

Multi-point Sources Continuous Release Inversion Based on Improved Four-Dimensional Variation Method

HAN CHAOSHUAI (✉ hanzaoshuai@163.com)

State Key Laboratory of NBC Protection for Civilian <https://orcid.org/0000-0002-0764-7654>

Xuezheng Zhu

State Key Laboratory of NBC Protection for Civilian

Jin Gu

State Key Laboratory of NBC Protection for Civilian

Qinwen Zuo

State Key Laboratory of NBC Protection for Civilian

Research Article

Keywords: Convection-Diffusion Equation, Multi-point Sources, Source Term Inversion, Four-Dimensional Variation, Numerical Simulation

Posted Date: October 21st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-989442/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Multi-point Sources Continuous Release Inversion Based on Improved Four-Dimensional Variation Method

Chaoshuai Han^{1,2}, Xuezheng Zhu^{1,*}, Jin Gu¹, Qinwen Zuo²

1 Institute of NBC Defence, Beijing, China

2 State Key Laboratory of NBC Protection for Civilian, Beijing, China

The corresponding author: Xuezheng Zhu, email: zxm4x@aliyun.com.

Funding: Ministry of Science and Technology of the People's Republic of China (2018YFC0214100)

Abstract: the continuous release of multi-point sources is one of the most common cases in the field of air pollution. In order to solve the problem of multi-point sources continuous release inversion, a source intensity and location estimation method based on improved Four-Dimensional Variation is proposed. Firstly, by constructing the diffusion equation of multi-point sources continuous release and the monitoring concentration matrix, the source term inversion model of multi-point sources continuous release is formed. Secondly, the joint transformation method and Taylor series upwind difference method are used to solve the convection-diffusion equation of multi-point sources continuous release, and the numerical solution format of the forward problem is formed through simulation analysis. Furthermore, based on the numerical solution scheme of the forward problem, a four-dimensional variation inversion algorithm for multi-point sources continuous release is constructed, and the flower pollination algorithm is used to improve the inversion accuracy and computational efficiency. Finally, the applicability, feasibility, advantages and disadvantages of the improved four-dimensional variation algorithm are verified by numerical simulation analysis. It is found that the improved four-dimensional variation algorithm can realize the source term inversion under three conditions: the location is known and the intensity is unknown, the intensity is known and the location is unknown, the intensity and the location are all unknown, which the accuracy and computational efficiency can basically meet the actual needs.

Keywords: Convection-Diffusion Equation; Multi-point Sources; Source Term Inversion; Four-Dimensional Variation; Numerical Simulation

Introduction

As an inverse problem of air pollution diffusion prediction, source term inversion (STI) has important practical significance for accurately predicting the temporal and spatial transmission of pollutants, effectively alleviating and curbing pollution sources, which is a key basic technology in the field of air diffusion (Andrew J. Annunzio et al., 2012). Source term information mainly includes source type, source location, source intensity, release time and release mode (continuous and instantaneous), where the complexity and nonlinear characteristics of diffusion process are important reasons for the difficulty of direct inverse derivation and solution of STI (Castelli S.T., 2017). At the same time, there can be no single point source pollution in the actual environment. The expansion from single source inversion to multi-source inversion has become the focus and difficulty of current research. Therefore, this paper carries out theoretical and Simulation Research on multi-point sources continuous

40 release inversion, and then carries out multi-point sources instantaneous and mixed release
41 inversion.

42 For multi-point sources continuous release inversion problem, the key and difficulty lies
43 in the model representation of multiple generalized Dirichlet functions in forward problem
44 solving and inverse problem tracing. In solving the forward problem of convection diffusion
45 equation, Wei T. et al. (2016) transformed the convection-diffusion equation into an integral
46 equation by using Laplace transform, and derived the approximate solution of the typical
47 diffusion equation by orthogonalization and solving Chebyshev polynomials. Based on the
48 relationship between Taylor series expansion and volume average flow, Pan Dartzi (2017)
49 reconstructed the local flow field with high-order piecewise polynomials, and then proposed a
50 spatial high-order finite volume method for solving convection diffusion equations. François et
51 al. (2021) proposed a new super conjugate collocation scheme based on the central finite
52 difference scheme for the forward solution of convection-diffusion equations. Wang Lei et al.
53 (2021) constructed an improved regularized lattice Boltzmann model to solve the convection-
54 diffusion equation with source term information by Chapman-Enskog analysis of the total
55 distribution function, which was found having higher stability and accuracy than other
56 traditional models. Different from others, Kazakov et al. (2021) used the method of
57 characteristic series and coefficient recursion to obtain the exact solution model of convection-
58 diffusion equation, and constructed the approximate solution algorithm by boundary element
59 method and dual reciprocity method.

60 In terms of tracing the source of the inverse problem of convection-diffusion equation,
61 Viktor et al. (2010) proposed a radial basis function meshless method based on finite difference
62 method and Tikhonov, which has the advantage of avoiding singular integral. Then Reza
63 pourgholi et al. (2019) further improved the stability of the method and used it to solve the
64 nonlinear convection-reaction-diffusion equation with unknown source function. Hamba et al.
65 (2012) studied the filtering effect of convection-diffusion equation from two aspects of
66 concentration field and concentration flux, and obtained the value range of filtering width
67 suitable for point source reverse identification. Sun C. L. et al. (2017) proposed a multi-source
68 inversion method based on error functional minimum and homotopy regularization to solve the
69 spatial correlation diffusion coefficient and source coefficient in multiple time fractional
70 diffusion equations. Liu C. S. et al. (2019) proposed a super-position of homogenization
71 functions method (SHFM) for source term identification of nonlinear convection-diffusion
72 equation under unknown boundary conditions, which was applied to one-dimensional diffusion
73 equation. Ruan et al. (2020) proposed a STI algorithm based on Laplace transform and
74 Tikhonov regularization. Through the uniqueness of the forward problem, the convergence of

75 the inverse problem and numerical simulation analysis, its feasibility in time fractional diffusion
76 equation was verified. Liu Hailiang (2021) proposed a direct discontinuous Galerkin (DDG)
77 method to characterize the non physical oscillation phenomenon in the convection-diffusion
78 equation, which has strong applicability to improve the stability of the inversion algorithm.

79 In recent years, with the rapid development of computer technology, the huge
80 computational demand brought by adjoint model has been effectively solved, which provides a
81 key foundation for the popularization and application of Four-Dimensional Variation (4DVAR).
82 Compared with regularization methods and other Assimilation Algorithms, 4DVAR has the
83 advantages of assimilating multiple continuous time windows and stronger universality, which
84 has been applied to various fields such as air pollution, tidal simulation, dust emission, Kuroshio
85 prediction and has attracted more and more attention (Genc D. et al., 2010; Zhang J. C. et al.,
86 2014; Jianbing Jin et al., 2018; Kim Sung min et al., 2019; Yasumasa Miyazawa et al., 2020).
87 In terms of chemical hazard diffusion STI, Kuo et al. (2016) used polynomial expansion method
88 to discretize the differential equation in two dimensions of time and space for the two-
89 dimensional time-varying heat source identification problem, and then solved the discretized
90 equations by 4DVAR, and finally obtained the ideal heat source function. Huang S. X. et al.
91 (2018) proposed a pollution source dynamic inversion and grid quantitative traceability
92 algorithm based on CAMx numerical prediction model, which can quickly and quantitatively
93 trace the emission source and contribution rate of air pollution in the next 7 days. Liu Y. et al.
94 (2017, 2019) used 4DVAR Cost Gradient Function and Truncated Total Teast Squares (TTLS-
95 VAR) to regularize and correct the error between diffusion prediction operator and observation
96 value, which reduced the error of inversion algorithm and improved the accuracy compared
97 with the basic 4DVAR algorithm. Hassan et al. (2020) proposed a quantitative traceability
98 algorithm combining improved variational iteration and Laplace transform, and applied it to the
99 inverse solution of fractional nonlinear convection-diffusion equation for the first time. Naus et
100 al. (2021) applied 4DVAR to the inversion of Methyl Chloroform (MCF) distribution in the
101 atmosphere, studied the interannual variation of MCF from 1998 to 2018, and then put forward
102 scientific and reasonable strategies for emission reduction regulation.

103 According to the former study, although the research and application of 4DVAR is
104 relatively mature, there are few related studies in the case of multi-point sources and continuous
105 release. It is of great research value to use 4DVAR to solve the inversion problem of multi-point
106 sources continuous release. In addition, although most 4DVAR algorithms are composed of
107 target functional, gradient model, numerical prediction tangent linear operator, background
108 error covariance matrix and observation error covariance matrix, the construction methods and
109 suitability of each module vary greatly in different fields and situations. How to build a suitable

110 4DVAR algorithm becomes the key to solve the inversion problem of continuous release of
 111 multi-point sources. Therefore, for the problem of continuous release inversion of multi-point
 112 sources, a source term inversion algorithm based on improved 4DVAR is proposed in this paper.
 113 Firstly, based on the joint transformation method and Taylor series upwind difference method,
 114 the analytical solution and numerical solution models of multi-point sources continuous release
 115 convection-diffusion equation are constructed respectively. Secondly, based on the research of
 116 the forward problem, the improved 4DVAR algorithm is designed from five dimensions:
 117 gradient model, numerical prediction tangent linear operator, background error covariance
 118 matrix, observation error covariance matrix and intelligent optimization iterative algorithm.
 119 Finally, the advantages and disadvantages of the improved 4DVAR algorithm are verified by
 120 numerical simulation and field tracer test.

121 1 Modeling of Multi-point Sources Continuous Release Inversion Problem

122 Usually, the diffusion of continuous release is the most common and practical. In the case
 123 of continuous release, the pollution source continuously and evenly releases hazardous gases
 124 or aerosols, and the released hazardous substances continue to diffuse along the wind direction.
 125 According to the law of mass conservation, combined with the convection diffusion equation
 126 and the calculation formula of dry and wet sedimentation effect (Pan Y. P. et al., 2015), without
 127 considering the chemical reaction between heterogeneous hazards, the basic equation of the
 128 migration mode of continuous release from multi-point sources is as follows:

$$129 \begin{cases} \frac{\partial C}{\partial t} = \nabla(\mathbf{D}\nabla C) - \mathbf{v}\nabla C - (v_d + I \cdot l)C + \sum_{i=1}^n Q_i \delta(|x - x_i| + |y - y_i| + |z - z_i|) \\ C(x, y, z, 0) = \varphi(x, y, z) \\ C(s, k, h, t) = 0 \\ (t \geq 0, 0 \leq x \leq s, 0 \leq y \leq k, 0 \leq z \leq h) \end{cases} \quad (1)$$

130 Where, $C=C(x,y,z,t)$ is the concentration at any point (x, y, z) in the space at time t ; $\mathbf{D}=(D_x,$
 131 $D_y, D_z)$ is the atmospheric diffusion coefficient tensor; $\mathbf{v}=(v_x, v_y, v_z)$ is the convective velocity
 132 vector; v_d is the aerosol dry deposition rate; I is the intensity of rain (snow); l is the removal
 133 rate of hazardous substances in the air, which is obtained according to the mass concentration
 134 of hazardous substances in rain (snow) particulate phase and atmospheric particulate phase; n
 135 is the number of hazard sources; Q_i is the source intensity of the i th continuous release point;
 136 (x_i, y_i, z_i) is the position coordinate of the i th continuous release point source; δ is Dirac
 137 function.

138 STI is the inverse tracing of the multi-point sources convection-diffusion equation in
 139 equation (1). However, due to the nonlinear characteristics of multi-point sources diffusion, the
 140 number, location coordinates and intensity expressions of hazard sources cannot be obtained by
 141 reverse derivation. Therefore, the inverse problem of hazard source number, location

142 coordinates and intensity identification needs to be formed according to the concentration
 143 distribution data of hazards at different times and combined with equation (1). Assume the
 144 hazardous substance monitoring concentration matrix of any monitoring point j (x_j, y_j, z_j) at
 145 each observation time is:

$$146 \quad c(x_j, y_j, z_j, T) = [c_1, c_2, \dots, c_M], \quad T = [t_1, t_2, \dots, t_m] \quad (2)$$

147 Where, T is the observation time array. In this way, the inversion problem of multi-point
 148 sources continuous release source term can be formed from equations (1) and (2). The
 149 parameters to be solved are ($n, Q[\], X_0[\], Y_0[\], Z_0[\]$). Next, the solution format of the forward
 150 problem is given, and then the STI problem will be solved by improved 4DVAR.

151 2 Solution of Forward Problem

152 2.1 Analytical solution based on joint transformation method

153 Assuming that the hazard is simultaneously released at multiple locations, the
 154 concentration of the hazard at any point (x, y, z) in the space at time t is determined by the sum
 155 of the diffusion concentration of each hazard source at that location at time t . Using Fourier
 156 transform and inverse Laplace transform (Karl Barnes, 2017), the analytical expression of
 157 multi-point sources continuous release convection-diffusion equation can be obtained, as shown
 158 in the following formula:

$$159 \quad C(x, y, z, t) = \frac{\sum_{i=1}^n \sqrt{\pi} Q_i}{2\sqrt{D_x D_y D_z}} \left\{ \frac{1}{(t)^{3/2}} \exp \left[-\frac{1}{4} \left(\frac{(x-x_i-v_x t)^2}{D_x t} + \frac{(y-y_i-v_y t)^2}{D_y t} + \frac{(z-z_i-v_z t)^2}{D_z t} \right) - (v_d + I \cdot l)t \right] + \int_0^t \frac{\tau}{(t-\tau)^{3/2}} \right. \quad (3)$$

$$\left. \exp \left[-\frac{(x-v_x(t-\tau)-x_i)^2}{4D_x(t-\tau)} - \frac{(y-v_y(t-\tau)-y_i)^2}{4D_y(t-\tau)} - \frac{(z-v_z(t-\tau)-z_i)^2}{4D_z(t-\tau)} - (v_d + I \cdot l)(t-\tau) \right] d\tau \right\}$$

