

Article

Emission Rate Estimation of Industrial Air Pollutant Emissions Based on Mobile Observation

Xinlei Cui, Qi Yu *, Weichun Ma and Yan Zhang

Department of Environmental Science and Engineering, Fudan University, Shanghai 200438, China; 21210740042@m.fudan.edu.cn (X.C.); wcma@fudan.edu.cn (W.M.); yan_zhang@fudan.edu.cn (Y.Z.)

* Correspondence: qiyu@fudan.edu.cn

Abstract: Mobile observation has been widely used in the monitoring of air pollution. However, studies on pollution sources and emission characteristics based on mobile navigational observation are rarely reported in the literature. A method for quantitative source analysis for industrial air pollutant emissions based on mobile observations is introduced in this paper. NO_x pollution identified in mobile observations is used as an example of the development of the method. A dispersion modeling scheme that fine-tuned the meteorological parameters according to the actual meteorological conditions was adopted to minimize the impact of uncertainties in meteorological conditions on the accuracy of small-scale dispersion modeling. The matching degree between simulated and observed concentrations was effectively improved through this optimization search. In response to the efficiency requirements of source resolution for multiple sources, a random search algorithm was first used to generate candidate solution samples, and then the solution samples were evaluated and optimized. Meanwhile, the new index S_{match} was established to evaluate the quality of candidate samples, considering both numerical error and spatial distribution error of concentration, in order to address the non-uniqueness of the solution in the multi-source problem. Then, the necessity of considering the spatial distribution error of concentration is analyzed with the case study. The average values of NO_x emission rates for the two study cases were calculated as 69.8 g/s and 70.8 g/s. The S_{match} scores were 0.92–0.97 and 0.92–0.99. The results were close to the online monitoring data, and this kind of pollutant emission monitoring based on the mobile observation experiment was initially considered feasible. Additional analysis and clarifications were provided in the discussion section on the impact of uncertainties in meteorological conditions, the establishment of a priori emission inventories, and the interpretation of inverse calculation results.

Keywords: source analysis; mobile observation; atmospheric pollutants; Gaussian plume model; CALPUFF



Citation: Cui, X.; Yu, Q.; Ma, W.; Zhang, Y. Emission Rate Estimation of Industrial Air Pollutant Emissions Based on Mobile Observation. *Atmosphere* **2024**, *15*, 969. <https://doi.org/10.3390/atmos15080969>

Academic Editor: Daniel Beysens

Received: 18 July 2024

Revised: 7 August 2024

Accepted: 12 August 2024

Published: 13 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Analysis of pollution sources based on environmental monitoring technology has always been an important method in various environmental protection supervision methods [1–3]. Analysis of air pollution sources can accurately identify and deeply analyze pollution sources, provide strong technical support and scientific decision-making basis for air pollution prevention and control [4,5], and then improve environmental quality through effective treatment measures. With the intelligent development of monitoring equipment and the further application of big data and artificial intelligence technology, air pollution source analysis technology is developing in a more efficient, accurate, and intelligent direction [6,7]. At present, some key pollution sources of key enterprises have installed real-time online monitoring equipment, but in fact, most enterprises or non-key pollution sources still lack real-time monitoring technology. Combined with the current situation of enterprise pollution source supervision, real-time online monitoring technology cannot cover all pollution sources of all enterprises, so it is still necessary to establish a

relatively general quantitative source analysis method suitable for most pollution sources of most enterprises.

The quantitative source analysis that takes the emission rate and other source information as the research objective is called Source Term Estimation (STE), which often requires inverse calculation [8,9]. STE research can provide technical support for the emission control of industrial parks or enterprises [6,10], and it can provide a basis for the relevant departments to formulate the control policy at the same time. The air pollutant source analysis based on the inverse calculation strategy uses the atmospheric dispersion model to describe the quantitative relationship between the pollution source parameters and the concentration at the receptors, and then with the help of a variety of traceability algorithms to backpropagate the estimation of the source parameters from the observed concentration at the receptors [11–13]. The methods of inverse calculation can be divided into two types of algorithms based on optimization theory and statistical theory. Optimization theory algorithms are divided into direct search algorithms and indirect search algorithms based on the solution method. The indirect search algorithms mainly include four-dimensional variational assimilation [14,15], concomitant methods [9,16], the least squares algorithm [17], the inverse calculation of the Safer system [18], etc. This kind of algorithm is more complicated when the derivative of the adaptive function may not be solved, and it is easy to fall into the dilemma of local optimization. Direct search algorithms represented by simulated annealing algorithms [9], genetic algorithms [19,20], artificial neural network algorithms [21], and particle swarm algorithms [22] do not need to solve for the derivative but rather iteratively search the data space for many times until the function converges through different ways of selecting the parameters, and it is relatively easy to obtain the globally better solution [19]. Statistical theoretical methods based on Bayesian inference have been widely used in STE research [23,24], mainly including the Bayesian Monte Carlo random process and the Markov chain Monte Carlo method and its improved methods [25–27]. Each method of inversion has its own advantages and disadvantages. Compared with a single algorithm, the source analysis, which, when coupled with multiple algorithms, is more effective in balancing global and local search performance and can effectively improve the accuracy of inverse results. Araki et al. [28] combined simulated annealing and genetic algorithms to expand the search space to find local optimums, which were applied to optimize an air monitoring network. Zheng et al. [29] proposed an optimization algorithm based on a genetic algorithm and particle swarm optimization algorithm to identify the source parameters of nuclear weapons, and the resultant prediction accuracy is over 98%. Qiu et al. [26] developed a fast source estimation method based on artificial neural networks and particle swarm optimization algorithm and simulated prediction by training artificial neural networks through a large number of scenarios, which significantly improved the computational efficiency and the accuracy of the result. Kong et al. [30] integrated particle swarm optimization algorithm and Kalman filtering algorithm to study the inversion of leakage emission rate, expanding the online monitoring range while improving the accuracy of source localization.

The above algorithms usually search for a better solution strategically and progressively and gradually approximate the true solution in continuous iterations or statistical sampling of larger samples. Their computational time cost is high, especially when the dimensions of the variables to be solved are large. Unfortunately, the number of emission sources of some enterprises is really very large when observing on a mobile cruise for real enterprises. The target enterprise had 25 potential emission sources in the NO_x emission cases in this paper. The time cost of the above search strategy for such an inverse calculation scenario is huge. These progressive search strategies were not used in our source analysis scheme in order to meet the time-sensitive requirements of practical environmental management. Instead, a random search strategy was used to generate a set of candidate solutions, and the better solutions were selected from the candidate set. Random search is a numerical optimization method that does not require an improvement in the gradient of the problem, also known as the black-box method. It can be used to solve discontinuous

or differentiable functions. The algorithm was proposed in the 1960s [31] and has been continuously discussed and optimized since then. Random search is computed by iteratively moving to better positions in the search space [32], and these positions are randomly sampled from the computational domain. Random number generators are available in most computer program libraries, so random search is able to compute large-scale data problems efficiently through fast coding [33,34] and is suitable for many types of global optimization studies [35,36].

Quantitative estimation of air pollution emission rate based on mobile observation is a concern in this paper. Mobile monitoring devices such as mobile monitoring vehicles and unmanned aerial vehicles have been increasingly used in ambient air quality management in the past nearly three decades [37–39]. Mobile observation techniques can effectively quantify the spatial distribution of pollution concentrations over a short measurement time while ensuring data accuracy [40]. They can also achieve fine spatial coverage of air pollution over large scales (tens to hundreds of square kilometers) [41,42] to help managers quickly identify pollution scenarios. Compared with the long-term deployment of fixed monitoring stations, the mobile monitoring approach greatly improves the flexibility, mobility, and convenience of pollution monitoring [43,44]. A large number of source analysis studies mentioned above are mainly based on the data from fixed monitoring equipment. However, there are very few actual source analysis studies on mobile monitoring data. Ge [45] built an integrated mobile monitoring platform based on an ideal model and used a fluent model as well as an improved concomitant probability algorithm to realize automated reverse pollutant traceability at the park scale, but actual cases are still needed to test the emission rate results. A few practical case studies have shown that the complexity of uncertainty in source analysis can lead to large result errors. Zhou [46] estimated the emission rates of five emission sources in a chemical park using the least squares algorithm based on the monitoring concentration data from drones and monitoring vehicles. The relative errors of the emission rate results based on different diffusion models ranged from 21% to 340% overall. The authors attributed the large emission rate errors to a variety of reasons. In addition to the small amount of concentration measurement data, sensor errors, and simulation errors in diffusion models, wind direction deviation was an important factor. The high speed of the mobile monitoring vehicle led to local airflow disturbance, resulting in unrepresentative observed wind direction. It has also been mentioned in literature studies that there is a large uncertainty in the pulsation of local small-scale real-time meteorological conditions, so the influence of meteorological conditions should be paid attention when establishing source analysis methods [21,29,47–49]. The same problem was found in our analysis of the wind data observed with the monitoring vehicle. If the meteorological data used in the emission rate estimation deviates significantly from the meteorological conditions at the time of the actual emission, the error of the result is likely to remain high, which may be important for source analysis at the enterprise scale. We did an ideal experiment based on the simulated data. The results showed that the error of the simulated concentration when the wind direction deviated by 5° – 30° was about 30.2–120.4% relative to the simulated concentration under the reference wind direction. Zhou [46] further suggests that using the diffusion model with higher spatiotemporal resolution or using more accurate meteorological data, especially wind direction data, might yield more accurate source term estimates.

The method of pollution emission rate estimation based on mobile observation is developed in this paper, and two consecutive mobile observation experiments of a large enterprise are used as examples to introduce the method. The observation experiment and the inversion cases identified based on the observed data are presented at the beginning of Section 2. Then, the source analysis process and related technical methods after the observation experiment are also introduced. Similar to the study by Zhou [46], the difference in the effect of the deviation between the simulated and actual meteorological conditions was clearly demonstrated in the cases of this enterprise. Therefore, forward-calculated simulations and searches considering variable meteorological conditions were employed to

cope with this nuisance. The necessity and effectiveness of fine-tuned searches based on actual meteorological conditions are illustrated in Section 3.1 by comparing the results of inversion under different simulation strategies. The determination of a reasonable sample size for the random search is presented in Section 3.2. As mentioned earlier, the source analysis technique here was not based on an asymptotic search, but rather on the evaluation and optimization of the solution samples generated by the random search. Section 3.3 describes the solution set evaluation method, considering both numerical error and spatial distribution error of concentration. The results of inversion are compared with the reference emission list data and the actual emission rate data based on online monitoring at the end of Section 3, and the source analysis method is discussed. The conclusions of the study are given in Section 4.

2. Case Data and Methods

2.1. Mobile Observation Experiment

The experimental scheme of navigation observation usually detours the road outside the enterprise in order to collect as much pollutant concentration data as possible in the downwind direction of the target enterprise. There are many kinds of pollutants involved in navigation observation, and the concentration fluctuation of pollutants on the experimental road section is helpful for emission inversion. Therefore, it is necessary to identify the cases suitable for the inverse calculation according to the observed concentration data characteristics of different pollutants.

The NO_x monitoring samples received from a mobile observation experiment of Enterprise K on 27 October 2022 were used as examples to introduce the pollutant emission rate estimation method based on the observation data. This enterprise was a petrochemical enterprise located in Shanghai, China, covering an area of about 2.2 km². The data collection of the observation experiment was accomplished by using a mobile monitoring vehicle, and the monitoring equipment mainly included a vehicle-mounted atmospheric sampling system and a vehicle-mounted NO_x analyzer. Meanwhile, the global positioning system device was used to record the real-time coordinates and speed information of the vehicle during the experiment. During the navigation experiment, the weather was fine, and the dominant wind direction in the area where Enterprise K was located was northeast wind 1–2. The mobile observation experiment was conducted around Enterprise K. Two observations followed essentially the same route, and the experimental sequences were labeled K-1 and K-2 in that order. Figure 1 illustrates the route of the mobile observation experiment where the peak of NO_x concentration was observed and where the probable primary source was Enterprise K. The two observations passed through the road section at about 15:43–15:52 and 16:27–16:35 (Beijing time), respectively. Analyzed together with the real-time observation data from the adjacent meteorological stations, it was more likely that the concentration fluctuation of NO_x on this road section was from the emission of Enterprise K. At the same time, the possibility of emission contribution from the adjacent upwind Enterprise A and Enterprise B of Enterprise K could not be excluded.

