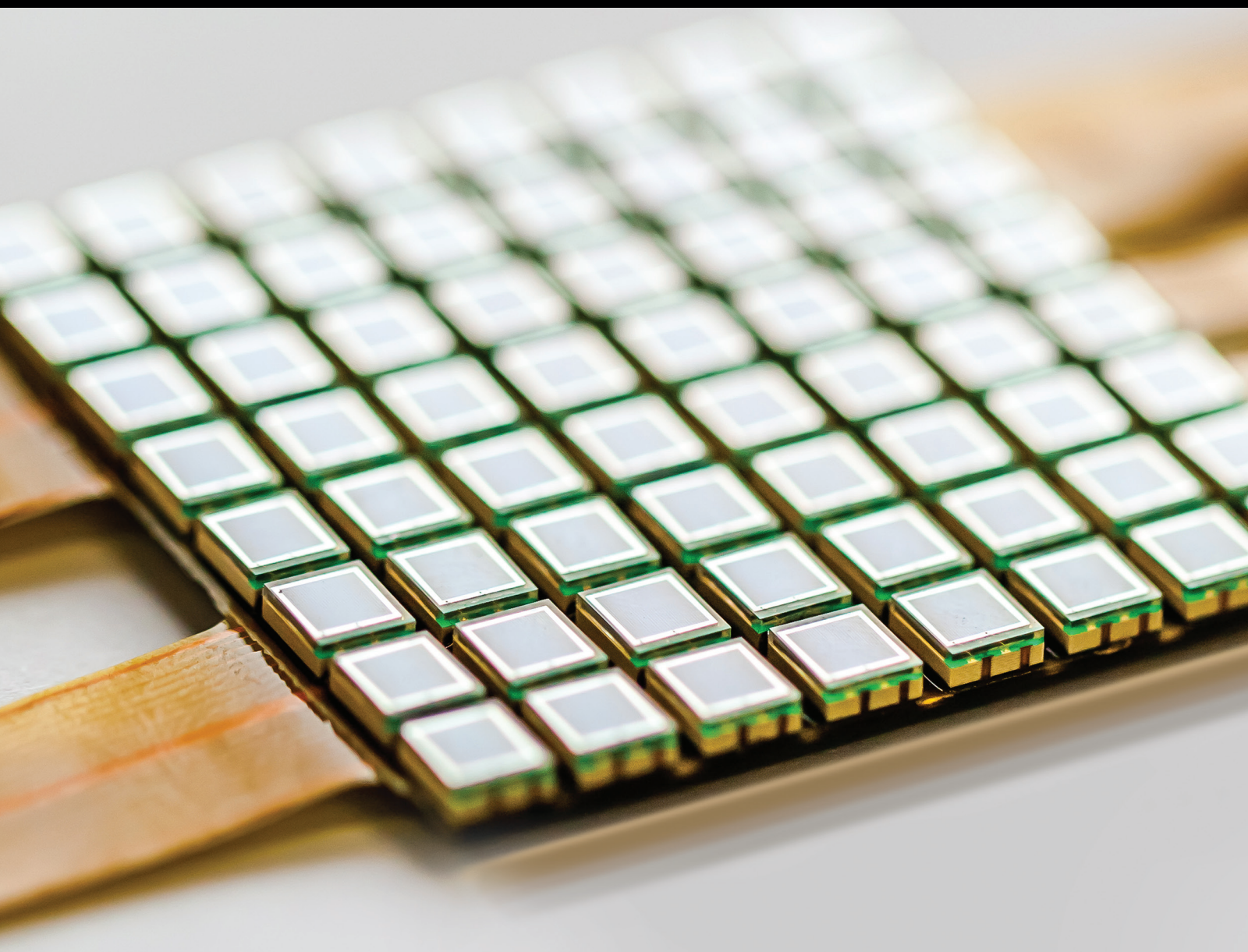
Journal of Sensors

# Sensors and Systems for Environmental Monitoring and Control

Lead Guest Editor: Jesús Lozano
Guest Editors: Constantin Apetrei, Mahdi Ghasemi-Varnamkhasti, Daniel Matatagui, and José P. Santos



# Sensors and Systems for Environmental Monitoring and Control

Journal of Sensors

# Sensors and Systems for Environmental Monitoring and Control

Lead Guest Editor: JesÃžs Lozano
Guest Editors: Constantin Apetrei, Mahdi Ghasemi-Varnamkhasti, Daniel Matatagui, and José P. Santos

# Editorial Board

# Contents

Hindawi
Journal of Sensors
Volume 2017, Article ID 6879748, 2 pages
https://doi.org/10.1155/2017/6879748

*Editorial*

# Sensors and Systems for Environmental Monitoring and Control

**Jesús Lozano,[1] Constantin Apetrei,[2] Mahdi Ghasemi-Varnamkhasti,[3] Daniel Matatagui,[4] and José Pedro Santos[5]**

[1] *University of Extremadura, Badajoz, Spain*
[2] *"Dunarea de Jos" University of Galati, Galati, Romania*
[3] *Shahrekord University, Shahrekord, Iran*
[4] *Universidad Nacional Autónoma de México (UNAM), Mexico City, Mexico*
[5] *Spanish Council for Scientific Research (CSIC), Madrid, Spain*

Correspondence should be addressed to Jesús Lozano; jesuslozano@unex.es

Received 19 October 2016; Accepted 19 October 2016; Published 30 October 2017



The detection and monitoring of odorants and/or chemical pollutants in the environment have become a major challenge to scientific community of modern developed and developing countries of the world. Billions of tons of organic and inorganic chemical pollutants are released into the air, water, and soil, annually resulting in widespread potential health hazards to plants, animals, and humans worldwide. Monitoring of environmental pollution is necessary and considered a crucial element in the assessment of air quality in cities and rural areas.

The field of environmental monitoring and control encompasses a broad range of activities, including the detection of a variety of gases—CO, $NO_x$, $NH_3$, and the particularly challenging case of $CO_2$. Sensing systems have been developed for all of these applications, but this Special Issue is focused on the employment of gas and liquid sensors, electronic noses/tongues, wireless sensor networks, and other systems to monitor and control airborne volatile compounds that are released when waste products are dumped in water, soil, or air.

The published papers deal with the development of sensors and wireless sensor networks for environmental applications. The first paper of this Special Issue addresses the water pollution source localization using wireless sensor networks. In this paper, a study on water pollution source localization is presented. Firstly, the source detection is discussed. Then the coarse localization methods and the localization methods based on diffusion models are introduced and analyzed, respectively, and both methods are compared.

The second paper is on the development of comprehensive water-quality monitoring system that employs a smart network management, nanoenriched sensing framework, and intelligent and efficient data analysis and forwarding protocols for smart and system-aware decision making.

The third paper presents an air pollution monitoring system for subway stations of Seoul (Korea) based on environmental sensors in order to preserve the health of commuters in the subway system. In this study, the accuracy of an instrument for particulate matter (PM) measurement using the light scattering method was improved with the help of a linear regression analysis technique to continuously measure the PM10 concentrations in subway stations. In addition, an air quality monitoring system based on environmental sensors was implemented to display and record the data of PM10, $CO_2$, temperature, and humidity.

The fourth paper presents an algorithm that is proposed to improve the localization of the unknown position nodes in wireless sensor networks by using fixed and mobile guide nodes (nodes with known position). To evaluate the proficiency, the proposed algorithm has been successfully studied and verified through simulation. Low cost, high accuracy, low power consumption of nodes and complete coverage are the

benefits of this approach, and long term in localization is the disadvantage of this method.

*Jesús Lozano*
*Constantin Apetrei*
*Mahdi Ghasemi-Varnamkhasti*
*Daniel Matatagui*
*José Pedro Santos*

Hindawi Publishing Corporation
Journal of Sensors
Volume 2016, Article ID 6385380, 5 pages
http://dx.doi.org/10.1155/2016/6385380

*Research Article*

# Improving Localization in Wireless Sensor Network Using Fixed and Mobile Guide Nodes

**R. Ahmadi,[1] G. Ekbatanifard,[2] A. Jahangiry,[3] and M. Kordlar[4]**

[1]*Department of Computer Engineering, Islamic Azad University, Rasht Branch, Rasht, Iran*
[2]*Department of Computer Engineering, Islamic Azad University, Lahijan Branch, Lahijan, Iran*
[3]*Department of Computer Engineering, Islamic Azad University, Qom Branch, Qom, Iran*
[4]*Young Researchers and Elite Club, Islamic Azad University, Tabriz Branch, Tabriz, Iran*

Correspondence should be addressed to R. Ahmadi; ramin_mcm2003@yahoo.com

Received 29 May 2016; Accepted 20 July 2016

Academic Editor: Constantin Apetrei



Wireless sensor network contains very large number of tiny sensors; some nodes with known position are recognized as guide nodes. Other nodes with unknown position are localized by guide nodes. This article uses the combination of fixed and mobile guide nodes in wireless network localization. So nearly 20% of nodes are fixed guide nodes and three nodes are intended as mobile guide nodes. To evaluate the proficiency, the proposed algorithm has been successfully studied and verified through simulation. Low cost, high accuracy, and low power consumption of nodes and complete coverage are the benefits of this approach and long term in localization is the disadvantage of this method.

## 1. Introduction

With recent advances in wireless electronics and transmission, designing and manufacturing of wireless sensors with low power consumption, small size, reasonable price, and various applications have become popular. These small sensors with the abilities such as receiving environmental information based on the sensor type and processing and sending data have caused the creation of networks type which is called wireless sensor networks. The issue of location evaluation or spatial coordinate of wireless sensors is called localization. In the past few years, localization in wireless sensor network has become a widespread researchable field. knowing the position of nodes in WSN is necessary for some applications and protocols, for example, in tracking and inquiry [1]. Due to the random distribution of nodes and even the mobility of them in some applications an appropriate localization algorithm is needed [2]. A lot of works have been done in the area of localization and related issues but various presented algorithms need extra hardware. For example, in [3] destination measurement hardware is required and [4] needs angle measurement hardware. Method used in [5] needs a movable hardware and in [6, 7] a send-range radio hardware is used. Generally, many localization schemes use fixed stations but in some others the position of sensor nodes is determined by mobile stations. In this article, a localization algorithm is proposed which uses fixed and mobile guide nodes to determine the position of nodes with remarkable speed and accuracy. The proposed algorithm is composed of two phases: in the first stage, the position of nodes will be determined by fixed guide nodes and in the second phase the position of the remaining unallocated nodes of the first stage will be determined by mobile guide nodes.

The rest of this paper is organized as follows: Section 2 presents related work. The proposed scheme will be described in Section 3. The simulation model and experiment results are presented in Section 4. Section 5 concludes this paper.

## 2. Related Work

Many localization methods contain two phases. In the first phase, the distances and angles of known position nodes and the node that is going to be localized are mapped. The first phase is known as the measurement phase. In the second phase, the distances or mapped angles are combined to determine the position of node. This phase is known

as positioning phase. Some existing methods for mapping phase are as follows: time differences of arrival signal [8], time of arrival signal [9], angle of arrival signal [10], and received signal strength [11]. Depending on the method used for mapping, an individual positioning technique is used in the second phase. The most widely used techniques of these techniques for second phase are trilateral approach [12], pentagon method [13], the maximum probability technique [14], and triangulation method [15]. For example, in [16] by using four mobile guide nodes the position of unknown node is acquired. In this algorithm nodes determine their positions by receiving signals from guide nodes. The guide nodes are equipped with global positioning system. This algorithm uses received signal strength in first phase and pentagon method in second phase. In [17], a method is presented that finds the unknown positions without using GPS, as follows: (1) having a distance and a direction (angle), (2) having two points and their angle, and (3) having the distance of three required points. A different method is represented in [18] which calculates the location of unknown nodes by weighting that is based on the connectivity with the known positioning nodes and placement with maximum weighted node. In this algorithm the adjacency matrix is used to determine neighborhood nodes before localization. In this network nodes can move partially. After completing location, the current map is compared with adjacency matrix, and if it is incompatible the node will locate on its original place by tensile and buoyancy equations.

## 3. The Proposed Algorithm

In this algorithm the network contains $N$ nodes and two types of nodes are used for localization: stationary guide node and mobile guide node. The proposed algorithm uses radio frequency to measure the distance between unknown node, mobile guide node, and located node by using two stages: in the first step, both stationary and mobile guide nodes are considered as stabile and do the locating. In the next step, the mobile guide nodes start to move and locate the remaining nodes of unknown position from the first step. To improve the speed of locating for each node in the network state flag of 1 or 0 can be defined. As a consequence, before starting positioning the identifier of guide nodes is one and for the other nodes it is zero. The identifier of located nodes will change from zero to one. In the second stage sensor nodes evaluate their identifier state by receiving signal from mobile stations. If the value of identifier is zero the node will participate in the process of locating; otherwise it would not do anything. The proposed algorithm is based on the assumptions below.

Fixed nodes are self-organized and establish contingency. Three mobile base stations broadcast their position's information to the sensor node. Mobile base stations are equipped with GPS and radio frequency (RF) transmitter compared to the fixed nodes that are equipped with RF receiver. The mobile station is not restricted to the energy.

*3.1. Number of Guide Nodes.* The number of nodes in the specified area is important, because if the number of guide nodes is high, it can cause conflict; otherwise locating time will be increased. In order to solve this problem, (1) is used. Consider

$$\mu(R) = \frac{N\pi R^2}{A}, \tag{1}$$

where $N$ is the number of nodes, $A$ is area, and $R$ is the maximum range of node radio communication.

Table 1: Value of $\alpha$ in different environments.

| Environment | $A$ |
|---|---|
| Outdoors | |
| Free | 2 |
| Jungle | 2/7–5 |
| Indoors | |
| Unobstructed | 1/6–1/8 |
| Obstacles | 4–6 |

*3.2. Distance Measurement.* The destination is measured by RSSI which is sent from fixed and mobile guide node. After receiving RF signal by the fixed node, fixed node will measure the received signal strength to measure the distance.

$$\begin{aligned}
E_{\text{Tx}}(K, D) &= E_{\text{Tx-ELEC}}(K) + E_{\text{Tx-AMP}}(K, D)\\
E_{\text{Tx}}(K, D) &= E_{\text{ELEC}} * K + \epsilon_{\text{AMP}} * K * D^{\alpha}\\
E_{\text{Rx}}(K) &= E_{\text{Rx-ELEC}}(K)\\
E_{\text{Rx}}(K) &= E_{\text{ELEC}} * K\\
E_{\text{Rx}} &= \frac{G \cdot E_{\text{Tx}}}{D^{\alpha}}\\
\text{BER} &= 0.5E^{(-0.5E_{\text{Rx}}/E_{\text{Tx}})},
\end{aligned} \tag{2}$$

where $E_{\text{Tx-ELEC}}$ and $E_{\text{Rx-ELEC}}$ are, respectively, one-bit transmit energy and one-bit received energy, $G$ and $\epsilon_{\text{AMP}}$ are radio signal amplification factor, $D$ and $K$ are, respectively, the distance between two nodes and the number of transmitted bits, BER denotes bit error rate, and $\alpha$ shows the dissipation factor of environment which can be obtained by Table 1.

Distance of two nodes can be calculated by using (2). Hence by using (3) and three points of mobile base station or three distances between fixed located nodes which cooperate in localization of other nodes and other nodes with unknown position, localization can be done.

$$\begin{aligned}
(x - x_1)^2 + (y - y_1)^2 &= d_1^2\\
(x - x_2)^2 + (y - y_2)^2 &= d_2^2\\
(x - x_3)^2 + (y - y_3)^2 &= d_3^2,
\end{aligned} \tag{3}$$

where $(x_1, y_1)$, $(x_2, y_2)$, and $(x_3, y_3)$ are locations of mobile stations, $(x, y)$ is the position of receiver node, and $d_1$, $d_2$, and $d_3$ are the distances between the mobile node and the fixed nodes.

*3.3. Measurements.* Five different kinds of measurement are used to evaluate the performance of localization algorithm.

*(1) The Average of Positioning Error.* The average difference between estimated location $(X_{e_i} - Y_{e_i})$ and the exact position $(X_i - Y_i)$ of all sensor nodes is specified.

$$\Delta E_{av} = \frac{\sum_{i=1}^{N} \sqrt[2]{\left(X_{e_i} - X_i\right)^2 + \left(Y_{e_i} - Y_i\right)^2}}{N}, \quad (4)$$

where $N$ represents the total number of fixed sensor nodes.

*(2) The Average of Energy Consumption.* It is defined as follows:

$$P_{av} = \frac{\sum_{i=1}^{N} E_i}{N}, \quad (5)$$

where $E_i$ illustrates the energy consumption of fixed node $i$.

*(3) The Average of Throughput.* It is defined as follows:

$$R_{av} = \frac{\sum_{i=1}^{N} R_i}{N}, \quad (6)$$

where $R_i$ is the throughput of fixed node $i$.

*(4) Location Duration.* The average time to locate all network nodes is called execution time. Time unit is considered as second. To find out the average time of a node location the values should be divided into $N$-$H$, where $H$ represented the number of guide nodes.

*(5) Success Rate of the Algorithm.* The success rate is the percentage ratio of successful tests number to the total tests number.

*3.4. Base Point Selection.* In Figure 1, the unknown node calculates its position based on three fixed or mobile guide nodes, $S_1$, $S_2$, and $S_3$. Unlike similar algorithms three mobile guide nodes are used in localization. The proposed algorithm avoids situations of placing guide nodes on a line or placing at least two guide nodes in the same position during localization.

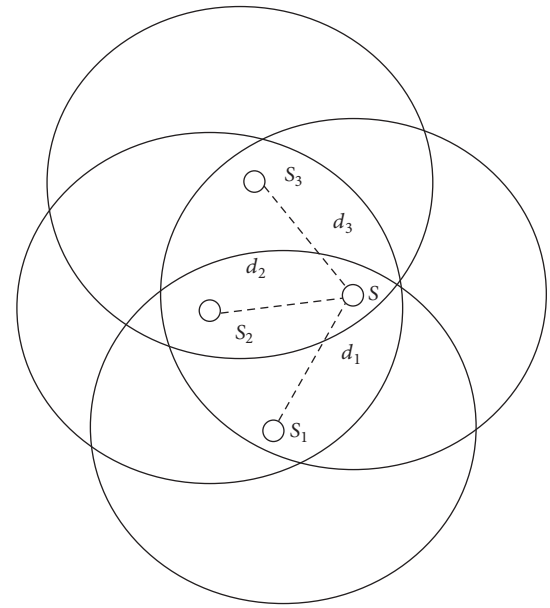



Figure 1: Calculation of node position using both fixed location nodes and mobile base station.

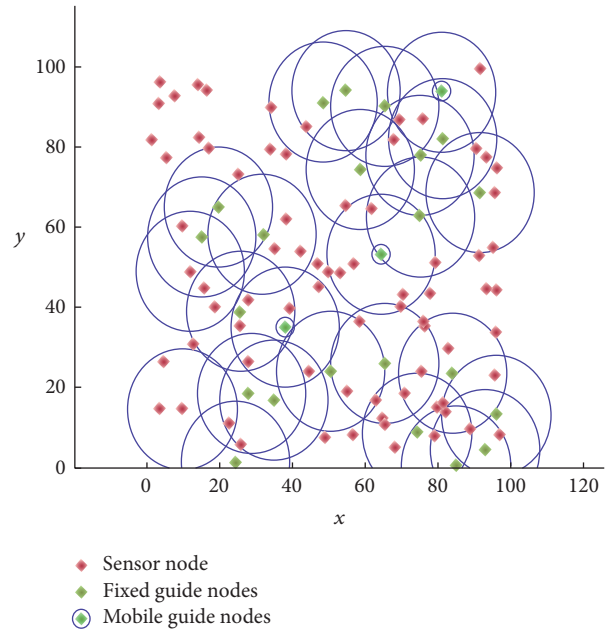



Figure 2: Nodes random distribution.

## 4. Simulation Results

In this section by using simulation we will describe some of the main concepts of the proposed algorithm such as error location, energy consumption, throughput, location duration, and success rate of the algorithm.

*4.1. Simulation Condition.* In order to evaluate the effects and efficiency of our proposed algorithm we implement it in MATLAB R2009a. The radio range of nodes is equal to $R$ and the number of guide nodes is shown by $H$. The values of $R$ and $H$ are given as input in the program. 100 nodes are placed randomly within a $100 * 100\,m^2$ area as shown in Figure 2. The mobile nodes are equipped with GPS, radio frequency, and capability of moving like curve from the middle region. Each fixed node is equipped with radio frequency receiver. The mobile nodes have no limited energy and all fixed sensor nodes need to store their energy. Experiments were performed for different values of $R$ and the results are presented in Figures 3–5. The results presented in each case are the mean of 30 repeated experiments. To show the efficiency of the proposed algorithm, we compare it with different algorithms; the results are shown in Figures 3–5. The first algorithm shows the localization with only fixed guide nodes. The second algorithm shows the localization with only

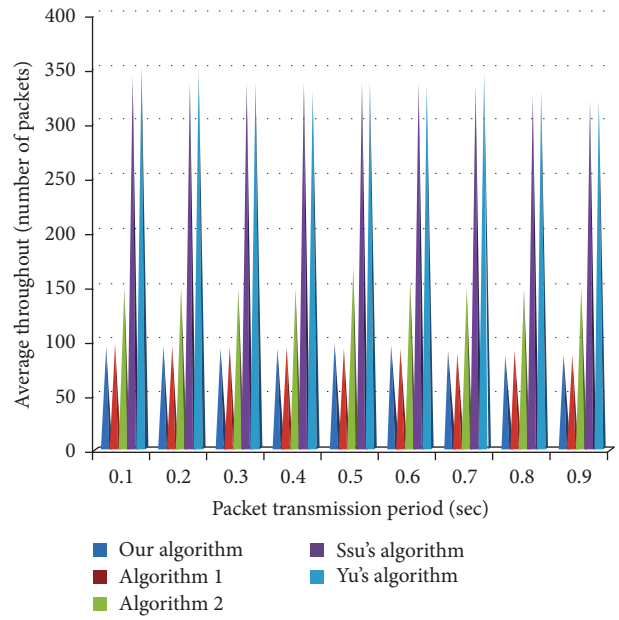



Figure 3: The average of throughput.

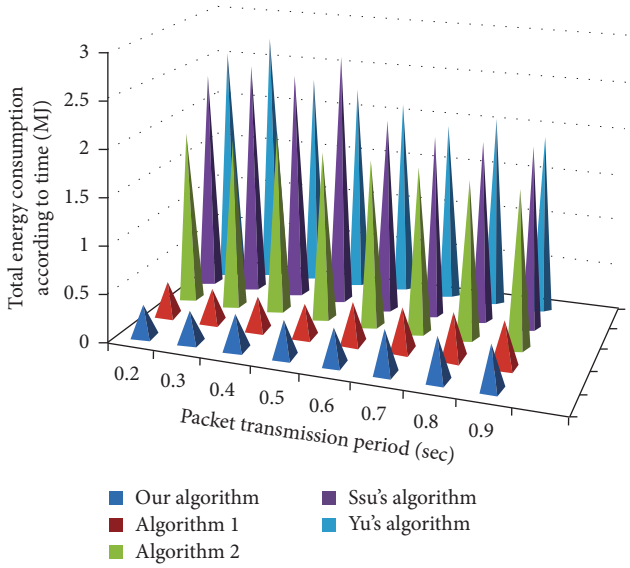



Figure 4: The average of energy consumption.

three mobile guide nodes and the next two algorithms are Ssu's and Yu's methods which are mentioned in [8, 19].

*4.2. Simulation Results.* In this part the simulation result will be shown.

Random distribution of nodes within a $100 * 100\ m^2$ area which are located by mobile and fixed guide nodes is shown in Figure 2.

*(1) Throughput.* The total throughput of the five algorithms is simulated and the results are shown in Figure 3. High performance of the proposed and the first algorithm is clear because they just need three distances to base points.

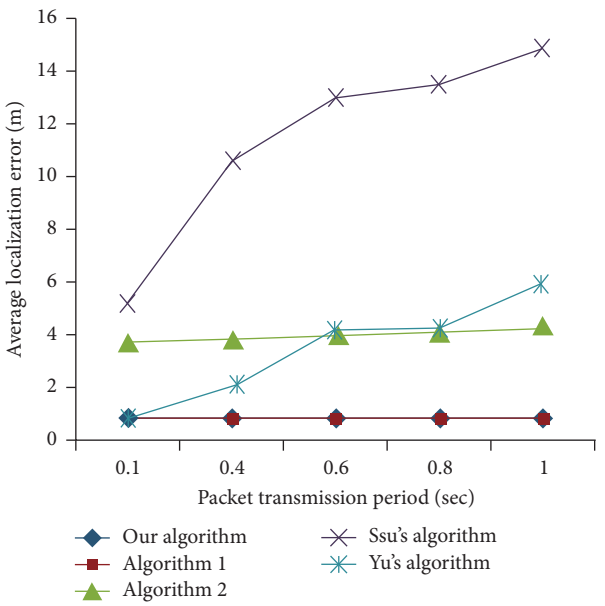



Figure 5: The average of localization error.

