

# **Adaptive Water Sensor Signal Processing: Experimental Results and Implications for Online Contaminant Warning Systems**

Y. Jeffrey Yang, Ph.D., P.E.<sup>1\*</sup>, Roy C. Haught<sup>1</sup>, John Hall<sup>2</sup>, Jeff Szabo, Ph.D.<sup>2</sup>, Robert M. Clark, Ph.D., P.E.<sup>3</sup>, Greg Meiners<sup>4</sup>

1. US EPA, National Risk Management Laboratory, 26W Martin Luther King Dr., Cincinnati, OH 45268
2. US EPA, National Homeland Security Research Center, 26W MLK Dr., Cincinnati, OH 45268
3. Environmental consultant, 9627 Lansford Drive, Cincinnati, Ohio 45242
4. Shaw Environmental and Infrastructure, Inc., 1600 Guest Street, Cincinnati, Ohio 45204

## **ABSTRACT**

A real-time event adaptive detection, identification and warning (READiw) system has two principal functions. First, signal treatment using adaptive algorithms reduces background noise and enhances contaminant signals, leading to accurate detection of water quality changes of as low as 1%. Second, its forensic classification technique relates changes of water quality parameters to the reactivity of contaminants and hereby their chemical classes. To test these detection functionalities, contaminant transport experiments in a pilot-scale single pass pipe were conducted for 16 herbicides and pesticide, inorganic and biological contaminants. Sensor outputs (free and total chlorine, chloride, pH, DO, conductivity, ORP, and turbidity) were analyzed with the adaptive procedures. The results show unique changes of water quality parameters and the reactivity differences among the tested contaminants, based on which an effective READiw system can be configured.

## **INTRODUCTION**

Intentional sabotage can introduce chemical, biological and nuclear contaminants as well as toxic industrial chemicals into a drinking water system (McKone et al., 2003; GAO, 2003). Detection of such adverse events is required to protect Nation's water infrastructure and help water utility and local authorities timely activate contingent management plans. It is also a prerequisite for advanced risk forecasting and consequence management using methods such as those described in Murray et al. (2006).

A contaminant warning system (CWS) uses commercially available sensor technologies to identify contamination effectively in a water distribution system (Dye, 2002; US EPA, 2005; Szabo et al, 2007). Conventional water quality sensors are commonly used for their availability, favorable capital and operational costs. The disadvantage is that they are non-selective responding to introduced contaminants as well as natural variations. This compromises the detection sensitivity and accuracy, leads to false negative and high rates of false positive detections, and undermines the usefulness of a CWS and its acceptance by the end users.

EPA's approach for improved detection is based on real-time event adaptive detection, identification and warning (READiw) methodology and associated data processing techniques.

---

\* Corresponding author: U.S. Environmental Protection Agency, NRMRL/WSWRD, Mail Stop 690, 26W. Martin Luther King Dr., Cincinnati, Ohio 45268. Tel, (513) 569-7655; Fax, (513) 569-8155. yang.jeff@epa.gov

Adaptive monitoring and data mining techniques have been used in other applications such as the Navy's ship board biological detection system (Solka et al., 2003), computer network intrusion detection (Eskin et al., 2000; Lee et al., 2000; Kanevski et al., 2002; and Lin et al., 2001), space craft data processing and flight control (Mills et al., 2003), and soil and groundwater sampling optimization for environmental restoration (Romanowicz and Young, 2002). The adaptive signal treatment and analysis statistically filter out background and instrumental noises, enhance contaminant signals, and use response patterns for contaminant classification.

This paper describes an adaptive technique, and presents the results of using the signal treatment to identify contamination events in the datasets that U.S.EPA Homeland Security Research Center generated for 16 chemicals and biological contaminants (Szabo et al, 2007). The tested contaminants (Table 1) include pesticides and herbicide (aldicarb, dicamba, and glyphosate), biological contaminants and growth media (*Escherichia coli*, sucrose, nutrient broth, terrific broth, and trypticase soy broth), alkaloid neuron-stimulants (nicotine, colchicine), laboratory chemicals (Dimethyl sulfoxide, potassium ferrocyanide, anhydrous sodium thiosulfate, and pentahydrate sodium thiosulfate), and inorganic chemicals (mercuric chloride, lead nitrate).

