Continuous versus Discrete Advection Adjoint
in Chemical Data Assimilation with CMAQ

Tianyi Gou, Adrian Sandu

Abstract

Data assimilation obtains improved estimates of the state of a physical system by combining imperfect model results with sparse and noisy observations of reality. In the four dimensional variational (4D-Var) framework data assimilation is formulated as an optimization problem, which is solved using gradient based optimization methods. The 4D-Var gradient is obtained by forcing the adjoint model with observation increments. The construction of the adjoint model requires considerable development effort. In the continuous approach the adjoint differential equations are discretized. In the discrete approach the numerical solution of the forward equations is differentiated. The two routes lead to different gradients.

In this paper we investigate numerically the effect of using discrete and continuous adjoints of the advection equation in chemical transport modeling. Continuous advection adjoints are easily implemented by calling the same advection subroutines as the forward model, with a sign change for the winds, and with rescaling the solution. Discrete advection adjoints involve the differentiation of a nonlinear, monotonic advection scheme like the piecewise parabolic method. The numerical experiments are carried out with CMAQ-ADJ. The results show that, while discrete advection adjoints are more accurate in point to point comparisons against finite differences, the continuous adjoints of advection perform better as gradients for optimization in 4D-Var data assimilation.

Keywords: Chemical data assimilation, continuous adjoint, discrete
Contents

1 Introduction ................................................. 3

2 The CMAQ-ADJ Chemical Transport Model ......................... 6

3 Continuous and Discrete Advection Adjoints ....................... 7

4 Validation of Advection Adjoints ................................ 9
   4.1 One-dimensional sensitivity studies .......................... 9
   4.2 Three dimensional sensitivity studies ......................... 10
       4.2.1 Validation of vertical advection adjoints ............... 11
       4.2.2 Validation of horizontal advection adjoints .......... 11
   4.3 Nonlinearity of PPM ...................................... 12

5 4D-Var Data Assimilation ...................................... 15
   5.1 One-dimensional advection 4D-Var in a twin experiment setting ......................................................... 16
   5.2 CMAQ-ADJ 4D-Var in a twin experiment setting ............. 19
   5.3 CMAQ-ADJ 4D-Var with real observation data ............... 22

6 Conclusions .................................................. 23
1. Introduction

The chemical composition of the atmosphere has important implications for urban and regional air quality, for human health, and for climate change. Complex atmospheric chemical transport models (CTMs) are designed to describe the fate and transport of atmospheric chemical constituents associated with the gas and aerosol phases. A close integration of these models with observations is essential for advanced atmospheric chemistry analyzes.

Chemical data assimilation optimally uses the observations (reality within errors) along with the model (imperfect representation of the processes and their connections) to produce a better estimate (in some optimal sense) of the chemical state of the atmosphere and of key parameters such as emissions and boundary conditions. Four dimensional variational data assimilation (4D-Var) provides an optimal control approach to integrating spatial and temporal distributed observations into models. In 4D-Var the data assimilation problem is formulated as an optimization problem, where the maximum likelihood estimates of uncertain parameters minimize the misfit between model predictions and observations, as well as the misfit between model predictions and a background estimate.

To solve the 4D-Var optimization problem one typically employs gradient based numerical procedures. In general, two methodologies can be used to obtain the required gradients. One is the direct method where each component of the gradient is calculated by performing a (tangent linear) model run; the direct method is not efficient when the number of model parameters is large. The second approach to obtaining the gradient is the adjoint method, which requires one forward model run followed by one backward adjoint model run. The adjoint method is effective when gradients are needed with respect to a large number of model parameters, e.g., when data assimilation seeks an optimal initial condition.

A central issue in 4D-Var data assimilation is the construction of the adjoint model. Two approaches are possible. In the continuous adjoint approach one first differentiates the underlying physical equations in an abstract setting, and then discretizes the resulting adjoint partial differential equations (PDEs). In the discrete adjoint approach one first discretizes the physical equations (using suitable numerical methods), then differentiates the discrete algorithm. Discrete adjoints are popular since they can be obtained by automatic differentiation [7]. In general the two ap-
proaches lead to different adjoint models.

The choice of the adjoint model can significantly impact the performance of the data assimilation system. Previous work in numerical weather prediction has pointed out that discrete adjoints of complex models, while fully consistent with the tangent linear model, are noisy and do not guarantee good performance for sensitivity and data assimilation studies [6, 10, 13, 19, 20, 22].

Advection describes the long range transport of pollutants under the wind fields and is a major component of chemical transport models. The most widely used numerical discretizations of the advection equation are monotonic (thus, preventing the formation of spurious wiggles), positive definite (thus, avoiding non-physical negative concentrations), and of order two or higher. Numerical discretizations with these properties are inherently nonlinear [8], even if advection itself is a linear process. Because of this nonlinearity the construction of adjoint advection models is challenging, and the continuous and the discrete adjoint formulations lead to different results. Moreover, many numerical advection schemes are not continuously differentiable with respect to model state or model parameters [18], and careful scrutiny needs to be applied to the way the sensitivities are defined.

Liu and Sandu [11] have studied the consistency of discrete adjoints with the adjoint advection PDE. They have concluded that inconsistencies arise wherever the forward method changes the computational pattern. Changes in the computational pattern include changes in upwinding direction, active limiting of the fluxes, active limiting of the slopes, as well as numerical boundary conditions (switching to a different discretization formula near a boundary). Thuburn and Haine [18] have shown that any advection scheme that satisfies a scaling property and has a continuous Jacobian is linear. Therefore, high order monotonic schemes with scaling property necessarily have a discontinuous Jacobian, and therefore non-trivial issues arise in the definition and construction of adjoint models.

Vukicevic and Steyskal [21] have performed 4D-Var experiments with a 2D advection system with both linear schemes (QUICK and LEAPFROG) and nonlinear schemes (MPDATA). The L-BFGS method was used for optimization, and different combinations of the forward and adjoint schemes were tested. The experiments have shown that when the discrete adjoint of the nonlinear MPDATA is used to provide the gradients, the optimization is likely to get trapped into local minima; the cost function values
remain relatively large, and different optimal points are obtained for dif-
ferent initial guesses. Much better results are obtained when the gra-
dients are provided by the continuous adjoint. The continuous adjoint
is free of noise and provides a coarse grain approximation of the gradi-
ent; the continuous approach provides an overall better approximation of
the continuous optimization problem, and leads to better optimal results.
Therefore, the authors recommend the use of continuous adjoints for data
assimilation problems involving advection.