The measurement resolution of the vehicle-mounted NO_x analyzer is 1 ppb, and the measurement error of the equipment is about $\pm 10\%$. The temporal resolution of the concentration data collected by the driving vehicle is 5 s, and the corresponding spatial resolution is about 20–30 m. However, the spatial resolution of atmospheric pollutant diffusion models is difficult to achieve such a high precision. The spatial resolution of the small and microscale diffusion models, which are more commonly used in the literature, is more than 100 m. If the spatial resolution of the source resolution is too large, the loss of spatial accuracy of navigation observation data will be too large. In order to perform the inversion, the continuous monitoring of the NO_x concentration on the road was spatially averaged at 50 m intervals, and the purple triangular symbols in Figure 1 indicate the starting and ending points of the road section taken in the inversion. A total of 50 source-resolved receptors were actually set after the spatial averaging process, and they were numbered in the order of the direction traveled by the observation vehicle.

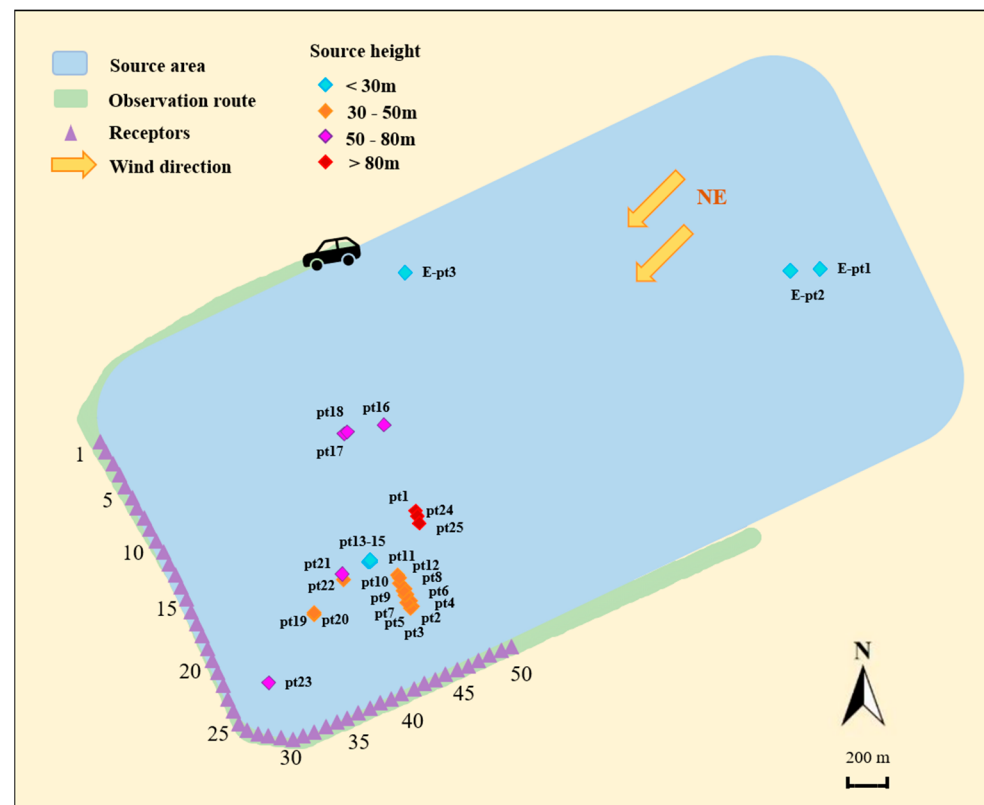


Figure 1. Navigation route and emission sources. The numbers 1–50 indicate the serial numbers of the receptors.

The overall concentration range of NO_x during the two observation cases was about $10.4\text{--}150.3\ \mu\text{g}/\text{m}^3$, and the maximum concentration of NO_x collected was about $77.3\text{--}112.7\ \mu\text{g}/\text{m}^3$ higher than the background concentration. Figure 2 shows the observed NO_x concentration along the route after deducting the background concentration. The high concentration values were concentrated in the range of 20–34 and 21–39 of the receptors, respectively, both located in the southwest direction outside the enterprise. The spatial distribution span of concentration peaks was about 750 m.

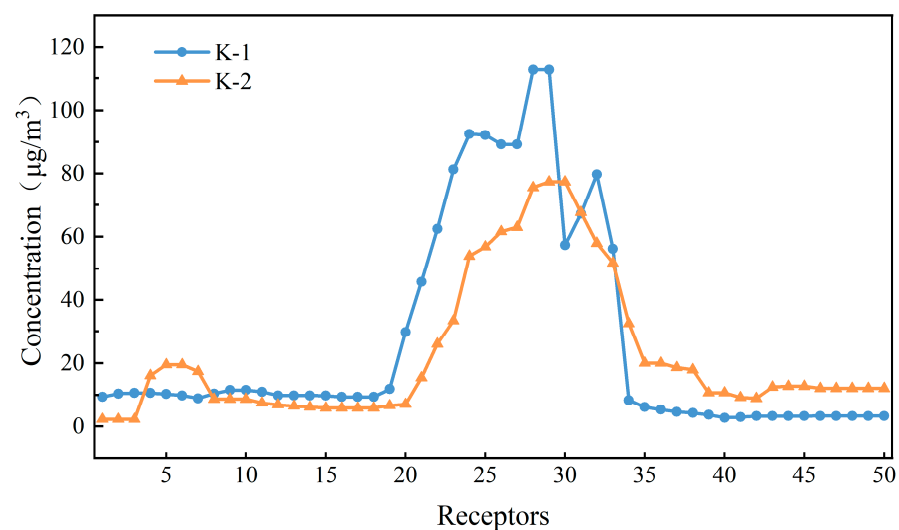


Figure 2. Observed concentrations of NO_x .

2.2. Process of Emission Rate Estimation

The observed concentration data can be regarded as the accumulation of concentration effects produced by independent emissions of multiple pollution sources for the source analysis problem of multiple pollution sources. Assuming that the number of pollution sources is n and the number of acceptor points is m , the estimation of pollutant emission rate based on mobile observation is essentially to find the emission rate solution set of Formula (1).

$$\begin{bmatrix} C_1 \\ \vdots \\ C_m \end{bmatrix} = \begin{bmatrix} H_{11} & \cdots & H_{1n} \\ \vdots & \ddots & \vdots \\ H_{m1} & \cdots & H_{mn} \end{bmatrix} * \begin{bmatrix} Q_1 \\ \vdots \\ Q_n \end{bmatrix} \quad (1)$$

where C represents the concentration at the receptors; H is the pollution source emission-receptors concentration–response matrix (matrix H for short), calculated by diffusion model; and Q indicates the emission rate of the pollution source. The source analysis in this study was to solve the emission rate matching the observed concentration C in Formula (1), hereinafter referred to as Q_{inv} . For the common industrial emission source analysis scenario, m is likely to be less than n , which makes the solution of Formula (1) not unique. However, even if m is large enough, the exact analytical solution of Formula (1) cannot be obtained if multiple pollution sources in a park or enterprise are distributed close to each other, considering the errors of monitoring equipment and diffusion model. Therefore, the actual source analysis research is more about seeking the optimal solution rather than the exact solution. It is worth noting that the more objects resolved using the inversion, the larger the dimension of the inversion problem, and the uncertainty of the result increases.

The inversion algorithms used by many scholars in source analysis studies are usually strategic stepwise searches, as mentioned earlier. However, it is difficult to meet the timeliness requirements of practical management work when the dimensionality of the solution problem is large. Therefore, an inversion scheme based on random search and solution set optimization was adopted in order to achieve a fast estimation of the emission rate. Figure 3 shows the emission rate estimation process adopted in this study, which mainly includes the following:

(1) Collect source list and set reference emission rates.

Source list collection means the comprehensive collection of a list of emission sources that may contribute to the concentrations corresponding at the monitoring site under real-time meteorological conditions. A complete source list is necessary for accurate source analysis results. Therefore, all relevant source information should be collected as much as possible. The reference emission rate is the input parameter used to calculate matrix H , hereinafter referred to as Q_{ref} . Source analysis is usually based on matrix H under the reference emission rate conditions to develop an inversion fit to the observed concentrations. So, it is recommended to use data that can represent the actual emission level of pollution sources as far as possible in the setting of reference emission rates, such as the emission data in the enterprise's environmental impact assessment report, pollutant discharge permit, pollution source survey, online monitoring data, and other materials in order to improve the efficiency and accuracy of inversion.

(2) Calculate matrix H .

Matrix H is calculated by using the atmospheric pollutant diffusion model based on the collected source list and reference emission rate under given meteorological conditions. It can reflect the quantitative relationship between the emission of pollution sources and the concentration response of acceptor points. The area for the forward dispersion simulation in this study was approximately $7.5 \times 7.5 \text{ km}^2$, and the issue of concern was near-field-based transport dispersion simulation. The accuracy of the near-field diffusion simulation is greatly affected by meteorological conditions, which means that the simulation scheme of matrix H will have a critical impact on the accuracy of the results of inversion. Therefore,

when determining the actual simulation scheme, the two simulation schemes of fixed meteorological condition and variable meteorological condition were compared and analyzed, as described in Section 2.2.2.

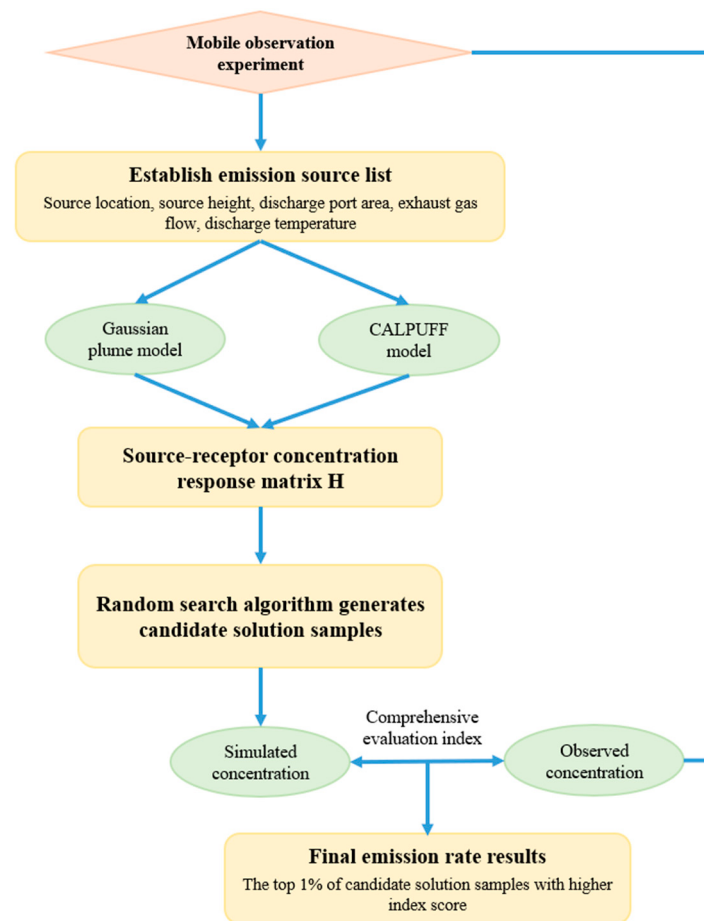


Figure 3. Emission rate estimation process.