160 Where, the definitions of parameters are the same as above.

161 2.2 Numerical solution based on Taylor series upwind difference

162 In the following, the difference method is used for numerical solution of the multi-point
 163 sources continuous release diffusion problem. For convection-diffusion equation, the upwind
 164 scheme is the best effect when the boundary conditions and parameters are fixed (He W. P. et
 165 al., 2004). Considering that the main influencing factor of chemical hazard diffusion is the
 166 convection term, in order to reduce the diffusion prediction error, Taylor series upwind
 167 difference is used to difference the equation (1). Assuming that $\Delta x, \Delta y, \Delta z$ are spatial discrete
 168 step size and Δt is time discrete step size, the difference equation is expressed as follows:

$$169 \quad \frac{C_{x,y,z}^{t+1} - C_{x,y,z}^t}{\Delta t}$$

$$170 \quad = \left(D_x + \frac{v_x \cdot \Delta x}{2} \right) \cdot \frac{C_{x-1,y,z}^t + C_{x+1,y,z}^t - 2C_{x,y,z}^t}{\Delta x^2} + \left(D_y + \frac{v_y \cdot \Delta y}{2} \right) \cdot \frac{C_{x,y-1,z}^t + C_{x,y+1,z}^t - 2C_{x,y,z}^t}{\Delta y^2}$$

$$171 \quad + \left(D_z + \frac{v_z \cdot \Delta z}{2} \right) \cdot \frac{C_{x,y,z-1}^t + C_{x,y,z+1}^t - 2C_{x,y,z}^t}{\Delta z^2} - v_x \cdot \frac{C_{x+1,y,z}^t - C_{x-1,y,z}^t}{2\Delta x} - v_y \cdot \frac{C_{x,y+1,z}^t - C_{x,y-1,z}^t}{2\Delta y} - v_z$$

$$172 \quad \cdot \frac{C_{x,y,z+1}^t - C_{x,y,z-1}^t}{2\Delta z} - (v_d + I \cdot l) \cdot C_{x,y,z}^t + f(x, y, z)$$

173 Where, $f(x, y, z) = \sum_{i=1}^n Q_i \delta(|x - x_i| + |y - y_i| + |z - z_i|)$. Get the difference equation
 174 about $C_{x,y,z}^{t+1}$:

$$175 \quad C_{x,y,z}^{t+1} = \Delta t \cdot \left\{ \begin{array}{l} (D_x + \frac{v_x \cdot \Delta x}{2}) \cdot \frac{C_{x-1,y,z}^t + C_{x+1,y,z}^t - 2C_{x,y,z}^t}{\Delta x^2} + (D_y + \frac{v_y \cdot \Delta y}{2}) \\ \cdot \frac{C_{x,y-1,z}^t + C_{x,y+1,z}^t - 2C_{x,y,z}^t}{\Delta y^2} + (D_z + \frac{v_z \cdot \Delta z}{2}) \cdot \\ \frac{C_{x,y,z-1}^t + C_{x,y,z+1}^t - 2C_{x,y,z}^t}{\Delta z^2} - v_x \cdot \frac{C_{x+1,y,z}^t - C_{x-1,y,z}^t}{2\Delta x} - v_y \cdot \\ \frac{C_{x,y+1,z}^t - C_{x,y-1,z}^t}{2\Delta y} - v_z \cdot \frac{C_{x,y,z+1}^t - C_{x,y,z-1}^t}{2\Delta z} - (v_d + I \cdot l) \cdot C_{x,y,z}^t + f(x, y, z) \end{array} \right\} + C_{x,y,z}^t \quad (4)$$

176 Let $p_x = \frac{D_x \Delta t}{\Delta x^2}$, $p_y = \frac{D_y \Delta t}{\Delta y^2}$, $p_z = \frac{D_z \Delta t}{\Delta z^2}$, $q_x = \frac{v_x \Delta t}{2\Delta x}$, $q_y = \frac{v_y \Delta t}{2\Delta y}$, $q_z = \frac{v_z \Delta t}{2\Delta z}$. Then the
 177 difference equation of $C_{x,y,z}^{t+1}$ is simplified to:

$$178 \quad C_{x,y,z}^{t+1} = (p_x + 2q_x)C_{x-1,y,z}^t + (p_y + 2q_y)C_{x,y-1,z}^t + (p_z + 2q_z)C_{x,y,z-1}^t \\ 179 + [-2(p_x + q_x + p_y + q_y + p_z + q_z) - (v_d + I \cdot l)\Delta t + 1]C_{x,y,z}^t + p_x C_{x+1,y,z}^t + p_y C_{x,y+1,z}^t + p_z C_{x,y,z+1}^t + \\ 180 \quad \Delta t \cdot f(x, y, z) \quad (5)$$

181 The initial boundary value condition can be discretized as follows:

$$182 \quad \begin{cases} C_{x,y,z}^0 = f(x, y, z) \\ C_{s,k,h}^t = 0 \end{cases} \quad (6)$$

183 In combination with equations (5) and (6), the concentration value iterative matrix \mathbf{C}^{t+1}
 184 of each grid point at time $t+1$ can be obtained as follows:

$$185 \quad \mathbf{C}^{t+1} = \mathbf{A}\mathbf{C}^t + \Delta t \cdot \mathbf{f} \quad (7)$$

186 Where, $\mathbf{C}^t = (C_{x_1,y_1,z_1}^t, C_{x_2,y_2,z_2}^t, C_{x_3,y_3,z_3}^t, \dots, C_{x_M,y_M,z_M}^t)^T$, which is the concentration
 187 value matrix of each grid point at time t ; Each point is numbered in the positive direction of x ,
 188 y and z axes, and M is the total number of grids; $\mathbf{f} = (f(x_1, y_1, z_1), f(x_2, y_2, z_2),$
 189 $f(x_3, y_3, z_3), \dots, f(x_M, y_M, z_M))^T$; $\mathbf{A} = [a_{ij}]_{M \times M}$, $i, j = 1, 2, \dots, M$. The value rule of a_{ij} is
 190 as follows:

$$191 \quad a_{ij} = \begin{cases} 0, & j < i - 3 \\ p_x + 2q_x, & j = i - 3, \\ p_y + 2q_y, & j = i - 2 \\ p_z + 2q_z, & j = i - 1 \\ -2(p_x + q_x + p_y + q_y + p_z + q_z) - (v_d + I \cdot l)\Delta t + 1, & j = i \\ p_x, & j = i + 1 \\ p_y, & j = i + 2 \\ p_z, & j = i + 3 \\ 0, & j > i + 3 \end{cases}$$

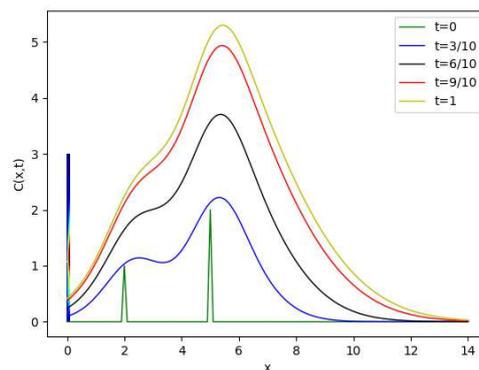
192 The truncation error of equation (7) is $O(\Delta t + \Delta x + \Delta y + \Delta z)$. According to the stability
 193 criterion of the difference scheme, the stability and convergence conditions of the difference
 194 expression are as follows:

$$195 \quad \begin{cases} p_x + q_x \leq \frac{1}{2} \\ p_y + q_y \leq \frac{1}{2} \\ p_z + q_z \leq \frac{1}{2} \end{cases} \quad (8)$$

196 Equation (8) is the normalization constraint on meshing.

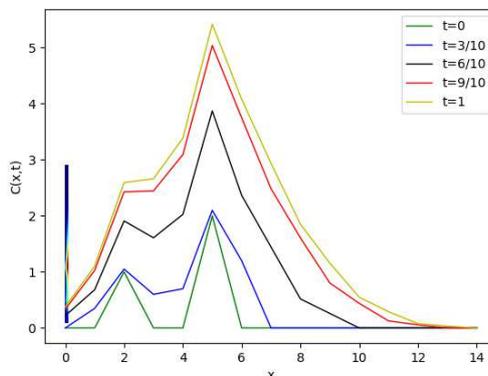
197 2.3 Numerical simulation analysis

198 Combined with the stability and convergence constraints of equation (8), taking one-
199 dimensional numerical simulation as an example, the analytical solution and numerical solution
200 are compared and analyzed. The parameter settings are as follows: spatial step size $dx = 1$,
201 spatial steps $N = 14$, time step size $dt = 1$, time steps $M = 10$, $Dx = 0.35$, $vx = 0.25$, the number
202 of hazard sources is 2, the coordinates are $[2, 5]$, and the intensity of hazard sources is $Q1 = 1$
203 and $Q2 = 2$ respectively. The calculation results are shown in Figure 1 and Figure 2.



204

205 Fig.1. Analytical simulation of one dimensional convection-diffusion equation



206

208 It can be seen from Figure 1 and Figure 2 that except for obvious differences at individual
209 points, the numerical solution and analytical solution are in good agreement as a whole. The
210 main reason for the difference is the truncation error $O(\Delta t + \Delta x)$ of the difference equation.
211 Next, considering the inverse problem on the basis of solving the forward problem, an improved
212 4DVAR algorithm for multi-point STI is constructed to estimate the source intensity and
213 position.

215 The mathematical expression of variational method refers to the functional extreme value
216 problem with constraints. Based on the theory of maximum likelihood estimation, it interpolates

217 the observed data to the preset mode lattice, and constructs a target functional by measuring the
 218 distance between the model field and the observation field (Fairbairn D. et al., 2014). Compared
 219 with other assimilation methods, 4DVAR realizes the assimilation of observed data in the whole
 220 assimilation time window, with higher prediction accuracy and wider application range. Based
 221 on the traditional 4DVAR, this paper constructs the gradient model, tangent linear operator of
 222 numerical prediction, background error covariance matrix and observation error covariance
 223 matrix suitable for multi-point sources continuous release, solves the gradient model by using
 224 the Improved Flower Pollination Algorithm (IFPA), and finally obtains the multi-point sources
 225 continuous release inversion algorithm.

226 3.1 Gradient model of multi-point STI

227 According to the multi-point sources continuous release convection-diffusion equation
 228 (equation (1)) and the 4DVAR assimilation principle, the multi-point source inversion target
 229 functional model under continuous release is constructed as follows:

$$230 J_{4D}(\mathbf{C}) = J_{4D}^b(c_0) + J_{4D}^r(\mathbf{C}) =$$

$$231 \frac{1}{2}(\sum_{i=1}^n \mathbf{c}_{i,0} - \mathbf{c}_b)^T \mathbf{B}^{-1}(\sum_{i=1}^n \mathbf{c}_{i,0} - \mathbf{c}_b) + \frac{1}{2}\sum_{t \in T} (\sum_{i=1}^n H_t \cdot \mathbf{c}_{i,t} - \mathbf{y}_t)^T \mathbf{R}^{-1}(\sum_{i=1}^n H_t \cdot \mathbf{c}_{i,t} - \mathbf{y}_t) \quad (9)$$

232 Where, \mathbf{c}_0 is the initial state vector of hazard sources; $\mathbf{c}_{i,0}$ is the concentration state
 233 vector of hazardous substance at the initial time of the i th hazard source; \mathbf{c}_b is the background
 234 field vector; H_t is the observation operator matrix at time t ; \mathbf{y}_t is the observation vector of
 235 hazardous substance concentration at time t ; $\mathbf{c}_{i,t}$ is the state vector of diffusion concentration
 236 at time t of the i th hazard source; \mathbf{B} is the background error covariance matrix; \mathbf{R} is the
 237 observation error covariance matrix; T is the assimilation time window; Other parameters are
 238 the same as above.

239 For equation (9), it is composed of background and observation, and the process of
 240 minimizing is also the superposition of these two parts, namely:

$$241 \nabla J_{4D}(c_0) = \nabla J_{4D}^b(c_0) + \nabla J_{4D}^r(c_0) \quad (10)$$

242 Calculate the partial derivative of $J_{4D}^b(c_0)$ in equation (9) directly to obtain its gradient
 243 model:

$$244 \nabla J_{4D}^b(c_0) = \mathbf{B}^{-1}(\sum_{i=1}^n \mathbf{c}_{i,0} - \mathbf{c}_b) \quad (11)$$

245 For the multi-point source continuous release diffusion model, $J_{4D}^r(\mathbf{C})$ involves many
 246 parameters and vectors, so it is difficult to obtain the partial derivative of \mathbf{c}_0 by direct
 247 derivation. Assuming that $J_{4D}^r(\mathbf{C})$ is bounded differentiable in Hilbert space, according to the
 248 first-order variational properties, we can derived:

$$249 \delta J_{4D}^r(\mathbf{C}) = \frac{d}{d\delta} J_{4D}^r(\mathbf{c}_0 + \delta \mathbf{c}_0)|_{\delta=0} = \langle \nabla J_{4D}^r(c_0), \delta \mathbf{c}_0 \rangle \quad (12)$$

250 Where, $\delta \mathbf{c}_0$ is the perturbation of \mathbf{c}_0 . Similarly, $\delta J_{4D}^r(\mathbf{C})$ is represented by $\nabla J_{4D}^r(\mathbf{c}_t)$

251 and $\delta \mathbf{c}_t$, which is shown as follows:

$$252 \quad \delta J_{4D}^r(\mathbf{C}) = \langle \nabla J_{4D}^r(\mathbf{c}_t), \delta \mathbf{c}_t \rangle = \langle \sum_{t \in T} H_t^T \mathbf{R}^{-1} (\sum_{i=1}^n H_t \cdot \mathbf{c}_{i,t} - \mathbf{y}_t), \delta \mathbf{c}_t \rangle \quad (13)$$

253 Where, \mathbf{c}_t is the diffusion concentration state vector of each hazard source at time t ; $\delta \mathbf{c}_t$
 254 is the perturbation of \mathbf{c}_t ; Other parameters are the same as above. Next, just build the
 255 relationship between $\delta \mathbf{c}_t$ and $\delta \mathbf{c}_0$ to get the $\nabla J_{4D}^r(\mathbf{c}_0)$ expression. Assuming that the adjoint
 256 model of numerical prediction from time t to time $t+1$ is $\mathbf{M}_{t+1} = (M_{1,t+1}, M_{2,t+1}, \dots, M_{m,t+1})$,
 257 m is the number of observation points, and \mathbf{a} is the set of meteorological parameters and
 258 background field parameters in the model, the relationship between \mathbf{c}_{t+1} and \mathbf{c}_t can be
 259 obtained as follows:

$$260 \quad \mathbf{c}_{t+1} = \mathbf{M}_{t+1}(\mathbf{c}_t, \mathbf{a}) \quad (14)$$

261 Assuming that $\mathbf{c}_t^* = \mathbf{c}_t - \delta \mathbf{c}_t$, which is the state vector without disturbance, then the
 262 equation (14) can be converted to:

$$263 \quad \begin{cases} \mathbf{c}_{t+1} = \mathbf{M}_{t+1}(\mathbf{c}_t^* + \delta \mathbf{c}_t, \mathbf{a}) \approx \mathbf{M}_{t+1}(\mathbf{c}_t^*, \mathbf{a}) + \mathbf{D}_{t+1} \cdot \delta \mathbf{c}_t = \mathbf{c}_{t+1}^* + \delta \mathbf{c}_{t+1} \\ \mathbf{D}_{t+1} = [D_{1,t+1}, D_{2,t+1}, \dots, D_{(x,y,z),t+1}, \dots, D_{m,t+1}]^T \end{cases} \quad (15)$$

