SAND2009-2445P

Sandia National Laboratories

Infrastructure Monitoring for Improved Security and Operations

Dr. SEAN A. McKENNA

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

Outline

- Background and Motivation
  – Problem (Intentional and Accidental Contamination Events)
  – Tools for Solution (Numerical Models, Statistics, Data Mining)
- Pre-Event Monitoring
  – Optimal Monitoring Locations
  – Filtering Background Variation
- Response During an Event
  – Real-Time Source Location Inversion
- Post-Event Restoration
  – Characterization of Magnitude and Extent
  – Is Restoration Complete

Infrastructure Monitoring

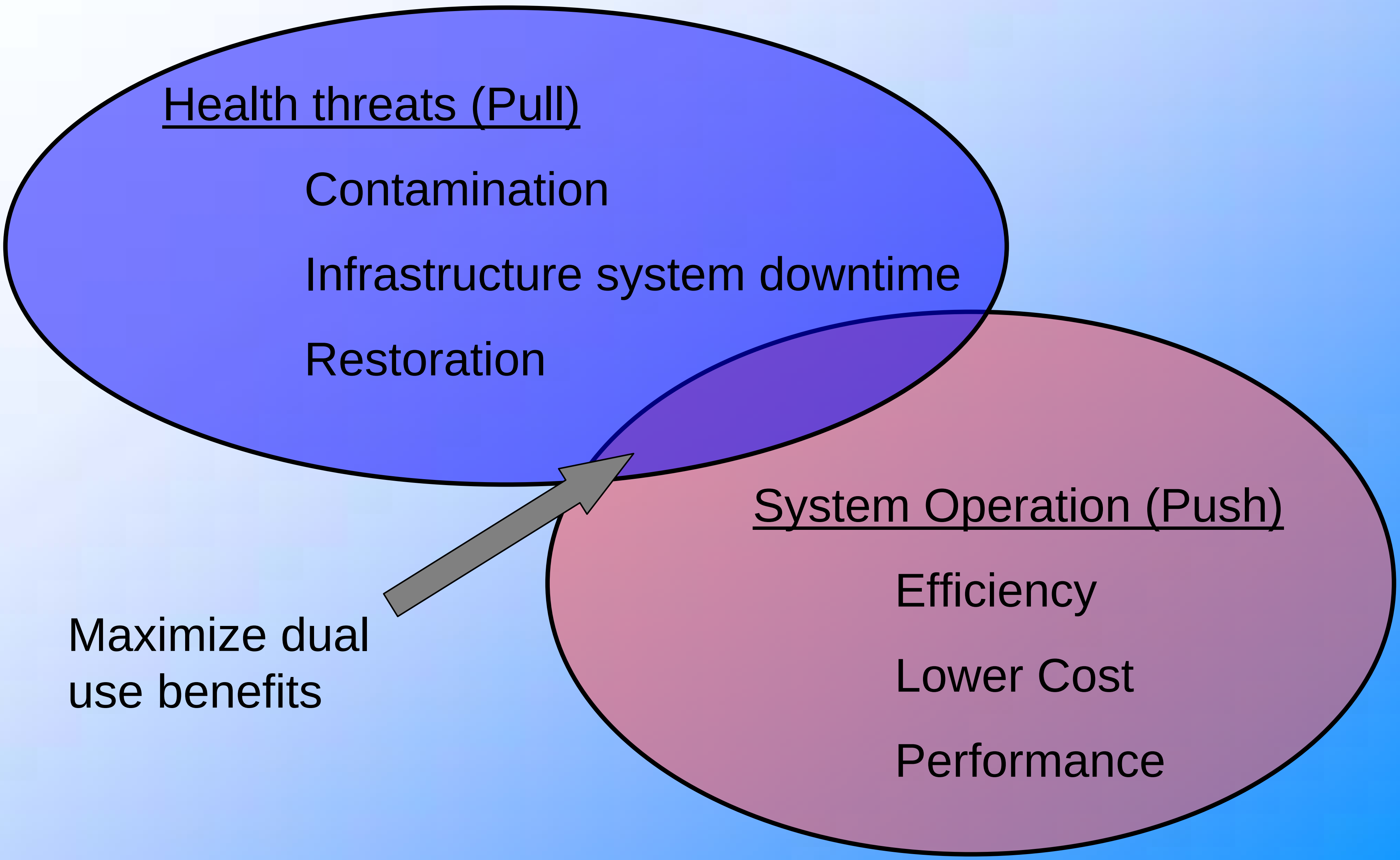
- Pre-event Planning
  - Risk Assessment
  - Monitoring Objectives
  - Optimal Monitoring Locations
  - Monitoring Sensors and Comm.
  - Background Characterization
  - Tuning Monitoring Algorithms
- Event Response
  - Source Location Identification
  - Operational Changes
- Post-Event Restoration
  - Characterization
  - Decontamination
  - Confirmation Sampling
  - Waste Disposal

Problem / Motivation

Large areal extent and customer focused design can create vulnerabilities in water distribution networks

Intentional attacks on US Gov't office buildings and postal facilities in Sept and Oct., 2001 led to costly restoration projects

Dual-Use Applications

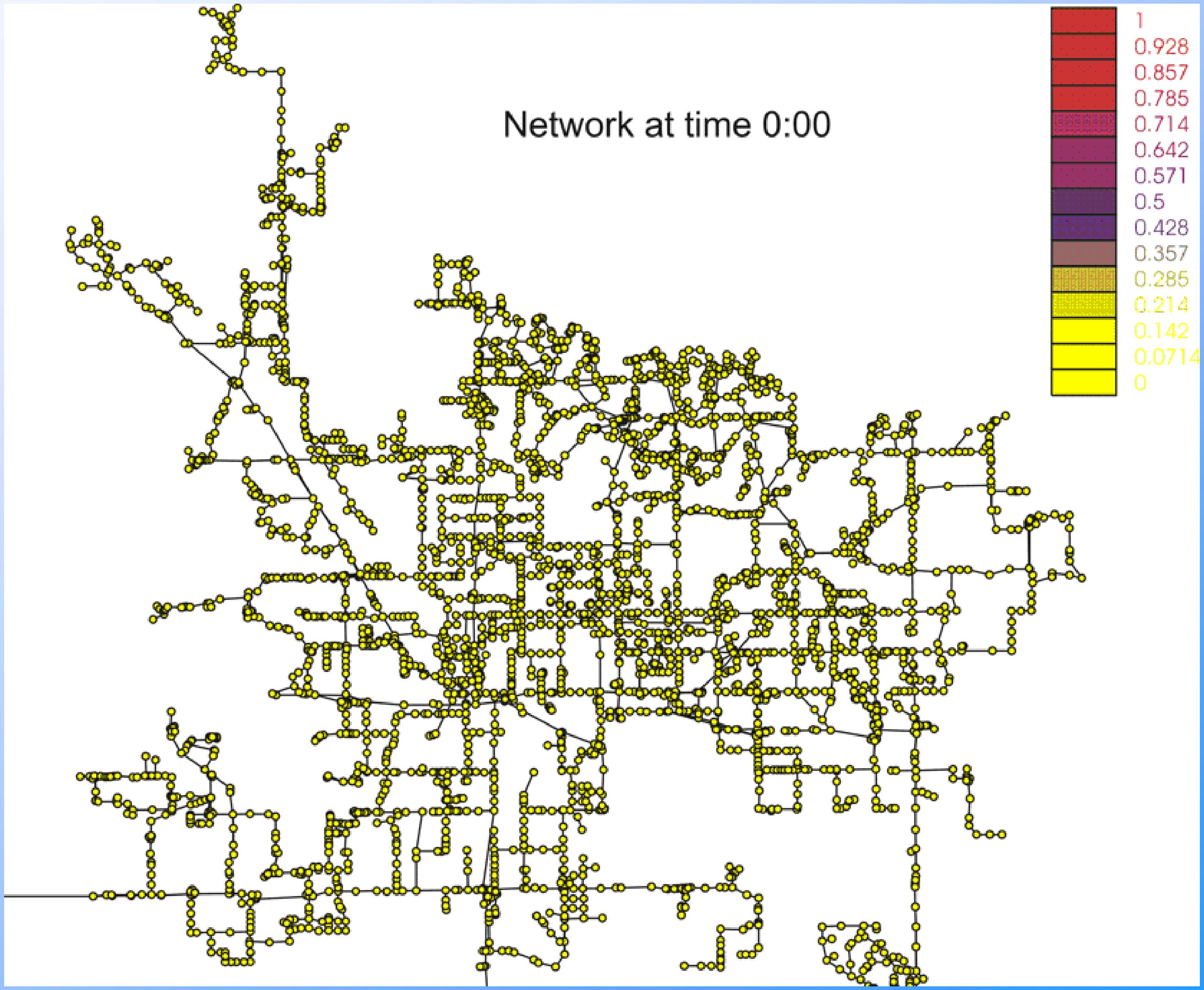


Outline

- Background and Motivation
  - Problem (Intentional and Accidental Contamination Events)
  - Tools for Solution (Numerical Models, Statistics, Data Mining)
- Pre-Event Monitoring
  - Optimal Monitoring Locations
  - Filtering Background Variation
- Response During an Event
  - Real-Time Source Location Inversion
- Post-Event Restoration
  - Characterization of Magnitude and Extent
  - Is Restoration Complete

Contamination Simulation

Unknown source location, strength and timing as well as unknown fine-scale demand patterns make it nearly impossible to predict results of contamination event



