

## Research Article

# Online Monitoring of Water-Quality Anomaly in Water Distribution Systems Based on Probabilistic Principal Component Analysis by UV-Vis Absorption Spectroscopy

**Dibo Hou, Shu Liu, Jian Zhang, Fang Chen, Pingjie Huang, and Guangxin Zhang**

*Department of Control Science and Engineering, State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China*

Correspondence should be addressed to Dibo Hou; [houbd@zju.edu.cn](mailto:houbd@zju.edu.cn)

Received 23 March 2014; Revised 25 May 2014; Accepted 25 May 2014; Published 19 June 2014

Academic Editor: Adam F. Lee

Copyright © 2014 Dibo Hou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study proposes a probabilistic principal component analysis- (PPCA-) based method for online monitoring of water-quality contaminant events by UV-Vis (ultraviolet-visible) spectroscopy. The purpose of this method is to achieve fast and sound protection against accidental and intentional contaminate injection into the water distribution system. The method is achieved first by properly imposing a sliding window onto simultaneously updated online monitoring data collected by the automated spectrometer. The PPCA algorithm is then executed to simplify the large amount of spectrum data while maintaining the necessary spectral information to the largest extent. Finally, a monitoring chart extensively employed in fault diagnosis field methods is used here to search for potential anomaly events and to determine whether the current water-quality is normal or abnormal. A small-scale water-pipe distribution network is tested to detect water contamination events. The tests demonstrate that the PPCA-based online monitoring model can achieve satisfactory results under the ROC curve, which denotes a low false alarm rate and high probability of detecting water contamination events.

## 1. Introduction

Drinking water is delivered through water distribution systems after being carefully treated to fulfill the requirements of the water-quality standards established by the government [1]. However, water distribution systems (WDS) are inherently vulnerable to both the corrosion of the metal pipe materials in the distribution systems and various exterior contaminants such as intentional sabotage or terrorist attacks. The contaminants in WDS are likely to pose significant threats to public health, such as the drinking water contamination event in Walkerton, Ontario, Canada, in 2000 [2]. The existing laboratory-based analytical methods or exceedance-criteria alarm methods cannot meet the needs of real-time, multiple-parameters, and high-accuracy detection of water events [3].

A water-quality event is defined as a time period over which water with anomalous characteristics is detected [4].

Conventional water-quality event detection methods monitor online events indirectly by detecting physical and chemical water-quality parameters such as pH, chlorine, conductivity, oxygen-reduction potential, and turbidity. Empirical evidence shows that common water-quality parameters are sensitive indicators of many contaminants, such as nicotine, arsenic trioxide, aldicarb, and *Escherichia coli*. However, using common water-quality parameters for anomaly monitoring involves manually operated chemical treatment, which takes relatively longer time, making it unsuitable for online monitoring application. Moreover, considering the aforementioned parameters, a sensor's electrical signal output can be significantly affected by multiple interferences such as instrument noise and baseline drift, which can generate signals resembling intentional contamination and in turn lead to a high rate of false detection [5].

Ultraviolet-visible (UV-Vis) spectroscopy uses absorption of the spectrum in visible and near-infrared (NIR)

ranges to detect chemicals in the water body. Compared with traditional water-quality online monitoring, UV-Vis spectral analysis technique includes advantages such as simple operation, no reagent, excellent repeatability, and rapid detection. Previous research has focused on the modeling of spectral data and water parameters by single wavelength UV spectroscopy. A wavelength of 254 nm is generally selected for its sensibility to total organic carbon (TOC) [6] and chemical oxygen demand (COD) [7], and 280 nm is chosen for that of the biochemical oxygen demand (BOD) [8]. To exclude the influences of turbidity and particles in the water sample, additional wavelengths were introduced as a sensitivity references. Pairs of 254 nm and 350 nm in addition to 254 nm with 580 nm have been applied for measuring the organic pollution in water samples [9]. Although both single wavelength and double wavelength are simple to operate, they are solely suitable for samples with simple constituents. With the demand to identify the enormous number of contaminants with broader spectrum coverage, a spectral analysis strategy has been extensively applied in which the information contained in UV-Vis spectra is used to the highest extent [10]. With the spectral analysis strategy, the rich information contained in the spectral data ideally compensates for the loss in compound-specific information [11]. However, most existing UV-Vis spectral analysis methods, either single wavelength or full-spectrum, only focus on the quantitative analysis of predetermined water-quality parameters or contaminants. In 2006, Langergraber proposed a qualitative analysis method to detect the abnormality in the water-quality series based on statistical analysis [11], but there is still no further research on the anomaly detection of water-quality contamination events.

Traditionally, many selective algorithms have been studied and developed for spectral analysis. Deconvolution methods were applied for wastewater quality monitoring [12], and a modified method was used for DOC estimation [13]. A rapid analytical method was proposed for oxygen demand (OD) in wastewater with artificial neural networks (ANN) [14], and a support vector machine (SVM) was introduced for UV spectral water-quality analysis [15]. Furthermore, principal component analysis (PCA) and partial least square (PLS) were combined for dimensionality reduction and determination of suspended solids, COD, and nitrates [16]. The characteristics of these traditional methods are summarized as shown in Table 1.

From the summary, it can be concluded that deconvolution, ANN, and SVM are suitable for quantitative analysis for water-quality and are possible to be applied to contaminants identification. As for PCA and PLS, although they achieve satisfying performance in dimensionality reduction of spectral data, the linear model of PCA and PLS is less flexible in dealing with external disturbance. Moreover, there is no application of approaches to draw a combination of information within each dimensionality of the principal components (PCs) for efficient anomaly detection. In this paper, we propose a water-quality event detection method that uses probabilistic principal component analysis (PPCA) together with multivariate monitoring chart to develop

TABLE 1: Summary of traditional approaches for water-quality event detection with UV-Vis.

