CANARY Overview: Capabilities and Demonstrations

WSI Pilot Cities Webinar

June 22nd, 2009

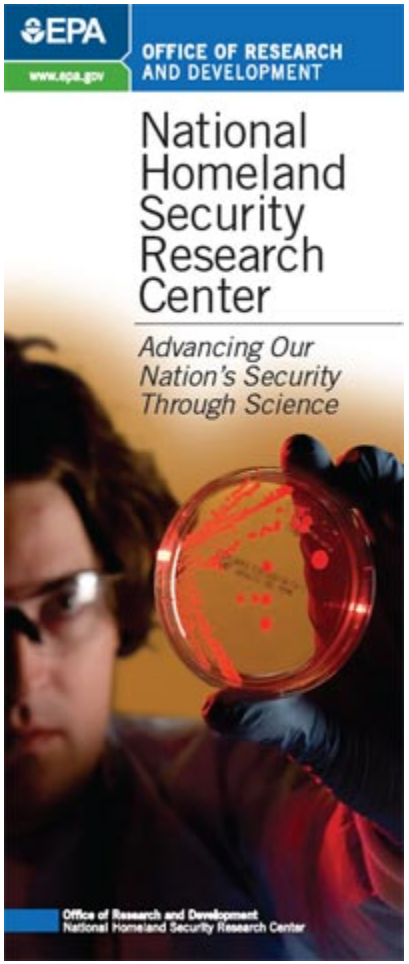
CANARY was developed through an InterAgency Agreement between the U. S. Environmental Protection Agency and Sandia National Laboratories. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Presentation Goals

- Present the CANARY event detection software
  - Quick overview of Capabilities
- Demonstrate use of CANARY
  - Video captures of CANARY results as interpreted in EDDIES middleware
- Answer questions from pilot utilities
- Query interest in more detailed CANARY training

EPA's National Homeland Security Research Center (NHSRC)

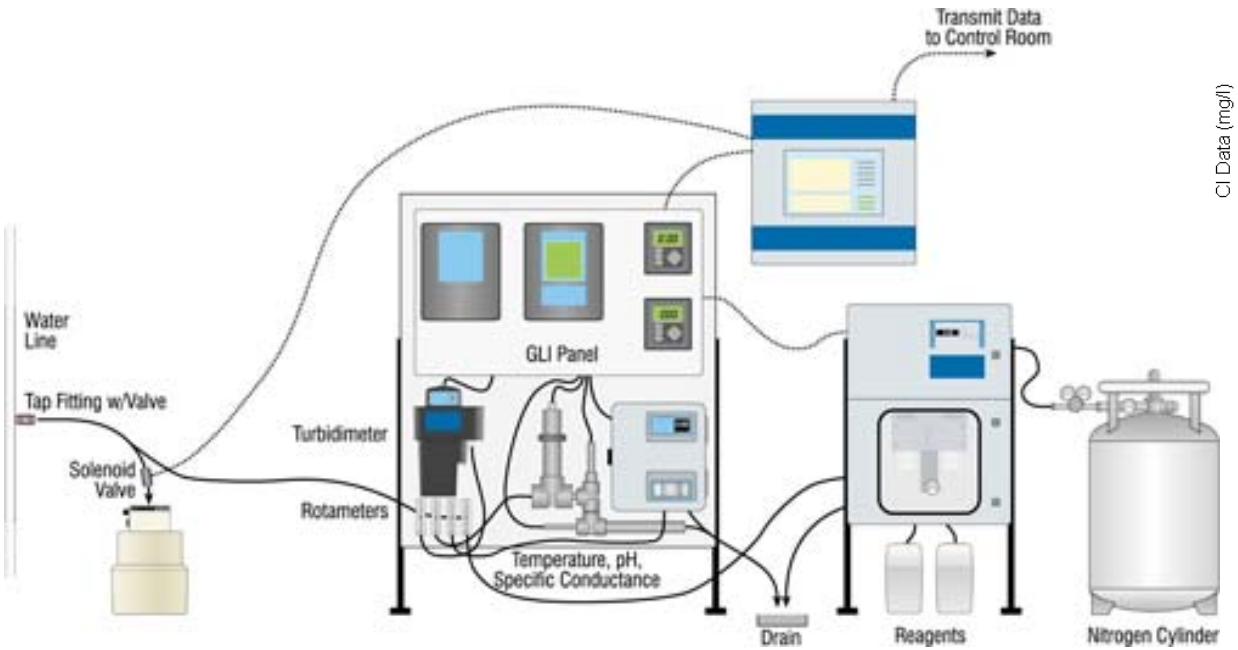
- Mission is to improve the nation's capability to:
  - Prevent, mitigate, and respond to terrorist attacks on our water and wastewater infrastructure
  - Respond to terrorist attacks affecting buildings and the outdoor environment

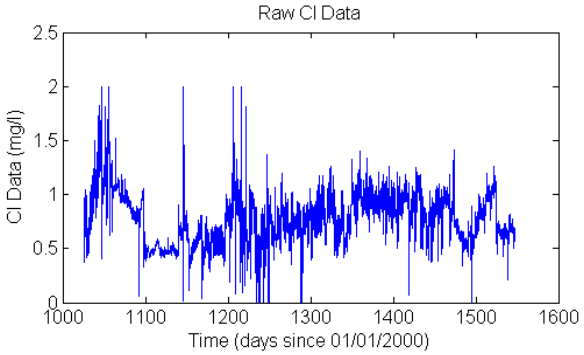


NHSRC Partnership with EPA's Water Security Division

- Identifying research needs
- Partnering with water utilities
- Communicating research results to stakeholders
- Designing and implementing WSI
  - NHSRC develops new tools and technologies for WSI
  - WSD manages implementation program

Online monitoring component of WSI





- Chlorine, TOC, conductivity and other water quality parameters change in the presence of many contaminants of concern

- Water quality data is noisy

- How can real events be reliably detected?

EPA Research on Event Detection

- Available proprietary “black box” methods could not be used to understand different algorithms and their performance in real-life field applications
- Partnered with American Water Works Association and 22 utilities (TEVA User’s Group) to identify research needs
- Collaborated with Sandia National Laboratories to address research questions:
  - What is the potential for an accurate event detection system with minimal false positives?
  - How can systems be evaluated in the absence of field data from real contamination events?

CANARY Software

- **Developed from 2003-present**
  - Tested off-line on data from 2 utilities
  - Tested on data from 20 contaminants tested in pipe loop studies
- **Real-time applications**
  - WSI pilot city since July 2007
  - EPA's T&E facility on data from field site
  - Singapore Public Utility
- **Documentation**
  - User's Manual (EPA, 2008)
  - Training workshop at 2008 AWWA Water Security Congress
  - Several published scientific articles and conference presentations on methodology and application

How can pilot cities use CANARY?

