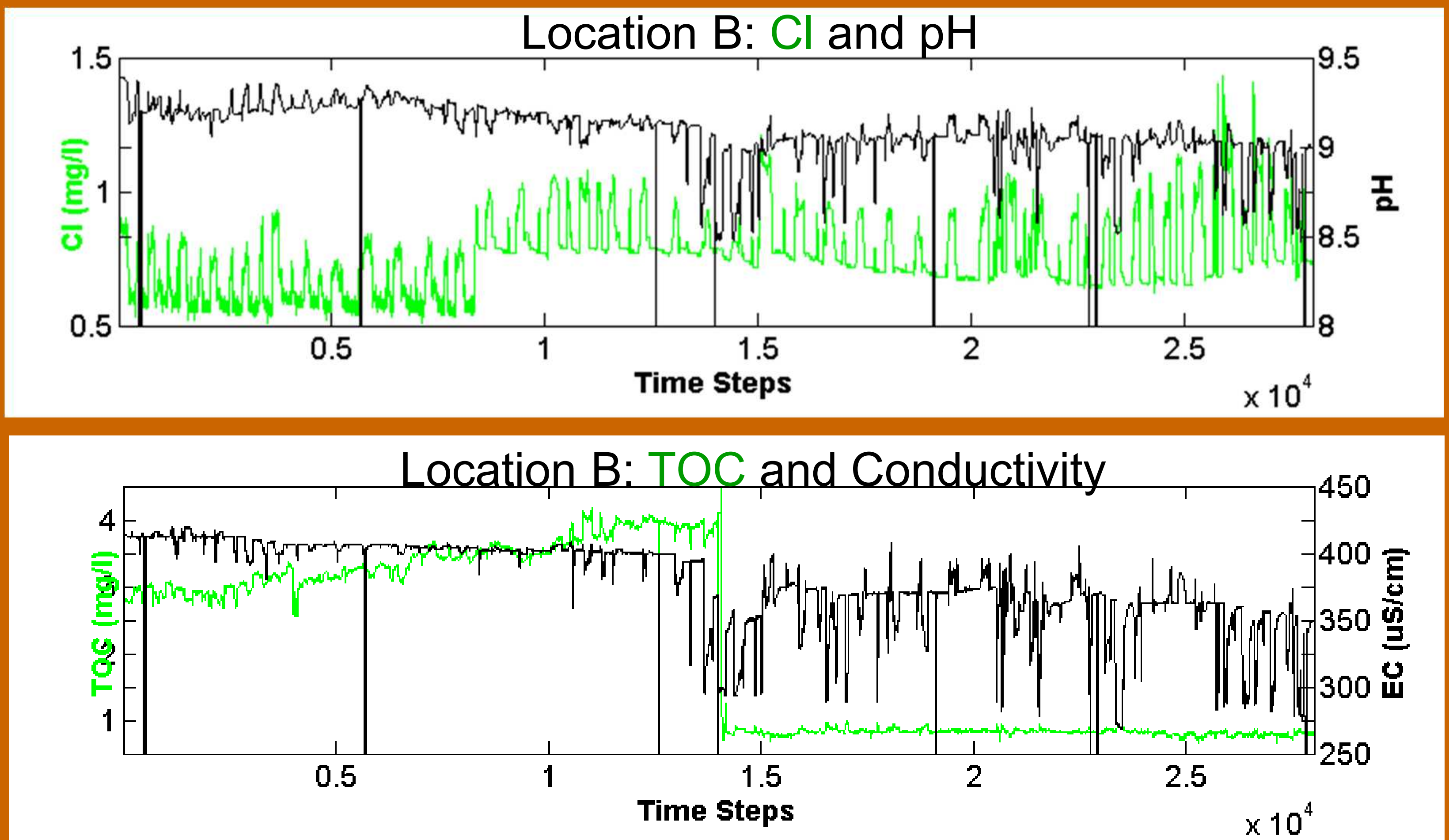
Case Study

- Examine two locations (A & B) in a distribution network in the USA
- Location B is strongly affected by the daily mixing of water from two different sources (groundwater and surface water)
- Available water quality data:
  - 2 minute sampling interval for 39 days (28,000 time steps)
  - Four signals: Cl, pH, Conductivity, TOC

