Bulgarian Chemical Communications, Special Edition F, (pp. 108 – 116) 2016

Chemical gas leakage source determination using distributed EM algorithm with Gaussian mixture model

Z. Yong[1,2], Z. Liyi[1,2*], W. Li[3], H. Jianfeng[2], B. Zhe[2]

[1]*School of Electronic Information Engineering, Tianjin University, Tianjin 300072, China*
[2]*The College of Information, Tianjin University of Commerce, Tianjin 300134, China*
[3]*The College of Information, Hebei University of Technology, Tianjin 300401, China*

Received February 12, 2016; Revised December 26, 2016

Chemical gas leakage source determination with sensor networks has become a research significance in the pollution environmental monitoring and security protection fields, which also known as gas leakage source parameters estimation. In this paper, we proposed a distributed EM algorithm for the chemical gas leakage source determination, and which was based on Gaussian Mixture model. Simulation results show that the proposed EM algorithm could determinate the gas leakage source localization and emission rate, Compare to the central EM algorithm, the distributed EM method was suggested because it can balance the accuracy performance and energy consumption in the sensor network, and it will get a significant reduction in the required numbers of sensor nodes and less energy to achieve the desired performance with less time, all of that was based on the dynamical adjusting scheme for computing sensor nodes selection.

Keywords: Chemical Gas Leakage Source Determination; Sensor Networks; Gaussian Mixture Model

INTRODUCTION

With the development of chemical industry, the hazardous chemical leakage, especially toxic gas leakage accidents occur occasionally. The leakage source's position and intensity determination has become the key problem to be solved urgently in the emergency rescue. Leakage source detection and determination generally means the process of discovering and tracking the spread of plume and finally determine the location and related parameters of the gas source, usually with "active" searching by sensor networks or mobile robots [1]. The research involves fundamental problems in information processing in sensor networks, detection and estimation, stochastic process, information entropy, artificial intelligence search and node routing planning in the field of information and automation [2-4]. And biomimetic olfaction and computational fluid dynamics and other studies are also related closely [5-7].

Scholars have made unremitting efforts and explorations on the study of leakage sources determination for many years, and achieved some research results. The results were mainly based on the stochastic process theory and probabilistic statistical estimation method, in which, the basic idea is to estimate the occurrence probability of the related leakage accidents. The accident occurred probability determination problem in the designated location, also generally known as dangerous leakage reconstruction [8,9]. The common methods such as Bayesian inference [10], minimum relative entropy (MRE) [11,12] and statistic induction were often used in the reconstruction of leakage accident inversion. Bayesian inference is the main research method of leakage accident reconstruction and source parameter determination.

At first, the Bayesian inference method makes use of the prior information of the likelihood function and the parameter to get the posterior probability distribution based on the known prior probability distribution. Secondly, the measurement result should be obtained to fit the posterior probability density function distribution, and finally the estimated value of the parameters to be obtained by sampling method. The current research mainly combines the Bayesian inference method with the stochastic Monte Carlo sampling method (MC) or the Markov chain Monte Carlo sampling method (MCMC) to achieve the estimation of source parameters [13,14].

Monte Carlo method was usually easy to converge to local optimal solution, especially when the initial value selected far away from the true value. The limitation is more serious in large-scale spatial hazardous chemical leakage determination studies, which will increase the difficulty of the method [15]. However, when the Bayesian inference method combining with the Monte Carlo method or the Markov chain Monte Carlo method with (the former is abbreviated as BMC, the latter is BMCMC), the probability distribution of iterative updating

* To whom all correspondence should be sent:
E-mail: zhangyong@tjcu.edu.cn

parameters could overcome the Monte Carlo method insufficient. The Markov chain Monte Carlo method (MCMC) extends the application of Bayesian inference to parameter inverse calculation. The Markov chain can be obtained by means of random search, so that the limit distribution of the Markov chain is the posterior probability density function [16,17]. Markov chains of sufficient length can guarantee that the sampling results close to the posterior distribution. Senocak [18], Chow[19] and Kosovic [20] realized that the combination of Bayesian inference and MCMC method can estimate the parameters of the release source with a set of concentration observations, and the parameters estimation results will reached based on the obtain posterior probability density function of the source parameters (position and intensity) by MCMC sampling. Since MCMC sampling usually takes several thousand iterations to converge to the posterior probability distribution, it is computationally intensive and computationally time-consuming, which is usually not sufficient in an accident contingency.

Keats[21-23],Yee[24] used the convective diffusion equation with the MCMC method to describe the source-receiver association to determine the likelihood function, which resulted in a significant increase in computational efficiency.

Keats [21] used BMCMC method to estimate the location and intensity of the source in a complex urban environment. The results show that the convective diffusion equation and the MCMC sampling method can be used to obtain accurate values of four parameters within a reasonable time. Keats[22] combine the inverse Lagrangian stochastic model with Eulerian type concomitant convection-diffusion equation further, and the source parameters can also be solved in BMCMC under non-conservation conditions quickly. Yee [24] established an adjoint model for Eulerian and Lagrangian type diffusion equations, respectively, in which, the concentration distribution can be calculated directly when the distribution of the source is given, and the computational efficiency can be improved significantly. Yee [25] deduces the source location and intensity with unknown number of leak sources. Guo [29] used an unsteady concomitant transfer model and an advanced numerical scheme (finite volume method) based on adaptive mesh encryption to perform the reconstruction of leakage source in a three-dimensional urban environment. Numerical results show that the application of the unsteady state transfer equation and the MCMC method is very effective, and the introduction of the non-stationary inversion method can significantly improve the accuracy of the leakage source location in the wind direction.