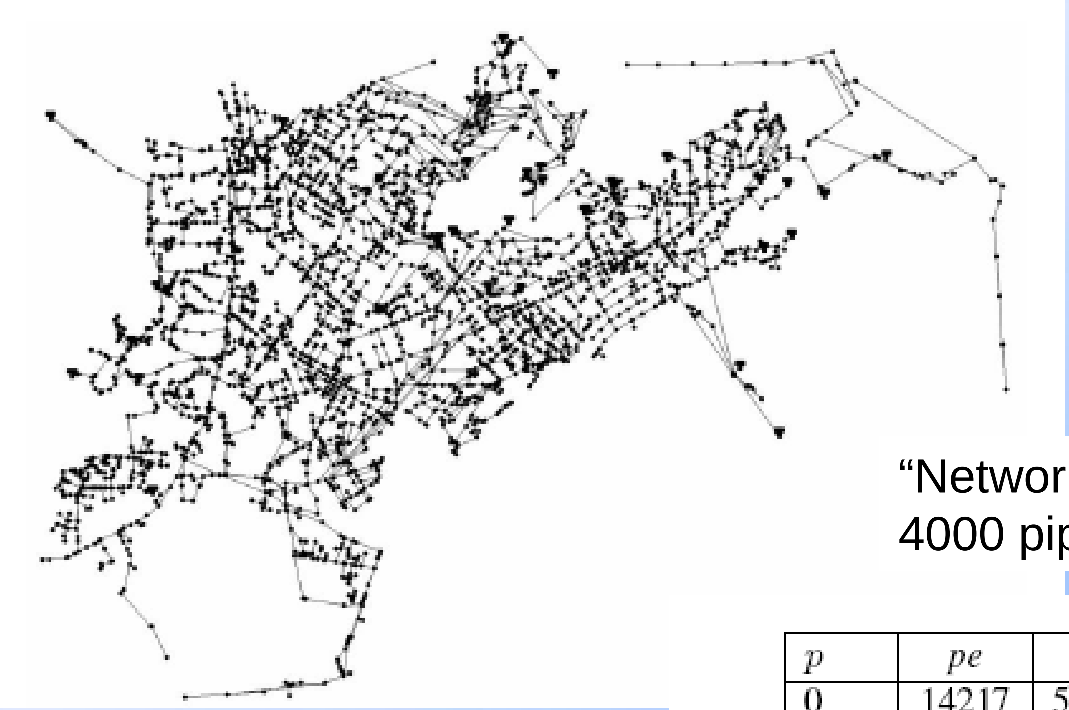
Sandia National Laboratories

Location Optimization

- Discrete optimization on a network
  - Multiple competing and/or overlapping objectives
  - Exact solutions vs. approximate solutions through heuristic solvers
  - Imperfect sensors, uncertain demands, changing over time

2008 Finalist INFORMS Franz Edelman Prize for Achievement in Operations Research and the Management Sciences – US EPA "Reducing Security Risks in American Drinking Water Systems"

Location Optimization: Example



“Network 2”: 3000 junctions, 4000 pipes, 50 tanks/reservoirs

p	pe	td	vc	mc	nfd	ec
0	14217	5760	11667200	3.90e+13	1.0	344376
5	1709	3218	162640	2.71e+13	0.47	38822
10	1061	2860	66241	2.44e+13	0.41	22062
50	347	2028	13675	1.74e+13	0.29	6382
100	205	1632	7549	1.42e+13	0.23	3604
500	50	124	1527	2.51e+12	0.0	754
1000	14	11	272	7.18e+10	0.0	84
2000	0	0	0	0	0.0	0

Table 2: Optimal values of design objectives for a range on p on Network2.

P = number of sensors

pe = people exposed

td = time to detection (minutes)

vc = volume consumed (gallons)

mc = mass consumed (Organisms)

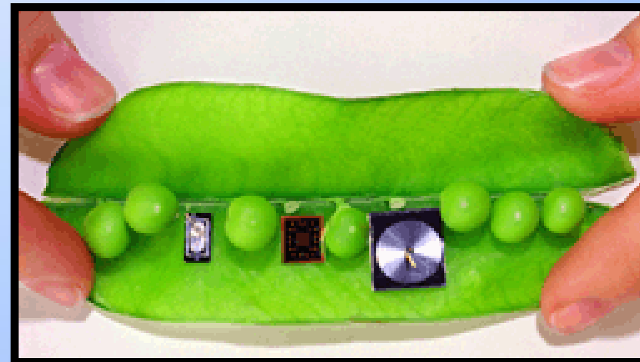
nfd = Proportion of events not detected

ec = length of pipe contaminated (feet)

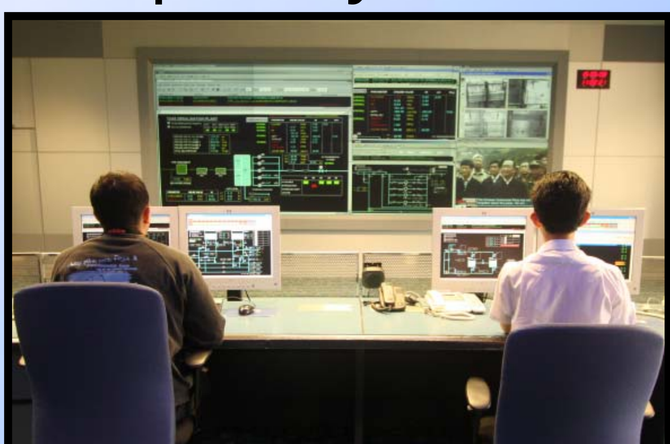
From: Watson, Hart, Geenberg (2005), ASCE EWRI

Water Quality Monitoring

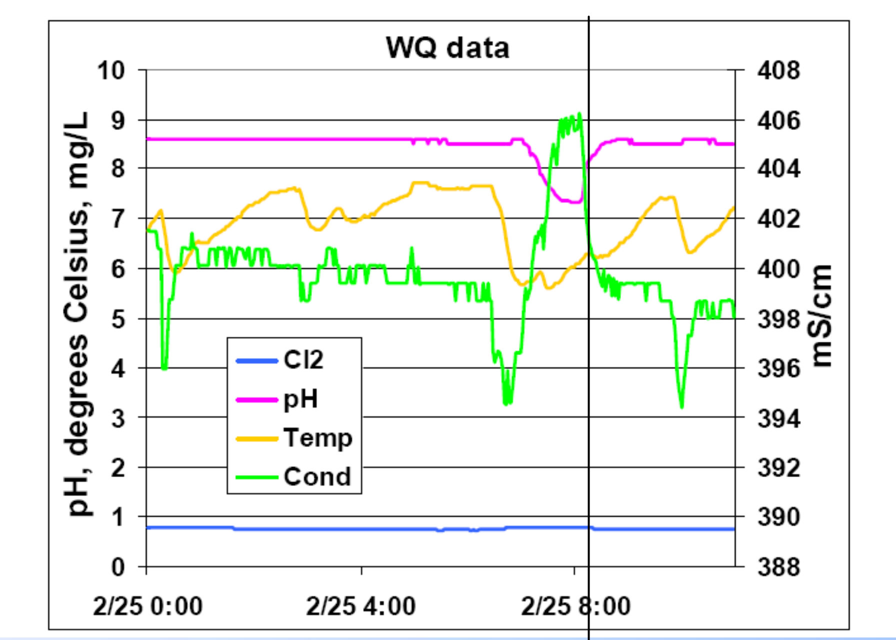
- The Goal: Inexpensive, robust, networked, compound specific, in-situ capability

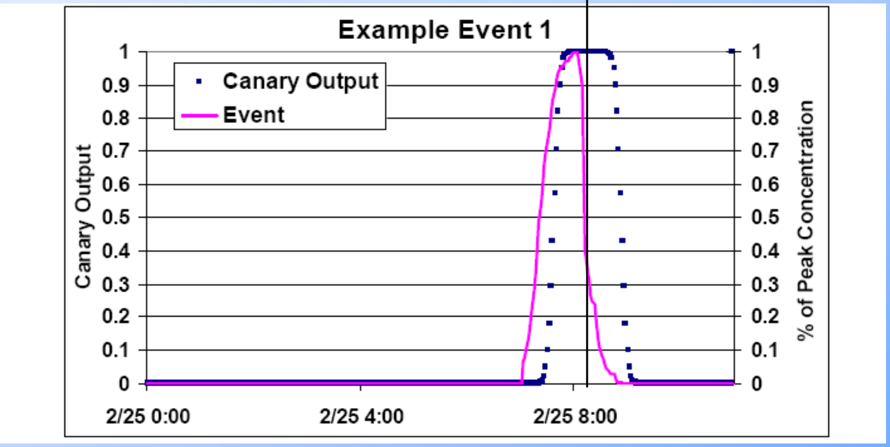


- The Reality: Expensive, robust, SCADA connected, basic water quality sensors with in-situ capability



CANARY Goals





Goal: Develop a tool that can analyze multivariate water quality signals in real-time and provide identification of periods of anomalous water quality