Table 1 Chemical and biological compounds tested in the pipe flow experiments

Contaminant	Class
Aldicarb	Pesticide
Dicamba	Herbicide
Dimethyl sulfoxide (DMSO)	Lab chemical
<i>E.coli</i> with nutrient broth	Bacteria in growth media
Glyphosate	Herbicide
Colchicine	Neurostimulant
Lead nitrate	Toxic inorganic
Mercuric chloride	Toxic inorganic
Nicotine	Neurostimulant
Nutrient broth	Biological growth media
Potassium ferrocyanide	Lab chemical
Anhydrous sodium thiosulfate	Lab chemical
Pentahydrate sodium thiosulfate	Lab chemical
Sucrose	Biological growth media
Terrific broth	Biological growth media
Trypticase soy broth	Biological growth media

## EXPERIMENTS AND SENSOR SIGNAL TREATMENT

### *Experiments and Composite Dataset*

Contaminant transport experiments were performed at the EPA's Test and Evaluation (T&E) Facility using a pilot-scale 335.4-m long fiberglass-lined straight ductile iron pipe with a 7.62-cm inside diameter (Fig. 1). The pipe flow rate in all tests was 83.3 L/min ( $Re \sim 25,000$ ). Constant pressure and flow velocity in the pipe were established at 68.9-82.8 kPa and 0.305 m/sec, respectively, using a 2841-L holding tank that gravity-drains water into the pipe. Feed water from the tank had a free chlorine residual of  $1.0 \pm 0.1$  mg/L. Contamination was simulated by injecting 10 liters of contaminant solution into the pipe flow at a rate 0.5 L/min (Fig. 1).

Water quality sensor stations were placed at 24.4-m and 335.4-m downstream from the contaminant injection port (Fig.1). A split stream was diverted to a sensor panel to measure conventional water quality parameters: total chlorine, free chlorine, chloride, oxidation-reduction-potential (ORP), dissolved oxygen (DO), turbidity, pH, and specific conductivity. Travel time from the pipe to sensors was <5 minutes. The sensors used in the experiment included an ATI 15 for free chlorine, a Hach Cl-17 for total chlorine, and a YSI-6920 multiprobe pipe sonde for oxidation reduction potential (ORP), specific conductivity, pH, turbidity, dissolved oxygen, and chloride. The Hach Cl-17 instrument uses the *n,n*-diethyl-*p*-phenylene

diamine (DPD) colorimetric method and has a 0.035 mg/L detection limit at 5% accuracy and precision. The free chlorine monitor (ATI 15) utilizes an agent-free membrane-covered polarographic sensor with automatic pH correction in a flow-cell configuration, and has a reported 0.5% or 0.02 mg/L detection accuracy. The instrument was calibrated in standard solution before the experiments. Other parameters were measured using the YSI-6920 pipe sonde. The data acquisition was performed according to approved EPA quality control project plans.

Datasets representing composite test runs for all 15 contaminants are valuable in examining the detection capability of any CWS detection methods. One approach is to arbitrarily compose individual experimental test data into a single sensor response spectrum, in which differences in baselines and calibrations between individual experiments are preserved. Each contaminant was tested in two separate runs, and no reconciliation of sensor output differences was

made in forming the composite dataset. This treatment, although arbitrary in nature, has potentially preserved influences from different sensor calibrations and sensor operations between repeated test runs and among the contaminants. The composite dataset can simulate potential complications in a field operation of sensor monitoring network, serving as a testing case to verify event detection algorithms and their ability in reducing false positive and false negative rates. Original data used to form the composite dataset are available in Szabo et al. (2007).

### Sensor Output Treatment

In general, a water quality sensor output ( $I^t$ ) consists of contributions from contaminants of interests ( $I_c^t$ ), natural background ( $I_b^t$ ), instrument noise ( $I_n^t$ ), drifting ( $I_d^t$ ), and random variance related to operations ( $\sigma$ ). The background ( $I_b^t$ ) is zero or negligible for compound-specific sensors, but can be significant for non-selective water quality sensors for which signal output and a target concentration ( $C_i$ ) are related by:

$$I^t = I_c^t + I_b^t + I_n^t + I_d^t + \sigma \quad (1)$$

$$C_i = \beta I_c^t + \theta_i \quad \text{or} \quad I_c^t = \frac{C_i - \theta_i}{\beta} \quad (2)$$

The linear instrument response factor  $\beta$  is defined in calibration against a set of standard solutions and the intercept  $\theta_i$  is often forced to zero. These two instrumentation parameters are

