PARAMETER IDENTIFICATION TECHNIQUES APPLIED TO AN ENVIRONMENTAL POLLUTION MODEL

Yuepeng Wang[a*]

[a,b,d]School of Mathematics and Statistics
Nanjing University of Information Science and Technology (NUIST), Nanjing, China, 210044

Yue Cheng[b], I. Michael Navon[c*] and Yuanhong Guan[d]

[c] Department of Scientific Computing
Florida State University, Tallahassee, FL 32306, USA

(Communicated by Kok Lay Teo)

Abstract. The retrieval of parameters related to an environmental model is explored. We address computational challenges occurring due to a significant numerical difference of up to two orders of magnitude between the two model parameters we aim to retrieve. First, the corresponding optimization problem is poorly scaled, causing minimization algorithms to perform poorly (see *Gill et al.,practical optimization, AP,1981,401pp*). This issue is addressed by proper rescaling. Difficulties also arise from the presence of strong nonlinearity and ill-posedness which means that the parameters do not converge to a single deterministic set of values, but rather there exists a range of parameter combinations that produce the same model behavior. We address these computational issues by the addition of a regularization term in the cost function. All these computational approaches are addressed in the framework of variational adjoint data assimilation. The used observational data are derived from numerical simulation results located at only two spatial points. The effect of different initial guess values of parameters on retrieval results is also considered. As indicated by results of numerical experiments, the method presented in this paper achieves a near perfect parameter identification, and overcomes the indefiniteness that may occur in inversion process even in the case of noisy input data.

1. **Introduction.** In the atmospheric boundary layer (ABL), air pollution has increasing importance in urban, industrialized regions. The pollutant transport is frequently affected by many dynamic processes such as advection, diffusion etc. due to the role played by turbulence of atmospheric motion. Many studies have been carried out to understand the mechanisms of the pollutant transport and to predict transport processes in the ABL by various means such as field measurements, laboratory experiments, analytical studies, and numerical models. For the purpose of pollutant prediction, numerical modeling is by far the most useful and effective method. Difficulties in estimating model parameters resulting from both laboratory experiments and field observations still prevent numerical modeling of

pollutant transport from achieving a high level of predictability. In order to tackle this problem, the authors have used Levenberg-Marquardt algorithm coupled with the generalized eigenfunction expansion method to solve an inverse problem and obtained a statistical agreement with the true value of the model parameters[25]. However there remain some challenges we face in practical application. First, the computation of sensitivities required for performing iterative algorithm can be very costly, rending it computationally impracticable for large-scale problems (as is the case for a typical meteorological model, e.g. 10^7 variables). This is particularly true if a straightforward way of computing the gradient (or sensitivity) is implemented in the following manner: perturbing each control variable in turn and estimating the change in the cost function; Second, it is indeed convenient to apply eigenfunction to expand unknown function (or generalized Fourier series expansion) in a certain direction. However, the eigenfunctions depend heavily on the boundary conditions of the problem. Computing the eigenfunction will become quite a difficult task when the boundary of the region of interest does not have a regular shape, which will affect the application of method itself and further extension. In this case, exploring and selecting a more effective method becomes very important. For such a situation, the variational adjoint method [14, 1, 16, 2] turns out to be the most adequate option.

The variational adjoint approach is commonly used to compute the sensitivity coefficients and the gradient of the cost function. The adjoint method of sensitivity computation is particularly useful and has proven its efficiency in operational meteorology and other branches of geosciences [18, 17]. The variational adjoint method can yield the exact gradient of the cost function with respect to the control variables by integrating the adjoint model only once backwards in time. Such a backward integration of the adjoint model is of similar complexity to a single integration of the forward model. Another key advantage of adjoint variational data assimilation is the possibility to minimize the cost function using standard unconstrained minimization algorithms[30].

In this paper, we provide an application of variational adjoint method to the retrieval of two model parameters namely the diffusivity and surface friction velocity. This theoretical exploration will not only present an alternative means of solving the problem, but, more significantly, provide an insight into practical measurements for some important model parameters. According to the requirement of variational adjoint method, we will transform the retrieval of the relevant parameters into an optimal control problem constrained by a partial differential equation (PDE), and the parameters to be estimated can be regarded as additional controls during the procedure of minimizing the cost functional[12, 27, 24]. Some challenges remain to be addressed. First, this optimization problem is poorly scaled, and minimization algorithms that fail to rescale properly will fail to converge [8]. This issue can be solved by proper rescaling. Second, difficulties arise from the presence of strong nonlinearity and ill-posedness which means that the parameters do not converge to a single deterministic set of values, but rather there exists a range of complementary combinations that produce the same model behavior. We will deal with this lack of identifiability by resorting to inclusion of a regularization term in the cost function.

The main contribution of this paper is to (a) develop a variational adjoint assimilation system using the simulated observational data from only two spatial points can retrieve the expected parameters, i.e., diffusivity and surface friction velocity related to the ABL; (b) numerical experiments reveal that the regularization term

included in cost function plays a key role in the process of estimation of parameters when noise is present in measurement data; (c) studying the effects of different initial guess values on the retrieval results. The outcome of this paper will contribute to the reduction of the practical field experimentation carried out for the determination of the relevant model parameters. The remainder of this paper is organized as follows. In section 2 we provide the optimal control system based on the adjoint method. In section 3, numerical simulation experiments are performed, and finally conclusions are provided in section 4.