Example application of CANARY in a US utility

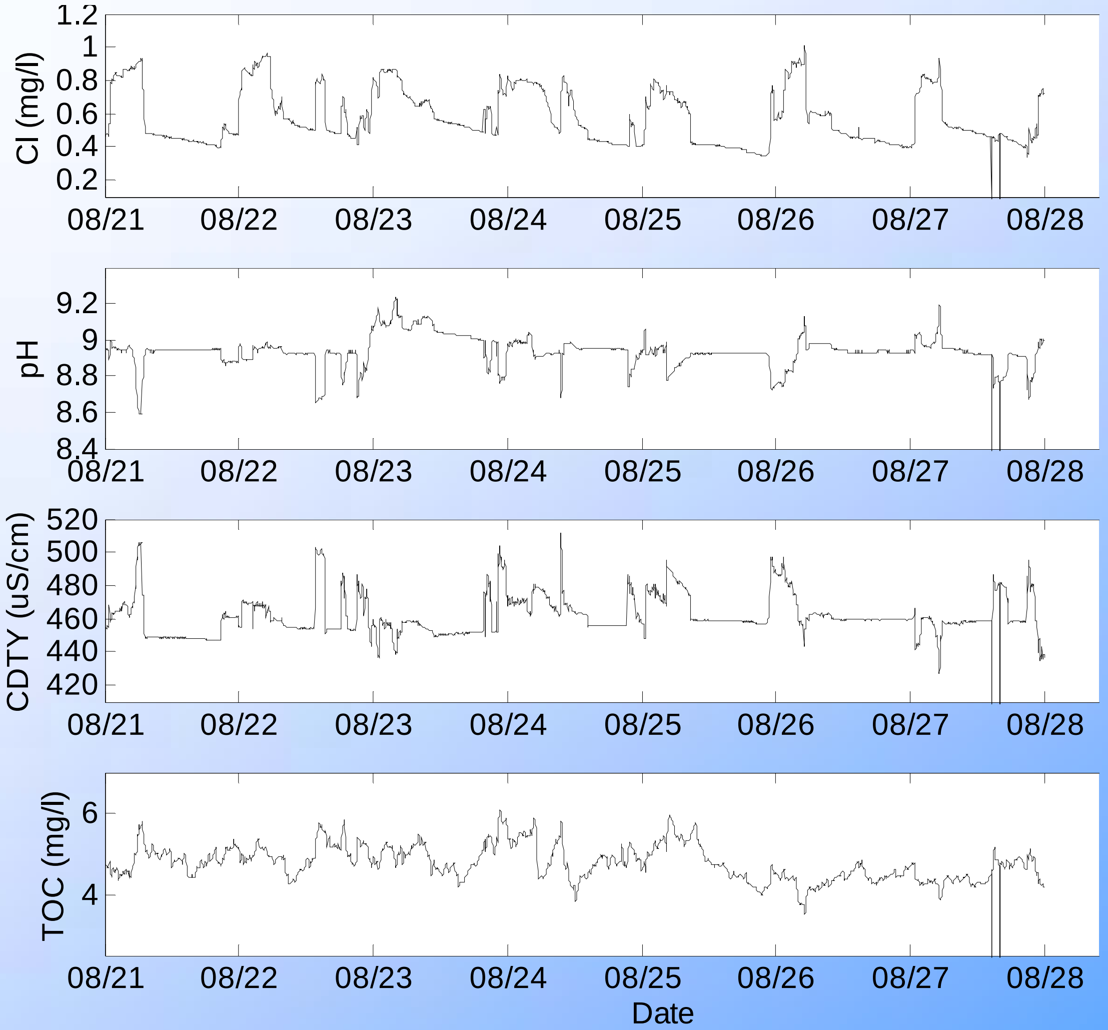
Four water quality signals

Results show unknown % of peak contamination and P(event) as identified by CANARY

Note the long stretch of zero probability prior to the event

Lag time in detection is adjustable

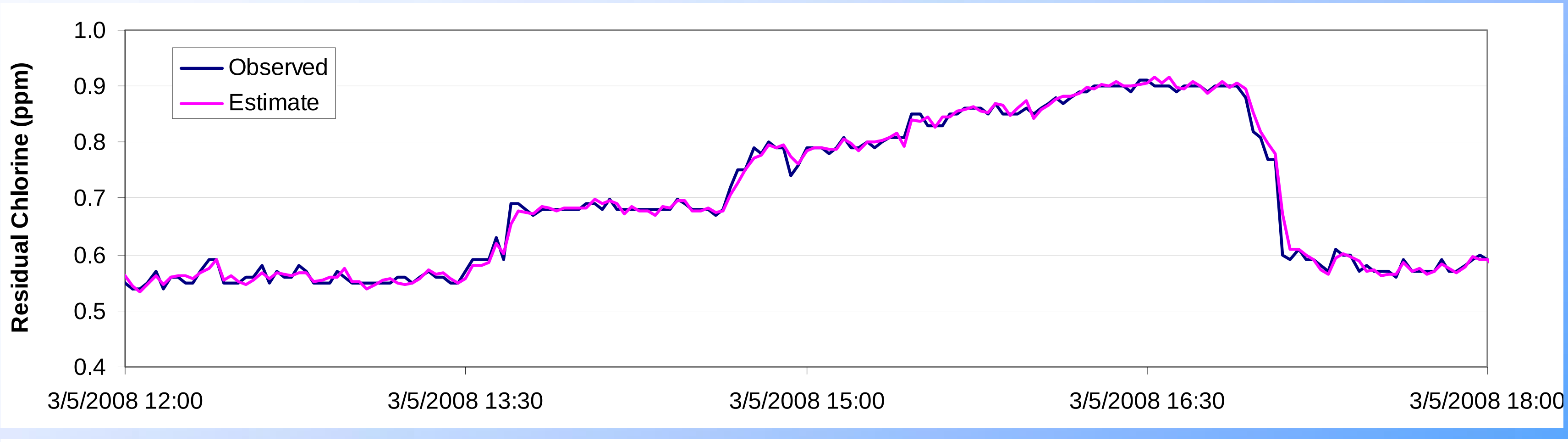
The Challenge

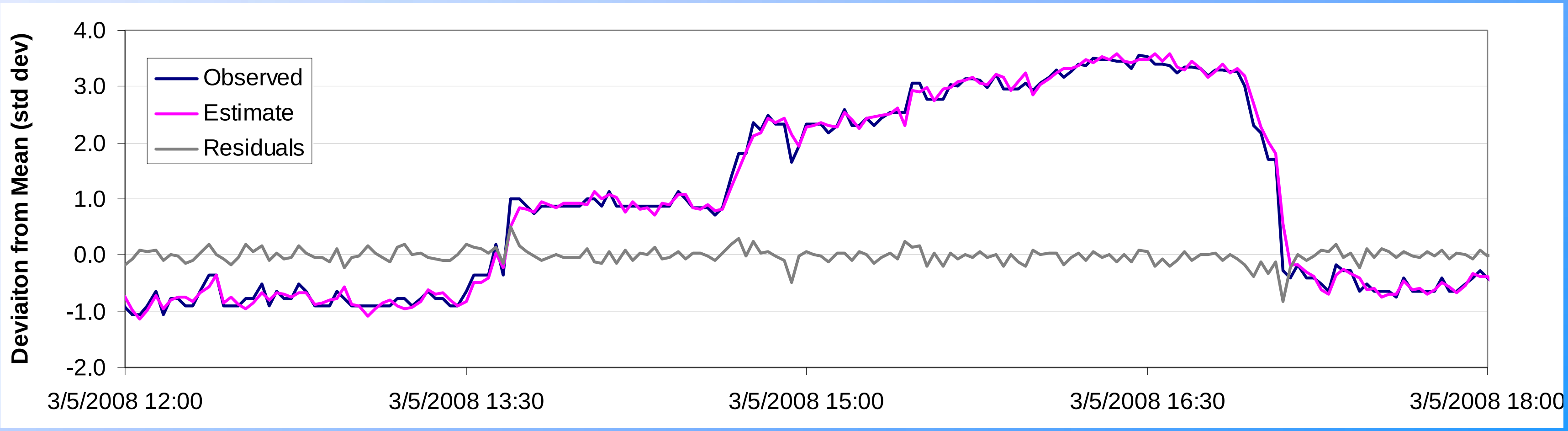


- Significant variation in background water quality within distribution systems
- Changes are on the order of changes caused by introduction of contaminants
- Causes include: noise in SCADA systems, changes in utility operations, mixing of multiple source waters in the network, etc.

Approach

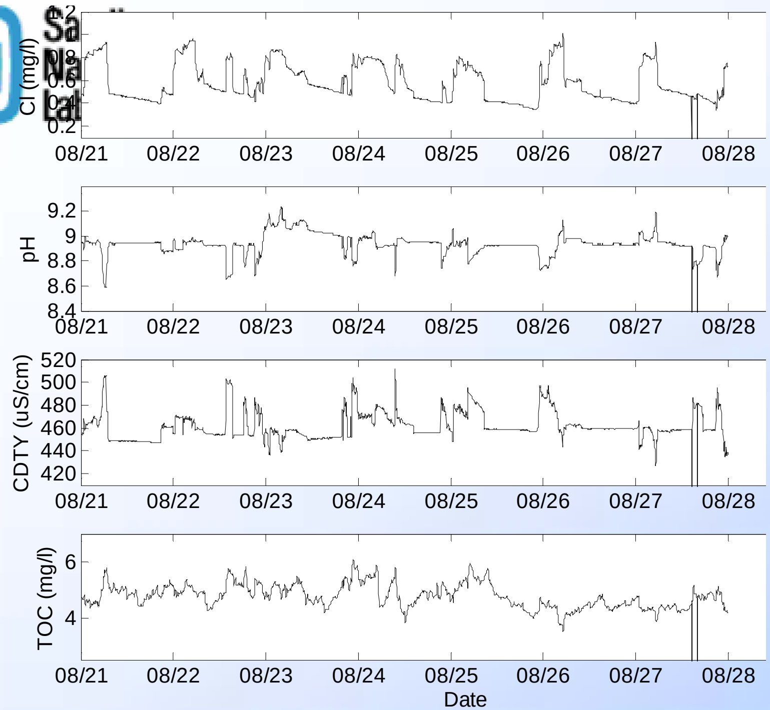
Identify a model with parameters optimized to predict next value of water quality. Look for significant differences (residuals) between prediction and next observation. Example application to 6 hours of chlorine data



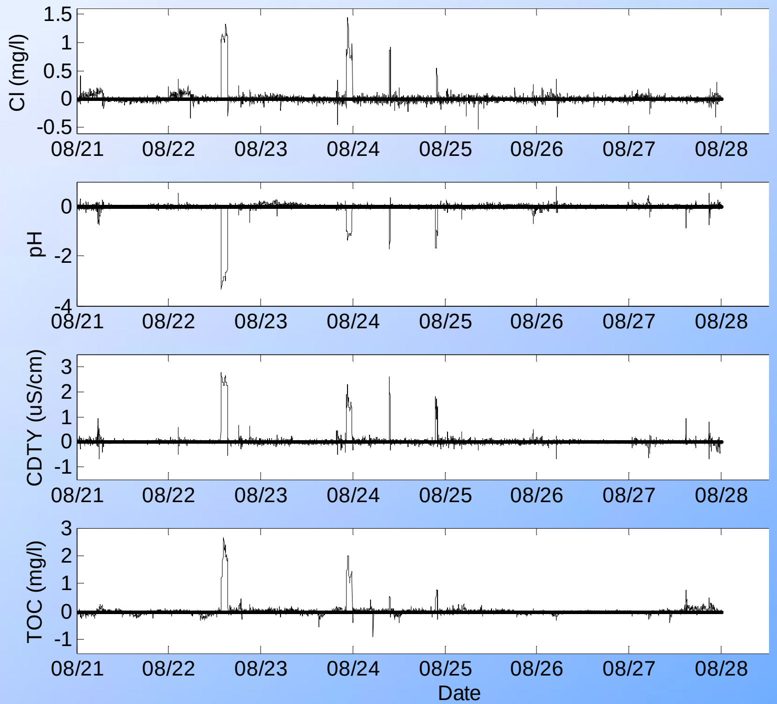


Signal Filtering

Raw Water Quality Data



Residuals (Obs – Predict)



Filtering algorithm removes expected background variation from data. Differences between expected and observed water quality (residuals) are retained

Residuals have units of standard deviations

Residuals are compared to a threshold and those exceeding the threshold are considered “outliers” otherwise “background”

How many outliers before an alarm is sounded? Or, how many outliers constitute an event?

