

CANARY Webinar: Algorithm Overview and Pattern Library Construction

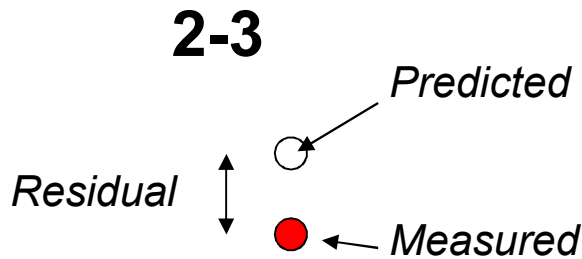
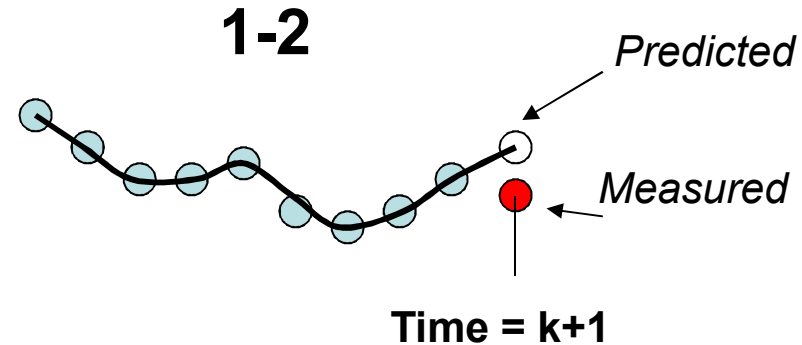
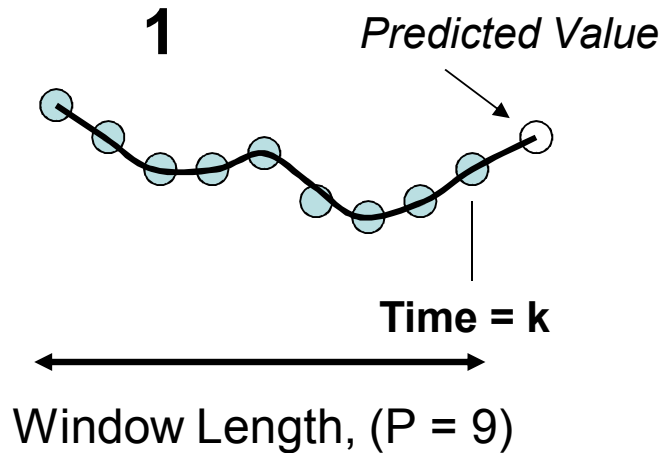
WSI Pilot Cities Webinar
September 17th, 2009

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.

Event Detection: Steps

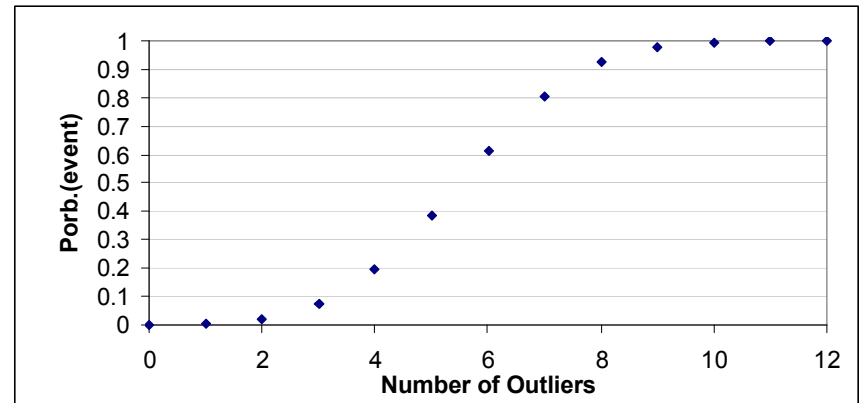
- Filter:
 - Use an adaptive filter to model background variations and predict next water quality value
- Compare:
 - Compare predicted and measured values for each time step (difference = residual)
- Combine:
 - Combine residuals across all water quality signals at a location to identify outliers in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

Event Detection: Steps



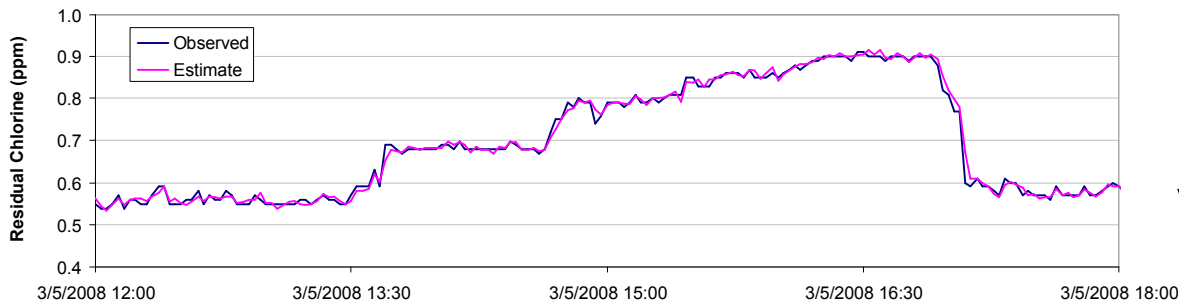
Compare the residual at each time step to a threshold. Those that exceed the threshold are “outliers”