2. Variational data assimilation system.

2.1. **Problem description.** The forward (or direct) problem used in this paper is the same as the one used in [25], which is based on the following 2-D mathematical model:

$$\frac{\partial c}{\partial t} + U(z)\frac{\partial c}{\partial x} = K_1\frac{\partial^2 c}{\partial x^2} + \frac{\partial}{\partial z}(K_2\frac{\partial c}{\partial z}) + s\delta(x - x_c)\delta(z - z_c)$$
$$(0 < x < A,\ 0 < z < H,\ t > 0) \qquad (2.1.1a)$$

subject to the initial-boundary conditions

$$\frac{\partial c}{\partial z}|_{z=0} = 0, \frac{\partial c}{\partial z}|_{z=H} = 0 \qquad (2.1.1b)$$

$$c|_{x=0} = 0, \frac{\partial c}{\partial x}|_{x=A} = 0 \qquad (2.1.1c)$$

$$c|_{t=0} = 0 \qquad (2.1.1d)$$

where c is the time-averaged pollutant concentration, and K_1 and K_2 are the diffusivities along horizontal and vertical directions, respectively, given by

$$K_1 = Const., \qquad (2.1.2)$$

and

$$K_2 = k(\mu_*)_0 H(\frac{z}{H})(1 - \frac{z}{H})(1 - 22\frac{z}{L})^{1/4} \qquad (2.1.3)$$

$U(z)$ is the velocity profile that depends on vertical height z like

$$U(z) = \frac{(\mu_*)_0}{k}\ln(\frac{z}{z_0}) + \ln[\frac{(1 + (u_0)^2)(1 + u_0)^2}{(1 + u^2)(1 + u)^2}]$$

$$+\frac{(\mu_*)_0}{k}[2(\arctan(u) - \arctan(u_0)) + \frac{2L}{33H}(u^3 - u_0^3)] \qquad (2.1.4)$$

and was assumed to be greater than zero here, while the parameters u and u_0 are given by

$$u = 1 - 22\frac{z}{L} \qquad (2.1.5)$$

$$u_0 = 1 - 22\frac{z_0}{L} \qquad (2.1.6)$$

k is the von Karman's constant with the value of 0.4, z_0 is the surface roughness and L is the Monin-Obukov length, aiming to construct the stability-dependent function $\frac{H}{L}$. When $\frac{H}{L}$ is greater than zero, the stratification is stable; while if $\frac{H}{L}$ is smaller than zero, the profile behavior is unstable. And $(\mu_*)_0$ is the surface friction velocity.

For the sake of illustrating our method, we select $K_1, (\mu_*)_0$ as the retrieved parameters due to their significant numerical difference of up to two orders of magnitude. Assume $K_1, (\mu_*)_0$ are given, the evolution of state variable with space and time can

be obtained by solving the problem (2.1.1). This pertains to the direct problem. On the contrary, given a set of prior measurements, we are required to determine the parameters K_1 and $(\mu_*)_0$, which belong to the inverse problem. The present research effort aims to retrieve the parameters K_1 and $(\mu_*)_0$ in the framework of variational adjoint data assimilation.

In order to determine simultaneously an optimal $K_1, (\mu_*)_0$ such that a reasonable solution satisfies the observed measurement data, we must first define an objective function:

$$J(K_1, (\mu_*)_0) = \frac{1}{2} \int_0^T \sum_{i=1}^{N_s} \int_\Omega (c(Q,t) - c^{obs}(Q,t))^2 \delta(Q - Q_i) d\Omega dt$$

$$+ \frac{\gamma}{2} \int_0^T \int_\Omega [(\frac{\partial c}{\partial x})^2 + (\frac{\partial c}{\partial z})^2] d\Omega dt \qquad (2.1.7)$$

which measures the discrepancy between the solution given by the original model (2.1.1) for the continuous assimilations and the known observation data. The spatial domain is referred to as $[0, A] \times [0, H]$. Here Q represents spatial position (x, z), and N_s stands for the number of spatial observational locations (in present study, $N_s = 2$). While the second term appearing in the RHS of (2.1.7) is referred to as regularization term which can stabilize the minimization process, and γ is a regularization parameter. This kind of regularization term which we add to the objective function J describes a priori constraint one would like the state function to be subject to, instead of including a constraint for the inverse solution itself like the one frequently used, $[K_1^2 + (\mu_*)_0^2]$. The main reasons for adopting this regularization procedure stem from the following considerations:

(1). The effect of $[K_1^2 + (\mu_*)_0^2]$ tends to seek for the minimum-norm solution. And when the noisy components in the observation data lead to the changes of solution which will induce larger H_0^1-seminorm of state function, we attempt to use the H_0^1-seminorm constraint of state function, rather than the minimum-norm constraint of inverse solution. It is shown experimentally later that imposing such a-priori constraint will lead to a stable solution within a suitable range of observational noise;

(2). From the point of view of physics, the H_0^1-seminorm is essentially a component part of energy functional derived in variational setting from the equations (2.1.1a-d). Addition of this term to the cost functional is equivalent to performing a bound constrained minimization of energy seminorm. As for the $[K_1^2 + (\mu_*)_0^2]$, it is feasible in mathematics to regard it as a constraint, but its physical meaning is unknown. So it is natural to select $\int_0^T \int_\Omega [(\frac{\partial c}{\partial x})^2 + (\frac{\partial c}{\partial z})^2] d\Omega dt \alpha$ as a regularized stable functional in the current study;

(3). The second term in the RHS of (2.1.7) can also serve as a forcing term of the adjoint equation (2.2.6) and the relevant smooth information of the state function $c(x, z, t)$ is passed on to the calculated gradient obtained using the adjoint equation.

For simplicity of notation, we denote $K_1, (\mu_*)_0$ by $P = [P_1, P_2]$. So solving $K_1, (\mu_*)_0$ from the problem (2.1.1) and the corresponding observation data will be transformed into the following optimal control problem: Find P^{opt} such that

$$J[P^{opt}] = minJ[P] \qquad (2.1.8)$$

where the control variable P is related to the state variable $c(x, z, t)$ via the problem (2.1.1). The classical way to solve a problem such as (2.1.8) is usually called gradient-type descent method. So the key issue is to find an efficient way to calculate the gradient, with gradient components denoted as $\nabla_{K_1} J$ and $\nabla_{(\mu_*)_0} J$ of J with respect to the parameters K_1 and $(\mu_*)_0$. Their calculation can be done by using variational adjoint method [31].

2.2. **Calculation of gradient and variational adjoint method.** In this section, we show that the calculation of gradient will be related to the derivation of the adjoint problem. The whole process can be divided into the following steps.