Observed Signals: Location A

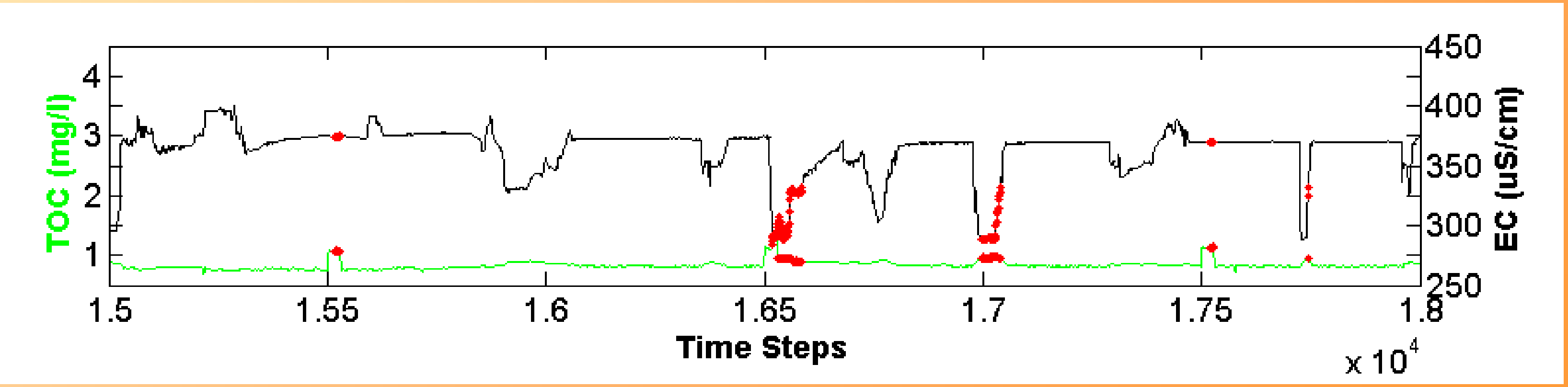


Observed Signals: Location B



Algorithm Training

- Background water quality conditions vary across the network
- Event detection algorithm parameters must vary to address background changes
- Examine results at both individual time steps and in terms of clusters of successive time steps



Sandia National Laboratories

Algorithm Training Results

Window Length (P)	Residual Statistics	Location A (σ)	Location B (σ)
P = 360	Mean (Std. Dev.)	0.427 (1.089)	2.079 (7.292)
P = 720	Mean (Std. Dev.)	0.169 (0.150)	0.463 (1.542)
P = 1080	Mean (Std. Dev.)	0.146 (0.123)	0.351 (1.146)
P = 1440	Mean (Std. Dev.)	0.135 (0.120)	0.292 (0.980)
P = 1800	Mean (Std. Dev.)	0.127 (0.118)	0.228 (0.720)

