Inverse Modeling of Aerosol Dynamics Using Adjoints: Theoretical and Numerical Considerations

A. Sandu,[1] W. Liao,[1] G. R. Carmichael,[2] D. K. Henze,[3] and J. H. Seinfeld[3]

[1]*Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, Virginia,USA*
[2]*Center for Global and Regional Environmental Research, University of Iowa, Iowa City, Iowa, USA*
[3]*Department of Chemical Engineering, California Institute of Technology, Pasadena, California, USA*

In this paper we develop the algorithmic tools needed for inverse modeling of aerosol dynamics. Continuous and discrete adjoints of the aerosol dynamic equation are derived, as well as sensitivity coefficients with respect to the coagulation kernel, the growth rate, and the emission and deposition coefficients. Numerical tests performed in the twin experiment framework for a single component model problem show that the initial distributions and the dynamic parameters can be recovered from time series of observations of particle size distributions.

INTRODUCTION

As our fundamental understanding of atmospheric particles and their transformations advances, novel computational tools are needed to integrate observational data and models together to provide the best, physically consistent estimate of the evolving state of the atmosphere. Such an analysis state better defines the spatial and temporal fields of key gas and particle phase chemical components in relation to their sources and sinks. This information is critical in designing cost-effective emission control strategies for improved air quality, for the interpretation of observational data such as those obtained during intensive field campaigns, and to the execution of air-quality forecasting. The development of the tools to integrate measurements and models is also critical to the challenge of a full utilization of the vast amounts of satellite data in the troposphere that are now becoming available, and which will become more prevalent in the coming years.

Assimilation of chemical information is only now beginning in air quality/chemistry arenas (Elbern et al. 1997; Sandu et al. 2003; Daescu et al. 2003; Sandu et al. 2004; Sandu et al. 2005), but offers the same motivations as those realized in the field of meteorology. Assimilation techniques can be utilized to produce three-dimensional, time varying optimal representations of the particle distributions in the atmosphere, that are consistent with the observed physical and chemical states.

Forward modeling of aerosols predicts the evolution of particle size distributions given the known parameters of the evolution (coagulation kernel, growth rate, emission rates, and deposition velocities) as well as the initial size distribution. Numerous numerical methods have been proposed in the literature for solving particle dynamics, and several are surveyed in Zhang et al. (1999). Methods for solving coagulation include the semi-implicit scheme (Jacobson 1999), Newton-Cotes quadrature (Sandu 2002), Galerkin (Pilinis 1990; Sandu and Borden 2003), and spline orthogonal collocation (Gelbard and Seinfeld 1978). Methods for solving the growth equation include the moving sectional approach (Kim and Seinfeld 1990; Lurmann et al. 1997; Jacobson 1997), finite volumes (Bott 1989), finite elements (Tsang and Hippe 1988; Meng et al. 1998), quintic splines (Nguyen and Dabdub 2001), semi-Lagrangian methods (Nguyen and Dabdub 2002), and other (Binkowski and Shankar 1995).

The inverse modeling problem consists of recovering the initial or emitted size distribution and the parameters of evolution given information about the time evolution of the system, for example periodic measurements of the number, surface, mass, or volume density. We take the variational approach, in which the inverse modeling problem is formulated as an optimization problem. In this paper the concept of "inverse modeling" is different than inverse in time simulation.

In this paper we develop the algorithmic tools needed for inverse modeling of particle dynamics. Continuous and discrete adjoints of the particle dynamic equation are derived, as well as sensitivity coefficients with respect to the coagulation kernel, the growth rate, and emission and deposition coefficients. For the derivations we consider spatially independent (box) models and

Address correspondence to Adrian Sandu, Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, VA 24060, USA. E-mail: sandu@cs.vt.edu

677

focus only on the physical particle dynamics aspects, that is, we do not treat chemical and thermodynamic transformations. We discussed inverse modeling of chemical kinetics in (Sandu et al. 2003), and data assimilation in 3D chemical–transport models (Sandu et al. 2004; Sandu et al. 2005). We have previously analyzed the adjoint equations for aerosols undergoing ideal thermodynamically driven condensational growth in (Henze et al. 2004), and more research is in progress to extend this work to nonideal systems. The methods developed here are a first step toward the goal of performing data assimilation for comprehensive three-dimensional particle chemistry and transport models.

The paper is organized as follows. An overview of the particle dynamics equations and of the forward modeling problem is given in Section The inverse modeling problem is formulated in Section Continuous adjoints are discussed in Section and discrete adjoints in Section A comprehensive set of numerical results on the coagulation and growth equation is presented in Section Section 6, and conclusions are drawn in Section Complete derivations of several continuous and discrete adjoint equations are given in Appendices A. through D.

AEROSOL DYNAMICS

To accurately study the effects of aerosols it is necessary to resolve aerosol number and mass distributions as a function of chemical composition and size. Three major approaches are used to represent the size distribution of aerosols: continuous, discrete and parameterized. In this paper we focus on continuous models (i.e., continuous size distributions and the general dynamic equations in continuous form) with particle size distributions being functions of particle mass (m) and time (t). The formulation follows mainly (Seinfeld and Pandis 1997). We also consider numerical discretizations of continuous models using the piecewise polynomial approach proposed by Sandu (Sandu 2004). The discussion in this paper focuses on physical transformations (growth, coagulation, sources, and sinks) and does not cover chemical transformations.

Number Density Formulation

We first consider the particle dynamic equation in number density formulation, with particle mass the independent variable. The size distribution function (number density) of a family of particles will be denoted by $n(m, t)$; the number of particles per unit volume of air with the mass between m and $m + dm$ is $n(m, t)\,dm$. This describes completely a population of single-component particles. The aerosol population undergoes physical and chemical transformations (Seinfeld and Pandis 1997) which change the number density in time:

$$\frac{\partial n(m, t)}{\partial t} = -\frac{\partial}{\partial m}[I(m, t)n(m, t)]$$
$$+ \frac{1}{2}\int_0^m \beta(m - m', m')n(m', t)n(m - m', t)\,dm'$$

$$- n(m, t)\int_0^\infty \beta(m, m')n(m', t)\,dm' + S(m, t)$$
$$- L(m, t)n(m, t). \tag{1}$$

with the initial and boundary conditions

$$n(m, t = t^0) = n^0(m), n(m = 0, t) = 0.$$

In this equation $I(m, t) = dm/dt$ is the rate of particle growth (e.g., due to condensation, evaporation, deposition and sublimation), $\beta(m, m')$ is the coagulation kernel, $S(m, t)$ is any source of particles of mass m (e.g., nucleation and emissions) and $L(m, t)$ is the first-order rate of removal of particles of mass m (e.g., by deposition).

Mass Density Formulation

We consider now particles composed of multiple chemical constituents $i = 1, 2, \ldots, r$. The total mass of a particle is the sum of masses of its r individual components

$$m = \sum_{i=1}^r m_i.$$

To describe the multicomponent population of particles one models the mass concentration distribution of each species i

$$q_i(m, t) = m_i n(m, t)[\mu g i\,\mu g^{-1} cm^{-3}],$$

with the total particle mass concentration distribution being

$$q(m, t) = mn(m, t) = \sum_{i=1}^r q_i(m, t).$$

The total mass density of component i in particles with masses between m and $m + dm$ is $q_i(m, t)\,dm$.

The growth (condensation/evaporation) rate of species i is $I_i = dm_i/dt$. It is customary to use the normalized growth rates

$$H_i(m, t) = \frac{1}{m}\frac{dm_i}{dt} = \frac{1}{m}I_i(m, t).$$

The total growth rate of particles of mass m is

$$H(m, t) = \frac{1}{m}\frac{dm}{dt} = \sum_{i=1}^r H_i(m, t).$$

The particle dynamic equation governs the evolution of the mass concentration distribution of each component $i = 1, 2, \ldots, r$ (Seinfeld and Pandis 1997).

$$\frac{\partial q_i(m, t)}{\partial t} = H_i(m, t)q(m, t) - \frac{\partial}{\partial m}[mq_i H]$$
$$+ \int_0^m \beta(m', m - m')q_i(m', t)\frac{q(m - m', t)}{m - m'}dm'$$

$$- q_i(m, t) \int_0^\infty \beta(m, m') \frac{q(m', t)}{m'} \, dm' + m_i S(m, t)$$
$$- L(m, t) q_i(m, t), \qquad [2]$$

where $m_i S(m, t)$ is the source term and and $L(m, t) q_i(m, t)$ is the removal term ($\mu g \, i \, \mu g^{-1} cm^{-3} \, s^{-1}$).

The initial and boundary conditions are

$$q_i(m, t = t^0) = q_i^0(m), \quad q_i(m = 0, t) = 0,$$
$$q_i(m = \infty, t) = 0.$$

The boundary condition at $m = \infty$ expresses the fact that there cannot be arbitrarily large particles (e.g., due to removal processes).

Discrete Formulation

In this section we exemplify the derivation of discrete particle dynamics models based on the number formulation of the dynamic equation (1) discretized in size with a piecewise polynomial approach as described by Sandu (2004). A completely similar approach for the mass density formulation is possible.

The finite dimensional approximation of the number distribution $n(m, t)$ is taken in the space spanned by the set $\{\phi_i\}_{1 \le i \le s}$ of piecewise polynomial basis functions. The dynamic equation discretized in size reads

$$n'(t) = G(t)n(t) + F(n(t)) + A^{-1}[B \times n(t)]n(t)$$
$$+ S(t) - L(t)n(t) \qquad [3]$$

with the initial condition by projecting the continuous initial condition of (1) onto the finite dimensional space:

$$n(t^0) = n^0.$$

The growth terms are discretized using a Discontinuous Galerkin approach where $F(n(t))$ is the Godunov flux difference and

$$G_{i,j}(t) = \int_0^\infty I(m, t)\phi_i(m)\phi_j'(m) \, dm$$

The coagulation terms are discretized using a collocation approach with collocation points $\{c_j\}_{1 \le i \le s}$. We have

$$A_{i,j} = \phi_j(c_i)$$
$$B_{i,k}^j = \frac{1}{2} \int_0^{c_j} \beta(m, c_j - m)\phi_i(c_j)\phi_k(c_j - m) \, dm$$
$$- \int_0^\infty \beta(m, c_j - m)\phi_i(m)\phi_k(c_j - m) \, dm,$$

and

$$\{[B \times n(t)] \cdot n(t)\}_i = n^T(t)B^i n(t).$$

A fully discrete particle dynamic equation is obtained by applying a time stepping algorithm to (3). For example the Forward Euler time discretization leads to

$$n^{k+1} = n^k + \Delta t\{G(t^k)n^k + F(n^k) + A^{-1}[B \times n^k]n^k$$
$$+ S(t^k) - L(t^k)n^k\}. \qquad [4]$$

where the numerical solution is $n^k \approx n(t^k)$.

The Forward Modeling Problem

In forward modeling one starts with a known initial distribution, $n^0(m)$ or $q_i^0(m)$, given boundary conditions and model parameters, $I(m, t)$ or $H_i(m, t)$, $\beta(m, m')$, $S(m, t)$, and $L(m, t)$. The evolution of the solution in time is completely described by the particle dynamic equation (1) or (2). In practice a numerical approximation, or discrete model of the form (4), is implemented and approximations of the distribution are computed at discrete time moments.

THE INVERSE MODELING PROBLEM

Variational methods provide an optimal control approach to the data assimilation problem. Four-dimensional variational (4D-Var) data assimilation allows the optimal combination of three sources of information: an a priori ("background") 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. The optimal analysis state is obtained through a minimization process to provide the best fit to the background estimate and to all observational data available in the assimilation window.