Step one, the sensitivity equation (or tangent linear equation). Let us consider the perturbation of P : $P \to \tilde{P} = P + \alpha \hat{P}$, where α is a perturbation parameter tending to zero, and the $\hat{P}$ stands for the direction of perturbation. Correspondingly, we have $U(z) \to \tilde{U}(z) = U(z) + \alpha \hat{U}(z)$, $K_2 \to \tilde{K}_2 = K_2 + \alpha \hat{K}_2$ and $c \to \tilde{c} = c + \alpha \hat{c}$, where $\hat{U}(z)$, $\hat{K}_2$, and $\hat{c}$ may be defined as follows:

$$\hat{U}(z) = \lim_{\alpha \to 0} \frac{\tilde{U}(z) - U(z)}{\alpha}, \hat{K}_2 = \lim_{\alpha \to 0} \frac{\tilde{K}_2 - K_2}{\alpha}, \hat{c} = \lim_{\alpha \to 0} \frac{\tilde{c} - c}{\alpha} \qquad (2.2.1)$$

$\hat{c}$ is the solution of the sensitivity equation

$$\begin{cases} \frac{\partial \hat{c}}{\partial t} + U(z)\frac{\partial \hat{c}}{\partial x} = K_1 \frac{\partial^2 \hat{c}}{\partial x^2} + \frac{\partial}{\partial z}\left(K_2 \frac{\partial \hat{c}}{\partial z}\right) + \hat{K}_1 \frac{\partial^2 c}{\partial x^2} - \\ \qquad\qquad\qquad\qquad \hat{U}(z)\frac{\partial c}{\partial x} + \frac{\partial}{\partial z}\left(\hat{K}_2 \frac{\partial c}{\partial z}\right) \\ \frac{\partial \hat{c}}{\partial z}|_{z=0} = 0 , \quad \frac{\partial \hat{c}}{\partial z}|_{z=H} = 0 \\ \hat{c}|_{x=0} = 0 , \quad \frac{\partial \hat{c}}{\partial x}|_{x=A} = 0 \\ \hat{c}|_{t=0} = 0 \end{cases} \qquad (2.2.2)$$

Step two, the change of the cost function. According to the definition of gradient, we have

$$J'[P; \hat{P}] = (\nabla_P J, \hat{P}) = (\nabla_{P_1} J, \hat{P}_1) + (\nabla_{P_2} J, \hat{P}_2) \qquad (2.2.3)$$

On the other hand,

$$J'[P, \hat{P}] = \lim_{\alpha \to 0} \frac{J[P + \alpha \hat{P}] - J[P]}{\alpha}$$

$$= \int_0^T \int_\Omega [\sum_{i=1}^{N_s} (c - c^{obs})\delta(Q - Q_i) - \gamma(\frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial z^2})]\hat{c}d\Omega dt \qquad (2.2.4)$$

Step three, the adjoint problem. Multiplying each side of the first equation in Eqs.(2.2.2) by $\lambda(x, z, t)$, and integrating the resulting equation on the domain $[0, A] \times [0, H] \times [0, T]$, we obtain

$$\int_0^T \int_0^A \int_0^H \lambda \frac{\partial \hat{c}}{\partial t} dzdxdt + \int_0^T \int_0^A \int_0^H \lambda U(z)\frac{\partial \hat{c}}{\partial x} dzdxdt$$

$$+ \int_0^T \int_0^A \int_0^H \lambda \hat{U}(z)\frac{\partial c}{\partial x} dxdzdt$$

$$= \int_0^T \int_0^A \int_0^H \lambda \hat{K}_1 \frac{\partial^2 c}{\partial x^2} dxdzdt + \int_0^T \int_0^A \int_0^H \lambda K_1 \frac{\partial^2 \hat{c}}{\partial x^2} dxdzdt$$

$$+ \int_0^T \int_0^A \int_0^H \lambda \frac{\partial}{\partial z}\left(\hat{K}_2 \frac{\partial c}{\partial z}\right) dzdxdt + \int_0^T \int_0^A \int_0^H \lambda \frac{\partial}{\partial z}(K_2 \frac{\partial \hat{c}}{\partial z}) dzdxdt \qquad (2.2.5)$$

Applying the integration by parts, and assuming that the following relations (adjoint problem) holds

$$
\begin{cases}
-\dfrac{\partial \lambda}{\partial t} - U(z)\dfrac{\partial \lambda}{\partial x} - P_1\dfrac{\partial^2 \lambda}{\partial x^2} - \dfrac{\partial}{\partial z}\left(K_2\dfrac{\partial \lambda}{\partial z}\right) = \sum\limits_{i=1}^{N_s}(c - c^{obs})\delta(Q - Q_i) \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad -\gamma(\dfrac{\partial^2 c}{\partial x^2} + \dfrac{\partial^2 c}{\partial z^2}) \\
\dfrac{\partial \lambda}{\partial z}\big|_{z=0} = 0, \dfrac{\partial \lambda}{\partial z}\big|_{z=H} = 0 \\
\lambda\big|_{x=0} = 0, \left[\lambda U(z) + P_1\dfrac{\partial \lambda}{\partial x}\right]\big|_{x=A} = 0 \\
\lambda\big|_{t=T} = 0
\end{cases}
\tag{2.2.6}
$$

Then the following equation can be obtained

$$
J'[P;\hat{P}] = -\int_0^T\int_0^A\int_0^H \frac{\partial c}{\partial z}\frac{\partial \lambda}{\partial z}\frac{\partial K_2}{\partial P_2}\hat{P}_2 dz dx dt
$$

$$
-\int_0^T\int_0^A\int_0^H \lambda\frac{\partial c}{\partial x}\frac{\partial U}{\partial P_2}\hat{P}_2 dx dz dt + \int_0^T\int_0^A\int_0^H \lambda\frac{\partial^2 c}{\partial x^2}\hat{P}_1 dx dz dt
$$

$$
= \int_0^T\int_0^A\int_0^H \lambda\frac{\partial^2 c}{\partial x^2}\hat{P}_1 dx dz dt
$$

$$
+\int_0^T\int_0^A\int_0^H \left(-\frac{\partial c}{\partial z}\frac{\partial \lambda}{\partial z}\frac{\partial K_2}{\partial P_2} - \lambda\frac{\partial c}{\partial x}\frac{\partial U}{\partial P_2}\right)\hat{P}_2 dz dx dt
\tag{2.2.7}
$$