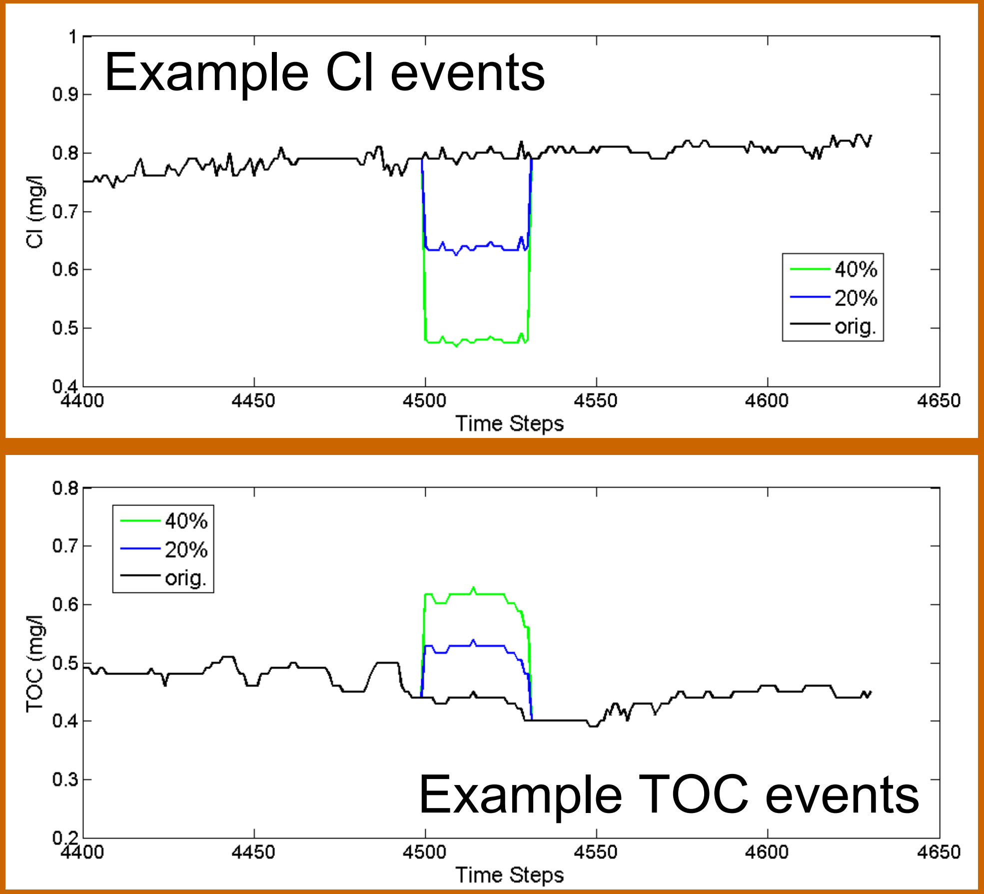
Window Length	False Positive Measures	Location A	Location B
P = 360	Time Steps (Clusters)	1182 (14)	3600 (54)
P = 720	Time Steps (Clusters)	8 (1)	1210 (25)
P = 1080	Time Steps (Clusters)	0 (0)	1032 (21)
P = 1440	Time Steps (Clusters)	0 (0)	695 (16)
P = 1800	Time Steps (Clusters)	0 (0)	557 (12)

Based on these results, window lengths of 1080 and 1800 were selected for locations A and B

Sandia National Laboratories

Adding Water Quality Events

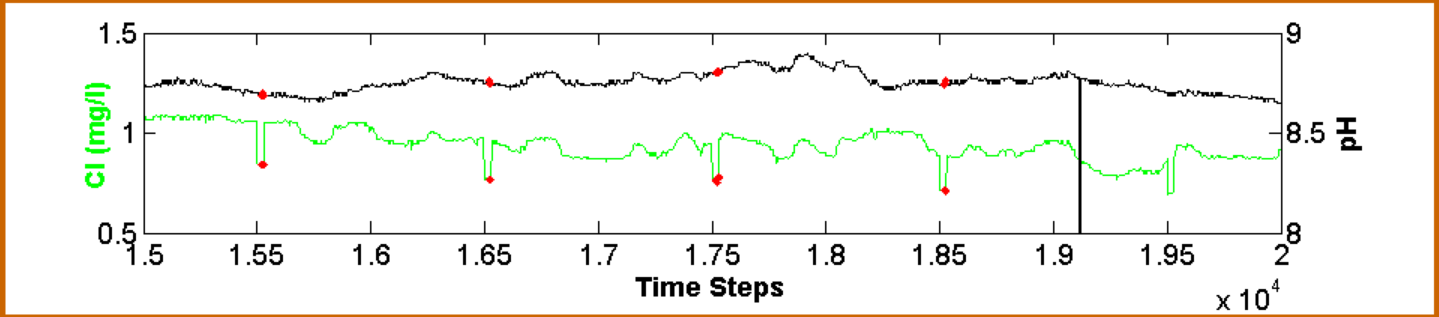
- Water quality events are added as deviations from the background measurements
  - Cl values decrease
  - TOC values increase
- Four event "strengths" are used corresponding to deviations of 20, 40, 60 and 80% of the background

