NSF Grant ATM-8814913
ARO contract DAAL03-86-K-0175
South African Weather Bureau

FITTING MICROPHYSICAL OBSERVATIONS TO A NUMERICAL MODEL THROUGH AN OPTIMAL CONTROL THEORY TECHNIQUE

by Johannes Verlinde

William R. Cotton, P.I.

Colorado State University

DEPARTMENT OF ATMOSPHERIC SCIENCE

PAPER NO. **493**

NUMERICAL SIMULATION OF CIRRUS CLOUDS - FIRE CASE STUDY AND SENSITIVITY ANALYSIS

by

Johannes Verlinde

Department of Atmospheric Science

Colorado State University

Fort Collins, CO 80523

Research Supported by

South African Weather Bureau

Dept. of Environment Affairs

National Science Foundation

under Grant ATM-8814913

Army Research Office

under contract DAAL03-86-K-0175

March 26, 1992

Atmospheric Science Paper No. 493

ABSTRACT

FITTING MICROPHYSICAL OBSERVATIONS TO A NUMERICAL MODEL THROUGH AN OPTIMAL CONTROL THEORY TECHNIQUE

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 also to the retrieval of three-dimensional wind fields from single Doppler radar observations.

This study investigates the feasibility of using four-dimensional variational assimilation for a complex discontinuous numerical model. Three test models were developed, a positive definite advection scheme, a one-dimensional liquid physics kinematic microphysical model with a positive definite advection scheme, 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.

All 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 suggest 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 (eg. reflectivity or liquid water path), a form is derived which shows that this can be accomplished. This is considered to be an advantage of this technique over others, since it is particularly suitable to remote sensing systems where only integral parameters or derivatives of model prognostic variables are observed.

Johannes Verlinde
Department of Atmospheric Science
Colorado State University
Fort Collins, Colorado 80523
Spring, 1992

ACKNOWLEDGEMENTS

First I would like to thank the people back home who have done so much to make it possible for me to come to Colorado for further education, especially Messrs. Groenewald and du Toit of the South African Weather Bureau. Then I would like to acknowledge those who have been more directly involved with my research. I thank my advisor, Dr. Cotton, for his support, encouragement and advice during all the years. My other committee members, Drs. Taylor, Schubert, Stephens and Rutledge also provided useful guidance during the course of the study. And then I would also like to thank Dr. Bringi whose desire for routines to assimilate multi-parameter radar data into cloud models was one of the big motivators for this study. I thank my wife Terri for her continued support and encouragement, especially at the times when I got tense and uptight.

Many people have contributed to my educational experience here. I have spent many enjoyable hours looking at radar data with Dr. Bringi. Piotr 'What's new' Flatau has contributed greatly in various ways. He re-stimulated my interest in analytical methods. He introduced me to the wonders of symbolic algebra, which eased the workload significantly. He also provided many programs, and technical assistance in their use. Dr. Taylor very patiently explained the intrinsic details of the adjoint technique to me, and again when I couldn't understand it, and he provided assistance at several rough spots. Drs. Walko and Tremback provided valuable assistance in my transition from observationist to modeler. They, along with Dr. Smolarkiewicz of NCAR, painstakingly shared their knowledge of advection to me. Dr. Errico from NCAR helped in the early stages. Jeff Copeland encouraged me with his discussions on adjoints. Mike Meyers stimulated me to organize my thoughts on microphysical parameterizations. Dr. Roux from CRPE/CNET-CNRS gave me the code of his dynamical retrieval algorithm. Although this may not be reflected in

the dissertation, it helped a lot to get my research going. Drs. Cram and Walko helped me with my first variational adjustment problems. Brenda Thomson provided technical assistance with LaTeX.

I was financially supported by the South African Weather Bureau, Dept. of Environment Affairs, during the course of my studies. The research was sponsored by the National Science Foundation under Grant ATM-8814913 and the Army Research Office under grant DAAL03-86-K-0175.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

INTRODUCTION

Model fitting to observations is one of the important tools available to scientists to help them understand complex physical phenomena. It may be considered to be fitting a conceptual model to a set of observations to describe how the physical process typically evolves, or it may be changing control variables of a numerical model of the same process, based on the observations, such that the numerical results closely reproduce the given set of observations. In atmospheric science examples of the first kind can be found in, for example, the Norwegian School frontal model (Bjerknes and Solberg, 1922), or the conceptual model of a mature squall-line system by Smull and Houze (1987). Analysis of a weather map is also a form of model fitting. Early attempts were all subjective, where the scientist used his own personal judgement, combined with his personal experience and knowledge, to produce a useful product. With the advent of computer technology the means became available to move away from subjective analysis towards objective analysis (Panofski, 1949). Three types of algorithms have been used in objective analysis: algorithms of spacial interpolation (Cressman, 1959; Barnes, 1973), statistical methods or linear regression algorithms (Gandin, 1965; Schlatter, 1975) and variational algorithms (Sasaki 1958). Spacial interpolation refers to the problem of filling data voids. The statistical methods combine two independent multivariates of the model state (the observations and the background model state) linearly in such a way as to minimize the error variance (Thacker 1987). The roots of variational analysis can be found in the eighteenth century with Euler, Lagrange and Gauss, and the development of the calculus of variations. However, due to the complexity of the problem it was only with the advent of computer technology that this theory was applied in atmospheric science (Sasaki, 1958). With variational assimilation model variables (the control

variables) are changed under physical constraints to fit the observations. The problem becomes much more complex when the observations to be assimilated are scattered in both time and space. The object is then to produce a regular, physically consistent four-dimensional representation of the state of the atmosphere from a heterogeneous array of *in situ* and remote instruments which sample imperfectly and irregularly in space and time, utilizing an imperfect model of the physical process (Daley, 1991).

The object of this dissertation is to investigate the feasibility of using a four-dimensional variational assimilation technique to incorporate observations of the physical processes involved in precipitation formation into a numerical cloud microphysical model. The means through which this is accomplished is a series of identical twin experiments. A kinematic bulk parameterized model of the liquid phase physics is developed. This model is used to create an observational data set in space and time. An iterative algorithm is derived through which the model can be fitted to the observations. The control variables selected for this study are the variables describing the state of the model at the initial time, but it may also include other elements of the model such as boundary conditions or parameterization constants. This algorithm is used to retrieve the original initial condition used in the observation run, starting from a constant initial condition. Factors influencing the rate of convergence are investigated.

The dissertation is set out as follows: In the remainder of the introduction the literature of four-dimensional data assimilation (FDDA) as applied in dynamic meteorology is surveyed. This is followed by a literature review of what has been done in the field of microphysical assimilation. The introduction is concluded by a discussion on what needs to be done. In the second chapter the optimal control theory algorithm is derived. In the third chapter the general problem of advection is discussed, and the FDDA technique is applied to a positive definite advection problem. In chapters four and five two liquid phase bulk parameterizations are discussed, along with their fitting algorithms. The final chapter discusses the results, and shows how these algorithms may be used when the observations are not of the prognostic variables of the model, but rather of some general functions of the prognostic variables (eg. reflectivity or integrated liquid water path). Suggestions for further research are also given.

1.1 Model fitting with dynamic meteorological models

Interpolation to fill data voids does not really fit into the realm of model fitting, since the direction is from the observations to the model (Thacker, 1987). This was the obvious first step for early analysis attempts. It is a single-time technique; since the focus here is on four dimensional assimilation, it will not be discussed any further.

As the science of numerical weather prediction advanced, the need for better initial conditions became critical. Numerical weather prediction classically is viewed as an initial value problem where the governing equations of fluid dynamics are integrated forward from fully determined initial values of the meteorological fields at some initial time (Morel, 1981). Morel concluded that single-time data-sets available to numerical forecasters are likely to remain incomplete and inaccurate, rendering inaccurate forecasts. Therefore any forecasting scheme must be initialized at time $t = 0$ by merging the new observations with currently estimated fields, computed on the basis of earlier observations collected at times $t \leq 0$. The data must be brought into the model in such a way that the model fields would remain internally consistent and in balance according to our understanding of the physical laws of nature, and yet in a sense optimally reflect the observations.

FDDA was attempted first by Charney et al. (1969). They suggested direct substitution of model values by observed values during integration of the model. The substitutions would take place at the proper time and position of the observation. However, it was found that the models frequently reject data which significantly differed from the model state, or that the introduction of the observations produced undesirable shocks in the model since it introduced unbalanced model states. Building on this Anthes (1974) and Hoke and Anthes (1976) developed the idea of Newtonian nudging (continuous dynamical assimilation). They added an additional forcing term, proportional to the degree of misfit, to the prognostic equations of an explicit dynamic model. The model state is then gradually "nudged" towards the observations. Advantages of this method are: It can be used to assimilate any data type (provided that observations are of model prognostic variables). It is conceptually and computationally simple. It allows for scale interactions and is readily adaptable to any domain scale. Disadvantages are: The nudging constant is assigned

in an application-dependent semi-arbitrary manner. It is not clear what constitutes the optimal assimilated data set. The use of accurate non-representative observations can significantly impact the analysis (Seaman, Lecture at the NCAR 1990 Summer Colloquium on Mesoscale Data Assimilation). The possibility also exists that when the nudging term is turned off the model can be in a physically unbalanced state. In that case the model may reject the information and return to its previously balanced state (Walko et al., 1989), or may become unstable.

Optimal interpolation is a widely used statistical approach to objective analysis. Schlatter (1975) proposed a general multi-variate statistical interpolation scheme whereby different fields may be simultaneously analyzed in terms of grid-point values. Its advantages are: It allows differentiation among observing systems and the incorporation of error information specific to each. It has the ability to estimate one variable from observation of another. It weighs observations relatively to be consistent with past atmospheric behavior. It produces estimates of the analysis error as a function of the distribution and accuracy of the data. Its disadvantages are that it is computationally quite expensive, and that scale-dependent correlation models require a long history of numerical forecasts for accurate determination of empirical coefficients (Schlatter, Lecture at the NCAR 1990 Summer Colloquium on Mesoscale Data Assimilation).

While analysis schemes like optimal interpolation schemes use information based on previous observations, it is not true FDDA. These analyses can be used in FDDA by techniques that assimilate analyses of data in an intermittent fashion. This can be done through either analysis nudging (Hoke and Anthes, 1976; Kuo and Guo, 1989), or through analysis—nonlinear normal mode initialization—forecast cycles (DiMego, 1988). These techniques are widely used in the assimilation procedures for large-scale prediction models. However, since this involves at least a two step process it was desired to find a dynamical data assimilation method, which would eliminate the objective analysis as well as the initialization procedures, where the model would act as an integrator of all data (Bengtsson, 1975).

The extension of optimal interpolation into the true four-dimensional assimilation realm is Kalman filtering (Kalman, 1960). In Kalman filtering the background error covariances (model error) are computed using the forecast model. For Kalman filtering knowledge of the error statistics of the observations as well as the forecast model is required. The error covariance matrix is propagated over the simulation period, and at each observation period assimilation is handled as for optimal interpolation (Thacker, 1987). Advantages of this method is that the model is not assumed to be perfect, and that for every estimate of the model state the error-covariance matrix is known (this is quite useful in assessing the reliability of the estimate). Disadvantages are that this is very expensive, and only the present (latest) state is based on all information available; earlier model states are not recomputed using subsequent data (Thacker, 1987).

Sasaki (1958) introduced the first variational analysis scheme to atmospheric science. This method uses dynamical integral constraints as part of the initialization. Initially, this was restricted to only the space domain, with no time variation. Sasaki (1969) and Thomson (1969) extended this theory to include time variation in the constraints. Although these studies pointed out the usefulness of the variational technique, the use of it was not fully pursued until a decade later. Variational methods have a great theoretical advantage over the previously mentioned methods, by providing exact consistency between the analysis and the dynamics, as is expressed by the constraint. However, their higher mathematical technicality and probable high computational cost have strongly limited their use (Le Dimet and Talagrand, 1986).

Talagrand (1981a,b) investigated the possibility of repeated data insertion during forward–backward integration cycles. Although these results indicated that this may be a feasible method, it could not be applied for general numerical prediction models, since terms like diffusion, or many physical parameterizations, are non-physical during backward integration, and may also become numerically unstable. Le Dimet and Talagrand (1986) suggested a technique, which they called the adjoint technique for data assimilation, to solve the classical variational adjustment problem posed by Sasaki (1958). They posed the problem in terms of control theory. If the solution of a numerical model, with

specified boundary conditions and parameterization constants, is a unique function of the initial condition, then it is possible to define a cost function (distance between the model predicted fields (or functions there-of) and the observed fields) which will also be only a function of the initial conditions, the control variables. The adjoint technique is a mathematical tool for the determination of the gradient of the cost function with respect to the initial conditions. Using this gradient, the initial conditions may be varied such as to reduce the cost. A full derivation of this method will be given in the next chapter. An alternative derivation of the adjoint method of data assimilation can also be found by formulating the variational problem in terms of Lagrangian multipliers (Thacker, 1988a,b). The control variables may also be expanded to include other elements of the numerical model such as boundary conditions and/or parameterization constants. In the terms adapted for variational assimilation (Sasaki, 1970), the model is a strong-constraint, although the effect of model errors can be included into the adjoint approach by including the error in the set of control variables and solving for it (Derber, 1989; Thacker, 1989).

The above techniques need to be evaluated in terms of the application of this study. Optimal interpolation does not really address the problem of FDDA, while Kalman filtering is prohibitively expensive. The various insertion techniques, though suitable for initializing forecasting models, have limited application as analysis tools and can even generate noise. The model is only striving towards the observations at all times. Thus, the precipitation evolution described by the model history does not represent the observed evolution, but rather some evolution corresponding to neither model nor real. This might conceivably be improved by doing forward–backward integrations, thus continually forcing the model closer to observations, but one is still left with physical interpretation of what one is doing during the backward integration, and there is no guarantee of convergence. These techniques also have the disadvantage that only observations of the prognostic variables can be included, whereas today significant parts of the observational data set derive from remote sensing instruments, which in general do not measure the common prognostic variables, but rather some function of them. This leaves the adjoint technique, which is

a true FDDA technique, which does provide a measure of the rate of convergence in the norm of the gradient, and also provides an estimate of the goodness of fit. It can also assimilate observations from remote sensing platforms.

1.2 Model fitting with microphysical models

In this section a literature review of the work that has been done in the field of fitting microphysical models to observations will be given. Attention will be focussed on the ways radar has been used for this purpose.

Radar data is most often used to infer microphysical information directly from the radar observables (Gunn et al., 1954; Atlas, 1957; Austin and Wexler, 1957; Battan, 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 utilizes 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 rawinsonde, cloud water content and precipitation observations. In the early 1980's Rutledge and Hobbs (1983; 1984) followed by Ziegler (1985; 1988) fitted microphysical numerical models to Doppler radar observed wind fields. The same basic 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, but only as a validating source. This technique therefore does not fit the definition of model fitting in the true sense of the word, but it was an important step in that direction in this field.

These models were used to study the 'seeder-feeder' process in warm-frontal rain bands (Rutledge and Hobbs, 1983), precipitation development in narrow cold-frontal rain-bands (Rutledge and Hobbs, 1984), a 'steady-state' convective storm (Ziegler, 1985), the electrification in an isolated mountain thunderstorm (Ziegler et al., 1986), precipitation processes in the stratiform region of tropical squall lines (Rutledge, 1986; Le Cam and Isaka, 1989), the stratiform region of a mid-latitude squall line (Rutledge and Houze, 1987), and a severe frontal rain-band (Rutledge, 1989).

Another study, closer to the definition of model fitting, was done by Hauser and Amayenc (1986). They developed a variational adjustment scheme for the retrieval of cloud water content and water vapor content. 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 rain water 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 contents 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 a dynamical retrieval routine to retrieve a model state that is internally consistent. In this study they moved away from using the reflectivity to force their model, but rather opted to use the reflectivity to validate their model. The microphysical model was based on the continuity equations for total water and precipitation content, utilizing bulk microphysics parameterization to describe interactions between different precipitation species. In this model they allowed both frozen and liquid phase physics. This model was used to study the precipitation physics in a tropical squall line.

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, eg. a developing cumulus cloud. They did an observing system simulation experiment (OSSE) for a convective cloud, and used algorithms of the Rutledge/Ziegler type to retrieve the microphysical fields. This study showed that the effect of non-steadiness not

only over- or under-estimates the microphysical and thermal fields, but can also misrepresent the evolution of the precipitation fields. It was concluded that the time history of the cloud is important, and that the direct assimilation of radar reflectivity observations will be needed if these processes are to be used in non-steady cases like typical convective storms. This then illustrated the need for further research in this field.

1.3 Setting the problem

It was concluded in Section 1.1 that the adjoint technique appears to hold the most promise as a FDDA technique that could assimilate time-varying radar observations into a cloud model. The necessity of a FDDA technique was discussed in Section 1.2. Therefore, this study will investigate the feasibility of applying the adjoint approach to the problem of microphysical retrieval.

The adjoint method has up to now mostly been used in dynamic models with dry dynamics and no parameterizations. Initial applications were for synoptic scale model initialization (Lewis and Derber, 1985; Courtier and Talagrand, 1987; Derber, 1989; Thacker, 1988b; Thacker and Long, 1988; Thacker, 1989; Tziperman and Thacker, 1989; Courtier and Talagrand, 1990; Marchuk and Skiba, 1990). Work is currently under way (Kapitza, 1991; Sun et al., 1991) on the much smaller scale problem of fitting non-hydrostatic models to single-Doppler radar observed radial wind-fields. This work holds the promise that continuous non-steady three dimensional wind, pressure and buoyancy fields can be retrieved from single- (or dual-) Doppler radar observations. The basis for continued work with a kinematic bulk parameterized microphysical model is thus being laid down. No work exists in the atmospheric science literature where the adjoint method is applied to a model with discontinuous physics.

This study will therefore focus on a kinematic bulk parameterized microphysical model. For simplicity only liquid phase physics will be considered. It will be assumed that continuous wind, pressure and buoyancy fields are available such as derived from multiple Doppler radar retrieval algorithms (Kapitza, 1991; Sun et al., 1991). To create a realistic evolution of these fields for the simulation experiments, the RAMS model

(Tripoli and Cotton, 1982; Tripoli, 1986) will be used, but with the microphysical package developed here. Identical twin experiments are then performed based on these fields. This work will primarily be concerned with the problem of whether the adjoint technique will work for a complex discontinuous system, therefore it will in general be assumed that the prognostic variables of the model can be observed. However, it will be shown how the theory can be extended to cases where the observations are general functions of the prognostic variables.

Chapter 2

VARIATIONAL ASSIMILATION OF OBSERVATIONS USING OPTIMAL CONTROL THEORY

In this chapter the basic mathematics and the philosophy of the variational approach to four-dimensional data assimilation will be discussed. Some discussion on the technical aspects of the construction of the adjoint model will also be given.

The stated problem is to produce a regular, physically-consistent four dimensional representation of a physical system, in this case the evolution of precipitation in a cloud, from a set of observations sampled in space and time. Specify a set of observations over the interval $(0, T)$, and assume that there is a numerical model describing the atmospheric processes observed. A model solution is sought that will fit the observations 'the best'. The goodness of fit is defined by a scalar cost function which, for any solution of the model over the interval $(0, T)$, measures the 'distance' between the model solution and the observations. The object is then to seek the particular model solution that will minimize this cost function.

Courtier and Talagrand (1990) suggest that in view of the complexity and the extremely large size of the corresponding minimization problem, the only practical way to numerically determine the minimizing solution is to take the model initial condition (and/or boundary conditions and/or parameterization constants) as the 'control variables' of the problem. Only the initial problem will be considered in this study. This can be done if the cost function is a unique function of the control variables. This is true since the model solution, assuming specified boundary conditions and parameterization constants, is uniquely determined by the initial conditions, and the observations are independent of the model. The gradient of the cost function with respect to the initial conditions can

then be determined, and a descent algorithm can be used to determine the initial condition that will minimize this cost function.

2.1 Derivation

The adjoint operator will be derived following Taylor (1991). It is necessary to derive the adjoint for the discrete version of the model. The model fields which are fitted to the observations are produced by numerical approximations of the continuous equations. By deriving the adjoint from the continuous model equations and then discretizing, errors in the gradient will be introduced. Assume that the numerical model which explicitly integrates, for given initial conditions, a set of differential equations, can be written in synthetic form as

$$\begin{aligned} \frac{d\mathbf{x}}{dt} &= \mathbf{f}(\mathbf{x}, t) \\ \mathbf{x}(0) &= \mathbf{x}^0, \end{aligned} \tag{2.1}$$

where $\mathbf{x} \in \mathrm{R}^n$ is the state vector, consisting of all the model prognostic variables defined at each grid point. This defines a unique solution $\mathbf{x} = \mathbf{x}(t)$ for $0 \leq t \leq T$. Note that bold faced lower case letters are used for vectors and bold face uppercase letters are used for matrices. Assume that observations of some function of the state vector $z = z(\mathbf{x})$ are available at N time periods, which for this derivation, need not be at set intervals. Let the set of observations be given by $\mathbf{o}(t_i) = \mathbf{o}^i$, $i = 0, 1, \ldots, N$. The model times corresponding to the observational times will be indicated by superscripts k_i.