- (3) Construct the candidate solution samples by using the random search algorithm.

The candidate solution sample was the ratio of the emission rate of inversion of each source to the reference emission rate since matrix H was calculated based on the reference emission rate. It was denoted as R in this paper and was calculated as shown in Formula (2).

$$R = \frac{Q_{inv}}{Q_{ref}} \quad (2)$$

The random search algorithm was used to build candidate solution sets. The random search method established in this study was to conduct a batch search of solution set samples, which was different from the commonly used progressive inversion search strategy [14–17]. Random sampling values were generated in the given range of emission rates, and different emission rate combinations formed different candidate solution samples. R took values between 0.001–1000 in the random search.

The more samples of candidate solutions, the wider the search domain, and the more favorable it is to obtain more accurate solutions. Therefore, the number of random number samples should be proportional to the number of emission sources to be analyzed. For this reason, it is necessary to test the sensitivity of the influence of sample size on the accuracy of source analysis to determine a reasonable random search sample size. The upper limit of the tested sample size was 200,000, and a total of nine sample sizes of 1000, 3000, 5000, 8000, 15,000, 50,000, 100,000, 150,000, and 200,000 were taken for the actual test in this paper.

- (4) Evaluate the solution samples, identify the better solution samples, and collect the results of source analysis.

The simulated concentrations were calculated using Formula (1) after obtaining samples of candidate solutions. The quality of each candidate solution sample was evaluated by comparing the simulated concentration with the observed concentration using an appropriate evaluation index. The optimal solution samples were then identified according to the evaluation results. The detailed introduction of the evaluation index is described in Section 2.2.3. The final emission rate results were obtained by selecting the top 1% of candidate solution samples with higher ratings for statistical analysis in order to avoid the error caused by overfitting. The reason why only the highest-rated solution samples were not selected for emission rate result statistics is explained in the analysis in Section 3.3.

2.2.1. Source List and Reference Emission Rate

The information in the emission permits and relevant environmental impact assessment reports was used to establish the list of emission sources and to determine parameters such as the height of the emission sources as well as the reference emission rate. Although some source online monitoring data were available for Enterprise K, they did not cover all outlets, and there were enterprises such as Enterprise A and Enterprise B that lacked source online monitoring data. Fortunately, the collected online monitoring data of pollution sources from Enterprise K were used for the validation analysis of the results of the inversion. The NO_x of Enterprise K mainly came from the emission of combustion sources such as cracking furnaces, thermal oxidation furnaces, steam superheating furnaces, hot kerosene furnaces, power centers, and waste gas incinerators. The 25 stationary emission outlets of Enterprise K, denoted as K-pt1–K-pt25, are shown as diamond-shaped points in Figure 1. In addition, further investigation of upwind sources revealed that Enterprise A and Enterprise B, which were adjacent to enterprise K, also had a number of NO_x emission outlets. They were the emission outlets of the heating furnace and incinerator, after checking the information such as the emission license, which were noted as E-pt1, E-pt2, and E-pt3 in this paper, as shown in the diamond-shaped point locations in Figure 1. Table 1 shows the list of NO_x emission sources for the inversion cases, and the overall height range of all emission sources was 15–120 m.

Table 1. Emission source list of NO_x .

Source Number	Height (m)	Source Number	Height (m)
K-pt1	120	K-pt15	24
K-pt2	50	K-pt16	65
K-pt3	50	K-pt17	65
K-pt4	50	K-pt18	65
K-pt5	50	K-pt19	50
K-pt6	50	K-pt20	50
K-pt7	50	K-pt21	80
K-pt8	50	K-pt22	50
K-pt9	50	K-pt23	68
K-pt10	50	K-pt24	120
K-pt11	50	K-pt25	120
K-pt12	50	E-pt1	25
K-pt13	15	E-pt2	25
K-pt14	15	E-pt3	23

The reference emission rates for each source are given in Figure 4. The sum of the reference emission rate of all sources was 76.2 g/s, and the reference emission rate from Enterprise K was 74.2 g/s. The highest value of the reference emission rates was in K-pt17, followed by K-pt18; the lowest values of the reference emission rates were found in K-pt13 and K-pt14; and the values of the emission rates of Enterprise A and Enterprise B were overall smaller than those of Enterprise K.

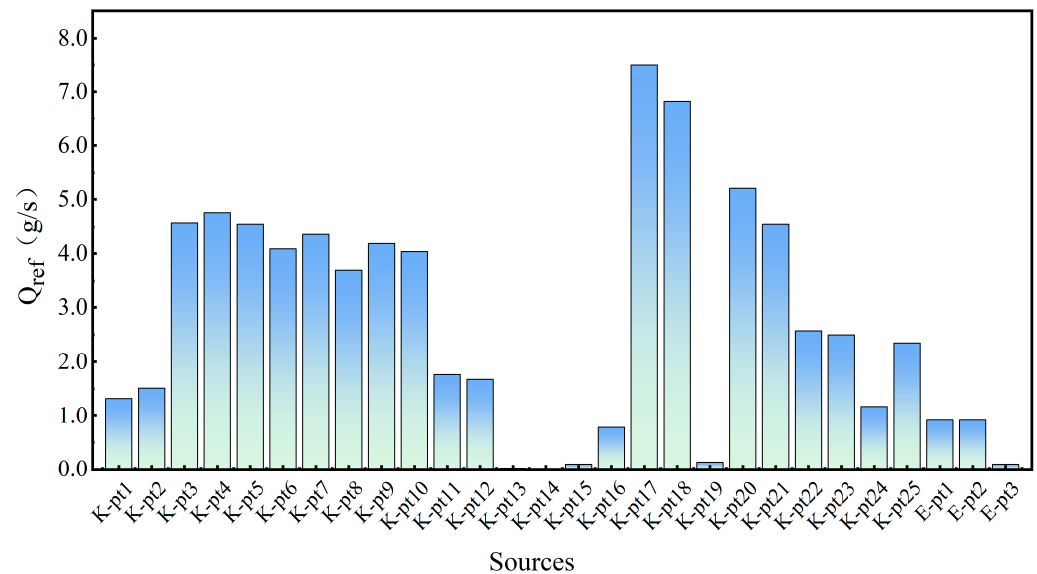


Figure 4. Reference emission rates of NO_x for candidate sources.

2.2.2. Diffusion Simulation and Modeling of Meteorological Conditions for Matrix H

There are many small-scale pollutant transport diffusion models in practical research and application, including the Gaussian diffusion model [50,51], AERMOD [52], CALPUFF [53], and CFD software (such as PHOENICS) [54]. Among them, CALPUFF and Gaussian plume models have significant advantages in small-scale and medium-scale diffusion simulation. The CALPUFF model can construct a relatively fine three-dimensional meteorological field under complex terrain conditions, which is suitable for diffusion simulation at scales of tens of meters to hundreds of kilometers, and it can also satisfy diffusion simulation needs of different types of pollution sources [55]. The Gaussian diffusion model is more accurate for small-scale atmospheric pollutant diffusion simulation results. It has strong computability, so it is easy to adjust the meteorological parameters for diffusion simulation [56].

The mobile observation routes are usually on the neighboring roads outside the enterprise's plant boundaries, resulting in a relatively short transport and diffusion distance of pollutants after leaving the emission source. The accuracy of the pollutant dispersion simulation is greatly affected by the meteorological conditions at such a spatial scale. Small fluctuations in the meteorological parameters may lead to significant errors in the emission rate results of inverse calculation. We first adopted a simulation scheme based on WRF/CALPUFF and Gaussian plume modeling for forward calculation under actual meteorological conditions due to this effect. When the inversion accuracy could not be improved, further attempts were made to fine-tune meteorological parameters based on actual meteorological conditions and to compare the inversion effects of these two simulation schemes. The former was referred to here as diffusion simulation with fixed meteorological conditions and the latter as diffusion simulation with variable meteorological conditions.

Fixed Meteorological Condition

In the simulation scheme with fixed meteorological conditions, two methods, WRF/CALPUFF and Gaussian plume modeling, were used to complete the simulation of meteorological fields and pollutant dispersion under actual meteorological conditions, respectively.

CALPUFF is a non-stationary Lagrangian diffusion model system that can simulate the migration, diffusion, and transformation of a variety of pollutants and is widely used in the field of environmental impact assessment [50]. The CALPUFF model mainly consists of a meteorological field module, CALMET, and a smoke mass diffusion module, CALPUFF. The CALMET module is used to provide the meteorological field required for pollutant dispersion, and the CALPUFF module is used to calculate matrix H. The core part of the

CALMET module consists of the micrometeorological field model and the diagnostic wind field. The micrometeorological field model describes the structure of the atmospheric boundary layer based on a parametric approach, and the diagnostic wind field is based on the mass-conserving continuum Formula for the diagnosis of the wind field. The CALPUFF module takes the source list in the simulation area and the three-dimensional wind field simulated by CALMET as input data and combines them with the relevant information of the receptors to obtain the concentration response at each receptor site in the computational domain.

The simulation scheme of fixed meteorological conditions adopts the CALMET module to calculate the simulated wind field from 15:00 to 17:00 on 27 October 2022 (Beijing time). With the meteorological field simulated by WRF as input, the grid resolution is 1 km. CALMET is used to achieve downscaling meteorological field simulation. Considering slope flow and terrain blocking effect, the initial guessing field is adjusted to reflect complex terrain and land use characteristics, and the grid resolution is refined to 50 m. The computational domain of CALPUFF was centered on the pollutant source area. The simulation area was set up with a grid number of 150×150 with a grid resolution of 50 m in the horizontal direction and was divided into ten layers totaling 1000 m in the vertical direction. The time resolution of the simulation output matrix H was 1 h. The two mobile observations occurred in two adjacent hours, so the two matrix H results might be similar.

The Gaussian plume model used the meteorological parameters calculated by CALMET as input to realize the simulation scheme calculation of fixed meteorological conditions. For both cases, K-1 and K-2, the actual wind direction was around 45° , and the wind speed was around 3.0 m/s. The introduction of the Gaussian plume model is described in the next section.

Variable Meteorological Condition

The simulation scheme with variable meteorological conditions used the Gaussian plume model for the simulation of meteorological fields and pollutant dispersion. The Gaussian plume model is mainly used to simulate the diffusion process of atmospheric pollutants that are continuously and constantly released from emission sources under steady-state conditions [56]. The model is based on the following assumptions: (1) The concentration distribution of pollutants in the y- and z-axis directions conforms to the normal distribution; (2) The wind speed is uniform and stable in all the space; (3) The strength of pollutant sources is continuous and uniform; and (4) The pollutant mass is conserved in the diffusion process. For continuous emissions from elevated point sources under constant meteorological conditions, the pollutant concentration $C(x, y, z)$ at any point downwind can be calculated by the Gaussian plume Formula (3) after considering the total reflection of the plume at ground level [49].

$$C(x, y, z) = \frac{Q}{2\pi U \sigma_y \sigma_z} e^{\frac{-y^2}{2\sigma_y^2}} \sum_{n=-k}^{+k} \left\{ e^{-\frac{(z+2nh-H_e)^2}{2\sigma_z^2}} + e^{-\frac{(z+2nh+H_e)^2}{2\sigma_z^2}} \right\} \quad (3)$$

where Q is the emission rate per unit time, mg/s; x, y are the vertical and horizontal distances between the point and the axis of the mean wind direction passing through the exhaust outlet in the horizontal plane, m, respectively; σ_y is the horizontal lateral diffusion parameter perpendicular to the mean wind direction, m; σ_z is the vertical diffusion parameter, m; U is the mean wind speed at the exit of the exhaust outlet, m/s; h is the thickness of the mixing layer, m; H_e is the effective source height, m; H is the geometric height of the exhaust outlet from the ground, m; and ΔH is the smoke lifting height, m.