To be specific, consider the model (1) which describes the evolution of the number density of a particle population given the initial distribution $n^0(m)$ and the parameters p of the problem ($I(m, t)$ or $H_i(m, t)$, $\beta(m, m')$, $S(m, t)$, and $L(m, t)$). This equation is called the *forward model* and describes the (forward in time) evolution of particle density. The general inverse modeling problem consists of finding the initial distribution and the parameter values which minimize an integral cost functional

$$\mathcal{J}(n^0, p) = \int_{t^0}^T \int_0^\infty J_0(n(m, t)) \, dm \, dt. \qquad [5]$$

The functional depends on the initial distribution and the parameter values implicitly, through the mapping of (n^0, p) to $n(m, t)$ provided by the forward model. The minimization procedure needs the functional value $\mathcal{J}$ and its gradients with respect to the model parameters $\nabla_{n^0}\mathcal{J}$ and $\nabla_p\mathcal{J}$. For given values of (n^0, p) the functional value is obtained by a forward integration of the forward model, while the derivatives can be efficiently obtained through adjoint modeling, as will be explained in the following section.

Data Assimilation

Of particular interest in this paper is the data assimilation problem, where the observations $y^1 \ldots y^N$ are available at the discrete times $t^1 \ldots t^N$. The observed quantities are functions of the solution

$$y(t) = h(n(m, t)).$$

Observations can be particle number or mass densities, optical properties, and so on. Data assimilation uses the information from observations to estimate the values of a finite set of parameters p (which can provide a parameterization of the initial solution n^0, or determine the forward model parameters I or H, β, S, and L). The cost functional is defined to measure the mismatch between model predictions and observations

$$\mathcal{J}(n^0, p) = \frac{1}{2}(p - p_B)^T B^{-1}(p - p_B)$$
$$+ \frac{1}{2}\sum_{k=1}^{N}(y^k - h(n^k))^T R_k^{-1}(y^k - h(n^k)). \quad [6]$$

The first term penalizes the departure of the parameters from the apriori estimate p_B, and is called the background term. The matrix B is the error covariance associated with background terms and R_k are error covariances associated with observations. The parameters p, which initially have values p_B, are adjusted such that the value of the functional is minimized.

CONTINUOUS ADJOINTS OF THE DYNAMIC EQUATION

In this section we define the adjoint equations for the model (1) that offer a computationally efficient approach to computing the gradients of the cost functional (5) with respect to the model parameters and initial conditions. Note that the functional depends on n^0 and p implicitly, since the forward model evolution uniquely determines $\{n^k\}_{k \geq 1}$ from n^0 and p. In the continuous formulation one derives the adjoint of the original (continuous in time and size) dynamic equation, to obtain the derivative of the exact solution. For a computer implementation both the forward and the adjoint model are solved using the numerical methods of choice.

Number Density Formulation

The *adjoint equation* of the tangent linear model of equation (1) for the cost functional (5) defines the evolution of adjoint variables

$$\frac{\partial \lambda(m, t)}{\partial t} = -\frac{\partial \lambda(m, t)}{\partial m} I(m, t)$$
$$- \int_m^\infty \lambda(m', t)\beta(m, m' - m)n(m' - m, t)\, dm'$$
$$+ \lambda(m, t)\int_0^\infty \beta(m, m')n(m', t)\, dm'$$

$$+ \int_0^\infty \lambda(m', t)n(m', t)\beta(m', m)\, dm'$$
$$+ \lambda(m, t)L(m, t) - \frac{\partial J_0}{\partial n}(m, t)$$
$$= \mathcal{M}^*\lambda(m, t) - \frac{\partial J_0}{\partial n}(m, t), \quad t^0 \leq t \leq T, \quad [7]$$

with the following boundary and initial conditions

$$\lambda(m, t = T) = 0, \quad \lambda(m = 0, t) = 0, \quad \lambda(m = \infty, t) = 0.$$

A complete derivation of the adjoint equations and the sensitivity relations is given in Appendix A.

The adjoint variables are initialized at the final time T. Equation (7) is defined such that, when integrated backwards in time from T to t^0, the value of the adjoint variable at t^0 is the sensitivity of the cost functional with respect to initial values $\lambda(m, t^0) = \nabla_{n^0}\mathcal{J}$. Thus the adjoint Equation (7) provides a method to obtain the gradients needed in the minimization of (5).

The Equation (7) depends on the state of the forward model $n(m, t)$. Consequently, one needs first to integrate the forward model (1) and save the state at all times, then use this information during the backward integration of (7).

The sensitivities of the cost functional (5) with respect to forward model parameters are (see Appendix A):

$$\nabla_{n^0}\mathcal{J} = \lambda(m, t^0) \quad [8a]$$
$$\nabla_{\beta(m,m')}\mathcal{J} = \int_{t_0}^T \left(-\lambda(m, t) + \frac{1}{2}\lambda(m + m', t)\right)$$
$$\times n(m, t)n(m', t)dt \quad [8b]$$
$$\nabla_{I(m,t)}\mathcal{J} = \frac{\partial \lambda(m, t)}{\partial m}n(m, t) \quad [8c]$$
$$\nabla_{S(m,t)}\mathcal{J} = \lambda(m, t) \quad [8d]$$
$$\nabla_{L(m,t)}\mathcal{J} = -\lambda(m, t)n(m, t) \quad [8e]$$

All the sensitivity coefficients (8a)–(8e) of the cost functional can be obtained by one forward in time integration of the forward model (1), followed by a single backwards in time integration of the adjoint model (7). When solving the data assimilation problem each iteration for the minimization of the cost functional (5) requires one evaluation of the cost functional and one evaluation of its gradient, therefore we have one forward and one backward integration per optimization step.

Data Assimilation

For data assimilation the functional (6) is defined at discrete time moments and uses a discrete set of observations. Formally, the function J_0 in (6) and the corresponding forcing term in (7) contain delta functions at measurement times t^k. For an equivalent formulation consider t_-^k and t_+^k to be the moments immediately before and immediately after t^k. The

adjoint model (7) is defined on each subinterval $[t^{k-1}, t^k]$ for $k = N, N-1 \ldots 1$ as

$$\frac{\partial \lambda(m,t)}{\partial t} = \mathcal{M}^* \lambda(m,t), \quad t_+^{k-1} \le t \le t_-^k,$$

$$\lambda(m, t_-^k) = \lambda(m, t_+^k) - \left(\frac{\partial h}{\partial n}(m, t^k)\right)^T R_k^{-1}(y^k - h(n(m, t^k))),$$

$$\lambda(m, t_+^N) = 0, \quad \lambda(m, t^0) = \lambda(m, t_+^0)$$
$$+ \left(\frac{\partial p}{\partial n}(m, t^0)\right)^T B^{-1}(p - p_B),$$
$$\lambda(m = 0, t) = 0, \quad \lambda(m = \infty, t) = 0. \qquad [9]$$

Note that instead of delta forcing terms in this formulation we have explicit jumps of the adjoint variable between intervals (across the observation times). The derivative of p with respect to n may be nonzero for parameters used to define the initial distribution.

Mass Density Formulation

For multiple components the cost functional is defined as

$$\mathcal{J} = \int_{t^0}^T \int_0^\infty J_0(q_1(m,t), q_2(m,t), \ldots, q_n(m,t)) \, dm \, dt. \quad [10]$$

A complete derivation of the adjoint equation and sensitivity coefficients for the multicomponent, mass density formulation (2) is given in Appendix B. The adjoint equation is:

$$\frac{\partial \lambda_i(m,t)}{\partial t} = -\sum_{j=1}^n \lambda_j(m,t) H_j(m,t) - mH(m,t)\frac{\partial \lambda_i(m,t)}{\partial m}$$
$$- \int_0^\infty \frac{\beta(m,m')}{m'}[\lambda_i(m+m',t)$$
$$- \lambda_i(m,t)]q(m',t)\,dm'$$
$$- \int_0^\infty \frac{\beta(m',m)}{m}\sum_{j=1}^n[\lambda_j(m+m',t)$$
$$- \lambda_j(m',t)]q_j(m',t)\,dm'$$
$$+ \lambda_i(m,t)L(m,t) - \frac{\partial J_0}{\partial q_i}(m,t), i = 1, \ldots, n,$$

with the initial and boundary conditions:

$$\lambda_i(m, t = T) = 0, \quad \lambda_i(m = 0, t) = 0.$$

For data assimilation the adjoint equation can be redefined similar to (9). The sensitivity equations are:

$$\nabla_{q_i^0}\mathcal{J} = \lambda_i(m, t_0) \qquad [11a]$$

$$\nabla_{\beta(m,m')}\mathcal{J} = \sum_{i=1}^r \int_{t_0}^T \left[\lambda_i(m+m',t)q_i(m',t)\frac{q(m,t)}{m} \right.$$
$$\left. - \lambda_i(m,t)q_i(m,t)\frac{q(m',t)}{m'}\right]dt \qquad [11b]$$

$$\nabla_{H_i(m,t)}\mathcal{J} = \lambda_i(m,t)q(m,t)$$
$$+ \sum_{j=1}^n mq_j(m,t)\frac{\partial \lambda_j(m,t)}{\partial m} \qquad [11c]$$

$$\nabla_{S(m,t)}\mathcal{J} = \sum_{i=1}^r m_i \lambda_i(m,t) \qquad [11d]$$

$$\nabla_{L(m,t)}\mathcal{J} = -\sum_{i=1}^r \lambda_i(m,t)q_i(m,t) \qquad [11e]$$

The minimization procedure employed for data assimilation requires at each step one evaluation of the cost functional (10) and one evaluation of its gradient (11a)–(11e). Therefore each optimization step is completed using one forward in time integration of the forward model (2), followed by a single backwards in time integration of the adjoint model (11).

DISCRETE ADJOINTS OF THE DYNAMICS EQUATION

In the discrete approach the numerical discretization (4) of the the particle dynamic equation is considered to be the forward model. This is a pragmatic view, as only the numerical model is in fact available for analysis. The adjoint of the discrete model (4) is formulated and solved. The approach amounts to computing the derivatives of the numerical solution, rather than approximating the derivatives of the exact solution. In this section we exemplify the derivation of discrete adjoint techniques based on the number formulation of the dynamic equation discretized with a piecewise polynomial approch (Sandu 2004) and the Forward Euler method (4).

Direct Approach

Taking the adjoint of the discrete equation (4) leads to a method to propagate the adjoint variables backwards in time. The general derivation of the discrete adjoints, including the case of a higher-order explicit Runge Kutta time stepping, is presented in Appendix C. The cost functional is defined in terms of the discrete model state at times t^j, usually multiples of the integration time step

$$J = \sum_{j=1}^N J_0^j(n^j). \qquad [12]$$

The adjoint equation of (4) reads

$$\lambda^k = \lambda^{k+1} + \Delta t\{G^T(t^k) + F_n^T(n^k) + [(B + B^T) \times n^k]A^{-T}$$
$$- L(t^k)\} \cdot \lambda^{k+1} + \left(\frac{\partial J_0^k(n^k)}{\partial n^k}\right)^T \delta_k^{obs},$$
$$k = N, N-1 \ldots 0 \qquad [13]$$
$$\lambda^N = 0. \qquad [14]$$

The factor δ_k^{obs} is one if t^k is a time point used in the definition of J, and zero otherwise. The adjoint variable at t^0 gives the gradient of the cost functional (12) with respect to the initial distribution,

$$\lambda^0 = \nabla_{n^0} J.$$

Note that this is the derivative of the numerical solution as used in the definition of (12), as opposed to the continuous adjoint formulation (8a), where λ^0 defines the derivative of the continuous solution.