In this paper we investigate numerically the implications that the type
of advection adjoint has on the accuracy of derivatives (compared against
finite differences) and on the solution of the 4D-Var data assimilation in
the context of chemistry and transport modeling. The atmospheric CTM
used in this study is the Community Multiscale Air Quality (CMAQ)
modeling system [1, 2, 4]. CMAQ is a powerful third generation air
quality modeling tool which is capable of modeling important air qual-
ity issues such as tropospheric ozone, acid deposition, fine particles and
visibility degradation and is used to predict the state of the airborne pol-
lutants [1, 2, 9]. CMAQ-ADJ (http://people.cs.vt.edu/~asandu/
Software/CMAQ_ADJ), the adjoint model of the CMAQ, has been de-
veloped through a collaboration between Virginia Tech, University of
Houston, and Caltech [9, 16, 17]. CMAQ-ADJ has the added capa-
bilities of performing sensitivity analysis, and of carrying out 4D-Var data
assimilation. The current implementation of CMAQ-ADJ considers the
advection, diffusion, emission, deposition, and gas-phase processes. The
continuous adjoint approach has been used for advection processes, while
the discrete adjoint approach has been used for other processes including
chemistry process and diffusion process.

In this paper, we report on the construction and properties of discrete
advection adjoints for CMAQ. The CMAQ advection is implemented us-
ing the the monotonic piecewise parabolic method (PPM)[3], and the cor-
responding discrete adjoint is obtained via automatic differentiation. We
investigate numerically the impact of the discrete and continuous advec-
tion adjoints on the performance of sensitivity analysis and data assimila-
tion. We conclude that, while the discrete adjoints provide more accurate
sensitivities (when compared against finite differences), the continuous
adjoints allow for better optimization solutions in 4D-Var data assimila-
tion. To the best of our knowledge this is the first study to assess the
impact of different types of advection adjoints in the context of chemistry
and transport modeling.

This paper is organized as follows. Section 2 introduces the mathematical formulation of chemical transport modeling, the CMAQ model, and its adjoint CMAQ-ADJ. Section 4 validates the implementation of discrete advection adjoints in one- and three-dimensional tests, and explains the differences between continuous and discrete sensitivities by the nonlinearity of the forward advection solver. Section 5 shows the results of several 4D-Var data assimilation tests with synthetic and with real observation data, and using both continuous and discrete advection adjoint gradients in the numerical optimization. Section 6 summarizes the conclusions of this work.

2. The CMAQ-ADJ Chemical Transport Model

Atmospheric chemical transport models (CTMs) are designed to describe the fate and transport of atmospheric chemical constituents associated with the gas and aerosol phases. Atmospheric CTMs describe the physical and chemical processes that govern the evolution of airborne pollutant fields. They solve the following atmospheric mass balance equation [9]:

\[
\frac{\partial c_i}{\partial t} = -u \cdot \nabla c_i + \frac{1}{\rho} \nabla \cdot (\rho K \nabla c_i) + \frac{1}{\rho} f_i(c) + E_i, \quad t \in [0, T],
\]

\[
c_i(t^0, x) = c_i^0(x),
\]

\[
c_i(t, x) = c^\text{IN}_i(t, x) \quad \text{for} \quad x \in \Gamma^\text{IN},
\]

\[
K \frac{\partial c_i}{\partial n} = 0 \quad \text{for} \quad x \in \Gamma^\text{OUT},
\]

\[
K \frac{\partial c_i}{\partial n} = V_i \text{dep}_i c_i - Q_i \quad \text{for} \quad x \in \Gamma^\text{GR}, \quad \text{for all} \quad 1 \leq i \leq s.
\]

We refer to the system (1a)–(1e) as the forward model. To simplify the presentation, in this paper we consider as parameters the initial state \( c^0 \) of the model; it is known that this does not restrict the generality of the formulation. The solution of the forward model \( c = c(t, c^0) \) is uniquely determined once the model parameters \( c^0 \) are specified.

Here \( u \) is the wind field vector, \( K \) the turbulent diffusivity tensor, \( \rho \) the air density in \( \text{moles/cm}^3 \), and \( c_i \) the mole-fraction concentration of chemical species \( i \) \( (1 \leq i \leq s) \). The density of this species is \( \rho c_i \text{ moles/cm}^3 \).
$V_{dep}^i$ is the deposition velocity of species $i$, $Q_i$ the rate of surface emissions, and $E_i$ the rate of elevated emissions for this species. The rate of chemical transformations $f_i$ depends on absolute concentration values; the rate at which mole-fraction concentrations change is then $f_i(\rho c)/\rho$.

In addition to the processes represented in (1a)–(1e), CMAQ also includes cloud, plume-in-grid, and aerosol processes. Details of all the science processes can be found in [1, 2].

CMAQ uses a time split approach to solve the mass balance equations (1a)–(1e). This means that, during each time step, individual science processes are solved for in succession; each science process is discretized with an appropriate family of numerical methods. Consequently, an adjoint model of CMAQ can be constructed from the adjoint models of individual science processes; during a (backward) time step, adjoint science models are solved for in succession, and in reverse order compared to the forward model. The time split approach offers the advantage that adjoints can be constructed independently for each science process. Different adjoint types can be employed for different processes, for example, a discrete adjoint can be used for diffusion while a continuous adjoint is used for advection.

CMAQ-ADJ (http://people.cs.vt.edu/~asandu/Software/CMAQ_ADJ), the adjoint model of the CMAQ, has been developed through a collaboration between Virginia Tech, University of Houston, and Caltech [9, 16, 17]. CMAQ-ADJ has the added capabilities of performing sensitivity analysis and 4D-Var data assimilation. The CMAQ-ADJ v. 5.4.5 used here considers the advection, diffusion, emission, deposition, and gas-phase processes. The continuous adjoint approach has been used for advection processes, while the discrete adjoint approach has been used for other processes including chemistry process and diffusion process.

In the next section we report on a new implementation of the discrete advection adjoint process in CMAQ-ADJ.

### 3. Continuous and Discrete Advection Adjoints

The implementation of advection in CMAQ is done using a dimensionally split approach on the Cartesian grid. The one dimensional advection problems in the $x$, $y$, and $z$ directions are solved in succession during each forward time step. Each one dimensional advection problem is discretized using the piecewise parabolic method (PPM).
Following the directional split approach, a three-dimensional advection adjoint can be constructed by successively solving the adjoint advection in the $z$, $y$, and $x$ directions (respectively) during each reverse time step. Both continuous and discrete approaches can be employed to construct the one-dimensional advection adjoint models. We next discuss each of these approaches.