Step four, the gradient of cost function with respect to $P = [K_1, (\mu_*)_0]$**.** Comparing the RHS of eqs. (2.2.3) and (2.2.7), the gradient of cost function with respect to the unknown parameters is followed:

$$
\frac{\partial J}{\partial P_1} = \int_0^T\int_0^A\int_0^H \lambda\frac{\partial^2 c}{\partial x^2}dx dz dt
\tag{2.2.8}
$$

$$
\frac{\partial J}{\partial P_2} = \int_0^T\int_0^A\int_0^H \left(-\frac{\partial c}{\partial z}\frac{\partial \lambda}{\partial z}\frac{\partial K_2}{\partial P_2} - \lambda\frac{\partial c}{\partial x}\frac{\partial U}{\partial P_2}\right)dz dx dt
\tag{2.2.9}
$$

Step five, gradient test. The gradient obtained above by the adjoint method [20] should satisfy numerically the relation below

$$
\tilde{\rho}(\alpha) = \frac{J(P + \alpha\hat{P}) - J(P)}{\alpha < \nabla J(P), \hat{P} >}
\tag{2.2.10}
$$

where the $\nabla J(P) = [\frac{\partial J}{\partial P_1}; \frac{\partial J}{\partial P_2}]$, denoted as g . And the variation of function with perturbation α can be seen in Figure 1.

Having the gradient of the cost function at our disposal, we can use various gradient-based unconstrained minimization algorithms. Compared with the steepest descent method and Newton method, evident advantages of the conjugate gradient method are its convergence rate, and low memory requirements [19, 22]. To obtain the optimal parameter $P = [K_1, (\mu_*)_0]$, the main steps of the current nonlinear conjugate gradient (NCG) method include:

a). Start with an initial guess P^g , if the J less than a given ε (stopping criterion), then stop, otherwise continue the next step;

b). Solve the direct problem (2.1.1) and store intermediate steps;

c). Solve the adoint problem (2.2.6), and considering eqs. (2.2.8) and (2.2.9), we obtain the desired gradient $g^{(k)} = (\frac{\partial J}{\partial P_1}, \frac{\partial J}{\partial P_2})^{(k)}$;

d). Determine the conjugate descent direction $D^{(k)} = -g^{(k)} + \beta_{k-1}D^{(k-1)}$ by
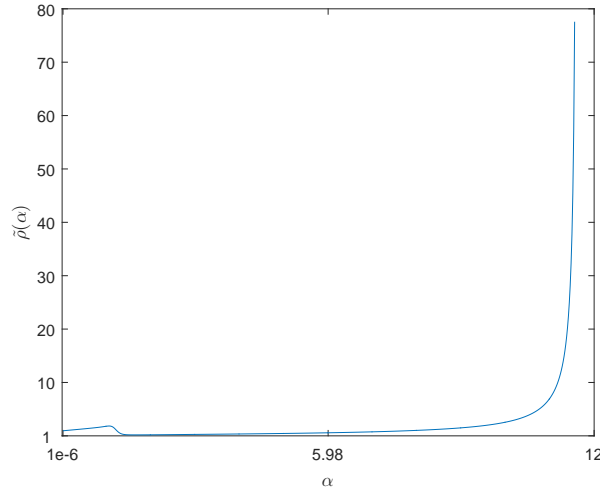$\beta_{k-1} = \frac{g^{(k)T}(g^{(k)} - g^{(k-1)})}{g^{(k-1)T}g^{(k-1)}}$ (Polak-Ribiere);

FIGURE 1. Gradient test

TABLE 1. Comparison between the methods of **NCG, ncg, lbfgs** and **tn** with different initial guess. The true parameters are (0.38,50).

Method / Cases	NCG		ncg		lbfgs		tn	
	Iter.	Result	Iter.	Result	Iter.	Result	Iter.	Result
[0.3;40]	2	[0.3751; 50.6269]	2	[0.3759; 50.7398]	5	[0.3752; 50.7179]	6	[0.3753; 50.7524]
[0.15;18]	3	[0.3794; 50.0722]	2	[0.3962; 52.4113]	11	[0.3791; 50.1205]	4	[0.3797; 50.0491]
[0.01;1.2]	3	[0.3771; 50.4490]	2	[0.6661; 89.1588]	10	[0.3773; 50.4361]	4	[0.3534; 54.4259]

e). Choose the step size $\rho^{(k)} = argmin J(P^{(k-1)} + \rho^{(k)}D^k)$ by linesearch;