- Availability
  - Software available to all WSI pilots (Murray.Regan@epa.gov)
- Training
  - Future webinars planned on configuration, training, and use of algorithms and clustering
- Model for use
  - Utilities should use contractors or staff to install, configure, and troubleshoot problems
  - Software bugs will be fixed by EPA/SNL
- Future of CANARY
  - EPA and SNL will maintain software for several years
  - Consultants will be trained to provide customer service
  - Vendors will be engaged to incorporate CANARY into existing SCADA or data management tools

What is CANARY?

- Water quality event detection software that will:
  - Continuously monitor water quality data streams at multiple sensor locations
  - Alert the water quality analyst to periods of anomalous water
  - Provide maximum flexibility in choosing the event detection sensitivity at each monitoring station
  - Disregard water quality values observed during periods of sensor malfunction or hardware alarm
  - Recognize recurring water quality patterns to reduce false alarms
  - Efficiently create graphics of results for analysis and presentation

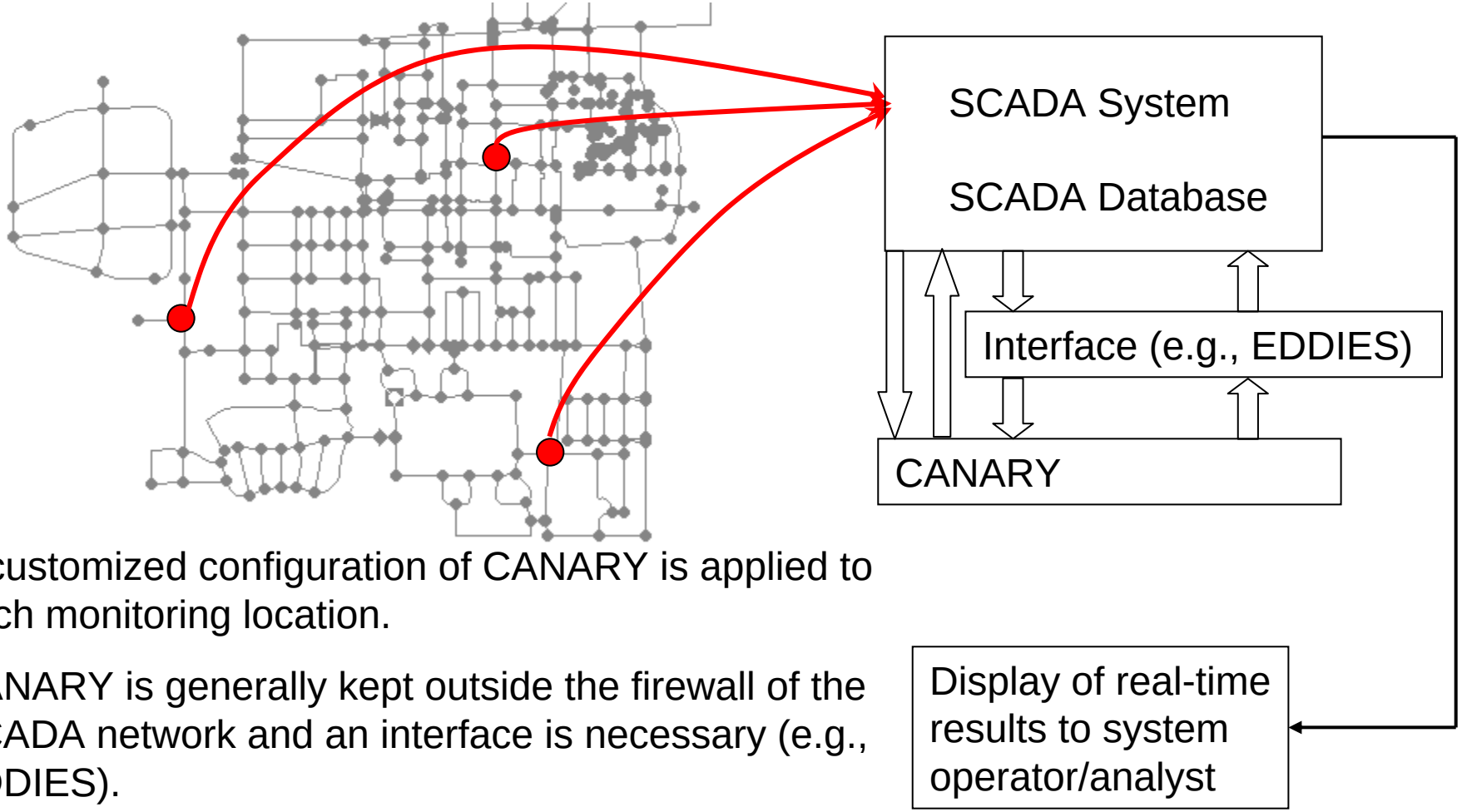
CANARY Design Philosophy

- Provide real-time indication of unexpected (anomalous) water quality at a monitoring station
  - Not trying to provide an indication of the cause of that anomaly
  - Provide some reference information in the form of pattern matching to recurring, previously seen water quality changes
- Each monitoring station (location) is analyzed independently
  - Does not require a calibrated network model
  - Does not require a complicated rule set for operations at each utility
- Work with heterogeneous mix of sensors
  - Different parameters, different manufacturers, easy to add and test new hardware within the event detection system
- Provide an open source software platform for Event Detection Systems (EDS)
  - Enable future algorithm development and testing

A Typical Application

- A typical application is CANARY running on a server with direct/indirect connection to a realtime SCADA database
  - Could be multiple instances of CANARY – running on each of multiple SCADA systems handling different parts of network
- Multiple monitoring stations and multiple water quality signals at each station
  - Typically 10-20 stations and 4 to 6 signals at each station (e.g., Cl, pH, CDTY and turbidity monitored at all stations)
  - Largest application to date is simultaneous event detection on 70 stations each with 3 to 5 water quality signals
- SCADA update (sampling interval) is in the range of 2 to 10 minutes
  - All signals and all stations have the same sampling interval

Connection to Network



A customized configuration of CANARY is applied to each monitoring location.

