A METHOD FOR RETRIEVING THE CUMULUS ENTRAINMENT RATE FROM GROUND BASED OBSERVATIONS

by

Timothy J. Wagner

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Atmospheric and Oceanic Sciences)

at the

UNIVERSITY OF WISCONSIN-MADISON

2011

A METHOD FOR RETRIEVING THE CUMULUS ENTRAINMENT RATE FROM GROUND-BASED OBSERVATIONS

submitted to the Graduate School of the
University of Wisconsin-Madison
in partial fulfillment of the requirements for the
degree of Doctor of Philosophy

By

Timothy J. Wagner

Date of final oral examination: May 2, 2011

Month and year degree to be awarded: May 2011

The dissertation is approved by the following members of the Final Oral Committee:

David D. Turner, Assistant Professor, Atmospheric and Oceanic Sciences

Steven A. Ackerman, Professor, Atmospheric and Oceanic Sciences

Ralf Bennartz, Associate Professor, Atmospheric and Oceanic Sciences

Ankur R. Desai, Assistant Professor, Atmospheric and Oceanic Sciences

Larry K. Berg, Scientist, Pacific Northwest National Laboratory

Abstract

The entrainment of drier environmental air into cumulus clouds affects the impact that these clouds have on the environment by modifying their radiative, microphysical, and thermodynamic characteristics. Entrainment is a difficult parameter to observe directly, and heretofore has been obtained from occasional aircraft penetrations.

To increase the range of cumulus entrainment rate observations, an algorithm for retrieving the cumulus entrainment rate from ground-based remote sensing observations has been developed. This algorithm, called the Entrainment Rate In Cumulus Algorithm (ERICA), uses the suite of instruments at the Southern Great Plains (SGP) site of the United States Department of Energy's Atmospheric Radiation Measurement (ARM) Climate Research Facility as inputs into a Gauss-Newton optimal estimation scheme. The forward model in this algorithm is the Explicit Mixing Parcel Model (EMPM), a cloud parcel model that treats entrainment as a series of discrete mixing events. Output from EMPM is used to calculate quantities that can be observed from the surface, including effective radius and liquid water path. The entrainment rate in EMPM is adjusted iteratively until the modeled output converges to the observations. Sensitivity testing and error and information content analysis show that ERICA is a robust method for obtaining accurate measurements of entrainment rate without the drawbacks of aircraft observations.

Results from a three-month trial of ERICA show significant variability of the entrainment rate of clouds in a single day and from one day to the next. The mean value from this analysis corresponds well with prior knowledge of the entrainment rate.

Acknowledgements

This document represents the end of an era. Some might say that it feels like it was just yesterday that they graduated college and started graduate school; I don't really feel that way. I've grown so much as a scientist, an educator, and a person since then that it feels like a significant amount of time has elapsed. I write this acknowledgement section as I face the precipice of a new life, both personally and professionally. There are new adventures to be had, an I'm extremely excited about getting the opportunity to engage in them.

Many people have helped me during this process of growth, and I only have space to thank a few of them here. My primary advisor, Dave Turner, provided just the right amount of guidance over the past few years: always willing to share ideas when I got stuck, but never micromanaging my progress. Since I began working on this project, I've screamed for oxygen on the top of a peak in the Andes, strolled along the fog-soaked streets of San Francisco, floated down the Seine, and gazed in awe at medieval cathedrals. I have Dave to thank for integrating me into the scientific community at these international conferences and for trusting me to run a field project when I didn't even trust myself.

For years, I've had the pleasure to work with Steve Ackerman on both research and education projects. It was Steve, along with Wayne Feltz, who first saw potential in me as an undergrad and gave me the chance to prove myself as a graduate student. I learned so much about how to work with large lecture courses in the trial-by-fire that was Intro to Weather and Climate with Steve. Both Steve and Dave recognized early on what my goals in life were, and were happy to accommodate me in taking and teaching the classes that

would help lead me to that goal. The other members of my committee, Ankur Desai, Ralf Bennartz, and Larry Berg also provided numerous comments that improved the final dissertation and will be invaluable when it comes time for revising and publishing this work. Larry, especially, provided great insight over meals at conferences.

Thanks go to my parents, who never expressed to me any concern that they had about me still being in school. Thanks also to my officemates, who never showed annoyance when I had to get up and pace. Most importantly, however, thanks go to my wife, Erin. We met as grad students in the same incoming class, quickly became friends, and soon became much more than that. Wherever I go in life, I know that she will always be with me, supporting and encouraging me in the way that only she can. A mere dissertation acknowledgement is little recompense for all that she has offered me over the years.

This era is over. The new one begins in Omaha this summer. To all who helped me get here, my sincerest thanks.

Tim Wagner

May 2011

Table of Contents

1. Introduction 1
2. Cumulus entrainment and its impact 6
   a. The evidence for entrainment 6
   b. Impact of entrainment on the cloud and environment 9
   c. Quantifying and measuring entrainment 11
   d. The need for entrainment measurements 13
3. Retrieving the entrainment rate 16
   a. Optimal estimation 16
   b. The Entrainment Rate In Cumulus Algorithm (ERICA) 22
4. The Explicit Mixing Parcel Model 26
5. Observations and Instrumentation 34
   a. Site Selection 34
   b. Atmospheric Emitted Radiance Interferometer (AERI) 37
   c. Raman Lidar 39
   d. Millimeter wavelength Cloud Radar 41
   e. Aerosol Observing System 46
   f. Microwave radiometer 47
6. Algorithm Implementation 50
   a. Quantitative measures of uncertainty and information content 50
   b. Stochastic sensitivity 52
   c. Cloud identification and selection 55
7. Case study and validation 57
   a. 18 June 2009 57
   b. Validation 60
8. Multi-day analysis 70
   a. Introduction and descriptive statistics 70
   b. Aircraft comparison and validation 73
   c. Dependence on other variables 74
   d. Impact of the first guess 75
9. Summary and outlook 91
10. References 96

Table of Figures

Figure 1.1: Scale drawing of a typical 12 km forecast model grid cell and the distribution of fair weather cumulus that can occur within that domain 5
Figure 3.1: Generalized flowchart of the Entrainment Rate In Cumulus Algorithm 24
Figure 4.1: Diagram of the triplet mapping procedure. 27
Figure 4.2: Illustration of the impact of an eddy on a scalar field with a uniform gradient..28
Figure 5.1: Map of the state of Oklahoma, noting the position of the Southern Great Plains site (SGP) relative to the two largest cities in the state 35
Figure 5.2: Typical spectrum recorded by the Atmospheric Emitted Radiance Interferometer (AERI). 38
Figure 5.3: Scatter plot of the logarithm of effective radius (in microns) observed by the Cloud Airborne Spectrometer (CAS) in CHAPS against the radar reflectivity (in dBZ) calculated from that volume 45
Figure 6.1: Plot of the retrieved entrainment rate for a single cloud (observed at 2142 UTC on 18 June 2009) with both the entrainment rate and the turbulent mixing in EMPM set to include stochastic modeling (described as *Full Random*) in the text 54
Figure 6.2: As in Figure 6.1, but for randomness only in the turbulent mixing. Note the reduced degree of variability in both the measurements and the error (described as *Mix Only*) in the text. 54
Figure 7.1: Total Sky Imager (TSI) photographs of cloud cover at the SGP site on 18 June 2009. 58
Figure 7.2: Time series of microwave radiometer retrieval of cloud liquid water path over the ARM SGP site on 18 June 2009. 61
Figure 7.3: Retrieved values of the entrainment rate as a function of time for 19 cumulus clouds over the SGP central facility on 18 June 2009 62
Figure 7.4: Plot of aircraft-observed liquid water content (in g m^{-3}) for 18 June 2009 as a function of aircraft position, showing the "stacked triangle" pattern common to numerous RACORO flights. 65
Figure 7.5: South elevation view of Figure 7.4 66
Figure 8.1: Distribution of retrievals by day for the May-June 2009 dataset 78
Figure 8.2: Histogram of times of valid retrieval for the ERICA-processed May-June 2009 dataset illustrated in Figure 8.1 79
Figure 8.3: Histogram of ERICA-retrieved entrainment rates for cumulus clouds over the ARM SGP site from May through July 2009 80
Figure 8.4: As in Figure 8.3, but for the error from the ERICA retrievals 81
Figure 8.5: Scatter plot of aircraft observed entrainment rate (x-axis) against mean ERICA retrieved entrainment rate (y-axis) 82
Figure 8.6: Time series of bin-averaged entrainment rates (in km^{-1}) for the May-June 2009 dataset 83
Figure 8.7: Scatter plot of observed effective radius (in microns) against the retrieved entrainment rate (in km^{-1}). 84
Figure 8.8: As in Figure 8.7, but for liquid water path (in g m-2). 85
Figure 8.9: As in Figure 8.7, but for vertical velocity (in m s-1) 86

Figure 8.10: Scatter plot of ERICA-retrieved entrainment rates of the May-June 2009 dataset for two separate first guesses.....87
Figure 8.11: Histogram of the degrees of freedom (a measure of information content and a unitless parameter) for the May-June 2009 retrievals. 88
Figure 8.12: Scatter plot of degrees of freedom (on the x-axis) and error of the retrieval (on the y-axis, in km-1).....89
Figure 8.13: As in Figure 8.12, but for the retrieved entrainment rate (in km-1).....90

Table of Tables

Table 5-1: ARM SGP instruments used in ERICA.....35
Table 5-2: Boundary Layer and General Operational Mode for SGP MMCR..... 43
Table 7-1: Variances of mixing ratios in the 18 June 2009 case study flights.....70
Table 7-2: Entrainment rate observations for 18 June 2009.....69
Table 8-1: Descriptive statistics of entrainment rate retrievals for May-July 2009.....71
Table 8-2: Correlation coefficients and probability between retrieved entrainment rate and selected variables.....76

1. Introduction

Cumulus clouds are a key component in the thermodynamic and radiative balance of the earth's troposphere. The thermal plumes that create cumuli help transport latent heat from the surface to the free atmosphere, while the condensation and evaporation processes provide a means to redistribute latent heat throughout the troposphere. In addition, cumuli are responsible for a net radiative cooling of the earth, due to reflecting more solar radiation than absorbing terrestrial radiation (Berg et al. 2011). Cumuli are widespread throughout the world; while common in both continental and marine environments, there are subtle differences in their structure that are a function of the environment in which they form. For example, continental cumuli tend to have higher droplet number concentrations and smaller droplet sizes than marine cumuli (Rogers and Yau 1989) due to an increased number of cloud condensation nuclei over land than over the sea. Liquid water content, however, is about the same for both continental and marine cumuli. A result of this is that marine cumuli are more likely to precipitate than continental cumuli, as precipitation growth processes favor having fewer larger droplets over many smaller ones (Fan et al. 2007).

The macrophysical and microphysical properties of cumulus clouds are highly dependent on the dynamic and thermodynamic conditions that surround them. This complicated interaction between continental cumuli and their environment make these clouds an important part of any effort to numerically model the atmosphere. However, their small spatial domain insures that they are smaller than the grid spacing of forecast

and climate models. For example, at the time of this writing, the grid spacing of the Weather Research and Forecasting Non-hydrostatic Mesoscale Model (WRF-NMM), run by the National Weather Service for operational forecasting, is 12 km. The spatial domain of a single continental cumulus is on the order of 1 km while separations between them are on the order of 3 km (Cotton et al. 2011). On a quiescent day, several dynamically independent cumuli can be found within the spatial extent of a single grid cell (Figure 1.1).

Traditionally, boundary-layer cumulus have been classified according to their morphology. The World Meteorological Organization notes three species of cumulus: humilus, marked by little vertical extent; mediocris, with moderate vertical extent and small protuberances on their tops; and congestus, with significant vertical extent often accompanied by a characteristic cauliflower appearance (WMO 1975). Instead of a morphological classification, Stull (1985) proposed one that divided cumuli into forced, active, and passive categories based on the stage of development. Cumuli are initially in the forced stage of development: as moist plumes ascend from the mixed layer and penetrate the stable layer condensation occurs but the plume remains negatively buoyant relative to its surroundings. Should the updraft be sufficiently vigorous to penetrate the level of free convection (LFC), the cloud is now able to grow with its own buoyancy and form circulations that are separate from the original plume. This marks the active stage of development, and it contains both mediocris and congestus cumuli. Finally, once the thermal has died off and the cloud begins to decay, the cloud enters the passive stage. Passive cumuli have no interaction with the mixed layer, and unlike the other stages they have a ragged base as a result of the decay process.

The active stage contains a particularly vexing problem to the modeling community. Cumuli mix with the environment through the entrainment process, wherein dry environmental air is engulfed by and assimilated into the moist cloud environment. This significantly impacts the cloud by reducing its liquid water content, lowering its buoyancy, increasing its decay rate, and altering its microphysical characteristics. These processes, in turn, affect the impact that the cloud has on its environment. Therefore, any serious effort to accurately model the atmosphere must account for cumulus entrainment, but as the clouds are sub-gridscale phenomena all of their processes must be parameterized instead of directly modeled.

In order to evaluate existing cumulus parameterizations and develop new ones, robust measurements of how much entrainment is taking place under a variety of environmental conditions are desired. Measurements of the entrainment rate, however, are not easy to come by. In the literature, examples of calculating the entrainment from research aircraft penetrations are found (e.g. Raga et al. 1990), but the expense and liability associated with such missions relegates them to rare occasions. As shown in Chapter 7 of this work, the magnitude of the error associated with these observations is also quite large, on the order of the observation itself. The paucity of robust, near-continuous observations means that little is known about the behavior of entrainment, its relationship to surface fluxes or upper-level dynamics, or even its basic probability density. There is a demonstrated need for another method of measuring entrainment so that models can be validated, parameterizations can be evaluated, and forecasts can be improved for both the short term and climatological timescales.

This work introduces a new method of measuring entrainment: the Entrainment Rate In Cumulus Algorithm (ERICA). By integrating ground-based remotely-sensed observations with a cumulus parcel model, ERICA is able to retrieve the entrainment rate without the use of in situ aircraft-borne instruments. This enables the development of a robust array of entrainment observations covering all parts of the seasonal and diurnal cycle. The datasets that are required by ERICA date back for the past four years, and thus there is a broad assortment of conditions for investigation.

The rest of this work is devoted to discussing ERICA and the impetus behind its development in greater detail. Chapter 2 presents a discussion on entrainment, including a look at the evolution of its conceptual model, its impact on the cloud and environment, and how to quantify it. In Chapter 3, an outline of ERICA is presented along with a mathematical explanation of its operation. The forward model that contains all of the cloud physics at work in ERICA is described in depth in Chapter 4, while Chapter 5 introduces the instruments and datasets in use. Chapter 6 remarks on implementation issues, including error analysis, stochastic sensitivity, and cloud identification. A case study that uses ERICA to evaluate the entrainment rate of a typical cumulus field is presented in Chapter 7, along with airborne observations of that same field. Chapter 8 is devoted to a two-month long climatology of entrainment rates for continental cumulus. The conclusions and future outlook of this work are discussed in Chapter 9.

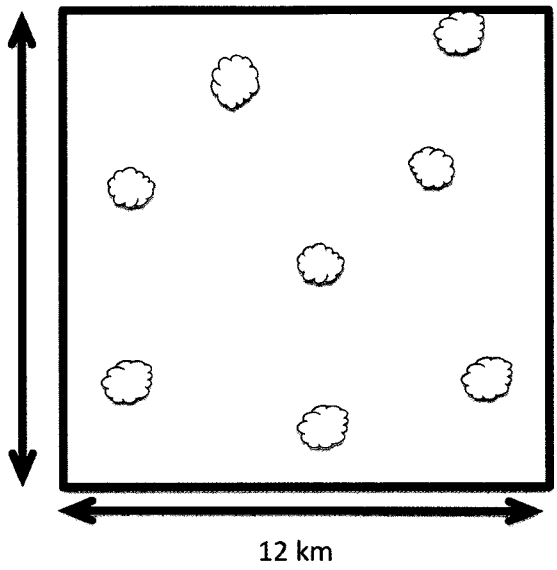

Figure 1.1: Scale drawing of a typical 12 km forecast model grid cell and the distribution of fair weather cumulus that can occur within that domain. The clouds are 1 km in extent and are separated by a distance that is on the order of 3 km.

2. Cumulus entrainment and its impact

a. The evidence for entrainment

Entrainment of environmental air into cumulus clouds has been recognized as an important part of cloud dynamics from the earliest days of in situ cloud liquid water observations. Since then, the mechanisms through which air is assimilated into cumuli have been a subject of debate. In a way, the conceptual model of entrainment has come full circle from its earliest days. Highlights of the literature are presented here; detailed reviews from the perspective of observational experiments and thermodynamic theory and modeling have been written by Blyth (1993) and Siebsma (1998) respectively.

Stommel (1947) was the first to hypothesize that cumulus clouds entrained environmental air as they ascended in order to explain thermodynamic observations that were inconsistent with the assumption of moist adiabatic ascent for convective parcels. Additional evidence for entrainment in these initial in situ observations included inhomogeneity on all scales down to the centimeter scale, regions of low liquid water content (LWC) distributed randomly throughout the cloud sample (Squires 1958a), and no preference for reduced LWC near cloud edge nor systematic variations in LWC to the middle (Warner and Squires 1958). Furthermore, the LWC was less than the value predicted by adiabatic theory (Warner 1955) and the in-cloud temperature lapse rate was close to the environmental lapse rate rather than the saturated adiabatic lapse rate (Squires 1958b).

The earliest models of entrainment assumed lateral mixing, in which an ascending undiluted parcel engulfs a constant stream of environmental air from beside the cloud as it ascends and mixes it instantaneously (e.g. Squires and Turner 1962); once a parcel mixes its buoyancy is reduced and it no longer ascends. However, observational studies showed that lateral entrainment models could not accurately reproduce both LWC and the cloud top height with one entrainment rate (Warner 1970). This led to the cloud-top entrainment hypothesis (Telford 1975), in which entrainment only occurs at the equilibrium level at maximum possible cloud height. Evaporative cooling of parcels at cloud top leads to penetrative downdrafts that mix throughout the cloud. Paluch (1979) supported this theory with empirical evidence through the use of mixing line diagrams; if two parcels are mixed together, a diagram of a conserved variable in the new parcel will lie on a straight line that connects the two thermodynamic states of the original parcels. It is possible to find the origin of the entrained air using this technique, and Paluch applied her technique to sailplane penetrations of high plains cumuli to determine that the entrained air usually originated from an altitude slightly below the top of the cloud. Kitchen and Caughey (1981) used tethered balloon observations to diagram the flow within cumulus plumes and found a characteristic overturning at cloud top in the direction of the mean wind shear, creating a P-shaped flow that also bolstered the hypothesis of cloud top entrainment.

For decades, this explanation dominated the analysis and modeling of cloud mixing. Yet, as Siebesma (1998) noted, "[This conceptual model] is rather odd since it suggests that clouds would only entrain once they have reached their maximum cloud top."

Furthermore, additional studies using the Paluch technique found that the origin of entrained air was typically very close to the level where the observations were taken (e.g. Fankhauser et al. 1982, Raymond and Wilkening 1982, Gardiner and Rogers 1987). When Blyth (1988) found similar conditions in summertime cumuli in Montana, he proposed the theory of ascending cloud top entrainment: rather than entraining from the cloud's maximum possible height, the cloud instead entrains from the level of the ascending cloud top as it grows. It still verified the mixing-line analysis by Paluch (1979) yet accounted for the origin of entrained air at different levels.

Still, there were elements of cumulus observations that could not be explained with this model. Taylor and Baker (1991) showed that the mixing line predictions would still verify if most cloud parcels are near their level of neutral buoyancy. They proposed a buoyancy sorting mechanism, in which newly-mixed parcels ascend or descend to the level of their new buoyancy. The lack of observed buoyant parcels is easily explained by the fact that vertical velocity is proportional to buoyancy; parcels that have a large positive (negative) buoyancy rapidly ascend (descend) to a new level of neutral buoyancy as the cloud settles into a state of buoyant equilibrium.

The current conceptual model of cumulus entrainment (Taylor and Baker 1991, Grabowski 1993, Krueger et al. 1997) echoes the earliest models of lateral entrainment with key differences. Convective scale eddies within the cloud subsume discrete units, or "blobs," of environmental air as they ascend. These blobs are stretched and compressed via turbulent deformation into successively thinner filaments until the Kolmogorov scale is reached and diffusion homogenizes the air at the molecular level. Due to the turbulent

mixing, the parcels are not necessarily homogenized at the macroscopic level, thus explaining the observed variability in LWC. While these parcels are mixing laterally as they ascend, they differ from the primitive plume models in that the mixing is neither instantaneous nor does it only involve undiluted parcels. A method to model cumulus in light of this conceptual framework is described in further detail in Chapter 4.

b. Impact of entrainment on the cloud and environment

Much work has been done on quantifying the microphysical effects of entrainment. One of the early impetuses for studying cumulus entrainment was to resolve the lack of agreement between models and cloud probe observations. It was found that in addition to underestimating the LWC, adiabatic models of cumulus predicted much narrower droplet size distributions than observed and were also unable to resolve the bimodal distribution that is often seen by in situ probes. Warner (1973) attempted to show the impact of entrainment on size distributions with a laterally entraining classic plume model, but was unsuccessful in recreating the distributions observed by Warner (1969). As a result, Warner determined that mixing was not responsible for the observed broadening. However, later studies reexamined the problem without the limitations of homogeneous mixing. By using the modern framework for parcel ascent and mixing, both one-dimensional parcel models (Su et al. 1998) and three-dimensional cloud models (Lasher-Trapp et al. 2005) have shown that entrainment is responsible for the observed broadening in droplet size spectra.

It has also been hypothesized that entrainment increases the rate at which large droplets grow. Baker et al. (1980) proposed that reduced competition for water vapor in diluted parcels would result in increased supersaturation and thus faster droplet growth rates for the remaining droplets when compared to adiabatic parcels. As condensational droplet growth is linearly proportional to the droplet diameter, these larger droplets would grow even faster and could even reach embryonic raindrop size (Telford and Chai 1980). Since then, numerous authors have found results that counteracted (e.g. Jonas and Mason 1983) or bolstered these findings (e.g. Cooper 1989) and the impact of entrainment on droplet growth rate is still a matter of debate.

Entrainment also affects the state of cumuli by importing new cloud condensation nuclei (CCN) into the parcel. Pontikis and Hicks (1993) modeled warm maritime cumulus droplet spectra using a parcel model in order to determine the impact of entrainment on effective radius, r_e:

$$r_e = \frac{\int n(r) r^3 dr}{\int n(r) r^2 dr} \tag{2.1}$$

where r is the droplet radius and $n(r)$ is the droplet count in the radius r to $r+dr$. Due to the dependence of the effective radius on the cube of droplet radius, the development of a few large droplets will force an increase in the effective radius. They also found that cloud effective radius was found to be independent of entrainment at a given cloud level if additional cloud condensation nuclei were not entrained, providing additional evidence against the impact of entrainment on droplet growth rate. However, when secondary

activation effects from entrained CCN were included, a wide variety of effective radii are produced from a given liquid water content. In a similar vein, Krueger et al. (2006) used a parcel model to determine the influence of entrained CCN on cloud droplet spectra. The model runs that included entrained CCN better matched observations of cloud droplet number concentration and mean droplet radius than the control runs.