Approach	Benefits	Confines
Single/double wavelength	Well-developed, simple	Vulnerable to external influences
Deconvolution	Combination of mechanisms	Require preacknowledge of contaminant category
ANN or SVM	Nonlinear models	Large number of samples are required
PCA with PLS	Linear decomposition with feature extraction	Lack of flexibility

a comprehensive indicator of PCs with a more flexible probabilistic model.

## 2. Methodology

*2.1. Probabilistic Principal Component Analysis.* PCA is a well-studied multivariate statistical dimensionality reduction of the observation data. This method has been applied to various fields such as image processing, data compression, time series analysis, and pattern recognition [17]. For a set of  $d$ -dimensional observation vectors,  $\{t_k \in \mathcal{R}^d, k = 1, 2, \dots, n\}$ , the PCs are obtained through eigen decomposition or singular value decomposition (SVD) by searching for the direction with the highest variance. PPCA, however, acquires its PCs through a probabilistic approach, with an expectation-maximization (E-M) algorithm for parameters estimation, which makes PPCA a more flexible method for satisfactory dimensionality reduction [18].

Latent factor analysis relates the observation to the latent variable, which is denoted as

$$t = Wx + \mu + \varepsilon. \quad (1)$$

$W = (\omega_1, \omega_2, \dots, \omega_q)$  is a  $d$  by  $q$  matrix that relates the observations to latent variables. Vector  $\mu \in \mathcal{R}^d$  represents the mean of the observation variable  $t$ , and  $\varepsilon$  is the noise or error variable.

In Tipping and Bishop's original PPCA model, latent variables  $x$  are defined to be independent and conform to normal distribution  $x \sim N(0, I)$ , where  $I \in \mathcal{R}^{q \times q}$  indicates the identity matrix. The distribution of the noise variable  $\varepsilon$  is introduced as  $\varepsilon \sim N(0, \Psi)$  where  $\Psi = \sigma^2 I$  is presumed. Moreover, when  $\sigma = 0$ , PPCA degenerates to PCA. Hence, PCA is a special PPCA under a specific circumstance.

*2.2. Multivariate Monitoring Chart.* Statistical process control (SPC) concepts and methods are widely applied to industrial practical process [19]. The main objective of SPC is to monitor the industrial processes and verify their controllable states. Several popular control charts for monitoring single quality characteristic have been developed, such as Shewhart, cumulative sum (CUSUM), and the exponentially weighted moving average (EWMA) charts [20]. In practice,

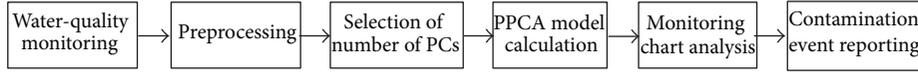


FIGURE 1: Schematic of proposed probabilistic principal component analysis- (PPCA-) based monitoring method.

industrial processes simultaneously require multiple quality characteristics correlated with corresponding measurements. Therefore, a multivariate statistical monitoring (MSPM) approach is developed, within which various multivariate control charts have been introduced [21]. Combined with PPCA, probabilistic models provide a novel view of such issues [22].

As  $x \sim N(0, I)$  is presumed, the squared Mahalanobis norm of  $x$  could be proved to conform to  $\chi_{(q)}^2$  distribution. The latent variable  $x$  cannot be acquired simply by operating eigen decomposition of the covariance matrix as PCA algorithm does. Thus, the latent variable  $x$  is substituted by its estimation,

$$\tilde{x} = E(x | t) = M^{-1} \cdot W^T \cdot t, \quad (2)$$

which is expected to conform to  $\|\tilde{x}\| \sim \chi_{(q)}^2$ . Here  $M$  is defined as  $M = W^T \cdot W + \sigma^2 \cdot I$ . Hence the following formula is used for normal monitoring:

$$\|\tilde{x}\| \leq \|x\|_{\text{lim}}, \quad (3)$$

where  $\|x\|_{\text{lim}}$  indicates the control limit for an event. Any point exceeding the control limit will be potentially regarded as anomaly or an event.

**2.3. Water-Quality Contamination Event-Detecting Model Based on PPCA and Monitoring Chart.** In this study, PPCA and the multivariate monitoring chart are combined to develop an early warning detecting method for online water-quality monitoring. The proposed method has six stages: (1) water-quality monitoring; (2) preprocessing; (3) calculation of the number of principal components; (4) PPCA model calculation; (5) monitoring chart analysis; and (6) contamination event reporting. Figure 1 shows the stages of the proposed method.

**2.3.1. Preprocess for the Algorithm.** Spectral online monitoring of water-quality relies on the online spectrometer sensors located at the essential parts of the distribution network, such as the entrance and termination points of the distribution network. Such sensors store and return the spectral data  $T$  for water-quality analysis in real time.

For every updated sliding window, after initially being collected, the spectra data  $T$  should be first normalized to the normalized observation variable  $t$  to dampen the noises from various resources during the spectrometer detecting process. In addition, the baseline of the background water may drift after a lengthy detection process, which can lead to a higher false alarm rate and deteriorate the performance of the monitoring system. Hence, normalization plays a quite critical role in the data processing.

Then, it is quite essential to decide the dimensions of the latent variable  $x$ . In the PPCA model, the approximation of observation variable  $t$  is the sum of the transformation part from latent variables  $Wx$ , the mean of the observation variables  $\mu$ , and the noise part  $\sigma^2 I$ .

