

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

**Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600**

UMI[®]

UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

ENSEMBLE CLOUD MODEL APPLICATIONS

TO

FORECASTING THUNDERSTORMS

A Dissertation

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

Doctor of Philosophy

By

Kimberly Laurence Elmore

Norman, Oklahoma

2000

UMI Number: 9962975

UMI[®]

UMI Microform 9962975

Copyright 2000 by Bell & Howell Information and Learning Company.

**All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.**

**Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346**

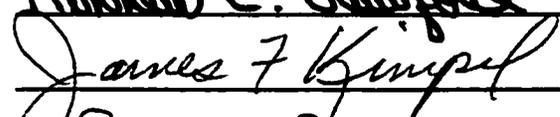
**© Copyright by Kimberly Laurence Elmore 2000
All Rights Reserved.**

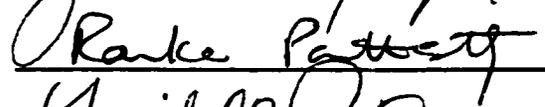
ENSEMBLE CLOUD MODEL APPLICATIONS
TO
FORECASTING THUNDERSTORMS

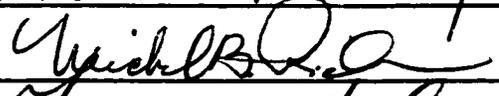
A Dissertation APPROVED FOR THE
SCHOOL OF METEOROLOGY

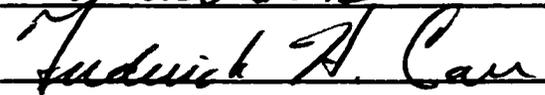
BY



Kenneth C. Crawford


James F. Kimpel


Ranke Patten


Michael B. O.


Frederick A. Carr

Acknowledgements

I could not have completed, much less begun, this work without help from many others. I thank my Ph.D. committee, and in particular, I thank my Committee Co-Chairs: Dr. Dave Stensrud and Dr. Ken Crawford. When Dr. Stensrud, of his own volition, visited my wife following our daughter's birth, my suspicions about his exemplary character were confirmed. He has graciously provided essential access to computing facilities, without which this work could not have proceeded. Dr. Stensrud craftily provided insight and dispensed wisdom by asking insightful questions. When I faltered, he directed me either with more questions, or by suggesting I investigate various references. He gently reminded me of what I already knew and recognized when I was genuinely dumbfounded. Dr. Stensrud has been, and continues to be, an exemplary scientist and his integrity is unmatched. Without his help, this entire topic would remain unknown to me.

Dr. Crawford has seen to it that my ideas and thoughts are stated clearly and cogently. I have found this experience to be both humbling and enlightening, which is exactly as it should be. Dr. Crawford has helped ease the inevitable bureaucratic machinations that mark progress through any Ph.D. program. During our first meeting, after I introduced my research topic, he said he felt his job would be, as much as possible, to clear away bureaucratic roadblocks. I shall forever cherish his closing words, which were: "You're too old to put up with a lot of crap." I now realize that he applies this maxim to his work, life, and students.

I wish to thank Dr. Mike Richman, who has enthusiastically served as a member of my Ph.D. committee. His patience is appreciated more than he knows. This work must, of necessity, depend heavily on statistical techniques. Dr. Richman always cheerfully made

time to listen to my concerns, answer my questions, and repeat explanations as many times as needed. Dr. Richman needed to repeat many explanations many times.

As should always be the case, my committee has been helpful and supportive, ready to be used to help me attain my goal. Dr. Carr coaxed me to perform verification analyses that proved very helpful. Seeing how easy it is to become dazzled by techniques, Dr. Kimpel gently reminded me to remain aware that this work is, first and foremost, about *meteorology*. Dr. Parthasarathy has always been helpful and supportive whenever we have met. His critical questions and overall good humor are deeply appreciated.

Partial support for this research was provided through a grant from the U. S. Weather Research Program of the National Atmospheric and Oceanic Administration. The National Severe Storms Laboratory, and in particular Mr. Mike Eilts, has graciously provided the balance of funding and support for this work. Being able to pursue a Ph.D. nearly full-time, with a minimum of additional job requirements or duties, has been a god-send. Mr. Eilts has lent encouragement at every step along the way, even though pursuing this Ph.D. has prevented me from performing other, needed duties for the Lab. I am deeply grateful for his patience and support.

Dr. Lou Wicker generously provided his numerical cloud model to me for this project. Conversations with Dr. Wicker have helped enhance immeasurably my insight and understanding into various numerical issues concerning cloud models and techniques.

Dr. Harold Brooks has also been a pivotal influence. He has helped me conceptualize much of what I needed to present within this work. His insight into cloud modeling has also been quite useful.

Dr. Charles Doswell has also been most supportive and helpful. While this topic is admittedly close to his heart, he has remained very approachable, kindly offering objective, insightful views whenever I have asked for them.

Mr. Phillip Spencer provided, and happily modified, his Near Storm Environment software, which is essential for extracting soundings from the AWIPS 212 gridded data. Much extra time would have been needed to create the capabilities he so blithely provided. He has also been, and continues to be, a good and patient friend.

Ms. Pam MacKeen has provided more help than she can imagine with her modifications to SCIT. Without her help, it would have been impossible to extract accurate storm lifetime estimates from the WSR-88D level II data. She is a valued friend, confidant, and sounding board. Her support and friendship are deeply valued.

Mr. Geoff Manikin was the sole source of the gridded Mesoeta 0000 UTC analyses, which do not normally leave the NCEP facility. He diligently chased down the data I needed, and made it freely available to me for this work.

Dr. Wes Wilson, of MIT/Lincoln Laboratory, has been most helpful with his mathematical insights. Our long, late-night conversations were extremely helpful. His candid assessments of some of my early work have saved me many public embarrassments.

I am fortunate in that I still have both of my parents, and they have been very supportive from the beginning. My father, in particular, has almost become more excited about finishing this entire process than I am.

Last, but by no means least, I wish to thank my wife, Dr. Pam Wilson. Her support, tolerance, suggestions, insight, understanding, and companionship, all lovingly given, made it possible to continue during the darkest times.

Table of Contents

Acknowledgements	iv
List of Abbreviations, Acronyms, and Symbols	ix
Abstract	xii
Chapter 1: Introduction	1
Chapter 2: Background	6
2.1 <i>Thunderstorm forecast applications</i>	6
2.2 <i>The current state of short-range thunderstorm forecasting</i>	10
2.3 <i>Cloud-scale numerical modeling</i>	14
2.4 <i>Cloud models used for forecasting</i>	19
2.5 <i>Numerical model ensembles</i>	24
Chapter 3: Methods and Techniques	32
3.1 <i>Introduction</i>	32
3.2 <i>Generation of ensemble forecasts</i>	33
3.3 <i>Ensemble output</i>	36
3.4 <i>Verification data</i>	39
3.5 <i>pdf estimates of storm lifetime</i>	42
3.6 <i>Comparing pdf's</i>	43
3.7 <i>Confidence intervals for pdf's</i>	47
3.8 <i>Dominant behavior modes</i>	48
3.9 <i>Contingency table analysis</i>	52
3.10 <i>Initial conditions verification</i>	55
Chapter 4: Results	59

4.1 <i>Days used in analysis</i>	59
4.2 <i>Initial condition verification</i>	60
4.3 <i>General results</i>	69
4.4 <i>Long-lived vs. short-lived storms</i>	76
4.5 <i>Specific cases</i>	86
4.5.1 <i>30 August 1996</i>	86
4.5.2 <i>14 July 1995</i>	91
4.5.3 <i>6/7 June 1996</i>	98
Chapter 5: Conclusions	107
Appendix 1: Kernel Density Estimation	116
Appendix 2: Euclidean Principal Component Analysis	118
A2.1 <i>Introduction</i>	118
A2.2 <i>Euclidean similarity</i>	121
A2.3 <i>Euclidean similarity PCA using least squares scores</i>	124
A2.4 <i>Examples</i>	128
A2.5 <i>Discussion and conclusions</i>	136
Appendix 3: Kolmogorov-Smirnov Goodness-of-Fit Test	138
References	140

List of Abbreviations, Acronyms, and Symbols

θ	Potential temperature (K)
AGL	Above Ground Level (m, km)
ARPS	Advance Regional Prediction System
ARTCC	Air Route Traffic Control Center
ASL	Achieved Significance Level
AWIPS	Advanced Weather Interactive Processor System
BRN	Bulk Richardson Number
CAPE	Convective Available Potential Energy (J kg^{-1})
CFCF	Central Flow Control Facility
cdf	Cumulative Distribution Function
COMMAS	Collaborative Model for Mesoscale Atmospheric Simulation.
CSI	Critical Success Index
dBZ_e	Decibels Relative to a Reflectivity of $1 \text{ mm}^6 \text{ m}^{-3}$ (for liquid water scatterers)
DFW	Dallas/Ft. Worth, TX, International Airport Identifier
ECMWF	European Centre for Medium Range Forecasting
FAA	Federal Aviation Administration
FAR	False Alarm Ratio
H₀	Null Hypothesis
hPa	Pressure in hectoPascals
ITWS	Integrated Terminal Weather System

K	Temperature in Kelvins (Absolute Scale)
KFWS	Dallas/Ft. Worth, TX, WSR-88D Radar Identifier
KNQA	Memphis, TN, WSR-88D Radar Identifier
KS	Kolmogorov-Smirnov
L₁	Minkowski L₁ Norm (Manhattan distance, km)
L₂	Minkowski L₂ Norm (Euclidean distance, km)
LI	Lifted Index (K)
MEM	Memphis, TN, International Airport Identifier
meso-α	Spatial Scales from 300-3000 km
meso-β	Spatial Scales from 30-300 km
meso-γ	Spatial Scale from 3-30 km
MDR	Manually Digitized Radar
MSL	Mean Seal Level
NWS	National Weather Service
NCEP	National Centers for Environmental Prediction
PCA	Principal Component Analysis
pdf	Probability Density Function
POD	Probability of Detection
PPP	Pooled Permutation Procedure
q	Mixing Ratio of Water Vapor to Dry Air (g kg⁻¹)
r	Pearson's Correlation Coefficient
RUC	Rapid Update Cycle
SCIT	Storm Cell Identification and Tracking (algorithm)

SREH	Storm Relative Helicity
STAS	Spatial Temporal Atmospheric Sampling
<i>t</i>	Time (s, min, hr)
TMU	Traffic Management Unit
TRACON	Terminal Radar Control
TSS	Trues Skill Score
<i>u</i>	East-West Wind Component, $\frac{\partial x}{\partial t}$, for <i>x</i> positive to the East (m s⁻¹)
UTC	Universal Coordinated Time (hr)
<i>v</i>	North-South Wind Component, $\frac{\partial y}{\partial t}$, for <i>y</i> positive to the North (m s⁻¹)
VCP	Volume Coverage Pattern
VOR	Very High Frequency Omnidirectional Range
<i>w</i>	Vertical Wind Component, $\frac{\partial z}{\partial t}$, for <i>z</i> positive upward (m s⁻¹)
WSR-88D	Weather Service Radar, 1988, Doppler

Abstract

A cloud model ensemble forecasting approach is developed to create forecasts which describe the range and distribution of thunderstorm lifetimes that may be expected to occur on a particular day. Such forecasts are crucial for both anticipating severe weather and ensuring the smooth flow of air traffic at busy, hub airports. Storm lifetime is an important characteristic to examine because long-lasting storms tend to produce more significant weather, and have a greater impact on air traffic, than do storms with brief lifetimes.

Eighteen days distributed over two warm seasons are examined. Soundings valid at 1800 UTC, 2100 UTC and 0000 UTC, provided by the 0300 UTC run of the operational Mesoeta model from the National Centers for Environmental Prediction, are used to provide initial conditions for the cloud model ensemble. These soundings are from a 160×160 km square centered over the location of interest and are shown to represent a likely range of atmospheric states. A minimum threshold value for maximum vertical velocity within the cloud model domain is used to estimate storm lifetime. Forecast storm lifetimes are verified against observed storm lifetimes, as derived from the Storm Cell Identification and Tracking algorithm applied to WSR-88D radar data from the National Weather Service (NWS).

When kernel density estimates are applied to the pooled data set consisting of all 18 days, a vertical velocity threshold of 8 m s^{-1} results in a forecast probability density function (pdf) of storm lifetime which is closest to the observed pdf. Model results from all 18 days also reveal that the storm lifetime resulting from a given input sounding cannot be determined by analyzing the bulk sounding parameters, such as convective available

potential energy, bulk Richardson number (BRN), BRN shear, or storm relative helicity. Standard 2 × 2 contingency statistics reveal that, under certain conditions, the ensemble model displays some skill locating where convection is most likely to occur. Contingency statistics also show that when storm lifetimes of at least 60 min are used as a proxy for severe weather, the ensemble shows considerable skill at identifying days that are likely to produce severe weather. Because the ensemble model appears to have skill in predicting the range and distribution of storm lifetimes on a daily basis, the forecast pdf of storm lifetime is used directly to create probabilistic forecasts of storm lifetime, given the current age of a storm. Such a product could furnish useful information to Air Traffic controllers by providing guidance about how soon a storm is likely to affect (or cease to affect) air traffic at a specific location. Similarly, this product could provide NWS forecasters with guidance about how likely it is that a particular cell will affect a given community.

Chapter 1: Introduction

On any day with thunderstorms, even the fabled, ubiquitous “casual observer” notices that thunderstorm types and their associated characteristics are remarkably variable over a small area, ranging from ordinary, short-lived cells (Byers and Braham 1949) to long-lived supercells (Browning 1964; Houze 1993). Days with longer lived storms also are often conducive to organized convective activity on the mesoscale (Emanuel 1994; Chappell 1986; Weisman and Klemp 1986). When viewed in the broadest sense, an accurate forecast of this highly variable behavior is a very difficult problem.

This variability in thunderstorm types and characteristics is well known by forecasters, but there is no way at present for them to know beforehand the range of this variability or even to know the dominant behavior modes for any given day (Johns and Hart 1993). This information is only known with certainty after thunderstorms have developed and the event is in progress. Without foreknowledge about the dominant behavior modes, it is difficult to anticipate even the most coarse thunderstorm characteristics. In addition, the convective mode can change, altering the severe weather threat. If this evolution is not anticipated, the outcome can be disastrous (Schwartz et al. 1990).

Storm initiation processes are presently under close scrutiny and constitute a forecast problem different from a storm’s behavior once initiated. However, once a storm forms, forecasting how it will behave during the immediate future is very important to the public and various industries. Forecasters currently use any number of tools and techniques to deal with thunderstorms. Numerous integrated parameters derived from either modeled or observed soundings, such as Convective Available Potential Energy (CAPE),

Lifted Index (LI), and Bulk Richardson Number (BRN), along with many others, are used to anticipate storm behavior once storms form.

Various ways of using these parameters have been developed. For example, Johns and Doswell (1992) outline a general methodology for severe storms forecasting. Another system, using BRN shear and storm relative helicity (SREH), aids forecasters in differentiating between bow-echo type storms and tornadic supercell storms (Stensrud et al. 1997). A formal decision tree system can be used to diagnose the likelihood of thunderstorms and whether or not those storms will be severe, supercell or tornadic (Mills and Colquhoun 1998). Some parameters are driven by the thermodynamic structure of the atmosphere, others by only the kinematic structure of the atmosphere, and a few are combinations. While some parameters are applicable only in certain geographic regions, all share the drawback that they provide vague, qualitative guidance for the dominant convective mode.

Similarly, various attempts have been made to use radar reflectivity data to quantitatively forecast future storm behavior (Battan 1953; Wilson 1966; Dixon and Wiener 1993; Henry 1993; MacKeen et al. 1999). Most systems are based, in one way or another, on persistence. For non-supercell thunderstorms, persistence works well for time scales of only a few minutes. Once recognized as such, quasi-steady state supercell thunderstorms are, by definition, more amenable to persistence-based forecasts. Unfortunately, parameters derived from radar reflectivity are strongly affected by inhomogeneous, range-dependent sampling scales and, to a lesser extent, by range folding and calibration problems. Parameters derived from single Doppler radar velocities suffer from the same drawbacks as reflectivity-derived parameters along with velocity folding problems (Wood and Brown

1997). Since only the radial velocity component is detected, inherent ambiguities are presented by a single Doppler velocity field. In the final analysis, none of the cell-specific, trend-based systems perform particularly well because radar detects only the results of nonlinear processes integrated over an indeterminate, preceding period.

Cloud or storm-scale numerical weather prediction seems a more refined way to quantitatively forecast storm behavior. Since the 1950's, numerical weather prediction has become integral to large-scale weather forecasting and, since the 1970's, computing capability has allowed ever more complicated, higher resolution numerical models to be developed and implemented. However, storm-scale modeling, as well as mesoscale modeling, face fundamentally different problems than does modeling on larger scales. Large-scale motions are better understood than are those on smaller scales, and the theory used to deal with large scale processes is generally unified within concepts comprising, for example, baroclinic instability and quasigeostrophic theory. Similarly elegant and unified theories do not exist for either the mesoscale or the cloud scale.

Storm-scale models have been developed primarily as research tools, to help understand dynamical processes that lead to certain, noteworthy storm behaviors, such as, the dynamic processes that causes storms to split into left- and right-moving cells (Klemp and Wilhelmson 1978). Only during the past few years have storm-scale models been used to provide forecast guidance (Kopp and Orville 1994; Brooks et al. 1993; Wicker et al. 1997; Wang et al. 1996; Carpenter et al. 1997; Carpenter et al. 1998). With occasional exceptions, it remains unclear whether these applications have been wholly successful.

Using storm-scale models to produce forecast guidance has various problems, including forecasting the environment, initializing the model, the model itself, interpreting

the model output, and communicating that output to others. Problems also exist with the inherent variability in storm type and behavior over a small region. When a storm-scale model is used to forecast one individual thunderstorm, to which particular observed thunderstorm does it refer? Since observed storm behavior is so variable, how is a single modeled realization verified? When more sophisticated modeling is used, wherein numerous individual storms are forecast within a mesoscale time-varying environment, how literally should the results be interpreted? And if they are not interpreted literally, how are error bounds placed on these results? Unfortunately, these sticky questions are deferred to forecasters who must subjectively formulate their own interpretations. The important point is that deterministic forecasts do not explicitly nor objectively provide guidance on the possible forecast errors.

Deterministic forecasts ignore the results of Lorenz (1963), where he convincingly suggests that even though the atmosphere is intrinsically deterministic, it comprises a chaotic system. While the atmosphere is strictly bound by physical laws and processes, its inherent deterministic nature cannot be realized. Even if a perfect model of the atmosphere became available, the model would display sensitive dependence to its initial conditions. Hence, any error or inaccuracy in the initial conditions, no matter how small, ultimately results in a forecast that diverges from the actual evolution of the atmosphere. At best, then, a single numerical model run provides guidance which represents only one of a multitude of states the atmosphere might attain. To remain within the framework of Lorenz's work, no single model run should be literally interpreted as a forecast.

If a model run provides only one possible state, then how likely is the atmosphere to attain this state? What is the range of other, reasonable states it might also attain? While

no single numerical weather prediction model provides this information, a framework exists for dealing with this inherent indeterminacy (Epstein 1969; Leith 1974). A carefully crafted ensemble of numerical models, each initialized with a different set of initial conditions that are all considered reasonable for the particular situation, should be able to provide the range as well as the distribution of possible future states of the atmosphere.

Therefore, instead of running a single cloud scale model, it is proposed as a hypothesis that an ensemble consisting of many cloud model runs can be created to provide reliable guidance concerning the nature of the individual thunderstorms which develop. In a Monte Carlo fashion, each run is started with different initial conditions. It is desired that the set of various initial conditions span those that can reasonably be expected. These initial conditions are derived directly from an operational mesoscale model over a pre-defined area of interest. Using a mesoscale model in this manner sets a bound, which is unique to each day, on the range of initial conditions. This dissertation has as its unique objective to determine whether or not output from a cloud-model ensemble forecast can predict the range and distribution of thunderstorm lifetimes for a given day. The ensemble output is compared against various environmental parameters extracted from the mesoscale model, to determine whether or not the ensemble provides added value beyond what is otherwise available.

Chapter 2: Background

2.1 Thunderstorm forecast applications

The potential for severe weather is of maximum concern when thunderstorms are forecast. Such concern is especially appropriate for the general public, because severe weather preparedness significantly reduces the injuries and fatalities that often accompany severe weather. In some cases, various indices are used to anticipate severe weather, but their use can lull forecasters into false senses of security, especially knowing that weak indicators of severe weather do not necessarily translate into a lack of severe weather (Stensrud et al. 1997). The application of an ensemble cloud model helps bound the range of possible storm behaviors, and thus, reduces the likelihood that a severe weather event catches a forecaster (and hence, the public) by surprise (Brooks et al. 1992).

For weather-sensitive industries, an operational distinction does not exist between severe and non-severe thunderstorms. A prime example is aviation, and in particular the airline industry. While Federal Aviation Administration (FAA) regulations do not explicitly prohibit flight through thunderstorms, all major airlines independently impose such regulations. The obvious reason is passenger safety. Aside from damage due to airframe overstress that can occur in strong convection, relatively weak convection possesses turbulence that can displace objects in the cabin and possibly injure passengers. Aside from injury, overall passenger comfort is a significant concern (R. Bevington, United Airlines, personal communication). Finally, thunderstorm development and evolution represents an additional concern for airlines, and the FAA, due to the routing of air traffic into major hub airports.

To understand how weather impacts airline operations requires an understanding of how airlines routinely operate. Most major airline carriers in the United States utilize hub airports, which reduces the infrastructure needed at each possible destination. This operational strategy imposes very high capacity demands on a few large airports. Recognizing that some bottlenecks in capacity can be anticipated, the FAA maintains two facilities for traffic management: the Central Flow Control Facility (CFCF) in Washington, D.C., and the local Traffic Management Unit (TMU) contained within each Air Route Traffic Control Center (ARTCC). The CFCF makes strategic, long-term planning decisions while the TMU makes short-term, tactical decisions in response to weather. Each morning, the CFCF develops an overall strategic plan for dealing with anticipated capacity limits along particular routes and at hub airports, by considering all known limitations to system capacity, including weather. Flights are planned, rerouted and occasionally canceled based on decisions made by both the CFCF and the TMU's. Adverse weather at a single hub airport can easily disrupt airline travel on a national scale.

All major airports utilize similar arrival and departure geometries (Fig. 2.1). Departing flights typically exit the airspace by utilizing a broad area aligned in the cardinal directions. Arriving flights are directed into the immediate airspace over one of four possible corner posts, called arrival gates, located about 60 km from the airport. At roughly 3000 m above ground level (AGL), aircraft are directed through these gates by the Terminal Radar Control (TRACON) Facility. Each arrival gate is situated over a Very High Frequency Omnidirectional Range (VOR) Station and arrivals must fly directly over these VOR's. This requirement implies that aircraft traverse the arrival gates in single file. Separation is maintained between flights, not only to avoid collisions, but also to avoid wake

turbulence. For jet transport aircraft, this separation is typically 8 to 10 km in weather that is good enough to support visual approaches. Larger separations are required during instrument meteorological conditions, when aircraft cannot see each other. At regular intervals during the day, which are unique to each hub airport, the hub airports operate at maximum “good-weather” capacity, such that arrivals are directed through all four gates at the minimum allowable aircraft separation.

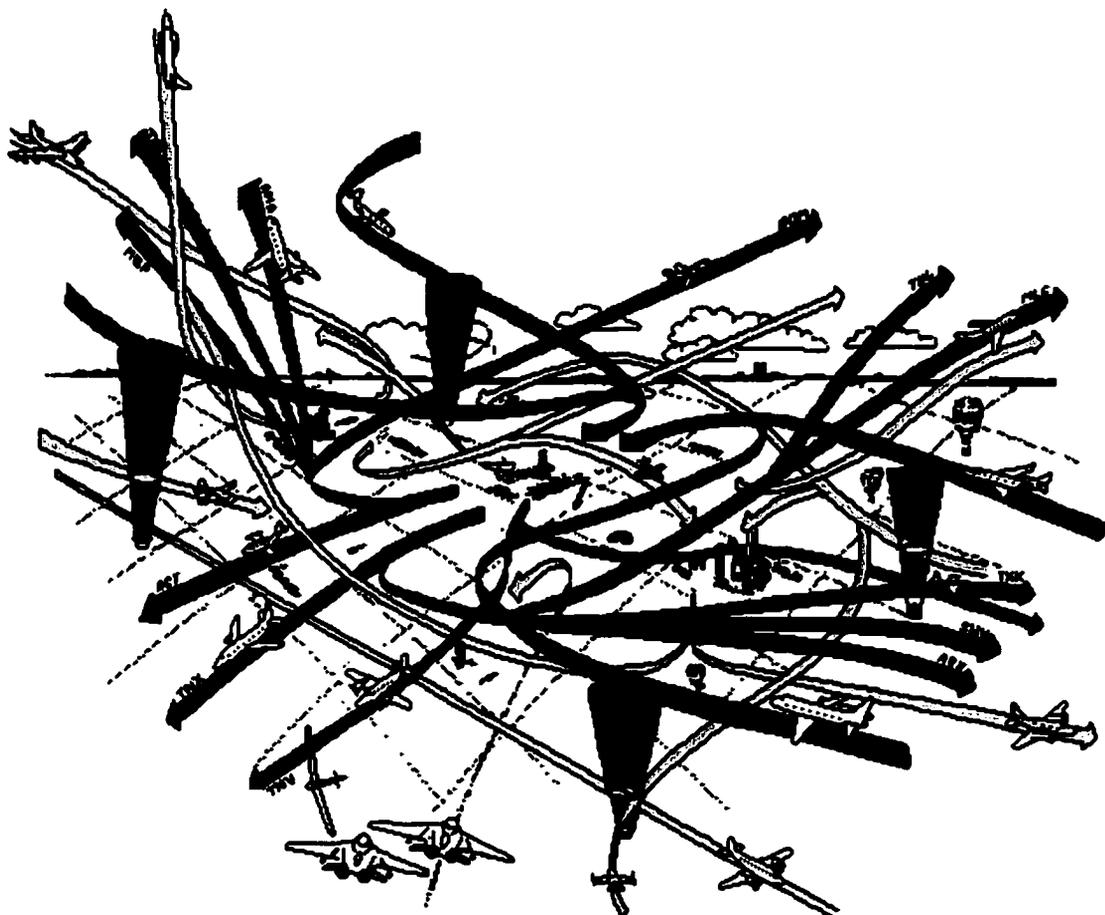


FIGURE 2.1. Schematic of Dallas/Ft. Worth terminal area traffic flow. The airport is near the center of the figure. Blue paths are for arriving flights, red paths are for departing flights, and yellow paths are for flights transitioning the terminal area to other destinations. The green cones show the radionavigation fixes that define each arrival gate. (J. Plummer, <http://web2.airmail.net/chuck/Flows.html> 1996)

Jet transport aircraft tend to avoid flight through radar echoes with intensities greater than 40 dBZ_e (Rhoda and Pawlak 1999). Consequently, if a thunderstorm with reflectivity above 40 dBZ_e is situated over an arrival gate, that gate becomes unavailable,

TABLE 2.1

Nature of Problem	Best Estimate of Costs
Diversion of Domestic Flights	\$5K/diversion
Missed Connections	\$150/passenger
Flight Cancellations	\$5K/cancellation
Delay (airborne, gate, and taxi-way combined)	\$30/minute
Value of Passenger Time (Aggravation Factor)	\$41/passenger hr
Downstream Delay factor	1.8

TABLE 2.1. Economic costs of weather-related operational issues. Values represent best estimates based on limited discussions with airline personnel, some of which occurred in 1994 at the conclusion of field demonstrations for the FAA's ITWS. Downstream delay factor is based on the mean downstream delay experienced by an air carrier during the remainder of its flying day, which is approximately 80% of the initial delay. Alternatively, the net delay for an aircraft over its flying day, due to an initial delay, is approximately 1.8 times the initial delay (from Stevenson 1997).

limiting the airport to 75% of its available capacity. Since arriving flights are positioned for a particular gate about 200 km from the airport, last minute diversions to different gates are not only costly, due to increased maneuvering and flight time, but also inconvenient to a wide range of individuals and companies. The difficulties of last-minute flight path changes are compounded because diverted flights must enter a traffic flow through an arrival gate that had already been planned and established. The costs involved with the operational problems created by various weather events are significant (Table 2.1; Stevenson 1997). Rerouting flights through a different arrival gate can result in late arrivals and missed flight connections. Should inclement weather reduce the actual airport capacity to something less than the demanded capacity, significant delays rapidly ensue causing further delays at other airports well removed from the weather-impacted one (Evans 1997).

Weather-induced problems continue long after the adverse conditions have dissipated. Because of these delays, other aircraft are held at their departing airports to avoid

diversions and in-flight holds. These latter flights are held until the weather feature causing the delay has ended. Such delays result in a lengthy period when the hub capacity is available but unused, because no aircraft are en route to the impacted gate before the weather threat has ended. The FAA recognizes this shortcoming in current operational procedures. As a result, the FAA is funding the Integrated Terminal Weather System (ITWS), a program charged with using weather information to make airport operations as efficient and safe as possible (Evans 1997; Wolfson 1997). Given some knowledge of when convective weather might be expected to end, flights could be en route so as to arrive as soon as the impacted arrival gate becomes available.

If thunderstorms result in delays, probabilistic forecasts of the thunderstorms' lifetime permit airline dispatch operations to use cost-loss analyses to determine the most economical strategy to deal with the delays. Probabilistic forecasts also may allow the TRACON facility to plan for the optimal arrival and departure of flights. The ARTCCs would benefit as well, as they could use available airways in an optimal fashion, even though some airways are being affected by convective activity. In turn, TMUs could plan for optimal traffic flow, while the CFCF could optimize national-scale strategic plans.

2.2 The current state of short-range thunderstorm forecasting

Even though the predictable time scale for thunderstorms is likely to be small, it is clear that certain industries, as well as the general public, benefit from improved short-range forecasts of storm behavior. With abundant radar data digitally available in real time, a logical approach to short-range forecasting is to use radar-derived time series of storm characteristics to predict storm behavior. Longevity studies, based upon digital

radar data, can examine how storm lifetimes are related to various other storm characteristics, such as size, echo intensity and echo top (e.g., Battan 1953; Wilson 1966; Henry 1993; MacKeen et al. 1999). These characteristics, a subset of the information used to make short range thunderstorm forecasts, have lead to new techniques for “thunderstorm nowcasting” (Wilson et al. 1998).

Thunderstorm nowcasting, and thunderstorm forecasting in general, challenge meteorologists daily. The basic characteristics of thunderstorms as observed by radar were documented by Battan (1952), who discovered how longer-lived cells are associated with the highest radar echo tops. Single-celled storms, that do not merge with other cells during their lifetime, typically last 20 min, but tend to last longer as their horizontal extent increases, up to 8 km in diameter (Battan 1952, 1953). There is no association between diameter and lifetime for storms with diameters greater than 8 km. These historical works demonstrated conclusively how radar may serve as a useful tool to predict thunderstorm evolution.

Nowcasts of thunderstorms generally are extrapolations made using two approaches: steady state assumptions (where time derivatives are assumed to be zero) and intensity/size trending (where time derivatives are a linear combination of past values and assumed to be constant; Wilson et al. 1998). Cross-correlation tracking, (Wilson 1966) is based upon steady state assumptions, and results are primarily dependent on the horizontal scale of the precipitating area. The steady state assumption is at the heart of the WSR-88D algorithm used by the National Weather Service (NWS), an algorithm that identifies and tracks storms. It is derived from the algorithm for Storm Cell Identification and Tracking (SCIT; Johnson et al. 1998) developed at the National Severe Storms Laboratory (NSSL).

The SCIT algorithm also uses linear intensity/size trending to extrapolate storm behavior. Unfortunately, these techniques perform poorly when used for forecasting storm characteristics (Henry 1993; MacKeen et al. 1999).

For example, a trend-based extrapolation tracker (TTAN; Dixon and Wiener 1993) is used to examine how storm size and storm volume help forecast the duration of an active storm (i.e., the “remaining storm lifetime,” Henry 1993). Henry defines storms as areas with reflectivity exceeding 35 dBZ_e and volumes exceeding 50 km³. Using data that are evaluated every 30 min, storms are divided into two categories: 1) complex storms that merge or split over their lifetime, and 2) simple storms that do not merge or split over their lifetime. Storms deemed supercells are excluded from Henry’s study. For simple storms, 85% are short-lived and dissipate in less than 30 min while only 12% of the complex storms dissipate in 30 min. Henry concludes that only large, intense, simple storms with volumes greater than 400 km³ and reflectivity greater than 55 dBZ_e, have a mean remaining storm lifetime greater than 30 min. Correlations between remaining storm lifetime and storm volume, and remaining storm lifetime and maximum storm reflectivity are weak. When the Pearson correlation coefficient (r) is used, the remaining storm lifetime is correlated with volume and reflectivity at levels ranging from $r = 0.39$ to $r = 0.52$. Hence, volume and reflectivity together explain, at most, only 25% of the total variance. Empirical forecasts of storm evolution fair no better. Even well-trained human forecasters who use research-quality Doppler radar to identify and track boundaries, and high resolution visual satellite images to characterize boundary layer cloud characteristics, have significant problems forecasting storm initiation and placement (Wilson and Mueller 1993).

In a similar study, MacKeen et al. (1999) examine 16 different radar-derived storm characteristics. These characteristics are: maximum reflectivity, vertically integrated liquid (VIL), cell volume, cell mass, maximum cell area at any height, height of storm top, height of storm base, maximum height of the 40 dBZ_e core, minimum height of the 40 dBZ_e core, probability of hail, probability of severe hail, maximum hail size, height of maximum reflectivity, height of center of mass, core aspect ratio (depth/width), and reflectivity ratio, defined as the maximum reflectivity/reflectivity at lowest tilt (Johnson et al. 1998; Witt et al. 1998). Remaining storm lifetime is examined by applying multivariate statistics to these 16 characteristics. To be declared a cell by the SCIT algorithm, the maximum reflectivity must be at least 40 dBZ_e and the cell must be wholly contained within an annulus between 30 km and 125 km from the radar. The temporal data sampling interval is 6 min (5 min) for Volume Coverage Pattern (VCP) 21 (VCP-11). A shorter temporal sampling interval means that MacKeen et al. better resolves short-lived storms than does Henry (1993). Storms that last less than two volume scans are ignored thus defining a 12 min (10 min) lower limit for storm lifetimes with VCP-21 (VCP-11). MacKeen et al. show that the best single-characteristic correlation with remaining storm lifetime is maximum reflectivity ($r = 0.36$). When all 16 characteristics are combined in a multiple, least-squares, linear regression the correlation improves to $r = 0.43$. Thus, with all 16 included, only 18.5% of the total variance is explained. All correlation coefficients are significant at the 99% level because of the large data set size. MacKeen et al. conclude that "... relationships between storm characteristics and remaining lifetime are not large enough to discriminate between short and long-lived storms." Thus, the resulting forecast guidance for both the public and industry is unacceptable.

2.3 Cloud-scale numerical modeling

Since radar extrapolation and trending performance leaves much to be desired, thunderstorm forecasting using cloud-scale numerical modeling is a next, logical step. Cloud scale models, notorious for the computational resources they consume, have seen little application to forecasting on the storm scale. Instead, they have been used primarily for simulation purposes, as an aid to understand how storms, particularly supercell storms, are organized and maintained. Because cloud models have proven useful for simulations, and because they can reproduce storm structures with remarkable fidelity, it is reasonable to investigate using them as a forecasting tool.

The first cloud-scale models used the anelastic mass continuity equation so all variables could be described by elliptical equations. Elliptical equations are convenient because elliptical solvers are very fast, and because acoustic waves are eliminated (Schlesinger 1975). Nevertheless, elliptical solvers become ill-conditioned when model grids become complex, e.g., grids on complex terrain, vertically stretched grids, and nested grids. Furthermore, because elliptical solvers are iterative, the computing resources needed to run cloud models are not known beforehand. These numerical difficulties led to the development of a more flexible form of the three-dimensional cloud model that uses the fully compressible mass continuity equation (Klemp and Wilhelmson 1978).

The fully compressible continuity equation requires explicitly solving the associated prognostic equations. However, this procedure has an added bonus: the computational resources needed for running the model can be estimated beforehand. Even so, explicit solutions exact a price, because the fully compressible equations retain numerically significant, but physically uninteresting, acoustic waves that quickly destroy the

model solution. Computer resources are wasted by explicitly solving for sound waves because much shorter time steps are needed than for the physically significant modes. Hence, the acoustic mode is split away and solved implicitly over shorter time steps. In addition, the cloud-scale models that use the fully compressible equations add another degree of sophistication by parameterizing subgrid-scale turbulence. The eddy kinetic energy equation is introduced as an additional prognostic equation and solved along with the other three component equations of motion. This approach yields parameterized subgrid-scale turbulence which is solved in a fashion consistent with the other prognostic variables. Thus, cloud-scale models represent the state-of-the-art in cloud modeling, and almost all of these models are at least partially derived from the original model of Klemp and Wilhelmson (1978).

The Collaborative Model for Mesoscale Atmospheric Simulation cloud model is used in the present study (COMMAS; Wicker and Wilhelmson 1995). It is the same one used in Brooks et al. (1993) and Wicker et al. (1997) and is, in principle, identical to the Klemp and Wilhelmson (1978) model. However, the code has been rewritten and made more amenable to the modern computing environment. This rewritten code takes advantage of the increased memory and sophisticated operating systems available on modern machines. The COMMAS code also incorporates improved numerical schemes. Improvements include a better advection scheme that is positive definite for moisture variables, monotonic for potential temperature advection, and has small numerical diffusion and phase errors.

Just as understanding parameterization schemes in mesoscale models help evaluate and use mesoscale model output (Cortinas and Stensrud 1995), understanding parame-

terizations peculiar to cloud models is a prerequisite to appropriate use of their results. Unlike mesoscale and larger scale models, cloud models are initiated with an unbalanced thermal perturbation (warm bubble) which is almost certainly an unrealistic representation of how deep, moist convection is naturally initiated. Previous research investigated inhomogeneously heated boundary layers, but this method is computationally expensive and provides results indistinguishable from the warm bubble initiation method (F. Proctor, personal communication). An unbalanced initiation leads to a “spin up” period for the model, as the model seeks to restore its own internal balance. Model results are dismissed during this period, a practice common with any numerical model that begins with an unbalanced initiation. Cloud models also display sensitivity to the initial warm bubble characteristics (Fig. 2.2; McPherson 1991; Brooks 1992; Lilly 1990). Such sensitivity can be an issue for forecasting purposes because the bubble characteristics needed to initiate deep convection are dependent upon the initial sounding. While not well documented in the literature, it is generally felt that as long as initial bubbles are not too far in excess of the strength needed to initiate deep convection, the qualitative results e.g., supercell vs. multicell modes, are generally insensitive to the initial bubble (Wicker et al. 1997; F. Proctor, personal communication).

The dynamical aspects of cloud modeling are substantially complicated by precipitation processes. Cloud models must deal explicitly with these aspects, because moist processes are the energy source for deep convection. One method of handling moist processes is to ignore the ice phase altogether, and invoke a warm-rain scheme that autoconverts water vapor to liquid cloud droplets (Kessler 1969). Once the mixing ratio of the cloud water exceeds a selected threshold (typically 1 g kg^{-1}) at temperatures below 0°C ,

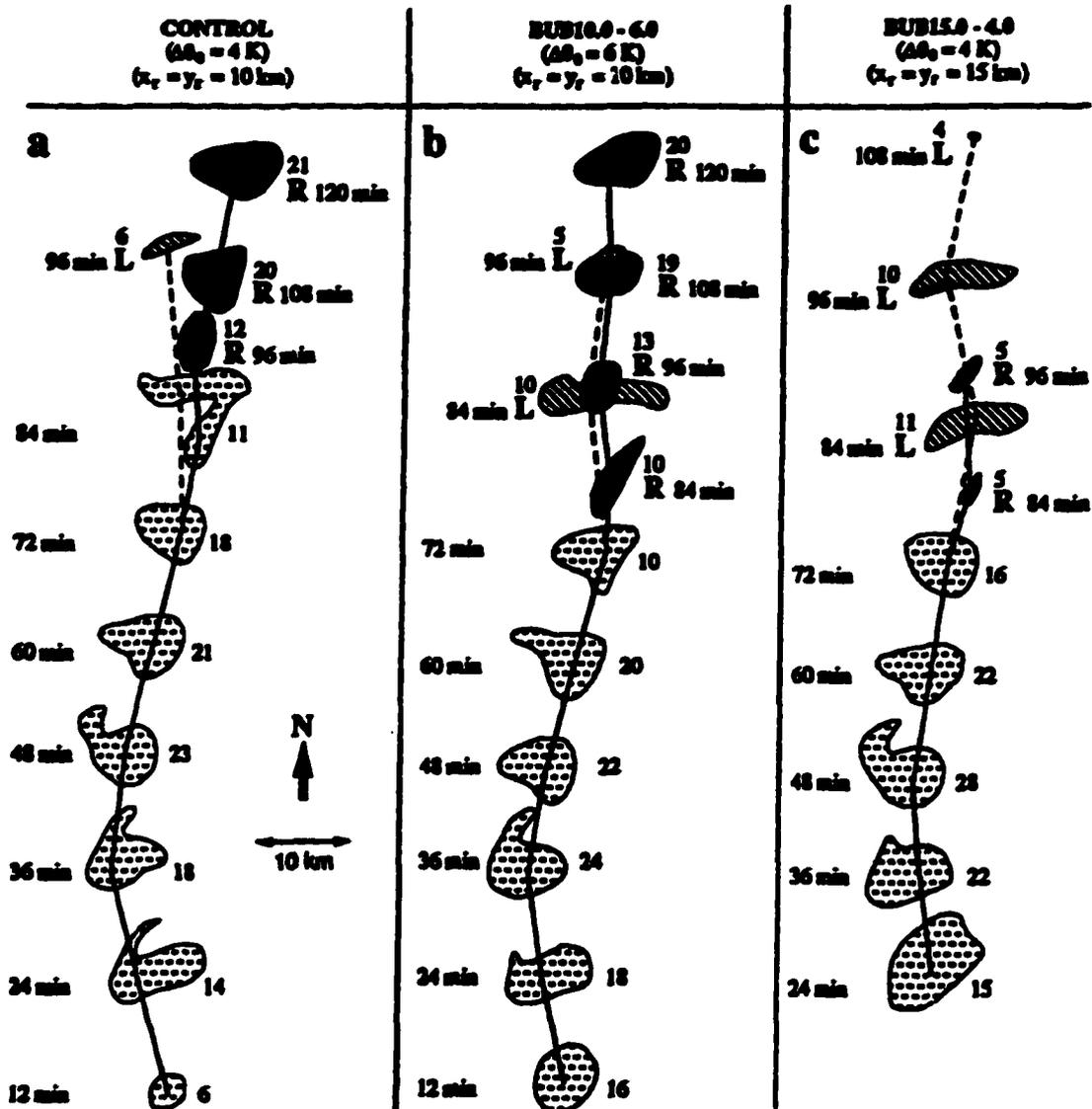


FIGURE 2.2. Evolution of updrafts from three different ARPS runs using identical soundings but different initial bubbles. All bubbles have a 1.4 km vertical radius. Panel a is a control run, using a bubble with 10 km horizontal radius and 4 K potential temperature excess. Panel b uses a bubble with a 10 km a 6 K potential temperature excess. Panel c uses a bubble with a 15 km and 4 K potential temperature excess. Stippled and hatched areas indicate updrafts $> 4\text{ m s}^{-1}$ at 4 km AGL. When a cell splits, the left and right moving cells are labeled with L and R, respectively. Times (in min) are times into the simulation. Numbers to the right of each area are the maximum updraft intensity at the associated time (from McPherson 1991).

these droplets autoconvert to raindrops with a predefined (Marshall and Palmer 1948) drop size distribution. In the Kessler scheme, latent heat release is accounted for in bulk by quantifying the latent heat of condensation of the water mass after it has been autoconverted to liquid. The Kessler scheme works as well as any other autoconversion scheme

(Houze 1993). Other methods explicitly address the water substance/ice phase process, and include: Proctor (1988), Tao et al. (1989), and Kopp and Orville (1994). While these latter approaches are more complete, they add considerable complexity and computational expense to the production of a cloud model forecast.

Unfortunately, autoconversion schemes ignore the latent heat transfer during phase changes that involve ice. This latent heat can be significant: the latent heat of freezing is about 13% of that for condensation, while the latent heat of deposition is 113% of that for condensation (Johnson et al., 1993). When explicit ice microphysics are invoked, the latent heat associated with ice processes is explicitly included in the cloud's overall heat budget. Since most precipitation from midlatitude thunderstorms begins as ice, significant thermal energy may be absent from model runs that do not consider ice (Rogers 1976, Johnson et al., 1993). This latent heat energy may be especially important when modeling or forecasting convection associated with low static stability and low relative humidity in the boundary layer (Wicker et al. 1997). In the extreme case, potential instability that depends upon latent heat release from ice processes is not realized. When convection depends upon latent heat from ice processes, autoconversion schemes suppress organized, deep convection.