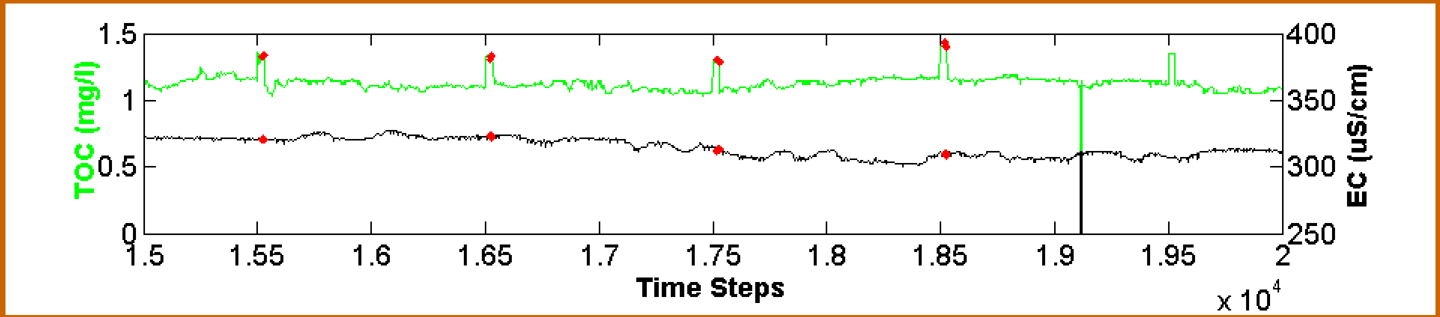


Length of each event is 1 hour (30 time steps)