These microscale effects in turn change the impact that the cloud has on the environment. For example, light extinction is proportional to the total droplet surface (Stephens 1978), and thus the spectral broadening and enhancement of the concentration of small droplets induced by the entrainment process results in greater solar extinction due to increased reflection.

c. Quantifying and measuring entrainment

As an ascending parcel entrains, its mass increases. This rate of mass change is used to quantify the amount of entrainment taking place through the fractional entrainment rate, λ, which is defined as the fractional rate of change of the parcel mass m per unit height z of ascent:

$$\lambda = \frac{1}{m}\frac{dm}{dz} \tag{2.2}$$

Typically, the entrainment rate is expressed in units of km^{-1}. For example, a cloud with an entrainment rate of 1.0 km^{-1} will entrain a quantity of air equal to 100% of its initial mass for every kilometer it ascends.

To measure the entrainment rate through in situ observations, it is necessary to measure an adiabatically conserved quantity. In the absence of mixing or precipitation, an adiabatically conserved quantity like moist static energy or total water mixing ratio remains constant throughout parcel ascent; the observed deviation from adibaticity for one of these quantities is proportional to the amount of air mixed into the parcel. This is shown in the mixing equation:

$$\frac{\partial \phi}{\partial z} = -\lambda(\phi - \phi_e) \tag{2.3}$$

where ϕ and ϕ_e represent an adiabatically conserved quantity inside and outside of the cloud respectively. Through the use of aircraft penetrations, it is possible to calculate the fractional entrainment rate from combined internal and external cloud observations. Raga et al. (1990) used this method on data obtained during flights through Hawaiian cumuli and found an average entrainment rate of 1.3 km^{-1}. Note that Equation 2.3 requires the assumption of a steady-state environment. As the lifespan of a given cumulus cloud is on the order of 30 minutes, significant changes may have occurred in the cloud's structure between aircraft penetrations especially if the flight plan involves penetrating multiple clouds on a single leg.

Siebsma and Cuijpers (1995) used a large eddy simulation (LES) coupled with assumptions about large-scale forcing to model tropical cumuli representative of those observed during the Barbados Oceanographic and Meteorological Experiment (BOMEX) in 1969. Through decomposition of the prognostic equations and application of a quasi-Boussinesq approximation for the actively growing clouds, they were able to obtain

characteristic entrainment rates from the LES output. For these tropical cumuli, they found representative entrainment values of roughly 3 km^{-1} for the whole updraft and 1.5 to 2.0 km^{-1} for the core. However, as their spatial resolution was 125 m, smaller clouds would not have been represented.

d. The need for entrainment measurements

The small size of cumuli ensures that they are sub-grid phenomena that must be parameterized in climate and numerical weather prediction models. Early attempts at cumulus parameterization (e.g. Squires and Turner 1962) assumed that the entrainment rate was inversely proportional to the radius of the thermal:

$$\lambda = \frac{\eta}{R} \tag{2.4}$$

where the non-dimensional proportionality constant η is often assumed to be equal to 0.2. This scheme found numerous operational uses, including the European Center for Medium-range Weather Forecasts (ECMWF) mass flux parameterization (Tiedtke 1989). As the thermal radius is on the order of 1 km, the result of this parameterization is that the assumed entrainment rates are roughly an order of magnitude less than the observations and model studies cited above. This creates a modeled environment in which cloud tops are much higher than in reality; subsequently moisture and heat mixing are too deep. Siebsma and Cuijpers (1995) postulated that the reason for the discrepancy in entrainment rates was due to the assumption that a typical cloud represents the entrainment rate for a

whole ensemble of clouds over a particular domain. To rectify this, they proposed that the ensemble entrainment rate was a weighted average of the entrainment rates of individual subsets of clouds defined by cloud top height. Modern forecast models, like the Weather Research and Forecasting (WRF, Skamarock et al. 2001) model use the Kain-Fritch convective parameterization, which assumes that the entrainment rate is proportional to the updraft mass flux and inversely proportional to the cloud radius (Kain 2004).

An issue with any parameterization scheme is the need to validate it against observations. The previous section noted two ways of obtaining the fractional entrainment rate: through aircraft penetrations at multiple levels or through LES modeling of a characteristic environment. Both approaches have drawbacks. The primary issue with aircraft observations is that they are not routine. Aircraft probes are only utilized during field projects and intensive observation periods (IOPs) as routine aircraft-based observations at multiple levels within the same cloud are too expensive for operational use. Furthermore, the light aircraft typically used for cumulus penetrations are unable to fly through supercooled liquid clouds due to icing concerns, which reduces the conditions and seasons in which evaluations can take place. These factors limit the range of entrainment observations and inhibit the ability to investigate connections between the entrainment rate and other environmental parameters. The modeling approach by Siebesma and Cuijpers (1995) did not attempt to retrieve entrainment rate for specific clouds. Instead, it used fields initialized by rawinsondes launched every 1.5 hours over the course of two days in order to create a characteristic steady state and inferred the entrainment rate from that.

Neither approach is able to generate measurements over a wide variety of seasons and times of day. A different measurement method is required.

The approach presented in the current work addresses the problems with these other approaches. By creating an entrainment rate product from remotely-sensed observations it is possible to have a near-continuous record of entrainment rate measurements regardless of aircraft status or field project operations. Such a product could be used to retrieve the entrainment rate from specific clouds, thus allowing the end-user to determine how large he or she wants the ensemble to be. Furthermore, since the necessary instruments have been recording observations for several years, no additional capital cost is required and a dataset based on historical observations can be readily produced for analysis. This product would facilitate the validation of existing cumulus entrainment parameterizations and enable the development of new ones by providing a bank of observations that encompasses a variety of cumulus-generating states, which would in turn eventually lead to more accurate weather and climate predictions.

3. Retrieving the entrainment rate

a. Optimal estimation

Remotely sensed observations can provide valuable information when in situ observations are too expensive or too difficult to fulfill the measurement requirements. Typically, remote observations are indirect; rather than measuring the desired variable, one instead measures a different quantity that can be used to approximate the desired value through mathematical manipulation. As a simple example, a space-borne radiometer does not directly measure surface temperature. Instead, it measures emitted radiance at one or more wavelengths that can be used to calculate the surface temperature through Planck's law provided the emissivity of the surface and the transmittance of the atmosphere can be assumed. The inversion is "trivial" in this case.

Most indirect measurements are not that simple. Many times the system is underdetermined, which results in an infinite set of possible solutions. This is the typical situation for profile retrievals where one is attempting to retrieve a continuous function and only has a finite number of observations with which to do so. The challenge is selecting the best possible solution from the limitless possibilities while accounting for the characteristics of the instrument, the noise in the observations, and the existing knowledge about the desired measurement. Additionally, these problems are usually non-linear and must be linearized about a base state. Problems such as these are known as inverse problems, and several techniques have been developed in order to solve them. Gauss-Newton Optimal Estimation is one such method, notable for its accommodation of non-

linear models as well as its ability to propagate model and measurement error throughout the inversion. This makes it possible to quantify the magnitude of the error for the retrieval.

The mathematical treatment of Gauss-Newton Optimal Estimation presented here is a streamlined version of the full treatment shown in Rodgers (2000). Fundamentally, the goal of an inverse retrieval is to transform a measurement vector, y, containing the values that can be measured into a state vector, x, that contains all of the quantities that are desired to be retrieved. The length of y and x are not necessarily identical. The two vectors are related to each other via a forward model F that encapsulates the physics of the relationship:

$$y = F(x) + \varepsilon \,, \tag{3.1}$$

where ε is the model error. Since the model is unable to perfectly represent the physics due to parameterizations, assumptions, and rounding, this ε vector is needed to account for the difference between reality and model-space. Of course, the analysis of the problem is computationally and conceptually much simpler if it can be linearized, therefore a first-order Taylor expansion about a base state x_0 is performed:

$$y = F(x_0) + \frac{\partial F(x)}{\partial x}(x - x_0) + \varepsilon \,. \tag{3.2}$$

It is convenient to define a matrix of derivatives, K, composed of elements representing the partial derivative of the forward model with respect to the state vector:

$$K_{ij} = \frac{\partial F_i}{\partial x_j}. \tag{3.3}$$

This K matrix is often called the Jacobian or the weighting function matrix. Making this substitution and rearranging the elements of the equation gives us the following relationship between y, x, and K:

$$y - F(x_0) = K(x - x_0) + \varepsilon. \tag{3.4}$$

Bayesian probability is fundamentally important to the process of retrievals as it determines the probability of one quantity given the state of another. In this case, given values in the y vector, Bayes's theorem can be used to find the probability of a range of values within x. By expressing the prior knowledge of the variable we desire (i.e., the state vector) in the form of a probability density function (or pdf), and using a forward model that transforms state quantities into quantities in the measurement space, Bayes's theorem can be used to create a new pdf that updates the prior pdf using information obtained by the measurement. Since there are an infinite number of possible solutions, this pdf is used to find the one that is the most likely given the measurements. Mathematically, Bayes's theorem is expressed as:

$$P(x \mid y) = \frac{P(y \mid x)P(x)}{P(y)} \tag{3.5}$$

where $P(x)$ is the prior pdf of the state, marking the knowledge of x before the measurement is made; $P(y)$ is the prior pdf of the measurement, marking the quantitative knowledge of the measurement before it is made; $P(y|x)$ is the conditional pdf of y given x, denoting the probability of a measurement in a certain range given a specific state value; and $P(x|y)$ is the conditional probability of x given y, giving the probability for a value of the state given a specific measurement. Quantifying this last value, $P(x|y)$, is the goal of a retrieval algorithm since the measurements are known and the states are not. By quantifying the other three probabilities in Equation (3.5) it is possible to ascertain that value. The value of $P(y)$ is obtained from knowledge of the noise level of the observations and also the covariance between these observations, while $P(x)$ is an estimate about the state of the atmosphere which often comes from climatology.

Gaussian distributions are used for these pdfs due to their mathematical simplicity and their ability to approximate many distributions that are pertinent to these types of problems. For a vector y, the Gaussian distribution is:

$$P(y) = \frac{1}{(2\pi)^{\frac{n}{2}} |S_y|^{\frac{1}{2}}} \exp\left[-\tfrac{1}{2}(y-\bar{y})^T S_y^{-1}(y-\bar{y})\right] \tag{3.6}$$

where S_y is the covariance matrix of the elements of y and the superscripts T and -1 denote the matrix transpose and inverse respectively. By taking Equation (3.6) and substituting Equation (3.4), which expresses y in terms of the linearized forward model and the error, the conditional pdf for y is obtained:

$$-2\ln P(y \mid x) = (y - Kx)^T S_\varepsilon^{-1}(y - Kx) + c_1 \tag{3.7}$$

where S_ε is the matrix describing the measurement error covariance and c_1 is a constant. Similarly, an expression for the probability of x can also be expressed in the form of a Gaussian distribution:

$$-2\ln P(x \mid y) = (x - x_a)^T S_a^{-1}(x - x_a) + c_2 \tag{3.8}$$

where x_a is the a priori state vector and S_a is its covariance matrix. Substituting these expressions into Equation (3.5), Bayes's Theorem, the conditional pdf of x is derived:

$$-2\ln P(x \mid y) = [y - F(x)]^T S_\varepsilon^{-1}[y - F(x)] + (x - x_a)^T S_a^{-1}(x - x_a) + c_3. \tag{3.9}$$

As the goal of these mathematical operations is finding the values of x that have the greatest probability of producing the observed values in y, the appropriate next step is to take the derivative of Equation (3.9) with respect to x and set it equal to zero:

$$0 = -\hat{K}(\hat{x})^T S_\varepsilon^{-1}[y - F(\hat{x})] + S_a^{-1}(\hat{x} - x_a) \tag{3.10}$$

where the ^ operator denotes an estimated quantity. In this case, $\hat{x}$ represents the maximum probability state for x, and Equation (3.10) represents it entirely in terms of values that are either known a priori or can be calculated with use of the forward model. The only remaining step is to use a numerical method that finds a vector $\hat{x}$ that minimizes the residual, or difference between the observations and modeled values. Newton's method for nonlinear regression for a generic vector equation g(x)=0 is expressed as:

$$x_{i+1} = x_i - [\nabla_x g(x_i)]^{-1} g(x_i). \tag{3.11}$$

Were the problem truly linear (or at least linear within the accuracy of the measurements), an iterative solution would not be necessary as the problem could be solved algebraically. By substituting the right hand side of Equation (3.10) into Equation (3.11), the following expression is obtained:

$$x_{i+1} = x_i - \{S_a^{-1} + K_i^T S_\varepsilon^{-1} K_i - [\nabla_x K_i^T] S_\varepsilon^{-1} [y - F(x)]\}^{-1} \{-K_i^T(x) S_\varepsilon^{-1} [y - F(x)] + S_a^{-1}(x - x_0)\} \quad (3.12)$$

Note the presence of a second derivative in the form of $\nabla_x K^T$, since K is the derivative of the forward model. This expression is computationally expensive to evaluate, but since it is small for moderately non-linear problems and its term becomes smaller as the system iterates due to the residual of y-F(x) approaching zero, it is safe to ignore this term for the purposes shown here. This yields the Gauss-Newton method:

$$x_{i+1} = x_i + (S_a^{-1} + K_i^T S_\varepsilon^{-1} K_i)^{-1} \{K_i^T S_\varepsilon^{-1} [y - F(x_i)] - S_a^{-1}(x_i - x_a)\}. \quad (3.13)$$

The last step is to rearrange this expression in terms of an iteration that departs from x_a. This yields the form of the Gauss-Newton method for optimal estimation that is used in the current study:

$$x_{i+1} = x_a + (S_a^{-1} + K_i^T S_\varepsilon^{-1} K_i)^{-1} K_i^T S_\varepsilon^{-1} [y - F(x_i) + K_i(x_i - x_a)]. \quad (3.14)$$

To stop the iterative process, some measure of convergence must be assessed at each step. One way to do this is to compare the difference in the retrieved variables from the current iteration to the previous one. If the change in the retrieved state vector is

substantially less than the magnitude of the error associated with the retrieval, then it is assumed that the retrieval has converged within its own error bars and additional iterations would bring no additional information. This is evaluated through the following expression:

$$(x_i - x_{i-1})^T (S_a^{-1} + K_i^T S_\varepsilon^{-1} K_i)(x_i - x_{i-1}) << n \qquad (3.15)$$

where n is the number of variables in the state vector. In this case the iterative process is stopped once Equation (3.15) is less than or equal to n/10.

b. The Entrainment Rate In Cumulus Algorithm (ERICA)

As described above, a forward model in a Gauss-Newton iterative retrieval is a system that maps variables in state space to the measurement space. By using a cloud parcel model as a forward model, a cloud characteristic that is difficult to observe directly, such as entrainment rate, can be transformed into parameters that can be observed, such as liquid water path and effective radius. If such a model is initialized with a guess of entrainment rate based on some prior knowledge of the situation, then Gauss-Newton optimal estimation can be used to adjust the entrainment rate guess until a reasonable estimate for its actual value is obtained. This is the conceptual framework behind the Entrainment Rate In Cumulus Algorithm (ERICA) that was developed for this investigation. In this case the forward model is the explicit microphysics version (Su et al. 1998) of the Explicit Mixing Parcel Model (EMPM, Krueger et al. 1997), a one-dimensional cloud parcel

model that treats entrainment as a series of discrete mixing events. EMPM is described in greater detail in Chapter 4 of this work.

ERICA begins by ingesting data about the cloud environment needed to execute EMPM. This includes cloud base temperature, pressure, and water vapor mixing ratio; vertical profiles of temperature and water vapor mixing ratio; and cloud vertical velocity. The instruments from which these observations are obtained are described in detail in Chapter 5. In addition to the observations, first guesses of entrainment rate and cloud droplet number concentration are required. ERICA ingests this data and assembles the Fortran-77 source code required to execute EMPM. Due to the explicit microphysics calculations, EMPM is computationally expensive to run. Each iteration of the forward model requires about six minutes of processing time on locally available hardware, with the majority of this time used to estimate the Jacobian using finite differences. In order to decrease the wall-clock time spent calculating the double-ended finite differences required to compute the derivatives that comprise the Jacobian, a "poor-man's parallelization" was undertaken. ERICA generates five different executable files, each with a slightly different entrainment rate or number concentration, and distributes them to separate nodes of a computer cluster where they run simultaneously. After the executions are complete, ERICA retrieves the data from the nodes and uses it to calculate the derivatives of quantities that can be observed (in this case the effective radius and liquid water path) with respect to the entrainment rate. A new guess for the state variables is calculated using Equation (3.14) and the cycle iterates until the convergence criteria (Equation 3.15) are met; this usually takes between six and ten iterations. A flowchart of this process is shown in Figure 3.1.

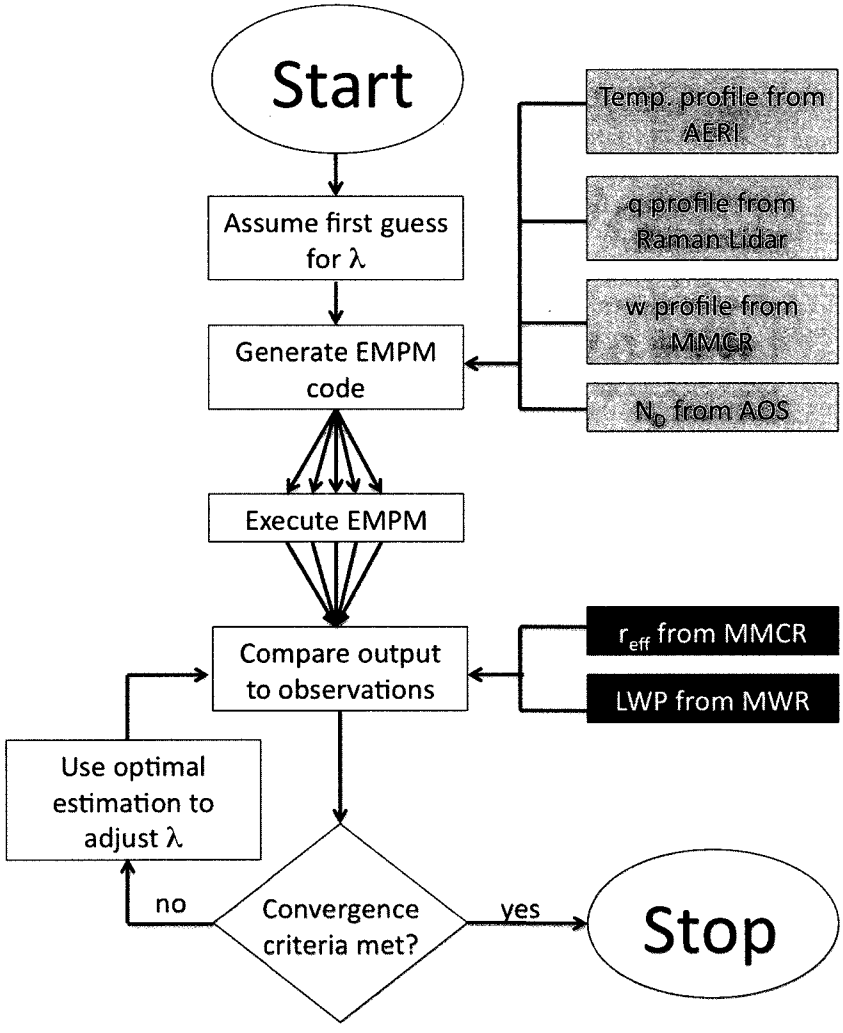

Figure 3.1: Generalized flowchart of the Entrainment Rate In Cumulus Algorithm, or ERICA. Light shaded boxes represent data sources used to run the model, while dark shaded boxes represent data sources used to make the optimal estimation adjustments. These data streams are described in detail in Chapter 5.

A key component of the optimal estimation calculation is the prior state vector and its associated covariance (x_a and S_a respectively). The a priori value of entrainment rate is

set at 1.5 km^{-1} based on values found in the literature (e.g. Raga et al. 1990). Due to the relative paucity of observations of entrainment rate, the variance is set at 1 km^{-1} in order to provide ample opportunity for the retrieval to move away from the a priori value in search of a solution; this also encompasses the range of reported entrainment rates.

An initial assumption was that the cloud droplet number concentration would have a significant impact on the EMPM calculations, and by extension the retrieved entrainment rate. An attempt was therefore made to retrieve that value alongside the entrainment rate. The a priori and variance values for droplet number concentration were taken from Cloud Airborne Spectrometer (CAS) data from the Cumulus Humilis Aerosol Processing Study (CHAPS, Berg et al. 2009), where instrumented aircraft penetrated cumulus clouds in central Oklahoma. The mean of the cloud peak concentrations was 523 cm^{-1} and the standard deviation was 300 cm^{-1}, resulting in a variance of 9×10^4. It was assumed that the two variables being retrieved are independent, therefore the off-diagonal elements of S_a were set to zero. Information content analysis, which was determined from the degrees of freedom of signal (a direct by-product of the Gauss-Newton method), showed that the forward model u was very little information on the cloud droplet number concentration in the observations, implying that at least for the forward model used in this particular realization. Because of this, the assumed prior variance of the cloud droplet number concentration has been set to 10^{-4} so that the first guess value, obtained from surface-based measurements of cloud condensation nuclei-sized aerosols, remains unmodified throughout the retrieval.

4. The Explicit Mixing Parcel Model

As described in Chapter 2, numerous approaches have been developed to model the mixing of environmental air with cloudy air parcels. In light of the evolution of the conceptual model for cumulus entrainment, a new generation of cloud parcel models has been developed to properly realize the finite-rate mixing postulated in the literature. Among these is the Explicit Mixing Parcel Model (EMPM), which serves as the forward model in this work's retrieval algorithm.

A parcel model tracks a single parcel beginning with its ascent from cloud base. The characteristics of the parcel at cloud base are specified, including temperature, pressure, droplet number concentration, and vertical velocity. Due to the lack of dynamics in such a model, surface conditions are not required to initialize it. At each time step, the parcel's altitude is increased according to the specified vertical velocity and the characteristics of the parcel are recalculated. This continues until a specified cut-off condition is met; in the case of the present work, this is once the parcel has entrained enough environmental air that its buoyancy has become negative.

The genesis of EMPM is found in a study of the role that cloud-top entrainment has in stratus clouds (Krueger 1993). A model that could separately treat turbulent mixing on a wide range of scales while also accounting for molecular diffusion was necessary to properly represent the interface between the stratus clouds and the plumes of environmental air that penetrated their tops. Krueger notes that the small-scale processes are key for this environment: molecular diffusion is insignificant until there is a sufficient

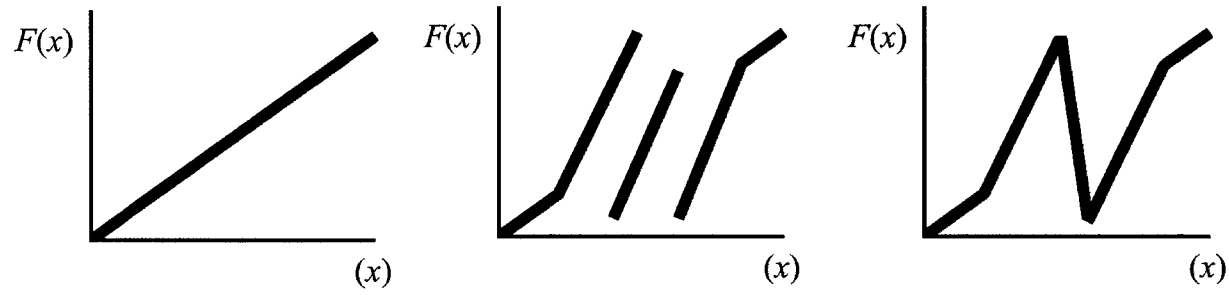

Figure 4.1: Diagram of the triplet mapping procedure. A scalar gradient (left) is compressed to three times its initial value. Three copies of this gradient are inserted into the space occupied by the original gradient (center), while the middle one is reversed to maintain a continuous function (right).

scalar gradient created by turbulent deformation while evaporation only occurs after cloud droplets and subsaturated air are brought into contact through the mechanism of molecular diffusion. While large-eddy simulation (LES) models had been used to investigate the structure of stratus, they typically underestimated the buoyancy of parcels by basing condensation on volume-averaged values for the relevant quantities. This approach was unable to account for the impact of the smaller scales on the cloud's evolution. Instead, a linear-eddy model (LEM, Kerstein 1989) approach was chosen.