Define the cost function as

$$E(\mathbf{x}^0) = \frac{1}{2} \sum_{j=0}^{N} \left[\mathbf{z}(\mathbf{x}^{k_j}) - \mathbf{o}^j\right]^T \mathbf{W} \left[\mathbf{z}(\mathbf{x}^{k_j}) - \mathbf{o}^j\right] \tag{2.2}$$

where $\mathbf{W}$ is a constant weight matrix and superscript T denotes a transpose. The gradient of the cost function with respect to the initial conditions is

$$\begin{aligned} \nabla_{\mathbf{x}^0} E(\mathbf{x}^0) &= \frac{1}{2} \sum_{j=0}^{N} \nabla_{\mathbf{x}^0} \left[\mathbf{z}(\mathbf{x}^{k_j}) - \mathbf{o}^j\right]^T \mathbf{W} \left[\mathbf{z}(\mathbf{x}^{k_j}) - \mathbf{o}^j\right] \\ &= \sum_{j=0}^{N} \left[\nabla_{\mathbf{x}^0} \left(\mathbf{z}(\mathbf{x}^{k_j}) - \mathbf{o}^j\right)\right]^T \mathbf{W} \left[\mathbf{z}(\mathbf{x}^{k_j}) - \mathbf{o}^j\right] \\ &= \sum_{j=0}^{N} \left[\nabla_{\mathbf{x}^{k_j}} \mathbf{z}(\mathbf{x}^{k_j}) \nabla_{\mathbf{x}^0} \mathbf{x}^{k_j}\right]^T \mathbf{W} \left[\mathbf{z}(\mathbf{x}^{k_j}) - \mathbf{o}^j\right] \end{aligned} \tag{2.3}$$

since $\nabla_{\mathbf{x}^0}\mathbf{o}^j = 0$. To get a usable expression for $\nabla_{\mathbf{x}^0}\mathbf{x}^{k_j}$ the chain rule must be applied backwards to the forward operator of the model, which in this case is just a simple forward upstream operator

$$\mathbf{x}^{k_j} = \mathbf{x}^{k_j-1} + \delta t\ \mathbf{f}(\mathbf{x}^{k_j-1}). \tag{2.4}$$

Taking the Jacobian of Eq (2.4) gives

$$\begin{aligned}
\nabla_{\mathbf{x}^0}\mathbf{x}^{k_j} &= \nabla_{\mathbf{x}^0}\mathbf{x}^{k_j-1} + \delta t\ \nabla_{\mathbf{x}^0}\mathbf{f}(\mathbf{x}^{k_j-1}) \\
&= \nabla_{\mathbf{x}^0}\mathbf{x}^{k_j-1} + \delta t\ \nabla_{\mathbf{x}^{k_j-1}}\mathbf{f}(\mathbf{x}^{k_j-1})\nabla_{\mathbf{x}^0}\mathbf{x}^{k_j-1} \\
&= \left[\mathbf{I} + \delta t\ \nabla_{\mathbf{x}^{k_j-1}}\mathbf{f}(\mathbf{x}^{k_j-1})\right]\ \nabla_{\mathbf{x}^0}\mathbf{x}^{k_j-1} \\
&= \mathbf{T}_{k_j-1}\nabla_{\mathbf{x}^0}\mathbf{x}^{k_j-1}
\end{aligned} \tag{2.5}$$

Realizing that $\nabla_{\mathbf{x}^0}\mathbf{x}^0 = \mathbf{I}$, the identity matrix, Eq (2.5) can be written as

$$\nabla_{\mathbf{x}^0}\mathbf{x}^{k_j} = \mathbf{T}_{k_j-1}\mathbf{T}_{k_j-2} \times \ldots \times \mathbf{T}_1\mathbf{T}_0. \tag{2.6}$$

Eq (2.3) can then be written as

$$\begin{aligned}
\nabla_{\mathbf{x}^0}E(\mathbf{x}^0) &= \mathbf{W}\left[\mathbf{z}(\mathbf{x}^0) - \mathbf{o}^0\right] + \sum_{j=1}^{N}\left[\nabla_{\mathbf{x}^{k_j}}\mathbf{z}(\mathbf{x}^{k_j})\nabla_{\mathbf{x}^0}\mathbf{x}^{k_j}\right]^T\mathbf{W}\left[\mathbf{z}(\mathbf{x}^{k_j}) - \mathbf{o}^j\right] \\
&= \mathbf{W}\left[\mathbf{z}(\mathbf{x}^0) - \mathbf{o}^0\right] + \sum_{j=1}^{N}[(\nabla_{\mathbf{x}^{k_j}}\mathbf{z}(\mathbf{x}^{k_j}))\mathbf{T}_{k_j-1} \times \ldots \times \mathbf{T}_0]^T\mathbf{W}\left[\mathbf{x}^{k_j} - \mathbf{o}^j\right] \\
&= \mathbf{W}\left[\mathbf{z}(\mathbf{x}^0) - \mathbf{o}^0\right] + \mathbf{T}_0^T \times \ldots \times \mathbf{T}_{k_1-1}^T\left\{\left[\nabla_{\mathbf{x}^1}\mathbf{z}(\mathbf{x}^1)\right]^T\mathbf{W}\left[\mathbf{z}(\mathbf{x}^{k_1}) - \mathbf{o}^1\right] + \right. \\
&\quad \mathbf{T}_{k_1}^T \times \ldots \times \mathbf{T}_{k_2-1}^T\left\{\left[\nabla_{\mathbf{x}^2}\mathbf{z}(\mathbf{x}^2)\right]^T\mathbf{W}\left[\mathbf{z}(\mathbf{x}^{k_2}) - \mathbf{o}^2\right] + \ldots\right. \\
&\quad \left.\left.\mathbf{T}_{k_{N-1}}^T \times \ldots \times \mathbf{T}_{k_N-1}^T\left\{\left[\nabla_{\mathbf{x}^{k_N}}\mathbf{z}(\mathbf{x}^{k_N})\right]^T\mathbf{W}\left[\mathbf{z}(\mathbf{x}^{k_N}) - \mathbf{o}^N\right]\right\}\ldots\right\}\right\}.
\end{aligned} \tag{2.7}$$

Defining $\mathbf{y}^j = \left[\nabla_{\mathbf{x}^{k_j}}\mathbf{z}(\mathbf{x}^{k_j})\right]^T\mathbf{W}\left[\mathbf{z}(\mathbf{x}^{k_j}) - \mathbf{o}^j\right]$ and $\mathbf{A}_{j-1} = \mathbf{T}_{k_{j-1}}^T \times \ldots \times \mathbf{T}_{k_j-1}^T$, Eq (2.7) can be written in shorthand notation as

$$\nabla_{\mathbf{x}^0}E(\mathbf{x}^0) = \mathbf{y}^0 + \mathbf{A}_0\{\mathbf{y}^1 + \mathbf{A}_1\{\mathbf{y}^2 + \ldots\{\mathbf{y}^{N-1} + \mathbf{A}_{N-1}\mathbf{y}^N\}\ldots\}\}. \tag{2.8}$$

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.*

Next the concept of the tangent-linear model is introduced, and it is shown how this relates to the gradient just derived. Let the numerical model be written as in (2.1). This

may be linearized by evaluating that expression at the point $\mathbf{x} + \delta\mathbf{x}$, where $\mathbf{x}$ again is defined as the state vector at a given time. After some manipulation this gives

$$\frac{d\delta\mathbf{x}}{dt} = \mathbf{f}'(\mathbf{x})\delta\mathbf{x}, \tag{2.9}$$

which describe the time evolution of a given perturbation on the model state $\mathbf{x}$ (Errico and Vukicevic, 1991). As previous, $\mathbf{f}'(\mathbf{x})$ is the Jacobian of the vector operator $\mathbf{f}$ evaluated at the point $\mathbf{x}$. *This is called the tangent linear model.* Comparing this to (2.5) and (2.8), it can be seen that the gradient is expressed in terms of the transpose of the tangent linear model.

2.2 A simple example

To illustrate how the adjoint model is computed, a simple example will be shown. Assume a single grid-point model which 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 rain drops are distributed according to the gamma distribution

$$f(D) = \frac{1}{\Gamma(\nu)}\left(\frac{D}{D_n}\right)^{\nu-1}\frac{1}{D_n}\exp\left(-\frac{D}{D_n}\right) \tag{2.10}$$

where D_n is the characteristic diameter of the distribution and Γ is the complete gamma function. The relationship between the characteristic diameter and the prognostic variables of the model can be determined from the definition of r,

$$\begin{aligned} r &= \int_0^\infty \frac{\pi}{6}\frac{\rho_l}{\rho_a} D^6 n f(D) dD \\ &= \frac{\pi}{6}\frac{\rho_l}{\rho_a}\frac{\Gamma(\nu+3)}{\Gamma(\nu)} n D_n^3 \\ &= C_2 n D_n^3, \end{aligned} \tag{2.11}$$

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

$$\begin{aligned} Z &= \int_0^\infty D^6 n f(D) dD \\ &= n D_n^6 \frac{\Gamma(\nu+6)}{\Gamma(\nu)} \\ &= C_1 n D_n^6, \end{aligned} \tag{2.12}$$

where C_1 is another constant.

The forward model can be described by

$$n^i = n^{i-1} + \delta t\ f_n(n^{i-1}, r^{i-1}) \tag{2.13}$$

$$r^i = r^{i-1} + \delta t\ f_r(n^{i-1}, r^{i-1}) \tag{2.14}$$

where f_n is the tendency equation for n and f_r for r. The matrices $\mathbf{T}_i$ are then given by

$$\begin{aligned} \mathbf{T}_i &= \mathbf{I} + \delta t\ \nabla_{\mathbf{x}^i}\mathbf{f}(\mathbf{x}^i) \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \delta t \begin{pmatrix} \frac{\partial f_n}{\partial n} & \frac{\partial f_n}{\partial r} \\ \frac{\partial f_r}{\partial n} & \frac{\partial f_r}{\partial r} \end{pmatrix} \end{aligned} \tag{2.15}$$

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

$$\nabla_{\mathbf{x}} Z = \left(\frac{\partial Z}{\partial n}, \frac{\partial Z}{\partial r}\right), \tag{2.16}$$

and

$$\frac{\partial Z}{\partial n} = C_1 D_n^6 + 6C_1 n D_n^5 \frac{\partial D_n}{\partial n} \tag{2.17}$$

$$\frac{\partial Z}{\partial r} = C_1 n D_n^5 \frac{\partial D_n}{\partial r} \tag{2.18}$$

where

$$\frac{\partial D_n}{\partial n} = -\frac{D_n}{3n} \tag{2.19}$$

$$\frac{\partial D_n}{\partial r} = \frac{D_n}{3r}. \tag{2.20}$$

Thus, Eq (2.16) can be written as

$$\nabla_{\mathbf{x}} Z = C_1 D_n^6(-1, 2). \tag{2.21}$$

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

$$\begin{pmatrix} n_*^i \\ r_*^i \end{pmatrix} = C_1 D_n^6 (Z_{mod} - Z_{obs}) \times$$

$$\begin{pmatrix} 1+\delta t\,\frac{\partial f_n}{\partial n} & \delta t\,\frac{\partial f_r}{\partial n} \\ \delta t\,\frac{\partial f_n}{\partial r} & 1+\delta t\,\frac{\partial f_r}{\partial r} \end{pmatrix} \begin{pmatrix} -1 & \\ & 2 \end{pmatrix} \begin{pmatrix} 1/w_n & 0 \\ 0 & 1/w_r \end{pmatrix} + \begin{pmatrix} n_*^{i+1} \\ r_*^{i+1} \end{pmatrix} \tag{2.22}$$

where all terms on the right hand side are evaluated at time $t = i$

2.3 Procedure

With these preliminaries the procedure to find the best fit of the model to the observations can be described. First the appropriate tangent linear model of the non-linear forward model needs to be derived. Now, with the cost (for a given initial condition) and the gradient of the cost (for the same initial condition) known, using a gradient based minimization algorithm, that initialization for which the cost is a minimum can be calculated via an iterative process (Taylor, 1991).

In this study the optimization code described by Buckley and LeNir (1983) is used to minimize the cost function. 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. The optimization routine provides an initial condition to the model, and requires a function evaluation (the cost function) and a gradient. The cost function and the matrices $\mathbf{T}_j$ are calculated from integrations of the forward and adjoint models.

There are two possible ways to approach the construction of the adjoint model. The commonly used approach is to take the numerical code for the forward model, and take it line by line, find the linearization of each line, take the transpose of it, and then integrate the code from the back to the front. This procedure is described in greater detail by Errico and Vukicevic (1991). The alternative procedure is to actually calculate the matrices $\mathbf{T}_j$ in Eq (2.8) during the forward integration. Both methods have advantages and disadvantages, and the decision which to use may be determined by the forward model used.

It is necessary during the adjoint integration to retrace exactly the same path as followed during the forward integration. This implies that for the construction of the adjoint code it is necessary that the outcome of each conditional statement can be determined from the information at hand during that stage of the adjoint integration, or else that information must be stored and be available. This complicates the engineering aspects of the construction of the first approach, but has no impact on the second since the matrix is calculated when the outcome of the conditional statement is known. However, the major disadvantage of the second approach may be the memory requirements. Depending on the complexity of the physics, but primarily on the advection scheme, it is necessary that several times the state vector be stored for each time step. With the first approach, depending on the evolution of the system, it may only be necessary to store a few state vectors during the course of the forward integration, since it is common to not construct a true tangent linear model, but to linearize around a typical or slowly evolving field. Advantages of the second approach are in the simplification of the construction of the adjoint, and the debugging of the adjoint code. Symbolic computing programs may be used to construct the code for the contribution of each physical process to the various matrix elements. Where the first approach requires that diagnosed variables be dealt with explicitly in the adjoint model, that problem doesn't appear in the second. Diagnostic variables calculated through iterative procedures may present difficult problems for the first approach. Since the adjoint model in the second approach consists of linear operations, this can efficiently be implemented on vector and parallel machines. Thus, several factors come into play when considering which approach to take. It involves trade-offs between such things like computer memory and speed, manpower input, complexity of the numerical model and physical processes modeled.

The second approach was selected for this study. It was decided that the problem studied will in general not require too large a grid domain and number of prognostic variables, that the physical process is discontinuous and therefore requires at least a true tangent linear matrix, that there is an iterative diagnostic variable in the forward model, and that ease for debugging the code was desired.

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

The next factor that then enters the picture is the construction of the state vector. This will impact the form of the matrices.

Suppose there are three prognostic variables in the model with N grid-points, $\mathbf{a}(\mathbf{x},t) = (a_1,\ldots,a_N)^T$, $\mathbf{b}(\mathbf{x},t) = (b_1,\ldots,b_N)^T$, and $\mathbf{c}(\mathbf{x},t) = (c_1,\ldots,c_N)^T$, each dependent on the state vector $\mathbf{x}$ and time. The time variation for each is described by

$$\frac{\partial f}{\partial t} = F(\mathbf{a},\mathbf{b},\mathbf{c}), \tag{2.23}$$

where F is a general (non)-linear operator. The state vector can be defined as

$$\mathbf{x} = (a_1,\ldots,a_N,b_1,\ldots,b_N,c_1,\ldots,c_N)^T \tag{2.24}$$

or as

$$\mathbf{x} = (a_1,b_1,c_1,\ldots,a_N,b_N,c_N)^T. \tag{2.25}$$

The first vector will result in a 3×3 block with each block matrix being $N \times N$. The single grid-point 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 multi grid-point physics will produce bands through all the block matrices.

The second vector will produce a single banded matrix. This structure of the matrix makes it easier to compute the gradient, and also is more cost efficient, since a single vectorizable loop can be constructed. This was the approach selected for this study.

Chapter 3

ON ADVECTION

One of the important problems in numerical modeling in Eulerian space is the numerical approximation of the advection terms in a stable and accurate way. When the advection of positive definite tracers (i.e. microphysical quantities, aerosols, etc.) are considered, additional constraints are added. This problem is further compounded when construction of an adjoint of the forward operator is considered. Since advection is the backbone of any numerical model, an overview of this problem will be given in this chapter. Cases of one- and two-dimensional advection and their adjoints will be considered. The feasibility of using different advection schemes in an assimilation model will be discussed by looking at a few examples of commonly used advection schemes.

3.1 General

Consider the following advection equation, written in flux form,

$$\frac{\partial \psi}{\partial t} = -\frac{\partial u\psi}{\partial x} - \frac{\partial v\psi}{\partial y} \tag{3.1}$$

where $\psi = \psi(t, x, y)$ is a non-diffusive scalar quantity, and $(u, v) = (u, v)(t, x, y)$ is the vector of the advecting velocity. ψ^n is the numerical approximation of the solution of Eq (3.1), defined at points $(t^n, x_{i,j})$, where $t^n = n\Delta t$ and $x_{i,j} = (i\Delta x, j\Delta y)$, $n = 0, \ldots, NT$; $i = 0, \ldots, NX$; $j = 0, \ldots, NY$; and Δt, Δx and Δy are constant time and spatial increments. The velocities are defined on a staggered grid, with $u_{i,j}$ defined at $x_{i+1/2,j}$ and $v_{i,j}$ at $x_{i,j+1/2}$. Let $F_{i,j}$ be the flux in the i-direction and $\alpha_{u_{i,j}} = u_{i,j}\Delta t/\Delta x$ the Courant number defined at $x_{i+1/2,j}$, while $G_{i,j}$ the flux in the j-direction and $\alpha_{v_{i,j}} = v_{i,j}\Delta t/\Delta y$ the Courant number defined at $x_{i,j+1/2}$. A numerical approximation for Eq (3.1) can then be

written as

$$\psi_{i,j}^{n+1} = \psi_{i,j}^{n} - \left[F_{i,j}^{n} - F_{i-1,j}^{n} + G_{i,j}^{n} - G_{i,j-1}^{n}\right]. \tag{3.2}$$

Leith (1965) however, has shown that this formulation is unstable, and has suggested a sequence of updating the orthogonal directions sequentially

$$\begin{aligned} \psi_{i,j}^{n+1/2} &= \psi_{i,j}^{n} - \left[F_{i,j}^{n} - F_{i-1,j}^{n}\right] \\ \psi_{i,j}^{n+1} &= \psi_{i,j}^{n+1/2} - \left[G_{i,j}^{n+1/2} - G_{i,j-1}^{n+1/2}\right]. \end{aligned} \tag{3.3}$$

This scheme was shown to be unstable for deformational flow (Petschek and Libersky, 1975). It was found that taking account of the divergence of the flow in each direction during calculations produces a stable scheme. Thus, the numerical approximation for the two-dimensional advection scheme is

$$\begin{aligned} \psi_{i,j}^{n+1/2} &= \psi_{i,j}^{n} - \left[F_{i,j}^{n} - F_{i-1,j}^{n}\right] + \Delta t \psi_{i,j}^{n} \frac{\partial u}{\partial x}\Big|_{i,j}^{n} \\ \psi_{i,j}^{n+1} &= \psi_{i,j}^{n+1/2} - \left[G_{i,j}^{n+1/2} - G_{i,j-1}^{n+1/2}\right] + \Delta t \psi_{i,j}^{n} \frac{\partial v}{\partial y}\Big|_{i,j}^{n}. \end{aligned} \tag{3.4}$$

Nothing has yet been said about how the fluxes are defined. This has another very important impact on the accuracy of the scheme. The fluxes can be calculated from Taylor expansion or polynomial fitting (Crowley, 1968; Tremback et al., 1987). References to the order of schemes refer to the order of spatial accuracy. Higher order schemes do well on maintaining the amplitude, but suffer from oscillations in the numerical solution due to the fact that certain fluxes are overestimated with respect to their analytical value. In contrast, the magnitude of the fluxes of the first-order schemes are underestimated (Smolarkiewicz and Grabowski, 1990). Thus, higher order schemes will produce negative values along the zero boundaries of the advected quantity, while a first order scheme, though remaining monotone, will suffer from heavy damping. Higher order schemes imply many more multiplications for the calculation of each flux, thereby increasing the cost.

Tremback et al. (1987) derived the equations for up to the tenth-order fluxes. Their expression for the second order flux is

$$F_{i,j} = \frac{\alpha_{u_{i,j}}}{2}(-\psi_{i,j} - \psi_{i+1,j}) + \frac{\alpha_{u_{i,j}}^{2}}{2}(-\psi_{i,j} + \psi_{i+1,j}) \tag{3.5}$$

and the fourth order flux on a constant grid is

$$\begin{aligned} F_{i,j} &= \frac{\alpha_{u_{i,j}}}{12}(\psi_{i-1,j} - 7\psi_{i,j} - 7\psi_{i+1,j} + \psi_{i+2,j}) \\ &\quad \frac{\alpha^2_{u_{i,j}}}{24}(\psi_{i-1,j} - 15\psi_{i,j} + 15\psi_{i+1,j} - \psi_{i+2,j}) \\ &\quad \frac{\alpha^3_{u_{i,j}}}{12}(-\psi_{i-1,j} + \psi_{i,j} + \psi_{i+1,j} - \psi_{i+2,j}) \\ &\quad \frac{\alpha^4_{u_{i,j}}}{24}(-\psi_{i-1,j} + 3\psi_{i,j} - 3\psi_{i+1,j} + \psi_{i+2,j}). \end{aligned} \tag{3.6}$$

In these equations, and all subsequent equations where there can be no confusion, the superscript n, indicating the time level, is left off to simplify the notation. The second dimension can similarly be constructed by changing the indices and using the proper Courant number. It is immediately apparent from these equations and Eq (3.3) or (3.4) that for this case of two-dimensional advection, the number of grid-points involved to change the value of a single grid-point is $(N+1)^2$, where N is the order of the advection scheme chosen. Though an improvement on the first order forward-upstream advection, it still has relatively high dispersion (in rotating cone tests they found maximum amplitude decreases from 10 to 0.56 for first order scheme, 4.57 for second order scheme and 8.04 for fourth order scheme), and these higher order schemes do not preserve the sign of the advected quantity.

These schemes are inherently linear and as such the adjoint is just the transpose of the advection matrix.

3.2 Positive definite schemes

Bott (1989) presented a modification to the Tremback et al (1987) scheme to make it positive definite. This scheme reduces the flux such that at maximum any box may be vacated. To keep the notation simple the one dimensional form is presented. Two dimensional advection follows directly from successive application. He defined the flux as

$$F_j = \frac{i_j^+}{i_j}\psi_j - \frac{i_j^-}{i_{j+1}}\psi_{j+1} \tag{3.7}$$

$$i_j^+ = \max\Big(0, I^+(\alpha_j)\Big) \tag{3.8}$$

$$i_j^- = \max\Big(0, I^-(\alpha_j)\Big) \tag{3.9}$$

$$i_j = \max(I_j, i_j^+ + i_{j-1}^- + \epsilon) \tag{3.10}$$

$$I_{i,j}^+(\alpha_j) = \sum_{k=0}^{l} \frac{a_{j,k}}{(k+1)2^{k+1}} \left[1 - (1 - 2\alpha_j^+)^{k+1}\right] \tag{3.11}$$

$$I_{i,j}^-(\alpha_j) = \sum_{k=0}^{l} \frac{a_{j+1,k}}{(k+1)2^{k+1}} (-1)^k \left[1 - (1 - 2\alpha_j^-)^{k+1}\right] \tag{3.12}$$

$$I_{i,j} = \sum_{k=0}^{l} \frac{a_{j,k}}{(k+1)2^{k+1}} \left[(-1)^k + 1\right] \tag{3.13}$$

where the abbreviations $\alpha_j^{\pm} = \pm(\alpha_j \pm |\alpha_j|)/2$ have been introduced and the factors a_k depends on the order of the advection scheme. For the second order scheme ($l = 1$) equivalent to the Tremback et al. (1987) scheme he derived $a_{j,0} = \psi_j$ and $a_{j,1} = \psi_{j+1} - \psi_j$, while the fourth order scheme has $a_{j,1} = 1/6(-\psi_{j+2} + 6\psi_{j+1} - 3\psi_j - 2\psi_{j-1})$, $a_{j,2} = 1/2(\psi_{j+1} - 2\psi_j + \psi_{j-1})$ and $a_{j,3} = 1/6(\psi_{j+2} - 3\psi_{j+1} + 3\psi_j - \psi_{j-1})$.

