Fitting Microphysical Observations of Nonsteady Convective Clouds to a Numerical Model: An Application of the Adjoint Technique of Data Assimilation to a Kinematic Model

JOHANNES VERLINDE* AND WILLIAM R. COTTON

Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado

(Manuscript received 23 November 1992, in final form 2 April 1993)

ABSTRACT

Rapid advances in the quality and quantity of atmospheric observations have placed a demand for the development of techniques to assimilate these data sources into numerical forecasting models. Four-dimensional variational assimilation is a promising technique that has been applied to atmospheric and oceanic dynamical models, and to the retrieval of three-dimensional wind fields from single-Doppler radar observations.

This study investigates the feasibility of using space–time variational assimilation for a complex discontinuous numerical model including cloud physics. Two test models were developed: a one-dimensional and a two-dimensional liquid physics kinematic microphysical model. These models were used in identical-twin experiments, with observations taken intermittently. Small random errors were introduced into the observations. The retrieval runs were initialized with a large perturbation of the observation run initial conditions.

The models were able to retrieve the original initial conditions to a satisfactory degree when observations of all the model prognostic variables were used. Greater overdetermination of the degrees of freedom (the initial condition being retrieved) resulted in greater improvement of the errors in the observations of the initial conditions, but at a rapid increase in computational cost. Experiments where only some of the prognostic variables were observed also improved the initial conditions, but at a greater cost. To substantially improve the first guess of the field not observed, some spot observations are needed.

The proper scaling of the variables was found to be important for the rate of convergence. This study suggests that scaling factors related to the error variance of the observations give good convergence rates.

To show how this technique can be used when observations are general functions of the prognostic variables of the model (e.g., reflectivity or liquid water path), a form is derived that shows that this can be accomplished. This is considered to be an advantage of this technique over other assimilation techniques, since it is particularly suitable to remote-sensing systems where only integral parameters or derivatives of model prognostic variables are observed.

1. Introduction

In our quest to gain an integrated understanding of precipitating weather systems, methods must be found to characterize both temporally and spatially the phase and properties of hydrometeors throughout entire cloud systems. Optimally, these methods will represent the latest experimental, theoretical, and observational knowledge. A method is investigated where the theoretical and experimental knowledge of liquid phase precipitation processes, as is represented in a numerical model, can be used to assimilate typical observations (radar, aircraft penetrations) of those processes.

Radar data are most often used to infer microphysical information directly from the radar observables (Gunn et al. 1954; Atlas 1957; Austin and Wexler 1957; Batten 1963; Bringi et al. 1986; Vivekanadan et al. 1990). The early work mostly tried to fit conceptual models to radar-observed reflectivity patterns. The recent work uses advances in radar hardware and theoretical work in radiative transfer to infer microphysical characteristics of the scatterers from the observed elements of the scattering matrix. Early quantitative work included efforts to determine precipitation efficiency over barriers (Elliot and Hovind 1964). In that work an orographic flow model and a water balance equation were used to determine precipitation efficiency based on upstream rainsonde, cloud water content, and precipitation observations. In the early 1980s Rutledge and Hobbs (1983, 1984), followed by Ziegler (1985, 1988), fitted microphysical numerical models to Doppler radar–observed wind fields. A similar approach was taken in both studies. They employed a detailed diagnostic numerical kinematical cloud model based on the bulk parameterized continuity equations.
for heat and water substance, allowing both frozen phase and liquid phase physics. Steady state was assumed, and the constant kinematic field derived from multiple Doppler radars at a given time was taken as the flow field. The model was then integrated forward in time until the modeled fields achieved steady state. These diagnosed fields were then compared to the radar reflectivity or aircraft-observed fields for validation. Thus, the reflectivity observations are not used as an information source, rather only as a validating source.

Hauser and Amayenc (1986) developed a variational adjustment scheme for the retrieval of cloud water and water vapor content, where the radar-observed radar reflectivity was used as an input into the scheme rather than for verification. Similar to the previous studies, they assumed a steady-state flow field derived from multiple Doppler radars. The numerical model that they fitted to the observations was based on the continuity equation for total water content. It was assumed that rainfall content could be deduced from the radar reflectivity and that temperature and pressure fields could be retrieved from a dynamic retrieval method (Roux et al. 1984). This model was used to study cloud water and water vapor content in a tropical squall line. In a follow-up study, Hauser et al. (1988) modified their approach somewhat, and combined it in an iterative scheme with the dynamical retrieval routine. In that study they used the reflectivity to validate rather than force their model.

Verlinde and Cotton (1990) investigated the importance of the steady-state assumption in application of the kinematic models to situations where that assumption is clearly violated—for example, a developing cumulus cloud. They performed an observing system simulation experiment (OSSE) for a convective cloud, and used algorithms of the Rutledge–Ziegler type to retrieve the microphysical fields. Their study showed that when Rutledge–Ziegler algorithms are applied to cases in which the steady-state assumption is violated the microphysical and thermal fields are overestimated or underestimated and the evolution of the precipitation fields may be misrepresented. In nonsteady cases, the time history of the cloud is important, and full use of all observations (radar reflectivity and radar-derived wind fields) is required if these algorithms are to be used. Their study thus illustrated the need to expand the current algorithms into time–space algorithms.

Four-dimensional variational (4DVAR) assimilation has become an area of intensive research and experimentation since the early papers of Lewis and Derber (1985), Le Dimet and Talagrand (1986), and Thacker and Long (1988). Applications have mostly been for the initialization of synoptic-scale dynamic models with few parameterizations of physical processes. Work is currently under way (Kapitza 1991; Sun et al. 1991; Sun 1992) on the problem of fitting nonhydrostatic models to single-Doppler radar–observed radial wind fields. This work holds the promise that continuous nonsteady three-dimensional wind, pressure, and buoyancy fields can be retrieved from single- (or dual-) Doppler radar observations.

The object of this study is to investigate the feasibility of using variational assimilation to incorporate observations of the physical processes involved in precipitation formation into a kinematic microphysical model. Since the processes modeled are discontinuous, this paper investigates how the variational assimilation algorithm will perform when the derivative of the model formally does not exist.

The means through which this was accomplished is a series of identical-twin experiments. Two kinematic bulk parameterized models of the liquid phase physics were derived: a one-dimensional model and a two-dimensional model. "Observation" datasets were created for the one-dimensional model by running the kinematic model forward in time using a prescribed vertical velocity, while for the two-dimensional model, the RAMS cloud model (Tripoli and Cotton 1982, 1989)—using the microphysical package described in this paper—was used to simulate the life cycle of a convective storm. The dynamic fields (velocity, pressure, and buoyancy fields) were saved at each time step. An iterative algorithm was derived through which the model could be fitted to observations. The control variables selected for this study were the variables describing the state of the model at the initial time, but they may also have been other elements of the model, such as boundary conditions or parameterization constants. This algorithm was used to retrieve the original initial condition used in the observation run, starting from a constant initial condition. Factors influencing the rate of convergence for this complex discontinuous model were investigated. This work should be considered a preliminary study to investigate whether this is a feasible approach to the problem, before using real observations.