In practice the continuous forward model (1) is solved numerically, and so is the continuous adjoint Equation (7). Therefore the continuous adjoint approach is in practice a hybrid approach. The operations of numerical discretization and adjoint do not commute in general, and consequently the numerical solution of (7) is different from the discrete adjoint (13). For data assimilation problems one needs the derivative of the numerical solution, that is, the discrete adjoints are in principle preferred. For sensitivity studies using the adjoint method one wants to approximate the sensitivities of the continuous model, and the continuous adjoint seems more appropriate.

Automatic Differentiation

Given a program that implements the forward model, automatic differentiation builds a new, augmented program, which computes the analytical derivatives along with the original program (Corliss et al. 2001; Griewank 2000). The derivatives are propagated using the chain rule and are accurate up to machine precision. Automatic differentiation can work in forward mode, in which case the tangent linear model is produced, and in reverse mode, in which case the discrete adjoint is produced. Several automatic differentiation tools are available including TAMC (Giering 1997; Giering and Kaminski 1998; Giering), Tapenade (Courty 2003), Adifor, ADOL-C, NAGWare Fortran 95, and so on.

Automatic differentiation has several highly attractive features. First, it requires only (an implementation of) the original forward model. The transformations needed to obtain the adjoint gradients are done completely automatically. This allows, in principle, to perform inverse modeling with any forward model, no matter how complicated. Moreover, parts of the model may be legacy code for which we do not have all the underlying details, and an analytical (continuous) adjoint of this submodels may be impossible to derive. Next, automatic differentiation produces discrete adjoints, therefore derivatives of the numerical solution, which are appropriate for optimization purposes.

In this paper we use the automatic differentiation tool TAMC (Giering 1997; Giering and Kaminski 1998) to generate the discrete adjoint derivatives.

The Role of Forward Numerical Method

Clearly, the formulation of the discrete adjoint equation depends not only on the particle dynamic equation itself, but also on the numerical method used to solve it (to define the discrete forward model). To illustrate the dependence of the discrete adjoint equation on the numerical method used to solve the forward problem we derive in Appendix D the discrete adjoint of the popular semi-implicit method (Jacobson 1999) for solving coagulation. One can compare the coagulation terms in (13) with the discrete coagulation adjoint (D.5). The adjoint derivation (D.5) is useful in its own right for performing data assimilation when coagulation is solved with the semi-implicit method.

NUMERICAL RESULTS

The Test Problem

For the numerical experiments we consider the test problem of aerosol coagulation and growth from (Gelbard and Seinfeld 1978), which admits an analytical solution. Let N_t be the total initial number of particles and V_m the mean initial volume. The initial number distribution is exponential, the coagulation rate is constant, and the growth rate is linear with the particle volume:

$$n^0(v) = \frac{N_t}{V_m} \exp\left(-\frac{v}{V_m}\right), \quad \beta(v, w) = \beta_0, \quad I(v) = \sigma_o v.$$
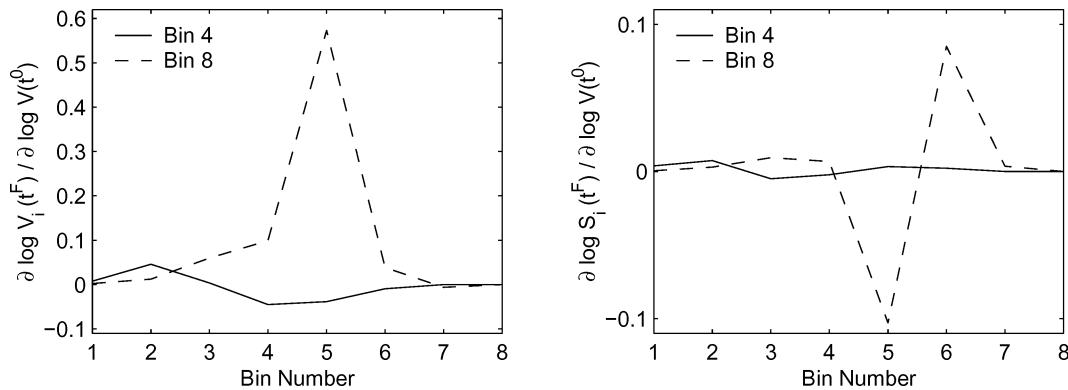

FIG. 1. The logarithmic sensitivities of the mean volume densities in bins 4 and 8 at final time with respect to the initial distribution bin averages (left) and bin slopes (right) for coagulation and growth.

TABLE 1
The logarithmic sensitivities of the mean volume densities in
bins 4 and 8 with respect to the values of σ and β for
coagulation and growth

|	$\partial \log V / \partial \log \sigma$	$\partial \log V / \partial \log \beta$
Bin 4	-0.288	-0.178
Bin 8	-0.303	-0.406

The analytical solution in (Gelbard and Seinfeld 1978) is

$$n^A(v,t) = \frac{4 N_t}{V_m(N_t\beta_o t + 2)^2} \cdot \exp\left(\frac{-2v \exp(\sigma_o t)}{V_m(N_t\beta_o t + 2)} - \sigma_o t\right).$$

We solve the dynamic equation for $\beta_o = 2.166 \times 10^{-6}\,\mathrm{cm^3\,h^{-1}}$ particles^{-1}, $\sigma_o = 0.02\,\mathrm{hour^{-1}}$, $N_t = 10^4$ particles, $V_m = 0.03\,\mu\mathrm{m}^3$. The value of β_o follows from (Jacobson 1999), and the value of σ_o is chosen such that coagulation and growth have effects of comparable magnitude. The size range is truncated to the volume interval $V_{\min} = 10^{-3}\mu\mathrm{m}^3$, $V_{\max} = 1\,\mu\mathrm{m}^3$. A piecewise linear discretization with 8 bins is employed for particle size as described in Section 5. Bin centers are log-uniformly distributed in the range $[V_{\min}, V_{\max}]$. The time interval under consideration is $[t_0 = 0, T = 48]$ h. The time discretization is performed with the second order Runge Kutta scheme (C.1) with the time step $\Delta t = 6$ min. The actual implementation is done in the (equivalent) volume concentration density formulation, which is also used for the presentation of results.

The experiments are carried out in the twin experiment framework. A reference run with the reference values for initial conditions, coagulation kernel, and growth rate is used to generate pseudo-observations $\{y_1 \ldots y_M\}$ of the number density, or a function of the number density

$$y_j^k = h_j(t^k, n^k), \quad 1 \le j \le M.$$
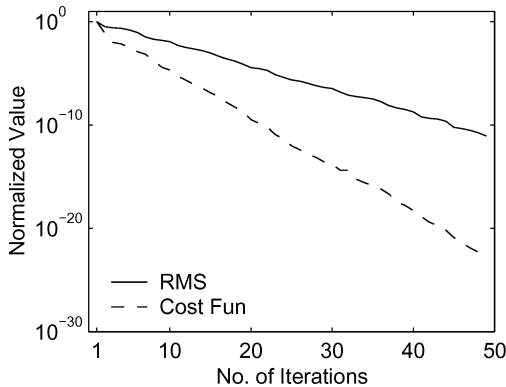
Each pseudo-observation is a vector of M components. For example, the number concentration is represented by a linear polynomial inside each bin. If we observe the number concentrations, the vector of observations will contain two values per observed bin (for the mean concentration and the slope inside the bin, or, equivalently, for two distinct concentrations inside the bin). The pseudo-observations are recorded at the end of each hour (i.e., after each 10 integration time steps) for the 48-h interval.

The model is re-run with perturbed values of the initial conditions, coagulation kernel, and growth rate, and the differences between model results and observations is used to define the cost functional

$$\mathcal{J}(n^0, \sigma, \beta) = \sum_{k=1}^{48} \sum_{j=1}^{M} \left(y_j^k - h_j(t^k, n^k)\right)^2 \qquad [15]$$

Note that this cost functional contains no background terms, which is justified by our apriori knowledge of the fact that the initial guesses are wrong, while the observations are correct.

The reference values of n^0, σ, and β are recovered by solving the minimization problem

$$\min_{n^0, \sigma, \beta} \mathcal{J}(n^0, \sigma, \beta)$$

In our experiments the number of independent observations is larger than the number of parameters to recover, and is likely that the minimization problem has a unique solution even in the absence of background terms. The gradient of the function (15) is obtained through adjoint modeling. The majority of the numerical experiments presented here are carried out using the discrete adjoints generated with TAMC. In Section 6.7 we compare the results obtained using continuous and discrete adjoints, and in Section 6.8 the results obtained with the discrete adjoints obtained analytically and via automatic differentiation.
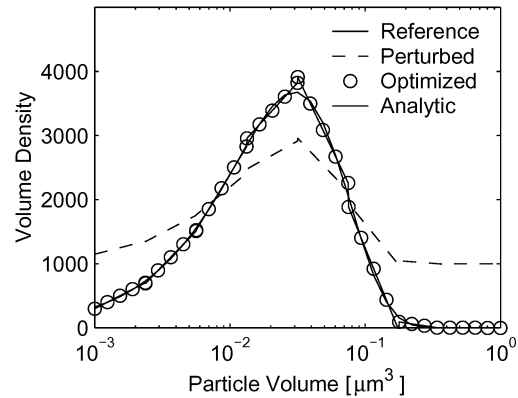

FIG. 2. Results for recovering the initial distribution for coagulation and growth from a complete set of observations.
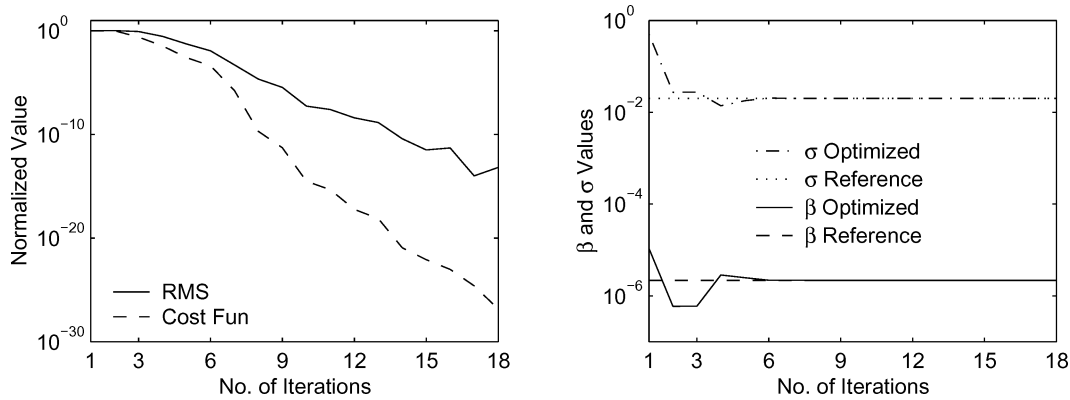
FIG. 3.   Results for recovering β and σ for coagulation and growth from a complete set of observations.

The optimization algorithm used is LBFGS (Nocedal 1980; Byrd et al. 1995). For the recovery of σ and β parameters we prescribe lower and upper bounds for the optimization as follows:

$$\sigma_{\text{lower}} = 0.0, \quad \sigma_{\text{upper}} = 0.5, \quad \beta_{\text{lower}} = 0.0, \quad \beta_{\text{upper}} = 10^{-4}.$$

These bounds are more than one order of magnitude larger than the reference values of the parameters (mimicking the real situation where the uncertainty in parameter values is large). Without using bounds to constrain the optimization process the combined initial distribution and parameters recovery is not convergent. This is due to the difficulty of the test where not only the state, but the also model are uncertain. Moreover, the uncertainty allowed in the model parameters is extremely large (as explained below, we start the optimization process with 5 times the reference coagulation kernel and 25 times the growth rate). The bounds should be set to describe the physically meaningful region in the parameter space. The use of bounds prevents the optimization process to search for false, unphysical solutions (e.g., $\beta < 0$).

