IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 9, 2019, accepted May 27, 2019, date of publication June 5, 2019, date of current version June 27, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2920838

A Wireless Sensor Network for Monitoring Environmental Quality in the Manufacturing Industry

QILONG HAN[1], PENG LIU[1], HAITAO ZHANG[1], AND ZHIPENG CAI[2], (Member, IEEE)
[1]College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China
[2]Department of Computer Science, Georgia State University, Atlanta, GA 30303, USA

Corresponding author: Haitao Zhang (zhanghaitao@hrbeu.edu.cn)

This work was supported in part by the China Numerical Tank Innovation Project, in part by the National Natural Science Foundation of China under Grant 61370084 and Grant 61872105, in part by the Harbin Science and Technology Talents Research Special Foundation under Grant G063316006, and in part by the Excellent Academic Leaders Project (The Key Techniques of Home-based Care Service System) 2016RAXXJ013.

ABSTRACT Urban industrial plant areas are highly concentrated, and air pollution is increasingly serious. The quantity of outdoor air quality monitoring sites is insufficient. Aiming at the above questions, related studies propose solutions that use relatively cheap equipment networking to collect pollution data and accurately analyze local monitoring information. In this paper, a new type of outdoor air quality monitoring system is studied and preliminarily practiced and has proven certain feasibility and applicability. The main contributions of this paper are: first, we improve the network layout by employing the Zigbee network, which is combined with factory characteristics, and collected data on carbonic oxide, nitrogen dioxide, sulfur dioxide, ozone, particulate matter, temperature, and humidity. And then, to establish the dilution coefficient and diffusion coefficient of pollution diffusion, we adopt air movement as the energy model and, by utilizing the method of pollution traceability, achieve the complete coverage pollution monitoring of the whole city by local monitoring sites. Finally, we propose an improved long short-term memory (LSTM) method to predict the pollution period of urban air quality. The experimental results show that the improved LSTM prediction model has strong applicability and high accuracy in the period prediction of pollution weather. Meanwhile, by analyzing the specific case in detail, we prove that air pollution in the city is mainly caused by the manufacturing industry. We conclude that it will make a great contribution to the atmospheric environment protection of cities by using weather quality prediction to dynamically adjust the production.

INDEX TERMS Environmental quality, wireless sensor network, Zigbee, pollution monitoring analysis, LSTM.

I. INTRODUCTION

In the past quarter century, with the development of the industrial technology and transportation industry has grown exponentially, the energy and fuel demand is increasing and urban air pollution is becoming more and more serious which has an important impact on atmospheric environment, human health, ecosystem and climate change [1]. There are various pollutants in the atmosphere, such as automobile exhaust, Nitrogen Oxides (NO_x), Carbonic Oxide (CO), Sulfur Dioxide (SO_2), Ozone (O_3), and Particulate Matter (PM), which may come from fixed locations, such as factory communities, or from mobile emissions point whenever, these pollutants may spread to a wide area for a short time and causing huge and irreparable damage [2], [3]. Past research has established the negative health effects originating from the exposure to these pollutants. Consequently, there is an urgent need to monitor and control the concentrations of these gaseous pollutants in factories and for large-scale deployment of sensor technologies to enable monitoring factory environmental quality (FEQ) with high temporal and spatial resolution in factories [4]. However, currently available solutions are scarce and too costly for large installations [5]. In particular, to date, all wireless gas sensor networks have been facing a trade-off between sensor node cost and data quality because there is currently no suitable, low-cost technology

The associate editor coordinating the review of this manuscript and approving it for publication was Tiago Cruz.

available for specific, quantitative chemical analysis [6]–[9]. Especially on this basis, the multi-source analysis of the interaction between various pollution sources in the city.

Since 2016, many sensor manufacturers have released new gas sensors, such as the SGPC10 of SENSIRION in Switzerland and the four-in-one gas sensor BME680 of BOSCH in Germany, which have smaller size, lower power consumption and more sensory content, such as VOC, humidity and pressure, etc. [10]. Due to the composition of pollutant gas is very complex, the pollution phenomenon is several variable with the effect of various environmental factors and the monitoring methods for pollutant gas are also different from each other. Pollutants are mainly divided into two categories are PM and chemical pollutants. Monitoring methods for PM are mainly divided into four types are filter membrane weighing method, TEOM method, β-ray absorption method and light scattering method [11]. (1) Filter membrane weighing method. The mass concentration of fine particulate is measured by separating fine particulate onto the filter membrane; (2) TEOM method. The known volume of air is separated through the sampling head with cutpoint diameter and stacked on the microbalance. The mass concentration is obtained by dividing the weight value and volume; (3) The β-ray absorption method. Using the β-ray attenuation principle, when β-ray irradiates the filter membrane with particle precipitation, the energy of the ray will decrease, the attenuation amount and attenuation degree can be used to measure and calculate the PM concentration; (4) Light scattering method, is the reverse application of Mie scattering theory. The mass concentration of fine particulate can be measured by measuring the magnitude of the scattered light signal based on the characteristics of the attenuation of light passing through the particles. The above detection methods and instruments are basically reliable and accurate, but there are still two shortcomings: (1) Due to the high cost of instruments and equipment, large-scale deployment of these instruments requires a large fund input. As a result, it is difficult for monitoring stations to cover residential communities, factories, schools and other areas, and the number of detection equipment is far from enough; (2) Due to the complexity of air quality and local geographical environment in outdoor areas, air quality is related to weather changes, traffic conditions and residents' activities. The current popular detector is only sensitive to the environment around the instrument and part of the air sampled by the instrument, rather than the overall environmental situation of the region, so the judgment of result of whole region is easy to be affected by the sudden increase of a small amount of particles and the inaccuracy of the sample air in the vicinity of the instrument. And for chemical pollutants, for example, the CO concentration was oftentimes inferred from the reading of a metal-oxide based, total volatile organic compound (TVOC) sensor [12]–[16], but the correlation between TVOC and CO has been shown to be weak. While Raman-based approaches may be used to detect many gases simultaneously, techniques based on absorption spectroscopy are the most promising candidates for reliable CO detection [17]–[20]. It has been found that a high degree of sensitivity and specific, quantitative detection can be achieved by tunable diode laser spectroscopy, albeit at high associated costs in terms of optical and computational infrastructure and maintenance [21]–[24]. Currently, the so-called non-dispersive infrared absorption spectroscopy (NDIR) tool is the most popular tool for CO monitoring, which does not require analytical grade concentration readings [25].

In recent years, with the development of computer vision technology and image processing technology, many domestic and foreign scholars have proposed the air quality evaluation theory based on image processing technology. Wong *et al.* [26] proposed photographs taken with digital single-lens reflex cameras, optical theory and visible light-wave bands are used as the algorithm of atmospheric characterization data. The relationship between the measured reflectivity and the reflectors from material surface and atmosphere is analyzed, and the concentration of PM10 in the atmosphere is estimated by regression analysis. A model of camera bright-ness and the physical environment characteristics of the scene is used [27]. In this model, the brightness of the camera is determined by the scattering of light in the atmosphere, and the scattering phenomenon itself is associated with the physical properties of the medium such as floating particles in the atmosphere, so the model can explain why the contrast of camera imaging in fog is lower than that in sunny days. This also suggests that the quality of images can be used to evaluate differences in air quality. By combining remote sensing image data with video surveillance image data, the good and bad degree of air quality in a certain area is monitored and analyzed, and the corresponding air quality classification algorithm is given [28]. The greatest benefit of using surveillance image data to evaluate air quality is its low cost. However, until now, there is seldom a calculation method to evaluate air quality completely through video surveillance image data. The main reason is that the image itself has obvious distortion during the shooting process, so it is difficult to evaluate the quality of the shooting environment through single image or comparison of multiple images. An implementation air quality estimation method based on color image processing technology is proposed [29]. By analyzing the relationship between PM2.5 and image quality degradation, the dark channel prior model and black point data in the image were used to establish a correlation model between light disappearance coefficient and PM2.5 concentration, so as to estimate PM2.5 concentration value. Hsin-HungHsieh proposed a method for estimating the concentration of particles using a consumer camera. Using the support vector machine (SVM) technology, the related data that may affect PM2.5, including the brightness value of the image gray image, the humidity value, the pixel information in the sky gray histogram, the contrast of the density distribution in the sky gray image and the histogram

information of the HSV mode image, and PM concentration model is set up. It is concluded that brightness, humidity and PM2.5 concentration are correlated [30].