264 Where \mathbf{D}_{t+1} is the tangent linear operator of $\mathbf{M}_{t+1}(\mathbf{c}_t, \mathbf{a})$. The relationship between
 265 $\delta \mathbf{c}_{t+1}$ and $\delta \mathbf{c}_0$ can be obtained from equation (15), which is shown as follows:

$$266 \quad \delta \mathbf{c}_{t+1} = \mathbf{D}_{t+1} \cdot \delta \mathbf{c}_t = \mathbf{D}_{t+1} \mathbf{D}_t \dots \mathbf{D}_1 \cdot \delta \mathbf{c}_0 \quad (16)$$

267 Substituting equation (16) into equation (13) and using the conjugate transpose theory, we
 268 can derived:

$$269 \quad \begin{aligned} \delta J_{4D}^r(\mathbf{C}) &= \langle \sum_{t \in T} H_t^T \mathbf{R}^{-1} \left(\sum_{i=1}^n H_t \cdot \mathbf{c}_{i,t} - \mathbf{y}_t \right), \mathbf{D}_{t+1} \mathbf{D}_t \dots \mathbf{D}_1 \cdot \delta \mathbf{c}_0 \rangle \\ &= \langle \sum_{t \in T} (\mathbf{D}_t \mathbf{D}_{t-1} \dots \mathbf{D}_1)^T H_t^T \mathbf{R}^{-1} (\sum_{i=1}^n H_t \cdot \mathbf{c}_{i,t} - \mathbf{y}_t), \delta \mathbf{c}_0 \rangle \end{aligned} \quad (17)$$

270 Combining equations (12) and (17), the partial derivative $\nabla J_{4D}^r(\mathbf{c}_0)$ of $J_{4D}^r(\mathbf{C})$ to \mathbf{c}_0 can
 272 be obtained:

$$273 \quad \nabla J_{4D}^r(\mathbf{c}_0) = \sum_{t \in T} (\mathbf{D}_t \mathbf{D}_{t-1} \dots \mathbf{D}_1)^T H_t^T \mathbf{R}^{-1} (\sum_{i=1}^n H_t \cdot \mathbf{c}_{i,t} - \mathbf{y}_t) \quad (18)$$

274 Combining equations (11) and (18), the partial derivative $\nabla J_{4D}(\mathbf{c}_0)$ of $J_{4D}(\mathbf{C})$ to \mathbf{c}_0 can
 275 be obtained:

$$276 \quad \nabla J_{4D}(\mathbf{c}_0) = \mathbf{B}^{-1} (\sum_{i=1}^n \mathbf{c}_{i,0} - \mathbf{c}_b) + \sum_{t \in T} (\mathbf{D}_t \mathbf{D}_{t-1} \dots \mathbf{D}_1)^T H_t^T \mathbf{R}^{-1} (\sum_{i=1}^n H_t \cdot \mathbf{c}_{i,t} - \mathbf{y}_t) \quad (19)$$

277 The above formula is the gradient model of multi-point STI.

278 3.2 Tangent linear operator of numerical prediction \mathbf{D}_{t+1}

279 The basis of finding tangent linear operator is to determine the numerical prognostic model
 280 of diffusion. Considering that the main factor affecting the diffusion of hazardous substances is
 281 the convection term, in order to reduce the diffusion prediction error, the numerical solution of
 282 convection-diffusion equation based on Taylor series upwind difference is used as the numerical
 283 prediction operator at the observation point (x, y, z) on time $t+1$, as shown in equation (7).
 284 According to the numerical prediction operator of each observation point at time $t+1$, the

285 numerical prognostic model \mathbf{M}_{t+1} of the whole state vector at time $t+1$ can be obtained,
 286 where $\mathbf{M}_{t+1} = (M_{1,t+1}, M_{2,t+1}, \dots, M_{m,t+1})$.

287 Assuming that the predicted concentration of hazardous substances at each observation
 288 point at time t is $\mathbf{c}_t = (c_{1,t}, c_{2,t}, \dots, c_{m,t})$, according to the definition of tangent linear operator,
 289 the numerical prediction tangent linear operator \mathbf{D}_{t+1} at time $t+1$ is obtained:

$$290 \quad \mathbf{D}_{t+1} = \begin{bmatrix} \frac{\partial M_{1,t+1}}{\partial c_{1,t}} & \frac{\partial M_{1,t+1}}{\partial c_{2,t}} & \dots & \frac{\partial M_{1,t+1}}{\partial c_{m,t}} \\ \frac{\partial M_{2,t+1}}{\partial c_{1,t}} & \frac{\partial M_{2,t+1}}{\partial c_{2,t}} & \dots & \frac{\partial M_{2,t+1}}{\partial c_{m,t}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial M_{m,t+1}}{\partial c_{1,t}} & \frac{\partial M_{m,t+1}}{\partial c_{2,t}} & \dots & \frac{\partial M_{m,t+1}}{\partial c_{m,t}} \end{bmatrix} \quad (20)$$

$$291 \quad \text{Where } \frac{\partial M_{(x,y,z),t+1}}{\partial c_t} = \begin{pmatrix} \dots, p_x + 2q_x, p_y + 2q_y, p_z + 2q_z, \\ 1 - 2(p_x + q_x + p_y + q_y + p_z + q_z) - (v_d + I \cdot l)\Delta t, \\ p_x, p_y, p_z, \dots \end{pmatrix}.$$

292 3.3 Background error covariance matrix \mathbf{B}

293 The background error covariance matrix is used to characterize the error covariance
 294 relationship between the background field and the true value. Assuming that the background
 295 field error of the chemical hazard concentration at each point in the diffusion space is unbiased,
 296 uncorrelated and isotropic, and the background error and observation error are independent of
 297 each other, the mathematical expression of the background error variance of the background
 298 field on the i th grid point is:

$$299 \quad \sigma_{bi}^2 = \overline{(c_{bi} - c_i^t)(c_{bi} - c_i^t)} \quad (21)$$

300 Where, c_i^t is the true concentration value of the i th grid point, and c_{bi} is the background
 301 field concentration value of the i th grid point; $\overline{\quad}$ is the average value (Varella H. et al., 2011).
 302 The background error covariance matrix \mathbf{B} of the background field between m observation
 303 points is:

$$304 \quad \mathbf{B} = \begin{bmatrix} \overline{(c_{b1} - c_1^t)(c_{b1} - c_1^t)} & \overline{(c_{b1} - c_1^t)(c_{b2} - c_2^t)} & \dots & \overline{(c_{b1} - c_1^t)(c_{bm} - c_m^t)} \\ \overline{(c_{b2} - c_2^t)(c_{b1} - c_1^t)} & \overline{(c_{b2} - c_2^t)(c_{b2} - c_2^t)} & \dots & \overline{(c_{b2} - c_2^t)(c_{bm} - c_m^t)} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{(c_{bm} - c_m^t)(c_{b1} - c_1^t)} & \overline{(c_{bm} - c_m^t)(c_{b2} - c_2^t)} & \dots & \overline{(c_{bm} - c_m^t)(c_{bm} - c_m^t)} \end{bmatrix}_{m \times m} \quad (22)$$

305 However, the background error covariance matrix (\mathbf{B} matrix) is a very challenging problem
 306 in both the field of weather prediction and the field of pollutant diffusion, while the core lies in
 307 the difficulty of obtaining the true concentration value of each grid point in the real environment,
 308 large dimension of state variables by meshing, and lack of sufficient data samples. Therefore,
 309 it is difficult to accurately describe \mathbf{B} matrix by equation (22). Generally, \mathbf{B} matrix is constructed
 310 by physical inference and rough statistics, where the typical construction methods include
 311 innovation vector (IV) analysis method, NMC method, potential vorticity characterization
 312 method, etc (Rutherford I, 1972; Liu YA et al., 2015; Cullen, 2003). In recent years, with the

313 development of computer technology, spherical wavelet analysis, set estimation method,
 314 covariant Lyapunov vector (CLV) and other methods have emerged (Chabot et al., 2017; Valler
 315 Veronika et al., 2019; Quinn C. et al., 2020). Compared with other methods, NMC has attracted
 316 more and more attention and application because of its better simplicity of operation,
 317 universality of scene and stability of results. Therefore, this paper selects the classical NMC as
 318 the basic method to construct \mathbf{B} matrix by adjusting the forecast period and increasing the
 319 background error correlation coefficient, which is suitable for multi-point sources continuous
 320 release inversion.

321 3.3.1 NMC basic model

322 NMC method is a construction method of background error covariance matrix proposed
 323 by Parrish and Derber in 1992. Compared with other methods, this method replaces the
 324 background error with the integral difference of prediction values at the same time and different
 325 prediction aging, which can effectively avoid the shortage of truth value and solve the
 326 calculation difficulty caused by too large dimension of state variables (Parrish D. F. et al., 1992).
 327 Assuming that \mathbf{c}_t^T is the predicted concentration state vector of observation points with time t
 328 and analytical aging T , $\mathbf{c}_t^T = (c_{1t}^T, c_{2t}^T, \dots, c_{mt}^T)$, Parrish and Derber convert the calculation of
 329 \mathbf{B} matrix into:

$$330 \quad \mathbf{B} = \frac{1}{2} \left[(c_{it}^{48} - c_{it}^{24})(c_{jt}^{48} - c_{jt}^{24}) \right]_{m \times m} \quad (23)$$

331 Where, c_i^{48} and c_i^{24} are the predicted concentration value of i th observation point at
 332 time t and analytical aging of $48h$ and $24h$ respectively; $1 \leq i \leq m, 1 \leq j \leq m$. Take $t = 48h$
 333 as an example, then

$$334 \quad \begin{cases} \mathbf{c}_{48}^{48} = M_{48 \leftarrow 0} \cdot \mathbf{y}_0, (t = 0) \\ \mathbf{c}_{48}^{24} = M_{48 \leftarrow 24} \cdot \mathbf{y}_{24}, (t = 24h) \end{cases} \quad (24)$$

335 Where, $M_{t_2 \leftarrow t_1}$ is the prognostic model of chemical hazard diffusion from time t_1 to t_2 ;
 336 \mathbf{y}_0 and \mathbf{y}_{24} are the observed concentration state vectors of the observation points at $t = 0$
 337 and $t = 24h$ respectively.

338 3.3.2 Correction strategy

339 (1) Adjust forecast period

340 In addition to extreme weather such as heavy rain and heavy snowfall, in case of major
 341 chemical explosion or leakage, the hazardous substances will quickly spread downwind in a
 342 few minutes to dozens of minutes and cause large-scale pollution, so it is inappropriate to use
 343 $48h$ and $24h$ as analytical aging and forecast time. Therefore, selecting the appropriate forecast
 344 period (analytical aging and forecast time) becomes the key to construct the \mathbf{B} matrix. If the
 345 forecast time is too long, it will affect the rescue and disposal. If it is too short, it is difficult to

346 obtain a relatively stable analytical aging. If the analytical aging is too large or too low, it will
 347 lead to a large difference between the two analysis field errors and the real background errors.

348 The main factors affecting the prediction period include: the action duration of hazardous
 349 substances, the timeliness requirements of emergency rescue or hazard disposal for diffusion
 350 prediction, the general disposal mode of hazardous substances, the data acquisition mode and
 351 the use characteristics of monitoring equipment. Comprehensively analyze the influencing
 352 factors, adjust the forecast time in equations (23) and (24) to 30min, and the analytical aging to
 353 10min and 20min. Then the predicted concentration state vector when the forecast time t is
 354 30min and the analytical aging are 10min and 20min respectively is:

$$355 \begin{cases} \mathbf{c}_{30}^{20} = \mathbf{M}_{30 \leftarrow 10} \cdot \mathbf{y}_{10}, (t = 10min) \\ \mathbf{c}_{30}^{10} = \mathbf{M}_{30 \leftarrow 20} \cdot \mathbf{y}_{20}, (t = 20min) \end{cases} \quad (25)$$

356 Where the parameters definition are the same as above.

357 (2) Correction based on background error correlation coefficient

358 In order to ensure the positive definiteness and smoothness of the \mathbf{B} matrix, the Gaussian
 359 formula close to the actual situation of chemical hazard diffusion is used to construct the
 360 background error correlation coefficient, and the background error covariance is formed by
 361 multiplying the background error and the correlation coefficient. Assuming any two lattice
 362 points x_1 and x_2 in the space, the background error correlation coefficient between two
 363 points $b(x_1, x_2)$ can be expressed as a function of its distance:

$$364 \quad b(x_1, x_2) = \text{dist}(x_1 - x_2) e^{-\frac{\text{dist}(x_1 - x_2)^2}{2}} \quad (26)$$

365 Where $\text{dist}(x_1 - x_2)$ is the distance between x_1 and x_2 (Ma Jianwen et al., 2013).

366 Thus, the expression of \mathbf{B} matrix can be obtained:

$$367 \quad \mathbf{B} = \frac{1}{2} [b(i, j) \cdot \overline{(c_{i30}^{20} - c_{i30}^{10})(c_{j30}^{20} - c_{j30}^{10})}]_{m \times m} \quad (27)$$

368 Where the parameters definition are the same as above.

369 3.4 Observation error covariance matrix \mathbf{R}

370 In 4DVAR, the observation error covariance matrix is used to characterize the error
 371 covariance relationship between the analysis field and the true value (Liu Y. A. et al., 2019).
 372 The mathematical expression of observation error variance for analysis field on i th grid point
 373 σ_{ri}^2 is:

$$374 \quad \sigma_{ri}^2 = \overline{(y_i - c_i^t)(y_i - c_i^t)} \quad (28)$$

375 Where, c_i^t is the concentration true value of i th grid point; y_i is the concentration
 376 observation value of i th grid point. Then the observation error covariance matrix (\mathbf{R} matrix) of
 377 analysis field between m observation points can be expressed as follows:

$$\mathbf{R} = \begin{bmatrix} \frac{(y_1 - c_1^t)(y_1 - c_1^t)}{(y_1 - c_1^t)(y_1 - c_1^t)} & \frac{(y_1 - c_1^t)(y_2 - c_2^t)}{(y_1 - c_1^t)(y_2 - c_2^t)} & \cdots & \frac{(y_1 - c_1^t)(y_m - c_m^t)}{(y_1 - c_1^t)(y_m - c_m^t)} \\ \frac{(y_2 - c_2^t)(y_1 - c_1^t)}{(y_2 - c_2^t)(y_1 - c_1^t)} & \frac{(y_2 - c_2^t)(y_2 - c_2^t)}{(y_2 - c_2^t)(y_2 - c_2^t)} & \cdots & \frac{(y_2 - c_2^t)(y_m - c_m^t)}{(y_2 - c_2^t)(y_m - c_m^t)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{(y_m - c_m^t)(y_1 - c_1^t)}{(y_m - c_m^t)(y_1 - c_1^t)} & \frac{(y_m - c_m^t)(y_2 - c_2^t)}{(y_m - c_m^t)(y_2 - c_2^t)} & \cdots & \frac{(y_m - c_m^t)(y_m - c_m^t)}{(y_m - c_m^t)(y_m - c_m^t)} \end{bmatrix}_{m \times m} \quad (29)$$

379 It can be seen that \mathbf{R} matrix contains statistical information about observation errors. For
380 chemical hazard alarm and monitoring equipment, the observation errors of each observation
381 point are distributed independently of each other. Assuming that the observation error is
382 unbiased and the influence of the observation operator and representative error on the
383 observation results is not considered, the \mathbf{R} matrix can be simplified into a diagonal matrix
384 about the instrument error variance, as shown in equation (30).

$$\mathbf{R} = \begin{bmatrix} \sigma_{r1}^2 & & & \\ & \sigma_{r2}^2 & & \\ & & \ddots & \\ & & & \sigma_{rm}^2 \end{bmatrix}_{m \times m} \quad (30)$$

386 3.5 Intelligent solution of gradient model based on IFPA

387 Aiming at the shortcomings of complex calculation for inverse matrix and easy to fall into
388 local convergence or non-convergence of Quasi Newton Algorithm (QNA) in the traditional
389 4DVAR, the Improved Flower Pollination Algorithm (IFPA) is used as the optimization
390 algorithm for the iterative solution of gradient model. IFPA is an efficient intelligent
391 optimization algorithm proposed by authors (Zhu X. Z., et al., 2021), where the Good-point Set
392 population, *Deb* feasibility comparison method and ε constraint method is integrated into
393 Flower Pollination Algorithm (FPA) (Yang X. S., 2012), which can comprehensively improves
394 the global optimization ability of FPA and the flexibility to deal with different constraint
395 problems. Compared with Particle Swarm Optimization Algorithm (PSO), Artificial Bee
396 Colony Algorithm (ABCA), Coevolutionary Differential Evolution Algorithm (CDE) and other
397 methods, IFPA is not limited by the unconnected feasible region, does not required to repeatedly
398 derive objective function gradient in process of solving, and significantly improves the
399 calculation rate. The basic steps of gradient model solution using IFPA are as follows:

400 Step 1: Define flower individuals (source intensity, quantity or location), set flower
401 population size N , pollination mode conversion probability P , maximum iteration times T and
402 other parameters to form the initial population.