As a measure of accuracy in the optimization process we consider the RMS difference between the reference and the optimized solutions:

$$\text{RMS} =$$

$$\sqrt{\frac{\delta_n \sum_{j=1}^{\text{ndof}} \left(\frac{n_{\text{ref}}^0[j] - n_{\text{opt}}^0[j]}{n_{\text{ref}}^0[j]} \right)^2 + \delta_\sigma \left(\frac{\sigma_{\text{ref}} - \sigma_{\text{opt}}}{\sigma_{\text{ref}}} \right)^2 + \delta_\beta \left(\frac{\beta_{\text{ref}} - \beta_{\text{opt}}}{\beta_{\text{ref}}} \right)^2}{\delta_n \cdot \text{ndof} + \delta_\sigma + \delta_\beta}}.$$

Here ndof is the number of degrees of freedom used in the discrete representation of the distribution, and δ_n (δ_σ, δ_β) equals one if we recover n^0 (σ, β), and is zero otherwise.

In the following experiments the initial guesses are obtained by providing a significant perturbation of the reference values. The perturbed values of model parameters are $\beta_p = 5\beta_o = 1.08 \times 10^{-5}$ cm^3 h^{-1}particles^{-1} and $\sigma_p = 25\sigma_o = 0.5$ h^{-1}.

Sensitivity Analysis

Adjoint calculations are a useful tool for sensitivity analysis. A widely used measure of the relative change in the functional
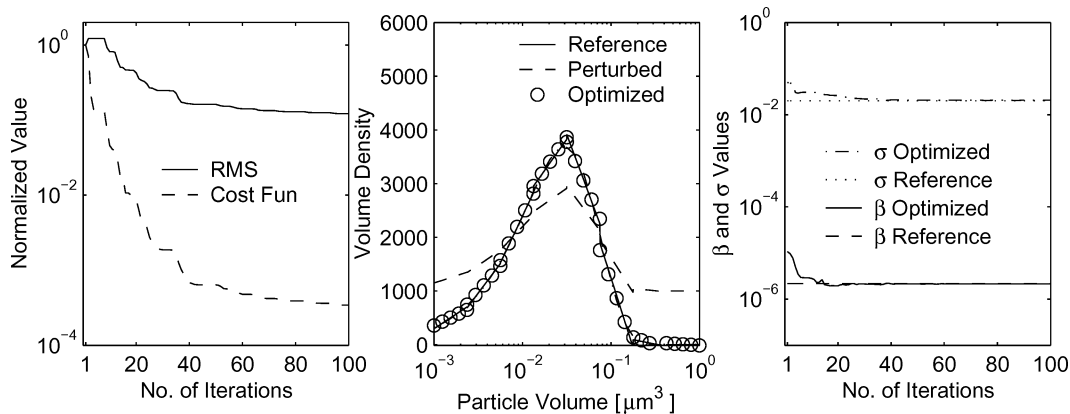

FIG. 4.   Results for recovering the initial distribution, β, and σ for coagulation and growth from a complete set of observations.
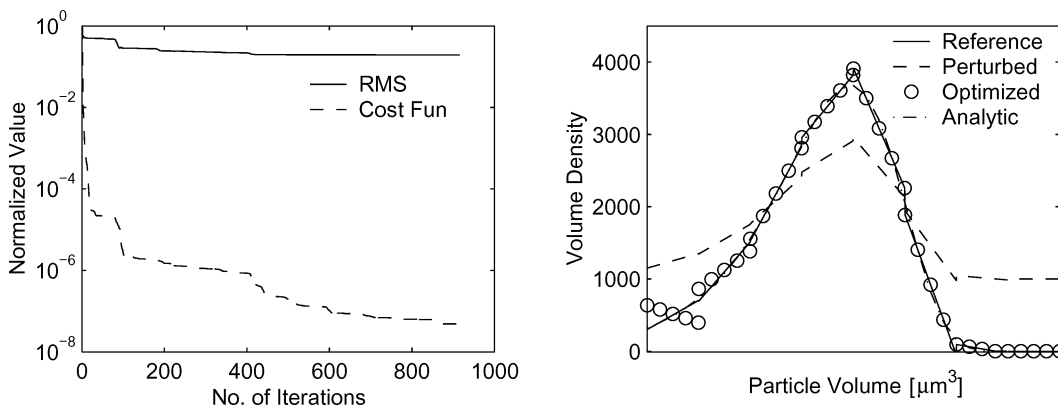
FIG. 5.   Results for recovering the initial distribution without observations at bins 1, 2, and 3 for both coagulation and growth.

due to relative changes in the input parameters is given by the logarithmic sensitivity coefficients

$$s = \frac{\partial \log J(t^F)}{\partial \log V(t^0)} = \frac{\partial J(t^F)}{\partial V(t^0)} \cdot \frac{V(t^0)}{J(t^F)}.$$

Figure 1 shows the logarithmic sensitivity coefficients of the mean volume densities in bins 4 and 8 at the end of the 48-h simulation interval with respect to the initial mean volume densities (left) and slopes (right). For both cases relative changes in the bin slopes have a considerably smaller influence on the final result than relative changes in the mean densities. The initial density of particles in bin 5 have the largest impact on the final density of large particles (bin 8). This means that after 48 h of evolution most of the large particles were formed through the growth and coagulation of particles that initially were in size bin 5. Conversely, the growth of particles from bin 5 to larger sizes has relatively small impact on the number density of particles in bin 4, hence, the sensitivities in bin 4 are small in comparison.

The logarithmic sensitivities with respect to the values of the growth parameter σ and the coagulation parameter β are given in Table 1. The influences of relative perturbations in these parameters are of the same magnitude for both bins. Note that all sensitivity coefficients for each bin were obtained via a single adjoint integration.

Data Assimilation with Complete Observations

We first consider a complete set of hourly observations, that is, all parameters of the solution (mean concentrations and slopes in each bin) are observed once every simulation hour. The test problem consists of the evolution under both coagulation and growth. Coagulation-only and growth-only test results are similar and are discussed briefly at the end of this section.

In Figure 2 the results for recovering the initial distribution are shown (the parameters β, σ assume reference values). The left plot shows the relative decrease of the cost function and RMS error with the number of iterations; the values of RMS and cost function are scaled by their initial value. We let the optimization routine run to convergence, and for the solution the RMS is decreased more than 10^{10} times from its starting value, and the cost function more than 10^{20} times. The right plot shows the exact, reference, perturbed, and optimized distributions at the initial time. The optimized distribution is visually identical to the reference one. We also show the analytical distribution
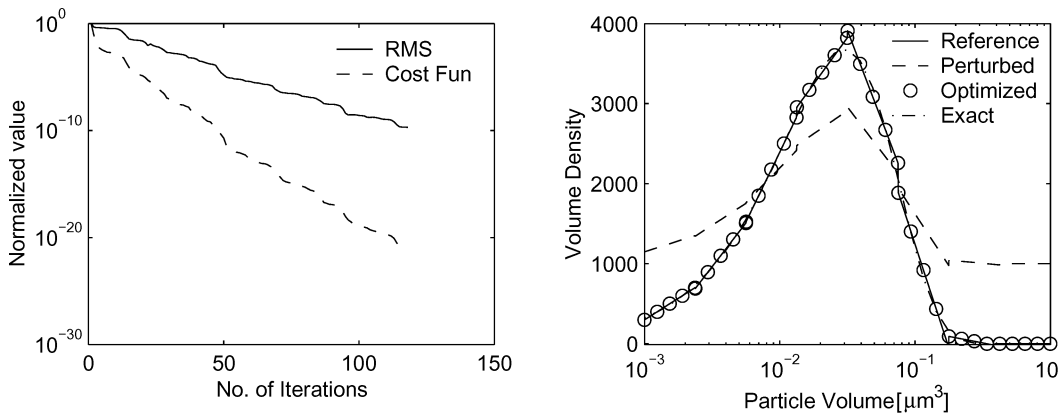

FIG. 6.   Results for recovering the initial distribution with observations at only even bins for both coagulation and growth.
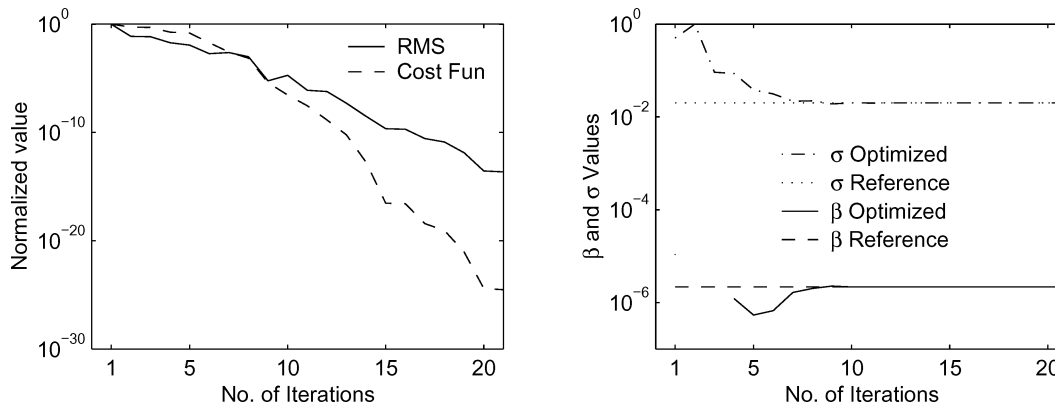
FIG. 7.   Results for recovering β and σ with observations at bin 1 for both coagulation and growth.

(round peak) which is well approximated by the numerical solution.

In Figure 3 the results for recovering β and σ are shown (the reference initial distribution is used in this experiment). The left plot shows the decrease of the cost function and RMS error with the number of iterations. The right plot shows the reference and optimized β, σ. As expected, the recovery of these 2 parameters require fewer iterations than the recovery of initial conditions (16 parameters).

In Figure 4 the results for recovering simultaneously the initial distribution, β, and σ are shown. This experiment is more challenging for the inversion procedure since perturbations now affect not only the initial distribution, but the dynamic equation itself (through β and σ). The change of these parameters affects the evolution of the model throughout the simulation time interval. The left Figure 4 plot shows the decrease of the cost function and RMS error with the number of iterations. Note the modest decrease of RMS and cost function, and the large number of iterations. The central plot shows the exact, reference, perturbed, and optimized distributions at the initial time. The optimized distribution is visually identical to the reference one. The right plot shows the reference and optimized β and σ, which are recovered accurately.

We have also performed separate experiments for growth only (recovering n^0, σ, and both) and coagulation only (recovering n^0, σ, and both) problems. The results are qualitatively similar to the ones presented in this section, and are not shown. The recovery of the parameters σ and β respectively is achieved in fewer than 10 iterations (RMS decreased more than 10^{10} times). The recovery of initial conditions is carried out in 15 iterations for coagulation, and 50 iterations for growth (RMS decreased more than 10^{10} times). For recovering both the initial distribution and the parameter (σ and β respectively) the optimization converges slower as expected.

Observations in Only Part of the Bins

In this section we present the performance of the data assimilation procedure when hourly measurements of number density are restricted to only a selected number of bins. We again consider the complete test problem with both coagulation and growth.