In this study, firstly we presented a design that built upon on the Zigbee idea, that is, using the Zigbee network to collect the pollution emission of an urban manufacturing industry. We took integration and network layout on the collection sensors with correlation pollution elements to collect pollution data and analyze pollution phenomena as a whole, instead of collecting single pollution index separately, we integrated this basic setup into a wireless sensor network to enable the deployment of sensor nodes capable of calculating the concentration of NO_2, CO, SO_2, temperature, humidity, and solid particulate matter. Using our wireless sensor network architecture, we demonstrated the long-term stable operation of pollution data collection with the networking method proposed in this study. The method opened up the future possibility of producing energy-saving, easy-to-manufacture pollution gas sensors that do not feature cross-sensitivities towards humidity. Because of the small size of the sensor nodes, their wireless internet connectivity, and their low cost, the system architecture can easily be adapted to different scenarios and employed on a large scale. Then, at the same time, due to the happening of the pollution is very complicated, and it is related to geography, meteorology, multi-polluting factors, manufacturing production and urban residents' lives. However, the existing methods have been relatively accurate, but the sampling categories and research pollution phenomena are relatively single. The formation, transmission and diffusion of pollution within the city and the production and living activities within the city should constitute a whole system. Therefore, we have studied the overall relationship between the production behavior of urban manufacturing industry and urban pollution. Compared with analyzing a single pollution phenomenon or a single pollution source separately, we prefer to conduct multi-source analysis through the interaction between individuals to solve complex pollution problems.

II. MATERIALS AND METHODS

Because of its importance in determining factory air quality, we took a gas sensor approach for CO by employing a concept that can easily be adapted for other gases, including NO_2, CO, and SO_2. There is evidence to suggest that correlations exist between the energy consumption of production and the energy expended on pollution control, with the quantity of flue gas also allowing for a production limit grade [31]. This made exhaust gas concentration the most important parameter in assessing FEQ. In order to enable large-scale deployment of the system at a low cost and at a small overall size, we made use of the RFD function, based on the original design, which made it possible to build a large network that could cover all enterprises in the whole jurisdiction while improving battery life and reducing system cost and power consumption through the combination of multi-hop transmission, FFD and RFD. In wireless sensor networks, the energy of the nodes is mainly consumed through communication, so choosing a communication chip with low power consumption and high performance was crucial to prolonging the life of the nodes [32]–[36]. In this design, CC2420 was selected as the main control chip of wireless transceiver module. The standard radio frequency transceiver, moreover, had to have very low power consumption and a stable performance to ensure the effectiveness and reliability of short-distance communication. Therefore, we chose the ATmegal28 chip as the processor. It had a high-performance, low-power 8-bit AVR microprocessor developed by ATmega. In addition, it had six power-saving modes that could be selected through software. Selecting these two chips could greatly help us reduce the power consumption of the nodes. Next, we had to extend the life of the node. To do this, an A/D converter was used to convert the current signal into a digital signal. Then, the integrated digital information was processed by the processor. Using an apparatus to simulate real-world conditions in the laboratory, we performed a calibration of the pollution gas sensor reading in dry synthetic air at 1bar pressure. We also installed the calibrated sensor in the real enterprise production environment. Data were transmitted via the repeater module, and sensing data on temperature, humidity, and pollution gas concentration were sent every 10min.

The sensor node design is depicted in Figure 3 and included wireless connectivity to make it possible to monitor the pollutant exposure of humans on a micro-scale, which is needed for next-generation studies on the health effects of pollutants on urban residents. The data from the whole network can also be fused to offer a comprehensive picture on a factory-wide level. Each node was equipped with a repeater to enable the transfer of the data to an internet application by means of an internet gateway. In the undisturbed environments, the physical range of the transmission was about 100m. We created a tool to help determine the energy consumption of production, the energy expended on pollution control, and the production rate of gaseous pollutants in factories. Moreover, data on the ratio of the above energy consumption metrics can be used and integrated into production systems to minimize energy expenditure during factory production. Using this system, we investigated the result of various gas detection methods in a city and made inferences on the development of classes of pollution in factories through a system deployed in a city in Northern China. The location of all sensor network components of one of them, which is a large-scale iron and steel enterprise, is depicted in Figure 1. Each sensor node was installed at about 1m height above ground. To ensure reliable data, transfer signal repeaters were installed. The factory was founded in 1990 and now has an annual production capacity of 2 million tons of iron, 2 million tons of steel, and 2 million tons of steel, with an annual output value of 10 billion yuan. The factory had a high production capacity and some ability to control pollution but not active production control. As such, its pollution discharge was affected by the enterprise's production capacity and by

TABLE 1. Sensor node parameter.

Instrument type	XHAQSN-808
XHAQSN-808 monitoring parameter	*PM, SO_2 , NO_2 , CO, temperature, humidity*
time resolution	*1 min*
Power supply	*Power (220 V) supply, solar power supply (12 V)*
overall dimension	*220*220*300 (Host)*
working environment	*T (-20–55) ° C, RH (15%–95%)*
communication mode	*GPRS, WiFi, Bluetooth*
Battery	*lead-acid battery*
working time	*Apparatus without outside power continuous operation of 30 days*
Weight	*2.4 Kg*
storage environment	*0 ° C–50 ° C, <90% RH*
product feature	*low cost, small volume, easy deployment*
	low power wireless communication design

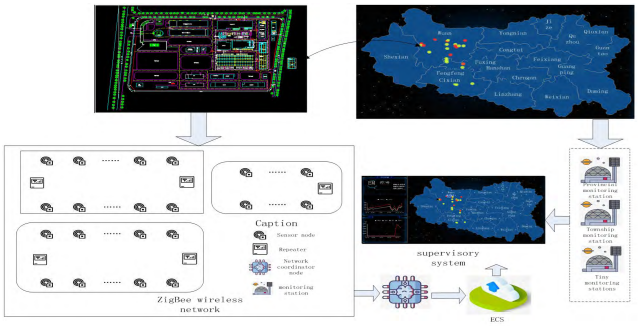

FIGURE 1. Overview of the system installation at the factory.

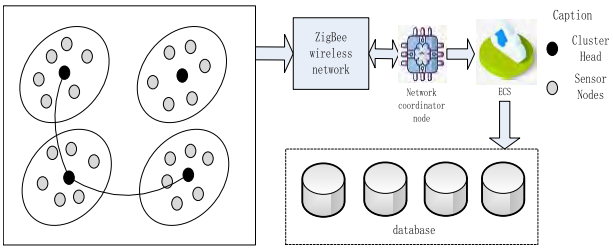

FIGURE 2. System structure diagram.

pollution control energy consumption. For investigating the influence of the ratio of production energy consumption to pollution control energy consumption (PE-PCR) on FEQ, 19 point locations were monitored, and each was equipped with a number of sensor nodes. Because of the materials deployed in the factory building, the signal wave distance was reduced and limited by the RFD information transmission distance. Therefore, an appropriate number of relay modules were installed according to the workshop area to ensure reliable signal transmission.