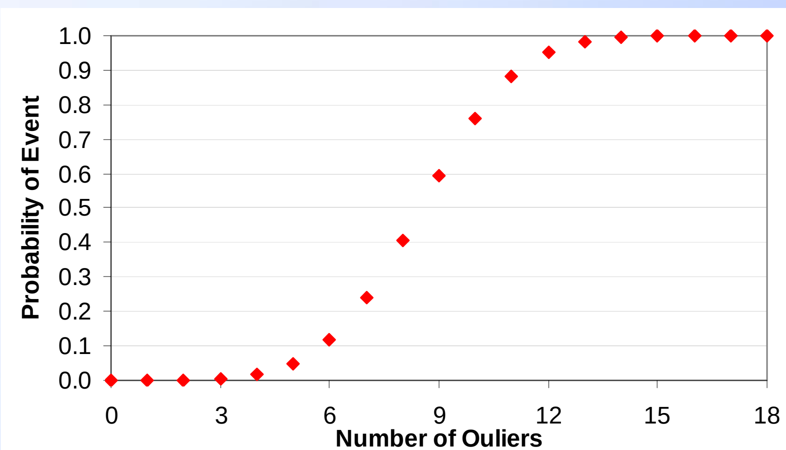
How would traditional approach of set point thresholds perform here?

Residual Aggregation

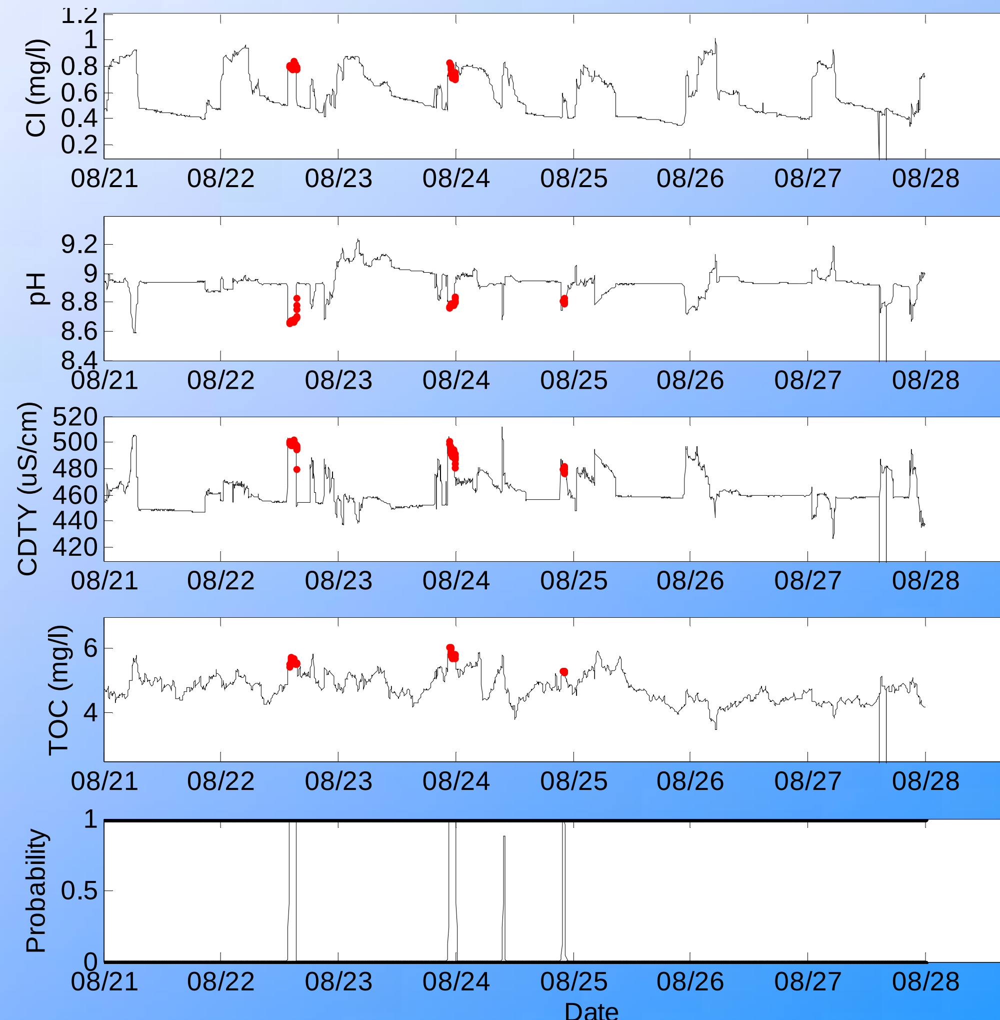
Binomial Event Discriminator (BED)

Each time step is a trial, *N*, outliers are considered failures, *R*, and there is a constant probability of failure at each time step:

P(event) = b(R; N,pfail)



Binomial failure model provides flexible means of calculating *P(event).* Compare *P(event)* to a probability threshold at every time step

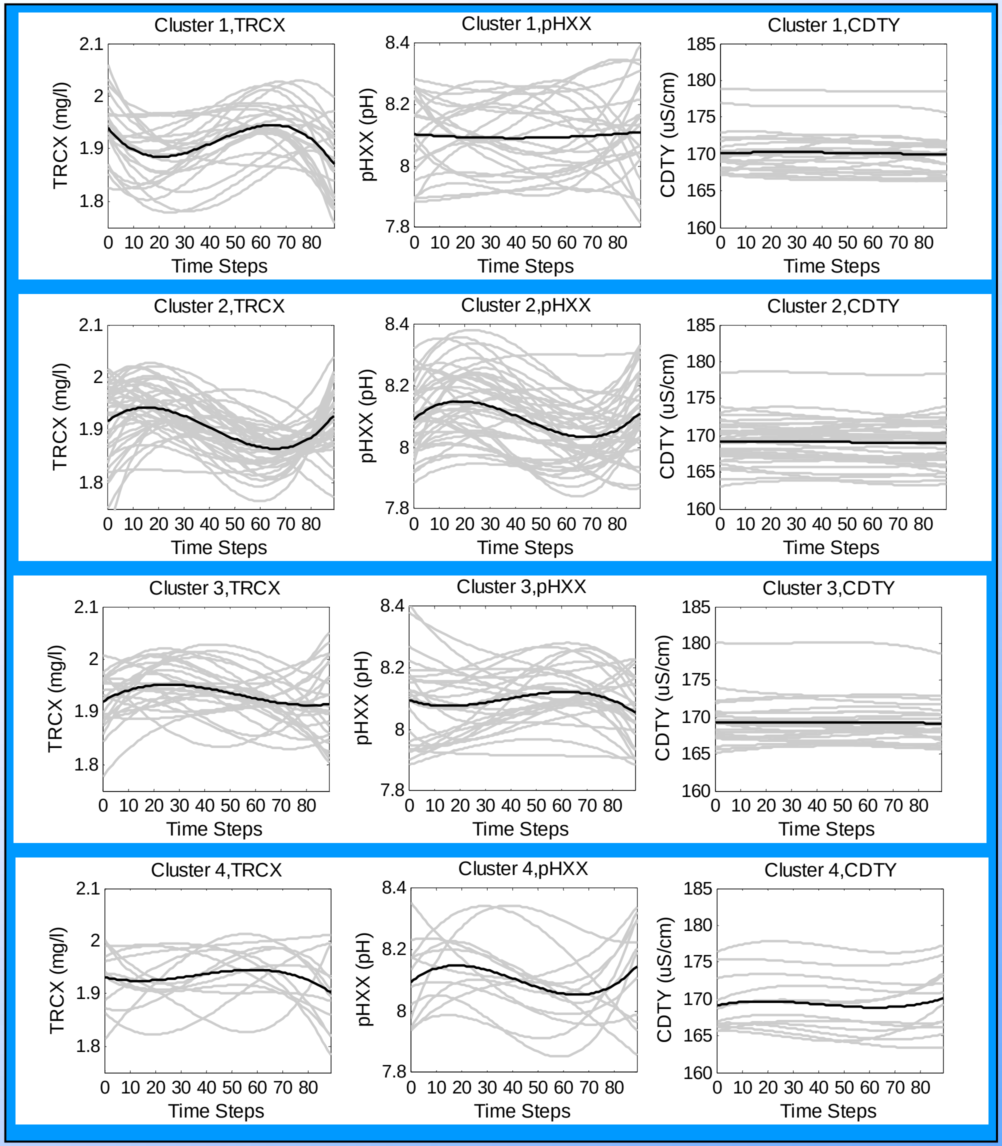


Outstanding Challenge

- False Positive Reduction
  - Changes in network operations cause changes in water quality that can lead to false alarms
  - Two ideas we are working on to reduce these:
    - Multivariate pattern matching for water quality
    - Integrating results from multiple stations into “network-wide” detection

Sandia National Laboratories

Pattern Matching



4 clusters (rows), 3 signals (columns)

Each measured pattern is shown along with the mean pattern for each identified cluster

Example application:

Build patterns using Jan 1st to April 30th data (120 days)

Test on data from May 1st to Aug 15th (100 days)