403 Step 2: Let $t=1$, $k=1$, traverse each population, calculate the $\nabla J_{4D}(c_0)$ value of each
404 flower in each population, analyze the stability and convergence of each flower, eliminate
405 unstable flowers and generate a new flower population.

406 Step 3: For the new flower population, calculate the $\nabla J_{4D}(c_0)$ value of each flower,
407 search for the best flower X_b^t .

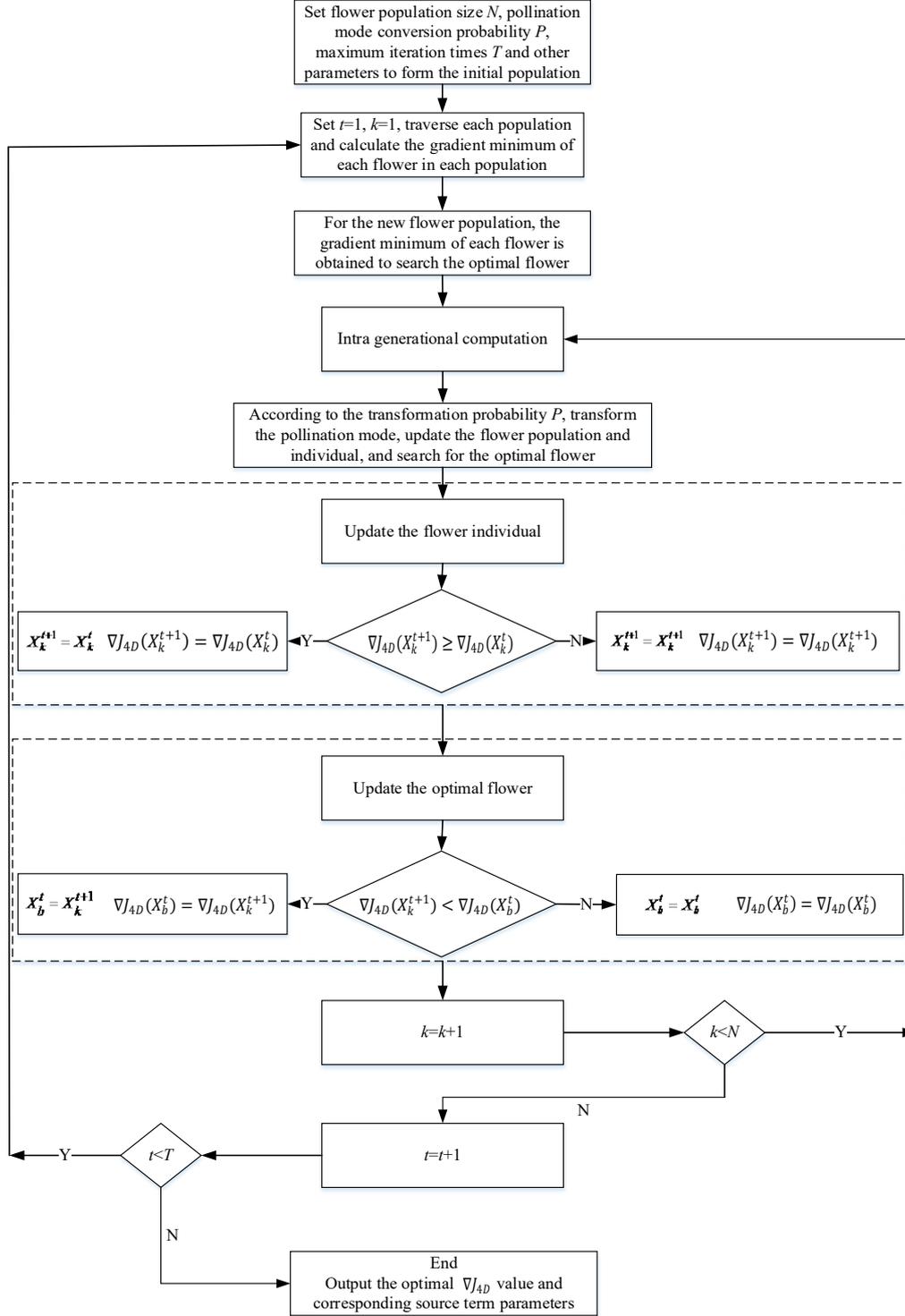


Fig.3 Basic process of gradient model optimization based on IFPA

Step 4: Perform intra generational computation.

Step 5: According to the transformation probability P , transform the pollination mode, update the flower population and individual, and search for the optimal flower.

Step 6: Update the flower individual. If $\nabla J_{4D}(X_k^{t+1}) \geq \nabla J_{4D}(X_k^t)$, then $X_k^{t+1} = X_k^t$, $\nabla J_{4D}(X_k^{t+1}) = \nabla J_{4D}(X_k^t)$; otherwise, $X_k^{t+1} = X_k^{t+1}$, $\nabla J_{4D}(X_k^{t+1}) = \nabla J_{4D}(X_k^{t+1})$.

415 Step 7: Update the optimal flower. If $\nabla J_{4D}(X_k^{t+1}) < \nabla J_{4D}(X_b^t)$, then $X_b^t = X_k^{t+1}$,
 416 $\nabla J_{4D}(X_b^t) = \nabla J_{4D}(X_k^{t+1})$; otherwise, $X_b^t = X_b^t$, $\nabla J_{4D}(X_b^t) = \nabla J_{4D}(X_b^t)$.

417 Step 8: $k=k+1$. if $k < N$, then go to step 4, otherwise go to step 9.

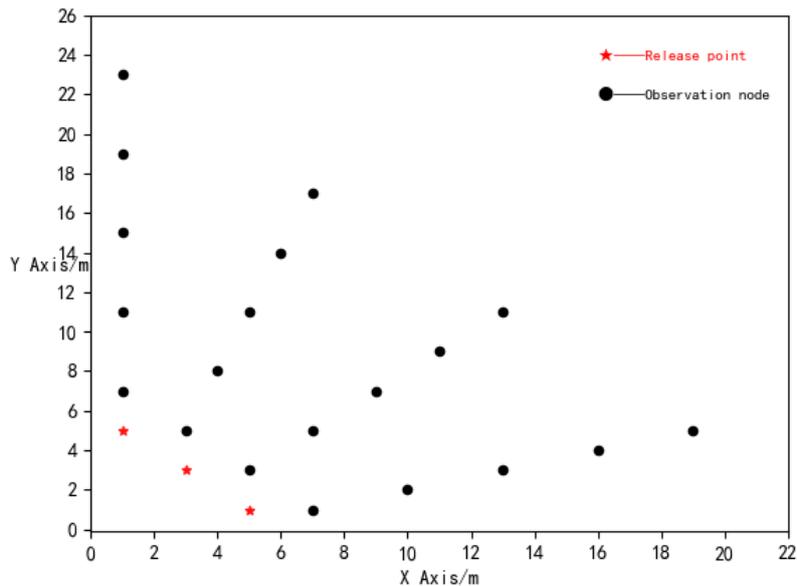
418 Step 9: $t=t+1$. If $t < T$, then go to step 2, otherwise go to step 10.

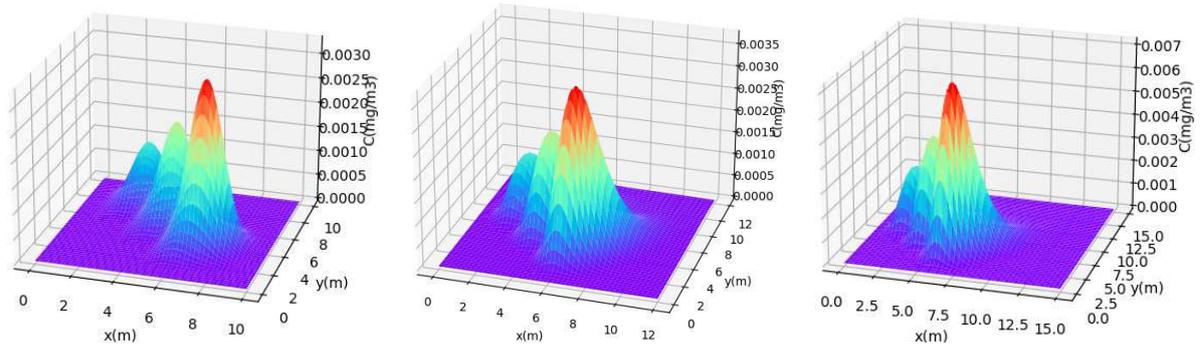
419 Step 10: At the end of population traversal and iterative calculation, output the optimal
 420 ∇J_{4D} value and corresponding source term parameters.

421 The basic flow of gradient model optimization based on IFPA is shown in Figure 3.

422 4 Simulation Case Analysis

423 In the numerical inversion in this section, set $D=(0.30,0.35,0.10)$, $v=(2,3,1)$, $v_d=0.002$,
 424 $I=3.6$, $l=0.001$; the spatial step size dx , dy and dz are all 1, the number of spatial steps $N=30$,
 425 the time step size $dt=0.1$, and the number of time steps $M=50$; The number of hazard sources is
 426 3, the coordinates are $X=[1,3,5]$, $Y=[5,3,1]$, $Z=3$, with the unit of m ; the intensities of hazard
 427 sources are $Q1=3$, $Q2=5$ and $Q3=8$ respectively, with the unit of mg/m^3 . In this paper, the
 428 analytical solution data of 20 nodes in height $2m$ at $t=1$, $t=3$ and $t=5$, which is plus 10%
 429 disturbance, are taken as the observed data. Through the improved 4DVAR algorithm, three
 430 kinds of situations such as source location known and inversion intensity, source intensity
 431 known and inversion location, source location and intensity all unknown are inverted and
 432 compared. In order to more intuitively analyze the feasibility of the algorithm in this paper, it
 433 is assumed that the background error and observation error of each observation point are in
 434 standard normal distribution in numerical simulation. The layout of observation nodes and
 435 diffusion results at each time are shown in Figure 4 and Figure 5.





438

440

Fig.5 Simulation chart of hazard diffusion at each time

441

442

443

444

445

446

447

448

449

450

451

It can be seen from Figure 4 that the observation nodes are regularly distributed at equal angles and equal intervals along the downwind direction of the release point, meeting the general layout requirements of dense in front and sparse in back and uniform arrangement. It can be seen from Figure 5 that due to the influence of wind speed and diffusion parameters, the effective observed data at $t=1$, $t=3$ and $t=5$ are controlled within the space from the origin to (10,10), (12,12) and (15,15) respectively, and the layout scheme of observation nodes can cover the diffusion area of hazardous substances at each time, which shows that it is correct and applicable to use the observed data of each node under the layout scheme as the input of the improved 4DVAR algorithm. The observed data of observation nodes at each time are shown in Table 1.

Tab.1 Observed data of observation nodes at each time

节点	1	2	3	4	5	6	7	8	9	10
t=1	4.386	8.091	1.301	1.086	5.455	5.473	1.533	8.015	1.265	4.925
	E-05	E-05	E-04	E-05	E-08	E-04	E-03	E-08	E-20	E-08
t=3	1.514	2.733	4.526	2.679	8.259	6.596	1.858	3.677	7.891	1.264
	E-04	E-04	E-04	E-05	E-07	E-04	E-03	E-07	E-10	E-04
t=5	2.812	5.075	8.403	5.006	1.949	1.390	3.919	8.311	2.375	3.481
	E-04	E-04	E-04	E-05	E-06	E-03	E-03	E-07	E-09	E-04
节点	11	12	13	14	15	16	17	18	19	20
t=1	1.349	7.636	0	2.079	4.479	0	0	0	0	0
	E-07	E-17	0	E-18	E-17	0	0	0	0	0
t=3	3.300	8.110	7.980	1.215	3.840	3.730	9.630	8.830	5.460	7.500
	E-04	E-10	E-13	E-05	E-05	E-12	E-19	E-07	E-06	E-16
t=5	9.033	2.636	1.999	3.115	1.064	5.803	1.549	1.824	9.431	1.003
	E-04	E-09	E-12	E-05	E-04	E-12	E-15	E-06	E-06	E-14

Coordinate of observation nodes: X=[1,3,5,7,1,4,7,10,1,5,9,13,1,6,11,16,1,7,13,19];

Y= [7,5,3,1,11,8,5,2,15,11,7,3,19,14,9,4,23,17,11,5]

452

4.1 Location known, inversion intensity

481 of $t=5$, the final phase transition position of observed data is close to the position of assimilation
 482 node, and there is an obvious distance between the final phase transition position of predicted
 483 data and observed data, indicating that there is still a certain gap between the inversion results
 484 and the actual situation due to the comprehensive effects of model disturbance error,
 485 background error and observation error, but because the disturbance effect of the third group
 486 observed data is small, the inversion accuracy will be significantly improved than before.