Because of the time-varying nature of the sensor data and the instability of the initial concentration field, it is necessary to perform a real-time determination method for the leakage source. For example, Johannesson [27] proposed a sequential Monte Carlo (SMC), The Monte Carlo (SMCMC) method to inverse the unstable dynamic system, which further extends the application of Bayesian inference method in the leakage source parameters determination problem. Chinese scholars were also use probability and statistics methods to fulfill the leakage source determination related research. Zhu [28] proposed a method based on Bayesian estimation theory, the ensemble Kalman smoothing and Kalman filtering method for the inverse problem of leakage source. Guo [29] used Bayesian inference combined with Markov chain Monte Carlo sampling method to calculate the gas source in urban area.

The Bayesian-Monte Carlo method has been widely used in the study of leakage source determination of various scales, but it needs to know the prior distribution of parameters at first, and the sampling process of parameter posterior distribution is extremely time consuming, so that the computational efficiency of the determination algorithm should be improved in the event of an happened emergency [30].

The inverse theory has been widely used in the source determination research of groundwater sources [31], earthquake sources [32], sound sources [33], heat sources [34]. In addition, most of the current research has been focused on the application of methods, and the analysis of diffusion patterns and its impact on the real-time and applicability of leakage source determination is seldom involved. Most of them used the static plume Gaussian model with a constant flow conditions, however, the actual flow is more time-varying and dynamic, and the static Gaussian model also has some limitations[35,36], so the real time performance was usually not meet the practical requirements.

Therefore, in this paper, we propose an integrated method for the gas leakage sources rapid determination problem based on the information processing technology of sensor network and the theoretical analysis of mixed Gaussian model. The proposed distributed EM algorithm with Gaussian Mixture models was considered for the distributed determination implementation because of the highly nonlinear diffusion model and the heavily noise corrupted sensor node's measurements. On the other hand, we also gave a computing sensor nodes modification method for the estimation performance improving and the energy consumption reducing.

The following structure: Section 2, the problem description is given. we analyzed the gas leakage diffusion Gaussian Mixture Model, proposed an distributed EM algorithm for chemical gas leakage source determination. In section 3, we compare the distributed EM method with the central EM method and analyze the simulation results. Section 4, conclusion.

PROBLEM DESCRIPTION

The gas leakage concentration information measured by sensor nodes in the sensor network is generally consistent with a diffusion model. Gaussian model and the model based on turbulent diffusion theory are usually used in the existing gas leakage source determination. Gas leakage concentration diffusion model usually can be described as a stochastic process and the source determination problem can also be known as a gas leakage diffusion model reconstruction problem.

In this paper, we assume that the flow environment is consistent with the Gaussian model distribution and based on the following application of turbulent gas diffusion model.

(1) The positive direction of the x axis is considered as the direction of the wind direction, without considering the obstruction of the obstruction and other effects, assuming that the flow environment is a stable and uniform airflow field;

(2) The main study is to determinate a gas source parameters of source coordinates $\boldsymbol{r}_s=(x_s,y_s)$ and the estimated value $\hat{q}$ of the source emission rate

(3) A rate of gas release from the gas source;

(4) N sensor nodes were used with a simple dynamic topology distributed in the square area, the location of each node itself was known.

The main goal is to design a determination method to achieve the gas leakage source parameters vector $\boldsymbol{\theta}$ estimation via stochastic process information processing and the estimator defined as:

$$\hat{\boldsymbol{\theta}}=[\hat{x}_s,\hat{y}_s,\hat{q}]$$

Where $[\hat{x}_s,\hat{y}_s]$ means the estimator of source coordinates, and $\hat{q}$ is the source emission rate estimator.

Gas Leakage Diffusion Model

In this paper, an approximate analytic model that adapted proposed by Ishida [10] was used:

$$c(\boldsymbol{r}_i,t)=\frac{q}{2\pi K}\frac{1}{d}\exp\left[-\frac{U}{2K}(d-\Delta x)\right] \qquad (1)$$

Where $c(\boldsymbol{r}_i,t)$ is the concentration at $\boldsymbol{r}_i=(x_i,y_i)$ and time t; q is the gas diffusion rate; K is the turbulent diffusion coefficient; U is the wind speed; $d=\sqrt{(x_i-x_s)^2+(y_i-y_s)^2}$ is the distance from any point to the gas source, and $\boldsymbol{r}_i=(x_i,y_i)$ is the current sensor node position.

$\Delta x=(x_s-x_i)\cos\alpha+(y_s-y_i)\sin\alpha$, α is the upwind direction angle with x-axis.

If we take the x-axis as the downwind direction, the equation (1) can be rewritten as:

$$c(\boldsymbol{r}_i,t)=\frac{q}{2\pi K}\frac{1}{x_i-x_s}\exp\left[-\frac{U}{2K}(d-(x_i-x_s))\right] \qquad (2)$$

Figure 2 shows the two-dimensional gas diffusion model under different wind speed and direction. Since the static model based on turbulent diffusion theory introduces the wind direction consideration, and more in line with the simulated environmental conditions. So in this paper, the gas source determination algorithm is mainly based on the Gaussian model and its Mixture models.