The Gaussian plume model can accept the input of simple meteorological parameters, so it is convenient to carry out forward diffusion simulation based on different meteorological conditions in inverse calculation. In the simulation scheme of variable weather conditions, the weather parameters, including wind speed, wind direction, and atmospheric stability, were fine-tuned based on the actual weather conditions. The wind speed

was set to change within the wind speed range corresponding to the actual wind level. The wind direction varied within $\pm 45^\circ$ of the actual wind direction data. The atmospheric stability was based on the CALMET output data and varies in its vicinity. Finally, the search range of the variable wind direction parameter was set to $45^\circ \pm 45^\circ$ with a search step of 5° . The range of simulated wind speed was set to 1.0–5.0 m/s with a search step of 1.0 m/s. And the atmospheric stability was set to A, B, C, D, and E in that order.

2.2.3. Candidate Solution Evaluation

Simulated concentration error evaluations such as relative error, root-mean-square error, and sum of squared errors [20,21,29] are the most commonly used evaluation methods in similar studies. However, it was found that the relative error had some shortcomings in the ability to interpret spatial characteristics through the analysis of the concentration distribution characteristics of a large number of candidate samples. The evaluation of the matching degree of concentration spatial distribution was added in the evaluation of candidate samples in order to make full use of the spatial distribution of pollutants to reveal the spatial distribution characteristics of pollution sources. Moreover, combining the spatial distribution error of concentration with the evaluation of numerical relative error, a new comprehensive evaluation index S_{match} of simulated concentration and observed concentration was constructed, which was defined as follows:

$$S_{match} = \frac{S_e + S_p}{2} \quad (4)$$

Among them, S_{match} is the comprehensive matching index of simulated and observed concentrations, with a value range of 0–1, and the larger the value, the higher the matching degree; and S_e is the index characterizing the consistency between the simulated and observed concentrations in terms of numerical magnitude, which is calculated as shown in Formula (5), where e is the relative error between the modeled and observed concentrations, calculated according to Formula (6); and C_{sim} and C_{obs} denote the simulated and observed concentrations, respectively.

$$S_e = 1 - e \quad (5)$$

$$e = \frac{C_{sim} - C_{obs}}{C_{obs}} \times 100\% \quad (6)$$

S_p is an index characterizing the peak overlap ratio between the simulated and observed concentrations, which is evaluated based on the agreement between the peak position and peak area of the simulated and observed concentrations and is calculated as shown in Formula (7).

$$S_p = \frac{1}{n} \sum_{i=1}^n \frac{A_{sim,i}}{A_{obs,i}} \quad (7)$$

where n denotes the number of concentration peaks; A_{obs} denotes the area of the peaks of the observed concentrations; and A_{sim} is the area of the overlap of the peaks of the simulated and observed concentrations. The evaluation of concentration peak matching was implemented using Python 3.9 programming.

3. Results

3.1. Comparison of Simulation Schemes for Matrix H

The concentration relative errors of 50,000 random samples were used as an example to analyze the differences between different matrix H simulation schemes. Figure 5 shows the statistical distribution of e under different simulation schemes. The sample size of the random search was relatively large, so the variation range of e is also relatively large. Take K-1 as an example. The value range of e in CALPUFF's simulation result was 67.0–97.7%. The value range of e of the Gaussian plume model was 47.7–169.8% under actual weather conditions, while the value range of e based on variable weather conditions was 17.0–124.4%. Emission rate estimation was used to find the solution with a small relative error of concen-

tration by comparing the minimum of e under different simulation schemes. It can be seen that in the inverse calculation results of the two cases, the relative error of concentration simulated by CALPUFF under actual meteorological conditions was the highest, and the minimum values of e of K-1 and K-2 were 67.1% and 54.8%, respectively. The relative error of corresponding concentration in the simulation scheme based on the Gaussian plume model with actual meteorological parameters was smaller, and the minimum value was 47.7% in case K-1 and 38.7% in case K-2. Although the actual meteorological conditions were referred to, the difference between the simulation results based on the two simulation schemes was very significant. It could be seen that the influence of meteorological data on the accuracy of inverse calculation is very important. A simulation scheme with variable meteorological conditions was used to further solve the problem because it was considered that the real-time wind was dynamically changing. Compared with the previous two simulation results, it was found that the minimum relative error of concentration in the simulation scheme with variable weather conditions was the lowest, and the minimum error of the two cases was 17.0% and 15.2%, respectively. From the comparison of the concentration error of the simulation results, the inverse calculation results considering the fine-tuning of meteorological parameters had been significantly improved.

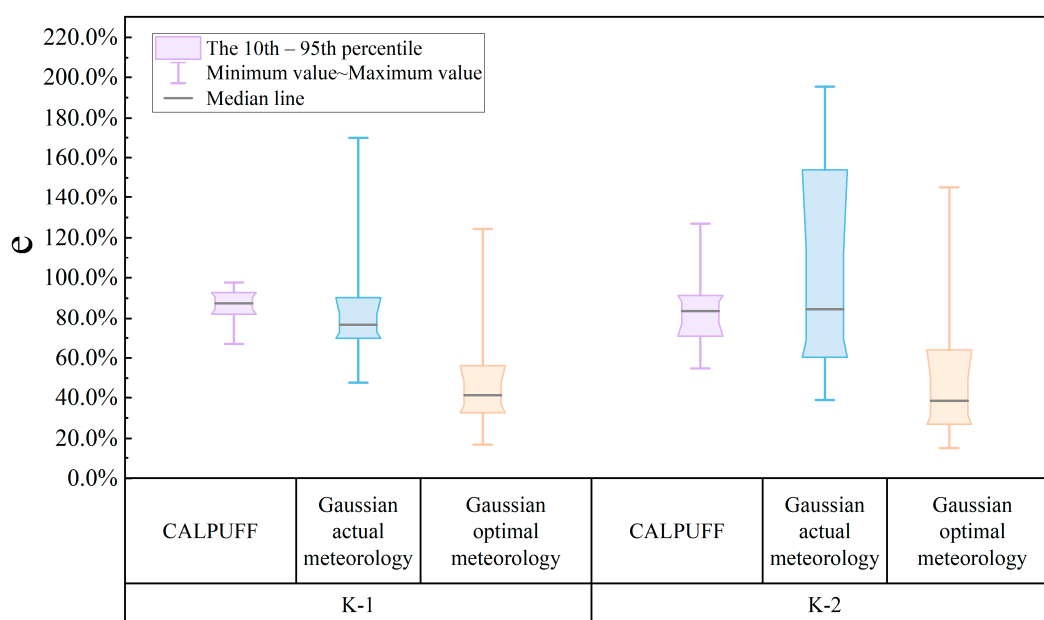


Figure 5. Relative errors for simulated concentrations.

In addition, it can also be clearly seen from the spatial distribution characteristics of the observed concentration of NO_x in the two cases that, although the times of the two mobile observation experiments were similar, the peak values of the observed concentration appeared in different positions, and the width of the concentration peaks were also different. From the observed concentration data given in Section 2.1, it can be seen that the observed concentration peak of K-1 appeared between receptors 20 and 34, and the concentration peak of K-2 appeared between receptors 21 and 39. This indicated that there were some fluctuations in the local wind direction during the mobile observation experiments. Although the difference in observed concentration data in the two cases was not only related to the fluctuation of meteorological conditions, the real-time dynamic change of meteorological conditions must be an important reason. In view of this, the e value of the solution sample showed an obvious decreasing trend when the simulation scheme based on variable meteorological conditions was conducted using a random search.

The differences in distribution patterns of matrix H were further explored in order to find the reasons for the difference between the concentration relative errors of simulation schemes based on fixed and variable meteorological conditions. The simulation results of

matrix H under the two simulation scenarios are given in Figure 6. Matrix H was based on the simulation results of Q_{ref} . Although there were some differences between the actual emission rate and Q_{ref} , the spatial distribution characteristics of the matrix H can be used to approximately assist in analyzing the potential contribution characteristics of the pollutant emission of enterprises to the concentration of pollutants in the environment. As can be seen in Figure 2, the observed peak concentration of NO_x in both cases occurred around receptor 27. The CALPUFF-based simulation results showed that part of the concentration peaks occurred around receptor 27 at the enterprise's daily emission levels. However, the location where concentration peaks were more likely to occur was likely to be around the receptors 33 to 50, stemming from the contribution of E-pt1 and E-pt2. In addition, the CALPUFF simulation results were in a relatively large gradient of concentration decrease near the peak of the observed concentration data (the receptor 27). It was difficult for matrix H of such a shape to obtain a concentration contribution consistent with the observed data at other receptors to the left and right of receptor 27, so the concentration error of the inversion candidate solution sample was relatively large. In contrast, in the results of the optimal meteorological parameters based on variable meteorological conditions, matrix H had some peaks near receptor 27, and the width of the simulated concentration peaks was also larger. Only matrices of this shape were more likely to receive a more desirable solution set in the inversion than the simulation results of CALPUFF.

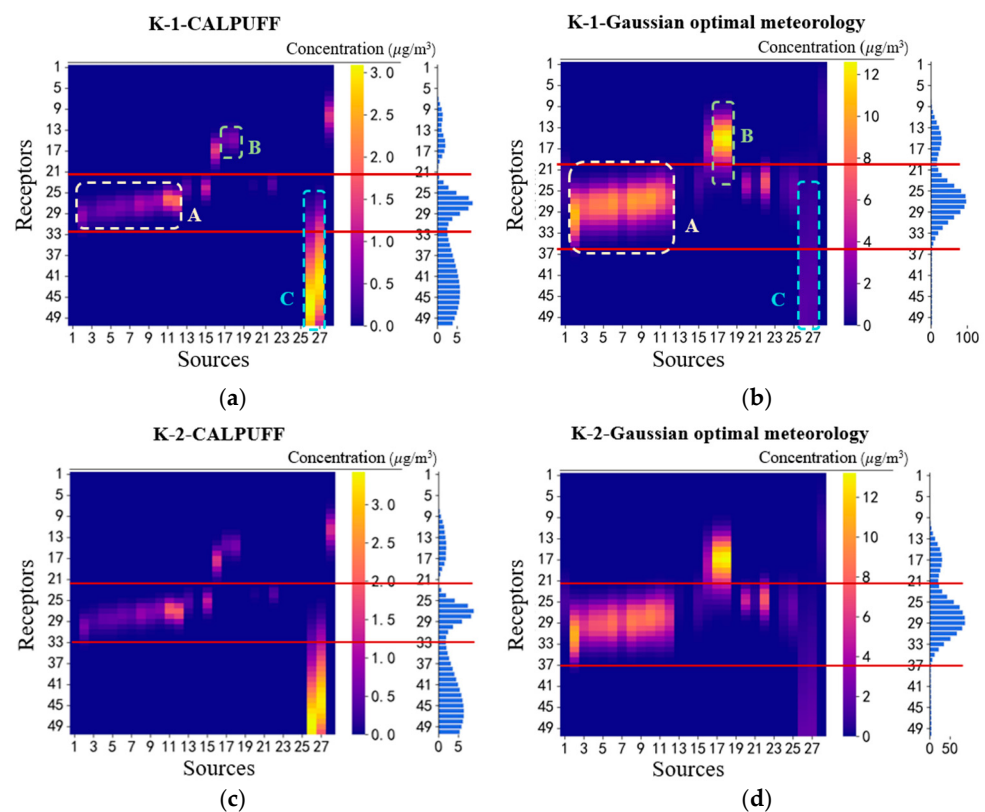


Figure 6. Matrix H outputs for different simulation schemes: (a) K-1 simulation results based on CALPUFF; (b) K-1 simulation results based on Gaussian plume model; (c) K-2 simulation results based on CALPUFF; (d) K-2 simulation results based on Gaussian plume model. The heat map on the left side of each subgraph represents the concentration–response values at each receptor for each source’s emissions, $\mu\text{g}/\text{m}^3$; the right-hand side bar graph shows the total concentration at each receptor due to emissions from all sources, $\mu\text{g}/\text{m}^3$; the location of the red line on the right side indicates the spatial extent corresponding to the peak of the concentration–response wave. Dashed box A represents the range of receptors affected by the emission of K-pt2–K-pt12, dashed box B represents the range of receptors affected by the emission of K-pt17–K-pt18, and dashed box C represents the range of receptors affected by the emission of K-pt26–K-pt27.