**Continuous advection adjoint.** Consider the one-dimensional forward advection PDE

$$\frac{\partial (\rho c)}{\partial t} = -\frac{\partial (u \cdot \rho c)}{\partial x}, \quad (2)$$

where $c(t,x)$ is the concentration vector (a function of time and space), $\rho(t,x)$ is the air density, and $u(t,x)$ is the $x$ component of the wind field. Equation (2) corresponds to fixed $y$ and $z$ coordinate values in a dimensional split framework.

The corresponding adjoint PDE [9] is

$$\frac{\partial (\lambda/\rho)}{\partial t} = -\frac{\partial (u \cdot \lambda/\rho)}{\partial x}, \quad (3)$$

where $\lambda$ is the adjoint variable. Note that (3) is itself an advection equation of the form (2), applied to $\lambda/\rho$ (instead of $\rho c$), and solved backwards in time. Formally the equation (3) can be solved forward in time, but with the wind field changed from $u$ to $-u$ to account for the different time direction.

The continuous adjoint is obtained by solving (3) numerically. To be specific, the continuous adjoint is implemented by calling the forward advection subroutine (PPM) with a reversed wind field ($-u$), and with the proper scaling applied to the adjoint variables.

**Discrete advection adjoint.** The discrete advection adjoint is obtained by differentiating the forward discrete advection model. The implementation is aided by the automatic differentiation tool T AMC (Tangent Linear and Adjoint Model Compiler, [7]). T AMC is applied to (one step of) the PPM algorithm to obtain (one step of) the corresponding discrete adjoint. Due to the nonlinearity of the PPM scheme, the discrete adjoint depends not only on the adjoint variable $\lambda$ but also on the forward concentration $c$. During the forward run concentrations are checkpointed into files at the beginning of each dynamic time step. These concentrations are read in at corresponding times during the adjoint step.
Modified discrete advection adjoint. PPM ensures monotonicity properties with the help of slope and curvature limiters. The reverse mode differentiation of these limiters may introduce discontinuities in the discrete adjoint solution [3]. The modified discrete adjoint is constructed as follows. First, the monotonicity characteristics of the PPM are removed by commenting out the steepening procedure (i.e., the limiters). The resulting algorithm is smooth and has no points of non-differentiability. This smooth (but non-monotonic) code is then processed by automatic differentiation to obtain the modified discrete adjoint code. Note that the original PPM algorithm continues to be used for the forward model runs, and that the modified discrete adjoint is not the exact discrete gradient anymore.

4. Validation of Advection Adjoints

In this section we validate the adjoints of vertical and horizontal CMAQ advection against finite difference results. We start with one-dimensional problems and continue with three-dimensional tests.

4.1. One-dimensional sensitivity studies

We first consider a test case based on the one dimensional advection PDE (2) and defined by the following parameters and initial and boundary conditions:

\[ x \in [0, 2\pi], \quad t \in [t_0 = 0, t_F = 1.17], \quad u = 1, \quad \rho = 1, \]
\[ c(x = 0, t) = c(x = 2\pi, t) = 1.0, \quad c_0(x) = c(x, t_0) = 1.0 + \sin(3x). \]

The spatial domain is discretized with \( N = 200 \) equally spaced grid points. The time step corresponds to a Courant Friedrichs Lewy (CFL) number of 0.5.

A cost function (a receptor) is defined as the sum of forward solutions at gridpoints \( x_i, a \leq i \leq b \), at the final time:

\[ \psi = \sum_{i=a}^{b} c(t_F, x_i). \]  

The adjoint advection PDE (3) has periodic boundary conditions, and the initial conditions

\[ \lambda(t_F, x_i) = \frac{d \psi}{dc(t_F, x_i)} = \begin{cases} 1 & \text{for } a \leq i \leq b, \\ 0 & \text{otherwise}. \end{cases} \]
Solutions of the adjoint system (3) at the initial time represent the sensitivities of the cost function (receptor) with respect to the initial conditions (source)

$$\lambda (t_0, x_i) = \frac{d \psi}{dc (t_0, x_i)}.$$ 

The continuous and the discrete advection adjoint subroutines from CMAQ-ADJ are used to solve the one dimensional test. Figure 1 shows the solutions of the continuous and discrete adjoint systems at initial time for the receptor (5) with $a = 60$ ($x_a = 0.3$) and $b = 90$ ($x_b = 0.45$). Both the continuous and the discrete adjoint systems have the same initial condition at the final time. The continuous adjoint solution in Figure 1(a) is monotonic, as expected, since the monotonic PPM scheme has been applied to the advection equation (3). Both the discrete adjoint solution (Figure 1(b)) and the modified discrete adjoint solution (Figure 1(c)), on the other hand, show spurious wiggles near the large adjoint solution gradients. This is an artifact of the nonlinearity of the PPM scheme; the wiggles come solely from differentiating the numerical solution and have no physical sensitivity interpretation.

4.2. Three dimensional sensitivity studies

We now test the advection adjoints in the full three-dimensional CMAQ-ADJ code, on a grid of $NCOLS \times NROWS \times NLEVS$ points. The tests are essentially one dimensional due to the directional split approach taken in CMAQ. During the tests all physical processes, except one-dimensional advection, are turned off. For a quantitative adjoint validation we compute the $R^2$ and RMS correlation factors between the same sensitivity coefficients computed by the adjoint and computed via finite differences. The $R^2$ and RMS correlation factors of two series $X$ and $Y$ of length $n$ are defined as:

$$R^2(X, Y) = \frac{\left( n \sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i \right)^2}{\left( n \sum_{i=1}^{n} X_i^2 - \left( \sum_{i=1}^{n} X_i \right)^2 \right) \left( n \sum_{i=1}^{n} Y_i^2 - \left( \sum_{i=1}^{n} Y_i \right)^2 \right)},$$

$$RMS(X, Y) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - Y_i)^2}.$$
4.2.1. Validation of vertical advection adjoints

The vertical adjoint validation is done pointwise against central finite difference sensitivities, as follows. We choose a source layer \(L_s\), a receptor layer \(L_r\), and a species \((S = O_3, \text{ozone})\). Sensitivities of the receptor concentration with respect to the source concentration, \(dc(L_r)/dc(L_s)\), are computed using both central finite difference approach and the adjoint sensitivity approach. We choose a challenging test where both the source \((L_s = 3)\) and the receptor \((L_r = 2)\) are located near the ground level.

Since only vertical advection is considered, each model column evolves independently; as a result, we obtain a set of \(NCOLS \times NROWS\) independent sensitivities coefficients. Each sensitivity coefficient is represented by a point in a scatterplot, where the \(x\) coordinate is the sensitivity value computed by the adjoint model, and the \(y\) coordinate is the value of the same coefficient computed by central finite differences.

