Research Article

Water Pollution Detection Based on Hypothesis Testing in Sensor Networks

Xu Luo[1] and Jun Yang[2]

[1]Department of Information Engineering, Zunyi Medical University, Zunyi 563000, China
[2]School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081, China

Correspondence should be addressed to Jun Yang; yangjun@wust.edu.cn

Water pollution detection is of great importance in water conservation. In this paper, the water pollution detection problems of the network and of the node in sensor networks are discussed. The detection problems in both cases of the distribution of the monitoring noise being normal and nonnormal are considered. The pollution detection problems are analyzed based on hypothesis testing theory firstly; then, the specific detection algorithms are given. Finally, two implementation examples are given to illustrate how the proposed detection methods are used in the water pollution detection in sensor networks and prove the effectiveness of the proposed detection methods.

1. Introduction

Water is the most important material to human's survival and valuable resource to industrial and agricultural production. With the development of economy and industry, more kinds of pollution materials are discharged into the water environment such as rivers and lakes, and more water pollution disasters have happened. Detecting the pollution timely is important for water conservation and is the precondition to locate and find the pollution source.

In most pollution monitoring and pollution source localization applications by using sensor networks, the criteria of the pollution detecting are that the nodes have pollution concentration values and the concentration values are larger than a given threshold, such as the works about the pollution monitoring [1–8] and the works about the pollution source localization [9–12].

Since there is an initial pollution concentration of normal production and life in water, when the sensor nodes have monitored relevant information, it cannot be deduced that there exists pollution generated by a pollution source. At the same time, in the water environment there are plankton, garbage, aquatic animals, plants, and so forth, which intervene in water pollution monitoring and bring disturbances

to the monitoring data. The decision threshold to determine whether there is pollution is difficult to be given properly in the simple source detection method.

In this paper, hypothesis testing is adopted to solve the water pollution detection problems. Firstly, a brief description of the monitoring sensor network and what problems there are in the pollution detection are given. Secondly, theoretical approaches to solve the detection problems are analyzed based on hypothesis testing. Thirdly, the specific detection algorithms are given. Finally, implementation examples are given to illustrate the proposed pollution detection methods.

2. Problem Statement

2.1. Network Deployment. The self-organizing sensor network is used in the water pollution monitoring. J (>5) sensor nodes are deployed in the monitoring area uniformly and the detail information of the pollutant to be monitored is known previously. The detection sensors which are stretched into water are identified. The locations of the nodes are fixed. The sensor nodes know their own positions and all static nodes in the network sample and store the concentration values
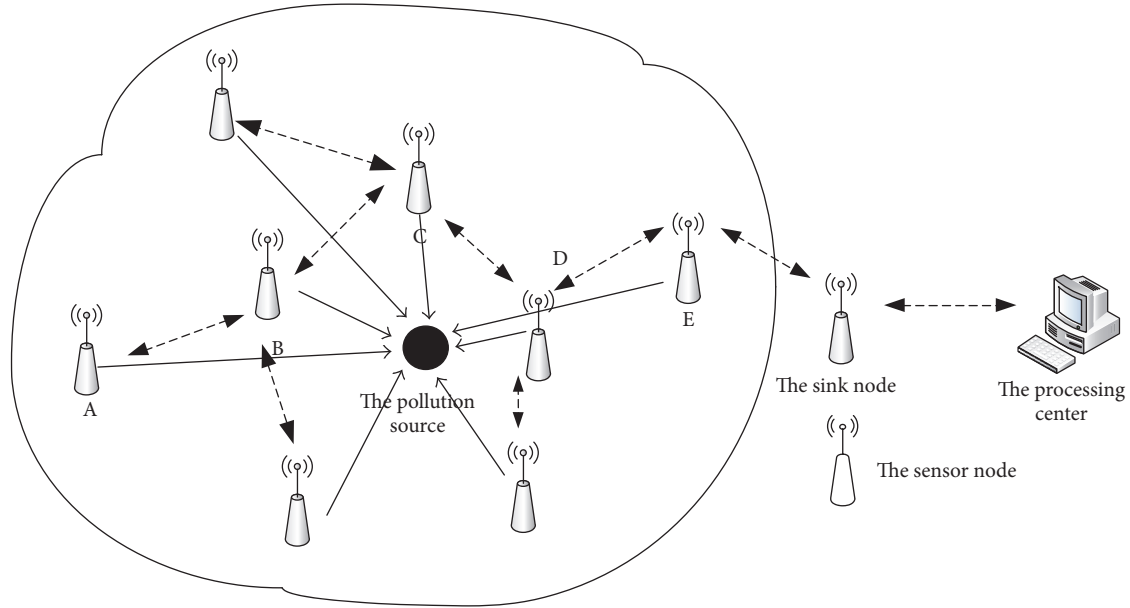
Figure 1: The sensor network for pollution detection.

synchronously with the same time interval. The background information such as the diffusion coefficient, the water depth, and the interval of sampling time is known previously. The monitoring information is routed to the sink node and processed by the data processing center. The network deployment is as shown in Figure 1.

