ENSEMBLE PREDICTION OF A SEVERE WEATHER EVENT: A STUDY OF THE 2009 SOUTHERN ONTARIO STORM

A THESIS SUBMITTED IN PARTIAL
FUFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

MASTER OF SCIENCE

GRADUATE PROGRAM IN
EARTH AND SPACE SCIENCE

YORK UNIVERSITY
TORONTO, ONTARIO

November, 2012

Abstract:

The tornadic storm of August 20, 2009 of Southern Ontario is studied using a numerical prediction model. It is found that a 3km resolution simulation works as well as a 1km resolution model to model the storms underlying physical processes relevant to supercell mesocyclone formation and storm propagation, although both models showed a significant phase bias in the storm system's squall line position. A 3m resolution ensemble of 20 members is used to model the storm system further, and it is found that the ensemble mean shows the same bias that the 1km and 3km resolution models exhibited. Investigation of ensemble perturbation growth rates from ensemble mean values reveals differing growth rates for baroclinic and convective modes. Ensemble-based sensitivity analyses reveal that there are strong correlations of squall line position with model variables up to 12 hours previously.

Dedication:

To my mother, Christina Bell, who's unfailing love and encouragement taught me to believe in myself, and who showed me that nothing is impossible. You will always be remembered Mum.

Acknowledgements:

This thesis would not have been possible without the invaluable direction of my advisor, Professor Yongsheng Chen. His generous willingness to share his time and expertise, and his patience with my efforts both in the background research and in the completion of this thesis, have made what would otherwise have been a difficult task a much easier and more successful pursuit. He has taught me much about how science works and is conducted, and has consistently been a very helpful and highly capable academic advisor.

I would also like to thank Prof. Peter Taylor, for initially suggesting to me that a graduate degree in Atmospheric Science would be a rewarding pursuit. It is to him, and to Professor Gary Klaassen, that I also owe thanks for both serving on my thesis review committee and for offering me much useful direction during my studies at York. I have also learned much from them, and am in their debt. I must also thank Professor Dong Liang for his time and his helpful suggestions on my review committee.

I am also grateful to my fellow students, Richard Liang, Zongqi Yu, Nan Miao and Sopan Kurkute for offering me helpful advice and input while I worked on my thesis.

My gratitude is also due to York University for its generous financial support in the completion of this degree.

I would also like to thank my superiors at IPC Energy, John Andrews and Terry Hawrysh, for their understanding in allowing me to pursue this degree part time while employed at IPC.

Most of all I would like to thank my family – my wife Jenny Bell, and my daughter Dana Bell – for their understanding for the time lost over the many weekends, evenings and mornings while I worked on this thesis, and to my father and mother, Ron Bell and Christina Bell, for providing me with their support and encouragement.

List of Tables:

List of Figures:

Table of Contents:

Motivation

The primary focus of this research study is to determine how well an ensemble of moderately high resolution (3km) numerical weather prediction (NWP) simulations can model an intense, deeply-convective mesoscale phenomenon. Attention will be paid to the performance of such an ensemble in comparison to that of a single higher resolution (1km) deterministic NWP simulation. The phenomenon that will be modeled is the tornadic squall line system which hit southern Ontario on August 20, 2009. A secondary focus will also relate to what other benefits can be realized by using an ensemble-based approach to modeling, over and above an expected increased forecast accuracy. Ideas associated with the relationship between the scale of features forecast, and their predictability will be explored as well.

This study will be broken down into three sections, and each these sections will be addressed in a separate chapter.

The first chapter will attempt to address the following question – are the physics that are being modeled at the 3km resolution the same as those being modeled by the 1km resolution simulation? This is an important consideration, for in both

cases convection will be fully resolved, with no model convective parameterization at these resolutions. Given the obvious intensity of convection of this storm, and that several studies (which we will present below) which have called into question the ability of NWP models to adequately model convection at these scales, it is important to establish that the same processes are being modeled by both the 1km and 3km models, before we start comparing the results of the 3km ensemble model to the 1km deterministic model.

Once we have established that the same physics are being modeled at both resolutions, the second chapter will attempt to determine how well the predictions of an ensemble of 3km resolution simulations perform when compared to that of a single 1km resolution simulation. Specifically, the performance of the ensemble mean's predictions will be investigated.

Finally, in the third chapter we will deal with another use of the ensemble – how the ensemble can be used for sensitivity analyses of modeled phenomenon. We will show here how model variables at much earlier times can correlate strongly with the final squall line position, even up to 12 hours previously in some locations. Such correlations are useful, for they indicate areas and variables where changes in observational data have significant impacts on our final squall

line position. If we should find that such areas of high correlation fall into areas where few observations have been taken, we would know that we could improve upon our model's accuracy by taking more observations in these areas, incorporating these results using the well-established practice of data assimilation. Note that our results will also show that data assimilation would be a necessary best step to improve upon our results, as we will see that both the high resolution and ensemble results show a similar significant phase error in the squall line position.

Note that, as mentioned above, issues concerning the relationship between feature scale and predictability will be addressed throughout the study. We will see that such issues are highly relevant to all of the considerations discussed above.

Finally, it should be noted that the use of ensemble methods for the investigation of highly convective mesoscale phenomena is a relatively new field of study. It is therefore hoped that the current study might help in some small way to add to the still somewhat limited body of knowledge related to this area of investigation.

1 Relative Performance of Low and High Resolution Deterministic Simulations

1.1 Introduction

As noted above, one of the main aims of this study will be to compare the performance of an ensemble of low resolution simulations to a higher resolution deterministic simulation. Before we investigate this issue (in Chapter 2), there is a concern with using this approach that must first be addressed. This concern relates to an uncertainty that the 3km resolution simulation is capable of capturing the same physical processes that the 1km resolution simulation is capable of. These concerns arise because of issues raised in several papers regarding the ability of simulations in this range to fully model convection (which we will elaborate upon further below). While it has been shown in this literature that convective parameterization is not used below a 4km resolution, there is some concern whether or not that, at resolutions at and above 1km, the direct explicit simulation of convection by the model is indeed correct.

Thus this chapter will present an examination of the ability of a 3km lower resolution simulation to correctly model the physical processes of our subject squall line system in comparison with those of the 1km resolution simulation.

This first step is important, for as mentioned above, our ultimate goal of this study will be to determine if an ensemble of 3km simulations can match or indeed exceed the performance of a single 1km deterministic simulation, and if our ensemble can offer benefits not afforded by a single deterministic simulation. For one to have confidence that the ensemble has an equal ability to adequately reflect the the reality of what is being simulated, one would expect that both the deterministic 1km high resolution simulation, and the 3km low resolution simulation[1] upon which the ensemble is based, would show similar abilities to capture the physics of the modeled event – in this case, a severe squall line formed on August 20, 2009 that spawned 11[2] separate tornadoes across Southern Ontario. Naturally, the characteristics that should be compared between the two simulations are those of a tornadic storm. For example, if the 1km simulation should capture strong vertical updraft velocities and vorticities along the squall line, but the 3km simulation should prove incapable of capturing or simulating these expected properties of the supercells of our squall line system, then any attempt to draw conclusions from an ensemble of such simulations would be highly questionable. Thus, our first step, and the first part of this study, will endeavor to answer the question of how well a 3km simulation is

1 We realize of course that in present NWP practice, a 3km resolution simulation is not regarded as a low resolution simulation – but for the purposes of this study, we will refer to the 3km resolution simulation as the low resolution simulation to contrast it with the higher resolution 1km deterministic simulation.

2 http://www.ec.gc.ca/default.asp?lang=En&n=714D9AAE-1&news=A4D26BC9-8777-4E5B-8258-C7213A24606C

able to model the expected physical properties of a squall line of severely convective tornadic supercells as compared to a deterministic 1km simulation, and to answer if even a 1km simulation is able to adequately model such properties. To frame this study as a whole, we will first provide a synoptic overview of the storm system being studied.

1.1.1 Synoptic Overview of Storm System

While it is not the purpose of this research project to conduct a thorough synoptic analysis of the storm event itself, an overview will nevertheless be helpful to provide some synoptic background of the event, and to provide some context which will help in our understanding of the storm's genesis and evolution.

The National Oceanic and Atmospheric Administration (NOAA) Storm Prediction Center (SPC)[3] predicted that between 0100Z and 1200Z on August 20, 2009 a deeply occluded low pressure system would remain situated over the upper Mississippi valley, with a cold front moving south-eastward across the mid Mississippi valley towards Indiana/Michigan, and a warm front running east-west across the southern Great Lakes. Boundary layer moisture content in the warm sector below the warm front was predicted as being very high. At this point models predicted an organized mesoscale convective system spreading northeastward across northern/central Indiana and portions of Michigan and Illinois.

Shown below in Figure 1 is a synoptic analysis for 00Z August 20, 2009. At this

3 The reports from which this synoptic overview is based are obtainable from http://www.spc.noaa.gov/products/outlook/, where putting "20090820" in the date entry field at the bottom of the page retrieves all of these reports.

point, the low pressure system associated with the storm system of this study was located over southern Manitoba, but was moving south-eastward towards the upper Mississippi valley.

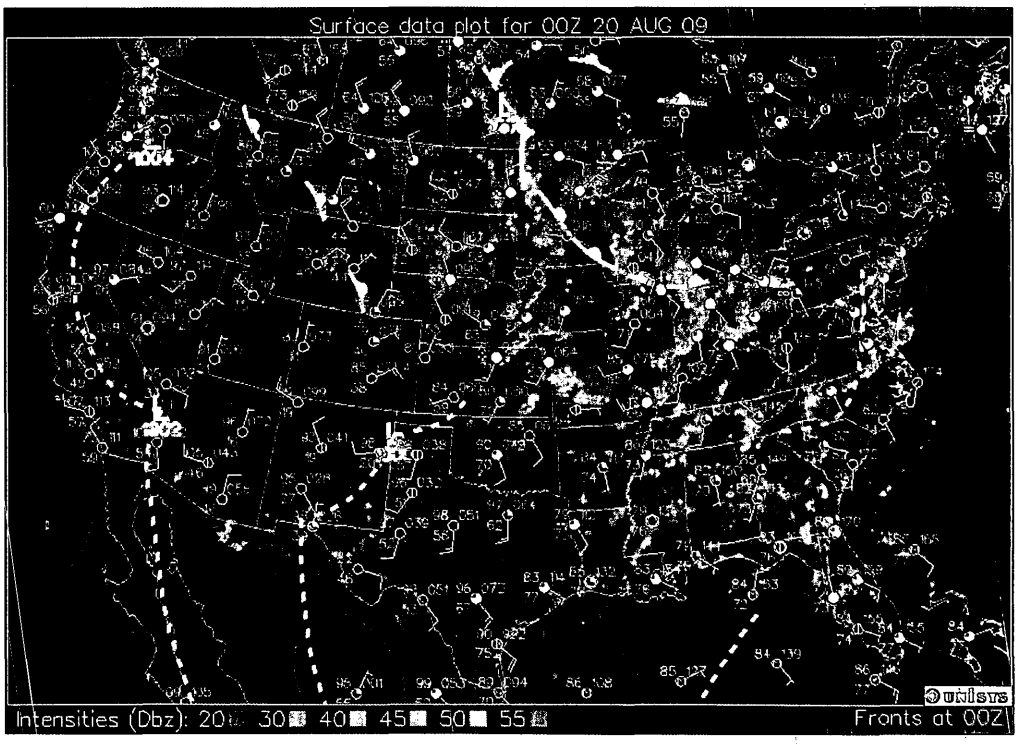


Figure 1: Surface analysis and radar reflectivity for 00Z, August 20, 2009. Sea level pressures are plotted in blue at intervals of 4hPa.

This storm system was then predicted to develop, between 1200Z August 20 and 1200Z August 21, an intensifying large scale upper-level trough located over the southern plains and running north to the Great Lakes, with a strongly sheared moist pre-frontal environment located over the middle Mississippi and lower Ohio River valleys, where it was noted that this environment would provide a potential for severe storms.

Clearing within a dry-slot across northern Indiana into lower Michigan was indicated as a possible factor in allowing for sufficient heating to allow for moderately strong destabilization. It was noted that within the warm sector the boundary layer was expected to remain very moist, and the low level hodographs were likely to be highly clockwise curved, with a large low-level shear. At this point the report indicated that this environment was supportive of the development of supercells, with the possibility of tornado generation.

Shown below in Figure 2 is the synoptic analysis for 1200Z, August 20. Note that by this time, the low pressure system previously situated over southern Manitoba has moved south-eastward over the upper Great Lakes and has become a detached low. Note also that the cold front near this low has become occluded, and runs north-south generally to the west of Lake Michigan roughly over the

Mississippi valley. Note also that a strong crescent-shaped band of reflectivity, identified by the SPC as an organized mesoscale convective system (MCS), is situated well to the east of this cold front.

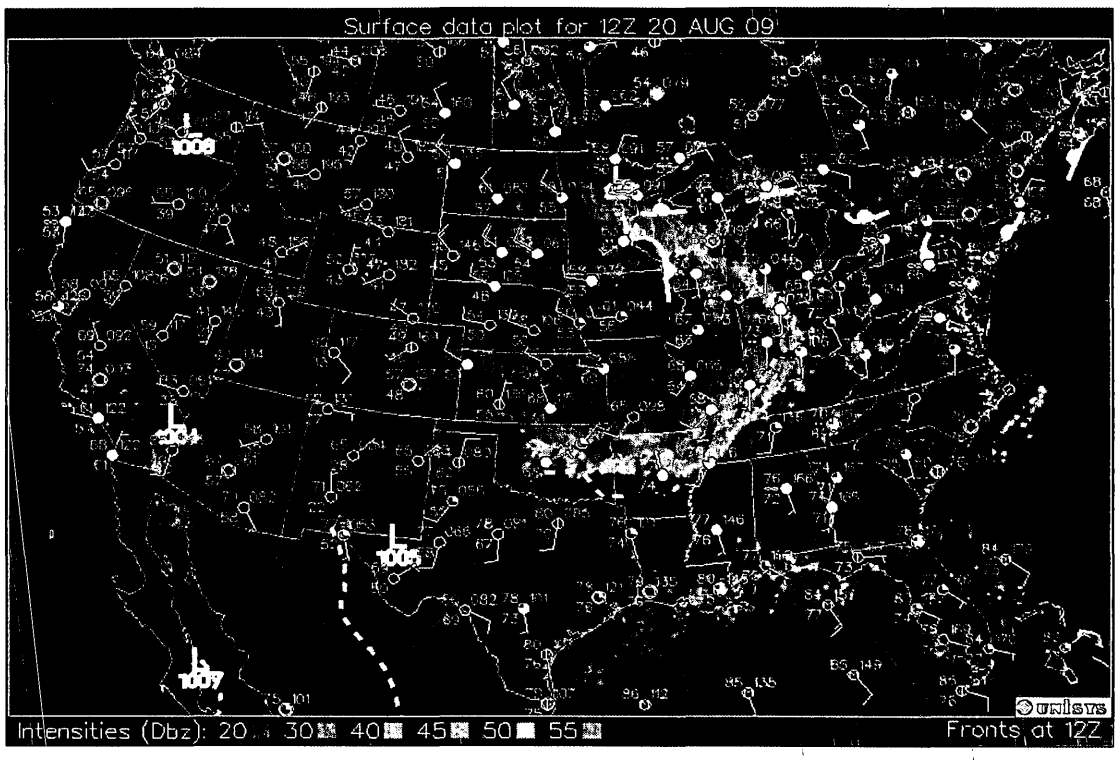
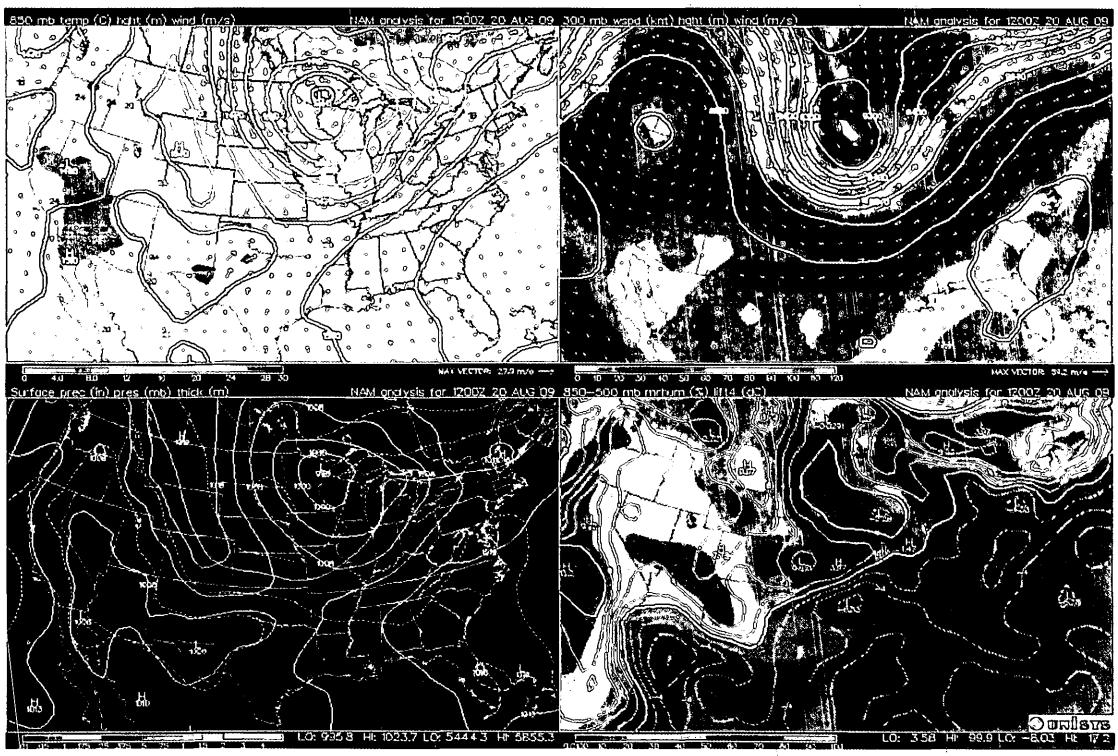


Figure 2: Same as Figure 1, but for 1200Z, August 20

Shown following in Figure 3 are output maps from a NAM model analysis at 1200Z. Notable in these maps is the presence of the developing upper level jet just to the south-east of the detached low in the 300 hPa wind speed map in the

top-right, and the high humidity of the environmental air in and to the east of this area, evident in the relative humidity plot in the lower right. Note also in the relative humidity plot the contours of the Lifted Index (LI) are also plotted. LI values of less than 0 are indicative of convective instability, and values of LI less than -4 are indicative of the possibility of severe thunderstorms[4]. While values of LI less than 0 do exist ahead of the cold front shown in Figure 1, severe thunderstorm LI values (less than -4) along the front and the pre-frontal reflectivity band exist only in north-eastern Oklahoma at this point.

4 See UNISYS upper air plot description at http://weather.unisys.com/upper_air/details.php



Top Left Panel: 850 hPa temperature (coloured shading) and 850 hPa geopotential (contours)
Top Right Panel: 300 hPa wind speed (coloured shading) and 300 hPa geopotential (contours)
Bottom Left Panel: Surface pressure (blue contours)
Bottom Right Panel: 850-500 hPa Integrated Relative Humidity (coloured shading) and Lifted Index (contours)

Figure 3: NAM analysis diagnostic plots for 1200Z, August 20

The NOAA SPC analyses predicted between 1630Z August 20 and 1200Z August 21 the presence of a trough confluence boundary, situated ahead of the cold front from central to south-eastern Michigan and north-western Ohio. It was predicted here that the combination of moderate instability with a largely unidirectional low-mid level flow of 35-50 knots, and an effective bulk shear of 35

knots, would be supportive of storms with supercell structures and persistent line segments capable of producing damaging winds and hail. Shown below in Figure 4 is the NOAA SPC analysis for 00Z, August 21. Note that the detached low has moved only slightly to sit over northern Lake Superior, and that both the cold front and the squall line ahead of the front have moved to the east. At this point, the squall line runs north-south over Lake Ontario and is situated well to the east of the Toronto area – a fact which will be of some consequence for this study and will be discussed further later in this report.

Figure 5 shows a satellite image at the same time. Clearly evident in this figure is the dry-slot mentioned above running south from Michigan through Illinois and Ohio.

Finally, the NAM Analysis plots for 00Z, August 21st are presented in Figure 6. Of significance in these plots are the strengthened upper level jet evident in the 300hPa wind speed plot at the top right, running north-northeast from Illinois up through Michigan and between Lake Huron and Lake Superior. Also of importance is the increased humidity in the southern Ontario area evident in the humidity plot at the lower right, along with the decreased LI values in this area. It is significant that at 00Z on August 21, the entire southern Ontario area has LI

values well below -4, again indicative of severe thunderstorm activity.

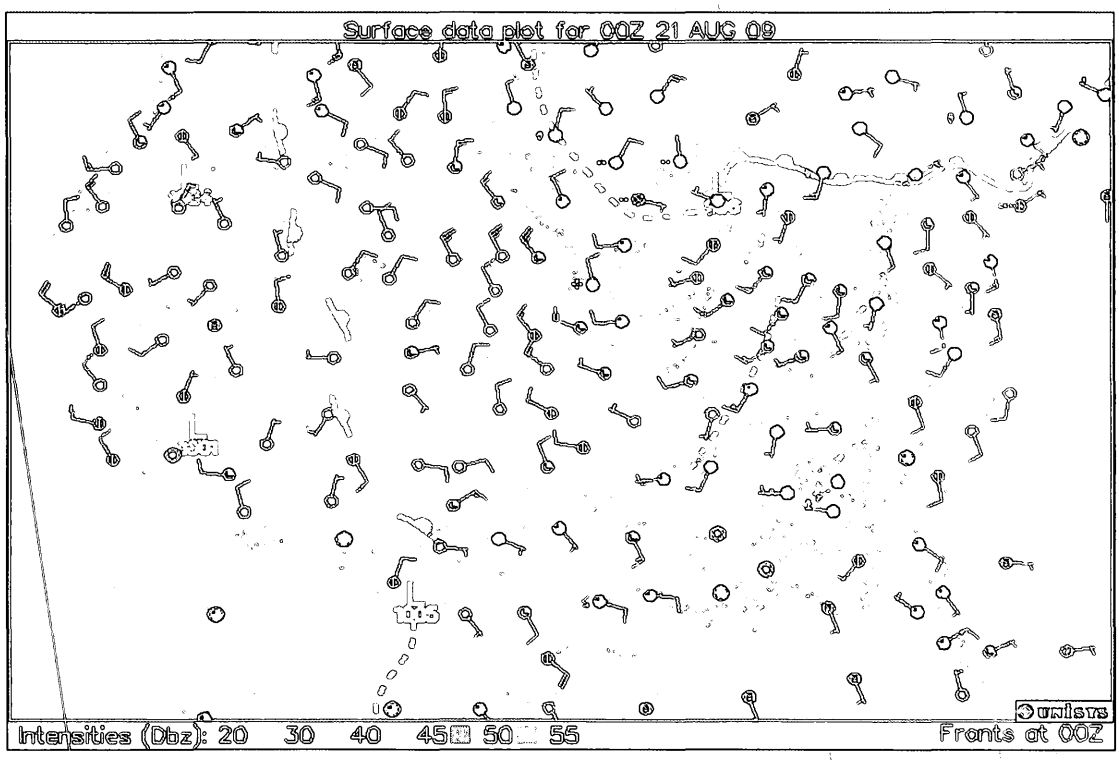


Figure 4: Same as Figure 1, but for 00Z, August 21

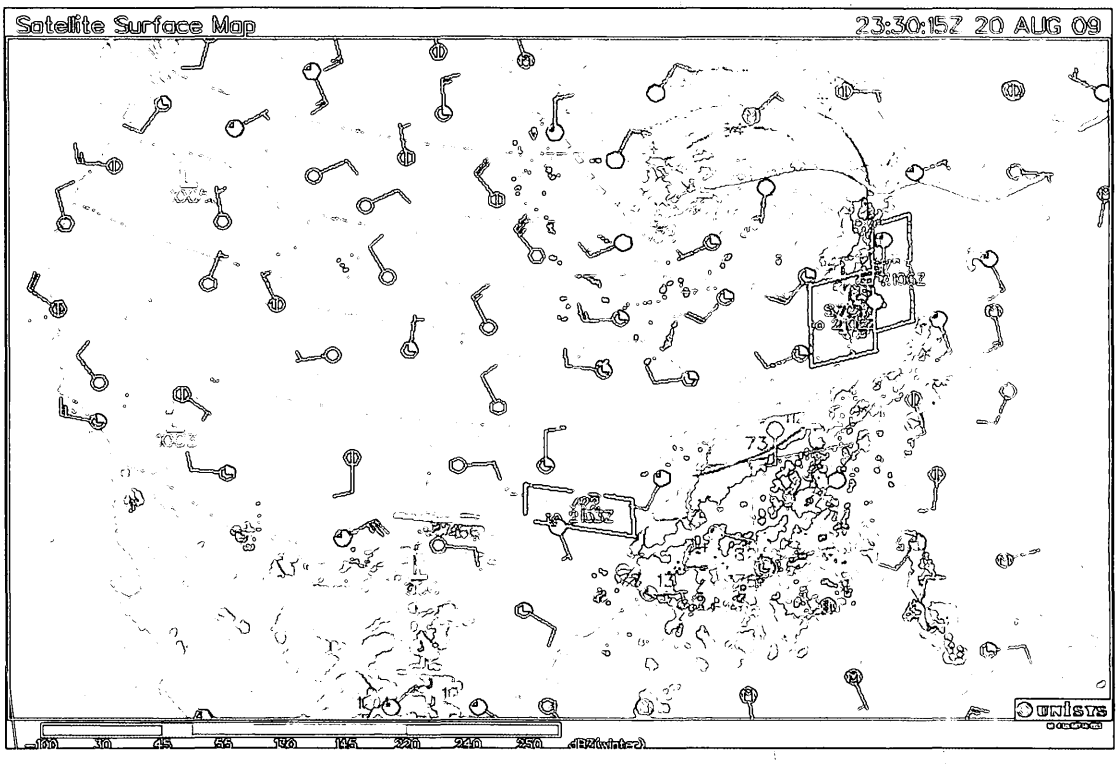


Figure 5: Satellite image and surface analysis at 00Z, August 21

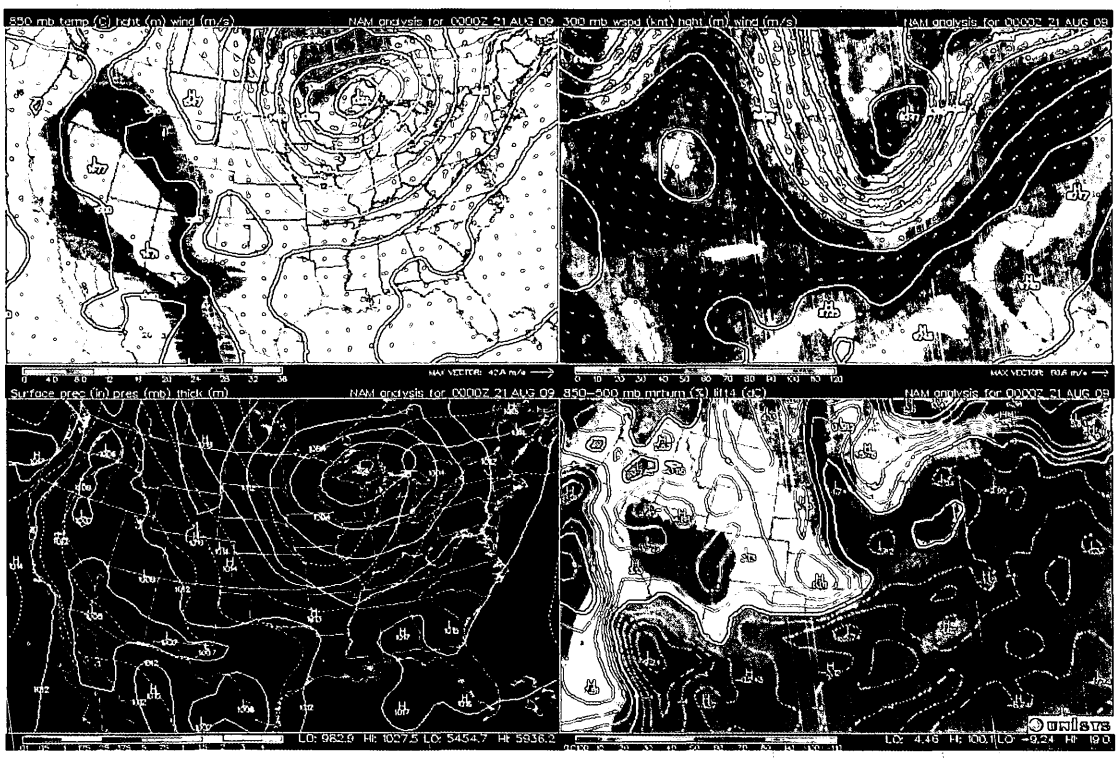


Figure 6: Same as Figure 3, but for 00Z, August 21

In summary, the synoptic analysis of the storm system featured in this study showed a squall line developing well ahead of a cold front, possibly at a trough confluence boundary. Ahead of the cold front in the warm sector, was a warm and moist air mass of high instability down to the boundary layer. As the storm system progressed, a strongly convective band/squall line formed at the confluence boundary, with the convection driven by the low level jet and its

associated high magnitude of low level wind shear. As expected, this storm system produced significant convection, and even resulted in tornadic activity. Indeed, as mentioned above, this storm produced 11 tornadoes across southern Ontario, an unusually large number for a single storm system.