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

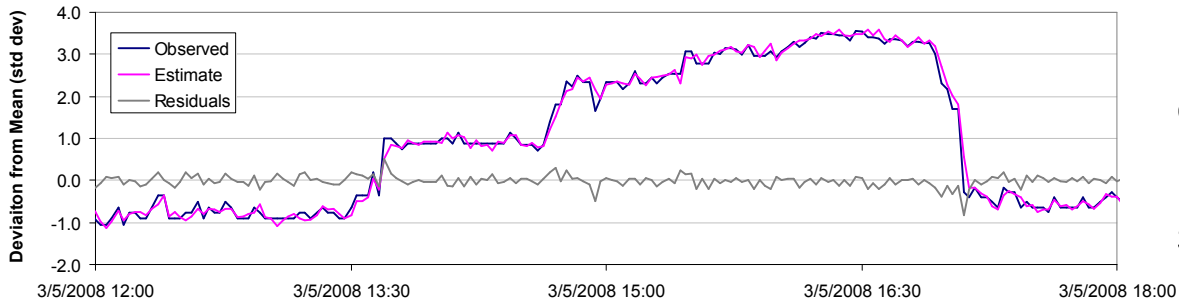


Adaptive Filtering Results

- 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)



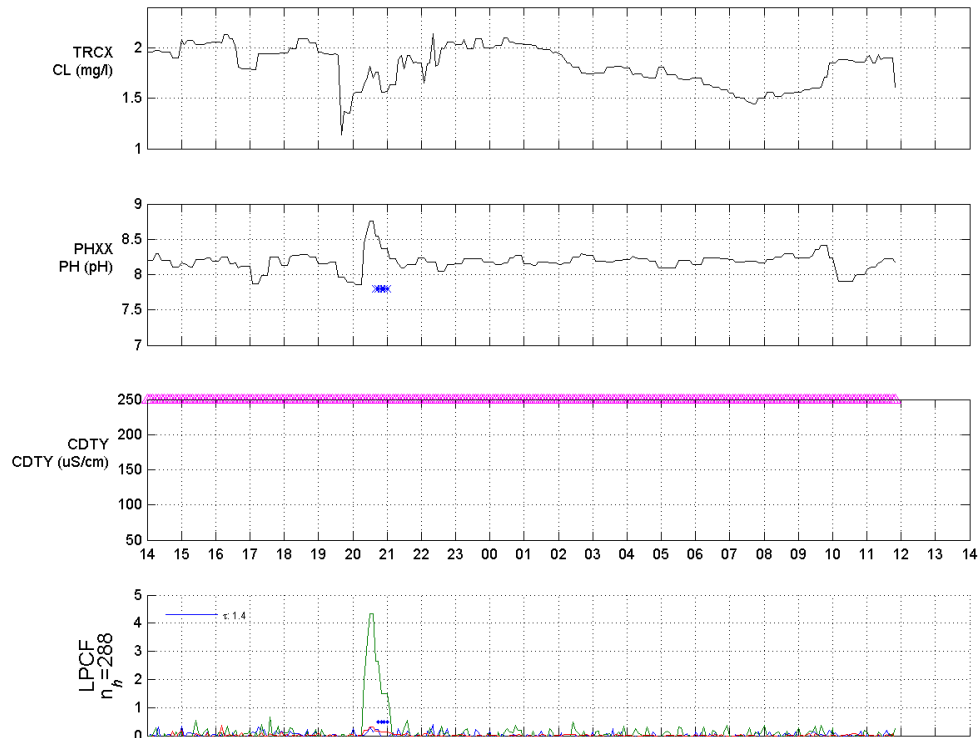
Observed and predicted chlorine values with residuals (std dev. space)

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”)

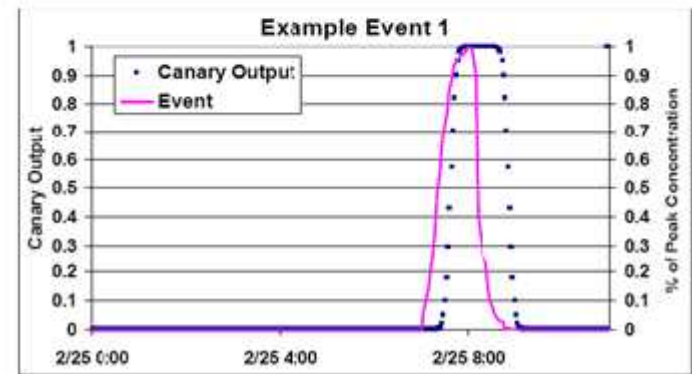
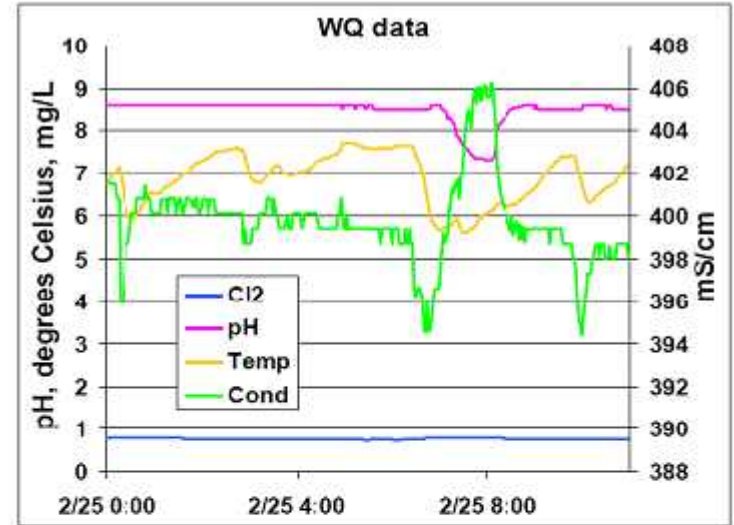
One day of analysis, three water quality signals, residuals are shown in bottom graph. Residuals only rise above 1.0 at one time.