Ignoring ice has other effects. Analyses show that 50% of the moisture entering the updraft of a squall line reaches the ground, 40% evaporates in downdrafts, and 10% is injected into the anvil portion of the cloud (Cotton and Anthes 1989). The Kessler scheme likely produces too much precipitation at the surface, and creates a precipitation efficiency that is too high, because it does not include ice processes. Because no water substance is converted to ice in a Kessler scheme, the excess precipitation produced can create cold

outflows too quickly that are too strong, leading to an anomalously short cell lifetime (Johnson et al., 1993). At the very least, the Kessler scheme by itself cannot account for 10% of water mass exiting the storm system through the anvil. Because ice processes produce less dense hydrometeors than do warm processes, the vertical distribution of precipitation loading is different between results obtained using the Kessler scheme versus results obtained using a full ice scheme. As a result, the maximum updraft strength and the overall storm structure are affected (Jewitt et al., 1990; Johnson et al., 1993, Straka and Rasmussen, 1998). Nevertheless, the characteristics that qualitatively define a supercell storm within a cloud scale simulation, e.g., a positive Pearson's r correlation between $w > 0$ and vertical vorticity (ζ) > 0 , are relatively insensitive to whether the microphysical processes are described by the Kessler scheme or an explicit ice scheme (Klemp and Wilhelmson 1978; Weisman and Klemp 1982; Weisman and Klemp 1984; Wicker and Wilhelmson 1995; Jewitt et al., 1990; Johnson et al., 1993; Straka and Rasmussen, 1998).

A final, operational consideration remains. Tests with the Advanced Regional Prediction System (ARPS) model show the execution time for runs that include ice parameterizations is 30% to 40% longer than those which use the Kessler scheme (Xue et al. 1995). The COMMAS model suffers similarly extended execution times when ice parameterizations are included (Wicker et al. 1997). Because execution time is an important consideration of ensemble model runs, only the Kessler scheme is used in this work.

2.4 Cloud models used for forecasting

Historically, running a cloud model in real time and then interpreting the output requires computational facilities and expertise that are not widely available. As greater computational capabilities have become more widespread, cloud scale models are run

more often. Yet, assessing and interpreting the real-time cloud model output remains problematic. For example, useful insight about atmospheric processes can be gleaned from synoptic and mesoscale forecasts by examining horizontal cross sections of model results. Because conceptual models are available at such scales (quasigeostrophic theory, for example), the kinematic structure and dynamic processes at work at any given time can be readily interpreted. Such conceptual frameworks do not exist for examining cloud scale model output. As a result, rarely are cloud models used in forecast applications.

The ARPS is used for real-time forecasting applications, though not as a cloud-scale model (Carpenter et al. 1997). When configured for real-time forecasts, the ARPS uses a 9 km horizontal grid spacing, augmented with an adaptively-nested 3 km grid. As a consequence of the grid spacing, the ARPS cannot accurately resolve meso- γ cloud-scale motions or characteristics; yet ARPS can capture many important meso- β characteristics. Background fields for model initialization are provided from the NCEP Rapid Update Cycle (RUC) model enhanced with additional data from the Oklahoma Mesonet, wind profilers, upper air data, and various WSR-88Ds. In the three cases presented by Carpenter et al., model resolution precludes storm-scale comparisons. Although similarities in meso- β scale features are apparent, spatial position errors on the order of one or two counties are common as are temporal errors of 1-2 hr. Clearly, in these three cases, the fact that convection occurs anywhere close to the actual location and time is a significant achievement.

An additional ARPS case (16-17 June 1997) is examined for sensitivity to initial conditions (Carpenter et al. 1998). Here, the model results are significantly altered and improved upon by using an initial cloud analysis. To test for further model sensitivity to initial conditions, an ensemble of seven ARPS runs is made on this case. The ensemble is

created using seven different analysis times, two of which use cloud analyses, similar to a lagged average forecast (Sivillo et al. 1997). A plot of the ensemble of 30 dBZ_e contours at 3 km mean sea level (MSL, not shown), is used to note where all the contours from all runs overlap. Where overlap exists is where ARPS is relatively insensitive to the initial conditions. Where overlap does not exist is where ARPS is sensitive to initial conditions. Because not all runs overlap everywhere, the ARPS model displays spatially and temporally variable sensitivity to initial conditions.

In a different experiment, a 2-D, slab-symmetric, cloud-scale model is applied during the North Dakota Thunderstorm Project. The model is used to help forecast when storms with hail or strong low level wind shear are likely. The model is also used to discriminate between convective and non-convective precipitation (Kopp and Orville 1994). The model uses a 200 × 200 m resolution across a 20 km × 20 km domain. The main advantage of using a 2-D model is that substantially less computational resources are required than for an equivalent 3-D model. However, 2-D models cannot reproduce storms that rotate, which are a significant operational concern. In this experiment, an area of interest is chosen and the observed morning sounding from the closest rawinsonde station is used to initiate the model. To produce convection, a slight amount of random variation in the boundary layer temperature field is imposed and the boundary layer is heated by a simple, time invariant, heat flux technique. Eventually, convection occurs. Additionally, a forecaster subjectively adds an estimated boundary layer convergence value as a model input parameter. A 2-D model applied in this fashion shows some skill, with the particular skill value depending upon the parameter being verified: precipitation type, precipitation occurrence, cloud type, or cloud top height. Even so, overall model performance

is rated mediocre while some forecasts are completely unsatisfactory due to inaccuracies and uncertainties in the initial environment.

In another example, a fully 3-D numerical cloud model is employed in the Storm Type Operational Research Model Test Including Predictability Evaluation (STORMTIPE) project, which is designed to forecast the gross characteristics of storms, such as life span and rotation (Brooks et al. 1993; hereafter ST-91; Wicker and Wilhelmson 1995). The ST-91 model uses 1500 m horizontal grid spacing and a vertically stretched grid. The lowest grid level is 400 m AGL, and the vertical grid spacing is 600 m near top of the model grid. Forecast soundings are created interactively by a human forecaster using observations, model forecast soundings, and the forecaster's best judgement. The resulting sounding is used to represent a horizontally homogeneous atmosphere as the model is initiated. Because the model boundaries are open, modeled storms cannot interact with other storms or with the mesoscale environment. Results indicate that the modeled storm can be quite sensitive to the initial environmental conditions. In one case, a change of only 1 K at 700 hPa makes the difference between no storm developing, and a supercell storm evolving. This sensitivity also is shown by Crook (1996), who finds that boundary layer temperature differences of ± 1 K and/or mixing ratio differences of ± 1 g kg⁻¹ make operationally significant differences in the modeled storms. Currently, the measurement accuracy for temperature and moisture is ± 1 K and ± 1 g kg⁻¹, respectively (Mueller et al. 1993). Consequently, even if the cloud model is perfect, the accuracy needed for initial conditions is not attainable (Crook 1996; Mueller et al. 1993; Brooks et al. 1992, 1993).

A similar experiment (STORMTIPE-95, hereafter ST-95) uses soundings from the Eta model, produced by the National Center for Environmental Prediction (NCEP). The goal is to help operational forecasters anticipate storm behavior (Wicker et al. 1997). A major difference between ST-95 and ST-91 is that no human forecaster is used to create an initial sounding. Rather, based on the 0000 UTC forecast from the 1200 UTC Eta run, a human forecaster extracts a sounding nearest to the location deemed most likely to produce severe convection. This sounding is used as input to the cloud model. Another difference between ST-91 and ST-95 is that ST-95 utilizes two grids: a coarse grid and an automatically generated, nested, adaptive high resolution grid that is placed over regions in the domain where convection develops. Both the coarse and nested grids use 35 vertical levels with 300 m vertical spacing near the surface and 700 m spacing near the top of the model domain. The coarse grid uses 4 km horizontal spacing while the nested grid uses 1.3 km horizontal spacing. Only six cases are examined. When the forecast sounding has convective indices similar to those observed, the model demonstrates skill in predicting the mode of convection. Thus, ST-95 produces results similar to ST-91. The Eta soundings typically forecast too much convective instability and wind shear to be present at 0000 UTC. As a result, forecast storms are too intense. Results from both ST-91 and ST-95 indicate that a single model run is insufficient for an accurate forecast of convective storms because of the model's sensitivity to variations on the mesoscale. To address the influence of environmental variability on convective modes, ST-95 proposes an experiment in which a Monte Carlo approach is used to produce a variety of initial conditions for the model (as in Brooks et al. 1992).

While the results outlined above appear promising, it is not clear that using single model runs produce accurate forecast guidance. Certainly, none of the approaches described provide explicit guidance about the range or variability that should be expected on any given day. Yet, verification analyses of modeling results reveal the existence of profound sensitivity to initial conditions. Verification issues aside, it is clear that, even if perfect cloud models existed, the required observational accuracy and density are insufficient for a deterministic forecast. Thus, the forecast problem is deterministically intractable. Because the goal is to create a forecast of thunderstorm behaviors over a particular region on a particular day, a Monte Carlo or ensemble approach is a logical approach to investigate. The output of such an ensemble helps create probability forecasts for gross convective characteristics, such as lifetime, motion, or even persistent cloud-scale mesocyclones. To date, no attempt has been made to use cloud models in an ensemble to capture the inherent variability of thunderstorms over small regions. Doing so requires multiple runs of a cloud-scale model initialized with a range of conditions considered likely for a given day (ideas explicitly mentioned in Brooks et al. 1992; Kopp and Orville 1994; Crook 1996; Wicker et al. 1997; and MacKeen et al. 1999).

2.5 Numerical model ensembles

Posing thunderstorm forecasts in terms of an ensemble explicitly recognizes the inherent uncertainty. In this framework, the question at hand is reduced to: What kind of thunderstorms are possible at a specific location today? This question is non-trivial because thunderstorms are sensitive to processes on scales smaller than can be operationally observed; furthermore, variations in parameters that are significant on the storm scale

are obscured by observational errors (Brooks et al. 1992; Crook 1996; Wilson et al. 1997). In addition, knowledge of thunderstorm processes is incomplete, so the governing equations used to describe them in any model are necessarily incomplete. These considerations are compounded by Lorenz's (1963, 1965, 1969, 1993) findings that the atmosphere constitutes a chaotic system with a finite, scale-dependent, predictable period. Accordingly, errors in initial conditions, regardless of how small, impose a limit on how far into the future a skillful forecast can be made. For example, large-scale motions have a longer predictable time scale than do small-scale motions (Lorenz 1969). This predictability limit holds even if the governing equations are known exactly.

Recognizing the significance of the results documented by Lorenz (1963, 1965), Epstein (1969) addresses predictability in a probabilistic sense using a stochastic dynamic forecasting technique. Atmospheric fluid motion is considered to be completely deterministic, even though it is patently impossible to observe the atmosphere in sufficient detail or with sufficient accuracy to realize this determinism.

Let the equations describing these motions (also considered to be deterministic and perfect) be written in vector form as:

$$\dot{\mathbf{X}} = G(\mathbf{X}(t)), \quad (2.1)$$

where the dot indicates the total derivative with respect to time, G is a nonlinear forecast operator, and \mathbf{X} is a multivariate, D -dimensional vector describing each state variable at each model grid point. The initial condition for \mathbf{X} is given by \mathbf{Y} , where \mathbf{Y} (typically) has a lower dimensionality than \mathbf{X} . In a deterministic system, an operation, e.g., objective analysis, is applied to \mathbf{Y} to yield a "best" first guess of \mathbf{X} at time $t = 0$. If the resulting first guess field is $\hat{\mathbf{X}}$, then

$$\hat{\mathbf{X}} = A(\mathbf{Y}) \quad (2.2)$$

where A is the analysis operator. Then, (2.1) is integrated over time to provide a forecast, \mathbf{X}_f at time $t = t_f$ and

$$\mathbf{X}_f - \hat{\mathbf{X}} = \int_{t_0}^{t_f} G(\mathbf{X}(t)) dt. \quad (2.3)$$

An alternative interpretation is that the analyzed observations ($\hat{\mathbf{X}}$) constitute a multivariate probability density function (pdf), $\varphi(\mathbf{X}; t_0)$, defined over the entire phase space. Here, φ must possess the properties defining a pdf which are: 1) $\varphi \geq 0$ for all \mathbf{x} and t , and 2) $\iint \dots \int \varphi(\mathbf{X}; t) dx_1 dx_2 \dots dx_D = 1$. Visualizing a multidimensional pdf is relatively easy up to three dimensions. For a one dimensional pdf, the integral constraint requires that the area beneath the pdf be unity. In two dimensions, the integral constraint requires that the volume beneath the pdf be unity and for three dimensions, the integral constraint requires the contained mass be unity. To use the alternative interpretation, the initial analysis (2.2) and the forecast (2.3) must be expressed probabilistically, each with their own pdf's. Even though it is not explicitly included in the formulations above, φ is affected by inaccuracies and parameterizations used in the model equations comprising the forecast function, G

When cast as probabilities, the integrand of (2.3) is expressed as

$$\frac{\partial \varphi}{\partial t} + \nabla \cdot (\dot{\mathbf{X}} \varphi) = 0, \quad (2.4)$$

which is commonly referred to as the Liouville equation (Ehrendorfer 1994). In Eq. 2.4, ∇ is taken over D dimensions. This equation is based on the principle that the total

amount of probability is conservative and equal to one, much like the principle that mass can be neither created nor destroyed. So, the probability of all parameters having some value is always one. With appropriate boundary conditions (such as $\phi \rightarrow 0$ as $\dot{\mathbf{X}} \rightarrow \pm\infty$) and the initial value $\phi(\dot{\mathbf{X}}; t_0)$, directly integrating (2.4) presents no intrinsic mathematical difficulties. If the atmosphere is viewed as a chaotic system displaying sensitive dependence to initial conditions ($\dot{\mathbf{X}}$), then the resulting integrated pdf becomes more and more diffuse with time (Lorenz, 1993). One interpretation is that, as $(t_f - t_0)$ increases, any single, deterministic forecast becomes less likely to be close to the true atmospheric state at t_f . However, using Epstein's approach, if D parameters are needed at L grid points, it is necessary to explicitly solve (2.4) for D^L terms at each time step, assuming no covariance exists between the individual variables. Covariances act to reduce the number of independent pdf's, making the effective number of independent parameters something less than D , depending on the value of the covariance. Clearly, numerical integration of (2.4) is intractable for a reasonable number of parameters on a grid of any appreciable size. Yet, with this approach, Epstein establishes the conceptual framework for ensemble forecasting.

Recognizing the computational problems posed by Epstein, Leith (1974) proposes the forecast pdf be discretized by using a finite Monte Carlo sample comprised of a set, or ensemble, of deterministic forecasts. Each ensemble member uses different, independent, identically distributed (equally likely) initial conditions. If $\mathbf{X}(t_f)$ consists of all ensemble members, then $\phi(\mathbf{X}; t)$ is approximated by a cloud of discrete points in D -dimensional phase space which changes shape and becomes more distorted and diffuse with time. The pdf $\phi(\mathbf{X}; t)$ is most easily visualized by viewing the pdf of a single variable as it changes

in time (Fig. 2.3). For the D -dimensional case, the overall mean position of the cloud yields the best estimate of the future state of the atmosphere at time t_f . After some time t_C , the cloud becomes essentially stationary with time representing the climate of the ensemble, thus indicating that all predictability (beyond t_C) has been lost.

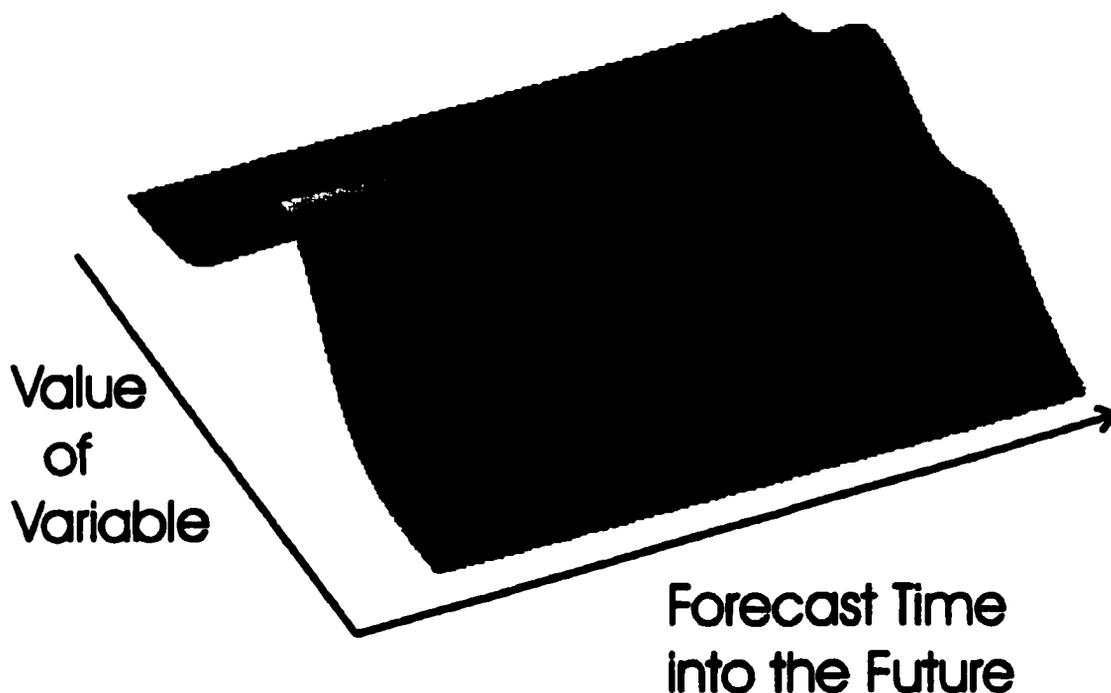


FIGURE 2.3. Hypothetical probability distribution for an arbitrary variable, such as temperature, at a point in space. Forecast valid time increases from left to right. Note that the width of the distribution generally increases as time increases. In this example, an initially Gaussian distribution ultimately becomes bimodal, which could occur due to uncertainty related to a frontal passage (from Sivillo et al. 1997).

With a perfect forecast model, using independent, identically distributed initial conditions results in independent, identically distributed forecasts. However, in numerical models, small-scale random perturbations are dissipated by processes such as filtering or internal gravity waves, and so don't affect the larger scales. Consequently, the forecasts are no longer independent, identically distributed because covariances are created between ensemble members. Hence, various techniques have been developed to ensure that the

resulting forecasts are most likely to span the range of non-zero $\varphi(\mathbf{X};t)$ with the fewest number of members.

Du et al. (1998) determine that as few as five members account for 65% of the improvement in the mean quantitative precipitation forecast compared to a single control forecast. However, reliable estimates of the ensemble forecast skill cannot be obtained with an ensemble this small. The ensemble prediction system used at European Centre for Medium Range Forecasting (ECMWF) uses up to thirty-two member ensembles. An ensemble that consists of thirty-two members provides a reliable bound on the control error, depending on the parameter considered, though the amount of improvement in the resultant forecast as the number of members is increased depends strongly upon the statistic used to measure the control error (Buizza and Palmer 1998).

Ensemble perturbations can be generated in at least two ways: unconditional or Monte Carlo methods, where the perturbations are not dynamically constrained, and dynamically conditioned methods, such as the breeding of growing modes (Kalnay and Toth, 1993) or singular vectors (Buizza, 1997). Both methods perturb the analysis fields, but one Monte Carlo-like method perturbs observations prior to the analysis step (Houtekamer and Derome, 1995). Dynamic conditioning methods do not necessarily produce good samples of the forecast probability distribution. Instead, these methods heavily sample the wings of the forecast probability distribution. Hence, dynamically-conditioned initial conditions result in ensemble members that span as large a phase space as possible in an attempt to use the spread as a predictor of forecast skill (Anderson 1996b). Discussion continues about the merits of dynamically conditioned methods. It is generally accepted that, at least for synoptic and hemispheric scale ensembles, perturbations that

lack dynamic conditioning produce less satisfactory results because many of the resulting perturbed modes are rapidly damped within the models.

The types of perturbation methods discussed highlight that the effort in ensemble development has been directed towards hemispheric and synoptic scale ensembles. Much less work has been accomplished regarding how to appropriately build perturbed initial conditions for models that operate on the mesoscale. So far, methods include varying the model parameterizations of physical processes (primarily precipitation) and Monte Carlo approaches (Du et al. 1997; Stensrud et al. 1999). Apparently, cloud models have not been considered ensemble candidates, even though current state-of-the-art cloud models have been subject to intense development for more than twenty years (e.g., Klemp and Wilhelmson 1978; Proctor 1988; 1989; Xue et al. 1995; Wicker and Wilhelmson 1995). This period is nearly as long as mesoscale numerical models have been under development. Since no work has addressed cloud model ensembles, how best to perturb initial conditions has not been determined. Certainly for cloud models, Monte Carlo perturbations are very attractive; yet, a straightforward method does not exist to condition cloud model perturbations dynamically.

The discussions of Brooks et al. (1992), where they speculate about how ensemble cloud models might be beneficial when used in operational forecasting, inspires the present research. Crook (1996), Wicker et al. (1997), and Straka and Rasmussen (1998) also speculate that ensemble cloud models may be the next logical step to take. While mesoscale ensemble forecasting studies are being pursued (Carpenter et al. 1998; Stensrud et al. 1999), none of them provide explicit insight into the range of behavior for individual storm cells. Since the scientific underpinnings of cloud models are apparently sound, and

computing advancements are sufficient for running many cloud models simultaneously, the time is right to determine whether or not cloud model ensembles can provide beneficial insight into convective storm characteristics.

Chapter 3: Methods and Techniques

3.1 Introduction

This chapter describes, in some detail, the statistical tools and techniques needed to examine the ensemble model output. Such tools are essential because ensemble models, and especially ensemble *cloud* models, produce Brobdingnagian quantities of output. Practical, and certainly operational, considerations dictate that the output data cannot be used effectively without some post-processing operations. The tools developed in this chapter also facilitate statistical tests that quantitatively compare the ensemble model output to observations. For example, one stated goal of this study is to produce a forecast of the range and distribution of thunderstorm lifetimes. Thus, a way to characterize the range and distribution of modeled lifetimes, which are defined by a pdf, is needed. Because there is no reason to believe that thunderstorm lifetimes conform to any particular theoretical distribution on a day-to-day basis, a method to estimate the parent pdf that produces the observed sample must be developed. Because the form or nature of the parent population distribution is unknown, the statistical techniques are non-parametric. The modeled and observed pdf's are always different, but are the differences *statistically* significant? To judge, quantitatively, how well the forecast and observed pdf's compare requires methods to statistically compare two different pdf estimates.

To use a methodology that can be generally applied, the behavior of cells produced by the model ensemble needs to be characterized. Significant modes (for example, long- vs. short-lived, strong updraft intensity vs. moderate updraft intensity) exist that characterize the general evolution of cells in the ensemble. Methods to extract such modes must be developed.

How well do the forecast soundings, which are used as the initial conditions for the ensemble model, compare to the best estimate of the true soundings? “Best estimate” is invoked because no observed soundings are available. If there are biases, are the biases statistically significant? Do the forecast soundings adequately capture the best estimate of the actual environmental variability? Statistical techniques to answer these, and other, related questions, are developed in this chapter. Hence, the final results can be quantitatively discussed without parallel discussions that contain details of the analysis technique.

3.2 Generation of ensemble forecasts

Ensemble techniques generally use a set of initial conditions created by perturbing a base state. However, the best procedure for perturbing a base state is unclear when thermodynamic and kinematic sounding variability is considered over a relatively small region where thunderstorms occur. Consequently, initial conditions for this ensemble are, instead, generated using the variability inherent in time and space over a given region. Initial conditions are provided by soundings extracted from the NCEP Mesoscale Eta forecasts (Black 1994). The Mesoeta is provided on the Advanced Weather Interactive Processing System (AWIPS) 212 grid. One sounding is used for each ensemble member. The AWIPS 212 grid has a 40 km horizontal grid spacing and a 25 hPa vertical grid spacing. Each ensemble member consists of a 2 h COMMAS cloud model run, initialized using a different Mesoeta model sounding. This ensemble work is centered on two ITWS testbed sites: Memphis, TN, (Fig. 3.1) and Dallas–Ft. Worth, TX (Fig. 3.2). Radar data from the associated WSR-88Ds (KNQA and KFWS, respectively) are used for verifica-

Memphis Ensemble Region and Landmarks

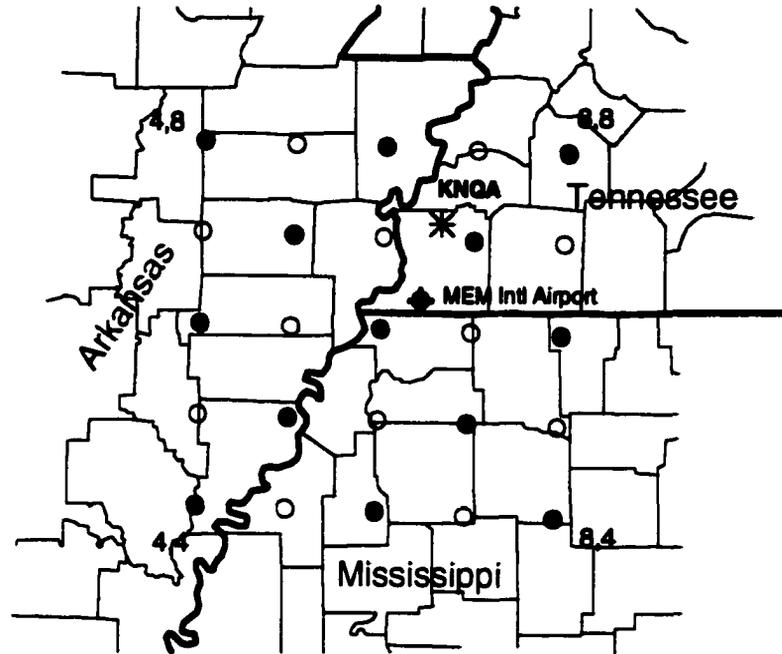


FIGURE 3.1. The Memphis area ensemble region with landmarks. Heavy blue lines are state boundaries, and lighter black lines outline counties. Filled spots are AWIPS 212 grid points from which Mesoeta soundings are extracted. Empty circles are AWIPS 212 grid points that are not used. The blue asterisk is the radar location, and the magenta symbol with four peripheral tics shows the airport location. Grid points are numbered by coordinate pairs, starting with 4,4 in the lower left, 4,8 in the upper left, 8,4 in the lower right and 8,8 in the upper right.

tion. These sites are convenient to use because MacKeen et al. (1999) provides a thorough analysis of convective cell lifetimes for the days examined.

All initial soundings are extracted from a 5×5 grid centered on the area of interest, and are derived from operational 0300 UTC Mesoeta model runs in 1995 and 1996. While forecast fields are available at three hour intervals from the Mesoeta, only the 15 h forecast (valid at 1800 UTC), the 18 h forecast (valid at 2100 UTC), and the 21 h forecast (valid at 0000 UTC) are used to create a daily ensemble. In one case (7 June 1996), the 21 h (0000 UTC), 24 h (0300 UTC) and 27 h (0600 UTC) forecasts are used. Soundings from three adjacent forecast times over the study domain implicitly recognizes temporal and

Dallas-Ft. Worth Ensemble Region and Landmarks

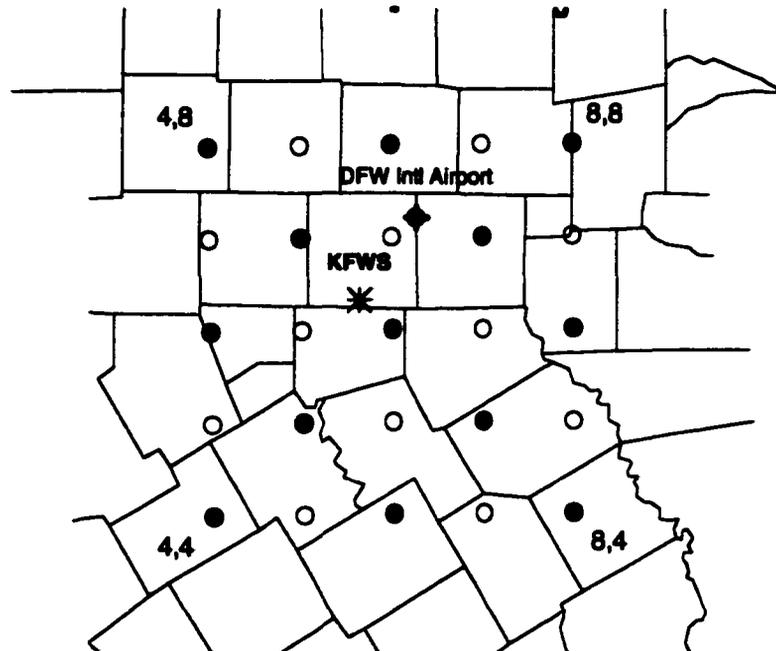


FIGURE 3.2. Same as Fig. 3.1, but for the Dallas/Ft. Worth ensemble region.

spatial uncertainties in the forecast soundings. Initial conditions of the ensemble created by this method are called Spatial Temporal Atmospheric Sampling (STAS) and are fundamentally different from perturbations imposed upon a base state.

At each forecast time, only every other sounding is extracted, so that 13 of the 25 available soundings are used. Because thirteen soundings are extracted from each of three forecast periods, each ensemble consists of thirty-nine separate cloud model runs or members. Consequently, the ensemble output is intended to be applicable across the entire region over a nine hour period. Because only cases with convection are used, an analysis that examines how useful this ensemble might be to determine unconditional probability of convection forecasts is precluded.

The ensemble is produced on a ten processor Silicon Graphics Power Challenge computer. Eight of the ten available processors are used simultaneously, with each proces-

processor dedicated to a single cloud model run. Instead of running each individual cloud model on multiple processors, this method results in perfect scaling for the overall ensemble and results in a faster execution time. On average, each cloud model requires about 130 min of processor time. Hence, eight cloud model runs can be completed every 130 min. Because a maximum of eight processors are used simultaneously, the ensemble is effectively divided into five sets. Each set requires the same processing time, so an entire ensemble is typically run in 650 min (~11 hr). The 0300 UTC Mesoeta output is typically available by 0800 UTC, so the full ensemble is complete by 1900 UTC. Thus, even in a non-operational environment, the ensemble output can be available within an operationally useful time frame. Obviously, faster processors, or more processors, will further reduce the overall execution time.

3.3 Ensemble output

Each COMMAS run extends to 2 hr, with output available for graphical display every 15 min. Because the cloud model is initiated with an unbalanced, warm bubble, non-physical, transient responses occur early in the model run. To make sure that any resulting cells have lifetimes beyond the eddy turnover time, output from the first 30 min is ignored. Storm lifetimes are based upon data extracted after this period. These data are available every 63 s, and consist of maximum and minimum values of various parameters, but only the maximum vertical velocity, w , and the maximum parameterized reflectivity, in dBZ_e , are retained and analyzed. With these parameters, cell lifetimes are defined as the time period over which the values of w or reflectivity exceed a predetermined threshold. For a given run, reflectivity and w typically provide different estimates of cell lifetime. Runs

that do not produce deep convection (i.e., have no parameter value above the predefined threshold) are not used to estimate cell lifetimes.

The definition of a cell lifetime implicitly depends upon a domain maximum w or reflectivity, which leads to the implicit assumption that only one cell is active at any given time. However, it is possible for more than one cell to occur within the domain of a cloud model. If outflow from the first cell triggers additional convection serially, such that w

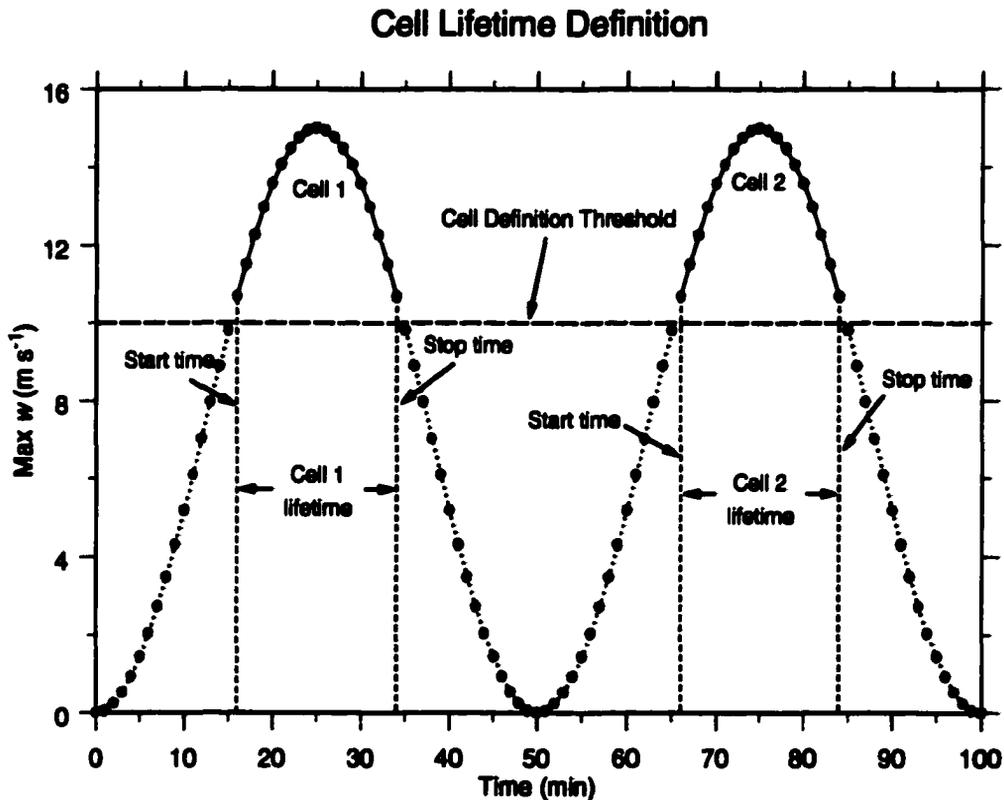


FIGURE 3.3. Schematic representation of how cell lifetime is determined. The data interval is 1 min and is shown by grey dots. Solid traces show the period for which cell lifetime is counted, and dotted traces show the period for which cell lifetime is not computed. The threshold that defines the existence of a cell shown by the horizontal dashed line at $w \geq 10 \text{ m s}^{-1}$. Vertical short-dashed lines show the time intervals for which cell lifetime is counted.

decays below the cell-definition threshold, and then recovers to a value above that threshold, *two*, separate cells are counted, each with separate lifetimes (Fig. 3.3)

The model domain (which is separate from the study domain) is 70 km × 70 km, which is large enough to contain more than one active cell. Errors can result when a domain maximum is used to define cell lifetimes and more than one cell is active within a single domain. For example, if two cells are simultaneously active, only the strongest counts for lifetime estimates. If a strong, long-lived cell masks a weaker, shorter-lived cell, a Type A error results (Fig. 3.4). This condition violates the one-cell-at-a-time

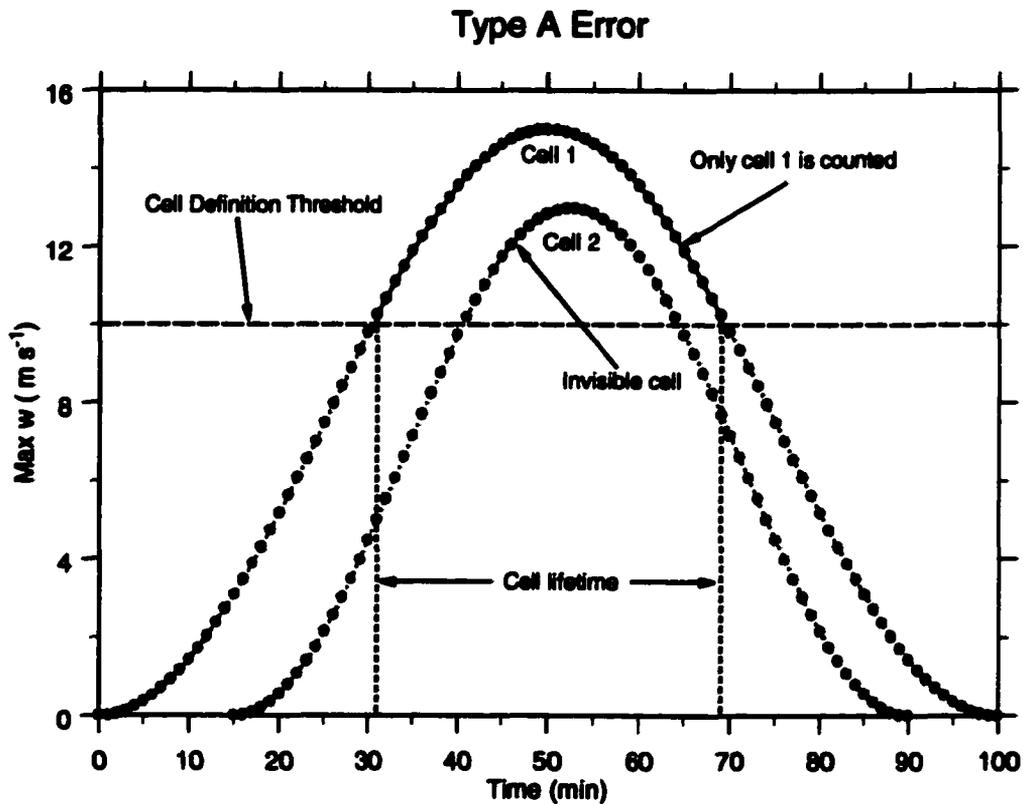


FIGURE 3.4. An example of a Type A error, where a strong, long-lived cell obscures a weaker, shorter-lived cell.

assumption and tends to bias cell lifetime estimates to be long. Both cells meet the existence criteria, but the second is essentially invisible to the analysis method. Type A errors occur in about 10% of the cases in this work.

Type B errors occur when storms follow each other serially in such a way that the first storm remains above threshold long enough for a second storm to intensify above the chosen threshold (Fig. 3.5). If the first cell then decays below the threshold, the lifetime

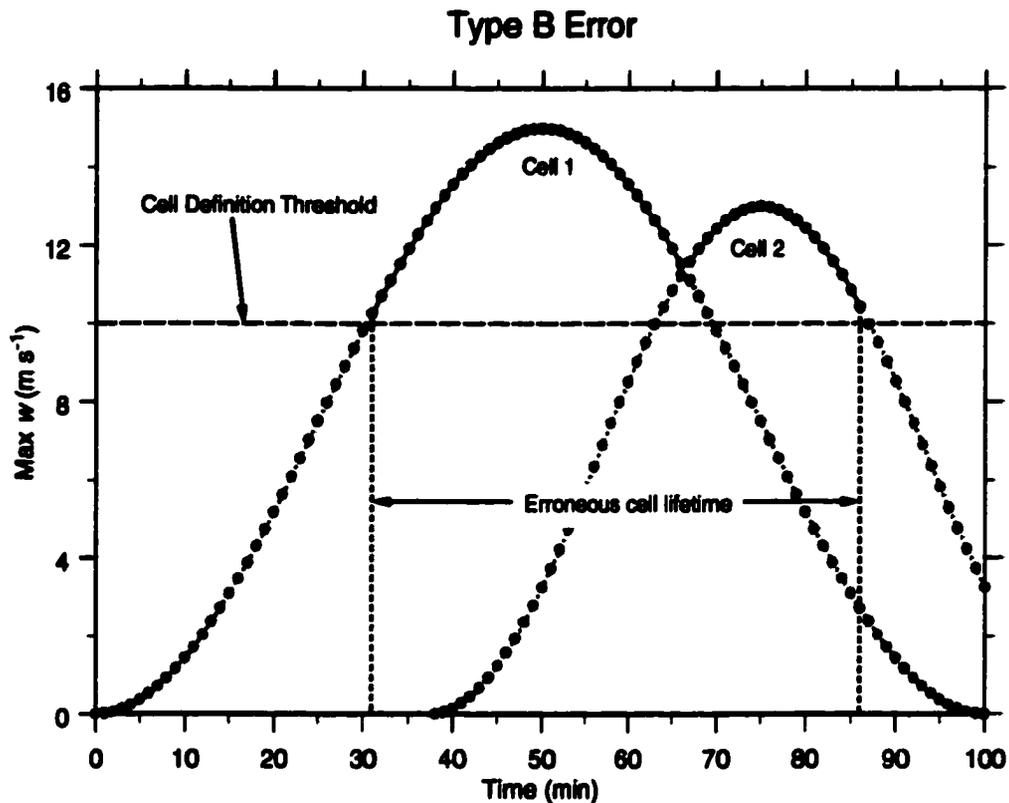


FIGURE 3.5. Depiction of Type B error, where two separate cells overlap temporally in such a way that they are counted as one, and appear as an anomalously long-lived cell.

estimate is an amalgam of two, separate cells within the same domain. Type B errors can lead to a significant overestimate of cell lifetime and are potentially more serious than Type A errors. Fortunately, Type B errors are rare, and none affect the storm lifetimes in the study data set.

3.4 Verification data

Level II archive data from the appropriate WSR-88D closest to the region of interest are used for verification. This verification data set is a superset of the data presented in

MacKeen et al. (1999). In total, eighteen days with convection are used, and eight of these are characterized by some kind of severe weather event (hail, wind or tornado). One day is characterized by a severe squall line, and the rest of the days are characterized by isolated convection. As in MacKeen et al. (1999), storms included in the verification data set must meet three criteria: 1) their lifetime is at least 12 min (10 min for VCP 11); 2) the maximum reflectivity must be at least 40 dBZ_e; and 3) the storm track is wholly contained within an annulus of 30 km to 125 km from the radar. This range interval is chosen to reduce radar sampling problems that can affect the SCIT algorithm when storms are too close to or too far from the radar. Relative to an airport, weather within this annulus has a large impact on airport capacity (MacKeen et al. 1999).

Archive data are replayed through the SCIT algorithm, which identifies cells and various parameters associated with them, and writes the derived characteristics to a text file (Johnson et al. 1998). Among other values recorded every volume scan are the cell identification number, maximum reflectivity within the cell, cell location (in range and azimuth from the radar), and the clock time for each volume scan. Collectively, these values define the observed cell lifetime for individual cells to be that period over which maximum reflectivity is at least 40 dBZ_e.

While parameters of the SCIT algorithm are tuned to reduce mis-association errors in these data, such errors still occasionally occur. A mis-association error occurs when either a new identification number is assigned to a pre-existing cell or a pre-existing identification number is carried to a new cell. Because cell lifetimes are linked to cell identification numbers, mis-association errors affect the statistics of cell lifetimes when spuriously short- and long-lived cells are created. For the parameters of the SCIT algo-

rithm used in this work, mis-association errors affect about 5% of the cells identified by the SCIT algorithm. To counteract this source of error, cell tracks are manually verified, and cells that are clearly mis-associated are removed from the verification data set. Hence, the overall impact is negligible because far fewer than 5% of the cells in the verification data set are affected by mis-association errors.

Because cloud model runs are terminated after 2 h, the longest storm lifetime that can be modeled is constrained to be 90 min. In contrast, observed cell lifetimes have no such intrinsic limit. To alleviate this inconsistency, observed storms that last longer than 90 min are truncated to a maximum lifetime of 90 min, which effectively treats all storms equivalently. This decision may not be appropriate for individual storms, because a storm that lasts 90 min may or may not be operationally equivalent to one that lasts 180 min. Even so, this artifice serves to identify gross characteristics.

Conversely, data are extracted from the model approximately every minute after the spin-up period, so the minimum detectable cell lifetime is 1 min. Yet, such a short-lived storm can not be observed if the WSR-88D is used as the verification source. Consequently, for consistency with observations, a modeled storm must last at least 6 min before it is counted. Where appropriate, this storm is assigned a minimum lifetime of 12 min for cases verified with VCP-21 data and 10 min for data verified with VCP-11 data. For example, storms with a modeled lifetime of 6 to 12 min are assigned a lifetime of 12 min if the verification data is collected in VCP-21.

3.5 pdf estimates of storm lifetime

Ensemble models applied to larger spatial scales typically reduce the ensemble forecast guidance to a mean of all the ensemble members. In contrast, this work attempts to forecast the pdf of storm lifetimes. To verify a forecast pdf, an observed pdf is needed. While histograms are a way to estimate a pdf, they suffer from many drawbacks. The choice of the bin width used to construct a histogram can significantly affect the histogram. Histograms are not continuous functions, because derivatives are undefined at each bin boundary, and derivatives are zero within the bins. When identical data are used, histograms constructed with different quantization intervals appear radically different. This characteristic, in particular, can seriously mislead the untrained observer. An example is the 1 min cell lifetime data interval provided by the cloud model and the 5 or 6 min data interval from the WSR-88D. To overcome this limitation, kernel density estimates are employed.

Observed and simulated cell lifetimes are transformed into non-parametric pdf's with a Gaussian kernel density estimator (Silverman 1986). The process is outlined here, with a more formal development provided in Appendix 1. For cell lifetimes, the true domain pdf lies on the interval $[0, \infty \text{ min}]$. Thus, the domain of the pdf is infinite, yet it must be placed on a finite domain for computational purposes. This computational domain is $[0, 100 \text{ min}]$. It extends to 100 min to avoid problems inherent in density estimators when the density is not zero near the end of the domain of the estimated pdf, which is called the support interval. The end value of 100 min is an arbitrary choice and primarily dependent upon the kernel bandwidth; yet it works well for this analysis.