1.1.2 Comparison of 1km and 3km Resolution Simulations

With the storm system's synoptic background established, we return to our main focus of this chapter – the comparison of the high and low resolution simulations of this event.

As discussed above, the primary focus of this study is to investigate how well an ensemble of moderately high resolution (3km) simulations can model an intense mesoscale[5] meteorological phenomenon such as the August 20, 2009 squall line system of this study, and compare the predictive performance of this ensemble to a single higher resolution (1km) deterministic model. But, as we also mentioned, we must first address concerns that a 3km resolution simulation (which again we shall hereafter refer to as the low resolution simulation) is capable of capturing the same physical processes that the 1km resolution simulation (the high resolution simulation) is capable of. As we mentioned, these concerns arise from issues addressed in several papers which have questioned the ability of simulations at this range of resolution to adequately model convection explicitly. We will now discuss and elaborate upon these matters below.

---

5 Note that while officially (according to the AMS), mesoscale is defined as covering features of scale from 2km to 2000km in extent, it must be remembered that in order to resolve features of size L, the grid spacing should be of size L/6 or less to resolve such features (Shamrock, 2004). Thus, our 1km resolution model will be considered a mesoscale model for this study.

1.1.2.1 Relative Abilities of 1km and 3km Resolution Simulations to Capture Physics of Supercell/Squall Line Activity

The current literature brings into question the ability of even our high resolution (1km) simulation to adequately capture the physical processes necessary for an accurate simulation of deep convective processes (Dudhia 2011, Zhang and Zhang 2011). As these authors have commented, while a 4km resolution appears adequate to resolve convection, as low as a 250m resolution might be required to fully resolve the deep convective processes of the supercells in the squall line being studied. Thus, it remains an open question whether or not all of the physical processes are being as well resolved at a 3km resolution as are resolved at a 1km resolution.

In order to address this question of whether or not the correct and same physical processes are being captured by both simulations, it will be necessary to look for certain characteristics of the squall line most relevant to a deep convective process. Particular attention should be paid to characteristics of the supercells which are believed to be constitute the squall line being studied – specifically, to the simulated vertical velocities and storm-relative helicities, since high values of these variables are associated with supercell activity (Weisman and Rotuno, 2000). Differences between the two simulations will be looked for with regard to

these variables, as it is expected that any differences between the capacities between the two simulation resolutions will manifest themselves in mesoscale processes that at least partially underlay supercell evolution.

Evidence for supercellular characteristics should be also looked for in the evolution of the storm cells over time. As Fankhauser et al (1992) have noted, it is unusual for an ordinary thunderstorm cell to last for more than 20-30 minutes, and longer lived cells are therefore symptomatic of supercells. Supercells, aside from being relatively long-lived phenomena, are also prone to cell splitting, where a single supercell will over time split into two distinct cells. Both simulations will also then be examined for evidence of this cell splitting phenomenon, as well as for evidence of structures that remain coherent for more than 1 hour. Such evidence is strongly indicative of the presence of the supercells that are believed to have spawned the 11 tornadoes that were generated by the August 20 squall line event.

In addition, both simulations should be compared with regard to their predicted rates of rainfall and their predictions of maximum radar reflectivity, as both of these variables are of course indicative of the precipitation that accompanies squall lines. In particular, attention will be paid to the degree of continuity of the

simulated squall lines, and to the locations of the squall lines. Total rainfall will also be a factor to consider, as there is some question, as noted by Dudhia (2011), as to how well NWP simulations are able to predict rainfall rates during convective activity with resolutions in the 1km – 3km range.

Following this, to get a complete picture of the modeled dynamics, a center of particularly intense convection will be chosen from an area in both models which shows a high value of vertical velocity, and zonal cross-sections will be taken through this point. We will look for evidence of environmental air which is predisposed to convective activity, which is most easily understood in terms of Convective Available Potential Energy (CAPE) – a measure of the amount of potential energy held in the air typically necessary to fuel convection. Before we can have convection however, it is necessary first to lift the air by some mechanism past the the level of free convection (LFC) where the air is free to convect. The vertical motion of the air before we have reached this point is inhibited by the environmental Convective Inhibition (CIN), in a sense the total energy necessary to overcome to initiate convection. After reaching the LFC however, convection is free to commence[6]. However, since the squall line itself, based on the synoptic analysis alluded to above appears not to have formed at a

6 Note that this is of course a purely thermodynamically driven convective mechanism – we will see later that the convective activity of the supercells has also a strong dynamically driven mechanism as well. Nevertheless, thermodynamic mechanisms will be important for supercell convection as well.

front, but instead appears to be pre-frontal, then evidence for some lifting mechanism other than frontal lifting will have found from the simulation results, perhaps from some low level convergence condition, or from the lifting that would be generated at low levels by the gust front, or the tilting effect which we will discuss in more detail below. In any event, if lifting characteristics are similar in both models, then further evidence will be available that the same driving dynamics are being captured in both simulations. There will therefore be some justification in comparing the results of an ensemble of low resolution simulations to a single deterministic simulation, since the same physics are being captured by both.

1.1.2.2 Storm Relative Helicity

Another important parameter to consider when dealing with supercells is Storm Relative Helicity (SRH), first introduced by Davies-Jones (1990). This parameter measures the degree to which there is a horizontal, streamwise component of vorticity which matches up with the direction of horizontal flow relative to the storm's motion, and is given by the following formula:

$$H = -\int_0^h \hat{k}\cdot\left(\left(\vec{V}-\vec{c}\right)\times\frac{d\vec{V}}{dz}\right)dz \qquad (1)$$

where V is the velocity of the wind, c is the direction of storm motion, and h is the top height to which this quantity is integrated to – typically taken to be 3000m. It can be interpreted as a parameter equal to the area swept out between 0 and h on the hodograph by a vector with one end held at c, in other words, the area swept out by the hodograph relative to the storm's velocity vector. Davies-Jones (1990) found this parameter to be a promising tool for the prediction of the tornadic potential of a supercell storm. Having high values of this parameter then would add further evidence that the simulations are indeed successful. We will therefore present an analysis of the simulations' predictions of this parameter below.

1.1.2.3 Supercell Evolution

Finally, some mention should be made regarding the evolution of the supercells. Supercells are of course a class of deeply-convective mesoscale structures that are strongly associated with the presence of high vorticity and strong updrafts, and which are known to be responsible for generating the majority of tornadoes.

They have several features that set them apart from other mesoscale storm structures associated with deep convection, such as single cells and multicell storm systems. For instance, as we have alluded to above, unlike single cell and multicell storm structures that obtain their convective energy by lifting warm, moist and high CAPE air above the level of free convection (LFC) as the storm's gust front is pushed underneath the moist air at the storm's leading edge, supercells are not generated by the interaction of a gust front with high CAPE air. Rather, the lift of this air results as a dynamic process involving the 'tilting' of highly sheared low level air by updrafts in the system.

Wallace (2006) offers a useful description of this process. He states that what is necessary first is a highly sheared environment, not necessarily exhibiting any turning of the wind with height (veering or backing). Assume then that we have

an isolated deep convective updraft develop for some reason, perhaps due to a more typical convective storm. As this updraft is lifted upward, it carries with it the horizontally sheared air, tilting it so that vertical vorticity is induced into the flow as a vortex couplet, as can be seen in part a) of Figure 7. On either side of the updraft, these counter-rotating vortexes induce negative pressure perturbations at mid levels, which not only intensify the upward pressure gradients at the base of the updraft and thereby intensify it, but as well cause the single updraft to split into two halves which are then each be pulled towards the vortexes – effectively splitting the single storm into two halves, each of which will thereafter generate its own vorticity couplet. The net effect of this sequence is the gradual evolution and splitting of individual supercells over time. This effect is shown in part b) of Figure 7. Note that while a downdraft is shown in part b) and is very often seen associated with such storm systems, Wallace (2006) states that the downdraft itself is not responsible for this splitting, and indeed splitting is observed in numerical simulations even when the microphysical processes that are responsible for downdrafts are removed. This splitting phenomenon seems to be largely a result of the creation of the counter-rotating vortexes by the updrafts and is purely a dynamic phenomenon not associated with subscale or thermodynamic processes in the system. Thus, one should expect splitting in the presence of strong updrafts and vorticity in storms, even in the absence of storm-

generated downdrafts. Further, the vortexes are not only responsible for splitting the supercell, but they are also responsible for generating the pressure perturbations which in turn further lift moderate or high CAPE air at the storm front. Thus, a supercell can be seen to be a self-propagating system requiring only the presence of low-level, highly-sheared air and moderate CAPE environment for its evolution.

Note that other interactions are important to this process. The process discussed above, without any veering or backing with height (with a straight hodograph), will generate supercell pair members of equal strength moving to the right and left of the mid level steering flow[7]. As Markowski and Richardson (2011) and Davies-Jones (1984) have noted, with veering of the wind with height (as is common with supercells), streamwise advection of the vortex couplet will push the right half of the vortex couplet into the updraft, and the left half out. As they also note, it is the correlation of the vorticity and updrafts in the supercell that is responsible for strengthening the supercells and extending their lives (up to 4 hours), giving them much longer lives than other mesoscale storms. Thus the right half supercell will strengthen in veering, highly sheared air, and the left half will weaken. This process will favour the generation of right-half supercells that will

7 Note however that with any such lack of veering or backing, the resulting hodograph will be uncurved and thus the SRH will necessarily be zero – indicating a lack of tornadic-generating potential.

tend to move to the right of the mid-level steering flow of the storm system. Thus, right-moving supercells will tend to be associated with veering, highly-sheared environments – environments where the hodographs are long and curved concave downwards, especially in the lower levels.

It has also been noted that CAPE values need not be especially high for supercell production as suggested in the above description - indeed Finley (1988) notes that supercells have been reported with CAPE values as low as 800 J/Kg. Of course, this is a result of the fact that supercells gain a large portion of their energy from the production of vertical vorticity through the tilting of strong low level environmental shear. Their updrafts are produced as a result of the dynamics of the pressure gradients resultant from this interaction, so high CAPE values are not a necessity to contribute the thermodynamic energy to drive these updrafts.

It should also be noted that there are many more processes at work in the evolution of a supercell than are outlined in the above description, which is admittedly highly simplified and involves only a prediction of the mesocyclone from the larger scale dynamics of the environment. Many of these processes that have not been covered are thermodynamic in nature, which naturally involve interactions at a significantly smaller scale than that of the supercell as a whole,

and are associated with such processes as evaporative cooling and downdraft generation. Given that the mesocyclones of supercells are typically sized on the order of 3-8 km in diameter (Markowski and Richardson, 2011), we would expect these sub-storm-scale processes to be below the scale of both the high and low resolution models to resolve adequately. Thus for instance, given that the downdrafts of a storm result from the evaporative cooling and descent of air within the storm (Finley, 1998), and that these downdrafts naturally result from processes occurring at scales smaller than are resolvable by our simulations, we should not be surprised if our simulations fail to predict these downdrafts. In addition, given that tornadoes are thought to be the result of the descent to ground level of mid-level mesocyclone vorticity by these downdrafts (Finley, 1998), we should also not expect our simulations to be capable of explicitly modeling tornadoes either. Our mesocyclone, driven dynamically by properties of the large scale environment, should however be resolvable, and thus we should expect to see both vertical vorticity in our simulations, as well as the updrafts resultant from the vertical pressure perturbations produced by this vorticity.

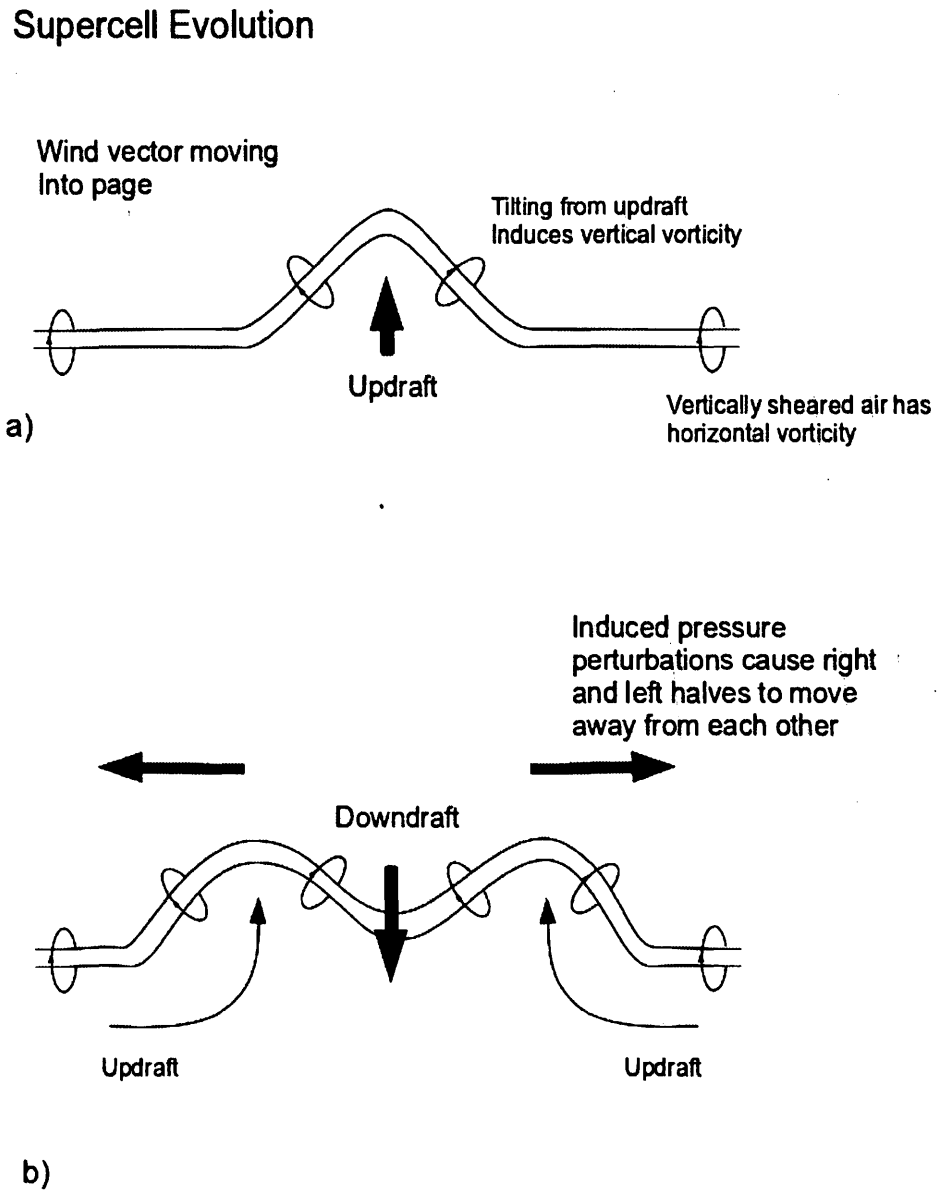


Figure 7: Evolution and splitting of a supercell storm

1.2 Methodology

Again, as outlined above, in order to address the question regarding whether or not a 3km resolution simulation in indeed able to capture the same physical processes as a 1km simulation, two experiments were carried out at resolutions of 3km and 1km. We will begin with a discussion of the model set-up for these experiments.

1.2.1 Physics Options and Model Domain Configuration

Both experiments, and indeed all of the experiments described in this study, were modeled using the latest version (3.3.1) of the Weather Research and Forecasting (WRF) model[8]. WRF is a fully compressible non-hydrostatic mesoscale model with a full suite of physics options (Shamrock et al., 2005). The particular options used for the experiments were shown in Table 1. Note that the control files used for the simulations, the 'namelist.input' files that WRF uses to control program execution, are both listed in Appendix A.

8 WRF was developed jointly by the National Center of Atmospheric Research (NCAR), the National Oceanographic and Atmospheric Administration (NOAA), the National Center for Environmental Prediction (NCEP), the Forecast Systems Laboratory (FSL), the Air Force Weather Agency (AFWA), the Naval Research Laboratory, and University of Oklahoma, and the Federal Aviation Administration.

All simulations utilized nest grids of increasing resolution. The 1km resolution simulations used 4 nested grids of horizontal resolution (in the x and y directions) of 27km, 9km, 3km and 1km, for the 1st, 2nd, 3rd and 4th domains respectively, while the 3km resolution simulation used the 1st, 2nd and 3rd domains of the 1km simulation, without the final 4th domain. The extents of these domains are showing in Figure 8 below.

It is important to note that in all of the simulations involved in this study, that two-way communication was utilized between the domains. What this means is that, not only are the higher resolution nested domains obtaining their initial and boundary conditions from their parent domains during the simulation (so for example, the 3rd domain would obtain its initial and boundary conditions from the 2nd domain nodes), but also the lower resolution parent domain will have its nodal values updated during the simulation using a weighted average of the corresponding adjacent higher resolution nodes in the child domain. In this way, the parent domain values gain from the increased resolution and accuracy of their child domains, and are updated due to this interaction.

Physics Scheme Type	Physics Scheme Name
Cloud Microphysics	Morrison Double-Moment Scheme
Surface Layer Physics	Monin-Obukhov Scheme
Land Surface Physics	Noah Land Surface Scheme
Planetary Boundary Layer Physics	Yonsei University Scheme
Cumulus Physics (only in domains 1 and 2)	Kain-Fritsch Scheme
Longwave Radiation Physics	Rapid Radiation Model Scheme
Shortwave Radiation Physics	Dudhia Scheme

Table 1: Physics schemes used in WRF simulations

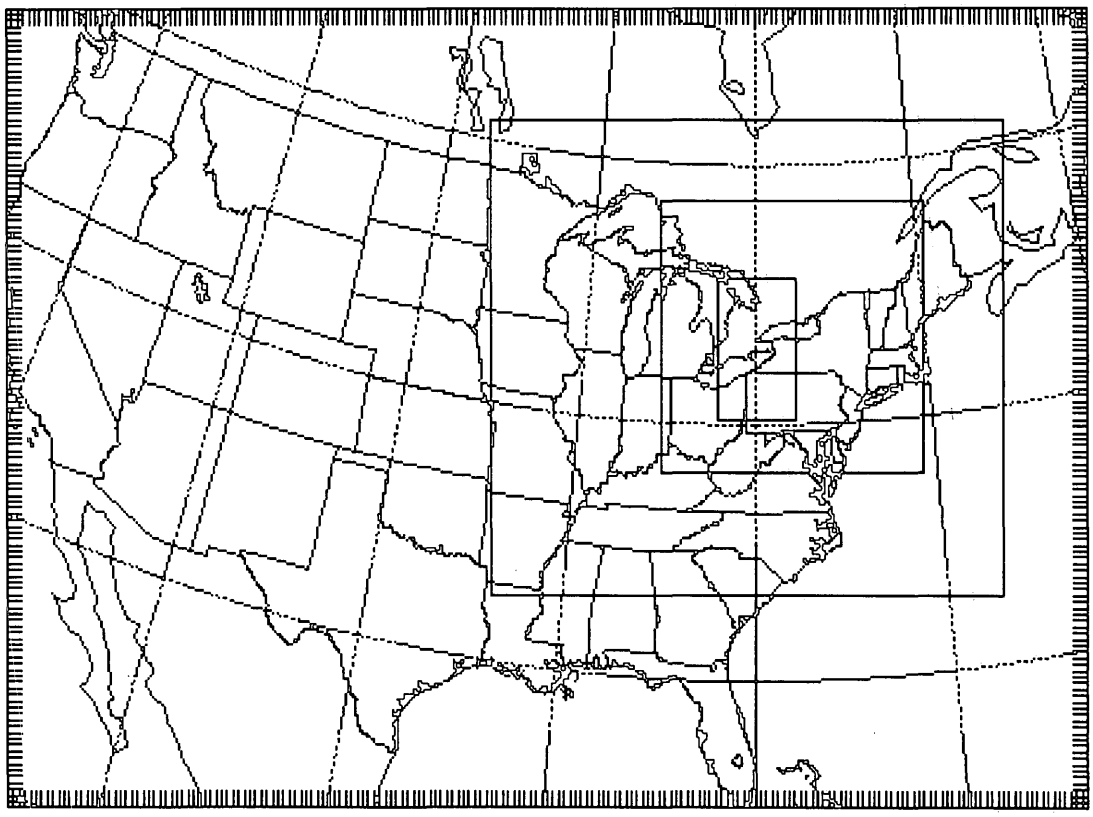


Figure 8: Domains used in WRF simulations

1.2.2 Initialization and Boundary Condition Data

The data used for the initial conditions and boundary conditions for the high and low resolution experiments was NCEP's GFS FNL Analysis. This data is available every 6 hours at a 1 degree by 1 degree resolution starting from 00Z, so data was obtained for August 20, 2009 0000Z, as well as 0600Z, 1200Z, 1800Z and August 21, 2009 0000Z. Note that for clarity, we will hereafter refer to the August 21, 2009 0000Z time as 2400Z.

The data was preprocessed using the WRF's WPS utilities, which both horizontally and vertically interpolate the data to the grid points specified by the user. For the present study, the horizontal grid points were specified as discussed above and as shown in Figure 8. WRF uses a terrain-following vertical coordinate called an η level, defined as follows:

$$\eta = \frac{(p - p_{top})}{(p_{bottom} - p_{top})} \qquad (2)$$

where p is the hydrostatic pressure, p_{bottom} is the pressure at the bottom of the air column (at ground level), and p_{top} is the pressure at the top of the modeled air

column. We chose 27 η levels, with the highest resolution of 60m near the ground and gradually stretching to 2000m aloft.

The simulations were configured to output model variables at 1 hour intervals, in a standard NetCDF format. NCAR's NCL language, which comes with a wide array of processing and plotting functions for atmospheric model data, was used to interpret these results. The results that were drawn from this processed data will be presented below.

1.3 Results and Discussion

1.3.1 Relative Capability of Capturing Squall Line Storm Dynamics

The primary goal of the first phase of this research project was as mentioned above, to establish that the low resolution 3km simulation resolves the same physics as the high resolution 1km simulation. Particularly, for the area where the dynamics are the most 'nonlinear ' in character – that is along the squall line - we are interested in the relative ability of the low resolution simulation to capture the correct evolution of the squall line as compared to the high resolution simulation.

We will begin the investigation of this question instead with an examination of the modeled distribution of radar reflectivity for both simulations, since radar reflectivity shows areas for which convection is the most intense, and it is these areas of intense convection that are the most challenging phenomena for a numerical model to simulate. Any significant deviations of the low resolution simulation's reflectivity from the observed reflectivity that were correctly simulated by the high resolution simulation, would be indicative of a failure of the low resolution simulation to resolve some of the physics that were being successfully resolved by the high resolution model. Thus we will begin first with this

examination of our models at both resolutions.

1.3.1.1 Simulation Radar Reflectivity

Shown below in Figures 9 and 10 are plots of the modeled compositie radar reflectivity for both the high and low resolution simulations. Plotted are the modeled reflectivities for the 3rd domain for both the high and the low resolution simulations. Note that while the high resolution simulation's highest resolution domain was in fact the 4th domain, for the two-way domain communication chosen for this study which we discussed above, values from the 3rd domain were obtained by an averaging process using adjacent 4th domain nodal values. Thus, the high resolution's 3rd domain values should reflect the more highly resolved physics of the 4th domain, and we should not expect significant differences between the two.

We can see from these figures that the high and low resolution models showed similar patterns of maximum radar reflectivity – a line of high radar reflectivity running generally north-south situated just west of the western end of Lake Ontario connected in the north to a more diffuse broad band of high reflectivity

running west-east just north of the Great Lakes.

The peak radar reflectivities are similar for both simulations – the high resolution storm predicted peak reflectivity values of 56.5 dBZ at 2400Z, while the low resolution storm predicted peak reflectivity values of 58.7 dBZ.

Figure 11 is an image showing the observed radar reflectivity pattern at 23:58Z. It can be noted from the observed radar plot that peak reflectivities are of a similar magnitude as the high and low resolution simulations – between 55 and 60 dBZ. Therefore, at least in terms of observed peak magnitudes, the simulations can be seen to be quite successful.

We can see clearly however from the observed reflectivity figure that there is a significant displacement of the observed squall line compared to the modeled squall line – with the observed squall line positioned approximately 80km to the east of the modeled squall lines (both squall lines delineated according to their peak values of maximum dBZ). It is important to note that although both the high and low resolution simulations showed this significant bias to the west in their positions as compared to the observed squall line, they showed a very similar position as compared to each other. This will be significant, for again, one of the

primary goals of the first phase of this study has been to establish that the low resolution 3km simulation is capable of modeling the same underlying processes as the high resolution 1km simulation. Clearly, for whatever physical processes were responsible for the phase error observed for the low and high resolution simulations, it is evident that their modeling was no more successful in the high resolution simulation than in the low resolution simulation.

One characteristic of the squall line however that was captured by the high resolution simulation which was not as well captured by the low resolution simulation was the degree of structure of the peak reflectivity areas. For the high resolution simulation, the structure of these areas is more finely delineated, with smaller and more contiguous features along the line than the lower resolution simulation – which again showed similar peak values of reflectivity, but in larger, and more dispersed 'cells' along the simulated squall line.

It is interesting to note that the greatest differences between the two simulations were in the areas of highest reflectivity – along the squall line and along the comma-shaped head to the north. There is however quite a lot of similarity between the two simulations along the eastern coast of the United States where reflectivity values are lower.

Thus, from these last two observations – the first observation being the similarity of observed bias of squall line position between the two simulations, and the second observation being the differences in the delineation of structural detail in the two simulations along the squall line – we can hypothesize that the differences in results appear to be dependent on the predictability of multiple length scales – the 3km simulation appears to be as successful as the 1km simulation in capturing the coarser grained physics driving squall line position – possibly because this is driven by an error in the large scale initial flow condition of our simulations – while the 3km simulation does not appear to be as successful at capturing the smaller scale structure of the individual cells along the squall line – because the processes responsible for such details are driven by smaller scale mesoscale phenomena, more successfully modeled by the finer-grained resolution of the 1km simulation. It must be reiterated however that the differences between the two model resolutions were evident only in the degree of structure or delineation of features in the squall line – values of variables such as maximum vertical velocity or maximum radar reflectivity were very much the same between the two simulations. This is the issue of scale mentioned earlier – the low resolution simulation appears to capture large scale features as well as high resolution simulation: squall line position for example, but not small scale

features, such as fineness of detail of structures in the squall line.

Again, we see that while the coarse, large scale structure – squall line position – is captured by the low resolution simulation as well as it is by the high resolution simulation, fine detail of the structure is modeled more readily by the high resolution simulation. It is important to note that specific detail of the higher resolution simulation does not necessarily correspond to an ability to model with higher accuracy actually occurring structures in the squall line.