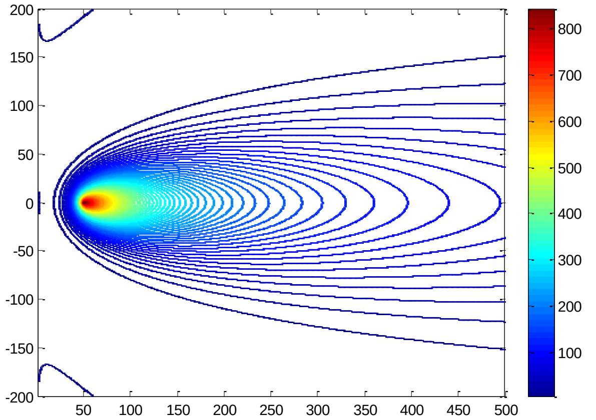


Fig. 1 Gas concentration diffusion in a 400m×500m rectangle area

Figure 1 shows the gas diffusion with equation (2), the gas leakage source location is (50, 0) m, the color bars stand for the diffusion concentration (unit is ppm). And figure 2 gives the diffusion with different wind speed and direction in the field.

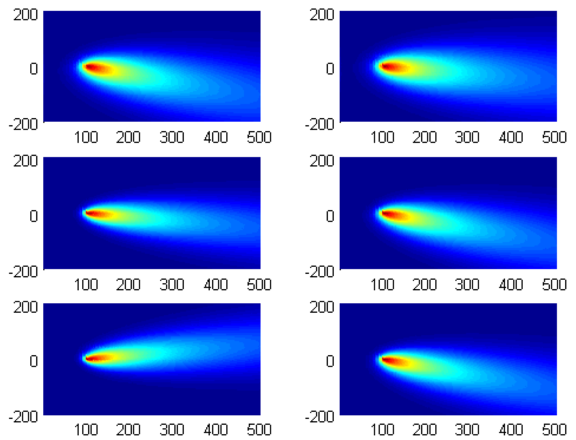


Fig. 2 Diffusion shape with different direction and speed

Observation Model

Based on the above gas leakage diffusion model, the measurement model at each sensor node is defined as following [8-9]:

$$z(\boldsymbol{r}_i, t) = c(\boldsymbol{r}_i, t) + b + e(\boldsymbol{r}_i, t) \quad (3)$$

Where $c(\boldsymbol{r}_i, t)$ is the chemical source concentration; b is a unknown constant bias term; and $e(\boldsymbol{r}_i, t)$ is the sensor's measurement noise subject to Gaussian distribution. Denoting $z_i(t) = y(\boldsymbol{r}_i, t)$, $e_i(t) = e(\boldsymbol{r}_i, t)$, and $\lambda(t) h_i(\boldsymbol{\theta}) = c(\boldsymbol{r}_i, t)$, we can rewrite (3) as

$$z_i(t) = \boldsymbol{H}_i(\boldsymbol{\theta}) \boldsymbol{x}(t) + e_i(t) \quad (4)$$

Where $\boldsymbol{H}_i(\boldsymbol{\theta}) = [h_i(\boldsymbol{\theta}), 1]$, $\boldsymbol{x} = [\lambda(t), b]^T$, and $\boldsymbol{\theta} = [x_s, y_s, q]$ represent the gas source parameters.

In order to use Bayesian inference method for the gas leakage source determination, we convert the space-time gas diffusion model into a discrete state space form, in which, the gas leakage diffusion process described by two equations and the observation state according to a classical Gauss-Markov model. The space-time continuous $c(\boldsymbol{r}_i, t)$ represented by a finite state vector $\boldsymbol{x}_k$ at discrete time t_k, the system equation is as follow:

$$\boldsymbol{x}_k = \boldsymbol{A}_k(\boldsymbol{\theta}) \boldsymbol{x}_{k-1} + \boldsymbol{B}_k(\boldsymbol{\theta}) \boldsymbol{w}_k \quad (5)$$

Where $\boldsymbol{A}_k(\boldsymbol{\theta})$ represent the system matrix, $\boldsymbol{B}_k(\boldsymbol{\theta})$ is the input matrix, $\boldsymbol{w}_k$ is the process noise with white zero-mean Gaussian distribution.

The observation equation:

$$\boldsymbol{z}_k^i = \boldsymbol{H}_k^i(\boldsymbol{\theta}) \boldsymbol{x}_k + \boldsymbol{v}_k^i \quad (6)$$

Where $\boldsymbol{z}_k^i$ is the observation of sensor node i at t_k , $\boldsymbol{H}_k^i(\boldsymbol{\theta})$ is the observation matrix, $\boldsymbol{v}_k^i$ is observation noise also with white zero-mean Gaussian distribution.

The noise $\boldsymbol{w}_k$ and $\boldsymbol{v}_k^i$ satisfy as following equations.

$$E[\boldsymbol{w}_k \boldsymbol{w}_l^T] = \boldsymbol{Q}_k \delta_{kl} \quad (7)$$

$$E[\boldsymbol{v}_k^i \boldsymbol{v}_l^{jT}] = \boldsymbol{R}_k^i \delta_{kl} \quad (8)$$

Where $\delta_{kl} = 1$, if $k = l$, and $\delta_{kl} = 0$ otherwise.