f). The $k-th$ iteration formulation can be expressed as $P^{(k)} = P^{(k-1)} + \rho^{(k)}D^{(k)}$. For a detailed description, please see Figure 2 below. Note that the adjoint model used in the next numerical experiments is derived through the *discretize-then-differentiate* approach[9, 7], in which one first discretizes the direct model (2.1.1), then differentiates it by hand and obtains the discrete version of the adjoint model. Different from the existing nonlinear conjugate gradient method (ncg)included in the published optimal tool–Poblano V1.0 [4], the present nonlinear conjugate gradient method (NCG) introduces the classical golden section search, and considers the update $\beta_{k-1}^{+} = \max\{0, \beta_{k-1}\}$ [22], which can be competitive for the current problem in comparison with the algorithm (ncg) , as well as unconstrained minimization algorithms (lbfgs) and (tn) included in Poblano. This can be found from the Table 1, which is derived based on the whole observation data and without taking into account the regularization term temporarily. This is sufficient to outline the differences between the NCG and other methods related to Poblano.
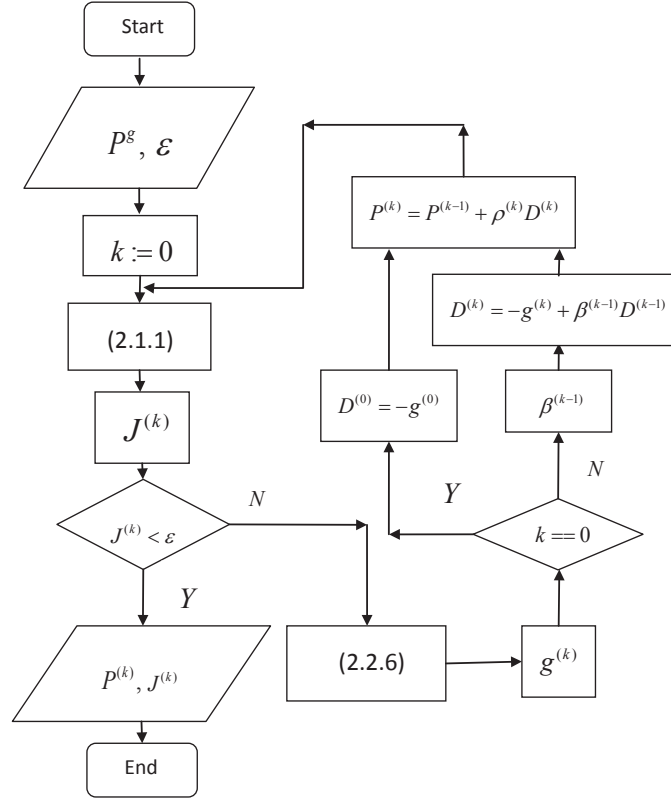
FIGURE 2. Flow chart of parameter estimation with NCG method

3. **Numerical experiments.** As mentioned above in Section 2, we are required to solve the direct problem (2.1.1). Since its analytical solution is very difficult to obtain, so we had to employ finite difference numerical discretization.

The Alternating Direction Implicit (ADI) method, first introduced by Peaceman and Rachford, is a finite difference method for solving the heat equation or the diffusion equation [23]. From then on, it is used frequently to the numerical solution of parabolic, hyperbolic [15] and elliptic partial differential equations. It can be viewed as an iterative method to solve a higher dimensional problem by solving a series of lower dimensional problems repeatedly. For the current 2D case, the ADI iteration process from n to $n+1$ can be separated into two parts, the x-axis sweeping at time level n and the z-axis sweeping at time level $n + \frac{1}{2}$. We use uniform grid in both space and time. It is assumed that the discrete functions are defined on an $N \times M$-grids in space, i.e. $\Omega = [0, A] \times [0, H]$. The following notation will be used: $x_j = j\Delta x$, $z_i = i\Delta z$, $t^n = n\Delta t$, $c_{ij}^n = c(x_j, z_i, t^n)$. Here $\Delta x = \frac{A}{N}$, and $\Delta z = \frac{H}{M}$. And the central difference scheme is used for the first- and second-order derivatives of c with respect to x. While the vertical diffusion term $\frac{\partial}{\partial z}(K_2 \frac{\partial c}{\partial z})$ is approximated

specially by the following formula at time level n,

$$\frac{1}{\triangle z^2}\{K_2(z_i + \frac{\triangle z}{2})c_{j,i+1}^n - [K_2(z_i + \frac{\triangle z}{2}) + K_2(z_i - \frac{\triangle z}{2})]c_{j,i}^n + K_2(z_i - \frac{\triangle z}{2})c_{j,i-1}^n\}$$

When performing the numerical computation, where diffusion terms and convection term are in time implicit and explicit, respectively. Based on the numerical scheme of the forward problem, the discretized version of the adjoint problem can be formed through careful modification of boundary conditions, forced terms and so on, and solved with almost the same routine as that of the forward problem. For the purpose of the numerical experiments, we assume that the true parameters are $K_1 = 50m^2s^{-1}$, $(\mu_*)_0 = 0.38ms^{-1}$, and in order to ensure the stability of numerical computation, the other related parameters are also set as, respectively, $A = 6000m$, $H = 1120m$, $(x_e, z_e) = (100m, 115m)$, $L = -71m$, $z_0 = 0.6m$, $M = N = 40$, $T = 560s$, $\Delta t = 56s$. The Dirac function $\delta(x - x_c)$ in source term [29] is approximated by

$$\delta_k(x - x_c) = \frac{k}{2\cosh^2(k(x - x_c))}. \tag{3.1}$$

In this paper, we take $k = 20$.

In order to retrieve the parameters P, we use simulated measurements in the forthcoming analysis, which can be obtained from the solution of the direct problem (2.1.1) at the two locations $Q_1(3,3) := Q_1(3\Delta x, 3\Delta z)$ and $Q_2(3,4)$, respectively. More concretely, given the true value of P as seen above, the corresponding solution $c(x, z, t)$ of the direct problem (2.1.1) is derived from which two time series $c^{obs}(Q_1, t^n) = c(Q_1, t^n)$ and $c^{obs}(Q_2, t^n) = c(Q_2, t^n)$ can be extracted to serve as observation data set. This can be regarded as a set of observation data without noise, called observation data I. Additionally, if noise is present, then another set of observation data, called observation data II, can be also obtained in the following manner:

$$c(3, I, t^n)_{noise}^{obs} = c(3, I, t^n) + \sigma * rand(1, n), \quad I = 3, 4 \tag{3.2}$$

where σ is referred to as the noise level.

We first present a verification of the convergence of the current algorithm using the observation data II, where $\sigma = 0.02$. The initial parameter guesses are set as $(1,1)$, $(10,1)$, $(30,10)$, $(20,3)$ and $(30,1)$, respectively. As seen from Figure 3, the parameter values will tend to evolve towards the ideal initial value $(50, 0.38)$ with minimization iterations, which shows that the algorithm is very robust with respect to variations in the initial parameter guesses, although the initial guesses for the parameters are far from the exact parameters. Next we will illustrate the role of regularization term in improving the accuracy of the retrieved result.