In a LEM, a fluid parcel is reduced to a single dimension. The single dimension allows for a very high spatial resolution, which means that all length scales pertinent to the problem can be explicitly calculated. This allows for the representation of the macroscale impact of turbulence, while the domain has been simplified enough for the microscale diffusion calculations to be calculated at all scales of flow. The LEM approach allows for much greater computational efficiency than would be possible with a full three-dimensional treatment. In this particular LEM, the turbulent rearrangement events are mapped to the

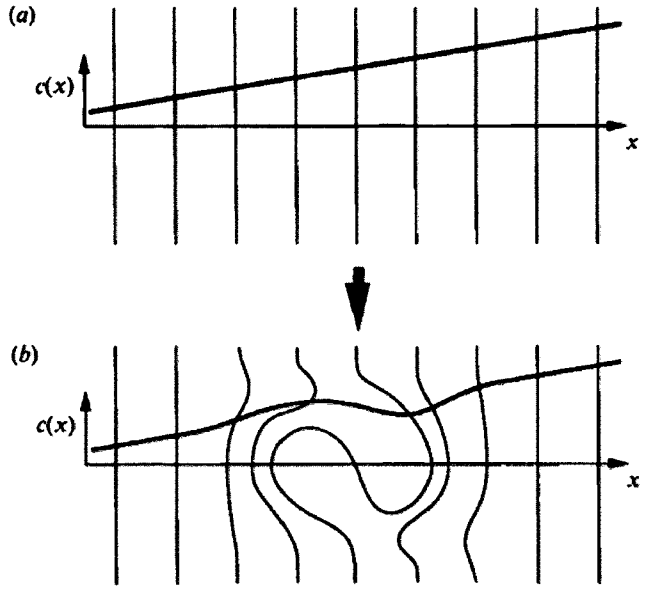

Figure 4.2: Illustration of the impact of an eddy on a scalar field with a uniform gradient. Part (a) denotes the state of the field before the eddy is applied. Part (b) shows the field after the eddy has impacted the gradient. Note the similarity of the new gradient to the triplet map illustrated in Figure 4.1. Image taken from Kerstein (1991).

pre-arrangement field through the use of the triplet mapping technique. With triplet mapping, the scalar field in a randomly selected portion of the domain is compressed to three times the initial gradient, which represents compressive strain. Three copies of the now-tightened scalar field are placed in the original segment, and the middle one is reversed so that the field remains continuous (Figure 4.1). This results in a field where scalar gradients and isopleth area exponentially increase with time, consistent with fluid theory (Batchelor 1952). Kerstein (1991) also notes a qualitative explanation for the use of triplet mapping. Given a two-dimensional scalar field with a uniform gradient, an eddy in that field will deform the scalar cross-section with respect to the horizontal dimension in a manner that resembles the triplet mapping procedure (Figure 4.2). The molecular diffusion equation,

$$\frac{\partial \phi}{\partial z} = D_M \frac{\partial^2 \phi}{dz^2} \tag{4.1}$$

where ϕ represents a scalar and D_M represents the molecular diffusivity, is numerically solved within the LEM in order to calculate the impact of diffusion on the various scalar fields.

Krueger et al (1997) extended the initial Krueger (1993) stratus cloud model to study entrainment in cumulus clouds. The primary change was to allow multiple entrainment events, and the new model was named the Explicit Mixing Parcel Model. EMPM differs from other models through the implementation of discrete entrainment events that bring air into the domain and finite-rate mixing in the LEM that mixes that air inhomogeneously throughout the parcel. This initial version of the EMPM accounted for cloud droplet microphysics through bulk schemes.

The implementation of discrete entrainment events, as opposed to the continuous stream used in early plume models, is the primary reason that EMPM was chosen to be the forward model for ERICA. Each entrainment event consists of inserting a segment of environmental air into the linear domain of the modeled parcel. While a real-world parcel grows in size as it ascends as an effect of entraining air, EMPM instead replaces a randomly selected sector of the parcel with the entrained air. This insures computational simplicity, and since the parcel is assumed to be statistically homogeneous this replacement process does not affect the statistics of the parcel as a whole. The fractional entrainment rate λ

governs the interval between entrainment events and is one of the user-definable parameters of the model. The mean height between entrainment events Δz_{avg} is defined as:

$$\Delta z_{avg} = \frac{1}{\lambda} \frac{d}{(D-d)} \tag{4.2}$$

where D is the length of the one-dimensional parcel domain and d is the characteristic size of the entrained blob of environmental air. There is no known distribution of entrained blob sizes in cumuli, but it is assumed that the blobs are distributed about a characteristic size that is determined by the dynamics of the cloud and its environment. To find a characteristic value, Krueger et al. (1997) modeled cumuli that had undergone aircraft observation penetrations and systematically varied the value of d until the modeled quantities measured the observed ones. It was determined that the best ratio of blob size to domain size was 0.1, which in turn is the value used in the present work. The random nature of entrainment is represented in the EMPM by choosing Δz from a Poisson distribution with a mean value of Δz_{avg}. Multiple runs of the EMPM can be averaged together to obtain a mean value for the various vertically integrated products required for the algorithm (i.e., effective radius and LWP); obviously, the more runs that are calculated, the more that the data will converge for the specified entrainment rate.

The finite-rate mixing is implemented through the aforementioned triplet map, thus rearranging and redistributing the newly entrained blob throughout the parcel. In the case of the EMPM, a segment size for the triplet map is randomly chosen from a distribution of segment sizes limited to a minimum of the Kolmogorov scale (the scale at which turbulence is identical for all flows) and a maximum of the integral length scale. Kerstein (1991) found

that the rate of turbulent rearrangement events was a function of both of these scales in a high Reynolds number (Re) flow. In addition, in flows with large Re, the linear eddy model results are statistically independent of Re (Krueger et al. 1997), making it possible to obtain accurate results without resolving down to the Kolmogorov scale.

Su et al. (1998) extended EMPM to explicitly calculate droplet microphysics in order to obtain cloud drop size distributions. The present work uses this extended version of the EMPM as the distributions are necessary to calculate cloud droplet effective radius and liquid water path. Therefore, this version is described here in greater detail. The radius of each individual cloud droplet is numerically calculated from the diffusional growth equation:

$$r\frac{dr}{dt} = \frac{S - A_1 + A_2}{A_3 + A_4} \tag{4.3}$$

where r is the droplet radius, S is the supersaturation ratio, and A1 through A4 are the curvature effect, solution effect, heat conduction, and vapor diffusion terms respectively. Each parcel is subdivided into grid cells that are 1.67 mm across, making it possible for these terms to be calculated using the conditions immediately surrounding each droplet which may be slightly different from one cell to the next due to the stochastic mixing.

The additional calculations introduced by the microphysical calculations come at a price. In order to run the model on existing infrastructure in a reasonable amount of time, it is necessary to shrink the parcel size by an order of magnitude; whereas the Krueger et al. (1997) edition used parcels hundreds of meters across, parcels in the extended EMPM used in this study are 10 m. Su et al. (1998) conducted a sensitivity test to determine the

impact of a reduced domain size on the drop size distributions. In comparing a 20 m parcel entraining a 2 m blob to a 100 m parcel entraining a 10 m blob, similar results were found. This insensitivity to domain size is a result of the lack of impact by entrainment on droplet growth or evaporation until scalar lengths have been reduced to the order of the diffusional scale. As a large blob needs to be broken down to a smaller one before evaporation has a significant impact, starting with a smaller domain merely means that the experiment is starting at a different point in the process.

The EMPM was initially validated against in situ observations of a Hawaiian trade cumulus cloud taken at several different altitudes during the Joint Hawaii Warm Rain Project (Raga et al. 1990). The measured and EMPM-modeled droplet spectral width and spectral evolution with height showed much better agreement than the results from either an adiabatic parcel model or an instant mixing model (Su et al 1998). Further analysis showed little variation in modeled spectra after changing the turbulent dissipation rate, but significant differences between observed spectra when compared to diffusion only and instant mixing model runs, implying that the finite-rate mixing model is an appropriate tool for the investigation of cumulus parcels.

Some modifications to EMPM were required for its use in ERICA. Continental cumulus have a greater cloud-droplet number concentration than the marine stratus and cumulus that EMPM was initially designed to study (Miles et al. 2000), so the parcel domain used in the present study had to be further reduced to 10 m from the 20 m used in the Su et al. (1998) sensitivity tests. The grid size is chosen so that the smallest turbulently rearranged segment is equal to the Kolmorgorov length scale. Due to the triplet map, six

grid cells are required for the finest level of rearrangement, necessitating a grid spacing of one-sixth of the Kolmorgorov scale. For the model realizations used in ERICA, the Kolmorgorov scale was set to 0.01 m, resulting in a grid spacing of 1.667 m^{-3}. The temporal resolution of the model is 0.75 s. Furthermore, EMPM was not designed to be run iteratively; it was necessary to include hooks in the source code so that it could be operated externally by a driver.

5. Observations and Instrumentation

a. Site Selection

For this proposed method to be carried out, a suitable location must be selected. Three criteria must be fulfilled for a site to be considered for this project's development. First, cumulus clouds must regularly be found over the site, insuring that a suitable range of cases can be observed. The site must also be well-instrumented, housing a variety of instruments capable of producing the necessary assortment of observations to characterize the cloudy and cloud-free air. Lastly, the site must also host occasional aircraft-borne observation experiments so that the product can be validated. The Southern Great Plains (SGP) site of the Atmospheric Radiation Measurement (ARM, Stokes and Schwartz 1994) Climate Research Facility is a natural choice for this role as it readily satisfies all of the necessary conditions.

ARM, a facility of the United States Department of Energy, operates a network of permanent and mobile ground-based remote sensing sites around the world in order to host instruments capable of providing a long-term record of radiative balance and cloud characteristics to assess changes that may impact the climate. Three permanent sites have been developed, each representing a separate climactic regime: the North Slope of Alaska (NSA) site, representing polar regions; the Tropical Western Pacific (TWP) site, representing the tropics; and the SGP site, representing the mid-latitudes. Of these, the SGP is the oldest and most-developed site in the network, having been established in 1992; its domain includes clusters of instruments distributed over 140000 km^2 in northern

Oklahoma and southern Kansas. At the heart of the SGP site is the Central Facility (CF), located southeast of Lamont, Oklahoma, United States (Figure 5.1). This 0.65 km^2 facility hosts dozens of remote sensing instruments capable of vertically probing the atmosphere; many of these instruments have been operating for well over a decade and have a near-continuous observation record during that time.

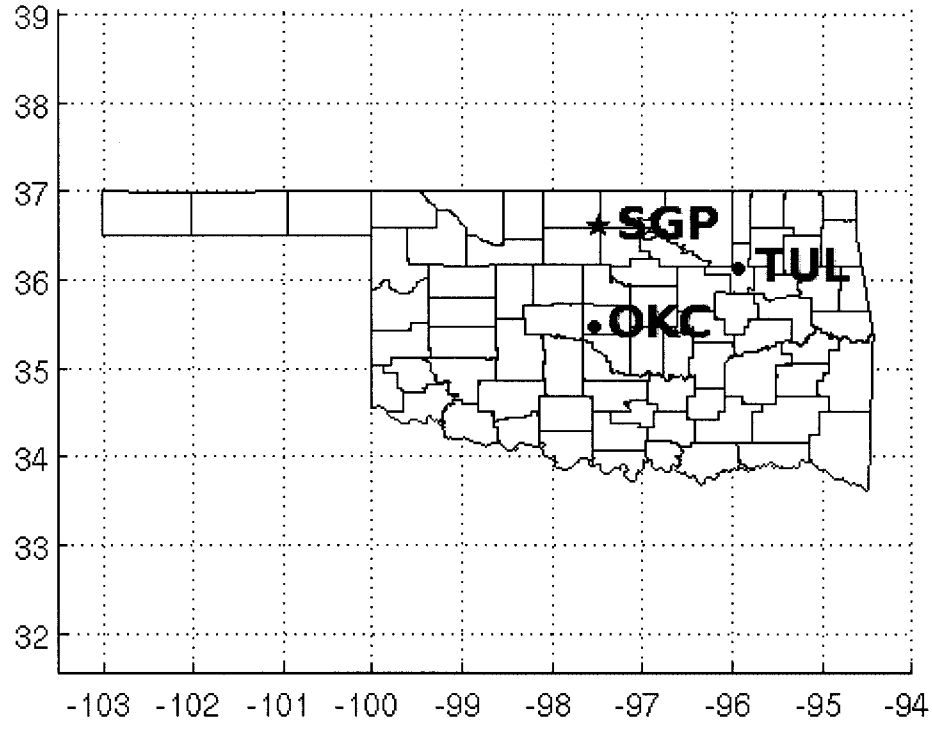

Figure 5.1: Map of the state of Oklahoma, noting the position of the Southern Great Plains site (SGP) relative to the two largest cities in the state, Oklahoma City (OKC) and Tulsa (TUL). The distance from SGP to OKC is 125 km; from SGP to TUL, it is 148 km.

Cumulus events over SGP are common enough to provide numerous opportunities for analysis. Lazarus et al. (2000) analyzed a ten-year dataset of human-made cloud type observations from SGP to determine the frequency of occurrence for various cloud types for the site as well as their diurnal cycle. The observations in this dataset were taken four times per day, at 0h, 6h, 12h, and 18h local standard time. During the summer, cumuli

were present during 10% of all local noon observations (including clear sky observations). This was the most common low cloud type during summer noon observations, besting stratus (1% of observations) and stratocumulus (5% of observations). The observed cumulus diurnal cycle was consistent with expectations for a buoyantly-forced cloud type, with the maximum occurring at local noon and the minimum at local midnight. This observation strategy may slightly underestimate the frequency of cumulus clouds, as the maximum radiative forcing will occur between the hours of 12h and 18h local time.

As previously described, the algorithm requires several different observations, including thermodynamic profiles, vertical wind speed, and cloud characteristics. All of these observations can be obtained from instruments installed at SGP. Table 5-1 introduces the key instruments used in this study, and the remainder of Chapter 5 is devoted to presenting these instruments in greater detail.

Table 5-1: ARM SGP instruments used in ERICA

Name	Measurement Used	Temporal Resolution	Use in ERICA
Atmospheric Emitted Radiance Interferometer (AERI)	Temperature profile	30 s	EMPM input
Raman Lidar (RL)	Water vapor mixing ratio profile	10 s	EMPM input
Millimeter-wave Cloud Radar (MMCR)	Vertical velocity profile, cloud effective radius	4 s	EMPM input, Output observation
Aerosol Observing System (AOS)	Surface cloud condensation nuclei concentration	1 hr	EMPM input
Microwave Radiometer (MWR)	Cloud liquid water path	20 s	Output comparison

b. Atmospheric Emitted Radiance Interferometer (AERI)

The Atmospheric Emitted Radiance Interferometer (AERI, Knuteson et al. 2004a, 2004b) is a self-calibrating ground-based interferometer that measures downwelling infrared emission from the atmosphere over the interval of 550-3300 cm^{-1} (3-18 μm) with a spectral resolution better than one wavenumber. Each spectral observation is calibrated against two blackbodies that are directly traceable to National Institute of Standards and Technology (NIST) calibrations, thus ensuring that the observation has a radiometric accuracy of better than 1% of the ambient radiance and a reproducibility of better than 0.2%. AERI is automated, field-hardened, and has successfully run unattended in numerous of harsh environments including the various ARM sites.

The fundamental design for AERI consists of splitting the incoming radiation beam onto two moving mirrors and then recombining it. As the path length between AERI's two moving mirrors changes, the recombined radiation exhibits destructive or constructive interference. A Fast Fourier Transform (FFT) can be used on the interferogram produced in this manner to reconstruct the original spectrum. A rotating scan mirror is used to change the radiation source; for each observation, AERI cycles between the upward sky view, a hot blackbody set at 333 K, and an additional blackbody at the ambient temperature. This hot/ambient approach to calibration insures a higher degree of accuracy for the near-surface radiation when compared to a traditional hot/cold calibration as the system is now calibrated about the environmental temperature. AERI radiances were originally obtained through a seven-minute observation cycle: three minutes on the sky view and two minutes on each of the blackbodies. This allowed ample

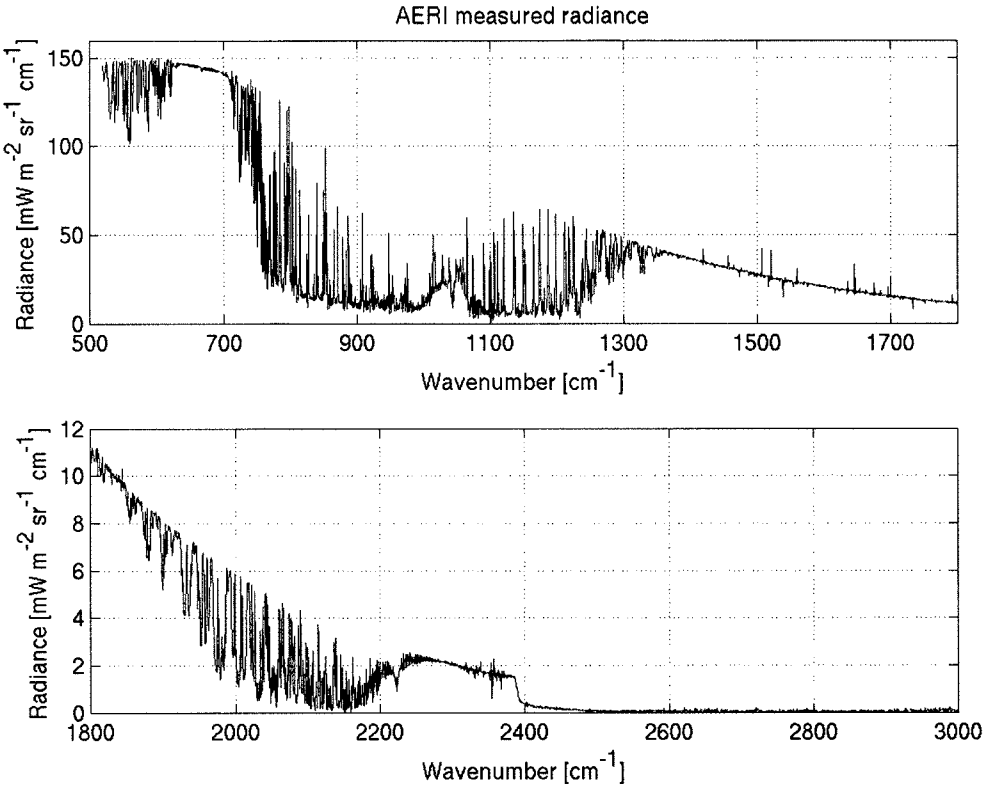

Figure 5.2: Typical spectrum recorded by the Atmospheric Emitted Radiance Interferometer (AERI). Because AERI is an upward pointing radiometer, spectral signatures are inverted when compared to downward pointing satellite instruments. For AERI, window channels show low radiance while absorption lines show high radiance values.

time produce noise-reduced spectra via averaging. With the introduction of a noise-filtering technique using principle component analysis (Turner et al. 2006), the temporal resolution has been increased to 30 seconds. An example of an AERI-observed spectrum is shown in Figure 5.2.

Through the use of a combined statistical and physical retrieval (Smith et al. 1999, Feltz et al. 1998, 2005), vertical profiles of temperature and moisture can be obtained from the observed radiance spectra. A first guess of temperature and moisture profiles for the "AERIPROF" retrieval algorithm is formed by blending profiles created from Rapid Update

Cycle 2 (RUC-2) numerical weather prediction model (Benjamin et al., 1998) output with statistical regression profiles created from hundreds of clear-sky rawinsonde launches at SGP; the more finely-resolved data from the regression are used in the lower troposphere where atmospheric structure itself is more detailed than can be captured by the model. A radiative transfer model is then used to calculate the spectrum from the first guess, and this output is compared to the observed spectrum. The profile guess is adjusted iteratively until the modeled spectrum converges on the observed one within a specified tolerance. Profiles can only be retrieved during conditions when the sky is clear overhead as the infrared emittance from clouds and precipitation impact the observed infrared spectra; a collocated laser ceilometer is used to verify cloud conditions for each observation. This limitation does not impact the current study as the AERI observation period is fine enough to enable observations of cloud-free air between typically-spaced cumulus clouds. This enables the measurement of the temperature of the environmental air being entrained by these clouds. Retrieved profiles show good agreement with coincident radiosonde launches (Feltz et al. 2003), and time series of stability measurements calculated from AERI-observed profiles show promise for severe storm forecasting (Wagner et al. 2008).

c. Raman Lidar

Atmospheric molecules typically scatter radiation at the same wavelength as the incident radiation. A small fraction of the photons, however, are scattered at a different frequency based upon the vibrational and rotational oscillation of the scattering molecules as well as the frequency of the incident source. This phenomenon, named Raman scattering after its discoverer C. V. Raman, can be used to identify the molecules causing the

scattering by illuminating the scatterers with a known wavelength and then measuring the weak wavelength-shifted return. Raman scattering can be used to measure the vertical profile of water vapor mixing ratio, defined as the ratio of the mass of water vapor to the mass of dry air. Since diatomic nitrogen gas (N_2) is well-mixed throughout the troposphere, the ratio of water vapor mass to N_2 mass is proportional to the water vapor mixing ratio. By using a lidar to measure backscattered radiation at the incident wavelength as well as the ratio of the backscattered radiation at the Raman wavelengths for N_2 and water vapor, the water vapor mixing ratio can be retrieved (Turner and Whiteman 2002).

Since 1998, the ARM SGP Raman lidar has operated continuously and automatically throughout the diurnal cycle, including greater than 90% availability from September 2004 to the present day. The lidar emits a beam at 355 nm at 30 Hz; the Raman scattering wavelength for N_2 at this wavelength is 387 nm while the Raman scattering for water vapor is at 408 nm. In order to reduce noise the system averages pulses to form a 10 s base temporal resolution, while the maximum vertical resolution is 7.5 m. Additional averaging in post-processing can be done in both the spatial and temporal dimensions to further enhance the signal-to-noise ratio. We have elected to decrease the vertical resolution to 75 m while maintaining the 10 s temporal resolution in order to insure observations of the tropospheric air between cumuli. Individual moist plumes leading to cumulus clouds can be resolved by the SGP Raman lidar.

The detection channels of the ARM SGP lidar system, especially the 408 nm water vapor channel, are sensitive to solar radiation. During nighttime this poses no problems,

but during the day the magnitude of the solar radiation, which is a noise source, can be much larger than the Raman scattering signal. To reduce the influence of the solar background, both narrow and wide fields of view (FOV) and narrowband filters are used to insure that only radiation at the desired wavelengths and the proper direction is admitted to the receiver. The use of the dual FOVs results in daytime profiles reaching a maximum altitude of 5 km with less than ~20% uncertainty (Ferrare et al. 2006). To account for the dual FOV, an overlap correction must be applied. Additionally, uncertainty in the Raman cross section of the constituent molecules (Penney and Lapp 1976) makes it impossible to absolutely calibrate the system. Instead, it calibrated against a collocated microwave radiometer (Turner and Goldsmith 1999). The typical accuracy of the Raman lidar water vapor mixing ratio profile is better than 5% (Turner and Goldsmith 1999).

d. Millimeter wavelength Cloud Radar

The Millimeter-wave Cloud Radar (MMCR, Moran et al. 1998) is a ground-based upward-pointing short-wavelength (8.6 mm) Doppler radar designed to observe cloud dynamics and structure. In operation at the ARM SGP site since 1996, it provides observations of power and vertical velocity that can be used to remotely obtain information about cloud droplet sizes and vertical motion within the cloud respectively. In addition, the spectral width of the velocity is also recorded which can be used to determine information about the distribution of velocities within the sample volume.