In Figure 5 the results for recovering the initial distribution without observations at bins 1, 2, and 3 are shown. The left plot shows the decrease of the cost function and RMS error with the number of iterations. Note the very large number of iterations needed, and the fact that, while the cost function decreases
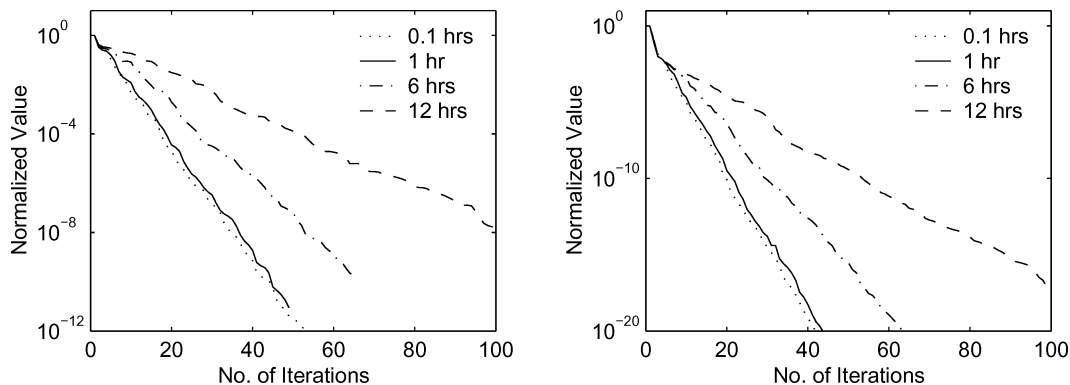

FIG. 8.   Results for recovering the initial distribution from a complete set of observations with different frequencies. RMS error on the left, and cost function value on the right.
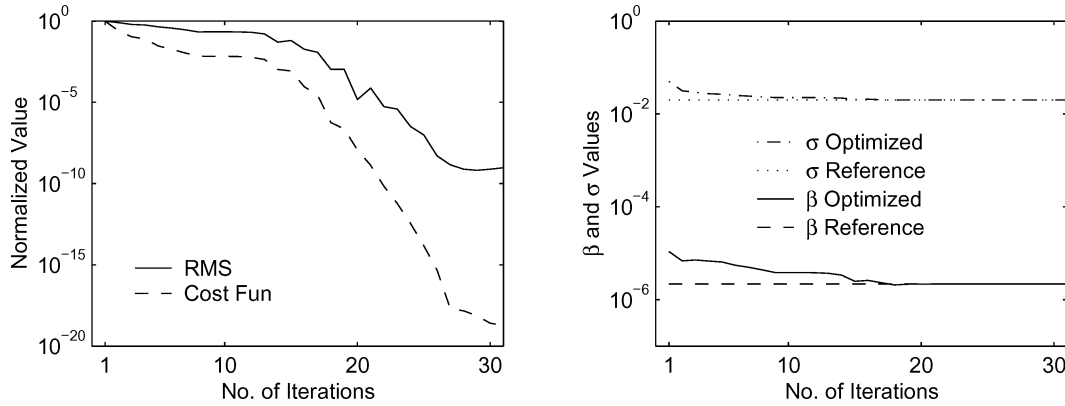
FIG. 9.   Recovering β and σ from observations of total surface density.

considerably, the RMS remains bounded below to about 10^{-1}. The right plot gives insight into this behavior. The initial distribution is recovered well overall, but the optimized values in bins 1 and 2 are inaccurate. For example, at bin 1 the bin average is correct but the slope is not. This is somewhat expected as the information in bins 1, 2, and 3 is not explicitly available to the optimization process. Additional experiments (not shown here) were carried out. The initial profile is recovered correctly with missing observations in bins 4 and 5 but shows consistent inaccuracies when observations in a large number of bins are missing. The conclusion is that it is possible to recover to a good extent the distribution of finer particles from the evolution equations, provided that enough observations of coarse particle are available.

Figure 6 presents the results for recovering the initial distribution with observations at only even bins. The information provided by even bins is sufficient for an accurate recovery of the initial conditions. The complementary experiment (not shown here) of providing observations in only the odd-numbered bins leads to a good recovered initial profile, with some inaccuracies in the last bin (number 8).

The recovery of both parameters β and σ with observations at only bin 1 was tested, and the results are shown in Figure 7.

The right panel shows the reference and optimized β and σ; note that, although the initial perturbations are very large, the reference values are recovered accurately with only several iterations. Additional experiments showed that the optimization process does not converge when observations are provided only in the large size bins (5–8), therefore the information in the lower size bins seems more important for the recovery of evolution parameters.

The Impact of Observation Frequency

To assess the importance of observation frequency we repeat the experiments with observations in all bins available every 0.1 h, 1 h, 6 h, and 12 h. Figure 8 shows the RMS (left) and cost function (right) decrease with the number of iterations for these cases. Increasing the observation frequency below 1 hour does not seem to benefit the assimilation. Less frequent observations each 6 and 12 hours respectively impact the assimilation performance, but in both cases the optimization process converges. Note that for complex problems which include fast processes (chemistry and thermodynamics) a high observation frequency (below 1 hour) is expected to have a positive impact.

The results in Figure 8 should be compared with the results in Figures 5 and 6. The assimilation with relatively frequent, but
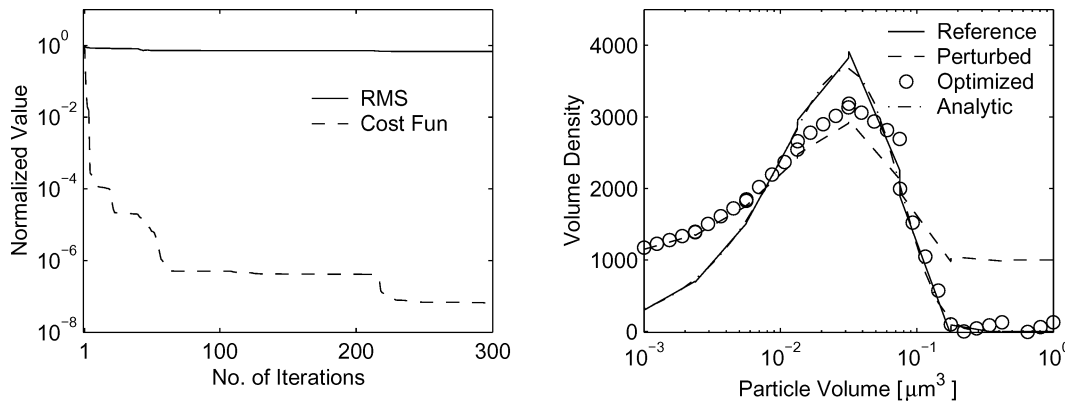

FIG. 10.   Recovering the initial distribution from observations of total surface density.
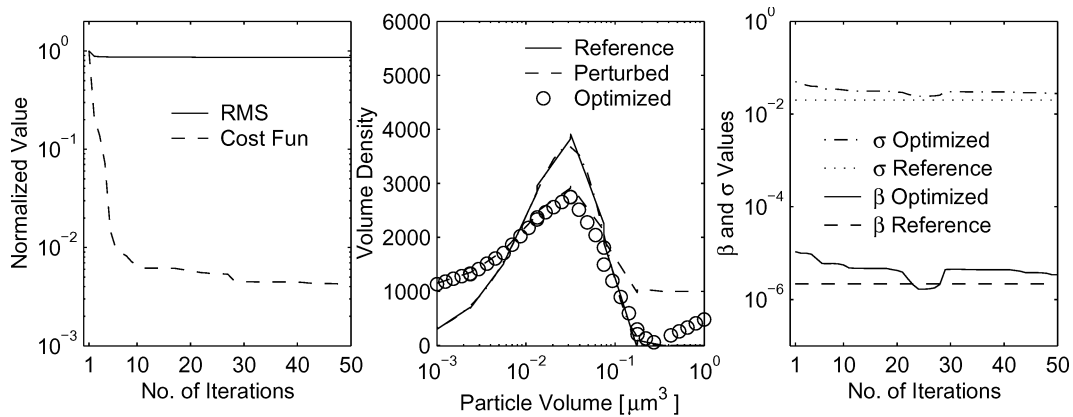
FIG. 11.   Recovering the initial distribution, β, and σ from observations of total surface density.

incomplete observation is more difficult than the assimilation with infrequent but complete observations for the model problem under consideration.

Observations of Total Surface Density

Several tests were performed to recover the model parameters from hourly observations of total particle number, volume, and surface area per air volume, and for coagulation, growth, and coupled models. The results are qualitatively similar, therefore it is sufficient to only present the total surface density observations case for coupled coagulation and growth.

Figure 9 shows the results for recovering β and σ (with the evolution of the optimized values on the right). The parameters are recovered accurately after a relatively small number of iterations.

The recovery of initial distributions is shown in Figure 10. While the cost function decreases as expected, the RMS stays flat for very large iteration numbers. The initial distribution in the large size bins is recovered fairly accurately, while the distribution in the lower size bins is hardly changed from the initial guess (perturbed value). This shows that in the logarithmic representation the small size bins span a much smaller physical

volume range, and therefore contribute very little to the total surface density. Consequently total surface (number, volume) density observations contain more information on the large size bin densities (which are recovered) and little information on the small bin densities.

Finally Figure 11 shows the results for simultaneously recovering the initial distribution, β, and σ. The optimized solution shows assimilation of the initial density in largest bin sizes, and an improvement in the β and σ values. The total surface density does not contain sufficient information for a complete recovery, and these parameters are still relatively far from the reference values, as are the initial densities in the lower bins.

The results obtained with the current initial density are typical, that is, data assimilation of other initial density profiles will have the same qualitative behaviour.

Discrete versus Continuous Adjoints

To compare the influence of the choice of discrete versus continuous adjoint gradient on the performance of the optimization process we consider the growth-only test problem and repeat the data assimilation experiment (for both the initial density and σ) using a continuous adjoint gradient of the cost function. The
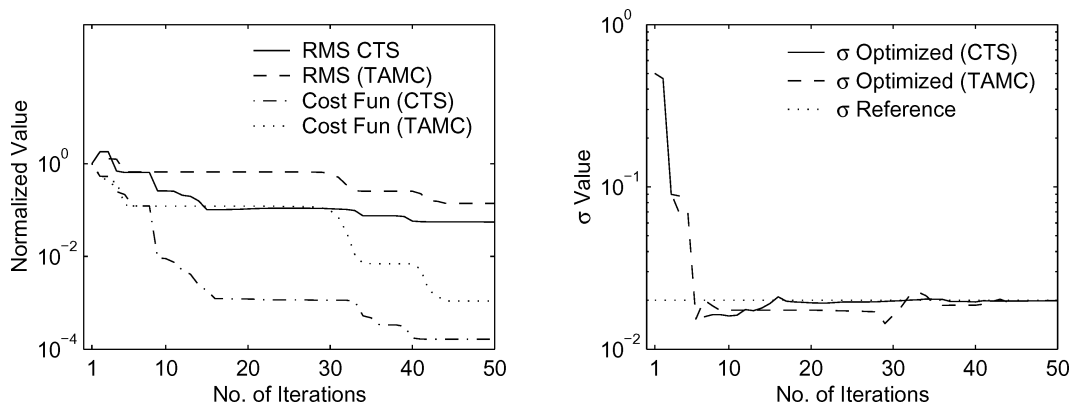

FIG. 12.   Recovering the initial distribution and σ for growth using continuous and discrete adjoint gradients. The continuous formulation leads to a better performance of the optimization process.
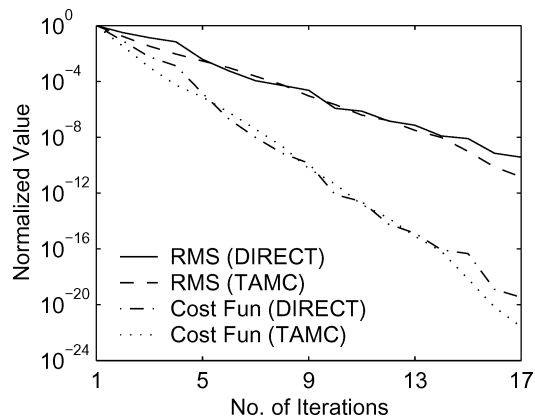
FIG. 13. Recovering the initial distribution for coagulation using two implementations of the discrete adjoint gradients. The automatic differentiation and the direct implementation lead to similar performance.

continuous adjoint of the growth equation is also a growth (advection) equation, and the same Discrete Galerkin numerical solver was applied (note that in the volume density formulation used here the advection speed, or the growth rate, is constant). The results are shown in Figure 12. The continuous adjoint gradients lead to a faster convergence of the optimization process, as can be seen from the faster decrease in both cost function and RMS with the number of iterations (left plot). The value of the growth rate σ is also recovered in fewer iterations (right plot). This result is somewhat surprising as one would expect the derivative of the numerical solution, that is, the discrete adjoint, to be more accurate for optimization purposes.