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



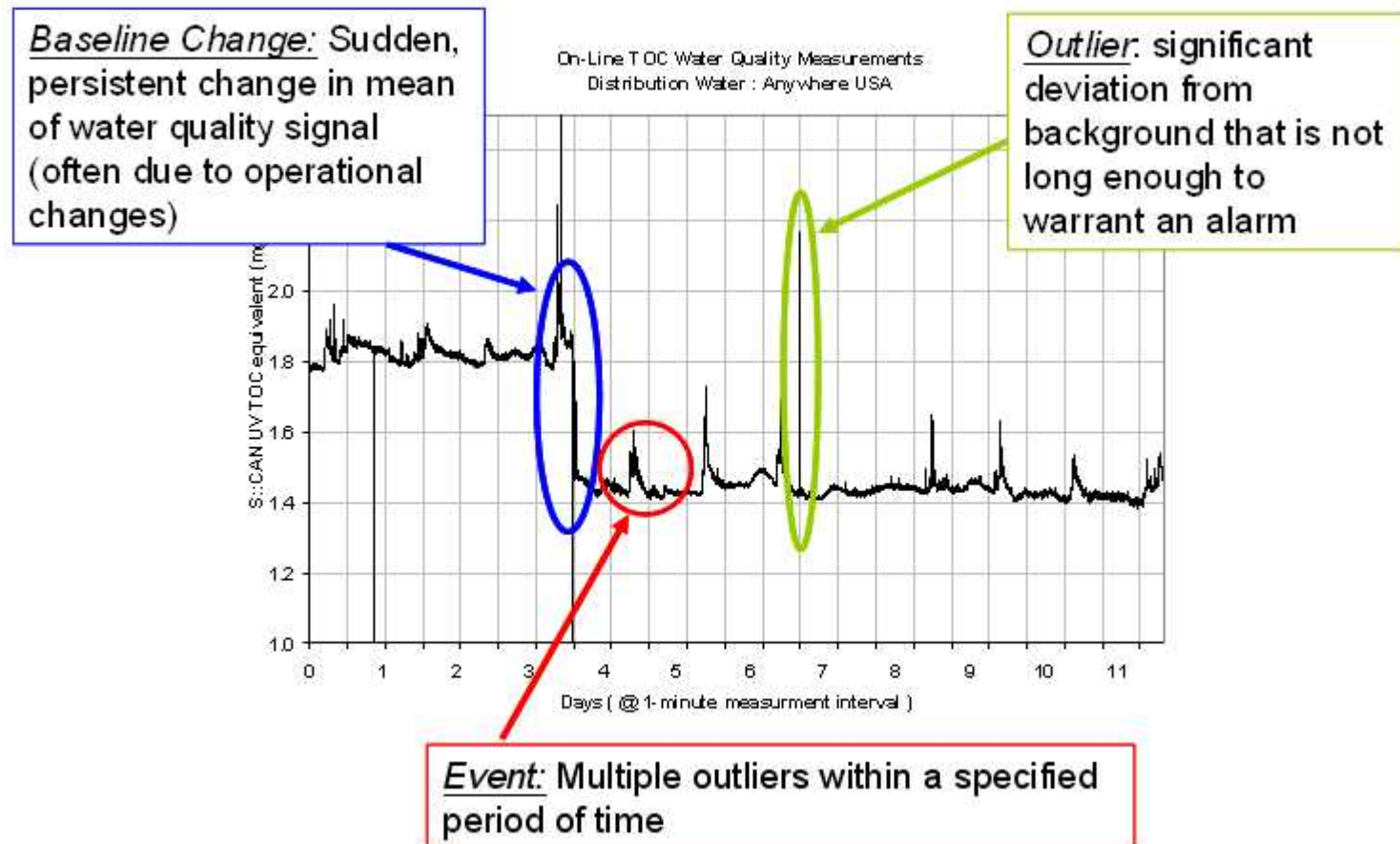
Aggregation: Binomial Event Discriminator