2. The variational problem

The problem may be viewed as that of finding the initial conditions that minimize a (scalar) cost function or distance between the model and the observations over the time interval considered. With boundaries given and constant parameterization coefficients, the cost function is only a function of the initial conditions. The minimization problem then requires the determination of the cost function (through the forward integration of the model) and the gradient of the cost function with respect to the initial conditions (through the integration of the adjoint model).

It is necessary to derive the adjoint model from the discrete version of the forward model. Assume that the
numerical model that explicitly integrates a set of differential equations can be written in synthetic form as

$$x_k = f(x_{k-1}, t) \quad k = 0, 1, 2, 3, \cdots, M,$$

(2.1)

where $x_k \in R^n$ is the state vector, consisting of all the model prognostic variables defined at each grid point. Assume that observations of some function of the state vector $z = z(x)$ are available at $N$ time periods, which, for this derivation, need not be at set intervals. Let the set of observations be given by $o(j) = o^j, j = 0, 1, \cdots, N$. The model times corresponding to the observational times will be indicated by superscripts $k_j$.

Define the cost function as

$$E(x^0) = \frac{1}{2} \sum_{j=0}^{N} [z(x^j) - o^j]^T W[z(x^j) - o^j],$$

(2.2)

where $W$ is a constant weight matrix (in general it may be a full matrix, but in this study it is considered a diagonal matrix) and superscript $T$ denotes a transpose. The gradient of the cost function with respect to the initial conditions is

$$\nabla_{x^0} E(x^0) = \frac{1}{2} \sum_{j=0}^{N} \nabla_{x^0} [z(x^j) - o^j]^T W[z(x^j) - o^j]$$

$$= \frac{1}{2} \sum_{j=0}^{N} \nabla_{x^0} z(x^j) \nabla_{x^0} [z(x^j) - o^j]^T W[z(x^j) - o^j].$$

(2.3)

To obtain a usable expression for $\nabla_{x^0} x^j$, the chain rule must be applied backward to the forward operator of the model, which in this example is just a forward upstream operator,

$$x^j = x^{j-1} + \Delta t f(x^{j-1}),$$

(2.4)

where $\Delta t$ is the time step. Taking the Jacobian of Eq. (2.4) gives

$$\nabla_{x^0} x^j = T_{k_{j-1}} \nabla_{x^0} x^{j-1},$$

(2.5)

where $T_{k_{j-1}} = I + \Delta t \nabla_{x^0} f(x^{j-1})$ is the linearization of the model about the model state at time $t_{k_{j-1}}$, also called the tangent linear matrix. Realizing that $\nabla_{x^0} x^0 = I$, the identity matrix, Eq. (2.5) can be written as

$$\nabla_{x^0} x^j = T_{k_{j-1}} T_{k_{j-2}} \cdots T_{k_1} T_0.$$

(2.6)

Equation (2.3) can then be written as

$$\nabla_{x^0} E(x^0) = y^0$$

$$+ A_{y^0} [y^1 + A_{y^0} [y^2 + \cdots (y^{N-1} + A_{y^0} y^N) \cdots]].$$

(2.7)

Thus, it can be seen that the gradient of the cost function can be expressed as a linear operator. This operator has been called the adjoint model. A simple example of how this operator may be constructed is given in appendix A.

In this derivation it was assumed that the model is perfect. Derber (1989) developed a variational assimilation method that takes the forecast model error into account; thus, this problem can be alleviated in this technique.

3. Procedure

The optimization code described by Buckley and LeNir (1983) was used to minimize the cost function. [See Navon and Legler (1987) or Navon et al. (1992) for a discussion on descent algorithms for large-scale minimization problems in meteorology.] This routine is a mixture between the quasi-Newton and the conjugate-gradient methods, and it makes optimum use of the available declared memory. It starts as a quasi-Newton method and switches to conjugate gradient when memory runs out.

There are different ways to approach the construction of the adjoint model. A commonly used approach is to linearize the numerical code of the forward model line by line, take the transpose of it, and integrate the constructed code from the final to the initial time. This procedure is described in greater detail by Errico and Vukicevic (1992) or Chao and Chang (1992). Another approach is by using Lagrangian multipliers (Thacker and Long 1988). An alternative procedure is to actually calculate the matrices $T_j$ and execute the linear operation given in Eq. (2.8). This last approach was selected for this study. The following are the two advantages of this approach.

1) The construction of the adjoint is greatly simplified. Symbolic computing programs may be used to construct the code for the contribution of each physical process to the various matrix elements.

2) The linear operator Eq. (2.8) can be effectively implemented on vector and parallel machines.

The following are the two disadvantages of this approach.
1) Disk space requirements. This application required 2.5 times the state vector at each time step. Since the study was dealing with discontinuous physics, the simplifying assumption of a slowly varying base state was not possible, and the model had to be linearized about the base-run state vector at each time step. This involved trade-offs between disk space and computation time.

2) Depending on the complexity of the advection and diffusion scheme, the in-core memory requirements may be very expensive. For the second-order forward upstream advection scheme and fixed-coefficient diffusion scheme selected for this study, the matrices $T_i$ required 21 times the state vector size in memory allocation.

The construction of the adjoint model was tested at all levels of code during the development phase. The Taylor expansion test presented by Navon et al. (1992) was used to test the adjoint of the complete model.

Throughout this study it was assumed that the prognostic variables of the model can be observed. The Jacobian $\nabla_{x^0} z(x^0)$ is then equal to the identity matrix.

The form of the matrices depends on the definition of the state vector. Assume there are three prognostic variables in the model with $N$ grid points—$a(x, t) = (a_1, \ldots, a_N)^T$, $b(x, t) = (b_1, \ldots, b_N)^T$, and $c(x, t) = (c_1, \ldots, c_N)^T$—each dependent on the state vector $x$ and time. The state vector can be defined as

$$x = (a_1, \ldots, a_N, b_1, \ldots, b_N, c_1, \ldots, c_N)^T,$$

or as

$$x = (a_1, b_1, c_1, \ldots, a_N, b_N, c_N)^T.$$

The first vector will result in a $3 \times 3$ block matrix, with each block matrix being an $N \times N$ matrix. The single-gridpoint-dependent physics will contribute terms to the diagonals of all the block matrices, depending on how the various prognostic variables are related in the physical process. Advection will produce bands through the diagonal block matrices, while multigridpoint physics will produce bands through all the block matrices.