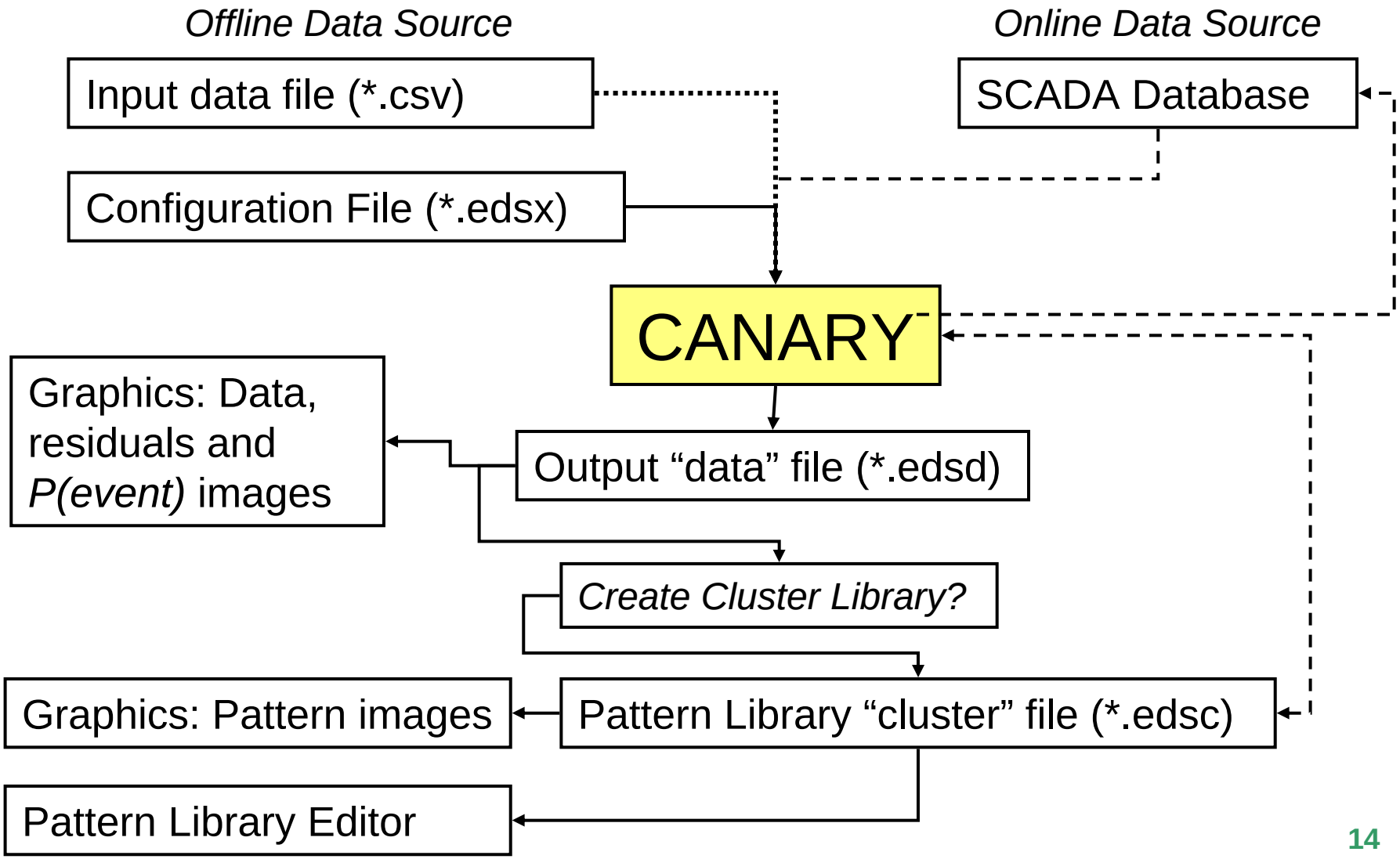
CANARY is generally kept outside the firewall of the SCADA network and an interface is necessary (e.g., EDDIES).

CANARY passes *P(event)*, alarms, pattern match information, etc back to SCADA for operator viewing.

CANARY: Offline and Online

- CANARY can be run in two different modes
- Offline:
  - Used on historical data to develop parameters for a particular site
  - Pattern libraries are constructed from historical data
  - Data source is a comma separated value (*.csv) file
  - Final output in the form of text files and graphics
- Online:
  - Data source is connection to realtime SCADA database
  - Output is sent back to SCADA system
  - Pattern libraries are updated automatically
  - Same text file and graphics outputs as in offline mode can be created

CANARY Data and Processing Flows



CANARY Capability Overview

- Base functionalities
  - SCADA connections and data quality indications
  - Automated and adaptive recognition of background variation
  - Residual classification
  - Binomial event discriminator (Outlier, Event, Baseline Change)

- Recent Developments
  - Set point algorithm
  - Consensus algorithms (combining results of two or more event detection algorithms)
  - Pattern recognition (offline library construction, online matching, can include hydraulic data as direct inputs)

SCADA connections

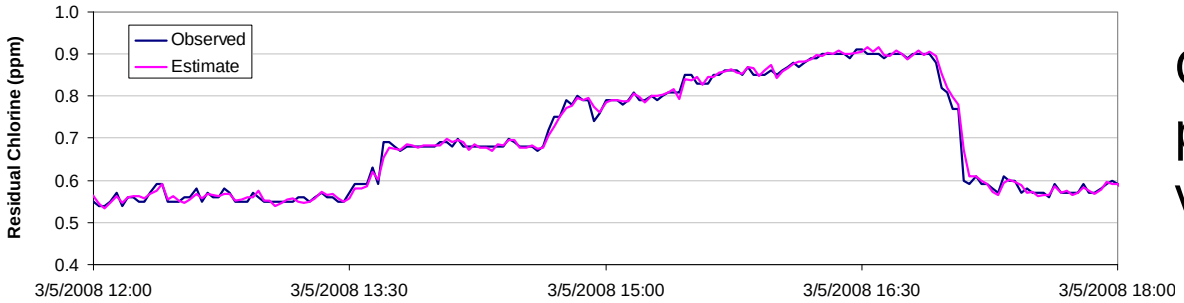
- CANARY has been designed to work with generic databases, or to connect to SCADA systems via the EDDIES software (available from EPA Office of Water)
- EDDIES provides a middleware layer that:
  - Allows EDS software to be held outside a firewall and outside direct access to the SCADA database system for improved security
  - Provides graphical representations of alarms, probabilities, and possible causes.