Gas Leakage Diffusion Gaussian Mixture Model

In the gas leakage accident area, the gas concentration measurement information is mainly composed of the gas diffusion noise and the wind turbulence noise with the sensor network. Through statistical analysis of the sensor observation, we can see that the diffusion noise usually follows the stationary Gaussian distribution, while the turbulence noise has strong nonlinearity, which could not be directly described by Gaussian distribution. In this paper, we assume that the gas diffusion model is a Gaussian Mixture Model(GGM), which composed of two different Gaussian distributions. The linear combination is expressed as follows:

$$\begin{aligned} f(\boldsymbol{z}; \boldsymbol{\theta}, \alpha_m) &= \sum_{m=1}^{2} \alpha_m p_m(\boldsymbol{z}; \boldsymbol{\theta}) \\ &= \alpha_1 p_1(\boldsymbol{z}; \boldsymbol{\theta}) + \alpha_2 p_2(\boldsymbol{z}; \boldsymbol{\theta}) \end{aligned} \quad (9)$$

Where m denotes that the model is composed of m different Gaussian distributions, $p_1(\boldsymbol{z}; \boldsymbol{\theta})$ denoting the diffusion noise subject to Gaussian distribution, α_1 and α_2 are the weighting coefficient, and satisfies $0 \le \alpha_1, \alpha_2 \le 1$， $\alpha_1 + \alpha_2 = 1$。

When $\alpha_2 = 0$, the GMM model can be expressed by Gaussian model, which indicating that there is no wind turbulence in the environment and the noise in the spill area is relatively stable.

When $\alpha_2 \neq 0$, it belongs to the GMM model, and with the α_2 increase, indicates that the environment has a strong effect of turbulence, which has weighting factors for the leakage source parameters determination.

The parameters to be determinate include the weighting factors α_1 , α_2 , and the leakage source parameters $\boldsymbol{\theta}$.

Distributed EM Algorithm

The observed environment information by sensor nodes does not contain all the diffusion and turbulence information, so the concentration information was an incomplete data set. For the imperfect data set, the most commonly used method is EM algorithm. The EM algorithm is an effective tool to estimate the maximum likelihood of the incomplete data set. The core of the method is to convert the complex problem of maximization of the likelihood function into a series of simple solutions of the expected value and the maximum value. Function optimization problem, which greatly reduces the computational complexity.

EM algorithm is essentially an iterative algorithm. It is composed of two steps for iteration: E step (Expectation) and M step (Maximization).

Assuming the observed data set $\boldsymbol{z} = \{z_1, z_2 \cdots z_n\}$, the posterior distribution $p(\boldsymbol{\theta} | \boldsymbol{z})$ with unknown parameters $\boldsymbol{\theta}$ is complex and difficult to directly perform statistical calculations. If it is assumed that some unobserved latent data $\boldsymbol{z}_c$ are known, then a simple posterior distribution $p(\boldsymbol{\theta} | \boldsymbol{z}, \boldsymbol{z}_c)$ may be obtained, which can be used for various statistical calculations with the simplicity, thus a complex maximization problem into a series simple maximization problem.

Let $\hat{\boldsymbol{\theta}}_k$ to be the k-th estimator of the maximum likelihood estimate of the representation $\boldsymbol{\theta}$, EM iterative algorithm can be achieved by the following two steps:

Step E: Determine the average log likelihood function for the complete data set $\boldsymbol{z}_c$:

$$Q\left(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}_k\right)=\int \log\ p\left(\boldsymbol{\theta}\middle|\boldsymbol{z},\boldsymbol{z}_c\right)p\left(\boldsymbol{z}_c\middle|\hat{\boldsymbol{\theta}}_k,\boldsymbol{z}\right)d\boldsymbol{z}_c \qquad (10)$$

Step M: Find the largest $\boldsymbol{\theta}$ with log-likelihood function of complete data

$$\hat{\boldsymbol{\theta}}_{k+1}=\arg\max_{\boldsymbol{\theta}} Q\left(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}_k\right) \qquad (11)$$

The E and M steps are iteratively calculated to obtain the estimated sequence $\hat{\boldsymbol{\theta}}_1,\hat{\boldsymbol{\theta}}_2,\cdots\hat{\boldsymbol{\theta}}_k$ of unknown parameters $\boldsymbol{\theta}$. The final convergence value of the sequence is the maximum likelihood estimator of the unknown parameter $\boldsymbol{\theta}$.

EM algorithm is an iterative algorithm to maximize the expected value, which can be divided into two categories: centralized EM algorithm and distributed EM algorithm. It mainly used in the following two cases:

(1) The observation data is not complete, which due to the limitations of the observation process;

(2) The likelihood function does not resolve, or likelihood function expression is too complex, which will result in maximum likelihood function of the traditional estimation method failure.

As for the second one, the centralized EM algorithm is usually used. Although the method can get more accurate results in the actual calculation, the entire sensor network carries a lot of data communications. Its communication bandwidth and energy have higher requirements, or easily lead to partial network paralysis, especially the sink node and its neighbor nodes. On the other hand, a large number of data traffic congestion will also increase the response time and reduce network efficiency. Therefore, in this paper, we introduce a distributed EM algorithm to effectively solve the problems.

Distributed EM algorithm is an improvement on the centralized EM algorithm, whose main purpose is to reduce the energy consumption of the whole network without affecting the positioning accuracy. The maximum energy consumption in sensor network was the communication cost, the number of nodes involved in the calculation and the inter-node traffic will be reduced with the use of distributed computing method, thus the entire network time life will extend. We proposed the distributed EM algorithm to estimate the maximum likelihood point to achieve the leakage source, which is based on the leak gas diffusion Gaussian Mixture Model. The algorithm can be divided into two steps: solving the maximum likelihood function and the distributed EM algorithm:

(1) The maximum likelihood function

The process of solving the log-likelihood function is described as follows:

$$p\left(\boldsymbol{z}\middle|\boldsymbol{\theta}\right)=\prod_{n=1}^{N}\left(\alpha_1 p_1\left(\boldsymbol{z}\middle|\boldsymbol{\theta}\right)+\alpha_2 p_2\left(\boldsymbol{z}\middle|\boldsymbol{\theta}\right)\right) \qquad (12)$$

take a logarithm on the formula (12):

$$\log\ p\left(\boldsymbol{z}\middle|\boldsymbol{\theta}\right)=\log\alpha_1^N+\sum_{n=1}^{N}p_1\left(\boldsymbol{z}_n\middle|\boldsymbol{\theta}\right)+\log\alpha_2^N+\sum_{n=1}^{N}p_2\left(\boldsymbol{z}_n\middle|\boldsymbol{\theta}\right) \qquad (13)$$