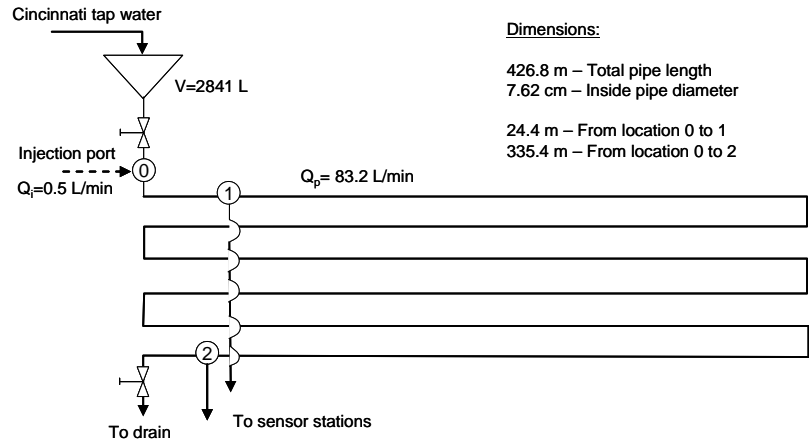


Figure 1 Pilot-scale single-pass drinking water pipe device used for experimental testing. Contaminant was introduced by injecting 10 liters of contaminant solution into the pipe at  $Q_i=0.5$  L/min.

independent of time, or instrumentation drifting occurs. After incorporating these concepts, sensor output concentration can be written as:

$$I_i^t = \beta \cdot C_i^t + \hat{m}F(t) + v(I_n^t, \sigma) \quad (3)$$

where  $\hat{m}$  is a constant matrix; the polynomial form  $F(t)$  is a function of time and represents combined sensor response from background ( $I_b^t$ ) and sensor drifting ( $I_d^t$ );  $v(I_n^t, \sigma)$  is normally distributed and independent of time to account for instrumental noise and operation-related random variances, which can be eliminated using time-series signal processing techniques. Eq.3 further indicates that the contaminant detection and its false rates are directly a function of relative strength between the time-dependant contaminant and background signals in a pipe flow system.

The component  $\hat{m}F(t)$ , with continuity in time, superimposes on the contaminant signal  $\beta \cdot C_i^t$ . For a successful contaminant detection, it is desired to separate these two signal components, suppress the non-contaminant signal portion  $[(\hat{m}F(t) + v(I_n^t, \sigma))]$ , and enhance the contaminant contributions. Yang et al. (2006) proposed an adaptive process to separate contaminant signal components in an adaptive time window ( $\bar{T}$ ) by identifying signal change points and differentiating signal variation types. This methodology is used in sensor data analysis in which concentrations are adaptively transformed into concentration ratios:

$$\Delta C_i^t = \frac{I_i^{t+1}}{I_0^{\bar{T}}} = 1 + \frac{\beta}{I_0^{\bar{T}}} C_i^{t+1} + \frac{[\hat{m}F(t)]^{t+1} - [\hat{m}F(t)]^{\bar{T}}}{I_0^{\bar{T}}} \quad (4)$$

In the pipe flow tests, controlled experimental conditions significantly damped the background variations. This simplification departs from field monitoring conditions, but makes it adequate to use the first 16 data measurements as the reference concentration ( $I_0^{\bar{T}}$ ). In processing the composite dataset of 15 contaminants, an adaptive procedure was adopted using a moving adaptive time window.

## RESULTS

### *Sensor Response Characteristics*

The free chlorine and total chlorine concentration data (Szabo et al, 2007) show U-shape residual loss curves on a concentration-time ( $c$ - $t$ ) plot for each sensor station (Fig. 2). The loss is the result of oxidation reaction between contaminants and the disinfectant. Correspondingly chloride as a final reaction product shows the concentration increase curve on the  $c$ - $t$  plots and a characteristic dispersion tail, which extends beyond the downstream boundary of the chlorine residual loss (Fig. 2). The tail is more pronounced for  $t > 2.8$  at the 335.4-m station. Similar V-shaped tails occurred for ORP that decreases as chlorine oxidants are depleted in the contaminant oxidation. Furthermore, pH changes occurred at the same  $t$ -range as for the other measured parameters. Detected and qualified pH changes are as low as approximately 1% using the adaptive technique.