**2.3.2. PPCA Model Calculation.** After the preparation, outer interferences are largely reduced by normalization. Then, estimation of the parameters  $W$  and  $\sigma^2$  can be acquired by the expectation-maximization (E-M) algorithm, which is shown in Figure 2. Initially, the log likelihood of observed data  $t(i)$  is given as  $\ell_C = \sum_{i=1}^N \ln p(t_i, x_i)$ , where  $p(t_i, x_i)$  represents the joint probability density between  $t_i$  and  $x_i$ . The values of  $w(i)$  and  $\sigma(i)^2$  were then initialized for the following maximization step. In this step, the expectation of  $\ell_C(i)$  is maximized with the initial parameters,  $w(i)$  and  $\sigma(i)^2$ , for the updated parameters,  $\bar{w}(i)$  and  $\bar{\sigma}(i)^2$ . The iteration process follows these E-M steps until termination is reached. Subsequently, the latent variables  $x$  are acquired with the latest parameter pair,  $W$  and  $\sigma^2$ , together with the originally collected observed data  $t$ .

**2.3.3. Monitoring Chart Analysis.** With the estimated parameters,  $W$  and  $\sigma^2$ , the latent variables can be generated, as demonstrated in Section 2.2. The Mahalanobis norm of the latent variables  $x$  is then employed as the indicator for the monitoring chart. Then, during real-time monitoring, newly detected points enter the sliding window and exclude the oldest ones in the window. Furthermore, control limit is decided by the distribution the Mahalanobis norm follows. However, with ROC as an evaluation method, fixed control limit is no longer used. Instead, a threshold changing from 0 to 1 to measure PD and FAR at the same time is employed.

**2.4. Performance of the PPCA Model.** The receiver operation characteristic (ROC) curve refers to the receiver operation characteristic curve, which originates from the evaluation of radar-receiving performance and is currently applied for the medical field, industrial quality control, and anomaly detection. Moreover, it employs several parameters to evaluate the algorithm's performance, such as probability of detection (PD), false alarm rate (FAR), and false classified rate (FCR). When applied to water-quality detection, PD represents the number of detected anomaly events out of the total number of anomaly events that actually occurred within a particular period. Similarly, FAR denotes the number of false alarms out of the total number of the alarms within a period. FCR demonstrates the number of those events without an alarm.

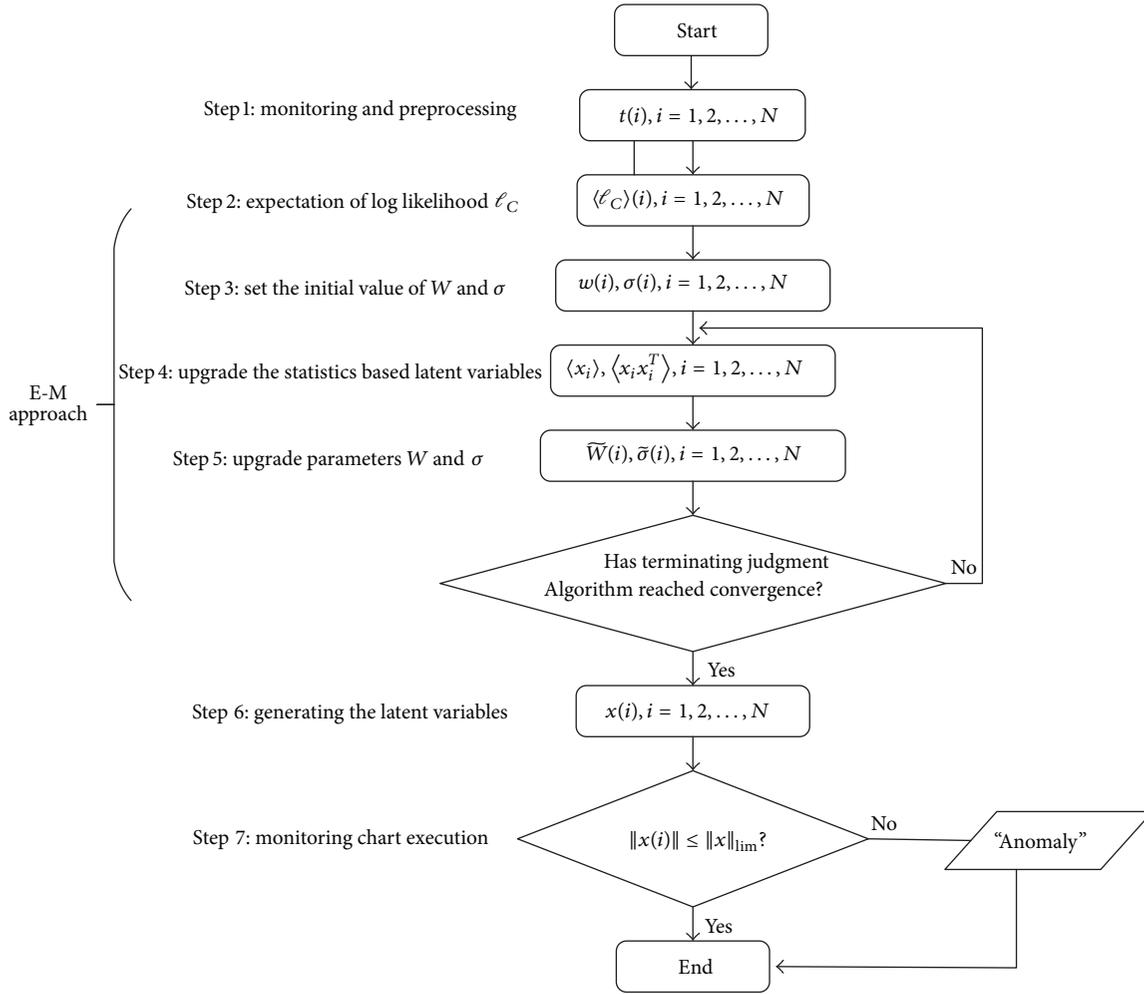


FIGURE 2: Real-time calculation at the time step  $k$  of probabilistic principal component analysis (PPCA)—monitoring chart detection method.