Direct versus Automatic Implementation of Discrete Adjoints

To assess the quality of the adjoints generated via automatic differentiation (TAMC) we compare their performance with a direct implementation of discrete adjoints as given by Equation (13). We consider the coagulation-only test case. The results presented in Figure 13 show that the optimization process performs similarly with both implementations, and requires a similar cpu time. This is expected since both approaches implement the same adjoint formulas, and the TAMC code is as efficient as the hand code for explicit time stepping.

CONCLUSIONS AND FUTURE WORK

In this paper we have developed the algorithmic tools needed for inverse aerosol modeling. Continuous and discrete adjoints of the integro-differential particle dynamic equation in number and mass concentration density formulations are derived. Formulas are obtained for the sensitivity coefficients of the solution with respect to the coagulation kernel, the growth rate, and emission and deposition coefficients.

The derivations are restricted to dynamics (without chemical and thermodynamic processes) and to zero dimensions (no transport processes). This is a first step toward performing data assim-

ilation for comprehensive three-dimensional particle chemistry and transport models.

Comprehensive tests have been carried out using a single component particle dynamics model in a twin experiment framework. Pseudo-measurements are generated from a reference model run. From hourly measurements of the particle size one can recover the initial distribution as well as the parameters of the model. Measurements in only some of the bins provide enough information to recover the initial distribution and parameters. Measurements of total surface, volume, or number density are useful to adjust the unknown model parameters and initial conditions, and provide information mainly for the large size bins. Caution must be exercised when extrapolating these conclusions to the case of real observations, which are influenced by chemical and thermodynamic processes in addition to particle dynamics, and are corrupted by measurement errors. Aerosol data assimilation using real observations will be the focus of future work.

The overall conclusion is that 4D-Var data assimilation is a feasible approach for particle dynamics models. Discrete adjoints can be obtained easily through automatic differentiation, but continuous adjoints are also a viable alternative.

Testing of the computational tools developed in this paper is planned for more realistic models and a variety of evolution conditions. Data assimilation for a growth problem with three-component aerosols is presented in the companion paper (Henze et al. 2004). Future work will extend the inverse modeling of aerosols to include chemical and thermodynamic processes. The techniques developed will be used ultimately to perform data assimilation in full three-dimensional models.

REFERENCES

Mathematics and Computer Science Division at Argonne National Laboratory and the Center for Research on Parallel Computation at Rice University, Automatic Differentiation of Fortran, http://www-unix.mcs.anl.gov/autodiff/ADIFOR/

Binkowski, F. S., and Shankar, U. (1995). The Regional Particulate Matter Model: 1: Model Description and Preliminary Results, *Journal of Geophysical Research*, 100:26,191–26,209.

Bott, A. (1989). A Positive Definite Advection Scheme Obtained by Nonlinear Renormalization of the Advection Fluxes, *Monthly Weather Review*, 117:1006–1115.

Byrd, A., Lu, P., and Nocedal, J. (1995). A Limited Memory Algorithm for Bound Constraint Optimization, *SIAM Journal on Scientific Computing* 16:1190–1208.

Carmichael, G. R., Daescu, D., Sandu, A., and Chai, T. (2003). Computational Aspects of Chemical Data Assimilation into Atmospheric Models, ICCS03, Melbourne, Australia, June 2003. In Springer Lecture Notes in Computer Science, Vol. 2660, Springer-Verlag GmbH, pp. 269–278.

Corliss, G., Faure, C., Griewank, A., Hascoet, L., and Naumann, U. (2001). Automatic Differentiation of Algorithms, from Simulation to Optimization Springer LNCSE series.

Courty, F., Araya, M., Vazquez, M., Pascual, V., Dervieux, A., Bellesso, N., Hascoet, L., and Greborio, R. (2003). Tapenade On-line Automatic Differentiation Engine. http://www-sop.inria.fr/tropics/index.html INRIA, France.

Daescu, D., Sandu, A., and Carmichael, G. R. (2003). Direct and Adjoint Sensitivity Analysis of Chemical Kinetic Systems with KPP: II—Validation and Numerical Experiments, *Atmospheric Environment* 37:5097–5114.

Elbern, H., Schmidt, H., and Ebel, A. (1997). Variational Data Assimilation for Tropospheric Chemistry Modeling, *J. Geophysical Research* 102(D13):15967–15985.

Gelbard, F. M., and Seinfeld, J. H. (1978). Coagulation and Growth of a Multicomponent Aerosol, *J. Colloid and Interface Science* 63(3):472–479.

Gelbard, F. M., and Seinfeld, J. H. (1996). Coagulation and Growth of a Multicomponent Aerosol, *J. Colloid and Interface Science* 63(3):472-479.

Giering, R. (1997). Tangent Linear and Adjoint Model Compiler, Users Manual Manual.

Giering, R., and Kaminski, T. (1998). Recipes for Adjoint Code Construction Article in ACM Trans. Math. Software.

Giering, R. Tangent linear and Adjoint Model Compiler, http://www.autodiff.com/tamc.

Griewank, A., Juedes, D., and Utke, J. (1996). ADOL–C, A Package for the Automatic Differentiation of Algorithms Written in C/C++ Article in ACM Trans. Math. Software. http://www.math.tu-dresden.de/wir/project/adolc/index.html

Griewank, A. (2000). Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation Frontiers in Applied Mathematics.

Henze, D. K., Seinfeld, J. H., Liao, W., Sandu, A., and Carmichael, G. R. (2004). Inverse Modeling of Aerosol Dynamics: Condensational Growth, *J. Geophys. Res.* 109:D14.

Jacobson, M. Z. (1997). Numerical Techniques to Solve Condensational and Dissolutional Growth Equations When Growth is Coupled to Reversible Reactions, *Aerosol Sci. Technol.* 27:491–498.

Jacobson, M. Z. (1999). *Fundamentals of atmospheric modeling*. Cambridge University Press, New York.

Kim, Y. P., and Seinfeld, J. H. (1990). Simulation of Multicomponent Aerosol Condensation by the Moving Sectional Method, *J. Colloid and Interface Science*, 135(1):185–199.

Lurmann, F. W., Wexler, A. S., Pandis, S. N., Musara, S., Kumar, N., and Seinfeld, J. H. (1997). Modeling Urban and Regional Aerosols - ii. Application to California's South Coast Air Basin, *Atmos. Environ.* 31:2695–2715.

.eng, Z., Dabdub, D., and Seinfeld, J. H. (1998). Size-Resolved and Chemically Resolved Model of Atmospheric Aerosol Dynamics, *J. Geophysical Research*, 103(D3):3419–3435.

Naumann, U., Christianson, B., and Riehme, J. NAGWare Fortran 95. Automatic Differentiation Differentiation Enabled Fortran Compiler Technology. http://www.nag.co.uk/nagware/research/ad_overview.asp

Nguyen, K., and Dabdub, D. (2001). Two-Level Time-Marching Scheme Using Splines for Solving the Advection Equation, *Atmos. Environ.* 35:1627–1637.

Nguyen, K., and Dabdub, D. (2002). Semi-Lagrangian Flux Scheme for the Solution of the Aerosol Condensation/Evaporation Equation, *Aerosol Sci. Technol.* 36:407–418.

Nocedal, J. (1980). Updating Quasi-Newton Matrices with Limited Storage, *Mathematics of Computation* 35:773–782.

Nocedal, J. Department of Electrical & Computer Engineering, Northwestern University. http://www.ece.northwestern.edu/ nocedal/lbfgs.html

Pilinis, C. (1990). Derivation and Numerical Solution of the Species Mass Distribution Equation for Multicomponent Particulate Systems, *Atmos Environ.* 24A(7):1923–1928.

Sandu, A. (2002). A Newton-Cotes Quadrature Approach for Solving the Aerosol Coagulation Equation, *Atmos. Environ.* 36:583–589.

Sandu, A., and Borden, C. T. (2003). A Framework for the Numerical Treatment of Aerosol Dynamics, *Applied Numer. Math.* 45:475–497.

Sandu, A., Daescu, D., and Carmichael, G. R. (2003). Direct and Adjoint Sensitivity Analysis of Chemical Kinetic Systems with KPP: I—Theory and Software Tools, *Atmos. Environ.* 37:5083–5096.

Sandu, A. (2004). Piecewise Polynomial Solutions of Aerosol Dynamics. Submitted.

Sandu, A., Daescu, D., Chai, T., Carmichael, G. R., Seinfeld, J. H., Hess, P. G., and Anderson, T. L. (2004). Computational Aspects of 4D-Var Chemical data Assimilation in Atmospheric Models. Darema, F. (ed.), Dynamic Data Driven Applications Systems, Kluwer Academic Publishers.

Sandu, A., Daescu, D., Carmichael, G. R., and Chai, T. (2005). Adjoint Sensitivity Analysis of Regional Air Quality Models. *J. Computational Physics* 204:222–252.

Seinfeld, J. H., and Pandis, P. N. (1997). *Atmospheric Chemistry and Physics. From Air Pollution to Climate Change*. John Wiley & Sons, New York.

Tsang, H., and Hippe, J. M. (1988). Asymptotic Behavior of Aerosol Growth in the Free Molecule Regime, *Aerosol Sci. Technol.* 8:265–278.

Zhang, Y., Seinfeld, J. H., Jacobson, M. Z., and Binkowski, F. S. (1999). Simulation of Aerosol Dynamics: A Comparative Review of Algorithms Used in Air Quality Models, *Aerosol Sci. Technol.* 31:487–514.