The unsatisfactory emission rate results based on matrix H of Figure 6a,c were initially understood as incomplete source list information collected. This means that there may be other sources of emissions that are not included in the inversion, so some of the observed concentration data may not be explained. However, the concentration error of the inverse results is still not satisfactory, even after confirming the completeness of the emission source list. Therefore, the simulation scheme of variable meteorological conditions was considered and adopted. The subsequent source analysis in this paper will only be based on the optimal meteorological condition parameters in the simulation scheme with variable meteorological conditions.

3.2. Influence of Sample Size for Random Research on Inversion Accuracy

Figure 7 shows the statistical results of e corresponding to different random search sample sizes. With the increase in the sample size, the trend of the error values at the 90th and 10th percentiles in the two cases was relatively smooth, but the changes in the maximum and minimum values of the errors were different. The maximum error of the candidate samples showed an increase in fluctuation with the increase in sample size, while the minimum error showed a decrease in fluctuation. The minimum concentration error of K-1 was the largest at sample size 1000, which was 20.4%. The minimum error of sample size 15,000 and 200,000 were close to each other, and there was a small decrease compared with the minimum error of sample size 1000. There was some fluctuation in the minimum error of the sample size in the range of 15,000–200,000. The minimum concentration error of K-2 was the largest at sample size 3000, which was 18.5%. The minimum error of sample sizes 50,000 and 200,000 were close, and they decreased slightly with respect to sample size 3000. The minimum error fluctuated at a sample size of 15,000, and the minimum error value was larger than the value at a sample size of 50,000. To sum up, the minimum concentration error of the solution set tended to stabilize when the sample size was larger than 50,000. The minimum error of the solution set generally tended to decrease when the sample size of the random search increased, so larger sample sizes should still be recommended. Therefore, in all subsequent analyses, the random sample size was set to 100,000.

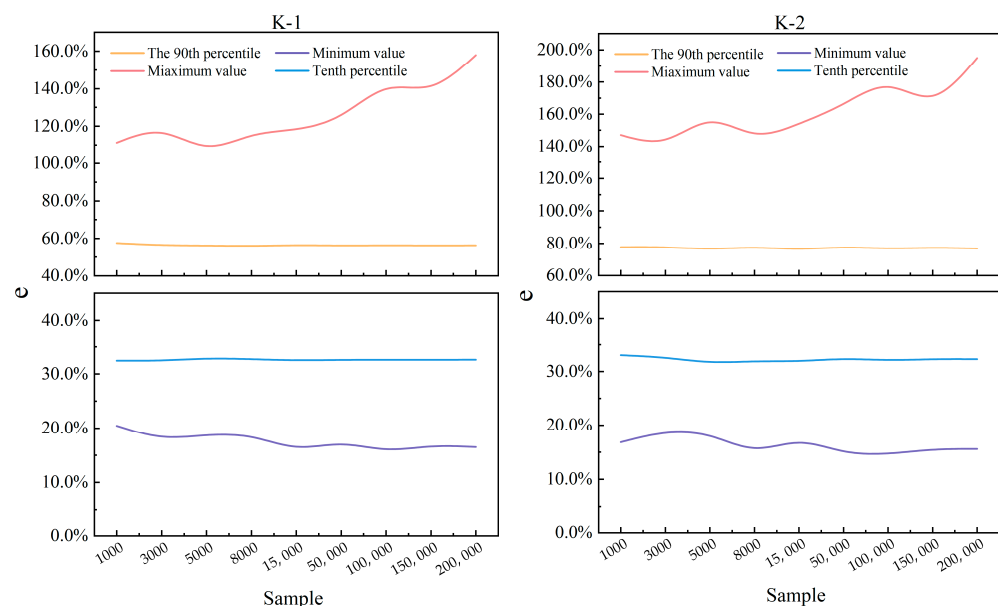


Figure 7. Relative errors for simulated concentrations varied with random search sample sizes.

3.3. Differences in Evaluation Indexes for Candidate Solution

Only the nearly 10,000 candidate solution samples with $S_e \geq 0.70$ ($e \leq 30.0\%$) were selected for analysis in this section, and the distribution of their evaluation metric values

is plotted in Figure 8, due to the large sample size. As can be seen in the top right corner of Figure 8, the minimum value of e for the obtained candidate solutions was about 15.3%, and the S_p value of this sample was 0.98. The closer to the upper right region of Figure 8, the smaller the number of solution samples was, but the distribution was also more concentrated. The S_p scores of solution samples in the interval $0.80 \leq S_e < 0.85$ were more centrally distributed in the range of 0.87 to 1.00, where 90% of the samples were concentrated in the interval of S_p values 0.94–1.00. This means that there was some consistency between the two indicators of the optimal sample. The solution samples with higher S_e scores also had higher S_p scores. Therefore, this part of the sample was also the sample with a higher S_{match} for the composite score, and their S_{match} was in the range of 0.92–0.99. In contrast, the number of samples in the ranges of $0.75 \leq S_e < 0.80$ and $0.70 \leq S_e < 0.75$ was much larger, and the larger the concentration error, the more the samples were. However, as the relative error increased, the difference in S_p metric scores between the samples continued to increase. Looking at Figure 8 from right to left, the S_p index of the samples at $0.75 \leq S_e < 0.80$ ranged from 0.84 to 1.00, with 92% of the samples having a S_p index greater than 0.90. In contrast, the S_p index of the samples at $0.70 \leq S_e < 0.75$ ranged from 0.74 to 1.00, with 90% of the samples having a S_p index greater than 0.85. This suggested that, as the consistency of the samples with S_e scores increased, the consistency of S_p also increased. It can also be found that there were differences between the two indexes by analyzing Figure 8 from the top down. The samples with the S_p score of 0.95–1.00 were the samples with the highest overlap rate, but the corresponding S_e scores were scattered and varied between 0.70 and 0.85.

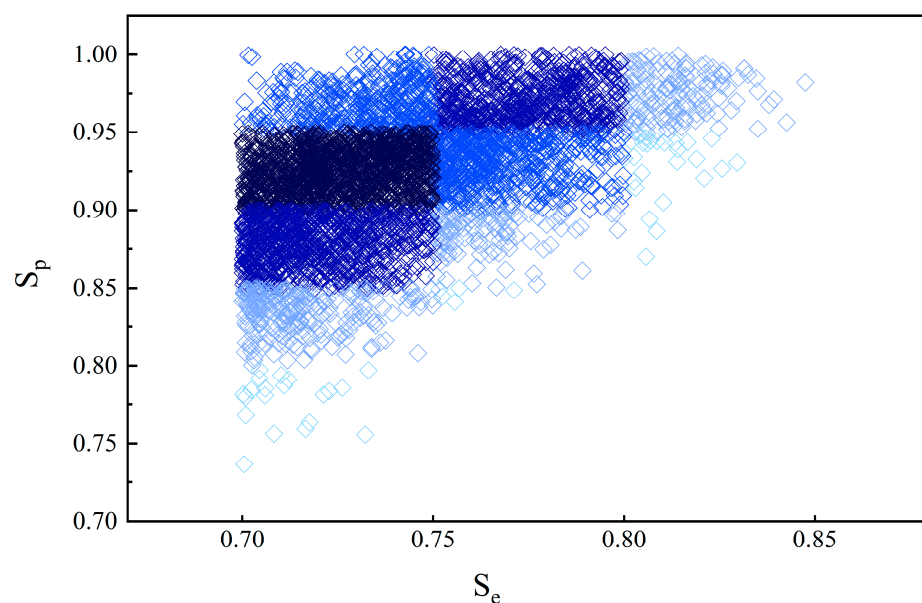


Figure 8. Distribution of evaluation index values for candidate solution samples. Darker colors indicate larger sample sizes.

Specifically analyzing the simulated concentration results corresponding to typical candidate solution samples could help us to further understand the differences between the two indexes. The results of comparing the simulated and observed concentrations for a series of solution samples are given in Figure 9. In each error group, the concentration distribution curve in green corresponded to a sample with a S_p score of about 1.00. The green samples in Figure 9a–c can be considered very close to the observed concentration curve if only the curve shape is concerned without considering the S_p score. However, the relative error of concentration corresponding to these three solution samples was actually quite different. Similarly, the orange sample in Figure 9a was similar to the green sample in Figure 9c, if only the concentration curve was concerned. In fact, the S_e index score of the former was better than S_p , while the S_p index score of the latter was better than S_e .

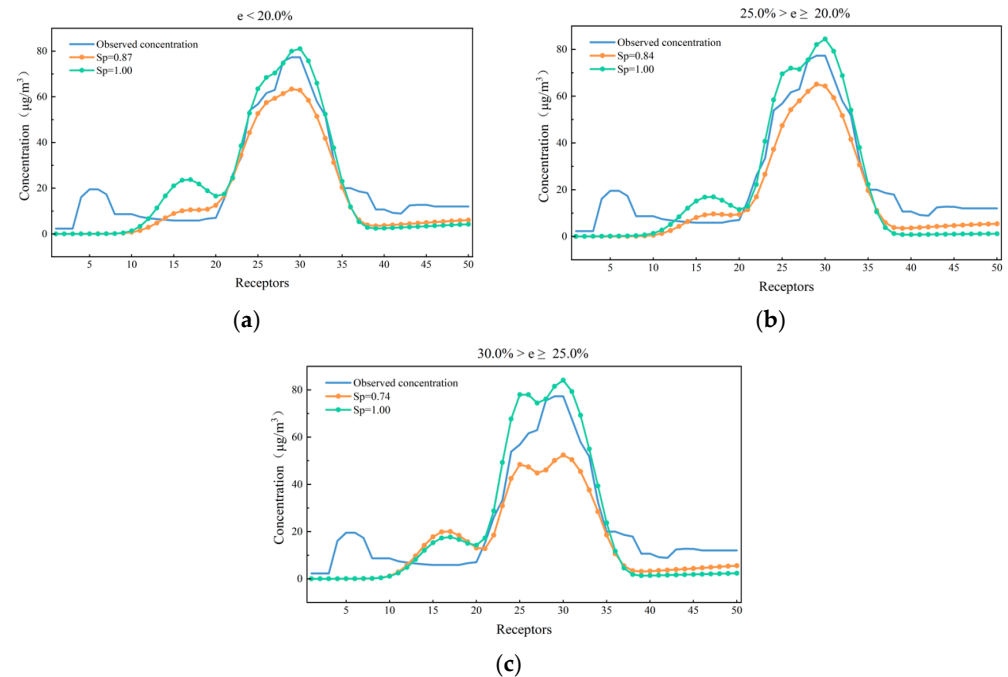


Figure 9. Simulated concentrations for solution samples with different S_e and S_p : (a) samples of $e < 20.0\%$; (b) samples of $25.0\% > e \geq 20.0\%$; (c) samples of $30.0\% > e \geq 25.0\%$.