For this work, the kernel density estimator is:

$$\hat{f}(x) = \frac{1}{(\sigma\sqrt{2\pi})} \sum_{i=1}^n e^{-\left\{\frac{1}{2}\left[\frac{(X_i-x)}{\sigma}\right]^2\right\}} \quad (3.1)$$

where $\hat{f}(x)$ is the density estimate at x ,

x = location of 1-D density estimate (101 locations are used at 1 min intervals),

X_i = observation location,

n = number of observations,

σ = bandwidth in units of standard deviations.

While the optimal bandwidth can be determined from theory, the pdf of the parent population must be known (Venables and Ripley 1997). For this study, the bandwidth, σ , is chosen to be 3 min because the grid upon which the pdf is analyzed is cell lifetime in minutes. This choice of bandwidth works well; it smooths most of the insignificant bumps and retains the real peaks (Fig. 3.6). If the same kernel density estimator is applied to both forecast and observed lifetimes, statistical comparison between the pdf's becomes straightforward. A kernel density estimate is nothing more than a unit-area histogram that is smoothed with a probability density function. The process is reminiscent of a 1-dimensional, 1-pass Barnes analysis (Barnes 1964, 1973) which uses the number of observations at locations X_i analyzed onto grid points at x .

3.6 Comparing pdf's

The difference between two pdf's has no specific definition. Accordingly, this work uses the Euclidean distance (Minkowski L_2 norm) between two pdf's to characterize the difference between them. An L_2 norm is preferable to a location metric, such as a

Histogram and Density Estimate for Cell Lifetime

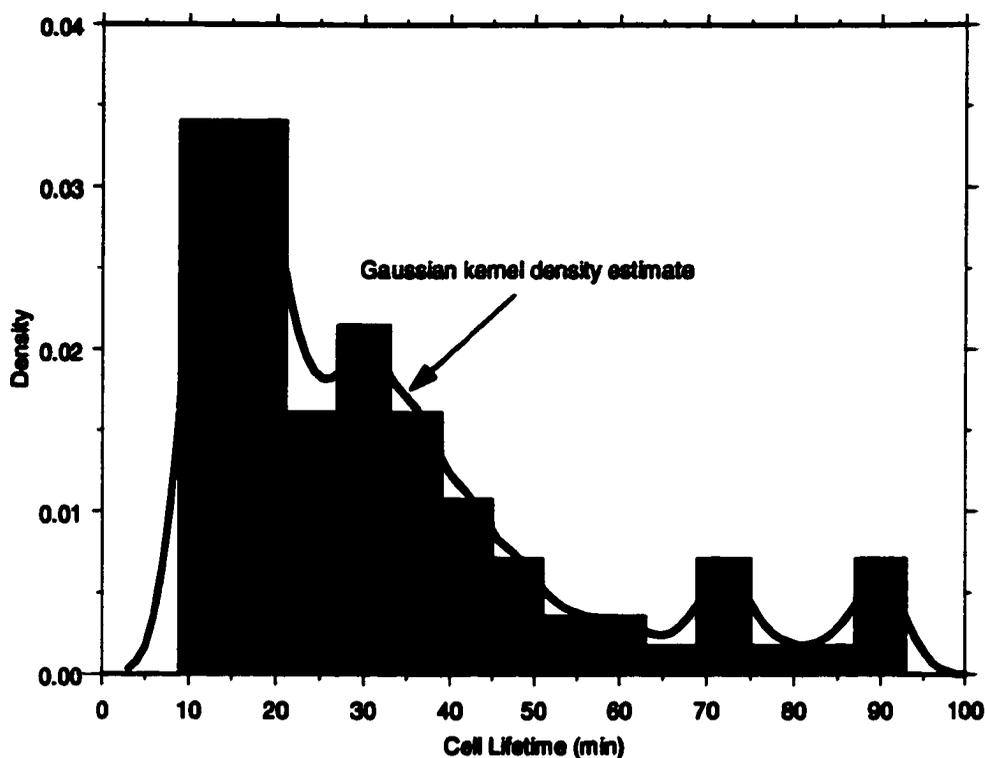


FIGURE 3.6. Cell lifetime histogram and corresponding Gaussian kernel density estimate. Data are observed cell lifetimes using WSR-88D and VCP-21 (6 min resolution). Histogram bars are centered on 6 min intervals. Bandwidth for kernel density estimate is 3 min.

mean or median, because the pdf's are multi-modal if more than one lifetime behavior mode exists. More importantly, multi-modal pdf's make location metrics difficult to interpret. If the distance between two pdf's is defined as d , then a perfect match between the pdf's occurs when $d = 0$. Thus, small values of d indicate that the pdf's are less different than do large values of d . Using d as a test statistic collapses an inherently multidimensional problem into a single dimension. The result is a non-unique measure because many different pdf's can yield the same value of d . Nevertheless, it is the most tractable way to proceed.

Almost certainly $d > 0$, but how large must d be before the pdf's are statistically different? Preisendorfer and Barnett (1983) discuss this problem. They develop a pool permutation procedure (PPP) and an auto-cross permutation procedure (APP) to determine the significance levels of the difference between two populations. Of these two, the PPP is the most applicable to the study data.

The general idea behind permutation tests is deceptively simple (Efron and Tibshirani 1993). Let F and O be two different pdf's where, for example, F is the forecast storm lifetime pdf, and O is the observed storm lifetime pdf. Let there be n samples from F and m samples from O , for a total of $m+n$ pooled samples. The null hypothesis, H_0 , states that the populations are indistinguishable, or simply $F = O$. If the null hypothesis is correct, any of the cell lifetimes could have come from either the forecast or the observed populations with equal probability. To start the PPP, first compute d_0 between the two original pdf estimates. Then, combine all of the $m+n$ samples and randomly remove a sample of size n without replacement to represent one population, and assign the m values that remain to the second population. Compute d between these two new populations and repeat the process a large number of times to create a number of d values. Once many values for d have been computed, calculate the likelihood of achieving a value of d_0 (Fig. 3.7). The achieved significance level (ASL) is the point where d_0 intersects the cumulative distribution function (cdf; Efron and Tibshirani 1993). This is also the probability of erroneously rejecting H_0 .

Significance levels for permutation tests tend to be more accurate than the straight bootstrap (Monte Carlo) significance levels (Efron and Tibshirani 1993). But, as in any resample technique, the results are somewhat variable. If another set of samples were

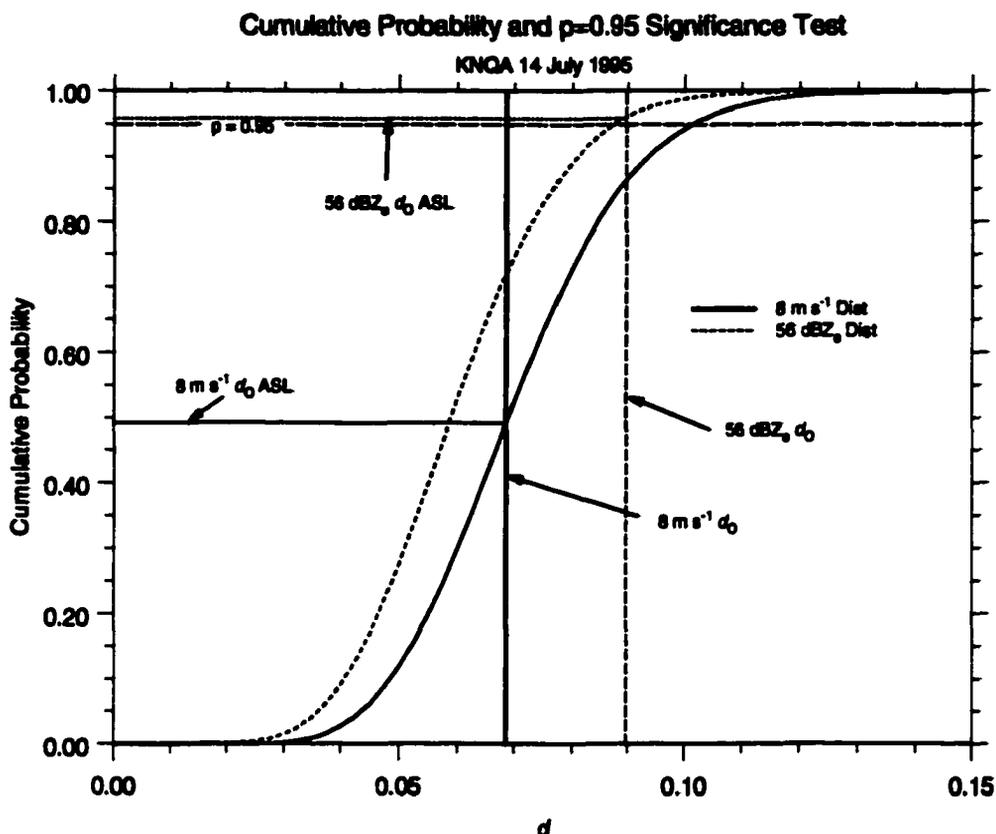


FIGURE 3.7. Cumulative density function (cdf) of d values for both w and reflectivity thresholds of cell lifetimes. The dotted lines and traces are for reflectivity while the solid lines and traces are for w . The vertical lines show d_0 . For reflectivity, the ASL = 0.959, and for w the ASL = 0.492. In one case (reflectivity), d_0 intersects the cdf above the $p = 0.95$ level (shown by the labeled dashed line). Consequently, the observed and forecast pdf's of storm lifetime are distinct with 95% confidence. In the other case (w), d_0 intersects the cdf below the $p = 0.95$ level (shown by the labeled, solid line). Thus, the observed and forecast pdf's of storm lifetime are indistinguishable with 95% confidence.

taken, a slightly different ASL might result. To determine how variable the ASL is given the number of trials, Efron and Tibshirani (1993) define a coefficient of variation to be a function of the desired significance level and the number of trials:

$$cv_B(\text{ASL}) = \sqrt{\left[\frac{(1-A/A)}{B} \right]}, \quad (3.2)$$

where A is the desired significance level, B is the number of trials, and $cv_B(\text{ASL})$ is the Monte Carlo error effect on the estimated significance level. This work uses 5000 trials,

and provides $cv_B(\text{ASL}) = 0.0032$ (0.32%) for $p = 95\%$. Hence, if the entire PPP process is performed many times, the resultant ASL will vary by about $\pm 0.16\%$. Consequently, a high confidence in the ASL is warranted.

3.7 Confidence intervals for pdf's

Cell lifetimes from both the model and the observations represent samples from two larger, unknown populations. Confidence intervals for these pdf's, which show how well each population pdf is represented, are created with bootstrap resamples (Efron and Tibshirani 1993). Bootstrap resampling assumes that the data to be resampled adequately represent the parent population, because the probability of observing a particular value is equal to the proportion of the data set that contains that value (this is the *sufficiency* criterion, Efron and Tibshirani 1993).

To create confidence bounds for the cell lifetime pdf, a set of storm lifetimes is resampled, *with replacement*, and an estimate of the kernel density pdf is made using the resampled set. Each resampled set is the same size as the original set. Each of the resulting pdf's represent a plausible sample from the population of cell lifetimes. Storm lifetimes are resampled 1000 times, the number needed for reliable confidence-interval estimates of a statistic (Efron and Tibshirani 1993). Then, the 2.5 and 97.5 percentiles are established at each discrete location of the estimated pdf. These confidence limits constitute an envelope that bound the true cell lifetime pdf with 95% confidence (Fig. 3.8). When two pdf's are plotted together, the pdf's are indistinguishable with 95% confidence where their confidence bounds overlap.

95% Confidence Interval for Cell Lifetime Kernel Density Estimate

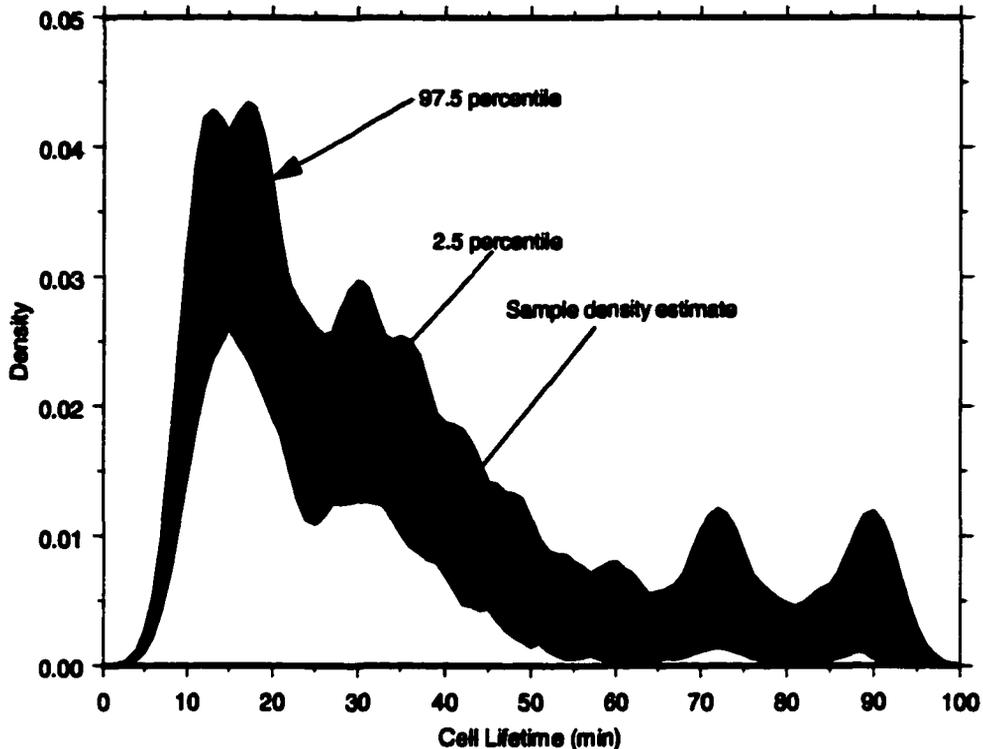


FIGURE 3.8. Bootstrap confidence intervals for the kernel density estimate in Fig. 3.6. The heavy dashed line is the sample density estimate, provided for comparison. Neither the 2.5 percentile nor the 97.5 percentile constitute a pdf. The true pdf of storm lifetime, which must enclose a unit area, exists within these limits with 95% confidence.

Bootstrap resampling also provides some insight into the stability of the pdf estimate. If the 2.5 percentile bound extends to zero density, then the original cell lifetime sample is not sufficient for reliable bootstrap confidence intervals. In such a case, the pdf estimate is not reliable.

3.8 Dominant behavior modes

A forecaster will have a better appreciation of how convection might behave if the convective modes contained within the ensemble forecast are identified (Anderson 1996a). If modes also are extracted from observations, then qualitative comparisons

between forecast and observed storm behavior is possible. Each model ensemble member is associated with a time series of maximum reflectivity and maximum w values. If these time series are arranged in a data matrix in which each column (variable) represents an individual time series, it is possible to perform a principal component analysis (PCA) that linearly combines those time series that display common characteristics (Richman 1986). To identify modes, whether reflectivity or w , for individual runs that overlay each other is more important than whether or not they vary together. To this end, Euclidean distance is a useful measure for classifying modes. Although Euclidean distance has not been used previously in PCA, cluster analysis uses Euclidean distance extensively to identify or group data or entities that are similar (Anderberg 1973; Gong and Richman 1994). Therefore, a PCA technique based upon Euclidean distance is developed (Appendix 2).

All of the extracted modes are based upon the apparent “closeness” of the time series to each other. Consequently, a group of time series that overlay each other or, put visually, that overlap or form a cloud, are linearly combined to form a single time series that represents the collection of similar time series. Hence, a cloud of w time series that last for the entire 90 min and have similar amplitude are linearly combined into a single mode (Figs. 3.9a and b). Results indicate that this measure is the best choice for identifying similar modes of storm behavior (see Appendix 2).

The extracted modes possess an artifact unique to the chosen similarity matrix, which is apparent when one mode ends while others remain active. In general, when one mode ends, at least one remaining mode shows an upward perturbation (Fig. 3.9b). This occurs because the similarity matrix is built with a distance-based measure and the retained modes explain as much of the total *distance* as possible. When one mode ends,

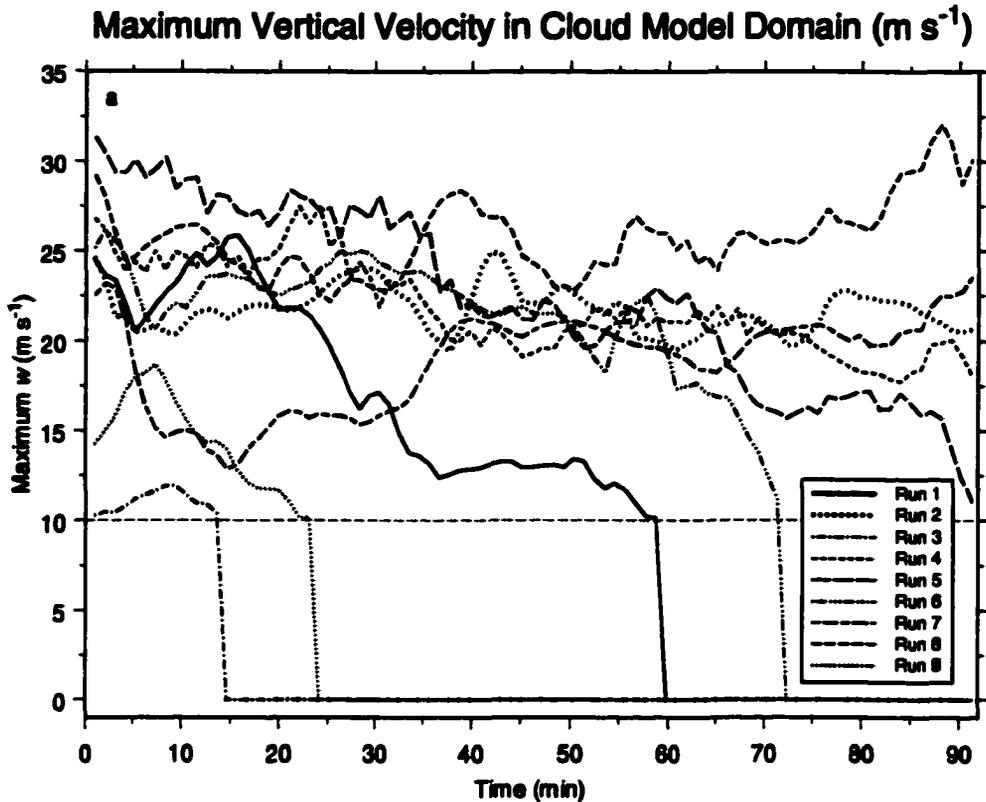


FIGURE 3.9A. Time series of vertical velocity from nine cloud model runs. Values represent the largest vertical velocity anywhere in the cloud model domain over a period of 92 min. The x -axis is the time in minutes, and the y -axis is vertical velocity in m s^{-1} . The dashed line at 10 m s^{-1} shows the threshold value used to determine cell lifetime for this example. Note that there are qualitatively three modes: one high-amplitude mode lasting the entire 92 min, a second moderate-amplitude mode lasting 60 to 73 min, and a low-amplitude, short mode that lasts 14 to 23 min.

the modes that remain must account for the “left-over” Euclidean distance by requiring as much distance as possible between the ending mode and those modes that remain. This distance is created by an abrupt upward deviation in whatever modes remain. Such an artifact can lead to nonphysical values, such as reflectivity values greater than 70 dBZ_e . Modeled modes do not show this characteristic as strongly, mainly because they have finer temporal discretization.

Another characteristic of Euclidean similarity is that the resultant modes do not retain the original amplitude of the data from which they are derived because the data

Euclidean Similarity Least Squares Modes

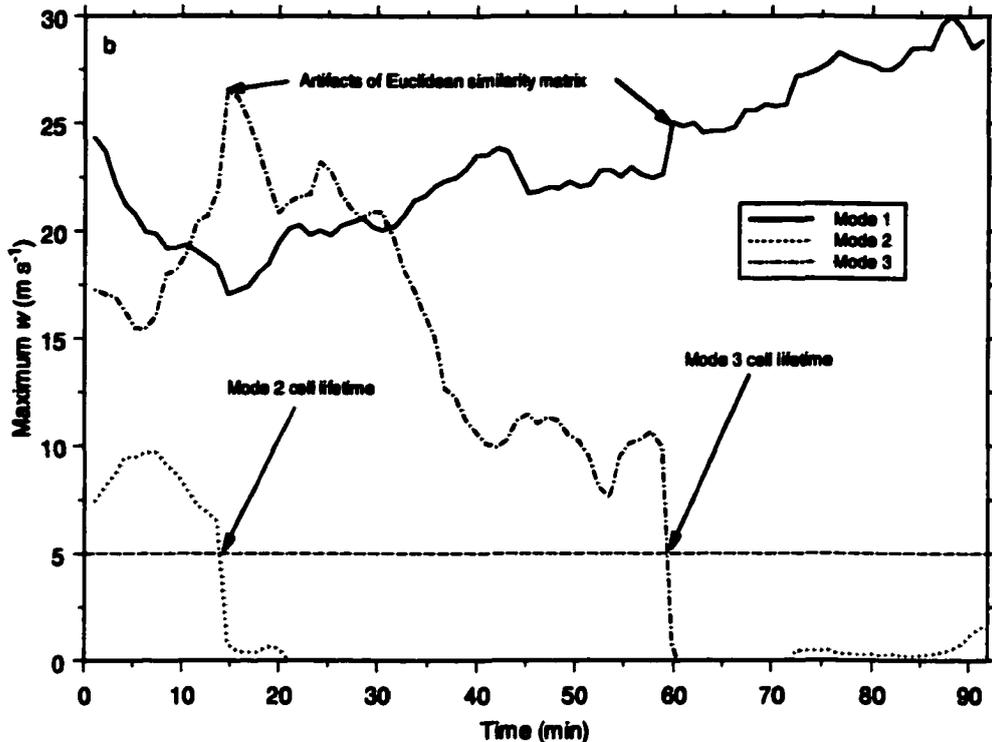


FIGURE 3.9B. The first three least-squares modes from a Euclidean-similarity PCA of the vertical velocities in Fig. 3.9a. The x-axis is time, and the y-axis is vertical velocity. The dashed line at 5 m s^{-1} shows the cell lifetime threshold used for the *modes*. Note that these modes are not filtered.

needed to do so are distributed in PCs that have been discarded (or, alternatively, cast into other dimensions of the eigenspace). If PCA is considered in a signal analysis paradigm, because some PCs (data) are discarded, some signal is discarded as well. The few retained components cannot recreate the total similarity contained in the original signal (the full data set). Hence, determining cell lifetime is problematic. Fortunately, another inherent characteristic of Euclidean similarity is that the resulting modes display either a sign change or rapidly decrease towards zero at some point. Either the sign change or the rapid decrease toward zero suffices to characterize the lifetime of each mode. This behavior is

not significantly affected by the amplitude of the mode and is used to define the cell lifetime.

The retained modes are filtered to remove the transient artifacts, especially since these artifacts can lead to non-physical values. A local, linear, least squares filter (Friedman 1984; Cleveland 1994) is applied to the modes in post processing to display the final modal structure. This filter effectively removes the short-time scale variations and simultaneously preserves the main features of interest in the recovered modes.

3.9 Contingency table analysis

Contingency table statistics are used to determine if the location of modeled storms, produced from the various Mesoeta soundings, provide information about where actual storms are observed. To simplify the geometry, the Minkowski L_1 norm (also known as a “Manhattan” or “taxicab” distance) is used to define the verification region (Fig. 3.10). This distance creates a diamond shaped area around each ensemble grid point. For the AWIPS 212 grid, a 40 km L_1 distance suffices. In this case, the vertex of each diamond is 40 km from an ensemble grid point. The L_1 distance is given by

$$L_1 = |\Delta x| + |\Delta y|, \quad (3.3)$$

where Δx is the x -direction distance from the grid point to a storm and Δy is the y -direction distance from the grid point to a storm. Put simply, the task is to determine if a cell is observed within a 40 km L_1 distance from each grid location where a sounding is used to initialize a cloud model forecast. A correct forecast occurs when the model produces deep convection at a particular grid point and a storm is observed by radar in the box defined by the L_1 distance from that point. A “miss” is defined when a storm occurs in the box and no

DFW Ensemble Region Overlaid with 40 km L_1 Areas

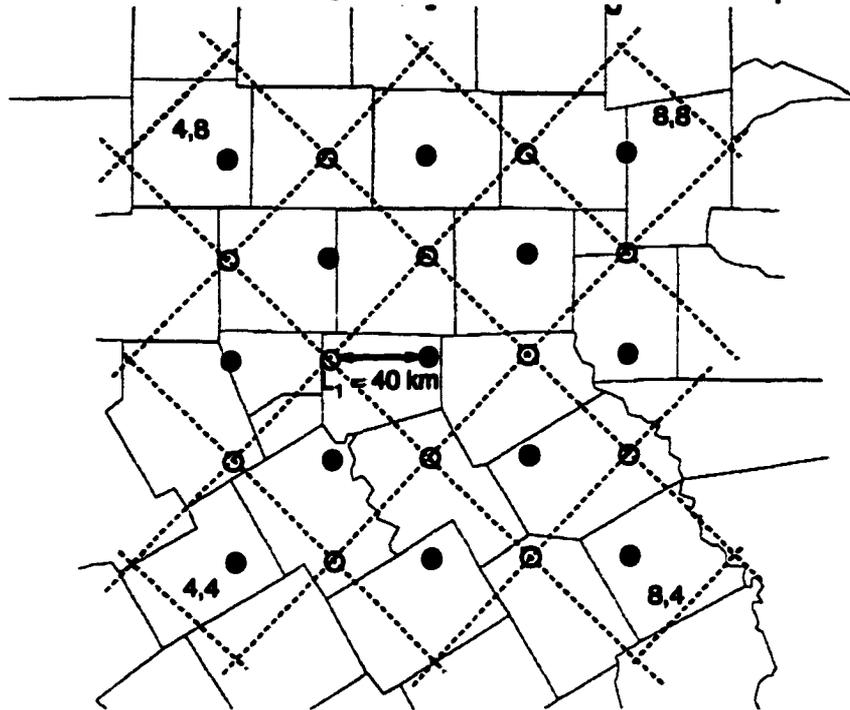


FIGURE 3.10. DFW ensemble region overlaid with 40 km L_1 distances which are shown by dotted lines. L_1 regions are centered on ensemble grid points. For clarity, county boundaries are grey.

storm is modeled. A false alarm is defined when modeled convection occurs but none is observed in the box around the grid point. A correct “no event” forecast (“correct null”) is defined when no modeled convection occurs and no convection is observed. The 2×2 statistics considered for the study data are probability of detection (POD), false alarm ratio (FAR), critical success index (CSI), and the True Skill Score (TSS; Wilks 1995). All of these scores represent an attempt to collapse a multidimensional problem into single numbers, and all suffer deficiencies. Wilks (1995) discusses the weaknesses and Marzban (1998) discusses at length the characteristics of these and other measures.

Many notations are used to describe 2×2 contingency tables. Here, a is the number of correct forecasts (“hits”), c is the number of missed detections (“misses”), b is the number of times convection is forecast and does not occur (“false alarms”), and d is the

number of correct nulls, which means no convection is forecast and none is observed (Table 3.1).

Table 3.1.

		Observed	
		Yes	No
Forecast	Yes	<i>a</i>	<i>b</i>
	No	<i>c</i>	<i>d</i>

TABLE 3.1. A basic 2 × 2 contingency table. The various skill scores are given by:

$$\text{POD} = \frac{a}{a + c} \quad (3.4)$$

$$\text{FAR} = \frac{b}{a + b} \quad (3.5)$$

$$\text{CSI} = \frac{a}{a + b + c} \quad (3.6)$$

and

$$\text{TSS} = \frac{ad - bc}{(a + c)(b + d)}. \quad (3.7)$$

The TSS possesses some appealing characteristics. Random forecasts, where a random forecast is based on the same relative frequency as the observed event frequency, and constant forecasts, such as “no convection,” receive a zero score. Also, the contribution of correct “no event” and correct “yes” (hit) forecasts increase as the event is more or less likely, respectively. Hence, forecasts of rare events are not discouraged based solely on their low relative frequency.

3.10 Initial conditions verification

Knowledge about the accuracy of the initial soundings used to initialize ensemble members helps cast the ensemble results within a proper context. Questions about the forecast soundings must be addressed, such as do forecast soundings capture the correct mesoscale variability? Do forecast soundings contain biases and, if so, what are their characteristics? These questions can be addressed directly with some common, simple statistical tools.

Because observed soundings are not available from either the DFW site or the MEM site, the Mesoeta analysis is used as the best guess of actual conditions in these regions. Mesoeta analyses in 1995 and 1996 occur twice daily, at 0000 UTC and 1200 UTC. The only common analysis and forecast time is 0000 UTC. Hence, 0000 UTC forecast soundings are compared to 0000 UTC analysis soundings for each day in the study data set. Ensembles are available for 18 different days, but only on 17 of those days can 0000 UTC analyses be recovered from NCEP (0000 UTC analyses for the 0300 UTC Mesoeta are not usually distributed). Because thirteen 0000 UTC soundings are used for each day, a total of 221 soundings are available for verification.

First, a t-test is used to determine if the difference between forecast and verification soundings is significantly different from zero (Blum and Rosenblatt 1972). The t-test assumes, as a null hypothesis, that there is no difference between the mean forecast and mean verification soundings. Tests are performed for θ (potential temperature), q (mixing ratio), u , v , and wind speed, at the surface and all levels from 950 hPa to 100 hPa, in 25 hPa increments. Results are displayed as the 95% confidence interval plotted against pressure for the forecast 0000 UTC sounding minus the verification 0000 UTC sounding (Fig.

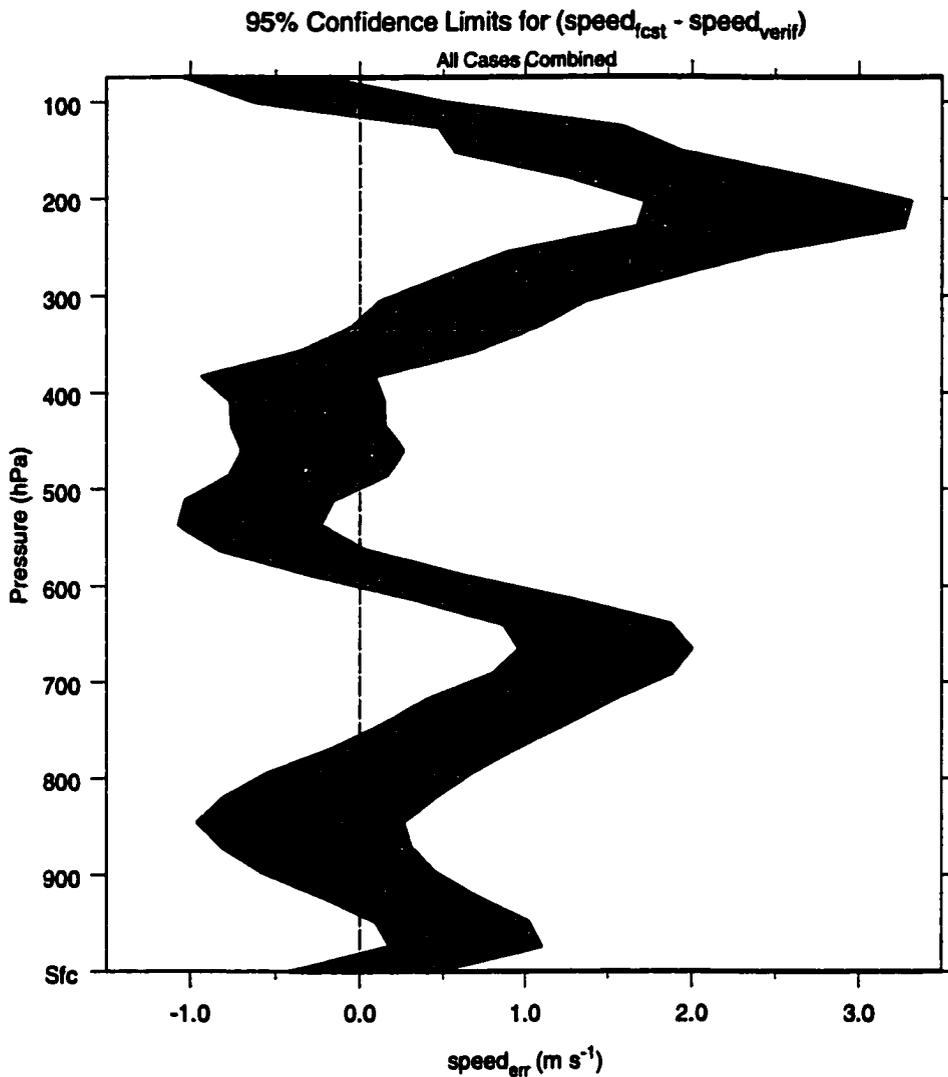


FIGURE 3.11. 95% confidence interval for wind speed errors (forecast speed - analysis speed). Dashed line is the zero-error line. When the dashed line is within the shaded region, the difference between the forecast and verification wind speed (or whatever parameter is considered) is insignificant with 95% confidence.

3.11). If this confidence interval contains the zero error value at some level, then, with 95% confidence, the null hypothesis cannot be rejected and the error is insignificant at that level. This same test can be used to determine the significant differences between soundings that produce long-lived storms and soundings that do not.

While a daily mean sounding is not used to create the ensemble initial conditions, a daily mean sounding is useful to help characterize the variability of the individual sound-

ings that comprise the ensemble initial conditions. Consequently, daily deviation soundings are produced from a daily mean forecast sounding and a daily mean verification sounding derived from the 0000 UTC analysis. A distribution of deviation values for the forecast and verification soundings is produced and compared via a two-sample Kolmogorov-Smirnov (KS) goodness-of-fit test (see Appendix 3). Descriptively, the KS test determines if two distributions are different based upon sample size and the maximum vertical separation between the sample cdfs (Fig. 3.12). The KS test can be applied when three conditions are met: 1) the two samples are random, 2) the two samples are mutually independent¹, and 3) the data are measured on at least an ordinal scale. Finally, KS test results are exact if the parent pdf's are continuous which is the case with the deviation values. All these conditions are met with these data.

While a PPP could be used to compare the parent distributions, it is not necessary because the perturbation temperatures are not artificially discretized, i.e., data values are not artificially cast into 1 degree increments. Why not, then, use the KS test to compare forecast and observed cell lifetime distributions? In the case of cell lifetimes, the parent pdf is continuous, but the sample is artificially discontinuous because cell lifetimes are discretized to either 1 min (forecast), 5 or 6 min (WSR-88D data). This discretization artificially masks the continuous nature of the parent pdf, which can cause significant error in KS test results. Moreover, the true number of elements in a sample, which is a pivotal part of the KS test, are discarded by the kernel density estimation process, because, regardless of how many elements comprise a sample, a kernel density estimate is always composed

1. This condition may not be strictly met for these data, but is probably met closely enough. If the two distributions are not mutually independent, the KS test will tend to falsely accept H_0 , that the distributions are indistinguishable.

KS Goodness-fo-Fit CDF Comparison

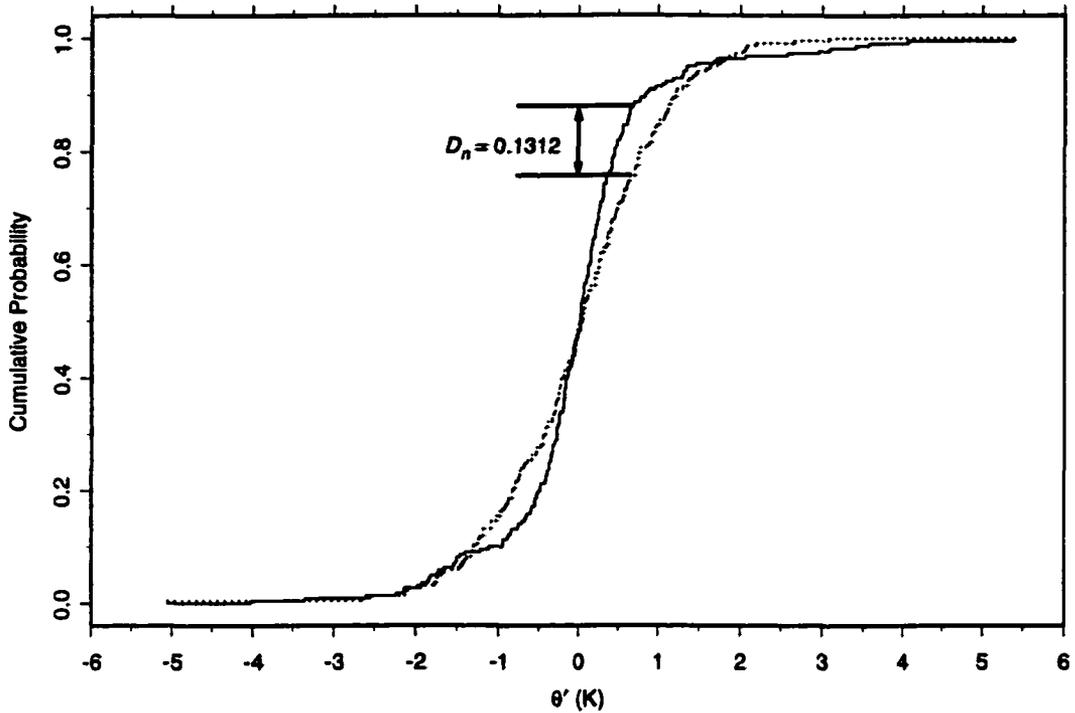


FIGURE 3.12. Kolmogorov-Smirnov goodness-of-fit computation for the perturbation temperature (θ') at 950 hPa. The solid curve is for the verification θ' , and the dotted curve is for the forecast θ' . The maximum vertical distance between these two curves is D_n . For this example, the two distributions are different with 95% confidence ($p = 0.037$).

of the same number of discrete points. So, the KS test cannot be reliably used after a kernel density estimate has been performed.

This completes the description of the various methods and techniques used to create and analyze the ensemble model output. While these techniques are statistical in nature, they constitute a good, expedient way to gain physical insights into the meteorology contained within the cloud model ensembles. With all this in place, ensemble results follow.

Chapter 4: Results

4.1 Days used in analysis

A total of eighteen days are investigated: 5 days from the 1995 warm season and 13 days from the 1996 warm season (Table 4.1). Sixteen of these days use data from the

Table 4.1. Days Used in Analysis

<i>Date</i>	<i>Location</i>	<i>Severe Weather Type</i>
6 June 1995	MEM	Wind, Hail, Tornado
14 July 1995	MEM	Wind
17 July 1995	MEM	Wind
17 August 1995	MEM	none
19 August 1995	MEM	Wind
28 May 1996	MEM	Wind, Hail
6/7 June 1996	DFW	Wind, Hail
13 June 1996	MEM	none
22 June 1996	MEM	none
23 June 1996	MEM	none
29 June 1996	MEM	none
8 July 1996	MEM	Wind, Hail
16 July 1996	MEM	none
8 August 1996	MEM	none
12 August 1996	MEM	none
17 August 1996	MEM	none
30 August 1996	MEM	none
21 October 1996	DFW	Wind, Hail, Tornado

Memphis, TN, region while the remaining two days use data from the Dallas/Ft. Worth, TX, region. According to the Storm Prediction Center's (SPC) Smooth Log (a quality controlled version of Storm Data, Doswell 1985), 8 of these 18 days had severe weather of

some kind within the region of interest, which means that either surface winds in excess of 25 m s^{-1} , hail greater than 1.9 cm in diameter occurred, or a tornado occurred.

4.2 Initial condition verification

Data from the 0000 UTC Mesoeta analysis are available for 17 of the 18 cases (a 0000 UTC Mesoeta analysis is not available for 6 June 1995). For each forecast sounding, five parameters, θ , q , u , v , and wind speed, all valid at 0000 UTC, are verified using soundings derived from the next day's 0000 UTC analyses. A combined analysis using all days pooled together is presented along with a focussed effort on three specific days. Most verification studies use RMS error to compare forecasts with analyses (or observations), but RMS errors do not provide information about statistical significance. However, 95% confidence intervals derived from a t-test provide insight into statistically significant forecast errors. The t statistic tests the null hypothesis, H_0 , that no difference exists between the forecast and analyzed parameter. Strictly applied, the t-test is parametric because the parent distributions are assumed to be normal, but the t statistic is tolerant of non-normally distributed data.

Potential temperature (θ) has a warm bias below 700 hPa, a moderate cool bias between 700 and 525 hPa, a slight warm bias from 450 to 300 hPa, and a strong warm bias above 225 hPa (Fig. 4.1a). At the surface, the temperature bias is +3 K (too warm). This bias decreases to slightly over +1 K above 950 hPa until near 700 hPa. This type of bias has been noted previously (Marshall 1998), and is a consequence of the land-surface parameterization scheme used in the Mesoeta model during the 1995-1996 period. In simple terms, the sensible heat flux in the Mesoeta model is too large. One consequence of

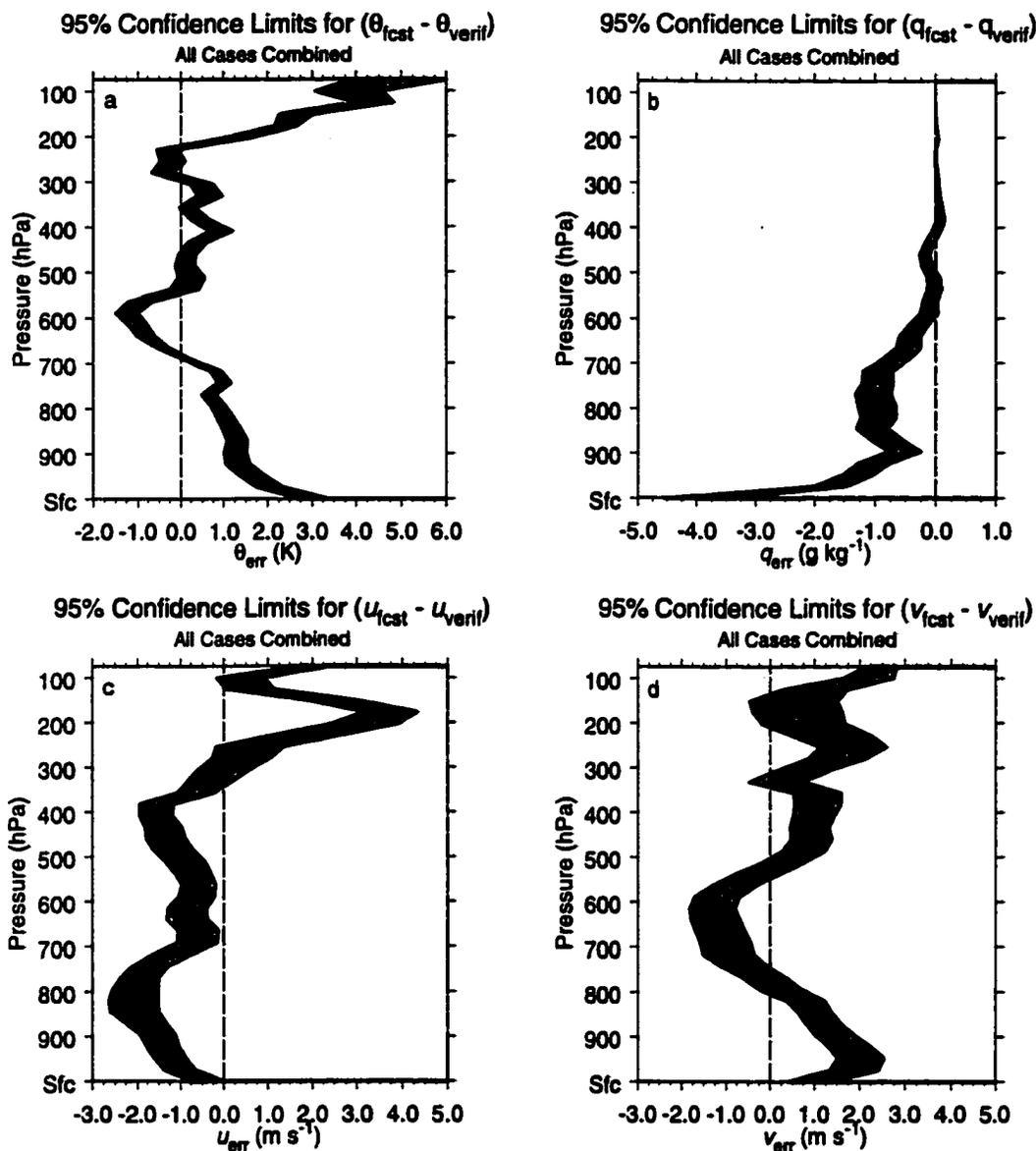


FIGURE 4.1. 95% confidence interval for a) θ_{err} , b) q_{err} , c) u_{err} , d) v_{err} , and e) wind speed error. The vertical dashed line is zero error. When the confidence interval (shaded region) contains the zero line, the error is insignificant at the 95% confidence level.

this erroneous heat flux is a modeled moisture flux that is less than the actual surface fluxes of latent heat. In turn, the low-level mixing ratio is reduced in the model.