Therefore they are less affected by internal transfer of the package compared to the three other projects.

*(2) The Average of Energy Consumption.* The main energy consumption is during packet receiving and transmission. The calculations consume less energy which can be ignored. The average energy consumption of the proposed method in comparison to the four other algorithms is shown in Figure 4. The simulation results show that the energy usage in the proposed and the first algorithms is less than the other three, because, in the proposed and the first algorithms, fixed sensor node is required to receive the position which is distributed by three actual based points, while, in the others, fixed nodes consistently receive the massages to calculate station's locations.

*(3) The Average of Localization Error.* In all localization algorithms, the sensors position is calculated based on points, received by the packets from mobile and fixed guide nodes. Therefore internal packet transmission is strongly influenced by accurate positioning. Figure 5 shows the average of localization error on different times. Simulation results show that the proposed and the first algorithms are highly efficient compared to the others.

## 5. Conclusion and Future Works

This paper presents an algorithm with a good performance that locates sensor nodes by using mobile and fixed guide nodes. To evaluate the performance, four algorithms are presented and their performance is compared with the proposed algorithm. The results of simulation indicate that, by increasing the density (number of $R$), error decreases. The proposed and the first algorithms are similar in throughput, average energy consumption, and average positioning error and much better than the three other algorithms. In terms

of location duration the proposed algorithm has better performance than the second algorithm but lower performance in comparison to the first algorithm. The success rate of the proposed algorithm is extremely better compared to the first algorithm. Due to the fairly good performance of the algorithm especially in success rate of locating, it can be used to simulate a network with the real transmission radius, when more information and details of environment are needed. For the future work, by integrating two stages of the proposed method and simultaneously localization of fixed and mobile nodes the location period will greatly reduce. The problem that may likely appear will be the overlap of fixed and mobile guide nodes. In order to solve the problem, fixed guide nodes can be used for locating intermediate nodes and mobile guide nodes can be used for other nodes.

## Competing Interests

The authors declare that they have no competing interests.

## References


[1] Y. Shang and W. Ruml, "Improved mds-based localization," in *Proceedings of the IEEE Conference on Computer Communications (INFOCOM '14)*, March 2004.

[2] G. Yu, F. Yu, and L. Feng, "A three dimensional localization algorithm using a mobile anchor node under wireless channel," in *Proceedings of the IEEE International Joint Conference on Neural Networks—IEEE World Congress on Computational Intelligence (IJCNN-WCCI '08)*, June 2008.

[3] P. Kułakowski, J. Vales-Alonso, E. Egea-López, W. Ludwin, and J. García-Haro, "Angle-of-arrival localization based on antenna arrays for wireless sensor networks," *Computers and Electrical Engineering*, vol. 36, no. 6, pp. 1181–1186, 2010.

[4] I. F. Akyildiz, W. Su, Y. Sankara subramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002.

[5] M. M. Noel, P. P. Joshi, and T. C. Jannett, "Improved maximum likelihood estimation of target position in wireless sensor networks using particle swarm optimization," in *Proceedings of the 3rd International Conference on Information Technology: New Generations (ITNG '06)*, pp. 274–278, IEEE, Las Vegas, Nev, USA, April 2006.

[6] P. P. Joshi and T. C. Jannett, "Performance-guided reconfiguration of wireless sensor networks that use binary data for target localization," in *Proceedings of the 3rd International Conference on Information Technology: New Generations (ITNG '06)*, pp. 562–565, IEEE, Las Vegas, Nev, USA, April 2006.

[7] Y. Shang, W. Ruml, Y. Zhang, and M. P. J. Fromherz, "Localization from mere connectivity," in *Proceedings of the 4th ACM International Symposium on Mobile Ad Hoc Networking & Computing (MobiHoc '03)*, pp. 201–212, Annapolis, Md, USA, June 2003.

[8] T. He, C. Huang, B. M. Blum, J. A. Stankovic, and T. Abdelzaher, "Range-free localization schemes for large scale sensor networks," in *Proceedings of the ACM Conference on Mobile Computing and Networks (MobiCom '03)*, pp. 81–95, San Diego, Calif, USA, September 2003.

[9] A. Mesmoudi, M. Feham, and N. Labraoui, "Wireless sensor networks localization algorithms: a comprehensive survey," in *Proceedings of the nternational Journal of Computer Networks & Communications (IJCNC '13)*, vol. 5, 2013.

[10] K.-F. Ssu, C.-H. Ou, and H. C. Jiau, "Localization with mobile anchor points in wireless sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 54, no. 3, pp. 1187–1197, 2005.

[11] V. Chaurasiya, R. L. Lavavanshi, S. Verma, G. C. Nandi, and A. Srivastava, "Localization in wireless sensor networks using directional antenna," in *Proceedings of the IEEE International Advance Computing Conference (IACC '09)*, pp. 131–134, IEEE, Patiala, India, March 2009.

[12] Y. Ohta, M. Sugano, and M. Murata, "Autonomous localization method inwireless sensor networks," in *Proceedings of the 3rd IEEE International Conference on Pervasive Computing and Communications Workshops*, pp. 379–384, Kauai Island, Hawaii, USA, March 2005.

[13] K. Langendoen and N. Reijers, "Distributed localization in wireless sensor networks: a quantitative comparison," *Computer Networks*, vol. 43, no. 4, pp. 499–518, 2003.

[14] M. Anlauff and A. Sünbül, "Deploying localization services in wireless sensor networks," in *Proceedings of the 24th International Conference on Distributed Computing Systems Workshops (ICDCSW '04)*, pp. 782–787, IEEE, Tokyo, Japan, March 2004.

[15] H. Chen, B. Liu, P. Huang, J. Liang, and Y. Gu, "Mobility-assisted node localization based on TOA measurements without time synchronization in wireless sensor networks," *Mobile Networks and Applications*, vol. 17, no. 1, pp. 90–99, 2012.

[16] J. Shu, C. Yan, and L. Liu, "Improved three-dimensional localization algorithm based on volume-test scan for wireless sensor networks," *Journal of China Universities of Posts and Telecommunications*, vol. 19, no. 2, pp. 1–6, 2012.

[17] P. Desai, N. Baine, and K. Rattan, "Fusion of RSSI and TDoA measurements from wireless sensor network for robust and accurate indoor localization," in *Proceedings of the International Technical Meeting of the Institute of Navigation (ITM '11)*, pp. 223–230, San Diego, Calif, USA, January 2011.

[18] Y. Zhu, B. Zhang, F. Yu, and S. Ning, "A RSSI based localization algorithm using a mobile anchor node for wireless sensor networks," in *Proceedings of the International Joint Conference on Computational Sciences and Optimization (CSO '09)*, pp. 123–126, April 2009.

[19] Y. Zhang, Y. Chen, and Y. Liu, "Towards unique and anchor-free localization for wireless sensor networks," *Wireless Personal Communications*, vol. 63, no. 1, pp. 261–278, 2012.

Hindawi Publishing Corporation
Journal of Sensors
Volume 2016, Article ID 9528050, 11 pages
http://dx.doi.org/10.1155/2016/9528050

*Research Article*

# A Study on Water Pollution Source Localization in Sensor Networks

**Jun Yang[1] and Xu Luo[2]**

[1]*Department of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan, Hubei Province 430081, China*
[2]*Department of Information Engineering, Zunyi Medical University, Zunyi, Guizhou Province 563000, China*

Correspondence should be addressed to Xu Luo; silyaseln@live.cn

Received 19 May 2016; Revised 20 August 2016; Accepted 22 August 2016

Academic Editor: Mahdi Ghasemi-Varnamkhasti



The water pollution source localization is of great significance to water environment protection. In this paper, a study on water pollution source localization is presented. Firstly, the source detection is discussed. Then, the coarse localization methods and the localization methods based on diffusion models are introduced and analyzed, respectively. In addition, the localization method based on the contour is proposed. The detection and localization methods are compared in experiments finally. The results show that the detection method using hypotheses testing is more stable. The performance of the coarse localization algorithm depends on the nodes density. The localization based on the diffusion model can yield precise localization results; however, the results are not stable. The localization method based on the contour is better than the other two localization methods when the concentration contours are axisymmetric. Thus, in the water pollution source localization, the detection using hypotheses testing is more preferable in the source detection step. If concentration contours are axisymmetric, the localization method based on the contour is the first option. And, in case the nodes are dense and there is no explicit diffusion model, the coarse localization algorithm can be used, or else the localization based on diffusion models is a good choice.

## 1. Introduction

Water pollution, one of the accident-prone man-made disasters, is attaining more and more attention. The pollution source localization in water is of great importance in water conservation. There are many existing water pollution source detection and localization methods, such as robots under water and artificial detection. However, underwater robots are expensive and prone to failure and thus cannot keep working. And artificial detection is time-consuming and vulnerable to water terrain and weather conditions. As a result, sensor networks are applied in the pollution source localization to overcome the deficiencies of the two methods. The advantages of sensor networks include the following: the node distribution is relatively dense; the monitoring range is large; and the monitoring is not restricted by geographical locations [1, 2].

The problem of water pollution source localization is how to locate the pollution source based on the known parameters such as node locations, sampling times, and sensing values of nodes. In the pollution source localization, the pollution source detection is the premise of the source localization. Only when the pollution source has been detected can the monitoring values of nodes be used in the pollution source localization.

In this paper, a study on the water pollution source localization in sensor networks is presented. The localization problem is discussed theoretically and practically. Firstly, the source detection problem is studied. Then, different water pollution source localization methods are introduced and analyzed. Finally, different source detection and localization methods are tested and compared in the experiments.

## 2. Pollution Diffusion in Water

In most cases, water pollution disasters are caused by the static source which discharges the sewage clandestinely.

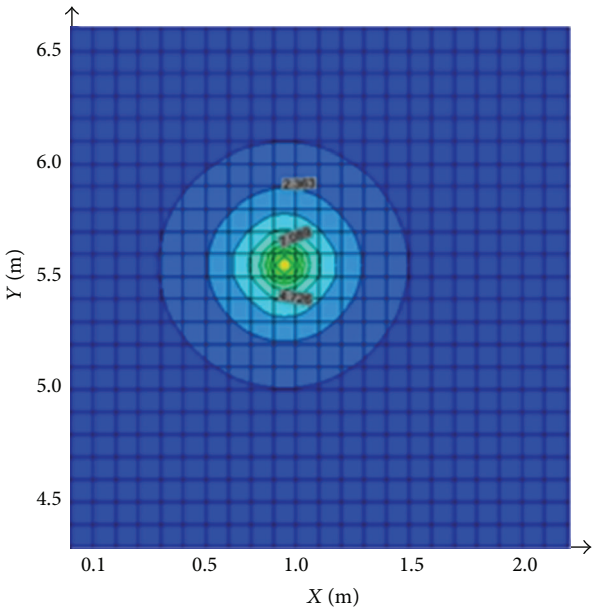



Figure 1: The concentration field without boundary constraints.

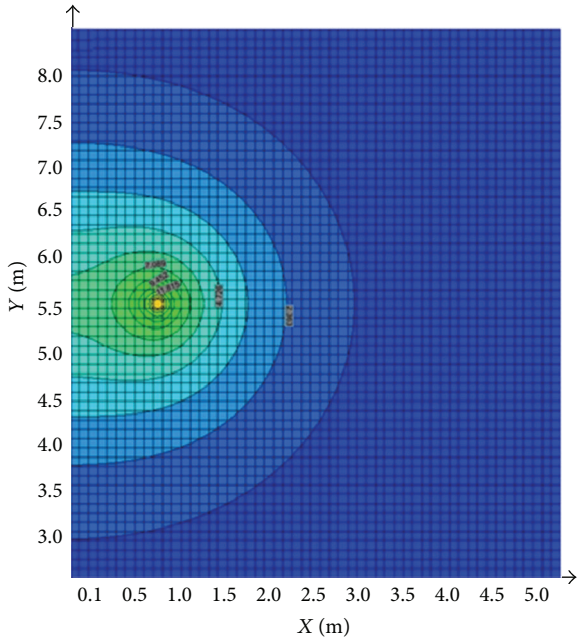



Figure 2: The concentration field with a boundary constraint.

Before choosing a source localization method, the physical processes of the pollution diffusion must be known.

The diffusion of pollutants in static water is slow, while in flow water the pollutants migrate with water and the diffusion is relatively faster.

In different backgrounds, the diffusion is different as well. In this paper, three typical examples are displayed: the diffusion without boundary constraints in static water, the diffusion with a boundary constraint in static water, and the diffusion with water flow. The diffusion with a boundary constraint is different from the diffusion without boundary constraints. Figures 1 and 2 show the diffusion simulations in MODFLOW [3, 4] which is standard software for the hydrological simulation of pollution diffusion. As shown in the figures, when the diffusion is not affected by the boundary, the concentration contours in the diffusion field are approximate circles; and as time goes on, the diffusion is influenced by the boundary, and the concentration contours deform.

An example of the case shown in Figure 1 is the following: in shallow water, there is an instantaneous source at $(\zeta, \eta)$, the mass of the pollutant is $M$, and the diffusion coefficient is $D$. The concentration at $(x_i, y_i)$ is [5]

$$C(x_i, y_i, t) = \frac{M}{4\pi D(t - t_0)} \cdot \exp\left\{-\frac{1}{4(t - t_0)D}\left[(x_i - \zeta)^2 + (y_i - \eta)^2\right]\right\}, \quad (1)$$

where $t_0$ is the initial diffusion time and $t$ is the current time.

An example of the case shown in Figure 2 is the following: in shallow water, the water depth is $f$, there is a continuous source at $(\zeta, \eta)$ with the mass flow rate of the pollutant $M_0$, and the diffusion coefficient is $D$. The concentration at $(x_i, y_i)$ is [5]

$$C(x_i, y_i, t) = \frac{M_0}{2f\sqrt{\pi D}}\left[\frac{1}{r(x_i, y_i)}\operatorname{erfc}\left(\frac{r(x_i, y_i)}{2\sqrt{D(t - t_0)}}\right) + \frac{1}{r^-(x_i, y_i)}\operatorname{erfc}\left(\frac{r^-(x_i, y_i)}{2\sqrt{D(t - t_0)}}\right)\right], \quad (2)$$

where $r(x_i, y_i) = \sqrt{(x_i - \zeta)^2 + (y_i - \eta)^2}$, $r^-(x_i, y_i) = \sqrt{(x_i + \zeta)^2 + (y_i - \eta)^2}$, and the erfc function with variable $\lambda$ is

$$\operatorname{erfc}(\lambda) = 1 - \operatorname{erf}(\lambda) = 1 - \frac{2}{\sqrt{\pi}}\int_0^\lambda e^{-\gamma^2} d\gamma. \quad (3)$$

In dynamic water, the contaminants migrate with flow water. An example is that, in shallow water, the water flow is along the $y$ direction with the flow rate $u$, there is an instantaneous source at $(\zeta, \eta)$, the mass of the pollutant is $M$, and the diffusion coefficient is $D$. The concentration at $(x_i, y_i)$ is [5]

$$C(x_i, y_i, t) = \frac{M}{4\pi D}\exp\left\{-\frac{1}{4tD}\left[(y_i - \zeta - ut)^2 + (x_i - \eta)^2\right]\right\}. \quad (4)$$

In this case, as the diffusion is affected by the flow, the pollution locates on one side of the source, and the concentration field is as shown in Figure 3.

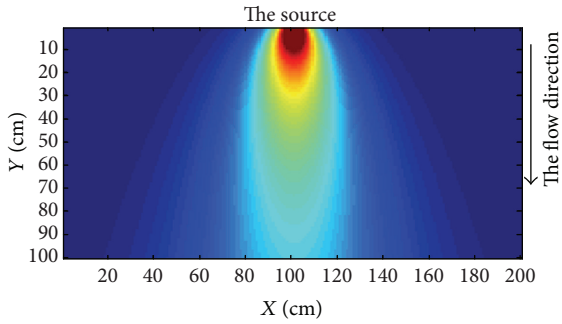



FIGURE 3: The concentration field in dynamic water.

The cases introduced above are the special ones in which the diffusion can be specified by explicit diffusion models. In most cases, the diffusion is influenced by many factors such as shearing flow, turbulent flow, and dispersion, and it is difficult to specify the diffusion processes by diffusion models with explicit expressions.

## 3. Pollution Source Localization

In this part, the background information about the sensor network is given firstly, followed by the pollution detection, and the pollution source localization methods are suggested at last.

*3.1. Network Deployment.* The self-organizing sensor network is used to monitor the pollution in the water. $N$ (>5) sensor nodes are deployed in the monitoring area uniformly and the type 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. After the initialization, the sensor nodes know their own positions. All static nodes in the network sample and store the concentration values synchronously with the same time interval. The background information such as the diffusion coefficient, the water depth, and the sampling time interval is known previously. The upper computer is the data processing center, and the monitoring information is routed to the sink node and processed by the data center. Literatures [6, 7] are the references for the specific self-organizing scheme and data routing scheme.

*3.2. The Detection Methods.* The detection problem is as follows: based on the known sampling data $\{\widetilde{C}(x_i, y_i, t_l),\ l = \ell, \ell+1, \ell+2, \ldots, \ell+L\}$ of node $(x_i, y_i)$, whether the node finds the water pollution source at time $t_{\ell+L}$ is tested. $\widetilde{C}(x_i, y_i, t_l) = C(x_i, y_i, t_l) + C_0 + e$, where $C_0$ is the initial pollutant concentration in water, $e \sim N(0, \delta^2)$ is the monitoring noise, and $C(x_i, y_i, t_l)$ is the theoretical concentration value.

*(A) The Simple Detection Method.* The current detection method available is the simple detection method, for example, in water pollution monitoring applications by using sensor networks [8–13]. In these efforts, the authors consider that if the nodes have monitoring values or the monitoring values are larger than a given threshold $\alpha'$, there is a pollution event.

Since there is an initial pollution concentration in normal production and life, 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 and plants, and so forth, which intervene in water pollution monitoring and bring about disturbances to the monitoring data. Therefore, it is difficult to determine an empirical threshold in the simple detection method. If the given value is less than the pollution concentration of normal production and living sewage, it will induce high false report rate. And if the given value is too large, the water area will be heavily stained when the network alarms the pollution risk.

*(B) The Detection Based on Monitoring Data.* To overcome the defects of the simple detection method, the detection method based on hypothesis testing can be used.

The present author once gave a simple detection method by using hypothesis testing [14] and in the work it is assumed that the distribution of noise $e$ is known. However, in the practical environment, $\delta$ is often unknown. The work in this paper can handle this problem and the specific method is as follows.

Investigate the mean value $\mu_k$ of the monitoring data $\widetilde{C}(x_i, y_i, t_{\ell+k}) - \widetilde{C}(x_i, y_i, t_\ell)$, $k = 1, 2, \ldots, L$. The binary hypotheses are given by

$$H_0^{(1)} : \mu_\kappa = 0,$$
$$H_1^{(1)} : \mu_\kappa \neq 0. \tag{5}$$

The test statistic is

$$G_1 = \frac{\overline{C}_\kappa}{S_\kappa \sqrt{L}}, \tag{6}$$

where $\overline{C}_\kappa = (1/L)\sum_{k=1}^{L}(\widetilde{C}(x_i, y_i, t_{l+k}) - \widetilde{C}(x_i, y_i, t_l))$ and $S_\kappa^2 = (1/(L-1))\sum_{k=1}^{L}(\widetilde{C}(x_i, y_i, t_{l+k}) - \widetilde{C}(x_i, y_i, t_l) - \overline{C}_\kappa)^2$.

If the concentration variation satisfies

$$|G_1| \geq t_{\alpha/2}(L-1), \tag{7}$$

that is, when $P(|G_1| \geq t_{\alpha/2}(L-1)) = \alpha$, reject $H_0^{(1)}$, and it is deduced that the node has detected the pollution source. Note that $\alpha$ is a given significance level and $t_{\alpha/2}$ is the $\alpha/2$ quantile of $t$-distribution.

In hypotheses testing, there are some empirical values of the significance level [15]. As the detection based on sensing data is to test the difference between sensing values, it does not care about the initial pollution concentration in normal production and living sewage, and the detection accuracy is not influenced by a single sample.

*3.3. The Source Localization Methods*

*3.3.1. The Coarse Localization Algorithms and the Localization Based on Diffusion Models.* The coarse localization

algorithms and the localization methods based on diffusion models are often used in water pollution source localization.

*(A) The Coarse Localization Algorithms*

*(1) The Maximum Monitoring Value Point Approach (MPA) [16].* As the sensor node with the maximum monitoring value is always very close to the pollution source, the location of the sensor node with the maximum monitoring value in the network is the source location.

*(2) The Earliest Detection Point Approach (EPA).* The source location is the location of the sensor node which detects the pollution the first time.

*(B) The Localization Algorithms Based on Diffusion Models.* The mathematical localization algorithms are based on the diffusion models, such as [17–20].

If $C(x_i, y_i, t_l, \zeta, \eta)$ is the theoretical concentration of node $(x_i, y_i)$ provided by the diffusion model, $\overline{C}(x_i, y_i, t_l) = C(x_i, y_i, t_l, \zeta, \eta) + \theta$ is the corresponding monitoring value with noise $\theta$, and $f_j(\zeta, \eta)$, $j = 1, 2, 3, \ldots$, is the related constraints of $(\zeta, \eta)$, under assumptions of whether the distributions of the measurement noise are known or not and the distributions are normal distributions or not, there are many estimation methods that can be available, such as Maximum Likelihood estimation [21], Bayesian estimation [22, 23], Extended Kalman filter [24], and Least Squares. The commonly used one is the Least Squares as follows:

$$\begin{aligned} \min_{\zeta,\eta} \quad & \sum_{i=1}^{n} \left[\overline{C}_i - C\left(x_i, y_i, \zeta, \eta\right)\right]^2 \\ \text{s.t.} \quad & f_j\left(\zeta, \eta\right), \quad j = 1, 2, 3, \ldots. \end{aligned} \tag{8}$$

The advantage is that this method is simple and can be applied in the practical applications when the distribution of $\theta$ is unknown.

*3.3.2. Analyses on the Localization Methods Above*

*(A) The Coarse Localization Algorithms.* The premise of the coarse localization algorithms is that small sampling errors occur to the MPA node and the EPA node. And the localization accuracy depends on the density of nodes. Theoretically, if the nodes are dense enough and there is a sensor node at any location in the water area, the pollution source localization would be very accurate.

For the coarse localization algorithms, when the pollution source is in the monitoring area (as shown in Figure 4), the location error is $0 \sim \sqrt{2}d$, where $d$ is the distance between the two farthest neighbor nodes. When all the nodes are far from the pollution source (as shown in Figure 5), the coarse localization algorithms fail to show the accurate estimation. Thus, the location errors are related to the distance of the source from the monitoring area.

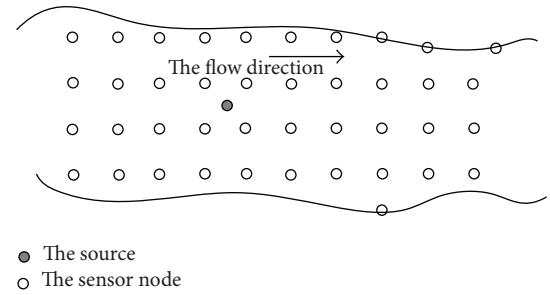



Figure 4: The pollution source in the monitoring area.

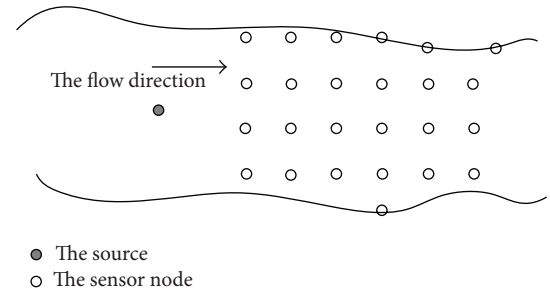



Figure 5: The pollution source off the monitoring area.

*(B) The Algorithms Based on Diffusion Models.* In the localization algorithms based on diffusion models, there are two key points.

*(1) Determining the Diffusion Model.* In practical applications, the diffusion is sophisticated. In many cases, there are no explicit mathematical models of the diffusion.

And, actually, one reason for estimation errors in the localization based on diffusion models is that the theoretical diffusion models are under ideal hypotheses and not accurate.

*(2) How to Solve the Mathematical Problem of the Localization.* For example, if the localization problem is a nonlinear Least Squares problem, there are many solving algorithms, such as the interior point trust-region method [25], Levenberg-Marquardt method [26], and Reflective Newton method [27]. The results are always different when different solving algorithms are used and the number of iterations in numerical calculation is different. In most cases, the unknown parameters are not only source positions but also the mass flow rate and the initial diffusion time, which bring about coupling interferences in the estimation.