Any positive definite advection scheme is by nature nonlinear, so that the adjoint operator is no longer trivial. The tangent linear operator of the second order Bott scheme will be presented next. For the application in the adjoint model it is necessary that the same path through conditional statements be followed during the backward integration of the adjoint model as was used during the forward integration. It is therefore necessary that all possible outcomes of the advection scheme be considered separately. All scenarios for Bott's formula for flux through the right-hand side the box marked J is given in Fig (3.1). Each case will be individually discussed.

Fig (3.1a) shows the case where the flux is out of the box through the right boundary, and into the box through the left boundary. The flux is given by

$$F_j = \begin{cases} \psi_j & \text{if } i_j^+ \geq \psi_j \\ i_j^+ & \text{if } i_j^+ < \psi_j \end{cases} . \tag{3.14}$$

The Jacobian matrix elements in this case will be

$$\mathbf{F}(j,j) = \begin{cases} 1 & \text{if } i_j^+ \geq \psi_j \\ \frac{\partial i_j^+}{\partial \psi_j} & \text{if } i_j^+ < \psi_j \end{cases} \tag{3.15}$$

$$\mathbf{F}(j,j+1) = \begin{cases} 0 & \text{if } i_j^+ \geq \psi_j \\ \frac{\partial i_j^+}{\partial \psi_{j+1}} & \text{if } i_j^+ < \psi_j \end{cases} . \tag{3.16}$$

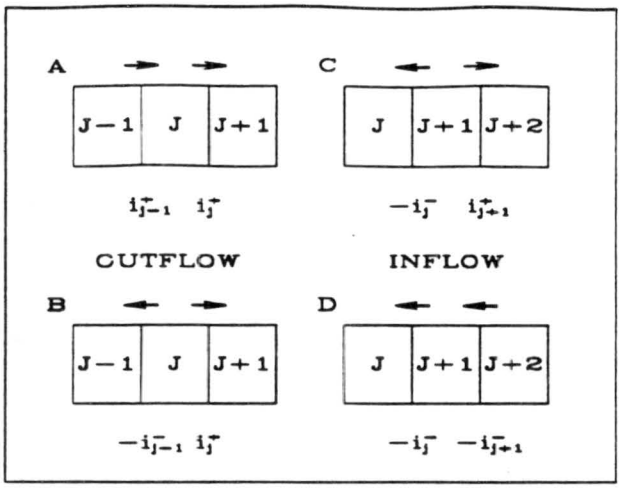

Figure 3.1: Possible scenarios from calculating the second order Bott (1989) fluxes. The arrows indicate the direction of the flux through the boundary.

The scenario in Fig (3.1b) is more complex. The box experiences divergent flow, and the flux is given by

$$F_j = \begin{cases} \frac{i_j^+ \psi_j}{i_j^+ + i_{j-1}^-} & \text{if } i_j^+ + i_{j-1}^- \geq \psi_j \\ i_j^+ & \text{if } i_j^+ + i_{j-1}^- < \psi_j \end{cases} \tag{3.17}$$

so that the Jacobian elements are given by

$$\mathbf{F}(j,j) = \begin{cases} \frac{\partial}{\partial \psi_j}\left(\frac{i_j^+ \psi_j}{i_j^+ + i_{j-1}^-}\right) & \text{if } i_j^+ + i_{j-1}^- \geq \psi_j \\ \frac{\partial i_j^+}{\partial \psi_j} & \text{if } i_j^+ + i_{j-1}^- < \psi_j \end{cases} \tag{3.18}$$

$$\mathbf{F}(j,j+1) = \begin{cases} \frac{\partial}{\partial \psi_{j+1}}\left(\frac{i_j^+ \psi_j}{i_j^+ + i_{j-1}^-}\right) & \text{if } i_j^+ + i_{j-1}^- \geq \psi_j \\ \frac{\partial i_j^+}{\partial \psi_{j+1}} & \text{if } i_j^+ + i_{j-1}^- < \psi_j \end{cases} \tag{3.19}$$

The Jacobian elements in general will be more expensive to calculate than the forward advection, since they frequently have more terms for each element.

When the right hand boundary is an inflow boundary, it must first be determined whether the donor box $(J+1)$ is experiencing divergent flow. If it does, as in Fig (3.1c),

the flux is given by

$$F_j = \begin{cases} \frac{-i_j^- \psi_{j+1}}{i_{j+1}^+ + i_j^-} & \text{if } i_{j+1}^+ + i_j^- \geq \psi_{j+1} \\ -i_j^- & \text{if } i_{j+1}^+ + i_j^- \geq \psi_{j+1} \end{cases} \tag{3.20}$$

so that the Jacobian elements are given by

$$\mathbf{F}(j,j) = \begin{cases} -\frac{\partial}{\partial \psi_j}\left(\frac{i_j^- \psi_{j+1}}{i_{j+1}^+ + i_j^-}\right) & \text{if } i_{j+1}^+ + i_j^- \geq \psi_{j+1} \\ -\frac{\partial i_j^-}{\partial \psi_j} & \text{if } i_{j-1}^+ + i_j^- \geq \psi_{j+1} \end{cases} \tag{3.21}$$

$$\mathbf{F}(j,j+1) = \begin{cases} -\frac{\partial}{\partial \psi_{j+1}}\left(\frac{i_j^- \psi_{j+1}}{i_{j+1}^+ + i_j^-}\right) & \text{if } i_{j+1}^+ + i_j^- \geq \psi_{j+1} \\ -\frac{\partial i_j^-}{\partial \psi_{j+1}} & \text{if } i_{j+1}^+ + i_j^- \geq \psi_{j+1} \end{cases} \tag{3.22}$$

$$\mathbf{F}(j,j+2) = \begin{cases} -\frac{\partial}{\partial \psi_{j+2}}\left(\frac{i_j^- \psi_{j+1}}{i_{j+1}^+ + i_j^-}\right) & \text{if } i_{j+1}^+ + i_j^- \geq \psi_{j+1} \\ 0 & \text{if } i_{j+1}^+ + i_j^- \geq \psi_{j+1} \end{cases} . \tag{3.23}$$

The final option is presented in Fig (3.1d). In that case the flux is given by

$$F_j = \begin{cases} -\psi_{j+1} & \text{if } i_j^- \geq \psi_{j+1} \\ -i_j^- & \text{if } i_j^- < \psi_{j+1} \end{cases} . \tag{3.24}$$

The Jacobian matrix elements are

$$\mathbf{F}(j,j) = \begin{cases} 0 & \text{if } i_j^- \geq \psi_{j+1} \\ -\frac{\partial i_j^-}{\partial \psi_j} & \text{if } i_j^- < \psi_{j+1} \end{cases} \tag{3.25}$$

$$\mathbf{F}(j,j+1) = \begin{cases} 1 & \text{if } i_j^- \geq \psi_{j+1} \\ -\frac{\partial i_j^-}{\partial \psi_{j+1}} & \text{if } i_j^- < \psi_{j+1} \end{cases} . \tag{3.26}$$

There are eight possible realizations for any given flux in this one-dimensional advection problem. Clearly, if two-dimensional sequential advection is considered, there would be 64 realizations that have to be considered, and for three dimensional sequential updating it would be 512. Although the formulation of each matrix element is not difficult, engineering aspects become complex for 2 or 3 dimensional cases when using sequential updating. Although sequential updating (Eq (3.4)) results in a stable algorithm, and also is less expensive than other options for forward integration (Tremback et al., 1987), the

complexity in formulating the adjoint may negate the advantages of having a positive definite advection scheme.

A different approach to this multi-dimensional problem was suggested by Boris and Book (1973). They suggested that the fluxes should be some combination of a lower order monotone scheme and a higher order scheme. This procedure is called the Flux-Corrected-Transport (FCT) schemes. The idea was further developed by Zalesak (1979), and by Smolarkiewicz (1983;1984;1989), Smolarkiewicz and Clark (1986) and Smolarkiewicz and Grabowski (1990). Consider Eq (3.2), where, in this case, F and G are fluxes calculated using an arbitrary higher order scheme. Let F^L and G^L be fluxes calculated by a low order monotonic scheme. The FCT scheme constructs the net transportive flux point by point (nonlinearly) as a weighted average of the flux computed by the low order scheme and that by the high order scheme. The weighting is done in a manner which insures that the high order flux is used to the greatest extent possible without producing oscillations (Zalesak, 1979). This is done in the following way: define a "anti-diffusive flux"

$$\begin{aligned} AF_{i,j} &= F_{i,j} - F^L_{i,j} \\ AG_{i,j} &= G_{i,j} - G^L_{i,j} \end{aligned} \tag{3.27}$$

These anti-diffusive fluxes are then limited such that

$$\begin{aligned} AF^C_{i,j} &= C \times F_{i,j} \times AF_{i,j} \\ AG^C_{i,j} &= C \times G_{i,j} \times AG_{i,j}, \end{aligned} \tag{3.28}$$

where the C-coefficients are determined from the set of constraints

$$0 \le CF, \qquad CG \le 1 \tag{3.29}$$

and

$$\psi^{MAX}_{i,j} \ge \psi^{n+1}_{i,j} = \Psi^{n+1}_{i,j} - \left(AF^C_{i,j} - AF^C_{i-1,j} + AG^C_{i,j} - AG^C_{i,j-1}\right) \ge \psi^{MIN}_{i,j}, \tag{3.30}$$

where Ψ denotes the solution given by the lower-order scheme, and

$$\psi^{MAX}_{i,j} = \max(\psi_{i-1,j}, \psi_{i,j-1}, \psi_{i,j}, \psi_{i+1,j}, \psi_{i,j+1},$$

$$\Psi_{i-1,j}, \Psi_{i,j-1}, \Psi_{i,j}, \Psi_{i+1,j}, \Psi_{i,j+1})$$

$$\psi_{i,j}^{MIN} = \min(\psi_{i-1,j}, \psi_{i,j-1}, \psi_{i,j}, \psi_{i+1,j}, \psi_{i,j+1},$$

$$\Psi_{i-1,j}, \Psi_{i,j-1}, \Psi_{i,j}, \Psi_{i+1,j}, \Psi_{i,j+1}). \tag{3.31}$$

The two conditions Eqs (3.29) and (3.30) lead to the following formulation of the constrained anti-diffusive fluxes (Smolarkiewicz and Grabowski, 1990)

$$\begin{aligned} AF_{i,j}^{C} &= \min\left(1, \beta_{i,j}^{\downarrow}, \beta_{i+1,j}^{\uparrow}\right) \max(0, AF_{i,j}) \\ &\quad + \min\left(1, \beta_{i,j}^{\uparrow}, \beta_{i+1,j}^{\downarrow}\right) \min(0, AF_{i,j}) \\ AG_{i,j}^{C} &= \min\left(1, \beta_{i,j}^{\downarrow}, \beta_{i,j+1}^{\uparrow}\right) \max(0, AG_{i,j}) \\ &\quad + \min\left(1, \beta_{i,j}^{\uparrow}, \beta_{i,j+1}^{\downarrow}\right) \min(0, AG_{i,j}) \end{aligned} \tag{3.32}$$

where

$$\begin{aligned} \beta_{i,j}^{\uparrow} &= \frac{\psi_{i,j}^{MAX} - \Psi_{i,j}^{n+1}}{A_{i,j}^{IN} + \epsilon} \\ \beta_{i,j}^{\downarrow} &= \frac{\Psi_{i,j}^{n+1} - \psi_{i,j}^{MIN}}{A_{i,j}^{OUT} + \epsilon}. \end{aligned} \tag{3.33}$$

and $A_{i,j}^{IN}$, $A_{i,j}^{OUT}$ are the absolute values of the total incoming and outgoing anti-diffusive fluxes (Eq (3.27)) from the (i,j)th grid box, respectively. The parameter ϵ is a small value, to ensure that the β-functions are defined at all times. This scheme is by definition stable, and is a general sign-preserving scheme. The algorithm for the application of this scheme is then as follows:

1. Compute F^L and G^L by a low order monotonic scheme
2. Compute F^H and G^H by a high order scheme
3. Compute the anti-diffusive fluxes (Eq (3.27))
4. Compute the low order time advance solution (Eq (3.2))
5. Limit the anti-diffusive fluxes (Eq (3.32))
6. Apply the limited anti-diffusive fluxes (Eq (3.2))

This scheme is computationally much more expensive than the Bott (1989) scheme.

The adjoint of this advection scheme will be considered by looking at the $\beta^{\downarrow}_{i,j}$ function. The term $A^{MAX}_{i,j}$ depends on any 1 of 16 combinations of velocities at the edges of the box, while the term $\psi^{MAX}_{i,j}$ may have any 1 of 10 realizations. Thus, just from these two terms, $\beta^{\downarrow}_{i,j}$ may have any one of 160 realizations. This clearly shows that the adjoint of this scheme will be even more complex than the Bott scheme.

This highlights an important issue when considering the use of an adjoint model for initialization. Specifically, desirable techniques in the current state-of-the-art models may lead to impractical adjoint models. Thus, a decision to use the adjoint technique of data assimilation may force compromises in the construction of the formulation of the forward model. The other alternative would be that an simplified model of the full complex forecast model may be used during the assimilation phase. Even though the initial fields may not be completely balanced in terms of the forecast model, this initialization should be an improvement on the spacial interpolation techniques and the horizontally homogeneous initialization procedures used today to initialize most non-hydrostatic models.

3.3 Application

The second order Bott scheme (Eqs (3.7—3.13)) was coded for a one dimensional advection problem to illustrate that even though the positive definite advection schemes in general result in complex adjoints models, that the technique as such does not fail.

An identical twin experiment was conducted where the model was first used to produce a set of observations, and then was run from a perturbed initial condition in conjunction with the adjoint model and the optimization code to retrieve the original condition.

The scheme was implemented for a 100 grid-point model with velocities defined on a staggered grid. The assimilation was done over a 100 time step interval, with $\delta t = 1$s, $\Delta x = 5$m, while the velocities were defined by

$$u_j = \begin{cases} -2.5 & \text{if } j < 48 \\ 1.25j - 62.5 & \text{if } 48 \leq j \leq 52 \\ 2.5 & \text{if } j > 52 \end{cases} , \tag{3.34}$$

such that the maximum Courant number was $\alpha = 0.5$. The base run was initialized with a normal distribution with mean 50.5 and a standard deviation of 10 (solid line in Fig (3.2a)).

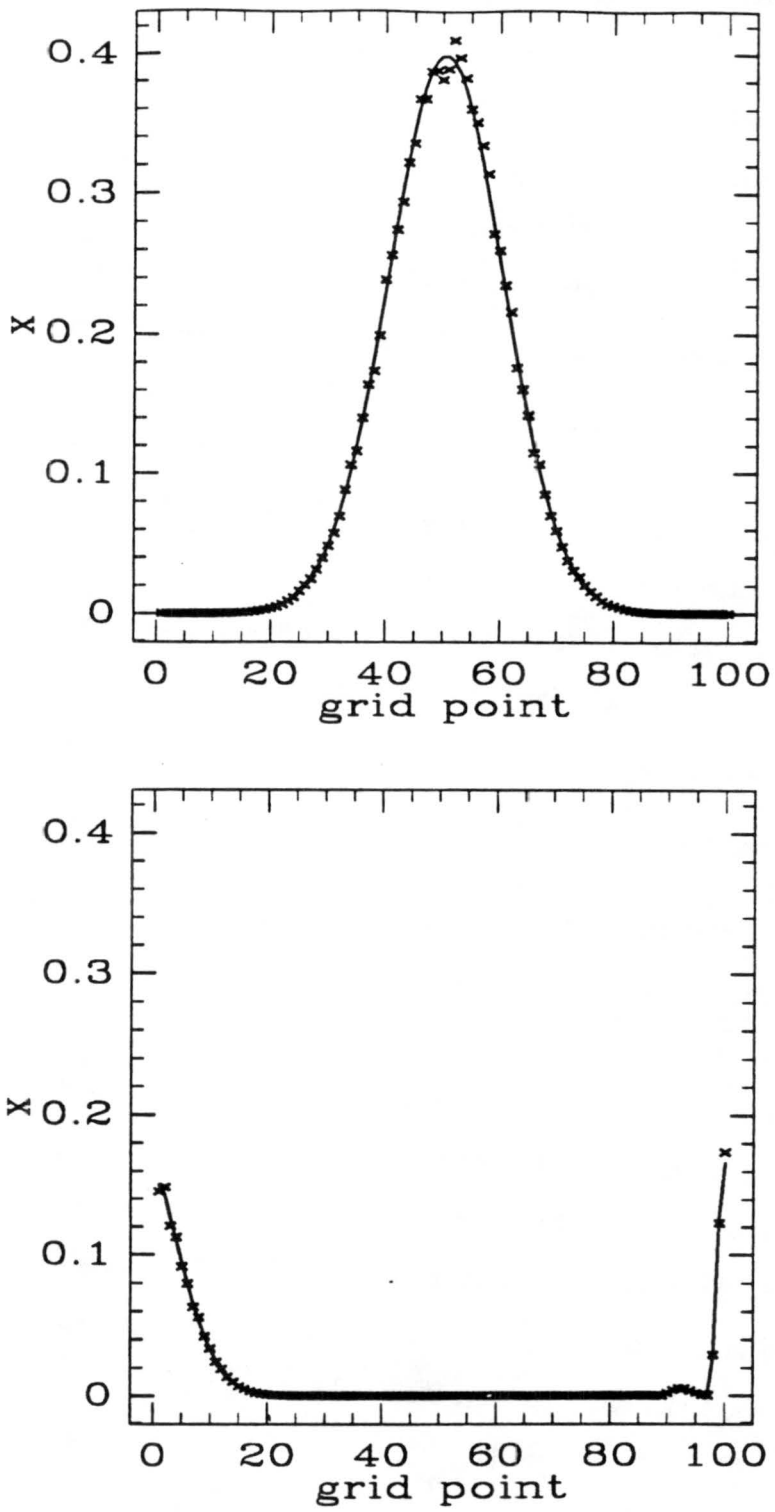

Figure 3.2: Initial (top) and final (bottom) true states of the advected fields (solid curves). The crosses indicate the observations used in the retrieval run.

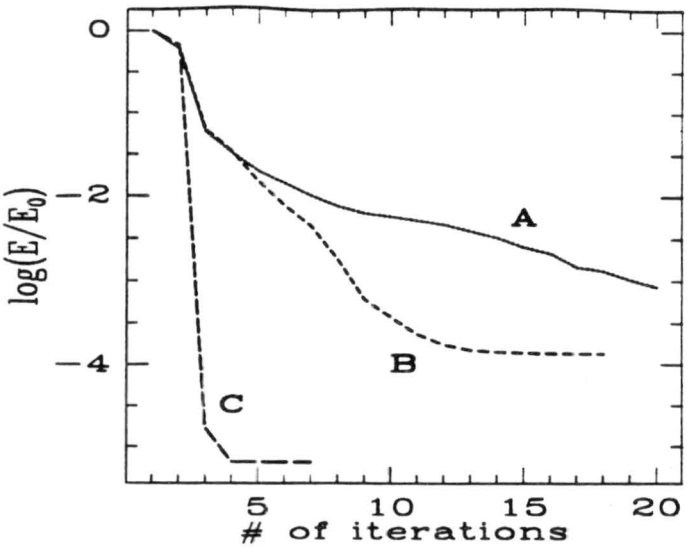

Figure 3.3: Cost function decline as a function of number of iterations for three experiments. The cost function was normalized by the initial cost for each experiment. The first experiment (A) used observations at every timestep, the second (B) used observations at every 10'th timestep, while (C) only used observations from the initial and final timestep.

'Observations' taken from this run were perturbed by a random value between 0 and 10% of the true value. The retrieval run was initialized with values varying randomly between 0 and 1. Results of three experiments, each using different amounts of observations, are shown in Fig (3.3).

These experiments show that a gross overdetermination of the number of degrees of freedom by the observations lead to slower convergence. It can also be seen that even for this non-linear problem, convergence can occur fast (4 iterations in (A)).

Thacker (1987) compared data assimilation to nonlinear regression, where it is essential that there should always be much more imperfectly sampled data than the number of adjustable parameters. Here it is shown again that compromises may have to be made between accuracy and cost. Another set of experiments were conducted to investigate Thacker's statement. The number of evenly spaced observations used in each retrieval run was increase from 2 to 11, and 20 iterations were allowed during each retrieval run. The results are presented in Table (3.1).

The marginal improvement of only 0.6% when using only the initial and final time as observations is no surprise considering Fig (3.2). It can be seen that at the final time

Table 3.1: Improvement in the error of the initial condition (as described by the improvement over the errors in the observations) as a function of the number of observations used in the retrieval

number of observations	% improvement
2	0.6
3	10.0
5	15.9
6	25.5
11	35.5

most of the higher values have been advected out of the domain, and with it the useful information. Small perturbations on the initial condition will have little or no impact on the final solution. Where in-between observations are used, however, it is noticed that with each increase in number of observations included, there is an increase in the improvement. Combining the results of Fig (3.3) and Table (3.1) gives an estimate of the cost-accuracy factor: To improve the initial condition by 35% required a fourfold increase in the computational cost.

Experiments with time-varying velocity fields exhibited similar trends and conversion rates as these shown here. It is therefore concluded that the formulation of the adjoint of the general Bott positive difinite scheme is feasible and useful, and that for the one-dimensional formulation it leads to realistic convergence rates in an initialization problem. It was also shown that the techniques produces improved initial conditions even with an increase in inaccurate observations, but that this comes at the cost of additional computational expense.

Chapter 4

A KINEMATIC ONE-DIMENSIONAL MODEL

A simple one-dimensional model was developed to investigate the feasibility of the optimal control theory approach where a complex physical process is modelled. The model is a warm rain microphysical model, with prognostic variables liquid water potential temperature θ_l, total water mixing ratio r_t, and cloud water mixing ratio r_r. Temperature T, cloud water mixing ratio r_c, saturation vapor mixing ratio r_s and air density ρ are diagnosed from the prognostic variables, while the wind and pressure fields are prescribed as 'observables'. Physical processes included in the model are condensation/evaporation (implicit in liquid water potential temperature), autoconversion of cloud water to rain water, collection of cloud water by rain water, and the sedimentation of rain water.