The initial guesses for the unknown parameters are taken as $(1,1)$, i.e. $P^{guess} = (1,1)$. In this situation, we consider four groups of experiments. i), using observation data I without regularization term ($\sigma = 0, \gamma = 0$); ii), using observation data I with regularization term ($\sigma = 0, \gamma \neq 0$); iii), using observation data II without regularization term ($\sigma \neq 0, \gamma = 0$); iv), using observation data II with regularization term ($\sigma \neq 0, \gamma \neq 0$). Table 2 presents the results obtained for the estimated parameters, wherein the descent process of the cost function with respect to the number of minimization iterations also provided, see Figure 4. In addition, the retrieval process of two parameters K_1 and $(\mu_*)_0$ with iterations can be shown in
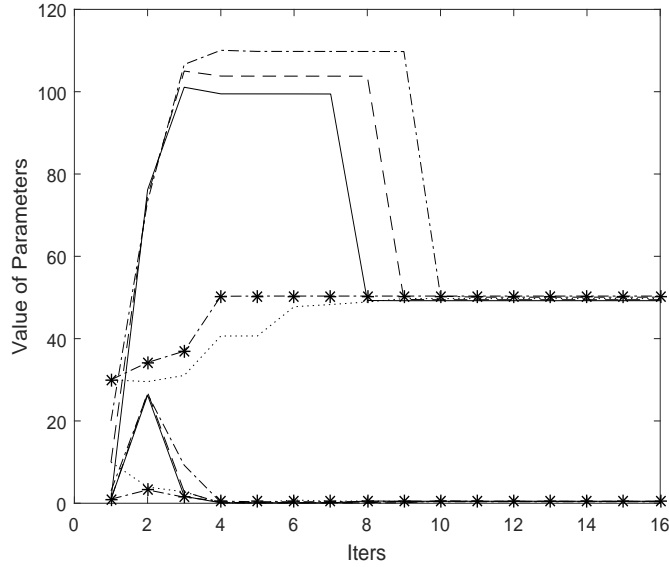
FIGURE 3. Evolution of the value of parameters with iterations for the initial guess values $(1,1)$(solid line -), $(10,1)$(dashed line - -), $(30,10)$(dotted line :), $(20,3)$(dotted-dashed line -.) and $(30,1)$(star dotted-dashed line *-.), respectively.

Figure 5. The stopping criterion, whichever occurs first, is set as

$$\frac{\left|(\mu_*)_0^{(k)} - (\mu_*)_0\right|}{|(\mu_*)_0|} < \varepsilon_1 \ or \ J^{(k)} < \varepsilon_2$$

We see that in the case of initial guess value $(1,1)$ the expected results are obtained by including the regularization term. In particular, when dealing with noisy data [11], employing the regularization technique will help us get the desired result, otherwise the recovery of the parameters may not be as good as it should be, see Figure 6, even when performing a larger number of iterations (not limited to 15 iterations as in the present test). Generally in practical application, the related parameters cannot be known in advance and no information is provided. With this objective in mind, another set of initial guess value of parameters are also taken into account, i.e., $(20, 20)$. We see that there is still a significant difference between the guess value and the true parameter $(50, 0.38)$. The retrieval results are presented in Table 2, which illustrates that the retrieval results match well the true value, as for the descent process of the cost function versus the minimization iterations, a similar behavior to the case of initial guess value $(1, 1)$ can be found (not shown here). This indicates that the current method has good stability. The successful model tests demonstrate that parameter estimation with variational adjoint data assimilation technique is feasible due to the addition of the regularization term allowing the identification of poorly known model parameters using the simulated observational data originating only from two spatial locations at different time instants.

TABLE 2. Results obtained with several sets of experiments.

Guessed value	Error level σ	Iters	J	Estimated Value	Regularization.Para.γ
(1,1)	0.00	12	2.339E-7	(49.4799,0.3974)	0
(1,1)	0.00	13	1.225E-6	(49.5557,0.3734)	3.5E-10
(1,1)	0.02	15	3.020E-6	(50.4796,0.2683)	0
(1,1)	0.02	14	4.900E-6	(50.2095,0.3864)	3.5E-10
(20,20)	0.02	5	4.1237E-6	(50.4308,0.3902)	3.0E-10
(20,20)	0.02	5	3.1178E-6	(50.0751,0.4118)	0
(20,20)	0.02	50	2.9000E-6	(49.1653,0.4669)	0

For the computational complexity of the current numerical algorithm of variational adjoint method, the most costly portion is related to the calculation of objective function (2.1.7) and its gradient (2.2.8-9) in each minimization iteration, while the key point requires one forward model run and additionally only a single integration of the adjoint dynamical model backwards in time to obtain the required gradient. Solving forward model (two sweeps in x and z direction, respectively) is measured by the computational operation count (flops), i.e., $O(nM^2)$, where n stands for the number of time levels, and M is the number of spatial grids in x and z direction, respectively. And we also note in current study that the integration of the adjoint model backwards in time can be equivalent to several forward model simulations and thus remains computationally expensive. This is mainly due to the forcing term in the adjoint model originating from the model-data differences, and therefore one needs to store the whole model trajectory to run the adjoint model. For the objective function, its operation count is approximated by $O(n)$ without taking into account the regularization term, and $O(nM^2)$ when the regularization term is included. When it comes to the gradients of the objective function with respect to two parameters K_1 and $(\mu_*)_0$, their operation counts are then $O(nM^2)$, respectively. The computation becomes expensive particularly when spatial mesh resolution increases. This may not be an active factor for the current method. However, with the fast development of the reduced-order techniques [3, 26], it is also possible to use this method in the framework of inverse problem of reduced-order model. This issue will be addressed in follow-up research efforts.