The CMAQ-ADJ numerical simulations are run for 20 hours, between 0 GMT and 20 GMT on the 183\(^{\text{rd}}\) day of 1999. Figure 2(a) shows the scattered plot for continuous adjoint sensitivity, and Figure 2(b) the scattered plot for discrete adjoint sensitivity. Table 1 presents the \(R^2\) and \(RMS\) coefficients that describe the correlation between adjoint and finite difference sensitivities.

Clearly the discrete adjoint provides a closer match to the finite difference values than does the continuous adjoint. This conclusion is supported by both the scattered plots of Figure 2 and by the \(R^2\) and \(RMS\) data of Table 1. The good correlation between discrete adjoint and finite difference values constitute a validation of the new discrete advection adjoint code added to CMAQ-ADJ.

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>(R^2(X, Y))</th>
<th>(RMS(X, Y))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous adjoint</td>
<td>Finite differences</td>
<td>0.43</td>
<td>0.49</td>
</tr>
<tr>
<td>Discrete adjoint</td>
<td>Finite differences</td>
<td>0.99</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Table 1: \(R^2\) and \(RMS\) values for the vertical advection sensitivities.

4.2.2. Validation of horizontal advection adjoints

Because \(x\) advection and \(y\) advection are equivalent (they are implemented in CMAQ by calling the same horizontal PPM routine, and provide similar test results), we only report validation results for the \(x\) ad-
vection in this section. Since all the physical processes, except \( x \) advection, are turned off, each row of gridpoints evolves independently and there are \( \text{NROWS} \times \text{NLAYS} \) independent advection simulations. Point-to-point sensitivities are generated for the species ozone as follows. For each row and vertical layer the source (of ozone perturbations at initial time) is located in column \( C_s \); the receptor measures the ozone concentration in column \( C_r \) at the final time. This test generates \( \text{NROWS} \times \text{NLAYS} \) pairs of sensitivities (finite difference sensitivity and adjoint sensitivity), one for each model row.

The numerical tests are carried out for a simulation time length of two hours, between 0 GMT and 20 GMT on the 183\textsuperscript{rd} day of 1999. For each row and vertical layer the source and the receptor are located at \( C_s = 25 \) and \( C_r = 26 \), respectively. Figure 3 presents the scatter plots of adjoint versus finite difference sensitivities. Figure 3(a) presents the scatter plots of adjoint versus finite difference sensitivities, and Figure 3(b) the scatter plot for the discrete adjoint sensitivity. \( R^2 \) and RMS coefficients that describe the correlation between adjoint and finite difference sensitivities are reported in Table 2. A comparison of the plots shows that the discrete adjoint results and modified adjoint results are in better agreement with the finite difference results than the continuous adjoint values. This good agreement validates the new discrete adjoint subroutines for horizontal advection added to CMAQ-ADJ.

\[
\begin{array}{|c|c|c|c|}
\hline
\text{X} & \text{Y} & R^2(X,Y) & \text{RMS}(X,Y) \\
\hline
\text{Continuous adjoint} & \text{Finite differences} & 0.77 & 0.17 \\
\hline
\text{Discrete adjoint} & \text{Finite differences} & 0.97 & 0.05 \\
\hline
\text{Modified discrete adjoint} & \text{Finite differences} & 0.86 & 0.11 \\
\hline
\end{array}
\]

Table 2: \( R^2 \) and RMS values for the horizontal advection sensitivities.

4.3. Nonlinearity of PPM

From previous sections, we found that discrete adjoint sensitivities agree better with finite difference sensitivities than their continuous adjoint counterparts. One of the reasons is that both discrete adjoint and finite difference approaches compute the sensitivities of the numerical solutions (with respect to changes in model inputs), while the continuous adjoint gives approximations of the sensitivities of the continuous solutions. These two sensitivities are different due to the nonlinearity of the
PPM scheme. In the following discussion, we explain in detail the link between the PPM nonlinearity and the discrepancy between continuous adjoint and finite difference sensitivities.

We consider a base problem defined by the advection equation (2) with the following parameters and initial and boundary conditions:

\[
x \in [0, 2\pi], \quad t \in [t_0 = 0, t_F = 2.36], \quad u = 1, \quad \rho = 1, \quad c(x = 0, t) = c(x = 2\pi, t) = 1.0, \quad c_0(x) = c(x, t_0) = 1.0 + \sin(x).
\]

The spatial domain is discretized with \( N = 200 \) equally spaced grid points. The time step corresponds to a Courant Friedrichs Lewy (CFL) number of 0.5. The exact solution of the base problem (6) is the base solution \( c(x, t) \). A base numerical solution is computed using the PPM scheme. Formally, the PPM process maps the numerical initial conditions to the numerical solution at the final time, \( c_F = \text{PPM}(c_0) \).

Next, we define a perturbed problem as follows. We modify the initial condition of (6) by adding a small Dirac delta perturbation of size \( 10^{-3} \) at \( x = 0 \) (corresponding to the first grid point). The discrete perturbed initial conditions are

\[
\hat{c}_0(x_i) = c_0(x_i) + \delta c_0(x_i), \quad i = 1, \ldots, N, \\
\delta c(x_1, t_0) = 10^{-3}; \quad \delta c_0(x_i) = 0, \quad i = 2, \ldots, N.
\]

The corresponding problem gives the perturbed solution \( \hat{c}(x, t) \). The numerical approximation of the perturbed solution is obtained by solving (2), (7) with the PPM scheme; at the final time the perturbed solution is \( \hat{c}_F = \text{PPM}(\hat{c}_0) \).

The difference between the perturbed and the base numerical solutions at the final time, \( \hat{c}_F - c_F \), is shown in Figure 4(a). This difference is wiggly, has both positive and negative values, and its support is dispersed over many grid points.

Since the advection PDE (2) is linear, the solution difference \( \delta c(x, t) = \hat{c}(x, t) - c(x, t) \) evolves according to (2) as well, starting from the initial condition \( \delta c(x, t_0) \). The delta perturbation is advected in time under the constant wind field, and the exact solution \( \delta c(x, t_F) \) is a delta function located at \( x = 2.36 \) (corresponding to grid point number 75). The perturbation propagation equation is the advection equation, and the initial condition is the perturbation value at \( t_0 \). A numerical solution for the perturbation propagated to the final time is obtained with the PPM scheme,
\[ \delta c_F = PPM(\delta c_0). \] The perturbation distribution at the final time is shown in Figure 4(b). As expected, the initial delta profile is diffused by the nonlinear PPM scheme, but the solution remains positive and wiggle-free.