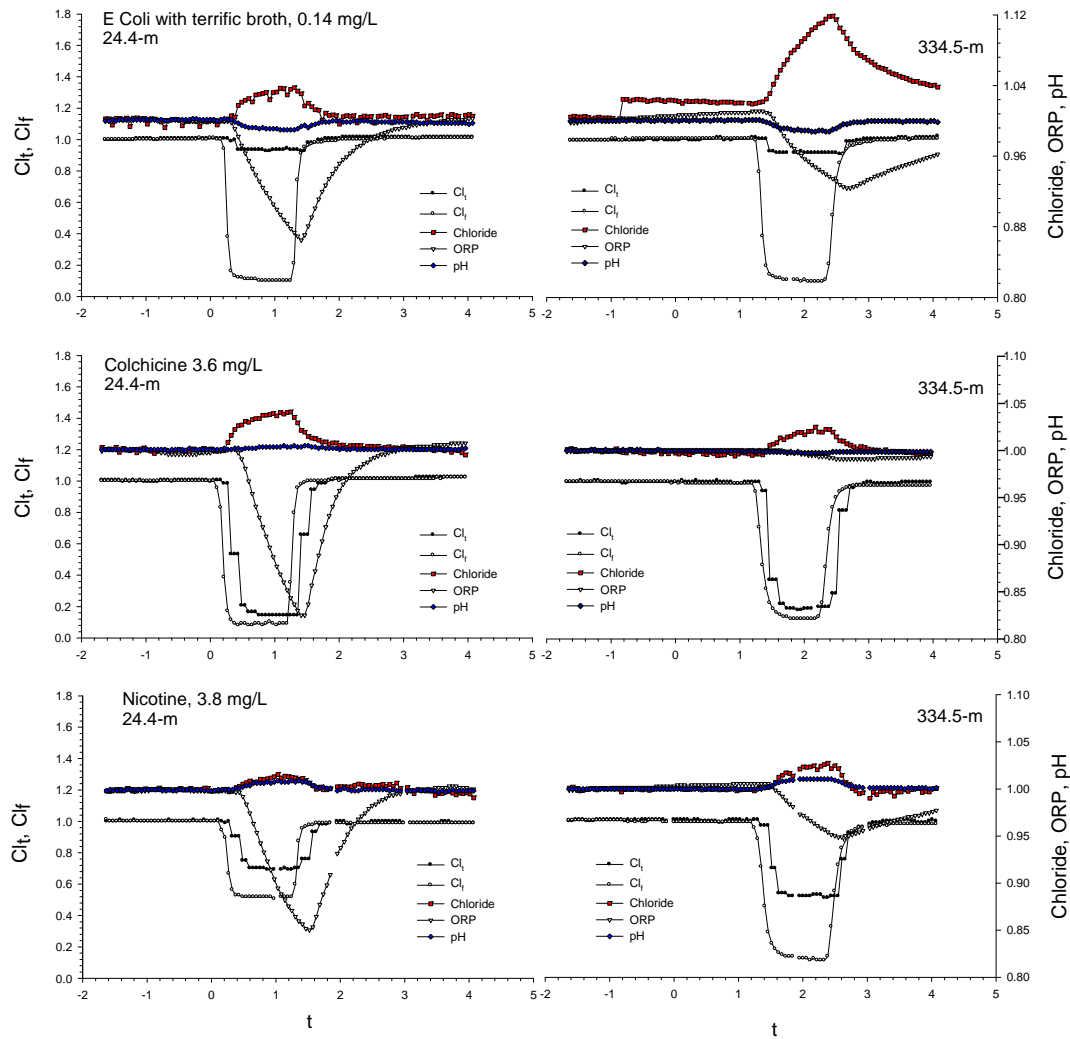


Figure 2. Sensor response patterns for *E.coli* and alkaloid colchicine and nicotine as detected at the 24.4-m and 334.5-m sensor stations. Y-axes are concentration ratios. X-axis is dimensionless time normalized to hydraulic retention time at the 334.5-m sensor station.

Chlorine residual loss  $\left[ \Delta c = 1 - \frac{C_t}{C_o} \right]$  is defined as concentration difference between the

background and the flat bottom of a residual loss curve on  $c$ - $t$  plots (Fig. 2). This calculation requires that  $C_t, C_o$  at time  $t$  and  $t=0$  be given by the residual concentration in the background and within the contaminant slug, respectively. Due to their variations in reactivity with chlorine, the tested contaminants have different  $\Delta c$  values and are clearly differentiated in the discrimination diagrams (Fig.3). The slopes on the plots are related to the kinetic reaction rates and the generation of chloramines during contaminant oxidation.

### Adaptive Detections

The artificially composed dataset for 15 contaminants has been processed adaptively. Selected results are shown in Figures 4 and 5 for free chlorine, total chlorine, specific conductivity, and pH. Results for the remaining sensor data (chloride, ORP, dissolved oxygen,

and turbidity) are not shown. For each parameter, direct sensor responses in absolute concentration values are plotted in the  $c$ - $t$  diagram along with the adaptively treated  $\Delta c$  values. Also shown in the figures are two injection events for each of the 15 tested contaminants. One injection lasted for 20 minutes.

#### Free and Total Chlorine

Differential sensor responses are observed for free chlorine and total chlorine in the experiment tests on the  $c$ - $t$  plots (Fig.4). Contaminants aldicarb, glyphosate, *E.coli* with nutrient broth, anhydrous and pentahydrate sodium thiosulfate all showed >80% free chlorine loss in the contaminant slug. Changes in total chlorine sensor responses as small as 3.15% occurred during the injection of several contaminants: dimethyl sulfoxide (DMSO), nutrient broth, terrific broth, and trypticase soy broth.