It is worth noting that the solution samples with the same S_e scoring level have a wider range of S_p values, which implies that the corresponding combinations of emission rates are also somewhat different for source analysis. The corresponding inverse calculated emission rates for the candidate solution samples in Figure 9 are shown in Figure 10. The S_p scores of the three groups of green samples in Figure 10 were close to 1.00, and their corresponding emission rates of inversion were relatively close to each other, which were 77.5 g/s, 72.6 g/s, and 73.5 g/s, respectively, although there was a large difference in the S_e scores of these three samples. The orange sample had the lowest S_p score among the samples with corresponding S_e levels, and its corresponding sum of emission rates of inversion were lower compared with those of the green sample, which were 51.4 g/s, 52.8 g/s, and 43.6 g/s, respectively. A combined comparison of Figures 9 and 10 revealed that the orange samples had a similar S_e level to the green samples, but their simulated concentration and emission rates of inversion were smaller than those of the green samples. It can be seen that the emission rates of inversion corresponding to samples with the same S_e score level also had a large range of values.

Based on the above analysis, the spatial matching degree of solution samples with similar relative errors was very different because the solution of source analysis is not unique. However, with the improvement of the quality of the solution samples, the consistency of S_e and S_p was improved. It infers that the spatial distribution matching degree is a useful supplement to solution quality evaluation. These two indexes can effectively evaluate the quality of candidate solution samples from different perspectives, so it is necessary to comprehensively evaluate the quality of solution samples by using S_e and S_p together.

3.4. Final Results for the Two Cases

The final emission rate estimation results were obtained by statistics of the top 1% solution samples with better index S_{match} score, as mentioned above. In the evaluation results of candidate solution sample quality, the S_{match} of the top 1% solution samples with higher comprehensive scores in the two cases were 0.92–0.97 and 0.92–0.99, respectively. The matching degree between the simulated concentration and observed concentration was high, and the simulated concentration corresponding to the optimal solutions is shown in

Figure 11. As can be seen from Figure 11, the simulated concentration of the better solution could basically reproduce the concentration peak near the receptors 20–35 on the mobile observation road section, which indicated that the source of most NO_x substances in this road section had been captured. The small concentration peak near the receptors 1–11 in case K-2 were not simulated, which would be further analyzed in the discussion section.

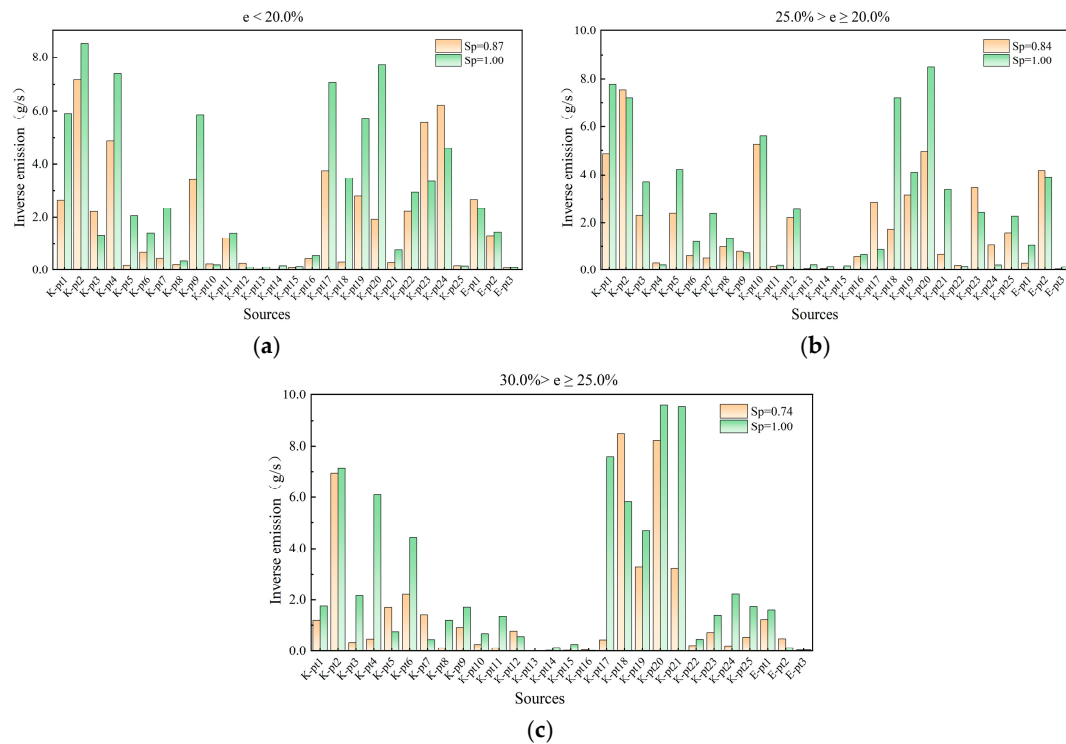


Figure 10. Emission rates for solution samples with different S_e and S_p : (a) samples of $e < 20.0\%$; (b) samples of $25.0\% > e \geq 20.0\%$; (c) samples of $30.0\% > e \geq 25.0\%$.

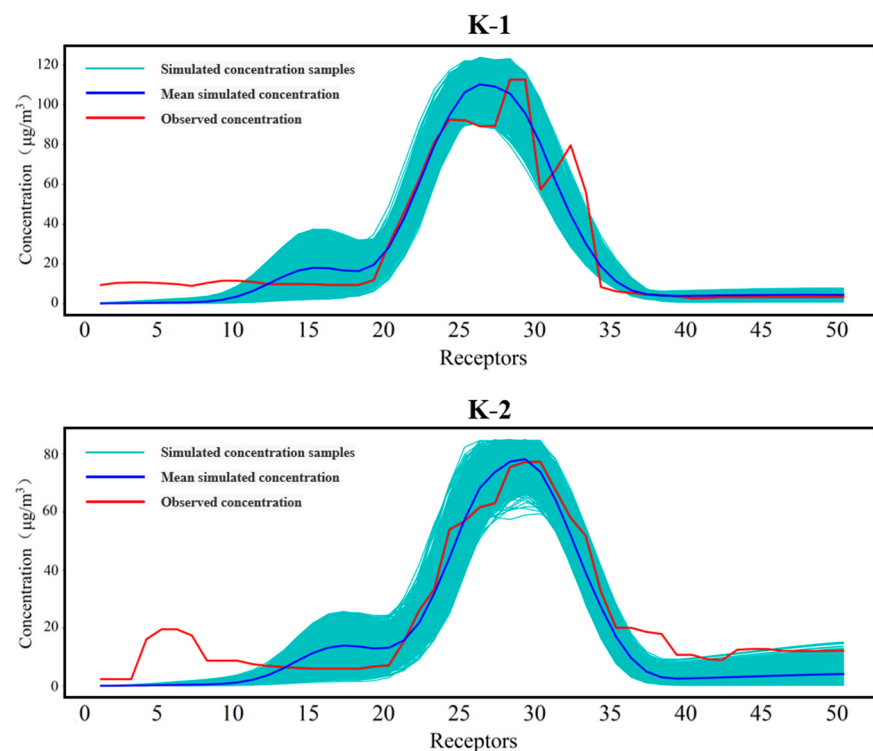


Figure 11. Simulated concentrations of the better solution for cases K-1 and K-2.

The emission rates of NO_x were solved using the inverse calculation according to the specific conditions of the two cases. The local NO_x emission rate of case K-1 was 38.9–108.3 g/s, and the average emission rate was 69.8 g/s, of which the NO_x emission rate from Enterprise K was 66.9 g/s. The total NO_x local emission rate of case K-2 was 38.3–114.2 g/s, and the average emission rate was 70.8 g/s, of which the NO_x emission rate from Enterprise K was 68.2 g/s.

As can be seen from the comparison of the emission rate estimation results for each source with the reference emission rate given in Figure 12, the emission rate estimation results for both cases were slightly smaller than the reference emission rate, and the R values for most of the sources were in the range of 0.80–1.20. For NO_x emissions from Enterprise K, case K-1 had an R value of about 0.90, and K-2 had an R value of about 0.92. The emission rate estimation results that were significantly higher than the reference emission rate were mainly the emission sources of Enterprise A and Enterprise B. A few sources with $R < 0.80$, such as K-pt17 and K-pt18, had significantly lower inverse emission rates than reference emission rates in both cases. The distribution of the NO_x concentration contribution under the reference emission rate condition is shown in Figure 6, where the emission of the sources K-pt17 and K-pt18 mainly affected the receptors numbered between 10–20. However, the observed concentration of NO_x on this road section was not high, so the emission rate obtained by reverse calculation was smaller than the reference emission rate. Some emission sources had online monitoring data, also given in Figure 12. Compared with the emission rate data of the online monitoring, the actual emission rates of K-pt17 and K-pt18 were much smaller than the reference emission rates used in the calculation, but the rates were closer to the emission rates of the inverse calculation in both cases. For many pollution sources with R values close to 1.00, such as K-pt2–K-pt12, their inverse emission rates were greater or less than the measured emission rates; there was no clear law. This was because the NO_x emission contributions of these sources have a high overlap rate on the mobile observation road section and were mainly concentrated around the receptors 20–35 (Figure 6). This showed that the emission rate solution was not unique, which is to say, NO_x emissions from these sources can be substituted for each other on the concentration of acceptor points.

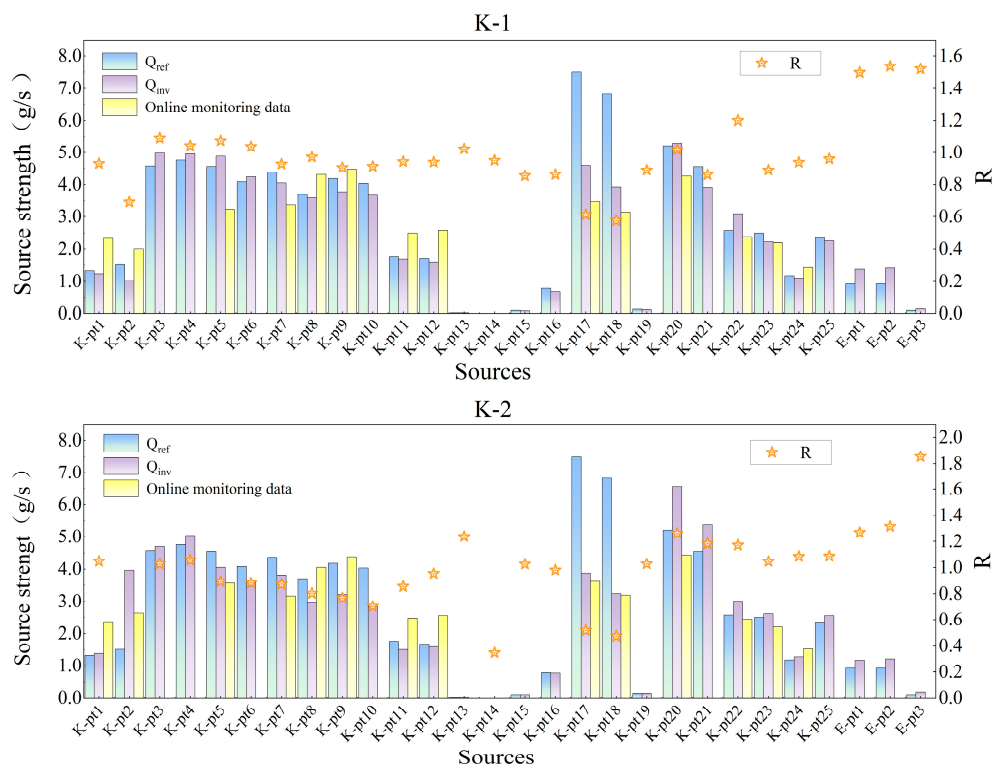


Figure 12. Q_{inv} compared with Q_{ref} and emission rates for online monitoring.