2.2. The Detection Problems. The pollution detection problem of the network is to detect whether the sensor network finds the pollution. More specifically, that is, J static nodes sample and store the concentrations uniformly with a time interval T. At the sampling time t_l, based on the samples $\widetilde{C}(x_i, y_i, t_l)$, $i = 1, 2, 3, \ldots, J$ of the J nodes determine whether there is pollution at a given significance level α in hypothesis testing.

The purpose of the pollution detection of the network is to detect the pollution timely.

The pollution detection problem of the node is that each node in the network determines whether it has accessed to the concentration information about the pollution source. More specifically, that is, all static nodes in the network sample and store the concentrations synchronously with a time interval T. At a given significance level α in hypothesis testing, it is determined whether the node (x_i, y_i) has found the water pollution at t_L based on the known samples $\widetilde{C}(x_i, y_i, t_l)$, $l = 1, 2, 3, \ldots, L$.

The change of the diffusion in the concentration field is slow. When the network finds the pollution, it is not that each sensor node has detected the pollution. With time passing, the sensor nodes having detected the pollution will be more and more. In the pollution source localization, a node can be used in the localization only when the node has detected the pollution.

3. Pollution Detection Based on Hypothesis Testing

In [13], a simple discussion about the water pollution detection is given by the present authors under the assumption that the distribution of the monitoring noise is normal and known previously. In this paper, the pollution detection problems are discussed in more general cases.

Assume that the initial pollutant concentration (the pollution concentration of normal production and living sewage) in water is C_0. If there is no diffusion source, $\widetilde{C}(x_i, y_i, t_l) = C_0 + e$, $i = 1, 2, 3, \ldots, J$. If there is some node (x_i, y_i) having detected the pollution, $\widetilde{C}(x_i, y_i, t_l) = C(x_i, y_i, t_l) + C_0 + e$, where e is the measurement noise of sensor nodes and $C(x_i, y_i, t_l)$ is the theoretical concentration value related to the pollution source.

Remark 1. The concentration $C(x_i, y_i, t_l)$ changes over time and at different locations. The specific forms of water pollution diffusion can be seen in literature [14].

3.1. Distribution Test. Under different statistical distributions of samples, the specific hypothesis testing problems are different. In the water pollution detection, the first is to determine whether the distribution of the observation noise is normal.

In the initial state, there are only a few nodes perceiving the pollution or there is no node perceiving the pollution. When there is no pollution, $\widetilde{C}(x_i, y_i, t_1) = C_0 + e$, and if the distribution of $\widetilde{C}(x_i, y_i, t_1)$ is normal, e is a normal variable.

In order to save cost, the number of sampling nodes is often limited in practical applications. The Shapiro-Wilk W

TABLE 1: The two groups of independent samples in the detection of the network.

Sample 1	$\widetilde{C}(x_1, y_1, t_l)$	$\widetilde{C}(x_2, y_2, t_l)$	$\widetilde{C}(x_3, y_3, t_l)$	$\cdots$	$\widetilde{C}(x_{\lfloor J/2 \rfloor}, y_{\lfloor J/2 \rfloor}, t_l)$
Sample 2	$\widetilde{C}(x_{\lfloor J/2 \rfloor+1}, y_{\lfloor J/2 \rfloor+1}, t_l)$	$\widetilde{C}(x_{\lfloor J/2 \rfloor+2}, y_{\lfloor J/2 \rfloor+2}, t_l)$	$\widetilde{C}(x_{\lfloor J/2 \rfloor+3}, y_{\lfloor J/2 \rfloor+3}, t_l)$	$\cdots$	$\widetilde{C}(x_J, y_J, t_l)$

test [15] method in the case of small samples can be used as the distribution test method here.