These sensor responses were correctly identified as contaminant events in the adaptive signal analysis (Fig.4). Compared to free chlorine, the composite sensor dataset for total chlorine contains large background variations. Some are due to sensor noise or background change, and some are the result of data sets from individual tests

arbitrarily integrated into the composite dataset. All 2268 chloride sensor measurements identified as background average at  $0.962 \pm 0.019$  mg/L

( $m \pm \sigma$ ). The smallest sensor response identified as a contaminant event is 0.03 mg/L or 2% ( $0.95$  mg/L and  $0.92$  mg/L for background and contaminant slug, respectively). This event identification for the lowest sensor response was made during the second terrific broth test (Fig.4). The difference is approximately 1.5 times of the background standard deviation ( $1.5\sigma$ ).

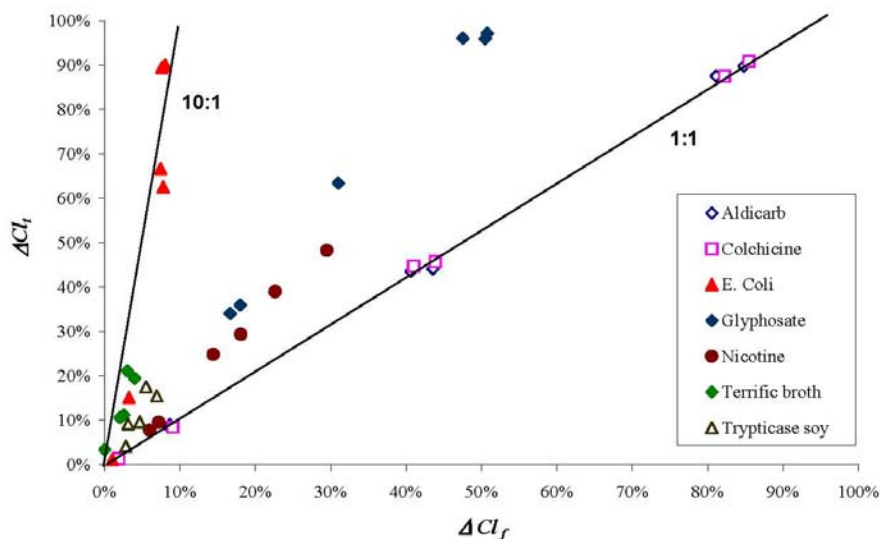


Figure 3 Forensic discrimination diagram using free and total chlorine loss for chemical and biological compounds at 3 concentrations in a real-time pipe flow testing. The differences among the contaminants reflect their reaction pathways and kinetic constants.

#### Specific conductivity and pH

Specific conductivity sensors registered small, but relatively significant increases during contaminant injections (Fig.5). The largest change occurred with potassium ferricyanide, shown by a peak of approximately 3.5-4.0 uS/cm from a background value of 348 uS/cm or a nearly 1% change. Sucrose registered a small increase in the conductivity, which was correctly identified in the adaptive analysis procedure. Conductivity signals generated from the nutrient broth injection are not characteristic of a contaminant slug in flow pipe, which led to it being excluded as a false signal in event identification. Other uncharacteristic instrument noises were observed in

glyphosate and trypticase soy broth injections (Fig.5). The pH data show large variations in background and between individual contaminant tests. For example, two anhydrous sodium thiosulfate injection events are marked by small pH changes of about 1% among the baseline variations. After adaptive corrections, the two events were enhanced among the flat baselines (Fig.5). False negative signals recorded in the *E.coli* and terrific broth tests were excluded from contaminant events largely based on the change curve geometry and adaptive event continuity in time.

#### ■ Chloride and ORP

Chloride sensor signals show large background variations, some of which were arbitrarily imposed in forming the composite dataset. They also contain an example of instrument failure at the time of second contaminant injection event for nutrient broth. Sensor response to the failure is representative of random variations of  $v(I_n^t, \sigma)$  in Eq.3, and can be identified using time