This brings us to a central point of contention of this study. There appears to be more than one scale at work here. There is a large scale, more 'linear' flow driving the system as a whole, which explains why the low resolution simulation works as well as the high resolution simulation. This large scale of flow is much larger than either the low or high simulation resolution scales, thus they work equally as well modeling this large scale flow since they are both much smaller than the scale of this flow and thus able to fully resolve it. There is also however, the squall line convective activity which is much more unstable (and hence sensitive to feedback and initial conditions) for which the scales are much smaller. Thus with this smaller scale instability, the extra degrees of freedom offered by the increased resolution of the high resolution simulation will offer

more opportunities along the squall line for the instabilities to manifest themselves. The convective cells will thereafter form more readily and in smaller areas, resulting in finer structure for the high resolution simulation. We will return again to this issue of scale in our ensemble analysis later in this study.

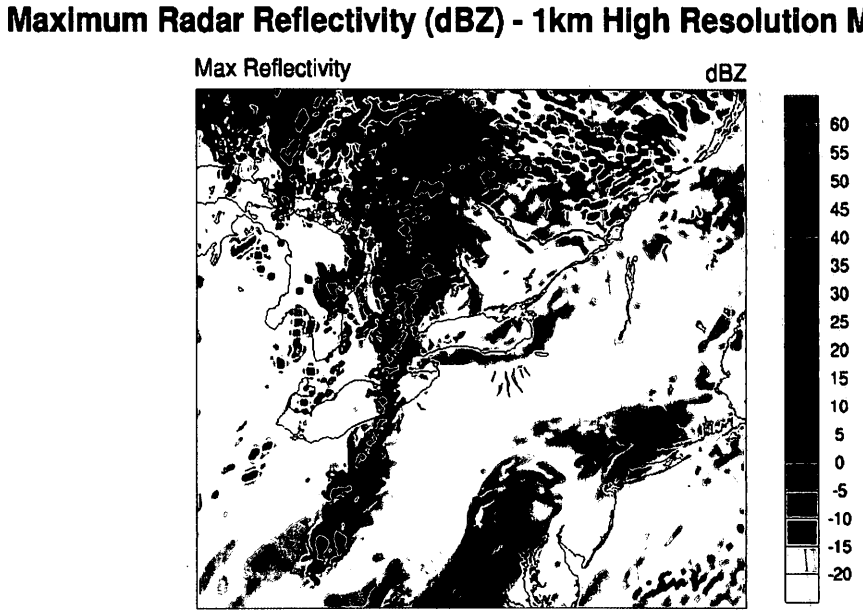


Figure 9: Composite radar reflectivity in high resolution simulation at 2400Z

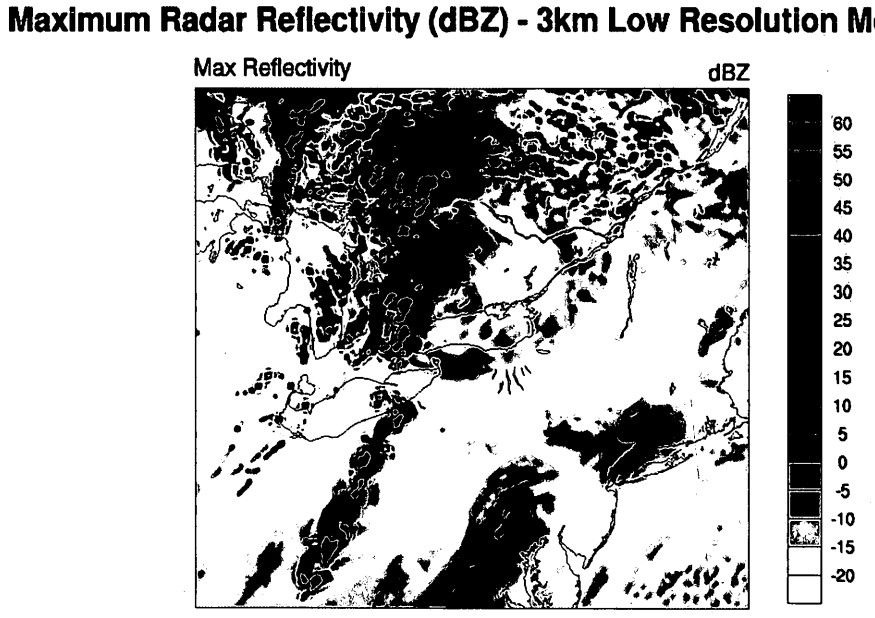


Figure 10: Composite radar reflectivity in low resolution simulation at 2400Z

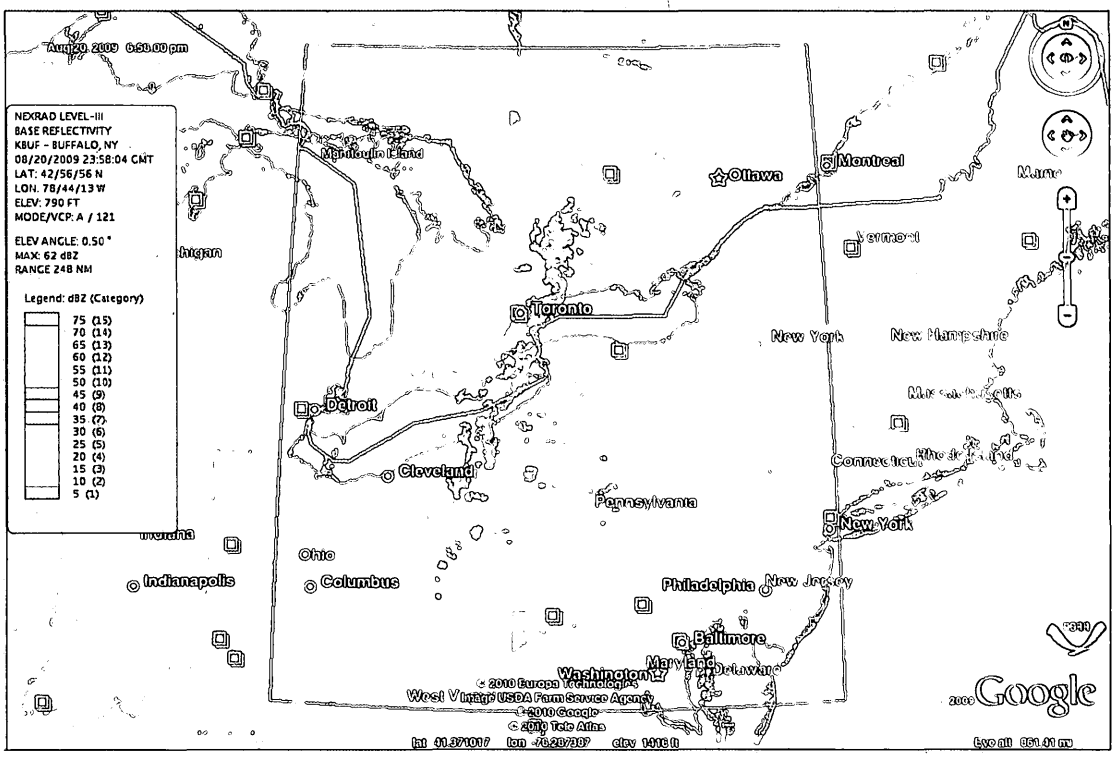


Figure 11: Base reflectivity from Buffalo NEXRAD radar at 2358Z

1.3.1.2 Storm Relative Helicity

Next to be considered is Storm Relative Helicity (SRH), which as discussed above, is a characteristic of the storm's environment that indicates a potential for the development of the tornadoes which were reported for the August 20, 2009 storm studied in this report.

Shown below in Figures 12 and 13 are plots of the 0-3km storm relative helicity (SRH) for the 3rd domain of both the high and low resolution simulations respectively.

We can see a similar pattern to that noted above for the radar reflectivity. Both simulations showed two regions of high helicity in roughly the same area – one just south-west of Lake Simcoe and the other on the eastern end of Lake Nippissing near North Bay. For the high resolution run, the area of high SRH southwest of Lake Simcoe ran in a line from Vaughn, Ontario upward to just south of Barrie and was broken into three smaller 'cells' of high helicity, while the 1km simulation showed only a single, larger 'cell' with high helicity. Thus again, we see that the areas of peak intensity are in roughly the same positions for the two simulations, but with more finely structured detail in the high resolution

simulation than the low resolution simulation.

Peak values of helicity predicted by the models at 2400Z were 1540.6 m^2/s^2 and 1909.7 m^2/s^2 for the high and low resolution models respectively. It should be noted that, as evident from Figures 12 and 13, that these values of high helicity only occurred in very localized regions, and that the general helicity values for the area were somewhat smaller than these values, just in excess of 800 m^2/s^2 in a larger area over the GTA for both simulations. Nevertheless, these helicity values are quite large, and indicative of an unusual set of circumstances.

Significantly, the two areas of high helicity for both simulations – just south-west of Lake Simcoe and at the eastern end of Lake Nippissing near North Bay, both reported tornadoes around the same time that these two figures correspond to (2400Z). Thus, there is a good chance that both simulations could have correctly predicted the tornadoes in these areas given sufficient grid resolution, especially since SRH is often noted as a good predictor for the supercell tornadic potential.

1.3.1.3 Vertical Velocity

Highly convective storms will also show areas of high vertical velocity. A successful model will therefore be able to be able to generate such high vertical velocities, and be able to position them correctly. In the following Figures (14 and 15), we show plots of vertical velocities at 2400Z for both the high and low resolution simulations.

We see from these figures that the areas of highest vertical velocity fall along the simulated squall line - the line running roughly north-south just west of the western end of Lake Ontario. Once again, the low resolution simulation places this squall line at roughly the same position as the high resolution simulation, but with somewhat less finely discriminated structure.

Peak vertical velocities at these levels were 16.64 and 19.10 m/s for the high and low resolution models respectively.

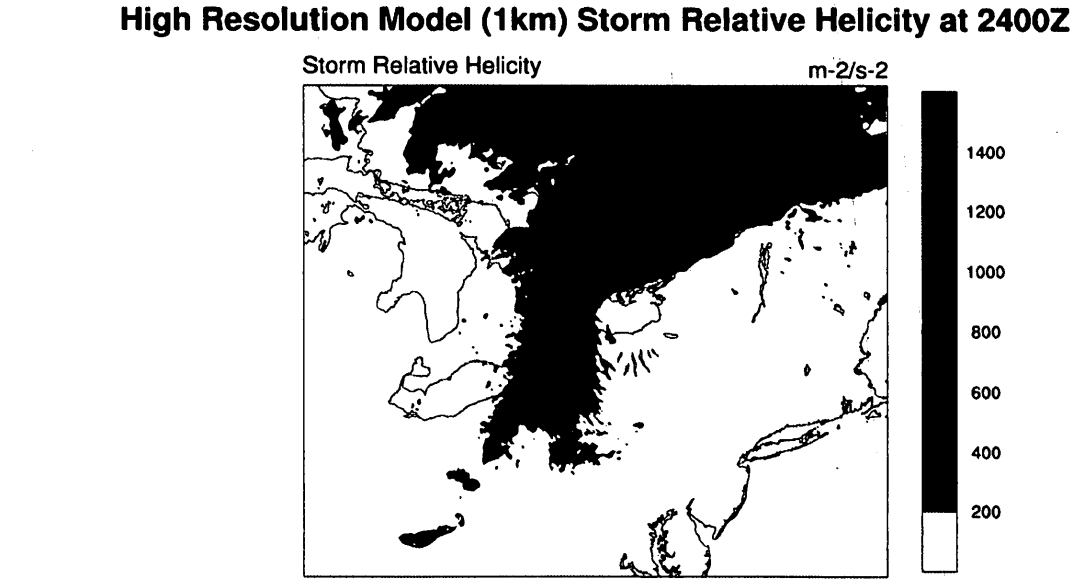


Figure 12: 0-3km storm relative helicity for high resolution simulation

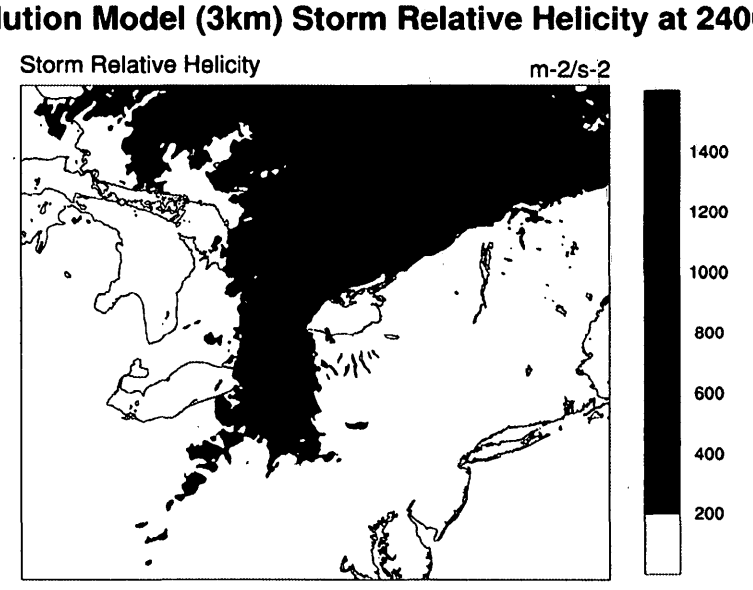


Figure 13: 0-3km storm relative helicity for low resolution simulation

Vertical Velocity at Eta=0.5 - 1km High Resolution Model

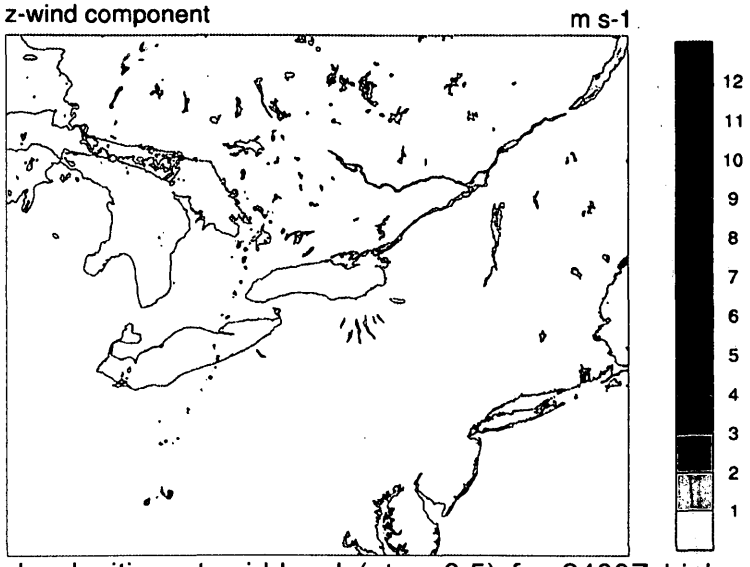


Figure 14: Vertical velocities at mid-level (eta =0.5) for 2400Z high resolution simulation

Vertical Velocity at Level 13 (Eta 0.5) - 3km Low Resolution Model

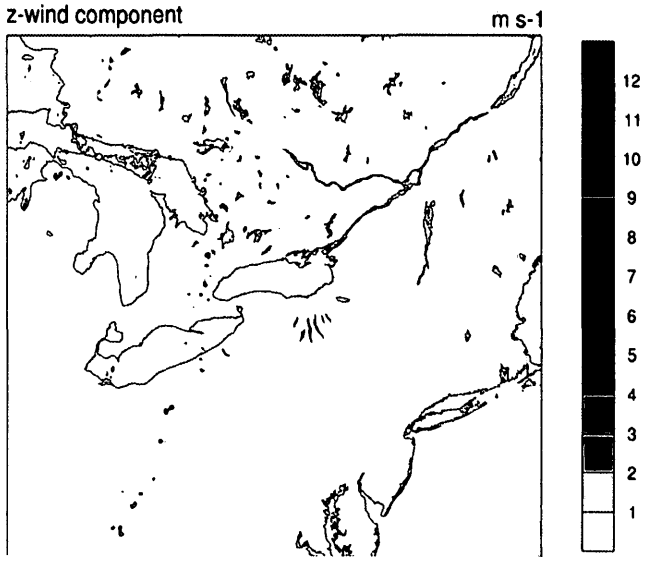


Figure 15: Vertical velocities at mid-level (eta=0.5) for 2400Z low resolution simulation

1.3.1.4 Longevity of Simulated Supercells

We will consider now the issue of longevity of the storm-cells simulated. As we noted in our Introduction, supercells are a relatively long-lived phenomenon. As we can see in Figures 16 and 17, which shows the progression of storm cells through the movement of mid level (level 13 where eta=0.5) vertical velocities from 1600Z to 1900Z for the high and low resolution simulations, the storm cells modeled are definitely long-lived phenomena. As stated in the Introduction, regular convective storm cells rarely live longer than 20-30 minutes, but supercells can live for several hours. As can be seen in the tracing of a single supercell over the 1600Z to 1900Z period (circled in the figures), this supercell stayed as a coherent structure for at least 3 hours. As well, we can see clear evidence of the splitting phenomenon mentioned in the Introduction, adding further evidence for the presence of supercells in our simulation.

Thus, based on the similar longevity of modeled storms, one could conclude that the 3km resolution simulation is indeed as capable of capturing the same physical processes contributing to storm longevity as the 1km simulation.

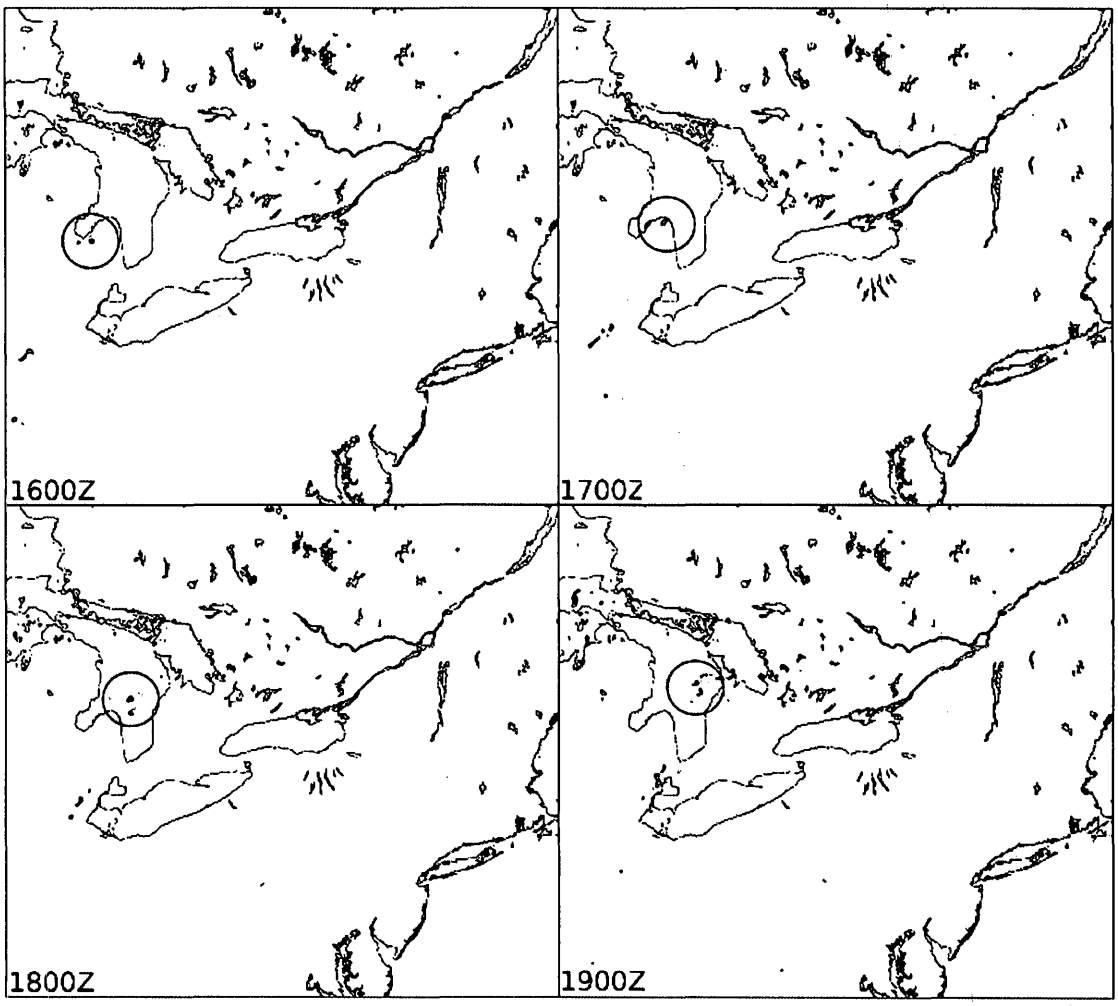


Figure 16: High resolution simulation mid-level (eta=0.5) vertical velocities

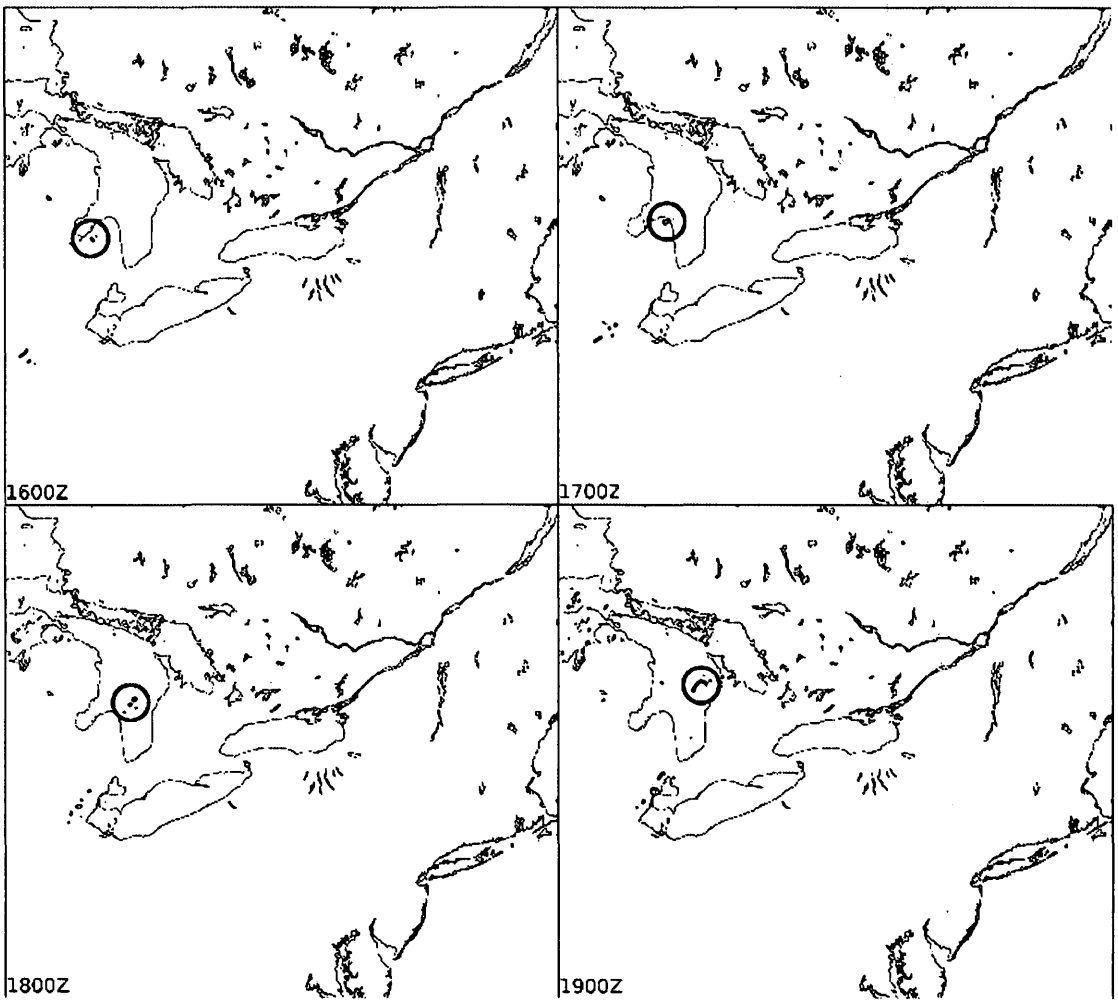


Figure 17: Low resolution simulation mid-level (eta=0.5) vertical velocities

1.3.2 A Detailed Analysis of One of the High Helicity Areas of the Simulations

In order to get a better understanding of the modeled processes at the two resolutions, we will look at a comparison of the two simulations from another perspective – from that of a vertical cross-sectional view through one of the areas of most significant convection – specifically, sections which run through the point 44.19° N, 79.71° W running along the 44.19° N latitude. Consider first the meridional and zonal wind sections for the high resolution simulation in figures 18, 19 and 20.

We can clearly see the jet running north at 10km altitude to the left of the meridional section in Figure 18, but quite significantly, there is also a lower level jet at approximately 79.7° W to 80.0° W also running north at around 2km height, which would therefore mean a high low level shear at this location just above ground level. Looking at the zonal winds in Figure 20, just to the east of the 79.7° W location there exists a low level band of air moving counter (east to west) to the main flow of the storm (west to east) just above ground level. This would result in a low level convergence at the same location, which could act as a lifting mechanism for the flow. This lifting mechanism would induce strong vorticity into the flow through the tilting of the highly sheared low level air

mentioned above.

We can see evidence for this low level shear in Figure 21, which shows a Skew-T diagram for this location. Also, we can see evidence for a moderately high CAPE for this location – 714 J/kg. Given that this moist air mass of moderate CAPE would be unstable in the presence of any lifting mechanism, and given that the low level convergence discussed above would act as just such a lifting mechanism, and that significant vorticity would be induced into the air mass by tilting of the highly sheared air by this lifting, we can understand this low level convergence would act as a self-perpetuating mechanism driving the formation of the supercells along the squall line. Lifted, high shear air would induce significant vorticity into the flow, driving the formation of supercells which, through the mechanism described in our Introduction would impart a strong vertical pressure gradient that would intensify lifting, pulling in adjacent low level moist air and resulting in further lifting immediately adjacent to the supercell, beginning the process again. It is important to note that this mechanism would only be mesoscale in extent. Thus, at least for the high resolution simulation, the lifting mechanism is clearly a mesoscale process.

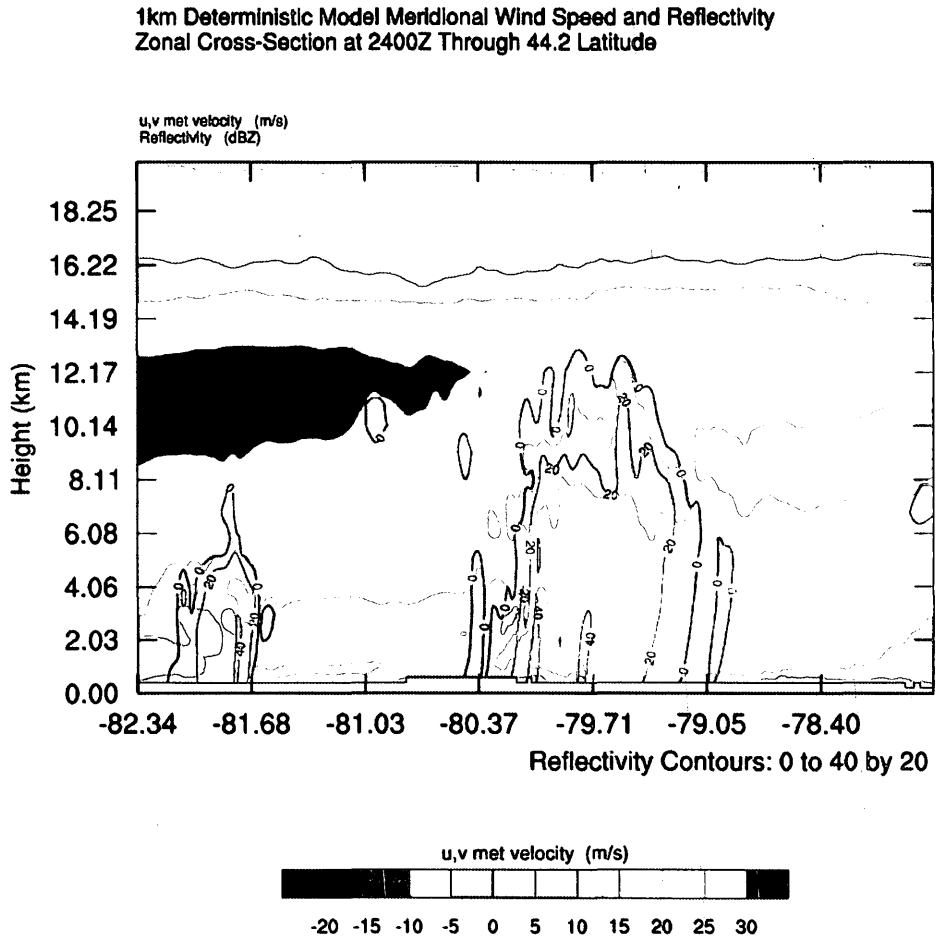


Figure 18: Meridional winds (color shaded) for the high resolution simulation through a vertical-zonal cross section crossing 44.19 N, 79.71 W at 2400Z. Simulated radar reflectivity (units: dBZ) is in contours at intervals of 20 dBz.