The second vector will produce a banded matrix. This structure of the matrix makes it easier to compute the gradient, and also is more cost efficient, since it can efficiently be vectorized on vector computers. This is the approach that was selected for this study.

The final comment in this section is a short discussion on the interface between the forward model and the adjoint model. This is regulated through the incorporation of scaling factors. The prognostic variables can vary by as much as 8–10 orders of magnitude. Optimization codes work best when all variables are in the range 0–1. It is therefore desirable to normalize the state vector. For the ease of programming it was decided to keep the forward model in terms of real variables (also the calculation of the matrix elements), and then to correct the gradient for the normalized state vector. Thus, the normalized initial vector passed by the optimization code to the model is multiplied by the normalization factor (weight), the forward integration is performed, and the tangent linear matrix is calculated. The adjoint integration is then performed. The relationship between the gradient calculated from Eq. (2.8) and that for the normalized state vector can be shown to be

$$\nabla_{x^0} E(x^0)|_{\text{norm}} = W \nabla_{x^0} E(x^0).$$

At this stage it would be appropriate to discuss the problems associated with scaling and convergence. Bad scaling of the variables will lead to slow convergence rates. Following Moore (1991) this may be explained as follows. When a certain subset of the unknowns are much smaller (larger) than the remainder of the set, the cost function [Eq. (2.2)] is very flat in certain directions. Consequently, the descent algorithm soon falls into a long flat valley with a slowdown in the rate of convergence. Ideally, one would like the gradient to be such that all variables converge to the solution at about the same rate. The optimum way to scale the system is to scale the gradient vector by the Hessian matrix of the cost function (Thacker 1989). Thacker (1989) discussed ways to compute the Hessian matrix for systems that are in steady state; however, in general it is prohibitively expensive to compute the Hessian. Moore (1991) suggests that the best way to scale the gradient is to multiply it by a diagonal matrix whose elements reflect the relative magnitudes of the unknown variables. Kapitza (1991) suggested that it is not clear a priori what the proper scaling for a particular problem is. As an extension from linear theory, the weights are often chosen as $W = \sigma^{-2}I$, where $\sigma^2$ is the variance of the measurement errors and $I$ is the identity matrix (e.g., Courtier and Talagrand 1990). These matters are investigated in this paper.

4. Description of models

Two kinematic bulk cloud microphysical models were developed to test the feasibility of this approach to fitting observations in time to a time-dependent model. A simple one-dimensional model was developed to test the concepts and to look for problem areas. The physics in the two-dimensional model was more complex (realistic). The extension to a three-dimensional model, though detailed, is straightforward. For simplicity, it was assumed in both applications that the model prognostic variables were observed. This is not a limiting assumption, as is shown in appendix A, but was made to simplify the code. Both these models predicted on liquid water potential temperature $\theta$, total water content $r$, and rainwater content $r_r$, with cloud water content $r_c$ a diagnostic variable. These variables
were chosen as prognostic variables since it is believed that current observational facilities may be able to provide a fair amount of information content. The work of Sun et al. (1991) holds the promise that potential temperature and pressure may be retrieved from radar velocity fields, while multiparameter radars (or even just radar reflectivity) are related to both parameters of the raindrop distribution (see appendix A). Liquid water potential temperature is dependent on the rainwater content and the potential temperature; thus, a radar, or other similar remote sensing instrument, potentially carries sufficient information content to attack this problem. The features common to both models are discussed first, and then specific features to each model. The tangent linear versions of the models are derived in full in Verlinde (1992).

a. Common features

1) Parameterization

It is assumed that cloud and rainwater are distributed according to the gamma distribution,

\[ f(D) = \frac{1}{\Gamma(\nu)} \left( \frac{D}{D_n} \right)^{\nu-1} \frac{1}{D_n} \exp \left( - \frac{D}{D_n} \right), \]

where \( \Gamma \) is the complete gamma function, \( D \) is the diameter, and \( D_n \) is a characteristic diameter.

The characteristic diameter \( D_n \) and the mean diameter \( D_m \) are related to the prognostic variables through the relationship

\[ D_n = \nu D_m \left( \frac{0.1 \rho_f}{\pi \rho_l n} \right)^{1/3}, \]

where \( \rho_l \) is the density of water, and \( r \) and \( n \) are the water content and mixing ratio, respectively. All constants and variables are defined in appendix B. The mass-weighted terminal velocity is given by

\[ V_{Tm} = c_e D_m^{5/6} \frac{\Gamma(p_m + p_v + \nu)}{\Gamma(\nu)}, \]

while the mean terminal velocity is given by

\[ V_T = c_e D_n^{5/6} \frac{\Gamma(p_v + \nu)}{\Gamma(\nu)}. \]

2) Liquid Water Potential Temperature and Total Water Content

The \( \theta_i, r_s, \) and \( r_r \) were chosen as prognostic variables. Cloud water mixing ratio and temperature are diagnosed assuming zero supersaturation. These two variables are interrelated through \( \theta_i \),

\[ T = \left( \frac{p}{p_0} \right)^{R_e} \theta_i \left[ 1 + \frac{\theta_0}{c_p \theta} (r_r + r_s) \right]. \]

\[ r_e = (r_r - r_s - r_r) H(r_r - r_s - r_r), \]

where the \( xH(x) \) is the Heaviside step function and the saturation vapor pressure is given by the formulation due to Murray (1967) as

\[ r_s = 610.78 \exp \left[ 17.27 \left( \frac{T - 273.16}{T - 35.86} \right) \right]. \]

Temperature and cloud water are recovered from a bisection iteration scheme of 11 iterations. Temperature is recovered to within 0.1° accuracy.

3) Sedimentation

Rain fallout is included through Lagrangian advection and subsequent redistribution, such that the change in rainwater mixing ratio due to precipitation is given by

\[ \frac{dr_r}{dt_{\text{prec}}} = - \frac{dV_{Tm} r_r}{dz}. \]

The change in number concentration of raindrops due to precipitation is given by

\[ \frac{dn_r}{dt_{\text{prec}}} = - \frac{dV_T n_r}{dz}. \]

The mean terminal velocity \( V_T \), as opposed to the mass-weighted mean terminal velocity \( V_{Tm} \), was chosen for the fallout of number concentration in an attempt to simulate the observed tendency of precipitation to reduce the mean diameter of the raindrop distribution at a given point.

It is assumed that cloud water has negligible terminal fall speed; thus, precipitation is made up entirely of rain fallout. The change in \( \theta_i \) is then given by Tripoli and Cotton (1981) as

\[ \frac{d\theta_i}{dt_{\text{prec}}} = - \frac{\theta_0}{c_p T \theta} \frac{dr_r}{dt_{\text{prec}}}, \]

while the change in \( r_i \) is just the precipitation flux from the box, or

\[ \frac{dr_i}{dt_{\text{prec}}} = \frac{dr_i}{dt_{\text{prec}}} \].