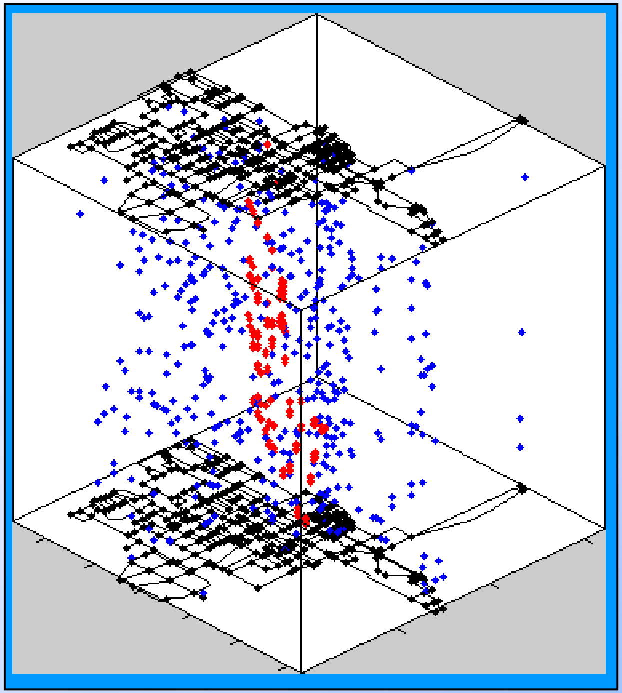
Without pattern recognition: 100 alarms sounded

With pattern recognition 14/11 alarms sounded – reduction of 86/89 percent

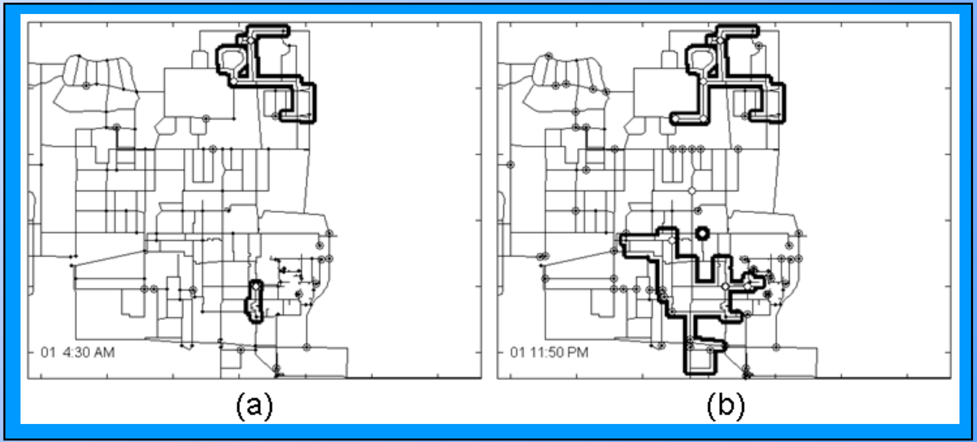
Vugrin, McKenna and Hart, 2009, Trajectory Clustering Approach for Reducing Water Quality Event False Alarms, ASCE Annual EWRI Conference, Kansas City, Missouri

Distributed Detection

- Goal: integrate multiple independent monitoring stations for *network-wide* event detection



Consider alarms to be a random point process in space and time and develop scan test to identify significant clusters of those alarms



Clusters identified under scenario of two simultaneous contamination events shown at time of detection (a) and 24 hours after injection (b)

Koch & McKenna, 2008, Distributed Network Fusion for Water Quality, ASCE World Environmental and Water Resources Congress, Honolulu, Hawaii, May 13-16.

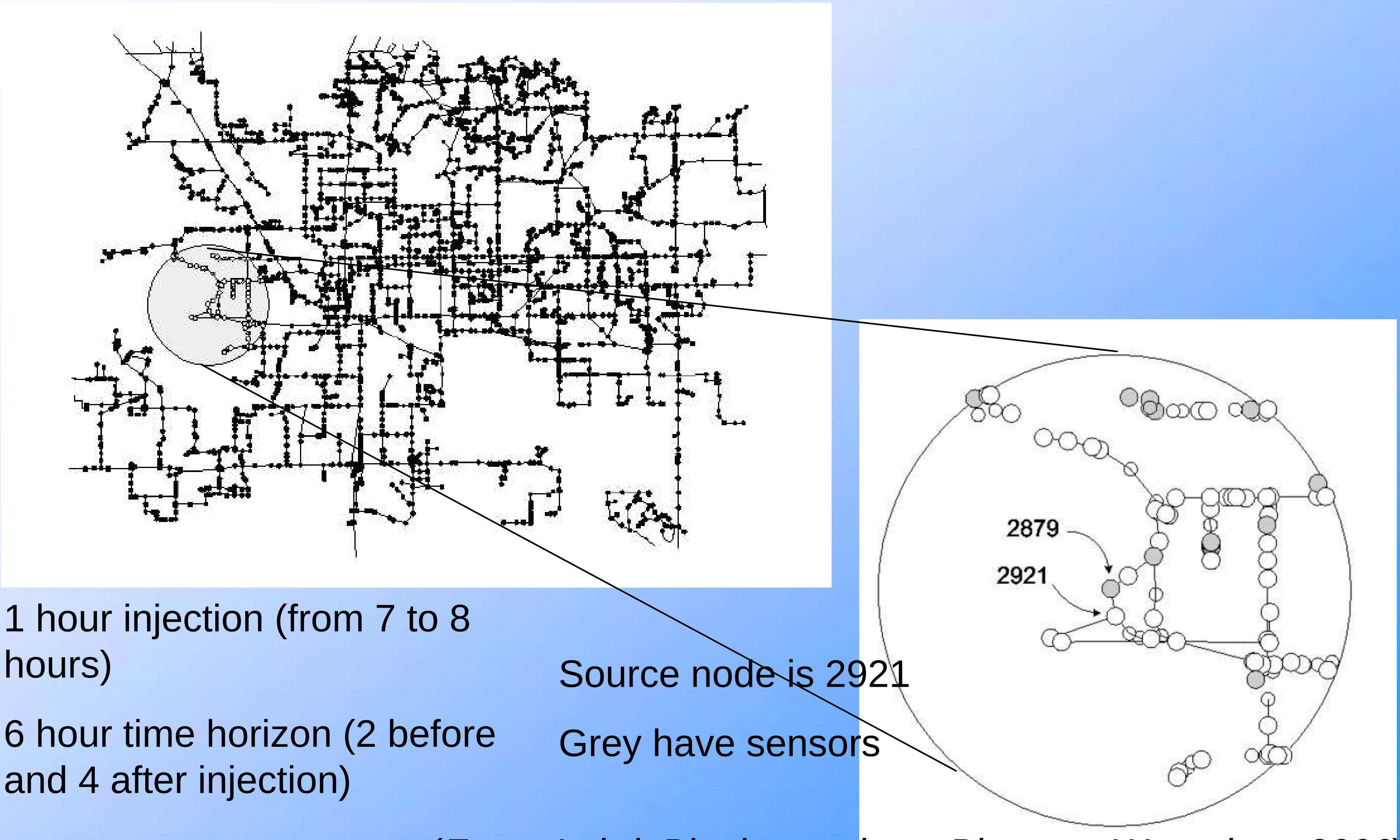
Sandia National Laboratories

Outline

- Background and Motivation
  - Problem (Intentional and Accidental Contamination Events)
  - Tools for Solution (Numerical Models, Statistics, Data Mining)
- Pre-Event Monitoring
  - Optimal Monitoring Locations
  - Filtering Background Variation
- Response During an Event
  - Real-Time Source Location Inversion
- Post-Event Restoration
  - Characterization of Magnitude and Extent
  - Is Restoration Complete

Example (Large Network)

Approximately 3500 nodes, 350 randomly placed sensors



1 hour injection (from 7 to 8 hours)

6 hour time horizon (2 before and 4 after injection)

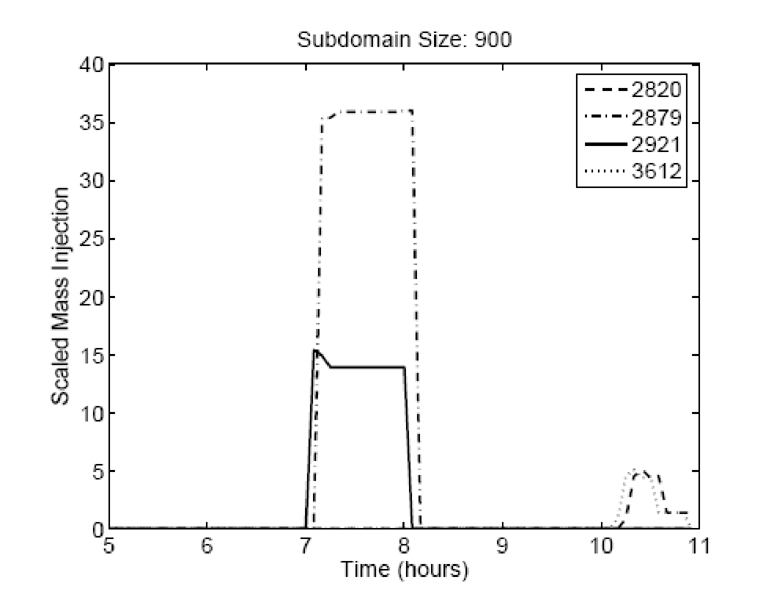
Source node is 2921

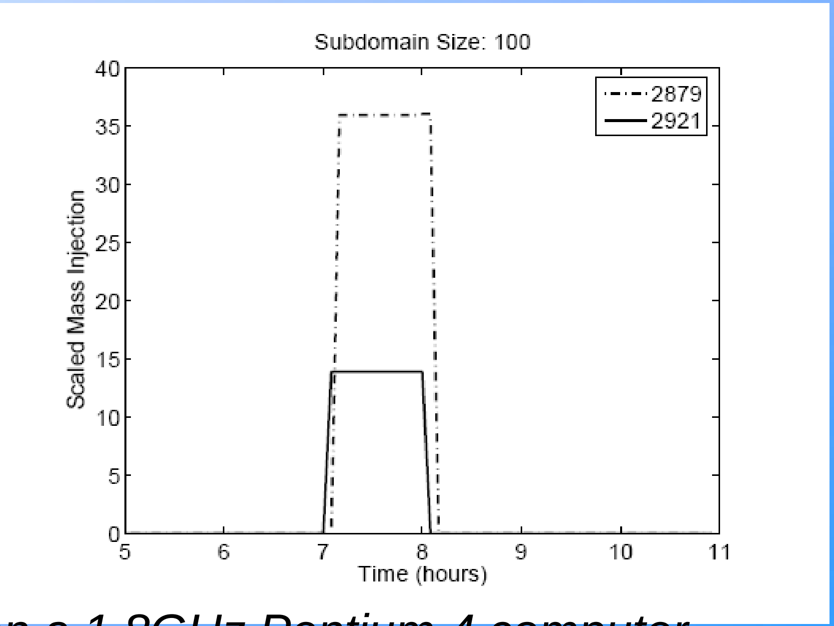
Grey have sensors

(*From: Laird, Biegler, and van Bloemen Waanders, 2006*)