Step 1. Remark the monitoring values of the J nodes at t_1 as $\widetilde{C}_i$, $i = 1, 2, 3, \ldots, J$. Order the values as follows:

$$[\widetilde{C}_1, \widetilde{C}_2, \widetilde{C}_3, \ldots, \widetilde{C}_J] \quad (\widetilde{C}_1 \leq \widetilde{C}_2 \leq \widetilde{C}_3 \leq \cdots \leq \widetilde{C}_J). \quad (1)$$

Step 2. The test statistic W is calculated as

$$W = \frac{\left(\sum_{i=1}^{J} a_i \widetilde{C}_i \right)^2}{\sum_{i=1}^{J} \left(\widetilde{C}_i - \overline{C} \right)^2}, \quad (2)$$

where $\overline{C} = (1/J) \sum_{i=1}^{J} \widetilde{C}_i$.

Step 3. At a given significance level of hypothesis testing α, if $W \leq W_\alpha$, the distribution is nonnormal; otherwise, the distribution is normal.

In the above steps, the values of a_i and W_α can be obtained by the method of table lookup [16].

The specific detection problems when the distribution of the sensing values is normal are different from the detection problems when the distribution is nonnormal. The detection methods in the two cases are discussed in the following.

3.2. The Pollution Detection of the Network

Case 1 (hypothesis testing under normal distribution). According to the nature of the normal distribution, if the node has detected no pollution at t_l, it can be deduced that (1) $\widetilde{C}(x_i, y_i, t_l) \sim N(C_0 + \mu, \delta^2)$, the noise $e \sim (\mu, \delta^2)$; (2) if $\widetilde{c}_1 = \min\{\widetilde{C}(x_i, y_i, t_l), i = 1, 2, 3, \ldots, J\}$, $(\widetilde{C}(x_i, y_i, t_l) - \widetilde{c}_1) \sim N(0, 2\delta^2)$.

Investigate the mean value μ_k of $C(x_i, y_i, t_l) - \widetilde{c}_1$, $i = 1, 2, 3, 4, \ldots, J$. The hypotheses are given by

$$H_0^{(1)}: \mu_\kappa = 0, \quad (3)$$
$$H_1^{(1)}: \mu_\kappa \neq 0.$$

As δ is unknown, the test statistic is

$$G_1 = \frac{\overline{C}_\kappa}{(S_\kappa \sqrt{J})}, \quad (4)$$

where $\overline{C}_\kappa = (1/J) \sum_{i=1}^{J} (\widetilde{C}(x_i, y_i, t_l) - \widetilde{c}_1)$ and $S_\kappa^2 = (1/(J-1)) \sum_{i=1}^{J} ((\widetilde{C}(x_i, y_i, t_l) - \widetilde{c}_1) - \overline{C}_\kappa)^2$.

At the significance level α [17], if

$$|G_1| \geq t_{\alpha/2} (J - 1), \quad (5)$$

that is, when $P(G_1 \geq t_{\alpha/2}(J - 1)) = \alpha$ reject $H_0^{(1)}$, it is deduced that there is a pollution source. Note that $t_{\alpha/2}$ is the $\alpha/2$ quantile of t-distribution [17, 18].

Case 2 (hypothesis testing under nonnormal distribution). While the sample distribution is nonnormal, it is difficult to verify what the specifying distribution of the values is. In this case, the Wilcoxon rank sum test is used directly [17].

Our test problem is to determine whether there is significant difference between the two groups of independent samples in Table 1. The hypotheses are

$$H_0^{(1)}: \mu_{S1} = \mu_{S2}, \quad (6)$$
$$H_1^{(1)}: \mu_{S1} \neq \mu_{S2}.$$

List the data in ascending order and allocate the ranks s_j according to the order. The significance level of hypothesis testing is α, and R_1 is the sum of ranks of Sample 1.

When

$$R_1 \leq C_L \left(\frac{\alpha}{2} \right)$$
$$\text{or } R_1 \geq C_U \left(\frac{\alpha}{2} \right) \quad (7)$$

reject H_0, there is a pollution source in the monitoring area. $C_U(\alpha/2)$ and $C_L(\alpha/2)$ are the upper tail value and lower tail value of the two-tailed rank sum test [17, 18].

3.3. The Pollution Source Detection of the Node.
Based on the monitoring values $\widetilde{C}(x_i, y_i, t_l)$, $l = 1, 2, 3, \ldots, L$ of node (x_i, y_i), determine whether the node has detected the pollution source.

Case 1 (hypothesis testing under normal distribution). If sample noise e is normal, when the node does not detect the pollution at t_l, there is $(\widetilde{C}(x_i, y_i, t_l) - \widetilde{C}(x_i, y_i, t_1)) \sim N(0, 2\delta^2)$.