Data Quality Indications

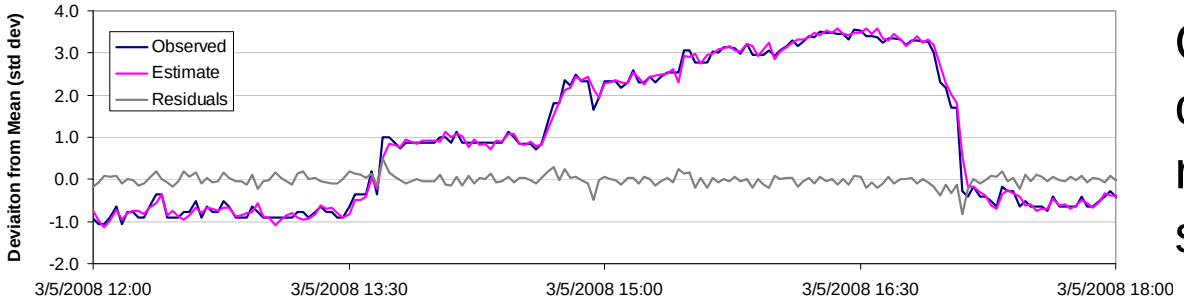
- Quality indicators from SCADA transmission monitoring or from hardware quality signals can be used to mask individual signals
  - Maintenance Required, Hardware Error signals, etc.
  - SCADA transmission line “Quality” tags
- Other SCADA signals can act as “calibration” indicators, turning off alarms at an entire station during maintenance
- Individual SCADA signals can also be assigned minimum and maximum possible data values to serve as malfunction warnings

Capability: Adaptive Signal Filtering

- Water distribution systems are noisy and water quality observations contain daily, weekly, seasonal and annual changes in addition to sensor and SCADA noise
- Predict this background variation so significant changes in signal are readily observed



Observed and predicted chlorine values (raw space)



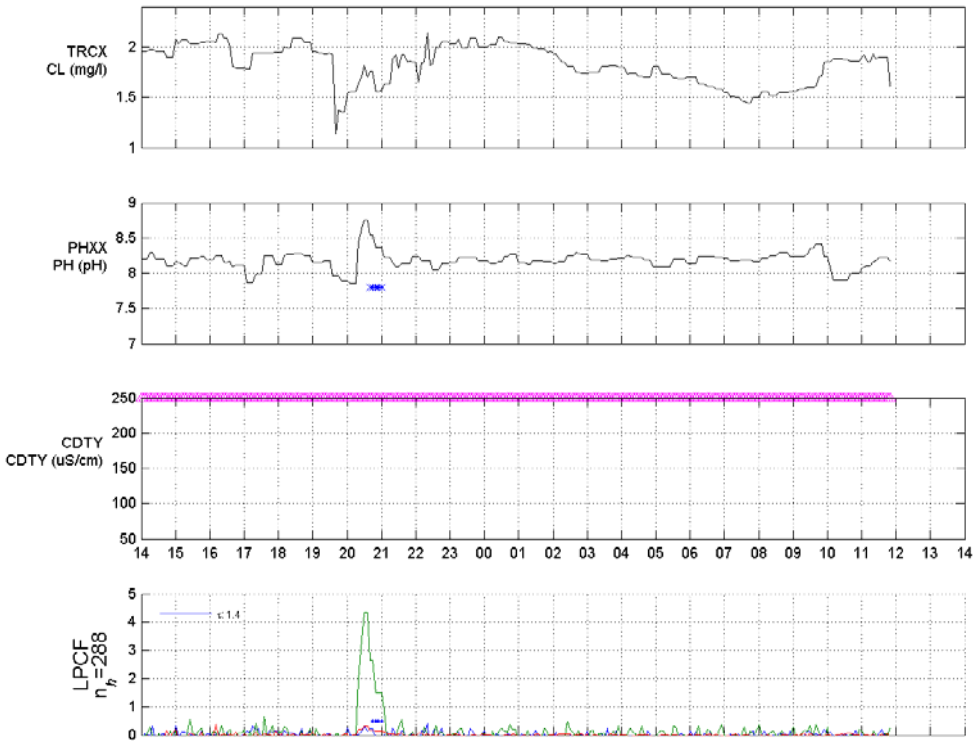
Observed and predicted chlorine values with residuals (std dev. space)

Capability: Residual Classification

- Water quality deviates from expected values. When is a deviation "significant"?
- Classification of residuals into those representative of background water quality and those that are not ("outliers")

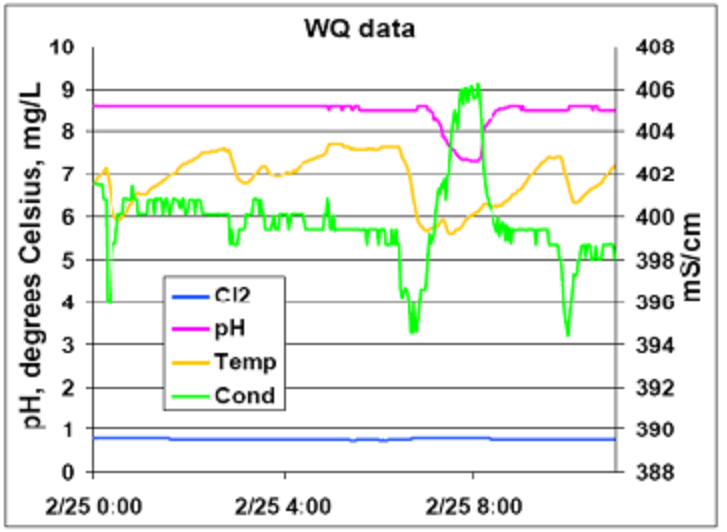
One day of analysis, three water quality signals, residuals are shown in bottom graph. Residuals only rise above 1.0 at one time.

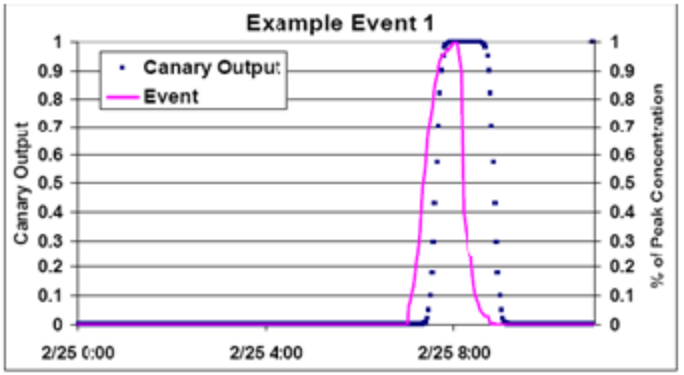
Cl is most variable, yet pH causes the large residuals – adaptive threshold recognizes relative change in signal