Large Network: Results

- Automatically break large network into smaller subdomains around sensor locations with contamination
- Solution determines how much mass came from every node in the subdomain and what time it was injected



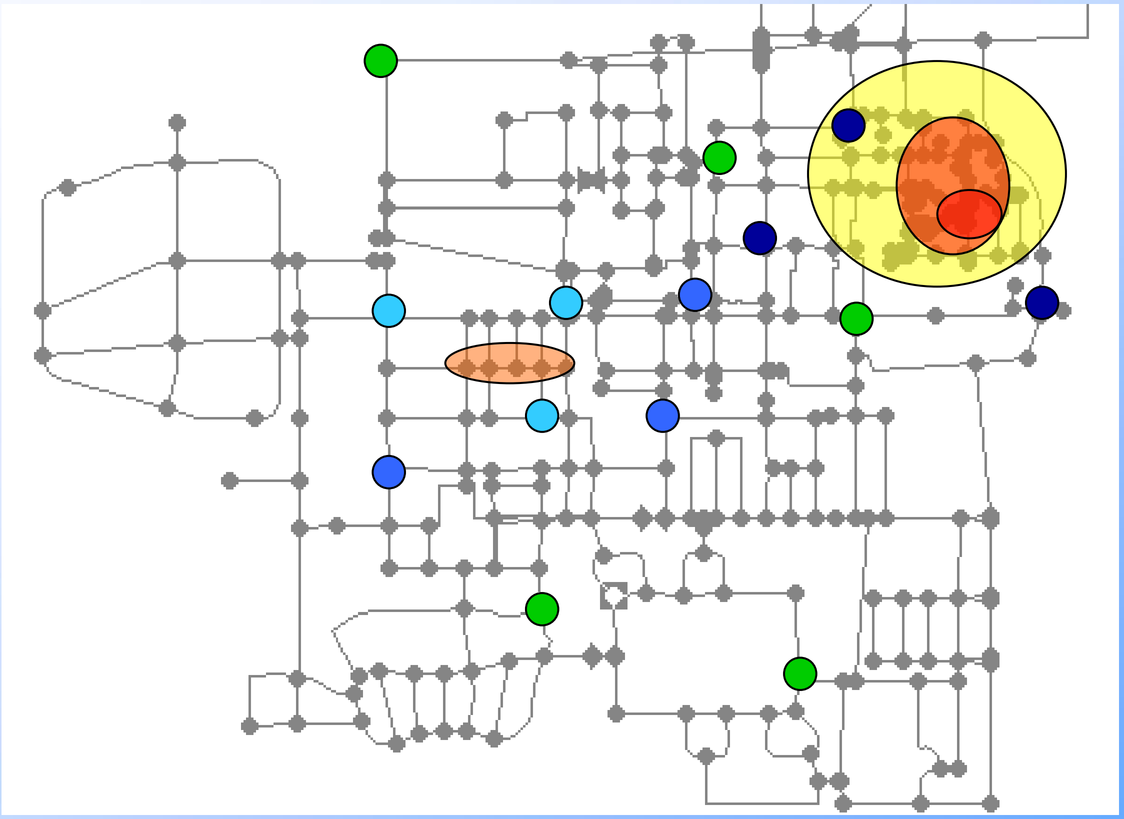


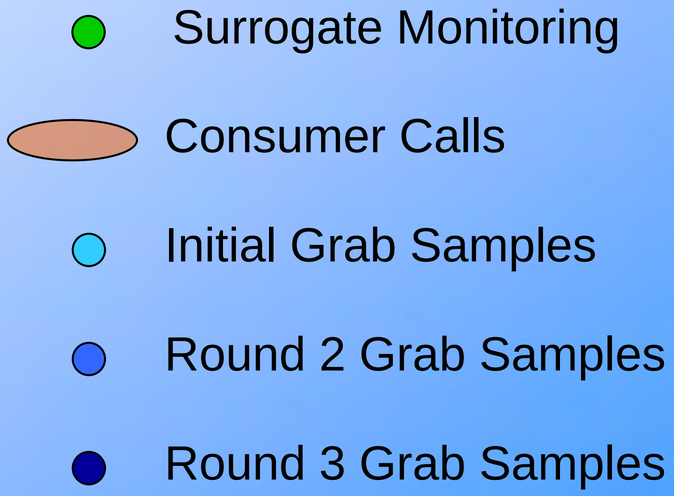
5 minute time steps, real-time solutions on a 1.8GHz Pentium 4 computer

(*From: Laird, Biegler, and van Bloemen Waanders, 2006*)

Using Discrete Samples

Simple schematic example of the process using a portion of an example network





Sandia National Laboratories

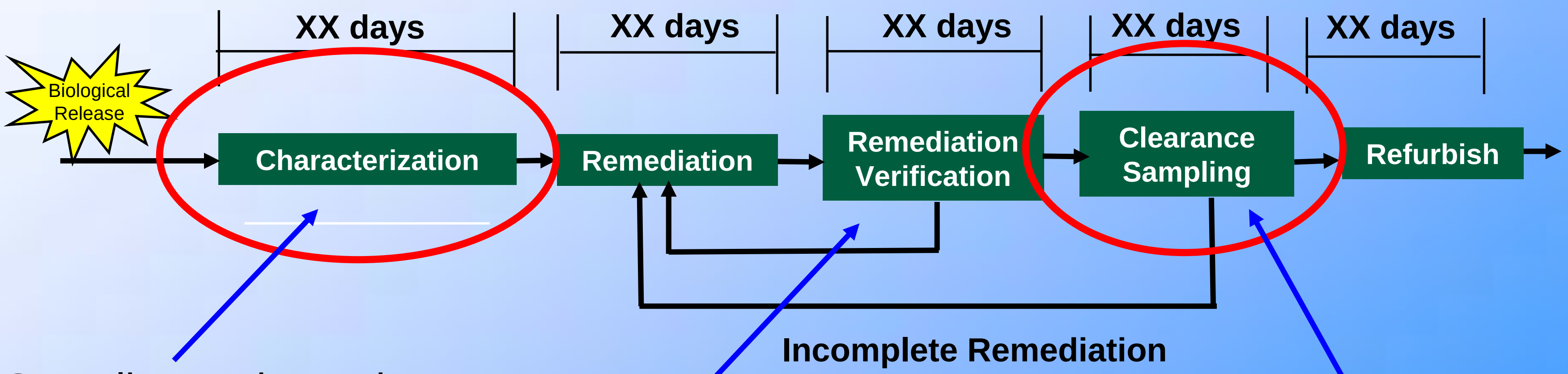
Outline

- Background and Motivation
  - Problem (Intentional and Accidental Contamination Events)
  - Tools for Solution (Numerical Models, Statistics, Data Mining)
- Pre-Event Monitoring
  - Optimal Monitoring Locations
  - Filtering Background Variation
- Response During an Event
  - Real-Time Source Location Inversion
- Post-Event Restoration
  - Characterization of Magnitude and Extent
  - Is Restoration Complete

Sandia National Laboratories

Facility Restoration

In facility restoration projects, characterization and clearance sampling are time consuming and expensive. If Remediation has failed, both remediation and clearance sampling must be redone



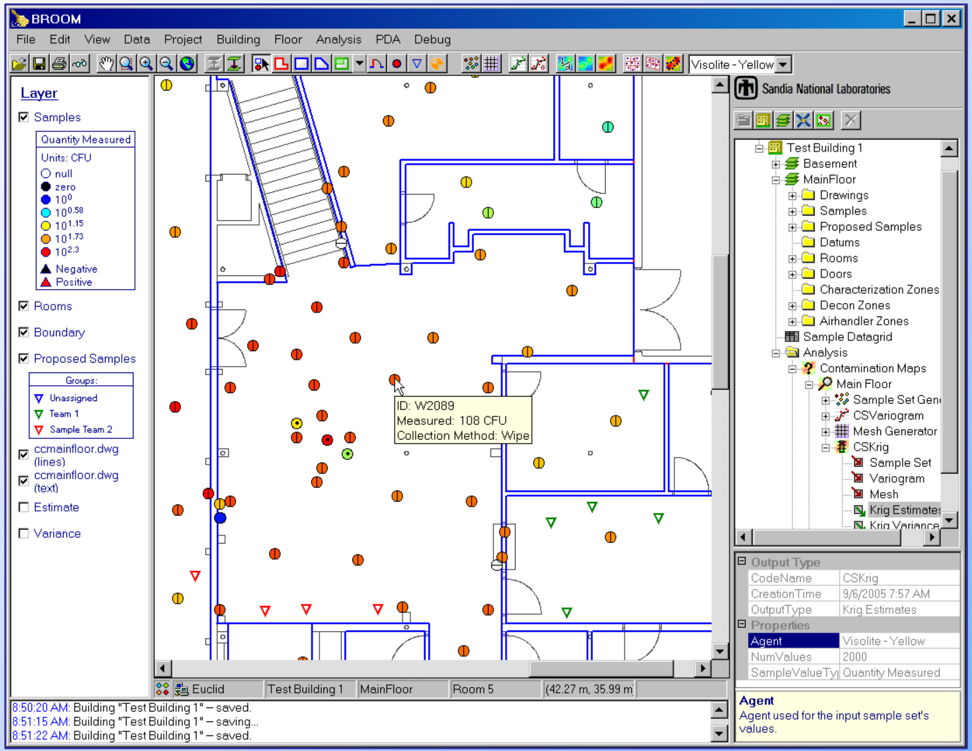
Sampling to determine magnitude and extent of contamination