4.1 The forward model

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

$$\frac{\partial q}{\partial t} = -w\frac{\partial q}{\partial z} + S_q \tag{4.1}$$

where q represents any of the three prognostic variables (θ_l, r_t or r_r) and S_q represents the sources and sinks for each respectively.

Advection is done with the Bott (1989) positive definite scheme described in Section (3.2) in Chapter 3. θ_l and r_t are conservative variables in the absence of precipitation, therefore the only source/sink for both these variables will be due to rain fallout. The source/sinks of rain are autoconversion of cloud to rain, collection of cloud by rain and rain fallout. It was assumed that the rain water was distributed in size according to

the distribution given by Marshall and Palmer (1948) ($n(D) = N_0 \exp(-\Lambda D)$, N_0 is the intercept and Λ is the slope).

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

$$\frac{dr_r}{dt}\Big|_{prec} = -\frac{dV_{Tm} r_r}{dz}, \tag{4.2}$$

where $V_T = c_v D_m^{p_v} \Gamma(p_m + p_v + 1)$ is the mass weighted mean terminal fall velocity. D_m is the mean diameter of the distribution. All constants used here and later in the text are defined in Appendix (A). It will be assumed that cloud water has negligible terminal fall speed, thus the precipitation is made up entirely by rain fallout. The source/sink term for θ_l is then given by Tripoli and Cotton (1981) as

$$\frac{d\theta_l}{dt}\Big|_{prec} = -\frac{l_v}{c_p T}\frac{\theta_l^2}{\theta}\frac{dr_r}{dt}\Big|_{prec} \tag{4.3}$$

while the source/sink term for r_t is just the precipitation flux from the box, or

$$\frac{dr_t}{dt}\Big|_{prec} = \frac{dr_r}{dt}\Big|_{prec}. \tag{4.4}$$

Here θ is the potential temperature and $\frac{dr_r}{dt}|_{prec}$ is the change in r_r due to rain fallout.

Collection is modeled using the accretion equation. From Flatau et al. (1989) the collection rate can be written in terms of the prognostic and diagnostic variables as

$$\begin{aligned}\frac{dr_r}{dt}\Big|_{col.} &= \left\{\frac{\pi}{4}\Gamma(p_v+2)c_v N_0 \left(\frac{p_0}{p}\right)^{0.4}\left(\frac{p}{\pi N_0 \rho_l R_d}\right)^{7/8}\right\}\left(\frac{r_r}{T}^{7/8}\right) r_c \\ &= C_1\left(\frac{r_r}{T}\right)^{7/8} r_c \end{aligned} \tag{4.5}$$

where C_1 is a constant in this model, and a collection efficiency of one has been assumed. The autoconversion rate is taken from Kessler (1969)

$$\frac{dr_r}{dt}\Big|_{auto} = \alpha\rho(r_c - r_0)\mathrm{H}(r_c - r_0) \tag{4.6}$$

where H is the Heaviside step function, α is the rate coefficient for and r_0 is a mixing ratio threshold for autoconversion, while ρ is the air density.

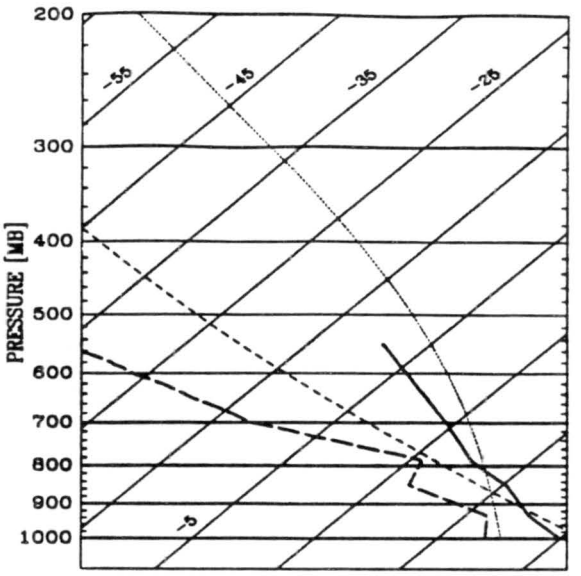

Figure 4.1: Sounding used to initialize the model. The solid line is the temperature profile, the long dashed line is the dew-point profile. The short-dash line is the 300 K potential temperature line, and the fine dotted line is the 335 K adiabat.

The temperature and the cloud water content can be diagnosed from the three prognostic variables. These two variables are interrelated through θ_l,

$$T = \left(\frac{p}{p_0}\right)^{R/c_p} \theta_l \left[1 + \frac{l_v}{c_p T}(r_r + r_c)\right] \tag{4.7}$$

$$r_c = (r_t - r_s - r_r)\mathrm{H}(r_t - r_s - r_r) \tag{4.8}$$

where the saturation vapor pressure is given by the formulation due to Murray (1967)

$$r_s = d \exp\left[a\left(\frac{T-b}{T-c}\right)\right], \tag{4.9}$$

where the constants are $d = 610.78$, $a = 17.2694$, $b = 273.16$ and $c = 35.86$. The temperature can be determined directly from the definition of θ_l if the cloud water content is known, but for the determination of the cloud water content the temperature is needed to determine the saturation mixing ratio. The diagnostic scheme is a simple bisection iteration scheme of 11 iterations, which determines the temperature to within 0.1° accuracy.

The model was initialized with the sounding from Fig (4.1). In this simulation the advective velocity w was held constant at $w = 2.5\ \mathrm{ms}^{-1}$. The model was integrated for 500

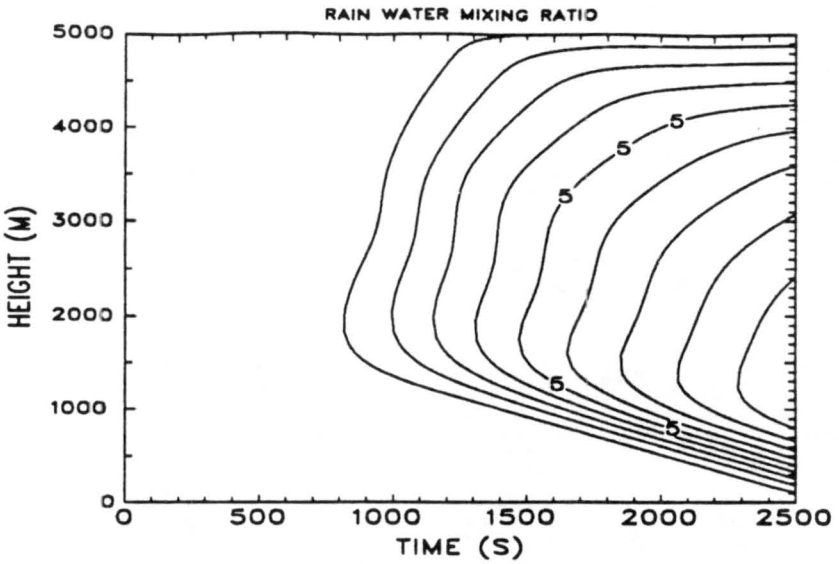

Figure 4.2: Time evolution of the rain water content with time. Contour intervals are given in g/kg.

time steps, with a time step of 5 seconds. The model domain consisted of 100 grid-points, with a 50 m grid-point separation. The outflow boundaries were open boundaries, while the inflow boundaries were held steady.

Figure (4.2) shows the time-height evolution of the rain water content in the simulation. The cloud begins to develop 160 second into the simulation, with the formation of the first rain delayed by another 150 seconds. The cloud base varied between 800 and 1000 m during the course of the simulation. The initial rain drops are very small, and are easily supported by the updraft. As the rain water content grows it begins to fall against the updraft. The first precipitation falls through the cloud base about 25 minutes after the initialization of the cloud. The precipitation shower reaches the ground 16 minutes later. The maximum rain water content reached in this simulation is 9.2 g/kg.

4.2 The tangent linear model

The cost function is as defined in Eq (2.2), where the state vector for this model was constructed in the form of Eq (2.25), with $f_1 = \theta_l$, $f_2 = r_t$ and $f_3 = r_r$. Thus, the tangent

linear matrix has a banded form. For this simulated experiment it is assumed that the three prognostic variables can be observed.

The contributions from the advection scheme has already been discussed in chapter 3, and will not be repeated here. The three prognostic variables add no further complications. The contributions of the forcing terms S_{θ_l}, S_{r_t} and S_{r_r}, and the contribution of the precipitation fallout will be derived.

First the derivatives of the diagnostic variables with respect to the prognostic variables will be derived. All the diagnostic variables are grid-point dependent; for clarity of the derivation the grid-point indicator was left out. Substitute Eq (4.9) into Eq (4.8), which, assuming $r_c > 0$, when substituted into Eq (4.7), results in a single equation for temperature, dependent only on the prognostic variables and itself,

$$T = \left(\frac{p}{p_0}\right)^{R/c_p} \theta_l \left\{1 + \frac{l_v}{c_p T}\left(r_t - d\exp\left[a\left(\frac{T-b}{T-c}\right)\right]\right)\right\}. \tag{4.10}$$

Now the derivatives may be taken. It is clear that the temperature has no dependence on the rain water mixing ratio, thus

$$\frac{\partial T}{\partial r_r} = 0, \tag{4.11}$$

while it can be shown that

$$\frac{\partial T}{\partial r_t} = \frac{\frac{l_v}{c_p}\left(\frac{p}{p_0}\right)^{R/c_p}\theta_l}{\left\{2T - \left(\frac{p}{p_0}\right)^{R/c_p}\theta_l\left[1 - \frac{l_v}{c_p}a\frac{(b-c)}{(T-c)^2}r_s\right]\right\}} \tag{4.12}$$

$$\frac{\partial T}{\partial \theta_l} = \frac{\left(\frac{p}{p_0}\right)^{R/c_p}\left[T + \frac{l_v}{c_p}(r_t - r_s)\right]}{\left\{2T - \left(\frac{p}{p_0}\right)^{R/c_p}\theta_l\left[1 - \frac{l_v}{c_p}a\frac{(b-c)}{(T-c)^2}r_s\right]\right\}}. \tag{4.13}$$

In the case where there is no cloud water the derivatives are

$$\frac{\partial T}{\partial r_r} = \frac{\frac{l_v}{c_p}\left(\frac{p}{p_0}\right)^{R/c_p}\theta_l}{2T - \left(\frac{p}{p_0}\right)^{R/c_p}\theta_l} \tag{4.14}$$

$$\frac{\partial T}{\partial r_t} = 0 \tag{4.15}$$

$$\frac{\partial T}{\partial \theta_l} = \frac{-\left(\frac{p}{p_0}\right)^{R/c_p}\left[T + \frac{l_v}{c_p}r_r\right]}{2T - \left(\frac{p}{p_0}\right)^{R/c_p}\theta_l}. \tag{4.16}$$

Since these derivatives are calculated during the forward integration of the model, it is always known *a priori* whether there is cloud water present, and thus which of the two sets of formulas to use.

The expression for r_c in terms of the prognostic variables is

$$r_c = r_t - r_r - d\exp\left[a\left(\frac{T-b}{T-c}\right)\right]. \tag{4.17}$$

Again, it can easily be seen that

$$\frac{\partial r_c}{\partial r_r} = -1 \tag{4.18}$$

while it can be shown that

$$\frac{\partial r_c}{\partial r_t} = 1 - \frac{b-c}{(T-c)^2}\frac{\partial T}{\partial r_t}ar_s \tag{4.19}$$

$$\frac{\partial r_c}{\partial \theta_l} = -\frac{b-c}{(T-c)^2}\frac{\partial T}{\partial \theta_l}ar_s. \tag{4.20}$$

With these results the derivatives of the source terms can easily be derived. The source terms for rain water mixing ratio are given by Eqs (4.5) and (4.6). The derivatives of Eq (4.5) are given in generic form by

$$\frac{\partial S_{r_r}}{\partial x}\Big|_{coll} = \frac{7}{8}C_1\left(\frac{r_r}{T}\right)^{-1/8}\left(\delta_{r_r}\frac{1}{T} - \frac{1}{T^2}\frac{\partial T}{\partial x}\right)r_c + C_1\left(\frac{r_r}{T}\right)^{7/8}\frac{\partial r_c}{\partial x} \tag{4.21}$$

where $\delta_{r_r} = 1$ when $x = r_r$ and zero otherwise. The derivatives of Eq (4.6) are given by

$$\frac{\partial S_{r_r}}{\partial x}\Big|_{auto} = \alpha\rho\left(\frac{\partial r_c}{\partial x} - (r_c - r_0)\frac{1}{T}\frac{\partial T}{\partial x}\right). \tag{4.22}$$

In the above equations x may be any of the prognostic variables (θ_l, r_t or r_r), and the grid-point indicators have again been omitted since these processes are single grid-point processes.

The terms related to precipitation fallout will be derived next. The discrete form of $S_{r_r}|_{prec}$ is given by

$$S_{r_r}(k)\Big|_{prec} = -\left(\frac{c_v\Gamma(p_m + p_v + \nu)}{\Delta z}\left(D_m^{p_v}(k+1)r_r(k+\nu) - D_m^{p_v}(k)r_r(k)\right)\right) \tag{4.23}$$

where k is the grid-point indicator and $\nu = 1$ for the Marshall-Palmer distribution. The derivatives are then given by

$$\frac{\partial S_{r_r}(k)}{\partial x(k)}\Big|_{prec} = -\frac{c_v\Gamma(p_m + pv + \nu)}{\Delta z} \times \left(p_v r_r D_m^{p_v-1}(k)\frac{\partial D_m(k)}{\partial x(k)} + D_m^{p_v}(k)\frac{\partial r_r(k)}{\partial x(k)}\right). \quad (4.24)$$

and

$$\frac{\partial S_{r_r}(k)}{\partial x(k+1)}\Big|_{prec} = -\frac{c_v\Gamma(p_m + p_v + \nu)}{\Delta z} \times \left(p_v r_r D_m^{p_v-1}(k+1)\frac{\partial D_m(k+1)}{\partial x(k+1)} + D_m^{p_v}(k+1)\frac{\partial r_r(k+1)}{\partial x(k+1)}\right) \quad (4.25)$$

Using these results, the derivatives for the source term for liquid water potential temperature can be written as

$$\frac{\partial S_{\theta_l}(k)}{\partial x(k+1)}\Big|_{prec} = -\frac{l_v}{c_p T(k)}\frac{\theta_l^2(k)}{\theta(k)}\frac{\partial S_{r_r}(k)}{\partial x(k+1)}\Big|_{prec} \quad (4.26)$$

and

$$\frac{\partial S_{\theta_l}(k)}{\partial x(k)}\Big|_{prec} = -\frac{l_v}{c_p T(k)}\frac{\theta_l^2(k)}{\theta(k)}\frac{\partial S_{r_r}(k)}{\partial x(k)}\Big|_{prec} - 2l_v\left(\frac{p_0}{p(k)}\right)^{R_d/c_p}\left(\frac{\theta_l(k)}{T(k)}\right)^2 \times \left(\frac{1}{\theta_l(k)}\frac{\partial\theta_l(k)}{\partial x(k)} - \frac{1}{T(k)}\frac{\partial T(k)}{\partial x(k)}\right)\frac{dr_r(k)}{dt}\Big|_{prec}. \quad (4.27)$$

This completes the list of derivatives needed to form the tangent linear matrix. Since all terms are linear, they are additive. In the construction of the code it was decided that due to the small size of the problem, that the matrix will be computed during the forward integration of the model and stored for the backward integration. The size of the matrix for this problem is 8 times the size of the state vector. It is also feasible to keep an array of all the tangent linear matrices in memory for a problem of this size, if speed is required. A routine was developed to take advantage of the special form of the matrix in the matrix vector multiplication of Eq (2.8).

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. In this model, and in the model in Chapter (5), the prognostic variables

can vary by as much as eight to ten orders of magnitude. Optimization codes work best when all variables are in the range zero to one. 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

$$\nabla_{\mathbf{x}^0} E(\mathbf{x}^0)\Big|_{norm} = (w_1, \ldots, w_N) \nabla_{\mathbf{x}^0} E(\mathbf{x}^0). \tag{4.28}$$

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: 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 in E, 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 which 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 $\mathbf{W} = \sigma^{-2}\mathbf{I}$ where σ^2 is the variance of the measurement errors and $\mathbf{I}$ is the identity matrix (eg. Courtier and Talagrand, 1990). These matters will be investigated in the following section.

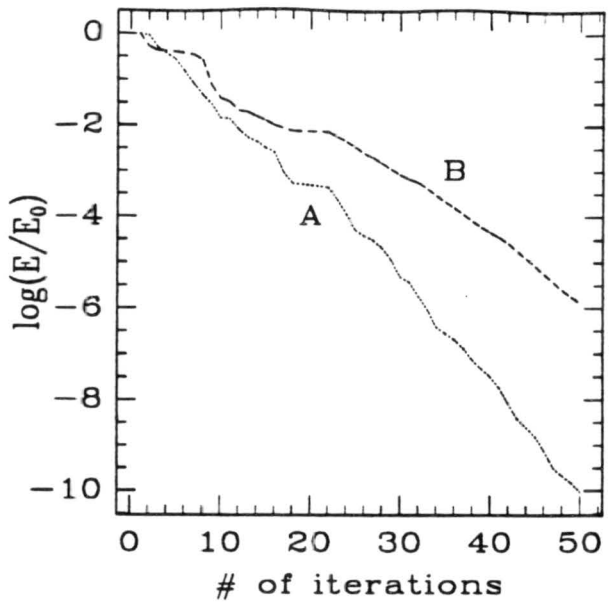

Figure 4.3: Cost function decline as a function of number of iterations for two experiments. The first experiment (A) has the weights set as $\Theta_l = 100$, $R_t = 0.04$ and $R_r = 0.025$, while the second (B) has the weights set to $\Theta_l = 300$, $R_t = 0.14$ and $R_r = 0.01$. Both experiments have the same startup initial condition.

4.3 Application

Several experiments were performed, using this model, to investigate factors influencing the rate of convergence. The first set of experiments were based on 'perfect' observations, while the second set were based on observations that were randomly perturbed.

4.3.1 Perfect observations

The experiments in this section all consist of running the forward model to create an observational data set. The assimilation is then done using different scaling factors, initial conditions and lengths of assimilation period.

Fig (4.3) shows results from two experiments using different sets of scaling factors. Experiment (A) has weights set to $\Theta_l = 100$ K, $R_t = 0.04$ and $R_r = 0.025$, while (B) has the weights set to $\Theta_l = 300$ K, $R_t = 0.14$ and $R_r = 0.01$. The assimilation period used here was from $t = 1500$s to $t = 2000$s (see Fig 4.2), where observations were available at $t = 1500$s, $t = 1750$s and $t = 2000$s. Profiles of constant θ_l, r_t and r_r were chosen as

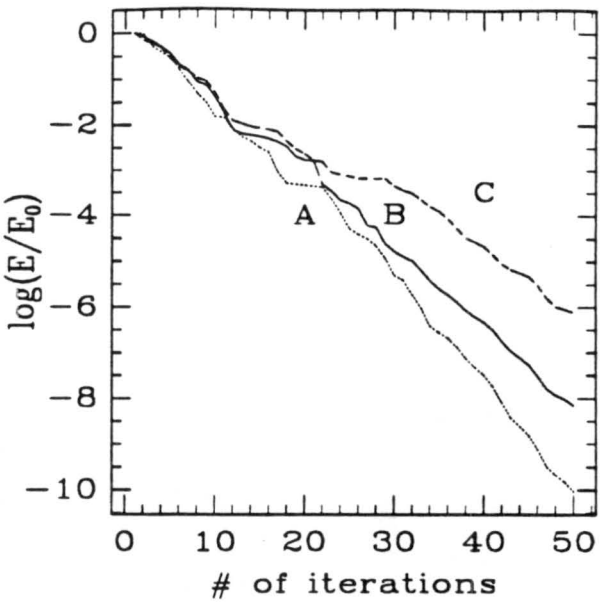

Figure 4.4: Cost function decline as a function of number of iterations for three experiments. Experiments (A) and (C) have the weights set as $\Theta_l = 100$, $R_t = 0.04$ and $R_r = 0.025$, while the third (B) has the weights set to $\Theta_l = 200$, $R_t = 0.04$ and $R_r = 0.025$. Experiments (B) and (C) have the same startup initial condition, while (A) has a different startup initial condition (see text for further details).

startup initial condition for the assimilation cycle. For both experiments the values were set as $\theta_l = 300$ K, $r_t = 0.01$ and $r_r = 0$. Scaling with the relative magnitudes of the unknown variables as suggested by Moore (1991) resulted in poor convergence rates.

Experiment (A) had the best results. The cost function was reduced by about ten orders of magnitude in 50 iterations. Further examination of the results revealed that the root mean square of the error in the initial conditions after 50 iterations were reduced to order -5 K for θ_l and order -7 g/kg 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) in 30 iterations.

The sensitivity of the convergence rate to the scaling of the control variables is demonstrated in Fig (4.3) by the difference in the results of experiments (A) and (B). The cost function only dropped by about 6 orders of magnitude in 50 iterations in (B), compared to the about 10 orders in (A). The associated RMS error for (B) was order -2 K for θ_l

and order -4 g/kg for the two mixing ratios. The cost in (B) can be reduced to the same as (A), but it takes many more iterations. Other experiments (not shown) give different (slower or faster) convergence.