- 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



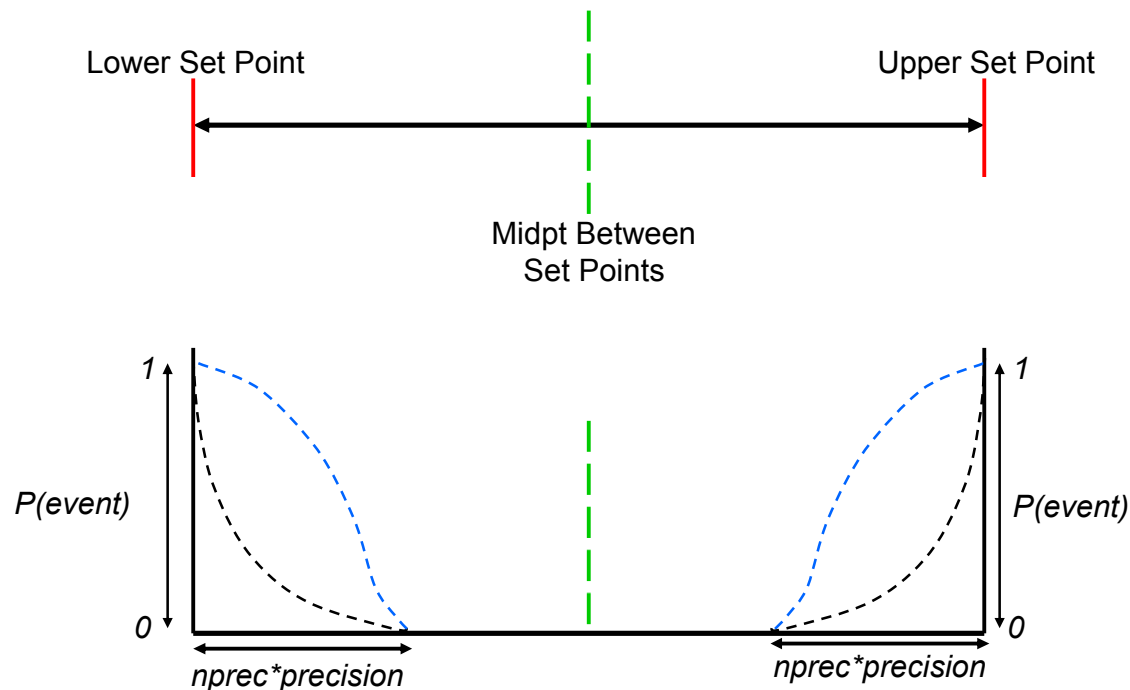
Aggregation: Binomial Event Discriminator

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



Set point algorithm

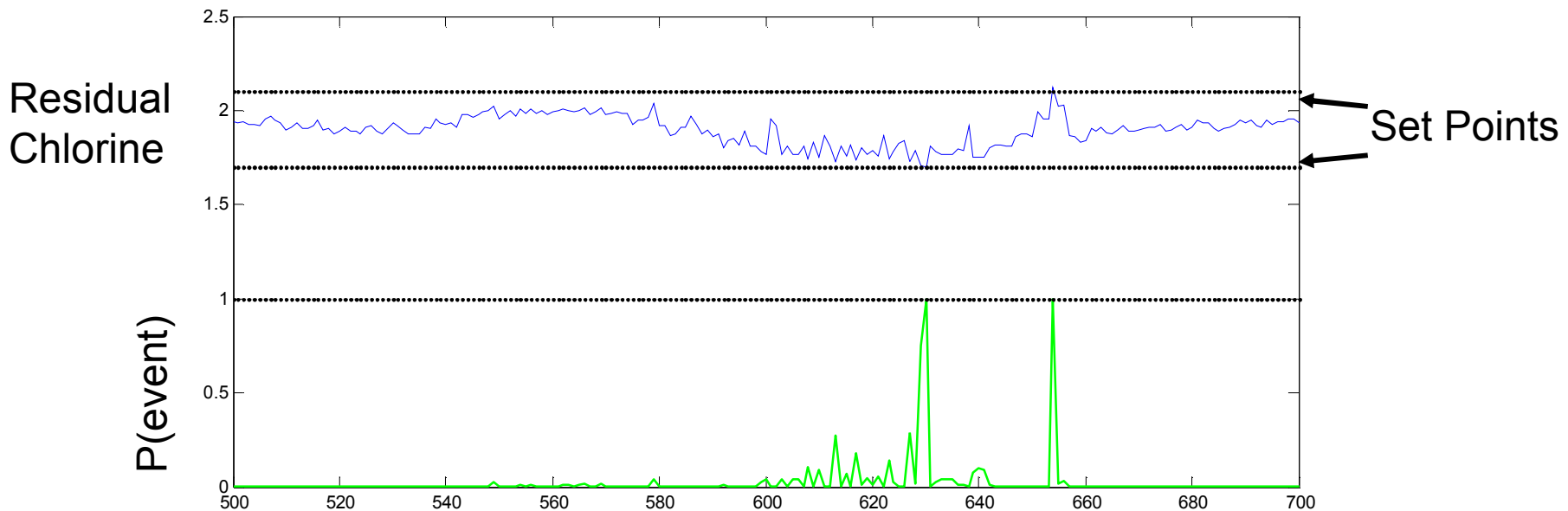
- Set point values (thresholds) are used by most utilities and provide alarm if water quality crosses a threshold
- CANARY provides a user-adjustable approach to provide an increasing probability of event as the water quality approaches a set point (takes into account the sensor precision)



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.