*3.3.3. The Localization Algorithms Based on the Contour.* Combining with the above analyses, there are many problems in the coarse localization algorithms and the localization based on diffusion models. In this paper, a localization algorithm based on the concentration contour is proposed when the concentration contours are axis-symmetric, like the contours shown in Figures 1 and 2. The localization method

is independent of the diffusion models and is discussed in the cases below.

*(A) The Source Localization Based on the Contour in Static Water.* The rectangular coordinate system is as shown in Figures 1 and 2; if there is a bank, the direction along the bank is $y$. Under the rectangular coordinate system, the symmetry axis is $y = \eta$. The location of the diffusion source $(\zeta, \eta)$ is on the axis of symmetry.

First, if there are two nodes $(x_1, y_1)$ and $(x_1, y_2)$ with the same $x$-coordinate value on the same contour, it can be obtained that

$$\eta = \frac{y_1 + y_2}{2}. \tag{9}$$

Second, even if there is concentration superimposed effect, the points far from the bank on the contour are still on a circle. Choose any $n$ points marked as $(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n)$ to locate the pollution source; one can obtain

$$\begin{aligned} (x_1 - \zeta)^2 + (y_1 - \eta)^2 &= r^2, \\ (x_2 - \zeta)^2 + (y_2 - \eta)^2 &= r^2, \\ &\vdots \\ (x_n - \zeta)^2 + (y_n - \eta) &= r^2, \end{aligned} \tag{10}$$

where $r$ is the circle radius. And it can be written as

$$\begin{aligned} x_1^2 - x_n^2 + y_1^2 - y_n^2 - 2(x_1 - x_n)\widehat{\zeta} - 2(y_1 - y_n)\eta &= 0, \\ x_2^2 - x_n^2 + y_2^2 - y_n^2 - 2(x_2 - x_n)\widehat{\zeta} - 2(y_2 - y_n)\eta &= 0, \\ &\vdots \\ x_{n-1}^2 - x_n^2 + y_{n-1}^2 - y_n^2 - 2(x_{n-1} - x_n)\widehat{\zeta} & \\ - 2(y_{n-1} - y_n)\eta &= 0. \end{aligned} \tag{11}$$

It can be obtained that

$$\widehat{\zeta} = \frac{1}{n-1}\sum_{i=1}^{n-1}\frac{\left(x_i^2 - x_n^2 + y_i^2 - y_n^2 - 2(y_i - y_n)\eta\right)}{2(x_i - x_n)}, \tag{12}$$

and the residual $\Lambda$ is

$$\begin{aligned} \Lambda = \sum_{i=1}^{n-1}\Big| & x_i^2 - x_n^2 + y_i^2 - y_n^2 - 2(x_i - x_n)\widehat{\zeta} \\ & - 2(y_i - y_n)\eta\Big|. \end{aligned} \tag{13}$$

Based on formulas (9)~(12), the whole localization algorithm is as follows.

*Assumptions*. The rectangular coordinate system is as shown in Figures 1 and 2. If there is a bank, the direction along the bank is $y$. There are some nodes with the same unidirectional coordinate value.

*Step 1.* Give a threshold $\beta$ and let the counting marks be $l = 1$ and $j = 1$.

*Step 2.* At sampling time $t_l$, connect any two points $(x, y)$ and $(x', y')$ when the concentrations $C(x', y', t_l)$ and $C(x, y, t_l)$ satisfy $|C(x', y', t_l) - C(x, y, t_l)| \le \beta$.

*Step 3.* If the number of connected nodes is larger than 4, it can be deduced that the nodes are in the same contour; go to Step 4. Otherwise, $l$ is adjusted to $l = l+1$ and return to Step 2.

*Step 4.* Let $N$ be the number of nodes which are on the same contour and obtained in Step 3. In the $N$ nodes, if there are two nodes $(x_1, y_1)$ and $(x_1, y_2)$ with the same $x$ position, the estimation of $\eta$ can be calculated as (9).

*Step 5.* Use the SL-n [28] algorithm to get the estimation; that is, in the $N$ points which are on the same contour, choose any $n$ points to get the estimation as (12). The $C_N^n$ results $(\widehat{\zeta}^{(j)}, \Lambda^{(j)})$, $j = 1, 2, 3, \ldots, C_N^n$, can be obtained.

*Step 6.* Search the minimum value in $\{\Lambda^{(j)}\}$ and let the corresponding location estimation of the minimum value be the ultimate estimation of $\zeta$.

*(B) The Source Localization Based on the Contour in Dynamic Water*

*Assumptions*. The rectangular coordinate system is as shown in Figure 3 and the $y$ direction is perpendicular to the bank. The monitoring area is mesh covered densely.

Under the rectangular coordinate, the symmetry axis is $x = \eta$. As the diffusion is affected by the water flow, the diffusion is unidirectional. Along the flow, the straight line between the two nodes in the same contour is parallel to the $x$-axis. In the innermost contour, if there are two nodes $(x_1, y_1)$ and $(x_1, y_2)$, one has

$$\begin{aligned} \zeta &= \frac{x_1 + x_2}{2}, \\ \eta &= y_1. \end{aligned} \tag{14}$$

The premise of the localization method based on the contour is that the concentration contours are axis-symmetric and there are enough nodes on the same contour. For the localization in static water, the selection nodes are in the outer contour and off the bank boundary. For the localization in dynamic water, if the pollution source in the monitoring area is as shown in Figure 4, under the mesh covered nodes deployment, the location error is $0 \sim \widetilde{d}$, where $\widetilde{d}$ is the distance between the neighbor nodes in the same row (line). If the pollution source is out of the monitoring area and is as shown in Figure 5, the location accuracy is related to how far the pollution source is from the monitoring area.

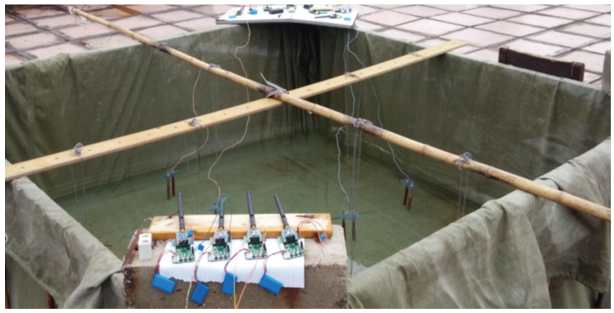


Figure 6: The experiment background of Experiment 1.

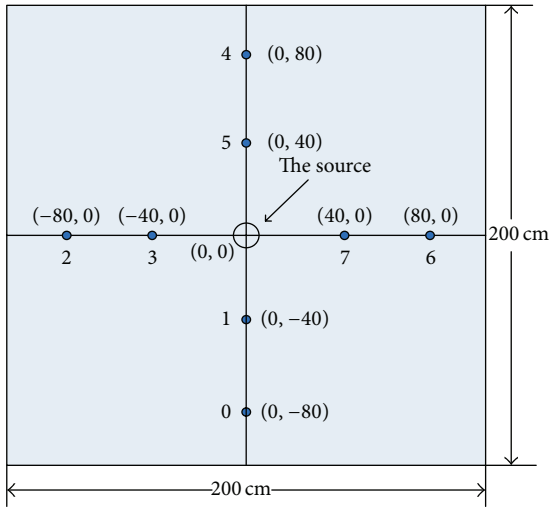



Figure 7: The locations of the source and the nodes.

## 4. Experiment Results and Discussions

### 4.1. Experiments

*Experiment 1* (the source localization in the concentration field without boundary constraints in static water).

*Background.* In shallow water, of which the size is 200 cm × 200 cm and the average depth is $f = 100$ cm, there is a continuous source at the center. Starting from $t_0 = 0$, the solution of $MgSO_4$ is discharged to the water. The background of the experiment is shown in Figure 6. The locations of the source and the sampling nodes are shown in Figure 7. In the initial state, the pollutant migrates with the solution flow. At some time, the diffusion would be stable, and the whole contaminated area can be deemed as a point source.

The diffusion process can be depicted by the diffusion model (1). The monitoring concentration values of different sensor nodes are shown in Table 1. The initial observation is at 5 s.

*The Detection Using Hypotheses Testing.* At different significance levels, the detection results are listed in Table 2.

*The Simple Method to Detect the Source.* For different thresholds, the detection results are listed in Table 3.

*Localization Using Different Methods.* The localization problem is to estimate the source location $(\zeta, \eta)$ based on the known information such as the node locations, the sampling times, the concentration samples of nodes, and the water depth. The localization results of different methods are shown in Table 4.

In Table 4, the localization based on the diffusion model is the Least Squares method as (8) with no constraint, and the data being used is the monitoring values at 20 s. The coarse localization result is the MPA point at 20 s. In the localization based on the contour, the threshold $\beta$ which connects the nodes on the same circle is 0.02 g/L, and the result is the average value of the localization result using nodes 0, 2, 4, and 6 and localization result using nodes 1, 3, 5, and 7. In the experiment, it can be seen that the performance of the localization based on the contour is the best and the coarse localization algorithm is the worst.

*The Results of Experiment 1.* The detection method using hypotheses testing is more stable. In the simple detection, in order to detect the pollution source timely, the threshold should be as small as possible. The performance of the localization method based on the contour is better than the coarse localization algorithm and the localization based on the diffusion model. The results of the localization based on the diffusion models vary with different initial values.

*Experiment 2* (the source localization in the concentration field with a boundary constraint in static water).

*Background.* In shallow water, of which the size is 10 m × 10 m and the average water depth is $f = 10$ m, apart from the impermeable bank $Y$, there is a continuous source at $(\zeta, \eta) = (1.05, 6.05)$ (m). The pollution solution is discharged to the water from $t_0 = 0$. The mass flow rate $M$ is 100 kg/h. The diffusion coefficient is $D = 1\ m^2/h$.

The diffusion can be depicted by the diffusion model (2). The experiment is studied in a MODFOLW simulation, and the simulation values are shown in Table 5.

*The Detection Using Hypotheses Testing.* At different significance levels, the detection results are listed in Table 6.

*The Simple Method to Detect the Source.* For different thresholds, the detection results are listed in Table 7.

Comparing Table 6 with Table 7, the same conclusions as Experiment 1 can be obtained.

*Localization Using Different Methods.* The localization problem is based on the known information such as the node locations, the sampling times, the concentration samples of nodes, the water depth, and the diffusion coefficient to estimate the source location $(\zeta, \eta)$. The localization results of different methods are as follows.

*The Coarse Localization Algorithm.* In this experiment, the EPA point and the MPA point are the same. If the nodes are

Table 1: The observations of different nodes in Experiment 1.

| Time (s) | Observations (g/L) of node 0 | Observations (g/L) of node 1 | Observations (g/L) of node 2 | Observations (g/L) of node 3 | Observations (g/L) of node 4 | Observations (g/L) of node 5 | Observations (g/L) of node 6 | Observations (g/L) of node 7 |
|---|---|---|---|---|---|---|---|---|
| 5 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00161 | 0.00000 | 0.00000 |
| 10 | 0.00000 | 0.00161 | 0.00000 | 0.01081 | 0.00000 | 0.01597 | 0.00000 | 0.00048 |
| 15 | 0.00000 | 0.01435 | 0.00000 | 0.01823 | 0.00000 | 0.02903 | 0.00000 | 0.00274 |
| 20 | 0.00000 | 0.03210 | 0.00000 | 0.02774 | 0.00000 | 0.03242 | 0.00000 | 0.01290 |
| 25 | 0.00000 | 0.04710 | 0.00000 | 0.04516 | 0.00000 | 0.04518 | 0.00000 | 0.02855 |
| 30 | 0.00000 | 0.05039 | 0.00000 | 0.04677 | 0.00000 | 0.05186 | 0.00000 | 0.04694 |
| 35 | 0.00000 | 0.07069 | 0.00000 | 0.06118 | 0.00000 | 0.07284 | 0.00000 | 0.07167 |
| 40 | 0.00000 | 0.09108 | 0.00000 | 0.08196 | 0.00000 | 0.09333 | 0.00000 | 0.09127 |
| 45 | 0.00000 | 0.10827 | 0.00000 | 0.09882 | 0.00000 | 0.10867 | 0.00000 | 0.10847 |
| 50 | 0.00000 | 0.12439 | 0.00000 | 0.10408 | 0.00000 | 0.12469 | 0.00000 | 0.12541 |
| 55 | 0.00000 | 0.13847 | 0.00000 | 0.12878 | 0.00000 | 0.13949 | 0.00000 | 0.13806 |
| 60 | 0.00000 | 0.15214 | 0.00000 | 0.13827 | 0.00000 | 0.15429 | 0.00000 | 0.15186 |
| 65 | 0.00065 | 0.16300 | 0.00065 | 0.14582 | 0.00097 | 0.16443 | 0.00097 | 0.16057 |
| 70 | 0.00113 | 0.16914 | 0.00113 | 0.15529 | 0.00129 | 0.17014 | 0.00161 | 0.16986 |
| 75 | 0.00210 | 0.17586 | 0.00210 | 0.16129 | 0.00242 | 0.19000 | 0.00194 | 0.17500 |
| 80 | 0.00306 | 0.18014 | 0.00306 | 0.16686 | 0.00274 | 0.18043 | 0.00323 | 0.18000 |
| 85 | 0.00403 | 0.18371 | 0.00403 | 0.17000 | 0.00355 | 0.18457 | 0.00355 | 0.18386 |
| 90 | 0.00500 | 0.18643 | 0.00500 | 0.17257 | 0.00532 | 0.18700 | 0.00468 | 0.18686 |
| 95 | 0.00597 | 0.18829 | 0.00597 | 0.17371 | 0.00581 | 0.18986 | 0.00597 | 0.18857 |
| 100 | 0.00726 | 0.18957 | 0.00726 | 0.17557 | 0.00661 | 0.19114 | 0.00726 | 0.18943 |
| 105 | 0.00806 | 0.19143 | 0.00806 | 0.17571 | 0.00855 | 0.19143 | 0.00823 | 0.19000 |
| 110 | 0.00984 | 0.19029 | 0.00984 | 0.17586 | 0.00984 | 0.19186 | 0.00952 | 0.19043 |
| 115 | 0.01113 | 0.19000 | 0.01113 | 0.17571 | 0.01081 | 0.19171 | 0.01097 | 0.19000 |
| 120 | 0.01194 | 0.18971 | 0.01194 | 0.17514 | 0.01161 | 0.19100 | 0.01242 | 0.18971 |
| 125 | 0.01323 | 0.18843 | 0.01323 | 0.17429 | 0.01339 | 0.20286 | 0.01371 | 0.18886 |
| 130 | 0.01500 | 0.18771 | 0.01500 | 0.17286 | 0.01468 | 0.18914 | 0.01548 | 0.18771 |
| 135 | 0.01613 | 0.19000 | 0.01613 | 0.17714 | 0.01597 | 0.18957 | 0.01613 | 0.18900 |
| 140 | 0.01726 | 0.19029 | 0.01726 | 0.17757 | 0.01758 | 0.18943 | 0.01758 | 0.18943 |
| 145 | 0.01871 | 0.18986 | 0.01871 | 0.17857 | 0.01919 | 0.18971 | 0.01887 | 0.18957 |

Table 2: The detection results at different significance levels in Experiment 1.

| 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) |
|---|---|---|---|---|---|---|---|---|
| 0.01 | 105 | 40 | 100 | 40 | 100 | 50 | 100 | 50 |
| 0.05 | 85 | 30 | 85 | 30 | 85 | 35 | 85 | 35 |
| 0.1 | 80 | 25 | 80 | 25 | 80 | 25 | 75 | 30 |

Table 3: The detection times for different decision thresholds in Experiment 1.

| Threshold (g/L) | Node 0 (s) | Node 1 (s) | Node 2 (s) | Node 3 (s) | Node 4 (s) | Node 5 (s) | Node 6 (s) | Node 7 (s) |
|---|---|---|---|---|---|---|---|---|
| 0.01 | 115 | 15 | 115 | 10 | 115 | 10 | 115 | 20 |
| 0.05 | — | 30 | — | 35 | — | 30 | — | 35 |
| 0.1 | — | 45 | — | 50 | — | 45 | — | 45 |

"—" represents having no result.

Table 4: The localization using different localization methods in Experiment 1.

| The coarse localization algorithm (m) | The localization based on the contour (m) | The localization based on the diffusion model (m) |
|---|---|---|
| (0, 40) | (0, 0) | (0.0205, 0) |

dense and there is a node close to the source, the accuracy of the coarse localization is guaranteed. Otherwise, there is a large localization error.

*The Localization Based on the Diffusion Model*. The nonlinear Least Square with the trust-region-reflective solving method is adopted in the localization based on the diffusion model. The observations at 4.0 h are applied. The localization results are shown in Table 8.

*The Localization Based on the Contour*. The threshold which connects the nodes on the same circle is 0.01 g/L. At 4.0 h, in the same contour, there are $p$ points with the same distance to the source and $q$ superimposed effect-influenced points with different distances to the source, and $p + q = 5$. The $p$ points are chosen from {(1.65, 6.85), (1.65, 5.25), (2.05, 6.05), (1.85, 5.45)}. The $q$ points are chosen from {(0.55, 7.25), (0.15, 4.85), (0.25, 7.25), (0.55, 4.85)}. The SL-4 ($n = 4$) method is used in the experiment and the localization results are shown in Table 9.

*The Results of Experiment 2*. The detection method using hypotheses testing is more stable, and the detection accuracy of the simple detection method depends on the precision of the given threshold. The accuracy of the coarse localization depends on the nodes density. The results of the localization based on the diffusion model vary with initial values and are not stable. In the localization based on the contour, the effect-influenced points bring about a larger location error; the more the effect-influenced points are, the worse the localization accuracy is.

*Experiment 3* (the source localization in dynamic water).

*Background*. In the water area, of which the size is [0, 10] m × [0, 20] m, there is a continuous source at the location $(\zeta, \eta) = (10, 0)$ (m). The pollution solution is discharged to the water with the mass flow rate $M_0 = 100$ kg/h from time $t = 0$. The interval of sampling time is 1 h. The water flow is along the $x$ direction with the flow rate $u = 1$ m/s. The diffusion coefficient is $D = 0.5\ \mathrm{m}^2/\mathrm{h}$.

The diffusion model is [27]

$$C(x, y, t) = \frac{M_0}{4\pi D} \int_0^t \frac{1}{t - \tau} \cdot \exp\left\{ -\frac{[x - \zeta - u(t - \tau)]^2 + (y - \eta)^2}{4D(t - \tau)} \right\} d\tau \tag{15}$$

which is related to the time integration and not an explicit model.

The simulation tool is MATLAB. The sample nodes are mesh grid deployed in the area [0, 10] m × [0, 20] m with the average distance between the neighbor nodes of 1 m.

*The Source Detection*. The same conclusions as Experiments 1 and 2 can be obtained.

*Localization Using Different Methods*. The localization problem is based on the known information including the node locations, the sampling times, and the concentration samples of nodes to estimate the source location $(\zeta, \eta)$. As there is no specific diffusion model, in this case, only the coarse localization and the localization algorithm based on the contour are tested. The experiment results are shown in Table 10.

In the experiment, the EPA point is the same as the MPA point, which is (9, 0) m. In the localization based on contours, the threshold $\beta$ which connects the nodes on the same circle is 0.02 g/L.

Table 5: The observations of different nodes in Experiment 2.

(a)

| Time (h) | Observations (g/L) of node (1.65, 6.85) | Observations (g/L) of node (1.65, 5.25) | Observations (g/L) of node (2.05, 6.05) | Observations (g/L) of node (1.85, 5.45) | Observations (g/L) of node (0.55, 7.25) |
|---|---|---|---|---|---|
| 0.01 | 0.0045 | 0.0043 | 0.0048 | 0.0043 | 0.0020 |
| 0.5 | 0.0435 | 0.0434 | 0.0431 | 0.0431 | 0.0292 |
| 1 | 0.0869 | 0.0869 | 0.0848 | 0.0856 | 0.0739 |
| 1.5 | 0.1212 | 0.1211 | 0.1170 | 0.1187 | 0.1124 |
| 2.0 | 0.1495 | 0.1494 | 0.1437 | 0.1462 | 0.1444 |
| 2.5 | 0.1738 | 0.1736 | 0.1667 | 0.1697 | 0.1716 |
| 3.0 | 0.1882 | 0.1878 | 0.1802 | 0.1833 | 0.1876 |
| 3.5 | 0.1913 | 0.1909 | 0.1832 | 0.1866 | 0.1912 |
| 4.0 | 0.1944 | 0.1940 | 0.1861 | 0.1896 | 0.1946 |

(b)

| Time (h) | Observations (g/L) of node (0.35, 7.55) | Observations (g/L) of node (0.35, 4.55) | Observations (g/L) of node (0.15, 4.85) | Observations (g/L) of node (0.25, 7.25) | Observations (g/L) of node (0.55, 4.85) |
|---|---|---|---|---|---|
| 0.01 | 0.0008 | 0.0006 | 0.0014 | 0.0015 | 0.0018 |
| 0.5 | 0.0161 | 0.0161 | 0.0273 | 0.0274 | 0.0293 |
| 1 | 0.0498 | 0.0498 | 0.0727 | 0.0728 | 0.0739 |
| 1.5 | 0.0826 | 0.0825 | 0.1120 | 0.1122 | 0.1123 |
| 2.0 | 0.1114 | 0.1111 | 0.1446 | 0.1449 | 0.1442 |
| 2.5 | 0.1366 | 0.1359 | 0.1720 | 0.1725 | 0.1711 |
| 3.0 | 0.1517 | 0.1506 | 0.1882 | 0.1888 | 0.1868 |
| 3.5 | 0.1551 | 0.1539 | 0.1918 | 0.1924 | 0.1903 |
| 4.0 | 0.1585 | 0.1571 | 0.1952 | 0.1960 | 0.1937 |

Table 6: The detection times at different significance levels in Experiment 2.

(a)

| Significance level | Node (1.65, 6.85) (h) | Node (1.65, 5.25) (h) | Node (2.05, 6.05) (h) | Node (1.85, 5.45) (h) | Node (0.55, 7.25) (h) |
|---|---|---|---|---|---|
| 0.01 | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 |
| 0.05 | 1.5 | 1.5 | 1.5 | 1.5 | 2.0 |
| 0.1 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 |

(b)

| Significance level | Node (0.35, 7.55) (h) | Node (0.35, 4.55) (h) | Node (0.15, 4.85) (h) | Node (0.25, 7.25) (h) | Node (0.55, 4.85) (h) |
|---|---|---|---|---|---|
| 0.01 | 3.0 | 3.0 | 2.5 | 2.5 | 2.5 |
| 0.05 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 |
| 0.1 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 |

*The Result of Experiment 3*. The performance of the localization based on the contour is better than the coarse localization algorithm.

*4.2. Performance Analyses Based on the Experiments.* All of the above experiment results show that the detection method using hypotheses testing is more stable, and the detection accuracy of the simple detection method is related to the given threshold. The simple detection method can be more timely but the decision threshold should be small. However, if the noise in the practical applications is considered, small thresholds may bring about large false alarm rates.

The simple localization methods can only be used when the nodes are deployed densely; otherwise, the localization error is possible. The results of the localization based on the diffusion models vary with different initial values and are not stable. Actually, in the numerical calculations, the variable boundaries are set previously to ensure the convergence in the iteration calculations. The performance of the localization method based on the contour is better than the coarse localization algorithms and the localization based on diffusion models when concentration contours are axisymmetric and most of the nodes participating in the localization are with the same distances to the source.

Table 7: The detection times for different thresholds in Experiment 2.

(a)

| Threshold (g/L) | Node (1.65, 6.85) (h) | Node (1.65, 5.25) (h) | Node (2.05, 6.05) (h) | Node (1.85, 5.45) (h) | Node (0.55, 7.25) (h) |
|---|---|---|---|---|---|
| 0.05 | 1 | 1 | 1 | 1 | 1 |
| 0.1 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 |
| 0.15 | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 |

(b)

| Threshold (g/L) | Node (0.35, 7.55) (h) | Node (0.35, 4.55) (h) | Node (0.15, 4.85) (h) | Node (0.25, 7.25) (h) | Node (0.55, 4.85) (h) |
|---|---|---|---|---|---|
| 0.05 | 1.5 | 1.5 | 1 | 1 | 1 |
| 0.1 | 2.0 | 2.0 | 1.5 | 1.5 | 1.5 |
| 0.15 | 3.0 | 3.0 | 2.5 | 2.5 | 2.5 |

Table 8: The localization results varying with the initial values.

| The initial values of $(\zeta, \eta)$ (m) | Localization results (m) |
|---|---|
| (1, 5.5) | (0.99, 6.03) |
| (0.55, 7.05) | (0.99, 6.03) |
| (0.55, 7.35) | (10, 6.81) |
| (1.55, 0.55) | (10, 6.80) |
| (0.55, 6.55) | (0.99, 6.03) |

Table 9: The localization results based on the contour.

| $p:q$ | Location error (m) |
|---|---|
| 4 : 1 | 0 |
| 3 : 2 | 0.14 |
| 2 : 3 | 0.43 |
| 1 : 4 | 0.55 |

Table 10: The localization results of different localization methods in Experiment 3.

| The coarse localization algorithm (m) | The localization based on the contour (m) |
|---|---|
| (9, 0) | (10, 0) |

From the results, it can be seen that, in the water pollution source localization, the detection using hypotheses testing works better in the source detection step. If concentration contours are axisymmetric, the localization method based on the contour is the primary selection. If the nodes are dense and there is no explicit diffusion model, the coarse localization algorithm can be used. When there is an explicit diffusion model and small errors can be guaranteed in the numerical calculations, the localization method based on diffusion models is a good choice.