To detect cloud droplets, a cloud radar must use a much shorter wavelength than a radar designed for precipitation observations due to the dependence of scattering on both

droplet radius and the incident wavelength. A spherical hydrometeor scatters incident radiation from a radar with an efficiency proportional to the size parameter x:

$$x = \frac{2\pi r}{\lambda} \tag{5.1}$$

where r is the radius of the hydrometeor and λ is the wavelength of the incident radiation. The result of this relationship is that radars that emit and receive shorter wavelengths of energy will show more sensitivity to smaller droplets than radars that utilize longer wavelengths due to increased scattering of the radar's energy by the droplets. MMCR uses a wavelength of 8.6 mm, roughly an order of magnitude shorter than the 10 cm wavelength of the National Weather Service Next-generation Radar (NEXRAD) systems used to track rain and snow. The wavelength translates into a frequency of 35 GHz; as this is in the middle of a window channel there is little interference by atmospheric emission or beam absorption. While a drawback of such a short wavelength is that moderate precipitation greatly attenuates the signal, this is not considered to be a liability in this study as the investigated cumuli are all non-precipitating. The peak transmitted power is 100 W.

MMCR cycles through several operational modes (Clothiaux et al. 1999). These operational modes differ from one another by changing the characteristics of the radar pulses in order to boost sensitivity to certain cloud types. The operational modes currently in use at the SGP site are the so-called "Boundary Layer" (1), "Cirrus" (2), "General" (3), "Precipitation" (4), and "Dual-polarization" (5) modes. The mode sequence used at the SGP site during the time of this study is:

1 2 1 3 1 4 1 3 1 2 1 3 1 5 1 3.

Each mode requires roughly two seconds for observations and processing; the entire sequence repeats itself in 32.5 seconds (Widener and Johnson 2005). This observation strategy results in frequent observations in the boundary layer mode, which has the best sensitivity to clouds in the lower atmosphere. Note that the boundary layer mode (1) is executed every other time in the sequence; this sequence was designed to capture the rapidly changing cloud conditions in the boundary layer.

In addition to the standard MMCR data stream, ARM provides a post-processed value-added product (VAP) named Active Remote Sensing of CLouds (ARSCL, Clothiaux et al. 2000). ARSCL combines MMCR with data from a micropulse lidar and a laser ceilometer to produce an accurate assessment of hydrometeor height and vertical motion by combining the strengths of each dataset. The laser instruments are generally insensitive to insects and virga, both of which provide significant returns in the MMCR dataset, and thus provide a much more accurate estimation of cloud base height, but they are easily attenuated by the cloud itself or by fog or light precipitation. MMCR is capable of penetrating clouds and measuring multiple cloud layers but isn't as accurate as the ceilometer at determining cloud base for clouds that are thin, high, composed of small reflectors, or are precipitating. The merged product created by the ARSCL algorithm includes identification of the cloudy pixels and their associated vertical velocity and droplet reflectivity. Since EMPM requires a specified vertical velocity, ARSCL provides an estimate of the vertical velocity within a cloud. Multiple data streams are produced, including a stream for each mode and a best estimate obtained by combining the various MMCR

modes. For this study, only ARSCL-processed observations from the Boundary Layer mode were used. This is due to this mode's sensitivity to cumulus cloud droplets and, with half of all observations, its frequency of observations as compared to the other modes. The characteristics of the Boundary Layer mode compared to the General mode are shown in Table 5-2.

Table 5-2: Boundary Layer and General Operational Mode for SGP MMCR

Parameter	Boundary Layer Mode	General Mode
Inter-pulse period	68 μs	106 μs
Pulse width	0.3 μs	0.6 μs
Gate spacing	0.3 μs	0.6 μs
Vertical resolution	45 m	90 m
Number of gates	110	167
Coherent averages	6	4
Spectral averages	10	12
FFT length	256	128
Dwell time	1.04 s	0.774 s
Observation and processing time	2.0 s	2.1 s
Estimated sensitivity at 5 km	-36 dBZ	-53 dBZ

The retrieval algorithm is attempting to adjust the entrainment rate in order to get agreement in cloud droplet effective radius, which can be determined from cloud radar reflectivity data. This is done through an empirical relationship between radar reflectivity

and effective radius. Fox and Illingworth (1997) calculated effective radius and radar reflectivity from observations of drop size distribution inside stratocumulus obtained from the Forward-Scattering Spectrometer Probe (FSSP). They found that there is a linear relationship between reflectivity in dBZ space and effective radius in log-10 space:

$$Z\,(\mathrm{dBZ}) = \alpha \log_{10}(r_e) + \beta. \tag{5.2}$$

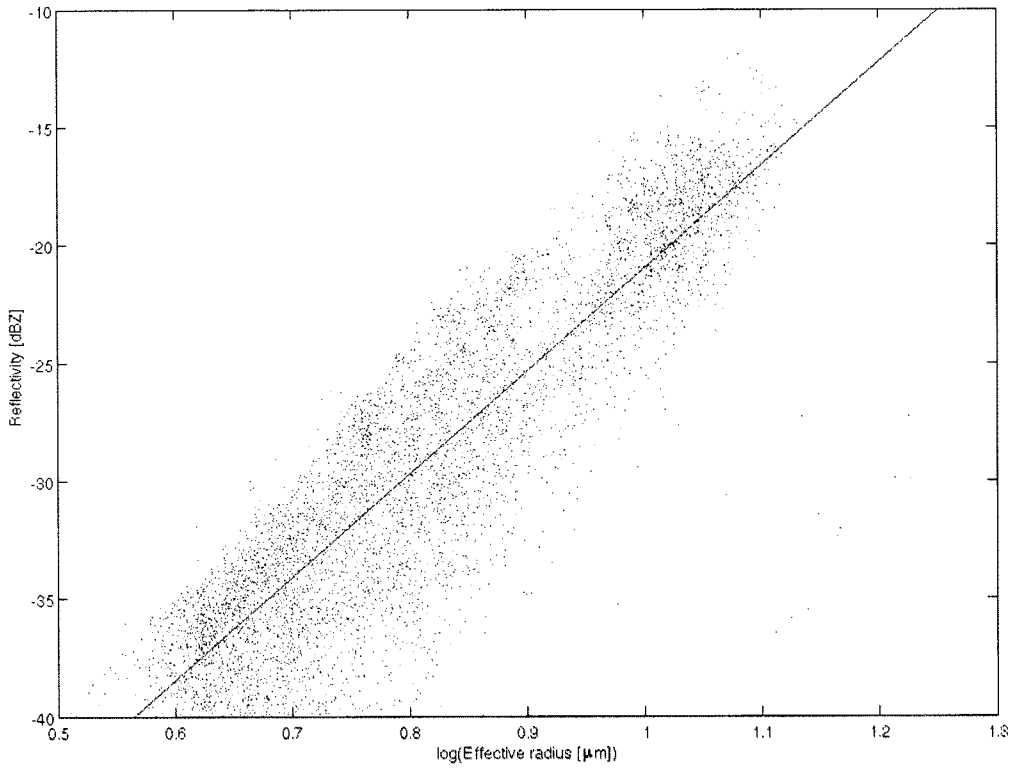

Figure 5.3: Scatter plot of the logarithm of effective radius (in microns) observed by the Cloud Airborne Spectrometer (CAS) in CHAPS against the radar reflectivity (in dBZ) calculated from that volume. The black line represents the linear fit of the data as described in the text.

For their dataset, they found the coefficients for α and β to be 40.9 and -64.2 respectively. The absolute error was found to be 1 μm, and the correlation was 0.83.

This technique was repeated for the current study using Cloud Airborne Spectrometer (CAS) data obtained during the Cumulus Humilis Aerosol Processing Study (CHAPS, Berg et al. 2009). During CHAPS, instrumented aircraft conducted research flights through cumulus fields over central Oklahoma in order to determine the impact of an urban aerosol plume on cumulus characteristics. Since CHAPS investigated cumuli in the same general area as the current investigation, it made for a natural source of aircraft data that was independent of any used for this work. By fitting a linear regression to the points with a calculated reflectivity greater than -40 dBZ, the coefficients for α and β were found to be 43.7 and -64.7 respectively with a correlation of greater than 0.89. These coefficients compare favorably to those found by Fox and Illingworth (1997). Figure 5.3 shows this retrieval for the CHAPS dataset. The mean difference between the actual and reflectivity-obtained effective radius is 0.087 μm with a standard deviation of 1.13 μm.

e. Aerosol Observing System

The droplet number concentration at cloud base is another parameter that EMPM requires in order to complete its calculations. Due to the well-mixed nature of the boundary layer on days in which cumuli are present, a surface measurement of cloud condensation nuclei (CCN) is a good approximation of the cloud droplet number concentration aloft.

Surface CCN at the ARM SGP site are measured by the Aerosol Observing System. This system contains several instruments that share a common intake, including nephelometers and a photometer. Air is ingested into the system through a stack opening located 10 m above ground level. If the air has a relative humidity of greater than 40%, it is

heated to remove enough moisture to lower it to that level. The air is then split via a manifold that directs the flow to the various instruments. Of chief interest for this work is the condensation nucleus counter (CNC), which measures the total concentration of particles with diameters between 0.01 μm and 3 μm. Air is drawn through the CNC at a rate of 30 l min^{-1}. Alcohol vapor is condensed onto the particles in the air stream, and the resulting droplets are counted by an optical particle detector (in this case, the TSI Model 3010, a commercially available small particle counter) operating at 780 nm. Observations are averaged together into one hour bins. Sheridan et al. (2001) conducted an analysis four years of CNC data from the SGP AOS. They found a minimum in condensation nuclei in the pre-dawn hours and a maximum in the middle and later afternoon, consistent with the diurnal cycle of turbulent mixing. Annual maxima are found in March and October.

f. Microwave radiometer

Microwave observations are commonly used for investigations of cloud liquid water path due to the long wavelengths as compared to the size of the cloud droplets; this results in negligible scattering of the microwave signal. The ARM SGP microwave radiometer (MWR) senses ambient radiation at 23.8 GHz (1.26 cm) and 31.4 (0.96 cm), frequencies dominated by water vapor absorption and cloud liquid water absorption respectively. Sky view observations are taken every 20 seconds. The field of view is frequency dependent, going from 5.9° at the lower channel to 4.5° at the higher one.

The ARM program performs a simple statistical retrieval of liquid water and water vapor from the MWR as part of its ordinary operation using site-specific relations between

the observations and collocated radiosondes. This approach has several drawbacks, however. Biases are present any time the observed conditions differ substantially from the mean. Furthermore, no information about the temperature of the cloud is ingested into the statistical retrieval. Some attempts were made to introduce fixes for these issues in a new statistical retrieval, but even greater accuracy was possible if a physical-iterative approach was adopted. This led to the development of the Microwave Radiometer Retrieval (MWRRET, Turner et al. 2007), an iterative optimal estimation algorithm that uses the two channels from the microwave radiometer to retrieve liquid water path and precipitable water vapor. Like the optimal estimation algorithm described in the present work, a forward model is key to mapping the atmospheric state (in this case, precipitable water vapor (PWV), cloud liquid water path (LWP), and thermodynamic profiles) to observations (the radiance observations at the two channels observed by the microwave radiometer). For its forward model, MWRRET uses the Monochromatic Radiative Transfer Model (MonoRTM, Clough et al. 2005), an extension of the Line-By-Line Radiative Transfer Model (LBLRTM, Clough et al. 1992). MonoRTM output is compared to the observations of brightness temperature at the two microwave frequencies, and iteration is undergone until convergence is achieved. As long as the thermodynamic structure of the atmosphere is known, either from radiosondes or from collocated profilers, the physical retrieval method is more accurate than the statistical method.

ERICA was initially designed with the ability to ingest both LWP and PWV observations into its retrieval. Information content analysis showed that PWV was not very sensitive to either entrainment rate or droplet number concentration. Furthermore,

accurate calculations of total column PWV from EMPM were difficult due to the lack of information above the cloud top. Therefore, it was decided to drop PWV entirely from the retrieval process.

6. Algorithm Implementation

a. Quantitative measures of uncertainty and information content

One of the advantages of using Gauss-Newton optimal estimation for an inversion problem is the ability to propagate the magnitude of the error of each of its constituent elements throughout the retrieval so that the final product has a known error. There are multiple sources of the error that must be accounted for. These include errors in the parameters used in the forward model, errors in the observations, and errors and approximations in the forward model itself, as well as the sensitivity of the model output to changes in the input parameters (Rodgers 2000).

The total error in an optimal estimation retrieval is represented by the covariance of the conditional probability of the state given the measurements (represented as $P(x|y)$ in Equation 3.5) and is a function of the Jacobian and the error and a priori covariance matrices:

$$S = (K^T S_\varepsilon^{-1} K + S_a^{-1})^{-1} . \tag{6.1}$$

The diagonal elements of this matrix represent the variance on the retrieved variables; the standard deviation about the mean for each retrieved term is easily obtained by taking the square roots of each of the diagonal elements and is used as the source of the error bars shown throughout the remainder of this work.

Special attention must be paid to the observation error covariance matrix term, S_ε. This term contains the covariance of the errors for two classes of observations: the measurement vector, containing the observations used to evaluate the forward model

output; and the input vector, which contains the observations that are used to initialize the forward model but are not retrieved as part of the inversion (these are often called model parameters). In ERICA, the model parameter vector consists of the cloud base temperature, cloud base moisture, cloud vertical velocity, and the environmental temperature and moisture profiles. The chain rule approach of Turner (2005) is used to propagate the uncertainty in the model parameters into the error covariance matrix S_ε:

$$S_\varepsilon = S_y + \sum_n S_n \tag{6.2}$$

where S_y is the error covariance matrix of the measurement vector:

$$S_y = \begin{bmatrix} \sigma^2_{R_{eff}} & 0 \\ 0 & \sigma^2_{LWP} \end{bmatrix} \tag{6.3}$$

and S_n is the covariance matrix for a given member n of the model parameter vector:

$$S_n = K_n^T \sigma_n^2 K_n \,. \tag{6.4}$$

The variance in n is estimated from the observations while the Jacobian about n is calculated using a finite difference of the forward model for a perturbed and unperturbed n. It is assumed that all errors are uncorrelated, so the off-diagonal elements of S_y are set to zero. Since components of S_ε are calculated using the forward model, S_ε has some non-zero off-diagonal elements. The variance values used for each element of S_n are taken directly from the characterized error for each instrument and are kept constant for all iterations and realizations.

An additional benefit of using the Gauss-Newton retrieval framework is a quantitative measure of the information content of the retrieval. One such measure of this is the degrees of freedoms of the signal (DFS), which tells how many independent pieces of information are in the measurements and thus how many variables can be retrieved (Rodgers 2000). DFS for a Gauss-Newton system is calculated as the trace of the averaging kernal matrix:

$$\hat{S} = (K^T S_\varepsilon^{-1} K + S_a^{-1})^{-1} \tag{6.1}$$

which, since ERICA is retrieving a single variable, will have a maximum possible value of one.

b. Stochastic sensitivity

As described previously, EMPM represents the turbulent nature of cumulus clouds with stochastic events: entrainment events occur at random intervals determined from a Poisson distribution centered on the specified entrainment rate, and parcels are rearranged randomly to in order to mix them through the triplet map mechanism described earlier. A concern, however, is the impact of the random numbers on the retrieved entrainment rate. Ideally a sufficiently large number of runs would be averaged together to find a mean characteristic entrainment rate. However, the computational expense is sufficiently large that a more efficient method to characterize the impact of the stochastic nature of the model was needed.

To determine what constitutes a sufficient number of runs, two batteries of tests were executed for a particular retrieval date and time. In the first battery, hereafter referred to as *Full Random*, randomness was utilized for both the entrainment intervals and the turbulent mixing. For the second battery, named *Mix Only*, the entrainment interval was fixed at the specified entrainment rate (as opposed to being selected from a distribution) while the parcels were allowed to stochastically mix as before. Both arrays of tests consisted of 17 runs, and they were initialized with the same series of random number generator seeds in order to facilitate direct comparisons between retrievals. The results for *Full Random* and *Mix Only* are shown in Figures 6.1 and 6.2 respectively. The mean entrainment rate for all the realizations in *Full Random* was found to be 0.769 km^{-1} while for *Mix Only* it was 0.763 km^{-1}. Based on these results, aforegoing assumption is that the mean entrainment rate for a large number of random cases initialized with the same conditions is the same whether or not the entrainment rate used in EMPM comes from a weighted distribution about the initial entrainment rate or is fixed exactly at the entrainment rate. The variability between the two series of tests, however, is substantially different. It was found that there was a much larger standard deviation about the mean for the *Full Random* corpus than there was for the *Mix Only*: 0.278 versus 0.046. The implication from this testing is that it is unnecessary to implement a large number of realizations into the algorithm; the mean entrainment rate for a large number of *Full Random* cases is equivalent to the entrainment rate from any one realization of a *Mix Only* case. Furthermore, since the ERICA is attempting to retrieve the entrainment rate, allowing the EMPM to essentially randomly perturb this value differently for each iteration was

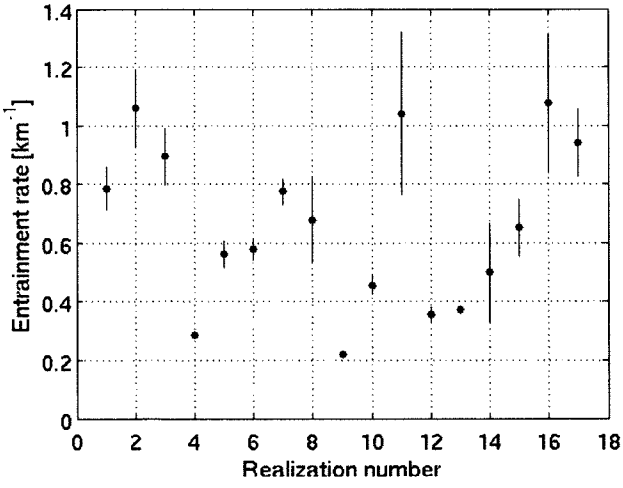

Figure 6.1: Plot of the retrieved entrainment rate for a single cloud (observed at 2142 UTC on 18 June 2009) with both the entrainment rate and the turbulent mixing in EMPM set to include stochastic modeling (described as *Full Random*) in the text. Note the significant variability in both retrieved entrainment rate and its associated error bars. The mean entrainment rate over this series is 0.769 km^{-1}.

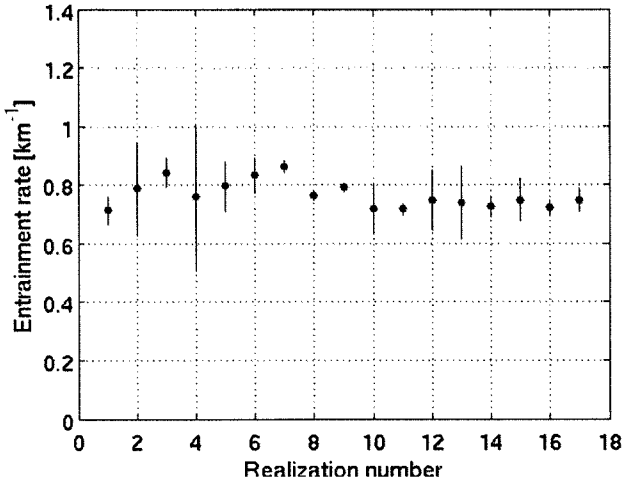

Figure 6.2: As in Figure 6.1, but for randomness only in the turbulent mixing. Note the reduced degree of variability in both the measurements and the error (described as *Mix Only*) in the text. For this series, the mean entrainment rate was 0.763 km^{-1}.

really a violation of the retrieval method and besides introducing additional noise was greatly compromising the ability of the retrieval to converge efficiently. Thus, the *Mix Only* approach was adopted, thereby greatly reducing the magnitude of computing resources required to retrieve the entrainment rate using ERICA.

c. Cloud identification and selection

Since one of the stated goals for the creation of ERICA is the development of a climatology of entrainment, an objective method for determining which times a retrieval should take place is necessary to avoid the impact of any selection bias on the resulting distribution. This set of criteria should use the instruments that are already being ingested into the algorithm so that simplicity can be maintained; with fewer instruments comprising the input stream, the likelihood that all necessary instruments will be operational for a given retrieval time increases. Cumulus cloud days were initially identified through visual inspection of total sky imager (TSI) movies. The TSI is a ground-based hemispheric imager that provides sky cover images every 30 s. Beginning and ending times for cumulus cloud events were recorded to provide a range of observation times for further investigation. Within that time span, the following objective process was undertaken.

Since the presence of liquid water is the defining characteristic of a cloud, the retrieved liquid water path from the microwave radiometer is used as the primary cloud-indicating measurement due to the reliability of its liquid water path observations and its ability to . For the purposes of this study, a cloud was determined as being overhead whenever there was a continuous series of liquid water path observations greater than 50

g m^{-2}; a series of elevated LWP observations separated by one observation less than 50 g m^{-2} is treated as two separate clouds. This minimum threshold in LWP was chosen so that the retrieved clouds would have enough vertical extent to enable multiple entrainment events in EMPM as well as to increase the probability that the core of the cloud is being sensed rather than an edge. Occasionally, clouds would be observed which maintained continuously elevated LWP for several minutes. Upon closer inspection, these clouds appeared to have the morphological characteristics of stratocumulus clouds. Therefore, a cutoff of 10 minutes for elevated LWP was initiated; no entrainment rate retrieval was attempted for clouds with a duration in excess of that value.

Identifying the cloud base is an important part of the ERICA retrieval, as EMPM begins its calculations from that level. Cloud base altitude is obtained from the Raman lidar, which determines it based on backscatter. Cloud base temperature and pressure are obtained by interpolating the Raman cloud base height to the AERI temperature profile, while cloud base water vapor mixing ratio is assumed to be the saturated value for the cloud base temperature. The vertical velocity comes from averaging the updraft velocities of pixels from times coincident with the MWR observations of cloud presence and heights greater than that of the observed cloud base. The uncertainties in all of these observations are integrated into the error measurements on the final product through the process described earlier in this chapter. Neither a minimum nor maximum restriction was placed on cloud base height.

7. Case study and validation

a. 18 June 2009

A case study is presented here to illustrate the characteristics of a cumulus field on a typical day as well as to demonstrate the ability of ERICA to retrieve the entrainment rate for individual clouds. Coincident with this investigation, a research airplane penetrated clouds near the SGP and recorded observations that can be used to retrieve the entrainment rate and provide information against which ERICA can be validated; these measurements are discussed in greater detail in Section b of this chapter.

A widespread cumulus field formed on 18 June 2009 above much of central Oklahoma and Northern Texas in the absence of any significant synoptic forcing. The primary cloud presence in the morning was a cirrus shield comprised of the remnants of an anvil that formed from a severe thunderstorm over the Texas panhandle the day before. This cirrus shield prevented the surface of the SGP site from receiving the maximum amount of insolation and thus inhibited surface diabatic heating. Persistent gulf flow insured that adequate moisture for convective clouds was present; the surface dew point was consistently above 18 °C all day. The 1200 UTC (7:00 AM CDT) sounding launched from the SGP central facility showed a small residual inversion remained from the previous night's radiative cooling; coupled with the reduced solar radiance, this was enough to inhibit convection throughout the morning.