The algorithm selected the optimal communication path and sent the data to the coordinator in the form of multiple hops. After the coordinator received the data, it verified the data. If it was correct, it returned the confirmation information

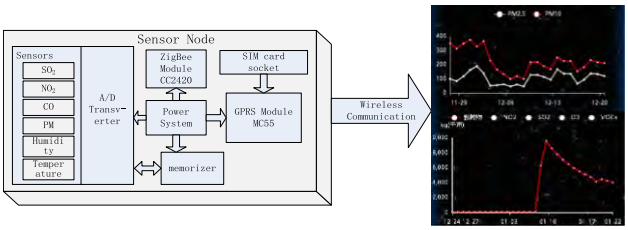

FIGURE 3. Sensor node topology.

to the sensor node according to the original path to realize handshake communication. If a checksum error occurred, a resend was required. If the sensor node did not receive a confirmation message, it would keep sending data until a confirmation message was received from the co-ordinator. When the coordinator received the correct data, it also performed protocol conversion and then transferred it to the monitoring platform through the network for data processing and analysis.

The concept of the individual sensor nodes was based on the use of micromachined sensor technology and internet connectivity in order establish a wireless network for online, in-situ factor environmental monitoring. The CO concentration was determined by listening to its concentration via the photoacoustic effect. Both temperature and humidity were determined using state-of-the-art microtechnology.

Each sensor module sent the current pollution gas concentration, speed, and pressure to the cloud-based ''EnControl'' platform every 10min. Outdoor conditions were monitored via weather stations in close proximity to the factory, whose data was available online through the weather underground web portal (i.e., the subordinate platform of meteorological department). The sensor node was a miniaturized instrument that monitored multiple parameters, such as PM, SO_2, NO_2, CO, temperature, and humidity in the atmosphere. The equipment used the flexible power take off method and could choose either a municipal power supply or solar power supply. The instrument selected various high precision sensors, such as electrochemical and optical, and the detection limit was low, the number of outputs was accurate, and the time resolution was high. In addition, the sensor node also had a small size and a low price, and was suitable for gridding and dense distribution. The equipment was based on wireless communication technology, and could realize confidential and safe communication with server and integrate a large quantity of environmental data into the ''cloud platform''.

III. THE IMPROVED POLLUTION ANALYSIS

In this section, we mainly proposed a pollution traceability method, and used LSTM to predict the period of each pollution element.

A. THE IMPROVED TRACEABILITY METHOD

In order to assess the air quality in the factory, we used the comprehensive reduced concentration of smoke as the main gas calculation parameter for exhaust emission. It is high cost

and not easy to achieve to city-wide coverage by sensors. Therefore, only 120 companies in the test city have installed our sensors. In order to achieve the purpose of pollution traceability to the whole city, predict the pollution diffusion of near-surface through these factories with sensors installed and analysis on pollution traceability and pollution cause to the whole city by the obtained knowledge. We find the main distribution area of factory blowdown chimney concentration and better height through the box-plot analysis, calculate the height interval is 100-120m. The height is set to 150m in the process of pollution traceability. Due to the near-surface height, the effect of atmospheric phenomena such as inversion layer on pollution diffusion can be neglected. In the process of modeling, ignore the effects of molecular motion, temperature and inversion layer, etc., assuming that the energy of pollution diffusion is only provided by air mass, we join geography which the whole urban area is divided into 22 thermal grids, emissions and precipitation data information, research on the pollutant diffusion, determine the air dilution factor and air diffusivity factor of the pollutants. The pollution comes from which thermal grid of monitoring alarm position is derived reversely based on wind direction, wind power and rainfall conditions, we determine which companies are most likely to be pollution sources according to the type of factory and its corresponding pollution list in the grid, and provide them to the local environmental protection department for a second detailed investigation. Where air dilution factor and air diffusivity factor are as shown in formula (1) and (2). Along with this value, we used speed and time to calculate the tailpipe emission and the background levels at each point location. Among the diverse methods used to calculate the air change rate (ACR), we chose the decay method according, to VDI 4300 (2001), which has been proven a feasible and effective way to determine the air change rates in scenarios like the ones explored in this work. The exhaust gas concentration upon venting using outdoor air had to converge to the global background concentration, which we assumed to be 400ppm. However, meteorology can contribute to the dilution and diffusion of polluted gases. As such, we used the equation which derived for the temporal evolution of the exhaust emission concentration to determine the effective back-ground concentration, Ca, the air dilution factor, ω (considering only rainfall), and the air diffusivity factor, λ, using the following fit function of the form:

$$C(\mathrm{t}) = \frac{(C_0 - C_\mathrm{a}) \cdot \left(\alpha e^{-\lambda \cdot \mathrm{t}} + \beta e^{-\omega \cdot \mathrm{t}}\right)}{2} + C_\mathrm{a} \qquad (1)$$

where CO is the initial concentration upon the start of the pollutant discharge at $\mathrm{t} = 0$, and α and β are the weight coefficients. A nonlinear regression ws applied to the measured raw data acquired in order to determine the background concentration. Based on the dry weather, we evaluated λ first, as follows:

$$C(\mathrm{t}) = (C_0 - C_\mathrm{a}) \cdot e^{-\lambda \cdot \mathrm{t}} + C_\mathrm{a} \qquad (2)$$

With a known air dilution factor and air diffusivity factor, the value of exhaust concentration could be used to determine the emission capacity of the factory. Apart from the flue gas concentration, the ability to produce and control pollution were important factors influencing the FQE. Oftentimes, there was an organized discharge of pollution during production. However, the mismatch between the ability to control pollution and the production capacity would still lead to the occurrence of pollution events. Further large-scale measurements in various factories would be necessary to establish a reliable and more precise pollutant discharge model for factories. The PE:PCE ratio of production energy consumption to pollution control energy consumption represents the basic energy consumption of factory sewage discharge with the presumption that the influence factor of energy consumption and sewage discharge was θ. The influence factor was easily adaptable to different cities, industries, and seasons in the year, and was represented as follows:

$$\mathrm{Dc} = \theta \cdot \frac{\sum \mathrm{E_{pr}}}{\sum \mathrm{E_{pl}}} \qquad (3)$$

where Dc represents the amount of pollutants discharged in a given time period, $\mathrm{E_{pr}}$ and $\mathrm{E_{pl}}$ represent the production energy consumption and the total power consumption for pollution control within the period, respectively. The value of θ had to be determined to calculate the relationship between energy consumption and pollution emission.

B. THE IMPROVED LSTM PERIOD PREDICTION METHOD

We adopted a more elaborate data processing process in the process of pollution period prediction, which is different from the off-group points and missing values processing in the above-mentioned pollution traceability method. This is because the traceability process is based on the scope of the thermal grid. We aim to perform relatively ambiguous analysis with a larger amount of real data. Although refined data processing can improve the data quality, it also reduces the recall rate of data. In the process of pollution prediction, it is necessary to seek higher accuracy. Therefore, we first delete the noise data caused by sensor error, pollution fluctuation and protocol analysis error by the box-plot. As shown in Figure 4, the adopted production factor data in factory is used as an example for off-group data analysis. Secondly, we treat each sensor's daily data as an individual. For each individual, we take the data from the first peak to the last peak but excluding the last value. To ensure the alignment of the data, using the cubic spline method to interpolate each individual to an equal length with 144 sample points. It is verified that the final prediction result can be improved by nearly two percentage points by the experiment. Finally, to meet signal processing requirements, partial waveform data such as current is transferred from frequency domain to time domain.