## 5. Conclusions

In this paper, the static pollution source localization including the source detection and the source localization in sensor networks is studied. The simple detection method, the coarse localization methods, and the localization methods based on diffusion models are introduced and analyzed theoretically. The detection method based on hypothesis testing and the localization method based on the contour are proposed. The performances of different detection and localization methods are tested and compared in the paper. Based on theoretical analyses and experiments, it can be concluded that, in the water pollution source localization, the detection using hypotheses testing is more convenient in the source detection step. If concentration contours are axisymmetric, the localization method based on the contour is the primary selection. In case the nodes deployment is dense and an explicit diffusion model is not available, the coarse localization algorithm can be used. Otherwise, the localization based on the diffusion model is a sound choice.

## Competing Interests

The authors declare that they have no competing interests.

## Acknowledgments

This research is sponsored by the National Natural Science Foundation (NNSF) of China under Grants 61463053 and 61501337.

## References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–114, 2002.

[2] Z. E. Kamal and M. A. Salahuddin, "Introduction to wireless sensor networks," in *Wireless Sensor and Mobile Ad-Hoc Networks*, B. D. Driss and A. Al-Fuqaha, Eds., pp. 1–18, Springer Press, New York, NY, USA, 2015.

[3] Aquaveo, "GMS manual," pp. 25–77, http://xmswiki.com/wiki/GMS:GMS_User_Manual_9.2.

[4] M. Q. Feng, B. M. Zheng, and X. D. Zhou, *Water Environment Simulation and Prediction*, Science Press, Beijing, China, 2008.

[5] Z. Z. Peng, T. X. Yang, X. J. Liang, and Z. S. Gu, *Mathematical Model of Water Environment and Its Application*, Chemical Industry Press, Beijing, China, 2007.

[6] H. Byun and J. Park, "A gene regulatory network-inspired self-organizing control for wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 11, no. 8, Article ID 789434, 2015.

[7] S. Rani and S. H. Ahmed, "Multi-hop reliability and network operation routing," in *Multi-hop Routing in Wireless Sensor Networks: An Overview, Taxonomy, and Research Challenges*, S. Rani and S. H. Ahmed, Eds., pp. 471–490, Springer Press, New York, NY, USA, 2015.

[8] 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.

[9] 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.

[10] B. O'Flynn, R. Martinez, and J. Cleary, "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.

[11] 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.

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

[13] J. Wang, X. Wang, X. Ren, and Y. Shen, "A water pollution detective system based on wireless sensor network," *Journal of Guilin University of Electronic Technology*, vol. 29, no. 3, pp. 247–250, 2009.

[14] 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*, pp. 2311–2315, Lijing, China, 2015.

[15] A. R. John, *Mathematical Statistics and Data Analysis*, Duxbury Press, New York, NY, USA, 2006.

[16] Y. Qing, A. Lim, K. Casey, and R.-K. Neelisetti, "Real-time target tracking with CPA algorithm in wireless sensor networks," in *Proceedings of the 5th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON '08)*, pp. 305–313, San Francisco, Calif, USA, June 2008.

[17] L. Fan, C. Li, and Y. Jun, "Diffusion source localization with a water-proof boundary," in *Proceedings of the 10th World Congress on Intelligent Control and Automation (WCICA '12)*, pp. 164–169, IEEE, Beijing, China, July 2012.

[18] J. Yang, L. Chai, and X. Luo, "Pollution source localization in Lake Environment based on wireless sensor networks," in *Proceedings of the IEEE 31st Chinese Control Conference*, pp. 6649–6653, Hefei, China, 2012.

[19] 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.

[20] 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.

[21] J. Xu, J. He, Y. Zhang, F. Xu, and F. Cai, "A distance-based maximum likelihood estimation method for sensor localization in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 12, no. 4, Article ID 2080536, 2016.

[22] T. Zhao and A. Nehorai, "Distributed sequential Bayesian estimation of a diffusive source in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 2213–2225, 2007.

[23] C. F. Huang, T. Hsing, N. Cressie, A. R. Ganguly, V. A. Protopopescu, and N. S. Rao, "Bayesian source detection and parameter estimation of a plume model based on sensor network measurements," *Applied Stochastic Models in Business and Industry*, vol. 26, no. 4, pp. 331–348, 2010.

[24] S. Wen, Z. Cai, and X. Q. Hu, "Constrained extended kalman filter for target tracking in directional sensor networks," *International Journal of Distributed Sensor Networks*, vol. 11, no. 5, Article ID 158570, 2015.

[25] T. F. Coleman and Y. Y. Li, "An interior trust region approach for nonlinear minimization subject to bounds," *SIAM Journal on Optimization*, vol. 6, no. 2, pp. 418–445, 1996.

[26] S. D. Shan, *A Levenberg-Marquardt method for large scale bound-constrained nonlinear least squares [M.S. thesis]*, Acadia University, Wolfville, Canada, 2006.

[27] T. F. Coleman and Y. Y. Li, "On the convergence of interior-reflective Newton methods for nonlinear minimization subject to bounds," *Mathematical Programming*, vol. 67, no. 2, pp. 189–224, 1994.

[28] X. Luo, L. Chai, and J. Yang, "SL-n iterative localization algorithm in wireless sensor networks," *Journal on Communications*, vol. 32, no. 5, pp. 129–131, 2011.

Hindawi Publishing Corporation
Journal of Sensors
Volume 2016, Article ID 3023018, 13 pages
http://dx.doi.org/10.1155/2016/3023018

*Research Article*

# System-Aware Smart Network Management for Nano-Enriched Water Quality Monitoring

**B. Mokhtar,[1,2] M. Azab,[2,3,4] N. Shehata,[2,5,6] and M. Rizk[1,2]**

[1]*Department of Electrical Engineering, Faculty of Engineering, Alexandria University, Alexandria, Egypt*
[2]*Center of Smart Nanotechnology and Photonics (CSNP), SmartCI Research Center, Alexandria University, Alexandria, Egypt*
[3]*Informatics Research Institute, City of Scientific Research and Technological Applications, Alexandria, Egypt*
[4]*ACIS, ECE Department, University of Florida, Gainesville, FL, USA*
[5]*Department of Engineering Mathematics and Physics, Faculty of Engineering, Alexandria University, Alexandria, Egypt*
[6]*USTAR Bioinnovations Center, Utah State University, Logan, UT, USA*

Correspondence should be addressed to B. Mokhtar; bassemmokhtar@gmail.com

Received 18 April 2016; Revised 19 July 2016; Accepted 20 July 2016

Academic Editor: Constantin Apetrei



This paper presents a comprehensive water quality monitoring system that employs a smart network management, nano-enriched sensing framework, and intelligent and efficient data analysis and forwarding protocols for smart and system-aware decision making. The presented system comprises two main subsystems, a data sensing and forwarding subsystem (DSFS), and Operation Management Subsystem (OMS). The OMS operates based on real-time learned patterns and rules of system operations projected from the DSFS to manage the entire network of sensors. The main tasks of OMS are to enable real-time data visualization, managed system control, and secure system operation. The DSFS employs a Hybrid Intelligence (HI) scheme which is proposed through integrating an association rule learning algorithm with *fuzzy* logic and weighted decision trees. The DSFS operation is based on profiling and registering raw data readings, generated from a set of optical nanosensors, as profiles of attribute-value pairs. As a case study, we evaluate our implemented test bed via simulation scenarios in a water quality monitoring framework. The monitoring processes are simulated based on measuring the percentage of dissolved oxygen and potential hydrogen (PH) in fresh water. Simulation results show the efficiency of the proposed HI-based methodology at learning different water quality classes.

## 1. Introduction

The detection of water quality parameters such as dissolved oxygen (DO) and potential hydrogen (PH) in aqueous media is important for wide variety of applications including environmental monitoring, biomedical research, and process control [1–3]. Compared to currently used techniques, fluorescence-based sensing technique has significant advantages over other procedures due to fouling avoidance, the fact that there is no need for a reference electrode, and the resistance to exterior electromagnetic field interferences [4–9]. One of the most promising optical nanostructures is ceria nanoparticles due to its oxygen capability storage, low-cost synthesis, and adequate sensitivity [10, 11]. This paper offers a comprehensive monitoring framework as an integration between nanotechnology and trustworthy wireless sensor networks.

The main objective of our work is to develop a complete sensing platform for real-time monitoring of two water quality parameters, DO concentration and PH value, in aqueous media [12, 13]. Our system, as shown in Figure 1, goes behind local, single location monitoring to a networked sensing of DO and PH concentrations at multiple locations, across streams, water treatment facilities, hydroponic farms, and aquafarms. The system comprises two main subsystems, which are the data sensing and forwarding subsystem (DSFS), and Operation Management Subsystem (OMS). The DSFS employs our proposed hybrid intelligence (HI) technique, which is embedded at powerful on-site computation nodes for efficient data analysis, classification, and

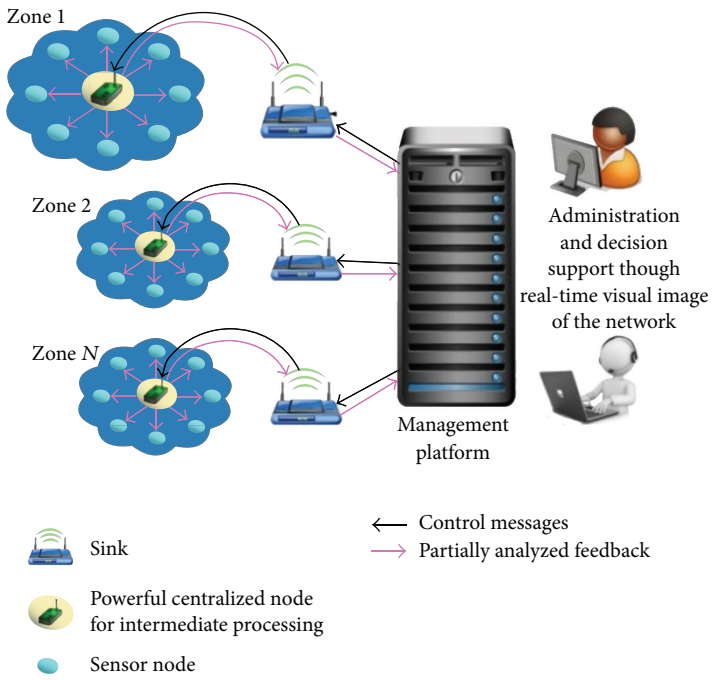



Figure 1: Comprehensive system architecture of data management and decision making.

forwarding processes. Powerful nodes receive the digitalized signal from one or more running nanosensors, merge the collected data with the geographical location of the sensing element(s), and run the HI techniques for learning data patterns and taking on-site decisions. Based on the weight of reached decisions, data might be forwarded to sinks and remote off-site management servers in a real-time fashion for further analysis with security extensions. The real-time management of the overall system with security policies' enforcement is performed by the OMS. The OMS operates based on learned patterns and rules of system operations projected from the DSFS to manage the entire network of sensors. The main tasks of OMS are to enable real-time data visualization, managed system control, and secure system operation.

As shown in Section 3, the main advantage of having such level of autonomic control and management is the ability to predict some of the sensor feedback based on the analysis of the overall network feedback [14–16]. Such ability facilitates optimizing the sensor power usage by controlling the activation periods of sensors leading to much better sensor utilization expanding the lifetime of the entire network and reducing the system cost. Further, such ability facilitates detecting and fixing/excluding any misbehaving or problematic sensing nodes that can massively reduce the system accuracy. Additionally, the automated system could definitely overcome the problems of manual data collection that requires extensive effort and time. This capability enables almost continuous real-time monitoring of DO with means to identify and address sensor drift helping researchers to construct test models based on the flow of the monitored water supply. Both DO and PH are sensed using fluorescence quenching technique depending on the reduction of visible emitted fluorescent optical intensity based on the change of the quencher concentration of dissolved oxygen or the chemical impact of PH in ceria nanoparticles itself.

Additionally, we present in this paper a HI-based scheme for data classification and forwarding in wireless sensor networks (WSNs) enabling energy-efficient, better utilized resources and optimized QoS operations. The main objectives of our proposed scheme, by adopting concepts in the information theory and Artificial Intelligence (AI) [17–19], are as follows:

(i) Building efficient reconfigurable information-driven data classification and forwarding methodology for WSNs.

(ii) Learning interesting routing-based attributes and generating forwarding models.

(iii) Enabling the capability of learning/expecting/detecting abnormal data flows and making classification and generating rules.

The organizing of the remaining sections of this paper is as follows. Section 2, Materials and Methods, discusses three subsections including the methodology of data sensing, forwarding subsystem with the proposed HI-based data classification/forwarding scheme, the experimental setup for sensing of DO and PH optical nanosensors, and the trustworthy Operation Management Subsystem. The obtained results are analyzed and discussed in Section 3. Section 4 concludes the paper and highlights our future work.

## 2. Materials and Methods

*2.1. Data Sensing and Forwarding Subsystem (DSFS).* The networking infrastructure of sensors' communication and data forwarding processes within the DSFS depends on having a network of wirelessly communicated sensors organized in clusters. The continuously growing interest and developed research in Wireless Sensor Networks (WSNs) lead to making such networks widely used in applications related to various fields, such as environmental monitoring [20]. WSNs provide the communication framework for such applications enabling data sensing, collection, and forwarding until reaching the final data destination center for further data analysis and decision making. As known, WSNs comprise limited energy and computation communicating nodes; thus, this leads to great challenges on the lifetime of such networks and consequently their reliability. So, there is a need to propose an efficient data forwarding scheme that considers the energy constraints in WSN and the importance of interesting data in order to route data that have large information gain. In the literature, many trials provided solutions for having efficient data forwarding and routing targeting energy saving at sensor nodes using opportunistic routing theory [21, 22]. Other researchers consider data routing and forwarding depending on data aggregation and clustering approaches [23]. Another work provides an effective data collection using a set of powerful mobile nodes which increases the probability of having available and reliable communication channels [24]. The inter distance and noise level between sensor nodes is sometimes considered for designing data forwarding strategies [25]. However, to the best of our knowledge, prior related work did not provide a data classification and forwarding scheme for WSNs leveraging the capabilities of simple hybrid intelligence techniques and information theory concepts to extract data semantics and learn interesting information. Accordingly, there will be an ability to control classification levels and amounts of data forwarded from areas of interest to remote data analysis and decision making locations. This will, consequently, aid in optimizing power consumption and memory utilization at sensor nodes. Leveraging concepts form information theory; this paper offers a novel hybrid intelligence-based scheme for data classification and forwarding in wireless sensing communication environments enabling energy-efficient, better utilized resources and optimized QoS operations.

As a case study, we run this scheme for a WSN directed to monitor water quality levels. The proposed methodology depends on profiling raw data readings, generated from a set of optical nanosensors, as profiles of attribute-value pairs. Then, data patterns are learnt adopting association rule learning algorithm clarifying the most frequent attributes and their related values. According to the discovered sets of attributes, a set of *fuzzy* membership functions are directed to produce a discrete sample space and a specific membership class for each attribute based on its value. Based on calculated attribute-dependent entropies and information gains, weighted probabilistic decision trees are built to help take decisions of data forwarding and to generate long-term *fuzzy* rules. One of the main objectives of our proposed scheme, by adopting concepts in the information theory and Artificial Intelligence (AI), is building efficient reconfigurable information-driven data classification and forwarding methodology for WSNs [17–19].

*2.1.1. Hybrid Intelligence-Based WSN-Based Data Classification and Forwarding Scheme.* Figure 2 shows the building blocks of the proposed scheme for data classification and forwarding for WSNs. Those blocks illustrate the main used algorithms and operations. We will discuss the comprised blocks shown in Figure 2 as follows.

(i) The shared memory: an accessible memory located at sensor nodes with powerful storage and processing resources. Data readings are stored as profiles of attribute-value pairs. Each data profile refers to a specific reading from a certain sensor node located at a location where that sensor is directed to execute a defined function, such as reading lead-related quality level at fresh water.

(ii) Association rule learning algorithm: a machine learning algorithm, which is adopted to learn data pattern at the shared memory and to extract the most frequent attributes located in the memory according to defined minimum support and confidence thresholds. Those attributes will be forwarded to the fusiliers in order to be classified based on attributes' values.

(iii) *Fuzzifiers*: operating engines which adopt *fuzzy* logic via defining a set of defined *fuzzy* membership functions to generate specific finite discrete classes for the attributes based on their values [26].

(iv) Statistical analysis: this defines a set of statistical operations that are performed on data profiles stored in the shared memory in order to aid in extracting some information leading to computing the information gains. For instance, a specific attribute will be analyzed to know how many times it is found in readings related to pollutants based on measured levels of dissolved oxygen and lead in fresh water at using specific types of sensors.

(v) Entropy and information gain calculation: it runs mathematical operations based on information concepts to calculate the information entropy due to classified attributes and the possible outcome classes after decision making. Then, it computes the information gains for each attribute with respect to the calculated class entropy. Accordingly, the attribute which has the greatest weight on making decisions will be known. Then, other attributes with weights ordered in a descending order will be used to form the decision tree. The next section will discuss an illustrative example to show how calculations are performed.

(vi) Decision trees and *fuzzy* rules generation: based on the above calculations, weighted entropy-based decision trees will be formed and used to make decision. According to these designed decision trees and possible attribute classes, a set of *fuzzy* rules can

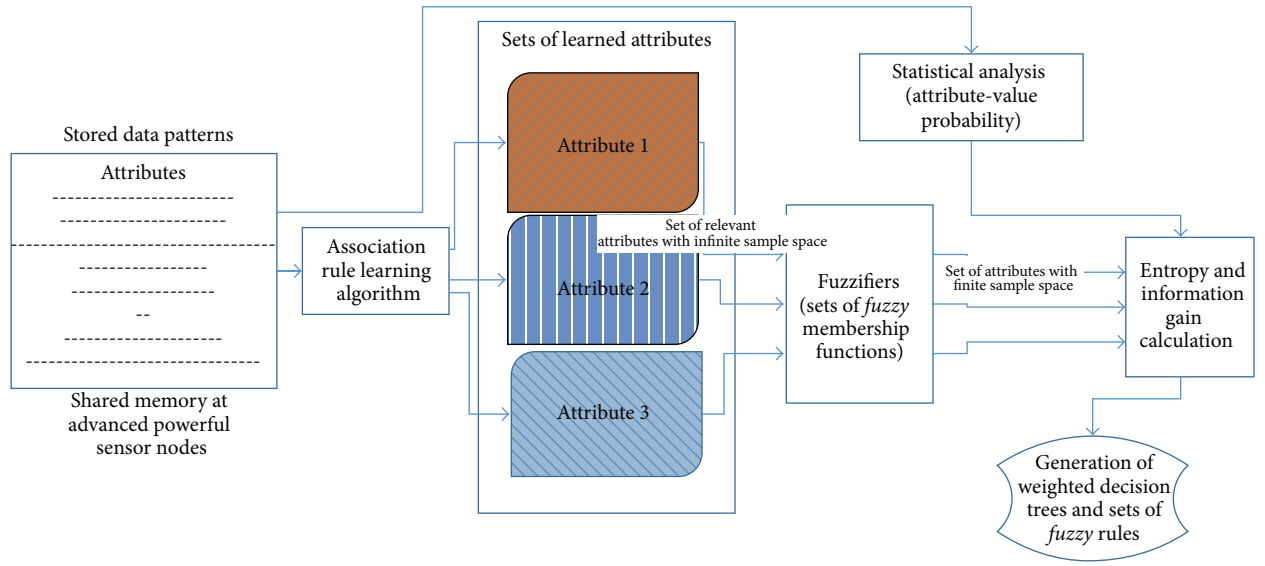



Figure 2: The proposed data classification and forwarding scheme.

be generated which shows the different cases that might be faced when getting a set of attributes within certain range of values.

*2.1.2. Overview of the Technical Insights of the Proposed Scheme for Efficient Water Quality Monitoring.* In this subsection, we will discuss the main concepts on which the proposed hybrid intelligence scheme for data classification and forwarding is developed. We assume a small-scale WSN comprising a set of communicating sensor nodes where some of those nodes, called advanced nodes, are with powerful computing and storage capabilities. This WSN is directed to collect data related to various classes of water pollutants and forward data to a final destination unit in order to make further data analysis and management. Collected data and readings from sensor nodes are stored in a shared on-demand accessible memory, where data entries are indexed according to sensors' identifications and their locations. Advanced nodes can play the role of cluster heads in case of forming a cluster hierarchy for the designed WSN. Also, advanced nodes analyze the patterns of stored data, where these data refer to readings about a certain pollutant. The following subsection will discuss building weighted classifiers, using *fuzzy* logic, entropy, and decision trees, which are adopted by sensor nodes [27].

The following points illustrate how the weighted classifiers are built and work in advanced sensor nodes. As discussed previously, readings from sensors are stored as attribute-value pairs in a shared memory.

(i) The *fuzzy* logic divides the registered attributes with an infinite real space into finite sets with discrete finite space.

(ii) According to obtained attributes' levels and expected related classes, a statistical analysis can be done which will show the percentage/probability of each attribute to a possible class.

(iii) Based on data pattern and using the entropy of all possible classes and the conditional entropy with respect to their related attributes, a *fuzzy* weighted decision trees can be built. Equation (1) calculates the class entropy:

$$H(C) = -\sum_{C=1}^{n} p_C \log_2 p_C, \tag{1}$$

where $p_C$ shows the probability of a certain class found in the analyzed patterns. It can be calculated based on the frequency of a certain class in data patterns. For example, if there are 10 data profiles where 5 of them refer to a water quality class of high level type, then the probability of high level quality class is 50%. Then, to know the weight of each attribute that will affect getting the main class, we will use the conditional entropy as described in

$$\begin{aligned} H\left(\frac{C}{A}\right) &= \sum_{a \in A} (p_a)\, H\left(\frac{C}{A=a}\right) \\ &= -\sum_{a \in A} (p_a) \sum_{C=1}^{n} p_{(C/(A=a))} \log_2 p_{(C/(A=a))}, \end{aligned} \tag{2}$$

where $a$ is a possible value for attribute $A$ and $H(C/(A = a))$ is the conditional entropy of having a certain class when attribute $A$ is of value $a$.

Assume that we have three *fuzzy* attributes (i.e., $A = 3$) that can affect the process of classification. We need to know the more affecting attribute that will be the first stage the classifier. This can be obtained from calculating the conditional entropy for all attributes and their possible values which can be defined from the *fuzzifiers* by which the attributes are with a finite sample space. For instance, there is an attribute $A_{\text{type}}$, such as sensor type, which can take two

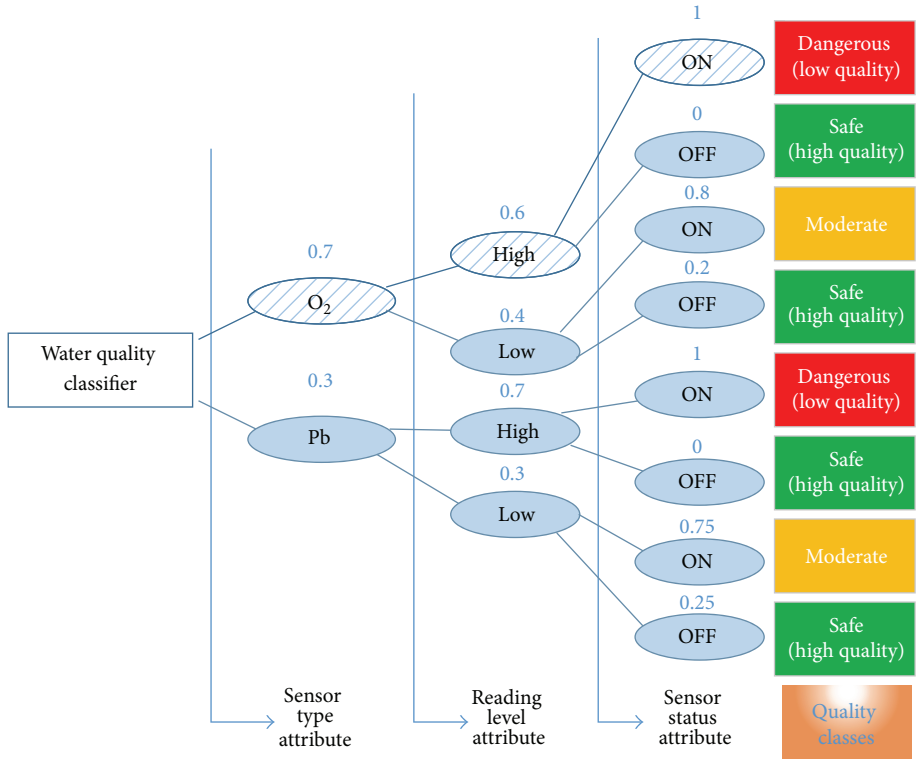



Figure 3: Example of a weighted decision tree.

different string values (i.e., $a = 2$), which are $X$ and $Y$. This is a binary attribute. From patterns, we found that the value of 20% of patterns which contain $A_{\text{type}}$ is $X$. Then, $(p_{a=X}) = 0.2$ and $(p_{a=Y}) = 0.8$.