The estimated value of $\boldsymbol{\theta}$ can be get from the solution of Partial Differential Equation:

$$\frac{\partial\log\ p\left(\boldsymbol{z}\middle|\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}_i}=0 \qquad (14)$$

It is difficult to obtain the estimator of $\boldsymbol{\theta}$ directly from Eq.(14), so we use the distributed EM algorithm to estimate the unknown parameters in the GMM model. According to Eq. (12) and Eq. (13), the EM iterative algorithm for Gaussian Mixture model parameters is:

$$\hat{\alpha}_{m,i+1}=\frac{1}{N}\sum_{n=1}^{N}\alpha_i\left(\boldsymbol{\theta}_m\middle|\boldsymbol{z}_n\right)=\frac{1}{N}\sum_{n=1}^{N}\frac{\alpha_m p_m\left(\boldsymbol{z}_n\middle|\boldsymbol{\theta}_m\right)}{\sum_{m=1}^{2}\alpha_m p_m\left(\boldsymbol{z}_n\middle|\boldsymbol{\theta}_m\right)} \qquad (15)$$

$$\hat{\boldsymbol{\theta}}_{m,i+1}=\frac{\sum_{n=1}^{N}\alpha_i\left(\boldsymbol{\theta}_m\middle|\boldsymbol{z}_n\right)\boldsymbol{z}_n}{\sum_{n=1}^{N}\alpha_i\left(\boldsymbol{\theta}_m\middle|\boldsymbol{z}_n\right)} \qquad (16)$$

Where $\alpha\left(\theta_m\middle|\mathrm{z}\right)=\frac{\alpha_m p_m\left(\mathrm{z}\middle|\theta_m\right)}{\sum_{m=1}^{2}\alpha_m p_m\left(\mathrm{z}\middle|\theta_m\right)}$ is the posteriori probability based on Bayesian theory and $\alpha\left(\boldsymbol{\theta}_1\middle|\boldsymbol{z}\right)+\alpha\left(\boldsymbol{\theta}_2\middle|\boldsymbol{z}\right)=1$。

By choosing the reasonable initial value, the maximum likelihood estimation can be obtained using Eqs. (15) and (16). EM algorithm is essentially an iterative algorithm, so two issues should be considered in the estimation of parameters:

(1) the convergence of parameter estimation;

(2) the convergence rate of parameter estimation.

To ensure the convergence of the EM algorithm, it is necessary that the iteration estimate value of the parameters should ensure the likelihood function to be monotonic. The convergence rate of the EM algorithm is usually related to the allowable error level of the initial value and the estimated value. The chosen initial value α_1, α_2 is a uniformly distributed random number.

Since the observation noise set can be decomposed into the diffusion environmental noise and the turbulence noise, the weighting coefficient satisfies the condition $\alpha_1 + \alpha_2 = 1$.

Therefore, we can improve the iteration of the weighting coefficient by choosing the distribution of $\alpha_1 \sim U(0,1)$, and $\alpha_2 = 1 - \alpha_1$, for the mean and variance convergence rate improvement can be found in the literature [8], simulation shows that this simplification can effectively reduce the number of iterations and improve operational efficiency than traditional EM algorithm .

Distributed EM algorithm is different from the centralized EM algorithm, it does not require all the network nodes to participate in the calculation at the same time, but select the appropriate number of nodes involved in the operation according to the accuracy of the algorithm performance needed, the selected node is called running node set. In this paper, we assume that the N-1operation node sends $\hat{\theta}_{k-2}$ and $\hat{\theta}_{k-1}$ to node N by communication, and if θ does not converge, it moves on to the next cycle, otherwise, the calculation will stop when the convergence is out of the loop.

Because the EM algorithm has the convergence, as for the distributed EM algorithm, we can say that it also has the convergence. However, due to the limited computing number of nodes and the noise interference, the convergence may be slow in the real calculation. There are two solutions: the number of sensor nodes can be adjusted in real time to improve the convergence rate, such as the use of four nodes involved in the calculation, if not converged in the next cycle, we can use five or six nodes involved in the calculation. This computing nodes increasing method can effectively increase the convergence speed, but also it will increase the energy consumption of the entire network; the other method is the threshold adjusting method, which means that you can increase the threshold to end the calculation if the operation has not converged for a long time. In this paper, the former method is used to solve the slow convergence problem. From the view of communication, the node with information in distributed method only communicates with the neighbor node, while the centralized method needs to communicate all the location information to sink, the former has great advantage in reducing the communication energy consumption. In the simulation experiment, the convergence of the distributed method is obvious and fast, and the problem of the slow convergence is only theoretically analyzed.

It can be seen that distributed EM algorithm is an effective method for the determination of leakage source, which can reduce the energy consumption of the whole network while ensuring high positioning accuracy with great significance for practical applications.