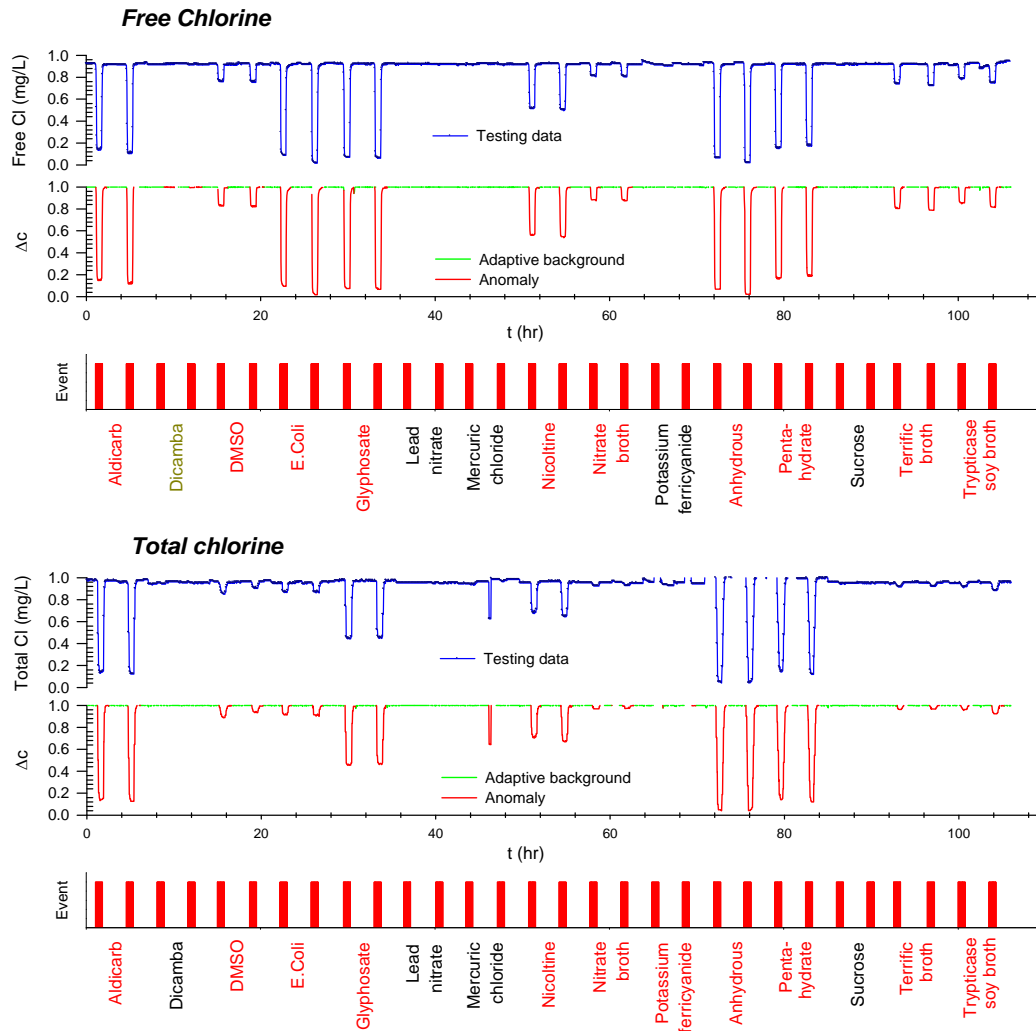


Figure 4 Free chlorine and total chlorine sensor output spectrum and adaptively treated data on the  $c$ - $t$  plot for the 15 contaminant composite dataset. Two contaminant injection events for each contaminant are shown for the duration of 20 minutes. Positively detected contaminants are marked in Red.

derivative signal treatment techniques. In nearly all experiments, chloride increased at the locations of the contaminant slug for aldicarb, *E.coli*, glyphosate, nutrient broth, potassium ferrocyanide, anhydrous and pentahydrate sodium thiosulfate, and trypticase soy broth. Two false positive events were identified for DMSO and sucrose. Such false identifications were corrected based on other sensor responses. The smallest sensor response change for a correct identification or the method resolution is 0.4 mg/L or 1.3% of chloride sensor output.