Capability: Binomial Event Discriminator

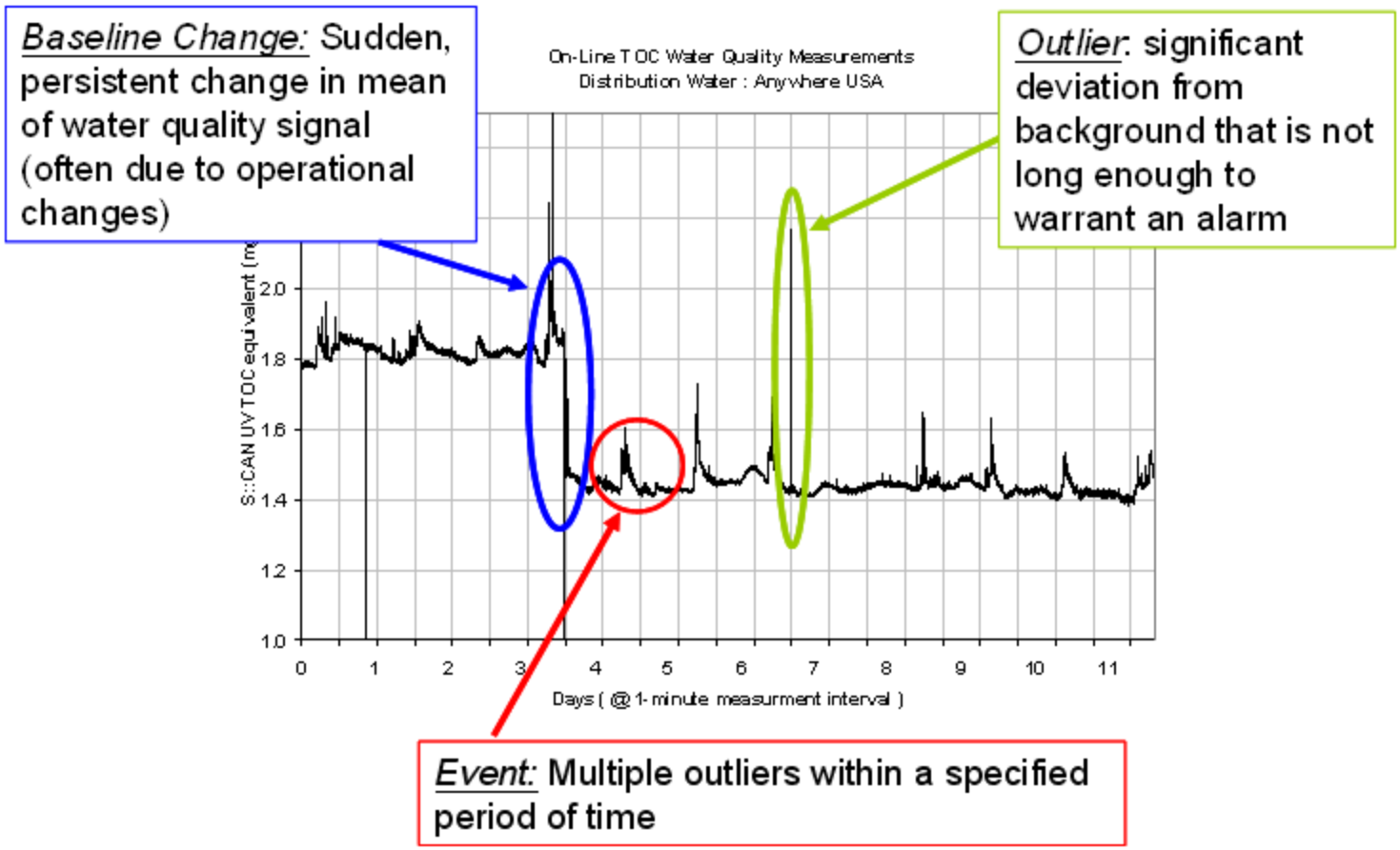
- Provide a continuous measure of the probability of an event and combine results from several consecutive time steps to suppress false positive alarms
  - Alarm on a significant change, but don't alarm on a spike in the SCADA system





Capability: Binomial Event Discriminator

- Use BED to identify three types of water quality changes:
  - Outlier, Event, Baseline Change



CANARY Demonstration

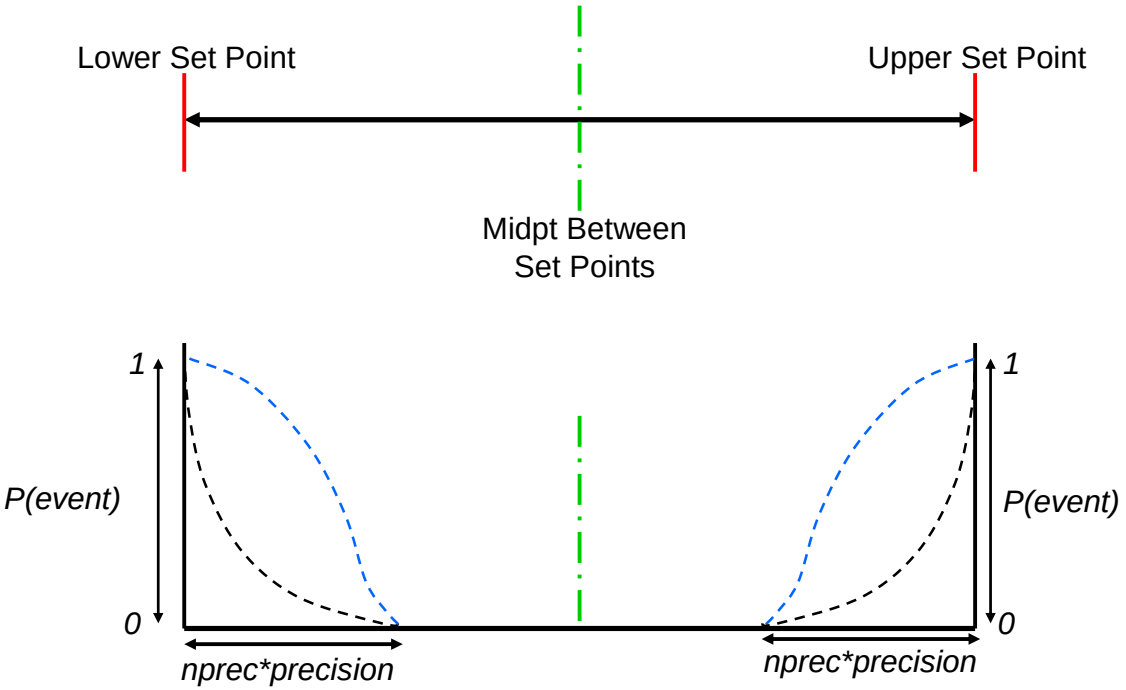
CANARY Capability Overview

- Base functionalities
  - SCADA connections and data quality indications
  - Automated and adaptive recognition of background variation
  - Residual classification
  - Binomial event discriminator (Outlier, Event, Baseline Change)

- Recent Developments
  - Set point algorithm
  - Consensus algorithms (combining results of two or more event detection algorithms)
  - Pattern recognition (offline library construction, online matching, can include hydraulic data as direct inputs)