487 In order to further assess the accuracy of inversion intensity, error and relative error are
 488 selected as comprehensive evaluation indexes, and the advantages and disadvantages of
 489 Kalman Filter (KF), Ensemble Kalman Filter (EnKF), 4DVAR and improved 4DVAR are
 490 analyzed. Among them, KF selects data of time $t = 3$ as assimilation data. The calculation
 491 expressions of sources intensity inversion error and relative error are as follows:

$$492 \begin{cases} \delta_{intension} = \sqrt{\sum_{i=1}^n (Q_{i1} - Q_{i0})^2 / n} \\ \epsilon_{intension} = \sqrt{\sum_{i=1}^n \left(\frac{Q_{i1} - Q_{i0}}{Q_{i0}} \right)^2 / n} \end{cases} \quad (31)$$

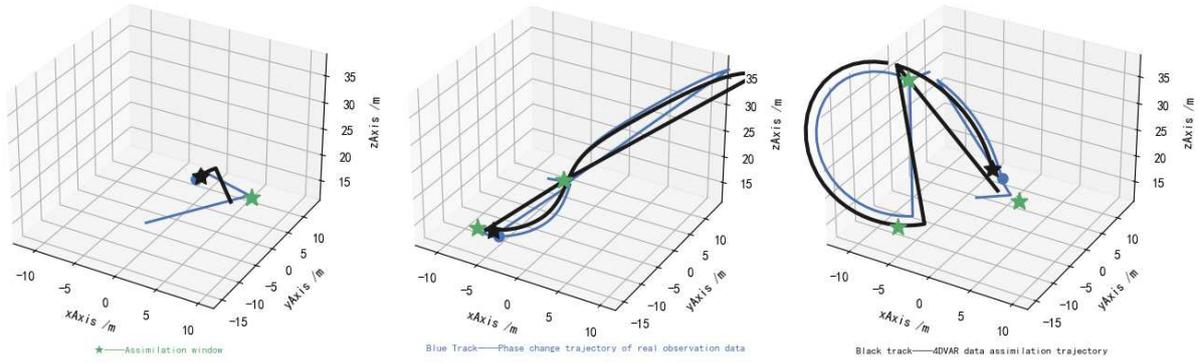
493 Tab.2 Advantages and disadvantages comparison of sources intensity inversion

STI algorithm	Intensity of No. 1	Intensity of No. 2	Intensity of No. 3	$\delta_{intension}$	$\epsilon_{intension}$
KF	4.5	3.7	6.4	0.8498	19.93%
EnKF	3.6	4.4	7.2	0.3887	17.38%
4DVAR	3.5	4.5	7.4	0.3091	14.08%
Improved 4DVAR	3.5	4.7	7.5	0.2560	6.26%

494 It can be seen from Table 2 that when the location of sources is known, EnKF, 4DVAR
 495 and improved 4DVAR can be assimilated through three groups of observed data, and KF can
 496 only be assimilated through one group of observed data. Therefore, the sources intensity
 497 estimation results of EnKF, 4DVAR and improved 4DVAR are significantly better than KF.
 498 Because 4DVAR and improved 4DVAR incorporate the numerical prediction model and tangent
 499 linear operator of convection-diffusion equation, compared with EnKF, which only assimilates
 500 and gathers the data of each observation window independently, the sources intensity estimation
 501 error can be reduced from about 0.40 before improvement to about 0.25, and the relative error
 502 can be reduced from about 17% before improvement to about 6%. At the same time, because
 503 IFPA is more global and stable than QNA, the sources intensity estimation error of improved
 504 4DVAR is smaller and closer to the real source intensity than 4DVAR.

505 4.2 Intensity known, inversion location

506 Based on the analysis in Section 4.1, set the intensity data of three hazard sources as known,
 507 and use the improved 4DVAR algorithm to invert the sources location. The inversion
 508 assimilation curve of sources location is shown in Figure 7.



509

510

a) $t=1$

b) $t=3$

c) $t=5$

511

Fig.7 Inversion assimilation curve of sources location

512

513

514

515

516

517

518

519

520

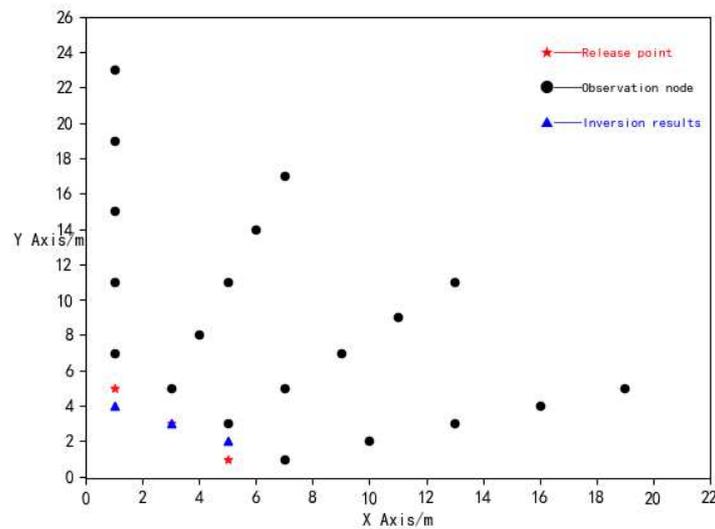
521

522

523

524

It can be seen from Figure 7 that when the sources location is unknown, the phase change trajectories of predicted data under three assimilation windows are basically consistent with the phase change trajectories of observed data, indicating that in this case, the improved 4DVAR algorithm can also better complete the data assimilation and inversion calculation. The final phase transition positions of predicted data and observed data under the three assimilation windows are basically coincide or close, indicating that for the unknown variables of 9 dimensions, IFPA can better complete the optimal value search and obtain a smaller $J_{AD}(C)$ value, which verifies the significant optimization potential of IFPA for multivariable global optimization. In addition, under the three time windows, the floating conditions of the final phase transition position of observed data and the position of assimilation node are basically consistent with Figure 6, which is caused by having the same model disturbance error, background error and observation error. The estimation result of sources location is shown in Figure 8.



525

526

Fig.8 Schematic diagram of sources location estimation

527

The location information of each source is estimated and analyzed by the four algorithms

528 in turn, where the calculation results are shown in Table 3. The calculation expressions of
 529 sources location inversion error and relative error are as follows:

$$530 \quad \begin{cases} \delta_{location} = \sum_{i=1}^n \sqrt{(x_{i1} - x_{i0})^2 + (y_{i1} - y_{i0})^2 + (z_{i1} - z_{i0})^2} / n \\ \epsilon_{location} = \sum_{i=1}^n \sqrt{\left(\frac{x_{i1} - x_{i0}}{x_{i0}}\right)^2 + \left(\frac{y_{i1} - y_{i0}}{y_{i0}}\right)^2 + \left(\frac{z_{i1} - z_{i0}}{z_{i0}}\right)^2} / n \end{cases} \quad (32)$$

531 Tab.3 Advantages and disadvantages comparison of sources location inversion

STI algorithm	Location of No. 1	Location of No. 2	Location of No. 3	$\delta_{location}$	$\epsilon_{location}$
KF	(0,5,3)	(3,3,2)	(2,4,3)	2.0809	168.65%
EnKF	(0,5,3)	(2,3,3)	(4,2,3)	1.1381	103.93%
4DVAR	(1,4,3)	(3,3,3)	(5,2,4)	0.8047	41.80%
Improved 4DVAR	(1,4,3)	(3,3,3)	(5,2,3)	0.6667	40.00%

532 It can be seen from Table 3 that on the basis of known source intensity, the error of KF
 533 calculation result is obviously large due to too many variables and nonlinear disturbance, and
 534 the error is mainly distributed in the position estimation of source No. 3. Compared with EnKF,
 535 the estimation accuracy of source location by 4DVAR and improved 4DVAR is significantly
 536 improved, the error can be reduced from 1.13 to 0.80 and 0.67, and the relative error can be
 537 reduced from 103.93% to about 40%. Compared with 4DVAR, the accuracy of sources location
 538 estimation by improved 4DVAR is not significantly improved, and the relative error difference
 539 between the two is small.

540 4.3 Location and intensity all unknown

541 Based on the above analysis, the number, location and intensity of hazard sources are set
 542 to be unknown, and the sources item information is estimated and analyzed through various
 543 algorithms, as shown in Table 4.

544 Tab.4 Advantages and disadvantages comparison of STI

STI algorithm	No. of sources	Sources location	Sources intensity	$\delta_{intension}$	$\epsilon_{intension}$	$\delta_{location}$	$\epsilon_{location}$
KF	1	(3,5,2)	14	-	-	-	-
		(1,5,2)	4				
EnKF	4	(2,3,3)	4	-	-	-	-
		(3,4,1)	4				
		(4,2,2)	4				
		(1,5,4)	5				
4DVAR	3	(3,4,3)	6	0.8165	23.57%	1.1381	57.36%
		(5,2,4)	9				
Improved 4DVAR	3	(1,5,4)	4				
		(3,4,3)	5	0.4714	11.87%	1.0000	55.56%
		(5,2,3)	7				

545 It can be seen from Table 4 that when the number, location and intensity of sources are
 546 unknown, KF and EnKF have errors in sources number inversion, indicating that they have

547 obvious disadvantages in multi-dimensional unknown number inversion and are difficult to
548 meet the traceability requirements. Both 4DVAR and improved 4DVAR can retrieve the number,
549 location and intensity of sources, which shows that the introduced numerical prediction tangent
550 linear operator can continuously assimilate the data at three observation times, and the inversion
551 effect is greatly improved compared with KF and EnKF. At the same time, compared with
552 4DVAR, both sources intensity inversion and sources location inversion are significantly
553 improved, and controlled within an acceptable range, indicating that it has a good assimilation
554 effect in multi-point sources continuous release inversion.

555 5 Verification of SF₆ tracer test

556 On the basis of numerical simulation analysis, the algorithm is further verified by tracer
557 test, and its adaptability and credibility in the field environment are analyzed.

558 5.1 Basic information of SF₆ tracer test

559 From 9 a.m. to 11 a.m. on May 22, 2016, the project team conducted two atmospheric
560 diffusion tracer tests in Xiangfang Township, Haixing County, Cangzhou City, where the tracer
561 was SF₆, the first test was used to construct the background error covariance matrix, and the
562 second test was used to verify the algorithm in this paper. The second test is as follows: the
563 number of release sources is two, the locations are the iron tower and next to YangCheng dam,
564 the heights are 100m and 10m respectively, the release amounts are 50kg and 27kg respectively,
565 and the release time is 40min. During the test, the dominant wind direction is northeast wind,
566 and the atmospheric stability is DDC.

567 The space of the experimental site was about 10km × 7km, a total of 4 fixed meteorological
568 stations, 20 portable meteorological observation points and 50 fan-shaped concentration
569 sampling points were set up to observe the data such as wind speed, temperature, humidity, air
570 pressure and tracer concentration in real time, as shown in Figure 9. The sampling time interval
571 was 10min and SF₆ samples were taken three times for each test, which were collected and
572 analyzed by *EM-1500 Portable Gas Sampler*, *LB-201-4L Aluminum Foil Sampling Bag* and
573 *Shimadzu GC-2010 Gas Chromatograph*.



Fig.9 Layout of release points and monitoring nodes

574

575

576

577

578

579

580

581

582

583

584

585

586

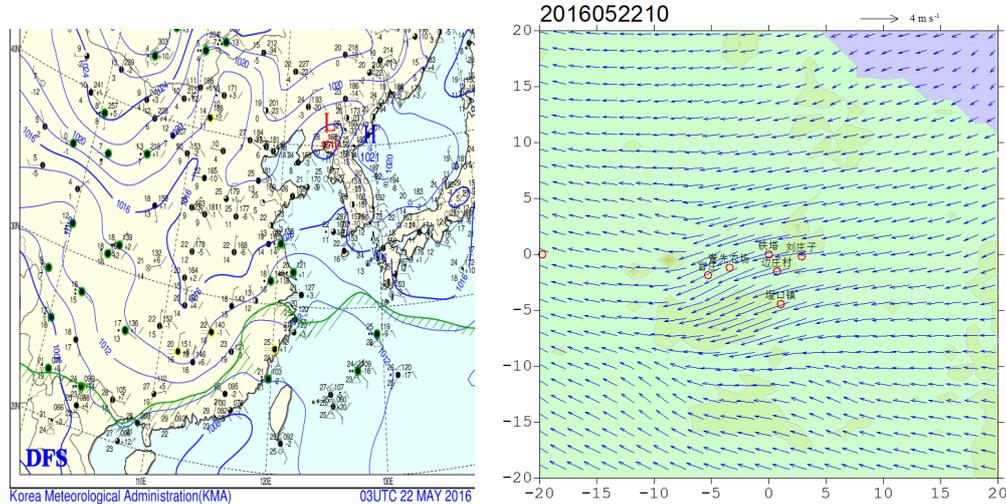
587

According to the data collected by fixed meteorological stations and portable meteorological observation points, the initial boundary conditions and wind field information of diffusion space are constructed based on WRF wind field diagnosis model. To make the calculations and the analyses convenient, East Longitude 117.636781-117.755706 and North Latitude 38.091503-38.154587 are taken as the four boundary points of coordinate axis, (117.636781, 38.091503) is taken as the origin of coordinate axis, due East direction is set as the positive direction of X axis, due North direction is set as the positive direction of Y axis, ground vertical upward direction is set as the positive direction of Z axis, and the longitude and latitude information of each point is converted to three-dimensional spatial coordinates, where the initial boundary conditions are set as shown in Table 5, the ground wind field information during the experiment is shown in Figure 10.

Tab.5 Initial boundary condition setting

Parameter	Direction / Position	Properties	Connotation	Parameter value
Wind speed	northeasterly wind	air inlet	simulate natural wind	wind direction 204° average wind speed 5.5m/s
External air pressure	X negative axis direction	air outlet	analog air pressure outlet	1019.2hpa
Release source coordinates	iron tower dam	release source location	simulate release point	(8406, 6322, 100) (8061, 6874, 10)
Release source information	Z positive axis direction	SF ₆ inlet	simulate SF ₆ emissions	20.83g/s, 40min 8.98 g/s, 40min
Turbulence information	X positive axis direction	turbulent inlet and outlet	turbulence intensity	4.02%, 4.15%

Dynamic viscosity	diffusion space	liquid viscosity	liquid viscosity coefficient	0.0000142Pa·s
Entrance area	X positive axis direction	air inlet	grid resolution	50m×50m
Setting of underlying surface and sidewall	all around	sidewall type	simulate underlying surface and sidewall	10km×10km×0.2km



(a) Ground weather situation

(b) Change of ground wind field

Fig.10 Wind field construction of diffusion space

5.2 Experimental simulation and analysis

5.2.1 Experimental parameter setting

Use the improved 4DVAR algorithm to inverse the whole diffusion process. The experimental parameters are set as follows: the number, location and intensity of sources are set unknown; the spatial steps dx , dy and dz are all 10, and the spatial steps are 1000, 1000 and 20 respectively; time steps $dt = 10$, number of time steps $M = 180$; σ_x , σ_y and σ_z is obtained from P - G diffusion curve; dry and wet sedimentation is not considered; the observation operator is the identity matrix; flower population size $N = 200$, pollination mode conversion probability $P = 0.8$, maximum iteration times $T = 500$. Since the sampling equipment and analysis equipment at each observation point are the same, the comprehensive error variance of the sampling equipment and analysis equipment is taken as the diagonal element of the observation error covariance matrix.