Table 2 gives four fundamental circumstances in actual water-quality, which are calculated as follows:

$$\begin{aligned}
 PD &= \frac{TP}{TP + FN} \times 100\% \\
 FAR &= \frac{FP}{TN + FP} \times 100\% \\
 FCR &= \frac{FP + FN}{FP + TN + TP + FN} \times 100\%.
 \end{aligned} \quad (4)$$

The ROC curve is introduced here to test the ability of the approach to discriminate between normal and anomalous water-quality by using a moving threshold from bottom to top. It employs PD and FAR as the axis and uses the area under the ROC curve to evaluate the performance of the algorithm.

### 3. Experiments

*3.1. Experiment Strategy.* The experiment was conducted in the mini drinking-water distribution system of water-quality detection and monitoring laboratory in Zhejiang University, Hangzhou, China. The structure of the distribution system

TABLE 2: Signs for four actual circumstances of water-quality.

	Actual water-quality normal	Actual water-quality abnormal
Evaluated water-quality normal	True negative (TN)	False negative (FN)
Evaluated water-quality abnormal	False positive (FP)	True positive (TP)

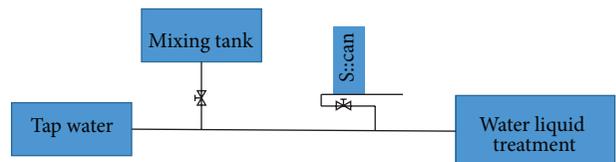


FIGURE 3: Experiments using the distribution device.

is shown in Figure 3. The main pipe is induced in the tap water through the distribution system. A valve on the main

pipe controls the flow at approximately 300 L/h, which is commanded by a control computer. A branch pipe joint is the main pipe at the point A for contaminant injection to simulate the water-quality event. Before injection, the contaminant was first dissolved in the mixing tank. A metering pump mainly controls the flow of contaminant injection by implementing the commands from the central control unit. Then, at point B, an additional branch pipe is inducted in the main flow at the section in which the various sensors are located. Afterward, the wastewater is eventually collected into a waste collection system.

During the contamination event simulation, several factors were required to achieve authentic water-quality events. The main stream kept flowing at the set flux, and the metering pump remained open during the simulations. Three groups of events were tested in respect to event severity. For each group, events lasted for varied periods. Furthermore, the injections of the contaminant were set at random times. In addition, spectrometer scanned the water flow at intervals of 0.5 min; thus, the spectrometer could collect two integral spectral datasets within a period of 1 min.

The contaminant used was phenol. Phenol and other phenolic compounds are common pollutants in natural water, particularly in areas adjacent to chemical plants. Phenol vapor and phenol itself are detrimental to human health because they can burn the skin or harm the central nervous system. Furthermore, because it is apparently soluble in water, a large amount of phenol in the water supply would be a huge threat to health.

Spectra within the wavelength range of 200–750 nm were generated when the main flow passed through the spectrometer within the pipe. The probe used was the Spectro:lyser, which is produced by S::can. The probe was submerged in the pipe and scanned the flow at intervals of 30 s without sampling. Unlike traditional cabinet analyzers, this probe is capable of online measurement with no consumption of chemicals. The metering pump injected the contaminant at a stable flow commensurate with the main flow during the experiment. The raw spectral data captured by the sensor were automatically stored and were able to be obtained directly.

To demonstrate the manner in which phenol contamination would influence the UV-Vis spectra, an example of the spectrum for the raw event is presented in Figure 4. Furthermore, to indicate this influence, the absorption at a single wavelength is also presented in the figure.

The data adopted in the following analysis was primarily based on the experiment operated from 06:21:30, 25 Jan., 2014 to 21:22:00, 25 Jan., 2014. Six water-quality events were simulated as different time span, with same concentration of phenol as 100 ppb (1 ppb is equal to  $1 \mu\text{g/L}$ ). The time span of each event was randomly set. And start time of each event is 10:15:30, 11:45:30, 15:25:30, 16:44:00, 18:15:00, and 20:04:30. As the system was continuously in operation, the status of equipment was relatively stable during the process. In addition, drinking water in a day stayed in similar status, contributing to less baseline drift.

When adding contaminant into the system, it is required to make phenol solution first. In order to simulate the event

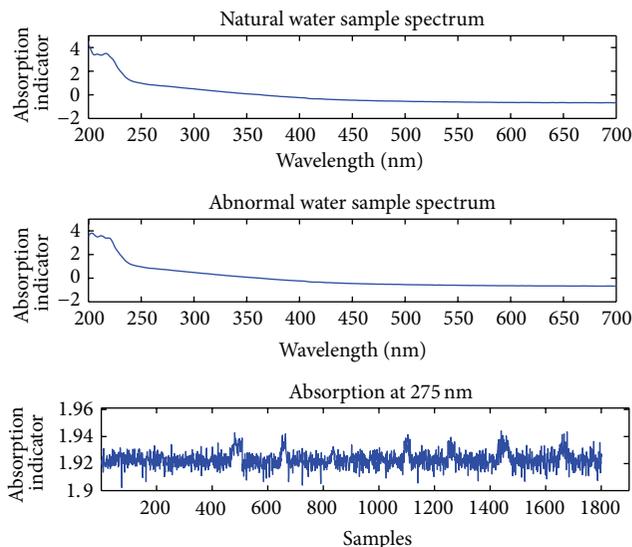


FIGURE 4: Online monitoring spectra.

equivalent to phenol event at the concentration of 100 ppb, the phenol solution of 5 ppm was added into the mixing tank, where the phenol solution and network water were fully mixed. As the contaminant was added to the system in a proportion of 1/50 to the main flow which was controlled by the central computer of the system, the contaminant finally entered the main flow which could be approximately deemed as with the concentration of 100 ppb.

The experiment of 50 ppb and 30 ppb for comparison followed the similar process.