Sampling to determine how well decontamination goals are met

Sampling to determine if restoration was successful

BROOM
(Building Restoration Operation Optimization Model)

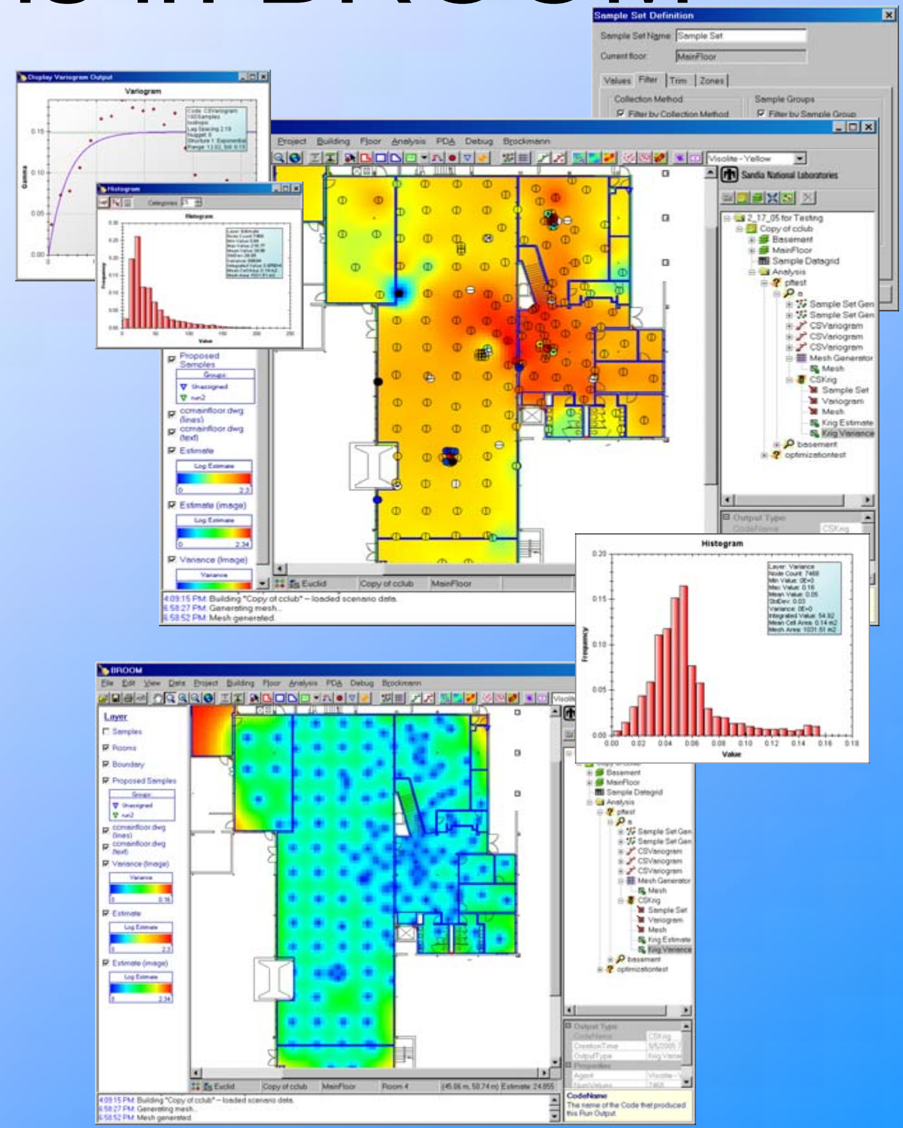
Interface for tracking and mapping sample locations and analysis results



PDA for use inside contaminated zone allows for electronic logging of sample information

Spatial Analysis in BROOM

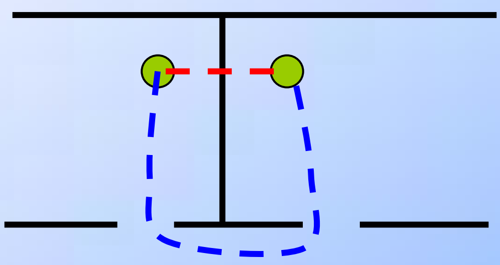
- Contaminant maps provide estimates of levels at unsampled points.
- Variance maps show level of confidence of contaminant estimates.
- Wide dynamic range: log and indicator transformation of sample sets.
- Integrated mass calculations give estimate of quantity of material released
- Ability to incorporate effects of walls and doors into mapping without use of a CFD model



Sandia National Laboratories

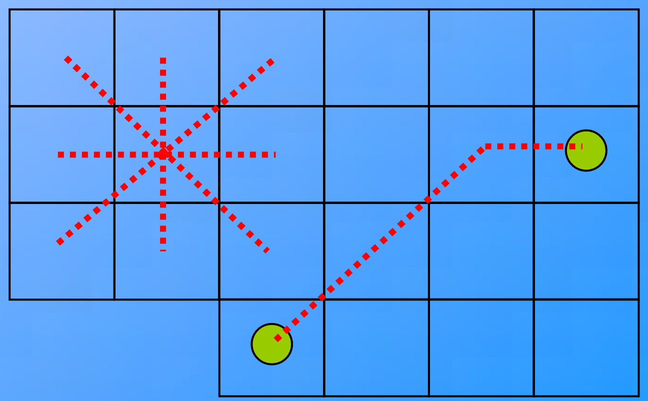
Mapping in Constrained Spaces

- Variogram/covariance is built using differences between sample values separated by a straight line distance
- Building architecture often precludes straight-line paths between samples – control volume flow models provide rough estimates of air-flow paths
- Can non-Euclidean distances improve the mapping process?



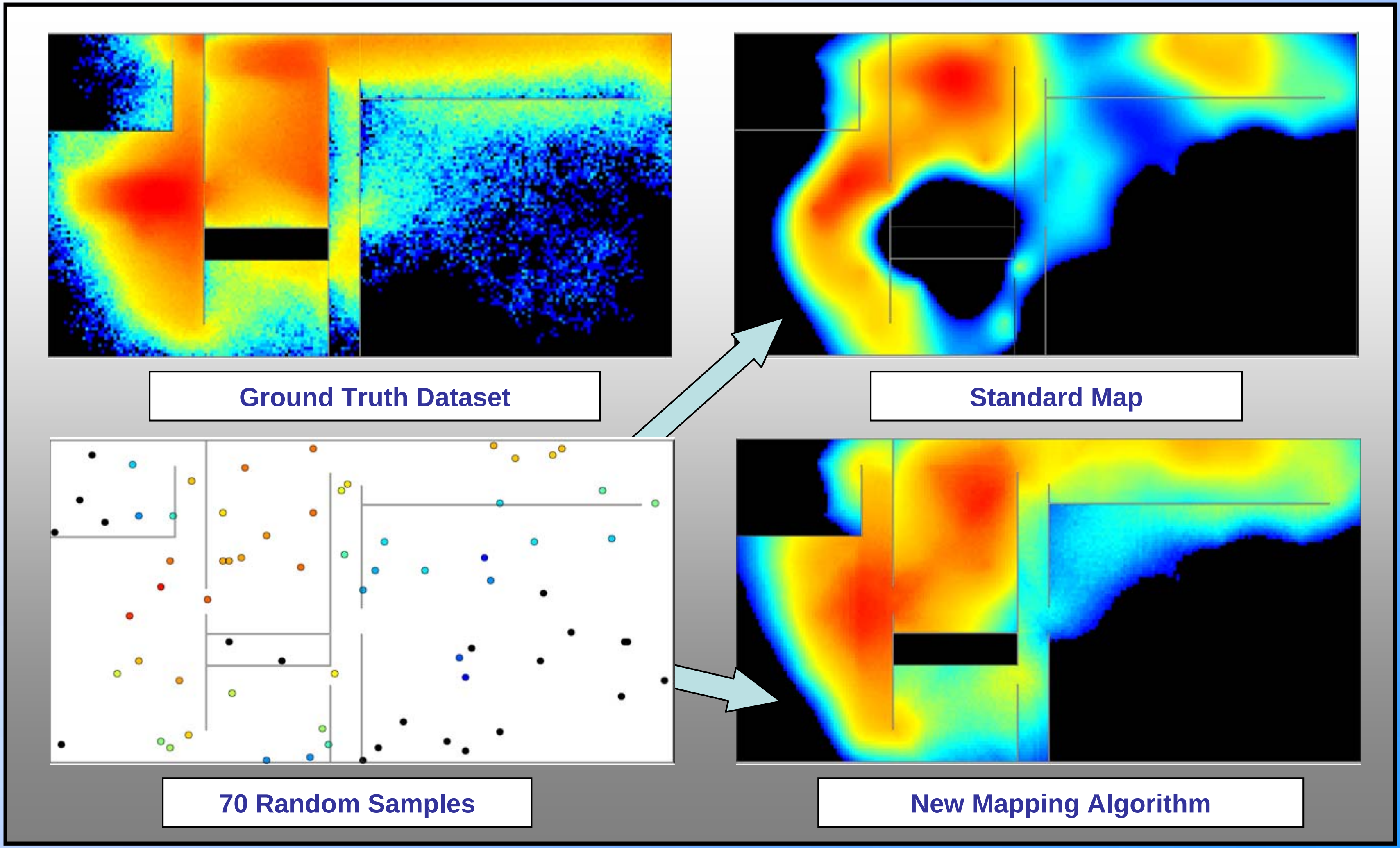
Grid-based approach to distance calculation

Dikstra's algorithm



McKenna and Finley, 2006, Mapping Contaminants in Buildings (invited presentation) First Annual Conference on Quantitative Methods and Statistical Applications in Defense and National Security, RAND Institute, Santa Monica, California, February 15-16th