In deep learning model, Recurrent Neural Network (RNN) introduces the concept of time sequence into the network structure design, which makes it more adaptable in

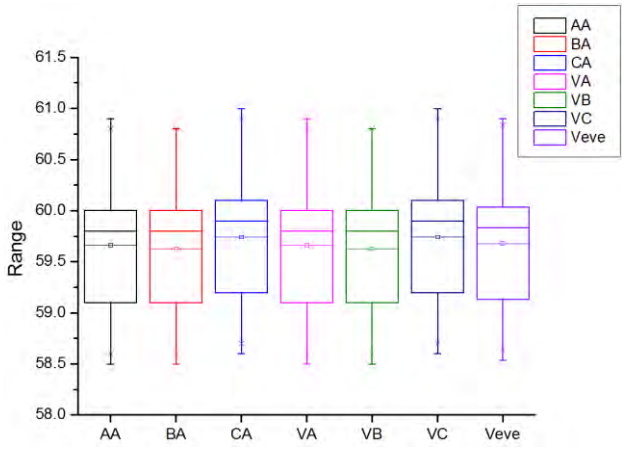

FIGURE 4. Box-plot analysis of factory electricity data.

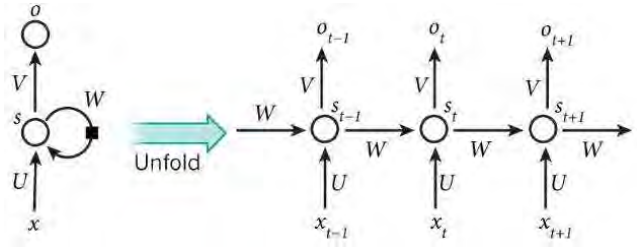

FIGURE 5. RNN model and cell structure in hidden layer.

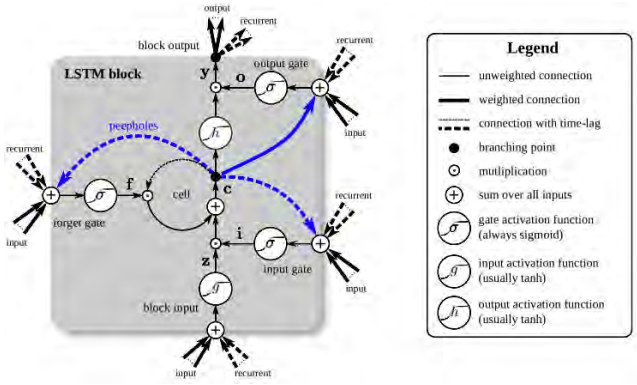

FIGURE 6. LSTM cell structure in hidden layer.

time series data analysis. Among numerous RNN variants, the long-short-term memory (LSTM) model compensates for gradient disappearance, gradient explosion and lack of long-term memory ability [37]. In recent years, LSTM has achieved good prediction results in the fields of fault series prediction [38], air pollution time-space forecast [39], short-term traffic flow prediction [40] and silicon content in hot metal prediction [41]. Compared with traditional RNN, LSTM can learn time series with long time span and automatically determine optimal time lag prediction, and more effectively utilize long-distance timing information. Like LSTM, recurrent neural network can almost seamlessly simulate problems with multiple input variables. This is a big advantage in time series prediction, and traditional linear methods are difficult to adapt to multivariate or multiple input prediction problems. We hope to obtain a prediction that includes the correlations between the attributes rather than a prediction that the attributes are fragmented. Therefore, according to the literature [42]–[45] we select eight indicators: PM2.5, PM10, SO_2, CO, O_3, NO_2, wind direct and wind speed for period prediction, and the data of the eight indicators are decomposed periodically to qualitatively validate our experience knowledge. Therefore, this paper selects LSTM method for pollution period prediction. The LSTM model replaces the hidden layer of RNN cells with LSTM cells to enable them to have long-term memory. After continuous evolution, the most widely used LSTM model cell structure [46]–[48] is shown in Figure 2, where z is input module, and its forward calculation method can be expressed as:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (4)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (5)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (6)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (7)$$

$$h_t = o_t \tanh(c_t) \quad (8)$$

where i is input gate, f is forgetting gate, c is cell state, o is output gate; W and b are the corresponding weight coefficient matrix and bias term respectively; σ and tanh are sigmoid and hyperbolic tangent activation function respectively [49]–[51]. The LSTM model training process uses a BPTT algorithm which is similar to the classical Back Propagation (BP) algorithm [52], it can be roughly divided into four steps: (1) calculate the output value of LSTM cells according to the forward calculation method; (2) inversely calculate the error term of each LSTM cell, including two back propagation directions by time and network level; (3) calculate the gradient of each weight according to the corresponding error term; (4) update the weights using a gradient-based optimization algorithm. Meanwhile, through the supervised learning framework, we provide the pollution collection values and the weather conditions of the previous time node for the neural network to predict the weather conditions in the next 48 hours.

Seasonal differences are strongly correlated with the spread of pollution and other pollution like coal-fired heating in winter causes more particles emissions happened. We added the seasonal difference markers and integer coding, if we understand the period prediction process as a classification process, the addition of a large amount of the same seasonal information will lead to over high purity of each subset in the seasonal features. Since we make predictions through the correlation between features, this will have an impact on our final prediction results, resulting in our prediction results being more inclined to the interference of seasonal features. In order to eliminate this purity effect in the process of adding

seasonal features, and we know short-term time series of spring and autumn are very similar through the data and the climate information of Wu'an, so sample D is divided by Gini coefficient into three sets of spring and autumn, winter, summer. We use seasonal features to divide the sample set, and the possible sets divided are as follows:

(1) divide point: ''Spring and Autumn'', divided subset: {Spring and Autumn}, {Winter, Summer}

(2) divide pointt: ''Winter'', divided subset: {Winter}, {Spring and Autumn, Summer}

(3) divide point: ''Summer'', divided subset: {Summer}, {Winter, Spring and Autumn}

For each partition mentioned above, the purity of the sample set D divided into two subsets D_1 and D_2 can be calculated.

$$\text{Gini(D, A)} = \frac{|D_1|}{|D|}\text{Gini}(D_1) + \frac{|D_2|}{|D|}\text{Gini}(D_2) \qquad (9)$$

Therefore, for a feature with multiple values, it is necessary to calculate each value as the divide point to divide the sample D, after divided the purity is Gini(D, A_i), where A_i is the possible value of feature A. Then, the division with the smallest Gini index is found from all possible divisions. The divide point of this division is the best division point for dividing the sample set D by feature A. It is found that the first division method is more suitable for our requirements through calculation. This is because although summer and winter are in a set, their series features are different. Even if the spring and autumn are placed in a set, due to the similar combination of internal time series features does not cause much impact.

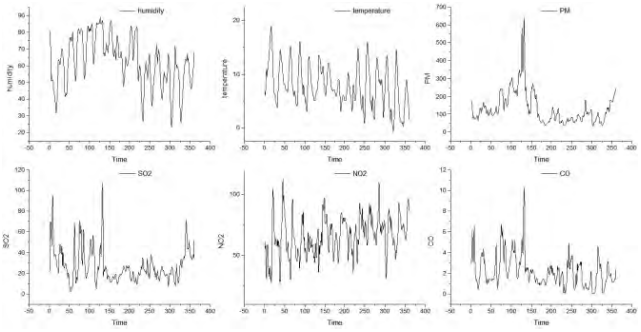

FIGURE 7. Environmental information data line chart.

IV. RESULTS

We analyzed data collected in November 2018. The raw data of the temperature calibration, humidity calibration, and gas-sensitive characterization are shown in Figure 7. We applied a band-pass Butterworth filter on the sampled data to eliminate out-band interference. The lower and upper cutoff frequency was set as 1Hz and 2Hz, respectively (the filter order was 4). After filtering, the extreme data were deter-mined by finding the peaks, as shown in Figure 8.