If the advection equation was solved exactly then \[ \delta c(x, t_F) = \tilde{c}(x, t_F) - c(x, t_F). \] However, due to the nonlinearity of the PPM scheme we have that

\[ \delta c_F = PPM(\delta c_0) = PPM(\tilde{c}_0 - c_0) \neq PPM(\tilde{c}_0) - PPM(c_0). \]

The advected delta perturbation (Figure 4(b)) corresponds to a continuous (forward) sensitivity approach where the numerical scheme is applied to solve the sensitivity equations. The difference between the perturbed and the base profiles (4(a)) corresponds to a finite difference approach to computing forward sensitivities. This finite difference approach approximates the derivatives of the PPM numerical solution, i.e., it approximates the discrete sensitivities.

Numerical results indicate that the continuous and the finite difference (discrete) sensitivities are very different. This is not an issue of the correctness of the implementation. Rather, it is a fundamental, algorithmic issue, due to the considerable nonlinearity of the PPM scheme.

Similar considerations hold for adjoint sensitivities: the nonlinearity of the PPM scheme leads to differences between the continuous and the discrete adjoints.

To illustrate this, consider the test problem (6) together with the (receptor) cost function (5). The source of perturbations at the initial time is located in the first grid cell \( x_1 = 0 \). We compute the sensitivity of the cost function with respect to perturbations in the initial condition, \( \partial \psi / \partial c(x_1, t_0) \). Recall that, if the test problem is solved exactly, the initial delta perturbation is advected to grid point 75. Therefore, the exact sensitivity value is

\[ \frac{\partial \Psi}{\partial c(x_1, t_0)} \bigg|_{\text{exact}} = \begin{cases} 1 & \text{if } a \leq 75 \leq b, \\ 0 & \text{otherwise}. \end{cases} \]

Two approaches are used to get \( \partial \psi / \partial c(x_1, 0) \). One is the finite difference approach (with an initial perturbation \( \delta c(x_1, 0) = 10^{-3} \)), and the other is the continuous adjoint approach. Table 3 shows the results for different choices of the receptor location \([a, b]\). The relative difference is obtained
by taking the difference between the two sensitivities, and dividing it by the exact value (one).

The results in Table 3 show that the agreement between the continuous adjoint and finite difference sensitivities depends strongly on the choice of the receptor. If receptors are chosen at grids where wiggles are present (e.g., \(a = 81\) and \(b = 86\)) the finite difference sensitivity is dominated by the numerical effects (the derivatives of the numerical solution), which leads to a large disagreement with continuous adjoint sensitivity. In particular, if a pointwise receptor is located on a wiggle, the finite difference result and the linearized (adjoint) result can be arbitrarily different; for example, the finite difference can give a negative result while the adjoint sensitivity is positive. A good agreement is observed when the receptors include the real change in the solution profile, and they leave out the spurious wiggles resulting from the nonlinearity of the scheme.

One practical way to alleviate the disagreement is to choose a large receptor area (e.g., \(a = 71\) and \(b = 97\)). In this case one "integrates" over the wiggles and cancels out their spurious contribution, while capturing the real change in the concentration profile (and therefore, in the cost function).

### Table 3: The sensitivity of the receptor cost function with respect to perturbations in the initial condition at the first gridpoint. Two computational approaches are compared: finite differences and continuous adjoint gradient.

<table>
<thead>
<tr>
<th>Receptor</th>
<th>Finite difference</th>
<th>Continuous adjoint</th>
<th>Relative difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a = 0) (b = 70)</td>
<td>0.00</td>
<td>2.68E-08</td>
<td>(\sim 0%)</td>
</tr>
<tr>
<td>(a = 71) (b = 97)</td>
<td>0.999</td>
<td>0.999</td>
<td>(\sim 0%)</td>
</tr>
<tr>
<td>(a = 98) (b = 202)</td>
<td>0.00</td>
<td>4.74E-25</td>
<td>(\sim 0%)</td>
</tr>
<tr>
<td>(a = 75) (b = 80)</td>
<td>1.01</td>
<td>0.956</td>
<td>-5.59 %</td>
</tr>
<tr>
<td>(a = 81) (b = 86)</td>
<td>-0.127</td>
<td>1.23E-08</td>
<td>12.7 %</td>
</tr>
<tr>
<td>(a = 75) (b = 85)</td>
<td>0.943</td>
<td>0.958</td>
<td>1.59 %</td>
</tr>
</tbody>
</table>

---

5. **4D-Var Data Assimilation**

Data assimilation is the process of fusing information from model predictions and observations in order to obtain better initial conditions, boundary conditions, or emission estimates. 4D-Var data assimilation optimally combines three sources of information: an a priori estimate of
the state of the atmosphere; knowledge about the physical and chemical processes that govern the evolution of pollutant fields as captured in the model; and observations of some of the state variables [15].

We apply 4D-Var data assimilation with CMAQ-ADJ to provide optimal estimates of the ozone initial conditions. 4D-Var data assimilation is posed as an optimization problem where the best estimate of initial conditions minimizes the following cost function:

$$J(c_0) = \frac{1}{2} \left( c_0 - c_0^B \right)^T B_0^{-1} \left( c_0 - c_0^B \right) + \frac{1}{2} \sum_{k=1}^{N} \left( H_k c_k - c_{k}^{\text{obs}} \right)^T R_k^{-1} \left( H_k c_k - c_{k}^{\text{obs}} \right).$$

(8)

The first term measures the mismatch between the actual initial conditions and the best prior estimate of them; $c_0$ is the initial concentration field of the control species, $c_0^B$ is the background (best prior estimate) concentration, and $B_0$ is the background error covariance matrix. The second term measures the model-observation mismatch; $c_k$ is the model prediction at time $t_k$, $c_{k}^{\text{obs}}$ is the vector of observations at $t_k$, $R_k$ is the observation error covariance matrix at $t_k$, and $H_k$ is the observation operator that maps the model space to the observation space.

The cost function (8) is minimized using a numerical optimization procedure. Here we employ L-BFGS, a limited memory quasi-Newton method for solving large scale optimization problems [12]. L-BFGS is an iterative method which requires, at each iteration step, the cost function value and its gradient at that point, and returns a better approximation to the optimal solution. The whole process continues until the convergence criteria are satisfied. The cost function and the gradient are obtained by one forward run, followed by one adjoint run of CMAQ-ADJ.