Let μ' be the average value of $\{\widetilde{C}(x_i, y_i, t_l) - \widetilde{C}(x_i, y_i, t_1)$, $l = 2, 3, \ldots, L.\}$. The hypotheses are given by

$$H_0^{(3)}: \mu' = 0$$
$$H_1^{(3)}: \mu' \neq 0. \quad (8)$$

The test statistic is

$$G_2 = \frac{\overline{C}'}{(S' \sqrt{L-1})}, \quad (9)$$

where $\overline{C}' = (1/(L-1)) \sum_{l=2}^{L} (\widetilde{C}(x_i, y_i, t_l) - \widetilde{C}(x_i, y_i, t_1))$ and $S'^2 = (1/(L-2)) \sum_{l=2}^{L} (\widetilde{C}(x_i, y_i, t_l) - \widetilde{C}(x_i, y_i, t_1) - \overline{C}')^2$.

If

$$|G_2| \geq t_{\alpha/2} (L - 2), \quad (10)$$

that is, when $P(|G_2| \geq t_{\alpha/2}(L - 2)) = \alpha$, reject $H_0^{(3)}$, and it is deduced that the node has detected the pollution. Here, $t_{\alpha/2}$

Table 2: The two groups of independent samples in the detection of the node.

Sample 3	$\overline{C}(x_i, y_i, t_1)$	$\overline{C}(x_i, y_i, t_2)$	$\cdots$	$\overline{C}(x_i, y_i, t_{\lfloor L/2 \rfloor})$
Sample 4	$\overline{C}(x_i, y_i, t_{\lfloor L/2 \rfloor+1})$	$\overline{C}(x_i, y_i, t_{\lfloor L/2 \rfloor+2})$	$\cdots$	$\overline{C}(x_i, y_i, t_L)$

is the $\alpha/2$ quantile of t-distribution, and α is the significance level of hypothesis testing [17, 18].

Case 2 (hypothesis testing under nonnormal distribution). The Wilcoxon rank sum test is used. Our test problem is to determine whether there is significant difference between the two groups of independent samples in Table 2.

The hypotheses are

$$H_0^{(4)}: \mu_{S3} = \mu_{S4}$$
$$H_1^{(4)}: \mu_{S3} \neq \mu_{S4}. \tag{11}$$

The same solving method as hypothesis testing problem (6) can be used in (11).

3.4. Sample Size Requirements in Detection

3.4.1. Basic Requirements. According to the basic sample number requirements of the hypothesis testing methods [17, 18], the basic sample size requirements in our detection methods are given as follows. In the distribution test, the number of samples J should be $3 \leq J \leq 50$. When the distribution of the sample noise is normal, there should be at least 4 samples in the pollution detection of the network and the pollution detection of the node, so $J \geq 4$ and $L \geq 5$. When the distribution is not normal, there should be at least 6 samples in the pollution detection of the network and in the pollution detection of the node, so the sample numbers should satisfy $J \geq 6$ and $L \geq 6$.

3.4.2. The Power of Tests

(A) t Test. In hypotheses test under the normal distribution, the OC function of t test of (3) and (8) is

$$\beta(\mu) = P_\mu \left(\text{accept } H_0 \right)$$
$$= P_\mu \left(-t_{\alpha/2}(n-1) < \frac{\overline{C}}{(s/\text{sqrt}(n))} < t_{\alpha/2}(n-1) \right), \tag{12}$$

where $\overline{C}/(s/\text{sqrt}(n)) = ((\overline{C} - \mu)/(\delta/\text{sqrt}(n)) + \mu/(\delta/\text{sqrt}(n)))/(s/\delta)$, $\overline{C}$ is $\overline{C}_\kappa$ in (4) and $\overline{C}'$ in (9), respectively, s is S_k in (4) and S' in (9), respectively, and $n = J$ in (4) and $n = L - 1$ in (9). μ is μ_k in (3) and μ' in (8), respectively.

In the pollution detection of the network, to reduce the cost, the number of nodes is always given previously. So there is a precondition; that is, the number of the sensor nodes is $n \leq N_1$, where N_1 is a given number. For the purpose of participating in the pollution source localization timely, there also should not be many sampling times in the pollution

detection of the node, and the maximum sampling number is also often given.

To reduce the probability of false alarm, under given thresholds $\beta, \theta > 0$, in the hypothesis testing problems (3) and (8), the significance level α should satisfy

$$P_\mu \left(\text{accept } H_0 \right)$$
$$= P_\mu \left(-t_{\alpha/2}(n-1) < \frac{\overline{C}}{(s/\text{sqrt}(n))} < t_{\alpha/2}(n-1) \right) \tag{13}$$
$$\leq \beta,$$

when $\mu \in H_1$ and $|\mu/\delta| \geq \theta$.