In Figure 7, it can be found that PM2.5 and CO and NO_2 are ''same high and low'', which further indicates that the

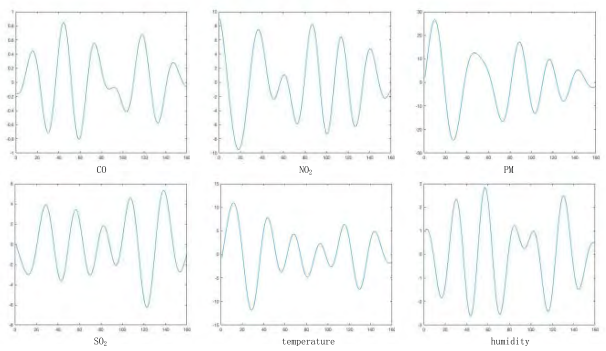

FIGURE 8. Time-domain chart.

correlation is very high. In general, it is easy to understand PM2.5 has the same high and low as PM10, but has the same high and low with CO, NO_2 need analysis. These two air pollutants are mainly related to automobile exhaust emissions and industrial emissions, the city already has strict restrictions. Therefore, it can only be explained that the possibility of industrial pollution is the greatest, combined with the above figure, which shows that urban pollution is related to industrial pollution.

The concept of the individual sensor nodes was based on the use of micromachined sensor technology and internet connectivity in order establish a wireless network for online, in-situ factory environmental monitoring. The CO concentration was determined by listening to its concentration via the photoacoustic effect. Both temperature and humidity were determined using state-of-the-art microtechnology.

Due to the high complexity of the environmental information monitored, the data curve showed both a sharp rise and a sharp decline, and the curve showed many burrs. Because the results exhibited almost no regularity compared with the original data, we used a Butterworth filter to process the existing data.

The fast Fourier transform of the sample data was carried out with a band-pass Butterworth filter to generate the time-domain chart of the sample data.

By filtering the CO, SO_2, NO_2, PM, temperature, and humidity data of the frequency domain chart and generating a time-domain chart, we eliminated the noise in the data and found better environmental monitoring information data in the November extremum position. The results showed that the pollutants' extremum position was concentrated roughly on the same date, whereas the temperature and humidity of the extremum position and location were almost at the opposite extreme of the pollutant concentration. This rise of temperature and humidity from pollutant diffusion and dilution played a positive role, a phenomenon that a curve similarity calculation also verified. Statistics for the extreme values are shown in Tables 2 and 3.

The number of extreme points in each index was counted, and the number of days with increased pollution was found to be approximately the same by counting the number of maximum points.

TABLE 2. Extreme point count.

Index	Count
CO	6
NO_2	5
PM	5
SO_2	5
Temperature	7
Humidity	6

TABLE 3. Extreme point statistics.

Number index	CO		NO_2		PM		SO_2		Temperature		Humidity	
	Points	Value	Points	Value	Points	Value	Points	Value	Points	Value	Points	Value
1	16	1.08	37	47	29	40	10	80	3	7.13	12	45.73
2	45	1.23	61	45	57	6	47	102	30	4.27	44	43.13
3	73	3.22	87	40	82	28	89	147	57	7	68	66.57
4	92	1.5	114	59	107	37	117	273	85	12.03	93	72.97
5	118	2.17	141	61	138	22	143	121	102	4.6	116	81.57
6	147	1.7							130	7	16	70.23
7									158	11.13		

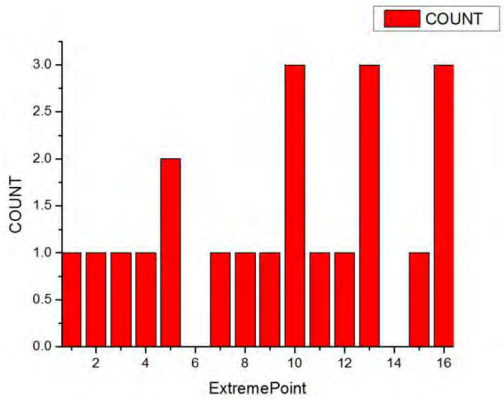

FIGURE 9. Extreme point statistics chart.

According to the specific location and data statistics of the extreme point of environmental information, it can be seen from the information in the table that the time period of pollution rise mainly occurred in the four frequency domains, namely [10-30], [40-50], [70-90] and [115-150].

According to the statistical data in Tables 2 and 3, the pollution time points were statistically analyzed through the histogram in Figure 9, which reflects the time area where the concentration of pollution increased in the urban monitoring jurisdiction. By comparing the statistical chart information in the figure with the statistical chart of factory production

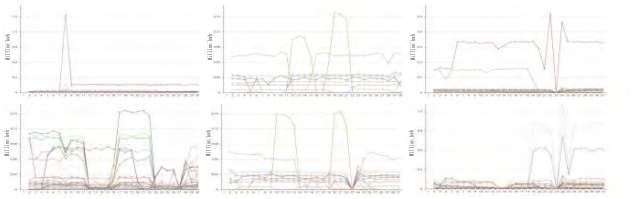

FIGURE 10. Line chart on monitoring data on production energy consumption.

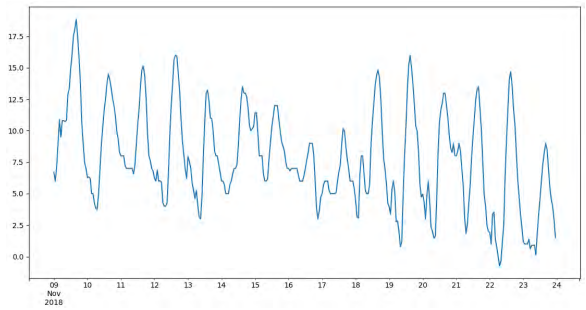

FIGURE 11. Temperature data curves after screening and interpolation. The moving average method was used to smooth the curves.

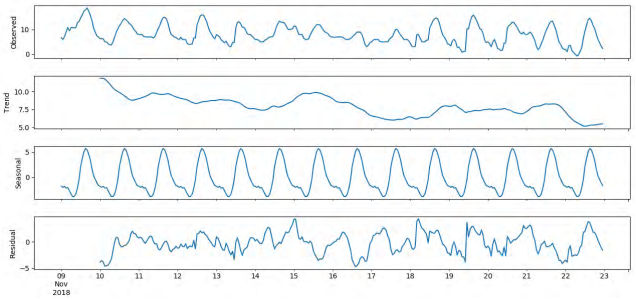

FIGURE 12. Periodic decomposition of temperature curves.

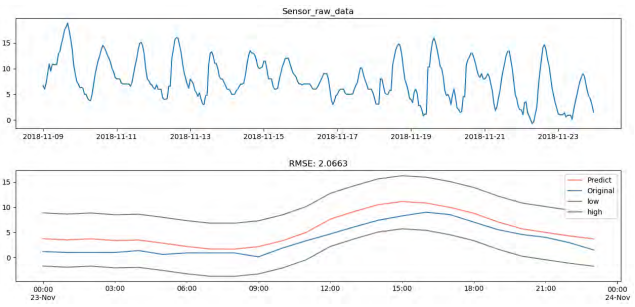

FIGURE 13. The test results from a certain day's data in the test set, where red is the prediction curve and blue is the original curve, and the cosine similarity of the two curves is 95.8%.

electricity in Figure 10, we found that the time area of increased pollution was consistent with the time area of increased plant production, and that the shapes of the two peaks were almost similar. Therefore, we could determine the effectiveness of the sensor node network, and the potential correlation between factory production and urban pollution, through the correlation between the two, and we found that

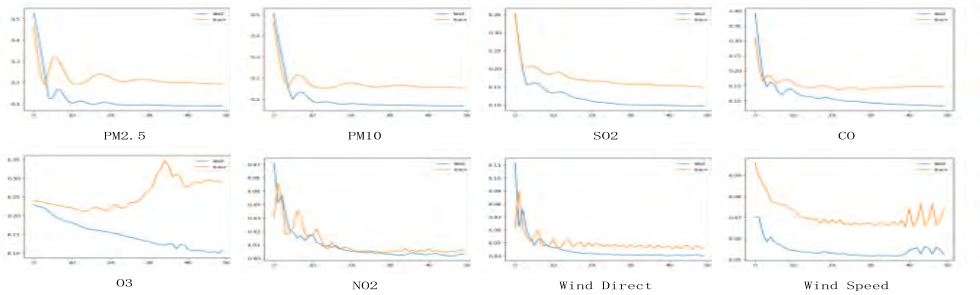

FIGURE 14. The root mean square error of the iterative process of eight environmental indicators, output is made every ten steps.

factory pollution had a great impact on the increase in urban pollution.