The very warm bias above 225 hPa suggests that the forecast tropopause height is too low. This upper-level error is unlikely to affect modeled storm lifetimes significantly.

Overall, these results indicate that the static stability forecast in the lower half of the troposphere is insufficient.

The same land surface parameterization that creates a warm bias in boundary layer temperature causes the forecast mixing ratio to be too low below about 575 hPa. Near the surface, the mixing ratio is

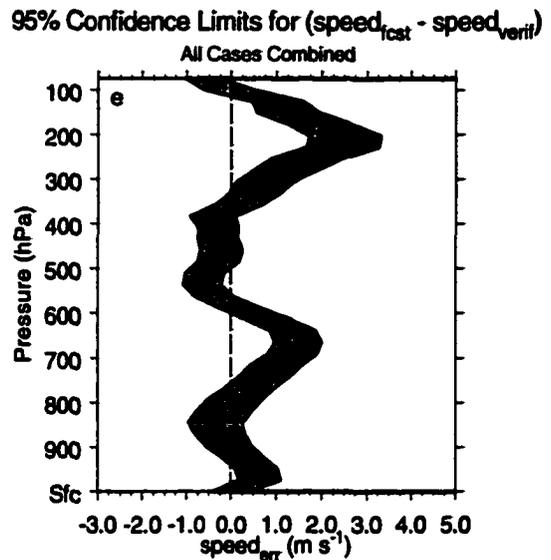


FIGURE 4.1. Continued.

too low by nearly 4 g kg^{-1} . This bias towards low values moderates to roughly 1 g kg^{-1} above 900 hPa. The excessively large sensible heat flux is particularly evident in summer, when evapotranspiration is a primary source of low-level moisture. Forecast boundary layer mixing ratios that are too small result in CAPE values that also are too small. Insufficient moisture also is forecast for mid-levels, which, when combined with surface temperatures that are too warm, translate to downdrafts that are too cold. These forecast biases, combined with the Kessler precipitation scheme, can lead to downdrafts originating in mid-levels that are both too cold and too strong. Depending upon the overall environment, this can lead to either storms with forecast lifetimes that are too short or storms that form a supercell-like organization too quickly, and hence last too long.

Errors in u and v are similar in magnitude (Fig. 4.1c and d). The u component contains too much easterly flow below 350 hPa, then sharply too much westerly flow above 250 hPa. The easterly bias is strongest (2 m s^{-1}) around 850 hPa. Errors in the v component forecast contain too much southerly flow below 800 hPa, too much northerly flow from 750 to 550 hPa, and too much southerly flow above that level. The strongest southerly bias

is about 2 m s^{-1} just above the surface, and the strongest northerly bias is about -1.5 m s^{-1} at 650 hPa (Fig. 4.1d). Errors in forecast speeds are near zero up to 850 hPa, but are positive (too strong) from 850 hPa to 600 hPa. Between 600 and 300 hPa, the errors in wind speed is near zero, but from 300 hPa to 100 hPa, the error in wind speed is as large as 2.5 m s^{-1} (Fig. 4.1e). Overall, the wind errors result from low-level southeasterly winds that are too strong, and mid-level winds that are too northerly.

In a similar fashion, 95% confidence intervals are constructed for forecasts of five bulk parameters that are known to be relevant to severe weather events. These bulk parameters are derived from the forecast soundings valid at 0000 UTC and verified against soundings from the next day's 0000 UTC analysis (Bluestein 1993; Davis-Jones et al. 1990). These parameters are CAPE, lifted index (LI), SREH, bulk Richardson number, and the shear from the bulk Richardson number (BRN shear; Bluestein 1993; Davies-Jones et al. 1990; Doswell 1985). This analysis helps answer questions about how well salient sounding features are forecast, regardless of how well any individual sounding level is forecast.

Consistent with the previous verification characteristics, a t-test on forecast mean CAPE vs. analyzed mean CAPE indicates that forecast CAPE is significantly less than the CAPE derived from the analysis (Fig. 4.2d). Results are similar for LI, where the forecast LI is significantly too positive (stable) compared to the LI derived from the analysis (Fig. 4.2a). These statistical results are expected primarily because the low-level mixing ratio forecast by the Mesoeta model is significantly less than the verification mixing ratio. Because these statistical results are derived from several days pooled together, there are

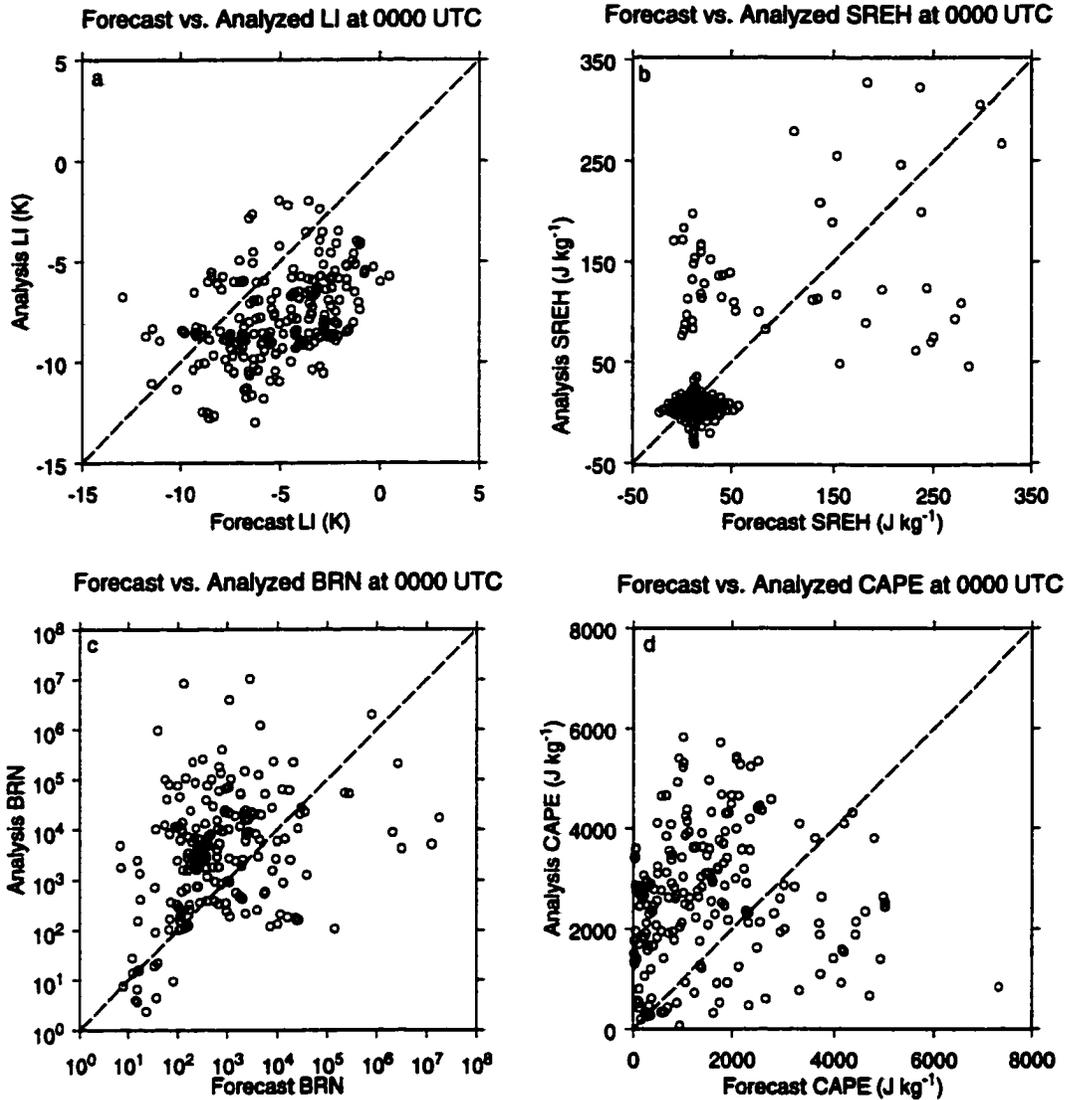


FIGURE 4.2. 0000 UTC forecast vs. 0000 UTC verification for a) lifted index (LI), b) storm relative helicity (SREH), c) bulk Richardson number (BRN), and d) convective available potential energy (CAPE). If the forecast and analyzed values agreed perfectly, they would lie on the dashed lines.

times when, for example, the warm bias is large enough to overcome the dry bias, which results in a forecast CAPE that is larger than the analysis CAPE.

At the 95% significance level, no difference exists between forecast and analyzed SREH (Fig. 4.2b). This result is at least partially an artifact of the many summer days used, which are typified by low-shear environments. The large data cluster centered on

zero SREH is clear evidence that most days are characterized by minimal shear. However, when soundings with either the forecast or the analysis SREH $> 50 \text{ J kg}^{-1}$ are compared, the analysis SREH is larger than the forecast SREH at the 95% significance level. Hence, on days with significant shear, the forecast SREH is too low. Because

Forecast vs. Analyzed BRN Shear at 0000 UTC

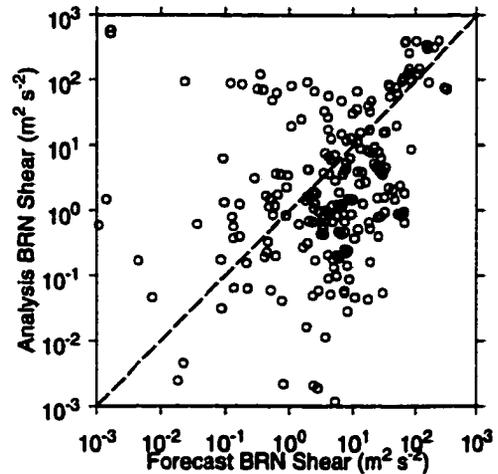


FIGURE 4.2E. Forecast BRN shear vs. verification BRN shear valid at 0000 UTC.

SREH is sensitive to storm motion, which is not estimated or forecast very accurately, clear evidence does not exist whether this bias error in SREH creates an associated bias error in storm lifetimes.

Because BRN ranges from 0 to 10^8 , BRN is transformed to $\log(\text{BRN})$ and a t-test is performed on $\log(\text{BRN})$. When transformed this way, the forecast BRN is significantly smaller than the corresponding analysis BRN (Fig. 4.2c). This result also holds true for $\log(\text{BRN shear})$, which, like BRN, suffers from numerical instability owing to the large range of values that are possible. When a t-test is performed on $\log(\text{BRN shear})$, which looks much the same as BRN, the forecast BRN shear is also significantly less than the analysis BRN shear. These results are consistent with the preceding verification analysis.

As in any verification exercise, these results do not necessarily apply to any particular sounding. These statistics serve only to help understand the nature of the forecasts. Further, whether the forecast values will be too high or too low is not known *a priori* from day to day. As always, when properly used, model output is guidance, and the forecaster must interpret any guidance as she sees fit.

Understanding how well a forecast model performs, on average, is only one aspect of the verification problem. Like any numerical model, ensemble forecasts are a hostage to their initial conditions. But, the ensemble model also depends upon the *range* of initial conditions it is provided. Because this particular cloud-model ensemble model depends upon forecast soundings, it is important to know how well these forecast soundings capture the range of variability that exists over the forecast region. How variable are the soundings extracted over such small regions? Is the forecast variability an accurate depiction of the analyzed variability? To answer these questions, the variability in the 0000 UTC forecasts is compared with the variability in the 0000 UTC analyses.

When perturbations from the forecast daily mean values are pooled, the overall variability is dependent on height. Potential temperature has its greatest variability in the boundary layer, and near the tropopause (Fig. 4.3a). Clearly, over a 160×160 km spatial region and a nine hour time span, at least 50% of the perturbation values fall outside a ± 1 K range, a range where significant variations in thunderstorm characteristics are observed (Crook 1996; Brooks et al. 1992). Mixing ratio is naturally height dependent as are any perturbations in mixing ratio. Hence, the greatest mixing ratio variability is in the boundary layer, where over 50% of the data have perturbation values outside a ± 1 g kg⁻¹ range, a range where significant variations in thunderstorm characteristics are observed (Crook 1996; Fig. 4.3b). Perturbation values of u and v are remarkably similar, with the smallest interquartile range (IQR) near the surface and the largest IQR near the tropopause (Figs. 4.3c and 4.3d). Both distributions are generally symmetric and similar in appearance, though the perturbation v has a larger negative extrema.

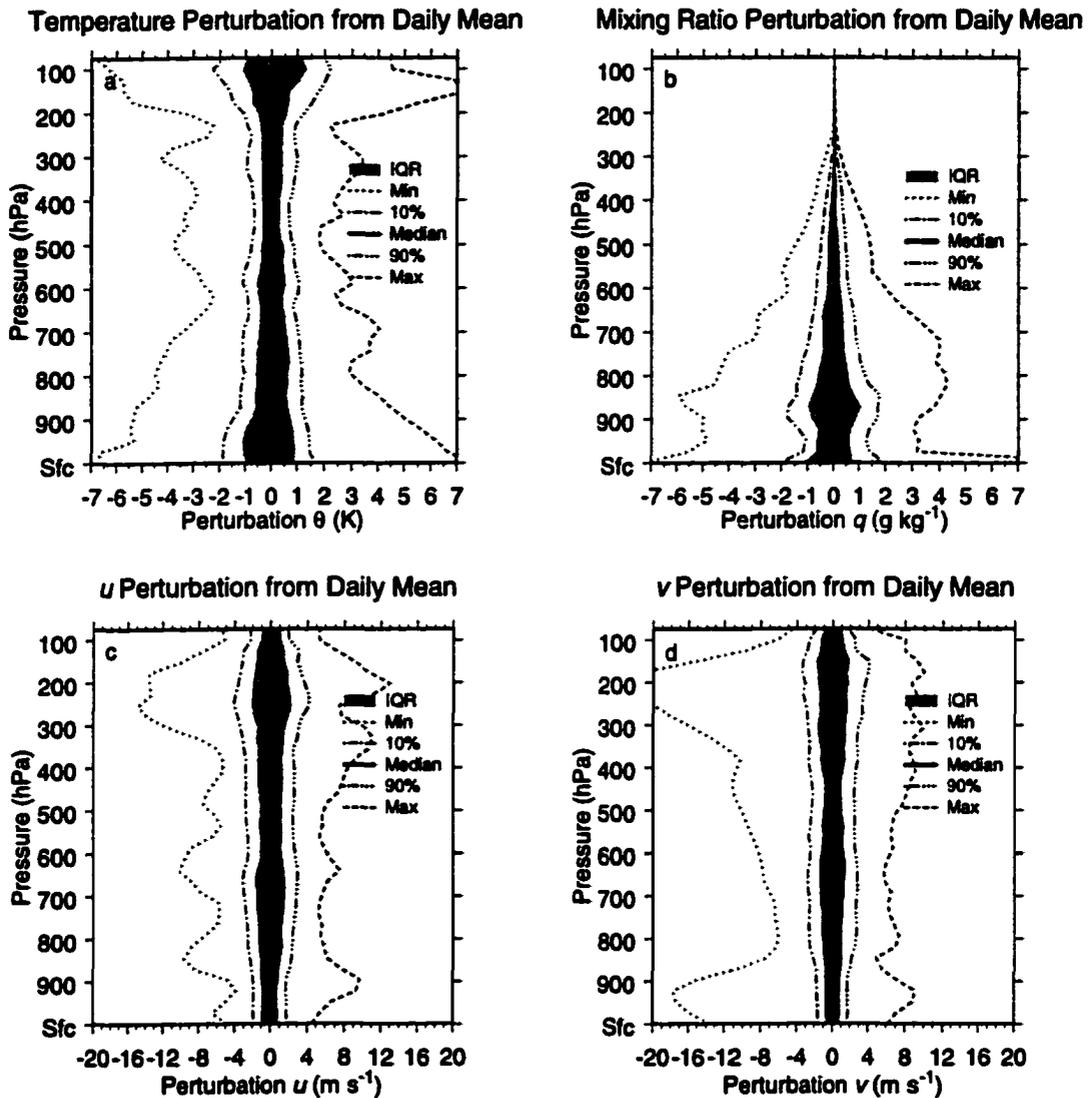


FIGURE 4.3. Distribution of forecast sounding parameters from daily means for all days, combined: a) potential temperature (θ), b) mixing ratio (q), c) u , and d) v . The solid line shows the median, and the grey region contains 50% of the data. The 10th and 90th percentiles are shown by dash-dot and dash-dot-dot-dot patterns (containing 80% of the data) as are maximum (dashed) and minimum (dotted) values.

A KS test is used to determine if the forecast and analyzed perturbation distributions at 0000 UTC are different at each sounding level. As implemented in all preceding statistical comparisons, 95% confidence limits are imposed. For perturbation θ , only one level in the boundary layer has different forecast and analysis perturbation distributions (Fig. 4.4a). Significantly different distributions also exist for a few layers above 600 hPa.

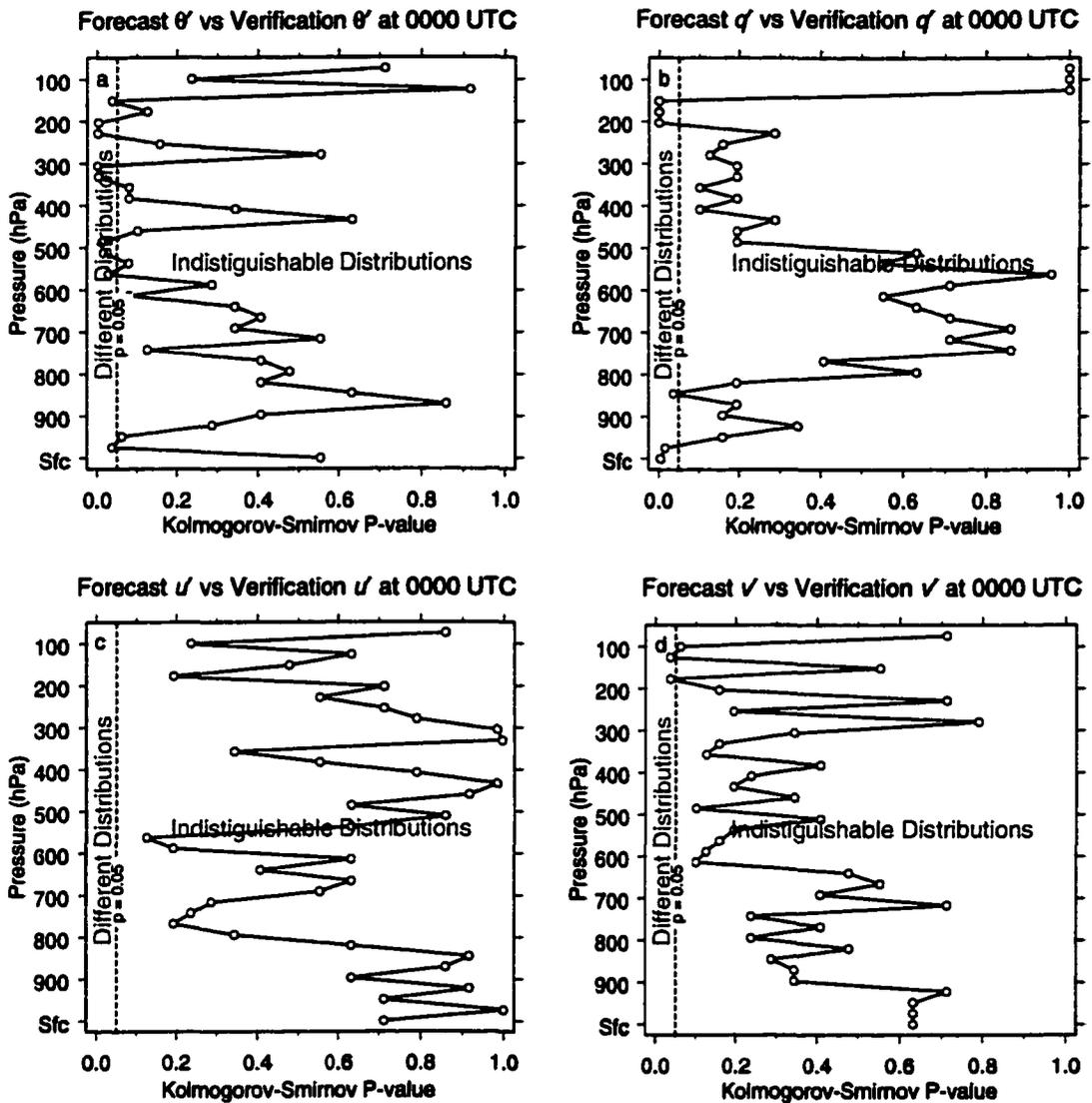


FIGURE 4.4. KS test results for the forecast vs. the analysis perturbation distributions valid at 0000 UTC for a) θ , b) q , c) u , and d) v . The vertical dashed line indicates the 95% significance level ($p=0.05$). Values to the left of this line indicate parent distributions that are different.

Except near the surface, the distribution of forecast perturbation mixing ratio (q') is indistinguishable from the distribution of analyzed q' for most of the troposphere (Fig. 4.4b). The KS results above 200 hPa have no meaning because mixing ratio is, for all practical purposes, zero at these levels. The distributions of forecast and analyzed perturbation u components are statistically indistinguishable at all levels (Fig. 4.4c). The same applies for perturbation v components (Fig. 4.4d). At two levels near the tropopause, the

forecast and analysis perturbation v distributions are different, but this result is expected by chance and is probably not significant.

These results are encouraging because the forecast and analysis dispersions are equal. This indicates that using a mesoscale model to generate the perturbations for a cloud-scale model is a reasonable approach.

Crook (1996) shows that variations of only ± 1 K and ± 1 g kg⁻¹ in the boundary layer can make the difference between whether or not a storm forms. The biases in the mean fields clearly exceed these limits. Given these verification results, why proceed with the ensemble cloud model exercise? If the goal is to explicitly (or deterministically) model particular thunderstorms at particular locations, the answer to this question is self evident and there is no reason to proceed. However, the variability exceeds the magnitude of biases (cf. Figs. 4.1 and 4.4). Hence, the goal to extract the maximum amount of data concerning the range and distribution of thunderstorm lifetimes from the Mesoeta may still be attainable. There is no reason why non-zero mean biases preclude success using an ensemble model.

4.3 General results

Results for all days combined are presented first, followed by results from individual days. Observed lifetime can be based only on observed reflectivity, but both w and reflectivity can be used to estimate forecast lifetime. However, a question arises: which parameter works best for forecasting the range and distribution of storm lifetimes? To determine the answer to this question, forecast results for all days are pooled into a single data set. The threshold for forecast lifetimes based on w is varied from 5 to 10 m s⁻¹ in 1 m

s^{-1} increments and, in a like manner, the threshold for forecast lifetime based on reflectivity is varied from 30 to 60 dBZ_e in 2 dB increments. This scenario provides 10 different thresholds for w and 16 different thresholds for reflectivity. Kernel density estimates are computed based on each threshold of each parameter, and compared to the kernel density estimate of the observed lifetime for the superset of all observed cells. As yet another measure of similarity, the Euclidean distance is computed between the forecast lifetime pdf's and the observed lifetime pdf. Ideally, as the Euclidean distance between the forecast and verification pdf decreases, the forecast pdf becomes a better representation of the observations.

A clear minimum for the pdf of storm lifetime based on w occurs at 8 m s⁻¹. Therefore, the best predictor for observed storm lifetime is w when a threshold of 8 m s⁻¹ is used (Fig. 4.5). When compared on a cell-by-cell basis, the 56 dBZ_e threshold tends to result in shorter lifetimes compared to the 8 m s⁻¹ threshold (Fig. 4.6). A least squares linear regression fit yields a slope less than 1, which indicates that given a lifetime based on 8 m s⁻¹, the expected lifetime based on 56 dBZ_e is shorter. In particular, when the 8 m s⁻¹ threshold yields a lifetime of 90 min, the 56 dBZ_e threshold yields lifetimes between zero and 90 min. Such a broad range indicates that, in some cases, a vertical velocity of 8 m s⁻¹ is maintained for 90 min, but the reflectivity never exceeds 56 dBZ_e.

The observed lifetime pdf is bimodal, with the strongest peak around 12 min and a secondary peak near 90 min (Fig. 4.7). In comparison, the pdf based on a w threshold of 8 m s⁻¹ also has a peak at 12 min and a broadened area at 23 min. A secondary peak, which is too large by a factor of three, exists at 90 min (Fig. 4.7a). In contrast, the reflectivity-

Euclidean Distance Between Forecast and Observed Cell Lifetime pdfs

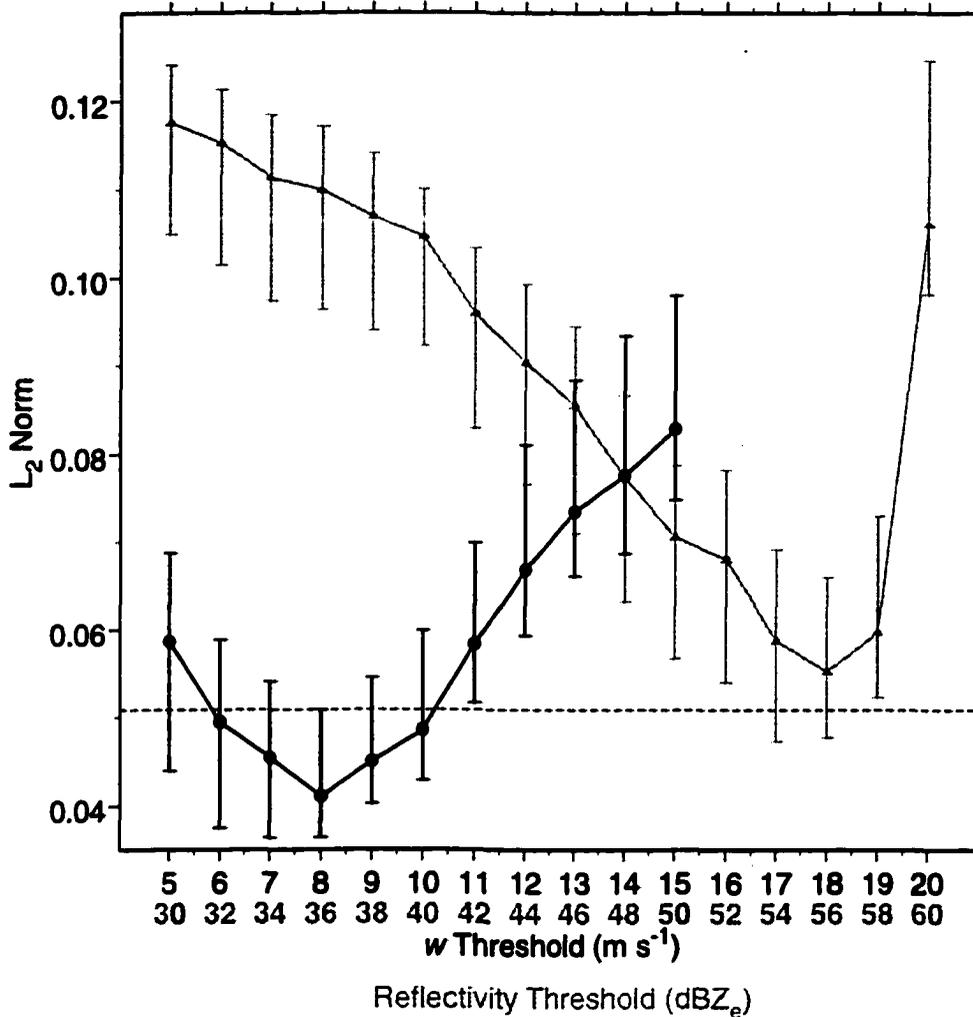


FIGURE 4.5. Euclidean distance between forecast and observed pdf's as a function of threshold for both w (in black) and reflectivity (in grey); 95% confidence limits are shown by capped, vertical bars. Results derived from any value whose confidence interval contains the dotted line are statistically indistinguishable from results derived from the 8 m s^{-1} threshold.

based forecast pdf has double-humped appearance between 12 and 30 min and a secondary peak at 90 min. The peak at 90 min is smaller than the pdf based on an 8 m s^{-1} threshold by about 50% (Fig. 4.7b). While there is no statistical difference between these two pdf's, the w -based pdf has more overlapping area in common with the observed pdf. Also, a 56 dBZ_e

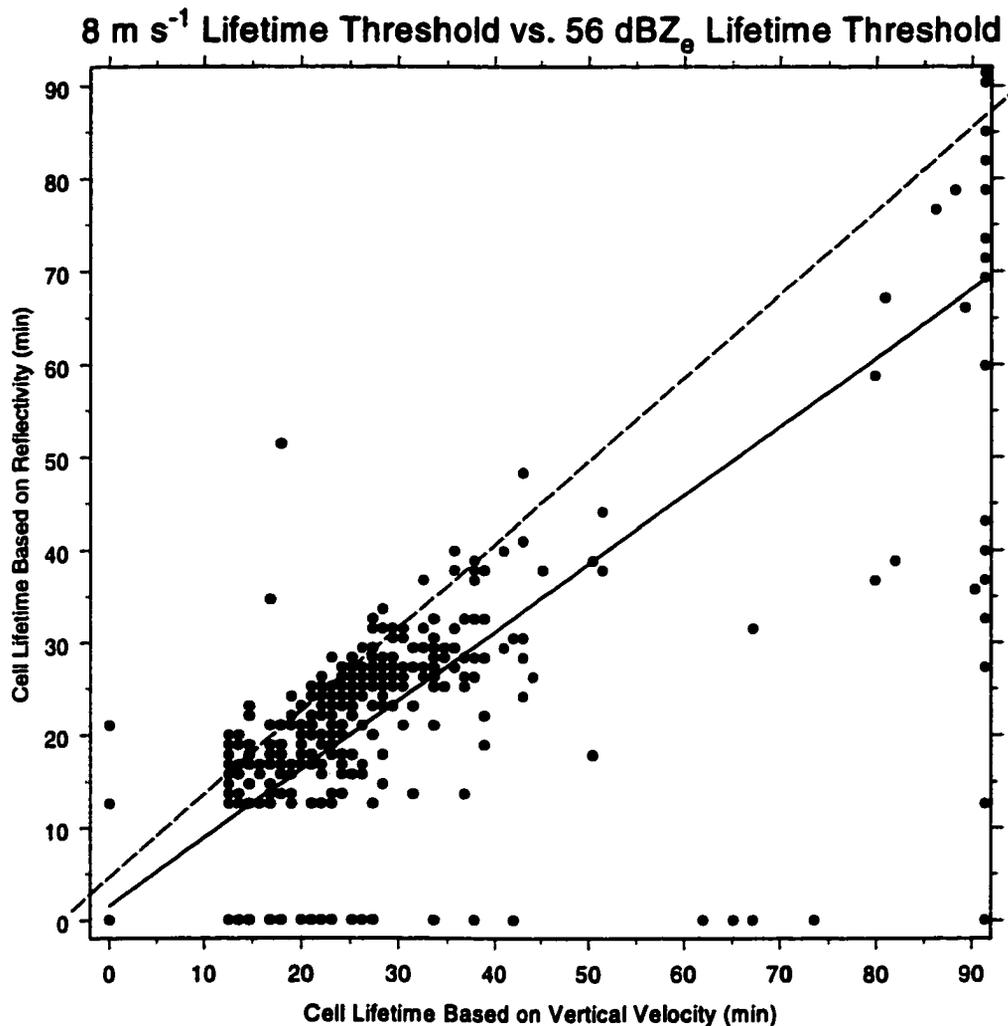


FIGURE 4.6. Cell lifetime threshold based on 8 m s^{-1} and 56 dBZ_e thresholds. The solid line is a linear regression fit to the data while the dashed line shows a line of perfect agreement.

threshold decreases the total number of cells that make up the pdf compared to the w pdf with an 8 m s^{-1} threshold; on some days, a 56 dBZ_e threshold cuts by half the number of cells that can contribute to a pdf forecast. Consequently, w -based forecast pdf's using the 8 m s^{-1} threshold are used for all subsequent analyses.

These pdf's provide insight into limits imposed by the particular cloud modeling approach used in the ensemble. The COMMAS cloud model is initialized with a homoge-

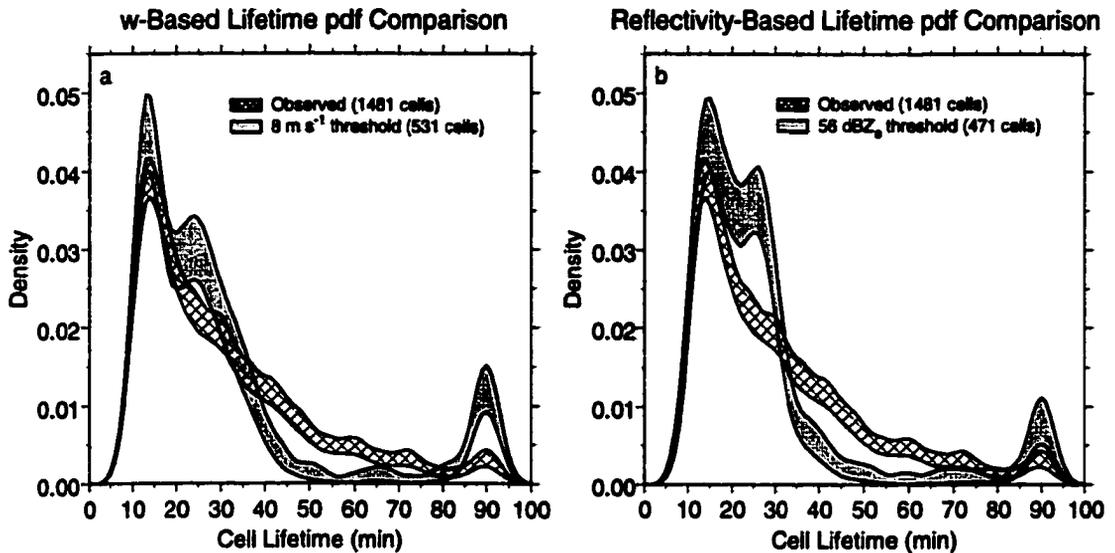


FIGURE 4.7. 95% confidence bounds for forecast (shaded) and observed (crosshatched) lifetime pdf's based on a) 8 m s^{-1} w threshold, and b) 56 dBZ_e reflectivity threshold. Where pdf's overlap, the pdf's are statistically indistinguishable with 95% confidence.

neous environment based on a single initial sounding. Thus, any convection that develops in the model is “trapped” in this environment because a simulated storm cannot propagate into different environmental conditions. A homogeneous environment provides one plausible explanation for the shape of the forecast pdf compared to the observed pdf. Any cell that forms in a marginal forecast environment cannot move into a more favorable forecast environment, even if a favorable environment is the obvious result of the forecast cell motion within the chosen Mesoeta sub-domain. In a like manner, the homogeneous environment may partially explain the excessive number of long-lived cells created by the ensemble model. A cell that starts in a favorable environment cannot propagate into a less favorable environment.

The 0000 UTC verification data suggest that the Mesoeta environment is, on average, less favorable to deep convection than is the real atmosphere. This characteristic is clearly seen in the resulting forecast pdf's of cell lifetime. Cloud scale models also are

well-known for their tendency to evolve storms too quickly, especially when using the Kessler microphysical parameterization (L. Wicker 1999, personal communication). Taken together, this tendency helps explain the dearth of cells with medium lifetimes.

The forecast data set consists of 531 cells and the observed data set consists of 1481 cells. Because the PCA is performed on the observed data using reflectivity and on the forecast data using vertical velocity, the two PCAs cannot be directly compared. However, the Euclidean similarity-based PCA extracts similar modes of cell lifetime from the set of observed cells and the set of forecast cells. Both PCAs produce a long-lived mode lasting 90 min, a medium-length mode and a short mode. The medium-length forecast mode is 37 min long and the medium-length observed mode is 39 min long. The

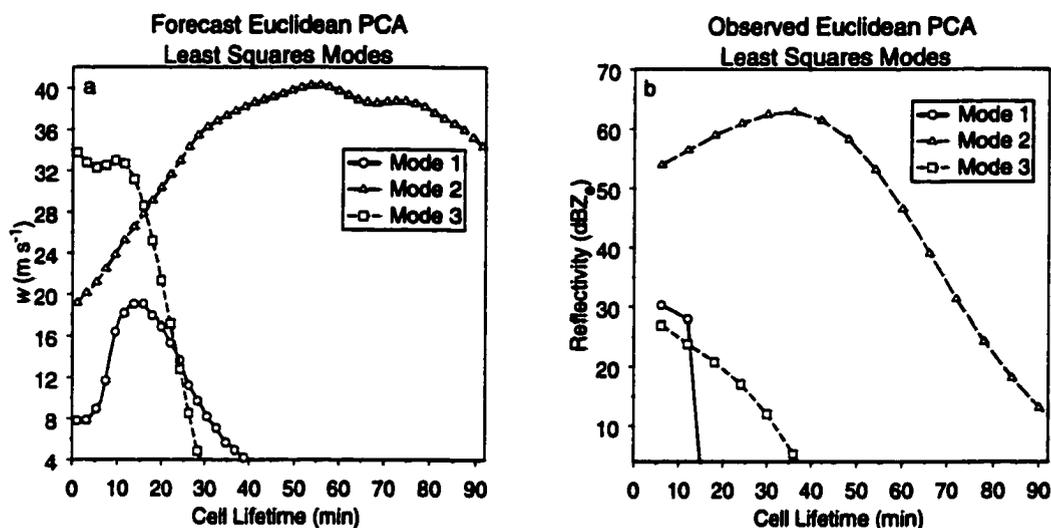


FIGURE 4.8. a) Forecast Euclidean PCA modes based on an $8 m s^{-1}$ threshold. b) Observed Euclidean PCA modes based on a $40 dBZ_e$ threshold.

short-lived forecast mode is 28 min long, while the short-lived observed mode is 15 min long (Fig. 4.8).

A 2×2 contingency table analysis yields further insight into the characteristics of the cloud-model ensemble. While substantial variability exists from day to day, overall

TSS = 0.064, which shows that the ensemble has little skill in helping to anticipate where convection might occur within the region (Table 4.2)

Table 4.2. 2 × 2 Contingency Table Statistics for Long-Lived PCA Modes

Parameter	All Days Combined	Days with Limited Convection	Reflectivity-Based Severe	w-Based Severe
POD	0.615	0.786	0.625	0.625
FAR	0.385	0.389	0.286	0.0
CSI	0.598	0.524	0.500	0.625
TSS	0.064	0.202	0.425	0.625

. Because the ensemble is not designed explicitly to examine or address the question of convective initiation, this result is expected. Such a lack of skill results because the COMMAS model tends to create deep convection at every grid point for at least one forecast sounding time. Too much deep convection is likely related to the warm bubble initiation procedure: identical warm bubbles are used, regardless of horizontal variations in the environment. However, four days occur when COMMAS does *not* produce deep convection at every grid point (“limited convection” days). For these days, TSS = 0.202, which indicates a substantial increase in the TSS. This increase suggests some skill in predicting where convection is likely on days with limited convection (Table 4.2).

Does the ensemble display any skill at identifying days when severe weather occurs? Again, 2 × 2 contingency table statistics help answer this question. For these scores, a “hit” occurs when the PCA applied to the ensemble output produces a long-lived mode (defined as longer than 60 min) and severe weather is observed. A false alarm occurs when the PCA produces a long-lived mode and no severe weather is observed, and a miss is observed when the PCA does not produce a long-lived mode and severe weather

is observed. A correct “no event” forecast is self evident. Similar to forecast lifetime pdf’s, two parameters can be used in the PCA: reflectivity and w .

Because only 8 days exhibit severe weather, these results must be interpreted cautiously. However, the TSS increases to above 0.425 for reflectivity-based lifetimes and 0.625 for w -based lifetimes. Because lifetimes based on reflectivity produce some false alarms, modes produced using a lifetime based on w yield a higher TSS. The only study that addresses a similar statistic is performed for temporal periods of one hour over a spatial domain the size of a single manually digitized radar (MDR) box, and is performed for all MDR boxes contained within severe thunderstorm and tornado watch boxes (Doswell et al. 1993). In contrast, the ensemble results presented here consist of a nine hour temporal period (instead of 1 hr periods) and uses a spatial area equivalent to 13 MDR boxes, instead of using single MDR boxes. Consequently, these scores are not directly comparable because a larger temporal and spatial domain inflates the TSS considerably compared to the results in Doswell et al. (1993). If the forecasts and events are independent and identically distributed, the TSS values reported in this study will be higher than those reported in Doswell et al. (1993), though how much higher is not clear. As expected, the TSS values reported here are larger than the normalized values reported in Doswell et al. (1993). Despite problems in comparing these two studies, the high TSS values suggests that there may be some utility in using a long-lived mode from the PCA as a severe weather indicator.

4.4 Long-lived vs. short-lived storms

When a large number of cloud model runs are available, the differences between soundings that create long-lived storms and those that create short-lived storms can be addressed. Do differences exist between soundings that produce long-lived and short-lived storms that, in hindsight, can be used to predict the resulting lifetimes? This question amounts to a sensitivity study and, while such an analysis is not an explicit goal of this work, the data are available to examine such sensitivities. If reliable indicators of storm longevity are revealed, then a need to run the cloud model ensemble may not exist, because an approximate lifetime can be extracted directly by examining certain sounding characteristics that may have been missed in earlier analyses of the Mesoeta model soundings (Stensrud et al. 1997).

To make sure that the demarcation between short and long-lived storms is distinct and without overlap, a maximum lifetime of 40 min is used to define short-lived storms, and a minimum lifetime of 60 min is used to define long-lived storms. Having such a clear distinction between short- and long-lived storms helps insure that any differences noted in the parent soundings are unmistakable.

Some characteristic in the perturbations may discriminate between long- and short-lived cells. Because the ensemble sometimes creates long- and short-lived storms based on data from the same day, an obvious approach examines differences in perturbation values of the raw sounding parameters, θ , q , u , and v . To facilitate this investigation, perturbation values θ' , q' , u' , and v' are constructed for each day. For example, $q' = \bar{q} - q$, where \bar{q} is the daily mean value at a particular level and q is the actual value at the same level from a single sounding. Perhaps long-lived storms possess statistically higher mixing

ratios near the surface. A t-test on the quantity $(\bar{q}'_{\text{long}} - \bar{q}'_{\text{short}})$, where \bar{q}'_{long} applies to long-lived storms and \bar{q}'_{short} applies to short-lived storms, is used to determine if the mean perturbations are significantly different from zero at any level. As done previously for the verification statistics, the 95% confidence interval for $(\bar{q}'_{\text{long}} - \bar{q}'_{\text{short}})$ is plotted as a function of height. If the confidence interval contains zero difference, no significant difference in the perturbation quantity for short- and long-lived storms exists. Unfortunately, when examined this way, perturbation soundings provide no discrimination for lifetime (Fig. 4.9).

A more sophisticated approach examines various parameter spaces constructed from the derived sounding parameters to determine if derived sounding parameters can help discriminate between long- and short-lived storms. Derived parameters clearly help forecasters anticipate the nature of convection that might occur. But, whether or not they are useful in discriminating if long- or short-lived storms will result is unknown.