**3.2. Probabilistic Principal Component Analysis.** The main objective of PCA is to reduce the dimensionality of the raw spectra for an acceptable number of PCs. These PCs contain the most essential water-quality information, with which water-quality monitoring performance could be significantly improved in accuracy and response speed.

When the latest data is detected, it is acquired through the sliding window, and the oldest data are simultaneously eliminated. For spectra in each window, to better qualify the data for event detection, the analysis was constrained in the section from 230 nm to 400 nm, since lower wavelength section might be significantly influenced by pure water and higher wavelength section possibly shows nonbeneficial information for the analysis. Normalization is then implemented in the preprocessing step to decrease the interferences from external environment. Subsequently, iteration of the EM algorithm is applied to the spectra data within the sliding window for the optimal PCs.

To evaluate the performance of PPCA model, the acquired PCs are used for reconstruction of the original spectra. The optimal dimensionality-reducing algorithm is designed to retain as much useful information as possible in the PCs. Moreover, when transforming back, the results should approximate the original spectra data to the closest extent. Figure 5 indicates the information retention ability of

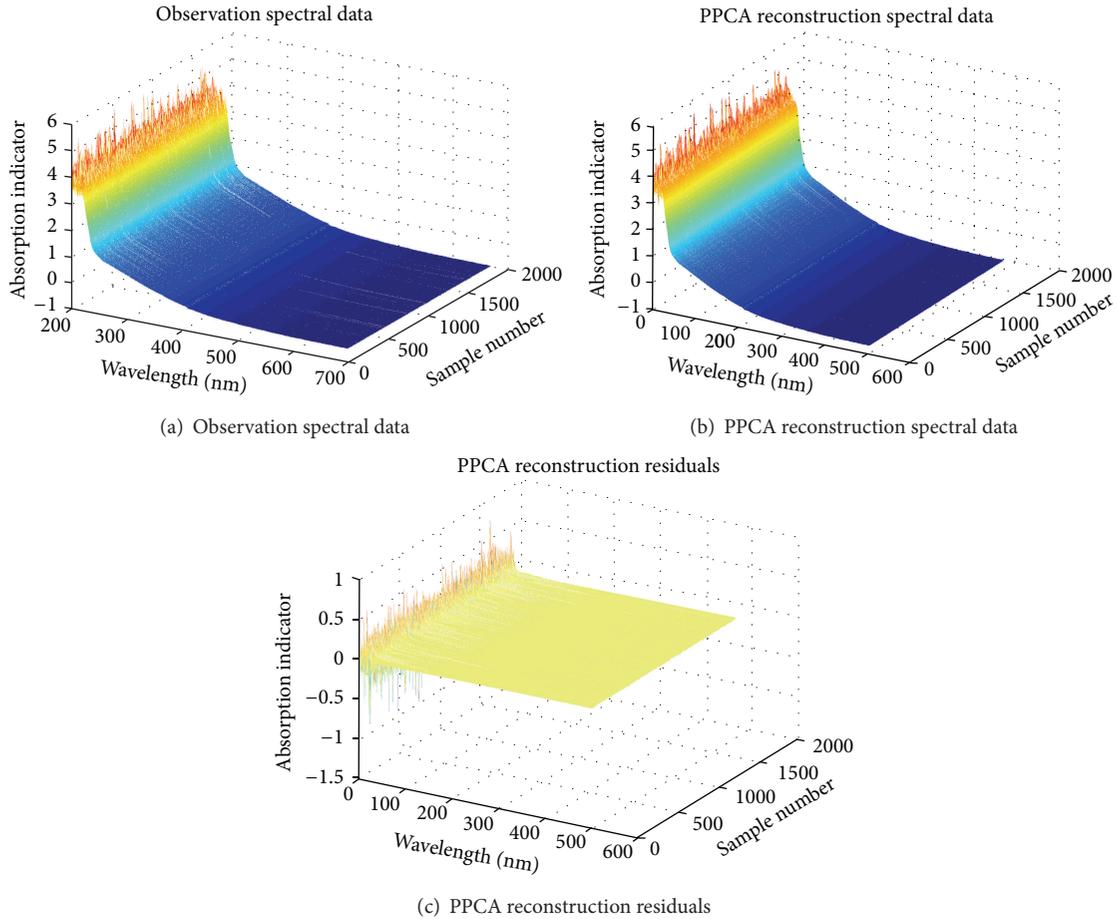


FIGURE 5: Probabilistic principal component analysis (PPCA).

the PPCA model. Figure 5(a) represents the original spectra data, and Figure 5(b) represents the results of reconstruction. Furthermore, for a more apparent illustration, the difference between Figures 5(a) and 5(b) is shown in Figure 5(c). By comparing these three figures, it is apparent to identify that the error between original and reconstruction mostly lies within the spectra range from 200 nm to 220 nm. The UV spectra in this range are most vulnerable to the absorption of pure water, and such section in spectra is always omitted in further analysis.

**3.3. SPC Monitoring Chart Implementation.** After acquiring the PCs, the multivariate monitoring chart was utilized to illustrate the identification of events. The Mahalanobis norm of the PCs for each spectrum was generated, as shown in Figure 6, where the yellow background indicates the actual event position.