The histogram of the position of the time point shows the extreme points of CO, NO_2, PM, and SO_2. In the figure, the main concentrated areas reflecting increases in pollution are divided into four time periods. After the inverse transformation in the frequency domain was converted into the time domain, the middle points of these four time intervals were roughly the 9th, 18th, 24th, and 30th days.

By monitoring the electricity consumption during six production processes of typical sewage discharge factories in the jurisdiction, it was found that the major production periods of large sewage discharge enterprises in November 2018 were concentrated in the following four sections: [4-12], [15-23], [24-25] and [27-30]. This performance was consistent with the time section when the concentration of pollution factors was increased in our urban environmental information monitoring analysis. Therefore, our proposed wireless sensor network was found to accurately reflect the changes in environmental pollution in the city and to determine the potential laws affecting factory production and urban pollution changes.

Through environmental data modeling and the study of historical data, we studied periodic trends in environmental factors. To forecast the future trends in pollution in the short term, we used the temperature condition. For example, we first observed the temperature history curve in Figure 11. We used the box-plot method to remove abnormal data from the corresponding attribute. We observed that the cyclical temperature curve was rough, and increased suddenly and sharply decreased scenario existed in the temperature curve. We also observed some abnormal cycles. The general cycle was high during the day, low in the morning and evening, and lowest at night. We divided the data into an 7:3 ratio for training and testing.

The first part is the original data curve, the second part is the trend curve of data decomposition, the third part is the data cycle curve, and the fourth part is the residual part of the data.

As can be seen from the second part of Figure 12, since November 2018, the temperature gradually decreased as the date increased, which was consistent with the temperate

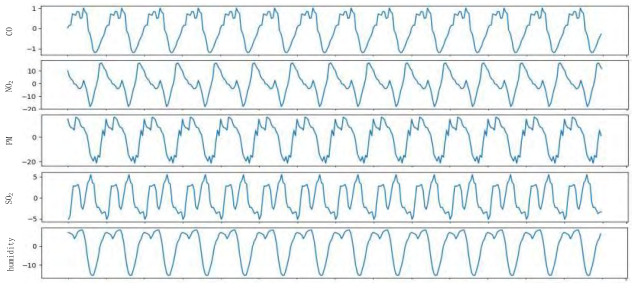

FIGURE 15. Curves of CO, NO_2, PM, SO_2, and humidity cycle.

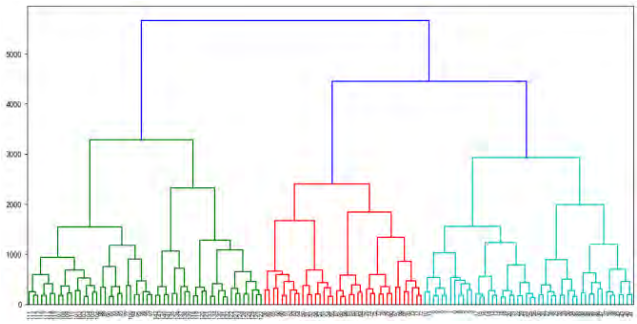

FIGURE 16. Hierarchical clustering results of pollution cases.

continental monsoon climate characteristics of this northern Chinese city. As the third part depicts, the periodicity of the temperature curve conformed to our previous conjecture and also conformed to the characteristics of urban daytime temperature.

In Figure 13, the red curve is the prediction curve, the blue curve is the original curve, the two gray curves are the predicted threshold region, and the confidence interval is the root mean square error. If the true curve exceeded the confidence interval, the abrupt position would be judged as abnormal pollution, and the alarm would be given. The periodic curves of the other elements are shown in Figure 15.

We obtained the pollution cases in the dataset through hierarchical clustering, and we conducted a detailed analysis on one of them. As is shown in Figure 16.

We interpret the pollution trend of wu'an through a visualization method of thermal map. The hourly air quality change

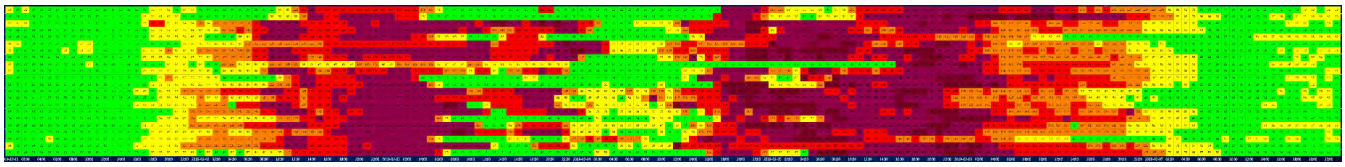

FIGURE 17. Thermoeleric chart of pollution diffusion.

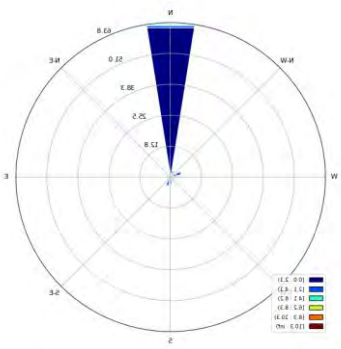

FIGURE 18. Statistical chart of wind direction in winter.

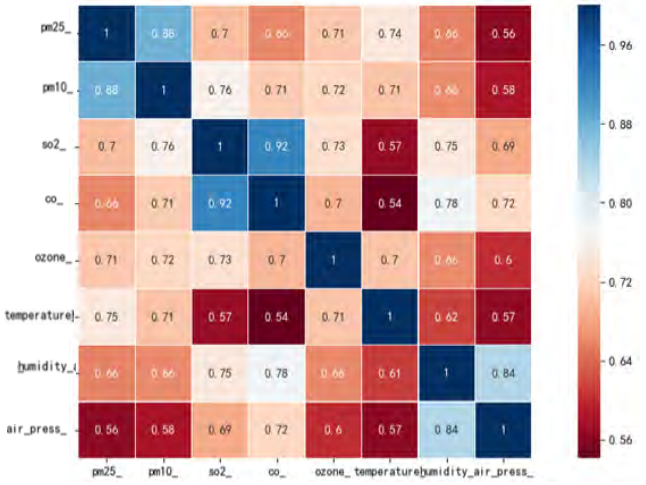

FIGURE 19. Correlation analysis chart of environmental indicators.

thermal map from November 2018 to January 2019 is a time series from left to right, the city is arranged from north to south according to the dimension, and it can be seen from the figure that the pollution trend is approximately oblique diagonal diffusion. Then the change in dark color in the figure is roughly equivalent to the pollution diffusion route. As is shown in Figure 17 that the two major over standard pollution incidents, the southern urban area mainly continues to be severe pollution, and the farther south, the higher pollution, especially in the second pollution incident, there is a serious pollution in the south. That is to say, the pollution has obvious geographical diffusion characteristic, and is roughly ''oblique diagonal'' diffusion. The impact of urban external pollution transmission is small.