Further insight in how scaling influence the convergence rate can be gained from Fig (4.4). In this figure, experiment (A) is the same as discussed above, while (C) has the same weights but a different startup initial condition ($\theta_l = 330$ K, $r_t = 10$ g/kg and $r_r = 1$ g/kg). (B) has the same startup initial conditions as (C), but different weights ($\Theta_l = 200$ K, $R_t = 0.04$ and $R_r = 0.025$). The startup initial conditions for (B) and (C) are further from the true initial condition, hence slower convergence rates may be expected. But it can be seen that the scaling which provided good convergence rates for (A) are not optimum for the different startup initial conditions of (B) and (C). Some insight in factors contributing to this can be gained by looking at the contribution of each of the three variable vectors to the cost function. For (A) the contributions to the cost for the first iteration were 1.32, 3.1 and 2.36 for the θ_l, r_t and r_r vectors respectively, for (B) these were 18.7, 3.9 and 3.67 respectively, and for (C) they were 4.67, 5.46 and 4.5. Comparing (B) to the other two, it can be noted that there the θ_l vector was the primary contributor to the cost, while in the (A) and (C) all variable vectors contributed approximately equally, but with θ_l contributing the least. Several other experiments (not shown) confirmed that when one of the variable vectors is the major contributor to the cost, convergence becomes slow. Thus the scaling should be such that each variable vector contributes similar values to the cost. The magnitude of the cost function is taken into account in the determination of the step length taken along the gradient by the minimization routine. When one variable dominates the cost, the step length is calculated to optimize that variable. However, this results in an overshoot for the other two variables, often resulting in an increase in the cost function instead of a decrease. To correct that, a shorter step length is taken, resulting in slower convergence rates. It was further found that with large initial magnitude of the cost function slow convergence rates may be expected. These results indicate that the often used practice of using weights related to the variance do indeed give best convergence.

In Fig (4.5) the impact of the length of the assimilation period is investigated. All these experiments used the same scaling and initial startup condition as experiment (A)

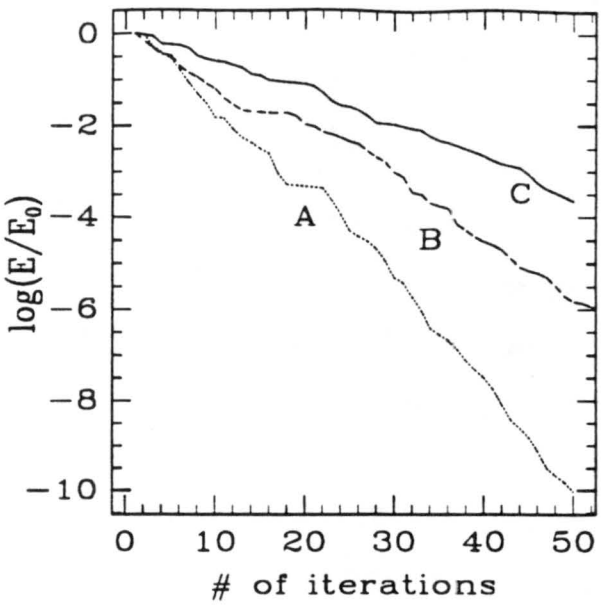

Figure 4.5: Cost function decline as a function of number of iterations for three experiments. All the experiments have the weights set as $\Theta_l = 100$, $R_t = 0.04$ and $R_r = 0.025$. Experiment (A) assimilates observations over 100 time steps, (B) assimilates observations over 200 time steps, and (C) assimilates observations over 300 time steps.

of Fig (4.3). In this case the assimilation was done over the time period t = 1500s to t = 2000s (A), t = 1000s to t = 2000s (B), and t = 500s to t = 2000s (C) (see Fig 4.2). 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 observations were used. This result should be expected since this is a highly non-linear model.

Up to now it was assumed that observations of all the prognostic variables are available. This may not always be the case. Fig (4.6) shows results of two experiments where no or little information on r_t was available. The startup initial condition for both cases were the true initial condition with an offset of 2 K in θ_l, 2 g/kg in r_t and 1 g/kg in r_r. The assimilation was done over the period t = 1500s to t = 2000s, with observations taken every 50 seconds. The first experiment (A) has no r_t observations, while the second (B) has r_t observations only at z = 750m and z = 2500m, but at all the observation

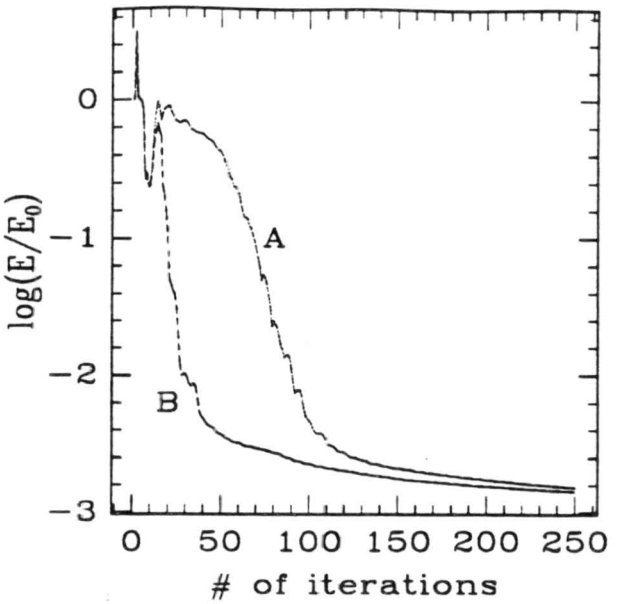

Figure 4.6: Decline of the error in r_t as a function of the number of iterations for two experiments. Both experiments assimilated data over 100 time steps, with observations at every 10'th step. In (A) no observations of r_t was used, while in (B) observations at only two of the hundred grid points were used. Both the experiments have the weights set as $\Theta_l = 100$, $R_t = 0.1$ and $R_r = 0.025$.

times. It was found that, even though r_t contributes nothing or very little to the cost, that the weight assigned to it remains important. This is so because there is a weight feedback through Eq (4.28). It was found that the principles derived earlier in this section worked fine. The weight for r_t had to be chosen such that the expected contribution of the r_t error was approximately the same as the cost of the other two variables. As can be expected, the convergence of r_t was much slower than the convergence of the other two variables which had full observations, but it can be seen that in both experiments the technique was tending towards 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_t drastically improved the rate of convergence in the early stages, but that after about 150 iterations both assimilations were approximately equal. The algorithm had significant problems correcting the r_t error at the upper boundary. This explains the convergence

of the two lines in the figure, at that stage the error statistics is dominated by the upper boundary and adjacent grid points of the initial condition. The inability of the technique to improve the upper few grid points can be explained by the setup of the model. Since the vertical velocity is held constant at a positive value, advection will quickly remove the initial field out of the domain. Thus observations at later times elsewhere in the domain will shed no information on what the gradient at those points should be. The only significant information that does contribute is observations of the upper boundary at the initial time. However, this is a limitation of this simple experiment, and not the technique as such. The opposite argument is valid for the lower boundary. Since those values are advected into the domain, even observations at the top of the domain at a much later time might add information on what the lower boundary should be. This results in good retrievals at the lower boundary.

4.3.2 Perturbed observations

The experiments in this section are all based on runs of the forward model where a random perturbation was added to the state vector before it was written to the observational data set. Maximum errors introduced to θ_l were ± 0.75 K, to r_t ± 1 g/kg and to r_r ± 0.35 g/kg. All experiments were conducted over the period $t = 1500$s to $t = 2000$, with the startup initial conditions of experiment (A) of Fig (4.6). Experiments using different number of observation times and weights were conducted.

Fig (4.7) shows the errors (compared to the 'truth') over the assimilation period after 50 iterations for each of the three prognostic variables. Both the maximum errors in the domain and the RMS errors of the domain are shown. These are contrasted with the errors in the observational data set. Results from three experiments are shown. (A) used three observation times, (B) used eleven while (C) used observations at every time step.

It can be seen that the statistical improvements in the initial conditions are marginal, but that later in the assimilation period significant improvements were obtained, in both the maximum and the RMS error. There was little improvement of the results when

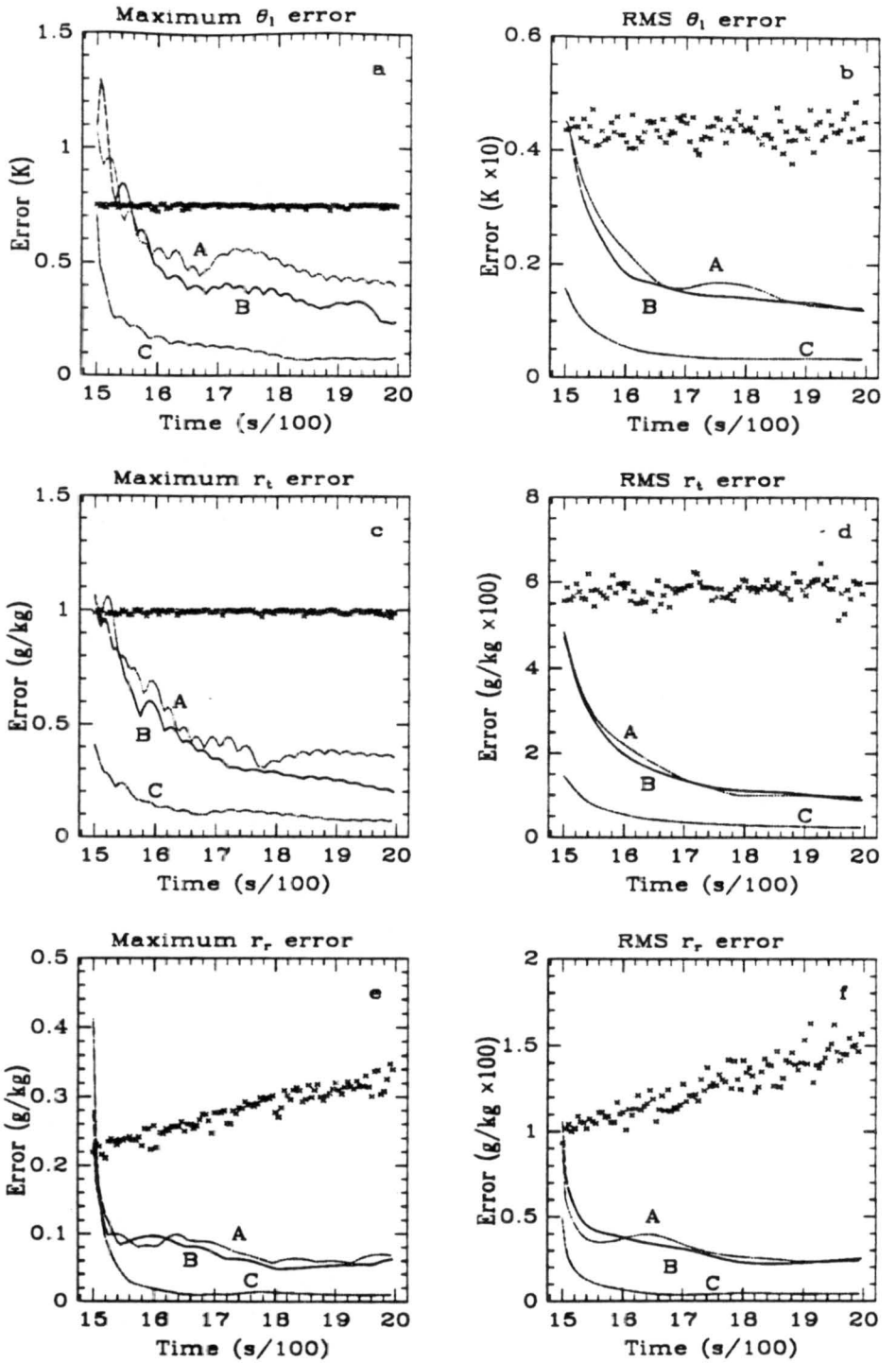

Figure 4.7: Errors over the assimilation period after 50 iterations for three experiments. The maximum error at each time step is displayed for θ_l, r_t and r_r in panels a, c and e respectively, and the root mean square error of the domain in panels b, d and f respectively. In experiment (A) 3 observation times were used, in (B) 11 observation times were used, while in (C) observations were taken at each time step. The crosses indicate the errors in the observational data set. The weights in all the experiments were set as $\Theta_l = 50$, $R_t = 0.04$ and $R_r = 0.01$.

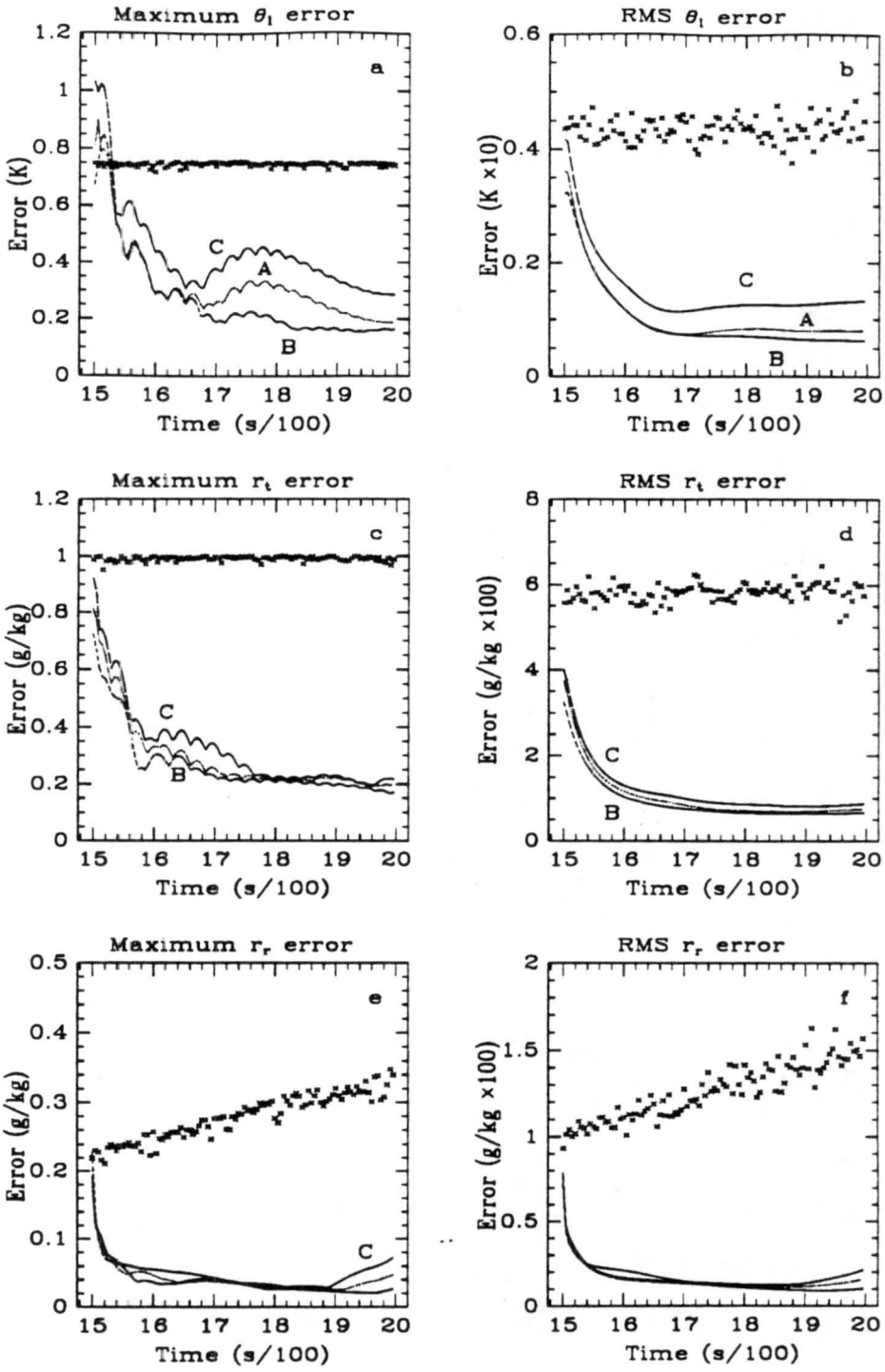

Figure 4.8: Same as Fig (4.7) except that here all experiments used 11 sets of observations. Experiment (A) has weights set to $\Theta_l = 50$, $R_t = 0.04$ and $R_r = 0.01$, (B) has them $\Theta_l = 25$, $R_t = 0.04$ and $R_r = 0.01$, and (C) has them $\Theta_l = 100$, $R_t = 0.04$ and $R_r = 0.01$.

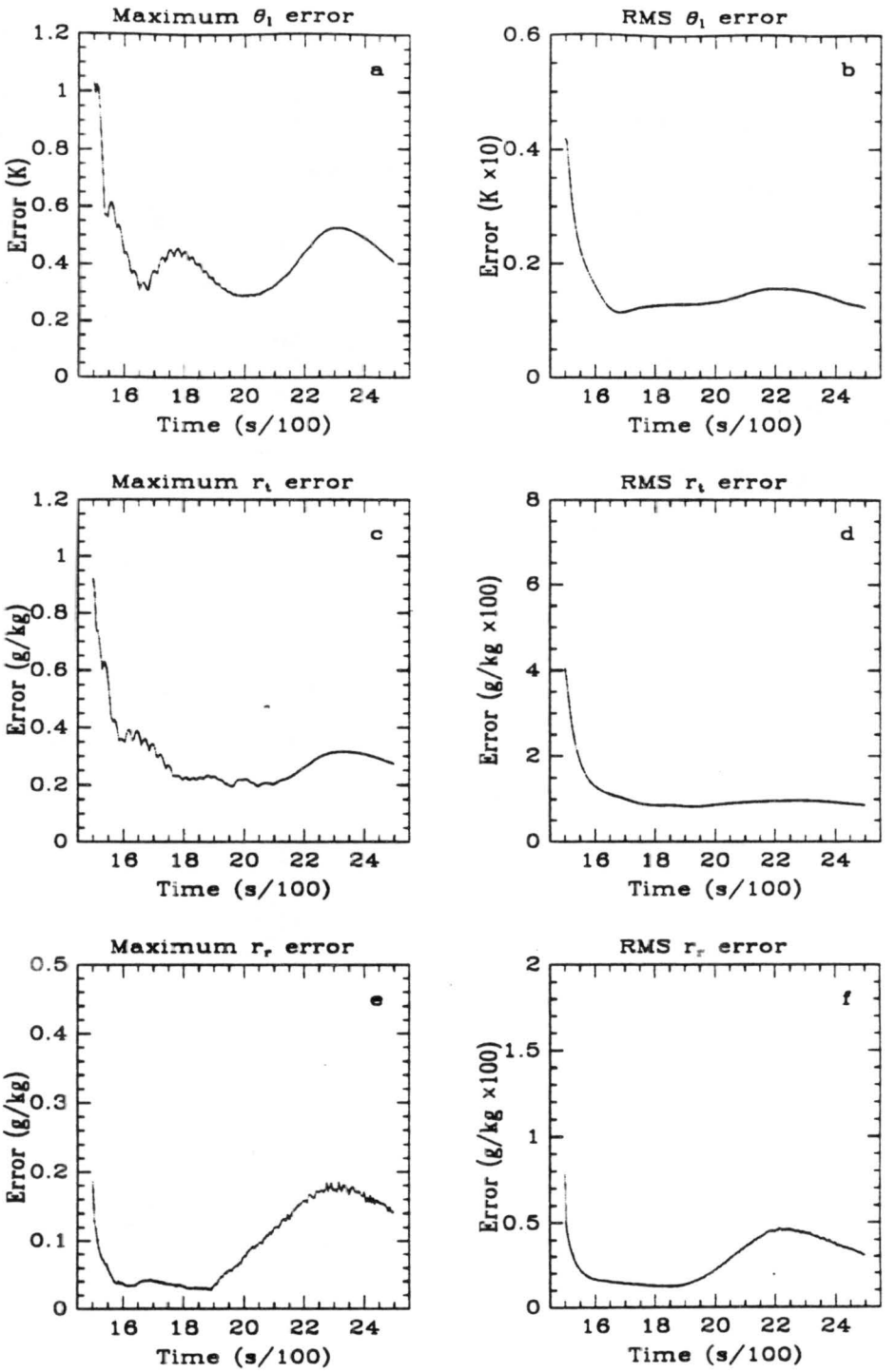

Figure 4.9: Maximum and RMS domain errors for model initialized with the initial conditions retrieved from experiment (C) (Fig 4.8).

increasing the number of observation times from 3 to 11, but there was a marked improvement when data from all time levels (101) were used. Problems with the upper boundary were again found when startup initial conditions further away from the 'truth' were used.

In Fig (4.8) another set of experiments were used. Here the weights of r_t and r_τ were kept the same as in Fig (4.7), but the weight of θ_l was doubled (C) and halved (B). By increasing the weight of θ_l, its contribution to the cost is decreased. It is shown that by changing the weights for one of the variables, different levels of accuracy can be obtained for that variable without negatively impacting the accuracy of the other variables. The window for allowable weights is quite narrow, however, since a further reduction or increase in the θ_l weights led to situations where the model converged very slowly.

The initial conditions retrieved from experiment (C) (Fig 4.8) were used to initialize a run of the forward model at $t = 1500$s. This run was integrated beyond the assimilation period to $t = 2500$s. Results of the error statistics are shown in Fig (4.9). It can be seen that as soon as the model integrates beyond the time where observations were available, the error statistics deteriorate. The improvement in the last 250s of integration is due to the simple model, which, as discussed previously, is able to retrieve the bottom boundary values with a greater degree of accuracy than other areas. These higher accuracy values are then advected into the model domain, resulting in the observed improvements.

Chapter 5

A KINEMATIC TWO-DIMENSIONAL MODEL

In Chapter 4 it was demonstrated that the optimal control theory approach is indeed feasible for application to complex physical problems. However, the model used was simplistic and unrealistic. In this chapter a more realistic model will be used.

A two-dimensional kinematic microphysical model is used. The microphysical parameterization only includes liquid phase physics. Rain water number concentration n_r has been added as a prognostic variable, thus together with liquid water potential temperature θ_l, total water content r_t, and rain water content r_r there are four prognostic variables. The diagnostic variables are temperature T, air density ρ, cloud water r_c, mean diameters for the cloud D_{mc} and rain water D_{mr} distributions and saturation vapor pressure r_s. The wind and pressure fields are prescribed as 'observables'.

The microphysical module of the kinematic model was built into a two-dimensional nonhydrostatic configuration of the RAMS cloud model (Tripoli and Cotton, 1982; Tripoli, 1986). Physical processes allowed in the model are condensation/evaporation (implicit in liquid water potential temperature), autoconversion of cloud water to rain water, collection of cloud water by rain water, selfcollection/breakup of the rain water distribution, and the sedimentation of rain.

5.1 The forward model

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

$$\frac{\partial q}{\partial t} = -ADV(q) + S_q \tag{5.1}$$

where advection and diffusion are combined in the term $ADV(q)$ for notational simplification as

$$ADV(q) = -u\frac{\partial q}{\partial x} - w\frac{\partial q}{\partial z} + K_x\frac{\partial^2 q}{\partial x^2} + K_z\frac{\partial^2 q}{\partial z^2} \tag{5.2}$$

and q represents any of the prognostic variables. S_q represents the respective sources and sinks.