4. **Conclusions.** Motivated by the work in [25] of parameter estimation, we have developed an alternative approach in the framework of variational adjoint data assimilation. Here when there is no priori information about the uncertain parameters to be relied on, we have successfully combined the variational adjoint data assimilation with a regularization technique to construct a scheme that is capable of recovering near-perfect parameter values, therefore improving our ability to predict future pollutant transport. To date the method has only been developed and tested in the current 2D convection-diffusion model that has two uncertain parameters to be determined, but the results of this study are extremely positive, which may be viewed as a first step towards solving practical optimal control problem
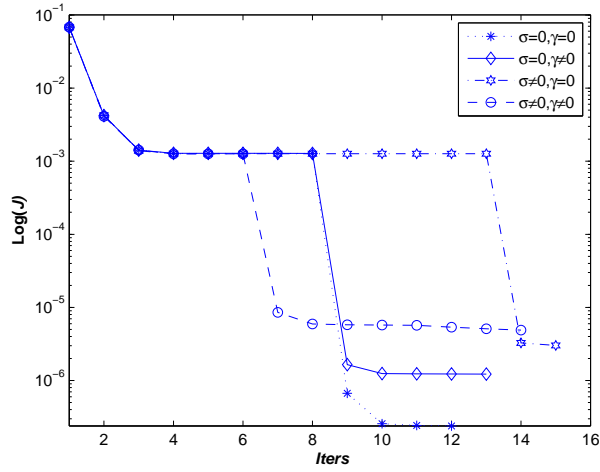
FIGURE 4. The evolution of the cost functional as a function of the number of minimization iterations for the case of initial guess value $(1, 1)$.
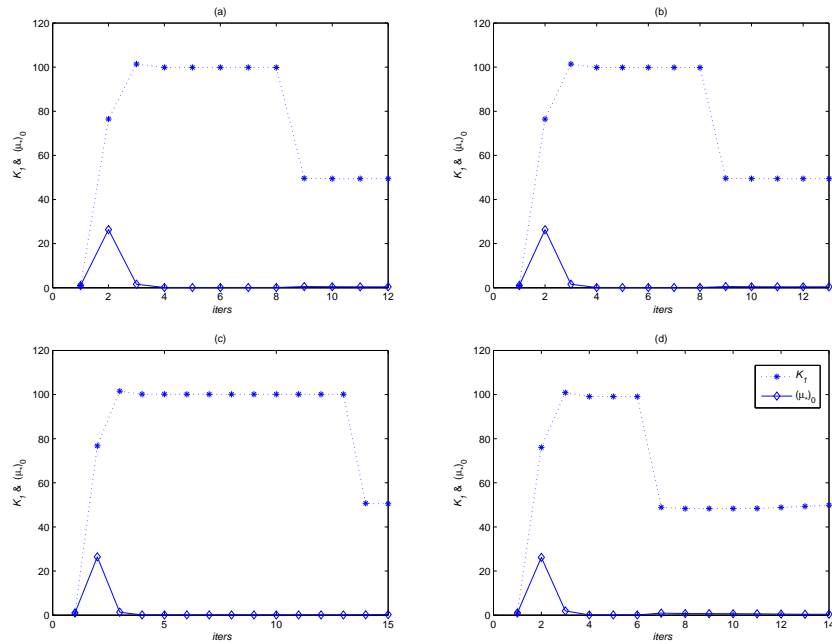

FIGURE 5. Retrieval process of K_1 and $(\mu_*)_0$ with iteration in different cases: (a). $\sigma = 0, \gamma = 0$; (b). $\sigma = 0, \gamma \neq 0$; (c). $\sigma \neq 0, \gamma = 0$; (d). $\sigma \neq 0, \gamma \neq 0$.

by using more realistic observation data and models in order to assess the practical utility of the method. In addition, particular attention should also be paid to the regularization technique. The combination of the Tikhonov regularization
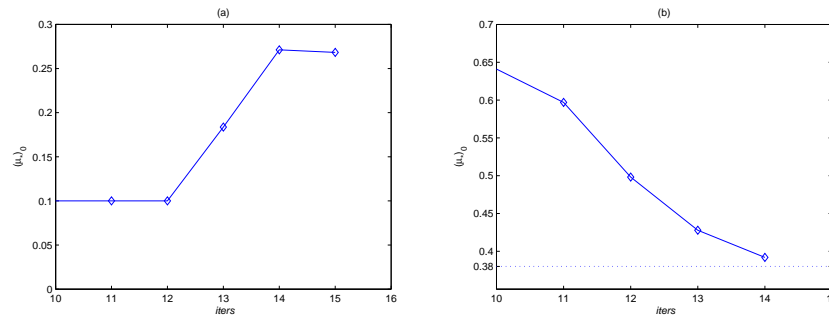
FIGURE 6. Change of $(\mu_*)_0$ magnified partially: (a). in Fig.5 (c) without regularization; (b). in Fig.5 (d) with regularization.

with the discrepancy principle can indeed overcome to some degree the difficulties arising frequently due to the strong nonlinearity and the ill-posedness which result in the parameters failing to converge to a unique value, yielding instead various combinations that result in similar model behavior (i.e. lack of identifiability). In our numerical experiments, when a typical minimization is encountered, which is characterized by a fast decrease in the cost function during the first few tens of iterations, followed by a slow decrease (almost flat) behavior, even if the solution keeps improving during this slow convergence period, practical considerations should prompt us to stop the minimization. As for the Tikhonov regularization, despite its success when equipped with an appropriate regularization parameter, it is yet a challenge how to determine this regularization parameter except when a clear corner of the L shape function is observable for L-curve method. As we all know, this is one of the most active fields of inverse problem research[10, 21, 13]. The regularization parameter in this paper was determined after a large number of trials and errors[6, 28], while some more advanced methods [28, 5] are not involved in current study. This topic will constitute the subject of future research.

REFERENCES

[1] Q. Chai, R. Loxton, K. L. Teo and C. Yang, A unified parameter identification method for nonlinear time delay systems, *Journal of Industrial and Management Optimization*, **9** (2013), 471–486.

[2] T. Chen and C. Xu, Computational optimal control of the Saint-Venant PDE model using the time-scaling technique, *Asia-Pacific Journal of Chemical Engineering*, **11** (2016), 70–80.