APPENDIX A. DERIVATION OF THE CONTINUOUS ADJOINT OF THE DYNAMIC EQUATION IN NUMBER DENSITY FORMULATION

In this section we consider the general dynamic equation in number density formulation, with particle mass the independent variable:

$$
\begin{aligned}
\frac{\partial n(m, t)}{\partial t} = & -\frac{\partial}{\partial m}[I(m, t)n(m, t)] \\
& + \frac{1}{2} \int_0^m \beta(m', m - m')n(m', t)n(m - m', t)\, dm' \\
& - n(m, t) \int_0^\infty \beta(m, m')n(m', t)\, dm' + S(m, t) \\
& - L(m, t)n(m, t). \quad\quad\quad\quad\quad\quad [A.1]
\end{aligned}
$$

where $S(m, t)$ is any source of particles of mass m and $L(m, t)n(m, t)$ is the first-order rate of removal of particles of mass m. The cost functional to be minimized is expressed in the general, integral form

$$
\mathcal{J} = \int_{t^0}^T \int_0^\infty J_0(n(m, t))\, dm\, dt. \quad\quad [A.2]
$$

We will use the Lagrangian multiplier method to derive the continuous adjoint equations and sensitivities with respect to all parameters. We define the cost functional as

$$
\begin{aligned}
\mathcal{J} = & \int_{t^0}^T \int_0^\infty J_0(n(m, t))\, dm\, dt - \int_{t_0}^T \int_0^\infty \lambda(m, t)(LHS_n \\
& - RHS_n)\, dm\, dt
\end{aligned} \quad\quad [A.3]
$$

Here LHS_n and RHS_n refers to the left side and right side of Equation (A.1), respectively. J_0 is the local cost functional component. We take variation of the functional (A.3) and obtain

$$
\begin{aligned}
\delta\mathcal{J} = & \int_{t^0}^T \int_0^\infty \frac{\partial J_0}{\partial n} \delta n(m, t)\, dm\, dt \\
& - \int_{t^0}^T \int_0^\infty \delta\lambda(m, t)(LSH_n - RHS_n)\, dm\, dt \\
& - \int_{t^0}^T \int_0^\infty \lambda(m, t)(\delta LSH_n - \delta RHS_n)\, dm\, dt.
\end{aligned}
$$

By exapnding the above equation, integrating by parts, and rearrranging, the variation $\delta\mathcal{J}$ can be written as:

$$\delta\mathcal{J} = \int_{t_0}^{T}\int_0^{\infty} \delta n(m,t)\mathcal{A}\,dm\,dt \qquad [\text{A.4a}]$$

$$+ \int_0^{\infty}\int_0^{\infty} \delta\beta(m,m')\mathcal{B}dm'dm \qquad [\text{A.4b}]$$

$$- \int_{t_0}^{T}\int_0^{\infty} \delta\lambda(m,t)(LHS_n - RHS_n)\,dm\,dt$$

$$- \int_0^{\infty} \lambda(m,t)\delta n(m,t)|_{t_0}^{T}dm \qquad [\text{A.4c}]$$

$$+ \int_{t_0}^{T}\int_0^{\infty} \delta I(m,t)\frac{\partial\lambda(m,t)}{\partial m}n(m,t)\,dm\,dt \quad [\text{A.4d}]$$

$$- \int_{t_0}^{T}(\delta I(m,t)n(m,t)\lambda(m,t)$$
$$+ \delta n(m,t)I(m,t)\lambda(m,t))|_0^{\infty}dt \qquad [\text{A.4e}]$$

$$+ \int_{t^0}^{T}\int_0^{\infty} \delta S(m,t)\lambda(m,t)\,dm\,dt \qquad [\text{A.4f}]$$

$$- \int_{t^0}^{T}\int_0^{\infty} \delta L(m,t)\lambda(m,t)n(m,t)\,dm\,dt \quad [\text{A.4g}]$$

where

$$\mathcal{A} = \frac{\partial J_0}{\partial n} + \frac{\partial\lambda(m,t)}{\partial t} + \frac{\partial\lambda(m,t)}{\partial m}I(m,t)$$
$$+ \int_m^{\infty} \lambda(m',t)\beta(m,m'-m)n(m'-m,t)\,dm'$$
$$- \int_0^{\infty} \lambda(m,t)\beta(m,m')n(m',t)\,dm'$$
$$- \int_0^{\infty} \lambda(m',t)n(m',t)\beta(m',m)\,dm' - \lambda(m,t)L(m,t)$$

and

$$\mathcal{B} = \int_{t_0}^{T}\left(-\lambda(m,t)+\frac{1}{2}\lambda(m+m',t)\right)n(m,t)n(m',t)\,dt \quad [\text{A.5}]$$

If the Lagrange multiplier λ is chosen as the solution of the *adjoint equation*

$$\frac{\partial\lambda(m,t)}{\partial t} = -\frac{\partial\lambda(m,t)}{\partial m}I(m,t)$$
$$- \int_m^{\infty} \lambda(m',t)\beta(m,m'-m)n(m'-m,t)\,dm'$$
$$+ \lambda(m,t)\int_0^{\infty} \beta(m,m')n(m',t)\,dm'$$
$$+ \int_0^{\infty} \lambda(m',t)n(m',t)\beta(m',m)\,dm'$$
$$+ \lambda(m,t)L(m,t) - \frac{\partial J_0}{\partial n}$$

then the term $\mathcal{A} = 0$.

To deal with the terms $-\int_0^{\infty}\lambda(m,t)\delta n(m,t)|_{t_0}^{T}dm$, $\int_{t_0}^{T}\delta I(m,t)n(m,t)\lambda(m,t)|_0^{\infty}dt$ and $\int_{t_0}^{T}\delta n(m,t)I(m,t)\lambda(m,t)|_0^{\infty}dt$, we impose the following boundary and initial conditions

$$\lambda(m,t=T) = 0 \qquad [\text{A.6a}]$$
$$\lambda(m=0,t) = 0 \qquad [\text{A.6b}]$$
$$n(m=+\infty,t) = 0 \Rightarrow \delta n(m=+\infty,t) = 0. \quad [\text{A.6c}]$$

Then the variation $\delta\mathcal{J}$ simplifies to

$$\delta\mathcal{J} = \int_0^{\infty} \delta n(m,t_0)\cdot\lambda(m,t_0)\,dm$$
$$+ \int_0^{\infty}\int_0^{\infty} \delta\beta(m,m')\cdot\mathcal{B}dm'dm$$
$$+ \int_{t_0}^{T}\int_0^{\infty} \delta I(m,t)\cdot\frac{\partial\lambda(m,t)}{\partial m}n(m,t)\,dm\,dt$$
$$+ \int_{t^0}^{T}\int_0^{\infty} \delta S(m,t)\cdot\lambda(m,t)\,dm\,dt$$
$$- \int_{t^0}^{T}\int_0^{\infty} \delta L(m,t)\cdot\lambda(m,t)n(m,t)\,dm\,dt$$

Therefore the values of the functional sensitivities are:

$$\nabla_{n^0}\mathcal{J} = \lambda(m,t_0) \qquad [\text{A.7a}]$$
$$\nabla_{\beta(m,m')}\mathcal{J} = \int_{t_0}^{T}\left(-\lambda(m,t)\right.$$
$$\left.+ \frac{1}{2}\lambda(m+m',t)\right)n(m,t)n(m',t)\,dt \quad [\text{A.7b}]$$
$$\nabla_{I(m,t)}\mathcal{J} = \frac{\partial\lambda(m,t)}{\partial m}n(m,t) \qquad [\text{A.7c}]$$
$$\nabla_{S(m,t)}\mathcal{J} = \lambda(m,t) \qquad [\text{A.7d}]$$
$$\nabla_{L(m,t)}\mathcal{J} = -\lambda(m,t)n(m,t) \qquad [\text{A.7e}]$$

APPENDIX B. DERIVATION OF THE CONTINUOUS ADJOINT OF THE DYNAMIC EQUATION IN MASS DENSITY FORMULATION

We consider the dynamic equation in mass density for components $i = 1, 2, \ldots, r$:

$$\frac{\partial q_i(m,t)}{\partial t} = H_i(m,t)q(m,t) - \frac{\partial}{\partial m}[mq_iH]$$
$$+ \int_0^m \beta(m',m-m')q_i(m',t)\frac{q(m-m',t)}{m-m'}dm'$$
$$- q_i(m,t)\int_0^{\infty} \beta(m,m')\frac{q(m',t)}{m'}dm'$$
$$+ m_iS(m,t) - L(m,t)q_i(m,t). \qquad [\text{B.1}]$$

where the source term $m_iS(m,t)$ and the removal term $L(m,t)q_i(m,t)$ have the same units. $m = \sum_{i=1}^{r} m_i$, $q(m,t) = \sum_{i=1}^{r} q_i(m,t)$, $H(m,t) = \sum_{i=1}^{r} H_i(m,t)$.

We use the Lagrange multiplier method to derive the continuous adjoint equation considering perturbations in all parameters. The cost functional is defined as

$$\mathcal{J} = \int_{t^0}^{T} \int_{0}^{\infty} J_0(q_1(m,t), q_2(m,t), \ldots, q_n(m,t)) \, dm \, dt$$

$$- \sum_{i=1}^{r} \int_{t_0}^{T} \int_{0}^{\infty} \lambda_i(m,t)(LHS_{qi} - RHS_{qi}) \, dm \, dt \quad [B.2]$$

Here LHS_{qi} and RHS_{qi} refers to the left side and right side of Equation(B.1), respectively. J_0 is the local cost functional component.

Taking the variation of Equation (B.2) gives

$$\delta\mathcal{J} = \int_{t^0}^{T} \int_{0}^{\infty} \sum_{i=1}^{r} \frac{\partial J_0}{\partial q_i} \delta q_i(m,t) \, dm \, dt$$

$$- \int_{t^0}^{T} \int_{0}^{\infty} \sum_{i=1}^{r} \delta\lambda_i(m,t)(LSH_{q_i} - RHS_{qi}) \, dm \, dt$$

$$- \int_{t^0}^{T} \int_{0}^{\infty} \sum_{i=1}^{r} \lambda_i(m,t)\delta(LSH_{q_i} - RHS_{qi}) \, dm \, dt \quad [B.3]$$

After integrating by parts and rearranging $\delta\mathcal{J}$ can be written in a compact form:

$$\delta\mathcal{J} = \sum_{i=1}^{r} \int_{t_0}^{T} \int_{0}^{\infty} \delta q_i(m,t) A_i \, dm \, dt$$

$$+ \int_{0}^{\infty} \int_{0}^{\infty} \delta\beta(m,m')\mathcal{B} \, dm' \, dm$$

$$+ \sum_{i=1}^{r} \int_{0}^{\infty} \delta q_i(m,t_0)\lambda_i(m,t_0) \, dm$$

$$+ \sum_{i=1}^{r} \int_{t_0}^{T} \int_{0}^{\infty} \delta H_i(m,t) \Bigg[\lambda_i(m,t)q(m,t)$$

$$+ \sum_{j=1}^{n} m q_j(m,t)\frac{\partial\lambda_j(m,t)}{\partial m} \Bigg] dm \, dt$$

$$+ \sum_{i=1}^{r} \int_{t_0}^{T} \int_{0}^{\infty} \delta S(m,t)\lambda_i(m,t)m_i \, dm \, dt$$

$$- \sum_{i=1}^{r} \int_{t_0}^{T} \int_{0}^{\infty} \delta L(m,t)\lambda_i(m,t)q_i(m,t) \, dm \, dt \quad [B.4]$$

where

$$A_i = \frac{\partial J_0}{\partial q_i} + \frac{\partial\lambda_i(m,t)}{\partial t}$$

$$+ \sum_{j=1}^{n} \lambda_j(m,t)H_j(m,t) + mH(m,t)\frac{\partial\lambda_i(m,t)}{\partial m}$$

$$+ \int_{0}^{\infty} \frac{\beta(m,m')}{m'}[\lambda_i(m+m',t) - \lambda_i(m,t)]q(m',t) \, dm'$$

$$+ \int_{0}^{\infty} \frac{\beta(m',m)}{m} \sum_{j=1}^{n}[\lambda_j(m+m',t)$$

$$- \lambda_j(m',t)]q_j(m',t) \, dm' - \lambda_i(m,t)L(m,t)$$

and

$$\mathcal{B} = \sum_{i=1}^{r} \int_{t_0}^{T} \Bigg[\lambda_i(m+m',t)q_i(m',t)\frac{q(m,t)}{m}$$

$$- \lambda_i(m,t)q_i(m,t)\frac{q(m',t)}{m'} \Bigg] dt \quad [B.5]$$

Imposing $A_i = 0$ leads to the adjoint equation

$$\frac{\partial\lambda_i(m,t)}{\partial t} = -\sum_{j=1}^{n} \lambda_j(m,t)H_j(m,t) - mH(m,t)\frac{\partial\lambda_i(m,t)}{\partial m}$$

$$- \int_{0}^{\infty} \frac{\beta(m,m')}{m'}[\lambda_i(m+m',t) - \lambda_i(m,t)]$$

$$\times q(m',t) \, dm' - \int_{0}^{\infty} \frac{\beta(m',m)}{m} \sum_{j=1}^{n}[\lambda_j(m+m',t)$$

$$- \lambda_j(m',t)]q_j(m',t) \, dm' + \lambda_i(m,t)L(m,t)$$

$$- \frac{\partial J_0}{\partial q_i}(m,t)$$

Then

$$\delta\mathcal{J} = \sum_{i=1}^{r} \int_{0}^{\infty} \lambda_i(m,t_0)\delta q_i(m,t_0)) \, dm$$

$$- \int_{0}^{\infty} \int_{0}^{\infty} \delta\beta(m,m')\mathcal{B} \, dm' \, dm$$

$$+ \sum_{i=1}^{r} \int_{t_0}^{T} \int_{0}^{\infty} \Bigg[\lambda_i(m,t)q(m,t)$$

$$+ \sum_{j=1}^{n} m q_j(m,t)\frac{\partial\lambda_j(m,t)}{\partial m} \Bigg] \delta H_i(m,t) \, dm \, dt$$

$$+ \sum_{i=1}^{r} \int_{t_0}^{T} \int_{0}^{\infty} \lambda_i(m,t)m_i\delta S(m,t) \, dm \, dt$$

$$- \sum_{i=1}^{r} \int_{t_0}^{T} \int_{0}^{\infty} \lambda_i(m,t)q_i(m,t)\delta L(m,t) \, dm \, dt$$