Set Point Algorithm Example

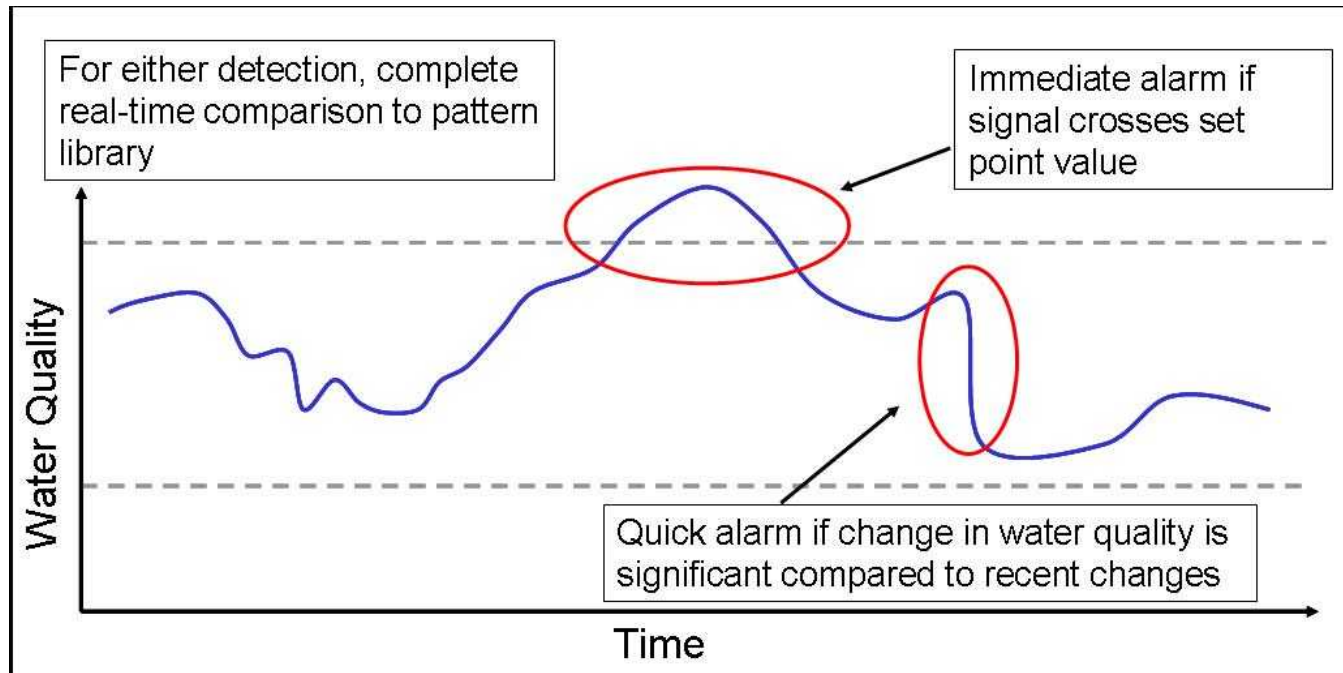
- Distance away from set point threshold is defined by user as a number of times of the precision of the sensor. Probability threshold can be set by user to provide as much advance warning as required



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

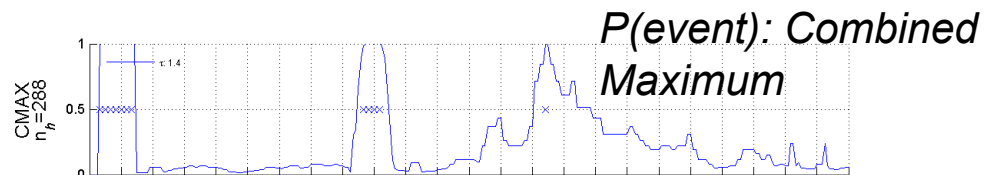
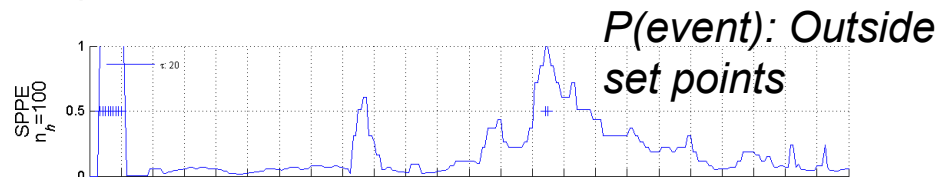
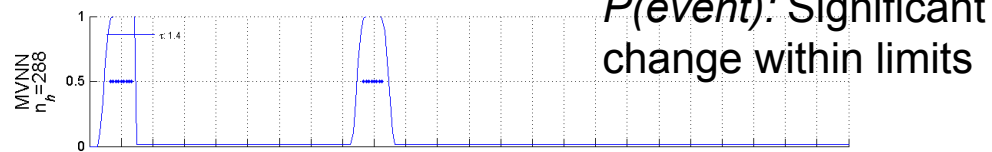
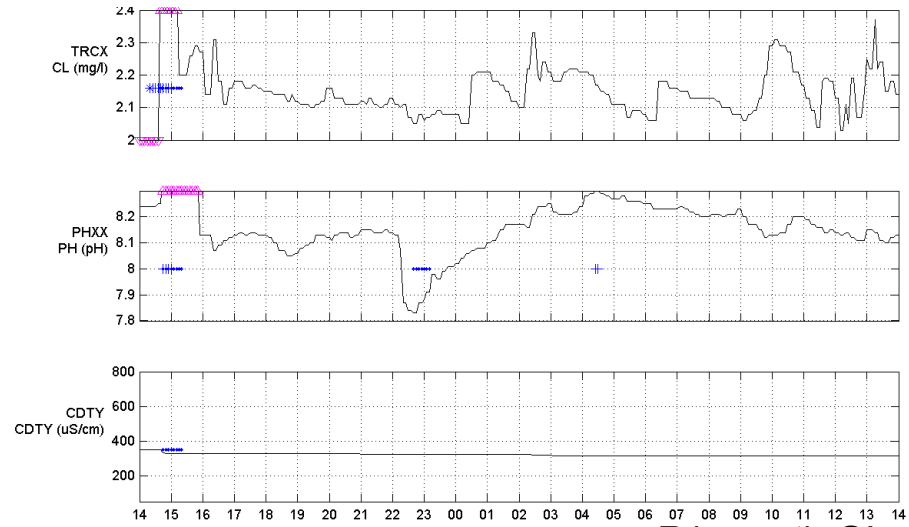
Consensus Algorithm