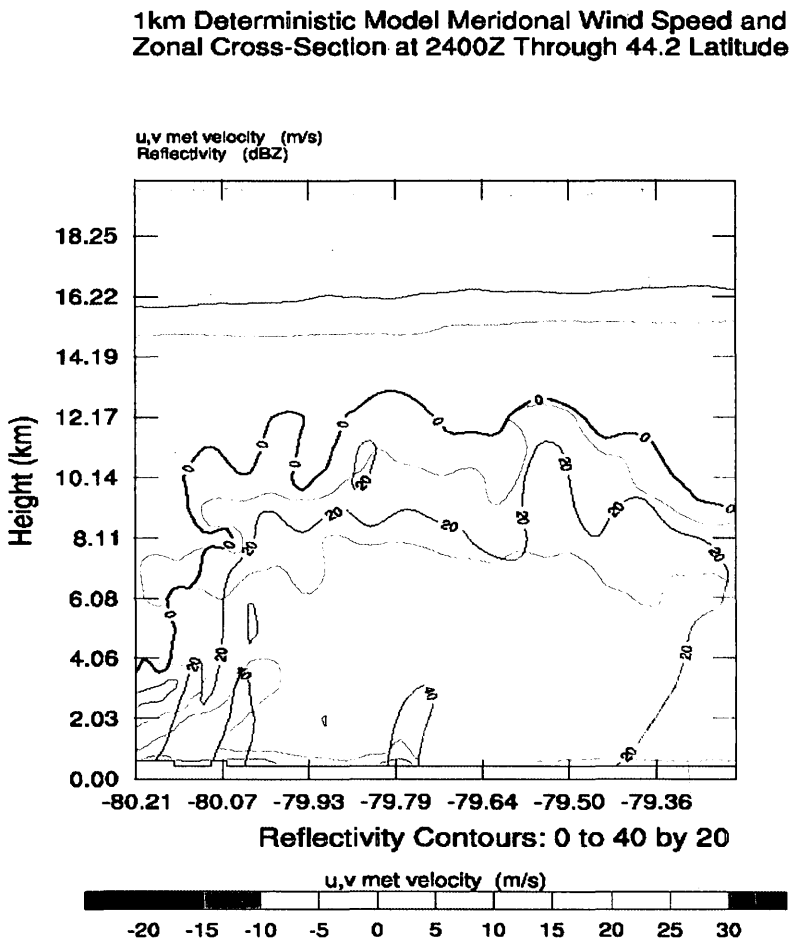


Figure 19: Same as Figure 18, but for a close up view near the squall line.

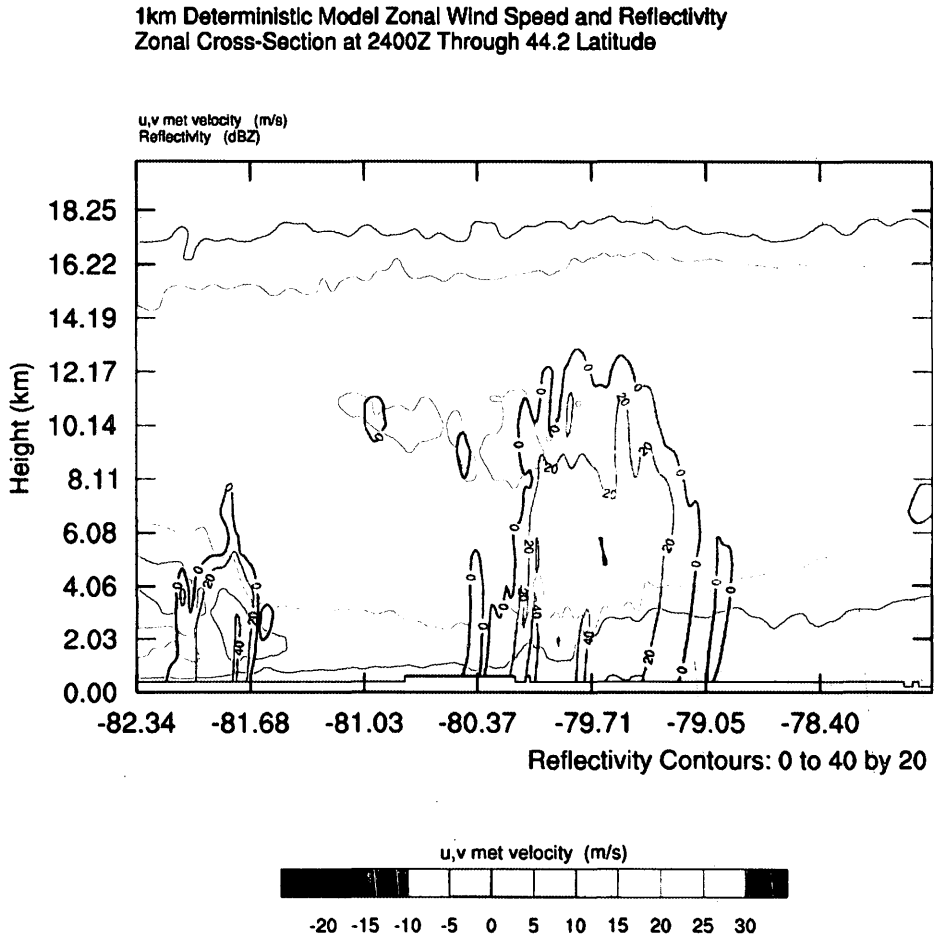


Figure 20: Same as Figure 18 but for zonal winds (color shaded).

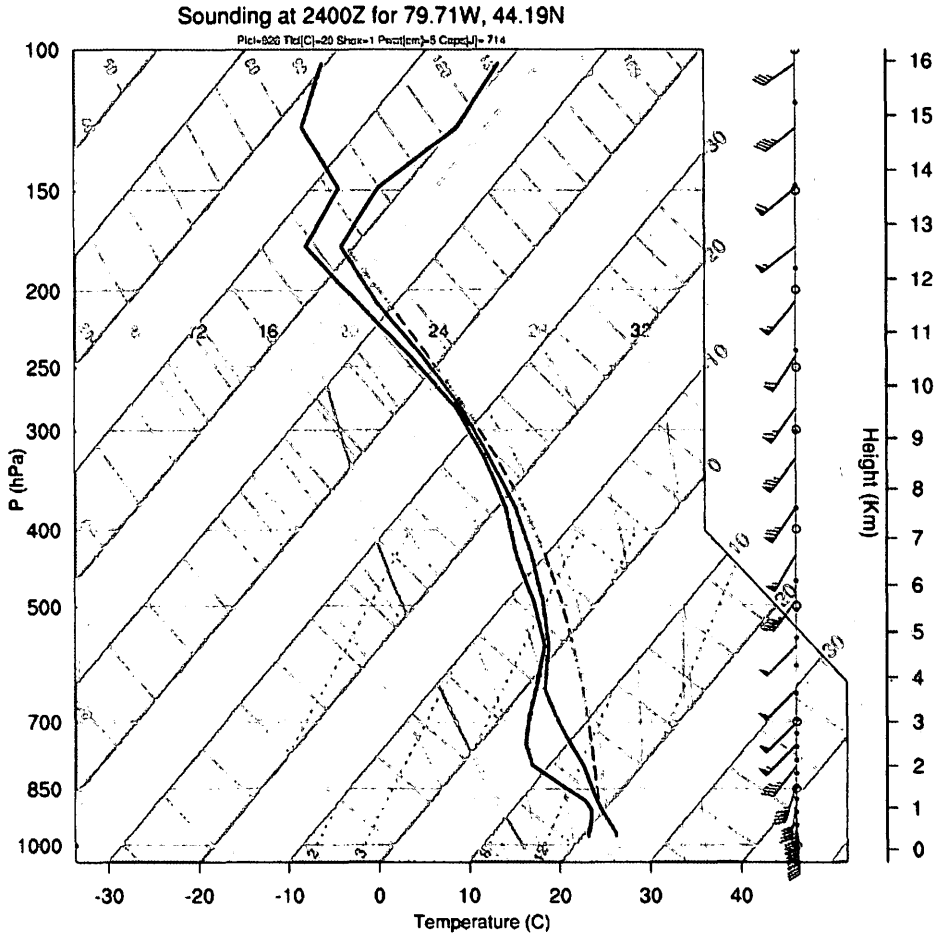


Figure 21: Skew-T diagram for high resolution simulation at 44.19 N, 79.71 W at 2400Z

The high resolution simulation then shows a clear mechanism for generating the supercells along the squall line. Is the same mechanism evident in the corresponding cross-sections for the low resolution simulation? Considering

Figures 22, 23 and 24, which show the low resolution cross sections for the meridional and zonal winds respectively, we see clear evidence for the same mechanisms – specifically a low-level jet with high wind shear at 2km level at approximately 79.7° W in the meridional cross section, with a corresponding east-west moving wind at ground level moving against the west-east moving air mass above up to this 79.7° W.

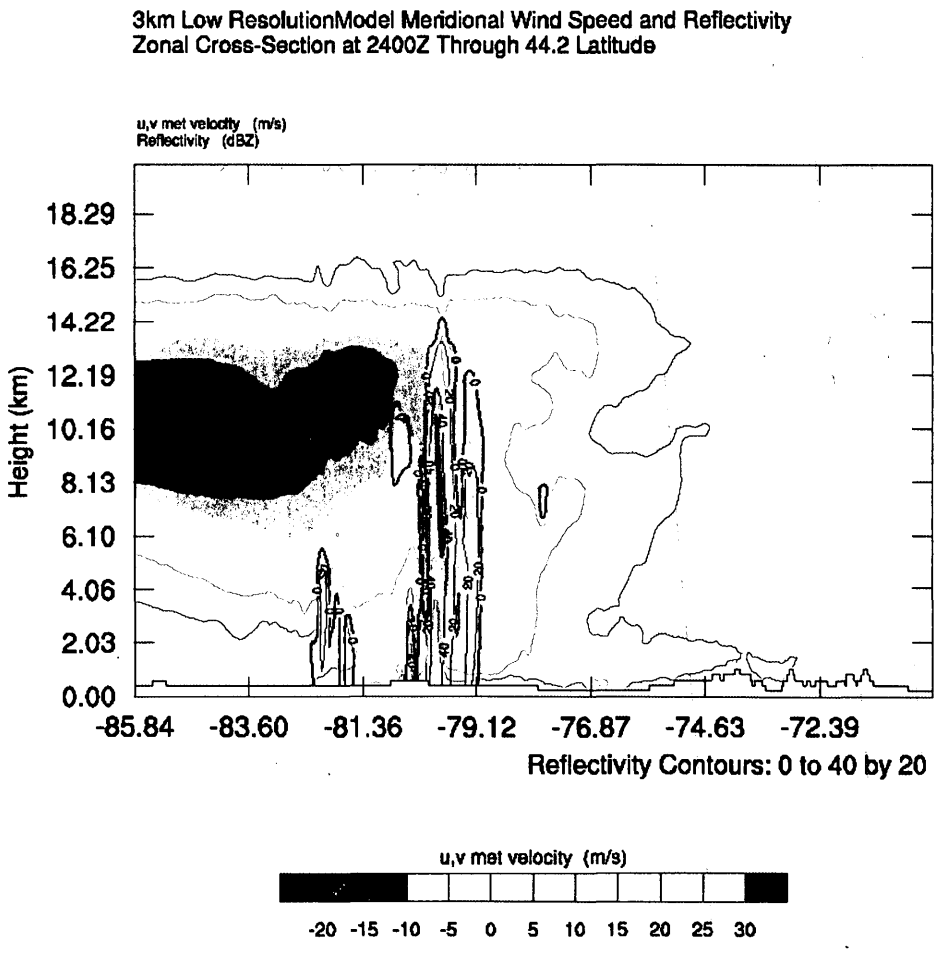


Figure 22: Meridional winds (color shaded) for the low resolution simulation through a vertical-zonal cross section crossing 44.19 N, 79.71 W at 2400Z. Simulated radar reflectivity (units: dBZ) is in contours at intervals of 20 dBz.

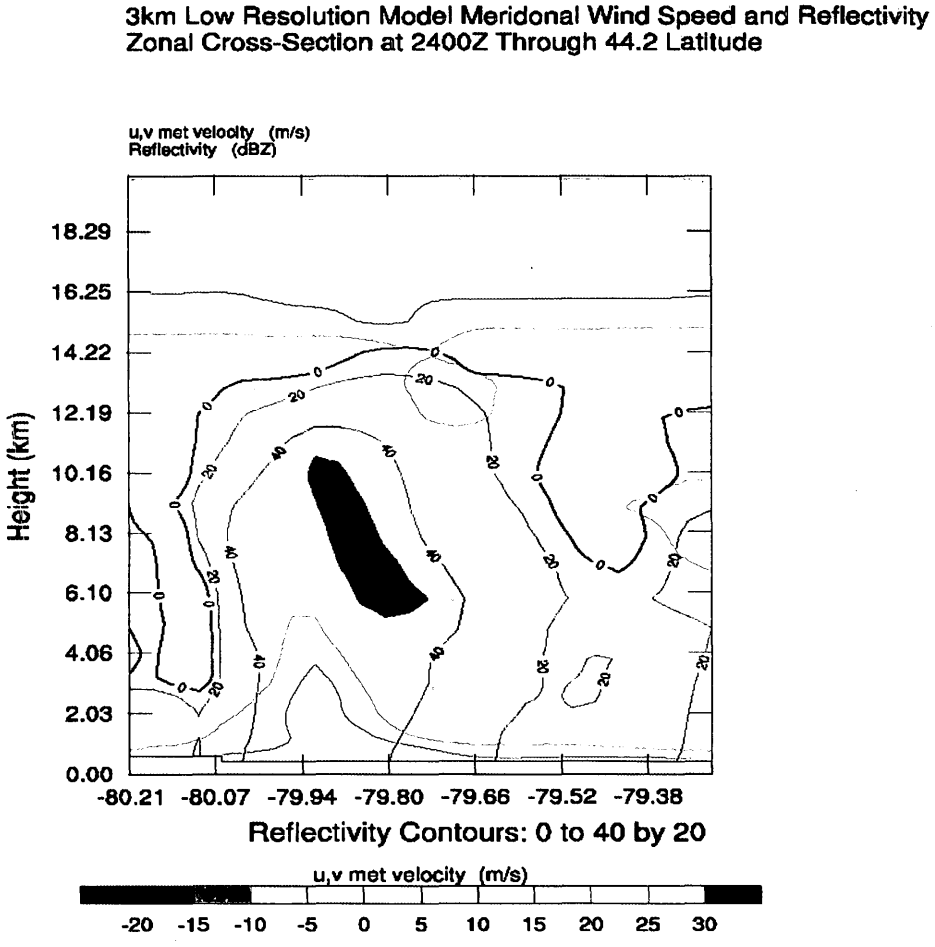


Figure 23: Same as Figure 22, but for a close up view near the squall line.

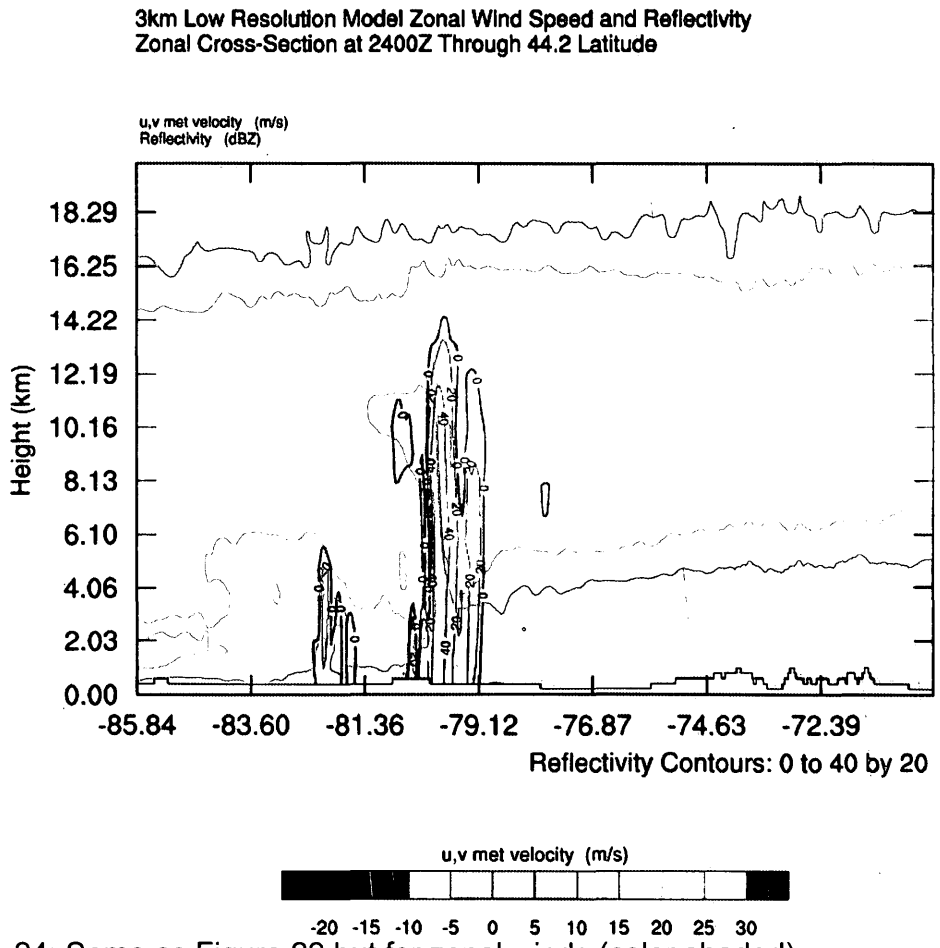


Figure 24: Same as Figure 22 but for zonal winds (color shaded).

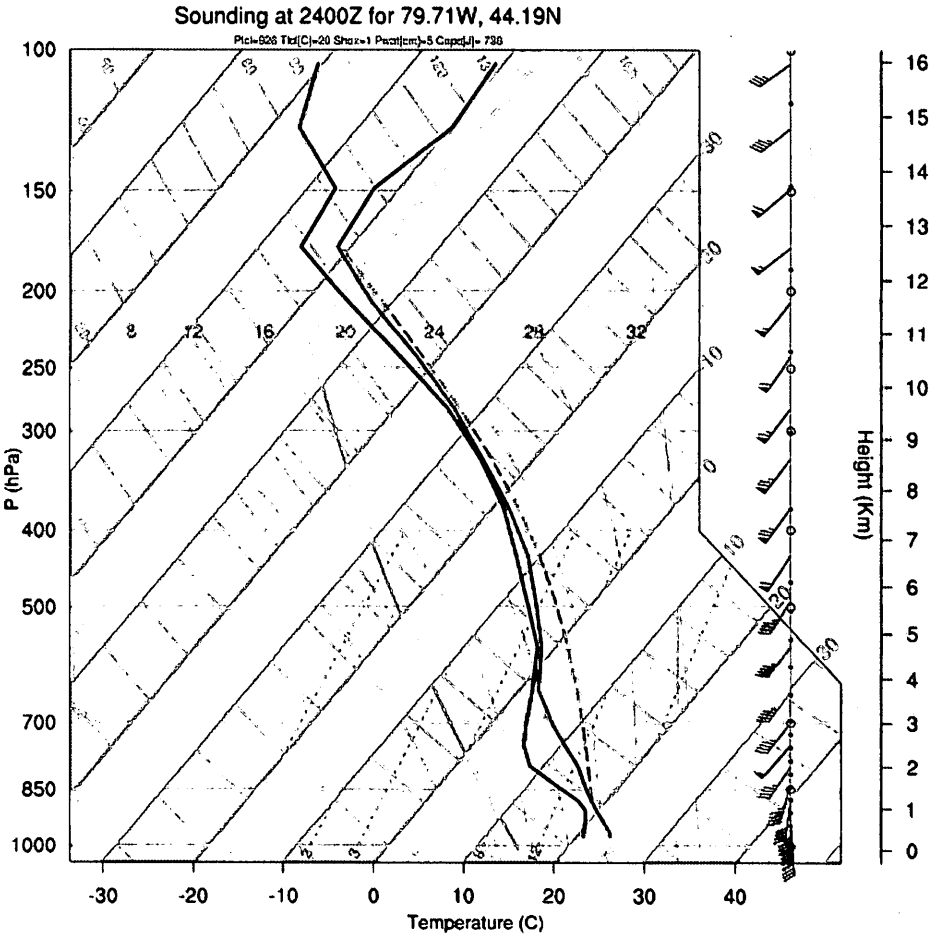


Figure 25: Skew-T diagram for low resolution simulation at 44.19 N, 79.71 W at 2400Z

1.4 Conclusions

So, as stated above, we can conclude that for the purposes of modeling squall line formation and position at 2400Z, the low resolution simulation appeared to work as well as the high resolution. However, as we have also noted, both the high and low resolution simulation showed a similar lag of approximately 80km to the west for this squall line position as compared to the actually observed squall line.

The high resolution simulation appears to simulate a more continuous squall line with finer structure than the low resolution simulation. This is seen in the simulation of radar reflectivity indicative of atmospheric convection.

We therefore surmise a possible mix of length scales at work contributing to the storms evolution – a large length scale driving squall line position captured equally well by the low resolution simulation as the high resolution simulation, but also a smaller, more 'nonlinear' length scale contributing to the formation of discrete supercell structures, and thus likely to take advantage of the increased degrees of freedom offered by the higher resolution, and thus more dependent on resolution.

Despite this however, peak values of radar reflectivity, storm relative helicity and vertical velocity were comparable if not higher, for the low resolution simulation, therefore adding to our contention that the low resolution simulation was as successful in capturing the basic dynamics of the storm's formation and evolution as the high resolution simulation.

The detailed cross-sectional analysis of the storm indicates that a low level jet at 2km was indeed a driver for the storm's evolution, creating a low level shear environment in the moist unstable air in the warm sector. While CAPE values were not especially high, the high shear values were enough to constitute a very high values of helicity which in turn contributed to the strong updraft vorticities that resulted in supercell evolution all along the squall line.

We can also conclude from our results that the same physical processes appear to be driving supercell evolution in both resolutions.

Finally, it should be mentioned that for both the low and high resolution simulations, mesocyclones (and therefore supercells) do appear to have been adequately simulated – in both cases, areas of high upward vertical velocity and vertical vorticity were predicted. However, significant downdrafts were not

simulated at either resolution. As we discussed in our Introduction however, this was to be expected given that these downdrafts were driven by sub-scale thermodynamic processes. Given that tornadoes were similarly of a scale smaller than our models are capable of resolving, we should not have expected tornadoes to be predicted by our model. Nevertheless, our simulations were capable of predicting the precursors for tornadic activity, the supercells themselves, and are thus certainly capable of predicting the at least the possibility of tornado genesis, which is the most that one could expect at these resolutions.

2 Convective-Scale Ensemble Methods

2.1 Introduction

We come then to the second phase of our study. Having established that the same physical processes are being captured by the low resolution model as are captured by the high resolution model, we can feel that one can use the results derived from an ensemble of low resolution simulations with some justification as a suitable basis of comparison with our single deterministic high resolution simulation – in other words, to put it more colloquially, that we would be comparing apples to apples, and not apples to oranges, as we would be for entirely different modeled physical processes. Again, this is important, for having done this, we can now regard the statistical products of our ensemble – the ensemble averages for instance – as being averages of the same physical processes as those of the high resolution simulation, and not averages of simulations which were too coarse to capture the same physics.

We will begin this Introduction with a general overview of those studies that have been carried out on the issues related to ensemble modeling in general – issues

such as nonlinearity and predictability, and the properties of the ensemble mean and its benefits, among others. Towards the end of this Introduction, we will introduce some studies which have been carried out specifically on the convection resolving ensemble modeling of convective phenomena. As noted in our general Introduction above, this is an important issue of considerable practical relevance, given the difficulties inherent in effectively modeling such strongly convective phenomena, and the great need to correctly forecast them given their often catastrophic impact.

We note again the apparent singularity of the meteorological condition being considered in this study – the tornadic storm system and the associated squall line of August 20, 2009 in southern Ontario. As noted above, this event spawned an unprecedented 11 tornadoes across southern Ontario and was therefore unusually strong in its intensity. Also, as we discussed above in the first phase of this study, the prime lifting mechanism driving the convection along the squall line would appear to be a mesoscale phenomenon, and thus driven by a more nonlinear set of dynamics than would be the case if more linear, quasi-geostrophic or synoptic scale processes were dominant.

Thus, as we mentioned above, it was decided for this study to carry out an

analysis focused only on this single event alone, given that it would be very difficult to find other storms which we could be sure were driven by exactly the same dynamics – certainly storms of this intensity are very rare, at least in Southern Ontario. It is recognized of course that an analysis of only a single event is more inconclusive than would be the case were multiple events analyzed. In a loose sense however, ensemble-based methods compensate for this loss of certainty by putting the statistical burden of certainty on a plurality of slightly differing simulations of a given single event, rather than on several different but similar events with similar (one hopes) underlying dynamics and physics. On important example of such an ensemble-based usage is the ensemble mean, which we will elaborate upon below.

We therefore will look at this study as a test case to see how much ensemble methods can tell us about a presumably highly-nonlinear, and somewhat singular atmospheric phenomenon, when multiple events of a similar nature may be unavailable for a more exhaustive study. There are of course also other practical benefits to this approach. It is our belief that these benefits, which will be elaborated upon below, will provide justification for this approach.

We will begin with a discussion of ensemble methods in numerical weather

prediction, and specifically what benefits are offered by such methods when compared to single, deterministic models. We will also discuss here the questions of predictability, and how predictability relates to questions of the scale, since such questions are central to this study.

2.1.1 Overview of the Development of Ensemble Methods

Ensemble methods had their beginning with Lorenz (1963), who had shown that nonlinear dynamical systems, like the atmosphere, have only a finite limit of predictability, where initial perturbations or errors in the system's starting state will over time multiply until any hope of predictability is lost – even round off errors in the system's starting state will eventually lead to completely different predictions of the system's final state, given enough time. The time at which such a complete loss of determinism of the system's final state occurs is a measure of the system's predictability – Lorenz estimated that the upper limit of the weather's predictability is about 2 weeks. Thus the well-known 'butterfly effect', where the flapping of a butterfly's wings in Brazil, could lead to the difference between calm weather or a hurricane in Texas at a sufficiently distant time in the future. While an extreme example, the unpredictability highlighted by this effect is certainly

relevant, for it is fundamentally this inherent unpredictability which accounts for the difficulties faced by all numerical weather prediction system today.