The source/sink terms for θ_l and r_t remain the same as in the previous chapter (Eq 4.3 and Eq 4.4), and are related to the precipitation flux. The source/sinks for rain water mixing ratio are autoconversion of cloud water to rainwater, collection of cloud water by rain, and precipitation flux, while the source/sinks for rain water number concentration are autoconversion of cloud water to rain water, selfcollection/breakup of raindrops and precipitation flux. The contributions to S_{r_r} will first be discussed, followed by the contributions to S_{n_r} and finally all the diagnostic relationships will be defined.

It is assumed that both the cloud water and rain water are distributed according to the gamma distribution

$$f(D) = \frac{1}{\Gamma(\nu)}\left(\frac{D}{D_n}\right)^{\nu-1}\frac{1}{D_n}\exp\left(-\frac{D}{D_n}\right), \tag{5.3}$$

where Γ is the complete gamma function, D is the diameter, and D_n is a characteristic diameter. It was assumed that $\nu = 3$ for both distributions. A constant cloud water number concentration ($n_c = 3 \times 10^8$ m^{-3}) is assumed. With these definitions the source terms can be discussed.

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

$$\left.\frac{dr_r}{dt}\right|_{auto} = \max(0, aut_1 D_{mc}^4 - aut_2)\max(0, aut_3 D_{mc} - aut_4)(\rho r_c)^2, \tag{5.4}$$

where $aut_i, i = 1\ldots4$ are constants given in Appendix (A). The precipitation scheme is the same as used in the previous chapter

$$\left.\frac{dr_r}{dt}\right|_{prec} = -\frac{dV_{Tm}r_r}{dz}, \tag{5.5}$$

where here V_{Tm} is the mass weighted terminal fall speed of the distribution. For this distribution $V_{Tm} = c_v D_{nr}^{p_v} \Gamma(p_m + p_v + \nu)/\Gamma(\nu)$.

Accretion is included from Flatau et al. (1990), but a lower limit for cloud droplet size was added to account for the drop-off in collection efficiency for small cloud droplets. Since this represented a formal discontinuity in the model, and therefore the adjoint formally doesn't exist, a spline was fitted over the discontinuity. Accretion can therefore be written as

$$\left.\frac{dr_r}{dt}\right|_{coll} = \begin{cases} 0 & \text{if } D_{mc} < 5\mu\text{m} \\ \frac{\pi}{4}\Gamma(p_v+2)c_v D_{nr}^{p_v+2} r_c n_r & \text{if } D_{mc} > 40\mu\text{m} \\ \frac{\pi}{4}\Gamma(p_v+2)c_v D_{nr}^{p_v+2} n_r f(r_c) & \text{if } 5\mu\text{m} \geq D_{mc} \leq 40\mu\text{m} \end{cases} \tag{5.6}$$

where

$$f(r_c) = (r_c - \gamma)^3 \left[\frac{\beta}{(\beta-\gamma)^3} - \frac{\gamma+2\beta}{(\beta-\gamma)^4}(r_c-\beta) + \frac{3(\gamma+\beta)}{(\beta-\gamma)^5}(r_c-\beta)^2\right]. \tag{5.7}$$

Here γ is the liquid water mixing ratio corresponding to the lower limit, and β to the higher limit. This spline was constructed so that the function values and the first and second derivatives equal that of the discontinuous function at the limits.

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

$$\left.\frac{dn_r}{dt}\right|_{auto} = autn \left.\frac{dr_r}{dt}\right|_{auto}, \tag{5.8}$$

where *autn* is a constant defined in Appendix (A). The change in rain water number concentration due to precipitation is given by

$$\left.\frac{dn_r}{dt}\right|_{prec} = -\frac{dV_T n_r}{dz}, \tag{5.9}$$

where $V_T = c_v D_{nr}^{p_v}$ is the mean terminal fall velocity. The mean terminal velocity, as opposed to the mass weighted mean terminal velocity, 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 rain drop distribution at a given point. The formulation for selfcollection was taken from Verlinde et al. (1990),

$$\left.\frac{dn_r}{dt}\right|_{self} = -\frac{\pi}{8\rho} c_v D_{nr}^{p_v+2} n_r^2 EC, \tag{5.10}$$

where the constant C is

$$\begin{aligned} C &= \sum_{n=0}^{2}\left[\frac{2}{\nu+n}\Gamma(\eta){}_2F_1(\nu+n,\eta;\nu+n+1;-1)\right. \\ &\qquad \left.-\Gamma(\nu+n)\Gamma(p_v+\nu-n+2)\right] \\ &\quad + \sum_{n=0}^{2}\left[\frac{2}{\nu+p_v+n}\Gamma(\eta){}_2F_1(\nu+p_v+n,\eta;\nu+p_v+n+1;-1)\right. \\ &\qquad \left.-\Gamma(\nu+p_v+n)\Gamma(\nu-n+2)\right] \end{aligned} \tag{5.11}$$

and

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

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

The collection efficiency E is defined as

$$E = \begin{cases} 1 & \text{if } D_{mr} < D_{br} \\ 2 - \exp[A(D_{mr} - D_{br})] & \text{if } D_{mr} \geq D_{br} \end{cases} \tag{5.12}$$

where D_{mr} is the mean diameter of the raindrop distribution, D_{br} is a cutoff (mean) diameter after which breakup will begin to be effective, and A is a constant (values for D_{br} and A are given in Appendix A). This formulation of the collection efficiency will reduce the collection efficiency exponentially from one as the mean diameter passes 600 μm, and will turn negative (increasing the number concentration) when the mean diameter exceeds 900 μm. As the mean diameter of the distribution increases due to selfcollection, breakup will take effect, and the mean diameter will oscillate towards a value of 900 μm. The equilibrium level for mean diameter was chosen to correspond approximately to the broad peak observed by Zawadski and De Agostinho Antonio (1988).

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

$$\begin{aligned} D_n &= \nu D_m \\ &= \left(\frac{0.1\rho r_x}{\pi \rho_l n_x}\right)^{1/p_m}. \end{aligned} \tag{5.13}$$

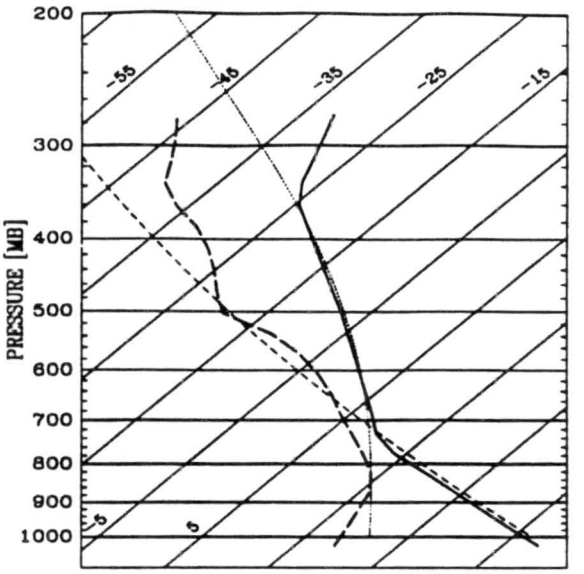

Figure 5.1: Sounding used to initialize the model. The solid line is the temperature profile, the long dashed line is the dew-point profile. The short-dash line is the 310 K potential temperature line and the fine dotted line is the 339 K moist adiabat.

The temperature, cloud water- and saturation mixing ratio are diagnosed from Eq (4.7) through Eq (4.9).

The sequential updating advection scheme with the deformation instability correction (Eq 3.4) was used with the second order fluxes (Eq 3.5), using the formulation of Tremback et al. (1987). This advection scheme was selected over a positive definite scheme due to the complexity of the adjoint of the positive definite schemes in two dimensions.

A two-dimensional non-hydrostatic configuration of the RAMS cloud model (Tripoli and Cotton, 1982; Tripoli, 1986) with the above described microphysics was used to produce wind and pressure fields. The simulation was run on a 20×20 grid domain, with a 5 km horizontal grid-point separation and a 500 m vertical grid-point separation. Open outflow horizontal boundary conditions were used, while horizontal inflow boundaries were kept at environmental values. A 5 second time-step was used. A wall was used as a top boundary. The model was initialized with the sounding from Fig (5.1) with a heat bubble of 1.5° at the center of the domain over four points horizontally and three points vertically.

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.

There are some incompatibilities between the kinematic model described above and the RAMS model. The non-hydrostatic version of the model uses a second order leapfrog advection scheme, and the diagnostic routine for temperature, cloud water and saturation vapor mixing ratio also produces slightly different results. The precipitation scheme in RAMS is also handled differently than in the kinematic model. It was felt that no conclusions regarding the usefulness of the retrieval technique would be possible if RAMS was used to produce the synthetic observational data sets, since failure to converge could very well be due to differences between the models. Therefore the time-varying wind and pressure fields of the RAMS simulation were used to drive the kinematic model to produce observational data sets for the identical twin experiment. The model prognostic fields at the final time-step are shown in Fig (5.2).

5.2 The tangent linear model

The state vector for this model was constructed in the form of Eq (2.25). The variables are arranged in order by grid-point, with the x-indexing taking priority over the z-indexing. The state vector is thus a 1600 element vector. The resulting matrix is a 1600×1600 matrix with non-zero elements on only 21 bands. The matrix can therefore be stored in memory as a 1600×21 array. Due to the size of the array it was decided not to store the matrix at every time-step, but to store sufficient information at every time-step such that all the matrix elements could easily be calculated during the backward integration. The following fields were then saved at every time-step: θ_l, r_t, r_r, n_r, r_s, r_c, T, ρ, D_{mc} and D_{nr}. This amounts to 2.5 times the state vector. The need for the iterative diagnostic routine is therefore eliminated during the backward integration.

The advection scheme chosen was a linear scheme with respect to the prognostic variables, and will not be presented in this document since it can easily be derived by writing the two steps of the forward scheme as a single equation. The handling of boundary conditions are tedious (the four corners and boundaries have to be treated individually),

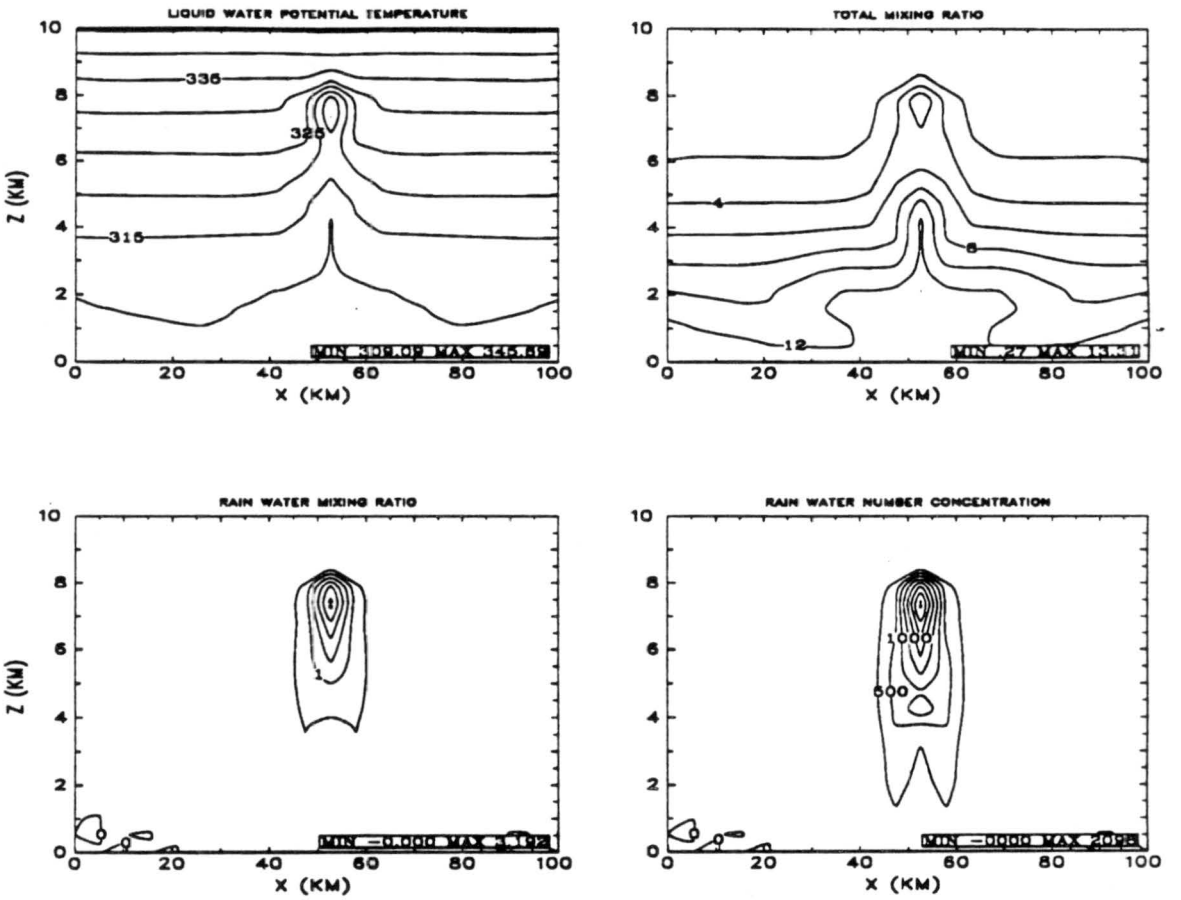

Figure 5.2: Model prognostic fields at the final time step. θ_l values are in °K, the mixing ratios are in g/kg, and the number concentration in $\#\mathrm{m}^{-3}$.

and it was found that the easiest and most accurate way to construct the matrix elements was to use symbolic algebra. For this dissertation a short Maple program (Char et al. 1991) was used to construct the Fortran code.

Next the physical parameterizations that are only dependent on a single grid-point will be considered. For clarity of presentation the grid indices will again be omitted. As the diagnostic routine for the determination of temperature and cloud water is not changed, the derivatives derived in Chapter 4 (Eqs (4.12—4.20)) remain unchanged as well. It should be noted that both temperature and cloud water are independent of the number concentration of the rain distribution, and those derivatives therefore are zero. The derivatives of all the processes will be presented in terms of these equations.

The generic derivative of Eq (5.4) is

$$\frac{\partial S_{r_r}}{\partial x}\Big|_{auto} = \frac{dr_r}{dt}\Big|_{auto}\left\{\left(\frac{aut_1 D_{mc}^3}{\max(\epsilon, aut_1 D_{mc}^4 - aut_2)} + \frac{aut_3}{max(\epsilon, aut_3 D_{mc} - aut_4)}\right)\times \frac{\partial D_{mc}}{\partial x} + \frac{2}{\max(\epsilon, \rho r_c)}\left(\rho\frac{\partial r_c}{\partial x} + r_c\frac{\partial \rho}{\partial x}\right)\right\} \tag{5.14}$$

where

$$\frac{\partial D_{mc}}{\partial x} = \frac{D_{mc}}{p_{mc}}\left(\frac{1}{max(\epsilon, r_c)}\frac{\partial r_c}{\partial x} - \frac{1}{T}\frac{\partial T}{\partial x}\right) \tag{5.15}$$

and

$$\frac{\partial \rho}{\partial x} = \frac{-\rho}{T}\frac{\partial T}{\partial x}. \tag{5.16}$$

Here ϵ is a small number used to prevent division by zero. The derivative for Eq (5.8) follows easily

$$\frac{\partial S_{n_r}}{\partial x}\Big|_{auto} = autn\frac{\partial S_{r_r}}{\partial x}\Big|_{auto}. \tag{5.17}$$

The derivative for Eq (5.10) can be written as

$$\frac{\partial S_{n_r}}{\partial x}\Big|_{self} = \frac{dn_r}{dt}\Big|_{self}\left(\frac{p_v + 2}{D_{nr}}\frac{\partial D_{nr}}{\partial x} - \frac{1}{\rho}\frac{\partial \rho}{\partial x} + \frac{2}{\max(\epsilon, n_r)}\frac{\partial n_r}{\partial x} + \frac{1}{E}\frac{\partial E}{\partial x}\right) \tag{5.18}$$

where

$$\begin{aligned}\frac{\partial D_{nr}}{\partial x} &= \nu\frac{\partial D_{mr}}{\partial x} \\ &= \frac{D_{nr}}{p_m}\left(\frac{1}{\rho}\frac{\partial \rho}{\partial x} + \frac{1}{\max(\epsilon, r_r)}\frac{\partial r_r}{\partial x} - \frac{1}{\max(\epsilon, n_r)}\frac{\partial n_r}{\partial x},\right)\end{aligned} \tag{5.19}$$

and

$$\frac{\partial E}{\partial x} = \frac{AE}{3}\frac{\partial D_{nr}}{\partial x}. \tag{5.20}$$

The derivative of Eq (5.6) is given by

$$\begin{aligned}\frac{\partial S_{r_r}}{\partial x}\Big|_{coll} &= \frac{\pi}{4}\Gamma(p_v+2)c_v D_{nr}^{p_v+2} g(r_c) \times \\ &\left(\frac{p_v+2}{D_{nr}}\frac{\partial D_{nr}}{\partial x} + \frac{1}{\max(\epsilon, n_r)}\frac{\partial n_r}{\partial x} + \frac{1}{\max(\epsilon, g(r_c))}\frac{\partial g(r_c)}{\partial r_c}\frac{\partial r_c}{\partial x}\right)\end{aligned} \tag{5.21}$$

where

$$g(r_c) = \begin{cases} 0 & \text{if } D_{mc} < 5\mu\text{m} \\ r_c & \text{if } D_{mc} > 40\mu\text{m} \\ f(r_c) & \text{if } 5\mu\text{m} \geq D_{mc} \leq 40\mu\text{m} \end{cases}. \tag{5.22}$$

This completes the single grid-point physics. Only precipitation is left. Since precipitation is constrained to one dimension, and precipitation is modelled the same in this model as in Chapter (4), Eqs (4.23—4.27) are valid for the rain water source term. The contribution Eq (5.9) is written in discrete form as

$$S_{n_r}(k)\Big|_{prec} = -\left(\frac{c_v}{\Delta z}\left(D_{nr}^{p_v}(k+1)n_r(k+1) - D_{nr}^{p_v}(k)n_r(k)\right)\right) \tag{5.23}$$

where k is the vertical index. The derivative may then be written as

$$\begin{aligned}\frac{\partial S_{n_r}(k)}{\partial x(k+1)}\Big|_{prec} &= -\frac{c_v}{\Delta z}\left(p_v n_r(k+1) D_{nr}^{p_v-1}(k+1)\frac{\partial D_{nr}(k+1)}{\partial x(k+1)}\right. \\ &\left. + D_{nr}^{p_v}(k+1)\frac{\partial n_r(k+1)}{\partial x(k+1)}\right)\end{aligned} \tag{5.24}$$

and

$$\frac{\partial S_{n_r}(k)}{\partial x(k)}\Big|_{prec} = -\frac{c_v}{\Delta z}\left(p_v n_r(k) D_{nr}^{p_v-1}(k)\frac{\partial D_{nr}(k)}{\partial x(k)} + D_{nr}^{p_v}(k)\frac{\partial n_r(k)}{\partial x(k)}\right). \tag{5.25}$$

This concludes the derivation of the tangent linear model. To illustrate what terms impact the gradient, a single step of the adjoint integration for one component will be shown. It is assumed that all processes are active at this grid-point, and weights are ignored.

The step for the adjoint variable $\delta\theta_l$ at the (i,k)'th grid-point is presented. All the values on the right hand side of the equation are computed at time $t = n$, and the result

(left hand side of the equation) is at time $t = n - 1$. For clarity of presentation the time level is not indicated in the equation. The abbreviation $ADV_{i,k}$ indicates the advective and diffusive tendency of θ_l at the point (i, k). The integration is performed from time $t = N$ (time of last observation) to time $t = 0$, the initial condition. The resulting vector is the gradient of the cost function with respect to the given initial condition. The step may be written as

$$
\begin{aligned}
\delta\theta_l\big|_{i,k} \;=\; & \delta\theta_l\big|_{i-1,k-1}\frac{\partial}{\partial\theta_l}\Big|_{i,k}(ADV_{i-1,k-1}) + \\
& \delta\theta_l\big|_{i,k-1}\left(\frac{\partial}{\partial\theta_l}\Big|_{i,k}(ADV_{i,k-1}) + \frac{\partial}{\partial\theta_l}\Big|_{i,k}(S_{\theta_l}\big|_{i,k-1})\right) + \\
& \delta r_t\big|_{i,k-1}\frac{\partial}{\partial\theta_l}\Big|_{i,k}(S_{r_t}\big|_{i,k-1}) + \delta r_r\big|_{i,k-1}\frac{\partial}{\partial\theta_l}\Big|_{i,k}(S_{r_r}\big|_{i,k-1}) + \\
& \delta n_r\big|_{i,k-1}\frac{\partial}{\partial\theta_l}\Big|_{i,k}(S_{n_r}\big|_{i,k-1}) + \delta\theta_l\big|_{i+1,k-1}\frac{\partial}{\partial\theta_l}\Big|_{i,k}(ADV_{i+1,k-1}) + \\
& \delta\theta_l\big|_{i-1,k}\frac{\partial}{\partial\theta_l}\Big|_{i,k}(ADV_{i-1,k}) + \delta\theta_l\big|_{i,k}\left(\frac{\partial}{\partial\theta_l}\Big|_{i,k}(ADV_{i,k}) + \frac{\partial}{\partial\theta_l}\Big|_{i,k}(S_{\theta_l}\big|_{i,k})\right) + \\
& \delta r_t\big|_{i,k}\frac{\partial}{\partial\theta_l}\Big|_{i,k}(S_{r_t}\big|_{i,k}) + \delta r_r\big|_{i,k}\frac{\partial}{\partial\theta_l}\Big|_{i,k}(S_{r_r}\big|_{i,k}) + \\
& \delta n_r\big|_{i,k}\frac{\partial}{\partial\theta_l}\Big|_{i,k}(S_{n_r}\big|_{i,k}) + \delta\theta_l\big|_{i+1,k}\frac{\partial}{\partial\theta_l}\Big|_{i,k}(ADV_{i+1,k}) + \\
& \delta\theta_l\big|_{i-1,k+1}\frac{\partial}{\partial\theta_l}\Big|_{i,k}(ADV_{i-1,k+1}) + \delta\theta_l\big|_{i,k+1}\frac{\partial}{\partial\theta_l}\Big|_{i,k}(ADV_{i,k+1}) + \\
& \delta\theta_l\big|_{i+1,k+1}\frac{\partial}{\partial\theta_l}\Big|_{i,k}(ADV_{i+1,k+1}). \qquad (5.26)
\end{aligned}
$$

This equation shows the opposite sense of movement of the adjoint integration.