(B) The Nonparametric Test. There are no explicit expressions of the test power in nonparametric tests. When the maximum number of samples is given, if we want to reduce the possibility of the I type error in the test, the possibility of the II type error often increases [19], so an appropriate significance level is necessary. In nonparametric tests, $\alpha = 0.05$ is often adopted.

4. The Detection Algorithms

Based on the theoretical research above, there are pollution source detection algorithms as follows.

Algorithm 1 (the pollution detection of the network).

Preconditions. The number of providing samples J is large enough and known. The samples at the first sampling time t_1 and the detection time t_l are known. The parameters β and θ which are related to the test power are given.

Step 1. Use Shapiro-Wilk W test to test the distribution of the sample noise according to (2) at the first sampling time t_1.

Step 2. If the distribution is normal, get the value range of the significance level according to (13), choose an any significance level α in the range, calculate the test statistic as (4), and go to Step 3. If the distribution is nonnormal, go to Step 4.

Step 3. When the test statistic satisfies the test criterion (5), there is pollution; otherwise, there is no pollution.

Step 4. List the data in ascending order and allocate the ranks according to the order. Calculate R_1 which is the sum of ranks of Samples 1 in Table 1. Look up the table to get the tail values in the rank sum test. When the sum R_1 satisfies the test criterion (7), there is pollution; otherwise, there is no pollution.

At time t_l, if the network does not detect the pollution, the pollution detection of the network will be made at t_{l+1}.

Algorithm 2 (the pollution detection of the node (x_i, y_i)).

Preconditions. The number of samples L is large enough and known. The samples at the first sampling time t_1 are known. The samples of the detection node (x_i, y_i) are known. The

parameters β and θ which are related to the test power are given.

Step 1. Use Shapiro-Wilk W test to test the distribution of the sample noise according to (2) at the first sampling time t_1.

Step 2. If the distribution is normal, get the value range of the significance level according to (13), choose an any significance level α in the range, calculate the test statistic as (9), and go to Step 3. If the distribution is nonnormal, go to Step 4.

Step 3. When the test statistic satisfies the test criterion (10), the node detects the pollution; otherwise, the node fails to detect the pollution.

Step 4. List the data in ascending order and allocate the ranks according to the order. Calculate R_3 which is the sum of ranks of Samples 3 in Table 2. Look up the table to get the tail values in the rank sum test. When the sum R_3 satisfies the test criterion (7), the node detects the pollution; otherwise, the node fails to detect the pollution.

5. Implementation Examples

Experiment 1. A simulation is carried out to test the proposed detection algorithms. The distribution of monitoring noises is normal.

Background. The size of the static shallow water is $10\,\mathrm{m} \times 10\,\mathrm{m}$, and the average depth of the water is $f = 10\,\mathrm{m}$. Apart from the coast of impervious Y, there is a continuous source at $(x_0, y_0) = (1.05, 5.55)\,(\mathrm{m})$. Starting from time $t_0 = 0\,\mathrm{h}$, the solution with pollutant is injected into the water. The mass of the pollutant is $M = 100\,\mathrm{kg}$. The diffusion coefficient is $D = 0.1\,\mathrm{m}^2/\mathrm{h}$. The nodes in the network sample the concentration uniformly with interval $T = 0.05\,\mathrm{h}$. The diffusion can be depicted by the model:

$$C(x, y, x_0, y_0)$$
$$= \left(\frac{M}{(4\pi D(t - t_0))}\right) e^{-[((x-x_0)^2 + (y-y_0)^2)/4D(t-t_0)]}, \quad (14)$$

where t is the current time. The locations of the monitoring nodes are (2.05, 5.55), (1.05, 4.55), (0.05, 5.55), (1.05, 6.55), (0.55, 6.05), (1.55, 6.05), (0.55, 5.05), (1.55, 5.05), (0.05, 6.25), and (1.05, 7.05). The monitoring values in the experiment are the simulation values of (14) adding the noise with a normal distribution $N(0, \delta^2)$. The given parameters are $\beta = 0.5$ and $\theta = 1$, and there are 10 nodes, $n = N_1 = 10$.

(A) The Pollution Detection of the Network. According to constraint (13) and the sample size table of t test in [18], it is can be deduced that $\alpha > 0.025$ under the given parameters β, θ, and δ in Table 3.

For different δ values and significance levels, detect the pollution at the initial observation time 0.01 h, and the results are shown in Table 3. The hypothesis testing detection method under the normal distribution is used, and in the

TABLE 3: The detection results of the network.