One common derived parameter space is maximum CAPE vs. BRN, an approach similar to that used in Weisman and Klemp (1982). CAPE vs. BRN is plotted for each initial sounding that results in deep convection, using distinctive symbols for short- and long-lived storms. If the combination of these two parameters can discriminate between long- and short-lived cells, distinct clusters of points for each type will appear. For long-lived cells, there is at least some success because long-lived cells tend to exist in the area where CAPE exceeds 500 J kg^{-1} and BRN is less than 60 (Fig. 4.10). Unfortunately, because many short lived-cells are mixed in with long-lived cells, and because a few long-lived cells exist for BRN between 60 and 100, this discrimination is deficient. The boundary

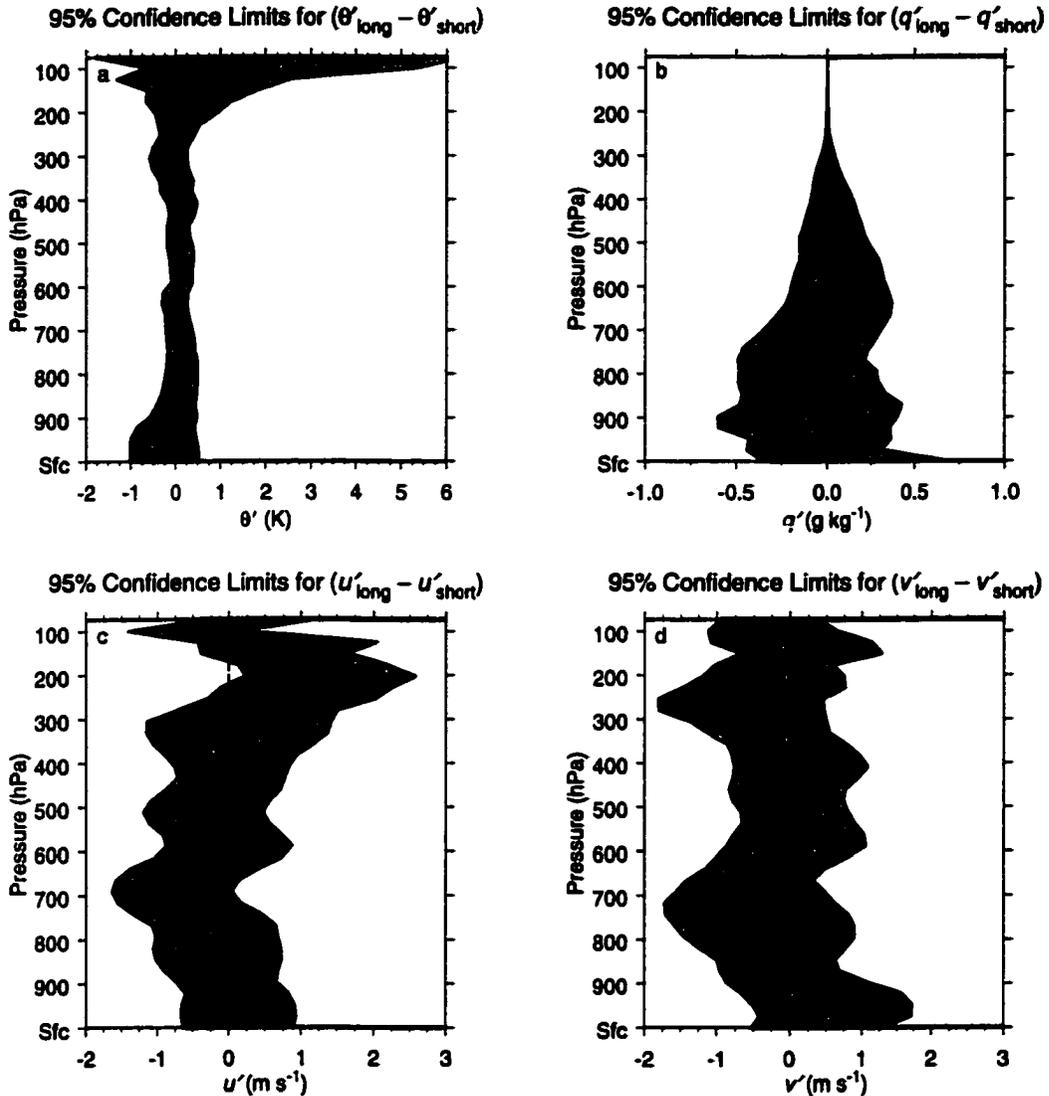


FIGURE 4.9. 95% confidence limits (shaded area) for the difference between perturbation a) θ , b) q , c) u , and d) v parameters associated with short- and long-lived storms. Perturbations are calculated using daily mean values. The vertical dashed line is the zero-difference line which, when contained in the shaded region, indicates no significant difference at the 95% confidence level.

between long- and short-lived cells becomes more diffuse when CAPE is less than 1000 J kg^{-1} , but considerably more distinct when CAPE exceeds 2000 J kg^{-1} . Consequently, limited discrimination is provided with the phase space defined by CAPE and BRN.

Another, related, parameter space compares the maximum CAPE to the BRN shear. Unfortunately, this combination provides no discrimination between short- and

long-lived storms because both are mixed together such that there is no distinct cluster that contains primarily one or the other (Fig. 4.11a). While short-lived cells occur primarily where BRN shear is less than 50 s^{-1} , several long-lived cells are mixed into this region. When only those days that produce both long-lived *and* short-lived cells are used, excluding all days when either only long- or short-lived cells result, no significant improvement is noted. In fact, the mixing of long- and short-lived cells is even more apparent (Fig. 4.11b).

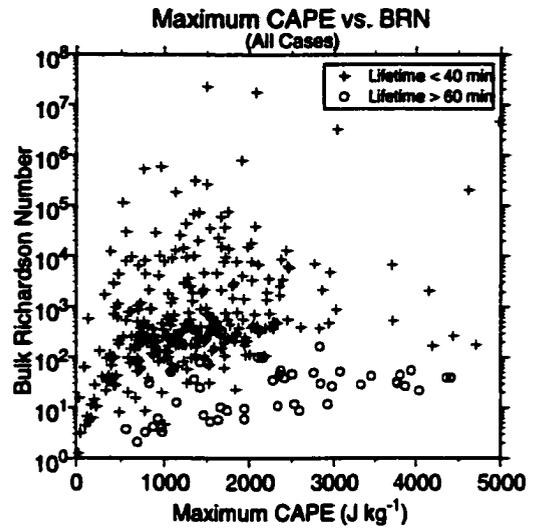


FIGURE 4.10. Forecast lifetime plotted as a function of max CAPE and BRN. The best discrimination is provided when CAPE is greater than 2000 $J\ kg^{-1}$ and BRN is less than 60.

significant improvement is noted. In fact, the mixing of long- and short-lived cells is even more apparent (Fig. 4.11b).

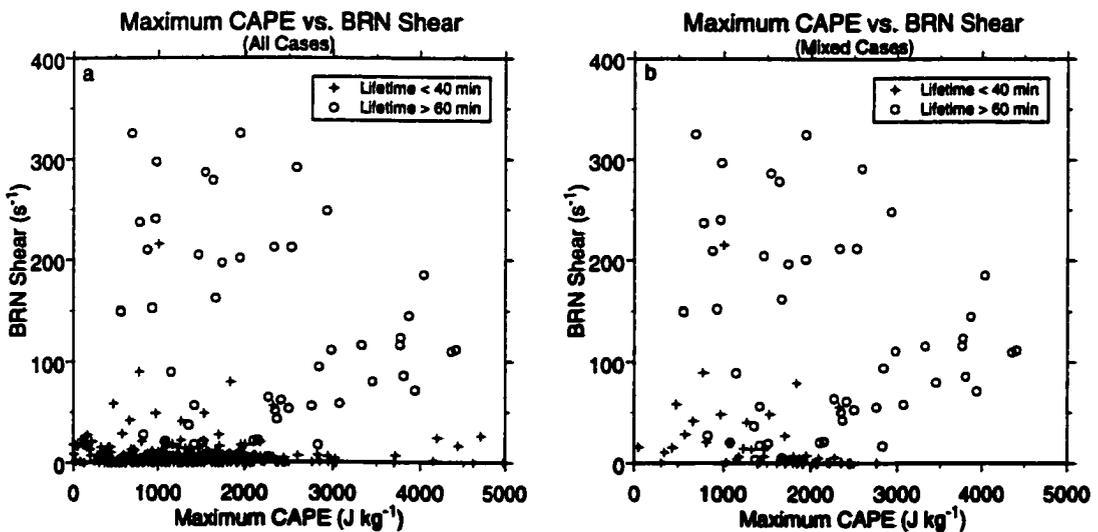


FIGURE 4.11. Same as Fig. 4.10, but plotting max CAPE vs. BRN for a) all days combined, and b) only those days that produced both short-lived and long-lived cells.

Storm type can be categorized based on observed storms with SREH as a parameter (Brooks et al. 1994). In one scenario, CAPE is plotted against SREH along with lines

of constant equivalent helicity index (EHI) ($EHI = CAPE \times SREH / 160,000$). EHI has been proposed as a discriminator between storms that produce strong tornadoes and those that produce violent tornadoes. Unfortunately, while modeled short-lived storms are not found where $EHI > 2.5$, long-lived storms tend to exist everywhere within this parameter space (Fig. 4.12a). Thus, this parameter space does only part of the job, because conditions that preclude long-lived cells are not identified.

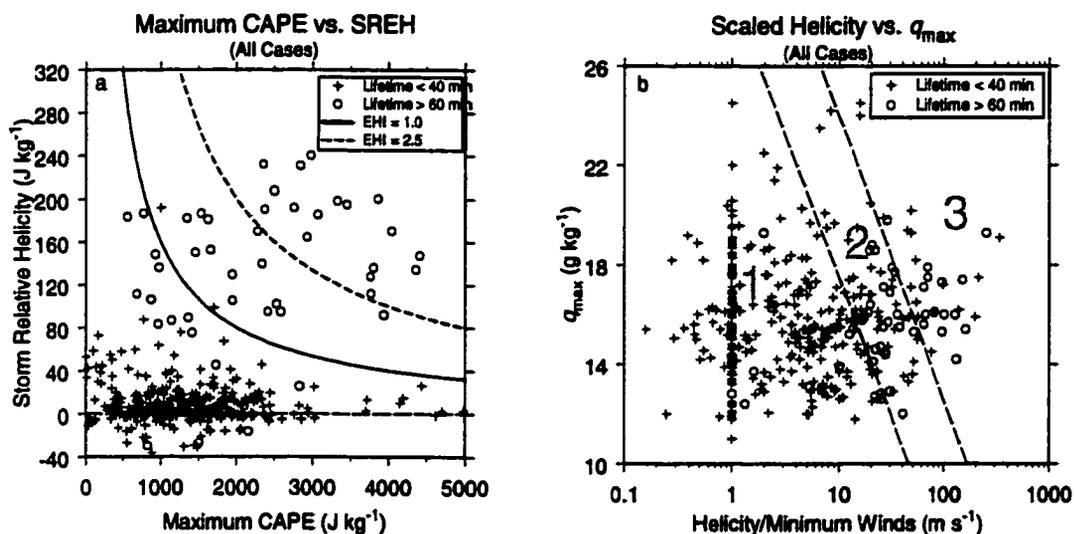


FIGURE 4.12. Parameter space graphs using SREH as a dimension. In a) maximum CAPE vs. SREH, the horizontal dashed line is the zero SREH reference line, while the solid and dashed curves are $EHI=1.0$ and $EHI=2.5$, respectively. In b) is shown SREH scaled by the minimum storm-relative wind vs. the maximum mixing ratio. Zone 1 contains observed LP supercells, zone 2 contains observed storms with tornadic mesocyclones, and zone 3 contains observed storms associated with extreme wind gusts (after Brooks et al. 1994).

Another variant on this parameter space uses SREH scaled by the minimum storm-relative wind as one dimension, and the maximum mixing ratio as the other dimension. One theory about low-level mesocyclogenesis concerns a balance between helicity and storm-relative, environmental winds at mid-levels. When SREH is scaled by the minimum storm relative wind, the two components of this balance are reduced to a single value (Brooks et al. 1994). Based on observations, this parameter space is further divided into

three zones. Storms in the Zone 1 environment tend to be of the low-precipitation (LP) supercell type. Zone 2 tends to contain tornadic storms, and Zone 3 tends to contain storms that create extreme wind events that are not associated with tornadoes (Fig. 4.12b). Clearly, this particular parameter space offers no discrimination between short- and long-lived storms that are modeled.

Other parameter spaces defined by perturbations from daily mean values are also examined. These additional parameter spaces are perturbation BRN vs. perturbation CAPE, perturbation BRN shear vs. perturbation CAPE, and perturbation BRN vs. perturbation SREH. Clearly, such perturbation values have meaning only for those cases when both long- and short-lived storms coexist. Unfortunately, none of these combinations offer any discrimination between long- and short-lived storms (not shown).

If very small differences in the parent soundings can produce substantial differences in the resulting storms, then the preceding results are not surprising. To show that modeled storms are, indeed, sensitive to very small differences in parent soundings, three cases are examined.

For the 6/7 June 1996 case, very small thermodynamic and kinematic differences below 700 hPa generate a storm with a 26 min lifetime and another storm with a 92 min lifetime (Fig. 4.13). Current sensing technology is accurate to ± 1 K, ± 1 g kg⁻¹ and ± 1 m s⁻¹ (Crook 1996), but below 700 hPa, the differences between these two soundings fall within the noise level of measurement sensors. Above 700 hPa, both soundings appear to be contaminated by parameterized convective processes within the Mesoeta model. Regardless of these convective parameterization effects, these soundings reveal the remarkable aspect

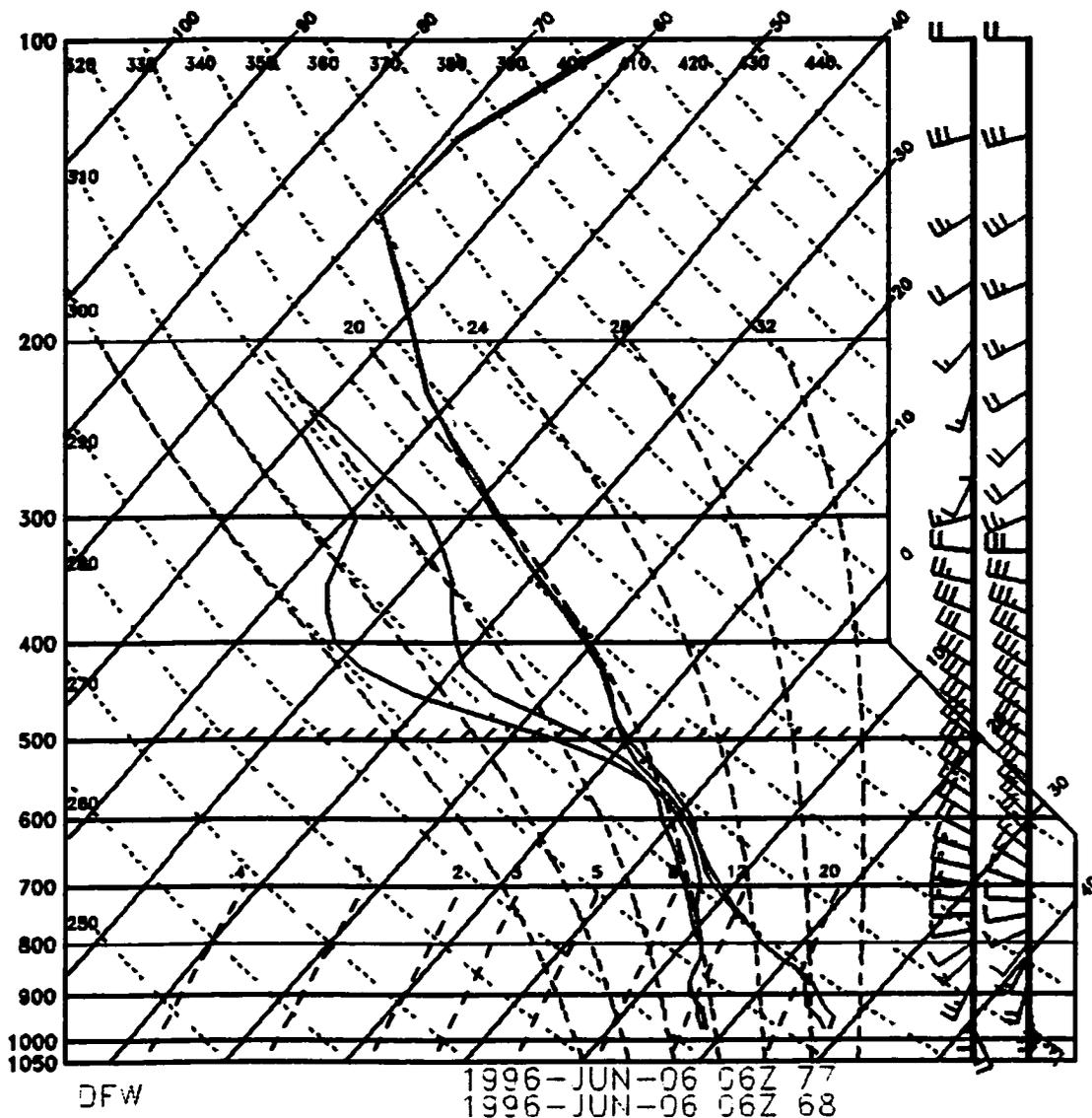


FIGURE 4.13. Soundings from the 6/7 June 1996 case, showing the sensitivity of storm lifetime to the parent soundings. The green sounding results in a storm that lasts the maximum possible time of 92 min, while the blue sounding results in a storm that lasts only 26 min. Below 500 mb, differences between these two soundings are less than 1 g kg^{-1} and 1 K .

that two soundings, containing what appear to be insignificant differences, result in storms with very different lifetimes.

A different situation occurs in the 8 July 1996 case (Fig. 4.14). The thermodynamic differences between these two soundings are everywhere less than 1 K and 1 g kg^{-1} , differences that cannot be resolved using current observation systems. In one case, a storm

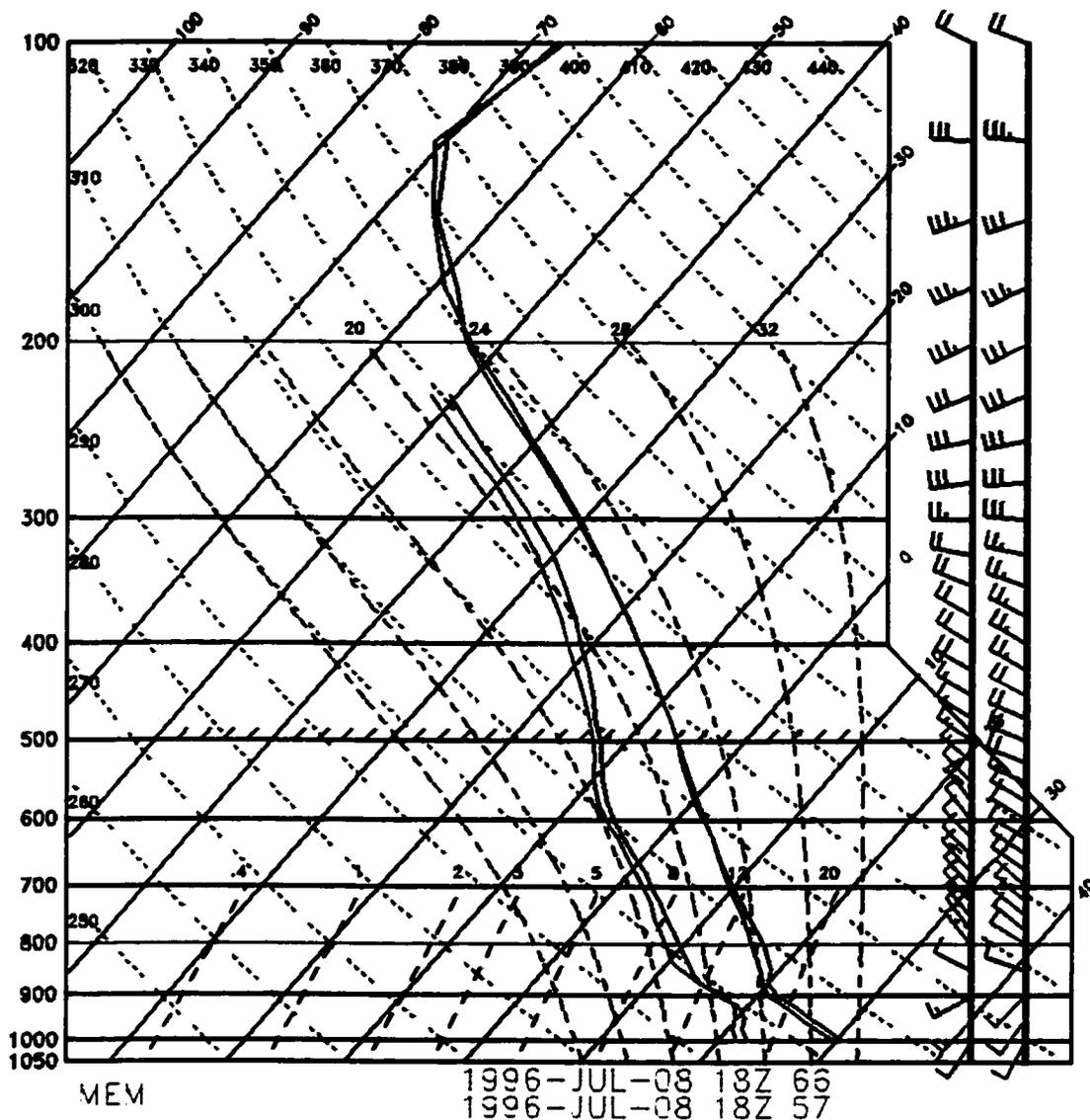


FIGURE 4.14. Same as Fig. 4.13, but for 8 July 1996. Here, the green sounding results in a cell lasting 62 min, while the blue sounding results in no convection. The only obvious difference is the slight inversion near 870 hPa in the sounding that produces no convection. The thermodynamic differences between these soundings are no greater than 1 K or 1 g kg^{-1} anywhere in soundings.

lasting 62 min is generated while, in the other case, no storm at all is generated. The sounding that does not generate a storm has a weak inversion centered near 850 hPa. In all likelihood, this sounding would result in deep convection if a stronger initial bubble were used. While this distinction is apparent upon inspection, it illustrates that the process of using a warm bubble to initiate convection in a cloud-scale model is inadequate. This case

also shows that the convective initiation process, which is poorly understood at present, is an important aspect of cloud-scale modeling.

Two soundings from the 14 July 1995 case are thermodynamically indistinguishable below 650 hPa, yet one produces a cell lasting 92 min while the other produces a cell lasting 25 min (Fig. 4.15). If the thermodynamic differences between these two soundings cannot be accurately measured, it is unlikely that they can be accurately forecast. Differences between the winds are $1\text{-}2\text{ m s}^{-1}$ below 650 hPa. However, current observing systems cannot resolve these kinematic differences. This case appears to be evidence of chaotic sensitivity, wherein infinitesimal differences in initial conditions result in disproportionately large differences in the outcomes. In many cases, soundings that possess differences which are easily measurable do *not* produce storms with significantly different lifetimes (not shown). This fact indicates that the cloud model is more sensitive in some environments than in others.

Evidently, there is no way to determine *a priori* how long a thunderstorm is going to last, even when the parent environmental sounding (instead of a less-representative proximity sounding) is known. Consequently, it appears that a cloud model is needed to estimate the storm lifetime given a particular sounding. Further, it appears that the modeled lifetime of a storm is sometimes highly sensitive to the initial sounding. Thus, it seems that an ensemble of cloud models is the only reliable method available to extract the range and distribution of storm lifetimes from mesoscale model output.

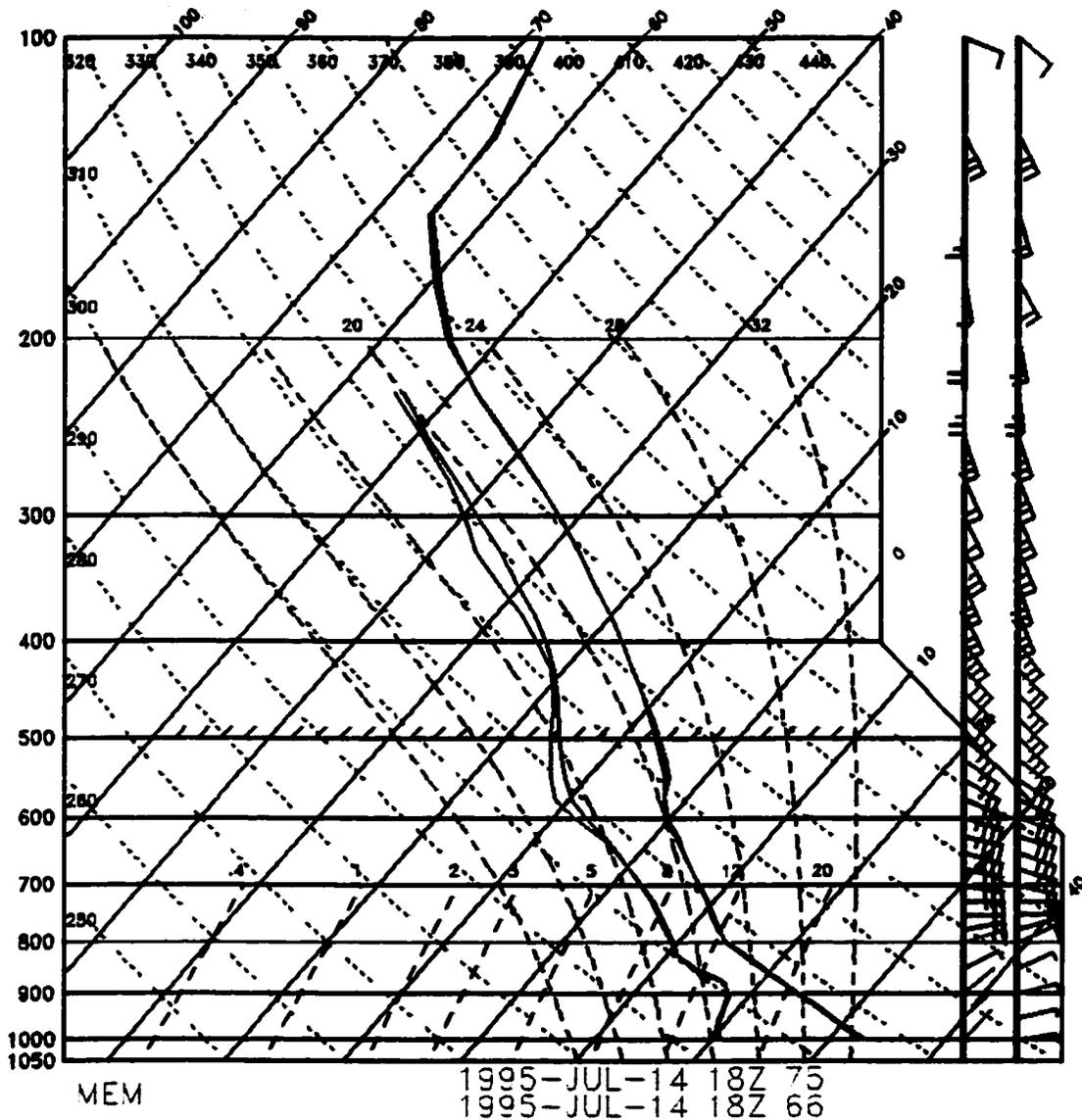


FIGURE 4.15. Same as Fig. 4.13, but for the 14 July 1995 case. The green soundings results in a cell that lasts the maximum of 92 min, while the blue sounding results in a cell that lasts only 25 min.

4.5 Specific cases

From the 18 available cases, three specific cases are presented. These typify the range of characteristics contained in all 18 cases. These representative cases help to illustrate daily ensemble model output, interpretation of the ensemble model output, and verification of the ensemble model forecasts.

4.5.1 30 August 1996

The 30 August 1996 case is from the Memphis, TN, area. While severe weather was not reported on this day, this case displays sensitivity to where convection occurs, which is valuable. However, the forecast lifetimes do not span a large enough range. Hence, the forecast and observed lifetime pdf's are different. Because the forecast lifetimes do not contain any long-lived cells, the forecast PCA modes also miss the long-lived mode in the observed data.

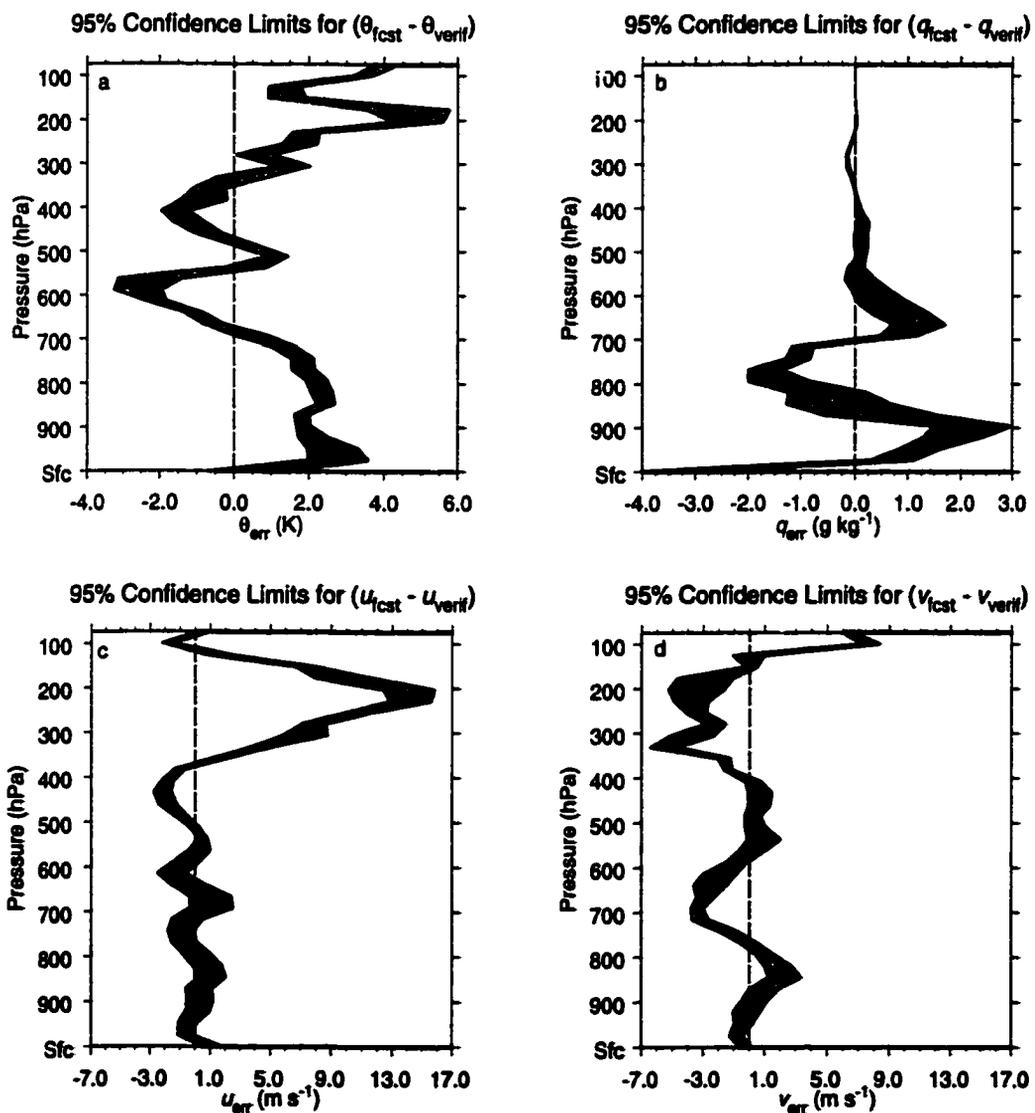


FIGURE 4.16. Same as Fig. 4.1, but for 30 August 1996.

The main errors in the Mesoeta forecasts of the θ field for this day indicate that the layer from just above the surface to 700 hPa is too warm by nearly 2 K, and too cold between 700 to 550 hPa by as much as 2 K. Smaller errors exist up to about 300 hPa, where θ is again too warm (Fig. 4.16a). This structure implies that the static stability is forecast to be insufficient (too unstable), in agreement with general characteristics of the Mesoeta verification from 1995 and 1996.

Errors in mixing ratio display a dry bias of 4 g kg^{-1} at the surface, which quickly reverses to a moist bias of $1\text{-}2 \text{ g kg}^{-1}$ from just above the surface to 850 hPa (Fig. 4.16b). Given this structure, surface-based CAPE estimates are low, but CAPE values resulting from a layer mean (across a depth of 100 hPa) mixing ratio are too high. Because there is a slight dry bias between 850 and 700 hPa, and because the Kessler precipitation parameterization is used, resulting downdrafts are likely to be too cold and too strong.

Errors in the u component are not significant between the surface and 525 hPa (Fig. 4.16c). However, errors in the v component are significant above 850 hPa, with a maximum northerly error of 3 m s^{-1} near 700 hPa (Fig. 4.16d). For both components, large errors occur near the tropopause level. In summary, Mesoeta forecast errors on this day tend towards insufficient static stability and insufficient surface-based moisture, but slightly excessive elevated boundary layer moisture. Whether forecast CAPE values are high or low depends upon how CAPE is calculated. Wind forecasts in the lower troposphere are good, which yield relatively accurate SREH values.

Given the mean forecast sounding, perturbation distributions can be computed and plotted, as in Fig. 4.3, but for individual days (Fig. 4.17). For 30 August 1996, at least half of all input soundings contain perturbations of $\pm 1 \text{ K}$ in potential temperature and $\pm 1 \text{ g kg}^{-1}$

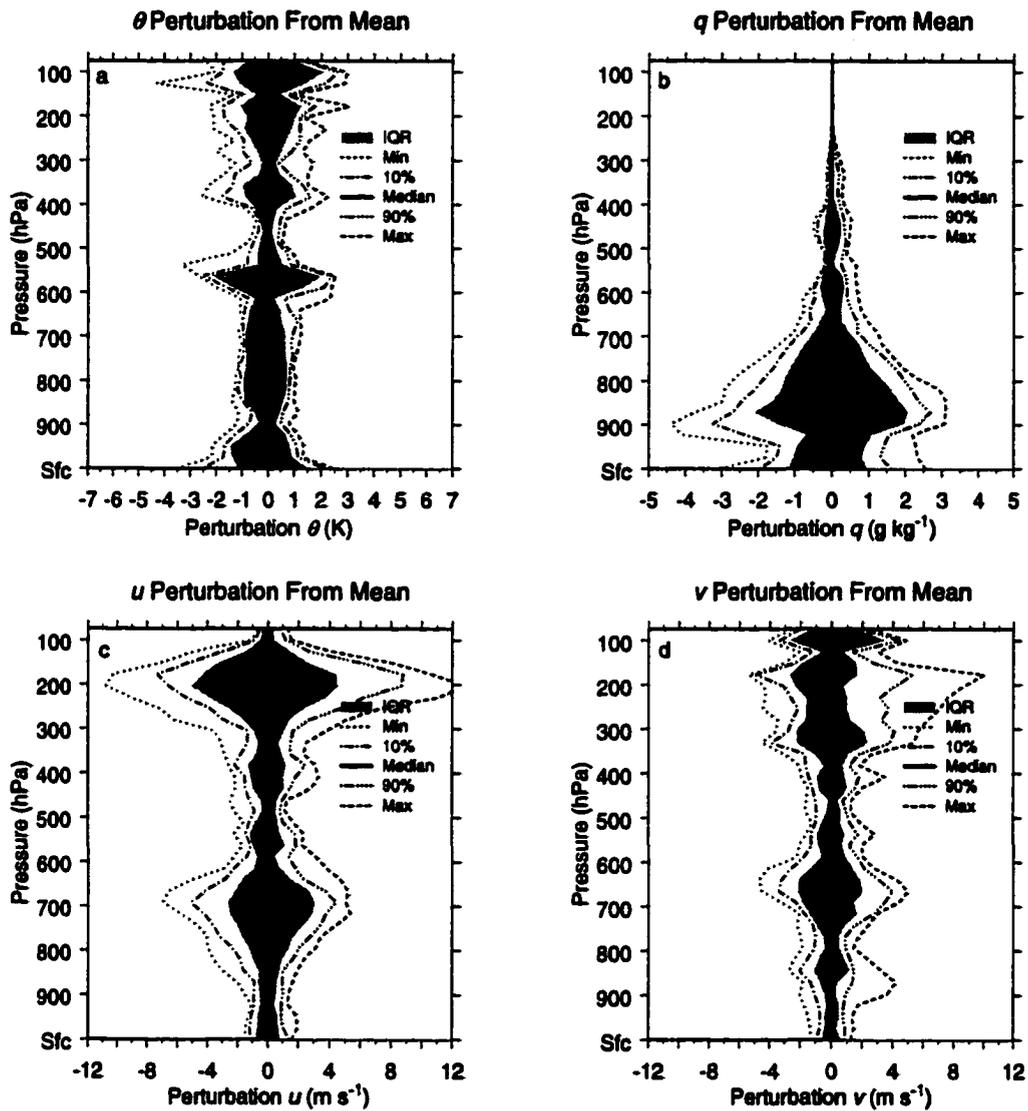


FIGURE 4.17. Same as Fig. 4.3, but for 30 August 1996.

¹ in mixing ratio. These small perturbations have been shown to be significant for the development of convection (Crook 1996). The perturbations remain large throughout the depth of the sounding, except for the fact that mixing ratio tends to zero as pressure decreases. In half of the soundings, perturbations in the u and v components are about 1 m s^{-1} in the boundary layer, and increase to $\pm 3 \text{ m s}^{-1}$ near the tropopause.

Because long-lived cells are not created within the ensemble, none appear in the modes recovered through PCA (Fig. 4.18). Three modes are needed to characterize the

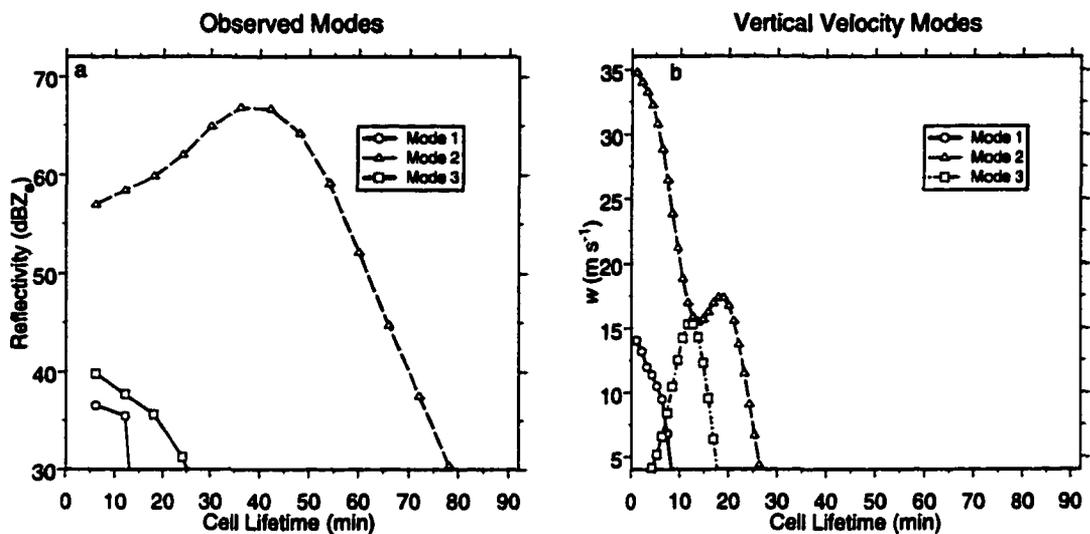


FIGURE 4.18. Observed PCA modes for a) observed reflectivity, and b) forecast vertical velocity modes. This ensemble fails to produce the lifetime range that is observed and so is under-dispersive. Modes resulting from observed data explain 79.9% of the Euclidean similarity while modes resulting from the forecast data explain 82.4% of the Euclidean similarity.

forecast data, and three modes are needed to characterize the observed data. However, the observed long-lived mode is not produced by the ensemble model. As a result, the forecast and observed lifetime pdf's show significant differences (Fig. 4.19). In particular, the observed lifetime pdf includes cells that last at least 90 min while none of the ensemble members lasted longer than 32 min. Consequently, the lifetime of storms created by this ensemble lack sufficient range, or is under-dispersive.

When applied to the reflectivity-based and w -based lifetime estimates, the PPP test indicates that the distribution of storm lifetimes based on either w or reflectivity are significantly different from the observed distribution of storm lifetimes (Fig. 4.20). This result is expected, given that the ensemble model results are under-dispersive.

95% Confidence Limits for Observed and Forecast Cell Lifetimes

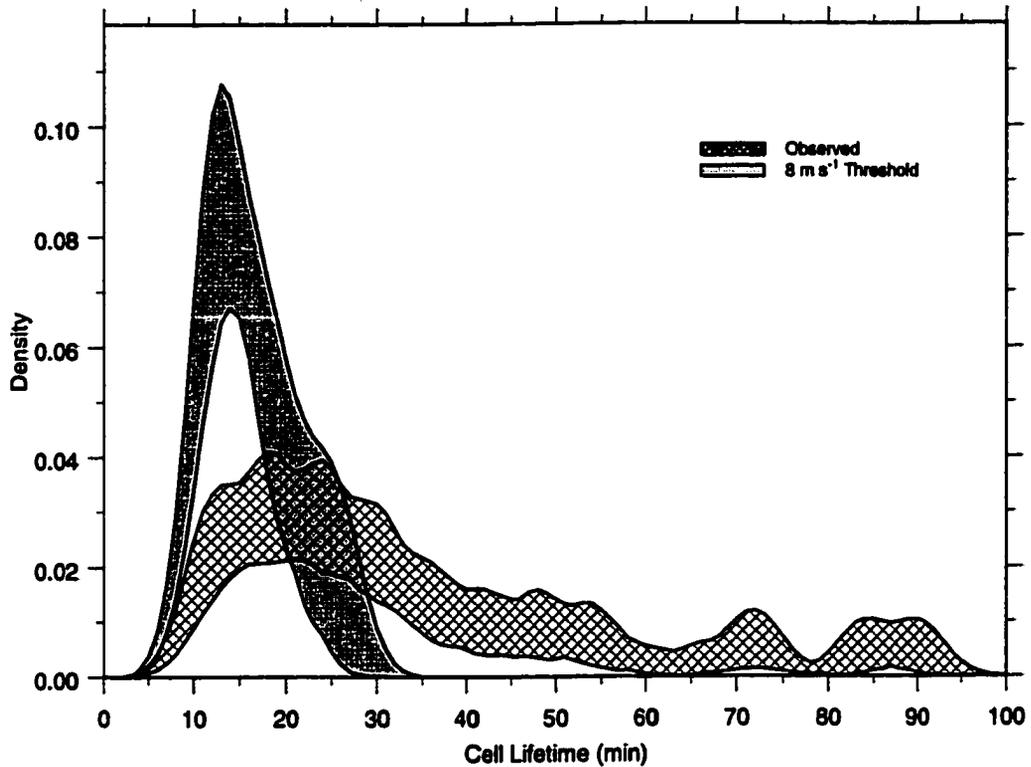


FIGURE 4.19. Forecast and observed pdf's as in Fig. 4.7a, but only for 30 August 1996.

On this day, observed thunderstorms are generally confined to a region south of an east-west line extending through the Memphis International Airport. This delineation is well represented by the ensemble (Fig. 4.21). The skill scores resulting from this forecast are: POD = 0.875, FAR = 0.222, CSI = 0.700, and TSS = 0.475. These statistics suggest that the ensemble output contains trustworthy spatial information about where convection will occur. If spatial discrimination is possible, useful information is available to air traffic control, which could plan to bring flights in from the north during the period covered by this forecast guidance.

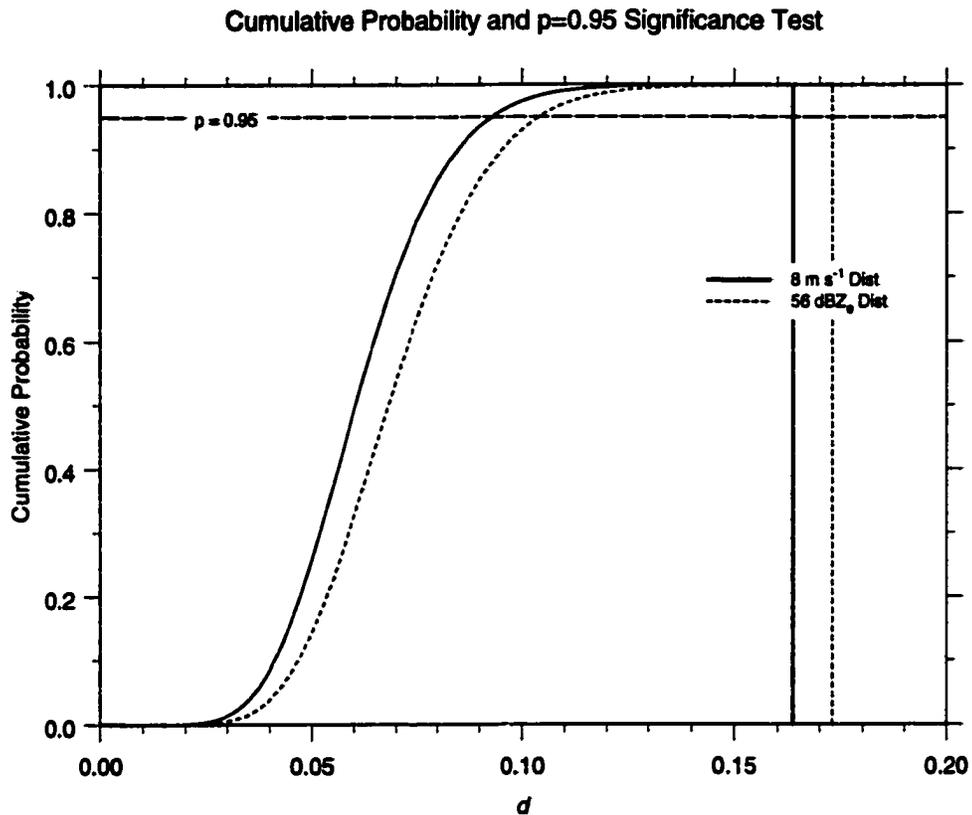


FIGURE 4.20. PPP results for both reflectivity and w -based lifetime estimates from the 30 August 1996 ensemble. Dashed lines are for the reflectivity-based estimates of storm lifetime, while solid lines are for the w -based lifetime estimates. Vertical lines show d_0 for each parameter. According to these results, regardless of the parameter used, the estimates of storm lifetime come from a population distinct from the population of observed storm lifetimes.

4.5.2 14 July 1995

Also from the Memphis area, the 14 July 1995 case is an example of good agreement between the forecast lifetime pdf and the observed lifetime pdf. Reinforcing this interpretation, the PPP test indicates that there is no statistical distinction between the parent population of forecast lifetime and the parent population of observed lifetime. The PCA modes for the forecast and observed lifetimes are similar. Unfortunately, this case does not provide any spatial information about where convection will occur.