Then, we will use (1) and (2) to calculate the information gain as stated in

$$\text{Information Gain} = I = I(C; A) = H(C) - H\left(\frac{C}{A}\right). \quad (3)$$

Based on the calculated gain, we can know the attributes with higher priorities and those attributes will be at the first stages or nodes in the decision tree-based classifier. The more the information gain we will have for an attribute, the higher the probability that such attribute will be in the first decision tree nodes.

For example, if $H(C) = 1.5$ bits and we have three attributes, which are $A_{\text{type}}$, $A_{\text{level}}$, and $A_{\text{status}}$, and each attribute has two possible string values, where $H(C/A_{\text{type}}) = 1.1$ bits, $H(C/A_{\text{level}}) = 0.7$ bits, and $H(C/A_{\text{status}}) = 0.5$ bits, this means that information gain according to $A_{\text{type}}$ is the largest one. So, the first node in the decision tree will be the sensor type attribute. Figure 3 shows an example of a decision tree:

(i) Some *fuzzy* rules from the built decision trees can be extracted [28]. Relying on those rules, data forwarding and routing decisions can be taken. For instance, a rule of a low level of quality can be extracted based on the designed decision tree as shown in Figure 3. As an example, if we have a set of ON dissolved $O_2$ sensors with high level readings, then this refers to a low quality level.

As a case study with an illustrative example, we discuss a numerical example which discusses the designed forwarding scheme. We assume that WSN is employed to forward quality-related data readings from a set of sensors to a final destination in a fresh water context like a river. The implemented WSN will comprise two main types of sensors, which are normal sensors and advanced sensors. The proposed scheme will be implemented in advanced sensor nodes, which possess powerful computing resources and communication capabilities. Advanced nodes will be able to register readings from normal nodes in a shared memory at them allowing other nodes to access and get more information (e.g., learning data patters concerning a certain pollutant during a specific period in the year). Readings will be represented by advanced nodes in their memories as data profiles of attribute-value pairs. The data profile will show the identification (ID) of the data source sensor node, type of sensor node, status of the sensor (ON, OFF, and maintenance), and level of readings (e.g., high or low).

Normal nodes will be equipped with two types of sensors that can provide readings about the percentage of dissolved oxygen and PH as indication to the quality of the fresh water in the studied river. The practical measurements of dissolved oxygen in fresh water show the following concentrations

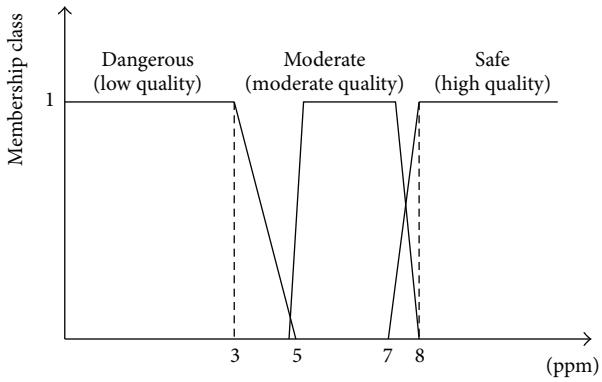



Figure 4: FMFs for dissolved oxygen concentration.

Table 1: Data profiles of sensors.

| ID | Sensor type | Reading level | Sensor status | Decision |
|---|---|---|---|---|
| 1 | DO | High (<3 ppm) | ON | Dangerous |
| 1 | PH | Low (5) | ON | Moderate |
| 2 | DO | No (6–8 ppm) | ON | Optimum |
| 2 | PH | Low (4) | ON | Moderate |
| 3 | DO | High (<3 ppm) | ON | Dangerous |
| 3 | PH | N/A | OFF | Maintenance |
| 4 | DO | High (<3 ppm) | ON | Dangerous |
| 4 | PH | Low (6) | ON | Moderate |
| 5 | DO | High (4 ppm) | ON | Dangerous |
| 5 | PH | No (8) | ON | Optimum |

where the last two ones can refer to a high degradation in the water quality due to pollution [29]:

(a) Safe (~8 parts per million or ppm).

(b) Moderate (5–7 ppm).

(c) Stressful (3–5 ppm).

(d) Dangerous (<3 ppm).

Concerning the concentration level of PH in a fresh water, we adopt the following three levels as indication to having acidic levels [29]:

(a) Low acidic (6–8).

(b) Moderate acidic (4–6).

(c) Highly acidic (<4).

The designed scheme in advanced nodes will adopt *fuzzy* logic that employs *fuzzy* membership functions (FMFs) to classify the readings levels of each sensor. For readings concerning the dissolved oxygen and the discussed concentration levels previously, FMFs are designed, as shown in Figure 4, to consider the following: the first concentration as best quality, the second concentration as moderate quality, and the last two concentrations as low quality. For the concentration level of absorbed lead, other FMFs are used to define the first concentration level as safe level or optimum quality of water and the second concentration level as moderate quality and the last concentration level as the lowest quality of water.

In this example, we assume that we have one final destination, four advanced nodes, and five normal nodes. Also, we have two types of sensors in each sensor node for measuring levels of DO concentration and PH. The data profile of a reading stored in a memory of an advanced sensor node will show the ID of the related sensor node, type of sensor, reading level, and status of the node.

Considering that each region in a river is covered by a set of sensor nodes with specific IDs for monitoring water quality, so each area will have data patterns (based on registered data profiles) that can be learned to know and detect the main pollutants that might exist and the expected consequences and recommend countermeasure for treatment.

In normal operation, we get two readings from each normal sensor node through the day. So, we expect 10 readings per day and 300 readings per month. As shown in Table 1, assume that we have for a region of interest in the river the following daily data profiles collected from five operating sensors through a certain month.

The advanced sensor nodes will learn the patterns of data profiles stored in their memories. Advanced nodes will learn Apriori-based association rule learning algorithm [30] for learning the main frequent attributes in the stored data profiles, which are sensor ID, sensor type, reading level, and sensor status. Additionally, advanced nodes will employ statistical analysis algorithm for generating some statistics about stored data profiles, such as how many profiles contain $O_2$ and high pollution level [31].

From Table 1, advanced nodes can build a weighted decision tree-based on entropy and information gain in order to classify data flows and know the more important flows. Hence, they can allocate more resources and offer more bandwidth.

For building the weighted decision tree, we will have the following procedures. There are 10 data profiles with three main attributes, which are sensor type, reading level, and sensor status. Also, there are main four decisions in the table which are maintenance, no pollution, low pollution, and high pollution. From (1), we can calculate the decision entropy as follows:

$$\begin{aligned} H(D) &= -\sum_{D=1}^{4} p_D \log_2 p_D \\ &= -p_{\text{Dangerous}} \log_2 p_{\text{Dangerous}} \\ &\quad - p_{\text{Moderate}} \log_2 p_{\text{Moderate}} \\ &\quad - p_{\text{Optimum}} \log_2 p_{\text{Optimum}} - p_{\text{Maint}} \log_2 p_{\text{Maint}} \qquad (4) \\ &= -0.4 \log_2 0.4 - 0.3 \log_2 0.3 - 0.2 \log_2 0.2 \\ &\quad - 0.1 \log_2 0.1 \\ &= 0.529 + 0.521 + 0.4644 + 0.3322 \\ &= 1.8466 \text{ bits.} \end{aligned}$$

Then, we will use (2) and (3) to compute the information gain for each interesting attribute of the main three we have.

For the first attribute: sensor type

$$\begin{aligned} I(D;A) &= H(D) - H\left(\frac{D}{\text{Type}}\right) \\ &= 1.8466 - \frac{|O_2|}{10}\text{Entropy}(O_2) \\ &\quad - \frac{|PH|}{10}\text{Entropy}(PH) \\ &= 1.8466 - 0.5\left(-0.8\log_2 0.8 - 0.2\log_2 0.2\right) \\ &\quad - 0.5\left(-0.6\log_2 0.6 - 2\times 0.2\log_2 0.2\right) \\ &= 1.8466 - 0.5\,(0.2575 + 0.4644) \\ &\quad - 0.5\,(0.4422 + 2\times 0.4644) \\ &= 1.8466 - 0.361 - 0.6855 = 0.8\ \text{bits}. \end{aligned} \tag{5}$$

Similarly for the other two attributes,

$$\begin{aligned} I(D;A) &= H(D) - H\left(\frac{D}{\text{Level}}\right) = 1.8466 \\ &\quad - \frac{|\text{High}|}{10}\text{Entropy}(\text{High}) - \frac{|\text{Low}|}{10}\text{Entropy}(\text{Low}) \\ &\quad - \frac{|\text{No}|}{10}\text{Entropy}(\text{No}) - \frac{|\text{NA}|}{10}\text{Entropy}(\text{NA}) \\ &= 1.8466 - 0.4\,(0) - 0.3\,(0) - 0.2\,(0) - 0.1\,(0) \\ &= 1.8466\ \text{bits}. \end{aligned} \tag{6}$$

$$\begin{aligned} I(D;A) &= H(D) - H\left(\frac{D}{\text{Status}}\right) = 1.8466 \\ &\quad - \frac{|\text{ON}|}{10}\text{Entropy}(\text{ON}) - \frac{|\text{OFF}|}{10}\text{Entropy}(\text{OFF}) \\ &= 1.8466 - 0.9\left(-0.444\log_2 0.444 - 0.333\log_2 0.333\right. \\ &\quad \left. - 0.222\log_2 0.222\right) - 0.1\,(0) = 1.8466 - 0.9\,(0.52 \\ &\quad + 0.5283 + 0.482) = 1.8466 - 0.9\,(0.566256) \\ &= 1.337\ \text{bits}. \end{aligned}$$

From the above results, we find that the higher information gain is obtained for the reading level attribute (1.8466 bits) and then for the sensor status (1.337 bits) and then for the sensor type (0.8 bits). This means that the advanced powerful sensor nodes will forward data to the final destination management units once they detect high reading level from operating sensor nodes without giving more weight to the sensor type or the main cause for the low water quality levels. Accordingly, a weighted decision tree can be formed and adopted by advanced nodes to take decision and forward data to the final destination. A set of rules can be generated according to the built decision tree such as the following: *if* there is a high level reading from an operating DO or PH sensor, *then* forward the data to the destination.

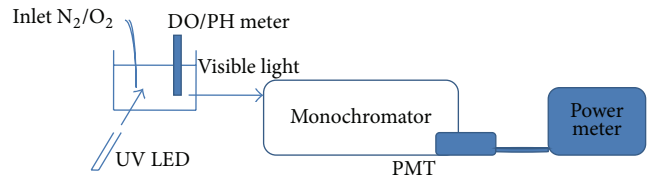



Figure 5: Experimental setup of optical sensing for dissolved oxygen.

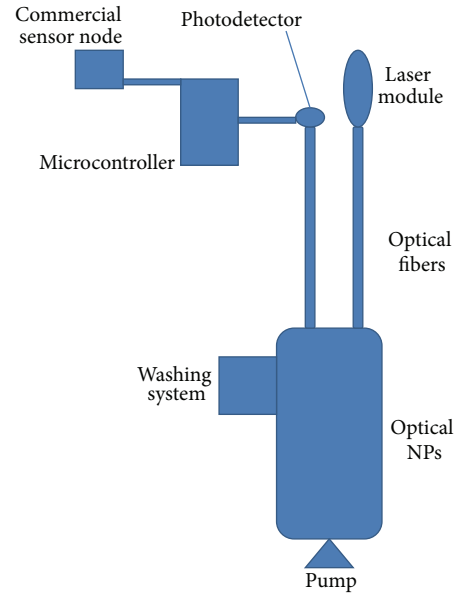



Figure 6: Suggested prototype design of the compact sensor.

*2.2. Optical Materials' Synthesis and Sensing Setup.* Ceria nanoparticles are prepared using a chemical precipitation technique for simplicity and relative low-cost precursors [32–34]. 0.5 g of cerium (III) chloride (heptahydrate, 99.9%, Aldrich Chemicals) is added to 40 mL deionized water as a solvent. The solution is stirred at rate of 500 rpm for 2 hours at 50°C with added 1.6 mL of ammonia. Then, the solution is stirred for 22 hours at room temperature. The synthesized ceria nanoparticles are characterized using UV-Vis spectroscopy (dual beam PG 90+) to detect the absorbance dispersion and consequent bandgap calculation.

The experimental test bed used to detect the change of the fluorescence intensity peak due to the change of DO concentration and PH value in the aqueous solution is shown in Figures 5 and 6. The fluorescence spectroscopy system consists of a near-UV LED of central wavelength of 430 nm as an excitation source exposed to a three-neck flask containing DI water solution of the synthesized ceria nanoparticles. Oxygen and nitrogen gases are fed through individual lines through double-holes cork placed into one of the necks on

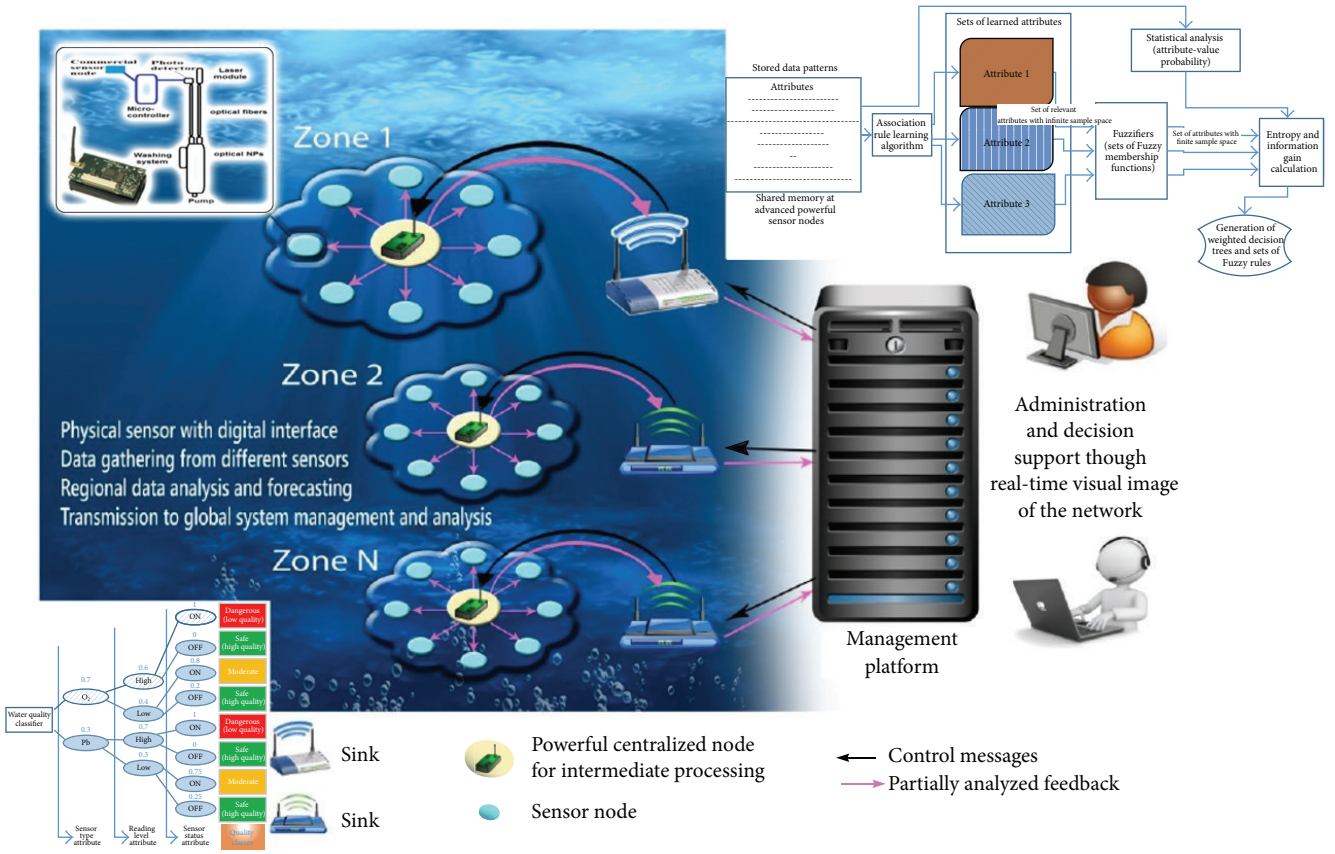



Figure 7: The sensor operation management framework.

the flask and controlled by a mass flow rate controller. The probe of a commercial DO meter (Thermo Scientific A500 with a measurement range up to 50 mg/L) is inserted in the second neck of the flask to measure DO concentration. The fluorescence signal is collected from the colloidal solution scanned by a Newport Cornerstone 1300 monochromator, positioned at a 90° angle to the excitation signal for minimum scattering effects. Then, a photomultiplier tube (Newport PMT 77340) is connected to a power meter (Newport Power meter 1915-R) for fluorescence intensity monitoring. Same setup is used for PH detection at normal DO concentration, by removing both oxygen and nitrogen inlets with varying added concentrations of acid (HCL). The value of PH is measured using same meter (Thermo Scientific A500) but with the optimum probe for PH detection.

*2.3. Autonomous Sensor Operation Management System.* Figure 7 is a block diagram of the sensor framework where the main sensing element is interfaced with an autonomously managed wireless sensor network for water quality monitoring. The sensing function is interconnected to the cyber layer for control, management, and monitoring.

The nanosensor is interfaced with an A2D chip, a powerful microcontroller chip, a GPS chip, and wireless transmission module. The microcontroller is programed to control the activation and deactivation of the sensing element and control the measurement configuration if necessary. The microcontroller receives its guideline from a remote management server. Each sensor has a unique identifier that is used in all transmissions along with the geographical location of the sensor in case of mobility. The system is built to be as generic as possible allowing more sensing elements to be attached to the same sensing elements if the application requires that.

The system is built to scale, where sensors are grouped into different zones. The sensor feedback and location are dispatched frequently based upon either event change and query or a predetermined schedule to a central data collection node at each zone. This node applies partial analysis and data grouping and dispatches a comprehensive zone status-report to the management server for further analysis and guidance. The algorithms used to establish such analysis are application dependent. We devised a simple case study with a simple model for the excremental study just to reflect the effect of such automation on the quality of the system output.

## 3. Results and Discussion

*3.1. Evaluation of HI-Based Data Classification and Forwarding Scheme.* In this subsection, we conduct a simulation

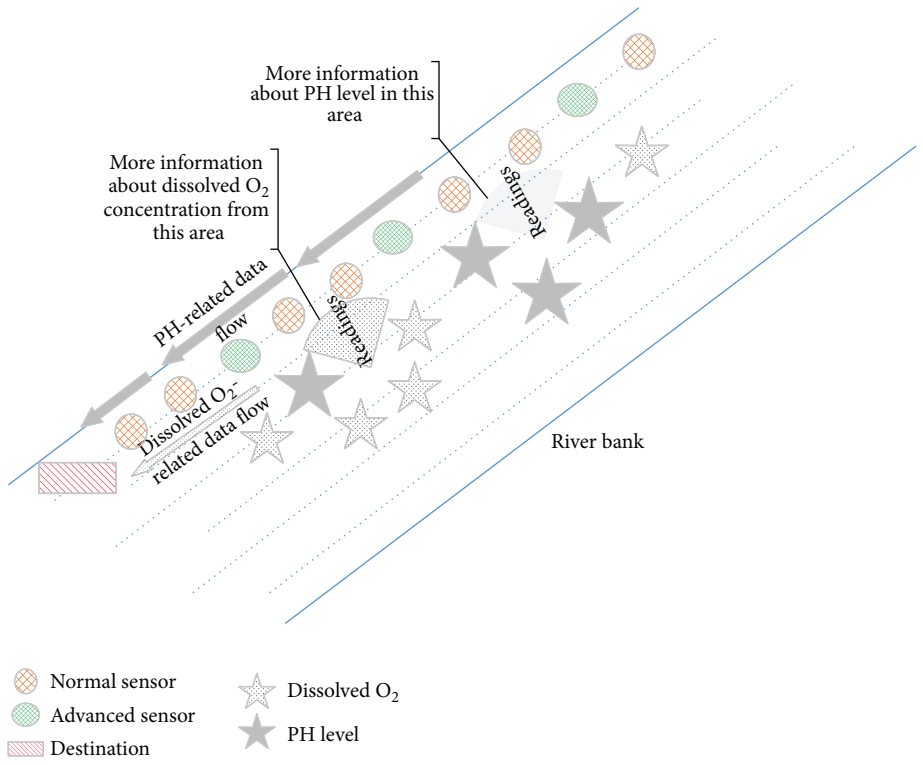



Figure 8: Simulation scenario of the proposed HI-based scheme for data classification.

scenario for evaluating the proposed HI-based scheme of data classification and forwarded data adopted by the powerful advanced nodes in the DSFS subsystem.

*3.1.1. Simulation Setup.* In this section, we provide a preliminary simulation study for the proposed classification and forwarding scheme. Using the discussed example in Section 2.1.2, we conducted a simple scenario of a WSN-based pollution monitoring and data forwarding network using Java based discrete time event-driven WSN simulator, called J-Sim [35]. We target a river as a network context where sensors will be distributed along a river bank. Figure 8 shows a layout of the scenario which depicts the network topology and communication amongst sensor nodes. We assume a linear WSN where sensors forward data to the next available sensors. We have two different types of sensors which are the normal node and the advanced node. The last type can access registered data patterns and decide whether the forwarded data are of a certain importance or not. According to the calculated information entropy and gain besides the constructed weighted binary decision trees, the advanced nodes will forward data to the final network destination. We assume that we have two types of pollutants that will result in various levels of pollution indicators which depend on measuring the concentrations of dissolved oxygen (DO) and PH levels in regions of interest in the river. In case of having low measurements of DO and low levels of PH, this means that we have indication of a combined low water quality level. On the other hand, if there are measurements' levels, from one sensor type, refering to a certain low water quality indication, this means that we have an explicit low water quality level. Table 2 shows the simulation parameters. Table 3 shows a set of key metrics which are used to study the offered QoS and network performance.

Figure 9 shows that advanced nodes allocate more resources and bandwidth for data flows concerning high level readings from dissolved $O_2$ sensors. The figure shows that, at early time, low water quality levels are detected and then huge amount of data are allowed to flow from normal sensor nodes to the destination. Also, some advanced nodes find high water quality levels based on measured dissolved $O_2$ concentrations. Once the water quality-related data patterns reach a certain information gain, advanced nodes allocate more resources and forward the related data to the final destination. So, in case of optimum water quality levels, less traffic is found since small information gain is calculated (i.e., we have in this case normal operations with low information entropy).

Figure 10 shows that there are higher data throughputs measured at the final destination in case of detecting low water quality levels in case of having one source (only dissolved $O_2$ concentrations) or with two sources (dissolved $O_2$ concentrations and PH levels). This is because more data packets will be forwarded amongst sensor nodes till reaching the destination.

Figure 11 depicts the response of the WSN to a low water quality level that exists widely in the studied context.

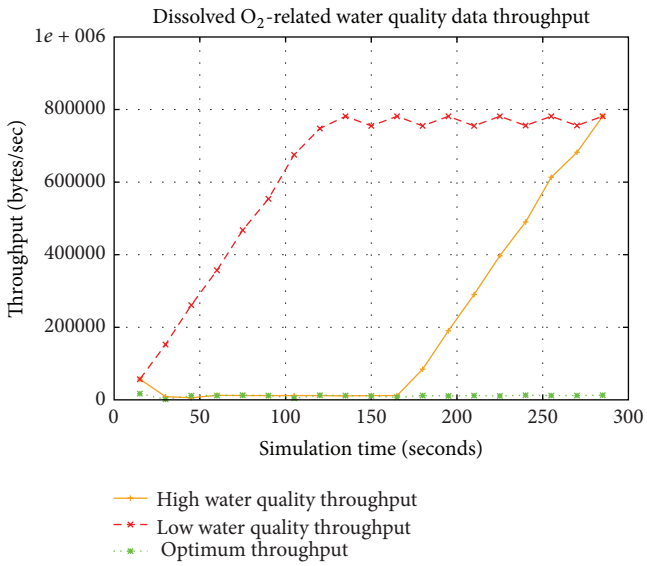



Figure 9: Data throughput based on detected dissolved oxygen-related water quality.