To complicate matters however, there is as well the fact this predictability limit depends also upon the state of the atmosphere – for some smaller scale (mesoscale) weather systems, which tend to have more nonlinear dynamics, the predictability time scale will in fact be even lower than for the less significantly nonlinear and slower evolving synoptic systems. Clearly, especially nonlinear mesoscale systems such as supercells and squall lines would fall into this category of short predictability processes. This point has been raised in the literature, where convective instability has been implicated in reducing the predictability of mesoscale processes in weather models (Hohenegger and Schar, 2007). Indeed, Hohenegger and Schar (2007) add that for some smaller-scale phenomena, the predictability horizons may be as short as 75 minutes.

Clearly then, modeling of mesoscale convective processes presents a significant challenge to the NWP community. In order to help deal with the general challenge presented by the loss of predictability of the nonlinear processes of weather modeling, Lorenz (1965) proposed that rather than running a single deterministic simulation, one might rather run an ensemble of weather models,

each differing from the other by a slight perturbation in initial conditions. Lorenz however had a simple model of 28 variables, and introduced an ensemble of 28 ensemble members.

There is of course wisdom in this approach – the approach of ensuring that the number of ensemble members equals the number of model variables. If one assumes for a moment that all model variables are independent, then one would need at least this many ensemble members, each with linearly independent initial perturbations, to ensure that the full space of possible perturbations realizable by the system is in fact spanned by the ensemble. By this reasoning, employing any smaller number of ensembles will run the risk that there will be directions in perturbation space that will be missed by ensemble members.

Of course, the number of model variables of any NWP system which attempts to accurately model the atmosphere is on the order of 10^7 (Kalnay, 2003), so clearly any attempt to run an ensemble with this many members is far from being practical. Equally clearly however, one has only to give the matter a little thought before one realizes that the model variables in any NWP model are far from being independent from each other. The many equations which drive evolution and balance of the model variables in a NWP model impose very significant

restraints on the independence of these variables, thereby lessening the span of perturbations realizable by the system, and therefore the number of ensemble members needed to span their phase space. As well, the system's restriction on the type of perturbations possible in the system will favor some of these realizable perturbations over others, making some dominant at a particular time and place in the system (the dominant modes of the error, or errors of the day) and some less dominant. While it is not the purpose of this study to get too deep into such matters, it is worth mentioning that studies have shown that most of the variation in atmospheric variation is in fact accounted for by relatively few dominant modes of variation, at least locally. Indeed, Patil et al. (2001) has shown that locally, the atmosphere can exhibit very low dimensions of variation – with values as low as 2.5 for a local measure of dimensionality called the Bred Vector dimension. As they have noted, such low levels of Bred Vector dimension indicate correspondingly low levels of atmospheric variation.

So, it would seem that one might be able to obtain useful results from an ensemble of number much less than the number of model variables of the system. One of the first to investigate this was Leith (1974), who found that, by using a 'Monte Carlo' approach to choosing ensemble members, considerable benefit could be obtained by using a relatively small number of ensemble

members. Indeed, according to Leith adequate accuracy for ensemble forecasts using conventional NWP models could be obtained with sample sizes as small as 8.

In his analysis, Leith was also able to show that a major benefit of using an ensemble for forecasting is that the ensemble mean has the property that, over a long time period, its variance is only slightly more than ½ of what the variance for a single forecast would be over the same period.

For u denoting the deviation of some modeled variable of interest, perhaps wind, from the climatology, and u_0 denoting the deviation of the true state of this variable from the climatology, then over a long enough, period, we will have a variance of this variable u from the true state u_0 of:

$$\langle (u-u_0)(u-u_0) \rangle = \langle uu \rangle + \langle u_0 u_0 \rangle - \langle uu_0 \rangle - \langle uu_0 \rangle \qquad (3)$$

Note however that over a long enough time u and u_0 are uncorrelated, so the last two terms disappear, and the variance of u approaches the variance of the climatology (which we will denote as U). The first and second terms become U, and thus the variance of u from the true state becomes:

$$\langle (u-u_0)(u-u_0) \rangle = 2U \qquad (4)$$

which is twice that of the climatology from the true state. So over a long enough time, any deterministic forecast has a variance of twice that of the climatology. Hence, the benefit of tempering a forecast towards a climatology (Leith, 1974).

If however, we use the ensemble mean instead of the deterministic forecast, our variance drops to (1 + 1/m) times the climatological variance – an automatic tempering towards climatology. To see this, consider the ensemble mean defined as:

$$\bar{u} = \frac{1}{m} \sum_{i=1}^{m} u_i \qquad (5)$$

then the variance of the mean from the true state is:

$$\langle (\bar{u}-u_0)(\bar{u}-u_0) \rangle = \langle \bar{u}\,\bar{u} \rangle + \langle u_0 u_0 \rangle - \langle \bar{u}\,u_0 \rangle - \langle u_0\,\bar{u} \rangle \qquad (6)$$

But since the mean u and u_0 become uncorrelated over time, the last two terms

approach 0. Also, substituting in the above equation for the mean u into the first term gives:

$$\langle(\bar{u}-u_0)(\bar{u}-u_0)\rangle = \frac{1}{m^2}\sum_{j=1}^{m}\sum_{k=1}^{m}\langle u_j u_k\rangle + \langle u_0 u_0\rangle \quad (7)$$

Since u_j and u_k are uncorrelated over a long enough period however, only those terms for which j=k affect the above sum, giving:

$$\langle(\bar{u}-u_0)(\bar{u}-u_0)\rangle = \frac{1}{m^2}\sum_{j=1}^{m}\langle u_j u_j\rangle + \langle u_0 u_0\rangle \quad (8)$$

or, more simply, noting that the first term is just 1/m times the variance for long lead times:

$$\langle(\bar{u}-u_0)(\bar{u}-u_0)\rangle = \left(1+\frac{1}{m}\right)\langle u_0 u_0\rangle = \left(1+\frac{1}{m}\right)U \quad (9)$$

So clearly, as we add more ensemble members, our variance gets closer and closer to the climatology – hence the automatic tempering to climatology for an ensemble mean forecast that we noted above.

2.1.2 Predictability and the Ensemble Mean

The ensemble mean has another benefit. It also acts as a filter which removes small scale unpredictable features at high resolution. This nonlinear filtering is important, because this unpredictability is manifested as an extreme sensitivity to initial conditions – simple round off error in the model quickly grows to the point where forecasts become useless. Having a mechanism such as the ensemble mean to filter out the unpredictable nonlinear modes helps remove this extreme sensitivity and make forecasts more reliable. Indeed, it has been noted that the ensemble mean is the optimal nonlinear filter which gives the best unbiased estimate of the true state of the atmosphere in an root-mean-squared (RMS) sense for a model that is identical with the real atmosphere and for an ensemble that is perfect (Szunyogh and Toth, 2002). However, even for an imperfect model and ensemble, the ensemble mean is still superior to a single control forecast as the forecast lead time increases (Toth and Kalnay, 1993). To understand why this is the case, we must understand how the perturbations of an ensemble will vary over time, and how this variation relates to the predictability of the ensemble mean.

Any ensemble is of course just a collection of initial condition states each with a

set of small perturbations from a given control state. These perturbations are ideally meant to equal the uncertainty in the analysis that one would obtain during a data assimilation step – in other words, the best possible estimate of our error in our model state assuming a known degree of uncertainty in our background error and a known degree of uncertainty in our observation error. So, what the perturbations should represent is the uncertainty that one would have in an operational context for the starting state of the atmosphere at the beginning of a simulation. With a perfectly predictable system - where there is a perfect correlation between the system's input state and its output state, and thus without any loss of determinism over time - we could track the evolution of our ensemble over the course of the model's evolution and know what the uncertainty in our final state would be simply by looking at how large the variation of final states is in our ensemble. For areas where there is a great variation of some particular variable, say 2m temperature for example, we can know that our predictions will be more uncertain in this location than at areas where there is a small variation in that variable. This variation is typically quantified by the ensemble spread, which is the RMS variation of each ensemble member from the ensemble mean, and is often taken as a measure of the unpredictability of a phenomenon – although this is a matter of some debate about which we will elaborate upon further below.

Thus, for areas where there is a great variation or spread in some particular variable, say again 2m temperature, there will be more uncertainty in our predictions for that variable at that location. We can however take the ensemble mean, which averages out these variations to give us the best possible estimate of this variable.

To see this, we will assume again that we have perfect predictability and a perfect ensemble. Now in any system, there will be dominant modes of variation (we will discuss this further below, but for now we need only note that some directions of variation will be followed by the system much more than others – these are the "errors of the day", or the dominant modes of variation of the system, which are time dependent and will vary with synoptic and mesoscale conditions). With a perfect ensemble, our ensemble member's perturbations will span these dominant modes of variation both along the vectors oriented in the direction of the dominant modes and opposite to them. In such a case for all modes of variation, our ensemble mean will therefore tend to fall midway between these spanning dominant modes, and so the ensemble mean should tend to fall at the center of this cloud spanning the most dominant modes of variation. Consider for illustrative purposes variations in only a single direction for a particularly dominant mode, with a sample size of N ensemble members. In

such as case, if each member has a variation (standard deviation) of σ, then the variation of the mean of these N members will tend to a variation (standard deviation) of $\sigma/N^{1/2}$ - a well known property of the standard deviation of the sample mean for an N member sample. This is why of course, for a perfect ensemble with perfect predictability, the ensemble mean converges to the truth as our ensemble size increases, when the initial state of the model equals the true state at that initial time.

Unfortunately of course, we neither have perfect predictability nor a perfect ensemble in practice. We will deal with this lack of perfect predictability in practice first.

2.1.3 The Relationship Between Scale and Predictability

For any NWP model, as we have noted above, we must have an inherent nonlinearity, since of course the advection term of the Navier-Stokes equations introduces an unavoidable nonlinearity, and any NWP model must of course contain the Navier-Stokes equations since these equations are part of the fundamental basis of any simulation of fluid flow. It is found however, that this nonlinearity is more pronounced at smaller scales than at larger scales – in other

words, sensitivity to initial conditions, the hallmark of this nonlinearity, is much more pronounced the smaller in scale we go. Thus, we have the larger scales of flow, baroclinic waves for example, much less susceptible to small variations in initial conditions than smaller scale features, such as for example, convective modes in the mesoscale. What this means then is that we can predict the large scale variations much more easily and for longer periods than the smaller scale features – the larger scales, although also nonlinear in nature, are much less susceptible to this nonlinearity (or sensitivity to initial conditions) than the smaller scales.

This lack of predictability is best understood in terms of the rate of growth of errors over time of variables that are reflective of the scale in question. For example, the variation of 500 hPa geopotential heights are reflective of synoptic variations in large scale flow (Hohenegger and Schar, 2007), and it is found that variations of 500 hPa geopotential tend to vary with a doubling time of about 2 days (the doubling time is of course the time it takes for a perturbation to double in size), while for variables like low level temperature, reflective of smaller scale modes of variation, the doubling time is only 2-4 hours (Hohenegger and Schar, 2007). These errors will not grow without bound, however, but will grow exponentially until they reach a maximum, at which point we will have reached

saturation. It is at this point where there is a complete loss of predictability of our system given our initial state – there will here be no correlation between our simulation end state and the initial state, and the variations of our end state will effectively be completely random. Again, Hohenegger and Schar (2007) estimate from their data that for the ECMWF EPS (Ensemble Prediction System) for 500 hPa heights, and for the European LM model for lowest level temperatures, saturation of the errors of 500 hPa heights and lowest level temperatures occur between 6 and 10 days, and 12 and 28 hours respectively. This latter result is extremely important in its relation to this study, for it appears to be stating that saturation of the 2m temperature can occur in shorter time span than is used for the simulations of this study – 24 hours. If saturation of the smaller modes occurs before the simulation finishes, what confidence can we have that there will be any correlation between our initial smaller-mode related perturbations and our final perturbations, as a fully saturated perturbation is effectively a random draw without any correlation to the initial perturbation?

The situation of course is not this simple. Note that the 12 to 28 hour saturation time is the average for low level temperatures over all LM runs of Hohenegger and Schar's study – for fully convective processes which are even more unpredictable, the situation is even worse. Short convective modes, manifested

in such features as individual thunderstorms, may indeed be unpredictable for time periods as little as 1 hour (Kalnay, 2003). However, Kalnay adds that organized convective activity such as squall lines and mesoscale convective systems which are forced by the larger scales can be predictable for much longer, even on the order of a day or less. As well, of course, these predictability limits will also vary from case to case, and will depend upon the dominant modes of variation of the system, hence the variations noted above in Hohenegger and Schar's results for the 500 hPa geopotential and low level temperatures, which reflect a range of conditions.

So what sort of predictability we should expect from the squall line system of the current study is an open and important question, and key to our understanding of this system. It will therefore be a question that we will consider further, and will be of relevance throughout this study.

2.1.4 Initial Ensemble Perturbation Size and Its Relationship to the Scale of the Modes Excited and the Saturation Amplitude of the Excited Modes

We should also mention the importance of ensuring that the perturbations chosen initially are large enough to span the modes related to the larger scale variations discussed above. This is important not only to ensure that the ensemble perturbations initially reflect the operational uncertainty of our initial state, but also to ensure that we excite more than just the small scale, convective modes of variation. As noted by Toth and Kalnay (1993) in their study, if initial amplitudes for the ensemble perturbations (using NCEP's bred vector method – Tracton and Kalnay, 1993) were initially set to be between 1 and 15 m for the 500 hPa geopotential height, then the scale of perturbations evolving from the initial ensembles during the simulation were of short, baroclinic waves and these perturbations grew at an exponential growth rate of 1.5/day. However, if the initial perturbations were scaled back to only about 10cm or less, then while these smaller perturbations grew much quicker initially, at a rate of about 5/day, they tended to saturate much more quickly and at much lower amplitudes, remaining at these low amplitudes without further growth[9]. Their results showed that this

9 It is important to note here that Toth and Kalnay are talking about growth of perturbations in a bred vector context – where perturbations are periodically rescaled back so that perturbation size equals an initial uncertainty. If a convective mode variation does not have time enough to align itself to one of the "higher energy" synoptic-scale modes of variation and grow along that mode for any appreciable time, but rather simply saturates in the convective mode before the next bred-vector iteration, then the perturbations will only remain at these lower energy convective mode amplitudes. Such an ensemble will be under-dispersive.

difference in saturation amplitude between the two cases was quite significant – while the larger scale , higher energy initial perturbations saturated at a value on the order of about 100% of climate variance, the smaller scale, lower energy perturbations saturated at a value on the order of about 1% of climate variance. Thus, with a set of initial perturbations which is too small to catch both high and low energy modes of variation, only the quickly saturating low energy convective modes will be excited, and not the higher energy baroclinic modes. We therefore need to be sure that our initial perturbations are sufficient to get past this low energy, small scale saturation barrier. In other words, the result from an ensemble of perturbations that are too small will be under-dispersive – the set of end states of the ensemble will vary only very slightly about their mean, and will in all probability not contain the true state. Thus, it is important that any ensemble employed has initial perturbations large enough to span these larger scale, higher energy modes.

So, as we have noted, the ensemble mean is the best estimate of the true atmospheric state that we have, given the information available from our ensemble. But one might ask, how certain can one be of that the mean correctly reflects this true state? What is the uncertainty that the ensemble mean is correct?

2.1.5 Ensemble Spread and the Spread/Skill Relationship

Typically, the ensemble spread has been used to make this determination. The ensemble spread is for any variable, equal to the root-mean-squared value of that variable among the ensemble members. It would seem evident that if there is little variation (and therefore spread) among the ensemble members then there should be little uncertainty about the true result and therefore the forecast should have a high skill. However, that there indeed exists in general a high correlation between ensemble spread and forecast skill is in fact a matter of some debate (Whitaker and Loughe, 1998).

On the affirmative side, one can consider a study has been carried out to answer this question for forecasts of precipitation (Verret et al., 2002), specifically for 12 hour probability of precipitation (PoP) forecasts. Here Verret et al have found that the spread can be used as a proxy for a confidence index of the skill of the ensemble mean, and that the spread-skill relationship for the 12-hour PoP remains statistically significant for 240 hours.

How significant in general however this relationship remains over time is an important issue for the purposes of this study, so we will discuss this matter in

more detail. Houtekamer (1993) has looked into the spread-skill relationship from a theoretical perspective which provides much useful insight. We will therefore present his analysis as follows.

We assume an ensemble of forecasts all verifying at the same time, and from this ensemble we determine the local ensemble spread S_M at a given point from the following lognormal distribution:

$$S_M = S_c \exp(\beta_M d_1) \qquad (10)$$

where β_M is a parameter which determines how much the spread varies, S_c is a constant[10], and d_1 is a normally distributed random variable of mean 0 and variance 1; N(0,1).

We also assume that the corresponding local forecast error η_M at a given point M is given by the following formula:

$$\eta_M = d_2 S_M \qquad (11)$$

10 Houtekamer's analysis does not include this constant, but his analysis has taken the maximum spread to be unity for simplicity, since he is only interested in the correlation between spread and forecast error, and the value of this constant has no bearing on the result. In the aim of correctness however, we have included this constant in the analysis, since the spread of any physical measurement necessarily should not be either dimensionless or in general necessarily 1 at maximum.

where again d_2 is also a N(0,1) distributed normal variable.

With this relationship established, Houtekamer (1993) shows that the correlation between the local spread and the local forecast error is given by:

$$\rho^2\left(S_M,\eta_M\right) = \frac{2}{\pi}\frac{1-\exp\left(-\beta_M^2\right)}{1-\frac{2}{\pi}\exp\left(-\beta_M^2\right)} . \qquad (12)$$

Also given in his study are two maximum likelihood estimates of S_M and β_M from data measurements of spread S_i:

$$\log\left(S_M\right) = \overline{\log S_i} \qquad (13)$$

$$\beta_M^2 = \overline{\left[\log S_i - \log\left(S_M\right)\right]^2} . \qquad (14)$$

These can be used to estimate S_M and β_M from known measurements of spread.

Once the parameter β_M a is known, the correlation between the spread and the forecast error is easily determined. Clearly, as β_M approaches negative infinity,

the correlation approaches 0 as well. With β_M near negative infinity, S_M is almost uniformly S_c over the domain with little to no variation, since the variations in d_1 are damped by the large negative β_M. With β_M large, variations in d_1 will have a much bigger effect, and S_M will vary much more over the domain. So β_M tells us how much S_M varies over the domain – it is a parameter which determines the variation of the spread, and as we can see, it also directly determines how much the spread correlates with the forecast error (equation 15). One can then see from inspection of equation 15 that particularly large positive or negative β_M values give us high values of the theoretically expected correlation. This gives us a theoretical understanding why anomalously high or low values of spread in recent studies have been found to be those with the strongest spread-error relationship (Grimt and Mass, 2003; Whitaker and Loughe, 1998).

2.1.6 Ensemble Signal to Noise Ratio

Another important relationship to consider is the ratio of the ensemble mean to the ensemble spread – called the signal to noise ratio. One can understand intuitively that if we have a low value of this ratio, then the variation of the quantity being modeled is much bigger than the value of the quantity being measured, and thus, the predictability of this quantity must be correspondingly

low. To help quantify this relationship Tang et al. (2008) have shown that the anomaly correlation (r) of a series of N predictions of some variable u, given by:

$$r = \frac{\sum_{i=1}^{n}\left(u_i^p-\mu^p\right)\left(u_i^o-\mu^o\right)}{\sum_{i=1}^{n}\sqrt{\left(u_i^p-\mu^p\right)^2}\sum_{i=1}^{n}\sqrt{\left(u_i^o-\mu^o\right)^2}} \qquad (15)$$

where p in the superscript denotes predicted values, and o denotes observed values, with μ^p and μ^o denoting their respective means, relates the signal to noise ratio s (the ratio of the mean to the spread or standard deviation of u), as follows:

$$r = \sqrt{\frac{s^2}{s^2+1}}\,. \qquad (16)$$

Thus, a signal to noise ratio of 0.5 for example, would correspond to an expected anomaly correlation of 0.447, which is not a very strong expected correlation between the observed and predicted values. As the signal to noise ratio increases, the correlation between our predicted and observed anomalies (r) increases, so a signal to noise ratio of 1 corresponds to an expected anomaly correlation of 0.71. For a signal to noise ratio of 2, the correlation has reached 0.89. Thus, an argument might be made that for a reasonably predictable

simulation, we would require that the signal to noise ratio be appreciably greater than 1. We will adopt this criterion (minimum signal to noise ratio of 1) for the purposes of this study as the minimum signal to noise ratio possible for a reasonably predictable phenomenon.

2.1.7 Previous Convection Resolving Ensemble Studies

As we have noted above, there have been only a few studies that have used ensemble methods for convection-resolving simulations of convective phenomena. Kong et al. (2007), studied the use of an ensemble of 3km simulations to model the development of a tornadic thunderstorm system – a study very similar to the study of this paper. As they noted, theirs was the first study to use full-physics storm scale ensemble simulations incorporating realistic terrain and realistic horizontally varying initial conditions with the assimilation of real observations (data assimilation). Since the study of this paper is also investigating a tornadic system using very similar methods, this study of Kong et al is very relevant. Note that one very significant difference between the current study and that of Kong's is that the current study does not employ data assimilation. Thus, any benefits realized in this study using ensemble methods

over a single deterministic study can be argued to be attributable to ensemble methods alone, rather than ensemble methods in combination with data assimilation methods.

However, there are also other differences between our study and that of Kong et al. Firstly, and most importantly perhaps, their study using a scaled lagged averaging forecasting (SLAF) technique to generate their ensembles (Ebiusuzaki and Kalnay - 1991). This technique uses staggered forecasts which start at different times but which all verify at the same time. The differences between the verifying forecasts are then used as the basis of perturbations which, when added to a control forecast, provide the ensemble for ensemble forecasting. Note that because error growth occurs as the staggered forecasts progress, earlier-initialized forecasts will have larger perturbations than later-initialized forecasts, which necessitates the need to scale the perturbations accordingly.

However, as Kong et al noted, the scaling technique used in traditional SLAF is inappropriate for fast growing convective perturbations, as perturbations are assumed to grow linearly with time. They therefore developed an amplitude-based scaling method for their study. In addition, presumably due to the fact that their ensemble was generated from staggered initializing forecasts, only 5

members were used in their study. While they did not make mention of the reasons for this relatively small sample size in their study, one may surmise that the reason may relate to the fact that it would be necessary to ensure that the staggered forecasts were not initializing too far into the past, as nonlinearities and possibly saturation of even the larger scale modes might then become an issue. They also note however that considerable benefits can be obtained from ensemble products even when the ensemble sample size is as small as 5, which is comparable to Leith's lower limit of 8 ensemble members, and could therefore be argued to be justifiable. Nevertheless, we can conclude generally from these considerations that generating ensemble perturbations for convective scale modeling presents a significant challenge. Hence the need to develop a computationally inexpensive alternative for generating ensemble perturbations for research/investigative purposes. In the approach of this study, to be presented in the Methodology section to follow, we use an existing ensemble – the Global Forecast System (GFS) Ensemble from the National Center for Environmental Predictions (NCEP) – to act as the basis from which the perturbations which generate the ensemble of this study will be interpolated. Note that while the approach of using existing ensembles to generate new ensembles has been used before, it should be remembered that the GFS ensemble uses a much coarser resolution (1 degree) than the ensemble

simulation resolution of this study (3km). Given this very significant difference in resolution, and that most ensemble prediction studies have been used for only for synoptic scale prediction using such coarser resolution ensembles as the GFS, and that only a few studies have been carried out using ensemble methods for convective scale prediction of strongly convective phenomena, the success of using such a coarse data set to generate a convective scale ensemble as in our study is an open question and is worth investigating further. Hence, the need to address this issue in the current study.

To continue, the study of Kong et al noted that for their ensembles, a delay in the initiation of convection was a problem – which they noted was due to model start up issues in the absence of initial data assimilation. Since the same conditions would also occur in the current study – a lack of data assimilation and start up/spin up issues after initialization, a delay in convection initiation for the current study may also be a problem. Obviously, such convection delay should be looked for in our results – especially given the phase lag evident in the low and high resolution single model runs of the first part of this study.

They also addressed the issue of the efficacy of using lower resolution data to generate ensemble perturbations – specifically, is it appropriate to generate

perturbations from coarse-grained initial conditions from an existing forecast and interpolate the data to the finer-grained resolution of our ensemble? Since as we have noted, and will see below, just such a data interpolation method is also used in the current study, this question will also be of importance.

Hanley et al (2011) also conducted a similar study of a convective scale ensemble simulation of a squall line over central Europe. While, like the Kong et al study, they were able to initiate convection for all ensemble members, they noted, also as in the Kong et al study, that there was a 1-2 hour delay in convection initiation among the most of the 28 members which resulted in a 50-200 km phase lag of the position of the simulated squall lines behind the actual squall line, with only a few of the ensemble members accurately positioning the squall line. It should be noted that the ensembles were generated using the Met Office Global and Regional Ensemble Prediction System (MOGREPS: Bowler et al 2008), which perturbs both initial conditions using an ensemble transform Kalman filter and model physics parameters.

They added that they believed that the speed of squall line development is linked to variations in the large scale dynamics. Thus, one could suppose that there would be a positive correlation between the initiation time or place of a simulated

squall line system and some large scale feature(s) of the model initiation environment. Again, this will be a key question which we will attempt to address in the third part of this study.

Ensemble methods for convective scale forecasting were also examined by Leoncini et al (2010) who noted that the growth rate of convective scale errors are around 10 times larger than those of the synoptic scale, and hence the tangent linear approximation used in non-ensemble based data assimilation techniques breaks down within a few hours. In other words, and as noted above, synoptic modes of variation (while inherently nonlinear due to the nonlinear term of the Navier-Stokes equation) are nevertheless less nonlinear than the convective modes of variation (and therefore retain some of the linearity needed by the tangent linear approximation). Thus synoptic modes of variation retain strong correlations for longer periods of time with initial conditions[11] than the convective modes. From this one could conclude that ensemble methods would appear to be much more effective for numerical weather prediction than single deterministic prediction simulations, as they have the ability, as we noted above, to filter out these small scale convective nonlinearities (which have lost any

11 We will look into correlations in more detail in the third part of this study, but it is worthwhile to note that formally defined correlation (see equation 24 in Chapter 3) presuppose a linear relationship. In other words, a significant correlation of a variable with initial condition variables is an indication of a retained linearity of this variable with initial conditions – a condition expected of the "less nonlinear" synoptic modes of variation.

useful information due to modal saturation), but still retain the useful information contained in the "less nonlinear" synoptic perturbations.

However, it should be noted in general that the unavoidable nonlinearities make conventional methodologies for perturbation generation, designed initially for large scale synoptic modes of variation – perhaps questionable when employed for convective prediction due to the loss of initial condition perturbation "information" due to convective mode saturation. Does this convective mode saturation problem negate the possibility of useful convective predictability using ensemble methods, or do the synoptic modes still retain useful predictability for such features as squall line position, if not for the placement of particular storm scale features past these convective saturation timescales? As a result of these concerns, the effectiveness of the ensemble used in the current study for convective prediction is a key question that needs to be addressed. Hence this issue remains one of the key concerns of this phase of our study.

2.1.8 Locating Squall Lines – Moisture Flux Convergence

Finally, a quick mention should be made of methods used to locate the squall

lines of the ensemble members in this study. We will use for our investigations of squall line position a quantity called moisture flux convergence (MFC), which is expressed in equation 17 below. This approach has been used in several studies in the past – for example Banacos and Shultz (2005) have noted, moisture flux convergence is an effective means of locating mesoscale convective boundaries (such as squall lines).

$$MFC = \nabla \cdot (w \vec{v}) \tag{17}$$

where *w* is the mixing ratio (in kg/kg) and v is the velocity.

2.2 Methodology

Our ensembles was generated using an existing analysis ensemble – the GFS ensemble of NCEP. The 20 available perturbed members of the GFS analysis ensemble were used, each verifying on August 20, 2009 at 0000Z, 0600Z, 1200Z, 1800Z and 2400Z (August 21st, 00Z). For each of these analysis times, the corresponding ensemble mean was calculated and subtracted from the ensemble members, to give a set of ensemble perturbations. Next, the 3km resolution simulation initial and boundary condition data sets from the first phase of the study were taken (our base state), and these perturbations were added to this base to give a set of 20 initial and boundary condition data sets for our ensembles. Note that by doing this, not only do we ensure that our initial condition perturbation sizes are the same as those used to initialize the GFS ensemble, but also, we ensure that our boundary conditions also change accordingly and are thus not all fixed to be the same for all ensemble members (which would otherwise tend to 'pin' and therefore restrict synoptic variations – Hohenegger and Schar, 2007). As well, some of the balance that exists between variables in the GFS ensemble will be transferred to our ensemble as well, assuming that the 'control' states for both are close. This will presumably help speed up the time needed for our simulations to 'spin-up' and attain the balance

needed for a successful simulation.