SIMULATION RESULTS AND VALUATION

It is assumed that the Gaussian Mixture Model distribution is composed of N1 (0, 0.5) and N2 (0, 4) two Gaussian noise random variables, the mixed weighting coefficients are $\alpha_1 = 0.75$ and $\alpha_2 = 1 - 0.75 = 0.25$. Since no Gaussian mixture distribution function is provided in the Matlab tool box, we use the Bernoulli test method to generate the observed data of the GGM model according to Eq. (9). In this paper, the simulation data length is N = 128 and the number of iterations is 20. GGM model parameters, after 20 iterations with (16), the estimated values are $\alpha_1 = 0.7493$ and $\alpha_2 = 0.2507$, it is clear that we get a good mixture of the weighted coefficients of the estimated value by the EM estimation.

In order to generate realistic environmental concentration data of the 2-D sensor field , We designed a simulation environment with MATLAB and VC++. As shown in Figure 3 and in Figure 4 the wind speed and direction in the fluent field was given.

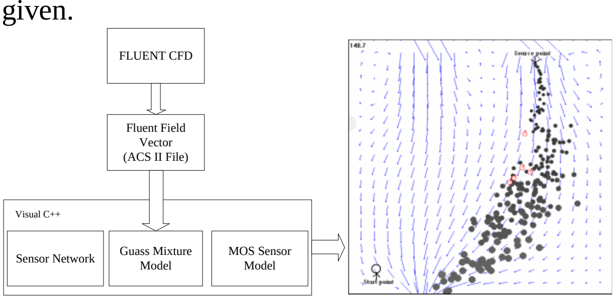


Fig. 3 The realistic environmental of 2-D sensor field

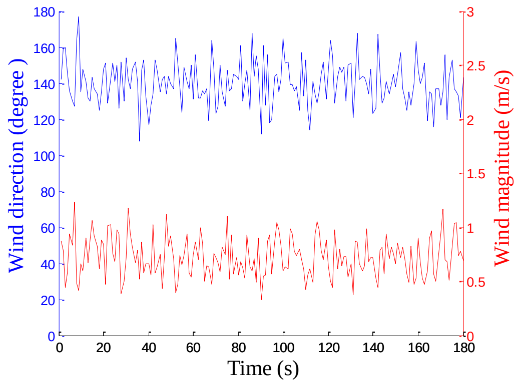


Fig. 4 The fluent field of wind speed and direction data

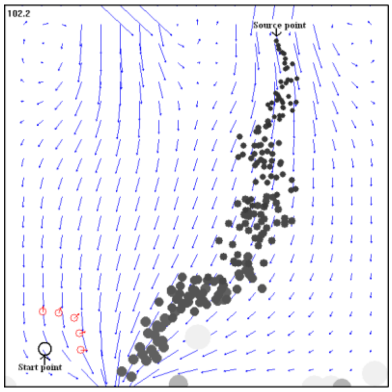
Figure 4 shows several typical moments of the gas leak source search and determination process in the indoor turbulence environment where the gas leak source is located at (7m, 9m). A red circle with a red arrow indicates the position and attitude of the sensor nodes. The length and direction of the blue arrows indicate the intensity and direction of the wind, respectively. After the gas leak source releases 100s (the number in the upper left corner of each subgraph indicates the time), the sensing node begins to diverge in different directions in the vicinity of the position (1m, 1m). When any node detects the gas concentration, the gas leakage source determination started. In Figure 5, the determination process with four time were provided

In order to analyze the performance of the distributed EM method, we compare to the central EM algorithm with the estimation errors of chemical gas leakage source’s coordinates and emission rate, and the determination speed is also considered.

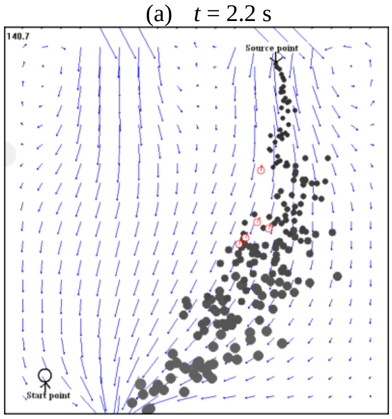
The trajectories of the chemical gas leakage source determination process with the distributed EM and central EM algorithms were given in Figure 6. The blue points represented the selected nodes that computed the parameters, and the circle with different radius surrounding means the sensor measurement. The source determination process could be started at any sensor node (such as the pink node) and iteratively move to the source field, until the final gas source determination was reached, (68,-45) of the central EM method and (52, 13) of the distributed EM method, which represented by the black star and the green triangle independently.

As for Figure 6, the distributed EM method was better than central EM method with less sensor nodes, faster speed and lower communication in the same time step.

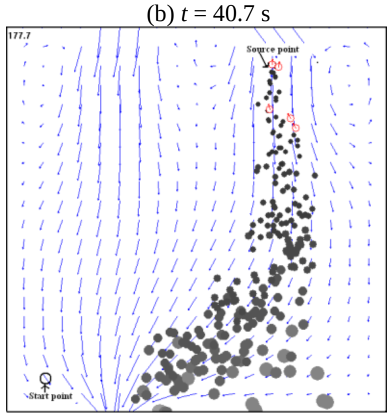
In order to compare the two algorithms more clearly between the determination accuracy and the selected sensors number, we give the performance of the distributed EM and central EM in Figure 7. The determination error of the gas leakage source coordinates and the emission rate with the execution time was shown respectively. The determination accuracy of the distributed EM were higher than the central EM with the same sensor nodes, but the execution time speed was not improved because that the sensor nodes selected was confirmed at each time step.