Figure 13 presents the contribution of each source to the concentration at receptors under the final emission rate estimation results, and it can be seen that the NO_x concentration at the receptors 20–35 in the optimal solution was mainly from the emissions of the sources K-pt2–K-pt12 and K-pt22.

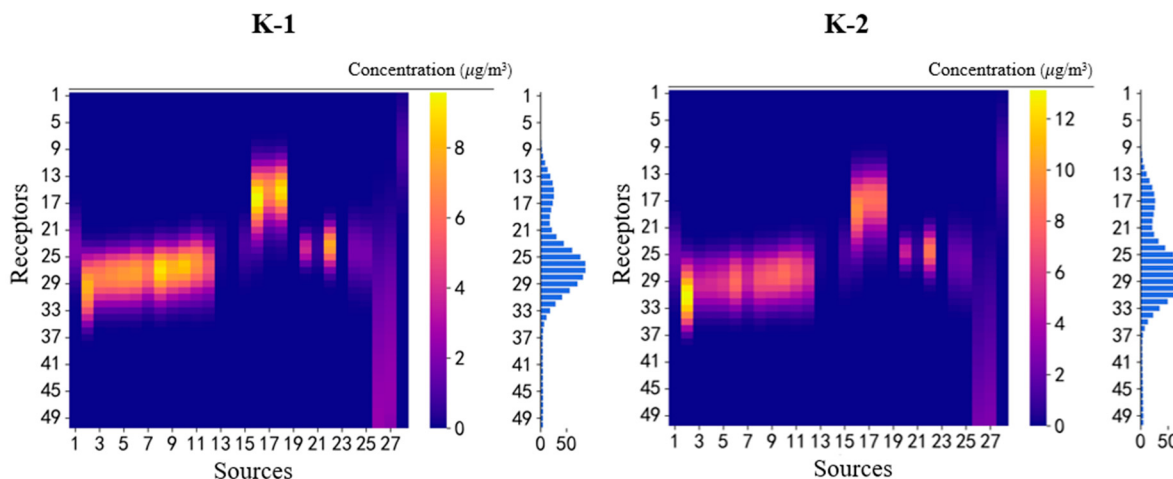


Figure 13. Concentration contributions from each source for final results. The heat map on the left side of each subgraph represents the concentration–response values at each receptor from each source’s emission under inverse calculated emission rate, $\mu\text{g}/\text{m}^3$. The bars on the right side represent the total concentration at each receptor under the inverse calculated emission rate, $\mu\text{g}/\text{m}^3$.

3.5. Discussion

In this study, mobile observation techniques are applied to quantitative source analysis methods in the hope of providing effective and powerful technical assistance for pollution control in the air environment. At present, there are very few actual source analysis studies based on mobile monitoring, and the error of the actual source analysis results of Zhou [46] in the previous stage was very large. Therefore, our study was very cautious when establishing the small-scale traceability method. Through comprehensive and in-depth analysis, combined with the intense fluctuation of local meteorological observation mentioned in literature [47–49], we carried out an ideal experiment test, and the results showed that the error of simulated concentration when the wind direction shifted by 5° – 30° was as high as 30.2–120.4%. With reference to the observations in Section 2.1 and the simulation results in Section 3.1, we were more convinced that the influence of fluctuations in meteorological conditions was worthy of attention in local microscale source analysis. In the study of Zhou [46], the meteorological data measured by the mobile sampling vehicle would be biased under the influence of factors such as vehicle orientation and speed, so the huge source analysis error was probably caused by the swing of meteorological conditions. The method adopted in this study was a random search based on fine-tuning of meteorological parameters to solve this effect. The final inverse result was very close to the real-time online data, which effectively improves the accuracy of source analysis, but this is obviously not the only method. CALPUFF only used the WRF/CALMET meteorological field as input in the simulation. If more accurate local ground meteorological observation data can be introduced, it will also be helpful to improve the inverse calculation accuracy. In terms of local microscale simulation, the recommended models also include not only the models used in this paper, but AERMOD, CFD, and other simulation methods have also been used in related studies. In conclusion, no matter what simulation methods are used, improving the simulation accuracy of meteorological fields should be a feasible direction to explore. However, for a small-scale real-time source analysis problem such as this study, there may still be deviation from the actual meteorological conditions. Thus, for forward diffusion simulations in inversion studies, it is not mandatory to use a simulation scheme based on variable meteorological conditions.

Furthermore, as already mentioned in Section 2.2, complete a priori information on potential sources is very important. The adjustment of the forward diffusion simulation scheme is not the only factor to be considered when the inversion effect of the emission rate is not ideal. If the error of source analysis is still unacceptable under the optimal simulation scheme, it is necessary to re-check whether the emission source is completely collected. The analysis in Section 3.4 mentioned the small NO_x concentration peak near the receptors 1–11 that was not reproduced in the inversion. This was not the result of an inherent error in the source analysis; it was also related to the emission source list created prior to the inversion. As described in Section 2.2.1, in the source analysis for Enterprise K, sources from Enterprise A and Enterprise B were added in order to create a complete list of emission sources. They were added to the list because they were considered likely contributors to the NO_x concentration peaks near receptors 20–35 under real-time meteorological conditions. In contrast, the NO_x concentration peak near receptors 1–11 was not considered related to the emissions from Enterprise K. Therefore, the source of this peak was not included in the list of emission sources and, thus, could not be explained by the results of inversion. It can also be seen from the simulation results in Figures 11 and 13 that NO_x emissions from Enterprise K, Enterprise A, and Enterprise B really did not affect the road section where receptors 1–11 were located.

In addition, through the analysis of the estimation results of source intensity in Section 3.4, it can be concluded that when pollution sources are distributed densely, the characteristics of the impact of these pollution sources on environmental concentration may be similar under any meteorological conditions. For example, the positions of K-pt2 to K-pt12 (or K-pt17 to K-pt18) were very close in Figure 1. These sources also had a high consistency in concentration contribution characteristics to receptors. As shown in Figure 6, the range of receptors generating the concentration response in dashed box A (or dashed box B) was very close. Although the range of receptors generating concentration response in dashed box C partially overlapped with dashed box A, the range of receptors corresponding to the two dashed frames was significantly different due to the different spatial distribution locations of the two groups of sources, K-pt2–K-pt12 and K-pt26–K-pt27. That is to say, the concentration contributions from these centrally distributed sources are interchangeable in terms of fitting simulated concentrations to observed concentrations. It may be difficult to achieve a refined differentiation of the emission contribution of each of these sources at present. Therefore, the inverted emission rates of individual sources in the final results, such as any of K-pt2–K-pt12 or K-pt17–K-pt18, may not be helpful to managers. Instead, the emission rate estimation results of inversion for sources with different emission impact characteristics, such as the sum of the emission rates for each of the two groups of sources, K-pt2–K-pt12 and K-pt17–K-pt18, as well as the sum of the emission rates for all the sources, have more reference value.

The inversion method we developed is based on obtaining the most complete source list data possible in the study area. Overall, the application of this inversion method has no regional limitation and is easy to promote in environmental management.

4. Conclusions

A set of emission rate estimation methods for air pollutants was explored and established based on mobile observation experiments. The whole research process included the identification of inversion cases, the comparison of forward diffusion simulation schemes, the testing and analysis of inversion search strategy, the evaluation of solution sets, and the statistics of results. The NO_x emission cases from a petrochemical enterprise in Shanghai were selected, and the pollution sources from related enterprises upwind were also included in the source list during the source analysis process. Finally, an inverse calculated algorithm based on random search and solution set optimization was adopted.

Fluctuations in meteorological conditions, especially wind direction, had a large impact on the simulation of dispersion at the enterprise scale. If the inversion error is still large under the premise of ensuring the completeness of the source list, the simulation

accuracy of the meteorological field should be improved. It is suggested to adopt the forward diffusion simulation scheme with variable meteorological conditions, which will fine-tune the meteorological parameters based on the actual meteorological condition in this paper. This method can restore the pollutant transport process during mobile observation to a greater extent and is more conducive to obtaining a solution close to the true emission rate. The search efficiency of the random search algorithm had a certain degree of randomness, but the minimum error of the simulated concentration tended to decrease in general with the increase in the sample size. Therefore, under the premise of improving computational efficiency, a larger sample size should be recommended. The consistency of the simulated concentration with the observed concentration in numerical and spatial distribution characteristics should be considered in the evaluation of the quality of the emission rate solution samples. The two indexes, S_e and S_p , can evaluate the quality of the candidate solution samples more efficiently from different perspectives, so it is necessary to adopt the comprehensive evaluation index S_{match} .

The NO_x local average emission rates for the two cases were 69.8 g/s and 70.8 g/s after inverse calculation. The observed concentrations in some road sections were not reproduced in the inversion, which is due to the fact that NO_x emissions from Enterprise K, Enterprise A, and Enterprise B really do not affect these road sections. The emission contributions of densely distributed pollution sources can be substituted for each other in terms of fitting simulated concentrations. The emission rates of emission sources with different emission impact characteristics have more reference value for environmental managers. In the results of inversion, the emission rates of most of the pollution sources were close to the daily emission levels, while the emission rates of individual sources differed from the historical monitoring data but were closer to the actual emission rate data from online monitoring. Overall, the results of the inversion reflected that the NO_x emissions of the target enterprises were at normal levels. It also initially proved that the source analysis method in this paper is feasible to be used in the daily monitoring and supervision of pollutant emissions.

The pollution emission rate estimation method needs to be further optimized through more case studies and tests in the future. The effect of the uncertainty of the meteorological field under the microscale conditions deserves further exploration because the fluctuation of the atmospheric flow field will increase the difficulty of accurate simulation of the diffusion model. Attempts can be made to realize the refined reconstruction of 3D meteorological fields using CFD models.

Author Contributions: Software, validation, formal analysis, data curation, writing—original draft preparation, visualization, X.C.; conceptualization, methodology, investigation, resources, writing—review and editing, supervision, project administration, Q.Y.; formal analysis, resources, writing—review and editing, supervision, project administration, W.M.; writing—review and editing, supervision, funding acquisition, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Natural Science Foundation of Shanghai Municipality, grant number 22ZR1407700.