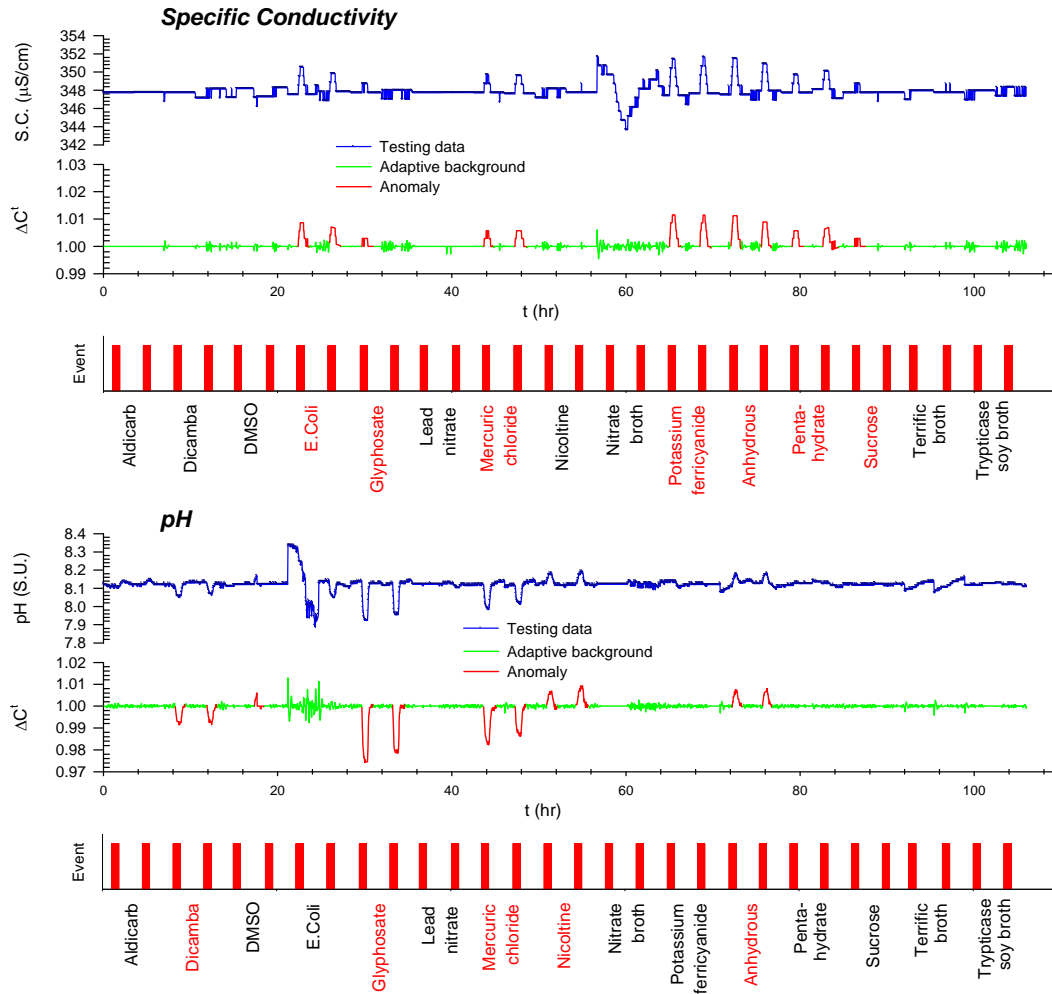


Figure 5 Specific conductivity and pH sensor output spectrum and adaptively treated data on the  $c$ - $t$  plot for the 15 contaminant composite dataset. Two contaminant injection events for each contaminant are shown for the duration of 20 minutes. Positively detected contaminants are marked in Red.

#### ■ DO and turbidity

In the dissolved oxygen composite dataset, the DO sensor signals contained extensive uncharacteristic and random variations that obscured the signals related to contaminant events. The noise also resulted in one false positive in the *E. coli* experiments. The adaptively treated sensor signals are randomly distributed around the baseline  $\Delta c = 1$ . Similarly the turbidity



signals are not indicative of the contaminants tested, except for lead nitrate, where the contaminant event was clearly identified.

## SUMMARY

The experimental results, through adaptive analysis, show positive identifications of the contaminant events (Table 2). Minimum sensor signal strength for a positive event identification is approximately 3-5% for chlorine and as little as 1% for pH and specific conductivity. The detection is based on the use of adaptive algorithms to suppress background variations and consequently enhance signals for change point detection on time-series plots. The data analysis technique successfully reduced the background variations and instrumental drifting into a flat line, transformed instrumental noises into oscillation patterns, and correctly identified all artificial anomalous changes in the composite dataset. The technique, however, was not capable of identifying and correcting data anomalies related to all instrumentation failures. False positives were observed for dimethyl sulfoxide and sucrose in pH, *E.coli* in DO, and mercuric chloride in chloride. When sensor detections for all parameters were analyzed together, such false identification could be detected and corrected.

The chlorine reactivity of a contaminant is its most characteristic property, based on which contaminant classes in reaction kinetics can be identified in forensic discrimination analysis. For example, aldicarb, glyphosate, and dicamba are pesticide and herbicides examined in this study. Aldicarb and glyphosate both caused significant and distinguishable free/total chlorine and chloride changes. However, unlike aldicarb, glyphosate generated a pH decrease of approximately 0.2 SU which was clearly defined in the adaptive signals (See Fig.5). The pH change reflects the hydrogen ion generation in the chlorine-induced oxidation of glyphosate (Brosillon et al., 2006), while aldicarb chlorination involves no hydrogen ion generation (Mason et al., 1990). Comparatively, the stable benzoic ring structure of dicamba makes it non-reactive with chlorine (Huston et al., 1999). Hydrolysis of the pesticide has been reported (Brosillon et al., 2006), and this was demonstrated as a clear pH anomaly in the adaptive monitoring (Fig.5). Collectively, the differential sensor responses are the basis for contaminant identifications.