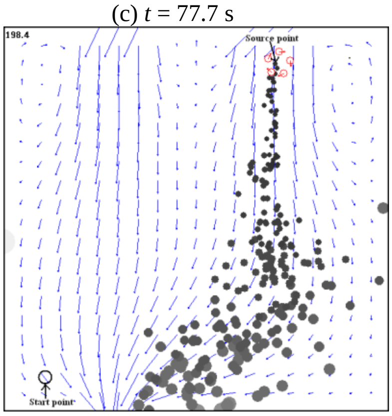
(a) $t = 2.2$ s



(b) $t = 40.7$ s

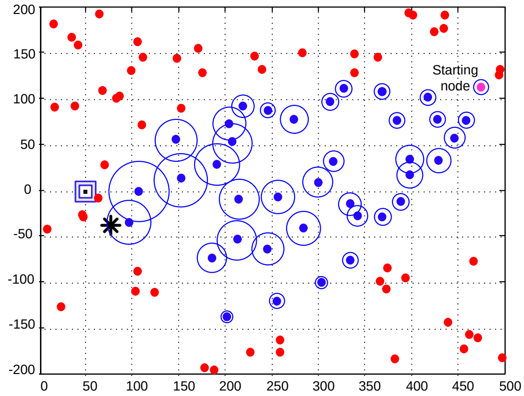


(c) $t = 77.7$ s

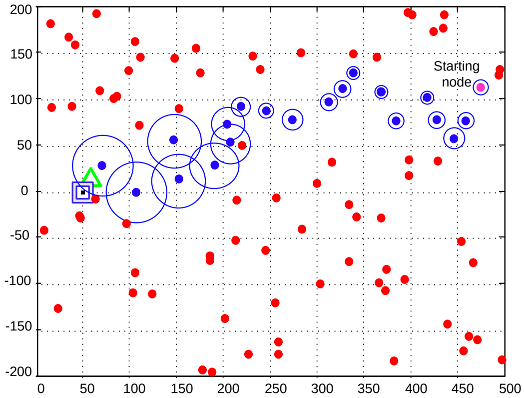


(d) $t = 98.4$ s

Fig. 5 The determination process



(a) Central EM parameter determination process



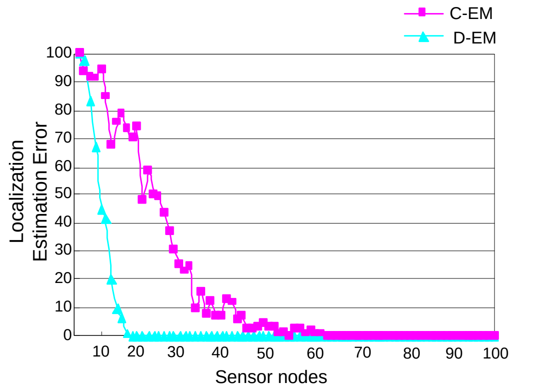
(b) Distributed EM parameter determination process

Fig. 6. An example of the distributed estimation algorithm

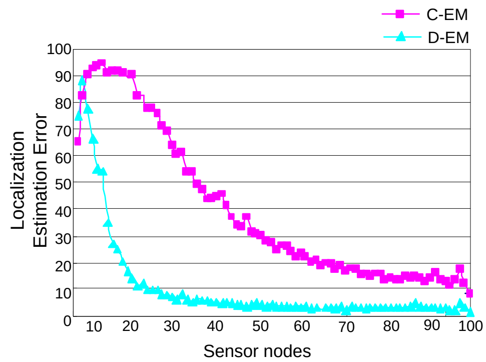
At last, we compared the different approach method with the energy consumption considering, the energy remaining results was shown with the sensor node selected numbers adjusting pattern in the Figure 8.

CONCLUDING

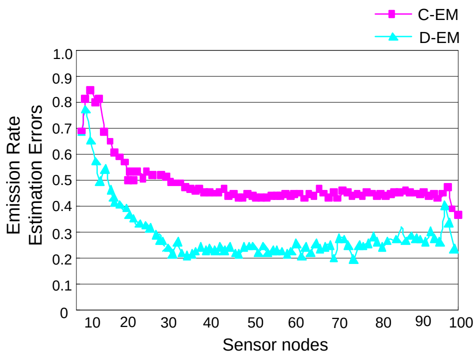
We have presented the distributed EM parameter determination method with Gaussian Mixture model analysis. At the same time, a dynamically adjusting method of sensor nodes selected based on the estimation covariance value was given for the balance of the determination performance and the sensor network consumption. The analysis of the two algorithms had presented with the simulation results. We can see that the distributed EM algorithm was better than central EM method and the nodes number affects the energy consumption and the bias estimation.



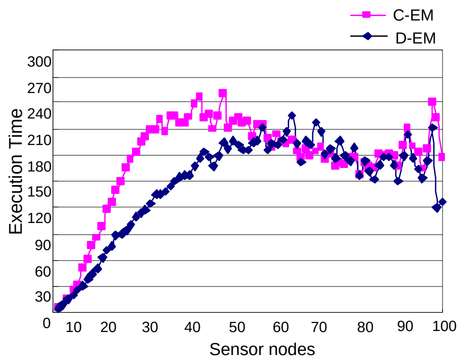
(a) X localization estimation error of the chemical source



(b) Y localization estimation error of the chemical source



(c) The emission rate estimation errors of the chemical source



(d) The execution time of the methods

Fig. 7 Determination error of coordinates and emission rate with the execution time