Cell and Report Locations for 30 August 1996

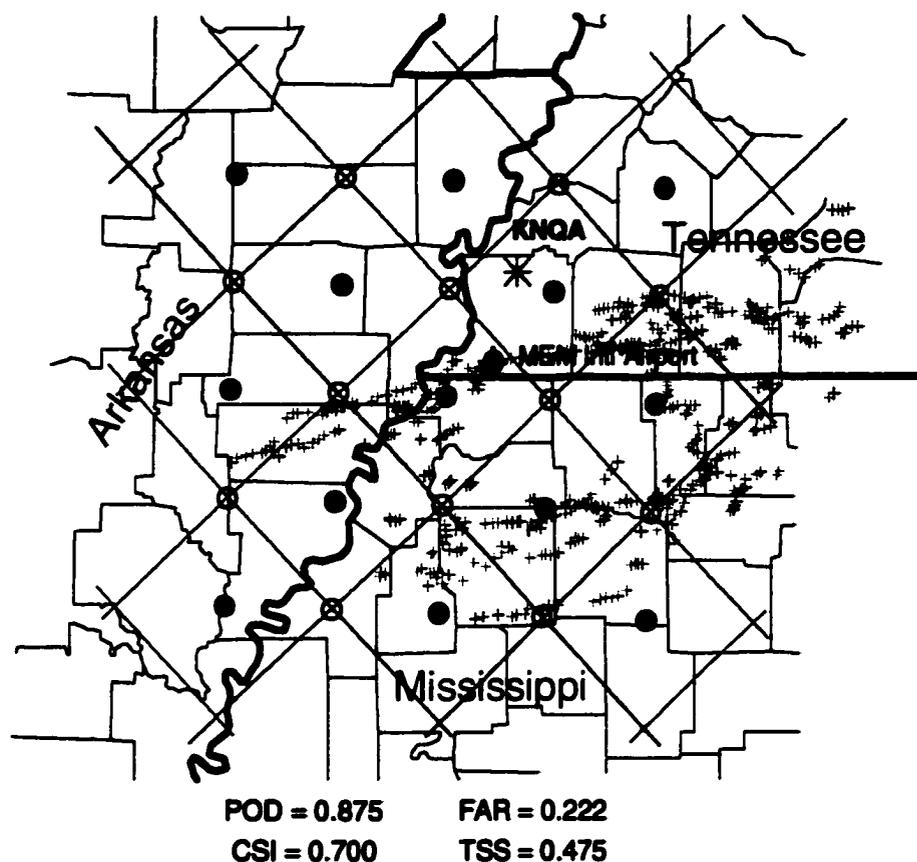


FIGURE 4.21. Observed storm locations for 30 August 1996. Grey crosses show locations for all storms that last 30 min or less, and light red crosses show the same for cells that last longer than 30 min. Green dots represent locations that generate soundings which result in deep convection (defined as a cell that lasts longer than 6 min with w at least 8 m s^{-1}) within the ensemble model, while black dots represent locations that generate soundings which do not generate deep convection. The light red lines show the 40 km L_1 distance from each location. Skill scores are shown at the bottom of the figure.

On this day, the Memphis area is south of a surface anticyclone that extends from the surface to at least 500 hPa. As a result, the area south of Memphis is dominated by easterly flow from the surface through mid-levels. Hence, the lower atmosphere is characterized by subtropical easterly flow. Verification results show that the θ field is forecast to be too warm at the surface, but too cool by 2 K just above the surface to 800 hPa (Fig. 4.22a). Between 800 and 700 hPa, errors in the θ field are insignificant. From 575 hPa to

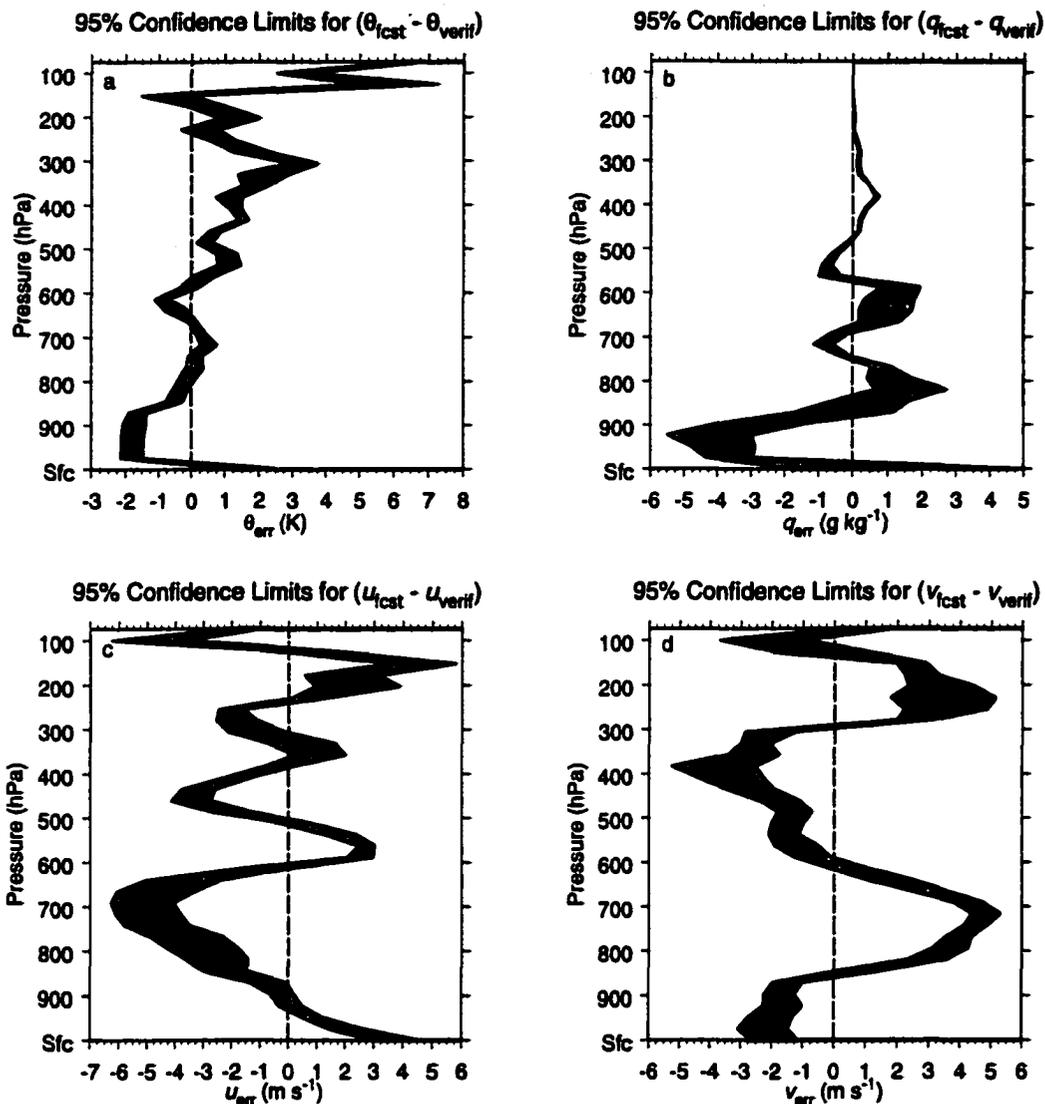


FIGURE 4.22. Same as Fig. 4.1, but for 14 July 1995.

the tropopause, the θ field is again too warm. Consequently, the forecast soundings have excessive static stability in the layer from 925 to 800 hPa. Mixing ratios are accurate at the surface, but too low by as much as $5\ g\ kg^{-1}$ from just above the surface to 850 hPa. Above 800 hPa, errors in the forecast mixing ratio are $\pm 1\ g\ kg^{-1}$ (Fig. 4.22b). Errors in the u and v components indicate southwest winds that are too strong from the surface to 900 hPa.

From 900 hPa to 500 hPa, southeast winds are too strong (Figs. 4.22c and d). This fact suggests that the forecast height gradient is too strong.

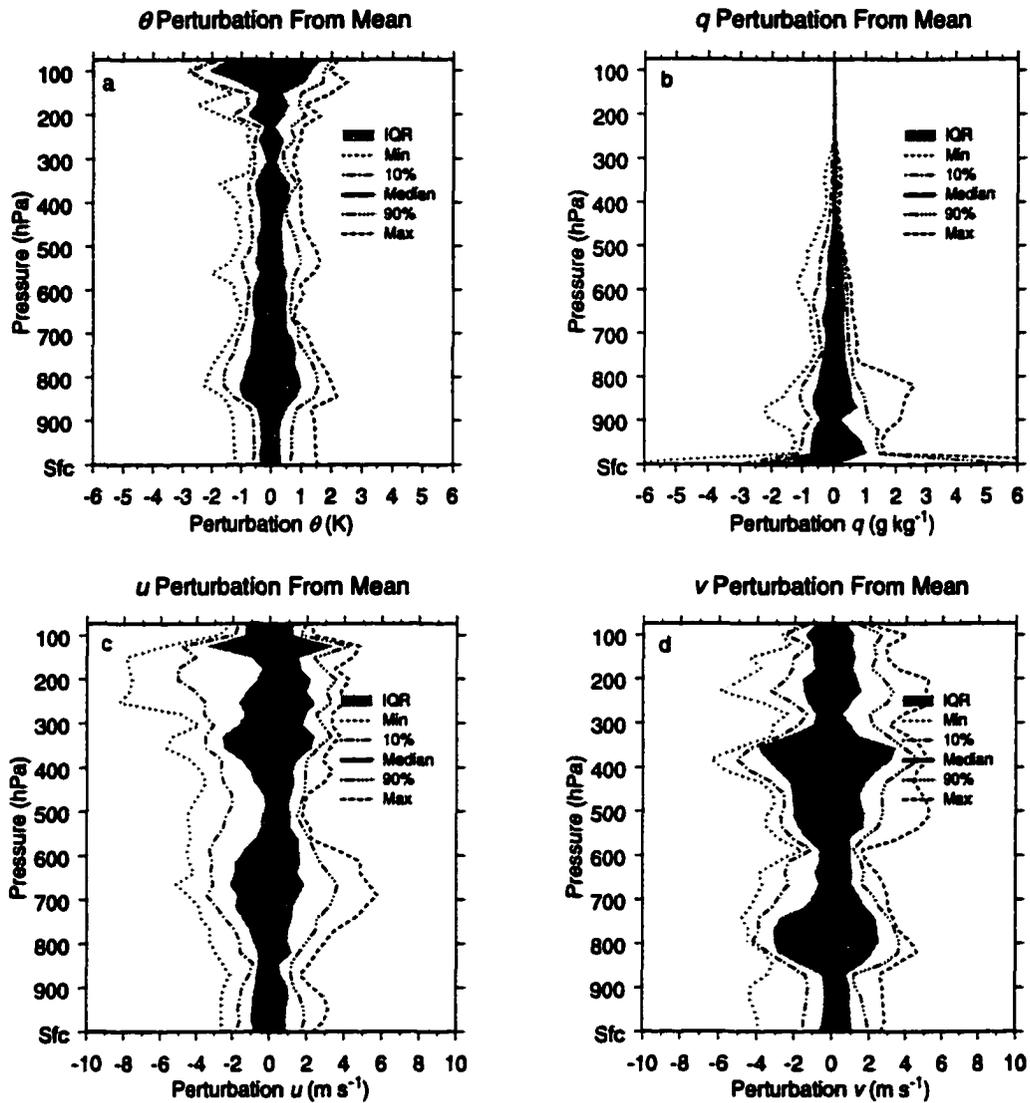


FIGURE 4.23. Same as for Fig. 4.15, but for 14 July 1995.

From the surface to 800 hPa, the initial conditions contain temperature perturbations that are generally less than ± 1 K, but above 800 hPa the range of perturbations exceeds ± 1 K (Fig. 4.23a). The perturbations in mixing ratio are skewed significantly from the surface to 975 hPa. This skewness indicates that most soundings display mixing ratios

less than the mean (hence forecasts that are too dry), while a few soundings display mixing ratios significantly above the mean. The perturbation range is at least $\pm 1 \text{ g kg}^{-1}$ from the surface to 850 hPa (Fig. 4.23b).

Perturbations in the u component are $\pm 1\text{-}2 \text{ m s}^{-1}$ throughout the depth of the atmosphere, but perturbations in the v component are $\pm 2\text{-}3 \text{ m s}^{-1}$ from 850 hPa to 700 hPa, and from 550 hPa to 350 hPa. The first maxima is co-located with the largest positive bias in the v component, where the 95% confidence interval widens. The second is co-located with the large negative bias in the forecast v component (Fig. 4.23c and d).

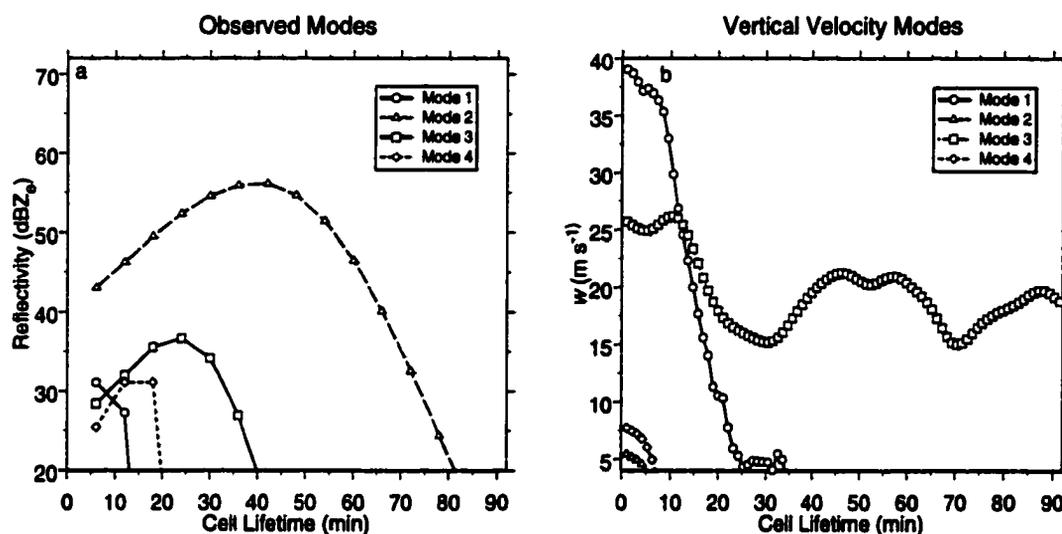


FIGURE 4.24. Same as fig. 4.18, but for 14 July 1995. Modes resulting from observed data explain 82.9% of the Euclidean similarity while modes resulting from the forecast data explain 85.1% of the Euclidean similarity.

The PCA modes from the ensemble suggest that the simulated storms display three behaviors: 1) lifetimes of less than 20 min and weak, 2) lifetimes of approximately 30 min and very intense, and 3) long-lived storms with moderate intensity. For forecast and observed storms, four modes characterize at least 80% of the Euclidean similarity (Fig.

4.24a and b). The forecast modes have lifetimes of 6, 8, 34 and 92 min, which qualitatively match the observed modes with lifetimes are 13, 20, 40 and 82 min.

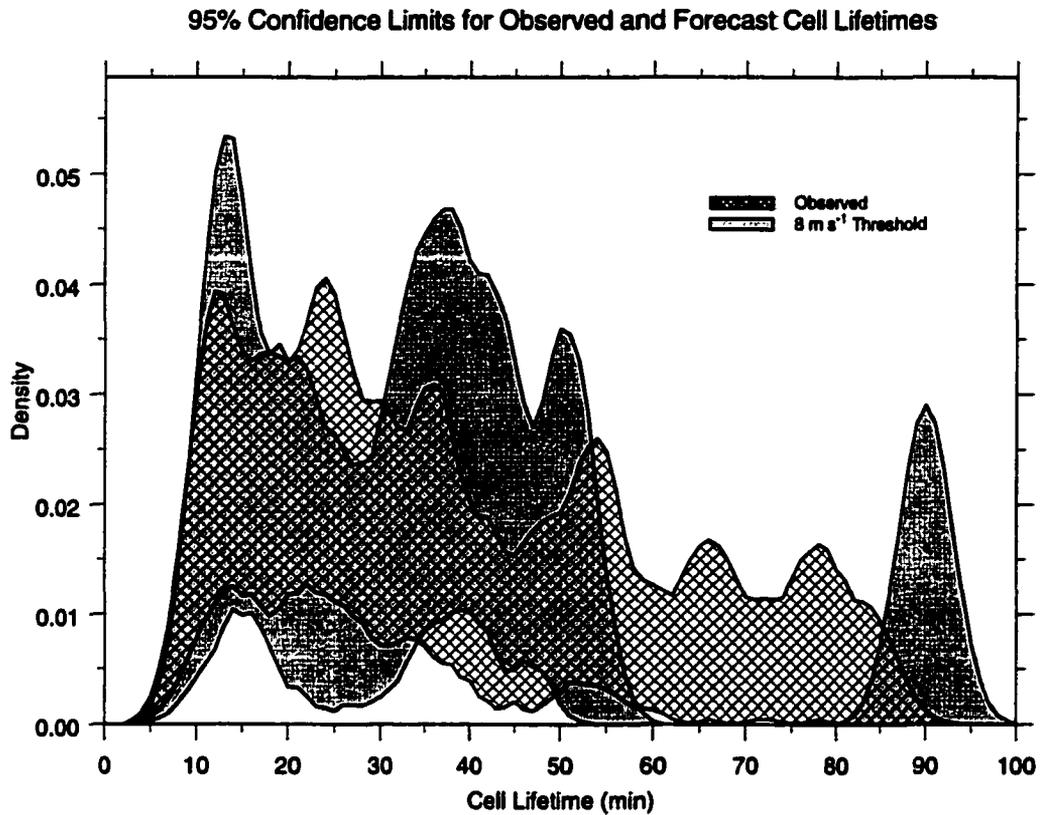


FIGURE 4.25. Same as Fig. 4.18, but for 14 July 1995.

The pdf estimates agree moderately well between those obtained from the ensemble versus those obtained from observations (Fig. 4.25). The only region without overlap is between 60 and 83 min. This fact means the ensemble did not produce any cells with lifetimes in this range. Also, the ensemble produced relatively few cells with lifetimes approaching 90 min. Thus, the 2.5 percentile limit for that region is zero. There is no disagreement, however, because the 2.5 percentile limit for the observed lifetime pdf is also zero. The PPP test indicates the parent population of lifetimes based on w is indistinguishable from the parent population of observed lifetime. Hence, for estimates of storm life-

times based on w , the PPP test indicates that the ensemble correctly forecast the pdf of observed storm lifetimes. However, this may represent an instance when the PPP test results incorrectly indicate that the null hypothesis should be accepted. If such an error has occurred, it may indicate that the PPP test lacks the power to discriminate between samples from different populations. For estimates of storm lifetime based on reflectivity, the observed and forecast parent populations are significantly different (Fig. 4.26).

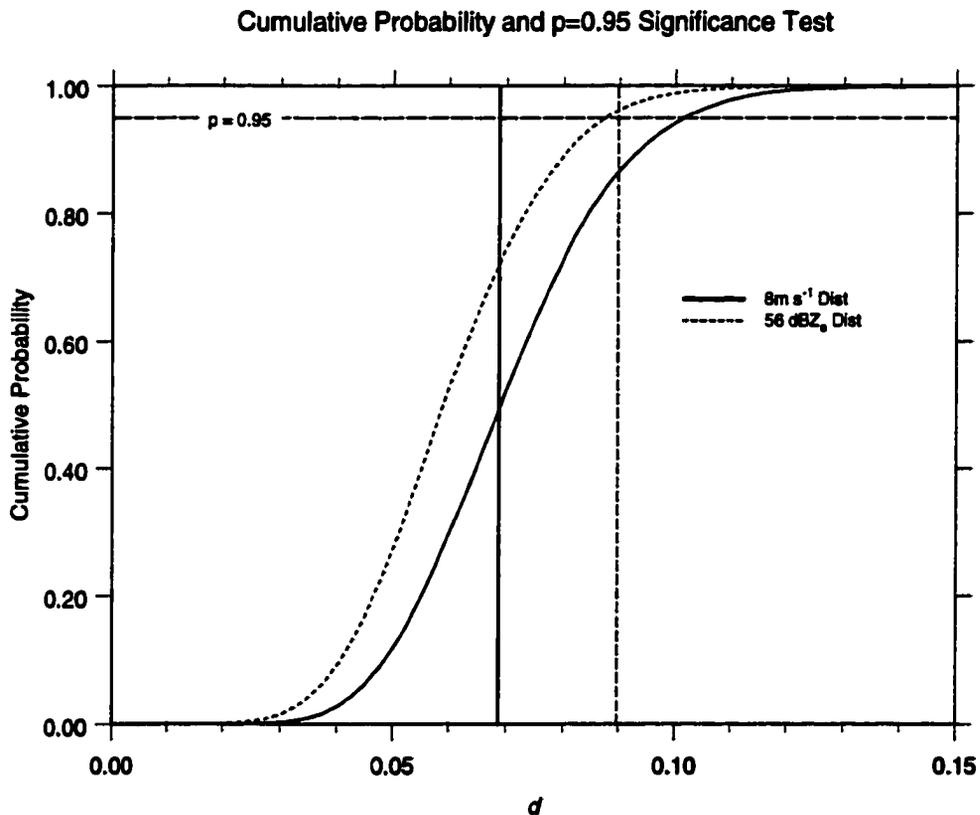


FIGURE 4.26. Same as Fig. 4.20, but for 14 July 1995.

Unfortunately, this ensemble also shows no ability to discriminate where convection may be expected, because deep convection results at every available sounding location (Fig. 4.27). Convection also occurs everywhere in fourteen of the eighteen cases. Because deep convection occurs at all sounding forecast locations, the TSS is zero for this

Cell and Report Locations for 14 July 1995

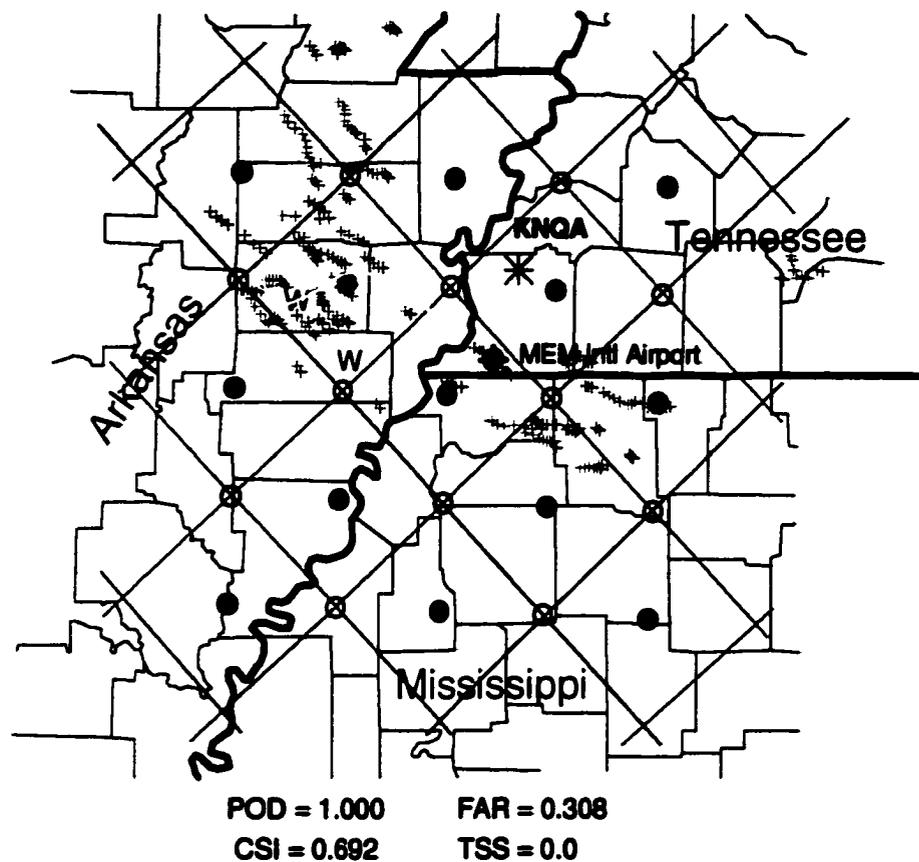


FIGURE 4.27. Same as Fig. 4.21, but for 14 July 1995. The red w symbol shows the location of a severe wind report between the hours of 1630 UTC 14 July 1995 and 0120 UTC 15 July 1995.

case. The ensemble model output contains long-lived (lifetime greater than 60 min) cells, and severe wind events are reported on this day.

4.5.3 6/7 June 1996

The 6/7 June 1996 case is one of two from the Dallas/Ft. Worth area. This case is distinctive because the forecast times used previously are not used on this day. Rather, forecast soundings for 0000 UTC, 0300 UTC, and 0600 UTC are used in the ensemble. When the standard forecast times are used, only two short-lived cells occur on the north-

ern-most grid points. Yet, a significant severe weather outbreak occurred on this day. The report times indicate severe weather occurs after 0000 UTC, as a short wave and associated dryline and cold front sweeps through the area. Hence, the times for the ensemble initial conditions center upon the times of reported events. Even so, a representative forecast pdf of storm lifetime is not possible because an insufficient number of cells are created by the ensemble. Yet, the PCA modes that result from the ensemble reveal a long-lived, possibly severe event, could have been foreseen.

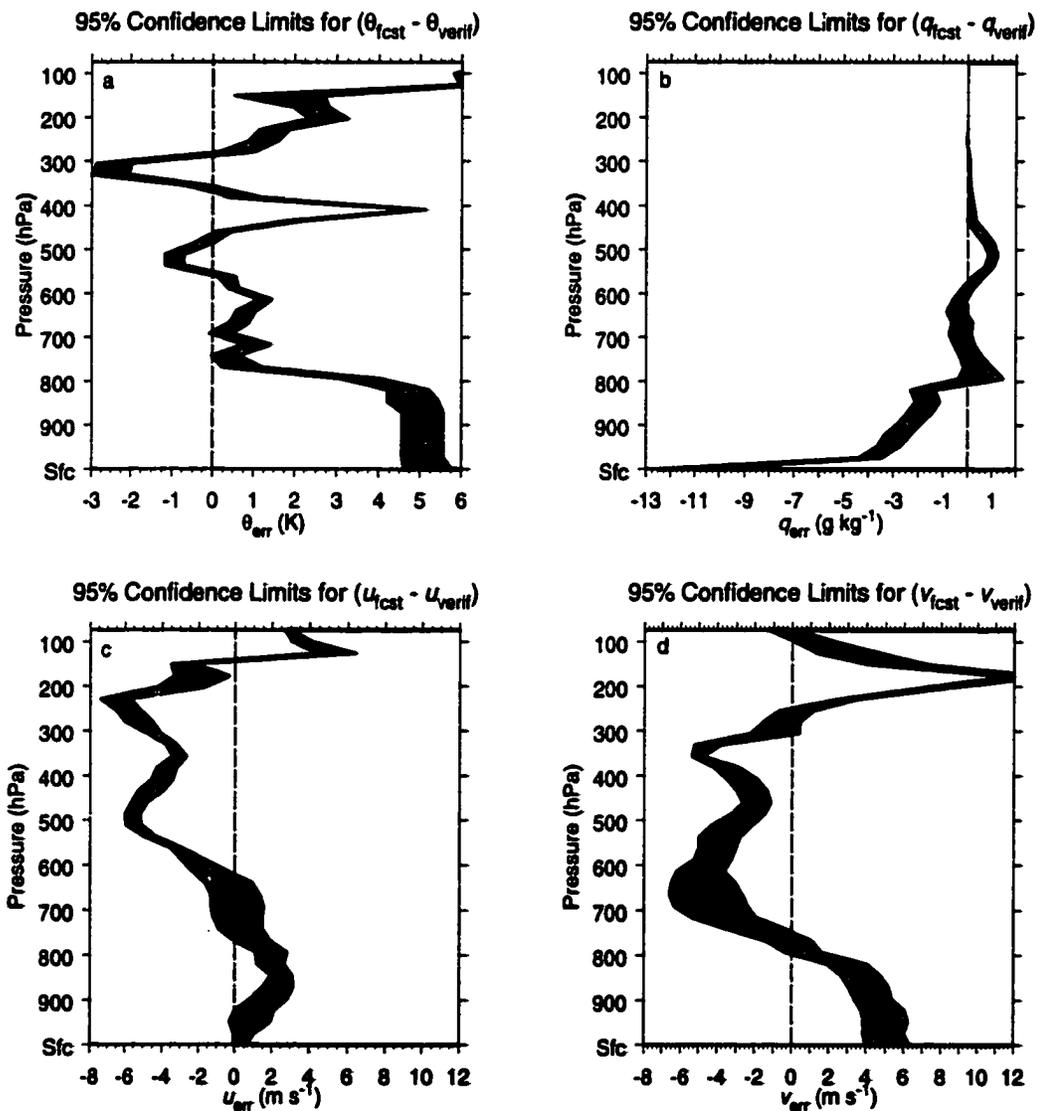


FIGURE 4.28. Same as Fig. 4.1, but for 6/7 June 1996.

On this day, the errors in the boundary layer Mesoeta forecasts have a large, warm, and dry bias, which suggests a possible phase error in the time of frontal passage. From the surface to 800 hPa, the bias in potential temperature is +5 K and the error in mixing ratio is -11 g kg^{-1} at the surface. The error in mixing ratio decreases to -4 g kg^{-1} just above the surface (Fig. 4.28). Above 800 hPa, the errors in mixing ratio are insignificant.

In contrast, the errors wind forecasts are less extreme. Errors in the u component are around $1\text{-}2 \text{ m s}^{-1}$ from the surface to 600 hPa, where the westerly component becomes too weak. Between 500 and 200 hPa the error in the u component is between -5 and -7 m s^{-1} (Fig. 4.28c). The error in the v component is roughly $+5 \text{ m s}^{-1}$ in the boundary layer. This error also suggests a timing error in the frontal passage. The errors in the v component are -5 m s^{-1} from 850 through 300 hPa. When combined, these kinematic and thermodynamic errors suggest that the forecast position of the short wave and associated cold front lags behind the actual positions.

A wide range of temperature perturbations characterize the initial conditions of the ensemble, which suggests that a significant temperature gradient exists over the region (Fig. 4.29a). The variability in mixing ratio is the same as for other days except for the layer between 800 hPa to 600 hPa (Fig. 4.29b). An elevated moisture gradient across the region is consistent with the dryline activity occurring during this period. The broad distribution of perturbations in the u and v components are also consistent with the frontal and dryline activity throughout the area during this period. The perturbations in the u component are particularly large at mid levels, which is also consistent with a short wave passing over the region during this time (Fig. 4.29c). The perturbations in the v component below

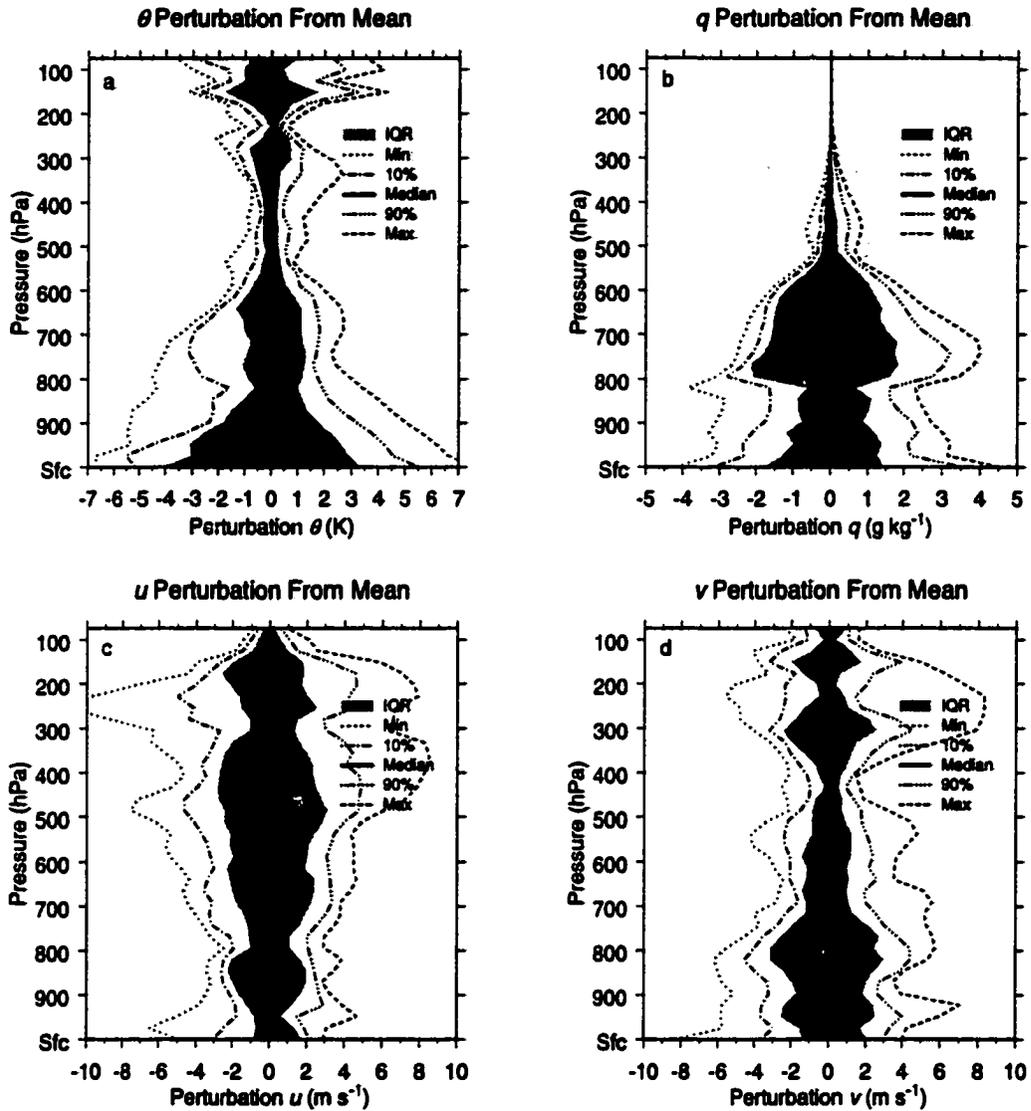


FIGURE 4.29. Same as Fig. 4.14, but for 6/7 June 1996.

800 hPa are more broad than for any other day examined, which is also consistent with the frontal activity recorded during this time (Fig. 4.29d).

The PCA modes recovered from the observations and the ensemble forecast reveal similar storm lifetime estimates. Two modes result from the observed data, and explain 82.5% of the Euclidean similarity (Fig. 4.30a). Two modes also result from the forecast data, and explain 79.8% of the Euclidean similarity (Fig. 4.30b). The short-lived storm

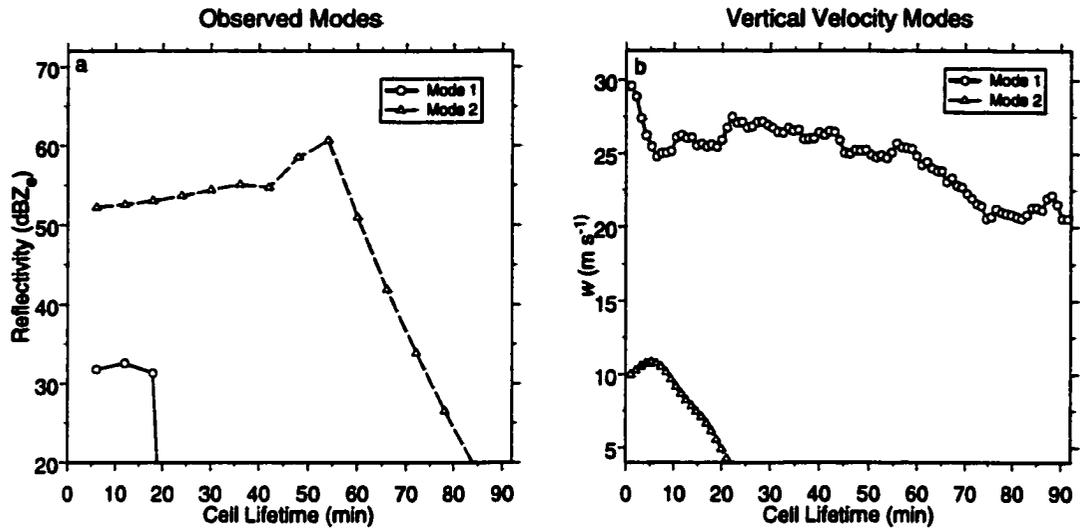


FIGURE 4.30. Same as Fig. 4.18, but for 6/7 June 1996. a) shows observed modes, which explain 82.5% of the observed Euclidean similarity, while in b) shows forecast modes, which explain 79.8% of the forecast Euclidean similarity.

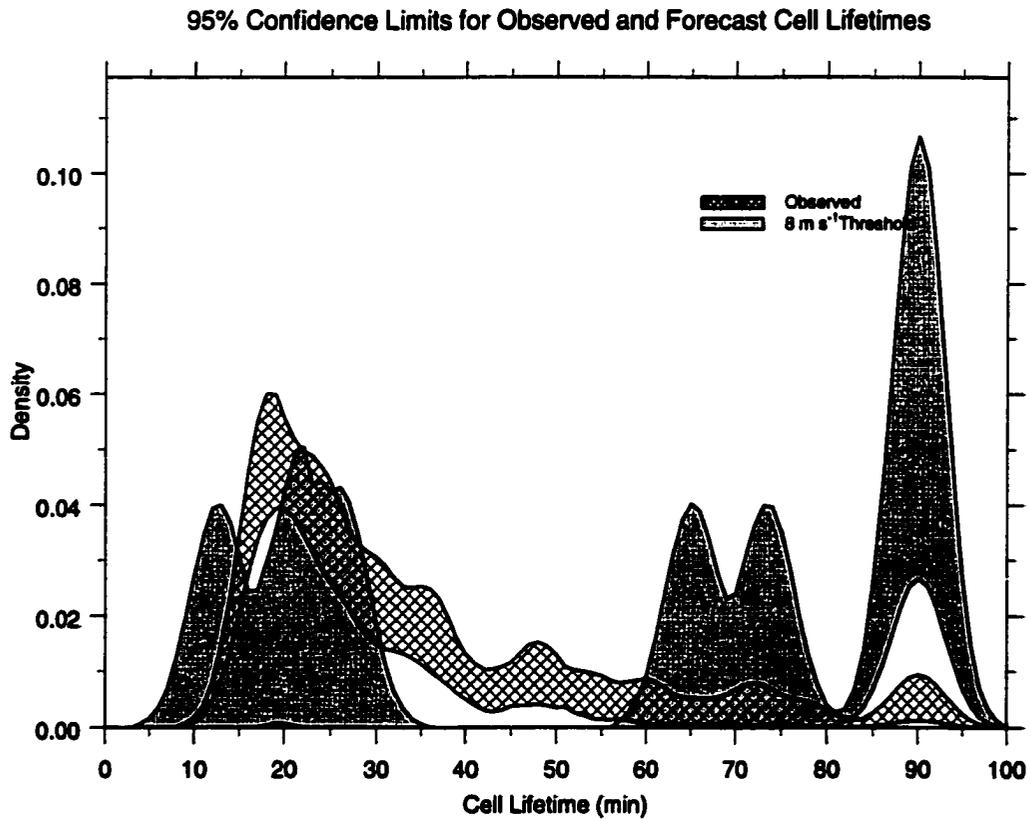


FIGURE 4.31. Same as Fig. 4.18, but for 6/7 June 1996.

lifetime mode differs by only 1 min between the observations and the forecast. Long-lived modes also exist in the observations and the forecast, though the long-lived mode resulting from the observed data shows a steady decay after 50 min. This decay occurs because the PCA constructs modes from a linear combination of the reflectivity time series. The observations contain only one cell with a lifetime greater than 90 min and several other cells with lifetimes between 50 and 80 min. When these observed time series of reflectivity are used to create the long-lived mode, a decreasing reflectivity value beyond 50 min results.

A comparison between the pdf's that describe forecast and observed lifetimes is not meaningful because there are not enough cells produced by the ensemble for reliable pdf estimates (Fig. 4.31). One indication that the sample size is too small is that the lower

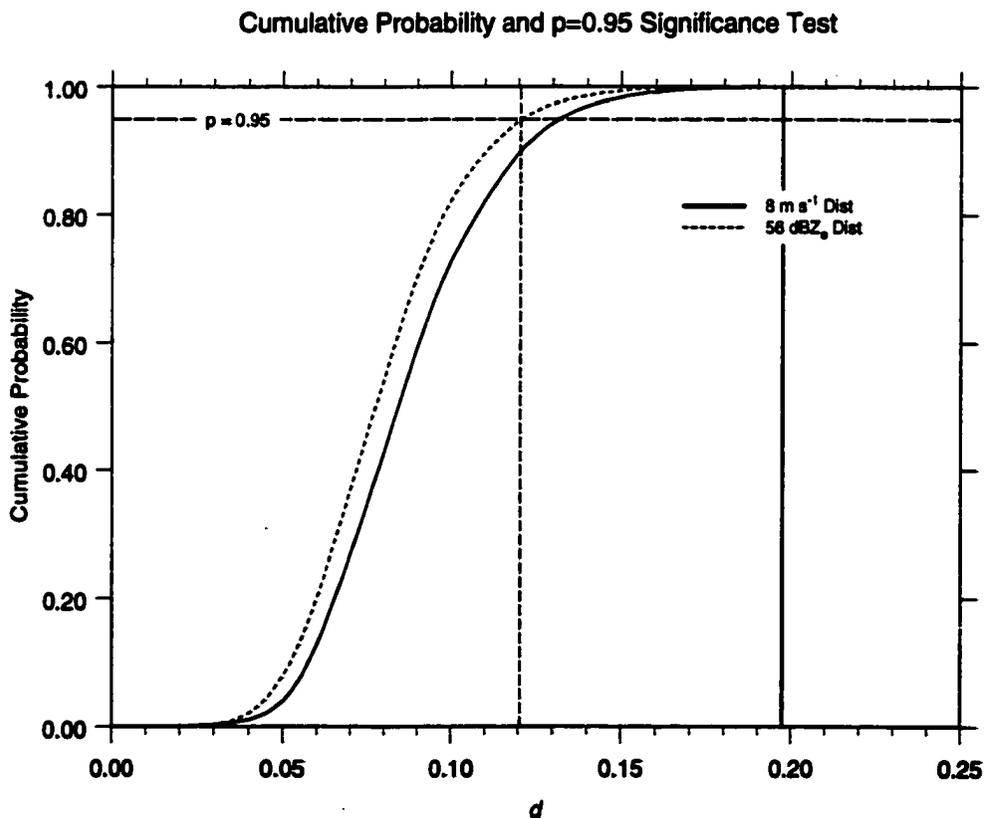


FIGURE 4.32. Same as Fig. 4.20, but for 6/7 June 1996.

bound for the forecast pdf is everywhere zero except around 90 min. Only ten cells are produced within the ensemble and, of these ten, five last 92 min. The other five cells have different lifetimes, which appear as bumps in the upper bound of the pdf estimate. Consequently, only six different lifetime estimates can be resampled. That the observed and forecast storm lifetime pdf's overlap is without question, but the nature of the parent population from which the sample of forecast lifetimes is drawn cannot be characterized with certainty.

Cell and Severe Report Locations for 6/7 June 1996

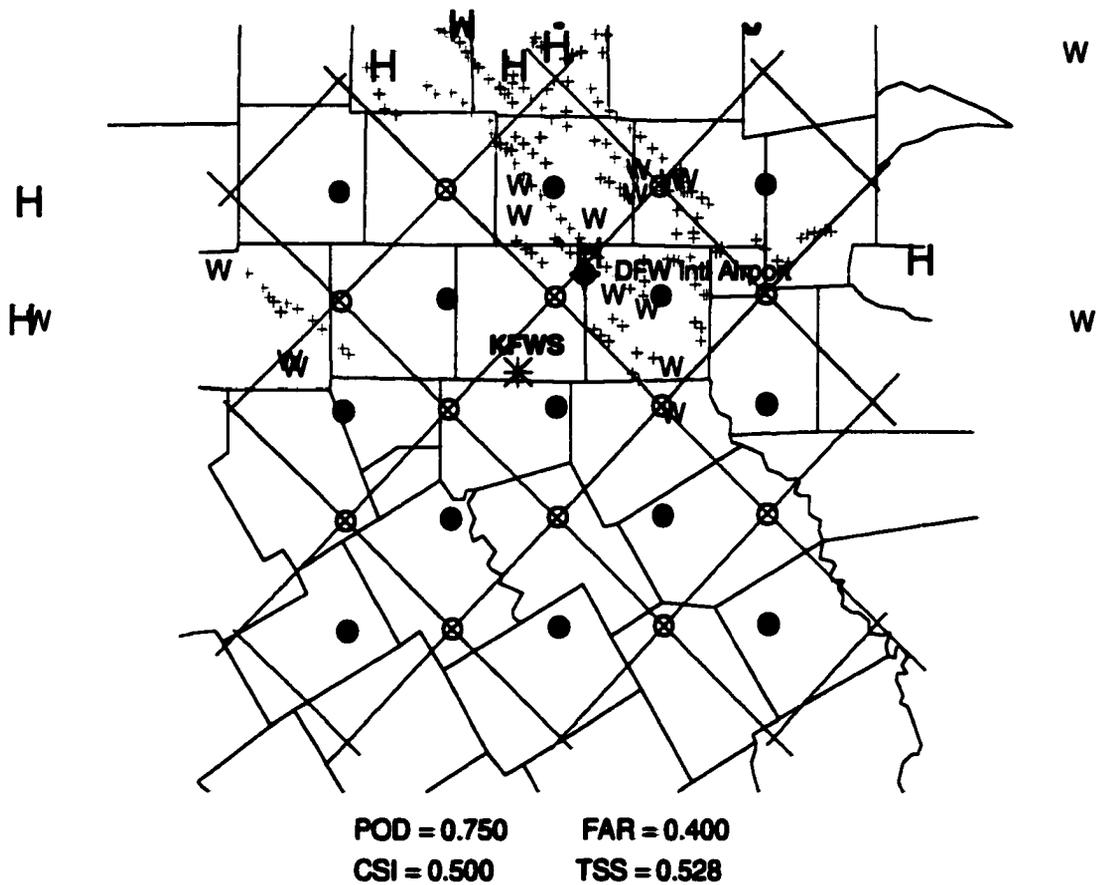


FIGURE 4.33. Same as Fig. 4.21, but for 6/7 June 1996. Red H's show location of hail reports that meet or exceed severe criteria.