δ	Significance levels	Detection results
0.01	0.03	100/100
0.01	0.05	100/100
0.01	0.1	100/100
0.1	0.03	100/100
0.1	0.05	99/100
0.1	0.1	100/100
0.5	0.03	99/100
0.5	0.05	100/100
0.5	0.1	100/100

table, "$X/100$" represents that in the 100 experiments the pollution is detected successfully X times.

The results show that the pollution can be detected by the network soon.

(B) The Pollution Detection of the Node. Detect whether the node has detected the pollution source based on the observed data of the node (1.05, 7.05). The monitoring data is as shown in Table 4. Compare with the simple detection method in which the criterion of whether the pollution source has been detected is that the monitoring value is larger than a given threshold, and the results are shown in Table 5.

In Table 5, "—" represents no result. Comparing the results in the table, it can be seen that the detection method using hypothesis testing is more stable if an appropriate significance level is chosen, and in the simple detection, to detect the pollution source timely the threshold should be as small as possible. But apparently, if the noise in the practical applications is considered, small thresholds may bring about large false alarm rates.

Experiment 2. A practical experiment is carried out to test the proposed detection algorithms.

Background. In water, of which the size is $200\,\mathrm{cm} \times 200\,\mathrm{cm}$, the average depth $f = 100\,\mathrm{cm}$. There is a continuous source at the boundary. Starting from $t_0 = 0\,\mathrm{s}$, the solution of MgSO4 is discharged to the water continually. The nodes deployment is depicted by Figure 2. The monitoring values of different sensor nodes in the experiment are shown in Table 6.

Distribution Verification. Both at time 5 s and 10 s, for different significance levels $\alpha = 0.01$, $\alpha = 0.05$, and $\alpha = 0.1$, the detection results all show that the distribution of monitoring data is not normal. So, the detection methods based on the Wilcoxon rank sum tests are used.

(A) The Pollution Detection of the Network. The given significance level is $\alpha = 0.05$, and the network detects the pollution at 30 s.

(B) The Pollution Detection of the Node. The significance level is $\alpha = 0.05$, and the time when nodes 0, 1, 2, 3, 4, 5, 6, and 7

Table 4: The monitoring values of node (1.05, 7.05) in Experiment 1.

Time(s)	T	$2T$	$3T$	$4T$	$5T$	$6T$	$7T$	$8T$	$9T$	$10T$
Observations (g/L)	0.008	0.009	0.010	0.015	0.017	0.019	0.081	0.217	0.423	0.748

Table 5: The detection results comparing with the simple detection method.

Significance levels	Detection time of t test	Threshold (g/L)	Detection time of the simple detection method
0.03	—	0.03	$7T$
0.05	$6T$	0.05	$7T$
0.1	$5T$	0.1	$8T$
0.15	$5T$	0.15	$8T$
0.2	$5T$	0.2	$8T$
0.5	$4T$	0.5	$10T$

Table 6: The monitoring values in Experiment 2.