Capability: Set point algorithm

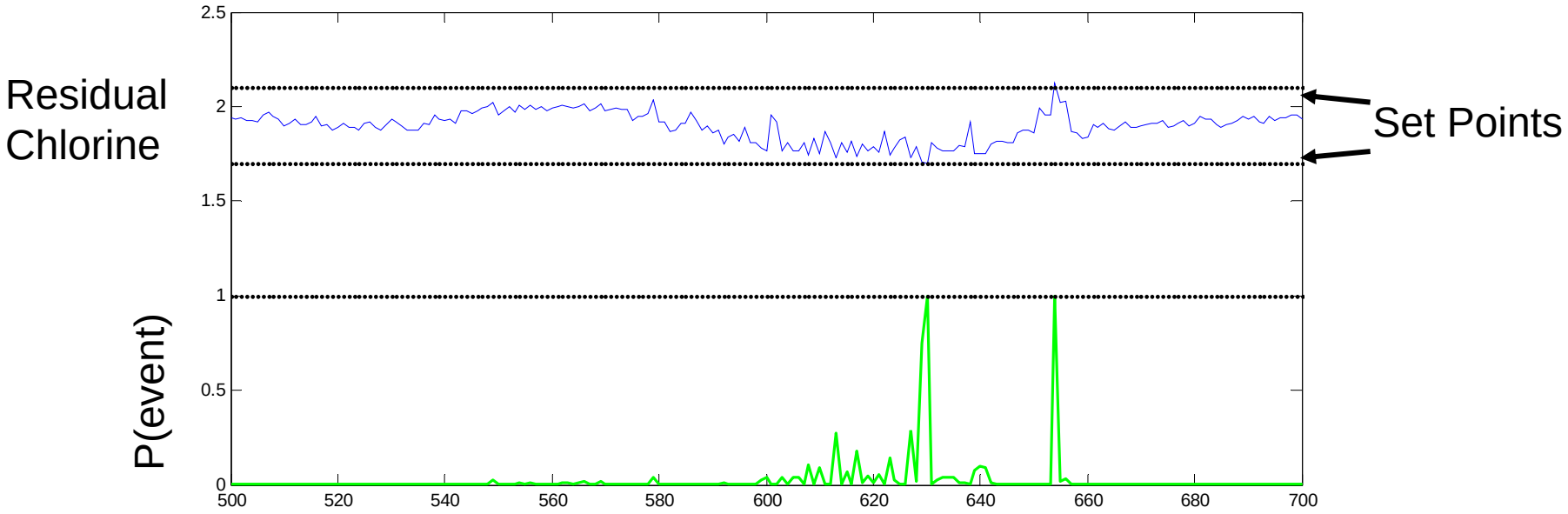
- Set point values (thresholds) are used by most utilities and provide alarm if water quality crosses a threshold
- CANARY provides a user-adjustable approach to provide an increasing probability of event as the water quality approaches a set point (takes into account the sensor precision)



Increase in *P(event)* as a function of proximity to set point is defined by exponential (black) or beta (blue) distribution as chosen by user.

Capability: Set point algorithm

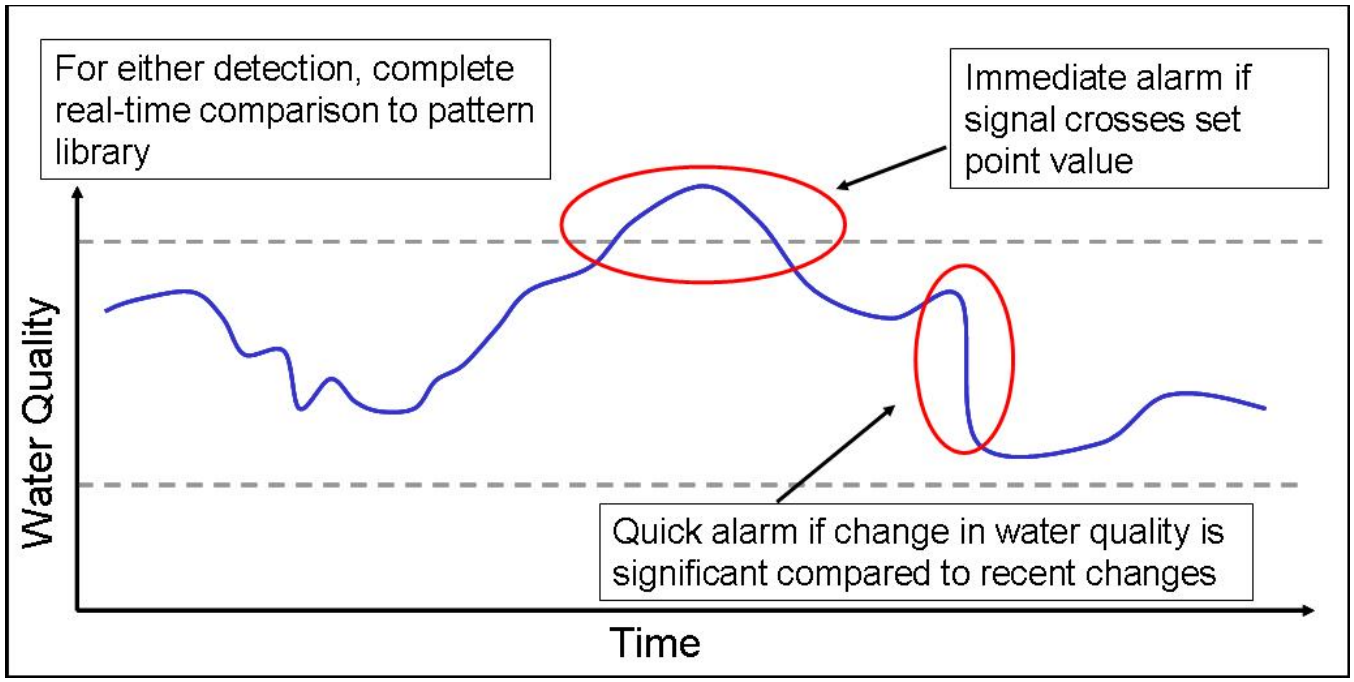
- Distance away from set point threshold is defined by user as a number of times of the precision of the sensor. Probability threshold can be set by user to provide as much advance warning as required



Experience to date shows that the set point algorithm is effective for early detection of a sensor that is slowly failing or slowly going out of calibration