Because the forecast lifetime pdf cannot be well estimated, it is no surprise that the PPP test shows significant differences between the observed and forecast parent distributions of storm lifetime (Fig. 4.32). Choosing a wider window in the kernel density function might result in a better comparison, but optimizing the kernel window width based on PPP test results assumes that the pdf of observed lifetimes is representative of the pdf of forecast lifetimes. This assumption is neither warranted nor defensible.

Despite these pdf results, this ensemble forecast has the exemplary quality that it captures the spatial distribution of convection well (Fig. 4.33). The ensemble generated long-lived cells where severe weather is reported, which means that, in this case, the ensemble is useful for severe weather guidance. The 2×2 statistics for this case are excellent, with POD = 0.750, FAR = 0.400, CSI = 0.500 and TSS = 0.528. Forecasters could use output from this ensemble to maintain an especially careful watch on the region from the Dallas/Ft. Worth area northward. Air traffic control could use output from this ensemble to prepare to route any incoming flights between 2230 UTC and 0730 UTC through the southern arrival gates, to avoid last-minute rerouting of aircraft.

Chapter 5: Conclusions

An ensemble cloud modeling system is developed to create forecasts of pdf's that describe thunderstorm lifetimes over a limited spatial and temporal domain. The forecast pdf's are valid for a 9 hr period across regions 160 km × 160 km in size and centered on either the Memphis or Dallas/Ft. Worth International Airports. These ensemble forecasts span 18 days during the summers of 1995 and 1996. Each ensemble consists of 39 cloud-scale model runs initialized with output from the operational Mesoeta forecast model. This technique maximizes the information available from the Mesoeta model. The ensemble approach presented in this work also facilitates objective, probabilistic forecasts of the range and distribution of storm lifetime. This approach can be extended to almost any observable thunderstorm characteristic. Ensemble models are particularly suited to forecast phenomena that can only be described using a pdf, such as thunderstorm lifetime over a given spatial domain and time period.

Consistent, nonparametric statistical techniques, based on the fundamental characteristics of pdf's, are developed and applied so that pdf's of storm lifetime can be extracted from both the cloud model ensemble and radar observations. These techniques allow direct comparisons between forecast and observed pdf's of storm lifetime. Only one parameter (reflectivity) is available for extracting lifetimes of observed storms, but two model parameters, vertical velocity and reflectivity, are available to define storm lifetime. Of these two model parameters, vertical velocity yields the best overall agreement between forecast and observed storm lifetimes.

Initial conditions for the ensemble model consist of soundings extracted from the operational Mesoeta forecast model, through a process called spatial and temporal atmo-

spheric sampling (STAS). The STAS process extracts soundings from the Mesoeta model over a nine hour temporal window and over a 160 km × 160 km spatial domain. Thus, STAS implicitly uses the spatial and temporal variations provided by the mesoscale forecast to generate the ensemble of initial conditions. STAS is a reasonable technique for generating a set of initial conditions because it uses the forecast variability for each day, which is itself variable, to define variability within the ensemble initial conditions. STAS also fits well into a concept that underlies ensemble modeling: namely, that the initial conditions (soundings, in this case) are all equally probable.

About 10 hr of clock time are required to complete the ensemble model run on a ten-processor SGI Power Challenge. Hence, even with outdated computational resources, this cloud-scale ensemble output is available within an operationally-useful time frame. Because ensemble models, and this one in particular, are ideally suited to parallel processing, a parallel cluster of 40 PCs running at 500 MHz could execute the entire ensemble in about 2 hr. Consequently, operational applications are feasible with modest, off-the-shelf hardware.

To further assist operational forecast activities, a new PCA technique is developed that combines the time series of any particular modeled parameter into a display that can be readily interpreted. Within a particular ensemble model, some members exhibit similar behavior. This PCA technique groups and displays common behavior patterns, or modes. In the case of cell lifetimes, this common behavior appears as a peak in the forecast pdf. But, a pdf of cell lifetimes provides only one characteristic of a two-dimensional data space consisting of time and amplitude. The new PCA technique uses Euclidean distance to build the parent similarity matrix, a metric that has not been previously applied to PCA.

This new PCA technique clearly captures the essence of ensemble members which display similar behavior.

While the Mesoeta is shown to produce the correct amount of dispersion or variance in the forecast fields at nearly all levels, the Mesoeta clearly displays pathological biases during the 1995 and 1996 summer seasons. When 0000 UTC analyses are used to verify the prior forecasts valid at 0000 UTC, the Mesoeta displays a dry and warm bias in the lowest model levels. These biases affect some derived sounding parameters, such as CAPE and LI, which are known to be associated with significant thunderstorm characteristics. However, given the ensemble model results, no evidence exists that complete information about storm lifetime can be garnered from these derived parameters. Various parameter spaces constructed from combinations of these derived parameters do not provide a reliable way to identify days when long-lived storms are absent. Thus, derived parameters do not help discriminate between days that do and do not produce long-lived storms. However, derived parameters may be useful for describing other storm attributes.

In several of the cases, modeled thunderstorm lifetimes are shown to be extraordinarily sensitive to variations in the initial soundings. However, such chaotic sensitivity is not always evident because, in several cases, thunderstorm lifetimes are remarkably *insensitive* to variations in the initial soundings. Because derived sounding parameters, and the various parameter spaces constructed from them, do not provide reliable insight into storm lifetimes, it is unknown whether the lifetime of a modeled storm will be disproportionately sensitive to the initial soundings. Therefore, a way does not exist to forecast the range and distribution of thunderstorm lifetimes short of running an ensemble of cloud models.

Because the ensemble tends to generate long-lived storms on days that experienced severe weather, some evidence exists that simulated storms with lifetimes longer than 60 min are a useful indicator of severe weather. In a similar vein, if convection does not occur at all the locations from which initial soundings are drawn (“limited convection”), the ensemble results display some skill in identifying where thunderstorms occur. However, only a handful of cases exist with either severe weather or limited convection. While these results are encouraging, more cases must be analyzed to know if storm lifetime is a useful proxy for severe weather occurrence, and whether there is any skill in identifying where convection is likely to occur.

Uncertainties in the forecast and observed pdf’s of storm lifetime are estimated with bootstrap resampling. When these uncertainties are considered, agreement between forecasts and observations are quite good, because in 17 out of 18 cases, the forecast and observed pdf confidence intervals have significant overlap. When considering the range of thunderstorm lifetimes, 50% of the forecast pdf’s captured the correct range, while 50% underestimated the range by varying degrees. However, there is some evidence that the PPP test may lack sufficient power to reliably discriminate between samples drawn from significantly different populations. When all the forecast and observed storm lifetimes are combined, the overall probability of storms with lifetimes between 35 and 75 min is too low.

The forecast pdf’s can be used to provide beneficial strategic guidance for planning the available capacity at airport terminals. For example, during periods when the probability of long-lived storms exceeds a predetermined threshold, arrivals can be appropriately scaled back in advance. If the ensemble displays skill at forecasting where storms

are most likely, plans for arriving flights can be refined to use the arrival routes that are most likely to remain free of convection.

Another straightforward application of the ensemble data is to create conditional cell decay forecasts based on the current age of a particular observed cell (Fig. 5.1). The

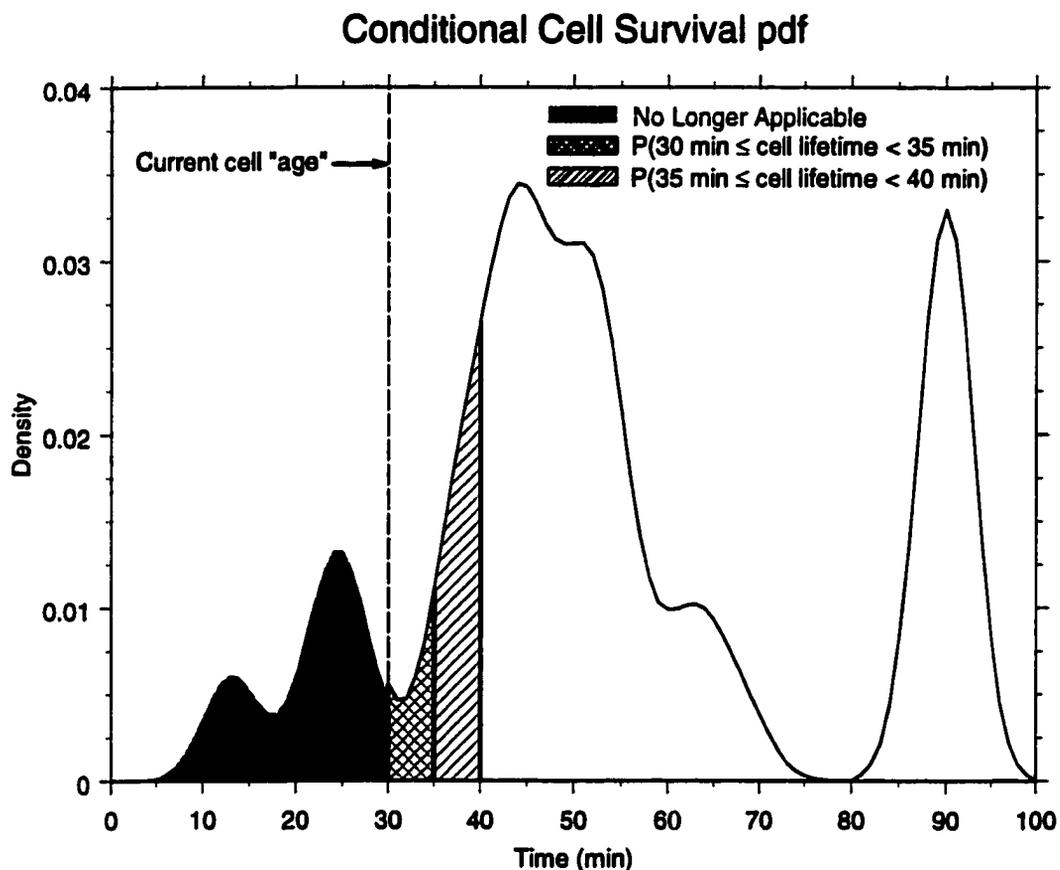


FIGURE 5.1. Conditional cell survival probability. The vertical dashed line shows the current age of the thunderstorm cell. The pdf to the right of this line has been rescaled such that the area beneath the pdf is unity. Grey area represents the part of the pdf that no longer applies because the cell has already existed for 30 min. The cross-hatched area yields the probability the cell has a lifetime between 30 and 35 min. The hatched area yields the probability that the cell has a lifetime between 35 and 40 min. The sum of the cross hatched and hatched areas yield the probability that the cell has a lifetime between 30 and 40 min.

conditional cell lifetime is given by $P(\text{cell lifetime} \mid \text{cell has survived for } t \text{ min})$, which is computed using the forecast pdf after it is adjusted for how long the cell has already survived. The forecast pdf is adjusted for the age of a cell by truncating that part of the forecast pdf from $t = 0$ to $t = \text{current cell age}$, and rescaling the remainder of the forecast pdf to

an area of unity. Thus, the probability that a cell will decay within a specified period of time beyond the current cell age can be calculated. Subtracting this cell decay probability

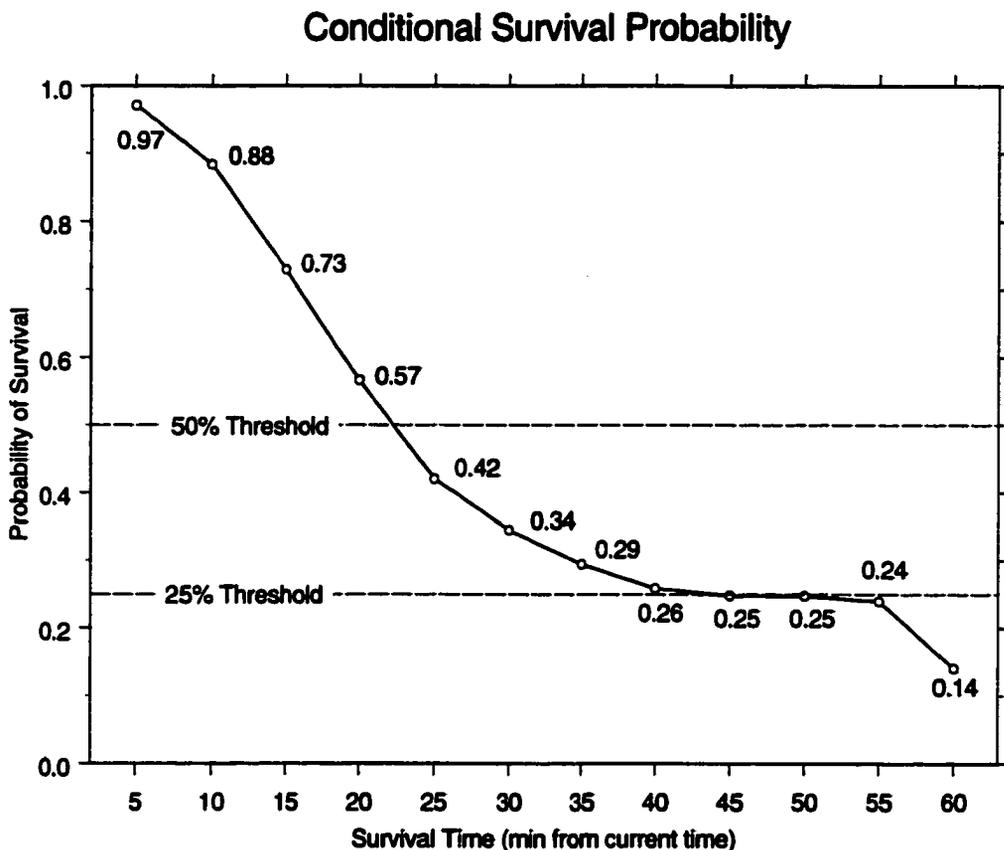


FIGURE 5.2. The conditional survival probability, computed from Fig. 5.1. The x-axis shows the cell survival time *beyond* the current cell age. The y-axis is the probability of survival beyond the current cell age. The horizontal dashed lines show probability thresholds of 50% and 25%.

from 1 yields the probability that the cell survives beyond a given age (Fig. 5.2). This information can be displayed easily as part of a storm track on a radar display. Colors can be used along the cell track to describe the likelihood that a cell survives to a particular part of the projected cell track (Fig. 5.3). Such a display furnishes useful information to Air Traffic controllers by providing guidance about how soon a storm is likely to affect (or cease to affect) an arrival gate. Similarly, it provides NWS forecasters with guidance about the likelihood that a particular cell will affect a given community.

Conditional Cell Survival Probabilities

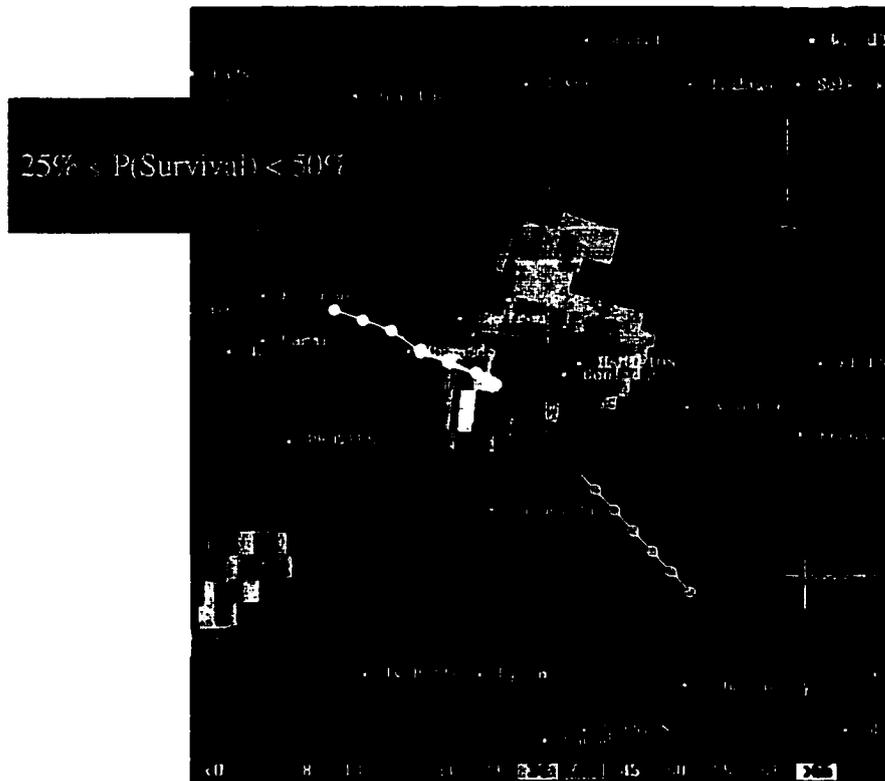


FIGURE 5.3. Conditional cell survival product. The previously observed cell positions are shown in white. Projected cell positions are shown in color. With at least 50% probability, the cell will survive to some point along the red part of the projected track, with probability between 25% and 50% that the cell will survive to some point along the yellow part of the track, and with probability less than 25%, the cell will survive to some point along the green part of the track.

Other characteristics derived from an ensemble can be encoded along the projected track, such as the probability of a mesocyclone, as defined by a positive correlation between vertical velocity and vertical vorticity. The projected cell tracks could be created with a 95% probability confidence interval using the pdf of cell motion for cells at various ages. Creating pdf forecasts has an elegant appeal because, once pdf's of a particular characteristic are constructed, creating probabilistic forecasts for that characteristic is easy.

While these results are encouraging, this work is preliminary and much remains to be done. The results raise as many questions as they answer. For example, the Mesoeta

model no longer exists, having been incorporated into the current Eta model. Various parameterizations within the Eta have been changed or modified. How these changes affect the cloud scale ensemble model is unknown. A broader question concerns which mesoscale model to use. One way to investigate this problem, and many other questions, is to simultaneously develop and maintain an archive of radar data from various sites and an archive of output from operational forecast models. The archived model output is then used to create variations of the cloud-scale ensemble, which are run and verified against the archived radar data.

Cloud scale models also are sensitive to the geometry and strength of the warm bubble used to initiate convection, but the nature of this sensitivity is not well understood. How convection is affected by the way it is initialized remains unknown. Warm bubbles clearly have weaknesses when used to initiate convection, but they remain the most economical initialization scheme available. How the initiation procedure affects the ensemble results remains unknown.

Cloud scale model results are known to be sensitive to the precipitation physics that are simulated (Jewitt et al. 1990; Johnson et al., 1993; Straka and Rasmussen, 1998). The cloud scale model applied to this ensemble uses the Kessler scheme to parameterize precipitation physics. Ample evidence exists (Wicker 1999, personal communication, Johnson et al., 1993) that the Kessler scheme can lead to unrealistic storm evolution by significantly accelerating certain cloud-scale processes, which can cause unrealistically short lifetimes. If, instead, a full ice microphysics package is used, more storms lasting between 35 and 75 min might occur, producing better overall forecasts of the pdf's of storm lifetime.

Days chosen for the ensemble runs were dependent upon the available verification data. This restriction also determined the nature of days examined with the ensemble to those with relatively weakly-sheared environments. Yet, most significant severe weather occurs in strongly sheared environments. How the ensemble performs in such cases is uncertain.

Finally, given the promising nature of the results so far, a field test is warranted. Obviously, all of the statistical tools are in place to perform manually such a test, and automating these tests, and other data processing tasks, is straightforward. However, to make a field test possible, a version of COMMAS that automatically adjusts the grid motion to keep storm cells within the center of the grid needs to be implemented.

Appendix 1: Kernel Density Estimation

Following the development by Silverman (1986), a histogram is constructed by placing n “boxes” of width $2h$, on the real line. Given the origin x_0 and the bin width h , the bins of the histogram are defined as the intervals $[x_0 + mh, x_0 + (m + 1)h)$ for positive and negative integers m . The interval is closed on the left and open on the right for definiteness. The histogram estimate of the pdf for a parameter x is then given by

$$\hat{f}(x) = \frac{1}{2hn} \times [\text{no. of } X_1, \dots, X_n \text{ falling in } (x - h, x + h)],$$

where X_1, \dots, X_n are the sample of n real observations whose density can be estimated. More transparently, define a weight function w by

$$w(x) = \begin{cases} \frac{1}{2} & \text{if } |x| < 1 \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A1.1})$$

Now, it is easy to see that the histogram estimating function can be written as

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} w\left(\frac{x - X_i}{h}\right). \quad (\text{A1.2})$$

This estimator can be generalized for any weight, or kernel, function K satisfying the condition that $\int_{-\infty}^{\infty} K(x) = 1$. K is almost always a symmetric pdf and commonly the Gaussian pdf. By analogy with the definition of the histogram estimator, the kernel density estimator is

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right), \quad (\text{A1.3})$$

where h is the window width (also called the smoothing parameter, or bandwidth). For this work, the kernel estimator is simply

$$\hat{f}(x) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{i=1}^n e^{-\left\{\frac{1}{2}\left[\frac{(X_i-x)}{\sigma}\right]^2\right\}}, \quad (\text{A1.4})$$

where $\hat{f}(x)$ = density estimate at x ,
 x = location of 1-D density estimate,
 X_i = observation location,
 n = number of observations,
 σ = bandwidth in units of standard deviations.

Note that $\hat{f}(x)$ is the *sum* of the kernel estimator, K , at the location x , over all the points, X_i , at which it is applied. Hence, in eq. A1.4, K , the kernel *function*, is

$\frac{1}{\sigma\sqrt{2\pi}} e^{-\left\{\frac{1}{2}\left[\frac{(x-\mu)}{\sigma}\right]^2\right\}}$, where x is replaced by X_i , and μ is replaced by x . Also, note that

$\int_{-\infty}^{\infty} \hat{f}(x) dx = 1$ and thus, in finite difference form, $\sum_{i=1}^n \hat{f}(x_i) = 1$, when summed over all

of the locations, x , at which kernel density estimates are made. For this work, the bandwidth is chosen to be 3 min (since the grid upon which the pdf is analyzed is lifetime in min). This bandwidth choice smooths most of the insignificant bumps while retaining the real peaks.

Appendix 2: Euclidean Principal Component Analysis¹

A2.1 Introduction

Eigentechniques have been widely used in meteorology since the 1950's. Three common variants are: Common Factor Analysis (CFA, Thurstone 1947), Empirical Orthogonal Functions (EOF, Lorenz 1956) and Principal Component Analysis (PCA, Hotelling 1933). EOF uses unit-length eigenvectors, whereas in PCA and CFA each eigenvector is weighted by the square root of its corresponding eigenvalue. Consequently, the weights represent the correlations or covariances between each variable and each principal component, depending upon which similarity matrix is employed (Jolliffe 1995). Any of the three techniques may be used as either a statistical modeling tool or as a diagnostic tool. Each eigentechnique is derived directly from a parent similarity matrix (also called a dispersion matrix in some texts) which typically consists of either a correlation or covariance matrix or, rarely, a matrix of cross-products. These similarity matrices are diagonalized such that eigenvalues and associated eigenvectors are identified, and eventually used in the physical interpretation phase of the analysis. Because the parent similarity matrix embodies the type of association desired, and defines the immediate starting point for the eigenanalysis by virtue of being diagonalized, the similarity measure used to build the parent similarity matrix is an important aspect of any eigentechnique. However, this choice is not always given the consideration it is due. Historically, the various similarity matrices are discussed and compared in meteorological literature. Examples include Craddock (1965), Kutzbach (1967, 1969) and Craddock and Flood (1969) who all favor the covariance matrix on grounds that it more accurately portrays the true variance structure. In con-

¹.Excerpted from a paper authored by Kimberly L. Elmore and Dr. Michael B. Richman, and submitted to *Monthly Weather Review*, 16 February 2000.

trast, Gilman (1957), Sellers (1957) and Glahn (1965) are proponents of the correlation matrix, claiming it puts all variables on equal footing, whereas Resio and Hayden (1975) and Molteni *et al.* (1983) find cross-products to have utility.

Table A2.1.

Similarity Measure	Vector Form	Matrix Form
Cross-Products	$p = \mathbf{x}^T \mathbf{y}$	$\mathbf{P} = \mathbf{X}^T \mathbf{X}$
Variance/Covariance	$s = \frac{[(\mathbf{x} - \bar{\mathbf{x}})^T (\mathbf{y} - \bar{\mathbf{y}})]}{(n - 1)}$	$\mathbf{S} = \frac{[(\mathbf{X} - \mathbf{M})^T (\mathbf{X} - \mathbf{M})]}{(n - 1)}$
Correlation	$r = \frac{\left[\frac{(\mathbf{x} - \bar{\mathbf{x}})}{s_x} \right]^T \left[\frac{(\mathbf{y} - \bar{\mathbf{y}})}{s_y} \right]}{(n - 1)}$	$\mathbf{R} = \frac{[(\mathbf{X} - \mathbf{M})(\sqrt{\mathbf{V}})^{-1}]^T [(\mathbf{X} - \mathbf{M})(\sqrt{\mathbf{V}})^{-1}]}{(n - 1)}$

TABLE A2.1. Vector and matrix forms for cross-products, variance/covariance and correlation. For the vector form, n is the length of the vector and $\bar{\mathbf{x}}$ ($\bar{\mathbf{y}}$) is the vector whose values consist of the mean of \mathbf{x} (\mathbf{y}). In the vector form of correlation, s_x (s_y) is the square-root of variance for the \mathbf{x} (\mathbf{y}) vector. For the matrix form, the \mathbf{X} is $n \times p$ (n rows and p columns), and \mathbf{M} (also $n \times p$) is the matrix whose i^{th} column is the mean of the i^{th} column of \mathbf{X} . \mathbf{V} is a $p \times p$ diagonal matrix whose p non-zero elements consist of the variance of each column of \mathbf{X} .

Depending upon the parent similarity matrix, eigentechnique results can have physically different meanings. Table A2.1 defines cross-products, covariance and correlation for both single column data vectors of length n and $n \times p$ data matrices. The correlation matrix groups variables together regardless of the amplitude of their variation or mean. By its nature, correlation provides no measure of how much parameters vary with each other, only that they do. For example, correlation does not address whether the *magnitude* of variation in one variable coincides with the *magnitude* of variation in another. Correlation addresses only relative variability relationship because the standardization puts all variances equal to unity. Often, such a relation is the most important aspect of an analysis. Because input data have been normalized to a zero mean and unit variance, cor-

relation is a dimensionless similarity metric; hence, it is appropriate for comparing variables with different units or scales.

The covariance matrix yields insight into how much variables change with respect to each other. Data that are transformed into a covariance matrix have been translated to zero mean. Strictly speaking, dimensions or scales are preserved with covariance, which means that applications that mix different units will emphasize those with units having the most variation. Moreover, analyses using a single variable with a large range over the domain (for example, sea level pressure at sites from the tropics to the poles) will emphasize these variables with a large variance. This can be a positive or a negative attribute, depending upon the specific nature of the analysis.

A cross-products (covariance without removing the means) similarity matrix also preserves units, and results are sensitive to the magnitudes of the means as well as co-association. A data vector that possesses a mean with large magnitude will dominate the eigenanalysis outcome. When the magnitudes of the means are similar, the eigenanalysis outcome is primarily determined by the covariance. Additionally, analyses utilizing the cross-products similarity matrix tend to have a first principal component that resembles the mean.

For these three similarity matrices, co-association plays an important role because the eigenanalysis, in some manner, identifies variables that change together. But might physical insight arise from another characteristic besides co-association?

Clearly, the investigator must carefully consider what is desired from an eigenanalysis and choose an appropriate parent similarity matrix that preserves the desired information. It is possible that co-association of the aforementioned types is not an important

issue. For example, objective identification and representation of variables that, when plotted together, are co-associated in that they appear visually close to one another may be important. A desirable result is to distill the data into a few, easily interpreted modes of behavior. To accomplish this result, a new similarity matrix, named Euclidean similarity (ES), is utilized. Euclidean similarity is inspired by the large body of literature in cluster analysis, which clearly demonstrates the effectiveness of Euclidean distance (ED), on which ES is based.

The next section introduces and defines ES. Section 3 demonstrates how modes of co-association are extracted with a PCA based on ES. Section 4 demonstrates results of PCA using ES in both S-mode and T-mode analyses, and these results summarized in Section 5.

A2.2 Euclidean similarity

As a motivational example, ES is used to extract vertical velocity (w) time series that exhibit similar behavior, and combine them into modes. In this example, the ensemble of w time series come from several cloud model runs, each started with slightly different initial conditions. Other time series can also be treated this way, such as precipitation, or the u and v wind components. Of course, extracting modes need not be limited to time series. For example, two-dimensional fields, such as spatial pressure or height patterns can be treated in an identical fashion.

In this case, the parameter to be examined is arranged in a data matrix that provides an S-mode analysis (Richman 1986), in which each column (variable) represents a vector. The vector may consist of an individual time series or, alternatively, spatial pattern.

Although the number modes is important, these ensemble modes must also be visualized (Anderson 1996). Once modes are identified, they must be displayed unambiguously, and the ED (or the Minkowski L_2 norm) between each data vector provides a natural way to proceed. Cluster analysis uses ED extensively to identify or group data or entities that are dissimilar (Anderberg 1973; Gong and Richman 1994). Euclidean distance is an attractive measure for identifying modes because whether the parameter comprising the individual vectors *vary* together is not as important as how closely they overlay each other.

Consider a $p \times n$ data matrix \mathbf{Z} composed of n columns (cases), each p elements long. The ED between the vectors z_i and z_j is

$$d_{ij} = [(z_i - z_j)^T (z_i - z_j)] \quad (\text{A2.1})$$

where d_{ij} is the distance between the vectors z_i and z_j . When computed for all i, j , and arranged in a matrix in an order identical to that obtained from matrix cross-products, this process results in a symmetric dissimilarity matrix of Euclidean distances, \mathbf{D} . For n cases, \mathbf{D} will be $n \times n$. This matrix has zeroes along the main or principal diagonal (because the distance between a vector and itself is zero) and has units identical to the input data. The difference between dissimilarity and similarity is orientation, so for a Euclidean *dissimilarity* metric, large values indicate a large Euclidean distance, whereas for a Euclidean *similarity* (ES) metric, large values indicate a small Euclidean distance. To create a similarity matrix, some simple operations must be applied to \mathbf{D} .

Define d_{\max} to be the maximum of all elements in \mathbf{D} . Normalize \mathbf{D} by d_{\max} such that no element in the new $\hat{\mathbf{D}}$ is greater than 1 by setting

$$\hat{\mathbf{D}} = \mathbf{D}[\text{diag}(1/d_{\max})], \quad (\text{A2.2})$$

where $[\text{diag}(1/d_{\max})]$ is the $n \times n$ diagonal matrix that consists of $1/d_{\max}$ along the minor diagonal. Let \mathbf{Q} be the $n \times n$ matrix for which all elements are 1 and use it to define a new similarity matrix, \mathbf{E} , such that

$$\mathbf{E} = \mathbf{Q} - \mathbf{D}. \quad (\text{A2.3})$$

These operations transform \mathbf{D} into a matrix of similarities, where the main or principal diagonal consists of 1's. Hence, \mathbf{D} mimics a correlation matrix. Euclidean similarity is dimensionless and, because ES is constructed to mimic correlation, an eigenanalysis preserves relative distances.

The PCA performed on ES creates a column-wise orthogonal loading matrix, though the loading vectors may be correlated. The Euclidean distance between the loading vectors is the primary structure imposed by the eigenanalysis. Let the columns of the loadings matrix \mathbf{A} be defined as $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$. The Euclidean distance between \mathbf{a}_1 and \mathbf{a}_2 is the largest of any pair of PC loadings. The Euclidean distance between \mathbf{a}_1 and \mathbf{a}_3 is the next largest, and so on to the distance between \mathbf{a}_1 and \mathbf{a}_n . The next largest distance between any two loadings is that between \mathbf{a}_2 and \mathbf{a}_3 , then that between \mathbf{a}_3 and \mathbf{a}_4 , and so on. The smallest distance between loadings is that between \mathbf{a}_{n-1} and \mathbf{a}_n .

An example that uses only two data vectors is presented as a simple introduction. Let the Euclidean distance between these two vectors be defined as d . Thus, the matrix of Euclidean distances is:

$$\mathbf{D} = \begin{bmatrix} 0 & d \\ d & 0 \end{bmatrix}.$$

Because $d_{\max} = d$, the matrix of normalized Euclidean distance, $\hat{\mathbf{D}}$ is:

$$\hat{\mathbf{D}} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Finally, this new matrix is converted to ES by element-wise subtracting it from \mathbf{Q} :

$$\begin{aligned} \mathbf{E} &= \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

The similarity matrix, \mathbf{E} , is the 2×2 identity matrix in this example. A PCA on this matrix is trivial because, by inspection, the eigenvalues are $\lambda_1 = \lambda_2 = 1$ and eigenvec-

tors are $\mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. This result shows that any two data vectors are as distinct as

possible (in a Euclidean distance sense) because they cannot be remapped to new orthogonal components that will result in a coordinate transformation that further separates them. This is quite different from, say, a variance/covariance-based PCA, where the data may be recast onto orthogonal coordinates such that the first loading (scaled eigenvector) represents the coordinate axis that explains the most variance, and the second loading defines the axis that explains the rest of the variance.

A2.3 Euclidean similarity PCA using least squares scores

The breadth of solutions and interpretations that can result from a single data set, based on different similarity measures, is briefly reviewed prior to the demonstration of pattern retrieval using scores derived from ES. This is intended to help emphasize the point that the investigator needs to deliberately address the question, *a priori*, of what is truly desired from the analysis. “Black box” approaches, which mean the blind application

of a given similarity matrix, may not lead to useful results. Moreover, such an approach may deprive the investigator of new insights.

Typically, the fundamental PCA equation is cast as

$$\mathbf{Z} = \mathbf{F}\mathbf{A}^T \quad (\text{A2.4})$$

where \mathbf{Z} is the ($p \times n$) data matrix, and, in the nomenclature of PCA, \mathbf{A} is the ($n \times n$) matrix of *loadings*, and \mathbf{F} is the ($p \times n$) matrix of *scores*. In PCA, the eigenvectors (\mathbf{V}) are scaled by the square root of their respective eigenvalues ($\Lambda^{1/2}$), which yields the matrix of loadings (\mathbf{A}). Despite the fundamental formula, the traditional manner in which \mathbf{V} is derived is through diagonalization of a similarity matrix \mathbf{E} , as

$$\mathbf{E} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T. \quad (\text{A2.5})$$

Any PC loading vector, \mathbf{a}_j , can have all of its elements multiplied by -1, because the signs of the loadings are arbitrary. These loadings may be considered weights that identify linear combinations of the scores that, as defined in the parent similarity matrix, behave similarly. Geometrically, the PCs define a new, orthogonal coordinate system into which the loadings are projected. If all of the eigenvectors are retained, the original data (and thus its total similarity) can be recovered exactly, although the original data may be standardized for certain analyses. If all are not retained, the original data (and its total similarity) can be recovered only approximately, yet the PC model yields the most efficient manner in which data can be expressed in a smaller number of dimensions.

The maximum variance and orthogonality constraints that act on a geometric domain can lead to a number of hindrances to physical interpretation of unrotated PC loadings (Richman 1986). These hindrances include merging of unique sources of variation on the leading PC (Karl and Koscielny 1982), high sampling errors if adjacent eigenvalues are similar (North et al. 1982), a set of predictable geometric patterns which are partly a function of domain shape (Buell 1975), and poor fit to the parent similarity matrix

(Richman and Gong 1999). Hence, the PC loading patterns may not optimally portray physical relationships embedded in the data. In fact, these patterns can be misleading if they are literally interpreted as the key modes of variation in the parent similarity matrix. Because the ability to portray physical relationships accurately is crucial to identifying modes of behavior in a PCA, coordinate transformations, called rotations, are often applied to the PCs. Once rotation is invoked, some of the characteristics of the loading and score matrices (in particular, orthogonality) no longer apply (Jolliffe 1995). Appropriate rotation does, however, enhance interpretability though rotation does not guarantee interpretability (Richman 1986; Cheng et al. 1995). How many (r , where $r < n$) PCs to retain for rotation is a somewhat subjective decision, and can be determined in numerous ways, i.e., by variance criteria, eigenvalue separation criteria, etc. In principle, the majority of the signal is captured in the first r PCs, and noise, which is contained in the $r+1, \dots, n$ PCs, is discarded. For the example here, enough PC's to explain at least 80% of the total ES are retained. For the examples that use the w time series from the cloud model ensemble, three PCs are required to meet this constraint which, collectively, explain 86.6% of the total ES.

To aid in obtaining physically interpretable results, a Varimax (Kaiser 1958) rotation is applied. The Varimax rotation criterion is the most widely accepted and employed orthogonal rotation, because it tends to produce, but does not guarantee, simplification of the unrotated loadings into easier to interpret results (Cooley and Lohnes 1971). Varimax tends to simplify by rigidly rotating the PC axes such that the variable projection (loadings) on each PC tend to be high or low (Cooley and Lohnes 1971), which is consistent with the physics needed for the definition of a small set of convective modes. These modes are used to describe the dominant behavior of convective storms out of an ensem-

ble of many cloud model runs. Varimax is an orthogonal rotation, which means that the original loading matrix, \mathbf{A} , is transformed via an orthogonal transformation matrix to the rotated loading matrix, \mathbf{B} . Mathematically,

$$\mathbf{B} = \mathbf{A}\mathbf{T}, \quad (\text{A2.6})$$

where \mathbf{T} is an orthogonal matrix such that

$$\mathbf{T}^T\mathbf{T} = \mathbf{I}, \quad (\text{A2.7})$$

and \mathbf{I} is the identity matrix. The Varimax method finds \mathbf{T} by iteratively maximizing the collective variance of the squared loadings for all the retained PCs.

Unlike typical PCA, where most of the physical interpretation is applied to the loadings, for this application physical interpretability arises from the scores (\mathbf{F}) recovered from the rotated loadings and original data. Recovery of modes from a small number of dimensions, $r < n$, uses a crucial PCA characteristic: PCA provides variables in the order necessary to allow linear least squares reconstruction of the data using the fewest possible terms. Modes are extracted through a least-squares formulation, because it is straightforward and optimal in an L_2 sense. Least squares scores are defined by

$$\mathbf{F} = \mathbf{Z}\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}, \quad (\text{A2.8})$$

where \mathbf{Z} is the original $p \times n$ matrix of w values. This yields r modes because, if \mathbf{Z} is $p \times n$ and \mathbf{B} is $n \times r$, where r is the number of retained PCs, then $(\mathbf{B}^T\mathbf{B})^{-1}$ is $r \times r$, \mathbf{F} is $p \times r$, which leaves r column vectors in the result, where each column vector represents a mode. The matrix represented by $\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}$ is the $n \times r$ matrix of least squares weights. The PCA model is closed because, for all cases, $\mathbf{Z} - \mathbf{F}\mathbf{A}^T = \mathbf{0}$. This relationship holds for both unrotated and rotated loadings, ensuring model closure. Note that the sign of the recovered

scores is arbitrary, dependent upon the arbitrary sign of the loadings. As such, all scores have been arbitrarily defined to start with positive values.

Another characteristic of any similarity metric, including ES, is that the resultant modes do not retain the original amplitude of the data from which they are derived. This arises because the data needed to do so fully are distributed in all PCs, including those that have been discarded, $r+1$, $r+2$, ..., n (or, alternatively, distributed into other, unused dimensions of the eigenspace). If PCA is cast in a signal analysis paradigm, because some PCs (data) are discarded, some signal is discarded as well. The few retained components cannot recreate the total similarity contained in the original signal (the full data set). Hence, determining cell lifetime from a w time series becomes problematic because the original threshold of 10 m s^{-1} may no longer be representative. Fortunately, when applied to these w time series, another inherent characteristic of ES based on rotated PCs is that each mode displays either a sign change or rapidly decrease towards zero at some point. Either the sign change or the rapid decrease toward zero suffices to characterize the lifetime of each mode. This behavior is not significantly affected by the amplitude of the mode and is used to define the cell lifetime.

A2.4 Examples

PCA may be based upon various similarity matrices. The chosen similarity matrix significantly affects the appearance of the resulting least-squares scores. Examples of PCA that use correlation, covariance, cross-products, and ES are developed and shown for the w time series. An example is also shown for a two-dimensional station pressure field. The pressure field example uses correlation and ES to demonstrate that a PCA based upon

a ES parent matrix will extract a pattern known to exist within the data, but that a PCA based on a correlation matrix extracts different patterns that must be interpreted differently.

The example that motivates this work uses a time series of the maximum w within the spatial domain for an ensemble of individual cloud model runs, each started with slightly different initial soundings. For these time series, w is available every 1.05 min and it is possible, at least for the sample case, to determine subjectively what the dominant modes are by visual inspection (Fig. A2.1). For example, let cell lifetime be defined as that period for which vertical velocity is at least 10 m s^{-1} . Subjectively, it is desired that

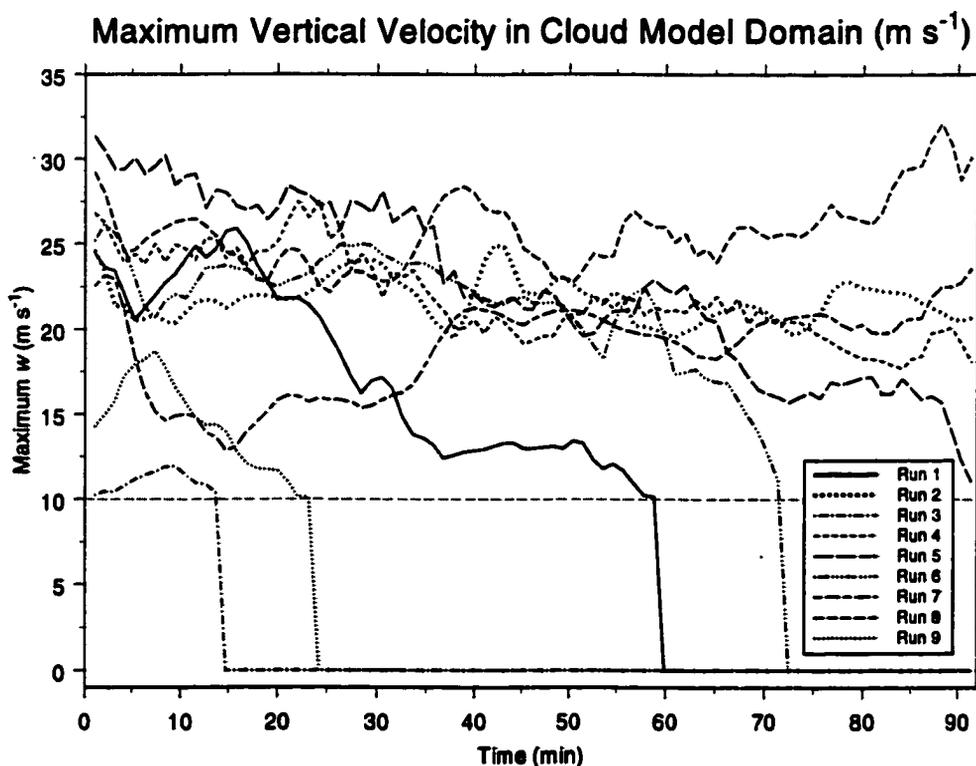


FIGURE A2.1. Vertical velocity time series from nine cloud model runs. Values represent the largest positive vertical velocity anywhere in the model domain, over a period of 92 min. The x-axis is time and the y-axis is vertical velocity in m s^{-1} . Dashed line shows the threshold for cell lifetime definition.

all of the extracted modes be based upon the apparent “closeness” of the time series to each other. Consequently, a group of time series that overlay each other or, put visually, that overlap or form a cloud, are linearly combined to form a single time series that represents a collection of similar time series. Accordingly, a cloud of w time series that last for the entire 90 min, and have similar amplitude, are linearly combined into a single mode.