## ACKNOWLEDGEMENT

Conclusions and opinions presented in this paper are those of the authors, and do not necessarily represent the position of the U.S. EPA and USDA. Mention of commercial products, trade names or services in the paper does not convey, and should not be interpreted as conveying official EPA approval, endorsement, or recommendation.

## REFERENCES

- Brosillon, S., D. Wolbert, M. Lemasle, P. Roche, and A. Mehrsheikh, 2006. Chlorination kinetics of glyphosate and its by-products: Modeling approach. *Water Research*, v.40:2113-2124
- Dye, D.H. (2002). Sensors for screening and surveillance. Conference on Technology for Preventing Terrorism, Hoover Institution National Security Forum, Stanford University.
- Eskin, E., M. Miller, Z-D. Zhong, G. Yi, W.-A. Lee, and S. Stolfo (2000) Adaptive model generation for intrusion detection. In Proc. ACMCCS Workshop on Intrusion Detection and Prevention, Athens, Greece.

- Government Accounting Office (2003). Expert's views on how future federal funding can best be spent to improve security, report to the Committee on Environment and Public Works, U.C. Senate. GAO-04-29, 69p.
- Huston, P.L., and J.J. Pignatello, 1999. Degradation of selected pesticide active ingredients and commercial formulations in water by the photo-assisted Fenton reaction. *Wat. Res.* vol. 33: 1238-1246.
- Kanevski, M., A. Pozdnukhov, S. Canu, and M. Maignan (2002). Advanced spatial data analysis and modeling with support vector machines. *Intl J. Fuzzy Systems*, 4(1): 606-615.
- Lee, W., S.J. Stolfo, P.K. Chan, E. Eskin, W. Fan, M. Miller, S. Hershkop, and J. Zhang (2000) Real-time data mining-based intrusion detection. *Proc. 2<sup>nd</sup> DARPA Information Survivability Conference and Exposition*.
- Lin, W., M.A. Orgun, and G.J. Williams (2001) Temporal data mining using hidden Markov-local polynomial model. *7<sup>th</sup> ACM SIGKDD Intl Conference on Knowledge Discovery and Data Mining (KDD-2001)*, San Francisco.
- Mason, Y., E. Coshen, and C. Rav-Acha, (1990). Carbamate insecticides: Removal from water by chlorination and ozonation. *Wat. Res.* 24(1), 11-21.
- McKone, T.E., B.M. Huey, E. Downing, and L.M. Duffy, (2003). Strategies to protect the health of deployed U.S. Forces – Detecting, characterizing, and documenting exposures. National Academy Press, Washington, D.C., 242p.
- Mills, C.S., G.Hines, K.R. Fowler, M.A. Garrison Darrin, R.F. Conde, and H.A.C. Eaton, (2003). Adaptive data analysis and processing technology (ADAPT) for spacecraft. *ESTC*, June, 2003.
- Murray, R., J. Uber, and R. Janke (2006). Model for estimating acute health impacts from consumption of contaminated drinking water. *J. Water Res. Planning and Management*, v.132: 293-299.
- Romanowicz, R.J., and P.C. Young, (2002). Data assimilation and uncertainty analysis of environmental assessment problems. *iEMSs 2002*, Lugano, Switzerland.
- Solka, J.L., E.J. Wegman, and D.J. Marchette (2003). Data mining strategies for the detection of chemical warfare agents. In H. Bozdogan, ed., *Statistical Data Mining and Knowledge Discovery*, p.57-92.
- Szabo, J.G., J.S. Hall and G. Meiners (2007). Water quality sensor responses to contamination in a single pass water distribution system simulator. EPA/600/R-07/001, Water Information Sharing and Analysis Center (WaterISAC) ( [www.waterisac.org](http://www.waterisac.org)).
- U.S.EPA, 2005. Technologies and techniques for early warning systems to monitor and evaluate drinking water quality: State-of-the-art review. ORD, NHRSC, Washington, DC.
- Yang, Y.J., R.C. Haught, J. Hall, J.A. Goodrich, and J. Hasan, 2006. Adaptive monitoring to enhance water sensor capabilities for chemical and biological contaminant detection in drinking water system. *Proc. of SPIE*, 6203, 6203K-1 to K-12.