Example Results: Location A

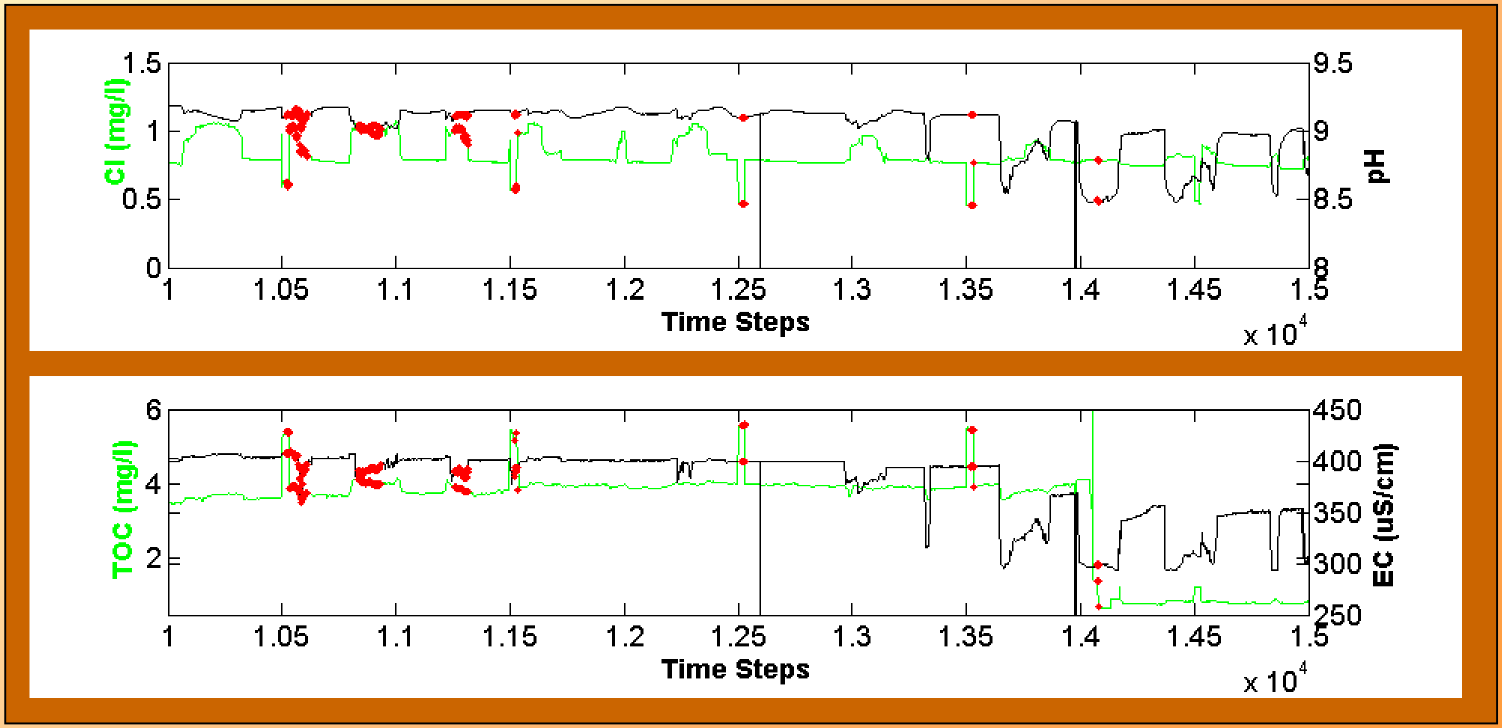
- Look at 5000 time steps (approximately 1 week) at Location A
- Events are marked as red dots





Example Results: Location B

- Look at 5000 time steps (approximately 1 week) at Location B
- Event strength = 40%
- Events are marked as red dots



Sandia National Laboratories

Evaluating Results

- A decision is made at every time step
- There are four possible results

	Estimated	Actual
– Correct Decision:	Backgrd	Backgrd
– Correct Decision:	Event	Event
– False Positive:	Event	Backgrd
– False Negative:	Backgrd	Event

Sandia National Laboratories

Results

- Greater than 99% correct at Location A and greater than 97% correct at Location B
- FN results decrease to 0.01% or 3 of approximately 27000 time steps examined

	Event Strength (%)	Correct (%)	FP (%)	FN (%)
Location A	20	99.13	0.49	0.38
	40	99.45	0.49	0.06
	60	99.50	0.49	0.01
	80	99.50	0.49	0.01
Location B	20	97.15	2.37	0.48
	40	97.37	2.43	0.20
	60	97.56	2.43	0.01
	80	97.56	2.43	0.01

Sandia National Laboratories

Summary

- On-line monitoring of surrogate parameters can be implemented now in the majority of distribution networks
  - Number of installed water quality monitors is increasing
  - Surrogate parameters react to the introduction of a broad range of contaminants
  - Processing of signals to recognize events above background variation is necessary

Sandia National Laboratories

Summary (Continued)

- Results of example application here show
  - Greater than 97% correct decisions at both locations
  - Reduction of false negatives to 0.01% for larger event strengths
- Future Work
  - Improved recognition of expected changes resulting from utility operations
  - Automated methods for setting algorithm parameters at each monitoring location
  - Distributed Detection: Integrating event detection results from multiple monitoring locations

Acknowledgements

- This work was support through an Interagency Agreement between Sandia National Laboratories and the US EPA National Homeland Security Research Center

- Contact: Sean A. McKenna samcken@sandia.gov