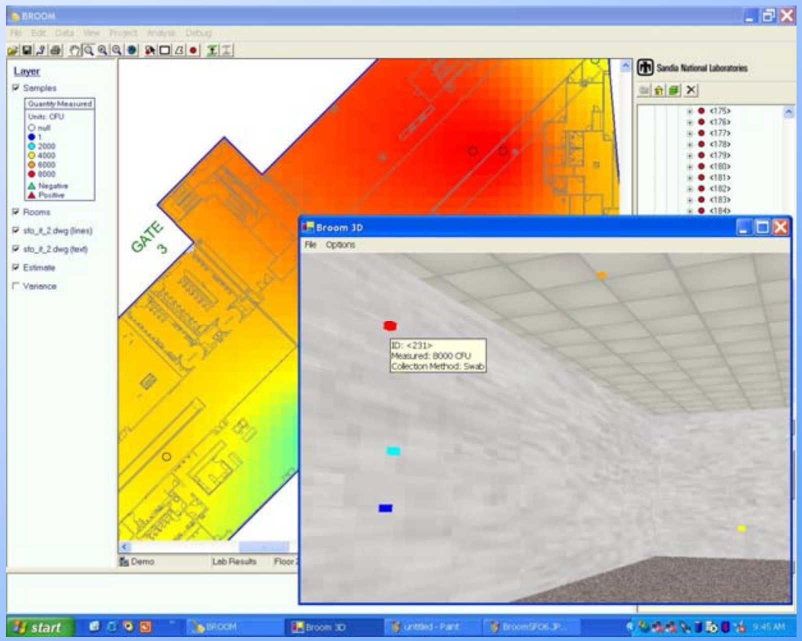
Advanced mapping algorithms in BROOM

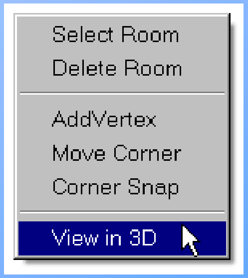


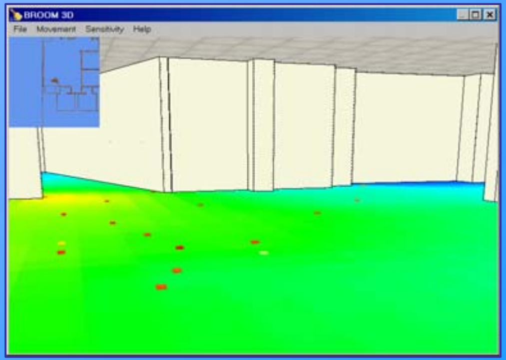
Sandia National Laboratories

Three Dimensional Analysis

- 2-D sample maps don't show vertically stacked samples
- Walls, furniture, HVAC vents commonly have stacked samples
- BROOM has interactive 3-D viewer to show closely spaced samples
- Use context menu from anywhere on map to get 3-D view of that part of the building





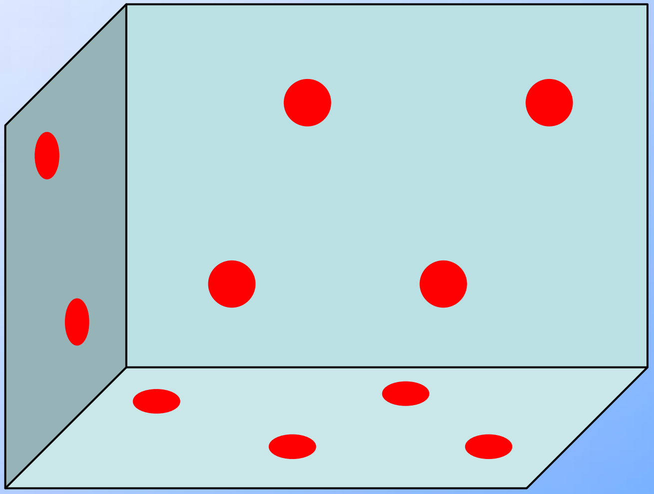


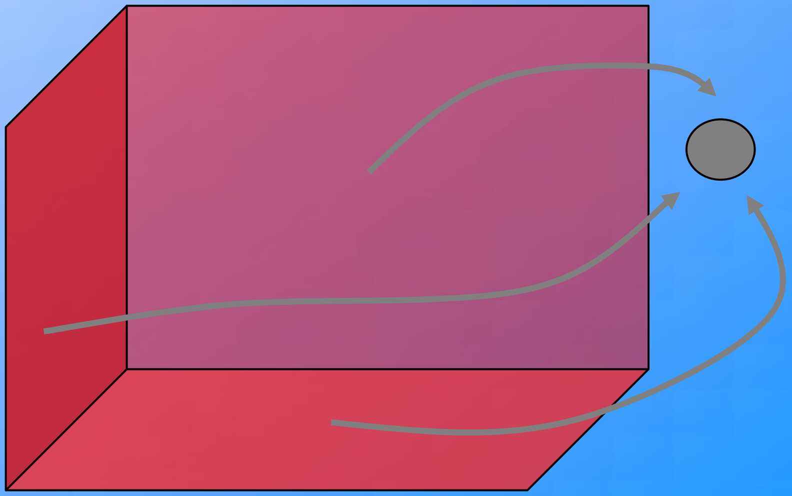
Outstanding Challenges

- Is the building clean?
  - How to prove a negative conclusion without taking thousands of samples?
- Ongoing work
  - “Aggressive air sampling” to reduce number of samples
  - Bayesian approach to incorporate prior information

Reducing Sample Numbers

- Currently we acquire multiple samples on room surfaces
- Ideally, have one sample per room volume that concentrates any contamination onto a single filter
  - *Bring the surfaces to the sampler (Concentrate)*





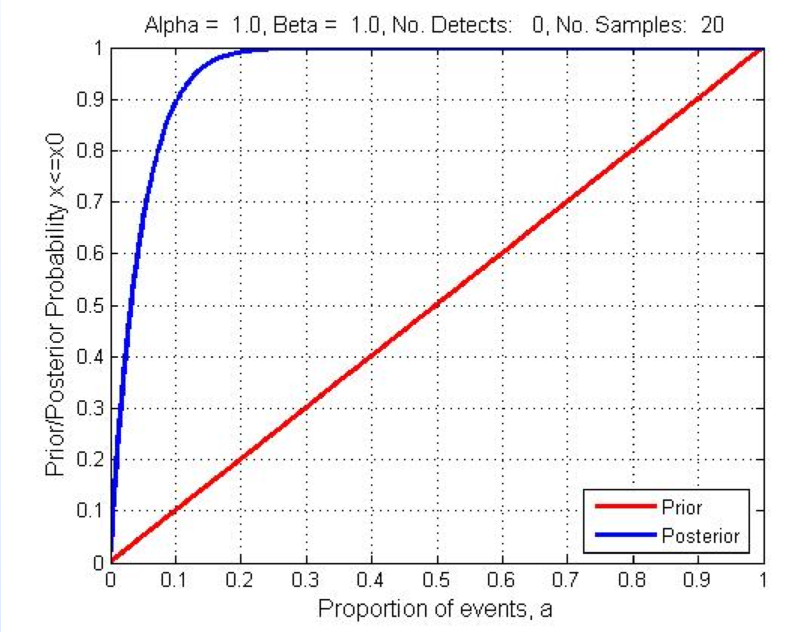
Incorporating Prior Information

Given *n* samples, all of which are negative, what is the probability that a positive sample could exist? Or what is the true proportion of possible samples that are positive

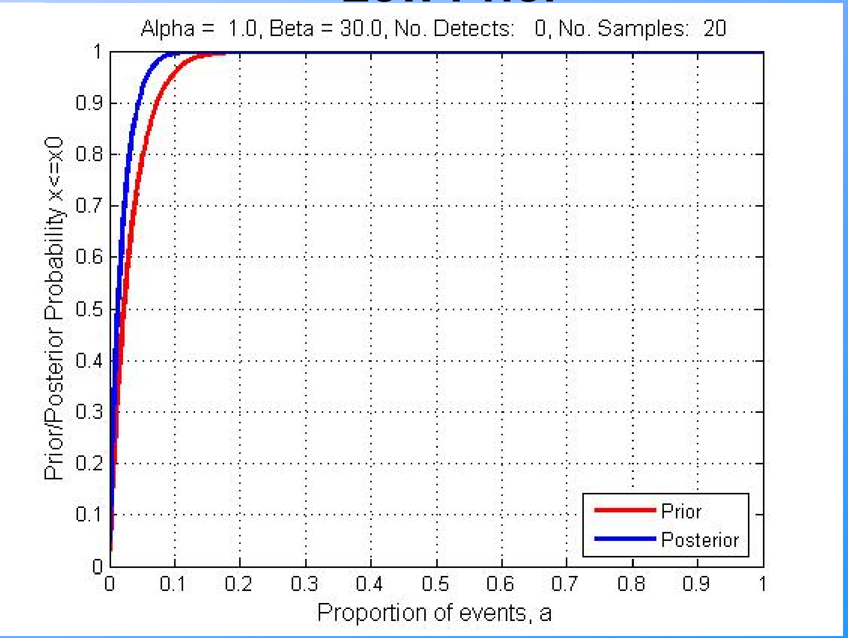
Use flow modeling and/or characterization and decontamination results to develop prior information on the distribution of the proportion of possible events

Graphs show prior and posterior distributions of the possible true proportion of positive samples, given that 20 negative samples have been acquired

Uninformed Prior



Low Prior



Summary

- Infrastructure monitoring and response are growing areas of research and industry
  - Engineers needed to understand background operating conditions and effects of those conditions on monitoring programs
  - A large number of unsolved problems (research areas driven by practical constraints)
- Uncertainty in operational elements of critical infrastructure requires stochastic approach to these problems

Sandia National Laboratories

Take Away Points

- The built environment is noisy and full of uncertainty
  - Sole reliance on deterministic modeling is not a viable option and statistical models need to be incorporated
- Operational efficiency and improved security are not exclusive sets
- Ability to work with large and real-time data sets needs to be part of engineering education
- What you learn today, may not be what you are doing tomorrow
  - Deep technical foundation in studies will allow you to attain more skills and apply those to an increasingly broad set of problems

Sandia National Laboratories

Acknowledgements

- This work was supported through:
  - Interagency Agreement between Sandia National Laboratories and the US EPA National Homeland Security Research Center
  - PUB (Singapore) Sandia water security project
  - US Department of Homeland Security

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