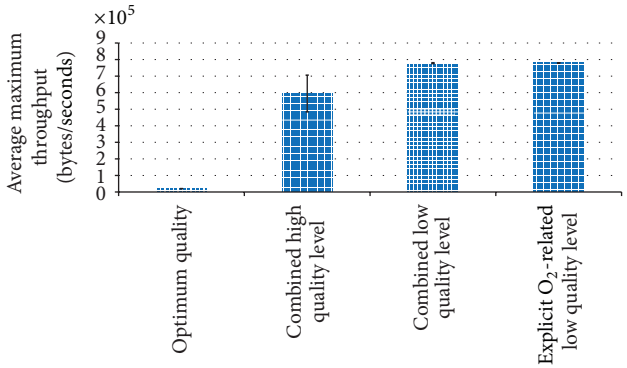



Figure 10: Average maximum throughput measured at running various simulation scenarios.

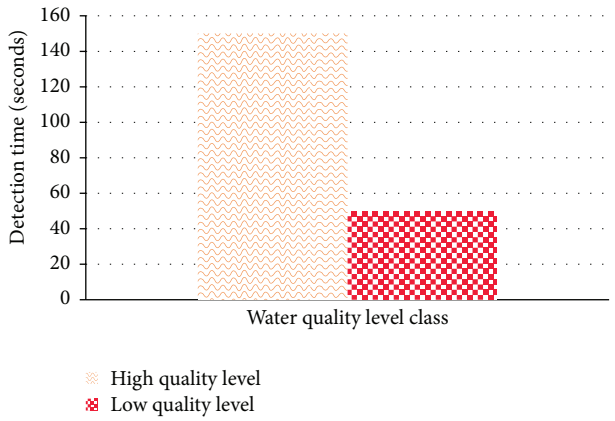



Figure 11: Detection time of dissolved oxygen-related water quality level.

Table 2: Simulation parameters.

| Simulation parameter | Value |
|---|---|
| Number of sensor nodes | 11 |
| Number of normal sensor nodes | 6 |
| Number of advanced nodes | 4 |
| Number of final destinations | 1 |
| Initial node power | 1000 watts |
| Packet size | 512 bytes |
| Packet time to live (TTL) | 255 seconds |
| Communication bandwidth | Variable {200, 1000, $10^4$, $10^6$} bytes/sec |
| Types of sensors | Two {dissolved $O_2$, PH} |
| Dissolved $O_2$ concentrations | Low, high |
| PH levels | Low, high |
| Number of defined low water quality classes | Three {low, high, no} |
| Status of operating sensors | ON, OFF |
| Expected number of data patterns | 8 |
| Pollution threshold | >50% of stored patterns |
| Patterns analysis and reasoning frequency | Every 15 seconds |
| Number of generated rules | 3 |
| Simulation time | 300 seconds |

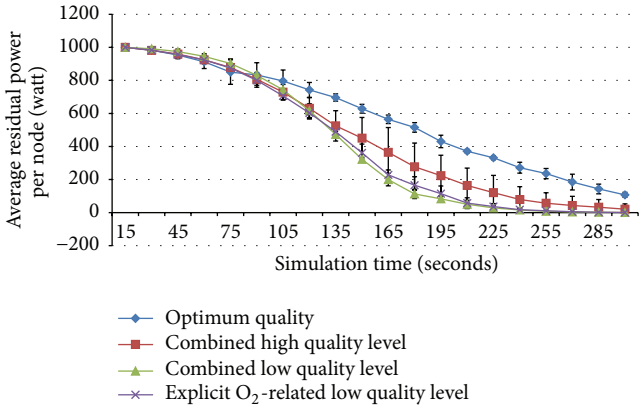



Figure 12: Average residual power per sensor node versus the simulation time.

Many advanced nodes are able to detect low dissolved $O_2$ concentrations at the studied context and, hence, they allow data flows to pass to the final destination.

We studied the average residual power at the first dead sensor node. As shown in Figure 12, the degradation in power increases as there are readings related to low water quality levels. Hence, more data packets will be forwarded to the final destination. Also, the figure shows that combined low water quality levels, which comprise readings related to low quality indicators based on the measured concentrations of dissolved $O_2$ and PH levels, result in more degradation in

Table 3: QoS and network performance metrics.

| Metric | Unit | Description |
|---|---|---|
| Received data throughput | Bytes/seconds | The amount of data received by the final destination |
| Average maximum throughput | Bytes/seconds | The average of maximum throughput captured at the final destination when running the same scenario for more than one time |
| Water quality level detection time | Seconds | Delay introduced by the time needed to measure amounts of data readings for determining a specific water quality level |
| Average residual power per sensor | Watt | The average amount of measured residual power at sensors after running the simulation scenario and forwarding data |

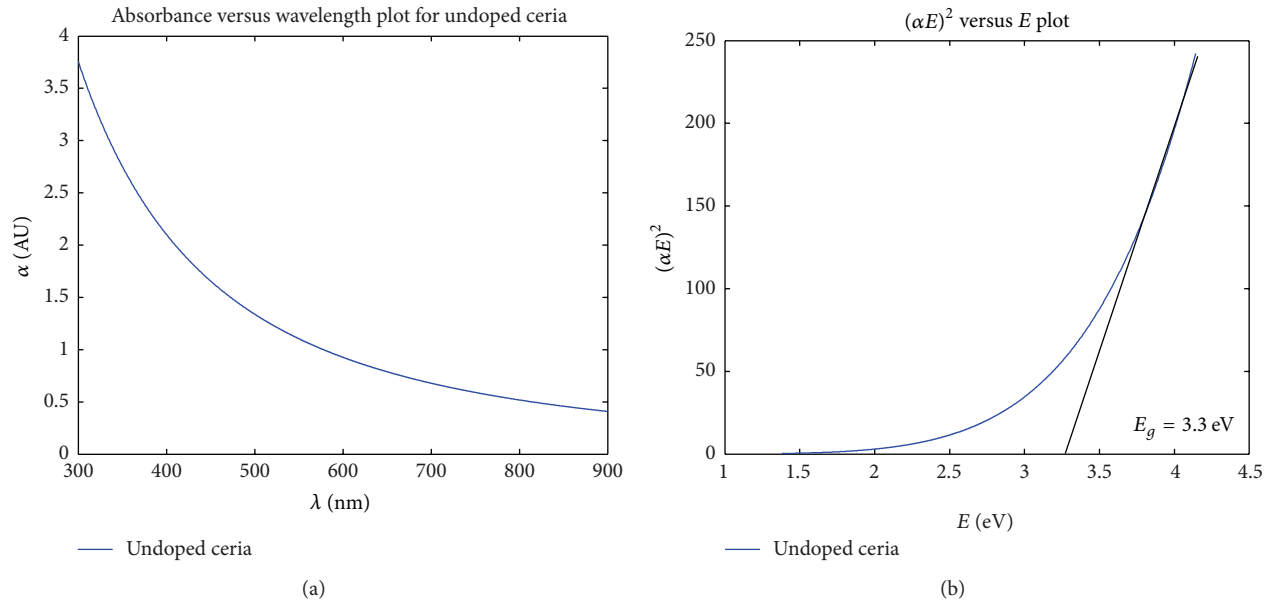



Figure 13: (a) Absorbance dispersion for the synthesized ceria nanoparticles and (b) direct allowed bandgap calculation.

the residual power level compared with the case of explicit low water quality level in case of having many measurements of low $O_2$ levels.

*3.2. Optical Nanoparticles Characterization.* Figure 13(a) shows the resulting absorbance dispersion of the synthesized nanoparticles. Based on the resulting absorbance measurements, the allowed direct bandgap semiconductor of the synthesized nanoparticles can be found through the following equation [36]:

$$\alpha E = A^* \left(E - E_g\right)^{1/2}, \tag{7}$$

where $A^*$ is a constant for the given material depending on its refractive index and effective masses of both electrons and holes, $E$ is the absorbed optical energy, and $E_g$ is the direct allowed bandgap energy. Then, $(\alpha E)^2$ is plotted with the absorbed optical energy, and the intersection with $x$-axis gives the value of bandgap energy as shown in Figure 13(b).

Figure 14 shows the change of the visible fluorescence emission intensity at 520 nm from the ceria nanoparticles with increasing DO concentration at normal PH, ~7, under

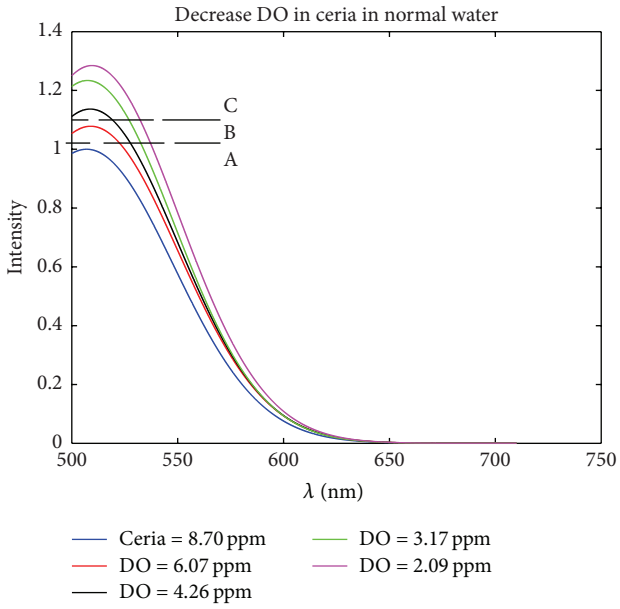



Figure 14: Visible fluorescence spectra at different DO concentration within neutral PH (~7).

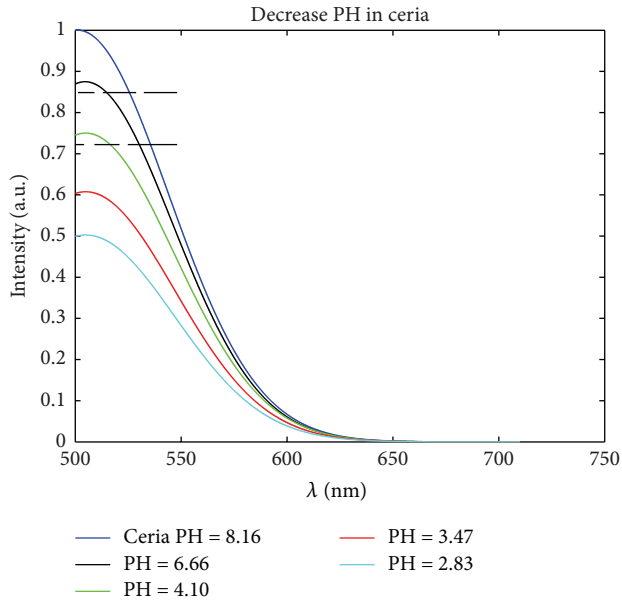



Figure 15: Visible fluorescence spectra at different PH levels due to increasing acid concentration within neutral DO (~8 ppm).

near-UV excitation. We speculate that this is due to a release of oxygen stored in the ceria lattice when the nanoparticles are introduced into the solution. Also, the emitted fluorescence emission is found to be reduced with increasing the added acidic concentration which consequently reduces the value of PH as shown in Figure 15, at normal DO concentration level (~8 ppm).

## 4. Conclusions

This work presented a smart network management system for efficient data classification and forwarding and decision making. The system comprised two main subsystems, a data sensing and forwarding subsystem (DSFS) and Operation Management Subsystem (OMS). The DSFS adopted a novel hybrid intelligence (HI) scheme for data classification and forwarding in wireless sensor networks integrating information theory concepts, machine learning, *fuzzy* logic, and weighted decision tress. Such adoption led to better energy consumption, improved resource utilization, and optimized QoS operation. Simulation scenarios discussed the performance of a WSN employing the proposed scheme for monitoring water quality indicators. Simulation results of the proposed HI-based data classification and forwarding scheme showed that the DSFS works efficiently with low time overhead and success in achieving small levels of power consumption for operating sensors. Additionally, we implemented the presented approach constructing a small test bed employing nano-enhanced sensing elements. Analyzing such elements showed a clear change in the visible fluorescence intensity with the variation of DO ratio. The reason is the developed oxygen vacancies concentration formed inside ceria nanoparticles which act as DO receptor. The entire system operation is managed by an automated management system. Results showed the clear effect of the smart, trustworthy, and automated data collection and analysis on the quality and accuracy of sensing. Having such system is essential for easy data consolidation and information extraction. Different, experiments were conducted to show the effect of having such automation in enhancing the sensor and the network lifetime even in presence of malicious nodes.

The proposed system is an adequate solution for a comprehensive automated management of DO sensors in the aqueous media. The automation platform within the OMS is built to be generic and can be easily modified to be used with wide variety of other applications and sensing elements achieving wide scope of applications. Our future work includes testing the proposed scheme experimentally for other applications and in large-scale scenarios.

## Competing Interests

The authors declare that they have no competing interests.

## Acknowledgments

The presented work is part of the awarded Grant "ARP2013 .R13.2" funded by Information Technology Industry Development Agency (ITIDA Egypt).

## References

[1] P. R. Warburton, R. S. Sawtelle, A. Watson, and A. Q. Wang, "Failure prediction for a galvanic oxygen sensor," *Sensors and Actuators, B: Chemical*, vol. 72, no. 3, pp. 197–203, 2001.

[2] M. A. Acosta, P. Ymele-Leki, Y. V. Kostov, and J. B. Leach, "Fluorescent microparticles for sensing cell microenvironment oxygen levels within 3D scaffolds," *Biomaterials*, vol. 30, no. 17, pp. 3068–3074, 2009.

[3] A. Mohyeldin, T. Garzón-Muvdi, and A. Quiñones-Hinojosa, "Oxygen in stem cell biology: a critical component of the stem cell niche," *Cell Stem Cell*, vol. 7, no. 2, pp. 150–161, 2010.

[4] C.-S. Chu and Y.-L. Lo, "Optical fiber dissolved oxygen sensor based on Pt(II) complex and core-shell silica nanoparticles incorporated with sol-gel matrix," *Sensors and Actuators, B: Chemical*, vol. 151, no. 1, pp. 83–89, 2010.

[5] W. C. Maskell, "Inorganic solid state chemically sensitive devices: electrochemical oxygen gas sensors," *Journal of Physics E: Scientific Instruments*, vol. 20, no. 10, pp. 1156–1168, 1987.

[6] R. Sanghavi, M. Nandasiri, S. Kuchibhatla et al., "Thickness dependency of thin-film samaria-doped ceria for oxygen sensing," *IEEE Sensors Journal*, vol. 11, no. 1, pp. 217–224, 2011.

[7] X.-D. Wang and O. S. Wolfbeis, "Optical methods for sensing and imaging oxygen: materials, spectroscopies and applications," *Chemical Society Reviews*, vol. 43, no. 10, pp. 3666–3761, 2014.

[8] L. Chen, S. Xu, and J. Li, "Recent advances in molecular imprinting technology: current status, challenges and highlighted applications," *Chemical Society Reviews*, vol. 40, no. 5, pp. 2922–2942, 2011.

[9] G. Mistlberger and I. Klimant, "Luminescent magnetic particles: structures, syntheses, multimodal imaging, and analytical

applications," *Bioanalytical Reviews*, vol. 2, no. 1, pp. 61–101, 2010.

[10] N. Shehata, K. Meehan, and D. E. Leber, "Fluorescence quenching in ceria nanoparticles: dissolved oxygen molecular probe with relatively temperature insensitive Stern-Volmer constant up to 50∘C," *Journal of Nanophotonics*, vol. 6, no. 1, Article ID 063529, 2012.

[11] N. Shehata, K. Meehan, M. Hudait, N. Jain, and S. Gaballah, "Study of optical and structural characteristics of ceria nanoparticles doped with negative and positive association lanthanide elements," *Journal of Nanomaterials*, vol. 2014, Article ID 401498, 7 pages, 2014.

[12] R. Ramamoorthy, P. K. Dutta, and S. A. Akbar, "Oxygen sensors: materials, methods, designs and applications," *Journal of Materials Science*, vol. 38, no. 21, pp. 4271–4282, 2003.

[13] A. Oczkowski and S. Nixon, "Increasing nutrient concentrations and the rise and fall of a coastal fishery; a review of data from the Nile Delta, Egypt," *Estuarine, Coastal and Shelf Science*, vol. 77, no. 3, pp. 309–319, 2008.

[14] M. Azab and M. Eltoweissy, "Bio-inspired evolutionary sensory system for cyber-physical system defense," in *Proceedings of the 2012 12th IEEE International Conference on Technologies for Homeland Security (HST '12)*, pp. 79–86, IEEE, Waltham, Mass, USA, November 2012.

[15] C. Hill and K. Sippel, "Modern deformation monitoring: a multi sensor approach," in *Proceedings of the 22nd FIG International Congress*, Washington, DC, USA, April 2002.

[16] E. A. Garich, *Wireless, automated monitoring for potential landslide hazards [M.S. thesis]*, Texas A& M University, 2007.

[17] B. Mokhtar and M. Eltoweissy, "Hybrid intelligence for semantics-enhanced networking operations," in *Proceedings of the 27th International Florida Artificial Intelligence Research Society Conference (FLAIRS '14)*, pp. 449–454, Pensacola Beach, Fla, USA, May 2014.

[18] B. Mokhtar and M. Eltwoeissy, "Hybrid intelligence for smarter networking operations," in *Handbook of Research on Advanced Hybrid Intelligent Techniques and Applications*, S. Bhattacharyya, P. Banerjee, D. Majumdar, and P. Dutta, Eds., IGI Global, 2015.

[19] B. Mokhtar and M. Eltoweissy, "Towards a data semantics management system for internet traffic," in *Proceedings of the 6th International Conference on New Technologies, Mobility and Security (NTMS '14)*, pp. 1–5, IEEE, Dubai, UAE, April 2014.

[20] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Computer Networks*, vol. 52, no. 12, pp. 2292–2330, 2008.

[21] J. Luo, J. Hu, D. Wu, and R. Li, "Opportunistic routing algorithm for relay node selection in wireless sensor networks," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 1, pp. 112–121, 2015.

[22] L. Pelusi, A. Passarella, and M. Conti, "Opportunistic networking: data forwarding in disconnected mobile ad hoc networks," *IEEE Communications Magazine*, vol. 44, no. 11, pp. 134–141, 2006.

[23] R. Kumar and N. Singh, "A survey on data aggregation and clustering schemes in underwater sensor networks," *International Journal of Grid and Distributed Computing*, vol. 7, no. 6, pp. 29–52, 2014.

[24] M. Di Francesco, S. K. Das, and G. Anastasi, "Data collection in wireless sensor networks with mobile elements: a survey," *ACM Transactions on Sensor Networks (TOSN)*, vol. 8, no. 1, article 7, 2011.

[25] K. Seada, M. Zuniga, A. Helmy, and B. Krishnamachari, "Energy-efficient forwarding strategies for geographic routing in lossy wireless sensor networks," in *Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems (SenSys '04)*, pp. 108–121, Baltimore, Md, USA, November 2004.

[26] J. J. Buckley and E. Eslami, *An Introduction to Fuzzy Logic and Fuzzy Sets*, Springer Science & Business Media, 2002.

[27] S. K. Debray, S. Kannan, and M. Paithane, "Weighted decision trees," in *Proceedings of the Joint International Conference and Symposium on Logic Programming (JICSLP '92)*, pp. 654–668, Washington, DC, USA, November 1992.

[28] K. Matiaško, J. Bohacik, V. Levashenko, and Š. Kovalík, "Learning fuzzy rules from fuzzy decision trees," *Journal of Information, Control and Management Systems*, vol. 4, no. 2, pp. 143–154, 2006.

[29] C. A. Atkinson, D. F. Jolley, and S. L. Simpson, "Effect of overlying water pH, dissolved oxygen, salinity and sediment disturbances on metal release and sequestration from metal contaminated marine sediments," *Chemosphere*, vol. 69, no. 9, pp. 1428–1437, 2007.

[30] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," *ACM SIGMOD Record*, vol. 22, no. 2, pp. 207–216, 1993.

[31] N. Vasconcelos and A. Lippman, "Statistical models of video structure for content analysis and characterization," *IEEE Transactions on Image Processing*, vol. 9, no. 1, pp. 3–19, 2000.

[32] H.-I. Chen and H.-Y. Chang, "Homogeneous precipitation of cerium dioxide nanoparticles in alcohol/water mixed solvents," *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, vol. 242, no. 1–3, pp. 61–69, 2004.

[33] N. Shehata, K. Meehan, I. Hassounah et al., "Reduced erbium-doped ceria nanoparticles: one nano-host applicable for simultaneous optical down- and up-conversions," *Nanoscale Research Letters*, vol. 9, no. 1, article 231, pp. 1–6, 2014.

[34] N. Shehata, K. Meehan, and D. Leber, "Study of fluorescence quenching in aluminum-doped ceria nanoparticles: potential molecular probe for dissolved oxygen," *Journal of Fluorescence*, vol. 23, no. 3, pp. 527–532, 2013.

[35] A. Sobeih, J. C. Hou, L.-C. Kung et al., "J-Sim: a simulation and emulation environment for wireless sensor networks," *IEEE Wireless Communications*, vol. 13, no. 4, pp. 104–119, 2006.

[36] J. Pankove, *Optical Processes in Semiconductors*, Dover Publications, New York, NY, USA, 1971.

Hindawi Publishing Corporation
Journal of Sensors
Volume 2016, Article ID 1865614, 10 pages
http://dx.doi.org/10.1155/2016/1865614

*Research Article*

# Air Pollution Monitoring and Control System for Subway Stations Using Environmental Sensors

**Gyu-Sik Kim,[1] Youn-Suk Son,[2] Jai-Hyo Lee,[3] In-Won Kim,[4] Jo-Chun Kim,[5] Joon-Tae Oh,[1,6] and Hiesik Kim[1]**

[1]*Department of Electrical and Computer Engineering, The University of Seoul, Seoul 130-743, Republic of Korea*
[2]*Research Division for Industry & Environment, Korea Atomic Energy Research Institute, Jeongeup 580-185, Republic of Korea*
[3]*Department of Mechanical Engineering, Konkuk University, Seoul 143-701, Republic of Korea*
[4]*Department of Chemical Engineering, Konkuk University, Seoul 143-701, Republic of Korea*
[5]*Department of Environmental Engineering, Konkuk University, Seoul 143-701, Republic of Korea*
[6]*Seoul Metropolitan Rapid Transit Co. Ltd., Seoul 133-851, Republic of Korea*

Correspondence should be addressed to Gyu-Sik Kim; gskim318@uos.ac.kr

Received 17 May 2016; Accepted 27 June 2016

Academic Editor: José P. Santos



The metropolitan city of Seoul uses more energy than any other area in South Korea due to its high population density. It also has high emissions of air pollutants. Since an individual usually spends most of his/her working hours indoors, the ambient air quality refers to indoor air quality. In particular, $PM_{10}$ concentration in the underground areas should be monitored to preserve the health of commuters in the subway system. Seoul Metro and Seoul Metropolitan Rapid Transit Corporation measure several air pollutants regularly. In this study, the accuracy of an instrument for PM measurement using the light scattering method was improved with the help of a linear regression analysis technique to continuously measure the $PM_{10}$ concentrations in subway stations. In addition, an air quality monitoring system based on environmental sensors was implemented to display and record the data of $PM_{10}$, $CO_2$, temperature, and humidity. Through experimental studies, we found that ventilation fans could improve air quality and decrease $PM_{10}$ concentrations in the tunnels effectively by increasing the air flow rate.

## 1. Introduction

People spend most of their time indoors—either at home, in the workplace, or in transit. Thus, there has been an increasing concern over indoor air quality (IAQ) and its effects on public health. The US Environment Protection Agency (EPA) reported that, in the US, the mean daily residential time spent indoors was 21 h, while the GerES II reported that this duration was 20 h in Germany. Thus, the IAQ has been recognized as a significant factor in the determination of the health and welfare of people [1]. The Korea Ministry of Environment (KMOE) enforced the IAQ act to control five major pollutants, including $PM_{10}$, $CO_2$, CO, VOCs, and formaldehyde in indoor environments. Out of these, the IAQ standard for $PM_{10}$ concentration is 150 $\mu g/m^3$. The IAQ is critical not only in buildings but also in underground areas and public transportation systems. Much effort has been made for the improvement of the IAQ in subway stations [2–5].

Among the various types of indoor environments, underground subway stations have especially unique features. The confined space occupied by the underground subway system can accumulate the pollutants entering from the outside in addition to those generated within the system. Therefore, it is likely that the subway system in the Seoul metropolitan area contains different types of hazardous pollutants due to the old ventilation and accessory systems [3, 6].