There is a great correlation between air pollution and meteorology. This is the meteorological data from November 2018 to January 2019, which is the main distribution month of the city in winter. It can be seen from the wind direction and wind speed statistical chart that during this period, is mainly the north wind of 0-2.1m, no obvious strong wind. This is consistent with the temperate continental monsoon climate of wu'an and the geographical information of northerly winds and northwest winds in winter. However, as shown in Figure 18, pollutants of pollution exceeding the standard twice are discharged from the southern region. Therefore, the north wind that often rises in winter hindering the diffusion of pollutants to some extent makes the pollution gas mainly concentrate in the southern region.

The Figure 19 above is the correlation analysis chart of various pollutions. The possible pollution sources are inferred by analyzing the correlation between main pollutants in the air, and because PM2.5 and PM10 are both particle pollution, only the radius is different, the correlation is very high, but there is no great practical significance in this study. SO_2 is mainly derived from the desulfurization link of the steel plant and automobile exhaust gas in the pollution source given by the local environmental protection department. CO mainly comes from the coal combustion of factories, heating department and automobile exhaust. While wu'an carry out vehicle restrictions, has little impact on automobile exhaust, so this time the pollution lock for factory emissions.

V. CONCLUSIONS AND FUTURE WORKS

Accurate monitoring and prediction of pollution and electricity consumption by enterprises can greatly relieve the pressure of urban pollution. Our study showed that in factories without active pollution monitoring, prediction, and an enterprise production monitoring system, it was very difficult to restrict the production of enterprises in the city and to forecast and alarm the pollution so as to improve the quality of urban environment. Based on our results, we concluded that in the real world, these pollutants would not only accumulate and spread in a city, but if a linkage were to form between several key polluted cities, the air quality in the corresponding region would become worse and worse, which would have a huge impact on the health of citizens living in the region. Cities, therefore, must take the initiative to monitor air quality and maintain a good quality of urban air environment. We also concluded that different degrees of restrictions on the production capacity of manufacturing

industry can effectively be used to curb the source of pollution without affecting the output value within the jurisdiction. Wireless sensor networks, a technology that is suitable for providing data on a variety of time and space scales, are the cornerstone of this work. However, these networks are usually limited by the cost of each individual sensor node. Our low-cost, reliable, and selective gas sensor nodes have been shown to provide a feasible alternative to these networks.

The limitations of our present work are also prospected as follows: (1) Improve the efficiency of data quality processing. In the process of data quality processing in complex environment background, some methods existing queuing or even conflict phenomenon, which leads to repeated interaction with the database. In the following work, we hope to propose a classification method to improve the efficiency of data quality processing by constructing multiple classifiers to simultaneously segment, identify and process data quality problems of time series data. (2) We cannot avoid the cold start problem when adding new factories or cities. Therefore, it is hoped to minimize the impact of this problem by transfer learning or data generation method, so as to reduce the data accumulation time when the system is formally applied. However, due to the differences in geography and meteorology between cities and the randomness of local meteorological changes, it is difficult to achieve high accuracy. Therefore, how to carry out high-precision data transfer will become one of our next research priorities. (3) We hope to further study the correlation between individual monitoring stations and the overall environmental system, and research on ecosystem level from the overall urban or intercity to further improve the accuracy of our air quality monitoring.

REFERENCES

[1] *Global Environmental Change*, World Health Org., Geneva, Switzerland, 2005.
[2] *Impact Fuel Changes Exhaust Evaporative Emissions*, Urban Air Quality Management Project, Ministry Environ. Natural Resour., Sri Lanka, Indian, May 2003, pp. 75–85.
[3] *Chemical Hazards*, World Health Org., Geneva, Switzerland, 2005.
[4] X. Luo and J. Yang, "A survey on pollution monitoring using sensor networks in environment protection," *J. Sensors*, vol. 2019, Jan. 2019, Art. no. 6271206.
[5] X. Luo and J. Yang, "Problems and challenges in water pollution monitoring and water pollution source localization using sensor networks," in *Proc. Chin. Autom. Congr. (CAC)*, Jinan, China, Oct. 2017, pp. 5834–5838.
[6] C. Sun, Y. Yu, V. O. K. Li, and J. C. K. Lam, "Multi-type sensor placements in Gaussian spatial fields for environmental monitoring," *Sensors*, vol. 19, no. 1, p. 189, Jan. 2019.
[7] Q. Han, B. Shao, L. Li, Z. Ma, H. Zhang, and X. Du, "Publishing histograms with outliers under data differential privacy," *Secur. Commun. Netw.*, vol. 9, no. 14, pp. 2313–2322, 2016.
[8] C. Sun, Y. Yu, V. O. Li, and J. C. K. Lam, "Optimal multi-type sensor placements in Gaussian spatial fields for environmental monitoring," in *Proc. 4th Int. Conf. Smart Cities*, Kansas City, MO, USA, Sep. 2018, pp. 420–429.
[9] N. Castell, F. R. Dauge, P. Schneider, M. Vogt, U. Lerner, B. Fishbain, D. Broday, and A. Bartonova, "Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates?" *Environ. Int.* vol. 99, pp. 293–302, Feb. 2017.
[10] J. Mitrovics, "Smart sensors for air quality monitoring: Concepts and new developments," in *Proc. IEEE Sensors*, Oct./Nov. 2017, pp. 1–2.
[11] J. C. Chow, "Measurement methods to determine compliance with ambient air quality standards for suspended particles," *J. Air Waste Manage. Assoc.*, vol. 45, no. 5, pp. 320–382, 1995.
[12] A. Abrardo and A. Pozzebon, "A multi-hop LoRa linear sensor network for the monitoring of underground environments: The case of the medieval aqueducts in Siena, Italy," *Sensors*, vol. 19, no. 2, p. 402, Jan. 2019.
[13] Q. Han, S. Liang, and H. Zhang, "Mobile cloud sensing, big data, and 5G networks make an intelligent and smart world," *IEEE Netw.*, vol. 29, no. 2, pp. 40–45, Mar./Apr. 2015.
[14] K. Zhang, Q. Han, and Z. Cai, "RiPPAS: A ring-based privacy-preserving aggregation scheme in wireless sensor networks," *Sensors*, vol. 17, no. 2, p. 300, 2017.
[15] A. Mecocci, G. Peruzzi, A. Pozzebon, and P. Vaccarella, "Architecture of a hydroelectrically powered wireless sensor node for underground environmental monitoring," *IET Wireless Sensor Syst.* vol. 7, no. 5, pp. 123–129, 2017.
[16] C. M. Imran, M. Aldukhail, N. Almezeini, and M. Alnuem, "Potential applications of linear wireless sensor networks: A survey," *Int. J. Comput. Netw. Commun. Secur.*, vol. 4, pp. 183–200, Jun. 2016.
[17] Z. He, Z. Cai, Q. Han, W. Tong, L. Sun, and Y. Li, "An energy efficient privacy-preserving content sharing scheme in mobile social networks," *Pers. Ubiquitous Comput.*, vol. 20, no. 5, pp. 833–846, 2016.
[18] Q. Han, Q. Chen, L. Zhang, and K. Zhang, "HRR: A data cleaning approach preserving local differential privacy," *Int. J. Distrib. Sensor Netw.*, vol. 14, no. 12, Dec. 2018, Art. no. 1550147718819938.
[19] Y. Liang, Z. Cai, J. Yu, Q. Han, and Y. Li, "Deep learning based inference of private information using embedded sensors in smart devices," *IEEE Netw.*, vol. 32, no. 4, pp. 8–14, Jul./Aug. 2018.
[20] Q. Han, Z. Xiong, and K. Zhang, "Research on trajectory data releasing method via differential privacy based on spatial partition," *Secur. Commun. Netw.*, vol. 2018, Nov. 2018, Art. no. 4248092.
[21] Q. Han, D. Lu, K. Zhang, X. Du, and M. Guizani, "Lclean: A plausible approach to individual trajectory data sanitization," *IEEE Access*, vol. 6, pp. 30110–30116, 2018.
[22] G. Marques and R. Pitarma, "An indoor monitoring system for ambient assisted living based on Internet of Things architecture," *Int. J. Environ. Res. Public Health*, vol. 13, no. 11, p. 1152, 2016.
[23] F. Salamone, L. Belussi, L. Danza, T. Galanos, M. Ghellere, and I. Meroni, "Design and development of a nearable wireless system to control indoor air quality and indoor lighting quality," *Sensors*, vol. 17, no. 5, p. 1021, 2017.
[24] P. Srivatsa and A. Pandhare, "Indoor air quality: IoT solution," *Int. J. Res. Advent Technol.*, vol. 19, no. 4, pp. 218–220, Mar. 2016.
[25] V. Wittstock, L. Scholz, B. Bierer, A. O. Perez, J. Wöllenstein, and S. Palzer, "Design of a LED-based sensor for monitoring the lower explosion limit of methane," *Sens. Actuators B, Chem.*, vol. 247, pp. 930–939, Aug. 2017.
[26] C. J. Wong, M. Z. MatJafri, K. Abdullah, H. S. Lim, and K. L. Low, "Development of air quality monitoring remote sensor using a digital SLR camera," in *Proc. IEEE Aerosp. Conf.*, Mar. 2008, pp. 1–6.
[27] S. G. Narasimhan and S. K. Nayar, "Contrast restoration of weather degraded images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 1, no. 6, pp. 713–724, Jun. 2003.
[28] C. J. Wong, M. Z. MatJafri, K. Abdullah, and H. S. Lim, "Temporal and spatial air quality monitoring using Internet surveillance camera and ALOS satellite image," in *Proc. IEEE Aerosp. Conf.*, Mar. 2009, pp. 1–7.
[29] H. Wang, X. Yuan, X. Wang, Y. Zhang, and Q. Dai, "Real-time air quality estimation based on color image processing," in *Proc. IEEE Vis. Commun. Image Process. Conf.*, Dec. 2015, pp. 326–329.
[30] L. Scholz, A. O. Perez, B. Bierer, P. Eaksen, J. Wöllenstein, and S. Palzer, "Miniature low-cost carbon dioxide sensor for mobile devices," *IEEE Sensors J.*, vol. 17, no. 9, pp. 2889–2895, May 2017.
[31] S. Knobelspies, B. Bierer, A. O. Perez, J. Wöllenstein, J. Kneer, and S. Palzer, "Low-cost gas sensing system for the reliable and precise measurement of methane, carbon dioxide and hydrogen sulfide in natural gas and biomethane," *Sens. Actuators B, Chem.* vol. 236, pp. 885–892, Nov. 2016.
[32] G. J. Papadopoulos and G. L. R. Mair, "Amplitude and phase study of the photoacoustic effect," *J. Phys. D, Appl. Phys.*, vol. 25, no. 4, pp. 722–726, 1992.
[33] J. Kneer, A. Eberhardt, P. Walden, A. O. Pérez, J. Wöllenstein, and S. Palzer, "Apparatus to characterize gas sensor response under real-world conditions in the lab," *Rev. Sci. Instrum.*, vol. 85, May 2014, Art. no. 055006.