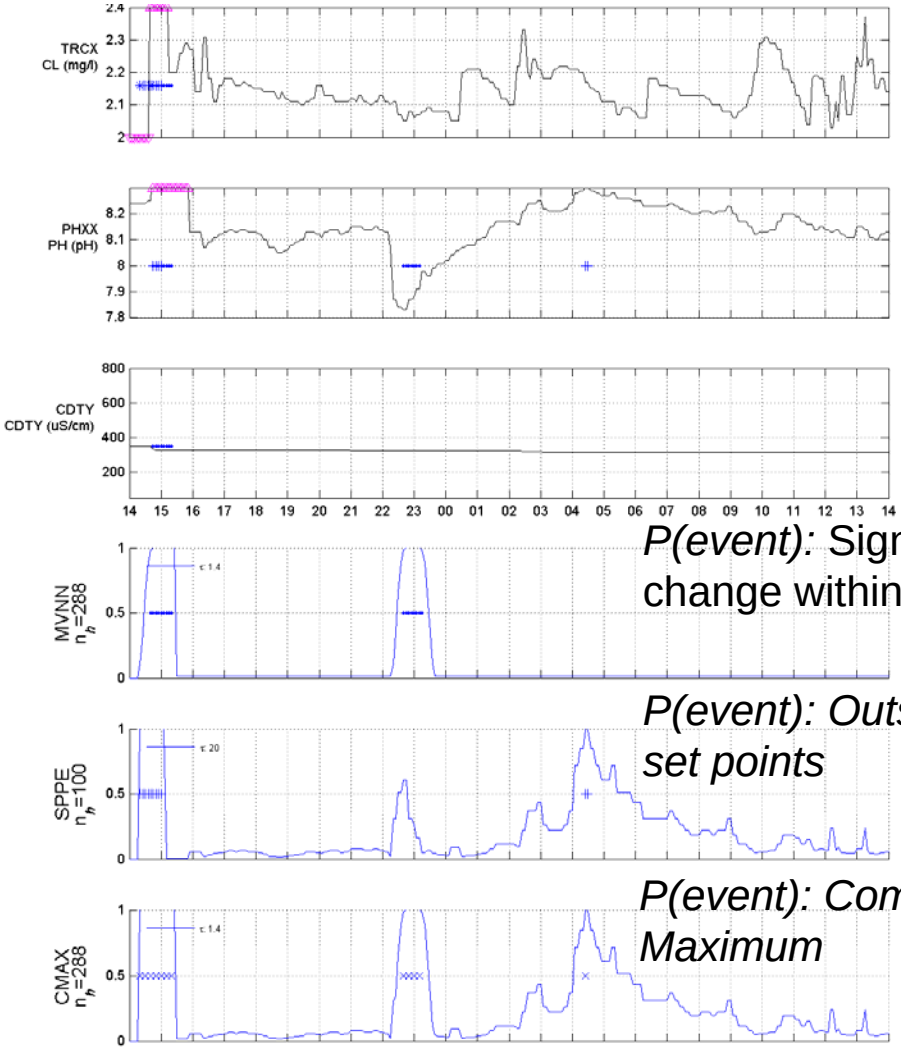
Capability: Consensus Algorithm

- Provide the advantages of multiple algorithms at the same time
- As an example, a combination that would identify significant changes within the set points as well as alarm if the water quality leaves the set points
- Implementation: User decides to retain the maximum or the average *P(event)* value from each algorithm



Capability: Consensus Algorithm

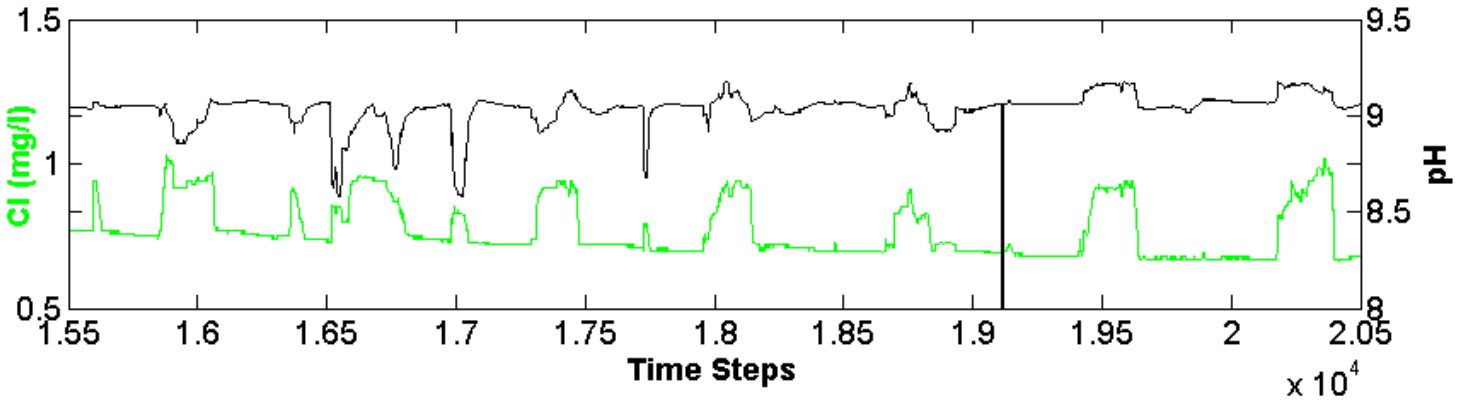
- Consensus algorithm simultaneously tracks significant relative changes within set point thresholds and provides warning of water quality approaching set point limits