As local noon approached, the shield's thickness and spatial extent began to erode. Along with the diurnal increase in radiative forcing, this allowed enough surface heating to

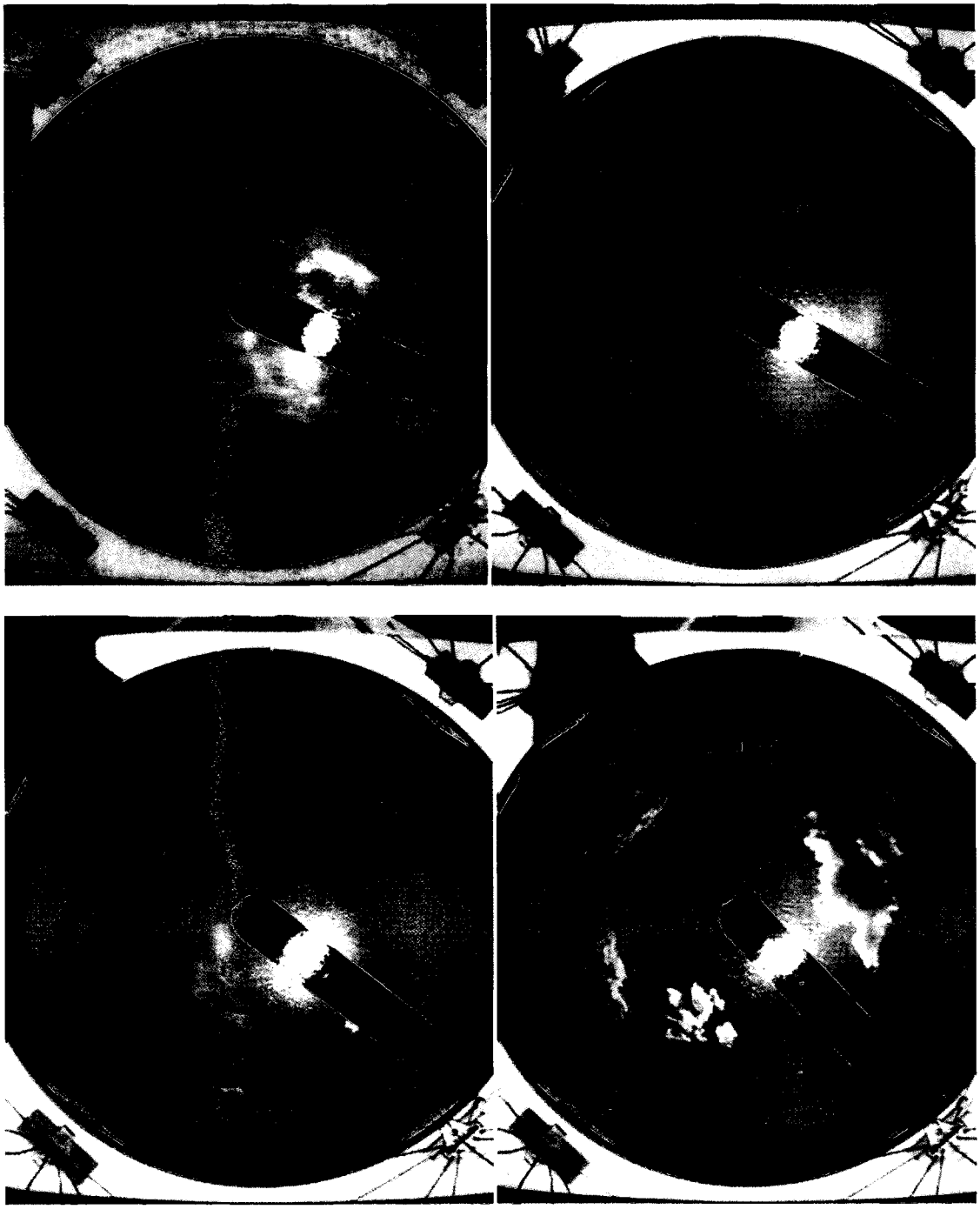

Figure 7.1: Total Sky Imager (TSI) photographs of cloud cover at the SGP site on 18 June 2009. Times represented are 1700 UTC (upper left), 1715 UTC (upper right), 1730 UTC (lower left), and 1745 UTC (lower right). These times represent local noon and fifteen minute intervals thereafter. Note the weakening of the cirrus shield coincident with the rise of cumulus convection.

enable convection. The first isolated cumulus clouds began to form at 1715 UTC (12:15 PM local time). Over the next forty-five minutes, cumuli rapidly developed over the entire sky while the cirrus deck dissipated entirely (Figure 7.1). Despite the warm surface temperature (in excess of 35 ^{o}C at local noon), deep convection did not occur. This was due to the presence of a capping inversion at the top of the boundary layer, which the 1800 UTC (1:00 PM local time) local sounding measured at 750 hPa. Copious numbers of cumulus subsequently formed above or were advected over the SGP central facility. As diurnal heating lessened throughout the afternoon, so did the cloud fraction. By 2200 UTC (5:00 PM CDT) the skies were clear.

ERICA was used to retrieve the entrainment rate for nineteen separate cumulus clouds that advected over the SGP central facility. Of these, fourteen contained sufficient information content to be classified as valid retrievals. The liquid water paths of the retrieved clouds are shown in Figure 7.2 while the entrainment rate retrievals and their associated uncertainties can be seen in Figure 7.3. The valid retrievals had a mean entrainment rate of 0.91 km^{-1} with a mean uncertainty of 0.28 km^{-1}. The standard deviation of the distribution was 0.37 km^{-1}, while it was 0.21 km^{-1} for the uncertainties.

A possible method of quality control for ERICA is to filter out retrievals where the known degrees of freedom of the signal (DFS) is less than some predetermined value. If only retrievals where the DFS is greater than 0.95 are examined, the mean and standard deviation for the entrainment rate are 0.85 km^{-1} and 0.20 km^{-1} respectively, which are values similar in magnitude as the total distribution. The uncertainties of the retrieved

entrainment rates are effectively the same in this smaller dataset: 0.26 km^{-1} for the mean and 0.20 km^{-1} for the standard deviation. It is possible to draw a conclusion

b. Validation

The Routine Atmospheric Radiation Measurement (ARM) Aerial Facility (AAF) Clouds with Low Optical Water Depths (CLOWD) Optical Radiative Observations (RACORO) campaign was an ARM field project in the first half of 2009 designed to regularly observe the internal structure and microphysics of boundary layer clouds. During the course of five months, an instrumented aircraft was regularly deployed to penetrate boundary layer clouds in a variety of conditions; approximately 240 research hours were flown during this campaign. These flights were based out of either Ponca City or Guthrie, Oklahoma, so that the suite of ground-based instruments at the SGP site could be used to augment the airborne observations. Observations taken by the airborne instrumentation included temperature, water vapor, vertical velocity, droplet and aerosol size distributions, liquid water content, and radiances. It was hoped that the long span of time devoted to the

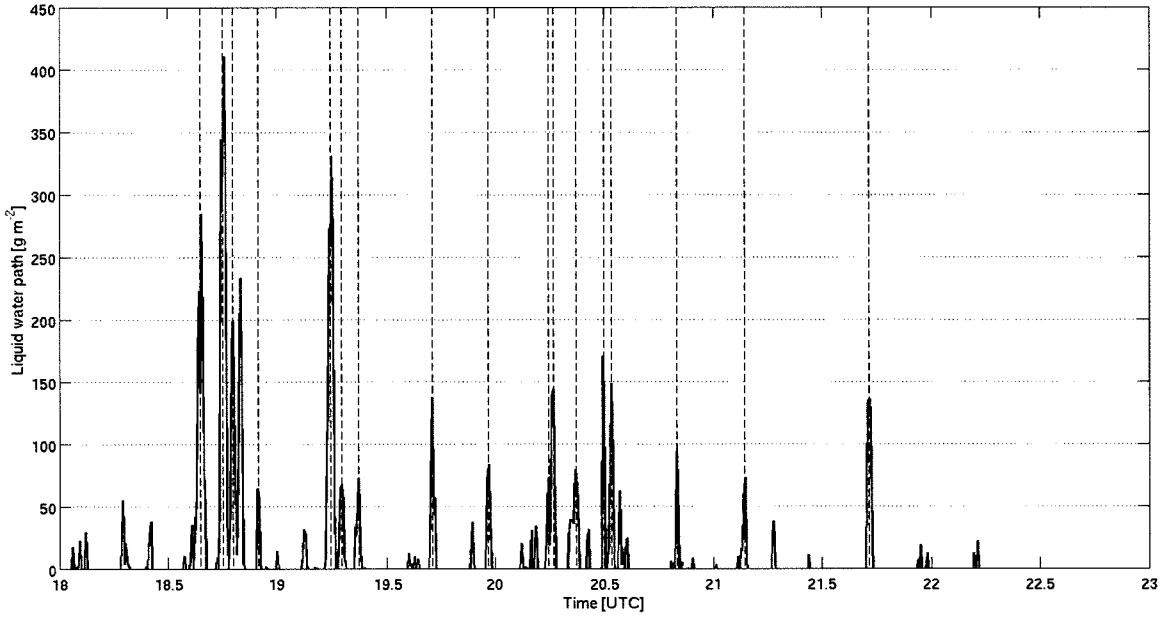

Figure 7.2: Time series of microwave radiometer retrieval of cloud liquid water path over the ARM SGP site on 18 June 2009. Units are in grams per square meter. The dashed blue lines denote the times of attempted retrievals.

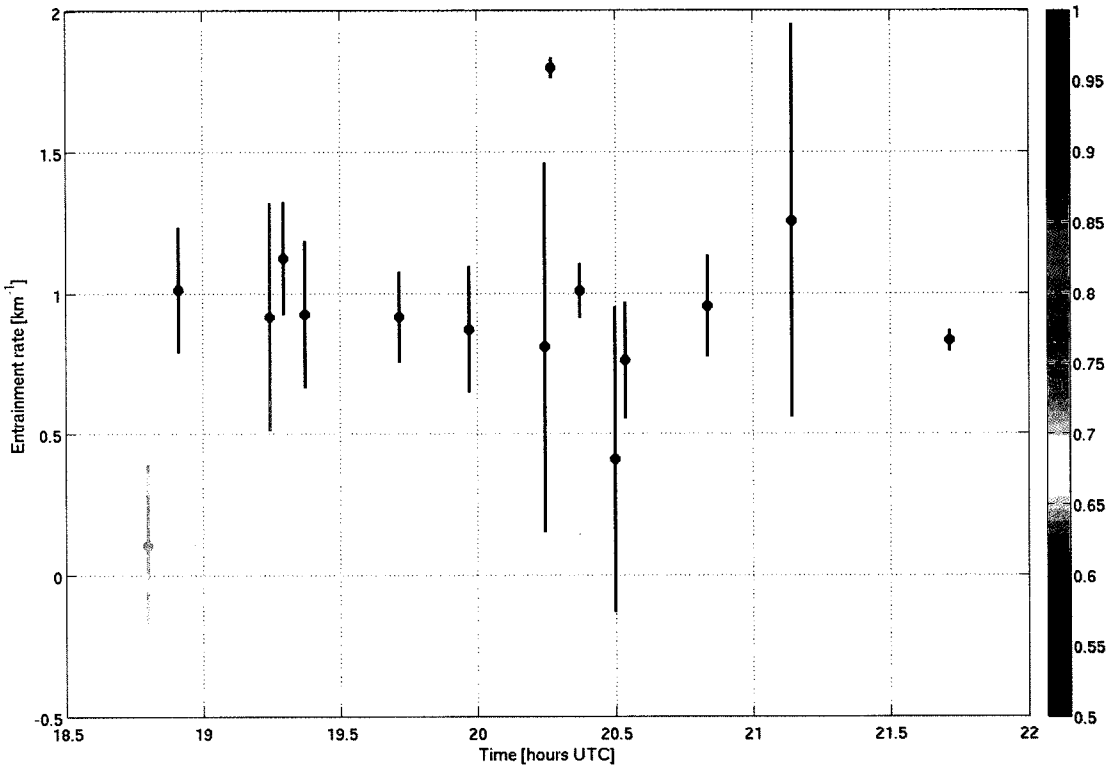

Figure 7.3: Retrieved values of the entrainment rate as a function of time for 19 cumulus clouds over the SGP central facility on 18 June 2009. Units are inverse kilometers. The central point represents the retrieved value, while the whiskers represent the uncertainty for that retrieval. The color of each retrieval plotted here represents the degrees of freedom present in the signal (a unitless parameter), and can be identified using the color bar to the right of the figure.

Project would allow for numerous different cloud types to be observed. In addition, several flights were flown on clear days in order to characterize turbulence patterns. In all, 59 different flights were performed and several of these flights included penetrations of cumulus fields, including one on 18 June 2009. It was reported that over 2000 separate cumulus clouds were sampled during RACORO (Yang et al. 2011).

The measurements taken during these research flights can be used to estimate the entrainment rate using Equation 2.3, which is reprinted here for convenience:

$$\frac{\partial \phi}{\partial z} = -\lambda(\phi - \phi_e) \quad . \tag{7.1}$$

For the purposes of validation in this study, the adiabatically conserved quantity ϕ to be measured at multiple levels is the total water mixing ratio (Q), which is simply the sum of the water vapor mixing ratio (Q_v) and liquid water mixing ratio (Q_l):

$$Q = Q_v + Q_l \quad . \tag{7.2}$$

The primary concern associated with this approach is the lack of multiple penetrations of the same clouds. The technique described in Equation 2.3 was developed for use in tropical deep convection, where cumulus depth is often in excess of 1 km and penetrations at multiple levels of the same cloud are common. By contrast, the aircraft used in the RACORO campaign frequently penetrated cumuli at just one level. However, on cumulus days it typically flew in what was named the "stacked triangle" pattern: it would fly a triangular pattern at a particular altitude, then fly another triangle at a slightly different

altitude, and so on until a significant depth had been profiled. These legs were spaced in the vertical coordinate such that one leg was right below cloud base and another was through the clouds. In the absence of condensation, Q_l is zero and Q reduces to Q_v. By using the observations of Q_v below the cloud and Q within the cloud, the vertical rate of change of Q can be calculated, and from that the entrainment rate can be determined.

The aircraft flight conducted on 18 June 2009 can be seen in Figure 7.3. The Twin Otter airplane departed the Guthrie, Oklahoma, airport at 1815 UTC (1:15 PM CDT) and flew above the clouds to the SGP Central Facility. A triangle was flown at 2900 m (9500 ft) with legs that were approximately 70 km long. After completing this triangle, the airplane flew a spiral from 2900 m to 450 m (1500 ft) and subsequently ascended to 2000 m (6500 ft), which was roughly 150 m below cloud base. Due to a lack of uniformity in cloud top or base altitude, the investigators decided to chase after the tops and bases instead of holding to predetermined altitudes for the remaining triangles. After completing three more triangles, the aircraft performed one more spiral and then returned to home base at 2356UTC (6:56 PM CDT).

The mean entrainment rate for the cumuli encountered on this flight was calculated at 1.82 km^{-1}; the data used to obtain this value came from the flight legs outlined in Figure 7.4. This is more than twice the mean entrainment rate from the ERICA retrievals, which was only 0.87 km^{-1}. Multiple reasons account for the observed discrepancy. First, the two systems are not measuring identical environments. ERICA measures the entrainment rate of individual clouds as they advect over the observation site while the airplane penetrations are taking place up to 20 km downwind from the SGP measurements.

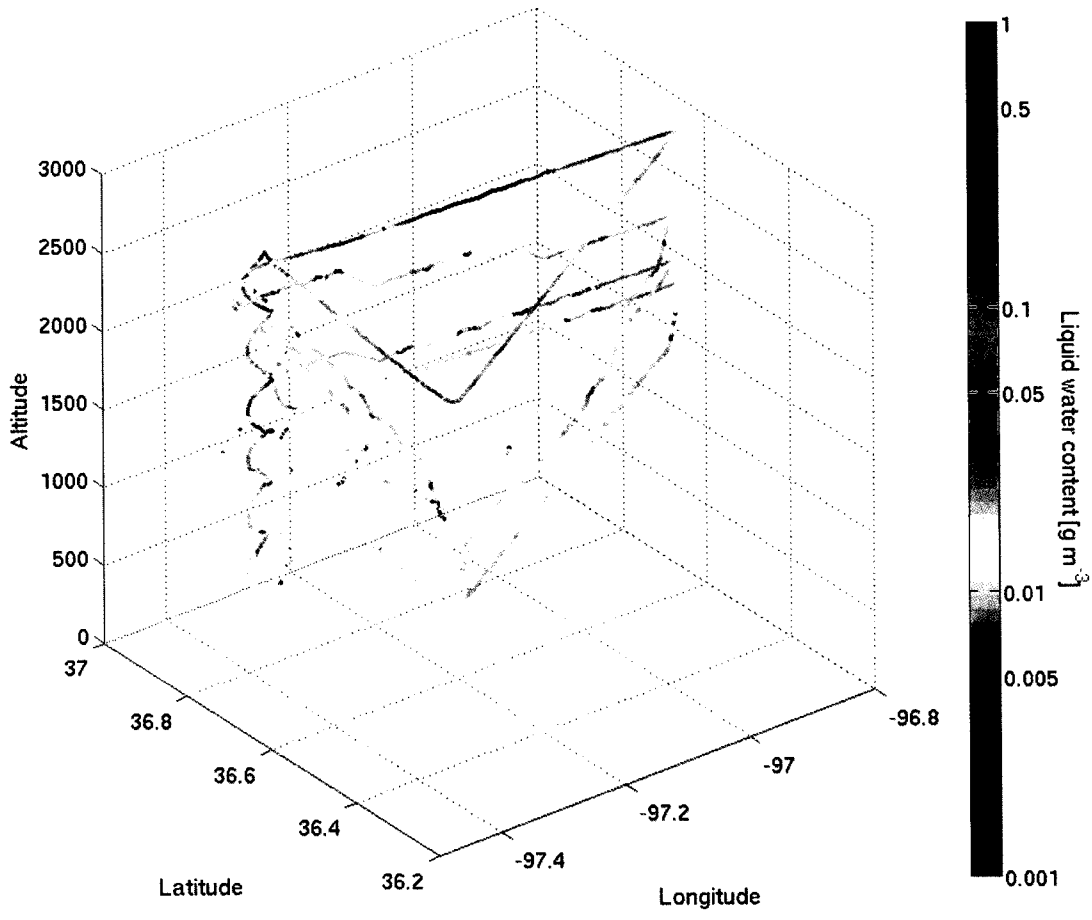

Figure 7.4: Plot of aircraft-observed liquid water content (in g m^{-3}) for 18 June 2009 as a function of aircraft position, showing the "stacked triangle" pattern common to numerous RACORO flights. Latitude and longitude are in degrees; altitude is in meters above ground level. The color of the line is proportional to the cloud liquid water content and corresponds with the color bar to the right of the figure. Note the logarithmic scale.

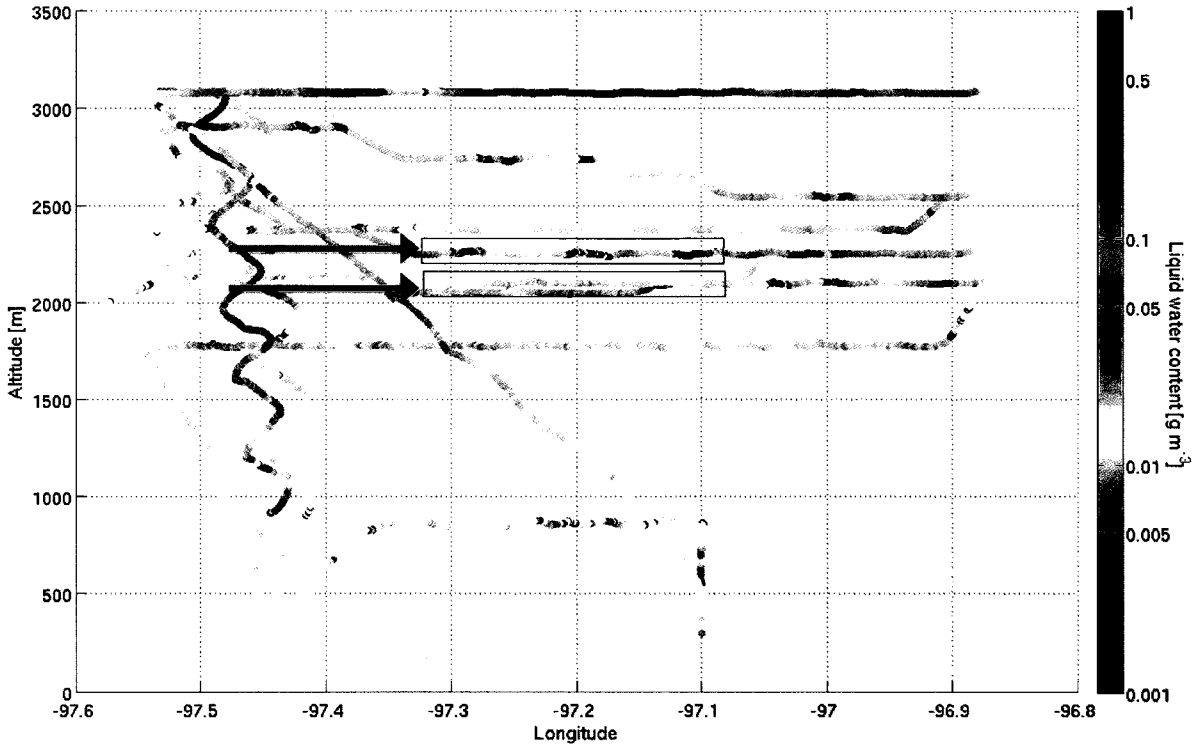

Figure 7.5: South elevation view of Figure 7.4. Legs used for the calculation of entrainment rate in this chapter are emphasized in black.