- 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



Consensus Algorithm Example

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

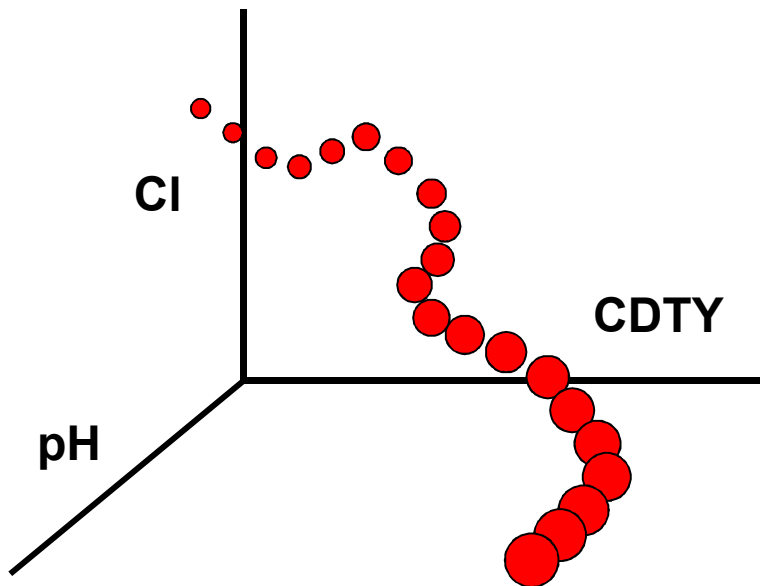


Pattern Matching

- Use the previously discussed algorithms to detect events that are expected
- Classify these detected events into different groups that are similar in some way
- The next time an event is detected, compare that water quality pattern to patterns in the library and look for a match

Water Quality Trajectory Concept

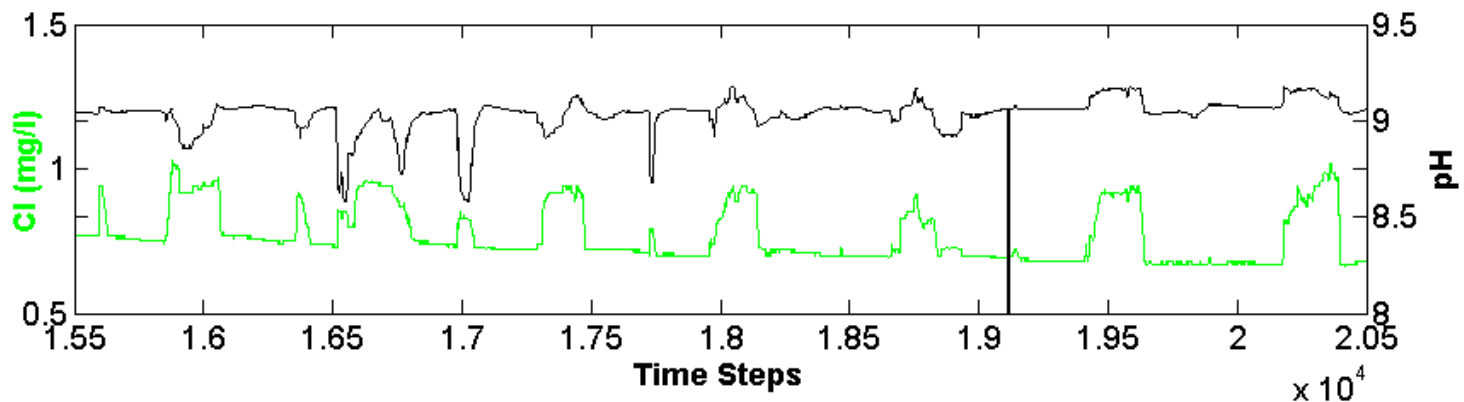
- Key element: Recognize that it is not classification of the water quality before and after an event that is important, it is classification of the *period of change* of the water quality



Implementation: Extend recently developed *trajectory clustering* approach to water quality data with realtime comparison