Here, \( \theta \) is the potential temperature.

b. One-dimensional model specifics

The prognostic variables in this model are \( \theta_i, r_i, \) and rainwater mixing ratio \( r_r \), while temperature and cloud water are diagnosed. It was assumed that the rainwater was distributed in size according to the exponential distribution [Eq. (4.1) with \( \nu = 1 \)]. This is equivalent to the Marshall–Palmer distribution (Marshall and Palmer 1948) when \( N_0 = n_r / D_n \) and \( \Lambda = 1 / D_n \). The cloud water is assumed to be monodispersed.
1) CONTINUITY EQUATIONS

The partial differential equations describing the changes in the prognostic variables are

\[
\frac{\partial q}{\partial t} = -w \frac{\partial q}{\partial z} + S_q, \quad (4.12)
\]

where \( q \) represents any of the three prognostic variables (\( \theta_l, r, \text{ or } r_r \)) and \( S_q \) represents the sources and sinks for each, respectively.

2) PHYSICAL PROCESSES

The physical processes allowed in this model are condensation and evaporation (implicit in \( \theta_l \)), autoconversion of cloud to rain, collection of cloud by rain, and sedimentation of rainwater. The autoconversion rate is taken from Kessler (1969):

\[
\frac{dr_r}{dt}_{\text{auto}} = \alpha \rho (r_c - r_0) H(r_c - r_0), \quad (4.13)
\]

where \( H \) is the Heaviside step function. Collection is taken from the accretion equation as given by Flatau et al. (1989) for the Marshall–Palmer distribution.

\[
\frac{dr_r}{dt}_{\text{coll}} = \frac{\pi}{4} n_c r_c V_T D_{or}^2 \Gamma (p_v + 3). \quad (4.14)
\]

3) NUMERICAL PROCEDURES

Advection was done with the Bott (1989) positive-definite advection scheme. In the simulation the vertical velocity was held constant at \( w = 2.5 \text{ m s}^{-1} \). The model was integrated for 500 time steps, with a time step of 5 s. The model domain consisted of 100 grid points, with a 50-m gridpoint separation. The outflow boundary was an open boundary, while the inflow boundary was held constant. Figure 1 shows the sounding used and the time–height evolution of the rainwater content in the simulation.

\( c \) Two-dimensional model

The prognostic variables in the two-dimensional (2D) model are \( \theta_l, r, r_r, \) and raindrop number concentration \( n_r \). Temperature \( T \), cloud water mixing ratio \( r_c \), and the characteristic diameters of the cloud water \( D_{nc} \) and rainwater distributions \( D_{nr} \) are diagnosed from the prognostic variables. Both the cloud water and rainwater are assumed to be distributed according to Eq. (4.1) with \( \nu = 3 \). A constant cloud-droplet number concentration \( (n_c = 3 \times 10^8 \text{ m}^{-3}) \) is also assumed.

1) CONTINUITY EQUATIONS

The partial differential equations describing the changes in the prognostic variables can symbolically be written as

\[
\frac{\partial q}{\partial t} = -u \frac{\partial q}{\partial x} - w \frac{\partial q}{\partial z} + K_x \frac{\partial^2 q}{\partial x^2} + K_z \frac{\partial^2 q}{\partial z^2} + S_q, \quad (4.15)
\]

where \( q \) represents any of the prognostic variables and \( S_q \) represents the respective sources and sinks. Sedimentation of precipitation is included in the source-sink terms.

2) PHYSICAL PROCESSES

Physical processes allowed in the 2D model are condensation and evaporation, autoconversion of cloud
to rain, collection of cloud by rain, self-collection and breakup of the rainwater, and sedimentation of rain.

For autoconversion, the Berry and Reinhardt (1974) parameterization was adapted. Ziegler (1985) has used this parameterization in a similar warm physics model. The formulation used in this study is

$$\frac{dr_i}{dt}_{\text{auto}} = \max(0, a_i D_{mc}^4 - a_2) \times \max(0, a_3 D_{mc} - a_4)(\rho r_c)^2.$$  \(4.16\)

where $a_i$, $i = 1 \cdots 4$ are constants.

Autoconversion not only adds mass to rain but also alters the number concentration. Still following Berry and Reinhardt (1974), this is written as

$$\frac{dn_r}{dt}_{\text{auto}} = a_n \frac{dr_i}{dt}_{\text{auto}},$$  \(4.17\)

where $a_n$ is a constant.

Accretion is included from Flatow et al. (1989), but a lower limit for mean cloud droplet size was added to account for the drop-off in collection efficiency for small cloud droplets. This introduced a first-order discontinuity in the model. A spline was fitted over an interval spanning the discontinuity such that, at the limits of this interval, the function values and the first and second derivatives were equal to that of the discontinuous function. Accretion was then written as

$$f(r_c) = (r_c - \gamma)^3 \left[ \frac{\beta}{(\beta - \gamma)^3} \frac{\gamma + 2\beta}{(\beta - \gamma)^4} (r_c - \beta) + \frac{3(\gamma + \beta)}{(\beta - \gamma)^5} (r_c - \beta)^2 \right].$$  \(4.19\)

Here, $\gamma$ is the liquid water mixing ratio corresponding to the lower limit $D_\gamma$, and $\beta$ to the upper limit $D_\beta$. The effect of the discontinuity on the assimilation process was tested by increasing and decreasing the interval of the spline.

The formulation for self-collection by rain was taken from Verlinde et al. (1990):

$$\frac{dn_r}{dt}_{\text{self}} = -\frac{\pi}{8\rho} c_v D_{mr}^2 n_r^2 EC,$$  \(4.20\)

where the constant $C$ is

$$C = \sum_{n=0}^{2} \left[ \frac{2}{\nu + n} \Gamma(\eta)_{2} F_{1}(\nu + n, \eta; \nu + n + 1; -1) - \Gamma(\nu + n) \Gamma(\nu + n + 2) \right]$$

$$+ \sum_{n=0}^{2} \left[ \frac{2}{\nu + p_v + n} \Gamma(\eta)_{2} F_{1}(\nu + p_v + n, \eta; \nu + p_v + n + 1; -1) - \Gamma(\nu + p_v + n) \Gamma(\nu + n + 2) \right].$$  \(4.21\)

and

$$\eta = p_v + 2\nu + 2.$$

Here, $_2F_1$ is the Gaussian hypergeometric function. Note that this formulation for the number concentration tendency differs slightly from that in Verlinde et al. (1990), since they derived the mixing ratio tendency.