## 4. Result and Discussion

**4.1. Selection of Moving Window Length.** An appropriate length for the moving window must be selected first. Technically, the simulation experiment requires two fundamental rules. (1)  $l$  should be such a value that two adjacent events

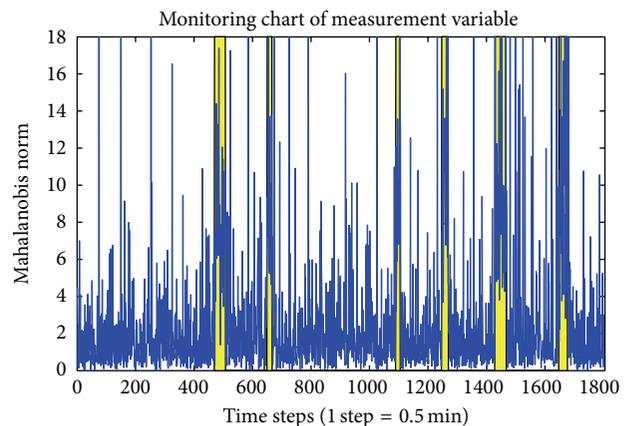
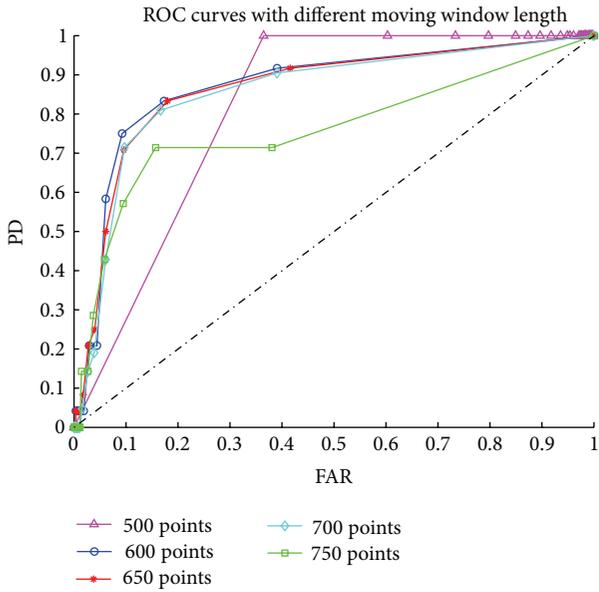
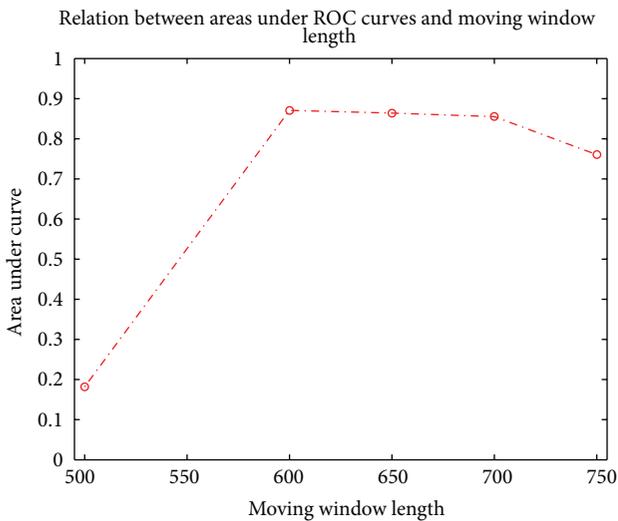


FIGURE 6: Monitoring chart of measurement variable.

not in the same window occur simultaneously. Initially, in real life, two pollution accidents rarely occur close to each other in terms of time and space. Furthermore, too many anomaly data in a window would detrimentally influence the effects of the background data. (2) A longer  $l$  does not equal a better performance. An excessively long window may include too many iterations of the E-M algorithm.



(a) Receiver operation characteristic (ROC) curves with various lengths of moving window



(b) Performance of various lengths of moving window

FIGURE 7: Selection of moving window length.

However, no simple method could be concluded for the optimal length of the moving window. Therefore, analysis was conducted to obtain an optimal moving window length for a supreme relation between PD and FAR. The testing data was the same as that in Section 4, consisting of approximately 1800 raw spectral data points. Figure 7(a) shows the relation curves between PD-FAR and the value of the moving window length with a constant number of four PCs. The relationship between the area under the ROC curves and the moving window length is illustrated in Figure 7(b). The figure explicitly illustrates that, with an increasing length, the area under the ROC curves rose initially and then fell. The optimal value of the moving window length was selected as that between 600 and 700 data points.

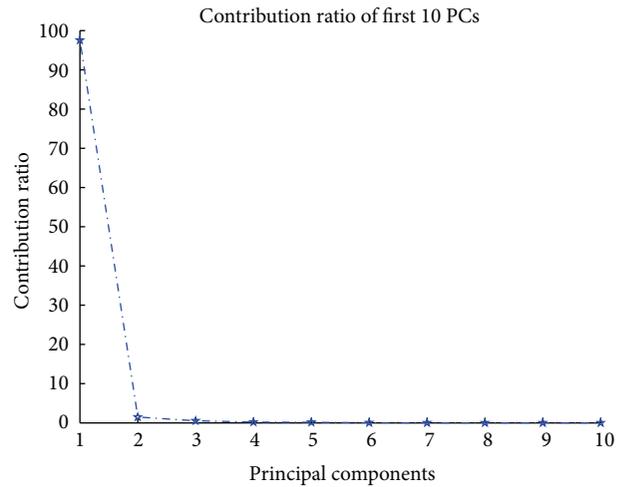


FIGURE 8: Contribution ratio of first 10 principal components (PCs).

**4.2. Selection of PCs Number.** A constant number of PCs are essential for a precise and reliable event detection method. A contribution rate-based approach for the number of PCs has been employed. However, this method is merely a theoretical basis for PC number selection; compatibility with actual data requires testing. Therefore, the relationship between PCs and the contribution rate is demonstrated in Figure 8 by listing the five most significant PCs in accordance with the contribution rate. It is obvious that the first PC has an overwhelming contribution ratio. Nonetheless, it is not reliable to indicate the water-quality event by a single dimension of a PC. Moreover, an excessively large or small PC dimensionality cannot achieve the optimal detection precision. This analysis was operated at the optimal window length, which was obtained in the previous section. The results demonstrate that more than two PCs have no significant variance. However, because a larger number of PCs delayed the speed of calculation, two were used for this dataset.