5.1. One-dimensional advection 4D-Var in a twin experiment setting

To investigate how the continuous and discrete advection adjoints impact the numerical optimization process, we first conduct 4D-Var twin experiments with the one-dimensional advection system. Four different scenarios are considered, as explained below.

We choose two reference initial conditions (to be retrieved through optimization in the twin experiment framework). The first initial condition
(IC₁) is smooth, while the second (IC₂) is non-smooth

\[ IC₁ : \quad c^\text{ref}_0(x) = 1 + \sin(3x), \quad (9a) \]

\[ IC₂ : \quad c^\text{ref}_0(x) = \begin{cases} 
2 & \text{for } x = k\pi/3, \ k \in \mathbb{Z}, \\
1 + \frac{\sin(3x)}{|\sin(3x)|} & \text{otherwise}. 
\end{cases} \quad (9b) \]

The background concentration, which also provides the starting point for the iteration, is a constant in space profile:

\[ c^B_0(x) = 1. \]

The background error covariance matrix is chosen as the identity matrix, which equates to considering a standard deviation of the background errors of 100%.

Two kinds of observations are used. The first set (O₁) measures the concentrations at 20 gridpoints evenly spaced throughout the \( N = 200 \) total computation grids. The second set (O₂) measures the concentrations at all \( N = 200 \) computation gridpoints

\[ O₁ : \quad c^\text{obs}_k = [c_k(x_1), c_k(x_{11}), \ldots, c_k(x_{181}), c_k(x_{191})]^T, \quad (10a) \]

\[ O₂ : \quad c^\text{obs}_k = [c_k(x_1), c_k(x_2), \ldots, c_k(x_{199}), c_k(x_{200})]^T. \quad (10b) \]

Synthetic observations are obtained by running the model starting from reference initial conditions, and saving the solution at different times. Observation error covariance matrices at any time are diagonal, with entries equal to 0.1.

The four different scenarios correspond to the four possible combinations of the two initial conditions (9) and the two observation sets (10). A different optimization is carried out in each case using the gradients provided by the continuous and by the discrete adjoint approaches. The maximum number of model runs is set to 150.

We analyze three different aspects of the results. First, the optimized solutions are compared visually against the reference solutions and against the initial guess. This gives a qualitative indication of the performance of data assimilation. Second, for a quantitative assessment, we show the decrease of the cost function with the number of model runs and with the number of iterations of the optimization procedure. Third, we consider the decrease of the root mean square (RMS) errors – of the optimal
solution against the reference one – with the number of model runs and iterations.

We first discuss the test with continuous initial conditions (9a). Figure 5 presents the reference initial condition, and the optimized initial conditions obtained with the discrete and with the continuous adjoint gradients. To reveal the fine features in the solutions we zoom in over the grid points 80–90. The case with sparse observations (10a) is shown in Figure 5(a). The continuous adjoint optimal solution is very accurate, and is indistinguishable from the reference solution. However, the discrete adjoint optimal solution shows non-physical wiggles of considerable amplitude over the peak of the reference solution. Near the solution peak the PPM limiters are active, and nonlinearity of the PPM scheme is the most pronounced. When the observations are dense (10b) both optimal solutions are visually identical with the reference solution, as shown in Figure 5(b).

The convergence of optimization iterations for the case with continuous initial conditions (9a) and sparse observations (10a) is analyzed in Figure 6. Each point represents one iteration of the L-BFGS optimization routine. The optimization converges nicely when continuous adjoint gradients are used. In 80 model runs the cost function decreases by a factor of $10^{12}$, and the RMS error decreases by a factor of $10^4$. When discrete gradients are used, the optimization initially makes steady progress (albeit at a slower rate than in the continuous adjoint case). The optimization stagnates after about 40 model runs. Each subsequent iteration (each marker on the graph) requires a considerable number of model runs, which indicates a larger number of line searches performed by L-BFGS. Therefore, it is clear that gradient obtained by continuous adjoint provides a more suitable descent direction for optimization. After reaching the plateau, neither the cost function nor the RMS error decrease any further. The corresponding values are relatively large, which is explainable by the fact that the optimal solution has wiggles and is not too close to the reference, see Figure 5(a). The continuous adjoint approach outperforms the discrete adjoint approach in regard to both convergence speed, and the quality of the optimal solution.

The convergence of optimization iterations for the case with continuous initial conditions (9a) and dense observations (10b) is analyzed in Figure 7. The performance of 4D-Var using the two approaches is similar. While the optimization with continuous adjoints converges slightly
faster, both gradients lead to similar decreases in the cost function and in the RMS error. This is in line with the results in Figure 5(b), where both optimal initial conditions are undistinguishable form the reference one.

The test cases where the reference solutions are not smooth lead to similar conclusions. We do not present their results in detail. For both dense and sparse observations, the initial conditions retrieved by the continuous adjoint approach better match the reference solution. The optimal initial condition obtained with discrete adjoints oscillates near the solution discontinuity points in the case with sparse observations. The discrete adjoint approach leads to a slower convergence of the optimization process both test cases.

In summary, the 4D-Var optimization provides more accurate analyses when the gradients are computed by the continuous adjoint approach. The optimal initial conditions obtained with discrete adjoint gradients have large errors near solution discontinuities, or at locations where the slope changes sign. The quality of the solution improves considerably when more information is available from observations. The faster convergence of optimization with continuous adjoints indicate that they provide better descent directions than discrete adjoints.

5.2. CMAQ-ADJ 4D-Var in a twin experiment setting

The first 4D-Var data assimilation test with CMAQ-ADJ is carried out in a twin experiment setting. The computational domain has $N_{COL} = 38$ columns, $N_{ROW} = 38$ rows, and $N_{LEV} = 6$ vertical layers. The active science processes are advection, diffusion, and chemistry. The assimilation time window length is twelve hours, between 00:00-12:00 CST on July 2, 1999.

We choose an array of equally spaced grid cells as observation locations. Specifically, observations are taken in the grid cells with row and column indices 1, 5, ..., 34, 38 and in the vertical layers 1, 3, 5. The measured species is ozone, and the control variables are ozone initial conditions. Synthetic observation values are obtained from a forward run starting from the reference initial condition $c_{0}^{\text{ref}}$. During this run the ozone concentrations at all the observation grid points are saved. The observation error covariance matrix is diagonal. Each entry corresponds to a variance of 1% of the reference concentration value in the corresponding grid cell and the corresponding time.
The background initial conditions, which also provide the starting point for the iterations, are obtained by perturbing the reference initial conditions: \( c_0^R = c_0^{\text{ref}} (1.3 + 9 \varepsilon) \) with \( \varepsilon = 0.001 \). The background error covariance matrix is diagonal. Each entry corresponds to a variance of 20% of the reference concentration value.