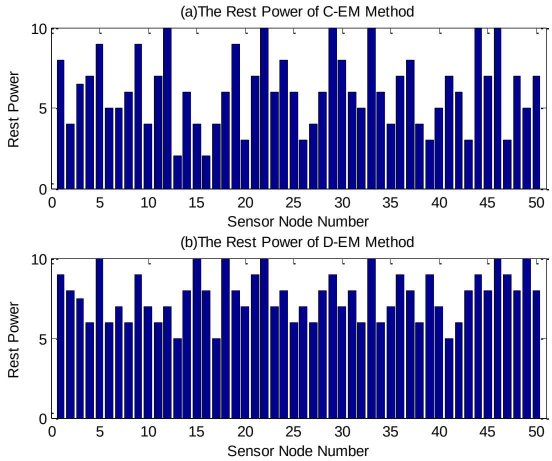


Fig. 8 The energy remaining results of different method

Acknowledgment: *The authors wish to thank for the financial support of Natural Science Foundation of China (61573253, 61271321), Tianjin Science and Technology Project (16YFZCGX00360), Tianjin Natural Science Foundation (16JCYBJC16400), The National Training Programs of Innovation and Entrepreneurship for Undergraduates (201610069007), National Natural Science Foundation Foster Project of Tianjin University of Commerce. The corresponding author is Professor Zhang Liyi.*

REFERENCES

1. A. Keats, E. Yee, F. S. Lien, *Environmental Modelling and Software*, **25,** 1000 (2010).
2. S. Mitra, S. P. Duttagupta, K. Tuckley, *IEEE International Conference on Industrial Technology*, 349(2012).
3. H. Ruhi, J. Charles, L. Ray, *Environmental Science & Technology*, **46**, 1739 (2012).
4. J. G. Li, Q. H. Meng, Y. Wang, *Autonomous Robots*, **30**(3), 281 (2011).
5. D. Q. Gao, W. Chen, *Sensors and Actuators B*, **120** (2), 584 (2007).
6. M. Ortner, A. Nehorai, A. Jerémic, *IEEE Transactions on Signal Processing*, **55**(6), 2520 (2007).
7. S. Balasubramanian, J. Amamcharla, Panigrahi S, *Food and Bioprocess Technology*, **5**(4), 1206 (2012).
8. D. Ucinski, M.Patan, *International Journal of Applied Mathematics and Computer Science*, **20**(3), 459(2010).
9. S. Capasso, P. Iovino, S.Salvestrini, *Chemosphere*, **73**(4),614 (2008).
10. K. S. Rao, *Atmospheric Environment*, **41**(33), 6964 (2007).
11. A. D. Woodbury, T. J. Ulrych, *Water Resources Research*, **32**(9), 2671(1996).
12. A. Woodbury, E. Sudicky, T. J. Ulrych, *Journal of Contaminant Hydrology*, **32**(1-2), 131 (1998).
13. X. P. Zheng, Z. Q. Chen, *Chinese Journal of Safety Science*, **20**(5), 29 (2010).
14. X. P. Zheng, Z. Q. Chen, *Journal of Loss Prevention in the Process Industries*, **24**(4), 293 (2011).
15. G. Cervone, P. Franzese, *Computers & Geosciences*, **36**(7), 902 (2010).
16. M. D. Sohn, M. J. Small, *Journal of Environmental Engineering*, **126**(10), 893 (2000).
17. S. S. Qian, C. A. Stow, M. E. Borsuk, *Ecological Modelling*, **159**(2-3), 269 (2003).
18. I. Senocak, N. W. Henuartner, M. B. Short, *Atmospheric Environment*, **42**(33), 7718 (2008).
19. F. K. Chow, B. Kosovic, S. Chan, *Journal of Applied Meteorology and Climatology*, **47**(6), 1553 (2008).
20. B. Kosovic, G. Sugiyama, S. Chan, *9th Annual George Mason University Conference on Atmospheric Transport and Dispersion Modeling*, 2005
21. A. Keats, E. Yee, F. S. Lien, *Atmospheric Environment*, **41**(3), 465 (2007).
22. A. Keats, E. Yee, F. S. Lien, *Ecological Modeling*, **205**(3-4), 437(2007).
23. A. Keats, E. Yee, *Atmospheric Environment*, **43**(3), 510 (2009).
24. A. Keats, F. S. Lien, *Journal of Wind Engineering and Industrial Aerodynamics*, **96**(10-11), 1805 (2008).
25. E. Yee, *Boundary-Layer Meteorology*, **127**(3), 359 (2008).
26. S.D. Guo, R. Yang, H. Zhang, *International Journal of Heat and Mass Transfer*, **52**(17-18), 3955 (2009).
27. G. Johannesson, B. Hanley, J. Nitao, *US: Lawrence Livermore National Laboratory*, 2004
28. J. Zhu, P. Wang, *Atmospheric Science*, **30**(05), 871 (2006).
29. S. D. Guo, R. Yang, *Journal of Tsinghua University (Science and Technology)*, **49**(05), 629 (2009).
30. A. Keats, E. Yee, F. S. Lien, *Environmental Modelling & Software*, **25**(9), 1000 (2010).
31. N. Tian, J. Sun, *Inverse Problems in Science and Engineering*, **19**(2), 181 (2011).
32. P. Mandal, U. Dutta, *Bulletin of the Seismological Society of America*, **101**(4), 1719(2011).
33. A. Levy, S. Gannot, *IEEE Transactions on Audio, Speech, and Language Processing* , **19**(6), 1540 (2011).
34. N. Tian, J. Sun, W. B. Xu, *InternationalJournal of Heat and Mass Transfer*, **54**(17-18), 4110 (2011).
35. J. Matthes, L. Groll, H. B. Keller, *IEEE Transactions on Signal Processing*, **53**(5), 1711 (2005).
36. P. Shuo, J. A. Farrell, *IEEE Transactions on System, Man, and Cybernetics, Part B*, **36** (5), 1068 (2006).