Consider the impact of the precipitation. In the forward integration of the model, the value of a given grid-point is determined by the value of the model variables at that grid point and the grid point *above* it (Eq 5.23). In the adjoint model, the influence of precipitation (through the S_q terms) on the adjoint variable at a given grid-point comes from the adjoint variables at that grid point and the grid point *below* it. *It has to be re-iterated at this point that the adjoint integration is not a backward integration of the forward model but that it is the integration of a different, but related, model.*

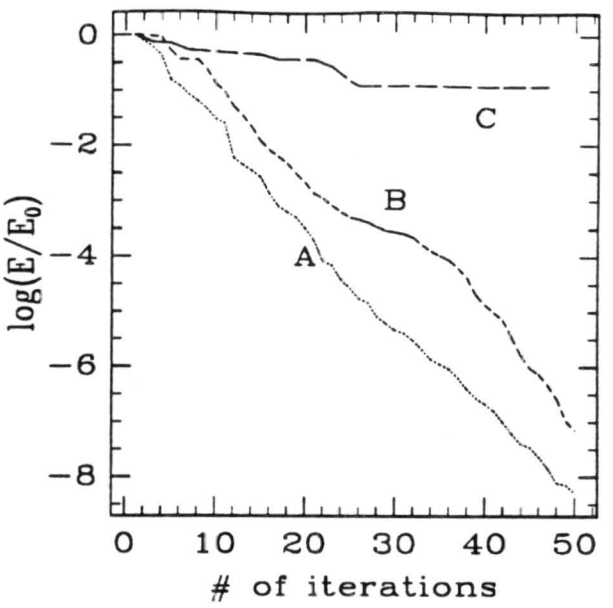

Figure 5.3: Experiments investigating the effect of different weights on the convergence rate. A has the weights set to $\Theta_l = 160$, $R_t = 0.16$, $R_r = 0.048$ and $N_r = 80000$, while B has the weight for θ_l doubled and C has it halved.

5.3 Application

The initial experiments were done with perfect 'observations'. All the results in this section are for experiments that used startup initial conditions of the environmental background with a constant perturbation inserted where the cloud was observed.

Typical mean errors for the variables were selected to decide on the scaling factors to use. The weights were further normalized by the number of grid-points. It was found that the correct scaling between the four prognostic variables were critical to the convergence rate, and that even small deviations resulted in significant lower convergence rates. The physics in this model include processes (eg. selfcollection/breakup) which will rapidly correct invalid physics (eg. rain distribution with a mean drop diameter of 1 m because the optimization code increased rain water 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 has happened, the algorithm never recovered in a reasonable number of iterations. The correct ratio between the weights was determined by placing

a very small perturbation on the startup initial condition. 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 further found that once the correct ratio for the weights was obtained, multiples of these weights provide good convergence for cases where the startup initial condition is further removed from the 'correct' initial condition. These scaling factors gave good convergence for different startup 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, that a set of typical weights can be determined which will work in most cases, and, where it does not provide good convergence, that it would be easy to adjust the weights based on a single run.

This is a good place to discuss some practical matters of constructing the retrieval algorithm. Absolute accuracy during the coding process is required. It is essential that every individual contribution to the matrix be tested to ensure that the forward code and the adjoint code are in agreement. Any error in either the forward or the adjoint code results in a gradient that is incorrect. 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 gradient 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 θ_l and r_t 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 indeterminable function value. The optimization code would then recalculate the initial condition by shortening the step.

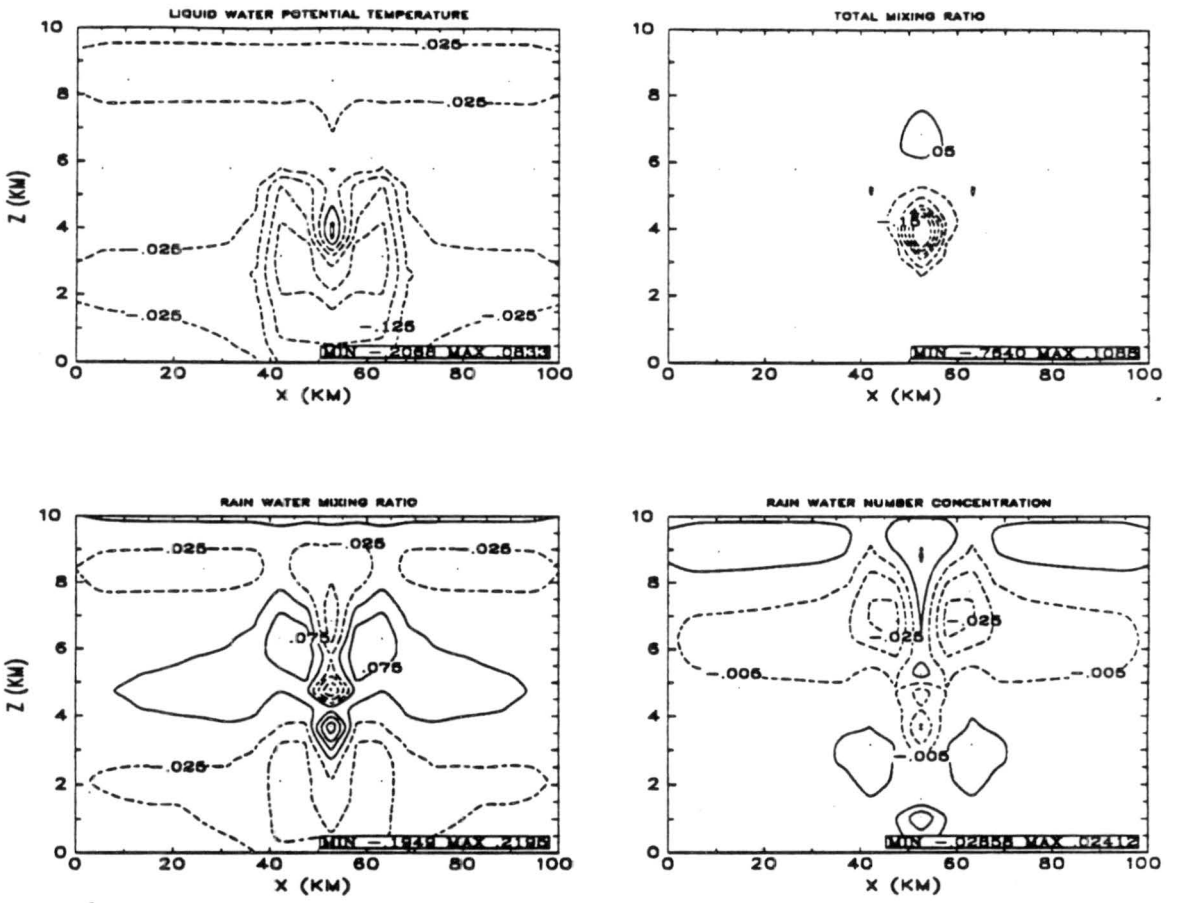

Figure 5.4: Contoured gradients for the individual variable vectors after the first iteration. Domain minima and maxima are listed in the lower left hand corner. Negative values are contoured with dashed lines, positive solid lines. The contour interval for liquid water potential temperature and rain water is 0.05, for total mixing ratio it is 0.1, and for rain water number concentration it is 0.01. The weights used in this experiment are $\Theta_l = 80$, $R_t = 0.16$, $R_r = 0.048$ and $N_r = 80000$.

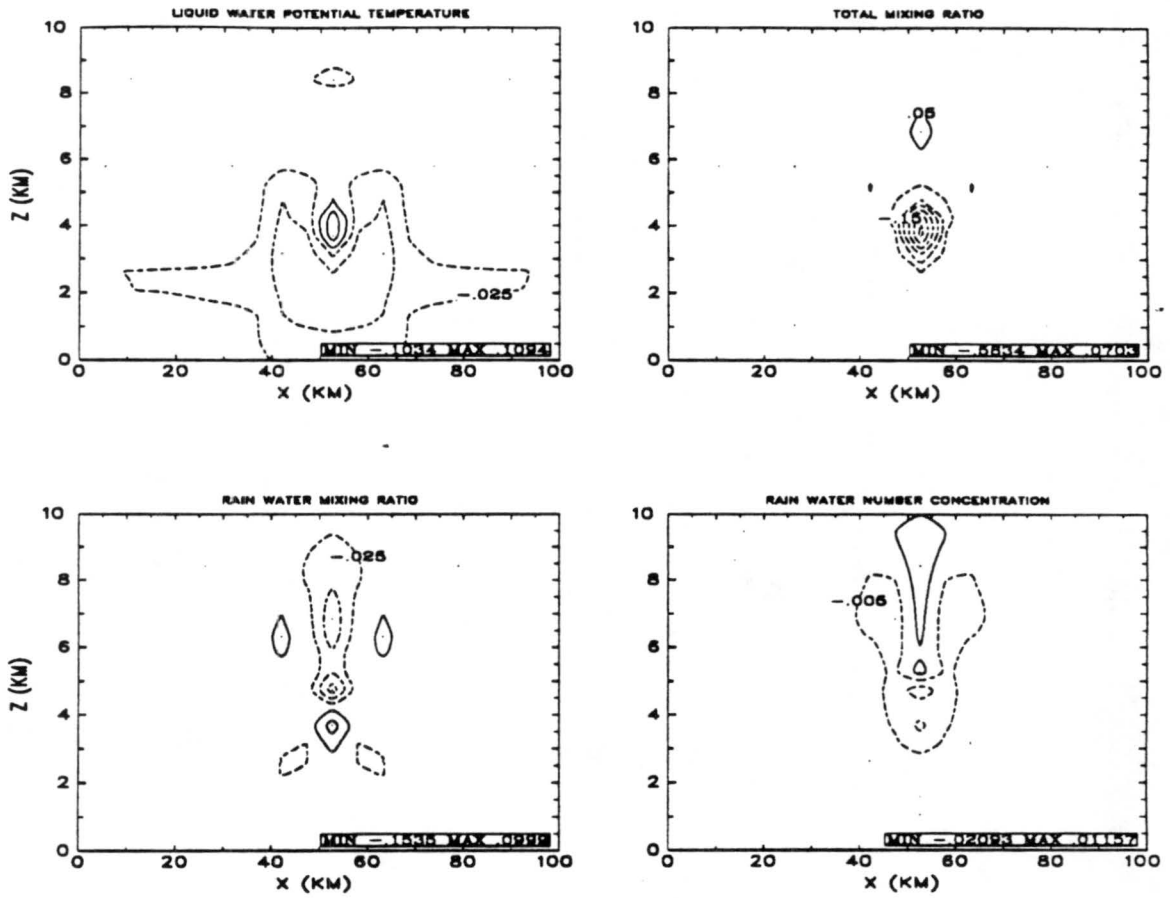

Figure 5.5: Same as in Fig (5.3), but with $\Theta_l = 160$.

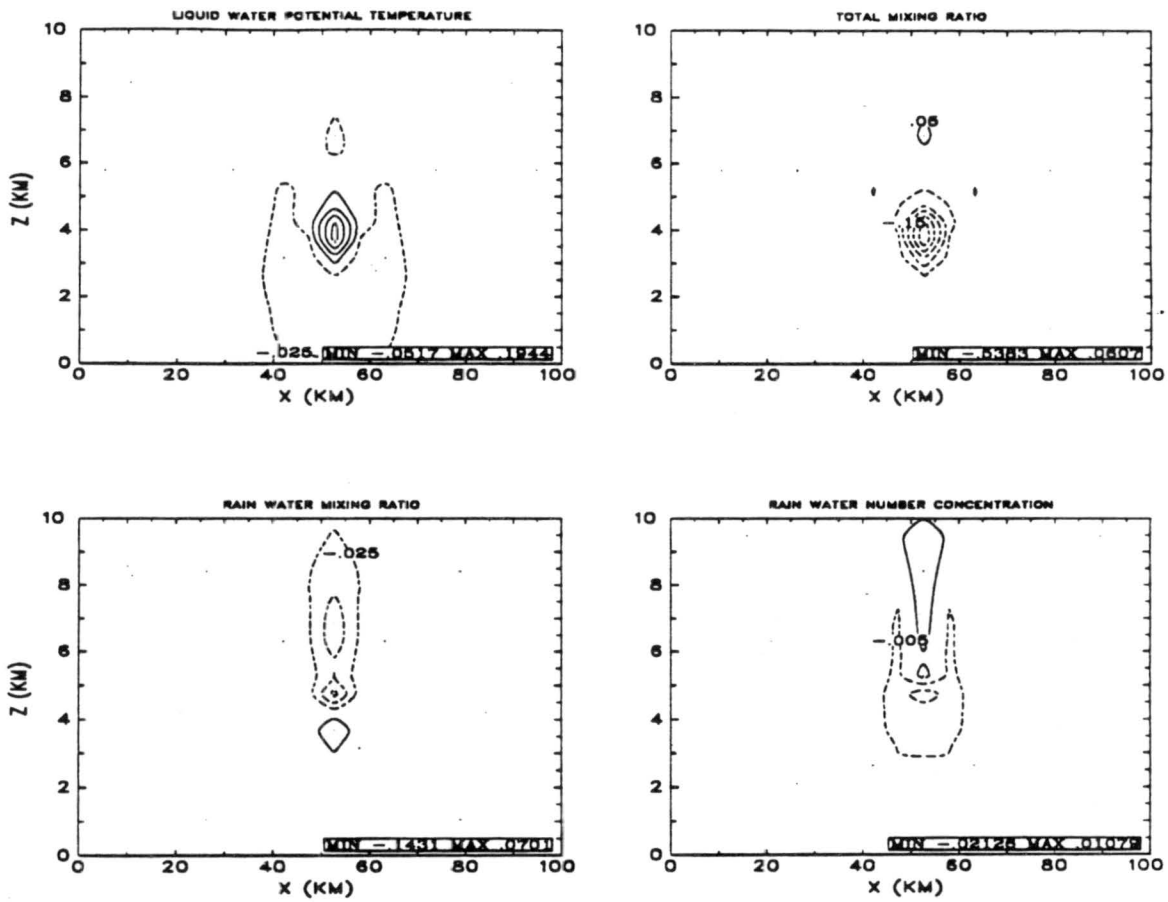

Figure 5.6: Same as in Fig (5.3), but with $\Theta_l = 320$.

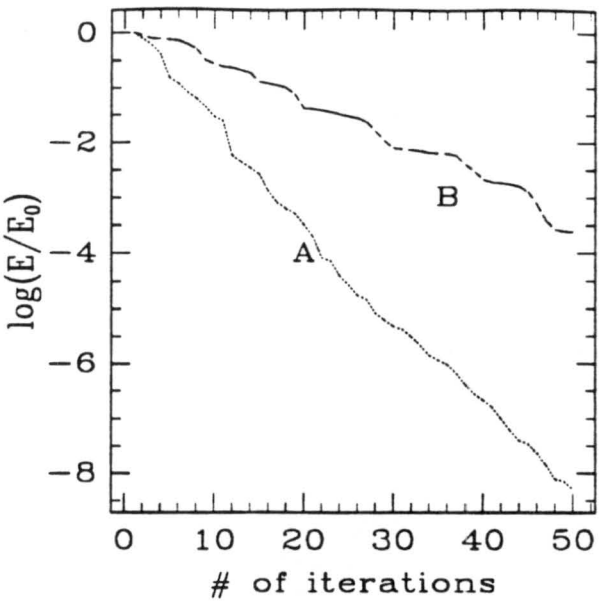

Figure 5.7: Experiments investigating the effect of the bounds of the spline fitted to Eq (5.6) on the convergence of the algorithm. A had the bounds set as in Eq (5.6), B had the top bound lowered from 40μm to 30μm.

The general weights determined for this study were $\Theta_l = 160$, $R_t = 0.16$, $R_r = 0.048$ and $N_r = 80000$. Fig (5.3) shows the impact on the convergence rate of changes in the ratio of the scaling factors. Figs (5.4—5.6) 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 θ_l weight from 160 to 320 had a minor impact on the overall structure and magnitude of the gradient, while reducing it to 80 resulted 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 also found that when the startup initial conditions were changed to the observed initial conditions with a set offset, much faster convergence rates were obtained. This can be explained by realizing that for this case, the adjustments needed to the initial conditions are uniform over the domain, resulting in an easier path for the optimization code.

One of the objects of this study was to investigate the impact of discontinuities in the forward model on the retrieval process. Eq (5.6) was constructed in such a fashion to test the results of a zeroth order discontinuity. It is known that there is a cutoff

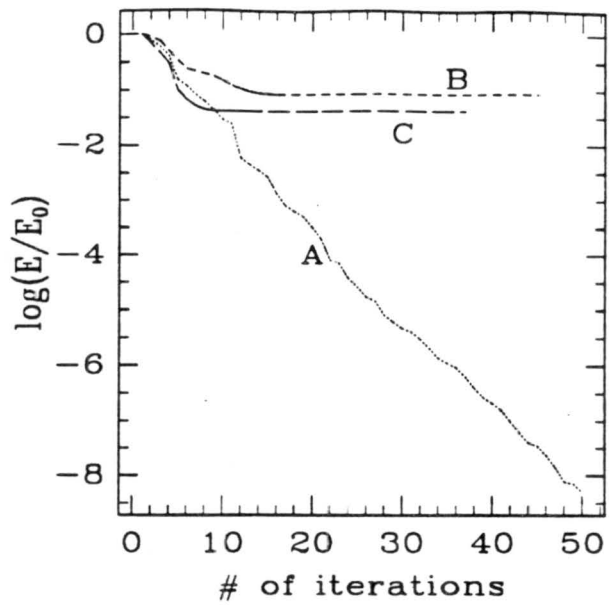

Figure 5.8: Experiments to investigate the neglect of physical processes in the retrieval algorithm on the convergence of the algorithm. A is the identical twin experiment, B has neglected accretion and selfcollection, while C has neglected only selfcollection.

where the collection efficiency of raindrops collecting cloud drops decreases to zero. Eq (5.6) 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 (5.7). 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 20μm the algorithm no longer converged. This indicates that zeroth order discontinuities will greatly impact retrieval algorithms, but also shows how this obstacle may be overcome. Most of the other physical processes, like phase changes, although discontinuous, started gradually such that to the numerical model they can be considered continuous, 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. Fig (5.8) compares results from two experiments where first accretion and selfcollection were neglected (B), and second only selfcollection was neglected (C) to the standard identical twin case. In these experiments the observations were created with all the physical processes included, while the retrieval runs were

Table 5.1: The errors in the initial conditions for different number of observations taken. Three experiments with different weights for r_r and n_r using 5 s time steps are included.

# of obs	θ_l max	θ_l rms ($\times 10^{-3}$)	r_t max ($\times 10^{-4}$)	r_t rms ($\times 10^{-6}$)	r_r max ($\times 10^{-4}$)	r_r rms ($\times 10^{-7}$)	n_r max	n_r rms
obs	0.50	14.8	5.00	10.0	1.57	4.55	171.49	0.45
3	0.61	12.9	6.23	9.65	1.59	6.87	169.93	0.59
5	0.61	11.4	4.94	8.81	1.55	6.96	171.02	0.63
13	0.50	7.85	3.63	7.47	1.47	6.13	176.72	0.56
25a	0.36	5.74	2.94	7.04	1.40	5.91	183.01	0.58
25b	0.39	5.74	3.22	7.03	1.50	4.43	179.19	0.47
25c	0.45	6.52	4.95	7.27	1.55	4.45	174.33	0.45

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 of up to 0.5 K for θ_l, 0.5 g/kg for r_t and 10% were added/subtracted to all the observations. Results after 50 iterations are presented in Table (5.1). In all cases the minimum was reached long before fifty iterations. All these retrieval runs used the observations at the initial time as the startup 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 θ_l and r_t, while there was a slight deterioration in the RMS errors for r_r and n_r. This can be explained

by realizing that observations of r_r and n_r only exist where there was precipitation, which only covers a small part of the domain. Over the rest of the domain the error in the initial r_r 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. In minimizing the cost function, the algorithm will thus favor θ_l and r_t, 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_r and n_r were decreased to increase their contribution to the cost. In experiment 25b the weights were set at $R_r = 0.024$ and $N_r = 40000$, while in 25c they were set at $R_r = 0.012$ and $N_r = 20000$. 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_r began increasing again, indicating that a further reduction in the weights would not further improve the results.

Chapter 6

DISCUSSION AND CONCLUSIONS

6.1 Discussion

The purpose of this study was to show that the adjoint technique of data assimilation can be applied to numerical models of complex physical processes. This was done using three different models; a positive definite advection model, a simplified one-dimensional kinematic microphysical model with a steady wind field and a two-dimensional kinematic microphysical model with an imposed time-varying wind field.

The matrix approach taken towards 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 linearized 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 needs 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 significant 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 individual 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 easily be

tested. In this case the tangent linear model and the adjoint model derive from the same code.

It is concluded, based on the discussion in Chapter 3, that despite the desirability of having a positive definite advection scheme, technical aspects of constructing the adjoint model for two- or three dimensions may outweigh the advantages. It was, however, shown with the one dimensional positive advection model that it is possible to construct the adjoint for such a scheme, and good convergence rates were found. It may therefore be necessary to construct simplified assimilation models in lieu of the state-of-the-art forecast model. This simplified model can then be used to assimilate observations to provide initial conditions for the forecast model. It is assumed that these initial conditions will require less spin-up time in the forecast model than the horizontally homogeneous initializations or interpolated analyses that are commonly used in mesoscale models today. An example of that would be a three-dimensional dispersion model with a positive definite advection scheme. The assimilation package then would be constructed with a standard advection scheme (eg. Eq 3.4) with the hope that the increased amount of information coming from the four-dimensional assimilation will provide initial conditions which will improve the forecast using the model with the positive definite advection scheme.

The importance of the proper scaling between variables was illustrated in Chapters 4 and 5. Four different views on how the scaling factors need to be determined was found in literature. First, it is not clear *a priori* what the proper scaling for a particular problem is, and that the proper weights thus need to be determined by testing the particular model (Kapitza, 1991). Second, the scaling should reflect the relative magnitudes of the unknown variables (Moore, 1991). Third, the scaling should be related to the variance of the observational error (Courtier and Talagrand, 1990). Fourth, the gradient should be scaled by the Hessian matrix of the cost function (Thacker, 1989; Moore, 1991). The last view, though theoretically correct, is prohibitively expensive. In this study we found that in each case 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 magnitude of the variable, scaling by the relative magnitude of the variables will give good results.