Recently, Platform Screen Doors (PSDs) were installed and are being used in many subway stations in Korea to prevent the diffusion of air pollutants into the subway stations and ensure the safety of the public. Some previous studies reported that the PM concentration in subway stations significantly reduced after the installation of PSDs [7, 8]. However,

they suggested that the PM concentration in the tunnels would be much higher due to the interruption of particle diffusion into the subway stations by the PSDs. Moreover, most of the ventilation fans may not be working properly because of their deterioration and the high operational cost. Therefore, the PM concentrations in the tunnels have probably been high for a long period of time.

The IAQ in the subway stations can be affected by many factors, such as the number of passengers, the outside conditions, and the natural ventilation rate [9, 10]. The management and monitoring of IAQ in subway stations have become an important issue of public interest [11–13]. Some environmental sensors are important for monitoring IAQ in subway systems and they provide the data needed for continuous online implementation. Sometimes, these sensors in subway stations suffer from poor quality of data and the unreliability of the sensor due to the highly deteriorated and polluted environment, which may cause the measuring instruments installed for monitoring to malfunction. The quality of the online measurement can determine the failure or success of IAQ monitoring and assessment. Although most researchers and practitioners agree with this opinion, very little attention has been given to the study of sensors in a realistic manner [14, 15].

In this study, the accuracy of the instrument for PM measurement using light scattering method was improved with the help of a linear regression analysis technique to continuously measure the $PM_{10}$ concentrations in the subway stations [16]. In addition, the air quality monitoring system based on environmental sensors was implemented to display and record the data on $PM_{10}$, $CO_2$, temperature, and humidity [17, 18]. Finally, for underground subway stations with natural ventilation, some ventilation fans were installed in order to improve the air quality in the tunnels. Through experimental studies, we found that the ventilation fans could improve the air quality in the tunnels and decrease the $PM_{10}$ concentration in the tunnels effectively by increasing their air flow rate.

## 2. Experimental Methods

*2.1. PM Measuring Instruments.* Particulate matter with an aerodynamic diameter less than 10 $\mu$m ($PM_{10}$) is one of the major pollutants in subway environments. The $PM_{10}$ concentration in the underground areas should be monitored to protect the health of the commuters in the underground subway system. Seoul Metro and Seoul Metropolitan Rapid Transit Corporation measure several air pollutants regularly. As for the $PM_{10}$ concentration, generally, measuring instruments based on $\beta$-ray absorption method are used. In order to keep the $PM_{10}$ concentration below a healthy limit, the air quality in the underground platform and tunnels should be monitored and controlled continuously. The $PM_{10}$ instruments using light scattering method can measure the $PM_{10}$ concentration once every several seconds. However, the accuracy of the instruments using light scattering method has still not been proven since they measure the particle number concentration rather than the mass concentration [19].

Table 1: Specifications of PM measuring instruments.

| Instrument | Ebam | Esampler | HCT-PM326 |
|---|---|---|---|
| Method | $\beta$-ray absorption | Light scattering | Light scattering |
| Range | 0~100 mg/m$^3$ | 0~65 mg/m$^3$ | 0~1 mg/m$^3$ |
| Sampling flow rate | 16.7 L/min | 2 L/min | 0.8 L/min |
| Sampling period | 60 min | 1 sec | 6 sec |

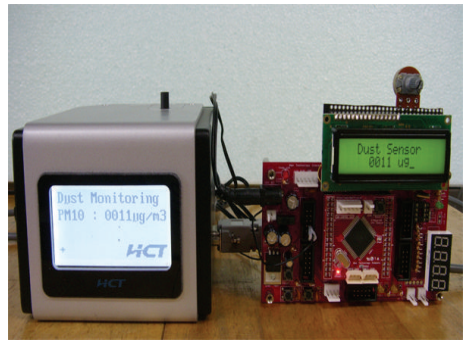



Figure 1: The PM measuring instrument HCT-PM326 connected to ATmega128(L) CPU board.

The purpose of this work is to study the accuracy improvement of the instruments which use light scattering method to continuously measure the $PM_{10}$ concentrations in the underground subway stations. One instrument using $\beta$-ray absorption method Ebam (Met one instrument, USA) and two different instruments using light scattering method, that is, Esampler (Met one instrument, USA) and HCT-PM326 (HCT, Korea), were installed on the platform at "Jegi" subway station of Seoul Metro line number 1 in order to evaluate their dynamic performances. The specifications of the PM measuring instruments are listed in Table 1. Figure 1 shows the PM measuring instrument HCT-PM326 connected to the ATmega128(L) CPU board, which is used to display the measured data and transfer them to an m2m wireless modem.

*2.2. $CO_2$ Sensor.* Of late, the monitoring of carbon dioxide ($CO_2$) has been considered very important for passengers and employees in underground subway stations due to the health risks associated with carbon dioxide exposure. For instance, headache, sweating, dim vision, tremors, and loss of consciousness are caused by exposure to high $CO_2$ concentration for a long time.

$CO_2$ gas sensors being used presently can be divided into two types. The first one is NDIR (Nondispersive Infrared) method and the second one is a chemical method. They are commonly available for monitoring $CO_2$ concentrations indoors. However, chemical $CO_2$ sensors have many limitations that prevent their application to a variety of practical fields. The obvious drawbacks of chemical $CO_2$ sensors are short-term stability and low durability. This is because chemical sensors are easily deteriorated by heterogeneous gases and minute particles in the ambient polluted air. On the

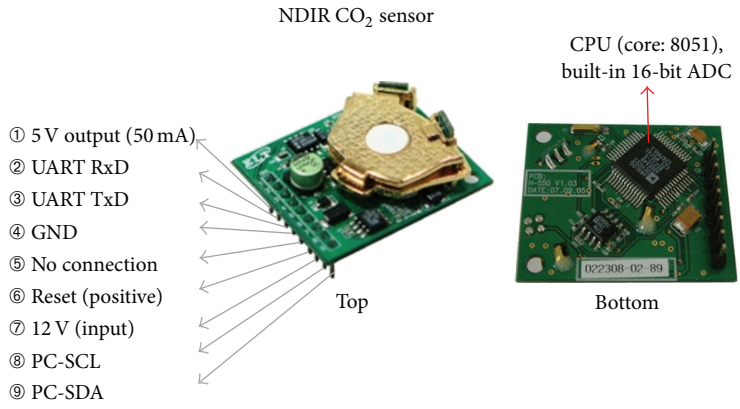



Figure 2: Pin assignment of the NDIR $CO_2$ sensor H-550 (ELT, Korea).

Table 2: Specifications of the NDIR $CO_2$ sensor H-550.

| | |
|---|---|
| Range | 0~50,000 ppm |
| Sensitivity | ±20 ppm ± 1% |
| Accuracy | ±30 ppm ± 5% |
| Response time (90%) | Within 30 sec |

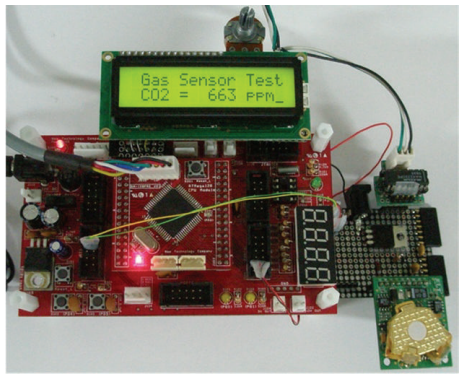



Figure 3: The NDIR $CO_2$ sensor H-550 connected to ATmega128(L) CPU board.

other hand, since the NDIR method uses the physical sensing principle, such as gas absorption at a particular wavelength [20], NDIR sensors are more advanced in terms of long-term stability and accuracy during $CO_2$ measurement. Hence, NDIR $CO_2$ sensors are the most widely applied for real-time monitoring of $CO_2$ concentration [21]. The pin assignment of the NDIR $CO_2$ sensor H-550 manufactured by ELT Co. Ltd., Korea, and its connection to ATmega128(L) CPU board are shown in Figures 2 and 3, respectively. The specifications of the NDIR $CO_2$ sensor H-550 are listed in Table 2.

*2.3. Temperature and Humidity Sensor.* Seoul Metropolitan Rapid Transit Corporation measures temperature and humidity as well as $PM_{10}$ and $CO_2$ in the waiting rooms of the subway stations regularly. Temperature and humidity can also be used for the analysis and prediction of IAQ in a subway station [22–24]. As for temperature and humidity sensors, the

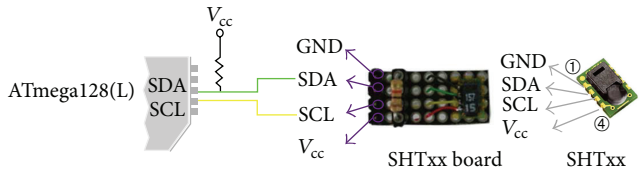



Figure 4: Pin assignment of the temperature and humidity sensor SHT11.

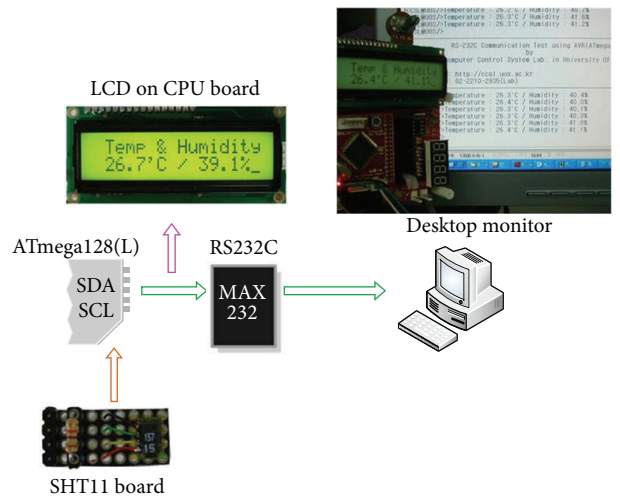



Figure 5: The temperature and humidity sensor SHT11 connected to the ATmega128(L) CPU board.

SHT11 manufactured by Sensirion was chosen in this study. It is a single chip relative humidity and temperature multisensor module, comprising a calibrated digital output. It is coupled to a 14-bit analog to digital converter and the 2-wire serial interface and internal voltage regulation, which allow easy and fast system integration. The pin assignment of SHT11 is shown in Figure 4. Figure 5 shows that the SHT11 is connected to ATmega128(L) CPU board, which transmits temperature

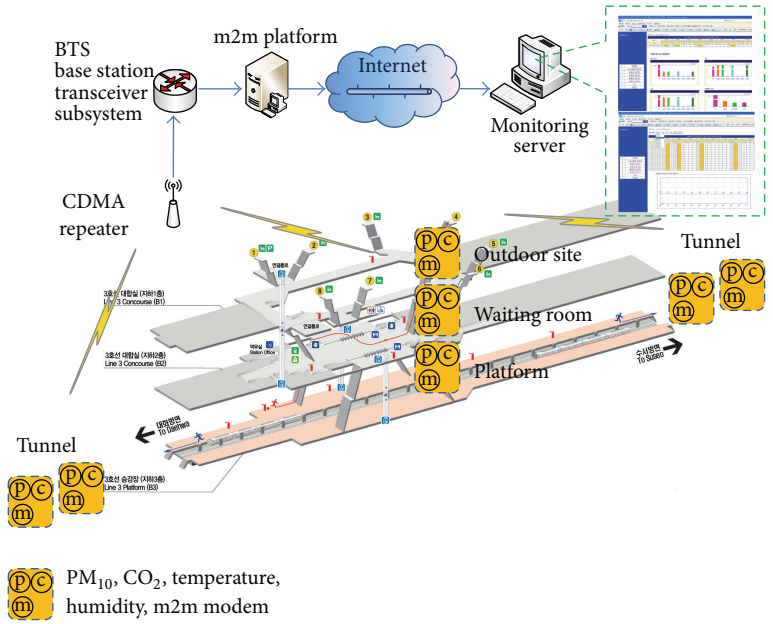



Figure 6: Air quality monitoring system for subway stations.

Table 3: Specifications of the temperature and humidity sensor SHT11.

| | |
|---|---|
| Temperature range/accuracy | −40~123.8°C/±0.4°C |
| Humidity range/accuracy | 0~100%/±3.0% |

and humidity data to the desktop PC using RS232C interface. The measuring range and accuracy of SHT11 are given in Table 3.

*2.4. Air Quality Monitoring System in a Subway Station.* This paper presents the implementation of an IAQ monitoring system, which uses sensor modules for measuring $PM_{10}$, $CO_2$, temperature, humidity, and a data processing module with CDMA (Code Division Multiple Access) communication capabilities for the transmission and management of the measurement results. The need for air quality measuring over large underground subway areas, such as waiting rooms, platforms, and tunnels, necessitates wireless connectivity for the measuring device. Wireless sensor networks represent a vast and active research area in which a large number of applications have been proposed, including indoor air quality monitoring and control [25], structural health monitoring [26], and traffic monitoring [27]. Figure 6 shows an air quality monitoring system for subway stations. The sensor and CDMA modules were installed at a waiting room, a platform, an outdoor site, and tunnels. MDT-800 (Telit, UK) is used for CDMA communication modules, while the MDT-800 is a complete modem solution for wireless m2m applications. The MDT-800 with a frequency band of about 800 MHz is ideally suited for real-time monitoring and control applications without the need for human intervention between remote machines and back office services. The measured air quality data were transmitted to the m2m platform via the CDMA Repeater and the BTS (Base station Transceiver Subsystem), eventually reaching the air quality monitoring server through the internet.

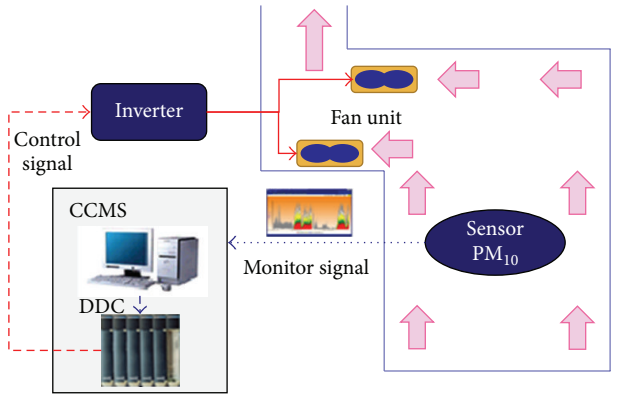



Figure 7: Ventilation fans and control systems installed at the natural ventilation point in the tunnel.

*2.5. Ventilation Fan System to Control IAQ in a Subway Station.* The PSDs improved the indoor air quality in subway stations. However, the air quality in the subway tunnels became worse. Therefore, ventilation systems are needed to improve the air quality in the subway tunnels. Some subway stations of Seoul Metro line number 1 had no ventilation systems for tunnels, "Jegi" subway station being one of them. In this study, we tried to improve the air quality by installing ventilation fans at the natural ventilation points in the tunnel at "Jegi" subway station. Figure 7 shows the ventilation fans and control systems installed at the natural ventilation points of the tunnel. The Central Control and Monitoring System (CCMS) monitored the $PM_{10}$ concentration in the tunnel. It

Table 4: Specifications of the ventilation fan.

| | |
|---|---|
| Input voltage | 220 V |
| Rated power | 600 W |
| Rated speed | 1,750 rpm |
| Static pressure | 15~20 mmAq |
| Air volume | 3,600 CMH |

Figure 8: Ventilation fans installed at the natural ventilation point in the tunnel at "Jegi" subway station.

controls the inverter of the ventilation fans so that if the air quality in the tunnel deteriorates and the $PM_{10}$ concentration increases more than the permissible limit, it can operate the ventilation fans using the Direct Digital Control (DDC). Figure 8 shows the ventilation fans installed at the natural ventilation point in the tunnel at "Jegi" subway station. The specifications of the ventilation fan are listed in Table 4.

## 3. Results and Discussion

*3.1. Accuracy Improvement of the PM Measuring Instruments HCT and Esampler.* A linear regression analysis method was used to improve the accuracy of HCT using the light scattering method. The data measured by this PM measuring instrument had to be converted to actual $PM_{10}$ concentrations using some factors. One instrument using $\beta$-ray absorption method (Ebam) and the other instruments using light scattering method (HCT, Esampler) were installed and measurements were taken on the platform of a subway station of Seoul Metro line number 1 for 5 days as shown in Figure 9. The measured $PM_{10}$ concentrations of Ebam and HCT are shown in Figure 10. The RMSE (Root Mean Square Error) in Figure 10 was 72.5227 $\mu g/m^3$. Using the linear regression analysis technique shown in Figure 11, the measured $PM_{10}$ concentration of HCT in Figure 10 can be corrected as shown in Figure 12. The RMSE in Figure 12 was 17.7128 $\mu g/m^3$. In Figure 11, the correlation coefficient $R^2$ was 0.9324. Next, the measured $PM_{10}$ concentrations of Ebam and Esampler are shown in Figure 13. The RMSE in Figure 13 was 94.7440 $\mu g/m^3$. Using the linear regression analysis technique shown in Figure 14, the measured $PM_{10}$ concentration of Esampler in Figure 13 can be corrected as shown in Figure 15. The RMSE in Figure 15 was 22.6132 $\mu g/m^3$. In Figure 14, the correlation coefficient $R^2$ was 0.8818. It can be seen in Figures

Figure 9: PM measuring instruments installed in "Jegi" subway station.

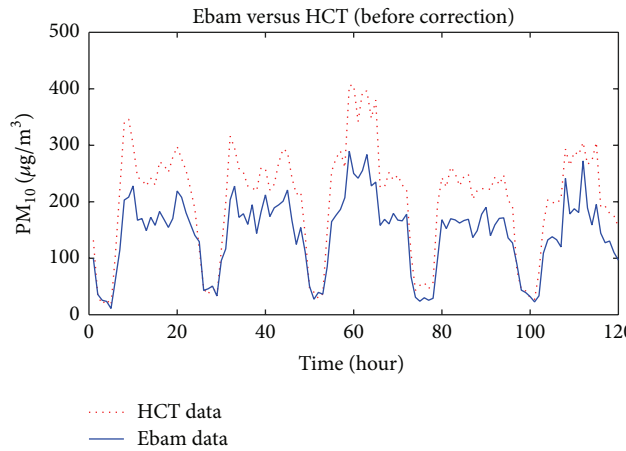



Figure 10: $PM_{10}$ concentrations of Ebam and HCT measured on the platform of "Jegi" subway station.

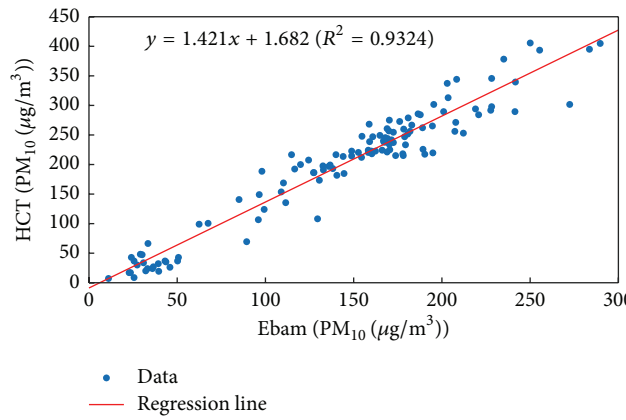



Figure 11: Linear regression analysis.

12 and 15 that the measured $PM_{10}$ concentrations of HCT and Esampler are very similar to that of Ebam if they are corrected using a linear regression analysis technique. This finding suggests that the PM measuring instruments using light scattering method can be used to measure and control the $PM_{10}$ concentrations of the underground subway stations. Because HCT is much cheaper than Esampler and is better in RMSE, we chose HCT to measure the $PM_{10}$ concentrations

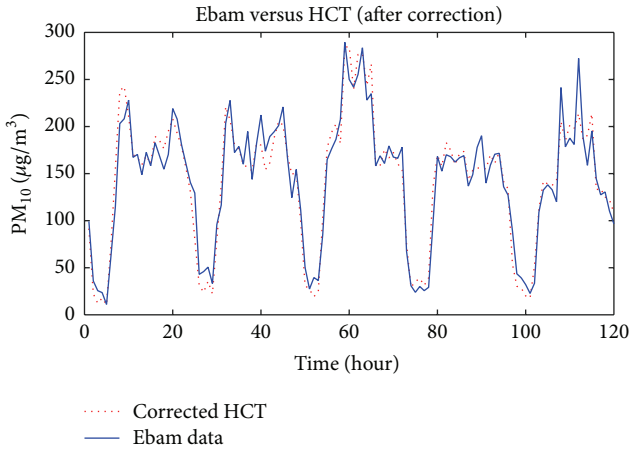



FIGURE 12: $PM_{10}$ concentrations corrected using a linear regression analysis.

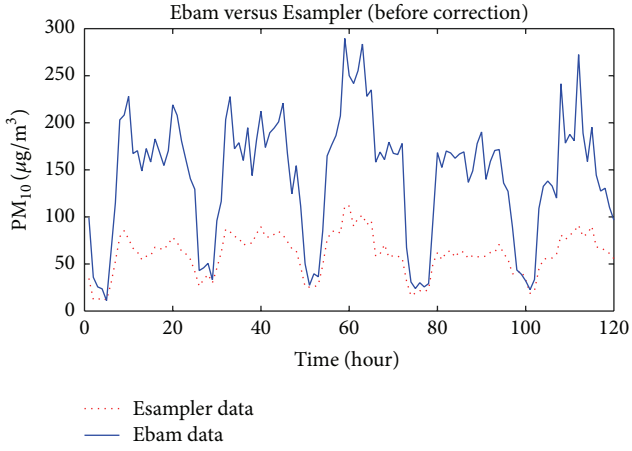



FIGURE 13: $PM_{10}$ concentrations of Ebam and Esampler measured on the platform of "Jegi" subway station.

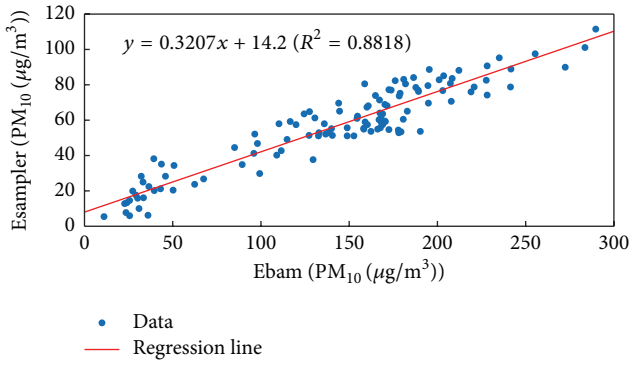



FIGURE 14: Linear regression analysis.

of the underground subway stations. However, because of Fe-containing particles, the particles in tunnel are heavier than those of same size in waiting room, platform, and outdoor site. So, the calibration of Figure 12 should be done for every location such as tunnel, waiting room, platform, and outdoor site.

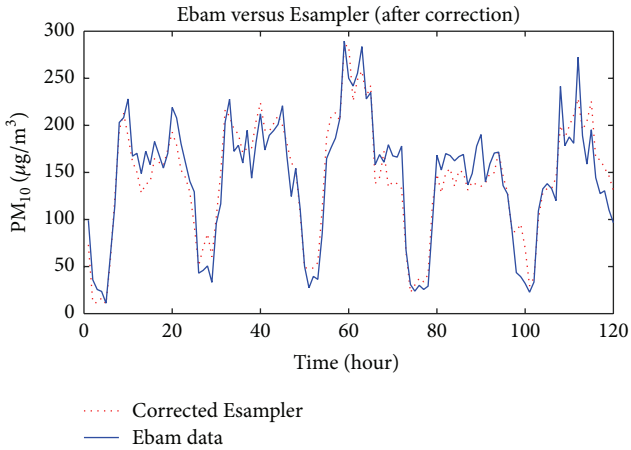



FIGURE 15: $PM_{10}$ concentrations corrected using a linear regression analysis.

*3.2. Monitoring of Air Quality in a Subway Station.* The air quality measuring instruments for $PM_{10}$, $CO_2$, temperature, and humidity were installed in the waiting room, platform, and tunnel and at the outdoor site of "Daecheong" subway station of Seoul Metro line number 3 as shown in Figure 16.