Pattern Matching

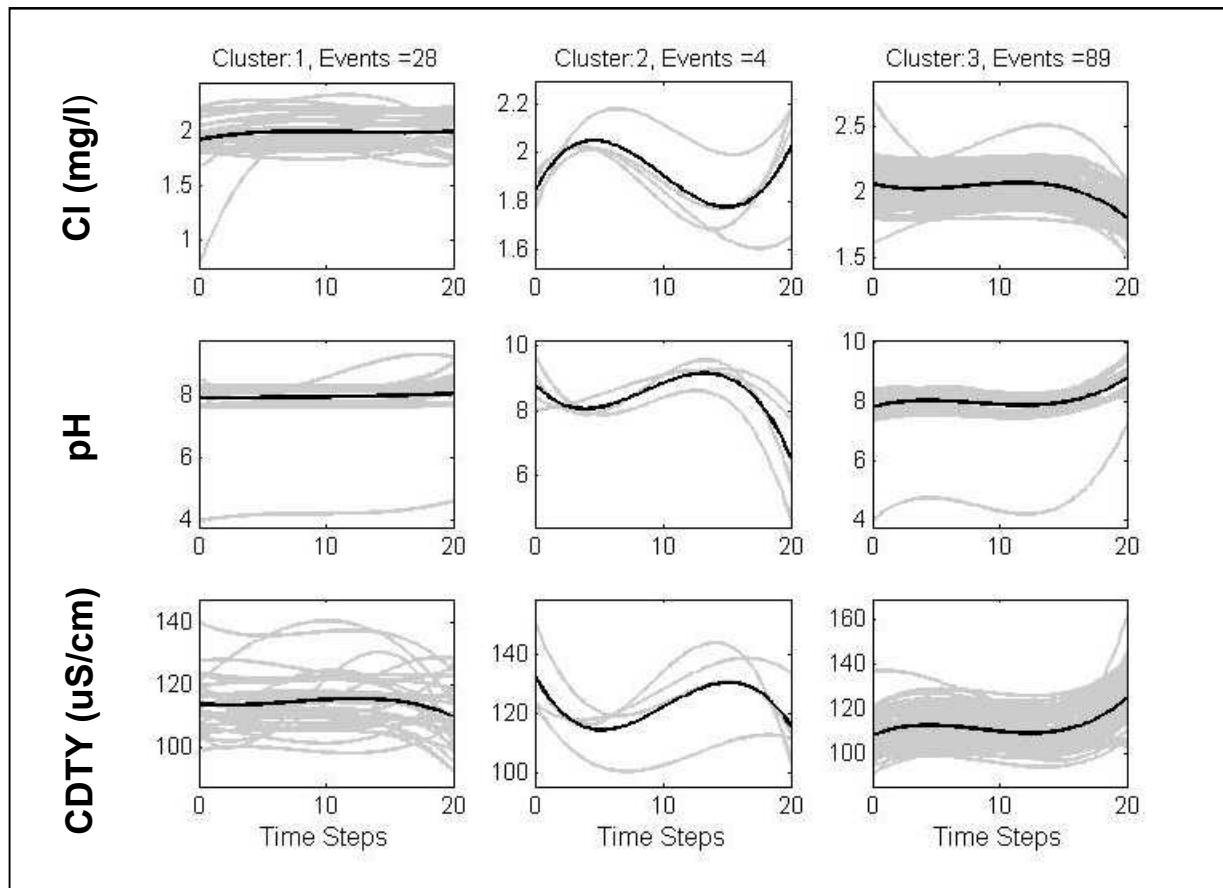
- Significant water quality changes can occur due to changes in network operations (draining tanks, pumps switching on and off, valve settings, etc.).
- CANARY has a pattern matching capability to recognize recurring, multivariate water quality patterns, store them in a library and then compare realtime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