Therefore,

$$\nabla_{q_i^0}\mathcal{J} = \lambda_i(m,t_0) \quad [B.6a]$$

$$\nabla_{\beta(m,m')}\mathcal{J} = \sum_{i=1}^{r} \int_{t_0}^{T} \Bigg[\lambda_i(m+m',t)q_i(m',t)\frac{q(m,t)}{m}$$

$$- \lambda_i(m,t)q_i(m,t)\frac{q(m',t)}{m'} \Bigg] dt \quad [B.6b]$$

$$\nabla_{H_i(m,t)} \mathcal{J} = \lambda_i(m,t)q(m,t)$$

$$+ \sum_{j=1}^{n} mq_j(m,t)\frac{\partial \lambda_j(m,t)}{\partial m} \qquad [\text{B.6c}]$$

$$\nabla_{S(m,t)} \mathcal{J} = \sum_{i=1}^{r} m_i \lambda_i(m,t) \qquad [\text{B.6d}]$$

$$\nabla_{L(m,t)} \mathcal{J} = -\sum_{i=1}^{r} \lambda_i(m,t)q_i(m,t) \qquad [\text{B.6e}]$$

APPENDIX C. DERIVATION OF THE DISCRETE ADJOINT OF THE DYNAMIC EQUATION

For an ordinary differential equation

$$y' = f(t,y), \quad t^0 \le t \le T, \quad y(t^0) = y^0,$$

and a cost functional

$$J = \int_{t^0}^{T} J_0(y(t))\,dt$$

the continuous adjoint equation reads

$$\lambda' = -f_y^T(t,y) - \frac{\partial J_0}{\partial y}(t), \quad t^0 \le t \le T, \quad \lambda(T) = 0.$$

Here f_y denotes the Jacobian of f with respect to y. Note that the adjoint equation depend on the state variable y of the forward equation.

For the Forward Euler discretization and a discrete in time cost function

$$y^{k+1} = y^k + \Delta t f(t^k, y^k), \quad J = \sum_{k=1}^{N} J_0^k(y^k),$$

the discrete adjoint equation is obtained from the interpretation of adjoint variables as derivatives of the cost functional and the application the transposed chain rule

$$\lambda^k = \left(\frac{\partial J}{\partial y^k}\right)^T$$

$$\frac{\partial J}{\partial y^k} = \sum_{j=k}^{N} \frac{\partial J_0^j(y^j)}{\partial y^k} = \frac{\partial J_0^k(y^k)}{\partial y^k} + \sum_{j=k+1}^{N} \frac{\partial J_0^j(y^j)}{\partial y^k}$$

$$= \frac{\partial J_0^k(y^k)}{\partial y^k} + \sum_{j=k+1}^{N} \frac{\partial J_0^j(y^j)}{\partial y^{k+1}}\left(\frac{\partial y^{k+1}}{\partial y^k}\right)$$

$$= \frac{\partial J_0^k(y^k)}{\partial y^k} + \frac{\partial J}{\partial y^{k+1}}\left(\frac{\partial y^{k+1}}{\partial y^k}\right)$$

$$= \frac{\partial J_0^k(y^k)}{\partial y^k} + \frac{\partial J}{\partial y^{k+1}}(I + \Delta t f_y(t^k, y^k))$$

$$\lambda^k = \lambda^{k+1} + \Delta t f_y^T(t^k, y^k)\lambda^{k+1} + \left(\frac{\partial J_0^k(y^k)}{\partial y^k}\right)^T \delta_k^{\text{obs}}.$$

In the above derivation we used the causality, that is, the fact that y^k does not depend on y^j for $j < k$. In practice the times where the cost functional is defined are multiples of the discretization time step, and the factor δ_k^{obs} equals 1 if t^k is a time used in the definition of the cost functional, and equals 0 otherwise.

The numerical experiments presented in this paper use the two stage, second order, strongly stable, explicit Runge Kutta method

$$y^A = y^k + \Delta t f(t^k, y^k)$$
$$y^{k+1} = \frac{1}{2}(2y^A + \Delta t f(t^{k+1}, y^A)). \qquad [\text{C.1}]$$

The discrete adjoint can be derived using an approach similar to the one used for Forward Euler

$$\lambda^A = \lambda^{k+1} + \frac{1}{2}\Delta t f_y^T(t^{k+1}, y^A)\lambda^{k+1}$$
$$\lambda^k = \lambda^A + \Delta t f_y^T(t^k, y^k)\lambda^A$$

The discrete adjoints for the particle dynamic equation are obtained by applying this formulation with the Jacobian f_n^T of the semidiscrete Equation (3)

$$f_n^T(t,n) = G^T(t) + F_n^T(n(t)) + [(B + B^T) \times n(t)]A^{-T} - L(t),$$

where

$$\{[(B + B^T) \times n(t)]\}_{i,j} = \{n^T(t)(B^i + (B^i)^T)\}_j.$$

APPENDIX D. DERIVATION OF THE DISCRETE ADJOINT OF THE SEMI-IMPLICIT METHOD

In this section we derive the discrete adjoint of the popular semi-implicit method for solving the coagulation equation. In this approach the particles are divided in s bins, and all particles in bin i have the same volume v_i. The coagulation of particles of volumes v_i and v_j forms particles of volume $V_{i,j} = v_i + v_j$. These particles are distributed to the nearest bins, a fraction $f_{i,j,k}$ going to bin k, and the reminder going to bin $k-1$. The fraction coefficients are defined as:

$$f_{i,j,k} = \begin{cases} \dfrac{v_{k+1} - V_{i,j}}{v_{k+1} - v_k}\dfrac{v_k}{V_{i,j}} & v_k \le V_{i,j} < v_{k+1} \quad k < N_B \\ 1 - f_{i,j,k-1} & v_{k-1} < V_{i,j} < v_k \quad k > 1 \\ 1 & V_{i,j} \ge v_k \qquad\quad k = N_B \\ 0 & \text{all other cases} \end{cases}$$

The number concentration in bin k at time t^ℓ is denoted by n_k^ℓ. The concentration changes during one time step $t^{\ell+1} = t^\ell + \delta t$

by the formula

$$n_k^{\ell+1} = \frac{n_k^{\ell} + (\Delta t/v_k)\sum_{j=1}^{k}\sum_{i=1}^{k-1} f_{i,j,k}\, v_i\, \beta_{i,j}\, n_i^{\ell+1}\, n_j^{\ell}}{1 + \Delta t \sum_{j=1}^{s}(1 - f_{k,j,k})\beta_{k,j} n_j^{\ell}}. \quad [D.1]$$

This equation is solved iteratively for $k = 1 \ldots s$.

To derive the discrete adjoint consider again the transposed chain rule

$$\lambda^{\ell} = \left(\frac{\partial J}{\partial n^{\ell}}\right)^{T},$$

$$\frac{\partial J}{\partial n^{\ell}} = \frac{\partial J}{\partial n^{\ell+1}}\frac{\partial n^{\ell+1}}{\partial n^{\ell}} \Rightarrow \lambda^{\ell} = \left(\frac{\partial n^{\ell+1}}{\partial n^{\ell}}\right)^{T}\lambda^{\ell+1} = S^{T}\lambda^{\ell+1}. \quad [D.2]$$

From (D.1), we can derive that

$$\frac{\partial n_k^{\ell+1}}{\partial n_p^{\ell}} = \delta_{k,p} + \frac{\Delta t}{v_k}\sum_{j=1}^{k}\sum_{i=1}^{k-1} f_{i,j,k} v_i \beta_{i,j}\frac{\partial n_i^{\ell+1}}{\partial n_p^{\ell}} n_j^{\ell}$$

$$+ \frac{\Delta t}{v_k}\sum_{i=1}^{k-1} f_{i,p,k} v_i \beta_{i,p} n_i^{\ell+1}$$

$$- \Delta t \sum_{j=1}^{s}(1 - f_{i,j,k})\beta_{k,j} n_j^{\ell}\frac{\partial n_k^{\ell+1}}{\partial n_p^{\ell}}$$

$$- \Delta t(1 - f_{k,p,k})\beta_{k,p} n_k^{\ell+1} \quad [D.3]$$

Introduce the notations

$$S_{k,\ell} = \frac{\partial n_k^{\ell+1}}{\partial n_p^{\ell}}$$

$$C_{k,p} = \frac{\Delta t}{v_k}\sum_{i=1}^{k-1} f_{i,p,k} v_i \beta_{i,p} n_i^{\ell+1} - \Delta t(1 - f_{k,p,k})\beta_{k,p} n_k^{\ell+1}$$

$$D_{k,i} = \begin{cases} -\dfrac{\Delta t}{v_k}\sum_{j=1}^{k} f_{i,j,k} v_i \beta_{i,j} n_j^{\ell} & \text{for } i = 1 \ldots k-1 \\[2ex] 1 + \Delta t \sum_{j=1}^{s}(1 - f_{k,j,k})\beta_{k,j} n_j^{\ell} & \text{for } i = k \\[2ex] 0 & \text{for } i = k+1 \ldots s \end{cases}$$

The Equation (D.3) can be written as

$$D_{k,k}S_{k,p} = \delta_{k,p} - \sum_{i=1}^{k-1} D_{k,i}S_{i,p} + C_{k,p} \Rightarrow DS = I + C. \quad [D.4]$$

From (D.2) and (D.4)

$$\lambda^{\ell} = S^{T}\lambda^{\ell+1} = (I + C^{T})D^{-T}\lambda^{\ell+1},$$

or equivalently

$$D^{T} z = \lambda^{\ell+1}, \quad \lambda^{\ell} = z + C^{T} z.$$

Since D^{T} is an upper triangular matrix the system in z can be solved by a backwards substitution, in the same fashion the forward equation is solved. The complete adjoint formula is therefore

$$z_k = \frac{\lambda_k^{\ell+1} - \sum_{r=k+1}^{s}\left((\Delta t/v_r)\sum_{j=1}^{r} f_{k,j,r} v_k \beta_{k,j} n_j^{\ell}\right) z_r}{1 + \Delta t \sum_{j=1}^{s}(1 - f_{k,j,k})\beta_{k,j} n_j^{\ell}}$$

$$\text{for } k = s, s-1 \ldots 1 \quad [D.5]$$

$$\lambda_p^{\ell} = z_p + \sum_{k=1}^{s}\left((\Delta t/v_k)\sum_{i=1}^{k-1} f_{i,p,k} v_i \beta_{i,p} n_i^{\ell+1}\right.$$

$$\left. - \Delta t(1 - f_{k,p,k})\beta_{k,p} n_k^{\ell+1}\right) z_k \quad \text{for } p = 1 \ldots s.$$