5.2.2 Calculation of STI

Based on the setting of experimental parameters, according to the improved 4DVAR solution and inversion process, the tracer diffusion test is calculated and simulated to solve the number, location and intensity of sources under $\nabla J_{4D}(c_0)$ minimum. Since only 42 nodes monitored SF_6 concentration data during the test, the observed data of these 42 nodes were used for data assimilation. The optimization assimilation curve of STI is shown in Figure 11, and the results are shown in Table 6.

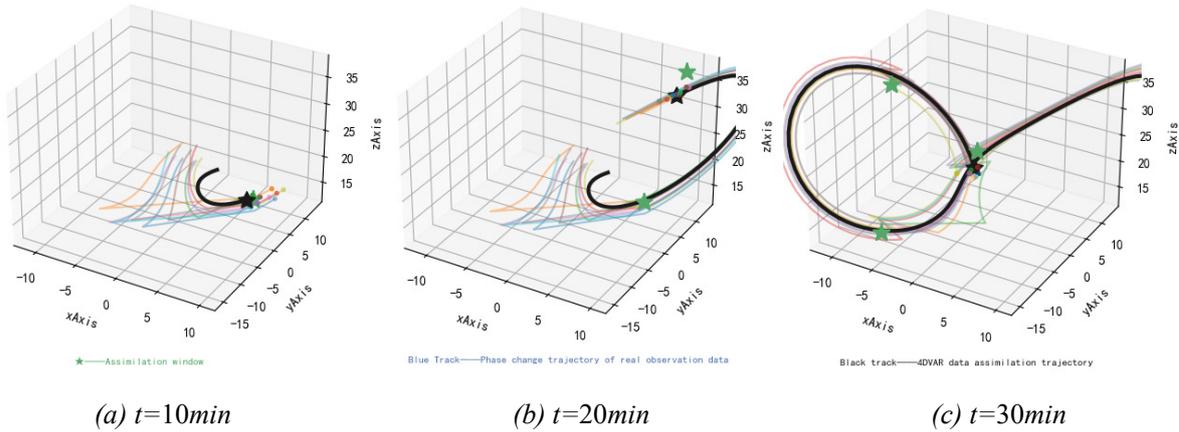


Fig.11 STI optimization assimilation curve

610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626

It can be seen from Figure 11 (a) that the first group of observation data has obvious discreteness and randomness. There is an obvious difference between the phase transition trajectory of prediction data and observation data, but the final phase transition positions of predicted data and observed data basically coincides, which verifies the stability and convergence of the improved 4DVAR algorithm. It can be seen from Figure 11 (b) that the phase change trajectory of predicted data in second time window is basically consistent with that of observed data, but there is a certain distance between the final phase transition positions of predicted data and observed data, which shows that the improved 4DVAR algorithm can quickly adjust the update mode and mutation operation. As can be seen from Figure 11 (c), under the third time window, the phase transition trajectory of predicted data is basically consistent with that of observed data, and the final phase transition positions of predicted data and observed data is basically close, indicating that after three times of data assimilation, $\nabla J_{4D}(c_0)$ has basically tended to the optimal solution.

Tab.6 STI results based on improved 4DVAR

STI algorithm	No. of sources	Sources location	Sources intensity	$\delta_{intension}$	$\epsilon_{intension}$	$\delta_{location}$	$\epsilon_{location}$
Improved 4DVAR	2	(8697, 6478, 60) (8160, 6690, 40)	20.54g/s 15.62g/s	3.3232	36.98%	271.8383	170.12%

627
628
629
630
631
632
633
634

It can be seen from Table 6 that the relative error of sources intensity is 36.98%, and the relative error of sources location is 170.12%, which basically meets the actual demand for sources seeking. In practical work, on this basis, active accurate traceability can be carried out through on-site reconnaissance or UAV search.

In order to better analyze the assimilation performance of improved 4DVAR algorithm, on the basis of STI, the predicted data by analytical solution, predicted data by improved 4DVAR assimilation and real observed data are compared, as shown in Table 7 and Figure 12. The calculation expression of concentration difference is:

635

$$\begin{cases} d_1 = |c_1 - c_3|/c_3 \\ d_2 = |c_2 - c_3|/c_3 \\ d_3 = d_2 - d_1 \end{cases} \quad (33)$$

636

Where, c_1 is the predicted concentration value by analytical solution; c_2 is the predicted concentration value by improved 4DVAR assimilation; c_3 is the real observed concentration value; d_1 is the difference between c_1 and c_3 ; d_2 is the difference between c_2 and c_3 ; d_3 is the increased degree of concentration prediction accuracy of 4DVAR assimilation relative to analytical solution.

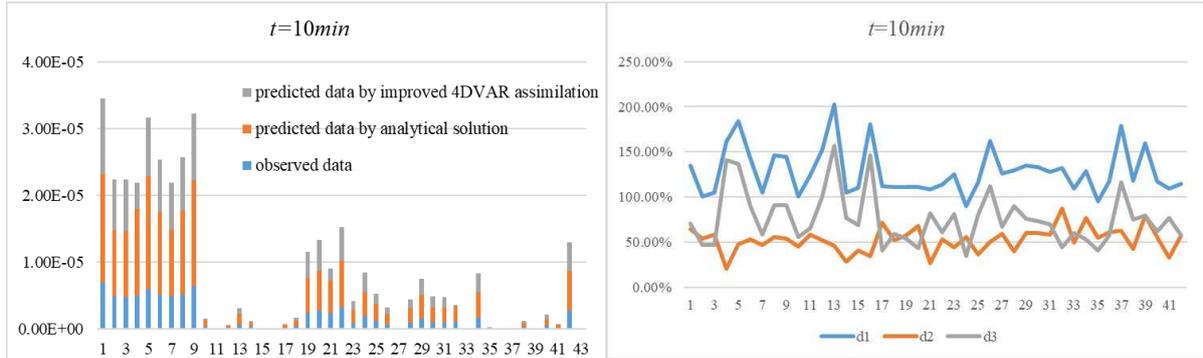
637

638

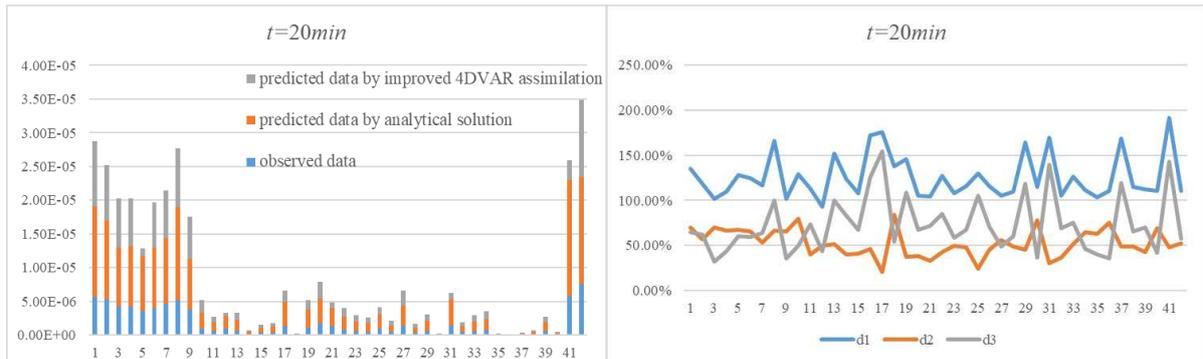
639

640

641



642



643

644

645

646

(a) Concentration data analysis

(b) prediction error analysis

Fig.12 Comparative analysis of predicted data by analytical solution, predicted data by improved 4DVAR assimilation and real observed data

Tab. 7 Results comparison of predicted data by analytical solution, predicted data by improved 4DVAR assimilation and real observed data

Monitoring		10min					Monitoring		10min				
nodes	$c_1/(\text{kg}/\text{m}^3)$	$c_2/(\text{kg}/\text{m}^3)$	$c_3/(\text{kg}/\text{m}^3)$	$d_1/\%$	$d_2/\%$	$d_3/\%$	nodes	$c_1/(\text{kg}/\text{m}^3)$	$c_2/(\text{kg}/\text{m}^3)$	$c_3/(\text{kg}/\text{m}^3)$	$d_1/\%$	$d_2/\%$	$d_3/\%$
1	1.62E-05	1.14E-05	6.92E-06	134.82	64.45	70.36	22	6.98E-06	4.99E-06	3.27E-06	113.45	52.67	60.79
2	9.88E-06	7.59E-06	4.93E-06	100.47	54.06	46.41	23	1.99E-06	1.27E-06	8.85E-07	124.72	43.98	80.74
3	9.88E-06	7.64E-06	4.83E-06	104.59	58.29	46.30	24	3.63E-06	2.97E-06	1.92E-06	89.56	55.10	34.46
4	1.30E-05	3.95E-06	4.97E-06	161.12	20.55	140.57	25	2.50E-06	1.57E-06	1.16E-06	115.97	35.83	80.14
5	1.69E-05	8.79E-06	5.95E-06	183.97	47.61	136.36	26	1.63E-06	9.38E-07	6.23E-07	162.17	50.45	111.73
6	1.24E-05	7.81E-06	5.12E-06	142.32	52.44	89.87	27	1.76E-08	1.24E-08	7.79E-09	125.67	58.77	66.90
7	9.97E-06	7.12E-06	4.87E-06	104.83	46.33	58.49	28	2.19E-06	1.33E-06	9.56E-07	129.28	39.65	89.63
8	1.26E-05	7.97E-06	5.13E-06	146.10	55.35	90.76	29	3.59E-06	2.44E-06	1.53E-06	134.83	59.49	75.34
9	1.58E-05	9.95E-06	6.46E-06	144.36	54.00	90.36	30	2.30E-06	1.58E-06	9.86E-07	133.05	60.15	72.90
10	8.68E-07	2.39E-07	4.33E-07	100.51	44.78	55.73	31	2.27E-06	1.57E-06	9.94E-07	128.15	58.36	69.79
11	1.87E-08	1.33E-08	8.37E-09	123.89	58.37	65.53	32	2.45E-06	1.31E-07	1.06E-06	132.02	87.64	44.39
12	2.94E-07	1.77E-07	1.16E-07	152.11	51.99	100.12	33	4.80E-08	3.42E-08	2.29E-08	109.61	49.50	60.11
13	1.72E-06	8.29E-07	5.69E-07	202.63	45.81	156.82	34	3.80E-06	2.93E-06	1.66E-06	128.78	76.33	52.46
14	6.85E-07	2.39E-07	3.34E-07	104.98	28.44	76.54	35	9.09E-08	7.21E-08	4.67E-08	94.72	54.28	40.44
15	5.10E-08	3.42E-08	2.43E-08	110.00	40.91	69.09	36	2.09E-08	1.54E-08	9.63E-09	117.30	60.38	56.92
16	1.83E-08	8.78E-09	6.52E-09	180.74	34.66	146.07	37	2.26E-08	1.32E-08	8.11E-09	178.94	62.33	116.61
17	2.99E-07	2.41E-07	1.41E-07	112.03	71.17	40.86	38	5.68E-07	3.71E-07	2.61E-07	117.63	42.33	75.29
18	7.82E-07	5.63E-07	3.71E-07	110.79	51.76	59.03	39	2.62E-08	1.82E-08	1.01E-08	159.41	79.70	79.70
19	5.19E-06	3.87E-06	2.46E-06	111.23	57.33	53.90	40	9.66E-07	6.92E-07	4.45E-07	116.99	55.51	61.48
20	5.90E-06	4.70E-06	2.80E-06	110.71	67.86	42.86	41	4.10E-07	1.32E-07	1.96E-07	109.35	32.79	76.56
21	4.91E-06	1.74E-06	2.36E-06	108.17	26.37	81.80	42	5.93E-06	4.34E-06	2.76E-06	114.55	57.24	57.31

Monitoring		20min					Monitoring		20min				
nodes	c ₁ /(kg/m ³)	c ₂ /(kg/m ³)	c ₃ /(kg/m ³)	d ₁ /%	d ₂ /%	d ₃ /%	nodes	c ₁ /(kg/m ³)	c ₂ /(kg/m ³)	c ₃ /(kg/m ³)	d ₁ /%	d ₂ /%	d ₃ /%
1	1.34E-05	9.69E-06	5.69E-06	135.17	70.33	64.83	22	1.92E-06	1.21E-06	8.46E-07	127.46	42.88	84.58
2	1.16E-05	8.32E-06	5.32E-06	118.80	56.39	62.40	23	1.38E-06	9.91E-07	6.61E-07	107.93	49.78	58.15
3	8.69E-06	7.31E-06	4.31E-06	101.61	69.62	31.99	24	1.20E-06	8.22E-07	5.55E-07	115.54	48.13	67.40
4	8.95E-06	7.10E-06	4.28E-06	109.36	66.12	43.23	25	2.13E-06	1.15E-06	9.28E-07	129.39	23.93	105.47
5	8.11E-06	1.15E-06	3.56E-06	128.12	67.62	60.49	26	9.85E-07	6.64E-07	4.56E-07	115.87	45.47	70.40
6	9.01E-06	6.63E-06	4.00E-06	124.97	65.49	59.48	27	2.97E-06	2.26E-06	1.45E-06	104.93	56.12	48.81
7	9.88E-06	6.99E-06	4.56E-06	116.69	53.21	63.48	28	7.59E-07	5.39E-07	3.63E-07	109.18	48.56	60.62
8	1.38E-05	8.62E-06	5.19E-06	166.31	66.05	100.26	29	1.58E-06	8.70E-07	5.98E-07	164.08	45.51	118.57
9	7.58E-06	6.24E-06	3.76E-06	101.56	65.78	35.78	30	5.84E-08	4.84E-08	2.72E-08	115.10	78.29	36.81
10	2.33E-06	1.83E-06	1.02E-06	129.33	79.82	49.51	31	3.86E-06	9.94E-07	1.43E-06	169.89	30.54	139.35
11	1.31E-06	8.60E-07	6.16E-07	112.84	39.71	73.13	32	9.13E-07	6.07E-07	4.46E-07	104.96	36.30	68.66
12	1.86E-06	4.90E-07	9.67E-07	92.76	49.31	43.45	33	1.41E-06	9.37E-07	6.20E-07	126.53	51.07	75.45
13	1.65E-06	9.93E-07	6.55E-07	151.45	51.53	99.92	34	1.60E-06	1.25E-06	7.56E-07	110.96	64.81	46.14
14	3.40E-07	2.13E-07	1.52E-07	123.40	40.27	83.13	35	8.81E-08	7.06E-08	4.32E-08	103.64	63.24	40.40
15	7.41E-07	5.01E-07	3.56E-07	108.20	40.72	67.48	36	5.05E-08	4.20E-08	2.40E-08	110.36	74.99	35.37
16	9.45E-07	5.08E-07	3.47E-07	172.03	46.43	125.61	37	1.80E-07	1.00E-07	6.73E-08	168.25	49.15	119.09
17	3.65E-06	1.60E-06	1.33E-06	175.43	20.51	154.92	38	3.24E-07	2.25E-07	1.51E-07	114.56	48.81	65.75
18	1.35E-07	1.04E-07	5.67E-08	137.68	83.88	53.80	39	1.26E-06	8.48E-07	5.96E-07	112.07	42.28	69.79
19	2.66E-06	1.48E-06	1.08E-06	145.98	36.97	109.01	40	1.99E-07	1.60E-07	9.43E-08	110.60	69.21	41.39
20	3.66E-06	2.46E-06	1.78E-06	105.36	38.12	67.23	41	1.71E-05	3.02E-06	5.85E-06	191.65	48.34	143.31
21	2.70E-06	8.91E-07	1.32E-06	104.34	32.58	71.76	42	1.59E-05	1.15E-05	7.54E-06	110.35	52.56	57.79