Second, the mean rates for a particular day were calculated from different types of data. The ERICA entrainment rate is simply the mean of the individual cloud entrainment rates while individual rate retrievals are impossible with this in situ technique. Rather, the aircraft rate is really the entrainment rate of an average cloud. One should not necessarily expect the two means to be the same. Finally, there is little justification in trusting the reliability of the airplane measurements. The following analysis demonstrates how the inherent variability of water vapor and liquid water observations as measured by a moving airplane lead to an extremely large uncertainty in the resulting entrainment rate. Equation 2.3 can be expressed in terms of finite differences as:

$$\frac{Q_1 - Q_0}{z_1 - z_0} = -\lambda(Q_1 - Q_e) \tag{7.3}$$

where Q_1 is the total water mixing ratio inside the cloud, Q_0 is the total water mixing ratio below the cloud, Q_e is the total water mixing ratio between clouds, z_1 is the altitude of the leg at cloud level, and z_0 is the altitude of the leg below the cloud. Solving Equation 7.3 for λ and expanding the product in the denominator yields the following expression:

$$\lambda = \frac{Q_0 - Q_1}{z_1 Q_1 - z_1 Q_e - z_0 Q_1 + z_0 Q_e}. \tag{7.4}$$

For a given parameter x that is a function of u, v, w, where all three are independent of each other, the following expression can be used to find the variance in x (Bevington and Robinson 1992):

$$\sigma_x^2 = \sigma_u^2\left(\frac{dx}{du}\right)^2 + \sigma_v^2\left(\frac{dv}{du}\right)^2 + \sigma_w^2\left(\frac{dw}{du}\right)^2 \tag{7.5}$$

where σ^2 of a variable is the variance of that variable. Through the use of Equation 7.4 in Equation 7.5 and liberal use of the quotient rule, an analytical function describing the variance in λ can be obtained:

$$\sigma_\lambda^2 = \sigma_{Q_0}^2\left(\frac{1}{(z_1 - z_0)(Q_1 - Q_e)}\right) + \sigma_{Q_1}^2\left(\frac{-1}{(z_1 - z_0)(Q_1 - Q_e)} - \frac{(Q_0 - Q_1)(z_1 - z_0)}{[(z_1 - z_0)(Q_1 - Q_e)]^2}\right)^2 + \sigma_{Q_e}^2\left(\frac{-(Q_0 - Q_1)(z_0 - z_1)}{[(z_1 - z_0)(Q_1 - Q_e)]^2}\right) \tag{7.6}$$

For this case, the variances can be calculated directly from the data being used in the calculation. The data values for *Q1* were obtained from the observations of the total water mixing ratio where the LWC was greater than 0.01 g m^{-3}. The values for Q_0 were taken from the observations from the cloud level leg where the LWC was less than 0.01 g m^{-3}, while the Q_e values was taken from the leg below cloud base where the vertical velocity was at least one standard deviation above the mean in order to capture the values associated with the updrafts. The variances associated with these values were used in Equation 7.6 and are shown in Table 7-1. Using this error analysis, the standard deviation about the mean of the aircraft observed entrainment rate was determined to be 2.17 km^{-1}, a larger magnitude than the mean itself. At first glance, this into question what, if any, scientifically relevant information can be obtained from the airborne observations. However, the assumption of independence between the constituent variables is not necessarily true; one would expect the total water mixing ratio in the updraft and within

the cloud to be correlated. Therefore, the uncertainty value presented here can only represent the uppermost value possible while the true uncertainty is somewhat less. Nevertheless, ERICA is a more robust method than airborne in situ observations when it comes to measuring the entrainment rate and has a well-characterized uncertainty. Results for this case study are summarized in Table 7-2.

Table 7-1: Variances of mixing ratios in the 18 June 2009 case study flights

$\sigma^2_{Q_0}$	1.10
$\sigma^2_{Q_1}$	1.02
$\sigma^2_{Q_e}$	0.86

Table 7-2: Entrainment rate observations for 18 June 2009

	ERICA	**Airplane**
λ [km^{-1}]	0.91	1.82
Error [km^{-1}]	0.28	2.17

8. Multi-day analysis

a. Introduction and descriptive statistics

As part of an effort to characterize this algorithm's performance and understand the variability of the entrainment rate in cumulus clouds, ERICA was used to generate a three-month dataset of entrainment rates. This dataset encompasses the months of May, June, and July 2009, which coincides with peak cumulus formation at the SGP site as well as occasional validation flights from the RACORO campaign. In all, 86 retrievals were performed over the course of twelve days. Four of the days included in this analysis also included RACORO flights. Clouds were selected for entrainment rate retrieval according to the objective process outlined in Chapter 6. Figure 8.1 shows the number of successful retrievals completed on each of these days. (To better facilitate comparisons, all figures for this chapter are located together in starting on Page 78.) The mean number of retrievals per day was 7.2, but the retrievals are not evenly distributed. This results in a fairly substantial standard deviation of 8.95. The times of day during which the retrievals took place can be seen in Figure 8.2. Note that the peak time of cumulus formation coincides with the time of maximum surface temperature, which is expected for a cloud driven by surface-based buoyancy: as diurnal heating increases (decreases) the number of positively-buoyant moist plumes increases (decreases). It is interesting to note, however, that the peak in retrievals is one hour later than the peak in cumulus cloud fraction over SGP found by Berg and Kassianov (2007), which could imply that clouds are smaller in spatial extent at that time of day. No retrievals were undertaken at night, even though

cumuli may occasionally linger past sunset. The Total Sky Imager (TSI), used to identify what times to initiate retrievals, is fundamentally a visible-wavelength camera and thus only operates during daylight hours. Since no information about the presence of nocturnal cumuli is known, all the retrievals presented here were processed for daytime clouds.

The histogram of retrieved entrainment rates for this dataset is shown in Figure 8.3 while the histogram of the error magnitudes is shown in Figure 8.4. The mean entrainment rate for this dataset is 1.02 km^{-1} +/- 0.33 km^{-1}. Table 8-1 displays some of the fundamental statistics used to describe the entrainment rate distribution:

Table 8-1: Descriptive statistics of entrainment rate retrievals for May-July 2009

Mean	1.02 km^{-1}
Median	1.01 km^{-1}
Standard Deviation	0.44 km^{-1}
Skewness	1.29
Kurtosis	9.13
Mean error	0.33 km^{-1}
Median error	0.22 km^{-1}
Standard deviation of the error	0.44 km^{-1}
Skewness of the error	0.758
Kurtosis of the error	2.28

Not all retrievals were successful. Several conditions prevented ERICA from commencing or completing a retrieval. The most common issue was a lack of necessary data at a particular time. While the ARM SGP instruments generally have excellent uptime, five separate instruments are required for ERICA to operate. If only one instrument is down, the retrieval cannot operate. The data product that was unavailable the most often was the temperature profile from the AERI. On days where the cumulus field was really a broken stratocumulus deck, the gaps between the clouds were frequently small enough that AERI was unable to get a valid temperature profile without cloud contamination. While work to retrieve AERI observations in the presence of clouds is progressing (Löhnert et al. 2009), these data are not yet operational. This led to several days in which only one to three retrievals were successful despite several hours of cumulus cover over the SGP site. Current and future instrumentation at the SGP site measure many of the same quantities as the instruments used in ERICA. A planned upgrade to the ERICA algorithm will be making provisions for integrating these data streams into the retrieval as backups when one of the primary instruments is unavailable. The decision was made to not include them at this time for this particular study in order to have a consistent error profile for intercomparisons. Occasionally, a retrieval would have the necessary ingredients to begin but would not be able to converge on a solution. If there is little sensitivity in the modeled space to the conditions at that time, the retrieval may move slowly in one direction or instead veer haphazardly among values. In either case, the system is unlikely to converge at a solution. If convergence is not achieved in eight iterations, ERICA assumes that a solution cannot be found for the current conditions and aborts the retrieval.

b. Aircraft comparison and validation

Entrainment rates and error bars for the four flights during the study period coincident with ERICA retrievals were calculated according to the procedure outlined in the previous chapter. The mean entrainment rate for all aircraft flights was calculated to be 1.42 km^{-1} with a standard deviation of 0.40 km^{-1}. By comparison, the mean ERICA retrieved entrainment rate for those days was 1.02 km^{-1} with a standard deviation of 0.39 km^{-1}. A plot of entrainment rates from the airplane versus contemporaneous ERICA retrievals is shown in Figure 8.5.

As before, the ERICA retrievals and aircraft observations, while of the same order, have substantial differences. The reliability of aircraft observations can again be called into question. Using the same error analysis outlined in the 18 June case study yields a mean error of 3.56 km^{-1} while the standard deviation is 2.60 km^{-1}. In every case, the magnitude of the error bar is larger than that of the measurement itself. By comparison, the mean and standard deviation of the error for ERICA for the four comparison days is 0.39 km^{-1} and 0.1448 km^{-1} respectively. While one would prefer to have a more robust tool for validation, no other observations of the entrainment rate currently exist. A possible avenue for future validation would be to use a cloud-resolving or large eddy simulation model to simulate cumulus clouds with known environments and entrainment rates, and then use ERICA to retrieve the entrainment rate from this synthetic data.

c. Dependence on other variables

To determine what, if any, relationship there may be between time and entrainment rate, the dataset was divided into ten bins based on the time of the retrieval. Times were not adjusted to account for variations in the diurnal cycle; since the dataset straddles the summer equinox the maximum displacement of local solar noon over the three month period is only ten minutes. The symmetric distribution of cloud times as shown in Figure 8.2 means that the bins will be smaller the closer to the center they occur. For this dataset, the amount of time encompassed by these bins ranges from 2.2 hours on the right edge of the temporal distribution to less than 15 minutes at the center. Each bin contains either eight or nine retrievals.

Once the data had been binned, the mean and standard deviation of the entrainment rate in each bin was calculated. The results are shown in Figure 8.6. Somewhat surprisingly, there is a clear linear dependence of entrainment rate on time of day. Surface temperature and boundary layer turbulence are correlated (Stull 1988), so one would expect an increase in entrainment rate with time up to the time of maximum surface temperature and maximum cumulus formation. However, the entrainment rate keeps increasing beyond that time. A possible explanation is that the clouds viewed in the late afternoon are no longer in the active stage of their live cycle where they entrained prodigious quantities of environmental air. Now that they are in the passive stage, they are still entraining air, but at a lesser temporal rate due to decreased turbulent kinetic energy. However, since the fractional entrainment rate measures the over amount of air entrained for a particular cloud, it will continue to increase as time goes on.

Other observed variables exhibit little correlation to the retrieved entrainment rate. Scatter plots of entrainment rate with effective radius (Figure 8.7), liquid water path (Figure 8.8), and vertical velocity (Figure 8.9) are shown at the end of this chapter. Table 8-2 displays the correlation coefficients and associated significance of entrainment rate and these parameters.

Table 8-2: Correlation coefficients (r) and probability (p) between retrieved entrainment rate and selected variables

	r	p
Effective radius	-0.095	0.19
Liquid water path	0.13	0.12
Vertical velocity	0.15	0.084

d. Impact of the first guess

The discussion of Gauss-Newton optimal estimation in Chapter 3 notes the importance of the first guess to the retrieval process. A cursory inspection of the entrainment rate distribution shows that there is a bias towards retrieved values of 1 km^{-1} that coincides with the assumed first guess for the algorithm. In practice, this means that the displacement of the adjusted entrainment rate after one iteration of the algorithm often falls within the error bars associated with the measurement. In these cases there is not enough information present in the observations to justify any further adjustment of the entrainment rate and so the retrieval ceases its execution. To determine what impact the first guess may have on the retrievals, the dataset was run again with the first guess

changed to 1.25 km^{-1}, and the results are shown in Figure 8.10. This secondary dataset has a mean of 1.09 km^{-1} and a standard deviation of 0.38 km^{-1}, which compares well to the original dataset. Inspection of Figure 8.7 shows that the second dataset exhibits behavior that is very similar to the first: the first guess is "good enough" by the standards of the objective adjustment, and so little adjustment is undertaken. The correlation between the two datasets is 0.62. However, if one only looks at the retrievals that are outside of the range defined by the first guess plus or minus the mean error, the correlation then rises to 0.97. This lends credence to the robust nature of ERICA, since the retrieval will converge on an entrainment rate regardless of the starting point as long as the real-world entrainment rate is outside of the range defined by the retrieval error.

e. Degrees of freedom of the retrievals

Overall, the entrainment rate retrievals in this dataset exhibited high values for degrees of freedom for the signal, with the distribution strongly skewed towards 1 (Figure 8.11). The median value of this distribution was 0.95 and the 25th percentile was 0.76. These high values are yet another reason for confidence in the entrainment rate retrievals produced by ERICA: as the degrees of freedom represent the sensitivity of the system to the entrainment rate, the high values indicate that the information content of the forward model is high and that the algorithm is successfully retrieving its intended parameters.

The degree of freedom is subsequently related to the error on the retrieved product. If the information content is low, the retrieved product will naturally have a large error bar as the system is unable to adjust to a better retrieval. When the information content is

high, the system is able to iterate towards a solution with a high degree of confidence. The correlation between the degrees of freedom and the error for this dataset is -0.81, and is illustrated in Figure 8.12. There is less of a correlation between the information content and the entrainment rate itself, which is calculated to be -0.43 and shown in Figure 8.13.

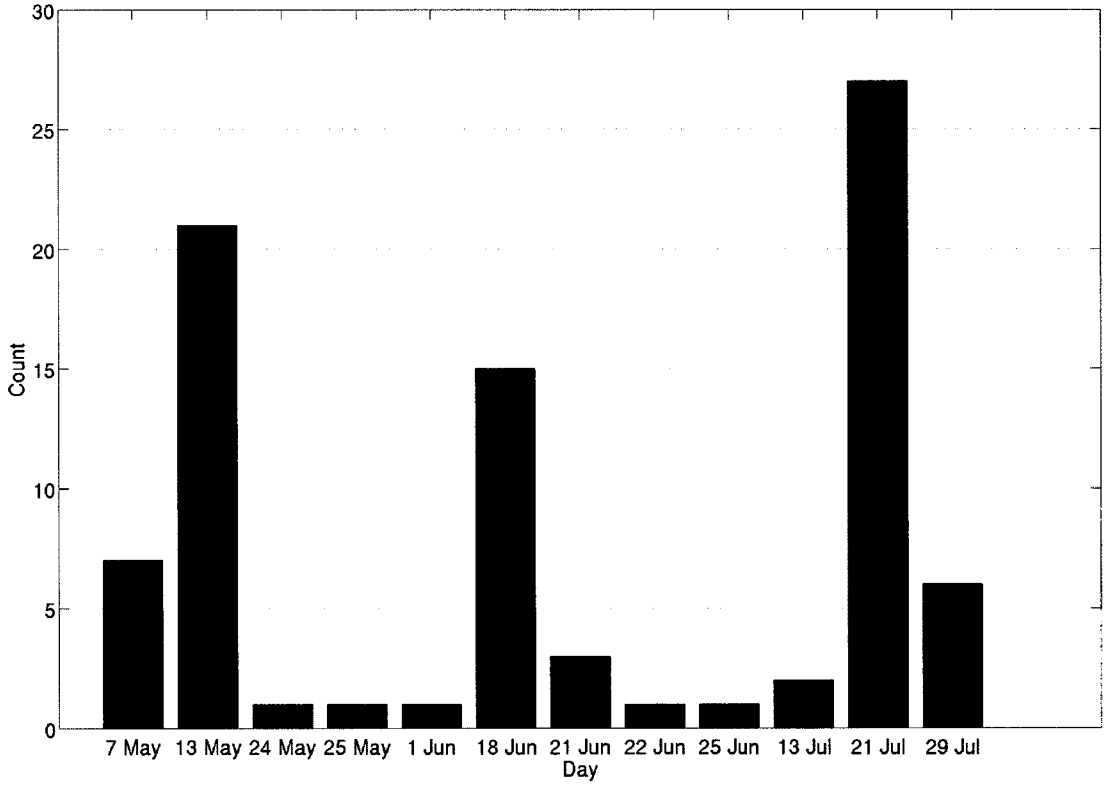

Figure 8.1: Distribution of retrievals by day for the May-June 2009 dataset. The mean number per day is 7.2 with a standard deviation of 8.95.

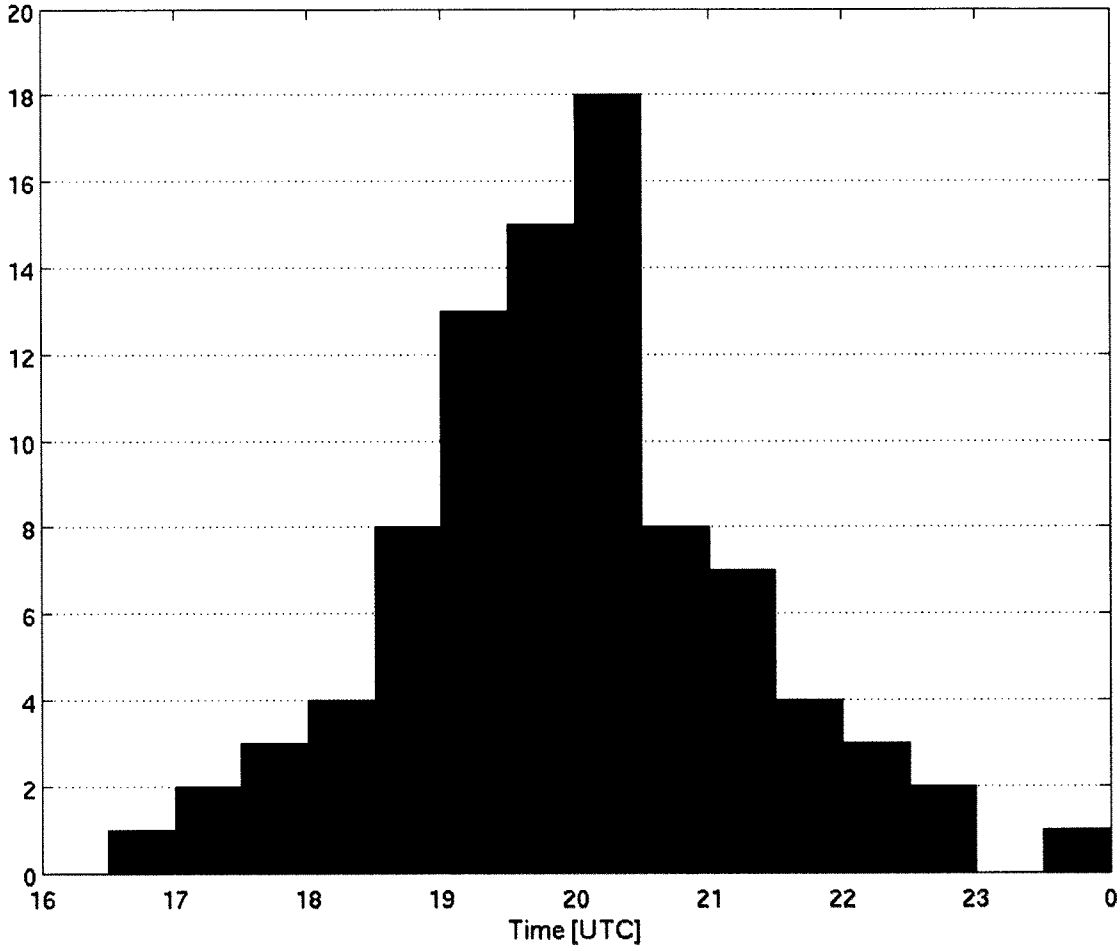

Figure 8.2: Histogram of times of valid retrieval for the ERICA-processed May-June 2009 dataset illustrated in Figure 8.1. Data is sorted into 0.5 hour bins.

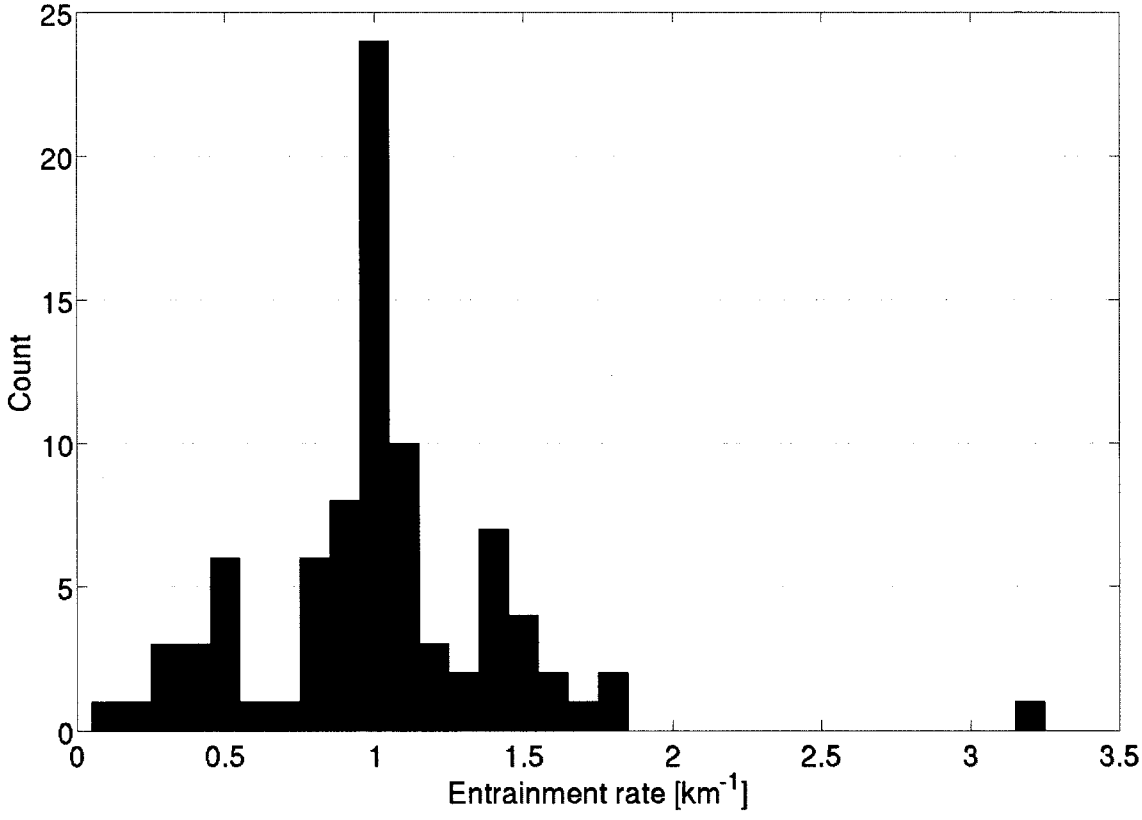

Figure 8.3: Histogram of ERICA-retrieved entrainment rates for cumulus clouds over the ARM SGP site from May through July 2009. Units are inverse kilometers.

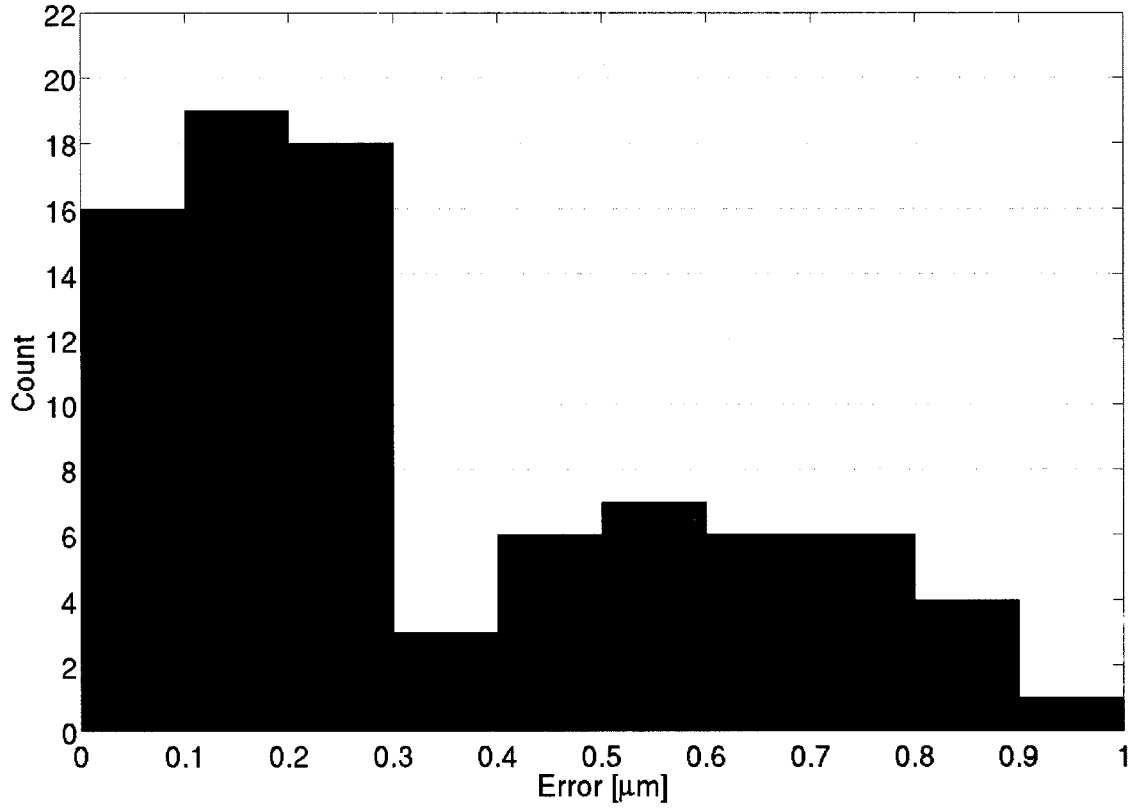

Figure 8.4: As in Figure 8.3, but for the error from the ERICA retrievals.