The collection efficiency $E$ is defined as

$$E = \begin{cases} 1, & D_{mr} < D_b; \\ 2 - \exp[A(D_{mr} - D_b)], & D_{mr} \geq D_b \end{cases}.$$  \(4.22\)

$D_b$ is a cutoff (mean) diameter after which breakup will begin to be effective, and $A$ is a constant (values for $D_b$ and $A$ are given in appendix B). When the mean diameter of the distribution exceeds 600 $\mu$m, breakup is activated and the collection efficiency decreases exponentially from one. It turns negative (increasing the number concentration) when the mean diameter exceeds 900 $\mu$m. The mean diameter will oscillate toward a value of 900 $\mu$m. The equilibrium level for mean diameter was chosen to correspond approximately to the broad peak observed by Zawadzki and de Agostino Antonio (1988).

3) NUMERICAL PROCEDURES

The sequential updating advection scheme with a deformation instability correction was used with second-order fluxes according to Tremback et al. (1987). This advection scheme was selected over the positive-definite scheme used for the one-dimensional model due to the complexity of the adjoint of the positive-definite schemes in two dimensions.

The simulation was run on a $20 \times 20$ grid domain, with a 5-km horizontal gridpoint separation and a 500-m vertical gridpoint separation. Open outflow horizontal boundary conditions were used, while horizontal inflow boundaries were kept at environmental
values. A 5-s time step was used. A wall was used as a top boundary.

For the observation run, the model was initialized horizontally homogeneous with no condensate. The pressure and wind fields from the RAMS simulation were used. The model was integrated for 480 time steps, during which a vigorous single-cell cloud formed. The wind and pressure fields were written to a file at each time step. Figure 2 shows the model prognostic fields at the final time step.

5. Results

a. Introduction

Identical-twin experiments were performed. The models described were initialized and integrated forward in time. Observations were taken at set intervals. Assimilation of observations was attempted over time intervals well into the simulation period. The first results are from experiments where perfect observations were taken.

b. One-dimensional identical-twin experiments

Figure 4a shows the cost function decline as a function of iteration number for three experiments with the one-dimensional model, while Fig. 4b shows the decline of the gradient for the same experiments. The assimilation period used here was from \( t = 1500 \) s to \( t = 2000 \) s (see Figs. 1a,b), where observations were available at \( t = 1500 \) s, \( t = 1750 \) s, and \( t = 2000 \) s. Profiles of constant \( \theta_i, r_i \), and \( r_t \) were chosen as start-up initial conditions for the assimilation cycle. The true initial profiles are presented in Fig. 3. Experiment A had initial start-up conditions of \( \theta_i = 300 \) K, \( r_t = 10 \).
be placed on the assimilation algorithm in regard to the length of the assimilation period. In Fig. 5 the impact of the length of the assimilation period is investigated using the one-dimensional model. All these experiments used the same scaling and initial start-up condition as experiment A, discussed before. In this case the assimilation was done over the time period \( t = 1500 - 2000 \) s (A), \( t = 1000 - 2000 \) s (B), and \( t = 500 - 2000 \) s (C) (see Fig. 1). In each case the observations were taken at the initial, final, and midpoint time. A drop-off in convergence rate can clearly be seen, with higher convergence rates for shorter assimilation periods. The convergence rates for the longer assimilation periods did not significantly differ when more frequent observations were used. This result should be expected.

Fig. 3. True \( \theta_m, r_m, \) and \( r_r \) profiles at the initial time of the assimilation run.

\[ g \text{ kg}^{-1}, \] \[ r_m = 1 \text{ g kg}^{-1}, \] \[ \text{and weights set to } \Theta_m = 100 \text{ K}, \] \[ R_m = 0.04, \] \[ \text{and } R_r = 0.025. \]

The cost function was reduced by about 10 orders of magnitude in 50 iterations, while the gradient was reduced by 4 orders of magnitude. Further examination of the results revealed that the root-mean-square of the error in the initial conditions after 50 iterations was reduced to order \(-5 \) K for \( \theta_m \) and order \(-7 \) g kg\(^{-1}\) for the two mixing ratios, while the maximum error was reduced to order \(-5 \) for all variables. The maximum error dropped to acceptable levels (below 0.1 K or 0.1 g kg\(^{-1}\)) in 30 iterations.

Further insight into how scaling influences the convergence rate can be gained from Fig. 4; experiment C has the same weights as A but a different start-up initial condition (\( \Theta_m = 330 \) K, \( r_m = 8 \) g kg\(^{-1}\), and \( r_r = 0 \) g kg\(^{-1}\)); experiment B has the same startup initial conditions as C but different weights (\( \Theta_m = 200 \) K, \( R_m = 0.04, \) and \( R_r = 0.025 \)). The start-up initial conditions for experiments B and C are farther from the true initial condition; hence, slower convergence rates may be expected. But it can be seen that the scaling that provided good convergence rates for experiment A are not optimum for the different start-up initial conditions of experiments B and C. These results indicate that the common practice of using weights related to the variance does indeed give best convergence. Scaling with the relative magnitudes of the unknown variables as suggested by Moore (1991) resulted in poor convergence rates.

It is hoped that routines like this may be used to assimilate radar-observed data into numerical cloud models. Since radars take finite times scanning a cloud volume, and the time taken for each volume scan involves compromises on the accuracy, constraints may

Fig. 4. (a) Cost function and (b) gradient decline as a function of number of iterations for three experiments with the one-dimensional model. Experiments A and C have the weights set as \( \Theta_m = 100, \) \( R_m = 0.04, \) and \( R_r = 0.025 \), while the third (experiment B) has the weights set to \( \Theta_m = 200, R_m = 0.04, \) and \( R_r = 0.025. \) Experiments B and C have the same start-up initial condition, while experiment A has a different start-up initial condition.
since this is a highly nonlinear model. These models lose their predictability skill over longer periods, resulting in complicated cost functions. [For a more detailed discussion see Sun (1992) and Zou et al. (1992)]. This indicates that faster scan rates, which would provide the same information content in a shorter time period, would be beneficial to algorithms of this type.