[3] J. Du, I. M. Navon, J. Zhu, F. Fang and A. K. Alekseev, Reduced order modeling based on POD of a parabolized Navier–Stokes equations model II: trust region POD 4DVAR data assimilation, *Comput. Math. Appl.*, **65** (2013), 380–394.

[4] D. M. Dunlavy, T. G. Kolda and E. Acar, *Poblano v1.0: A Matlab Toolbox for Gradient-Based Optimization,* SANDIA Report, 2010.

[5] H. W. Engl, Discrepancy principles for Tikhonov regularization of ill-posed problems leading to optimal convergence rates, *J. Optim. Theory Appl.*, **52** (1987), 209–215.

[6] C. Farhat, R. Tezaur and R. Djellouli, On the solution of three-dimensional inverse obstacle acoustic scattering problems by a regularized Newton method, *Inverse Problems*, **18** (2002), 1229–1246.

[7] R. Giering, Tangent linear and adjoint biogeochemical models, in *Inverse Methods in Global Biogeochemical Cycles* (eds. P. Kasibhatla, M. Heimann, P. Rayner, N. Mahowald, R. G. Prinn and D. E. Hartley), American Geophysical Union: Washington DC, 2000, 33–47.

[8] P. E. Gill, W. Murray and M. H. Wright, *Practical Optimization*, Academic Press, 1981.

[9] M. Gunzburger, Adjoint equation-based methods for control problems in incompressible, viscous flows, *Flow Turbulence Combust.*, **65** (2000), 249–272.

[10] P. C. Hansen, *Rank-Deficient and Discrete Ill-posed Problems, Practical optimization*, SIAM, 1998.

[11] M. J. Hossen, I. M. Navon and D. N. Daescu, Effect of random perturbations on adaptive observation techniques, *International Journal For Numerical Methods in Fluids*, **69** (2012), 110–123.

[12] S. X. Huang, W. Han and R. S. Wu, Theoretical analysis and numerical experiments of variational assimilation for one-dimensional ocean temperature model with techniques in inverse problems, *Science in China D*, **47** (2004), 630–638.

[13] D. Krawczyk-stando and M. Rudnicki, Regularization parameter selection in discrete ill-posed problems–The use of the U-curve, *Int. J. Appl. Math. Comput. Sci.*, **17** (2007), 157–164.

[14] Q. Lin, R. Loxton, C. Xu and K. L. Teo, Parameter estimation for nonlinear time-delay systems with noisy output measurements, *Automatica*, **60** (2015), 48–56.

[15] W. J. Liu, G. Li and L. H. Hong, General decay and blow-up of solutions for a system of viscoelastic equations of Kirchhoff type with strong damping, *Journal of Function Spaces*, **2** (2014), article ID 284809, 21pp.

[16] R. Loxton, K. L. Teo and V. Rehbock, An optimization approach to state-delay identification, *IEEE Transactions on Automatic Control*, **55** (2010), 2113–2119.

[17] B. Malengier and R. V. Keer, Parameter estimation in convection dominated nonlinear convection-diffusion problems by the relaxation method and the adjoint equation, *Journal of Computational and Applied Mathematics*, **215** (2008), 477–483.

[18] I. M. Navon, Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography, *Dynam. Atmos. Oceans*, **27** (1998), 55–79.

[19] I. M. Navon and D. M. Legler, Conjugate-gradient methods for large-scale minimization in meteorology, *Monthly Weather Review*, **115** (1987), 1479–1502.

[20] I. M. Navon, X. Zou, J. Derber and J. Sela, Variational data assimilation with an adiabatic version of the nmc spectral model, *Monthly Weather Review*, **120** (1992), 1433–1446.

[21] T. Nieminen, J. Kangas and L. Kettunen, Use of Tikhonov regularization to improve the accuracy of position estimates in inertial navigation, *International Journal of Navigation and Observation*, **2011** (2011), 1–10.

[22] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, 1999.

[23] D. W. Peaceman and H. H. Rachford Jr., The numerical solution of parabolic and elliptic differential equations, *J. Soc. Ind. Appl. Math.*, **3** (1955), 28–41.

[24] B. Protas, T. R. Bewley and G. Hagen, A computational framework for the regularization of adjoint analysis in multiscale PDE systems, *Journal of Computational Physics*, **195** (2004), 49–89.

[25] R. B. Storch, L. C. G. Pimentel and H. R. B. Orlande, Identification of atmospheric boundary layer parameters by inverse problem, *Atmospheric Environment*, **41** (2007), 1417–1425.

[26] Y. P. Wang, I. M. Navon, X. Y. Wang and Y. Cheng, 2D Burgers equation with large Reynolds number using POD/DEIM and calibration, *International Journal For Numerical Methods in Fluids*, **82** (2016), 909–931.

[27] Y. P. Wang and S. L. Tao, Application of regularization technique to variational adjoint method: A case for nonlinear convection–diffusion problem, *Applied Mathematics and Computation*, **218** (2011), 4475–4482.

[28] Y. W. Wen and A. M. Yip, Adaptive Parameter Selection for Total Variation Image Deconvolution, *Numer. Math. Theor. Meth. Appl.*, **2** (2009), 427–438.

[29] J. C.-F. Wong and J.-L. Xie, Inverse determination of a heat source from natural convection in a porous cavity, *Computers and Fluids*, **52** (2011), 1–14.

[30] Y. Zhu and I. M. Navon, Impact of parameter estimation on the performance of the FSU global spectral model using its full-physics adjoint, *Monthly Weather Review*, **127** (1999), 1497–1517.

[31] X. Zou, I. M. Navon and F. X. LeDimet, An optimal nudging data assimilation scheme using parameter estimation, *Quarterly Journal of the Royal Meteorological Society*, **118** (1992), 1163–1186.

E-mail address: eduwyp@163.com
E-mail address: chengyue_10@126.com
E-mail address: inavon@fsu.edu
E-mail address: guanyh@nuist.edu.cn