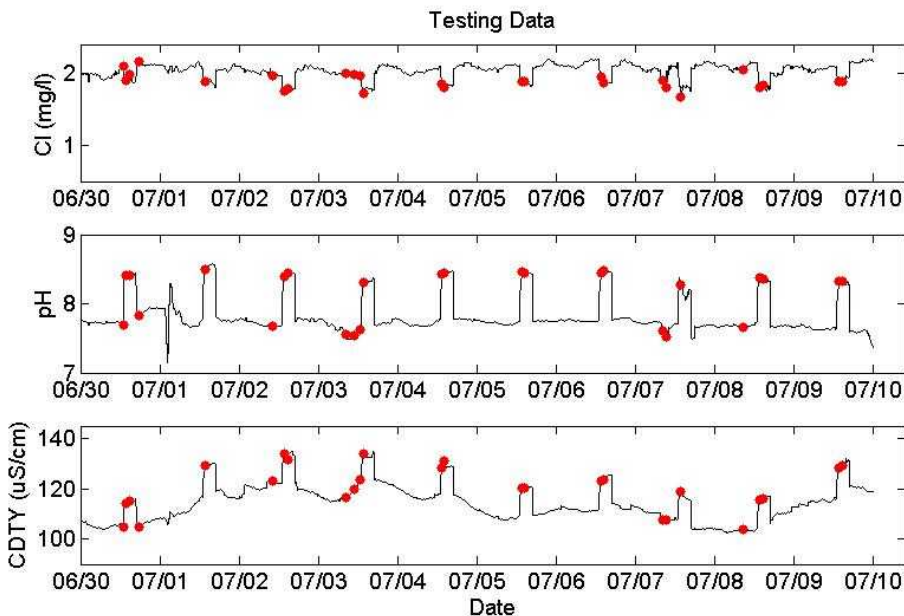
Example Pattern Matching Results

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

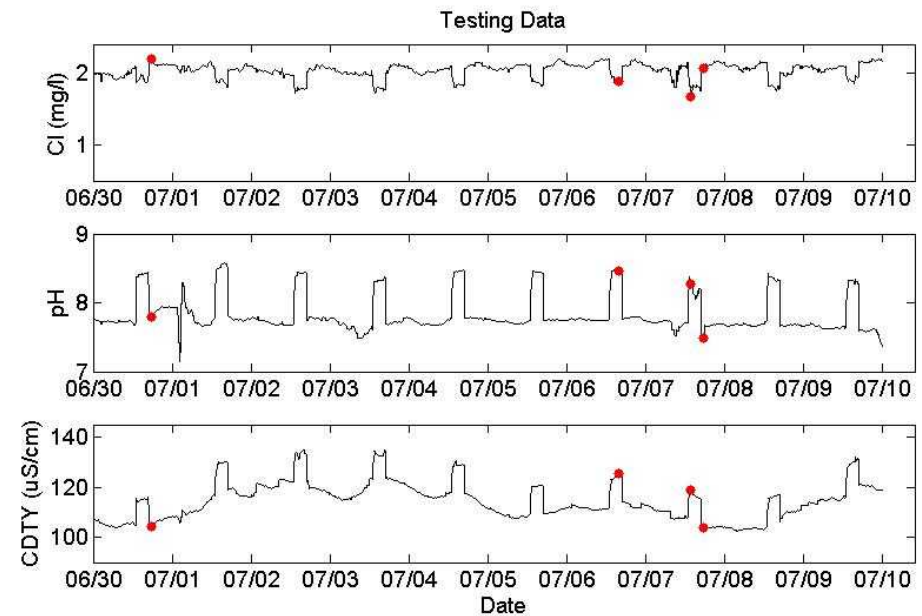
Pattern Matching Off: 65 alarms sounded

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

No Pattern Matching

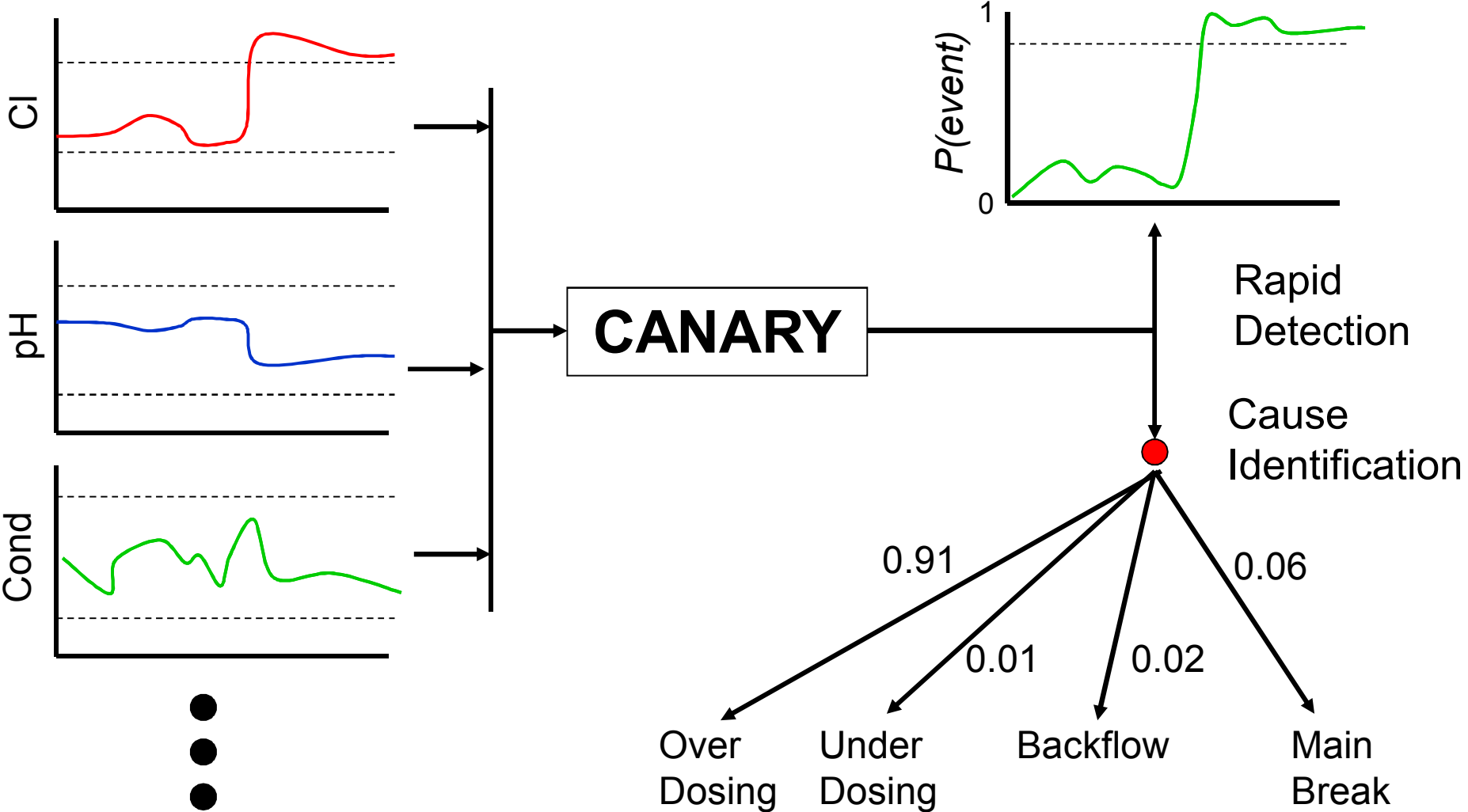


Pattern Matching



Example Results for 10 days of data

Recognizing Previous Patterns



Identification template matching based on historical data

Live Demos

- 1) Construction of a pattern library
 - Run CANARY with event detection algorithm parameters to create an edsc file
 - Analyze events in the edsc file to determine which ones to keep
- 2) Visualize and edit the resulting pattern library

Q&A Session

Submitted Question

What is causing the event and is this something that should be included in a pattern library?