Three different optimization experiments are carried out. Each uses a different type of advection gradient, based on continuous adjoint, discrete adjoint, and modified discrete adjoint respectively. In all cases (the same) discrete adjoints are used for the diffusion and chemistry processes. The L-BFGS optimization is carried out for at most 26 model runs.

Figure 8 presents the absolute background and analysis errors. The background error is the difference between the background initial condition and the reference initial condition. The analysis errors are the differences between the optimized initial conditions given by 4D-Var and the reference initial condition. The background error (Figure 8(a)) is decreased after data assimilation. The optimization with continuous adjoints gradients yields a very accurate analysis, as seen in Figure 8(b). The optimization with discrete adjoints gradients leads to the analysis error presented in Figure 8(c). While this error is smaller than the background error (Figure 8(a)), the discrete adjoint analysis is less accurate than the continuous adjoint analysis.

Figure 9 presents the convergence speed of the numerical optimization process. Specifically, Figure 9(a) shows the decrease of the cost function with the number of runs. Figure 9(b) shows the decrease of the root mean square error (difference between the initial conditions found by optimization and the reference solution) with the number of model runs. The optimization converges considerably faster in both metrics when continuous advection adjoints are used. The optimization with discrete advection adjoints attains a plateau characterized by a large RMS error; this may indicate that the solution converges to a local minimum. When the modified discrete adjoint is used the RMS error actually increases.

We have seen in Figure 4 that the continuous adjoint yields smooth gradients, while the discrete adjoint approach leads to wiggles in the computed gradients. To investigate how the gradient smoothness of gradient affects optimization procedure, we add artificial diffusion to the adjoint fields after adjoint advection process to smooth out the gradient. The smoothed discrete adjoint gradients are then used in the optimization process.
Specifically, after each directional solution step of the discrete adjoint advection, we solve a diffusion equation

\[ \frac{\partial \lambda}{\partial t} = K \frac{\partial^2 \lambda}{\partial x^2}, \]  

(11)

where \( \lambda(x, y, z, t) \) is the adjoint variable field and \( K \) is the constant artificial diffusion. This extra step smooths the gradient by spreading out the wiggles created by the discrete adjoint model. Larger \( K \) values lead to smoother but more inaccurate gradients. This is illustrated in Table 4.

The cost function is computed for the background initial conditions \( c_B^0 \).

When a small perturbation \( \Delta c_0 = 10 \varepsilon c_0^{\text{ref}} \) is added to the initial conditions, the cost function changes with \( \Delta J = J(c_B^0 + \Delta c_0) - J(c_B^0) \). The cost function change is approximated to first order by the dot product of the adjoint variable times the perturbation, \( \Delta J \approx \lambda^T_0 \cdot \Delta c_0 \). The results in Table 4 the discrete adjoint with artificial diffusion matches well the finite difference value for \( K \leq 10^3 \), but becomes more and more inaccurate as \( K \) becomes larger.

<table>
<thead>
<tr>
<th>Artificial diffusion</th>
<th>Finite difference</th>
<th>Discrete adjoint</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K = 0 )</td>
<td>290.08</td>
<td>291.56</td>
<td>0.51 %</td>
</tr>
<tr>
<td>( K = 10^2 )</td>
<td>290.08</td>
<td>291.32</td>
<td>0.43 %</td>
</tr>
<tr>
<td>( K = 10^3 )</td>
<td>290.08</td>
<td>284.34</td>
<td>-2.02 %</td>
</tr>
<tr>
<td>( K = 10^4 )</td>
<td>290.08</td>
<td>219.95</td>
<td>-31.88 %</td>
</tr>
<tr>
<td>( K = 10^5 )</td>
<td>290.08</td>
<td>56.07</td>
<td>-417.33 %</td>
</tr>
</tbody>
</table>

Table 4: Finite difference validation of discrete adjoint gradients after using artificial diffusion. Changes in the the cost function \( \Delta J \) are computed by finite differences \( (J(c_B^0 + \Delta c_0) - J(c_B^0)) \) and by the first order adjoint sensitivity \( (\lambda^T_0 \cdot \Delta c_0) \).

Figure 10 illustrates the convergence of L-BFGS optimization with different levels of artificial diffusions applied to discrete adjoint of advection sensitivities. The performance is similar for all values of \( K \); regardless of the level of artificial diffusion applied, the optimization does not converge as fast as with continuous adjoint. The converged solution with the smallest RMS error is obtained for \( K = 10^4 \), while the smallest cost function value is obtained for \( K = 0 \) (no artificial diffusion). For \( K = 10^5 \) the gradient is very inaccurate and the convergence is noticeably slower than with smaller artificial diffusion coefficient values.
In summary, smoothing gradients by applying artificial diffusion everywhere is not helping the optimization process. We can regard the continuous adjoint approach with the PPM scheme as a selective application of artificial diffusion in areas where the adjoint variable has the largest variation (i.e., where the PPM limiters become active). This selective diffusion benefits the optimization more.

5.3. CMAQ-ADJ 4D-Var with real observation data

In this section 4D-Var data assimilation is carried out with the full CMAQ-ADJ adjoint. The real observation data are ground level ozone concentrations provided by the AIRNOW network over Texas (http://airnow.gov). The assimilation window is from 18:00 UTC to 24:00 UTC on September 1st, 2006. All model inputs (meteorological fields, emissions, boundary conditions) are real data for the simulation period and domain under consideration. The control variables are ozone initial conditions. Due to the relatively short window length the boundary conditions have a limited effect on model predictions. We investigate the impact of continuous and discrete adjoint gradients in the 4D-Var application. For each observation location (in space and time) there are multiple ozone concentration values available: the AIRNOW measurements, the model prediction during the initial free run (the background), and the model predictions using optimized initial conditions (the 4D-Var analyses obtained using continuous or discrete adjoint gradients).

The correlation coefficient between the CMAQ predicted ozone concentrations and AIRNOW measurements is $R^2 = 0.22$ when the background initial conditions are used. Figure 11 presents the scatter plots of the analyses versus observations. The correlation coefficient between observations and model predictions is $R^2 = 0.69$ when the simulation starts from the continuous adjoint optimal initial conditions. The correlation is $R^2 = 0.73$ for the discrete adjoint optimal initial conditions. Data assimilation considerably improves the model-observation agreement when compared to the free run case. The discrete and continuous gradients have comparable performance in this assimilation test, with the discrete adjoint leading to a slightly better agreement between model predictions and observations.