**4.3. Performance of PPCA Method.** According to the PPCA-based MSPC theory, the 95% control limit is expected to be 5.99. Based on the control limit, the false alarm rate could be calculated as 5.32%. Although it is not ideally equal to 5% as expected, the false alarm rate as 5.32% could be deemed as acceptable. To further evaluate the performance of PPCA-based approach, ROC is used in the remaining part.

To illustrate the merits of the PPCA-based approach, other dimensionality-reducing algorithms are utilized here for comparison. ROC curves for PPCA, PCA, and independent component analysis (ICA) approaches are illustrated in Figure 9. Apparently, the PPCA-based approach generated a significantly higher PD-FAR value.

Furthermore, to test the performance of this PPCA-based approach in discriminating water-quality events under various concentrations of contaminants, three events with concentrations equal to 30 ppb, 50 ppb, and 100 ppb are compared in Figure 9 as well. When applying the PPCA-based approach to these three sets of observations, the

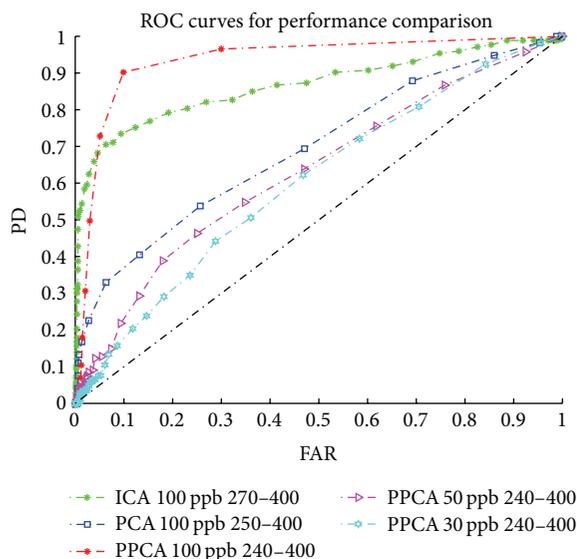


FIGURE 9: Receiver operation characteristic (ROC) curves for comparison.

parameters for calculation, including the sliding window length  $l$  and number of PCs, were fixed at the optimal values determined in previous sections. As expected, a higher concentration is related to a more satisfying ROC. Moreover, for concentrations of at least 100 ppb, the PPCA-based model demonstrated reliable and practical qualifications for water-quality event identification from background spectra data. However, no such qualifications were obtained for lower concentration events.

**4.4. Error Analysis.** As demonstrated by the ROC theory, PD indicates the actual event that is successfully identified while FAR indicates the normal status that is recognized as event by mistake. Generally, errors are possibly introduced in the pretreatment process, the PPCA model application process, and MSPC application process.

Figure 5(c) has clearly exhibited that errors introduced by PPCA model are mostly concentrated within the section below 220 nm. However, the disturbance in this wavelength range is mostly influenced by absorption of the pure water and has been eliminated from the analysis. Figure 9 indicates another consequential error source as the concentration of the contaminants. Since events with low contaminant concentration cause the contaminant effective information contained in the spectra immersed within the background noise, which deteriorates the relationship between PD and FAR. Figure 10 shows the various PD-FAR relationships with a variety of spectra sections. As one of the UV absorption peak of phenol exists around 240–250 nm, and lower wavelengths could be possibly influenced by the absorption of pure water, the spectra coverage for analysis is proven to be another error source. A too wide coverage deteriorates the possibility to identify the event correctly and increases the alarm rate by false; on the other hand, a too narrow coverage loses some of

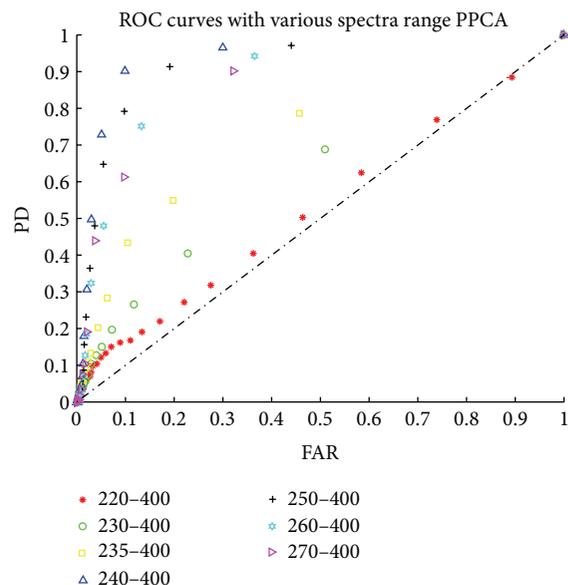


FIGURE 10: Receiver operation characteristic (ROC) curves with various spectra range.

essential spectra information which further introduces errors as well.

## 5. Conclusions

An effective spectral online monitoring of water-quality events approach, which is based on PPCA algorithms and the multivariate monitoring chart, is proposed in the paper that employs a spectral online monitoring technique to timely and accurately detect the water-quality events. This method substitutes the traditional PCA-based dimensionality-reducing algorithm with a more flexible PPCA-based approach, which is able to precisely extract the most essential information contained in the observation spectra. Combined with the multivariate monitoring chart, this method provides a reliable and flexible online monitoring method of water-quality events. Some critical values, such as the sliding window length and number of the PCs, are discussed in this paper. The results obtained demonstrate that this PPCA-based online monitoring approach for water-quality events has a higher PD-FAR value compared with traditional counterparts of PCA and ICA. When tested under various concentrations of contaminants, the approach shows reliable qualifications for 100 ppb or worse situations. Nonetheless, when testing with the lower concentrations, its performance deteriorates significantly, which is likely due to the inherent nature of UV-Vis monitoring such that it is not suitable for low-density pollution detection. In addition, the potential error sources are analyzed. With this approach, the probabilistic model extends the traditional linear analysis by providing more flexibility. On contrast to single and double wavelength water-quality detection approaches that require preacquired knowledge of the contaminant category, the PPCA-based

approach covers a wide range of spectrum and identifies events through a more comprehensive analysis.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Acknowledgments

This work was funded by the National Major Projects on Control and Rectification of Water-Body Pollution of China (no. 2008ZX07420-004) "Research and Application of Water Quality Security Evaluation and Early-warning Technologies," the National Natural Science Foundation of China (no. 41101508) "Research on Water Quality Event Detection Methods based on Time-Frequency Analysis and Multisensor Data Fusion," and the Fundamental Research Funds for the Central Universities (no. 2013FZA5011) "Research on Intelligent Detection and Evaluation of Water Quality Contamination Events."