Up to now it was assumed that observations of all the prognostic variables are available. This may not always be the case. Figure 6 shows results of two experiments with the one-dimensional model where little or no information on \( r_f \) was available. The start-up initial conditions for both cases were the true initial condition with an offset of 2 K in \( \theta_f \), 2 g kg\(^{-1}\) in \( r_s \), and 1 g kg\(^{-1}\) in \( r_r \). The assimilation was done over the period \( t = 1500-2000 \) s, with observations taken every 50 s. The first experiment (A) has no \( r_f \) observations, while the second (B) has \( r_f \) observations only at \( z = 750 \) and 2500 m at all the observation times. It was found that even though \( r_r \) contributes nothing or very little to the cost function, the weight assigned to it remains important. This is so because there is an impact of the weight assigned to each variable to the gradient through Eq. (3.3). It was found that the principles derived earlier in this section worked fine. As can be expected, the convergence of \( r_f \) was much slower than the convergence of the other two variables that had full observations, but it can be seen that in both experiments the technique was tending toward the original initial condition. It is one of the strengths of the adjoint technique that it is able to provide useful information about something that is not observed at all. It can also be seen that even a small amount of information on \( r_f \) drastically improved the rate of convergence in the early stages but that after about 150 iterations both assimilations were approximately equal.

c. Two-dimensional identical-twin experiments

The experiments with the two-dimensional model were done in a similar fashion. The RAMS model was used to simulate a vigorous convective cloud, using the microphysical parameterization described above. Wind and pressure fields were saved at each time step. These time-varying fields were used to drive the kinematic model. The assimilation was done over a time interval during the growth to mature stage of the convective cloud, 100 time steps after the initialization of the simulation. The assimilation run was initialized with the horizontally homogeneous initialization of the RAMS run, but with constant \( \theta_f, r_s, r_r, \) and \( r_f \) fields in the approximate area where the cloud was observed. Results from the one-dimensional experiments were used to choose scaling factors for the two-dimensional model.

Typical mean errors for the variables were selected to decide which scaling factors to use. The weights were further normalized by the number of grid points. The general weights determined for this study were \( \theta_f = 160, R_s = 0.16, R_r = 0.048, \) and \( N_f = 80,000 \). Figure 7 shows the impact on the convergence rate of changes in the ratio of the scaling factors. Figures 8–10 shed some light on what is going on. These figures show the gradient after the first iteration for the three experiments. It can be seen that changing the \( \theta_f \) weight from 160 to 320 had a minor impact on the overall structure and magnitude of the gradient, while reducing it to 80 re-

---

**Fig. 5.** Cost function decline as a function of number of iterations for three experiments with the one-dimensional model. All the experiments have the weights set as \( \theta_f = 100, R_s = 0.04, \) and \( R_r = 0.025 \). Experiment A assimilates observations over 100 time steps, experiment B assimilates observations over 200 time steps, and experiment C assimilates observations over 300 time steps.

---

**Fig. 6.** Decline of the error in \( r_r \) as a function of the number of iterations for two experiments' one-dimensional model. Both experiments assimilated data over 100 time steps, with observations at every 10th step. In experiment A no observations of \( r_f \) were used, while in experiment B observations at only 2 of the 100 grid points were used. Both the experiments have the weights set as \( \theta_f = 100, R_s = 0.1, \) and \( R_r = 0.025 \).
sulted in significant changes in the magnitude of the gradients, mostly concentrated in a few spots. This resulted in an overshoot in those places, with lower convergence rates.

It was found that the correct scaling between the four prognostic variables was critical to the convergence rate and that even small deviations resulted in significantly lower convergence rates. The physics in this model include processes (e.g., self-collection/breakup) that will rapidly correct invalid physics (e.g., rain distribution with a mean drop diameter of 1 m because the optimization code increased rainwater content significantly while reducing the number concentration at the same time), causing large gradients. It was found during the course of the study that once this happened the algorithm never recovered in a reasonable number of iterations.

A guideline to obtain the correct ratio between the weights is to start with a simulated experiment where the model is used to provide observations. Introduce a very small perturbation of known variance in the start-up initial condition. Use these variances as the initial weights. Under these conditions the optimization algorithm should start in the steep gradient area surrounding the minimum, and should converge. Minor adjustments can be made to the weights to find the optimum ratio. It was found that once the correct ratio for the weights is obtained, multiples of these weights provide good convergence for cases where the start-up initial condition is further removed from the “correct” initial condition. These scaling factors gave good convergence for different start-up initial conditions as long as they all had similar magnitudes in their initial errors. This would indicate that once the algorithm has been constructed and tested, a set of typical weights can be determined that will work in most cases, and where it does not provide good convergence, it would be easy to adjust the weights based on a single run.

Fig. 7. Experiments investigating the effect of different weights on the convergence rate for the two-dimensional model. Experiment A has the weights set to $\theta_i = 160$, $R_e = 0.16$, $R_i = 0.048$, and $N_i = 80,000$, while experiment B has the weight for $\theta_i$ doubled and experiment C has it halved.

One of the objects of this study was to investigate the impact of discontinuities in the forward model on the retrieval process. Equation (4.18) was constructed to test the results of a zeroth-order discontinuity. It is known that there is a cutoff where the collection efficiency of raindrops collecting cloud drops decreases to zero. Equation (4.18) was constructed with this in mind. When the upper and lower limits were the same (discontinuity) there was no convergence. Results of two different experiments are displayed in Fig. 11. It can be noted that as soon as the interval over which the spline was fitted is reduced the convergence rate drops off significantly. When the upper limit was further reduced to 20 $\mu$m, the algorithm no longer converged. This indicates that zeroth-order discontinuities will greatly impact retrieval algorithms, but it also shows how this obstacle may be overcome. The other physical processes, like phase changes, although discontinuous, started gradually such that they can be considered continuous to the numerical model, and therefore, present no problems to the algorithm.

Next some experiments were conducted to investigate the impact of the neglect of some physical processes on the retrieval. Figure 12 compares results from two experiments where first accretion and self-collection were neglected (experiment B), and second, only self-collection was neglected (experiment C), in the standard identical-twin case. In these experiments the observations were created with all the physical processes included, while the retrieval runs were done with the above processes neglected in both the forward and adjoint model. It can be seen that the neglect of some physical processes greatly inhibits the recapture of the original initial condition. This is no surprise, but it should be reflected in attempts to retrieve useful microphysical initial fields from real observations with faulty models. It may be necessary to “tune” the parameterization constants of the model by including them as unknowns in order to improve the fit over an assimilation period. For analysis purposes where time and cost limitations are not as likely to be a serious factor, the model would need to be improved.

Finally, some experiments were performed with this model where random errors with magnitudes up to 0.5 K for $\theta_i$, 0.5 g kg$^{-1}$ for $r_e$, and 1 g kg$^{-1}$ for $r_i$ were added to all the observations. Results after 50 iterations are presented in Table 1. In all cases the minimum was reached long before 50 iterations. All these retrieval runs used the observations at the initial time as the start-up initial conditions. The experiments using 3, 5, and 13 sets of observations, as well as experiment 25a using 25 sets of observations, used the standard set of weights. There was general improvement in the rms
errors for $\theta_l$ and $r_\tau$, while there was a slight deterioration in the rms errors for $r_\tau$ and $n_r$. This can be explained by realizing that observations of $r_\tau$ and $n_r$ exist only where there was precipitation, which covers only a small part of the domain. Over the rest of the domain the error in the initial $r_\tau$ and $n_r$ fields are thus zero, not only reducing the rms error of the observations but also resulting in a much lower contribution by these two variable vectors to the cost function. In minimizing the cost function, the algorithm will thus favor $\theta_l$ and $r_\tau$, reducing their contribution to the cost function at the expense of the other two variables. To investigate if these results could be countered, the weights for $r_\tau$ and $n_r$ were decreased to increase their ($r_\tau$, $n_r$) contribution to the cost function. In experiment 25b the weights were set at $R_r = 0.024$ and $N_r = 40\,000$, while in 25c they were set at $R_r = 0.012$ and $N_r = 20\,000$. These changes resulted in an improvement in the results for these two variables at only moderate cost to the other two variables. It can, however, be seen that the rms error for $r_\tau$ began increasing again, indicating that a further reduction in the weights would not further improve the results.