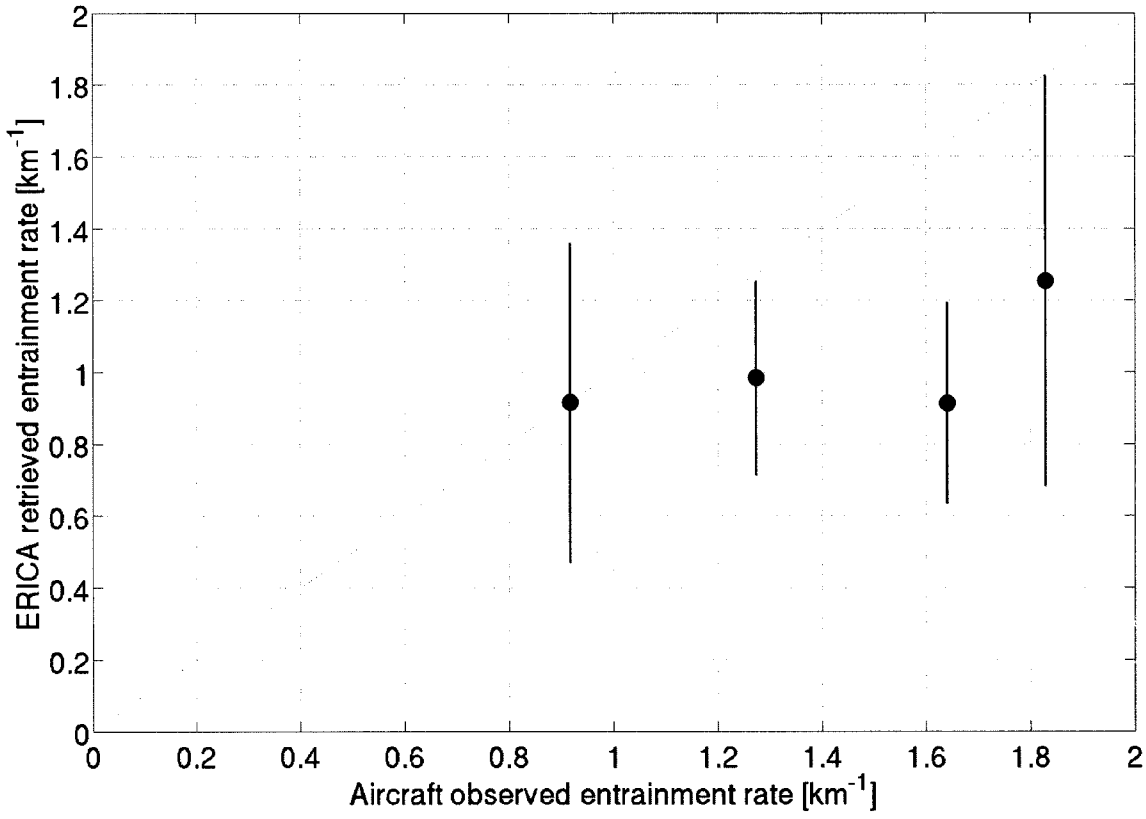

Figure 8.5: Scatter plot of aircraft observed entrainment rate (x-axis) against mean ERICA retrieved entrainment rate (y-axis). Units are inverse kilometers. The thin black lines represent the mean error of the ERICA retrievals for each day; the error bars from the aircraft observations are too large to be included on this figure (the mean value was 3.6 km^{-1}). The diagonal line represents a 1:1 fit.

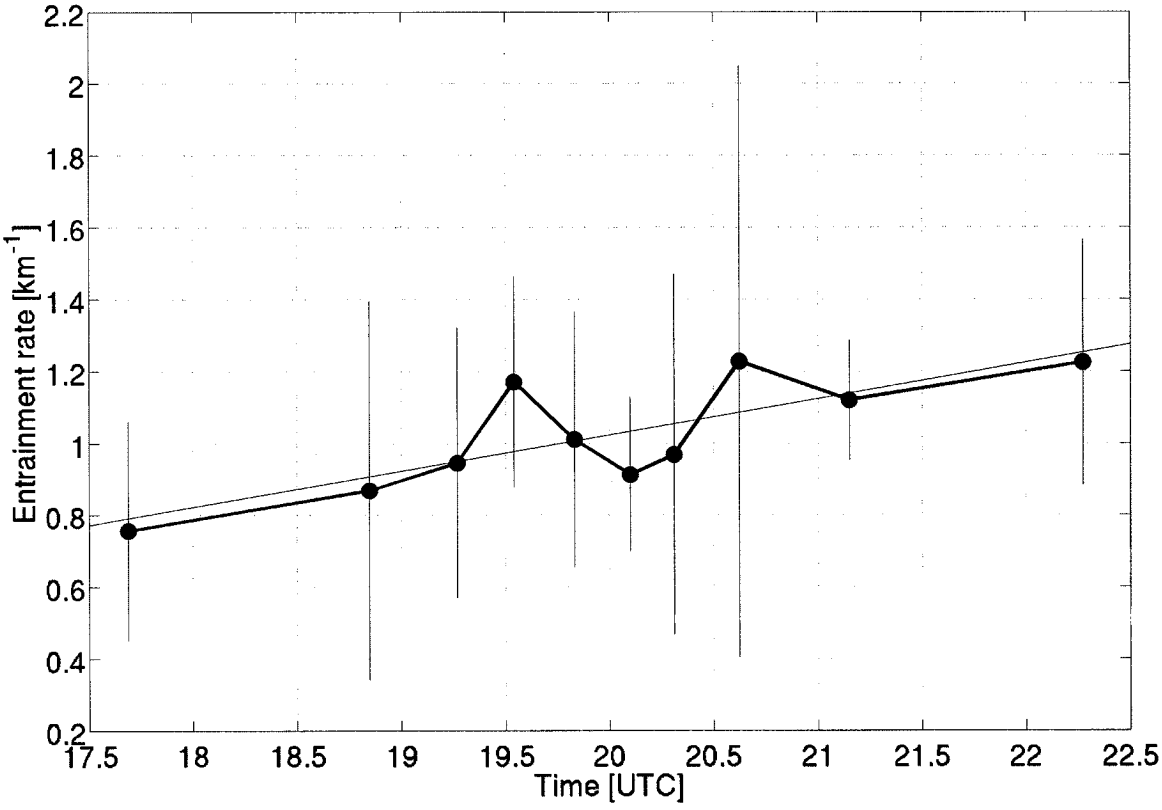

Figure 8.6: Time series of bin-averaged entrainment rates (in km^{-1}) for the May-June 2009 dataset. Each bin represents one decile of the data according to time of day. The thick black line represents the mean, the thin black lines represent the standard deviation of the data in each bin, and the red line represents the linear fit of the means.

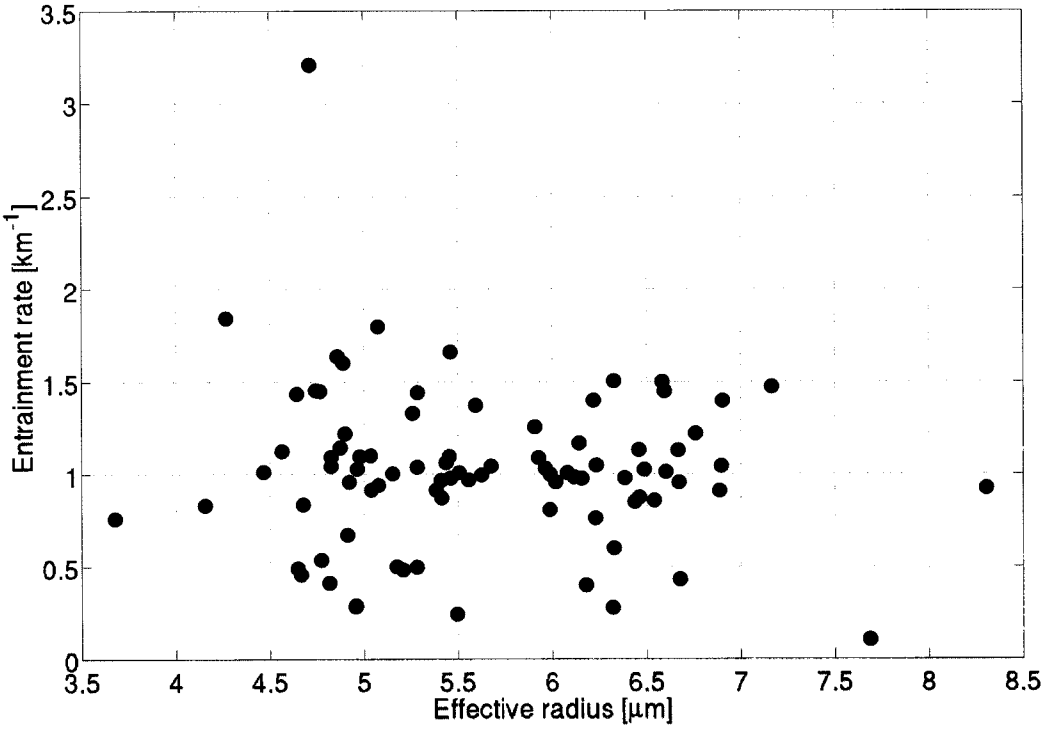

Figure 8.7: Scatter plot of observed effective radius (in microns) against the retrieved entrainment rate (in km^{-1}).

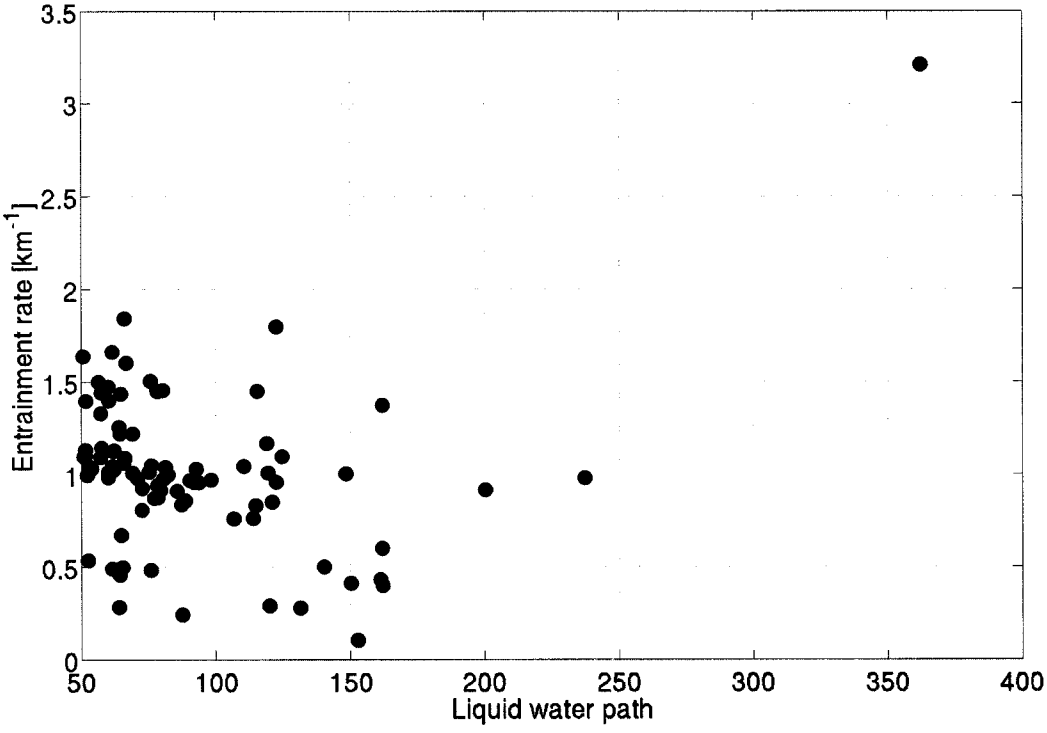

Figure 8.8: As in Figure 8.7, but for liquid water path (in g m^{-2}).

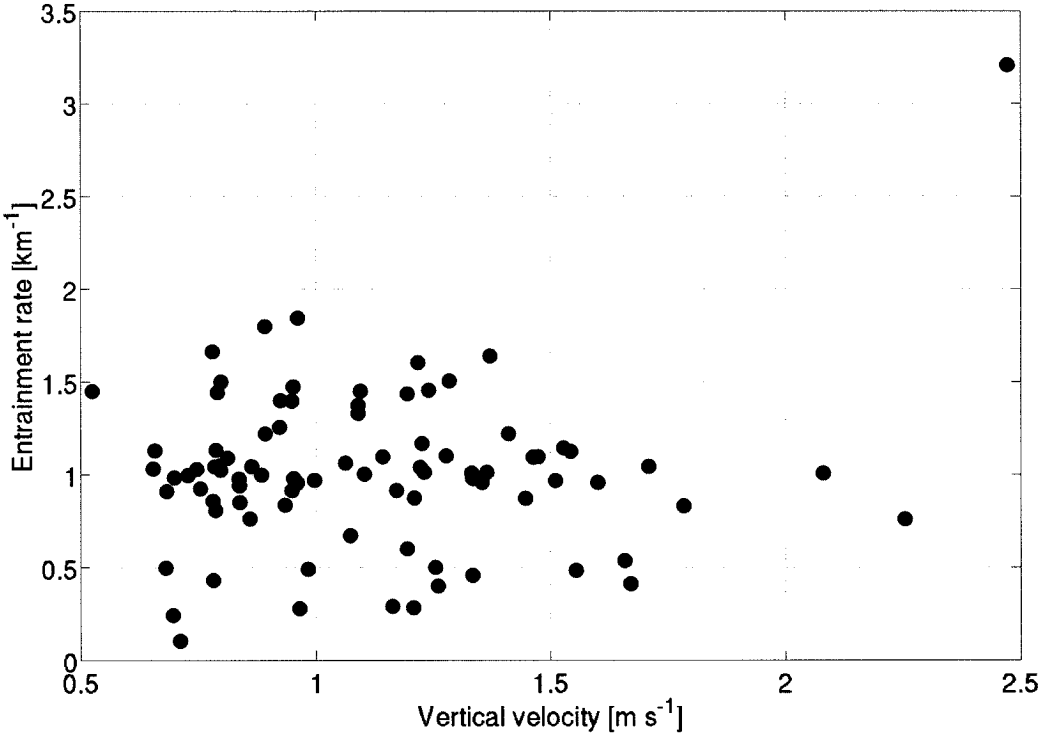

Figure 8.9: As in Figure 8.7, but for vertical velocity (in m s^{-1}).

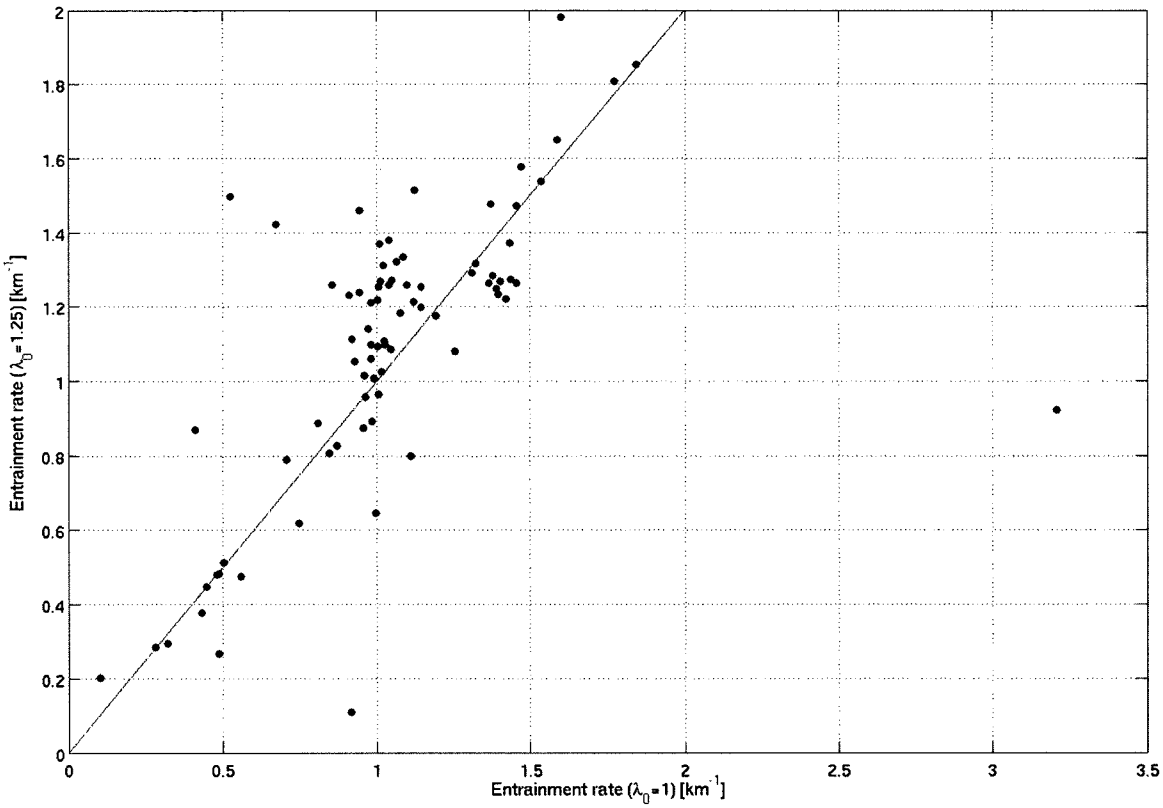

Figure 8.10: Scatter plot of ERICA-retrieved entrainment rates of the May-June 2009 dataset for two separate first guesses. Units are inverse kilometers. The x and y axes mark first guesses of 1.0 and 1.25 km^{-1} respectively. The diagonal line represents a 1:1 fit.

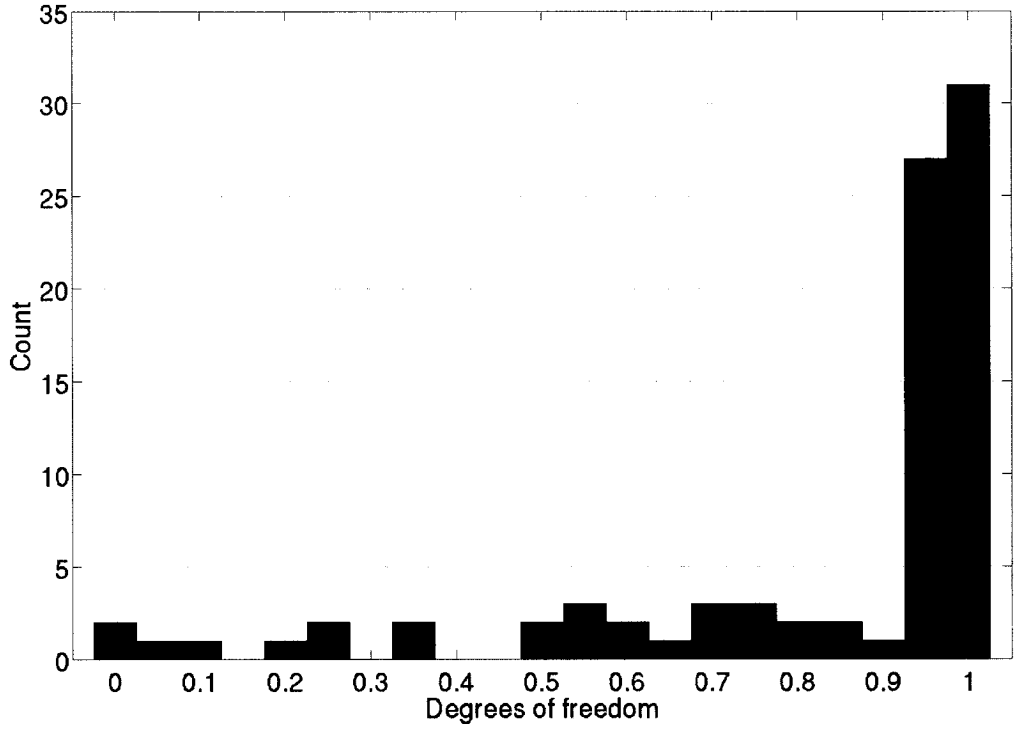

Figure 8.11: Histogram of the degrees of freedom (a measure of information content and a unitless parameter) for the May-June 2009 retrievals.

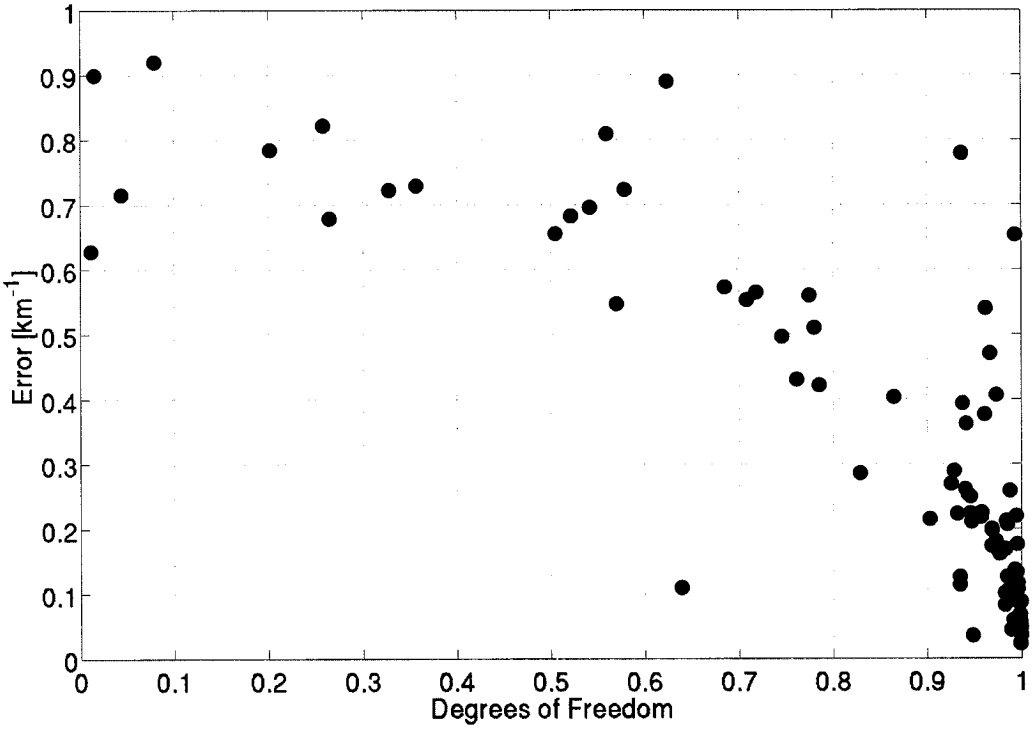

Figure 8.11: Scatter plot of degrees of freedom (on the x-axis) and error of the retrieval (on the y-axis, in km^{-1}).