Data Availability Statement: The datasets presented in this article are not readily available because the data are part of an ongoing study. Due to the intellectual property rights of the partners, the data involved in this study is not publicly available.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Adams, M.D.; Kanaroglou, P.S. Optimized monitor reduction for an industrial air pollution monitoring network. In Proceedings of the 13th International Conference on Environmental Science and Technology, Athens, Greece, 5–7 September 2013.
2. Efthimiou, G.C.; Kovalets, I.V.; Venetsanos, A. An optimized inverse modelling method for determining the location and strength of a point source releasing airborne material in urban environment. *Atmos. Environ.* **2017**, *170*, 118–129. [[CrossRef](#)]
3. Singh, K.S.; Kumar, P.; Turbelin, G. Uncertainty characterization in the retrieval of an atmospheric point release. *Atmos. Environ.* **2017**, *152*, 15234–15250. [[CrossRef](#)]

4. Hong, Z.Y.; Li, M.Z.; Wang, H. Characteristics of atmospheric volatile organic compounds (VOCs) at a mountainous forest site and two urban sites in the southeast of China. *Sci. Total Environ.* **2019**, *657*, 1491–1500. [[CrossRef](#)] [[PubMed](#)]
5. Hsu, C.-Y.; Wu, P.-Y.; Chen, Y.-C. An integrated strategy by using long-term monitoring data to identify volatile organic compounds of high concern near petrochemical industrial parks. *Sci. Total Environ.* **2022**, *821*, 153345. [[CrossRef](#)]
6. Wang, J.; Wang, B.; Liu, J. An inverse method to estimate the source term of atmospheric pollutant releases. *Atmos. Environ.* **2021**, *260*, 118554. [[CrossRef](#)]
7. Liu, Y.; Yu, Q.; Huang, Z. Identifying key potential source areas for ambient methyl mercaptan pollution based on long-term environmental monitoring data in an industrial park. *Atmosphere* **2018**, *9*, 501. [[CrossRef](#)]
8. Cheng, K.; Zhao, X.; Zhou, W. Source term estimation with deficient sensors: Traceability and an equivalent source approach. *Process Saf. Environ. Prot.* **2021**, *152*, 131–139. [[CrossRef](#)]
9. Thomson, L.C.; Hirst, B.; Gibson, G. An improved algorithm for locating a gas source using inverse methods. *Atmos. Environ.* **2007**, *41*, 1128–1134. [[CrossRef](#)]
10. Bieringer, P.E.; Rodriguez, L.M.; Vandenberghe, F. Automated source term and wind parameter estimation for atmospheric transport and dispersion applications. *Atmos. Environ.* **2015**, *122*, 206–219. [[CrossRef](#)]
11. Huang, Z.; Wang, Y.; Yu, Q. Source area identification with observation from limited monitor sites for air pollution episodes in industrial parks. *Atmos. Environ.* **2015**, *122*, 1–9. [[CrossRef](#)]
12. Huang, Z.; Yu, Q.; Ma, W.; Chen, L. Surveillance efficiency evaluation of air quality monitoring networks for air pollution episodes in industrial parks: Pollution detection and source identification. *Atmos. Environ.* **2019**, *215*, 116874. [[CrossRef](#)]
13. Xue, Y.; Cui, X.; Li, K. Statistical source analysis of recurring sulfur dioxide pollution events in a chemical industrial park. *Atmos. Environ.* **2023**, *296*, 119564. [[CrossRef](#)]
14. Wang, C.; An, X.; Hou, Q. Development of four-dimensional variational assimilation system based on the GRAPES–CUACE adjoint model (GRAPES–CUACE-4D-Var V1.0) and its application in emission inversion. *Geosci. Model Dev.* **2021**, *14*, 337–350. [[CrossRef](#)]
15. Xu, X.; Wang, J.; Henze, D.K. Constraints on aerosol sources using GEOS-Chem adjoint and MODIS radiances, and evaluation with multisensor (OMI, MISR) data. *J. Geophys. Res. Atmos.* **2013**, *118*, 6396–6413. [[CrossRef](#)]
16. Wang, C.; An, X.; Zhai, S. Tracking sensitive source areas of different weather pollution types using GRAPESCUACE adjoint model. *Atmos. Environ.* **2018**, *175*, 154–166. [[CrossRef](#)]
17. Lushi, E.; Stockie, M.J. An inverse Gaussian plume approach for estimating atmospheric pollutant emissions from multiple point sources. *Atmos. Environ.* **2009**, *44*, 1097–1107. [[CrossRef](#)]
18. Gilbert, E.J.; Khajehnajafi, S. Estimation of Toxic Substance Release. U.S. Patent US6772071B2, 14 November 2002.
19. Allen, C.T.; Young, G.S.; Haupt, S.E. Improving pollutant source characterization by better estimating wind direction with a genetic algorithm. *Atmos. Environ.* **2007**, *41*, 2283–2289. [[CrossRef](#)]
20. Wang, J.; Zhang, R.; Li, J. Locating unknown number of multi-point hazardous gas leaks using Principal Component Analysis and a Modified Genetic Algorithm. *Atmos. Environ.* **2020**, *230*, 117515. [[CrossRef](#)]
21. Kia, S.; Nambiar, M.K.; Thé, J. Machine Learning to Predict Area Fugitive Emission Fluxes of GHGs from Open-Pit Mines. *Atmosphere* **2022**, *13*, 210. [[CrossRef](#)]
22. Mao, S.; Hu, F.; Lang, J. Comparative Study of Impacts of Typical Bio-Inspired Optimization Algorithms on Source Inversion Performance. *Front. Environ. Sci.* **2022**, *10*, 894255. [[CrossRef](#)]
23. Hosseini, B.; Stockie, J.M. Bayesian estimation of airborne fugitive emissions using a Gaussian plume model. *Atmos. Environ.* **2016**, *141*, 122–138. [[CrossRef](#)]
24. Kopka, P.; Wawrzynczak, A.; Borysiewicz, M. Application of the Approximate Bayesian Computation methods in the stochastic estimation of atmospheric contamination parameters for mobile sources. *Atmos. Environ.* **2016**, *145*, 201–212. [[CrossRef](#)]
25. Wang, Y.; Huang, H.; Huang, L. Evaluation of Bayesian source estimation methods with Prairie Grass observations and Gaussian plume model: A comparison of likelihood functions and distance measures. *Atmos. Environ.* **2017**, *152*, 519–530. [[CrossRef](#)]
26. Qiu, S.; Chen, B.; Wang, R. Atmospheric dispersion prediction and source estimation of hazardous gas using artificial neural network, particle swarm optimization and expectation maximization. *Atmos. Environ.* **2018**, *178*, 158–163. [[CrossRef](#)]
27. Septier, F.; Armand, P. A bayesian inference procedure based on inverse dispersion modelling for source term estimation in built-up environments. *Atmos. Environ.* **2020**, *242*, 117733. [[CrossRef](#)]
28. Araki, S.; Iwahashi, K.; Shimadera, H.; Yamamoto, K.; Kondo, A. Optimization of air monitoring networks using chemical transport model and search algorithm. *Atmos. Environ.* **2015**, *122*, 22–30. [[CrossRef](#)]
29. Zheng, Y.; Wang, Y.; Wang, L. Identification Method of Source Term Parameters of Nuclear Explosion Based on GA and PSO for Lagrange-Gaussian Puff Model. *Atmosphere* **2023**, *14*, 877. [[CrossRef](#)]
30. Kong, Y.; Guan, M.; Zheng, S. Locating Hazardous Chemical Leakage Source Based on Cooperative Moving and Fixing Sensors. *Sensors* **2019**, *19*, 1092. [[CrossRef](#)]
31. Solis, F.J.; Wets, R.J.B. Minimization by random search techniques. *Math. Oper. Res.* **1981**, *6*, 19–30. [[CrossRef](#)]
32. Spall, J.C. Introduction to Stochastic Search and Optimization. Estimation, Simulation, and Control (Spall, J.C.; 2003) [book review]. *IEEE Trans. Neural Netw.* **2007**, *18*, 964–965.
33. Bergstra, J.; Bengio, Y. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305.

34. Dyer, M.E.; Frieze, A.M. Computing the Volume of Convex Bodies: A Case Where Randomness Provably Helps. *Proc. Symp. Appl. Math.* **1991**, *44*, 123–169.
35. Vavasis, S.A. Complexity Issues in Global Optimization: A Survey. In *Handbook of Global Optimization*; Springer: New York, NY, USA, 1995; pp. 27–41.
36. Zlochin, M.; Birattari, M.; Meuleau, N.; Dorigo, M. Model-Based Search for Combinatorial Optimization: A Critical Survey. *Ann. Oper. Res.* **2004**, *131*, 373–395. [[CrossRef](#)]
37. Miller, D.J.; Actkinson, B.; Padilla, L. Characterizing Elevated Urban Air Pollutant Spatial Patterns with Mobile Monitoring in Houston, Texas. *Environ. Sci. Technol.* **2020**, *54*, 2133–2142. [[CrossRef](#)]
38. Van den Bossche, J.; Peters, J.; Verwaeren, J. Mobile monitoring for mapping spatial variation in urban air quality: Development and validation of a methodology based on an extensive dataset. *Atmos. Environ.* **2015**, *105*, 148–161. [[CrossRef](#)]
39. Järvi, L.; Kurppa, M.; Kuuluvainen, H. Determinants of spatial variability of air pollutant concentrations in a street canyon network measured using a mobile laboratory and a drone. *Sci. Total Environ.* **2023**, *856*, 158974. [[CrossRef](#)]
40. Solomon, A.P.; Vallano, D.; Lunden, M. Mobile-platform measurement of air pollutant concentrations in California: Performance assessment, statistical methods for evaluating spatial variations, and spatial representativeness. *Atmos. Meas. Tech.* **2020**, *13*, 3277–3301. [[CrossRef](#)]
41. Wallace, J.; Corr, D.; DeLuca, P. Mobile monitoring of air pollution in cities: The case of Hamilton, Ontario, Canada. *J. Environ. Monit.* **2009**, *11*, 998–1003. [[CrossRef](#)]
42. Zwack, L.M.; Paciorek, C.J.; Spengler, J.D. Modeling spatial patterns of traffic-related air pollutants in complex urban terrain. *Environ. Health Persp.* **2011**, *119*, 852–859. [[CrossRef](#)]
43. Padilla, L.E.; Ma, G.Q.; Daniel, P. New methods to derive street-scale spatial patterns of air pollution from mobile monitoring. *Atmos. Environ.* **2021**, *270*, 118851. [[CrossRef](#)]
44. Shairsingh, K.K.; Jeong, C.; Wang, M.J. Characterizing the spatial variability of local and background concentration signals for air pollution at the neighbourhood scale. *Atmos. Environ.* **2018**, *183*, 57–68. [[CrossRef](#)]
45. Fan, G. Reverse Traceability of Pollutants in Industrial Parks Based on Mobile Monitoring. Master's Thesis, Dalian University of Technology, Dalian, China, 2023.
46. Wang, Z. Research on Air Pollution Traceability in Chemical Park Based on Mobile Monitoring System. Master's Thesis, Zhejiang University, Hangzhou, China, 2023.
47. Dongqin, Z.; Gang, H.; Jie, S. A novel spatio-temporal wind speed forecasting method based on the microscale meteorological model and a hybrid deep learning model. *Energy* **2024**, *288*, 129823.
48. Statistics and Meteorology of Air Pollution Episodes over the South African Highveld Based on Satellite–Model Datasets. *J. Appl. Meteorol. Climatol.* **2017**, *56*, 1583–1594. [[CrossRef](#)]
49. Jiang, W.-M.; Liu, H.-N.; Zhang, N. *Air Pollution Meteorology*; Nanjing University Press: Nanjing, China, 2021.
50. Fernandes Rafael, S.; Lopes, D.; Coelho, S. The air pollution modelling system URB AIR: How to use a Gaussian model to accomplish high spatial and temporal resolutions. *Air Qual. Atmos. Health* **2021**, *14*, 1969–1988. [[CrossRef](#)]
51. Biggart, M.; Stocker, J.; Doherty, R.M. Street-scale air quality modelling for Beijing during a winter 2016 measurement campaign. *Atmos. Chem. Phys.* **2020**, *20*, 2755–2780. [[CrossRef](#)]
52. Siahpour, G.; Jozi, S.A.; Orak, N. Estimation of environmental pollutants using the AERMOD model in Shazand thermal power plant, Arak Iran. *Toxin Rev.* **2022**, *41*, 1269–1279. [[CrossRef](#)]
53. Rzesutek, M. Parameterization and evaluation of the CALMET/CALPUFF model system in near-field and complex terrain—Terrain data, grid resolution and terrain adjustment method. *Sci. Total Environ.* **2019**, *689*, 31–46. [[CrossRef](#)]
54. Jose, R.S.; Perez-Camanyo, J.L. High-resolution impacts of green areas on air quality in Madrid. *Air Qual. Atmos. Health* **2023**, *16*, 37–48. [[CrossRef](#)]
55. Roscioli, R.J.; Yacovitch, I.T.; Floerchinger, C. Measurements of methane emissions from natural gas gathering facilities and processing plants: Measurement methods. *Atmos. Meas. Tech.* **2015**, *8*, 2017–2035. [[CrossRef](#)]
56. Wu, Z.; Liu, C. Source depletion analogy for reactive plume dispersion over schematic urban areas. *Atmos. Environ.* **2018**, *190*, 226–231. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.