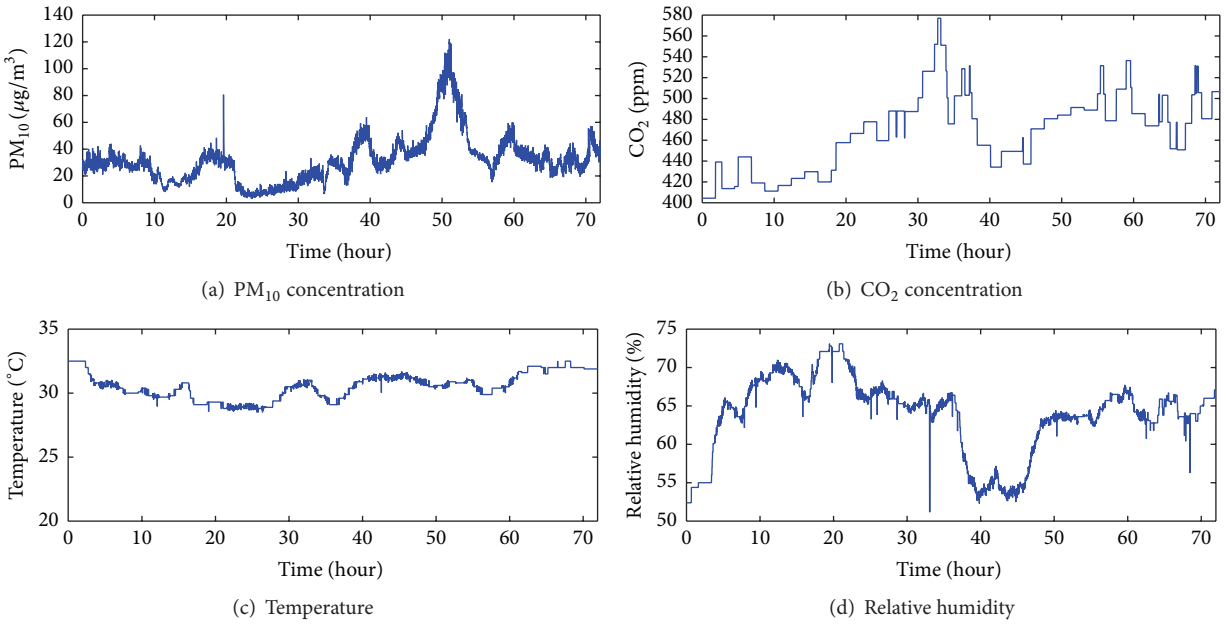
The $PM_{10}$, $CO_2$, temperature, and humidity data which were measured in the waiting room of "Daecheong" subway station for 3 days at a sampling interval of 30 sec are shown in Figure 17. As for the $PM_{10}$ concentration, it was kept under 120 $\mu g/m^3$, which met the KMOE's IAQ standard for $PM_{10}$ concentration (150 $\mu g/m^3$). As for $CO_2$ concentration, it was kept between 400 and 580 ppm, which met the KMOE's IAQ standard for $CO_2$ concentration (1000 ppm). The temperature of the waiting room was 29–32°C and the relative humidity was 53–73%. Figure 18 shows the results for the platform at "Daecheong" subway station. The $PM_{10}$ concentration was kept under 75 $\mu g/m^3$, which was lower than that in the waiting room. This was due to the PSDs, which blocked the dust from the tunnel. In addition, the ventilation fans on the platform were operated more frequently compared to those in the waiting room because passengers gather at the platform. The $CO_2$ concentration was 500–700 ppm, which was a little higher than that in the waiting room. It was because of the crowd of passengers. Temperature of the platform was 31–33°C and the relative humidity was 50–64%.

Figure 19 shows the results for the tunnel between "Daecheong" and "Hangnyeoul" subway stations. The $PM_{10}$ concentration was 40–400 $\mu g/m^3$, which was much higher than KMOE's IAQ standard for $PM_{10}$ concentration (150 $\mu g/m^3$). However, it was lower than 50 $\mu g/m^3$ when the train stopped for 4 hours (1:00 a.m.–5:00 a.m.). Most of the ventilation fans may not have been in working condition because of their deterioration and high running cost. Therefore, the $PM_{10}$ concentration in tunnels might have been high for a long period of time. The $CO_2$ concentration in the tunnel was 400–800 ppm, while the temperature was 27–31°C and the relative humidity was 60–80%.

Finally, the data for the outdoor site at "Daecheong" subway station are shown in Figure 20. The $PM_{10}$ concentration

(a) Waiting room (b) Platform (c) Tunnel (d) Outdoor site

Figure 16: Air quality measuring instruments installed at "Daecheong" subway station.




(a) $PM_{10}$ concentration (b) $CO_2$ concentration

(c) Temperature (d) Relative humidity

Figure 17: $PM_{10}$, $CO_2$, temperature, and relative humidity in the waiting room of "Daecheong" subway station.

was approximately 10–100 $\mu g/m^3$ except for two cases in which it was higher than 700 $\mu g/m^3$. The air quality of the outdoor site at "Daecheong" subway station was relatively good. The cause of the generation of two-pulse data is that the $PM_{10}$ concentration was measured using light scattering method. One of the demerits of light scattering method is that, sometimes, it gives very large values. Environmental sensors are important components for monitoring the IAQ in subway systems, as they provide the data needed for continuous online implementation. Sometimes, these sensors suffer from poor data quality and sensor reliability due to the hostile environment in the subway stations in which the measuring instruments are installed for monitoring. They may even fail for a long period of time as indicated in Figure 20(a). These failures could reduce the accuracy and reliability of the measurement, which may result in an erroneous control action and false perception regarding the performance of the monitoring system. Faulty sensors that have either completely or partially failed could provide incorrect information regarding monitoring and control. Therefore, many researchers have tried to prevent these problems [23, 28–30]. The $CO_2$ concentration of the outdoor site was 400–520 ppm, while the temperature was 24–35°C and the relative humidity was 50–90%. It can be seen that the relative humidity is inversely proportional to temperature.

*3.3. IAQ in the Tunnel Controlled Using Ventilation Fan System.* As mentioned earlier, the tunnel at "Jegi" subway station has no ventilation systems. Therefore, we installed ventilation fans at the natural ventilation points of the tunnel to improve the air quality, as shown in Figure 8. The CCMS monitors the $PM_{10}$ concentration in the tunnel and controls the inverter of the ventilation fans using the DDC. Figure 21 shows the experimental results of the $PM_{10}$ concentration of the tunnel at "Jegi" subway station with offline fan control for 25 days. At first, the ventilation fans were stopped for 5 days to investigate the effects of the ventilation fans. Subsequently, the speed of the fans was set to 60 Hz for 8 days and then to 70 Hz for 6 days. Finally, for 5 days, it was set to 75 Hz during rush hours and 60 Hz during off-peak hours to decrease the

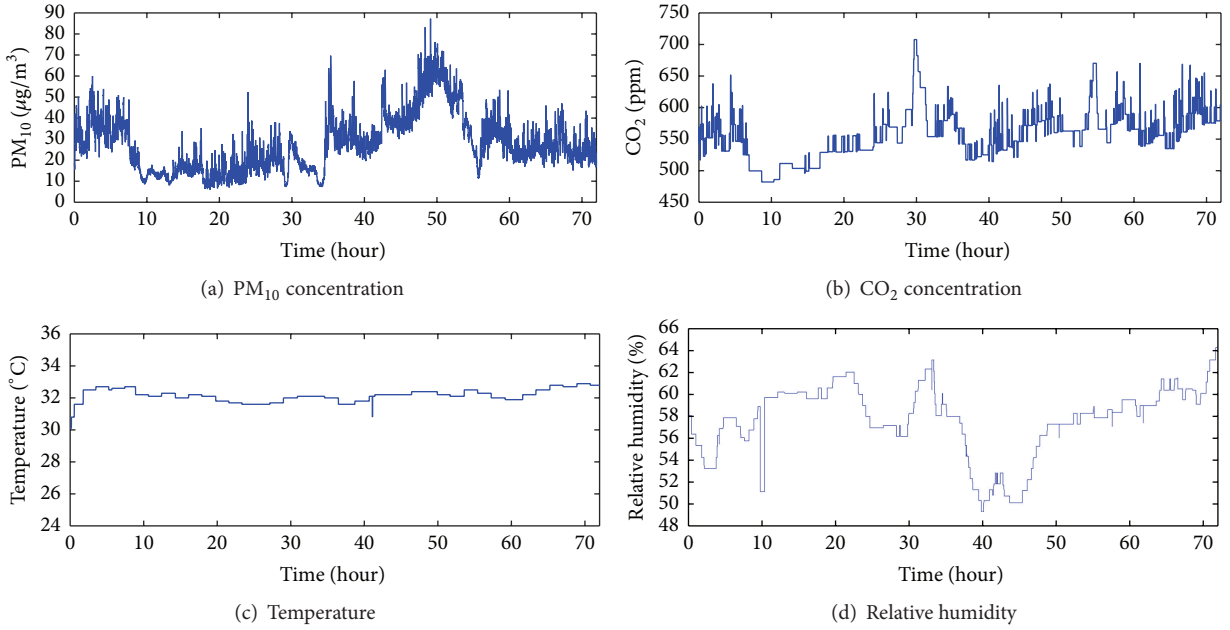



FIGURE 18: $PM_{10}$, $CO_2$, temperature, and relative humidity at the platform of "Daecheong" subway station.

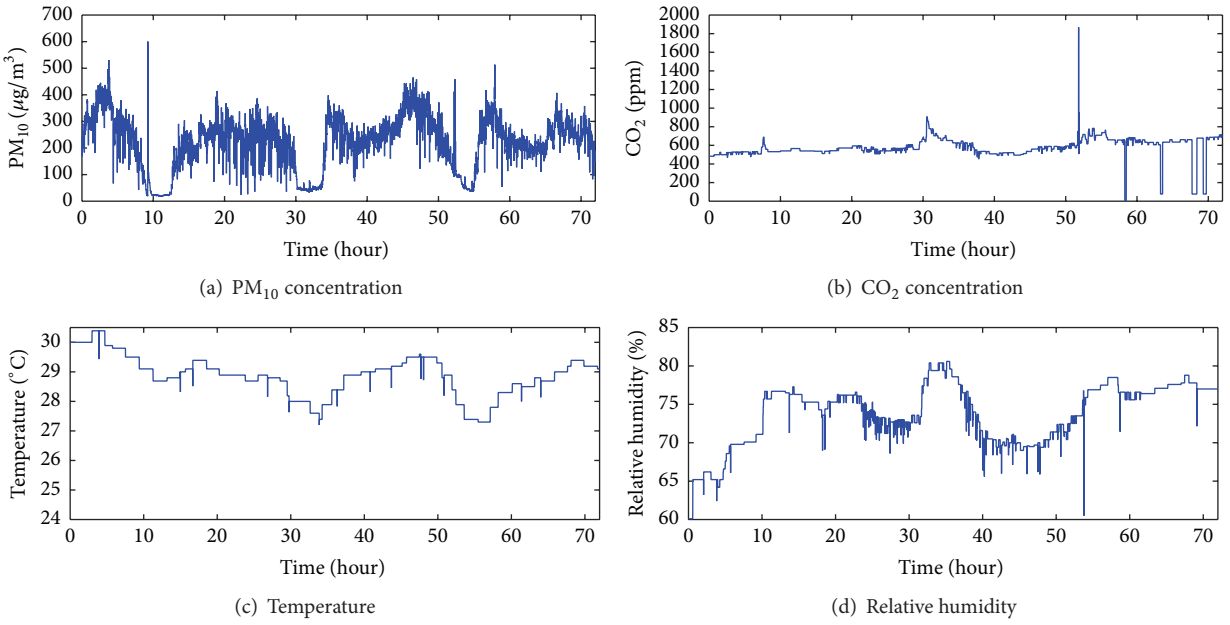



FIGURE 19: $PM_{10}$, $CO_2$, temperature, and relative humidity inside the tunnel between "Daecheong" and "Hangnyeoul" subway stations.

loss of electric energy. Through experimental studies, we found that the $PM_{10}$ concentration of the tunnel could be decreased by increasing the air flow rate of the ventilation fans (by increasing the speed of the ventilation fans). We also noted an abrupt increase in the $PM_{10}$ concentration when the fans were stopped for a day before the experiments were completed.

Figure 22 shows the experimental results of the $PM_{10}$ concentration in the tunnel at "Jegi" subway station with online fan control for 85 days. Initially, the ventilation fans were stopped for 2 weeks. Then, as shown in Figure 7, the DDC controlled the speed of the fans at 60, 70, or 75 Hz depending on the $PM_{10}$ concentration in the tunnel for 10 weeks. As shown in Figure 22, we found that the air quality in the tunnel could be improved gradually by using the ventilation fans, although it might cause the IAQ to be affected by the ambient air.

## 4. Conclusions

An air quality monitoring system based on environmental sensors was implemented to display and record the data of

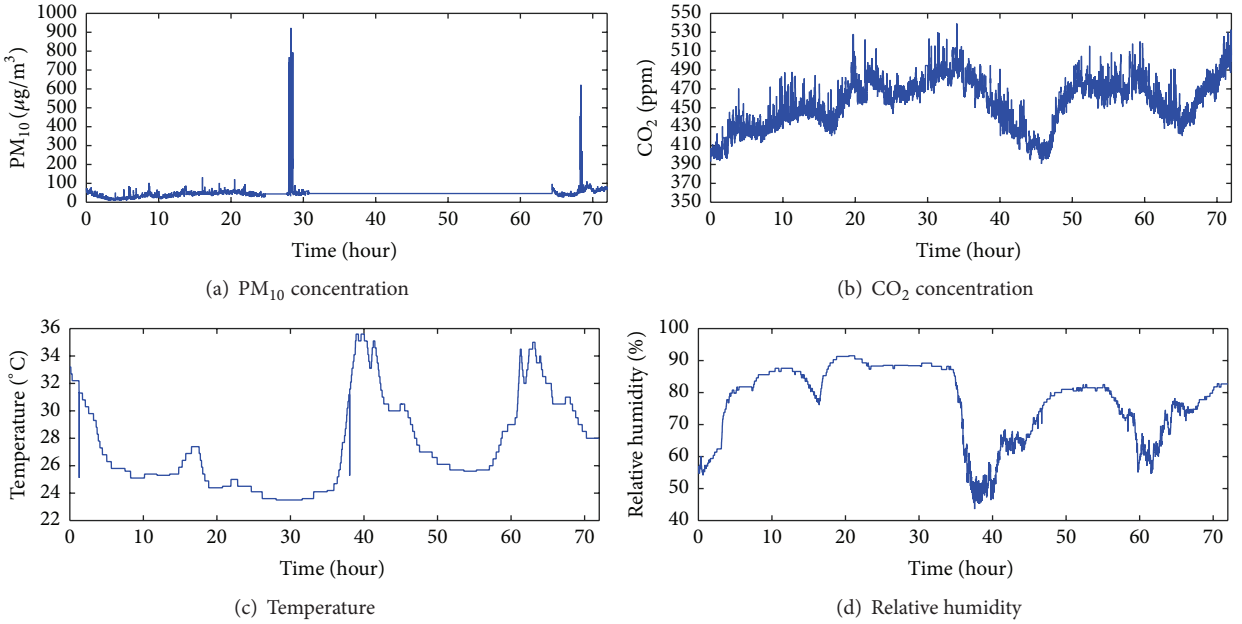



FIGURE 20: $PM_{10}$, $CO_2$, temperature, and relative humidity at the outdoor site of "Daecheong" subway station.

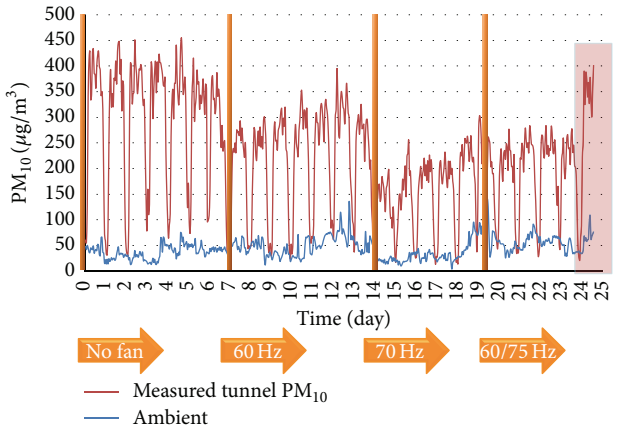



FIGURE 21: $PM_{10}$ concentration in the tunnel and the ambient air of "Jegi" subway station with offline fan control.

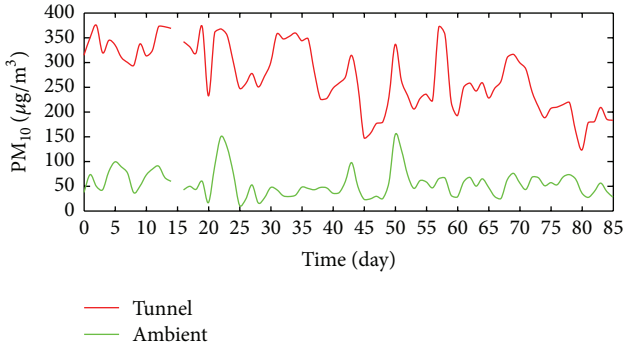



FIGURE 22: $PM_{10}$ concentration in the tunnel and the ambient air of "Jegi" subway station with online fan control.

$PM_{10}$, $CO_2$, temperature, and humidity of waiting rooms, platforms, tunnels, and outdoor sites at underground subway stations. The accuracy of the PM measuring instruments using light scattering methods was improved with the help of a linear regression analysis technique to continuously measure the $PM_{10}$ concentrations in the subway stations. Even though the accuracy was greatly improved, this approach had its demerits, such as the generation of very large measured data and the need to repeat the linear regression analysis every time the PM measuring instruments were moved to other places.

Ventilation fans were installed at the natural ventilation points in the tunnel to improve its air quality. Through some experimental studies, we found that the $PM_{10}$ concentration of the tunnel could be decreased by increasing the rate of air flow from the ventilation fans. Thus, the air quality of the tunnel could be improved by using the ventilation fans, although it might cause the IAQ to be affected by the ambient air. Therefore, some strategies for controlling the ventilation fan should be implemented such that the ventilation system should switch off when the air quality outside is very poor.

## Competing Interests

The authors declare that they have no competing interests.

## Acknowledgments

This work was supported by the 2014 Research Fund of the University of Seoul for Gyu-Sik Kim. Also, this work was supported by the Seoul R&D Program (CS070160) for Jo-Chun Kim.

## References


[1] http://www.euro.who.int/__data/assets/pdf_file/0009/128169/e94535.pdf.

[2] J. Song, H. Lee, S.-D. Kim, and D.-S. Kim, "How about the IAQ in subway environment and its management?" *Asian Journal of Atmospheric Environment*, vol. 2, no. 1, pp. 60–67, 2008.

[3] Y.-S. Son, Y.-H. Kang, S.-G. Chung, H. J. Park, and J.-C. Kim, "Efficiency evaluation of adsorbents for the removal of VOC and $NO_2$ in an underground subway station," *Asian Journal of Atmospheric Environment*, vol. 5, no. 2, pp. 113–120, 2011.

[4] K. J. Ryu, M. Juraeva, S.-H. Jeong, and D. J. Song, "Ventilation efficiency in the subway environment for the indoor air quality," *World Academy of Science, Engineering and Technology*, vol. 63, pp. 34–38, 2012.

[5] Y.-S. Son, T.-V. Dinh, S.-G. Chung, J.-H. Lee, and J.-C. Kim, "Removal of particulate matter emitted from a subway tunnel using magnetic filters," *Environmental Science and Technology*, vol. 48, no. 5, pp. 2870–2876, 2014.

[6] K. Y. Kim, Y. S. Kim, Y. M. Roh, C. M. Lee, and C. N. Kim, "Spatial distribution of particulate matter ($PM_{10}$ and $PM_{2.5}$) in seoul metropolitan subway stations," *Journal of Hazardous Materials*, vol. 154, no. 1–3, pp. 440–443, 2008.

[7] H.-J. Jung, B. Kim, J. Ryu et al., "Source identification of particulate matter collected at underground subway stations in Seoul, Korea using quantitative single-particle analysis," *Atmospheric Environment*, vol. 44, no. 19, pp. 2287–2293, 2010.

[8] K.-H. Kim, D. X. Ho, J.-S. Jeon, and J.-C. Kim, "A noticeable shift in particulate matter levels after platform screen door installation in a Korean subway station," *Atmospheric Environment*, vol. 49, pp. 219–223, 2012.

[9] S.-B. Kwon, Y. Cho, D. Park, and E.-Y. Park, "Study on the indoor air quality of Seoul metropolitan subway during the rush hour," *Indoor and Built Environment*, vol. 17, no. 4, pp. 361–369, 2008.

[10] S.-B. Kwon, D. Park, Y. Cho, and E.-Y. Park, "Measurement of natural ventilation rate in seoul metropolitan subway cabin," *Indoor and Built Environment*, vol. 19, no. 3, pp. 366–374, 2010.

[11] R. M. Harrison, A. R. Deacon, M. R. Jones, and R. S. Appleby, "Sources and processes affection concentrations of $PM_{2.5}$ and $PM_{10}$ particulate matter in Birmingham (U.K.)," *Atmospheric Environment*, vol. 31, no. 24, pp. 4103–4117, 1997.

[12] P. Aarnio, T. Yli-Tuomi, A. Kousa et al., "The concentrations and composition of and exposure to fine particles ($PM_{2.5}$) in the Helsinki subway system," *Atmospheric Environment*, vol. 39, no. 28, pp. 5059–5066, 2005.

[13] M. Braniš, "The contribution of ambient sources to particulate pollution in spaces and trains of the Prague underground transport system," *Atmospheric Environment*, vol. 40, no. 2, pp. 348–356, 2006.

[14] C. Rosen, L. Rieger, U. Jeppsson, and P. A. Vanrolleghem, "Adding realism to simulated sensors and actuators," *Water Science & Technology*, vol. 57, no. 3, pp. 337–344, 2008.

[15] C. K. Yoo, K. Villez, S. W. Van Hulle, and P. A. Vanrolleghem, "Enhanced process monitoring for wastewater treatment systems," *Environmetrics*, vol. 19, no. 6, pp. 602–617, 2008.

[16] G. A. Seber, *Linear Regression Analysis*, John Wiley & Sons, New York, NY, USA, 1977.

[17] D.-U. Park and K.-C. Ha, "Characteristics of $PM_{10}$, $PM_{2.5}$, $CO_2$ and CO monitored in interiors and platforms of subway train in Seoul, Korea," *Environment International*, vol. 34, no. 5, pp. 629–634, 2008.

[18] A. Kumar, H. Kim, and G. P. Hancke, "Environmental monitoring systems: a review," *IEEE Sensors Journal*, vol. 13, no. 4, pp. 1329–1339, 2013.

[19] H. Grimm and D. J. Eatough, "Aerosol measurement: the use of optical light scattering for the determination of particulate size distribution, and particulate mass, including the semi-volatile fraction," *Journal of the Air and Waste Management Association*, vol. 59, no. 1, pp. 101–107, 2009.

[20] B. S. Lee, K. H. Kim, Y. D. Yeom et al., "A study on $CO_2$/CO dual sensor module by NDIR method," *Advanced Engineering Forum*, vol. 2-3, pp. 541–546, 2012.

[21] J. Kwon, G. Ahn, G. Kim, J. C. Kim, and H. Kim, "A study on NDIR-based $CO_2$ sensor to apply remote air quality monitoring system," in *Proceedings of the International Joint Conference (ICCAS-SICE '09)*, pp. 1683–1687, August 2009.

[22] Y.-S. Kim, J. T. Kim, I.-W. Kim, J.-C. Kim, and C. Yoo, "Multivariate monitoring and local interpretation of indoor air quality in Seoul's metro system," *Environmental Engineering Science*, vol. 27, no. 9, pp. 721–731, 2010.

[23] H. Liu, M. Kim, O. Kang et al., "Sensor validation for monitoring indoor air quality in a subway station," *Indoor and Built Environment*, vol. 21, no. 1, pp. 205–211, 2012.

[24] J. Lim, Y. Kim, T. Oh et al., "Analysis and prediction of indoor air pollutants in a subway station using a new key variable selection method," *Korean Journal of Chemical Engineering*, vol. 29, no. 8, pp. 994–1003, 2012.

[25] H. C. Gi and G. S. Choi, "Design of service system framework for web-based monitoring and control of indoor air quality (IAQ) in subway stations," in *Proceedings of the International Conference on New Trends in Information and Service Science (NISS '09)*, pp. 659–663, July 2009.

[26] M. Z. A. Bhuiyan, J. Cao, and G. Wang, "Deploying wireless sensor networks with fault tolerance for structural health monitoring," in *Proceedings of the IEEE 8th International Conference on Distributed Computing in Sensor Systems (DCOSS '12)*, pp. 194–202, IEEE, Hangzhou, China, May 2012.

[27] M. Hulea, S. Folea, T. Leţia, and G. Moiş, "A collaborative approach to autonomous single intersection control," in *Proceedings of the 19th Mediterranean Conference on Control and Automation (MED '11)*, pp. 694–699, Corfu Island, Greece, June 2011.

[28] R. Dunia and S. J. Qin, "Joint diagnosis of process and sensor faults using principal component analysis," *Control Engineering Practice*, vol. 6, no. 4, pp. 457–469, 1998.

[29] S. Wang and Y. Chen, "Sensor validation and reconstruction for building central chilling systems based on principal component analysis," *Energy Conversion and Management*, vol. 45, no. 5, pp. 673–695, 2004.

[30] Y. Chen and L. Lan, "Fault detection, diagnosis and data recovery for a real building heating/cooling billing system," *Energy Conversion and Management*, vol. 51, no. 5, pp. 1015–1024, 2010.