Note that to do this it was necessary to interpolate the GFS ensemble perturbations to our model's higher horizontal resolution, as well as interpolate the perturbations also to our model's vertical resolution. Thus, each ensemble perturbation was interpolated using NCL to the same horizontal and vertical resolution of the base state low resolution 3km resolution simulation.

The simulations using the resulting ensembles as initial conditions and boundary conditions were then run for the full 24 hour period, from August 20, 00Z to August 21, 00Z – which again, for clarity, we will call 2400Z in the following analysis.

As noted above, it should be remembered that our method is not the typical method used to generate ensembles. Typically, fairly involved procedures are used, such as the Breeding Method (Tracton and Kalnay, 1993) used at NCEP. Indeed, this is the method that was used to generate the GFS ensemble used in the study. As has been noted in the literature, ensembles generally suffer from under-dispersion (Stensrud et al., 1997), so a great deal of research has been carried out to ensure that the ensembles used by the various centers (NCEP,

ECMWF, CMC, etc.) are indeed performing as needed, having their ensemble means remain as close to reality as possible, while at the same time having reality fall within the 'cloud' of ensemble members – or as Kalnay has noted (2003), having reality fall as a 'plausible member of the ensemble'. This is a significant challenge and involves a fair degree of work, so having a method to generate quickly an ensemble for study from an already existing ensemble could be useful for investigative and research purposes. Having a result from this study therefore which shows that this method generates an ensemble that is not under-dispersive, and generates accurate forecasts, would therefore hopefully be of some benefit.

2.3 Results and Discussion

We begin first with an examination of the behavior of the different ensemble simulations – how the perturbations of the initial conditions were carried through the simulation to result in differing final conditions of the model by the end of our simulation period – at 2400Z in our model time – or August 21st, 00Z.

2.3.1 Moisture Flux Convergence

We first turn to our moisture flux convergence (MFC) results, which as we have noted in our introduction, are a diagnostic means of locating mesoscale convective boundaries – squall lines (Banacos and Shultz, 2005).

We will consider then low level (850 hPa) moisture flux divergence, noting of course that moisture flux convergence is just the opposite of the divergence, so we are looking for negative values. Following in Figures 26 and 27 are two plots showing the 2400Z low level moisture flux convergence for two of the ensemble members – ensemble members 6 and 8 respectively.

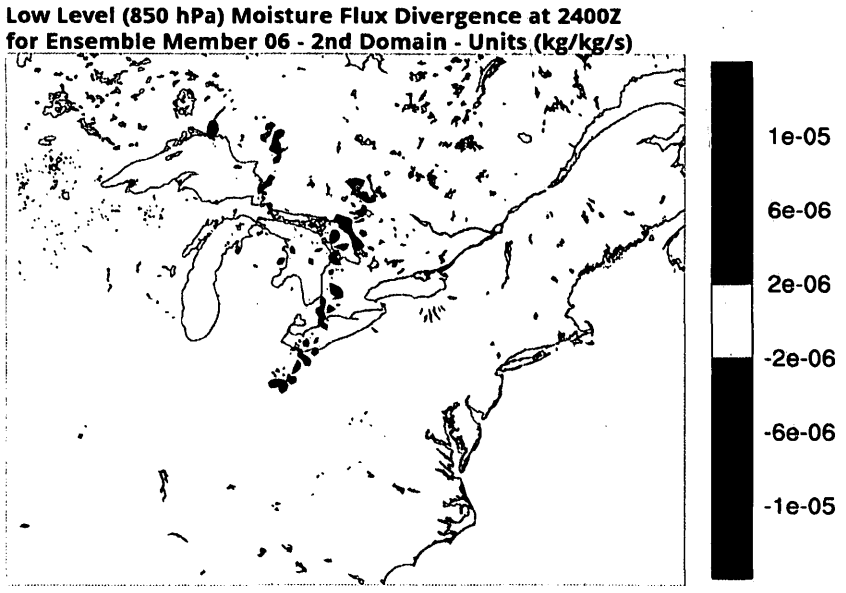


Figure 26: Low level moisture flux divergence for ensemble member 6

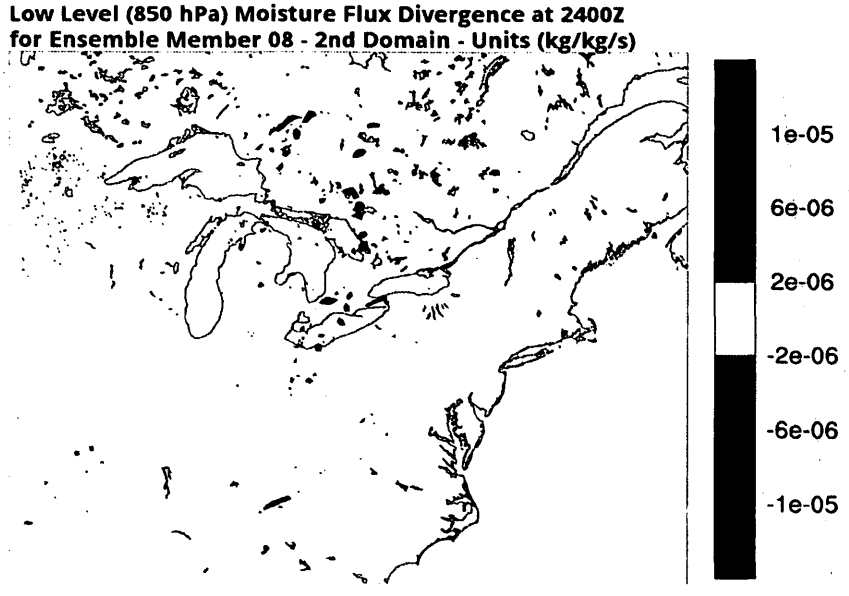


Figure 27: Low level moisture flux divergence for ensemble member 8

As we can see from these figures, there is considerable difference between the two with regard to the quality of the squall lines simulated. Ensemble member 6 shows a line of well-defined areas of convergence and divergence alternating along a line running roughly north-northeast through the west end of Lake Erie, while ensemble member 8 shows a much less well defined 'band' of smaller convergence areas at roughly the same position, with little moisture divergence. Note that all 20 of the ensemble members are presented in Appendix C, but for brevity of the current study, we will only present a few of the ensemble members in the main body to illustrate points we wish to draw attention to, as needed.

As we noted in our Introduction, areas of low level moisture flux convergence correspond fairly well with areas of low level mass convergence, so areas of moisture flux convergence should correspond with areas of upward air movement, and areas of divergence with areas of downward air movement. One could surmise then that ensemble member 6 simulated more well-defined downdraft cells, and stronger, more well-defined updraft cells. There is therefore quite a significant degree of difference between the ensemble members with respect to the strength and formation (and as we will soon discuss, position) of the generated squall-line.

Further, as we have noted in our Introduction, one question of importance for a generated ensemble is whether or not the ensemble is under-dispersive. As we have noted, a poorly generated ensemble which has too high a degree of correspondence between ensemble members will tend to be under-dispersive. As we have also noted, it is a non-trivial matter and a subject of considerable research to generate a well behaved ensemble, not subject to this and other deficiencies. Thus we arrive at one of the aims of the present thesis - to prove that the method used by the current thesis to generate the ensemble used for this study gives an ensemble that is sufficiently dispersive, and with the actual observed state falling within the ensemble. This issue will be covered in more detail below.

2.3.2 Squall Line Positions As Determined by Extreme Moisture Flux Convergence Regions and High Temperature Gradient Regions: How These Regions Relate in Position to the Actually Observed Squall Line

In order to answer this question of dispersion then, we consider again some of the plots that were obtained for moisture flux divergence discussed above, as well as corresponding plots of high temperature gradient values – which as we noted in our Introduction, are highest at the squall line boundary. Thus, the same areas which show high moisture flux convergence should also be those areas which show high temperature gradients as well – those areas with significant convective activity. For clarity in the temperature gradient plots, we have isolated those areas which have absolute temperature gradients more than 50% of the domain (2nd domain) peak absolute temperature gradient value. Each 'object' obtained (indicating a contiguous region of high temperature gradient value) corresponds to a high absolute temperature gradient – and should thus in theory correspond fairly well to squall line position. We will show in the following figures one example where the simulated squall line fell to the east of the actually occurring squall line (Ensemble member 4 - Figures 28 and 29) and one to the west of the squall line (Ensemble member 17 - Figures 30 and 31). It should be noted however, as could be expected from the westward biases of our 1km and

3km resolution simulations from first phase of this study, that there was also a significant westward bias of the ensemble members. Thus, we should expect a larger fraction of our ensemble members to have squall lines which fall to the west of the observed squall line, as opposed to the east. However, we should also expect for a well designed, non under-dispersive ensemble, to have some number that would fall to the east for a large enough ensemble – this would be a necessary condition for a sufficiently dispersive ensemble.

Note that in Appendix B the squall lines defined using this 'object-based' are plotted for the 20 ensemble members. While the 'objects' that correspond to the high temperature gradient areas are in fact determined objectively, as per the method outlined above, the linear extent of the squall lines, determined by choosing those two 'objects' which define the squall lines ends, were determined subjectively, and are thus admittedly subject to some interpretation. The resulting squall lines are presented in these plots in Appendix B. Also presented in Appendix C are the plots of 2400Z moisture flux divergence for the 20 ensemble members.

There are several things to note from Figures 28 to 31 for ensemble members 4 and 17. First, in both cases, when an area has a squall line position expected as

a result of high levels of moisture flux divergence/convergence, there is also a high level of temperature gradient as well at roughly the same position.

To summarize findings from the figures in Appendix B and Appendix C, we have the following results for estimated squall line positions (see explanations both above and in Appendixes B and C), both to the east of the observed squall line, and to the west of the actual squall line, which we can see in Figure 11 above, and which falls about 80km east of the western end of Lake Ontario at 2400Z:

Squall Line Position Relative to Observed Position[12]	Moisture Flux Divergence Squall Line Position Ensemble Members[13]	High Temperature Gradient Squall Line Position Ensemble Members[14]
West	1,2,6,7,9,10,12,13,14,16,17,18,19,20	1,2,6,7,8,10,12,13,16,17,18
East	4,15	4,9,15
Same Position	3,5,11	3,11
Uncertain/Indistinct Squall Line	5,14,19,20	8

Table 2: Positions of simulated squall lines relative to observed position using moisture flux divergence and high temperature gradient plots at 2400Z

12 The observed squall line position is simply the position as seen in Figure 11 – the linear cloud band running about 100km east of th western end of Lake Ontario.

13 This line is estimated at the best fit through the high moisture flux divergence/convergence linear band which runs roughly north/south through most of the plots of Appendix C. Of course, such an estimate is subjective, but in most cases, the band is fairly clear.

14 The squall lines here are just the gray lines drawn through the plots in Appendix B – see above and Appendix B for an explanation.

As we can see from Table 2, there is a definite westward bias of simulated squall line position for both the squall line positions from the moisture flux divergence, and for those from the high temperature gradient, so clearly the two methods generally agree – although there are some slight differences. Again however, there were nevertheless a few ensemble members which did fall to the east of the observed squall line, so it would certainly seem that the observed squall line position did appear to fall as a plausible member of the ensemble. This, then lends support to the contention that the ensemble used for this study satisfies one the main criteria of a 'good ensemble' (Kalnay, 2003) – that the observed 'cloud' of ensemble members encompasses the true state as well. Nevertheless, our bias/phase lag is exactly the same one mentioned by Kong et al (2007) and Hanley et al (2011) discussed in the Introduction, and is of comparable size to the bias noted by Hanley et al in their study (they reported a 50-150 km phase lag error for their ensemble members, while our ensemble mean had approximately an 80 km phase lag error, as did the high and low resolution simulations).

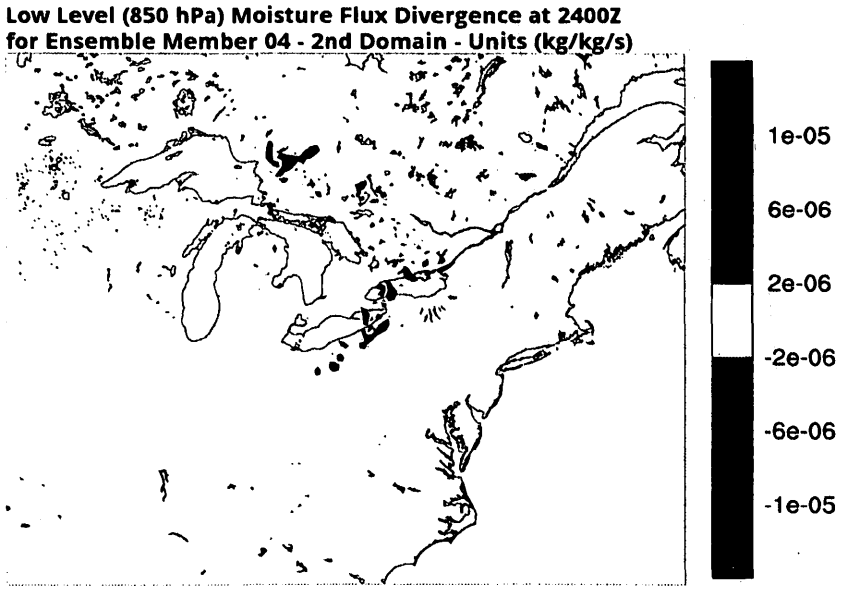


Figure 28: Low level moisture flux divergence for ensemble member 4 at 2400Z

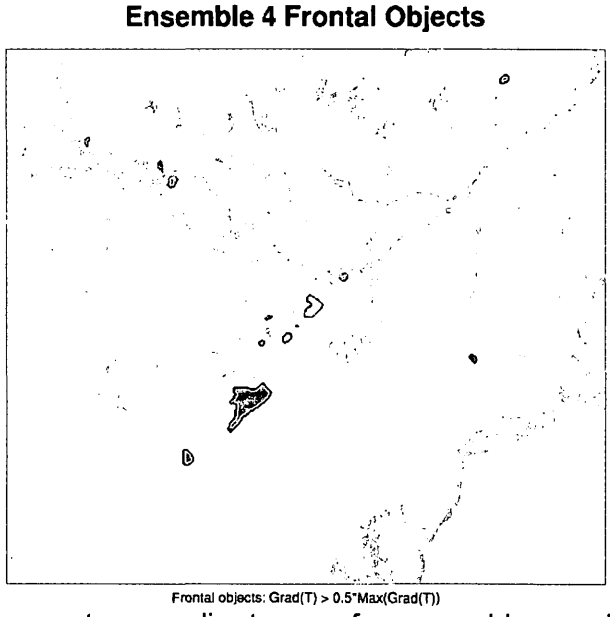


Figure 29: High temperature gradient areas for ensemble member 4 at 2400Z

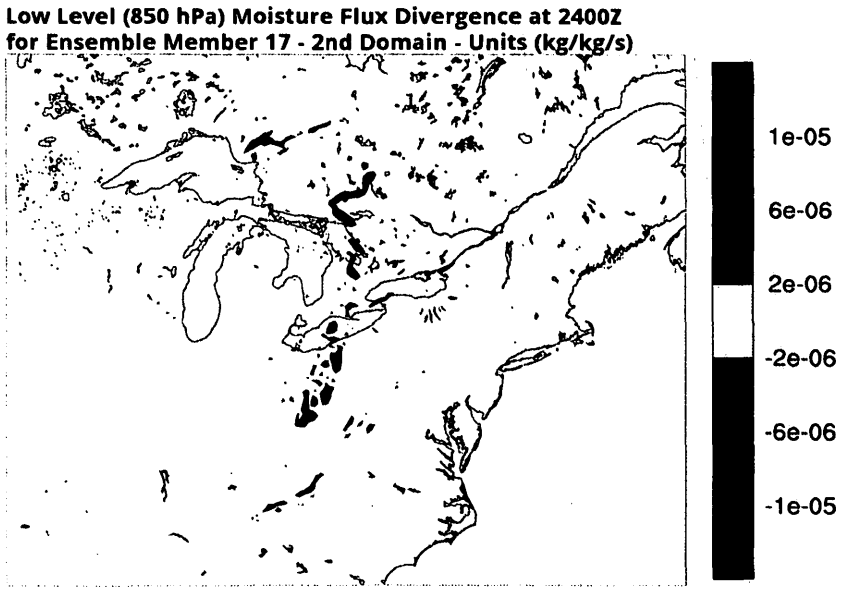


Figure 30: Low level moisture divergence for ensemble member 17 at 2400Z

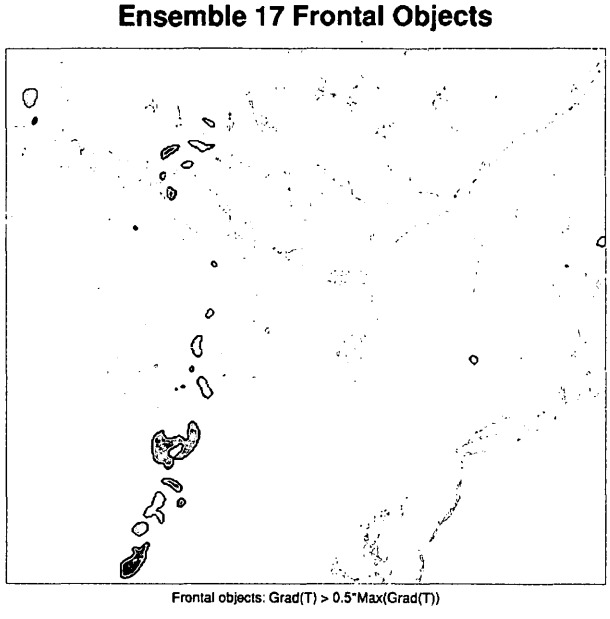


Figure 31: High temperature gradient areas for ensemble member 17 at 2400Z

There is however another reason, as noted in our Introduction, to be interested in the degree to which we have adequate dispersion of the ensemble members – again, that they diverge enough as the simulation progresses to envelop the true atmosphere state. As has been hypothesized by a number of authors (Hamil and Colluci, 1998; Stensrud et al., 1999; Wandishin et al., 2001), under-dispersion is believed to be a major cause of the inability of the ensemble spread to correlate accurately with the skill of the ensemble mean. Given that our ensemble does indeed manage to capture the actually occurring state, we therefore have further evidence that our method for generating the ensembles provides an ensemble spread that will have the ability to reflect accurately the skill of the ensemble mean forecast.

Thus, having shown that the ensemble we have created is not under-dispersive, we now turn to two of the most widely used products of the ensemble, the ensemble mean and the ensemble spread.

2.3.3 Vertical Velocity – Ensemble Mean and Ensemble Spread Relationship

We will consider the ensemble mean and ensemble spread here of mid level (eta=0.5) vertical velocity at 2400Z. The meteorological phenomenon which is the

focus of this study is of course the squall line/supercell system discussed in the Introduction, so it would be most informative to study variables which are most indicative of squall line/supercell activity, and vertical velocity is of course one of the most informative of these variables, as high vertical velocities are naturally associated with severe convective activity.

With this in mind we consider first Figures 32 and 33 which give the ensemble mean and ensemble spread vertical velocity at level 13 (approximately eta=0.5) at 2400Z. We will note of course immediately that for the ensemble mean vertical velocity plot (Figure 32), extreme values of the mean occur at the position of the squall line predicted by our high and low resolution simulations of the first part of this study. Thus, clearly the ensemble mean predicts the squall line at the same position as the high and low resolution simulations.

One should also note that the ratio of mean vertical velocity to spread vertical velocity (the signal to noise ratio), is on the order of 1. As we noted above, a low level of this ratio in an area where the ensemble mean is significantly greater than zero[15] is indicative of a higher degree of unpredictability in our simulation . Values along the squall line are on the order of a value of 1 for these vertical

15 The stipulation that the mean must be greater than 0 is important, otherwise for low ensemble mean areas the signal to noise ratio might be small simply because there is little vertical velocity in any of the simulations, which has nothing to do with predictability

velocity plots. This of course is right at the lower limit of what we have chosen to be representative of a reasonably predictable phenomenon.

Thus, using this criterion, the vertical velocity mean and spread plots can be argued to offer some evidence that the processes along the squall line appear to be somewhat unpredictable in nature.

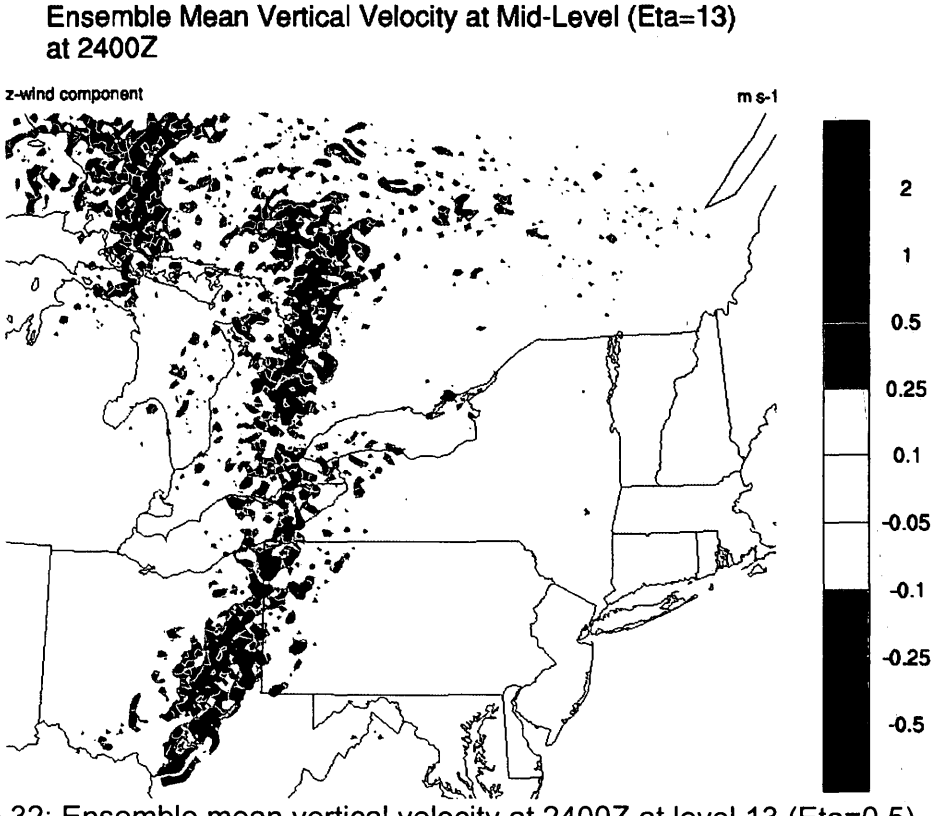


Figure 32: Ensemble mean vertical velocity at 2400Z at level 13 (Eta=0.5)

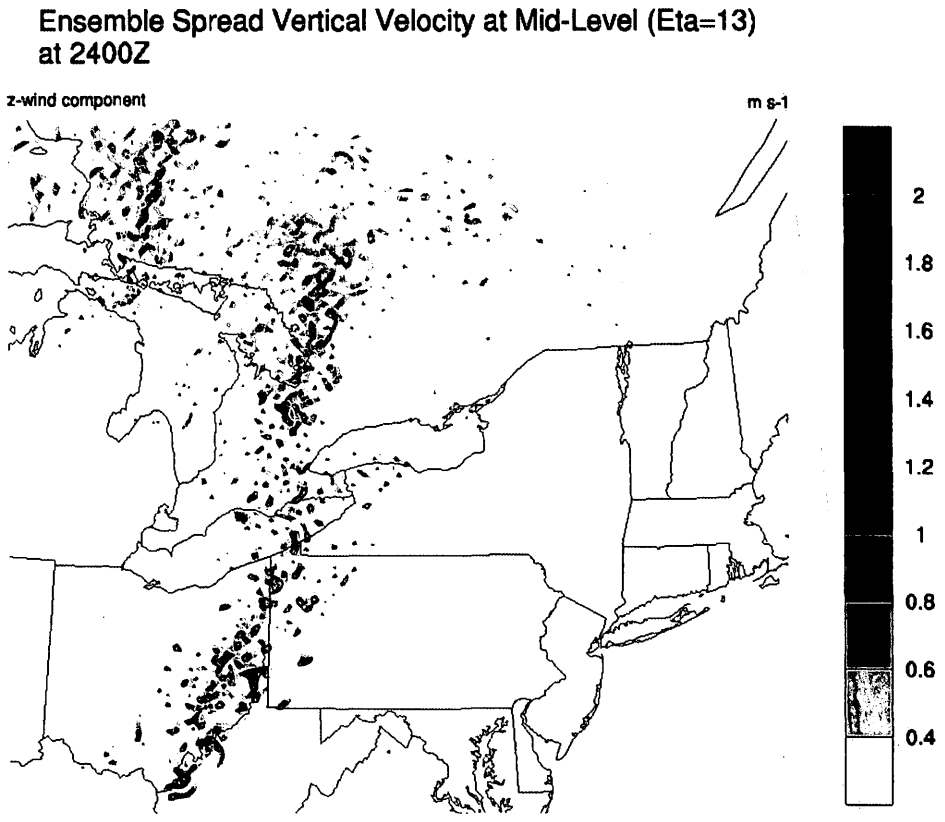


Figure 33: Ensemble spread vertical velocity at 2400Z at level 13 (Eta=0.5)

It is important to note here that the unpredictability that is being discussed here is being manifested as high values of ensemble spread (or low values of signal to noise ratios for appreciably non-zero ensemble mean values) for vertical velocity. Note that these high values of vertical velocity spread are distributed in very concentrated areas along the squall line, as evident in Figure 33. Thus, the unpredictability that we are dealing with here is an unpredictability for features in the mesoscale. This is important, for we will see later in Chapter 3 that

predictability remains significant for the larger synoptic scale features of the storm, specifically the squall line's position. Thus, issues of scale and their relationship to predictability (an important theme of this study) are evident in Figures 32 and 33 as well.

Note also that peak values of mean vertical velocity are only on the order to 2 m/s along the squall line – well under the 16.6 m/s and 19.1 m/s peak vertical velocities obtained for the high and low resolution simulations respectively in the first part of this study. Thus, it would appear that the ensemble mean has had the effect of smearing the peak values over a larger area, exactly the ensemble mean smearing effect noted in our introduction.

2.3.4 Ensemble Spread and Predictability of Rainfall Rate

Let us consider the ensemble spread of the 2400Z rainfall, given in Figure 34. As can be seen in this figure, the spread is highest along the squall line, as expected, since nonlinearity should be highest here. We might ask however, how indicative is this spread of the expected forecast error along here.

As we noted above, anomalously high values or low values of spread show better spread-skill relationships than values of spread which tend to fall in the middle, so we might expect that along the squall line, there would be a good correlation, and thus we could surmise a high degree of uncertainty along the squall line.

Houtekamer's theory allows us to quantify this correlation however. Using equations 13 and 14 above, we can estimate the β parameter for the ensemble spread distribution of Figure 34. Using the data from this distribution, we calculated the β parameter for all values of spread above the following cutoff thresholds[16] using equations 13 and 14 as follows:

Spread Threshold (mm/hr)	Calculated β Parameter
0.0	2.057
0.2	2.165
0.5	2.120
1.0	2.053
2.0	1.961
5.0	1.790

Table 3: Calculated β parameters for various cutoff thresholds of ensemble spread of rainfall at 2400Z

16 We use a cutoff threshold because we are interested in the behavior of spread along the squall line, where presumably our predictability is lowest, and hence the spread is highest. So by using the threshold, we exclude the low values of spread in the areas away from the squall line. We see from our results however, that for reasonable cutoff values, our β parameter is fairly insensitive to our choice of cutoff threshold.