6. Discussion

First some practical matters relating to the construction of the retrieval algorithm will be discussed. Ab-
solute accuracy during the coding process is required. It is essential that every contribution to the matrix be tested to ensure that the forward code and the adjoint code are in agreement. Any error in either code results in an incorrect gradient. It was found helpful to construct the code in such a way that individual processes, or any combination of processes, may be tested by numerical determination of the linearization of the forward model for comparison with the results of the adjoint model. It is almost impossible to determine from running the retrieval algorithm whether bad convergence is due to an incorrect gradient or to bad weights. Therefore, it is of great help if, at the stage of determining the weights, the user has confidence in the estimation of the gradient. It was further necessary to implement a trap at the beginning of each iteration to trap bad values for any of the variables. The optimization code as implemented has no knowledge of the physical limitations of the variables (although it certainly would be possible to include that). Values for $\theta_i$ and $r_i$ were constrained to remain greater than zero at the initial time to prevent the code from becoming unstable. When values lower than zero for these two variables were found, the algorithm returned to the optimization module, indicating that the initial conditions resulted in an indeterminate function value. The optimization code would then recalculate the initial condition by shortening the step length taken along the gradient.

The matrix approach taken toward the construction of the adjoint model worked well. Since the physical process modeled in this study changes rapidly with processes turning on and off, it was necessary to work with a true tangent linear model, that is, a model lin-
earized about each time step. To do the adjoint integration based on linearization about each time step requires knowing the value of all the variables in the model during the forward integration, or recalculating them, at each time step during the adjoint integration. The values of each variable at each time step during the forward integration need to be stored. This is the minimum storage requirement whether one is taking the matrix approach or the conventional approach. When the matrix approach is used, the memory requirements are higher than in the conventional approach, but a gain in cost can be made since the adjoint code can be effectively vectorized. The construction of the adjoint model is also simpler using the matrix approach. Symbolic algebra programs can be used to construct the contributions of each process to the tangent linear matrix, and a matrix–vector multiplication routine can be developed to take advantage of the organized structure of the matrix. Individual pieces of the code can also be tested easily. In this case the tangent linear model and the adjoint model derive from the same code.

The importance of the proper scaling between variables was illustrated. In this study we found that selecting scaling factors related to the variance of the observational error provided good convergence, while scaling by the relative magnitudes resulted in no convergence. Thus, we suggest that there is indeed a way to a priori select proper scaling factors for a given problem. This result does not invalidate Moore’s suggestion; in the case where the variance of the observational error for all the variables are similar with respect to the mag-
The results from the two kinematic models generally agreed in all aspects. It was found that with "perfect" observations both models are able to retrieve the original start-up condition to great accuracy in 50 iterations, or to acceptable accuracy in 30 iterations (see section 5). It was found that longer assimilation periods, as well as more observations, lead to slower convergence rates of the cost function. More observations resulted in better fits. The routine was able to retrieve information about a variable that was not observed, although up to 150 iterations were needed. Small amounts of information about that variable greatly improved the convergence rates, reducing the number of iterations by about half to 75 for comparable accuracy. The algorithm did not converge when a first-order discontinuity in the tendency function for one of the variables was included. It was necessary to construct a spline over a wide interval spanning the discontinuity to obtain convergence. The neglect of physical processes in the retrieval model resulted in greatly reduced accuracy of the retrieved model fields. This has implications when considering using similar models to assimilate real observations. There may again be two approaches toward this problem. First, develop a parameterized model where all essential physical processes are included. Construct the control problem with the parameterization coefficients included as control variables. This algorithm can then be applied to a well-observed case to determine optimal settings for the pa-

| Table 1. The errors in the initial conditions (two-dimensional model, perturbed observations) for different numbers of observations taken. Three experiments with different weights for $r$, and $n$, using 5-s time steps are included. |
|---|---|---|---|---|---|---|
| No. of observations | $\theta_i$ | $r_i$ | $r_r$ | $n_r$ |
| max | rms ($\times 10^{-4}$) | max | rms ($\times 10^{-4}$) | max | rms ($\times 10^{-4}$) | max | rms |
| Observations | | | | | | | | |
| 3 | 0.50 | 14.8 | 5.00 | 10.0 | 1.57 | 4.55 | 171.49 | 0.45 |
| 5 | 0.51 | 11.4 | 4.94 | 8.81 | 1.59 | 6.87 | 169.93 | 0.59 |
| 13 | 0.50 | 7.85 | 4.43 | 7.04 | 1.47 | 6.13 | 171.02 | 0.63 |
| 25a | 0.36 | 5.74 | 3.32 | 7.03 | 1.50 | 4.43 | 179.19 | 0.47 |
| 25b | 0.39 | 5.74 | 4.95 | 7.27 | 1.55 | 4.45 | 174.33 | 0.45 |
| 25c | 0.45 | 6.52 | 4.95 | 7.27 | 1.55 | 4.45 | 174.33 | 0.45 |
rameterization coefficients. The model with the parameterizations thus determined can then be used to assimilate observations. Second, if the object is more analysis and understanding of the physical processes, then more detailed physics would need to be included. When inaccurate observations were used, both models were able to improve the error statistics over the assimilation period.

It was shown that the adjoint technique can be applied to cloud microphysical models. This technique does allow the opportunity to combine the latest experimental knowledge, theoretical knowledge, and observational knowledge in a consistent way. It was shown how any observation, which can be expressed in terms of the prognostic variables, can be assimilated into a numerical model.

Acknowledgments. We gratefully acknowledge Dr. Jerry Taylor for his continued encouragement and interest, and his help in many ways. Dr. Peter Flatau introduced us to symbolic algebra, which greatly eases the pain of adjointing. Many discussions with Drs. Bob Walko and Craig Tremback also helped. Johannes Verlinde was supported by the South African Weather Bureau during the course of this study. The research was sponsored by the National Science Foundation under Grant ATM-8814913 and by the Army Research Office under Grant DAAL03-86-K-0175. We also would like to thank the three reviewers whose comments helped to improve this manuscript.