Monitoring		30min					Monitoring		30min				
nodes	c ₁ /(g/m ³)	c ₂ /(g/m ³)	c ₃ /(g/m ³)	d ₁ /%	d ₂ /%	d ₃ /%	nodes	c ₁ /(g/m ³)	c ₂ /(g/m ³)	c ₃ /(g/m ³)	d ₁ /%	d ₂ /%	d ₃ /%
1	1.42E-05	2.90E-06	6.60E-06	115.15	56.04	59.11	22	3.75E-06	2.21E-06	1.67E-06	124.24	32.08	92.17
2	1.64E-05	3.08E-06	7.69E-06	113.00	59.96	53.04	23	9.64E-07	6.78E-07	4.63E-07	108.06	46.36	61.70
3	1.47E-05	1.01E-05	6.87E-06	114.56	46.29	68.27	24	3.91E-06	2.08E-06	1.52E-06	157.51	36.84	120.67
4	9.84E-06	3.01E-06	4.82E-06	104.15	37.48	66.67	25	1.88E-06	8.81E-07	5.91E-07	218.14	49.08	169.06
5	1.28E-05	8.59E-06	5.64E-06	126.67	52.27	74.40	26	2.84E-07	1.83E-07	1.28E-07	121.09	42.21	78.87
6	1.19E-05	8.26E-06	5.52E-06	116.32	49.74	66.58	27	1.33E-06	8.37E-07	5.52E-07	140.45	51.81	88.64
7	7.99E-06	5.51E-06	3.71E-06	115.47	48.70	66.77	28	9.96E-07	7.04E-07	4.76E-07	109.17	47.90	61.27
8	7.62E-06	4.63E-06	3.31E-06	130.22	40.03	90.19	29	3.40E-06	1.96E-06	1.20E-06	183.50	64.00	119.50
9	8.58E-06	4.56E-06	3.29E-06	160.77	38.48	122.30	30	1.78E-06	1.10E-07	6.59E-07	170.15	83.32	86.82
10	3.02E-06	2.34E-06	1.41E-06	114.20	65.84	48.35	31	2.52E-07	1.17E-06	3.26E-06	92.27	64.01	28.26
11	1.66E-06	9.41E-07	6.33E-07	161.36	48.65	112.71	32	1.19E-06	3.61E-07	5.77E-07	106.10	37.48	68.62
12	1.27E-06	8.40E-07	5.63E-07	124.72	49.18	75.54	33	7.80E-07	4.95E-07	3.67E-07	112.30	34.76	77.54
13	1.34E-05	9.16E-06	6.17E-06	117.70	48.38	69.32	34	7.21E-07	5.08E-07	3.46E-07	108.68	47.11	61.58
14	1.97E-08	9.57E-08	2.87E-07	93.11	66.59	26.52	35	6.85E-07	4.39E-07	3.09E-07	121.62	41.96	79.65
15	1.84E-06	1.23E-06	8.71E-07	111.45	41.62	69.83	36	3.59E-07	2.27E-07	1.75E-07	104.78	29.69	75.09
16	1.56E-06	8.17E-07	5.68E-07	174.74	43.89	130.86	37	8.69E-08	5.45E-08	4.08E-08	113.29	33.75	79.54
17	6.00E-07	4.42E-07	2.97E-07	101.97	48.70	53.26	38	4.58E-07	3.34E-07	2.42E-07	89.11	38.00	51.11
18	1.87E-06	1.16E-06	7.94E-07	136.15	45.90	90.26	39	1.68E-06	1.06E-06	8.17E-07	105.53	29.80	75.73
19	3.00E-06	1.75E-06	1.25E-06	139.82	39.86	99.96	40	4.80E-07	3.03E-07	2.21E-07	116.80	36.95	79.85
20	1.97E-06	1.23E-06	8.49E-07	132.30	44.76	87.54	41	9.80E-09	2.64E-07	3.95E-07	97.52	33.33	64.19
21	1.29E-08	9.91E-07	1.34E-06	99.03	25.85	73.18	42	8.56E-06	5.99E-06	4.06E-06	111.09	47.54	63.55

X=[7897,7799,7750,7750,7701,7653,7628,7579,7530,7437,7339,7265,7192,7264,7019,7093,6969,6993,6374,6472,6569,6641,6739,6836,6932,7005,5864,5864,5670,559,5402,5304,5207,5085,4038,3843,3648,3453,3259,3064,2844,2770]; Y=[6071,6101,6101,6101,6131,6162,6223,6284,6315,5697,5789,5850,5942,6035,6157,6126,6341,6434,4517,4549,4581,4643,4674,4706,4830,4923,4329,4329,4296,4296,4294,4293,4293,4292,4284,4283,4281,4280,4278,4308,4368,4522]; Z=1.5

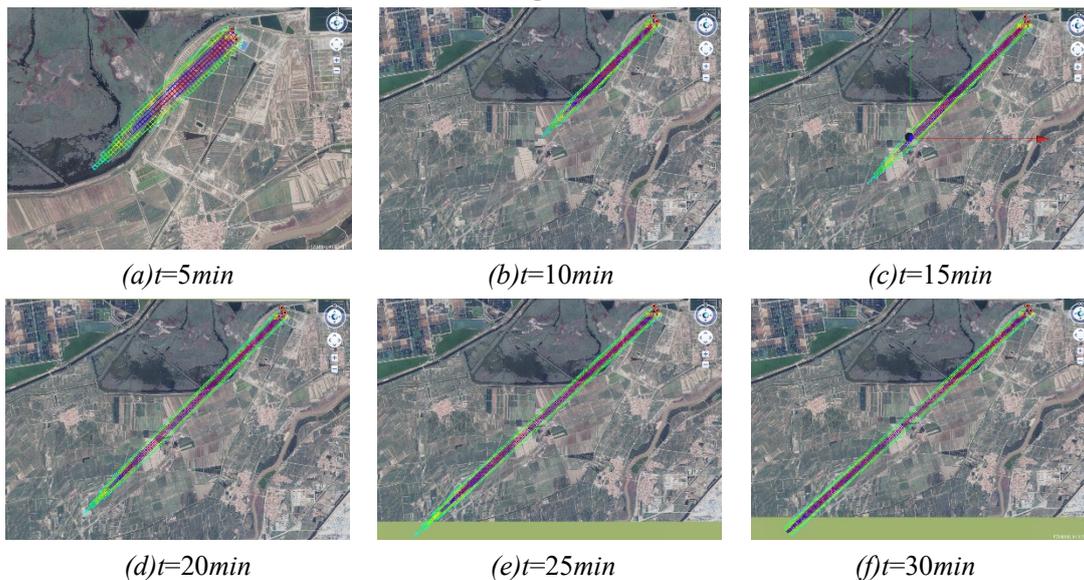
649 It can be seen from Table 7 and Figure 12 that the minimum value (89.11%) of d_1 occurs
 650 at monitoring point No.38 (30min), and the maximum value (218.14%) occurs at monitoring
 651 point No.25 (30min), with an average error percentage of 126.07%. The minimum value
 652 (20.51%) of d_2 occurred at monitoring point No.17 (20min), and the maximum value (87.64%)
 653 occurred at monitoring point No.32 (10min), with an average error percentage of 50.33%. The
 654 minimum value of d_3 (26.52%) occurred at monitoring point No.14 (30min), and the maximum
 655 value (169.06%) occurred at monitoring point No.25 (30min), with an average error difference
 656 of 75.74%. Compared with the analytical solution, the prediction accuracy of improved
 657 4DVAR algorithm is significantly improved.

658 5.2.3 Experimental process simulation

659 Based on the experimental parameter setting and STI calculation, the diffusion process,
 660 situation distribution and isosurface of tracer at different times are simulated and visualized.

661 (1) Simulation of diffusion process at different times

662 Figure 13 shows the simulated distribution of tracer diffusion concentration at 5min,
 663 10min, 15min, 20min, 25min and 30min along the wind direction.



666 Fig.13 SF₆ diffusion concentration at each time along the wind direction

669 As can be seen from Figure 13, with the increase of time, SF₆ gradually spreads to the
 670 downwind area along the wind direction. In case of interference from buildings or high walls,
 671 SF₆ will be transported with the air flow from both sides or above. As the underlying surface
 672 of the whole diffusion space is relatively flat, there is no slowdown or retention.

673 (2) Situation analysis at different times

674 Taking the height $h = 1.5m$ as an example, the situation distribution of tracer at 5min,
 675 10min, 15min, 20min, 25min and 30min along the wind direction is simulated and analyzed, as
 676 shown in Figure 14.

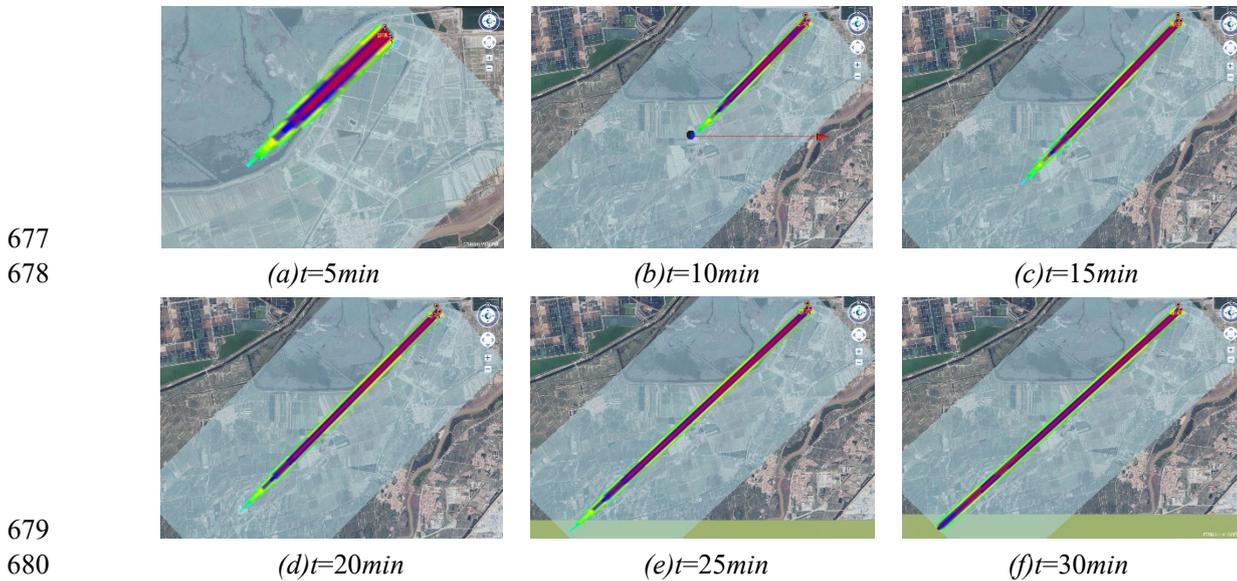


Fig.14 Distribution of SF₆ diffusion situation at each time

682 As can be seen from Figure 14, along the wind direction, with the increase of diffusion
 683 area, the SF₆ concentration gradually decreases and spreads outward. At the same time, under
 684 the continuous release of two point sources, the area close to the source is greatly affected by
 685 the source term and the situation is relatively complex. With the increase of distance, the impact
 686 of the source term is gradually weakened and the situation is gradually regularized.

687 (3) Isosurface diffusion simulation analysis

688 Still take the height of 1.5m as an example, extract the equivalent data at this height
 689 according to the extraction rules of severe pollution $0.001mg/m^3$, moderate pollution 0.00001
 690 mg/m^3 and mild pollution $0.00000001 mg/m^3$, and draw the equivalent surface, as shown in
 691 Figure 15.

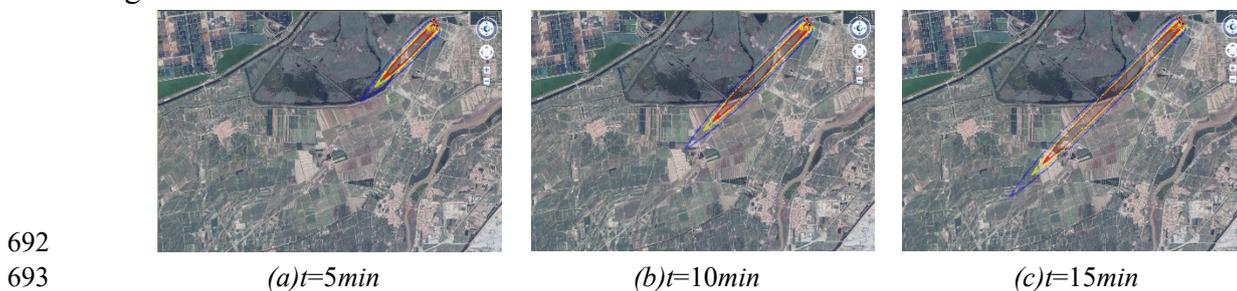




Fig.15 SF₆ concentration isosurface at each time

It can be seen from Figure 15 that the SF₆ concentration isosurface at each time presents a regular curve distribution, indicating that the main influencing factor is convection, and the underlying surface and turbulence are not obvious, which is due to the selected height and the flatness of the underlying surface.

5.3 Discussion of experimental results

5.3.1 Advantages and disadvantages analysis of inversion algorithm

In order to further verify the advantages and disadvantages of the inversion algorithm, KF, EnKF, 4DVAR and improved 4DVAR are used to calculate the experiments in this section. The number of experiments is set to 20, and the inversion results are compared and analyzed, as shown in Table 8.