Time(s)	Node 0 (g/L)	Node 1 (g/L)	Node 2 (g/L)	Node 3 (g/L)	Node 4 (g/L)	Node 5 (g/L)	Node 6 (g/L)	Node 7 (g/L)
5	0.00000	0.00000	0.00032	0.00000	0.00016	0.00000	0.00000	0.00000
10	0.00000	0.00000	0.00242	0.00000	0.00548	0.00000	0.00000	0.00000
15	0.00000	0.00000	0.00581	0.00000	0.01210	0.00000	0.00000	0.00000
20	0.00000	0.00000	0.00790	0.00000	0.01726	0.00000	0.00000	0.00000
25	0.00000	0.00000	0.01161	0.00000	0.03532	0.00000	0.00000	0.00000
30	0.00210	0.00000	0.01661	0.00050	0.04532	0.00000	0.00033	0.00000
35	0.01726	0.00000	0.05686	0.00415	0.06284	0.00242	0.01561	0.00000
40	0.05686	0.00000	0.08422	0.01445	0.09608	0.03048	0.03206	0.00000
45	0.06216	0.00000	0.10429	0.01877	0.12071	0.04081	0.05216	0.00000
50	0.08441	0.00000	0.12990	0.05186	0.13163	0.07324	0.06912	0.00000
55	0.09333	0.00000	0.14969	0.06784	0.15386	0.09186	0.08118	0.00000
60	0.11235	0.00000	0.16957	0.08324	0.17286	0.10673	0.09559	0.00000
65	0.12439	0.00000	0.18557	0.09804	0.17557	0.11449	0.11316	0.00000
70	0.12959	0.00016	0.19014	0.11694	0.17957	0.12418	0.12663	0.00000
75	0.13449	0.00129	0.19457	0.12245	0.18786	0.13122	0.13010	0.00081
80	0.13673	0.00177	0.20385	0.12561	0.19957	0.13449	0.13449	0.00258
85	0.14204	0.00242	0.20846	0.12888	0.20846	0.13990	0.13684	0.00290
90	0.14643	0.00403	0.21135	0.13224	0.21019	0.14122	0.13847	0.00435
95	0.14878	0.00516	0.21692	0.13786	0.21154	0.14714	0.14327	0.00613
100	0.15114	0.00532	0.21808	0.14133	0.21346	0.15071	0.15200	0.00694
105	0.15386	0.00726	0.22423	0.14776	0.21481	0.15343	0.15814	0.00823
110	0.15686	0.00790	0.22885	0.15300	0.21558	0.15714	0.16043	0.00790
115	0.16000	0.00839	0.23115	0.15829	0.21962	0.16214	0.16643	0.00839
120	0.16329	0.01000	0.23462	0.16343	0.22154	0.16543	0.16814	0.01097
125	0.16557	0.01194	0.23558	0.16986	0.22288	0.16814	0.17000	0.01161
130	0.16757	0.01339	0.23885	0.17214	0.22365	0.17386	0.17114	0.01403
135	0.16814	0.01484	0.23923	0.17329	0.22404	0.17414	0.17186	0.01597
140	0.16857	0.01516	0.23942	0.17386	0.22500	0.17429	0.17229	0.01613
145	0.16914	0.01613	0.23962	0.17414	0.22519	0.17457	0.17286	0.01694
150	0.16929	0.01774	0.24000	0.17457	0.22519	0.17457	0.17300	0.01774
155	0.16971	0.01887	0.24058	0.17486	0.22558	0.17471	0.17314	0.01887
160	0.17014	0.01919	0.24077	0.17514	0.22577	0.17486	0.17357	0.02871
165	0.17086	0.01935	0.24096	0.17543	0.22615	0.17514	0.17371	0.02968
170	0.17114	0.02048	0.24115	0.17571	0.22635	0.17529	0.17414	0.03000
175	0.17157	0.02065	0.24154	0.17614	0.22654	0.17557	0.17429	0.03016
180	0.17171	0.02097	0.24173	0.17629	0.22712	0.17571	0.17457	0.03065

TABLE 7: The detection time of different nodes in Experiment 2.

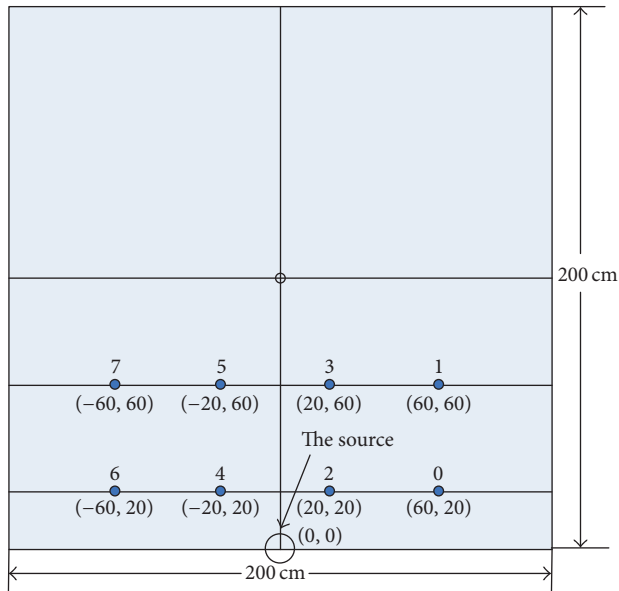| Significance level	Node 0 (s)	Node 1 (s)	Node 2 (s)	Node 3 (s)	Node 4 (s)	Node 5 (s)	Node 6 (s)	Node 7 (s)
0.05	50	95	40	50	40	50	50	100

FIGURE 2: The nodes deployment in Experiment 2.

find the pollution is shown in Table 7. The detection method is as (11).

The results show that the pollution can be detected by the nodes only when there are some increasing concentration samples.

From the results of the experiments above, it can been seen that, in the simple detection method, an appropriate decision threshold is hard to be given, so the pollution source detection by using hypothesis testing is more preferable. Whether the distribution of the sample noise is normal or not, the corresponding detection algorithms are available.