As we can see from Table 3, the β parameter calculation is fairly insensitive to our cutoff threshold, so choosing a threshold of 0.5 mm/hr seems appropriate. This insensitivity to the threshold indicates that the behavior of the variation of the spread is driven by the higher values of spread – those spread values along the squall line.

Choosing a representative value of β of 2.12 then, we can calculate an expected correlation between spread and forecast error from Equation 12, giving a correlation of ρ=0.796. This is a reasonably high correlation between ensemble spread and forecast error, so we can be fairly certain that our high spread values along the squall line are indeed indicative of a higher forecast error of the mean. To see how significant these high spread values are, it is important to again look at the variation of the ensemble mean to the spread – the signal to noise ratio, for rainfall.

Figure 35, gives the ratio of the ensemble mean rainfall to the ensemble spread for all points which received more than 0.5 mm/hr of rain at 2400Z. As noted above, where this ratio is small in those areas where the ensemble mean is significantly greater than zero, the noise or variation in rainfall is greater than the mean, meaning there is more uncertainty in the mean than in those places where

this ratio is higher. It is evident from this figure that along most of the squall line, the signal to noise ratio is on the order of 0.5 or less, so we should expect that predictability would be low here. Again, we must note that the predictability that we are referring to here relates to the scale of the phenomenon being studied – squall line associated rainfall distribution – which is mesoscale in character. Thus, our β parameter results inferred from our distribution indicate again that first that our spread skill relationship is fairly high (ρ=0.796), and therefore the high signal to noise ratios evident in Figure 35 indicate a true lack of predictability for mesoscale phenomena such as rainfall rates. These findings of low predictability along the squall line for rainfall of course corroborate our similar findings from our vertical velocity results.

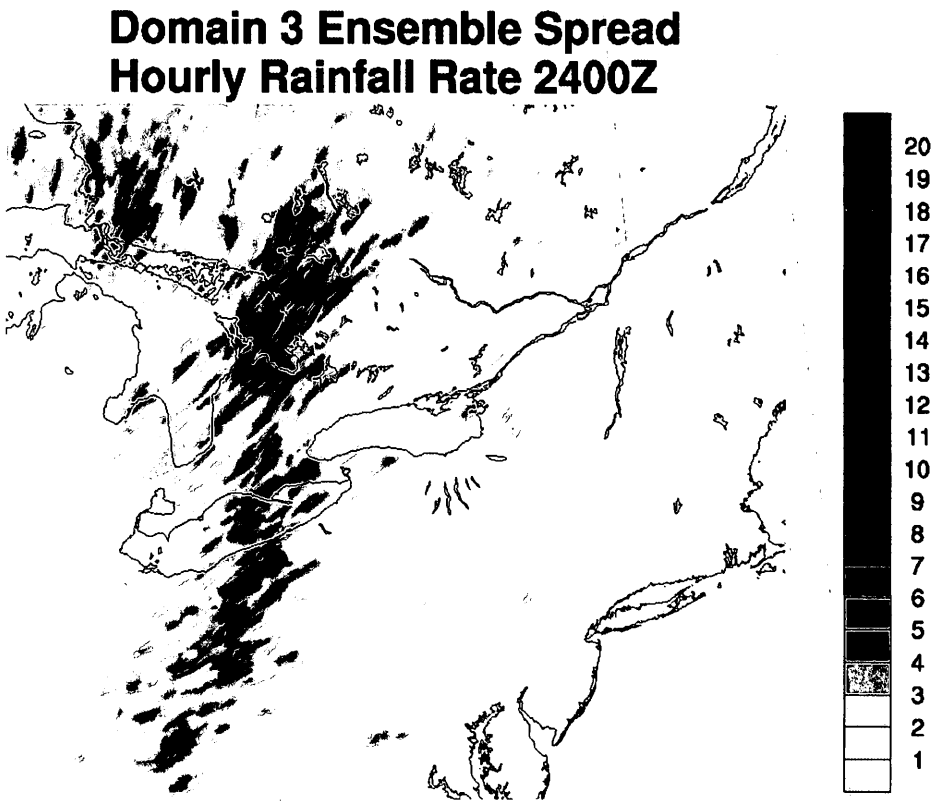


Figure 34: Ensemble spread hourly rainfall at 2400Z (mm/hr)

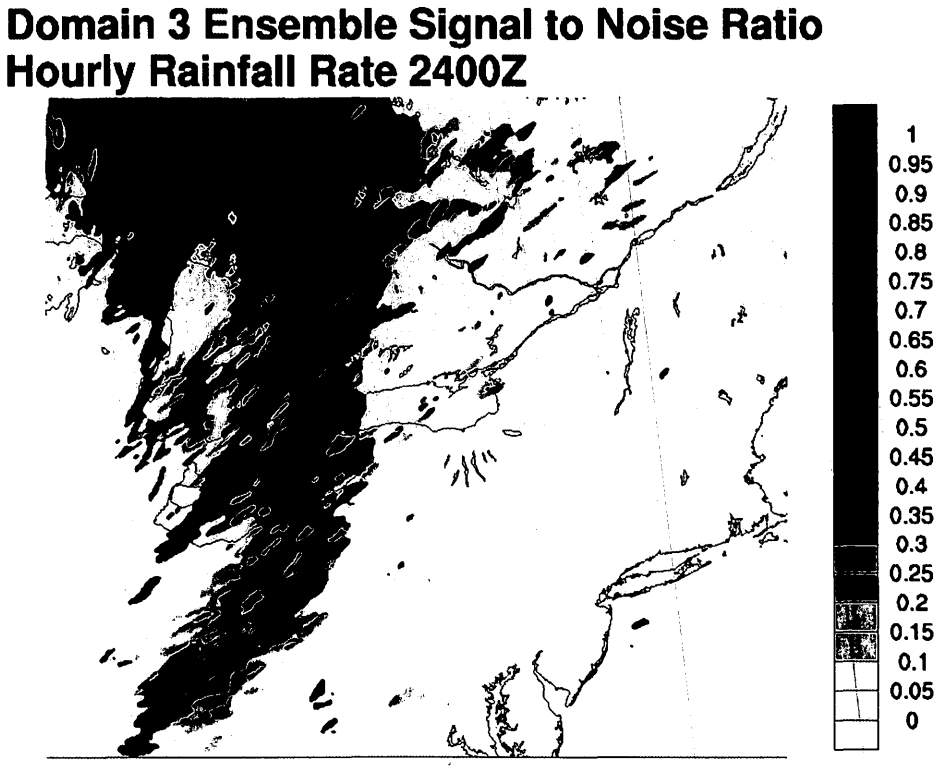


Figure 35: Ensemble signal to noise ratio for hourly rainfall at 2400Z (mm/hr)

2.3.5 Relative Predictive Abilities of the Deterministic, Ensemble Mean and Rescale Mean Distributions – Hit Rates and Root Mean Squared Rainfall Rate Errors Relative to Observed Values

In order to get a more quantitative understanding of the relative performance of the high resolution deterministic and ensemble mean distributions, we will consider the hit rate[17] of the rainfall rates over the period of simulation. Shown in Figure 36 is of the hit rate for both the ensemble mean distribution plotted against the hit rate for the deterministic simulation for the 1 mm/hr rain threshold. Following this is a similar plot in Figures 37 for the 2.5 mm/hr threshold, and a plot in Figure 38 for the 5 mm/hr threshold.

Clearly, as can be seen from these figures, the ensemble mean compared favorably with the deterministic simulation. Indeed, both distributions appeared to follow each other's performance fairly closely over the 24 hour simulation period.

One might surmise that this performance similarity has a lot to do with the similar westward bias of squall line position for the 2 simulations, since as noted above, both simulations showed a similar predicted position for the squall line. Naturally, having a bias in squall line position, along which a large proportion of the rainfall

17 The hit rate is defined as the ratio of the number of cases where observed values and predicted values were both above or below the threshold – in other words they agreed about rainfall being above or below the threshold - to the total number of cases in total.

is distributed, would have a large effect on the hit rate for rainfall. Thus, any one could argue that the differences between the three distributions are more than offset by their similar westward bias, making any meaningful quantitative comparison difficult.

A similar comparison was made with rainfall rates derived from radar reflectivity data. In this case, the rainfall rate product derived from radar reflectivity values were taken from the Buffalo Airport radar facility data set at 2400Z, and compared against the predicted rainfall rates from the high resolution deterministic and ensemble mean distribution. The results are presented below in Table 4:

Model	Root Mean Squared Rainfall Rate Error As Compared to Buffalo Radar Derived Data
High Resolution Deterministic Model	8.56 mm/hr
Ensemble Mean	7.68 mm/hr

Table 4: RMS error of predicted rainfall rates at 2400Z as compared to Buffalo RADAR data derived rates

We can see that from this result the ensemble mean appears to perform better than the deterministic mode. One might expect this, given the phase error of the

squall line position – with the ensemble mean having smearing of the rainfall pattern over a more distributed area than the deterministic distribution, the phase error will have less of an impact for a smeared distribution. The deterministic distribution however, was not able to benefit from this smearing effect, and so with its large phase error, had the highest RMS error. Naturally we have to be careful about drawing any conclusions from the above results however, as we are looking at only a single case.

It has to be admitted then that these results are somewhat inconclusive – the large phase error which was observed for the deterministic and ensemble mean distributions makes any meaningful comparison difficult, since the variations between the three are minimal in effect compared to the effect of their common phase error.

2.3.6 The Relationship Between the Scales of the Modes of Variation and Predictability

We come now to an issue that, as we discussed in the Introduction, is an important issue for this study - the contention that the smaller scales of variation, such as those relating to convective modes, are inherently less predictable than the larger scales of variation, such as those relating to baroclinic modes.

We can see this relationship more concretely if we consider Figures 47 and 48, which show the growth of 3rd domain root mean squared deviation from the ensemble mean of 2m temperature perturbations and 500 hPa geopotential perturbations respectively.

We can see from Figure 39 that the temperature perturbations drop from their initial value to a lower value at around 2 hours, before gradually increasing in size until they reach a maximum value at around 16 hours, after which there is no further growth but just a variation around this value. Of course this is only a single case, and so these results cannot be taken as conclusive, but the results are certainly similar to the results of Hohenegger and Schar (2007) mentioned above, who found that for low level temperature perturbations, there will be saturation of low level temperature perturbation heights between 12 and 28 hours. Further, they also noted that initially, temperature perturbations will drop to a low value after several hours, after which there will be growth to this point of saturation. This initial drop corresponds to the initial perturbations initially aligning themselves with the directions of dominant growth[18] - the directions of the dominant singular vectors - after which growth will follow these dominant singular

18 What happens is that the portions of the perturbations not aligned with a singular vector will quickly dissipate as gravity waves, thereby causing an initial drop in perturbation magnitude, while the portion of the perturbation aligned with a dominant singular vector will grow as expected for a dominant mode. Thus, the pattern resulting is as seen in our results, an initial drop of perturbation amplitude, followed by a growth of perturbation amplitude over time.

vectors (Kalnay, 2003). So the results of the ensemble temperature perturbation growth are indicative of the dominant modes of fast growing, unpredictable convective error growth. The 12-16 hour saturation time will be important as well, which we will discuss further below.

Figure 40 is less clear however, since the 500 hPa geopotential heights necessarily grow much more slowly than the low level temperature perturbations, and tend to saturate as we noted above between 6 and 10 days (Hohenegger and Schar, 2007). Thus, our 24 hour study period is much too short to see this effect. Nevertheless, we can see a distinct drop in perturbation height between 0 and 16 hours, after which their appears to be gradual growth of the perturbation heights with time (presumably after alignment with the dominant synoptic modes), although of course this growth is truncated soon after it began at our 24 hour cutoff time. These results, while again noting that we have only a single case here and therefore no definite conclusions can be drawn from them, are nonetheless suggestive of the more slowly growing, baroclinic modes of variation noted by Hohenegger and Schar (2007), and indeed appear to match those of their study.

It would appear that our ensemble was able to successfully capture both the

unpredictable, quickly-varying modes of variation suggestive of convection, as well as the more slowly varying and more predictable modes of variation associated with baroclinic activity. This conclusion can be made because, as Hohenegger and Schar (2007) have noted, baroclinic activity is captured by the 500 hPa geopotential RMS perturbations from the ensemble mean, and convective instability is captured by the low level temperature perturbations from the mean. As we mentioned in the Introduction, ensemble perturbations that start too-small in size quickly saturate (after around an hour) at unrealistically small perturbation values – and hence, as noted above, are under-dispersive. The slower growth of our perturbations evident for 2m temperature and 500 hPa geopotential heights would indicate then that our simulation's ensemble does not suffer from this problem. Thus the approach that we have taken to generate our higher resolution ensemble (from an existing lower resolution ensemble) appears to be successful in capturing both the synoptic and convective modes of variation, and should therefore be considered in this respect a successful ensemble.

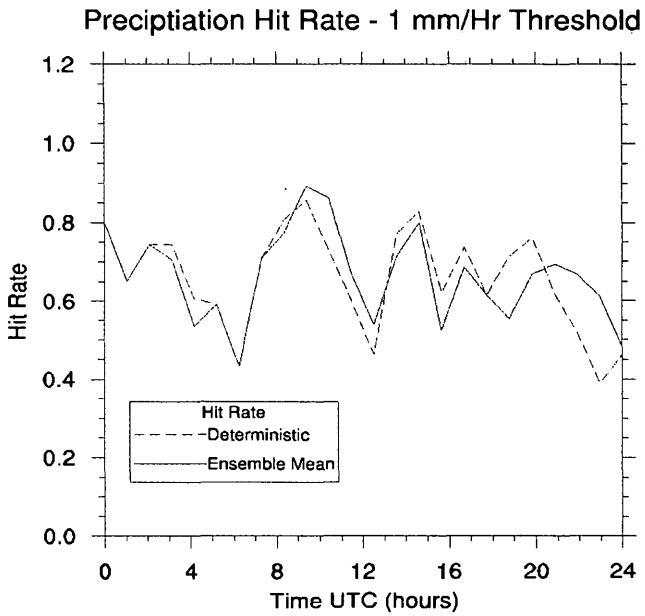


Figure 36: Hit rate for ensemble mean distribution – 1 mm threshold

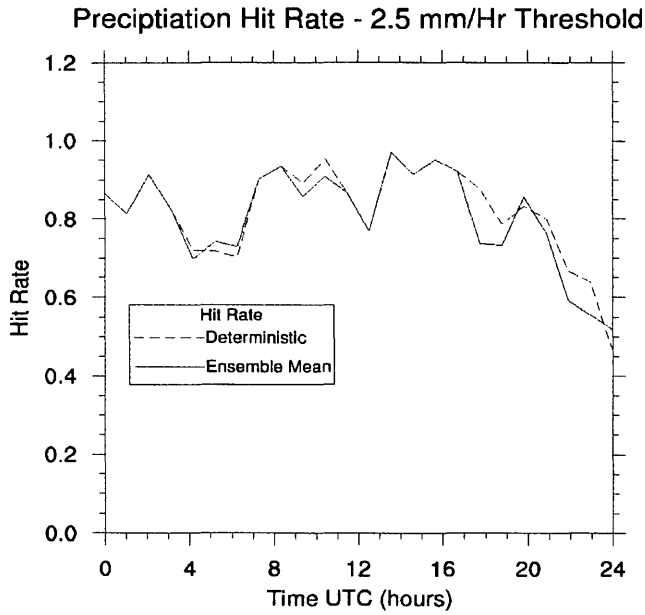


Figure 37: Hit rate for ensemble mean distribution – 2.5mm threshold

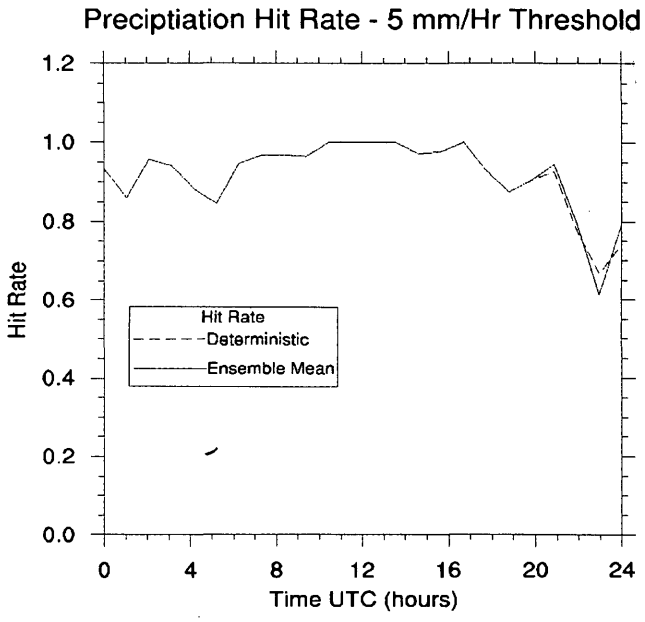


Figure 38: Hit rate for ensemble mean distribution – 5mm threshold

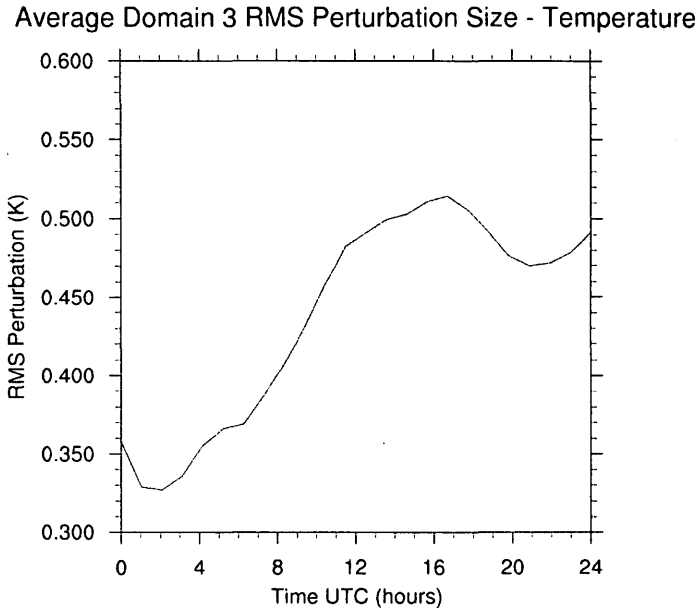


Figure 39: RMS perturbation size growth for domain 3 2m temperatures

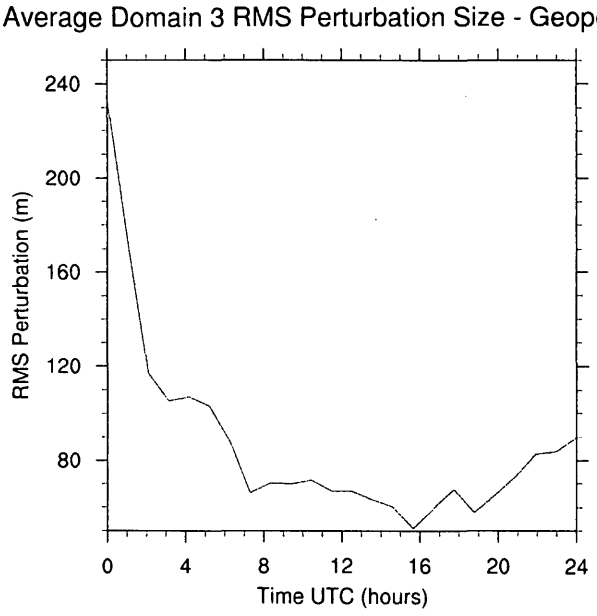


Figure 40: RMS perturbation size growth for domain 3 500 hPa geopotential

2.4 Conclusions

We have seen then that the ensemble mean has the same bias of squall line position that was seen in both the high and low resolution simulations – a general lag of the position of convection in relation to observations. Therefore, the use of the ensemble mean has not improved our simulation in this respect – westward bias is still a problem. It is worth noting that this is the same bias that has been seen in two studies of convective scale ensemble modeling discussed in our Introduction, which was hypothesized to be related to a delay in convection initiation.

However, we have also shown that the ensemble members are not under-dispersive – the true state of the atmosphere appears to be a plausible member of the ensemble, albeit somewhat biased to the east. Thus, at least in terms of dispersion, our ensemble generation method outlined in the Methodology appears to be a successful means of obtaining initial ensemble members and ensemble boundary conditions. This would lend evidence to the contention that the outlined methodology for creating a set of ensemble perturbations from a preexisting ensemble is a viable approach. Also, the results of this study allay to a certain extent the concerns raised in the Introduction regarding the use of low

resolution ensemble data for the creation of a higher resolution set of ensemble perturbations. It should be noted however that this study is only a single case, so no definitive conclusions can be drawn from its results. Nevertheless, our ensemble generation approach – creating a high resolution ensemble quickly and cheaply by adding the interpolated perturbations obtained from a pre-existing low resolution ensemble to a high resolution control – appears to have worked, and appears therefore to be worthy of further study in the proposed follow-up study employing ensemble data assimilation methods.

As we should have expected, unpredictability is highest along the squall line, with the signal to noise ratio either 1 or below along this line for vertical velocities, potential vorticities, and rainfall. Thus, we cannot expect that measurements of these quantities will have any real possibility of being accurately forecasted along the squall line – especially since there appears to be a great variability of these quantities along the squall line with domain maximum and minimum values occurring in very close proximity. Therefore, forecasting the genesis or evolution of specific structures in the squall line would be questionable at best.

While this modeled unpredictability is manifested as an increase in spread among our ensemble members, we have seen for the rainfall predictions that

there is a high likelihood, based on Houtekamer's theory, that this unpredictability also indicates an actual inherent unpredictability of the true atmospheric state.

We have also seen that there appear to be at least two scales of unpredictability relevant to our results – a smaller scale of unpredictability related to low level temperatures that reaches saturation (and thus loses all predictability) around 16 hours of simulation time, and a larger, synoptic scale of unpredictability related to mid-level geopotential heights that reaches saturation at time scales well past our 24 hour simulation period. Thus, those phenomena that relate to the larger, synoptic scale should remain predictable over the time scale of our study. Given that the ensemble mean and both the high and low resolution simulations all showed the same positioning of the squall line, there is a strong argument to be made for the fact that the squall line position is therefore driven by the predictable larger synoptic-scale flow, rather than from the unpredictable smaller scale – since all three methods agreed on this position, adding indirect evidence to their predictability at this scale.

3 Convective-Scale Ensemble Sensitivity Studies

3.1 Introduction

Having shown then that there is some predictability after 24 hours of simulation time (at least for our baroclinic modes of variation), we come then to the third phase of this study. With at least some level of predictability now established, can we establish that this predictability is useful for informing us about squall line position? In other words, is their a significant correlation of squall line position with model variables at earlier times – correlations that at we have noted earlier, might possibly be exploited to advantage, by helping optimally target those areas where additional observations in data poor areas would maximally improve upon our simulation results using data assimilation?

3.1.1 Correlation of Input Variables with Output States (Sensitivity Analysis) and Targeted Observations for Data Assimilation

Predictability means of course that there is some positive correlation between our input state and our output state. This correlation is important, for, as we mentioned above, one method widely used today to improve forecasts is data assimilation. It is not the purpose of this study to discuss data assimilation in too much detail, since data assimilation was not actually employed in this study, but it is worth mentioning that data assimilation improves upon a forecast by incorporating observational data into a forecast so that the resulting product, the analysis, is the optimal interpolation between the model's prediction and the observations in a least-squared-error sense, given a known covariance of model errors, and a known covariance of observational errors.

Usually, the observation error covariance is known fairly well, but the model error covariance, also called the background error covariance, is not as well known, and it is generally a non-trivial matter to determine this matrix. One group of methods frequently used to address this matter are called ensemble data assimilation methods, which use the covariance among the ensemble perturbations to estimate this model or background covariance.

So, for this type of data assimilation, it is helpful to have a well behaved ensemble – one that incorporates the true state as a 'plausible member of the ensemble' and at the same time is not under-dispersive. Thus, the results of the second phase of this study, which indicate that the ensemble has both these properties, would indicate that the ensemble generated for this study might be useful for a further improvement of our simulation's results using data assimilation. This is important, for as we have seen in the first phase, there was a significant westward phase error in the position of the squall line that was not removed by ensemble averaging, and remained the same in both the low and high resolution simulations. Given that all 3 simulations showed roughly the same error, we can surmise either two causes – either an error in the initial conditions (perhaps a large scale under-estimation in the background eastward wind field), that caused the squall line end up to far to the west, or a model deficiency that resulted in too slow a propagation of the squall line. However, given that the squall line position appears to require the presence of the low level jet for propagation (a conclusion from the first phase of our study), and that this low level jet results from the action of our large scale flow, it would seem much more likely that our phase error relates to an error in our initial large scale background flow, since propagation appears to require the presence of this low level jet. In either case however, data assimilation would improve our results by pulling the

simulation closer to reality.

One need only give the matter a little thought however to conclude that not all observations at earlier times will be equally effective to improve model predictions when assimilated. For instance, on purely intuitive grounds one would expect that observations taken along a squall line at some earlier time might have a larger impact on the predictions than observations taken well away from the squall line. This question has been addressed in some detail already in the literature (Ancel and Hakim, 2007). As we alluded to earlier, it is important to help locate these areas of maximum effectiveness for data assimilation when observations are not already present, as often there are no observations available where in fact they would do the most good. Thus, for a potentially catastrophic even such as a tornadic storm, knowing where to take additional radiosonde observations ahead of time to most effectively improve upon the model's predictions would be useful data indeed.

It has been found that ensemble sensitivity analysis is an effective means of finding these areas of maximum benefit for data assimilation (Bishop et al. 2001; Hamill and Snyder 2002) . Further, it has been noted by Ancel and Hakim (2007) that ensemble sensitivity analysis is effective in pinpointing those sensitivities

that relate to synoptic scale features (which we have seen may drive the squall line phase error of our simulations), while adjoint-sensitivity analysis is more effective in finding sensitivities to mesoscale features. Indeed, ensemble sensitivity analysis was first investigated by Torn and Hakim (2008), who examined the linear relationship between the 24 hour forecast of an extra-tropical cyclone position and the forecast's initial conditions, and they found that the linear relationships uncovered were related mainly to synoptic scale differences (mainly phase differences) in the primary synoptic structures. Thus, ensemble-based data assimilation would seem very well suited to improving upon our simulation's results, given that we are looking to improve upon synoptically driven phase differences in our simulation.

Data assimilation is therefore believed to be the most promising means of improving upon the performance of the forecast of this study, and so as we mentioned above, further work is planned to see if improvements can be effected using data assimilation. However, before such work can be done, those areas which would offer the most observational benefit will have to be targeted, and naturally, for the reasons mentioned above, a sensitivity analysis is the most promising first step for such targeting. Hence, the third phase of this study will endeavor to begin this work with an investigation of such sensitivity studies for

the storm system of this study.

Before such studies are carried out however, a few technical points must be noted. The first of these relates to spurious correlations that can arise from such sensitivity analyses. Bishop and Hodyss (2007) have noted that small ensemble size can contribute to significant error in the values of correlations obtained from an ensemble analysis.

Thus, for this reason it will be necessary to first filter out such sample-size-related correlation errors. Before we discuss how we will do this, we must first discuss how the sensitivity analyses will be conducted.