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 startup condition to great accuracy in fifty iterations, or to acceptable accuracy in thirty iterations. It was found that longer assimilation periods, as well as more observations, lead to slower convergence rates of the cost function. The routine was able to retrieve information about a variable which was not observed, although up to a 150 iterations was 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 towards 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 (derivation in next section). This algorithm can then be applied to a well observed case to determine optimal settings for the parameterization 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. More observations resulted in better fits, as was suggested by Thacker (1987).

6.2 Recommendations for further research

The observations to which a model such as this is to be applied, will typically not be the prognostic variables of the model. They would rather be some function of the prognostic variables. Throughout this study, however, it was assumed that all the prognostic variables of the model could indeed be observed. It remains to be shown that the general form for the adjoint derived in Chapter 2 will work. Application of the general form raises another interesting problem, namely, that of formulating the transformation function. The transformation function is that function which relates the actual physical quantities being observed to the model prognostic variables. In the example in Chapter 2 it was shown how reflectivity observations could be incorporated, however, in many other remotely sensed quantities this relationship is not altogether clear, especially when ice processes are included.

The same derivation from Chapter 2 may also be used to determine parameterization coefficients. Let the control vector be given by

$$\mathbf{v}^0 = (x_1^0, \ldots, x_N^0, p_1, \ldots, p_n) \tag{6.1}$$

where x_i^0, $i = 1, \ldots, N$ is the vector containing the initial conditions for the model, and p_i, $i = 1, \ldots, n$ is the vector of the constant parameterization coefficients that needs to be adjusted. A forward model for this state vector can be written in the form of Eq (2.4) by realizing that $f_{p_j} = 0$, $j = 1, \ldots, n$. The rest of the derivation remains the same. Thus, this technique may equally well be used to tune parameterization coefficients for optimal modelling of specific processes. This further allows for the investigation of possible changes in coefficients for different applications (eg. autoconversion coefficients in cumulus clouds vs in orographic cloud) when extensive observational data sets are available.

This study has answered the question whether the adjoint technique of data assimilation can be used to assimilate microphysical observations into a numerical cloud model using bulk parameterized microphysics. A more realistic test would be to use a high resolution explicit (bin) microphysical model to produce an observational data set. Bulk observations based on this can then be produced and used in the retrieval model. An

experiment such as this can give valuable insight into the necessary complexity required in microphysical modeling to obtain a realistic simulation. The obvious next application would then be to apply the algorithm to real observations. For a simpler application a one-dimensional version of the model could be applied in a situation where horizontal homogeneity can be assumed. Possible candidates would be from the recent FIRE II field experiment in Kansas where extensive observations of a liquid alto-stratus cloud were taken, the coming ASTEX experiment in the eastern Atlantic Ocean in the coming summer (looking at the development and breakup of stratus), or from one of the DOE ARM CART sites.

The general form of equations to parameterize physical processes can intuitively be determined with a good understanding of the process to be parameterized. If detailed observations of the physical process exist, then the technique described in this study can be used to determine the set of coefficients for the parameterization such that the model will produce a 'best' fit. This can be accomplished even if none of the observations are model prognostic variables (the variables in the parameterized form).

If this technique is to be used as an initialization tool for a forecast model, then it will be necessary to develop a simplified model (which does include all the physics in the detailed model) which would be used to assimilate the observations and provide the initial fields for the detailed forecast model. The technique may be used to develop the simplified model by adjusting coefficients such that the differences between the two models for the same situation are minimized. This would be done by running the detailed model for a given initial condition, and then doing a retrieval run with the same initial conditions, where the control variables include the coefficients that are to be adjusted, but not the initial conditions.

Adjoint models can also be used for sensitivity analysis of a measure of a forecast aspect (Marchuk and Skiba, 1976; Cacuci, 1981a; Cacuci, 1981b; Hall et al., 1982; Hall and Cacuci, 1983; Hall, 1986; Errico and Vukicevic, 1991). Given the problems this study had with discontinuities, it still needs to be determined how applicable information coming from models with discontinuities would be.

6.3 Conclusions

The following are the major conclusions of this study

- The adjoint technique can be applied to numerical models based on parameterized physics, provided that zeroth order discontinuities in the forcing terms are smoothed out.
- The matrix approach was found to be easy to implement and debug, and the code optimized well.
- There may be limitations in applying the adjoint technique for some state-of-the-art techniques due to technical difficulties in the construction of the adjoint resulting in inefficient code.
- It was found that scaling factors related to the error variance gave the best convergence rates.
- The algorithm was able to retrieve the initial field for a prognostic variable which was not observed from observations of the other variables. Small amounts of observations of that variable greatly improved the convergence rate.
- Neglect of specific processes active in the physics being observed lead to greatly reduced accuracies in the retrievals.

Appendix A

LIST OF CONSTANTS

Symbol	Description	Value
$aut1$	Const. autoconv.	9.74×10^{19} m^2 kg^{-2} s^{-1}
$aut2$	Const. autoconv.	10.8 m^6 kg^{-2} s^{-1}
$aut3$	Const. autoconv.	1124836. m^5 kg^{-2} s^{-1}
$aut4$	Const. autoconv.	20.27 m^6 kg^{-2} s^{-1}
$autn$	Const. autoconv.	3.5×10^6 m^{-3}
c_m	Coefficient for mass of rain	$\rho_l \pi / 6$
c_p	Spec. heat water vapor	1952 J K^{-1} kg^{-1}
c_v	Coeff. term. vel. of rain	132 m$^{1/2}$s^{-1}
l_v	Latent heat of vaporization	2.5×10^6 J kg^{-1}
p_0	Reference pressure	1×10^5 N m^{-2}
p_{mr}	Power coeff. for mass	3.0
p_v	Power coeff,, term. vel. rain	0.5
r_0	Mixing ratio threshold	7×10^{-4}
A	Const. collection efficiency	2300
C_1	Constant in collection equation	eq (4.5)
D_{br}	Cutoff diameter for breakup	6×10^{-4} m
N_0	Intercept, Marshall-Palmer distr.	8×10^6 m^{-4}
K_x	Horizontal diffusion coefficient	1.28
K_z	Vertical diffusion coefficient	5.9
α	Rate coefficient for autoconversion	0.0001 s^{-1}
β	Upper threshold for collection	2.23×10^{-2}
γ	Lower threshold for collection	4.36×10^{-5}
ρ_l	Density of liquid water	1×10^3 kg m^{-3}
ν	Shape parameter of gamma distr.	3

Bibliography

Anthes, R. A., 1974: Data assimilation and initilization of hurricane prediction models. *J. Atmos. Sci.*, **31**, 702—719.

Atlas, D., 1957: Radar measurements of precipitation growth. In *Artificial stimulation of precipitation.* Pergamon Press, 288—305.

Austin, P. M. and R. Wexler, 1957: Observations of space and time variations in the radar echo intensity of showers. In *Artificial stimulation of rain.* Pergamon Press, 271—276.

Barnes, S. L., 1973: Mesoscale objective analysis using weighted time-series observations. Technical Memorandum ERL NSSL-62, NOAA, Environmental Research Laboratories, NSSL, Norman, OK, 60 pp.

Battan, L. J., 1963: Relationship between cloud base and initial radar echo. *J. Appl. Meteor.*, **2**, 333—336.

Bengtsson, L., 1975: 4-dimensional assimilation of meteorological observations. Technical report, ICSU/WMO GARP Publ. Series No 15, 76 pp.

Berry, E. X. and R. L. Reinhardt, 1974: Analysis of cloud drop growth by collection: Part II. Single initial distributions. *J. Atmos. Sci.*, **31**, 1825—1831.

Bjerknes, J. and H. Solberg, 1922: The life cycle of cyclones and the polar front theory. *Geofys. Publ.*, **3**(1), 17 pp.

Boris, J. P. and D. L. Book, 1973: Flux-corrected transport. I. SHASTA, a fluid transport algorithm that works. *J. Comput. Phys.*, **11**, 38—69.

Bott, A., 1989: A positive definite advection scheme obtained by nonlinear renormalization of the advective fluxes. *Mon. Wea. Rev.*, **117**, 1006—1015.

Bringi, V. N., J. Vivekanandan, and J. D. Tuttle, 1986: Multiparameter radar measurements in Colorado convective storms. Part II: Hail detection studies. *J. Atmos. Sci.*, **43**, 2564—2577.

Buckley, A. and A. Lenir, 1983: QN-like variable storage conjugate gradients. *Mathematical Programming*, **27**, 155—175.

Cacuci, D. G., 1981a: Sensitivity theory for nonlinear systems. I. Nonlinear functional analysis approach. *J. Math. Phys.*, **22**, 2794—2802.

Cacuci, D. G., 1981b: Sensitivity theory for nonlinear systems. II. Extensions to additional classes of responses. *J. Math. Phys.*, **22**, 2803—2812.

Char, B. W., K. O. Geddes, G. H. Gonnet, B. L. Leong, M. B. Monagan, and S. M. Watt, 1991: *MapleV language reference manual.* Springer-Verlag, pp 267.

Charney, J., M. Halem, and R. Jastrow, 1969: Use of incomplete historical data to infer the present state of the atmosphere. *J. Atmos. Sci.*, **26**, 1160—1163.

Courtier, P. and O. Talagrand, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. II: Numerical results. *Quart. J. Roy. Meteor. Soc.*, **113**, 1329—1347.

Courtier, P. and O. Talagrand, 1990: Variational assimilation of meteorological observations with the direct and adjoint shallow-water equations. *Tellus*, **42A**, 531—549.

Cressman, G. P., 1959: An operational objective analysis system. *Mon. Wea. Rev.*, **87**, 367—374.

Crowley, W. P., 1968: Numerical advection experiments. *Mon. Wea. Rev.*, **96**, 1—11.

Daley, R., 1991: *Atmospheric data analysis.* Cambridge University Press, New York, 400 pp.

Derber, J. C., 1989: A variational continuous assimilation technique. *Mon. Wea. Rev.*, **117**, 2437—2446.

DiMego, G. J., 1988: The National Meteorological Center regional analysis system. *Mon. Wea. Rev.*, **116**, 977—1000.

Elliot, R. D. and E. L. Hcvind, 1964: The water balance of orographic clouds. *J. Appl. Meteor.*, **3**, 235—239.

Errico, R. M. and T. Vukicevic, 1991: Sensitivity analysis using an adjoint of the PSU/NCAR mesoscale model. Submitted to *Mon. Wea. Rev.*

Flatau, P. J., G. J. Tripoli, J. Verlinde, and W. R. Cotton, 1989: The CSU-RAMS cloud microphysical module: General theory and code documentation. Dept. of Atmos. Science Paper 451, Colorado State University, Ft Collins, CO 80523.

Gandin, L. S., 1965: *Objective analysis of meteorological fields.* Israel Program for Scientific Translations, Jerusalem, 242 pp.

Gunn, K. L. S., M. P. Langleben, A. S. Dennis, and B. A. Power, 1954: Radar evidence of a generating level for snow. *J. Meteor.*, **11**, 20—26.

Hall, M. C. G., 1986: Application of adjoint sensitivity theory to an atmospheric general circulation model. *J. Atmos. Sci.*, **43**, 2644—2651.

Hall, M. C. G. and D. G. Cacuci, 1983: Physical interpretation of adjoint functions for sensitivity analysis of atmospheric models. *J. Atmos. Sci.*, **40**, 2537—2546.

Hall, M. C. G., D. G. Cacuci, and M. E. Schlesinger, 1982: Sensitivity analysis of a radiative convective model by the adjoint method. *J. Atmos. Sci.*, **39**, 2083—2050.

Hauser, D. and P. Amayenc, 1986: Retrieval of cloud water and water vapor contents from Doppler radar data in a tropical squall line. *J. Atmos. Sci.*, **43**, 823—838.

Hauser, D., F. Roux, and P. Amayenc, 1988: Comparison of two methods for the retrieval of thermodynamic and microphysical variables from Doppler radar measurements: Application to the case of a tropical squall line. *J. Atmos. Sci.*, **45**, 1285—1303.

Hoke, J. E. and R. A. Anthes, 1976: The initialization of numerical models by a dynamic-initialization technique. *Mon. Wea. Rev.*, **104**, 1551—1556.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *J. Basic Eng.*, **82**, 35—45.

Kapitza, H., 1991: Numerical experiments with the adjoint of a nonhydrostatic mesoscale model. *Mon. Wea. Rev.*, **119**, 2993—3011.

Kessler, E., 1969: *On the distribution and continuity of water substance in atmospheric circulations.* Meteor. Monogr., No 32, Amer. Meteor. Soc., 84 pp.

Kuo, Y. H. and Y. R. Guo, 1989: Dynamic initialization using observations from a hypothetical network of profilers. *Mon. Wea. Rev.*, **117**, 1975—1998.

Le Cam, M. N. and H. Isaka, 1989: Retrieval of microphysical variables by a diagnostic modelling study: comparison between parameterized and detailed warm microphysics. *Tellus*, **41A**, 338—356.

Le Dimet, F. X. and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus*, **38A**, 97—110.

Leith, C. E., 1965: Numerical simulation of the earth's atmosphere. Applications in hydrodynamics, Vol 4. In *Methods in Compuational Physics.* Academic Press, 1—28.

Lewis, J. M. and J. C. Derber, 1985: The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus*, **37A**, 309—322.

Marchuk, G. I. and Y. K. Skiba, 1990: Role of adjoint functions in studying the sensitivity of a model of the thermal interaction of the atmosphere and ocean to variations in input data. *Bull. USSR Acad. Sci. Atmos. Oceanic Phys.*, **26**, 335—342.

Marchuk, G. I. and Y. N. Skiba, 1976: Numerical calculation of the conjugate problem for a model of the thermal interaction of the atmosphere with the oceans and continents. *Bull. USSR Acad. Sci. Atmos. Oceanic Phys.*, **12**, 459—469.

Marshall, J. S. and W. M. Palmer, 1948: The distribution of raindrops with size. *J. Meteor.*, **5**, 165—166.

Moore, A. M., 1991: Data assimilation in a quasi-geostrophic open-ocean model of the Gulf Stream region using the adjoint method. *J. Phys. Ocean.*, **21**, 398—427.

Morel, P., 1981: An overview of meteorological data assimilation. In *Dynamic meteorology data assimilation methods*, Bengtsson, L., Ghil, M., and Källén, E., editors, Springer-Verlag, New York, 5—16.

Murray, F. W., 1967: On the computation of saturation vapor pressure. *J. Appl. Met.*, **6**, 203—204.

Panofski, H. A., 1949: Objective weather-map analysis. *J. Meteor.*, **6**, 386—392.

Petschek, A. G. and L. D. Libersky, 1975: Stability, accuracy, and improvement of Crowley advection schemes. *Mon. Wea. Rev.*, **103**, 1104—1109.

Roux, F., J. Testud, M. Payen, and B. Pinty, 1984: West African squall-line thermodynamic structure retrieved from dual-Doppler radar observations. *J. Atmos. Sci.*, 41, 3104—3121.

Rutledge, S. A., 1986: A diagnostic modeling study of the stratiform region associated with a tropical squall line. *J. Atmos. Sci.*, **43**, 1356—1377.

Rutledge, S. A., 1989: A severe frontal rainband. Part IV: Precipitation mechanisms, diabatic processes and rainband maintenance. *J. Atmos. Sci.*, **46**, 3570—3594.

Rutledge, S. A. and P. V. Hobbs, 1983: The mesoscale and microscale structure and organization of clouds and precipitation in midlatitude cyclones. VIII: A model for the "seeder-feeder" process in warm-frontal rainbands. *J. Atmos. Sci.*, **40**, 1185—1206.

Rutledge, S. A. and P. V. Hobbs, 1984: The mesoscale and microscale structure and organization of clouds and precipitation in midlatitude cyclones. XII: A diagnostic

modeling study of precipitation development in narrow cold-frontal rainbands. *J. Atmos. Sci.*, **41**, 2949—2972.

Rutledge, S. A. and R. A. Houze, 1987: A diagnostic modeling study of the trailing stratiform region of a midlatitude squall line. *J. Atmos. Sci.*, **44**, 2640—2656.

Sasaki, Y., 1958: An objective analysis based on the variational method. *J. Meteor. Soc. Japan*, **36**, 77—88.

Sasaki, Y., 1969: Proposed inclusion of time variation terms, observational and theoretical, in numerical variational objective analysis. *J. Meteor. Soc. Japan*, **47**, 115—124.

Sasaki, Y., 1970: Some basic formalisms in numerical variational analysis. *Mon. Wea. Rev.*, **98**, 875—883.

Schlatter, T. W., 1975: Some experiments with a multivariate statistical objective analysis scheme. *Mon. Wea. Rev.*, **103**, 246—257.

Smolarkiewics, P. K., 1983: A simple positive definite advection scheme with small implicit diffusion. *Mon. Wea. Rev.*, **111**, 479—486.

Smolarkiewics, P. K., 1984: A fully multidimensional positive definite advection transport algorithm with small implicit diffusion. *J. Comput. Phys.*, **54**, 325—362.

Smolarkiewics, P. K. and T. L. Clark, 1986: The multidimensional positive definite advection transport algorithm: Further developments and applications. *J. Comput. Phys.*, **67**, 396—438.

Smolarkiewicz, P. K., 1989: Comment on "A positive definite advection scheme obtained by nonlinear renormalization of the advective fluxes". *Mon. Wea. Rev.*, **117**, 2626—2632.

Smolarkiewicz, P. K. and W. W. Grabowski, 1990: The multidimensional positive definite advection transport algoritm: Nonoscillatory option. *J. Comput. Phys.*, **86**, 355—375.

Smull, B. F. and R. A. Houze, 1987: Dual-Doppler radar analysis of a midlatitude squall line with a trailing region of stratiform rain. *J. Atmos. Sci.*, **44**, 2128—2148.

Sun, J., D. W. Flicker, and D. K. Lilly, 1991: Recovery of three-dimensional wind and temperature fields from simulated single-Doppler radar data. *J. Atmos. Sci.*, **48**, 876—890.

Talagrand, O., 1981a: On the mathematics of data assimilation. *Tellus*, **33**, 321—339.

Talagrand, O., 1981b: A study of the dynamics of four-dimensional data assimilation. *Tellus*, **33**, 43—60.

Taylor, G. D., 1991: Data assimilation and the adjoint method: An overview. Technical Report ISSN No. 0737-5352-22, CIRA, CSU, 13 pp.

Thacker, W. C., 1987: *Three lectures on fitting numerical models to observations.* GKSS Forschungszentrum Geesthacht GmbH, Geesthacht, FRG. External Report GKSS87/E/65, 64 pp.

Thacker, W. C., 1988a: Fitting models to inadequate data by enforcing spatial and temporal smoothness. *J. Geophys. Res.*, **93**, 10655—10665.

Thacker, W. C., 1988b: Large least-squares problems and the need for automating the generation of adjoint codes. *Computional solution of nonlinear systems of equations*, in Lectures in Applied Mathematics, Amer. Math. Soc., Providence RI.

Thacker, W. C., 1989: The role of the Hessian matrix in fitting models to measurements. *J. Geophys. Res.*, **94**, 6177—6196.

Thacker, W. C. and R. B. Long, 1988: Fitting dynamics to data. *J. Geophys. Res.*, **93**, 1227—1240.

Thompson, P. D., 1969: Reduction of analysis error through constraints of dynamical consistency. *J. Appl. Met.*, **8**, 738—742.

Tremback, C. J., J. Powell, W. R. Cotton, and R. A. Pielke, 1987: The forward-in-time upstream advection scheme: extension to higher orders. *Mon. Wea. Rev.*, **115**, 540—555.

Tripoli, G. J., 1986: *A numerical investigation of an orogenic mesoscale convective system.* PhD thesis, Colorado State Univ., Fort Collins CO 80523, 290 pp.

Tripoli, G. J. and W. R. Cotton, 1981: The use of ice-liquid water potential temperature as a thermodynamic variable in deep atmospheric models. *Mon. Wea. Rev.*, **109**, 1094—1102.

Tripoli, G. J. and W. R. Cotton, 1982: The Colorado State University three-dimensional cloud/mesoscale model - 1981. Part I: General theoretical framework and sensitivity experiments. *J. Rech. Atmos.*, **16**, 185—220.

Tziperman, E. and W. C. Thacker, 1989: An optimal-control/adjoint-equations approach to studying the oceanic general circulation. *J. Phys. Ocean.*, **19**, 1471—1485.

Verlinde, J. and W. R. Cotton, 1990: A critical look at kinematic microphysical retrieval algorithms. In *Conf. on Cloud Physics*, AMS, 453—457. San Fransisco.

Verlinde, J., P. J. Flatau, and W. R. Cotton, 1990: Analytical solutions to the collection growth equation: Comparison with approximate methods and application to cloud microphysics parameterization schemes. *J. Atmos. Sci.*, **47**, 2871—2880.

Vivekanadan, J., V. N. Bringi, and R. Raghavan, 1990: Multiparameter radar modeling and observations of melting ice. *J. Atmos. Sci.*, **47**, 549—564.

Walko, R. L., C. J. Tremback, and W. R. Cotton, 1989: Assimilation of Doppler radar wind data into a numerical prediction model: A demonstration of certain hazards. In *Preprints, 24th Conf. on Radar Meteor.*, AMS, Tallahassee, 248—250.

Zalesak, S. T., 1979: Fully multidimensional flux-corrected transport algorithms for fluids. *J. Comput. Phys.*, **31**, 335—362.

Zawadski, I. and M. de Agostino Antonio, 1988: Equilibrium raindrop distributions in tropical rain. *J. Atmos. Sci.*, **45**, 3452—3459.

Ziegler, C. L., 1985: Retrieval of thermal and microphysical variables in observed convective storms. Part I: Model development and preliminary testing. *J. Atmos. Sci.*, **42**, 1487—1509.

Ziegler, C. L., 1988: Retieval of thermal and microphysical variables in observed convective storms. Part II: Sensitivity of cloud processes to variation of the microphysical parameterization. *J. Atmos. Sci.*, **45**, 1072—1090.

Ziegler, C. L., P. S. Ray, and D. R. MacGorman, 1986: Relations of kinematics, microphysics and electrification in an isolated moutain thunderstorm. *J. Atmos. Sci.*, **43**, 2098—2114.