Based on the above subjective definition of a mode, three modes should result from the PCA. Our *a priori* expectations are that one should have a large amplitude that lasts for the entire length of the data series (92 min). This mode results from the similarity between Runs 2, 4, 5, 7 and 8. A second mode, that lasts about two-thirds of the available data series length, and with an amplitude close to the first mode might also be reasonably expected. This second mode is driven primarily by the similarity between Runs 1 and 6. A third mode that has a low amplitude and a brief duration, is also expected. The third mode will be driven primarily by the similarity between Runs 3 and 9.

Least-squares scores that are recovered from a rotated correlation-based PCA do not result in elements that can be physically interpreted as a vertical velocity time series (Fig. A2.2). This is because these scores represent the centered (to zero mean) and scaled (to unit variance) uncorrelated vertical velocity modes.. Given the nature of the modes that are desired, scores based on correlated behavior are not the desired result. Unfortunately, these scores provide no way to scale cell lifetime. Neither do these scores provide any way to extract information about the intensity of convective activity. Similar, though not identical, results are obtained from the covariance similarity matrix (Fig. A2.3).. Given the nature of the modes that are desired, scores based on correlated behavior are not the desired result. Unfortunately, these scores provide no way to scale cell lifetime. Neither do

Rotated Vertical Velocity Correlation Least Squares Scores

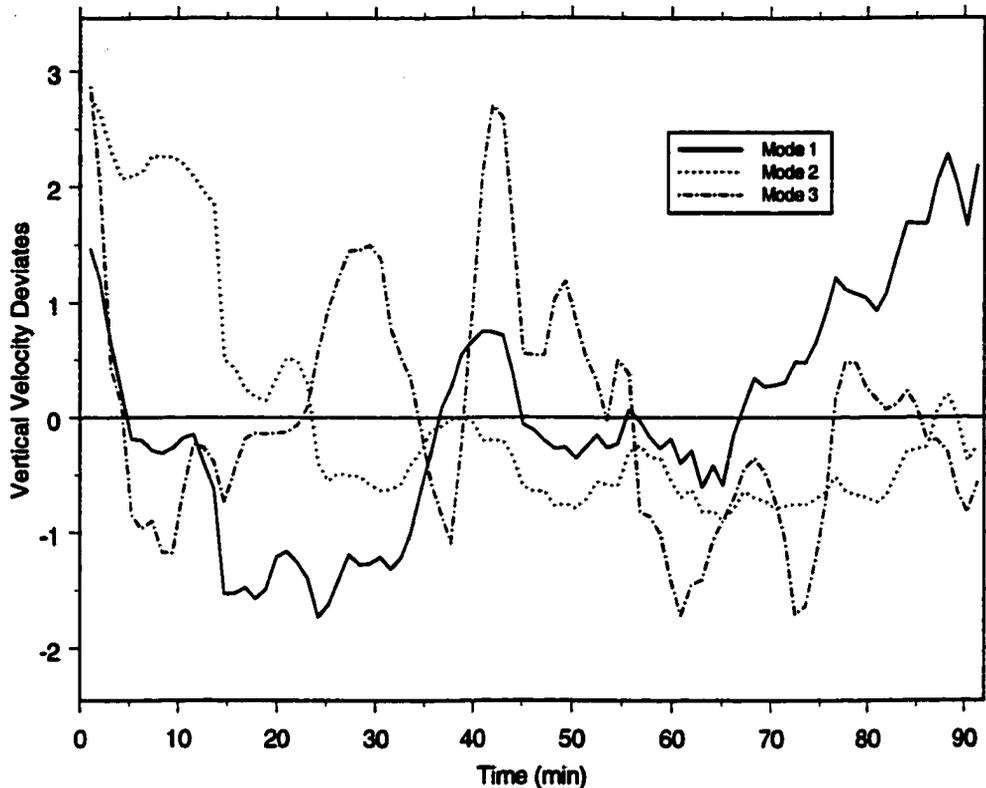


FIGURE A2.2. The first three least-squares modes for a correlation-based S-mode analysis of the vertical velocities in Fig. A2.1. The x-axis is time and the y-axis is centered, scaled vertical velocity. The solid line shows the zero reference.

these scores provide any way to extract information about the intensity of convective activity. Similar, though not identical, results are obtained from the covariance similarity matrix (Fig. A2.3). Again, these are the centered, uncorrelated vertical velocity scores. Certain segments appear similar to the original data (which is expected), but how these scores relate to updraft intensity or cell lifetime is not clear.

Least-squares score recovery with cross-products have some characteristics similar to the *a priori* expectations (Fig. A2.4).. Unfortunately, nothing that resembles vertical velocity magnitudes are preserved within the scores. It is possible to define cell lifetime with these scores by when the score magnitude rapidly decreases. Using this rule, Mode 1

Rotated Vertical Velocity Covariance Least Squares Scores

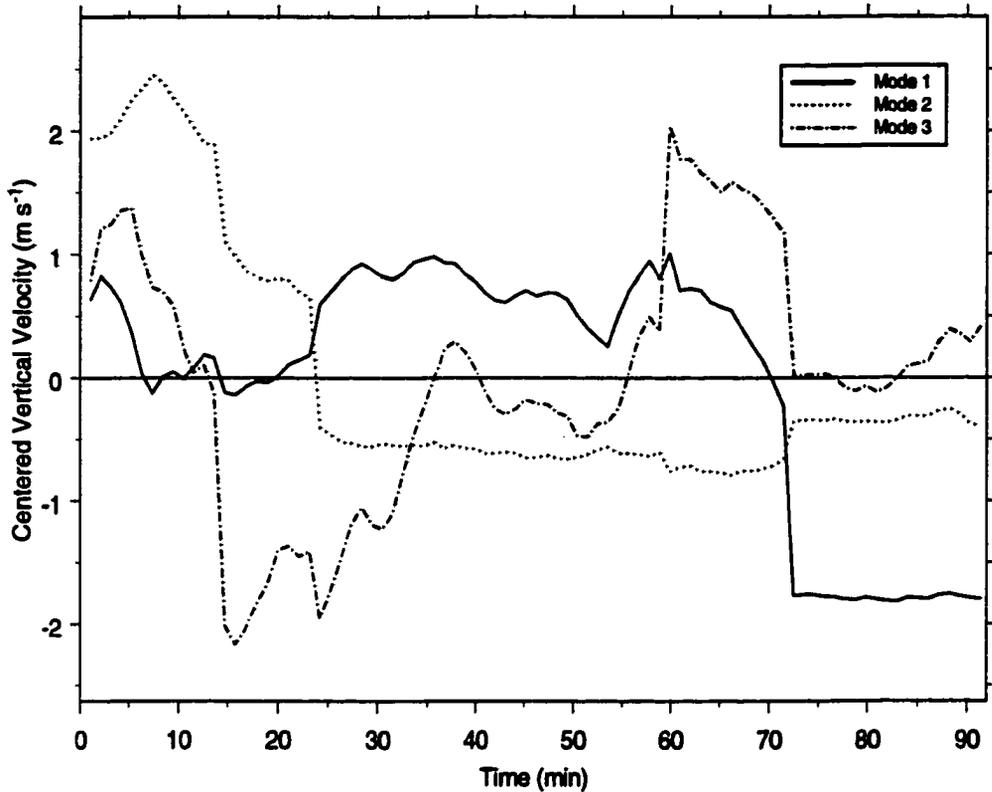


FIGURE A2.3. Same as Fig. A2.2, but for a covariance-based analysis. Here, the y-axis is centered vertical velocity, in m s^{-1} . Solid line shows the zero reference.

lasts the entire 92 min, which is an expected lifetime mode. Mode 2 length could be either 59 min or 72 min, depending upon whether the first sharp, but small, decrease or the second, larger decrease is used. Mode 3 could also be either 14 min or 23 min. However, information about relative updraft intensity is unavailable, which, along with cell lifetime, might be an important indicator of storm characteristics.

Least squares scores that result from ES meet the requirements of preserving both cell lifetime and relative vertical velocity magnitude information (Fig. A2.5).. Mode 1 is clearly a large-magnitude convective mode that lasts the entire 92 min. In this case, mode 2 is the short-lived, low-magnitude mode. Mode 3 lasts about 60 min and has a high

Rotated Vertical Velocity Cross-Products Least Squares Scores

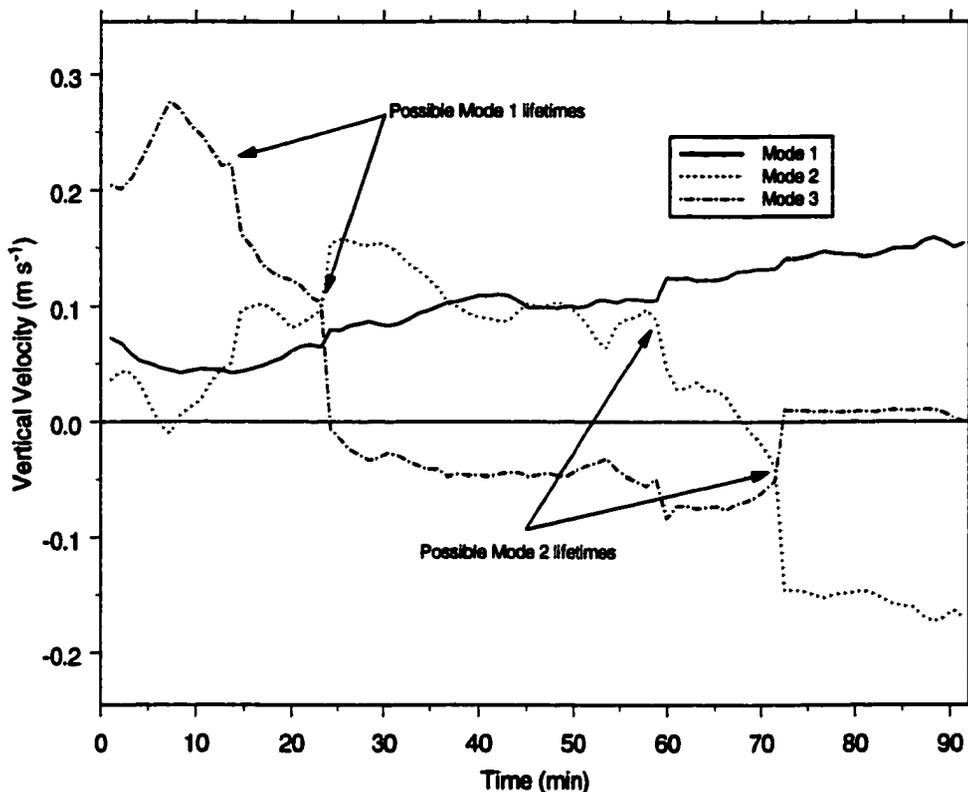


FIGURE A2.4. Same as Fig. A2.2, but for a cross-products analysis. Here, the y-axis is vertical velocity in m s^{-1} . Solid line shows the zero reference.

amplitude. As for cross-products, cell lifetimes can be defined in various ways. Here, cell lifetime is defined as that period when the least squares vertical velocity score is at least 5 m s^{-1} . Cell lifetime could also be defined as for cross-products, where the first large negative deviation signals the end of the storm. For this example, using either a 5 m s^{-1} threshold or the first large negative deviation results in equivalent cell lifetime estimates.

A second example uses a familiar meteorological parameter: surface station pressure. For this example, the data comes from the Oklahoma mesonet (Brock, et al. 1995). The data matrix consists of station pressures taken at 1 hr intervals for the entire month of October 1994, one column for each hour, which makes this a T-mode analysis. Because

Euclidean Similarity Least Squares Modes

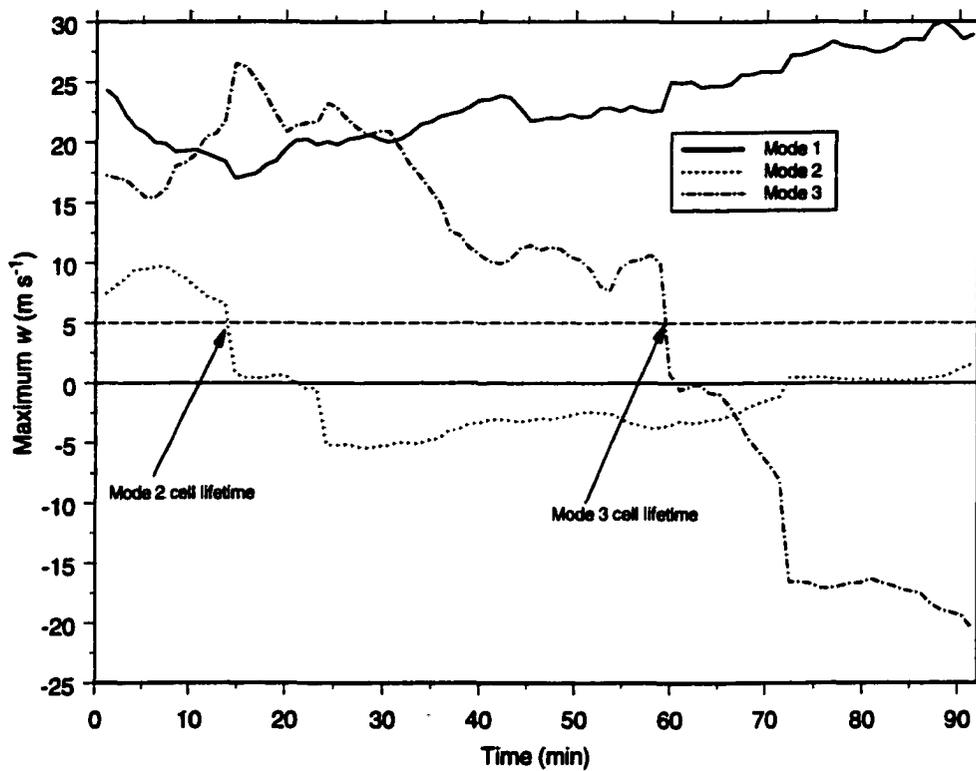


FIGURE A2.5. Same as Fig. A2.2, but for a Euclidean similarity-based analysis. Here, the y-axis is vertical velocity in m s^{-1} . Solid line shows the zero reference. Dashed line represents a 5 m s^{-1} threshold for cell lifetime definition, which is necessary because the full amplitude of the original data is not retained. See text for further details.

the data are station pressures, the overwhelming signal or pattern is driven by surface elevation (Fig. A2.6). Hence, the expected corresponding pressure mode should resemble closely the pattern of station elevations or, alternatively, the mean station pressure.

A PCA is performed using a ES matrix and a correlation matrix to provide a contrast. Two PCs are retained from the ES analysis, which explain 87% of the total ES. Alternatively, three PCs are retained from the correlation analysis, which explain 83% of the total variance. The retained PCs are rotated using the Varimax rotation algorithm. The least squares scores, which constitute modes, are recovered using Eq. A2.8.

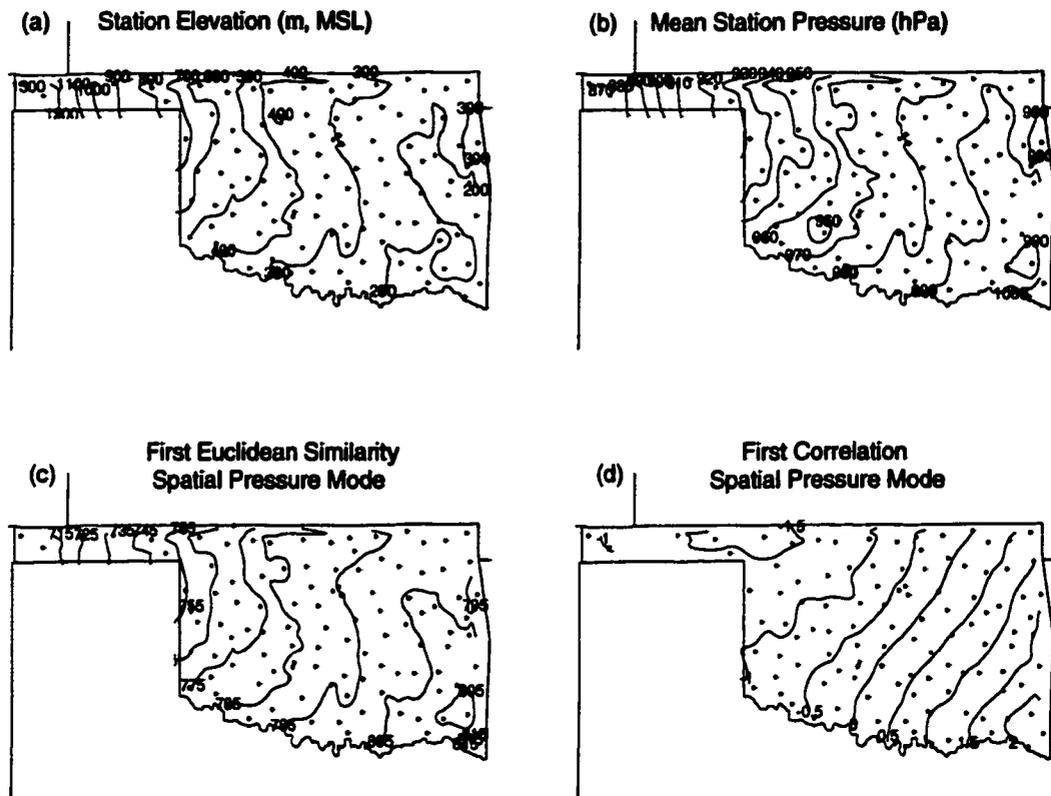


FIGURE A2.6. Oklahoma mesonet data as input to both a PCA based on ES, and a PCA based on correlation. Input data consists of hourly station pressures for the month of October 1994. Dots show individual mesonet station locations. a) Oklahoma mesonet station elevation in m; b) mean pressure, in hPa, over the Oklahoma mesonet for October 1994; c) the first modal pressure, in hPa, resulting from a PCA based on ES; d) the first modal pressure resulting from a PCA based on correlation. The field in d) is pressure normalized to a mean of zero and unit variance, and is thus dimensionless

The pressures that result from the PCA based on ES do not equal the station pressures because the total pressure similarity has been distributed to all n dimensions, and r dimensions have been retained. Despite this, the contour pattern formed by the first ES mode is strikingly similar to both the pattern of average station pressure and the pattern of station elevation. The values resulting from the PCA based on correlation cannot be interpreted as pressure, because each mode is normalized to a mean of zero and unit variance. The contour pattern formed by the first correlation mode shows a general gradient of nor-

malized pressure from northwest to southeast, but the lowest normalized pressure appears in northwestern Oklahoma, when in reality the lowest station pressure should be in the western Oklahoma panhandle. Overall, the correlation mode lacks the detail in the morphology and gradient of the isopleths contained in the ES mode. Based on the two examples provided, it is clear that a PCA based on ES can recover both one-dimensional and two-dimensional modes that are contained in the input data. Additionally, the recovered modes are easily interpreted in units that are native to the original input data. Because gradients, and their interpretation, are an important part of meteorological analyses, an ES alternative to the often-used correlation and covariance similarity matrices appears to have broad utility.

A2.5 Discussion and conclusions

Eigentechniques, such as PCA, are important, and commonly applied, tools for meteorological analysis. When PCA is performed appropriately it can lead to physical insight and understanding of large amounts of data. Traditionally, PCA has been applied to a parent similarity matrix based on either correlation, covariance, or occasionally, cross-products. This paper shows that the specific choice of the similarity matrix used to recover the PC loadings and scores can profoundly affect the results. Many meteorological data fields are distributed in both time and space, e.g., pressure, rainfall, and temperature. For single- and multi-dimensional problems embedded in such fields, our investigation illustrates that when a similarity matrix based upon Euclidean distance (ED) is constructed, the resulting PCs can lead to unique insights. Furthermore, these insights are consistent with physical factors known to control the behavior of spatial gradients in these fields.

In the PCA development presented here, rotated PC loadings and the resulting scores are used to recover these structures or co-associations, called modes. By applying the ED-based similarity matrix to data fields that have known modes of behavior, the results are shown to be reasonable. As contrasting examples, PCA scores that result from the correlation matrix, the covariance matrix, and cross-products matrix are also depicted and discussed. The ED-based results are also demonstrated to be valid for *both* S-mode and T-mode analyses. Hence, a PCA based on ED can recover both one- and two-dimensional scores. Another, useful characteristic of the modes that result from an ED-based PCA is that they tend to preserve physically-interpretable gradients (both the gradient and the direction of the gradient) within the original data fields.

This technique is intended to illustrate both the utility and flexibility of eigenanalysis. Therefore, investigators should not feel eigenanalysis is necessarily constrained by two or perhaps three popular similarity matrices, nor should the analysis step of choosing a similarity matrix be taken lightly. Instead, the large number of similarity matrices which are available, such as Mahalanobis distance, similarity based on the L_1 distance, and theta angle between entities (Anderberg 1973), must be considered. Hence, the choice of a similarity matrix is limited only by the desired output and the analyst's insight into the best procedure to maximize the underlying physics that can emerge from a PCA. As a bonus, such investigations constitute an opportunity to expand upon eigenanalysis as a diagnostic tool.

Appendix 3: Kolmogorov-Smirnov Goodness-of-Fit Test

The Kolmogorov-Smirnov goodness-of-fit (KS) test uses the empirical cdf of two different samples to determine, for some p-level, if the samples come from indistinguishable or different distributions. The following development follows that of Blum and Rosenblatt (1972).

Let X_1, \dots, X_n be independent, observed random variables with a common unknown cdf, F_O . For each real x , let $F_E(x)$ be the proportion of X_1, \dots, X_n satisfying $X_i \leq x$. F_E is called the empirical cdf based on X_1, \dots, X_n . The values of $F_E(x)$ (the proportion of $X_i \leq x$) should be close to $F_O(x) = P\{X \leq x\}$ for large n . In fact, since

$$nF_E(x) = \sum_{i=1}^n Y_i, \text{ where } Y_i = \begin{cases} 1 & \text{if } X_i \leq x \\ 0 & \text{otherwise,} \end{cases} \quad (\text{A3.1})$$

then $nF_E(x)$ is a binomial random variable with parameters n and $F_O(x)$. That $nF_E(x)$ is binomial is important because it allows the Chebychev inequality to be used directly. The Chebychev inequality is invoked as:

$$P\left\{|F_E(x) - F_O(x)| \leq k \sqrt{\frac{F_O(x)[1 - F_O(x)]}{n}}\right\} \geq 1 - \frac{1}{k^2} \quad (\text{A3.2})$$

where k is a constant. This states the probability that $|F_E(x) - F_O(x)|$ is within at least k standard deviations of zero, where the standard deviation is given by

$$\sqrt{\frac{F_O(x)[1 - F_O(x)]}{n}}.$$

Hence, when n is large,

$$P\left\{|F_E(x) - F_O(x)| \leq \frac{k}{2\sqrt{n}}\right\} \geq 1 - \frac{1}{k^2}, \quad (\text{A3.3})$$

showing that for large n , $F_E(x)$ is close to $F_O(x)$ with high probability if the two parent cdfs are indistinguishable. Let the null hypothesis be $H_0: F_E = F_O$. It seems reasonable to use some measure of the distance between F_E and F_O as a test statistic. That statistic is the maximum vertical distance, D_n , between F_E and F_O . Thus, D_n is real and within the interval $[0, 1]$.

For each α in $(0, 1)$, there exists a value, h_α , such that

$$\lim_{n \rightarrow \infty} P\{D_n \sqrt{n} \leq h_\alpha\} = 1 - \alpha, \quad (\text{A3.4})$$

where α is any real number in the interval $[0, 1]$. Typically, α is used to define the significance level, such as 0.05 for a 95% confidence level.

In fact,

$$\lim_{n \rightarrow \infty} P\{D_n \sqrt{n} \leq h_\alpha\} = 1 - 2 \sum_{j=1}^n (-1)^{j+1} e^{-2j^2 h_\alpha^2}, \quad (\text{A3.5})$$

and, from this expansion, h_α can be numerically calculated and the significance level determined.

To test the hypothesis H_0 , that the cdf, F_O , of the independent random variables X_1, \dots, X_n is indistinguishable from the empirical cdf, F_E , at approximately the α level, reject H_0 if the value of $D_n \sqrt{n}$ exceeds h_α . More generally, to test the hypothesis H'_0 that $D_n \sqrt{n}$ at approximately the α level, reject H'_0 if the value of $D_n \sqrt{n} > k \sqrt{n} + h_\alpha$.

The KS test can be used when the χ^2 test cannot because the KS test does not test against a theoretical, or parametric, distribution. The χ^2 test assumes that the source of the F_E cdf is the Gaussian pdf while the KS test does not depend on any assumptions about the nature of F_E .

References

- Anderberg, M. R., 1973: *Cluster Analysis for Applications*. Academic Press, 359 pp.
- Anderson, J. L., 1996a: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. of Climate*, **9**, 1518-1530.
- Anderson, J. L., 1996b: Selection of initial conditions for ensemble forecasts in a simple perfect model framework. *J. Atmos. Sci.*, **53**, 22-36.
- Barnes, S. L., 1964: A technique for maximizing details in numerical weather map analysis. *J. Appl. Meteor.*, **3**, 396-409.
- Barnes, S. L., 1973: Mesoscale objective analysis using weighted time-series observations. NOAA Tech, memo. ERL NSSL-62, National Severe Storms Laboratory, Norman, OK 73069, 60 pp. [NTIS COM-73-10781].
- Battán, L. J., 1952: Formation of precipitation in natural clouds indicated by radar. Preprints, *Third Conf. on Radar Meteorology*, Montreal, PQ, Canada, Amer. Meteor. Soc., A9-A16.
- Battán, L. J., 1953: Duration of convective radar cloud units. *Bull. Amer. Meteor. Soc.*, **34**, 227-228.
- Black, T. L., 1994: The new NMC mesoscale eta model: Description and forecast examples. *Wea. Forecasting*, **9**, 265-278.
- Bluestein, H. B., 1993: *Synoptic-Dynamic Meteorology in Midlatitudes, Vol. II: Observations and Theory of Weather Systems*. Oxford University Press, 594 pp.
- Blum, J. R and J. I. Rosenblatt, 1972: *Probability and Statistics*. W. B. Saunders Co. 549 pp.
- Brock, F. W., K. C. Crawford, R. L. Elliot, G. W. Cuperus, S. J. Stadler, H. J. Johnson, and M. D. Eilts, 1995: The Oklahoma mesonet: A technical overview. *J. Atmospheric and Oceanic Tech*, **12**, 5-19.
- Brooks, H. E., B. F. Jewitt, L. J. Wicker, J. M. Straka, D. Vigneux, R. B. Wilhelmson, 1990: the role of hail in numerical simulations of the 29 May 1986 Montreal storm. Preprints, *16th Conf. on Severe Local Storms*, Kananaskis Park, Alta., Canada. Amer. Meteor. Soc.
- Brooks, H. E., 1992: Operational implications of the sensitivity of modeled thunderstorms to thermal perturbations. *Preprints, Fourth Workshop on Operational Meteorol-*

- ogy, Whistler, BC, Atmospheric Environment Services/Canadian Meteor. and Oceanographic Soc., 398-407.
- Brooks, H. E., C. A. Doswell III, and R. A. Maddox, 1992: On the use of mesoscale and cloud-scale models in operational forecasting. *Wea. Forecasting*, **7**, 352-362.
- Brooks, H. E., C. A. Doswell III, and L. J. Wicker, 1993: STORMTIPE: A forecasting experiment using a three-dimensional cloud model. *Wea. Forecasting*, **8**, 352-362.
- Brooks, H. E., C. A. Doswell III, and J. Cooper, 1994: On the environments of tornadic and nontornadic mesocyclones. *Wea. and Forecasting*, **9**, 606-618.
- Brooks, H. E., C. A. Doswell III, and R. B. Wilhelmson, 1994: On the role of midtropospheric winds in the evolution and maintenance of low-level mesocyclones. *Mon. Wea. Rev.*, **122**, 126-136.
- Browning, K. A., 1964: Airflow and precipitation trajectories within severe local storms which travel to the right of the winds. *J. Atmos. Sci.*, **21**, 634-639.
- Buell, C. E., 1975: The topography of empirical orthogonal functions. Preprints, *Fourth Conf. on Prob. and Stats. in Atmos. Sci.*, Tallahassee, FL, Amer. Meteor. Soc. 188-193.
- Buizza, R., 1997: Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system. *Mon. Wea. Rev.*, **125**, 99-119.
- Buizza, R. and T. N. Palmer, 1998: Impact of ensemble size on ensemble prediction. *Mon. Wea. Rev.*, **126**, 2503-2518.
- Burgess, D. W., V. T. Wood and R. A. Brown, 1982: Mesocyclone evolution characteristics. Preprints, *12th Conf. on Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., 422-424.
- Byers, H. R. and R. R. Braham, 1949: *The Thunderstorm*. U.S. Govt. Printing Off., Washington, D.C., 187 pp.
- Carpenter, R. L. Jr., K. K. Droegemeier, G. M. Bassett, W. L. Qualley and R. Strasser, 1997: Project Hub-CAPS: Storm-scale NWP for commercial aviation. *7th Conf. on Aviation, Range and Aerospace Meteorology*, Long Beach, CA, Amer. Meteor. Soc., 474-479.
- Carpenter, R. L. Jr., K. K. Droegemeier, G. M. Bassett, K. Brewster, D. E. Jahn, J. Levit, M. Xue, W. L. Qualley and R. Strasser, 1998: Storm-scale NWP for commercial

- aviation: Results from real-time operational tests in 1996-1997. *12th Conf. on Numerical Weather Prediction*, Phoenix, AZ, Amer. Meteor. Soc., 213-216.
- Chappell, C. F., 1986: Quasi-stationary convective events. In *Mesoscale Meteorology and Forecasting*, P.S. Ray (Ed.), Amer. Meteor. Soc., 289-310.
- Cheng, X., G. Nitsche and J. M. Wallace, 1995: Robustness of low frequency circulation patterns derived from EOF and totaled EOF analysis. *J. Climate*, **8**, 1709-1713.
- Cleveland, W. S., 1994: *The Elements of Graphing Data*. AT&T Bell Laboratories, Murray Hill, NJ, 297 pp.
- Crook, N. A., 1996: Sensitivity of moist convection forced by boundary layer processes to low-level thermodynamic fields. *Mon. Wea. Rev.*, **124**, 1767-1785.
- Cooley, W. W. and P. R. Lohnes, 1971: *Multivariate Data Analysis*. John Wiley and Sons, Inc., New York, NY, 364 pp.
- Cortinas, J. V. Jr. and D. J. Stensrud, 1995: The importance of understanding mesoscale model parameterization schemes for weather forecasting. *Wea. and Forecasting*, **10**, 716-740.
- Cotton, W. R. and R. A. Anthes, 1989: *Storm and Cloud Dynamics*. Academic Press, New York, NY. 880 pp.
- Craddock, J. M. and C. R. Flood, 1969: Eigenvectors for representing the 500-mb geopotential surface over the northern hemisphere. *Quart. J. Roy. Meteor. Soc.*, **95**, 576-593.
- Davies, J. M., 1993: Hourly helicity, instability, and EHI in forecasting supercell tornadoes. Preprints, *17th Conf. on Severe Local Storms*, St. Louis, MO, Amer. Meteor. Soc., 107-111.
- Dixon, M. and G. Wiener, 1993: TITAN: Thunderstorm identification, tracking, analysis and nowcasting – a radar-based methodology. *J. Atmos. Oceanic Tech.*, **10**, 785-797.
- Doswell, C. A. III, 1985: *The Operational Meteorology of Convective Weather, Volume II: Storm Scale Analysis*. NOAA Tech. Memo. ERL ESG-15, 240 pp.
- Doswell, C. A. III, S. J. Weiss, and R. H. Johns, 1993: Tornado forecasting: A review. *The Tornado: Its Structure, dynamics, Prediction, and Hazards*, 557-571. Am. Geophys. Union, Washington, DC.

- Du, J., S. L. Mullen and F. Sanders, 1997: Short-range ensemble forecasting of quantitative precipitation. *Mon. Wea. Rev.*, **125**, 2427-2459.
- Efron, B. and R. J. Tibshirani, 1993: *An Introduction to the Bootstrap*. Chapman and Hall, New York, NY, 436 pp.
- Ehrendorfer, M., 1994: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part II: Applications. *Mon. Wea. Rev.*, **122**, 714-728.
- Emanuel, K. A., 1994: *Atmospheric Convection*. Oxford University Press, New York, NY, 580 pp.
- Epstein, E. S., 1969: Stochastic dynamic prediction. *Tellus*, **21**, 739-759.
- Evans, J. E., 1997: Operational problem of convective weather in the national airspace system. Preprints, *Convective Weather Forecasting Workshop*, Long Beach, CA, A1-A14.
- Freidman, J. H., 1984: A Variable Span Smoother. Technical Report No. 5, Laboratory for Computational Statistics, Department of Statistics, Stanford University. 30 pp.
- Gilman, D. L., 1957: Empirical Orthogonal Functions Applied to Thirty-Day Forecasting. M.I.T. Dept. of Meteorology Science Report No. 1. Contract AF19(604)-1283.
- Glahn, H. R., 1965: Objective weather forecasting by statistical methods. *The Statistician*, **15**, 111-142.
- Gong, X. and M. B. Richman, 1994: On the application of cluster analysis to growing season precipitation data in North America east of the Rockies. *J. of Climate*, **8**, 897-931.
- Harman, H. H., 1976: *Modern factor Analysis*. The University of Chicago Press, 487 pp.
- Henry, S. G., 1993: Analysis of thunderstorm lifetime as a function of size and intensity. Preprints, *26th Conf. on Radar Meteorology*, Norman, OK, Amer. Meteor. Soc., 138-140.
- Hotelling, H., 1933: Analysis of a complex of statistical variables into principal components, *J. Educ. Psych.*, **24**, 417-520.
- Houtekamer, P. L., J. Derome, 1995: Methods for ensemble prediction. *Mon. Wea. Rev.*, **123**, 2181-2196.
- Houze, R. A., 1993: *Cloud Dynamics*. Academic Press, New York, NY, 573 pp.

- Jewitt, B. F., R. B. Wilhelmson, J. M. Straka, and L. J. Wicker, 1990: Impact of ice parameterization on the low-level structure of modeled supercell thunderstorms. Preprints, *16th Conf. on Severe Local Storms*. Kananaskis Park, Alberta. Amer. Meteor. Soc., 275-280.
- Johns, R. and J. A. Hart, 1993: Differentiating between types of severe thunderstorm outbreaks: A preliminary investigation. Preprints, *17th Conf. on Severe Local Storms*, Saint Louis, MO, Amer. Meteor. Soc., 46-50.
- Johns, R. H. and C. E. Doswell III, 1992: Severe local storms forecasting. *Wea. and Forecasting*, **7**, 588-612.
- Johnson, J. T., P. L. MacKeen, A. Witt, E.D. Mitchell, G. J. Stumpf, M. D. Eilts, and K. W. Thomas, 1998: The storm cell identification and tracking algorithm: an enhanced WSR-88D algorithm. *Wea. And Forecasting*, **13**, 263-276.
- Johnson, D. E., P. K. Wang, and J. M. Straka, 1993: Numerical simulations of the 2 August 1981 CCOPE supercell storm with and without ice microphysics. *J. Appl. Meteor.*, **29**, 745-759.
- Jolliffe, I. T., 1995: Rotation of principal components: choice of normalization constraints. *J. Appl. Statistics*, **22**, 29-35.
- Kaiser, H. F., 1958: The Varimax criterion for analytic rotation in factor analysis. *Psychometrika*, **23**, 187-200.
- Kessler, E., 1969: On the distribution and continuity of water substance in atmospheric circulation. *Meteor. Monogr.*, **10**, no. 32, 83 pp, Amer. Meteor. Soc.
- Klemp, J. B. and R. B. Wilhelmson, 1978: The simulation of three-dimensional convective storm dynamics. *J. Atmos. Sci.*, **35**, 1070-1096.
- Kopp, F. J. and H. D. Orville, 1994: The use of a two-dimensional, time-dependent cloud model to predict convective and stratiform clouds and precipitation. *Wea. and Forecasting*, **9**, 62-77.
- Krause, A. and M. Olson, 1997: *The Basics of S and S-Plus*. Springer-Verlag, New York, NY, 242 pp.
- Karl, T. R. and A. J. Koscielny, 1982: Drought in the United States: 1895-1981. *J. Climatology*, **2**, 313-329.
- Kutzbach, J. E., 1967: Empirical eigenvectors and sea-level pressure, surface temperature, and precipitation complexes over North America. *J. Appl. Meteor.*, **6**, 791-802.

- Kutzbach, J. E., 1969: Large-scale features of monthly mean northern hemisphere anomaly maps of sea-level pressure. *Mon. Wea. Rev.*, **98**, 708-716.
- Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409-418.
- Lilly, D. K., 1990: Numerical prediction of thunderstorms—has its time come? *Quart. J. Roy. Meteor. Soc.*, **116**, 779-798.
- Lorenz, E. N., 1956: Empirical orthogonal functions and statistical weather prediction. Scientific Report 1, Statistical Forecasting Project, Dept. of Meteor., MIT, 49 pp. [NTIS AD 110268].
- Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130-141.
- Lorenz, E. N., 1965: A study of the predictability of a 28-variable atmospheric model. *Tellus*, **17**, 321-333.
- Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289-307.
- Lorenz, E. N., 1993: *The Essence of Chaos*. University of Washington Press, Seattle, WA. 227 pp.
- MacKeen, P. L., H. E. Brooks and K. L. Elmore, 1999: Radar-reflectivity derived thunderstorm parameters applied to storm longevity forecasting. *Wea. and Forecasting*, **14**, 289-295.
- Marshall, J. S. and W. M. Palmer, 1948: The distribution of raindrops with size. *J. Meteor.*, **5**, 165-166.
- Marshall, C. H. 1998: Evaluation of the New Land-Surface and PBL Parameterization Schemes in the NCEP Mesoscale ETA Model Using Oklahoma Mesonet Observations. M.S. Thesis, Univ. of Okla., Norman, OK, 176 pp.
- Marzban, C., 1998: Scalar measures of performance in rare-event situations. *Wea. and Forecasting*, **13**, 753-763.
- McPherson, R. A., 1992: *Predictability Experiments of a Numerically Modeled Supercell Storm*. M.S. Thesis, Univ. of Okla., Norman, OK, 128 pp.
- Mills, G. A. and J. R. Colquhoun, 1998: Objective predictions of severe thunderstorm environments: Preliminary results linking a decision tree with an operational regional NWP model. *Wea. and Forecasting*, **13**, 1078-1092.

- Molteni, F., R. Buizza, T. N. Palmer and T. Petroliaigis, 1996: The ECMWF ensemble prediction system: methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73-119.
- Mueller, C. K., J. W. Wilson and N. A. Crook, 1993: The utility of sounding and mesonet data to nowcasting thunderstorm initiation. *Wea. Forecasting*, **8**, 132-146.
- Murphy, A. H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281-293.
- Preisendorfer, R. W. and T. P. Barnett, 1983: Numerical model-reality intercomparison tests using small sample statistics. *J. Atmos. Sci.*, **40**, 1884-1896.
- Proctor, F. H., 1988: Numerical simulations of an isolated microburst. Part I: Dynamics and structure. *J. Atmos. Sci.*, **45**, 3137-3160.
- Proctor, F. H., 1989: Numerical simulations of an isolated microburst. Part II: Sensitivity experiments. *J. Atmos. Sci.*, **46**, 2143-2165.
- North, G. R., T. L. Bell, R. F. Calahan and F. J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.*, **110**, 699-706.
- Resio, D. T. and B. P. Hayden, 1975: Recent secular variations in mid-Atlantic winter extratropical storm climate, *J. Appl. Meteor.*, **14**, 1223-1234.
- Rhoda, D. A. and M. L. Polka, 1999: An Assessment of Thunderstorm Penetrations and Deviations by Commercial Aircraft in the Terminal Area. Project Report NASA/A-2, Lincoln Laboratory, Massachusetts Institute of Technology. Available through NTIS, Springfield, VA. 79 pp.
- Richman, M. B., 1986: Rotation of principal components. *International J. of Climatology*, **6**, 293-335, Roy. Meteor. Soc.
- Richman, M. B. and X. Gong, 1999: Relationships between the definition of the hyperplane width in the fidelity of principal component loading patterns. *J. Climate*, **12**, 1557-1576.
- Rogers, R. R., 1976: *A Short Course in Cloud Physics*. Pergamon Press, 227 pp.
- Schlesinger, R. E., 1975: A three dimensional numerical model of an isolated deep convective cloud: Preliminary results. *J. Atmos. Sci.*, **32**, 934-957.
- Schwartz, B. E., C. F. Chappell, W. E. Togstad and X. Zhong, 1990: The Minneapolis flash flood: Meteorological analysis and operational response. *Wea. and Forecasting*, **5**, 3-21.

- Sellers, W. D., 1957: A Statistical-Dynamic Approach to Numerical Weather Prediction. *M.I.T. Dept. of Meteorology Scientific report No. 2*, Statistical Forecasting Project, Contract AF19(604)-1566.
- Silverman, B. W., 1986: *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York, NY, 175 pp.
- Sivillo, J. K., J. E. Alquist and Z. Toth, 1997: An ensemble forecasting primer. *Wea. and Forecasting*, **12**, 809-818.
- Stensrud, D. J., J. V. Cortinas Jr. and H. E. Brooks, 1997: Discriminating between tornadic and nontornadic thunderstorms using mesoscale model output. *Wea. and Forecasting*, **12**, 613-632.
- Stensrud, D. J., J. W. Bao and T. T. Warner, 1999: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.*, in press.
- Stevenson, L., 1997: "Emerging Picture" of WSDDM's potential efficiency benefits based on field demonstrations. Draft Letter Report, available from Volpe National Transportation Systems Center, Washington, D.C.
- Straka, J. M. and E. N. Rasmussen, 1998: Thirty years of cloud modeling: does the emperor wear clothes? Preprints, *19th Conf. on Severe Local Storms*, Minneapolis, MN. Amer. Meteor. Soc., 342-347.
- Tao, W., J. Simpson and M. McCumber, 1989: An ice-water saturation adjustment. *Mon. Wea. Rev.*, **117**, 231-235.
- Thurstone, L. L., 1947: *Multiple Factor Analysis*. Univ. of Chicago Press, 535 pp.
- Toth, Z. and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317-2330.
- Venables, W. N. and B. D. Ripley, 1997: *Modern Applied Statistics with S-Plus*. Springer-Verlag, New York, NY, 548 pp.
- Wang, D. H., M. Xue, V. Wong, K. K. Droegemeier, 1996: Predictions and simulations of convective storms during VORTEX95. Preprints, *11th Conf. on Numerical Weather Prediction*, Norfolk, VA, Amer. Meteor. Soc.
- Weisman, M. L., and J. B. Klemp, 1982: The dependence of numerically simulated convective storms on vertical wind shear and buoyancy. *Mon. Wea. Rev.*, **110**, 504-520.

- Weisman, M. L., and J. B. Klemp, 1984: The structure and classification of numerically simulated convective storms in directionally varying wind shears. *Mon. Wea. Rev.*, **112**, 2479-2498.
- Weisman, M. L. and J. B. Klemp, 1986: Characteristics of isolated convective storms. In *Mesoscale Meteorology and Forecasting*, P.S. Ray (Ed.), Amer. Meteor. Soc., 331-358.
- Wicker, L. J., M. P. Kay, M. P. Foster, 1997: STORMTIPE-95: Results from a convective storm forecast experiment. *Wea. Forecasting*, **12**, 362-372.
- Wicker, L. J., and R. B. Wilhelmson, 1995: Simulation and analysis of tornado development and decay within a three-dimensional supercell thunderstorm. *J. Atmos. Sci.*, **52**, 2675-2703.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction*. Academic Press, New York, NY, 467 pp.
- Wilson, J. W., 1966: Movement and predictability of radar echoes. Environmental Science Services Administration Tech. Memo. IERTM-NSSL-28, 30 pp. [Available from National Information Service, Operations Division, Springfield, VA 22161].
- Wilson, J. W. and C. K. Mueller, 1993: Nowcasts of thunderstorm initiation and evolution. *Wea. Forecasting*, **8**, 113-131.
- Wilson, J. W., N. A. Crook, C. K. Mueller, J. Sun and M. Dixon, 1998: Nowcasting thunderstorms: a status report. *Bull. Amer. Meteor. Soc.*, **79**, 2079-2099.
- Witt, A., M. D. Eilts, G. J. Stumpf, J. T. Johnson, E. D. Mitchell and K. W. Thomas, 1998: An enhanced hail detection algorithm for the WSR-88D. *Wea. Forecasting*, **13**, 286-303.
- Wood, V. T. and R. A. Brown, 1997: Effects of radar sampling on single-Doppler velocity signatures of mesocyclones and tornadoes. *Wea. and Forecasting*, **12**, 928-938.
- Wolfson, M. M., C. K. Mueller and M. D. Eilts, 1997: Convective weather forecasting for FAA applications. Preprints, *7th Conf. on Aviation, Range and Aerospace Meteorology*, Long Beach, CA, Amer. Meteor. Soc.
- Xue, M., K. K. Droegemeier, V. Wong, A. Shapiro, K. Brewster, 1995: *ARPS Version 4.0 User's Guide*. Center for Analysis and Forecasting of Storms, Univ. of Okla, Norman, OK, 380 pp. [available from CAPS, U. of Okla., Norman, OK 73019].