CANARY Demonstration

Capability: Pattern Matching

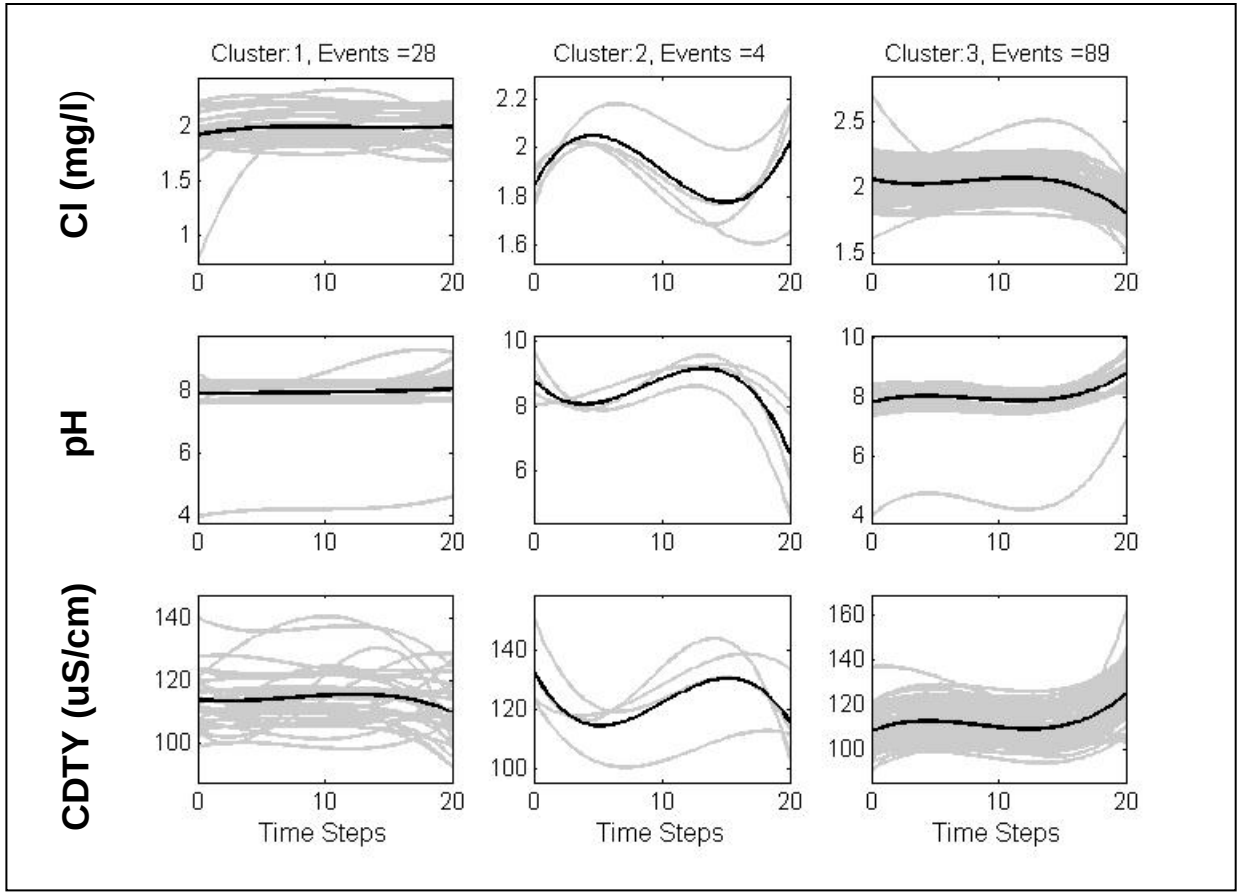
- Significant water quality changes can occur due to changes in network operations (draining tanks, pumps switching on and off, valve settings, etc.).
- CANARY has a pattern matching capability to recognize recurring, multivariate water quality patterns, store them in a library and then compare realtime water quality observations to that library to reduce false positives



One week of Cl and pH data showing recurring patterns of change. Note that the pattern changes across time.

Multivariate Pattern Library

Three patterns identified from training data. Regression model fit to data of each pattern member (grey lines) and average regression models (black) are shown



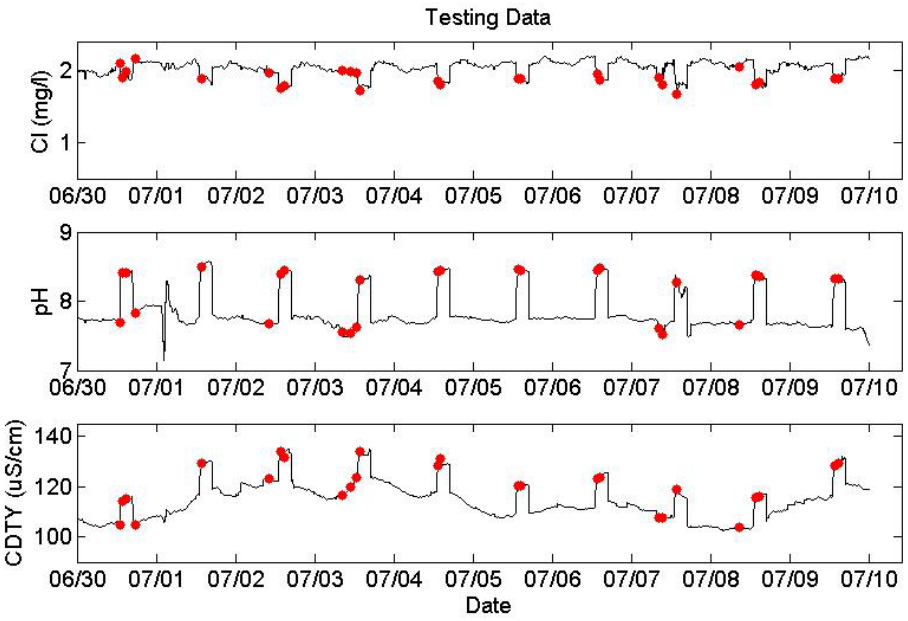
Example Pattern Matching Results

Analyzed 72 days of data with and without the pattern matching turned on

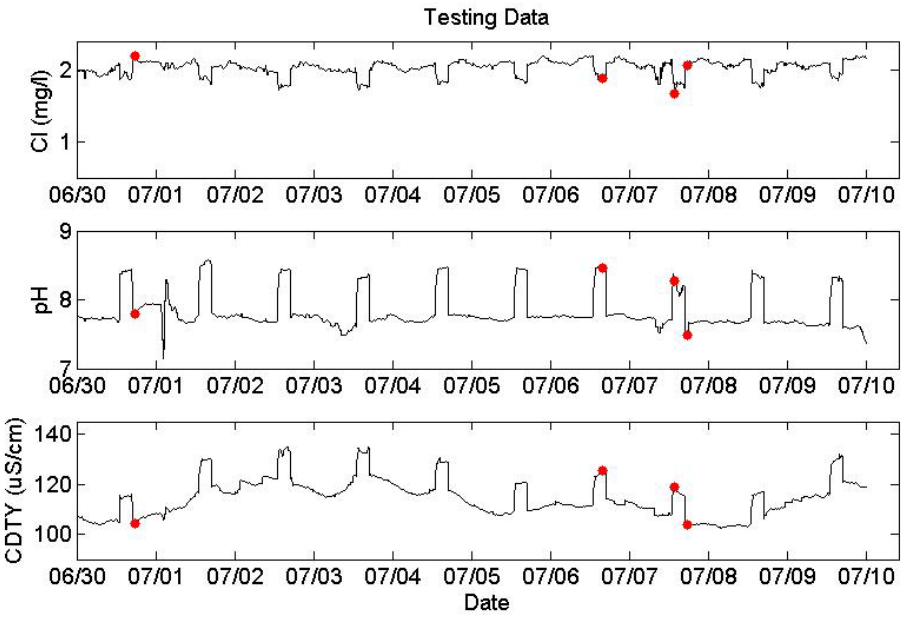
Pattern Matching Off: 65 alarms sounded

Pattern Matching On: 14 alarms sounded (79 percent reduction)

No Pattern Matching



Pattern Matching



Example Results for 10 days of data

Summary

- CANARY available to WSI participants: Murray.Regan@epa.gov
- Questions?
- Follow up