### References

- [1] US EPA, *Edition of the Drinking Water Standards and Health Advisories*, US Environmental Protection Agency, Washington, DC, USA, 2006.
- [2] R. Holme, "Drinking water contamination in Walkerton, Ontario: positive resolutions from a tragic event," *Water Science and Technology*, vol. 47, no. 3, pp. 1–6, 2003.
- [3] M. C. Sarraguça, A. Paulo, M. M. Alves, A. M. A. Dias, J. A. Lopes, and E. C. Ferreira, "Quantitative monitoring of an activated sludge reactor using on-line UV-visible and near-infrared spectroscopy," *Analytical and Bioanalytical Chemistry*, vol. 395, no. 4, pp. 1159–1166, 2009.
- [4] R. Murray, T. Haxton, S. A. McKenna et al., "Water quality event detection systems for drinking water contamination warning systems—development, testing, and application of CANARY," *EPA1600IR-IOI036*, US, 2010.
- [5] D. Kroll and K. King, "Real world operational testing and deployment of an on-line water security monitoring system," in *Proceedings of the 8th Annual Water Distribution Systems Analysis Symposium*, p. 118, August 2006.
- [6] R. A. Dobbs, R. H. Wise, and R. B. Dean, "The use of ultra-violet absorbance for monitoring the total organic carbon content of water and wastewater," *Water Research*, vol. 6, no. 10, pp. 1173–1180, 1972.
- [7] M. Mrkva, "Automatic U.V. control system for relative evaluation of organic water pollution," *Water Research*, vol. 9, no. 5-6, pp. 587–589, 1975.
- [8] S. K. E. Brookman, "Estimation of biochemical oxygen demand in slurry and effluents using ultra-violet spectrophotometry," *Water Research*, vol. 31, no. 2, pp. 372–374, 1997.
- [9] W. Bourgeois, J. E. Burgess, and R. M. Stuetz, "On-line monitoring of wastewater quality: a review," *Journal of Chemical Technology and Biotechnology*, vol. 76, no. 4, pp. 337–348, 2001.
- [10] H. El Khorassani, P. Trebuchon, H. Bitar, and O. Thomas, "A simple UV spectrophotometric procedure for the survey of industrial sewage system," *Water Science and Technology*, vol. 39, no. 10-11, pp. 77–82, 1999.
- [11] G. Langergraber, J. Van Den Broeke, W. Lettl, and A. Weingartner, "Real-time detection of possible harmful events using UV/vis spectrometry," *Spectroscopy Europe*, vol. 18, no. 4, pp. 19–22, 2006.
- [12] C. M. Tsoumanis, D. L. Giokas, and A. G. Vlessidis, "Monitoring and classification of wastewater quality using supervised pattern recognition techniques and deterministic resolution of molecular absorption spectra based on multiwavelength UV spectra deconvolution," *Talanta*, vol. 82, no. 2, pp. 575–581, 2010.
- [13] A. Escalas, M. Droguet, J. M. Guadayol, and J. Caixach, "Estimating DOC regime in a wastewater treatment plant by UV deconvolution," *Water Research*, vol. 37, no. 11, pp. 2627–2635, 2003.
- [14] S. Fogelman, H. Zhao, and M. Blumenstein, "A rapid analytical method for predicting the oxygen demand of wastewater," *Analytical and Bioanalytical Chemistry*, vol. 386, no. 6, pp. 1773–1779, 2006.
- [15] S. Du, X. Wu, and T. Wu, "Support vector machine for ultraviolet spectroscopic water quality analyzers," *Fenxi Huaxue*, vol. 32, no. 9, pp. 1227–1230, 2004.
- [16] N. D. Lourenço, F. Paixão, H. M. Pinheiro, and A. Sousa, "Use of spectra in the visible and near-mid-ultraviolet range with principal component analysis and partial least squares processing for monitoring of suspended solids in municipal wastewater treatment plants," *Applied Spectroscopy*, vol. 64, no. 9, pp. 1061–1067, 2010.
- [17] I. Jolliffe, *Principal Component Analysis*, Wiley, New York, NY, USA, 2005.
- [18] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *Journal of the Royal Statistical Society B: Statistical Methodology*, vol. 61, no. 3, pp. 611–622, 1999.
- [19] A. De Vries and B. J. Conlin, "Design and performance of statistical process control charts applied to estrous detection efficiency," *Journal of Dairy Science*, vol. 86, no. 6, pp. 1970–1984, 2003.
- [20] D. C. Montgomery, *Introduction to Statistical Quality Control*, John Wiley & Sons, New York, NY, USA, 2007.
- [21] T. Chen and Y. Sun, "Probabilistic contribution analysis for statistical process monitoring: a missing variable approach," *Control Engineering Practice*, vol. 17, no. 4, pp. 469–477, 2009.
- [22] D. Kim and I. Lee, "Process monitoring based on probabilistic PCA," *Chemometrics and Intelligent Laboratory Systems*, vol. 67, no. 2, pp. 109–123, 2003.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