Figure 12 presents the ozone time series (one value per simulation hour) at the locations of two AIRNOW stations. We see that model results match observations much closer after data assimilation. The continuous
and the discrete adjoint analyses lead to very similar ozone predictions at both locations.

6. Conclusions

In this paper we study the effect of using discrete and continuous adjoints of the advection equation in chemical transport modeling with the CMAQ-ADJ model. Continuous advection adjoints are easily implemented by calling the same PPM subroutine during the backward integration, with a sign change for the winds, and with rescaling the solution. The construction of the discrete advection adjoint involve the differentiation of the nonlinear, monotonic PPM advection scheme.

A one-dimensional advection test problem is considered first. While the continuous adjoint approach produces smooth, positive sensitivity fields, the discrete adjoint approach leads to gradients that are corrupted by non-physical wiggles. The results show that discrete adjoint sensitivities agree better with finite differences in point to point tests than continuous adjoint sensitivities. However, in the context of 4D-Var data assimilation, the optimization using continuous adjoint gradients converges faster, and to a more accurate optimal solution, than with discrete adjoint gradients.

The discrepancy between the continuous adjoint and finite differences is explained by the nonlinearity of PPM scheme. The location of the receptor (the definition of the cost function) is important: if the receptor only measures the field at the location of the spurious wiggles caused by nonlinearity of PPM, the continuous adjoint and the finite differences produce very different sensitivity values. On the other hand, the good agreement between finite differences and discrete adjoint sensitivities reveals that discrete adjoints capture the nonlinearity of PPM. In summary, the discrepancy between continuous and discrete adjoint sensitivities depends on the receptor location; large discrepancies are due to algorithmic, rather than implementation, issues.

The implementation and validation of the discrete advection adjoints in CMAQ-ADJ are presented. In point to point tests the discrete adjoints of both horizontal and vertical advection match considerably better the finite difference sensitivities than the corresponding continuous adjoints. A first 4D-Var data assimilation test is carried out in a twin experiment.
framework. The optimization converges considerably faster when continuous adjoint gradients are used; moreover, the quality of the analysis is vastly superior to that obtained by discrete adjoint gradients. Artificial diffusion used to smooth out discrete adjoint sensitivities has little impact on the performance of optimization. A second 4D-Var data assimilation test is carried out using real AIRNOW observations and real meteorological fields. The assimilation adjusts the ozone initial conditions. The optimization has similar performance with both the continuous and the discrete advection adjoints, and reduces considerably the errors compared to their background levels.

In summary, the choice of advection adjoint in chemical transport modeling impacts the results of both the sensitivity analyses and the data assimilation studies. Typical advection solvers like PPM are highly nonlinear, and the gradients obtained by the discrete and by the continuous adjoint approaches are different. The discrete advection adjoint sensitivities come out as more accurate in point to point validation tests against finite differences. However, the continuous advection adjoints seem to provide better descent directions for numerical optimization, and lead to superior performance of the 4D-Var chemical data assimilation system.

Acknowledgements

The authors thank Houston Advanced Research Center for supporting this work through award H98. A. Sandu was partially supported by the National Science Foundation through awards NSF DMS–0915047, NSF CCF–0916493, NSF OCI–0904397. The authors thank Dr. Peter Percell, Dr. Tianfeng Chai, and Dr. Daewon Byun for providing the data used in the experiments of Section 5.3, and for very useful discussions related to the topics of this paper.

References


Figure 1: Solutions of the adjoint system of one dimensional adjoint advection PDE using (a) the continuous adjoint approach, (b) the discrete adjoint approach, and (c) the modified discrete adjoint approach. All solutions are reported at the initial time $t_0$. The receptor area (5) at the final time $t_F$ is between the gridpoints $a = 60$ ($x_a = 0.3$) and $b = 90$ ($x_b = 0.45$).
Figure 2: Vertical advection test with CMAQ-ADJ. Scattered plots of (a) continuous adjoint sensitivity, and (b) discrete adjoint sensitivity versus finite difference values. The simulation was run for 20 hours, with the source layer $L_s = 3$ and the receptor layer $L_r = 2$. This test is challenging since both the source and the receptor are located near the ground level.

Figure 3: Horizontal advection test with CMAQ-ADJ. Scattered plots of (a) continuous adjoint sensitivity, and (b) discrete adjoint sensitivity versus finite difference values. The simulation was run for 2 hours, with the source $C_s = 25$ and the receptor $C_r = 26$ for each model row.
(a) Difference between the perturbed solution and the base solution at final time: $PPM(\hat{c}_0) - PPM(c_0)$.

(b) Delta perturbation advected to final time using the PPM scheme: $PPM(\hat{c}_0 - c_0)$.

Figure 4: Perturbation at the final time due to a delta perturbation in the initial condition at the leftmost gridpoint.
Figure 5: Zoom in on the reference solution initial condition, and on the optimal initial conditions obtained with the discrete and with the continuous adjoint gradients. One dimensional 4D-Var optimization for smooth initial conditions $IC_1$ (9a). The two panels present results for different observation sets.

(a) Sparse observations $O_1$ (10a).

(b) Dense observations $O_2$ (10b).
Figure 6: Convergence for the optimization process using discrete and continuous adjoint gradients. One dimensional 4D-Var optimization for smooth initial conditions $IC_1$ (9a), and sparse observations $O_1$ (10a).
Figure 7: One dimensional 4D-Var optimization for smooth initial conditions $IC_1$ (9a), and dense observations $O_2$ (10b). Convergence for the optimization process using discrete and continuous adjoints. (a) Cost function decrease with model runs, and (b) RMS error decrease with the number of model runs.
Figure 8: CMAQ-ADJ 4D-Var in the twin experiment setting. Errors in ozone concentrations at the initial time (differences between background/optimized and reference ozone concentrations).
Figure 9: Convergence of L-BFGS during CMAQ-ADJ 4D-Var optimization in the twin experiment setting.
Figure 10: Convergence of L-BFGS optimization for artificial diffusion. (a) cost function values vs. model runs (global view); (b) cost function values vs. model runs (local view); (c) RMS values vs. model runs (global view); (d) RMS values vs. model runs (local view).
Figure 11: Scatter plots for the correlation between AIRNOW observations and CMAQ-ADJ 4D-Var analyses using continuous adjoint and discrete adjoint gradients. The correlation between observations and the background model predictions is $R^2 = 0.22$. 

$R^2 = 0.69269$ 

$R^2 = 0.73561$
Figure 12: Time series plots of ozone concentrations at the locations of two AIRNOW stations. The locations are represented by the model grid indices (column, row, layer).