[34] A. Persily and L. de Jonge, "Carbon dioxide generation rates for building occupants," *Indoor Air* vol. 27, pp. 868–879, Sep. 2017.
[35] L. Scholz, A. O. Perez, B. Bierer, J. Wöllenstein, and S. Palzer "Gas sensors for climate research," in *Proc. 19th ITG/GMA-Symp. Sensors Measuring Syst.*, Jun. 2018, pp. 1–4.
[36] M. Campolo, A. Soldati, and P. Andreussi, "Forecasting river flow rate during low-flow periods using neural networks," *Water Resour. Res.*, vol. 35, no. 11, pp. 3547–3552, 1999.
[37] W. Xin *et al.*, "Application of singular spectrum analysis to failure time series analysis," *J. Beijing Univ. Aeronaut. Astronaut.*, vol. 42, no. 11, pp. 2321–2331, 2016.
[38] R. Li and R. Kang, "Research on failure rate forecasting method based on ARMA model," *Syst. Eng. Electron.*, vol. 30, no. 8, pp. 1588–1591, 2008.
[39] S. Rocco and M. Claudio, "Singular spectrum analysis and forecasting of failure time series," *Rel. Eng. Syst. Saf.*, vol. 114, pp. 126–136, Jun. 2013.
[40] M. das Chagas Moura, E. Zio, I. D. Lins, and E. Droguett, "Failure and reliability prediction by support vector machines regression of time series data," *Rel. Eng. Syst. Saf.*, vol. 96, no. 11, pp. 1527–1534, 2017.
[41] K. Xu, M. Xie, L. C. Tang, and S. L. Ho, "Application of neural networks in forecasting engine systems reliability," *Appl. Soft Comput.*, vol. 2, no. 4, pp. 255–268, 2003.
[42] X. Wang, J. Wu, C. Liu, S. Wang, and W. Niu, "A hybrid model based on singular spectrum analysis and support vector machines regression for failure time series prediction," *Qual. Rel. Eng. Int.*, vol. 32, no. 2, pp. 2717–2738, 2016.
[43] K. Greff, R. K. Srivastava, J. Koutnìk, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.
[44] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Netw.*, vol. 18, nos. 5–6, pp. 602–610, 2005.
[45] C. Y. Gizem, H. Mirisaee, G. P. Goswami, E. Gaussier, and A. Aït-Bachir, "Period-aware content attention RNNs for time series forecasting with missing values," *Neurocomputing*, vol. 312, pp. 177–186, Oct. 2018.
[46] U. D. Turdukulov, M.-J. Kraak, and C. A. Blok, "Designing a visual environment for exploration of time series of remote sensing data: In search for convective clouds," *Comput. Graph.*, vol. 31, no. 3, pp. 370–379, 2007.
[47] T. W. Wlodarczyk, "Overview of time series storage and processing in a cloud environment," in *Proc. IEEE Int. Conf. Cloud Comput. Technol. Sci.*, Dec. 2012, pp. 625–628.
[48] R. J. Fryer and M. D. Nicholson, "Assessing covariate-dependent contaminant time-series in themarine environment," *ICES J. Marine Sci.*, vol. 59, no. 3, pp. 1–14, 2002.
[49] Z. Hou, Z. Yu, and P. Shi, "Study on class of nonlinear time series models, and E. I. R. E. domain, "Study on a class of nonlinear time series models and ergodicity in random environment domain," *Math. Methods Oper. Res.*, vol. 61, no. 3, pp. 299–310, 2005.
[50] W.-S. Chung and S. Tohno, "A time-series energy input-output analysis for building an infrastructure for the energy and environment policy in South Korea," *Energy Environ.*, vol. 20, no. 6, pp. 875–899, 2016.
[51] T. Hoshiko, K. Yamamoto, T. Prueksasit, and F. Nakajima, "Time-series analysis of polycyclic aromatic hydrocarbons and vehicle exhaust in roadside air environment in Bangkok, Thailand," *Procedia Environ. Sci.*, vol. 4, no. 3, pp. 87–94, 2011.
[52] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio "Learning phrase representations using RNN encoder-decoder for statistical machine translation," Jun. 2014, *arXiv:1406.1078*. [Online]. Available: https://arxiv.org/abs/1406.1078