Tab.8 Comparative analysis of STI algorithms

STI algorithm	Average value of $\epsilon_{intension}$	Average value of $\epsilon_{location}$	Experimental success rate
KF	107.56%	230.62%	50%
EnKF	85.22%	198.38%	60%
4DVAR	65.32%	195.50%	90%
Improved 4DVAR	35.37%	158.04%	100%

It can be seen from Table 8 that the improved 4DVAR algorithm is higher than other algorithms in both inversion accuracy and inversion success rate. The relative error of sources intension estimation can be reduced from 107.56% to 65.32%, and the relative error of sources location estimation can be reduced from 230.62% to 195.50%. By improving the adaptability of 4DVAR, the relative error of sources intension estimation can be reduced from 65.32% to 35.37%, the relative error of sources location estimation can be reduced from 195.50% to 158.04%, the success rate can be increased from 90% to 100%, whose STI ability is significantly improved.

5.3.2 Advantages and disadvantages analysis of optimization algorithm

Although IFPA has good effect in system efficiency optimization, its convergence and stability for 4DVAR model calculation have not been effectively verified. Therefore, on the basis of section 5.3.2, Genetic Algorithm (GA), Particle Swarm Optimization (PSO) and FPA

720 are used to calculate the experiment for 20 times, and compared with Quasi Newton Algorithm
 721 (QNA) and IFPA, as shown in Table 9. The convergence curves of the objective functional
 722 under different algorithms are shown in Figure 16.

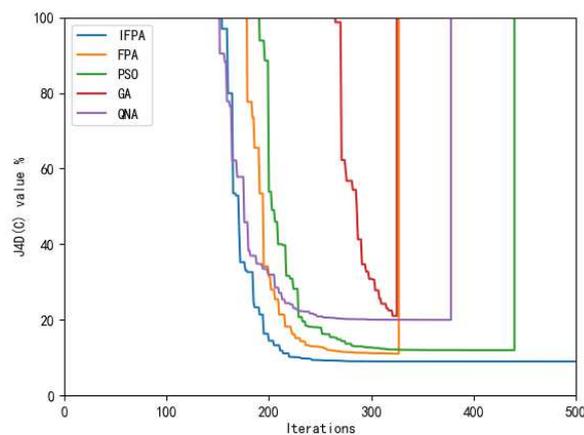
723 Tab.9 Advantages and disadvantages comparison of five optimization algorithms

Algorithms	Parameters setting	Average value of $\epsilon_{intension}$	Average value of $\epsilon_{location}$	Average value of $J_{4D}(C)$ minimum
QNA	$\epsilon=0.5$	65.32%	195.50%	20.3%
GA	$\xi=0.95, C_r=0.5$	52.50%	178.60%	15.7%
PSO	$\omega=0.95, C_1=0.5, C_2=0.5, V_m=1$	45.28%	170.10%	12.8%
FPA	Section 5.2.1	38.06%	162.50%	12.3%
IFPA	Section 5.2.1	35.37%	158.04%	9.3%

ϵ ----Allowable error, ξ ---- Variation factor, C_r ---- Crossover probability,

ω ---- Inertia factor, C_1, C_2 ---- Learning factor, V_m ---- Maximum speed

724 It can be seen from Table 9 that in the inversion of sources intensity, the inversion errors
 725 of IFPA, FPA, PSO and GA are effectively reduced compared with QNA, indicating that the
 726 intelligent optimization algorithm is an important way to improve the solution ability of
 727 4DVAR. In the sources location inversion, the position errors of the five algorithms are almost
 728 the same, and the average deviation is more than 150%, where the location of the hazard source
 729 can be further determined through the active traceability of the mobile platform. Finally,
 730 compared with the other four algorithms, IFPA has smaller error and higher accuracy, whether
 731 it is sources intension, sources location or target functional minimum, which shows obvious
 732 advantages in global optimization and optimization accuracy.



733 Fig.16 Convergence curve of target functional of five algorithms

734 As can be seen from Figure 16, IFPA, FPA and PSO are relatively high in convergence
 735 accuracy. In terms of convergence speed, IFPA is faster than other algorithms, and can find the
 736 optimal solution of the target functional in about 200 generations. In terms of algorithm stability,
 737

738 IFPA always searches smoothly after finding the optimal solution until the end of the iteration,
739 and the other four algorithms jump, where the execution error of algorithm cross mutation
740 strategy may be caused by too large calculation dimension, indicating that IFPA also has good
741 stability in the iterative solution of gradient model.

742 6 Conclusion

743 In this paper, the inverse problem of source intensity and source location estimation in
744 multi-point sources continuous release is discussed. A Taylor series upwind difference scheme
745 based on the solution of the forward problem is proposed. An improved 4DVAR algorithm is
746 used for effective numerical simulation inversion and tracer test verification, which show that
747 the improved 4DVAR algorithm can realize the STI under the conditions of unknown intensity,
748 unknown location, unknown intensity and location, and the accuracy, stability and convergence
749 of the algorithm can meet the traceability requirements. Next, the inversion of instantaneous
750 release and mixed release of multi-point sources will be studied.

751

752 **Conflict of interest:** The authors declare no conflicts of interest.

753 **Data Availability Statement:** The data used to support the findings of this study are
754 available from the corresponding author upon request.

755 REFERENCE

- 756 Andrew J. Annunzio, George S. Young, Sue Ellen Haupt., 2012. Utilizing state estimation to determine
757 the source location for a contaminant. *Atmospheric Environment*. 46, 580-589.
- 758 Castelli S.T., Tinarelli G., Reisin T., 2017. Comparison of atmospheric modelling systems simulating
759 the flow, turbulence and dispersion at the microscale within obstacles. *Environ. Fluid Mech.* 17
760 (5), 879-901.
- 761 Chabot V., Berre L., Desroziers G., 2017. Diagnosis and normalization of gridpoint background-error
762 variances induced by a block-diagonal wavelet covariance matrix. *Quarterly Journal of the Royal
763 Meteorological Society*. Vol. 143, No. 704, 1268-1279.
- 764 Cullen M. J. P., 2003. Four-dimensional variational data assimilation: A new formulation of the
765 background-error covariance matrix based on a potential-vorticity representation. *Q. J. R.
766 Meteorol. Soc.* 129, 2777-2796.
- 767 Fairbairn D., Pring S. R., Lorenc A. C., 2014. A comparison of 4DVar with ensemble data assimilation
768 methods. *Quarterly Journal of the Royal Meteorological Society*. Vol. 140, No. 678, 281-294.
- 769 François De l'Isle, Robert G. Owens, 2021. Superconsistent collocation methods with applications to
770 convection-dominated convection-diffusion equations. *Journal of Computational and Applied
771 Mathematics*. 391, 113367. <https://doi.org/10.1016/j.cam.2020.113367>.
- 772 Genç D., Yesilyurt Canan, Tuncel Gurdal, 2010. Air pollution forecasting in Ankara, Turkey using air
773 pollution index and its relation to assimilative capacity of the atmosphere. *Environmental
774 Monitoring and Assessment*. Vol.166, No.1-4, 11-27.
- 775 Hamba Fujihiro, Abe Satoshi, Kitazawa Daisuke, Kato Shinsuke, 2012. Filtering for the inverse
776 problem of convection-diffusion equation with a point source. *Journal of the Physical Society of*

777 Japan. 81, 114401. <http://dx.doi.org/10.1143/JPSJ.81.114401>.

778 Hassan K., Saima M., Izaz A., Poom K., Dumitru B., Muhammad A., 2020. Approximate analytical
779 fractional view of convection-diffusion equations. *Open Physics*. 18, 897-905.

780 He W. P., Feng G. L., Dong W. J., Li J. P., 2004. Comparison with solution of convection-diffusion by
781 several difference schemes. *ACTA PHYSICA SINICA*. Vol. 53, No. 10, 3258-3264. (in Chinese)

782 Huang S. X., Liu F., Sheng L., et al., 2018. Air pollution traceability based on adjoint method. *Science*
783 *Bulletin*. 63, 1594-1605. (in Chinese)

784 Jianbing Jin, Hai Xiang Lin, Heemink, A., Segers, A., 2018. Spatially varying parameter estimation for
785 dust emissions using reduced-tangent-linearization 4DVar. *Atmospheric Environment*. Vol. 187,
786 358-373.

787 Karl Barnes, 2017. *Mathematical Methods of Physics*. Magnum Publishing.

788 Kazakov A. L., Spevak L. F., 2021. Exact and approximate solutions of a problem with a singularity for
789 a convection-diffusion equation. *Journal of Applied Mechanics and Technical Physics*. Vol. 62, No.
790 1, 18-26.

791 Kim Sung-Min, Kim Hyun Mee, 2019. Forecast Sensitivity Observation Impact in the 4DVAR and
792 Hybrid-4DVAR Data Assimilation Systems. *Journal of Atmospheric & Oceanic Technology*. Vol.
793 36, No. 8, 1563-1575.

794 Kuo C. L., Liu C. S., Chang J. R., 2016. The modified polynomial expansion method for identifying
795 the time dependent heat source in two-dimensional heat conduction problems. *International*
796 *Journal of Heat and Mass Transfer*. 92, 658-664.

797 Liu C. S., Chen Y. W., Chang J. R., 2019. Solving a nonlinear convection-diffusion equation with source
798 and moving boundary both unknown by a family of homogenization functions. *International*
799 *Journal of Heat and Mass Transfer*. Vol. 138, 25-31.

800 Liu H. L., 2021. Analysis of direct discontinuous Galerkin methods for multi-dimensional convection-
801 diffusion equations. *Numerische Mathematik*. 147, 839-867. [https://doi.org/10.1007/s00211-021-](https://doi.org/10.1007/s00211-021-01183-x)
802 [01183-x](https://doi.org/10.1007/s00211-021-01183-x).

803 Liu Y., Li H., Sun S. D., Fang S., 2017. Enhanced air dispersion modelling at a typical Chinese nuclear
804 power plant site: Coupling RIMPUFF with two advanced diagnostic wind models. *Journal of*
805 *Environmental Radioactivity*. Vol. 175-176, No. 9, 94-104.

806 Liu Y., Liu X. J., Li H., et al., 2019. Numerical study on source term inversion of truncated total least
807 squares variation. *Nuclear power engineering*. Vol. 40, No. 1, 120-125. (in Chinese)

808 Liu YA, Huang HLA, Gao W, Lim Agnes H. N., Liu CS, Shi RH., 2015. Tuning of background error
809 statistics through sensitivity experiments and its impact on typhoon forecast. *Journal of Applied*
810 *Remote Sensing*. Vol.9, No.1, 096051.

811 Liu Y. A., Li Z. L., Huang M. L., 2019. Towards a data-derived observation error covariance matrix for
812 satellite measurements. *Remote Sens*. 11, 1770-1794.

813 Ma Jianwen et al., 2013. *Research and experiment of Data Assimilation Algorithm*. Beijing: Science
814 Press. (in Chinese)

815 Naus S., Montzka S. A., Patra P. K., Krol M. C., 2021. A three-dimensional-model inversion of methyl
816 chloroform to constrain the atmospheric oxidative capacity. *Atmos. Chem. Phys.*, 21, 4809-4824.

817 Pan Dartzi, 2017. A high-order finite volume method for solving one-dimensional convection and
818 diffusion equations. *Numerical Heat Transfer, Part B*. Vol. 71, No. 6, 533-548.
819 <https://doi.org/10.1080/10407790.2017.1326769>.

820 Pan Y. P., Wang Y. S., 2015. Atmospheric wet and dry deposition of trace elements at 10sites in Northern

821 China. Atmospheric Chemistry and Physics. Vol. 15, No. 2, 951-972.

822 Parrish D. F., Derber J. C., 1992. The National Meteorological Center's Spectral Statistical Interpolation
823 analysis system. *Mon Wea Rev.* 120, 1747-1763.

824 Quinn C., O'kane T.J., Kitsios V., 2020. Application of a local attractor dimension to reduced space
825 strongly coupled data assimilation for chaotic multiscale systems (Article). *Nonlinear Processes
826 in Geophysics.* Vol. 27, No. 1, 51-74.

827 Reza Pourgholi, Akram Saeedi, Abbas Hosseini, 2019. Determination of nonlinear source term in an
828 inverse convection-reaction-diffusion problem using radial basis functions method. *Iranian Journal
829 of Science and Technology, Transactions A: Science.* Vol. 43, 2239-2252.
830 <https://doi.org/10.1007/s40995-017-0379-6>.

831 Ruan Z. S., Zhang S., 2020. Simultaneous inversion of time-dependent source term and fractional order
832 for a time-fractional diffusion equation. *Journal of Computational and Applied Mathematics.* Vol.
833 368, 112566.

834 Rutherford I., 1972. Data assimilation by statistical interpolation of forecast error fields. *Atmospheric
835 Science.* 29, 9-15.

836 Sun C. L., Li G. S., Jia X. Z., 2017. Simultaneous inversion for the diffusion and source coefficients in
837 the multi-term TFDE. *Inverse Problems in Science and Engineering.* Vol.25, No.11, 1618-1638.

838 Varella H., Berre L., Desroziers G., 2011. Diagnostic and impact studies of a wavelet formulation of
839 background-error correlations in a global model. *Q. J. R. Meteorol. Soc.* 137, 1369-1379.

840 Veronika Valler, Jorg Franke, Stefan Brönnimann, 2019. Impact of different estimations of the
841 background-error covariance matrix on climate reconstructions based on data assimilation.
842 *Climate of the Past.* Vol. 15, No. 4, 1427-1441.

843 Viktor Popov, Tu Thanh Bui, 2010. A meshless solution to two-dimensional convection-diffusion
844 problems. *Engineering Analysis with Boundary Elements.* 34, 680-689.

845 Wang L., Yang X. G., Wang H. L., Chai Z. H., Wei Z. C., 2021. A modified regularized lattice Boltzmann
846 model for convection-diffusion equation with a source term. *Applied Mathematics Letters.* 112,
847 106766. <https://doi.org/10.1016/j.aml.2020.106766>.

848 Wei T., Xu M. T., 2016. An integral equation approach to the unsteady convection-diffusion equations.
849 *Applied Mathematics and Computation.* Vol.274, No.C, 55-64.

850 Yang X. S., 2012. Flower pollination algorithm for global optimization. *International Conference on
851 Unconventional Computing and Natural Computation.* 240-249.

852 Yasumasa Miyazawa, Max Yaremchuk;Sergey M. Varlamov, Toru Miyama, Kunihiro Aoki, 2020.
853 Applying the adjoint-free 4dVar assimilation to modeling the Kuroshio south of Japan. *Ocean
854 Dynamics.* Vol. 70, No. 8, 1129-1149.

855 Zhu X. Z., Han C. S., Liu R. Q., Yan G. H., Gu J., 2021. One universal method of complex system
856 reliability, maintainability, supportability, testability quotas design and trade-off based on
857 improved flower pollination algorithm. *Quality Reliability Engineering International,* Vol. 37, No.
858 4, 1524-1543.

859 Zhang J. C., Wang Y. P., 2014. A method for inversion of periodic open boundary conditions in two-
860 dimensional tidal models(Article). *Computer Methods in Applied Mechanics and Engineering.* Vol.
861 275, 20-38.