APPENDIX A

A Simple Example

To illustrate how the adjoint model is computed a simple example will be shown. Assume a single-gridpoint model that describes the evolution of a raindrop distribution through prognostic equations for two bulk parameters of the distribution, number concentration $n$, and mixing ratio $r$. The raindrops are distributed according to the gamma distribution [Eq. (4.1)]. The relationship between the characteristic diameter and the prognostic variables of the model can be determined from the definition of $r$,

\[
\begin{align*}
\left( \begin{array}{c} n^i \\ r^i \end{array} \right) &= \frac{C_tr}{C_n n^i} \left( Z_{\text{mod}} - Z_{\text{obs}} \right) \left( 1 + \frac{\Delta t}{\Delta n} \frac{\partial f_r}{\partial n} \right) \\
&+ \frac{\Delta t}{\Delta r} \frac{\partial f_r}{\partial r} \\
&= \frac{C_tr}{C_n n^i} \left( \begin{array}{c} (-r) \left( \frac{1}{w_r} 0 \right) + \left( \Delta t \frac{\partial f_r}{\partial r} \right) \\ 2n \left( \Delta t \frac{\partial f_r}{\partial r} \right) \end{array} \right)
\end{align*}
\]

where all terms on the right-hand side are evaluated at time $t = i$. The forward model can be described by

\[
\begin{align*}
n^i &= n^{i-1} + \Delta t f_n(n^{i-1}, r^{i-1}) \\
r^i &= r^{i-1} + \Delta t f_r(n^{i-1}, r^{i-1})
\end{align*}
\]

where $f_n$ is the tendency equation for $n$, and $f_r$ for $r$. The tangent linear matrices $T_i$ are then given by

\[
T_i = 1 + \Delta t \nabla_x f(x^i)
\]

where all the derivatives are evaluated at time $t = i$. The Jacobian $\nabla_x Z(x^i)$ needs to be determined. Since there is only a single observation, for this simple example this will in fact be a vector. So

\[
\nabla_x Z = \left( \frac{\partial Z}{\partial n}, \frac{\partial Z}{\partial r} \right) = \frac{C_tr}{C_n n^i} (-r, 2n).
\]

With these terms defined, a single step of the adjoint model can then be written as

\[
\begin{align*}
\Delta t &\frac{\partial f_n}{\partial n} \left( \begin{array}{c} 1/w_n 0 \\ 0 1/w_r \end{array} \right) + \left( \Delta t \frac{\partial f_n}{\partial r} \right) \\
&= \left( \begin{array}{c} 1 + \Delta t \frac{\partial f_r}{\partial n} \\ 2n \left( \Delta t \frac{\partial f_r}{\partial n} \right) \end{array} \right)
\end{align*}
\]

where $C_2$ is a constant.

Assume that the reflectivity of this distribution is observed. The functional relation between the model prognostic variables and the observed function is given by

\[
Z = \int_0^\infty D^6 n f(D) dD = n D^6 \frac{\Gamma(v + 6)}{\Gamma(v)} = C_1 n D^6_n,
\]

where $C_1$ is another constant.
## APPENDIX B

### List of Constants

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>Constant autoconvergence</td>
<td>$9.74 \times 10^{19}$ m$^2$ kg$^{-2}$ s$^{-1}$</td>
</tr>
<tr>
<td>$a_2$</td>
<td>Constant autoconvergence</td>
<td>$10.8$ m$^2$ kg$^{-2}$ s$^{-1}$</td>
</tr>
<tr>
<td>$a_3$</td>
<td>Constant autoconvergence</td>
<td>$124836$ m$^2$ kg$^{-2}$ s$^{-1}$</td>
</tr>
<tr>
<td>$a_4$</td>
<td>Constant autoconvergence</td>
<td>$20.27$ m$^2$ kg$^{-2}$ s$^{-1}$</td>
</tr>
<tr>
<td>$a_n$</td>
<td>Constant autoconvergence</td>
<td>$3.5 \times 10^8$ m$^{-3}$</td>
</tr>
<tr>
<td>$c_m$</td>
<td>Coefficient for mass of water drops</td>
<td>$\rho / \pi / 6$</td>
</tr>
<tr>
<td>$c_p$</td>
<td>Specific heat water vapor</td>
<td>$1952$ J K$^{-1}$ kg$^{-1}$</td>
</tr>
<tr>
<td>$c_v$</td>
<td>Coefficient for terminal velocity of rain</td>
<td>$132$ m$^{1/2}$ s$^{-1}$</td>
</tr>
<tr>
<td>$l_v$</td>
<td>Latent heat of vaporization</td>
<td>$2.5 \times 10^9$ J kg$^{-1}$</td>
</tr>
<tr>
<td>$p_0$</td>
<td>Reference pressure</td>
<td>$1 \times 10^5$ N m$^{-2}$</td>
</tr>
<tr>
<td>$\rho_m$</td>
<td>Power coefficient for mass</td>
<td>$3.0$</td>
</tr>
<tr>
<td>$\rho_r$</td>
<td>Power coefficient, terminal velocity of rain</td>
<td>$0.5$</td>
</tr>
<tr>
<td>$r_0$</td>
<td>Mixing ratio threshold</td>
<td>$7 \times 10^{-4}$</td>
</tr>
<tr>
<td>$A$</td>
<td>Factor in collection efficiency</td>
<td>$2300$ m$^{-1}$</td>
</tr>
<tr>
<td>$D_s$</td>
<td>Cutoff diameter for breakup</td>
<td>$6 \times 10^{-4}$ m</td>
</tr>
<tr>
<td>$D_s$</td>
<td>Lower limit of spline</td>
<td>$5 \mu$m</td>
</tr>
<tr>
<td>$D_s$</td>
<td>Upper limit of spline</td>
<td>$40 \mu$m</td>
</tr>
<tr>
<td>$K_x$</td>
<td>Horizontal diffusion coefficient</td>
<td>$6412$ m$^2$ s$^{-1}$</td>
</tr>
<tr>
<td>$K_z$</td>
<td>Vertical diffusion coefficient</td>
<td>$298$ m$^2$ s$^{-1}$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Rate coefficient for autoconversion</td>
<td>$0.0001$ s$^{-1}$</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Upper threshold for collection</td>
<td>$2.23 \times 10^{-2}$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Lower threshold for collection</td>
<td>$4.36 \times 10^{-5}$</td>
</tr>
<tr>
<td>$\rho_l$</td>
<td>Density of liquid water</td>
<td>$1 \times 10^3$ kg m$^{-3}$</td>
</tr>
<tr>
<td>$\nu$</td>
<td>Shape parameter of gamma distribution</td>
<td>$3$ or $1$</td>
</tr>
</tbody>
</table>

## REFERENCES