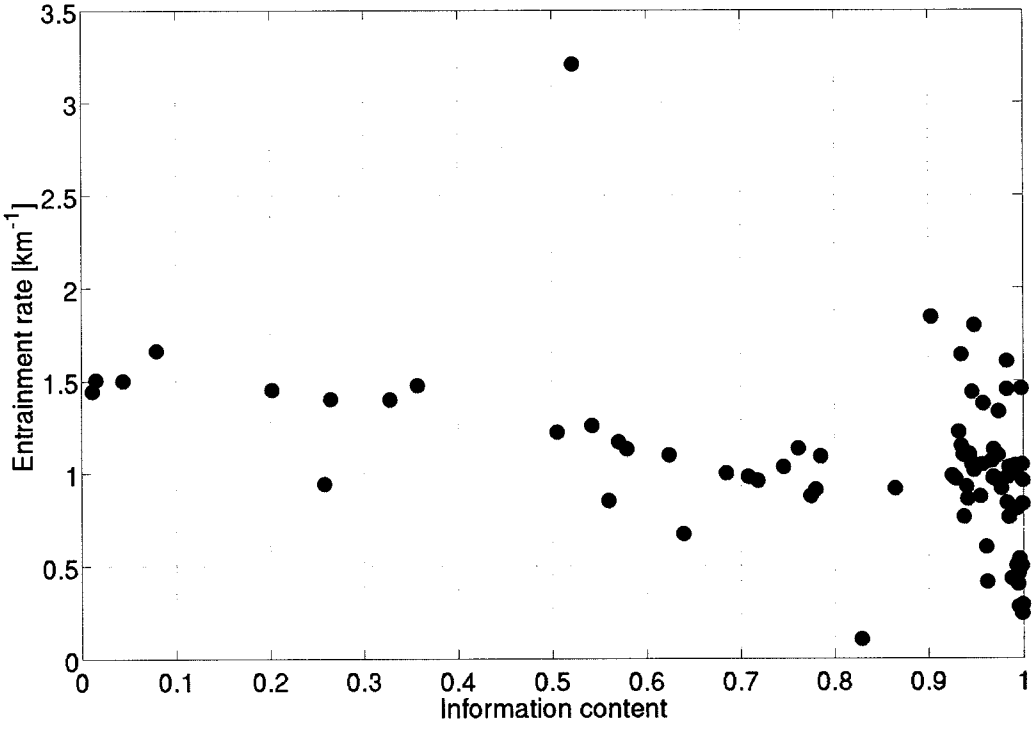

Figure 8.12: As in Figure 8.12, but for the retrieved entrainment rate (in km^{-1}).

9. Summary and outlook

Cumulus clouds mix with their environment throughout their life cycle. The entrainment of environmental air into the cloud radically alters its characteristics on both the microphysical and macrophysical scale. The entrainment is quantified through a quantity known as the fractional entrainment rate, which is a simple measurement of the amount of mixing that has taken place between a cloud and the air that surrounds it. The modeling community requires access to accurate, well-characterized measurements of the amount of mixing taking place for a particular cumulus cloud (or for a characteristic cloud in a cumulus field) in order to properly parameterize the impact that these small convective clouds have on the environment. However, accurate entrainment rate measurements have been difficult to obtain through existing means. Currently, knowledge of the entrainment rate comes from aircraft observations, which due to their expense are uncommon and do not represent a wide array of meteorological and climatological conditions. Furthermore, the error on the few available aircraft-borne measurements is of such a sufficiently large magnitude as to call into question the utility of these observations. Unfortunately, the literature makes little effort to characterize the error of these measurements, thus lending them an aura of accuracy that may not be deserved.

To rectify this situation, a remote sensing algorithm capable of retrieving the cumulus entrainment rate has been developed. This algorithm, named the Entrainment Rate In Cumulus Algorithm (ERICA), uses measurements from the suite of ground-based instruments at the Atmospheric Radiation Measurement (ARM) Southern Great Plains

(SGP) site as inputs to a Gauss-Newton optimal estimation retrieval. The core of this retrieval is the Explicit Mixing Parcel Model (EMPM), a one-dimensional parcel model that entrains discrete blobs of air and mixes them throughout the parcel at a finite rate.

ERICA has several key advantages over the airplane observations. Observations are far more frequent, for one. Since each of the instruments that ERICA requires has been collecting data for several years, a large back catalog of data is available for generating an analysis of entrainment in southern Great Plains cumulus. No funding is required to obtain the data necessary to generate an extensive climatology. Additionally, the characterized uncertainty of the ERICA observations is of a much smaller magnitude than the measurement itself, empowering end-users to use the entrainment product with confidence.

The entrainment rates produced by ERICA were used to investigate the characteristics of entrainment on both the one-day and three-month scales. The mean values are in line with the expected values for entrainment rate. Variability from one cloud to the next is significant, as the standard deviation about the mean of the measurements is over 40% of the size of the measurement itself. The preliminary analysis shows a weak increase in the cumulus entrainment rate as the afternoon progresses possibly do to enhanced mixing with increasing surface temperatures giving way to aged clouds that are mixing away entirely. None of the other basic measurements tested correlate with the entrainment rate.

Validation of the product is difficult. A small number of coincident aircraft datasets were available during the test period. The entrainment rate measurements from these

flights show that ERICA has a bias towards smaller entrainment rates than the airborne measurements indicate. However, the number of airplane measurements is so small and the size of the error is so large that it is difficult to place any confidence in comparisons that use them, but no other tools are available. Despite the relative lack of external validation data, it is still possible to maintain confidence in the retrieval's output due to the high values for information content and its propensity to converge on the same value regardless of where it begins.

There are many avenues available for enhancing and applying the ERICA retrievals. Obviously, a climatology over a term longer than the three months presented here is desired. With at least four years of data available, it would be desirable to process many years of data to see if the knowledge gained thus far is an outlier or if it is something real. This would also provide the chance to see what seasonal or annual cycles may be found in the entrainment rate, and would help with validating parameterizations. For example, the short term climatology presented here shows that the ECMWF parameterization described in Equation 2.4 is off by roughly a factor of five.

As described in Chapter 8, ERICA is currently only designed to ingest specific instruments and must fail when one of those instruments is unavailable. Current plans are to redesign the algorithm so that it can default to another instrument when one of its primary measurements is unavailable. This would require analysis of the error characteristics of each additional instrument but would also allow for many more retrieval days than are currently possible. An error analysis may even show that certain new

instruments can provide the necessary measurements with less noise than the existing ones, which would reduce the overall noise of the retrieval.

Another area for imminent investigation is the relationship between entrainment, sensible heat, and turbulent kinetic energy. The preliminary analysis conducted in Chapter 9 showed little dependence of entrainment rate on the basic surface-measured quantities, but additional insight may be gleaned from investigating more complex variables. Since cumulus are an important part of the interface between the boundary layer and the free troposphere, there may also be correlations between the entrainment rate and mid-troposphere dynamics which would provide another facet for study. Recent work (Wulfmeyer et al. 2010) has shown that the Raman lidar can also retrieve profiles of turbulent parameters, which could provide another arena for correlation with ERICA retrievals.

ERICA no longer has to be limited to the study of shallow continental clouds as it is now possible to investigate tropical deep convection. As part of the American Recovery and Reinvestment Act (ARRA, colloquially known as "the Stimulus"), ARM received $60 million dollars in infrastructure funding to upgrade instruments and add new ones to its system of observing sites. The site in Dawrin, Australia, now has all of the observations that ERICA requires to retrieve the entrainment rate. The deep convective clouds that form over the tropics arguably have a more significant impact on the global radiation balance than their continental brethren while suffering from the same lack of entrainment measurements. An investigation of the distribution of deep tropical convection

entrainment rates and a comparison to continental cumulus would be of great utility to the global climate modeling community.

The role of convective cloud entrainment in the radiative and thermodynamic balance of the atmosphere is an important one that up until now has been difficult to quantify and characterize. As a robust and accurate tool for measuring the entrainment rate, ERICA promises to enable studies that lead to new understandings of the atmosphere.

10. References

Baker, M. B., R. G. Corbin, and J. Latham, 1980: The influence of entrainment on the evolution of cloud droplet spectra: 1. A model of inhomogeneous mixing. *Quart. J. Roy. Meteor. Soc.*, **106**, 581-598.

Batchelor, G. K. 1952: The effect of homogeneous turbulence on material lines and surfaces. *P. R. Soc. London*, **213**, 349-366.

Benjamin, S. G., J. M. Brown, K. J. Brundage, B. E. Schwartz, T. G. Smirnova, and T. L. Smith, 1998: The operational RUC-2. Preprints, *16th Conf. on Weather Analysis and Forecasting*, Phoenix, AZ, Amer. Meteor. Soc., 249-252.

Berg, L. K. and coauthors, 2009: Overview of the Cumulus Humilis Aerosol Processing Study (CHAPS). *Bull. Amer. Meteor. Soc.*, **67**, 3336-3354.

Berg, L. K. and E. I. Kassianov, 2007: Temporal variability of fair-weather cumulus statistics at the ACRF SGP site. *J. Clim.*, **21**, 3344-3358.

Berg, L. K., E. I. Kassianov, C. N. Long, D. L. Mills Jr., 2011: Surface summertime radiative forcing by shallow cumuli at the Atmospheric Radiation Southern Great Plains site. *J. Geophys. Res.* **116**, doi:10.1029/2010JD014593.

Bevington, P. R., and D. K. Robinson, 1992: *Data reduction and error analysis for the physical sciences.* 2nd. Ed., McGraw-Hill, 328 pp.

Blyth, A. M., 1993: Entrainment in cumulus clouds. *J. Appl. Meteor.*, **32**, 626-641.

Blyth, A. M., W. A. Cooper, and J. B. Jensen, 1988: A study of the source of entrained air in Montana cumuli. *J. Atmos. Sci,* **45**, 3944-3964.

Clothiaux, E. E., and coauthors, 1998: The Atmospheric Radiation Measurement Program cloud radars: Operational modes. *J. Atmos. Oceanic Technol.*, **16**, 819-827.

Clothiaux, E. E., T. P. Ackerman, G. G. Mace, K. P. Moran, R. T. Marchang, M. A. Miller, and B. E. Martner, 2000: Objective determination of cloud heights and radar reflectivities using a combination of active remote sensors at the ARM CART sites. *J. Appl. Meteor.*, **39**, 645-665.

Clough, S., A., M. W. Shephard, E. J. Mlawer, J. S. Delamere, M. J. Iacono, K. Cady-Pereira, S. Boukabara, and P. D. Brown, 2005: Atmospheric radiative transfer modeling: a summary of the AER codes. *J. Quant. Spectrosc. Ra.*, **91**, 233-244

Clough, S. A., M. J. Iacono, and J.-L. Moncet, 1992: Line-by-line calculation of atmospheric fluxes and cooling rates: Application to water vapor. *J. Geophys. Res.*, **97**, 15761-15785.

Cotton, W. R., G. H. Bryan, and S. C. van den Heever, 2011: *Storm and Cloud Dynamics.* 2nd ed. Academic Press, 809 pp.

Cooper, W. A., 1989: Effects of variable growth histories on droplet size distributions. Part I: Theory. *J. Atmos. Sci.,* **46**, 1301-1311.

Fan, J., R. Zhang, G. Li, and W.-K. Tao, 2007: Effects of aerosols and relative humidity on cumulus clouds. *J. Geophys. Res.*, **112**, doi: 10.1029/2006JD008136

Fankhauser, J. C., I. R. Paluch, W. A> Cooper, D. W. Breed, and R. E. Rinehart, 1982: *Hailstorms of the Central High Plains.* Vol. 1, *The National Hail Research Experiment,* Colorado Associated University Press, 95-149.

Feltz, W. F., W. L. Smith, R. O. Knuteson, H. E. Revercomb, H. M. Woolf, and H. B. Howell, 1998: Meteorological applications of temperature and water vapor retrievals from the ground-based Atmospheric Emitted Radiance Inteferometer (AERI). *J. Appl. Meteor.*, **37**, 857-875.

Feltz, W. F., W. L. Smith, H. B. Howell, R. O. Knuteson, H. Woolf, and H. E. Revercomb, 2003: Near-continuous profiling of temperature, moisture, and atmospheric stability using the Atmospheric Emitted Radiance Interferometer (AERI). *J. Appl. Meteor.*, **42**, 584-597.

Feltz, W. F., H. B. Howell, R. O. Knuteson, H. M. Woolf, D. D. Turner, R. Mahon, T. D. Halther, and W. L. Smith, 2005: Retrieving temperature and moisture profiles from AERI radiance observations: AERIPROF value-added product technical description. DOE-ARM Tech. Rep. TR-066, 41 pp. [Available online at http://www.arm.gov/publications/tech_reports/arm-tr-066.pdf.]

Ferrare, R. A., and coauthors, 2006: Evaluation of daytime measurements of aerosols and water vapor made by an operational Raman lidar over the Southern Great Plains. *J. Geophys. Res.*, **111**, doi:10.1029/2005JD005836.

Fox, N. I., and A. J. Illingworth, 1997: The retrieval of stratocumulus cloud properties by ground-based cloud radar. *J. Appl. Meteor.*, **36**, 485-492.

Gardiner, B. A., and D. P. Rogers, 1987: On mixing processes in continental cumulus clouds. *J. Atmos. Sci.*, **44**, 250-259.

Grabowski, W. W., 1993: Cumulus entrainment, fine-scale mixing, and buoyancy reversal. *Quart. J. Roy. Meteor. Soc.*, **119**, 935-956.

Jonas, P. R. and B. J. Mason, 1982: Entrainment and the droplet spectrum in cumulus clouds. *Quart. J. Roy. Meteor. Soc.*, **108**, 857-869.

Kain, J. M., 2004: The Kain-Fritsch convective parameterization: an update. *J. Appl. Meteor.*, **43**, 170-181

Kerstein, A. R., 1989: Linear-eddy modeling of turbulent transport. Part II: Application to shear layer mixing. *Combust. Flame*, **75**, 397-413.

Kerstein, A. R., 1991: Linear-eddy modeling of turbulent transport. Part 6: Microstructure of diffusive scalar mixing fields. *J. Fluid Mech.*, **216**, 411-435.

Kitchen, M., and S. J. Caughey, 1981: Tethered-balloon observations of the structure of small cumulus clouds. *Quart. J. Roy. Meteor. Soc.*, **107**, 853-874.

Knuteson, R. O., and coauthors, 2004a: Atmospheric Emitted Radiance Interferomter. Part I: Instrument design. *J. Atmos. Oceanic Technol.*, **21**, 1763-1776.

Knuteson, R. O., and coauthors, 2004b: Atmospheric Emitted Radiance Interferomter. Part II: Instrument performance. *J. Atmos. Oceanic Technol.*, **21**, 1777-1789.

Krueger, S. K., 1993: Linear eddy modeling of entrainment and mixing in stratus clouds. *J. Atmos. Sci.*, **50**, 3078-3090.

Krueger., S. K., C.-W. Su, and P. A. McMurtry, 1997: Modeling entrainment and fine-scale mixing in cumulus clouds. *J. Atmos. Sci*, **54**, 2697-2712.

Krueger, S. K., P. J. Lehr, and C.-W. Su, 2006: How entrainment and mixing scenarios affect droplet spectra in cumulus clouds. *Proc. 12th Conf. Cloud Physics.* Amer. Meteor. Soc., Madison, WI.

Lasher-Trapp, S. G., W. A. Cooper, and A. M. Blyth, 2005: Broadening of droplet size distributions from entrainment and mixing in a cumulus clouds. *Quart. J. Roy. Meteor. Soc.*, **131**, 195-220.

Lazarus, S. M., S. K. Krueger, G. G. Mace, 2000: A cloud climatology of the Southern Great Plains ARM CART. *J. Climate*, **13**, 1762-1775.

Löhnert, U., D. D. Turner, and S. Crewell, 2009: Ground-based temperature and humidity profiling using spectral infrared and microwave observations. Part I: Simulated retrieval performance in clear-sky conditions. *J. Appl. Meteor. C.*, **48**, 1017-1032.

Miles, N. L., J. Verlinde, and E. E. Clothiaux, 2000: Cloud droplet size distributions in low-level stratiform clouds. *J. Atmos. Sci.*, **57**, 295-311

Moran, K. P., B. E. Martner, M. J. Post, R. A. Kropfil, D. C. Welsh, and K. B. Widener, 1998: An unattended cloud-profiling radar for use in climate research. *Bull. Amer. Meteor. Soc.*, **79**, 443-455.

Paluch, I. R., 1979: The entrainment mechanism in Colorado cumuli. *J. Atmos. Sci.*, **36**, 2467-2478.

Penney, C. M., and M. Lapp, 1976: Raman-scattering cross sections for water vapor. *J. Opt. Soc. Am.*, **66**, 422-425.

Pontikis, C. A., and E. Hicks, 1993: The influence of clear air entrainment on the droplet effective radius of warm maritime convective clouds. *J. Atmos. Sci.*, **50**, 2889-2900.

Raga. G. B., J. B. Jensen, M. B. Baker, 1990: Characteristics of cumulus bands off the coast of Hawaii. *J. Atmos. Sci.*, **47**, 338-355.

Raymond, D. J., and Wilkening, M. H., 1982: Flow and mixing in New Mexico mountain cumuli. *J. Atmos. Sci.*, **39**, 2211-2228.

Rodgers, C., 2000: *Inverse Methods for Atmospheric Sounding: Theory and Practice.* World Scientific, 238 pp.

Rogers, R. R. and M. K. Uay, 1989: *A short course in clud physics.* 3rd ed. Butterworth-Heinemann, 290 pp.

Sheridan, P. J., D. J. Delene, and J. A. Ogren, 2001: Four years of continuous surface aerosol measurements from the Department of Energy's Atmospheric Radiation Measurement Program Southern Great Plains Cloud and Radiation Testbed site. *J. Geophys. Res.*, **106**, 20735-20747.

Siebesma, A. P., 1998: Shallow cumulus convection. *Buoyant Convection in Geophysical Flows.* E. J. Plate, et al., Eds. Kluwer Academic, 441-486.

Siebsma, A. P., and J. W. M. Cuijpers, 1995: Evaluation of parametric assumptions for shallow cumulus convection. *J. Atmos. Sci.*, **52**, 650-666.

Skamarock, W. C., J. B. Klemp, and J. Dudhia, 2001: Prototypes for the WRF (Weather Research and Forecasting) model. Preprints, *Ninth Conf. on Mesoscale Processes*, Ft. Lauderdale, FL, Amer. Meteor. Soc., J11-J15.

Smith, W. L., W. F. Feltz, R. O. Knuteson, H. E. Revercomb, H. M. Woolf, and H. B. Howell, 1999: The retrieval of planetary boundary layer structure using ground-based infrared spectral radiance measurements. *J. Atmos. Oceanic Technol.*, **16**, 323-333.

Squires, P., 1958a: The spatial variation of liquid water and droplet concentration in cumuli. *Tellus.*, **10**, 256-261.

Squires, P., 1958b: Penetrative downdraughts in cumuli. *Tellus*, **10**, 381-389.

Squires, P. and J. S. Turner, 1962: An entraining jet model for cumulonimbus updraughts. *Tellus*, **14**, 422-434.

Stephens, G. L. 1978: Radiation profiles in extended water clouds. I: Theory. *J. Atmos. Sci.*, **35**, 2111-2122.

Stokes, G. M. and S. E. Schwartz, 1994: The Atmospheric Radiation Measurement (ARM) program: Programmatic background and design of the cloud and radiation test bed. *Bull. Amer. Meteor. Soc.*, **75**, 1201-1221.

Stommel, H., 1947: Entrainment of air into a cumulus clouds. *J. Meteor.*, **4**, 91-94.

Stull, R. B., 1985: A fair-weather cumulus cloud classification scheme for mixed-layer studies. *J. Appl. Meteor.*, **24**, 49-56.

Su, C.-W., S. K. Krueger, P. A. McMurtry, and P. H. Austin, 1998: Linear eddy modeling of droplet spectral evolution during entrainment and mixing in cumulus clouds. *Atmos. Res.*, **47-48**, 41-58.

Taylor, B. R. , and M. B. Baker, 1991: Entrainment and detrainment in cumulus clouds. *J. Atmos. Sci.*, **48**, 112-121.

Telford, J. W., 1975: Turbulence, entrainment, and mixing in cloud dynamics. *Pure Appl. Geophys.*, **113**, 1067-1084.

Telford, J. W., and S. K. Chai, 1980: A new aspect of condensation theory. *Pure Appl. Geophys.*, **118**, 720-742

Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Wea. Review.*, **177**, 1779-1800.

Turner, D. D., and J. E. M. Goldsmith, 1999: Twenty-four hour lidar water vapor measurements during the Atmospheric Radiation Measurement Program's 1996

and 1997 water vapor intensive observation periods. *J. Atmos. Oceanic Technol.*, **16**, 1062-1076.

Turner, D. D., and D. N. Whiteman, 2002: Remote Raman spectroscopy. Profiling water vapor and aerosols in the troposphere using Raman lidars. *Handbook of Vibrational Spectroscopy.* Vol 4. 2857-2878. J. M. Chalmers and P. R. Griffiths, Eds., Wiley and Sons.

Turner, D. D., 2005: Arctic mixed-phase cloud properties from AERI-lidar observations: Algorithm and results from SHEBA. *J. Appl. Meteor.*, **44**, 427-444.

Turner, D. D., R. O. Knuteson, H. E. Revercomb, C. Lo, and R. G. Dedecker, 2006: Noise reduction of Atmospheric Emitted Radiance Interferometer (AERI) observations using principal component analysis. *J. Atmos. Oceanic Technol.*, **23**, 1223-1238.

Turner, D. D., S. A. Clough, J. C. Liljegren, E. E> Clothiaux, K. Cady-Pereira, K. L. Gaustad, 2007: Retrieving liquid water path and precipitable water vapor from the Atmospheric Radiation Measurement (ARM) microwave radiometers. *IEEE Trans. Geosci. Remote Sens.*, **45**, 3860-3690.

Wagner, T. J., W. F. Feltz, and S. A. Ackerman, 2008: The temporal evolution of convective indices in storm-producing environments. *Wea. Forecasting*, **23**, 786-794.

Warner, J. 1955: The water content of cumuliform clouds, *Tellus*, **7**, 449-457

Warner J, 1969: The microstructure of cumulus cloud: Part I. General features of the droplet spectrum. *J. Atmos. Sci.*, **26**, 1049-1059.

Warner, J., 1970: On steady-state one-dimensional models of cumulus convection. *J. Atmos. Sci.*, **30**, 256-261.

Warner J., 1973: The microstructure of cumulus cloud: Part IV. The effect on the droplet spectrum of the mixing between cloud and environment. *J. Atmos. Sci.*, **30**, 256-261.

Warner J. and P. Squires, 1958: Liquid water content and the adiabatic model of cumulus development. *Tellus*, **10**, 390-394.

Widener, K. B. and K. Johnson. Millimeter Wave Cloud Radar (MMCR) Handbook. DOE-ARM Tech. Rep. ARM TR-081, 11 pp. Available online at http://www.arm.gov /publications/tech_reports/handbooks/mmcr_handbook.doc

World Meteorological Organization, 1975: *International Cloud Atlas.* WMO no. 407.

Wulfmeyer, V., S. Pal, D. D. Turner, and E. Wagner, 2010: Can water vapor Raman lidar resolve profiles of turbulent variables in the convective boundary layer? *Boundary Layer Meteor.*, **136**, 253-284.

Yang, H. J., G. McFarquhar, and H. Johnsson, 2011: Effect of aerosols on shallow cumuli sampled during RACORO. 2nd Ann. Science Team Meeting, Atmospheric Systems Research, San Antonio, TX.