6. Conclusions

Water pollution detection is important in the water environment monitoring. The pollution source detection problems of the network and of the node are discussed based on hypothesis testing. The sample size requirements in different detection problems are also analyzed. In implementation examples, the proposed pollution detection algorithms are tested. The effectiveness of the detection algorithms is proved. This work mainly focuses on theoretical detection approaches based on hypothesis testing. In the future work, more problems in the practical applications will be studied when the proposed detection algorithms are adopted, such as the optimized detection methods of the node related to large or small concentration variations, and the influences of the concentration variations on the statistical distribution in the distribution test step.

Disclosure

Competing Interests

Acknowledgments

References

[1] J. Wang, X. Wang, X. Ren, and Y. Shen, *A Water Pollution Detective System Based on Wireless Sensor Network*, Addison-Wesley, Harlow, UK, 3rd edition, 1999.

[2] P. Jiang, H. Xia, Z. He, and Z. Wang, "Design of a water environment monitoring system based on wireless sensor networks," *Sensors*, vol. 9, no. 8, pp. 6411–6434, 2009.

[3] E. V. Hatzikos, N. Bassiliades, L. Asmanis, and I. Vlahavas, "Monitoring water quality through a telematic sensor network and a fuzzy expert system," *Expert Systems*, vol. 24, no. 3, pp. 143–161, 2007.

[4] B. O'Flynn, R. Martinez-Catala, S. Harte et al., "SmartCoast: a wireless sensor network for water quality monitoring," in *Proceedings of the 32nd IEEE Conference on Local Computer Networks (LCN '07)*, pp. 815–816, Dublin, Ireland, October 2007.

[5] J. Heidemann, M. Stojanovic, and M. Zorzi, "Underwater sensor networks: applications, advances and challenges," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 370, no. 1958, pp. 158–175, 2012.

[6] J. X. He, J. Li, and H. W. Yan, "An architecture of geospatial sensor network for water quality monitoring," *Journal of Jilin University (Earth Science Edition)*, vol. 41, no. 4, pp. 1262–1266, 2011.

[7] N. Harris, A. Cranny, M. Rivers, K. Smettem, and E. G. Barrett-Lennard, "Application of distributed wireless chloride sensors to environmental monitoring: initial results," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 4, pp. 736–743, 2016.

[8] Y. Derbew and M. Libsie, "A wireless sensor network framework for large-scale industrial water pollution monitoring," in *Proceedings of the IST-Africa Conference and Exhibition (IST-Africa '14)*, pp. 1–8, Mauritius, May 2014.

 [9] J. Yang, L. Chai, and X. Luo, "Pollution source localization in lake environment based on wireless sensor networks," in *Proceedings of the 31st Chinese Control Conference*, pp. 6649–6654, Hefei, China, July 2012.

[10] X. Luo, L. Chai, and J. Yang, "Mobile diffusion source tracking in wireless sensor networks," *Control Theory and Applications*, vol. 24, no. 3, pp. 143–161, 2007.

[11] X. Luo, L. Chai, and J. Yang, "Offshore pollution source localization in static water using wireless sensor networks," *Acta Automatica Sinica*, vol. 40, no. 5, pp. 849–861, 2014.

[12] J. Yang, X. Luo, and L. Chai, "The localization of pollution source in steady diffusion state in river based on wireless sensor networks," *Chinese Journal of Sensors and Actuators*, vol. 27, no. 10, pp. 1423–1430, 2014.

[13] X. Luo, J. Yang, and L. Chai, "Water pollution source detection in wireless sensor networks," in *Proceedings of the IEEE International Conference on Information and Automation (ICIA '15)—In conjunction with IEEE International Conference on Automation and Logistics*, pp. 2311–2315, IEEE, Lijiang, China, August 2015.

[14] Z. Peng, T. Yang, and X. Liang, *The Water Environment Mathematical Model and Its Application*, Chemical Industry Press, Beijing, China, 2007.

[15] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3-4, pp. 591–611, 1965.

[16] B. Thomas, *Serious Stats: A Guide to Advanced Statistics for the Behavioral Sciences*, Palgrave Macmillan Press, London, UK, 2012.

[17] R. E. Walpole, R. H. Myers, S. L. Myers, and K. Ye, *Probability and Statistics for Engineers and Scientists*, Pearson Prentice Hall, Upper Saddle River, NJ, USA, 8th edition, 2007.

[18] Z. Shen, S. Xie, and C. Pan, *Probability and Mathematical Statistics*, High Education Press, Beijing, China, 2005.

[19] Sample size for nonparametric, January 2016, https://www.graphpad.com/guides/prism/6/statistics/index.htm?stat_sample_size_for_nonparametric_.htm.