3.1.2 Sampling Error of Correlation Results

First, correlations for some 'output' variable, which we will call V, will be correlated with some 'input' variable u(i,j,k,t), where i,j,k denote the location in space of the variable (the grid point indexes perhaps), and t denotes the point in time that this variable is considered. Note that V will be a single derived variable – such as zonal or meridional position of some feature of the squall line, and we will confine our interest to output variable values at 2400Z only. Our input

variable u however, as is evident from the indexes above, will be taken to be distributed both in space and time. So we will determine ensemble correlations of these time and space distributed input variables with our output variable, giving us a series of location dependent plots of correlations at different times, and these will help us target those areas (and times) which are the most correlated areas (and times) with our variable of interest. The correlations will be found using the following formula, noting that for simplicity of notation we have dropped the indexes on u but these are to be taken to be implied:

$$\rho(u,V) = \frac{\langle u'V' \rangle}{\sigma_u \sigma_V} \tag{18}$$

where

$$\begin{aligned} u' &= u - \langle u \rangle \\ V' &= V - \langle V \rangle \end{aligned} \tag{19}$$

and

$$\begin{aligned} \sigma_u &= \sqrt{\langle u'^2 \rangle} \\ \sigma_V &= \sqrt{\langle V'^2 \rangle} \end{aligned} . \tag{20}$$

Note that these averages are taken over the ensemble. Now, as we have mentioned, sampling error will be a concern for our study. In order to address this, error bounds on our correlations were determined using a Fisher transformation (von Storch and Zweirs, 2002). The Fisher transformation is used as follows – for those correlations ρ transformed by the Fisher transformation, which is given by:

$$z = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right) \qquad (21)$$

it is known that they are normally distributed with mean $z = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$ and with a standard error of $\frac{1}{\sqrt{N-3}}$, where N is the ensemble sample size.

With this known, one can determine error bounds on the correlation. For example, to find the lower bound on our sample correlation such that one can be 95% sure that the actual correlation is higher, we simply calculate z from our sample correlation, and then find out from a standard error function table where the 5% cumulative probability is situated, in terms of standard deviations, from the mean. Knowing this, we multiply this number of standard deviations by our

standard error $1/(N-3)^{1/2}$, subtract this amount from our determined z value, and then find the inverse this value using:

$$\rho = \frac{e^{2z}-1}{e^{2z}+1} \tag{22}$$

to get our 95% lower confidence bound on ρ. Of course, a similar method allows us to find our upper confidence bound as well.

3.2 Methodology

The first sensitivity analysis carried out was an analysis of the relationship between what we will term the rainfall weighted "center of gravity" (CG) of the rainfall at 2400Z over domain 2 (this being our output variable V as discussed in the Introduction), and the 850 and 500 hPa zonal and meridional winds, temperatures, relative humidities and potential vorticities (our input variables).

3.2.1 Rain-Weighted 'Center of Gravity' (CG) Correlations

We will have to explain of course what we mean by rain weighted CG. What this variable is is an attempt to mark the location of the "center" of the storm system by weighting each coordinate value (in the zonal and meridional directions) in the domain by the amount of rain falling at that location, and dividing by the total domain rainfall. Note that the equation used to find this CG is given as follows:

$$i_{avg} = \frac{\sum_{i=1}^{N_m}\sum_{j=1}^{N_n} r(i,j)i}{\sum_{i=1}^{N_m}\sum_{j=1}^{N_n} r(i,j)} \qquad (23)$$

where r(i,j) is the rainfall at grid point (i,j), and i_{avg} is the rain-weighted i value of the CG. Of course, j_{avg} is found in a similar way.

So shifts of the storm system either zonally or meridionaly will be manifested as changes in the zonal or meridional component of the CG. This CG is not a perfect measure however, since the rainfall is distributed also over areas other than the squall line, but it is believed that the analysis based on this output variable will be informative nonetheless. What will result from this analysis then is a set of graphs for correlations of the zonal position of this center of mass with 850 hPa and 500hPa zonal and meridional winds, temperatures, relative humidities and potential vorticities.

Further, in order to remove the sampling errors of our correlations discussed above, the 95% lower bound on the positive correlations will be calculated, and only those lower bound values which remain positive will be plotted as non-zero. Similarly, the 95% upper bound on all negative correlations will be calculated, and only those upper bound values which remain negative will be plotted as non-zero. Thus, what will remain in the plot will be areas of 0 correlation, with only a few areas for which we can be 95% sure that there is indeed either a positive or negative correlation between our input variable and either the zonal or meridional

component of our rain weighted CG - we call this the 95% reduced correlation in our plots and discussion.

We noted above however that the rainfall CG will not correspond perfectly with the location of the squall line. One reason we would tend to believe that this correspondence might not be perfect is again because rainfall will fall in other areas, such as the warm front which will tend to move northward as the squall line moves eastward. To deal with this issue, as second sensitivity analysis was carried out. In this analysis, an attempt was made to use an "object-based", more subjective method, to help pinpoint more definitively the squall line.

3.2.2 Temperature Gradient Object-Based Squall Line Correlations

This method involves finding all those areas that constitute a region of high 2m temperature gradient (in this case, high was taken to be 0.5 of the domain maximum temperature gradient), and then drawing a line through these high temperature gradient 'objects' which are determined by inspection to fall along the squall line. This of course is done subjectively as the position of the object based squall line is a matter of some interpretation. Examples of objects for

some ensembles through which these lines have been drawn have already been shown in Figures 29 and 31 above for our previous analysis of ensemble dispersion.

Shown in Figures 41 and 42 are these same two ensemble object plots with lines drawn through the objects that define their squall lines. As can be seen from these two figures, and those in the Appendix C for the other ensemble members, the length of the squall lines and their angles with respect to the vertical (for simplicity, all squall lines are assumed to be purely linear in extent) change considerable from member to member. Thus, in order to find some objective zonal characteristic with which they could be compared, it was decided to find the longitude where each object 'squall line' intersects with some given 'reference' latitude. This 'reference' latitude was taken to be the average mid-latitude point of the ensemble member 'object-based' squall lines, to ensure that intersection with these object 'squall lines' was maximally possible. The longitude of intersection of an ensemble member's squall line with this reference latitude was taken to be the best measure of the zonal position of the squall line – since, for example, taking instead the mid-point longitude for a very short 'squall line' which has its highest temperature gradients concentrated mostly along the northern extent of the line (and therefore has temperature gradient 'objects' only to the north) and

which also has a large tilt toward the horizontal, would give us an unrealistically eastward bias of mid-point longitude, or westward, for a southward dominated temperature gradient line. By taking the intersection of the *extended* object-based 'squall line' with a given reference latitude, we are able to resolve this difficulty and give ourselves a better idea of the representative longitudinal position of the squall line.

This intersection longitude then gives us our 'output variable' for our correlation. For this particular correlation study however, unlike the first study discussed above, we looked only at the correlation of this output variable with a single input variable – the latitude of the pressure low, the longitude of the pressure low, and intensity of the pressure low, and the mean zonal velocity at 00Z. Note then that this would give us only single numerical correlations for the 4 'input' variable considered. Naturally, equations 18 to 20 were used to determine these correlations.

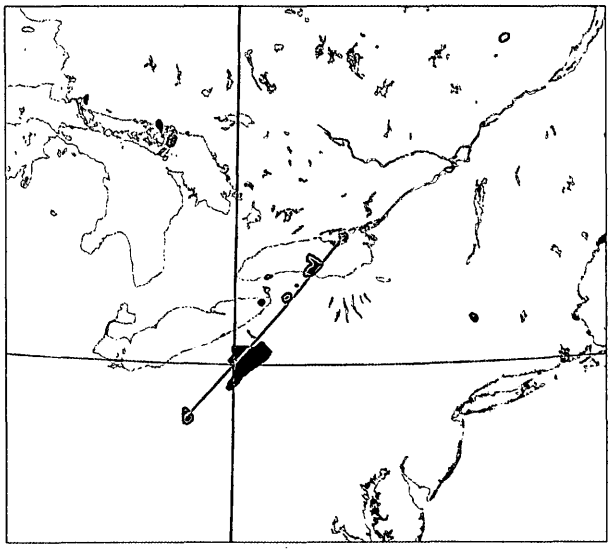


Figure 41: Ensemble member 4 squall line intersection with reference latitude

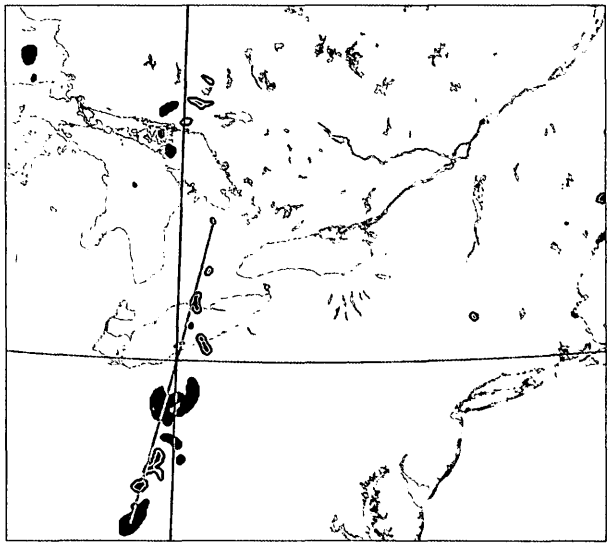


Figure 42: Ensemble member 17 squall line intersection with reference latitude

3.3 Results and Discussion

3.3.1 Rain Weighted CG Correlations

Shown in Figures 43 to 46 are the 95% reduced correlations (see the Introduction for an explanation of this term) of the 850 hPa zonal velocity with the rainfall-weighted CG longitude at 2400Z. These figures show these correlations at 1200Z to 2400Z over the 2nd domain.

If we look at Figure 46, we can see a large area of positive correlation of zonal velocity with CG longitude at 2400Z located right over the area where the squall line was predicted to be by the model, as well as to the east of this line over Lake Ontario. We can explain this correlation by noting that at 2400Z, a higher zonal wind in the background flow will tend to carry the squall line system further eastward, resulting in this positive correlation between zonal wind and CG longitude.

As explained in the introduction, the 95% reduced correlation means that we are 95% sure that the correlations are as high or higher than the indicated values in

Figure 46, so it is evident that there is a positive correlation between these variables, physical explanations aside. The fact that we can physically explain this relationship only makes the contention stronger, but in general having such an explanation is not a necessity, as the physical relationship for other variables may in fact be fairly subtle.

If we now track backward three hours to 2100Z (Figure 45), we can still see a positive correlation between the zonal wind speed and the CG longitude, but now this area of positive correlation is shifted westward – it is in fact tracking backward with the squall line system. Since this region has remained as a contiguous, integrated region of positive correlation, that tracks with a physical structure and ends up in a position for which we have a plausible and physically-based explanation of position correlation, we have no reason to expect that this prior positive correlation is not just as physically relevant and explainable.

Tracking back another 3 hours to 1800Z (Figure 44), we see that this region of positive correlation has increased in size and moved further westward.

Tracking back 6 hours more to 1200Z (Figure 43), we see that the positive correlation region has moved back further to the west and the south. Interestingly

however, we now have a region of negative correlation in the north, just north of Lake Huron. This is explainable however by noting that our low pressure system sits just to the west of the 2nd domain, between the areas of high and low correlation. With a stronger low, the cyclonic motion around this low will be stronger, increasing eastward zonal flow to the south (where our positive correlation area is) increasing westward zonal flow to the north (where our negative correlation area is). Thus, the pattern evident here could be interpreted as a positive correlation between cyclonic flow around our pressure low and the eastward position of the squall line, in other words, a positive correlation between the strength of the baroclinic disturbance at 1200Z and its eventual position at 2400Z, which is certainly plausible.

In any event however, physical discussions aside, the 95% reduction that we have effected on our correlations gives us a 95% certainty that these positive correlations are statistically significant anyway. Again, having a physical explanation helps, but such physical explanations are not necessary to establish our positive correlations – they have been established statistically.

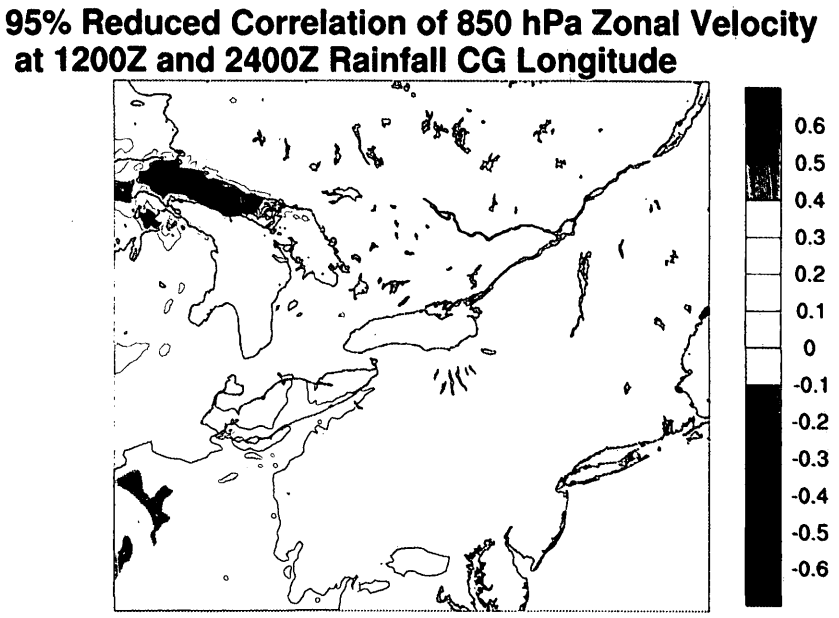


Figure 43: 850hPa 1200Z zonal velocity and 2400Z CG longitude correlations

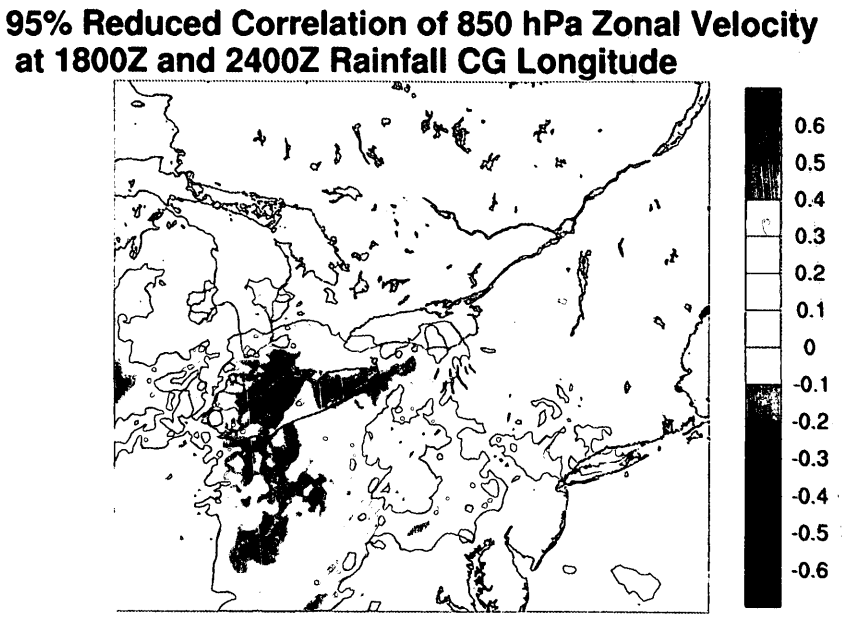


Figure 44: 850hPa 1800Z zonal velocity and 2400Z CG longitude correlations

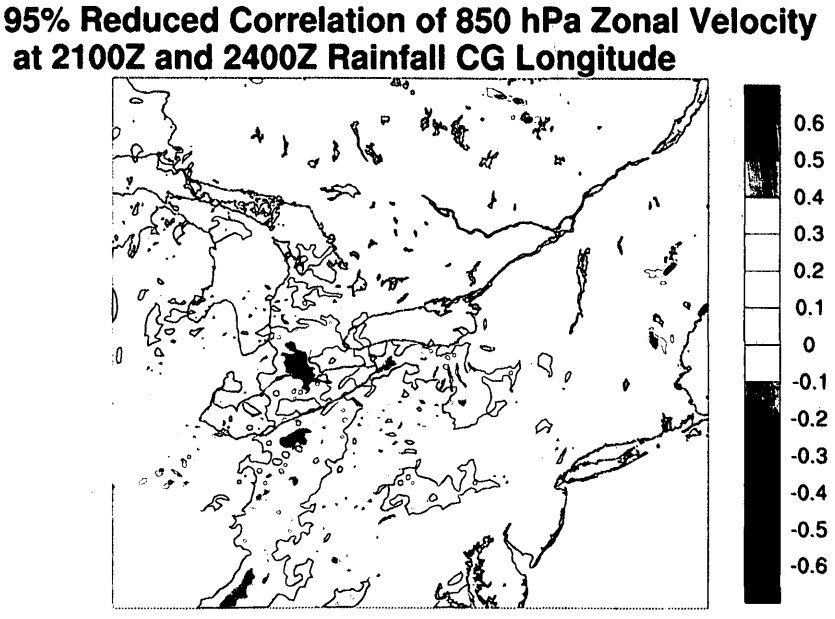


Figure 45: 850hPa 2100Z zonal velocity and 2400Z CG longitude correlations

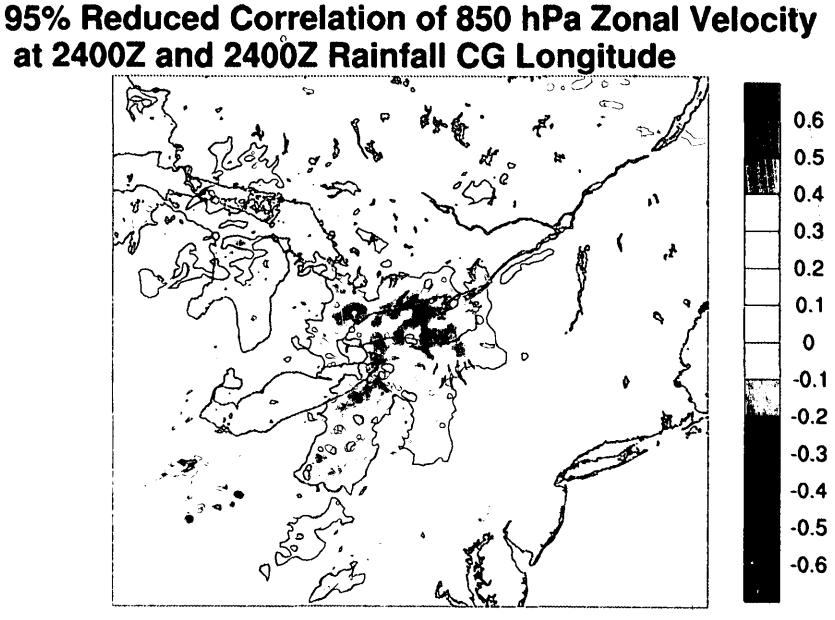


Figure 46: 850hPa 2400Z zonal velocity and 2400Z CG longitude correlations

Next, we will consider the correlations of 500 hPa humidity with 2400Z CG longitude. In order to understand the following correlation plots however, we will first have to consider the distribution of 500 hPa relative humidity along the squall line at 2400Z, Figure 47. We can see a fairly concentrated band of relative humidity here running along the squall line. What relationship might zonal shifts of this band of relative humidity have to the relative humidity/CG longitude correlation? To help answer this, consider Figure 48.

As we can see from the simplified explanation in this diagram, we can expect to have running on on either side of the relative humidity maximum band (along the squall line), negative and positive bands of correlation with CG longitude. In the center between these bands, we should expect a low or zero band of correlation. As we can see in the 2400Z plot of correlation for relative humidity, Figure 52, this was exactly what was seen. Thus physically, the 2400Z correlation of 500 hPa humidity with CG longitude makes sense. Note as well that the humidity/CG longitude correlation is not so simple once we get away from the squall line at 2400Z as there is no requirement that the variables be directly related like they are along the squall line at this time – the above arguments refer only to variations the squall line itself. Thus, the humidity distribution along the warm front (which was appreciable), may not have as direct a correlation with the

squall line position, so the large region of positive correlation to the north-east along the St. Lawrence is not as easily explainable.

As we step back 3 hours, we see in a situation similar to above, in Figure 51, that these correlation bands track with the squall line, and hence again appear to be physically meaningful. Back 6 hours at 1800Z, Figure 50, we have these bands tracking further to west. Back 12 hours, at 1200Z, Figure 49, we have only a single line of positive correlation running from the southwest to the northeast over the Great Lakes. The relationship between final squall line position and humidity position is not as direct over 12 hours, so it is probably not surprising that we do not see parallel bands of positive and negative correlation anymore.

However, we can see that even 12 hours beforehand, we can be 95% certain that the correlation between relative humidity and CG longitude remains significant, as there are regions where the correlation exceeds 0.6 even after we take the 95% lower error bound on correlation.

This is a significant point, for it is just this sort of long-term positive correlation that would indicate those areas that would be most effective as data input regions for improvements to the model using data assimilation.

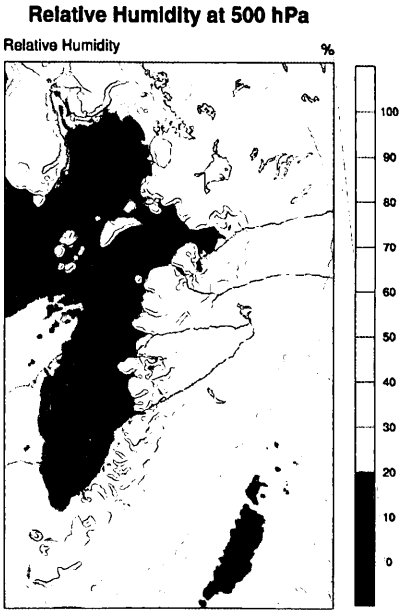


Figure 47: Relative humidity at 500 hPa along 2400Z squall line

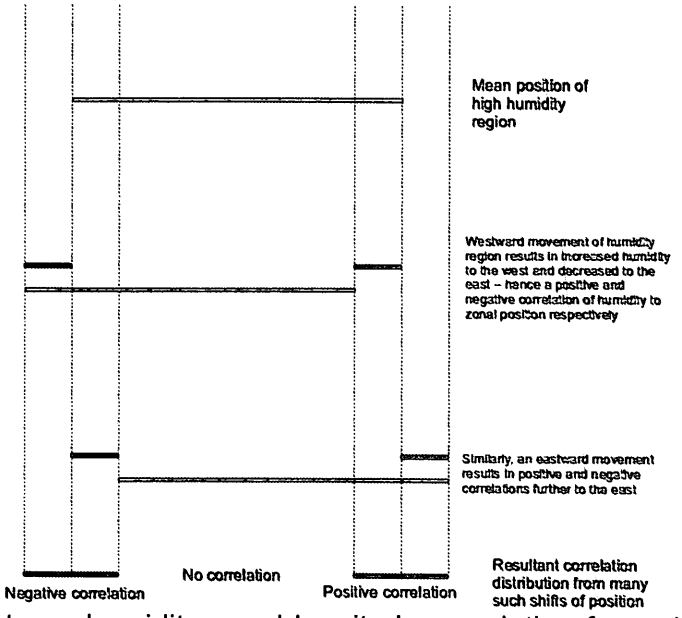


Figure 48: Effects on humidity-zonal longitude correlation from shifts in humidity position

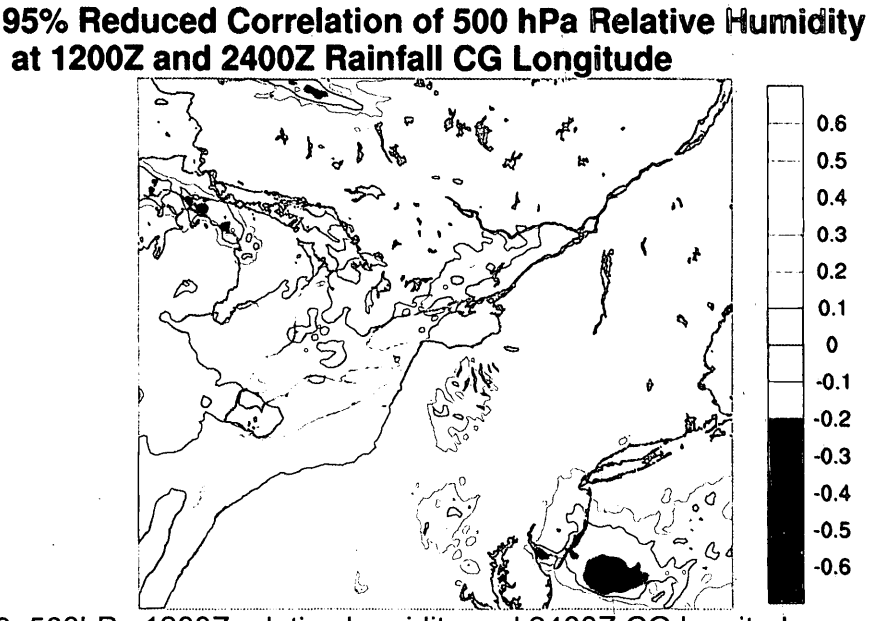


Figure 49: 500hPa 1200Z relative humidity and 2400Z CG longitude correlations

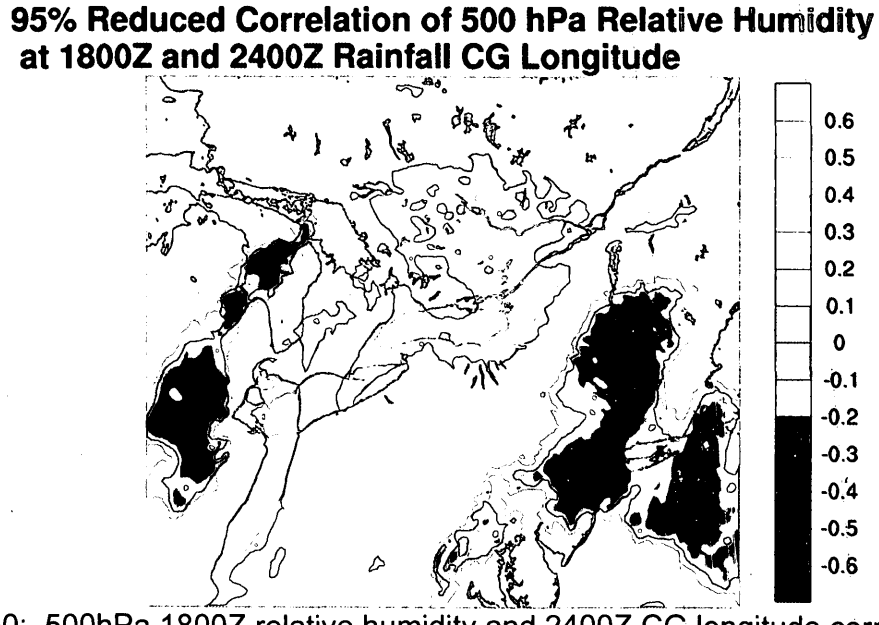


Figure 50: 500hPa 1800Z relative humidity and 2400Z CG longitude correlations

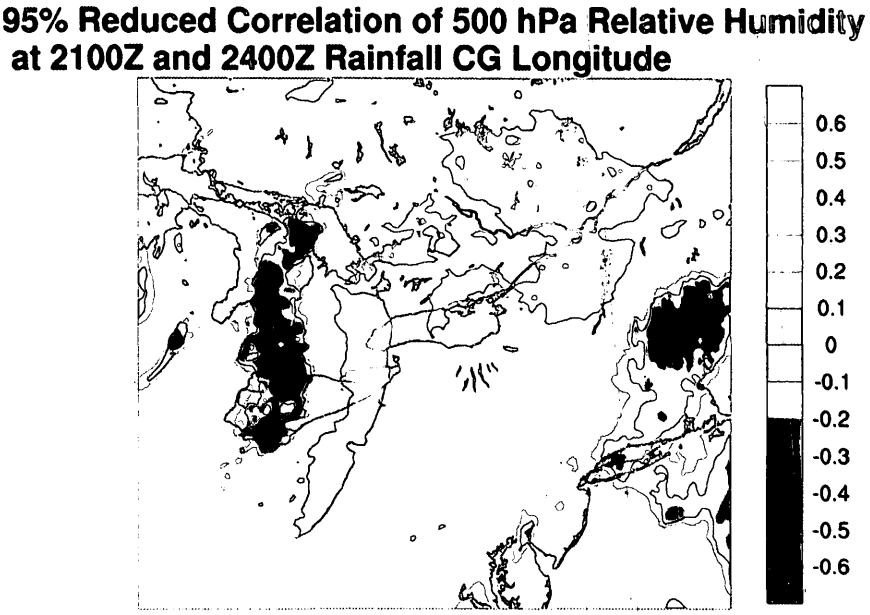


Figure 51: 500hPa 2100Z relative humidity and 2400Z CG longitude correlations

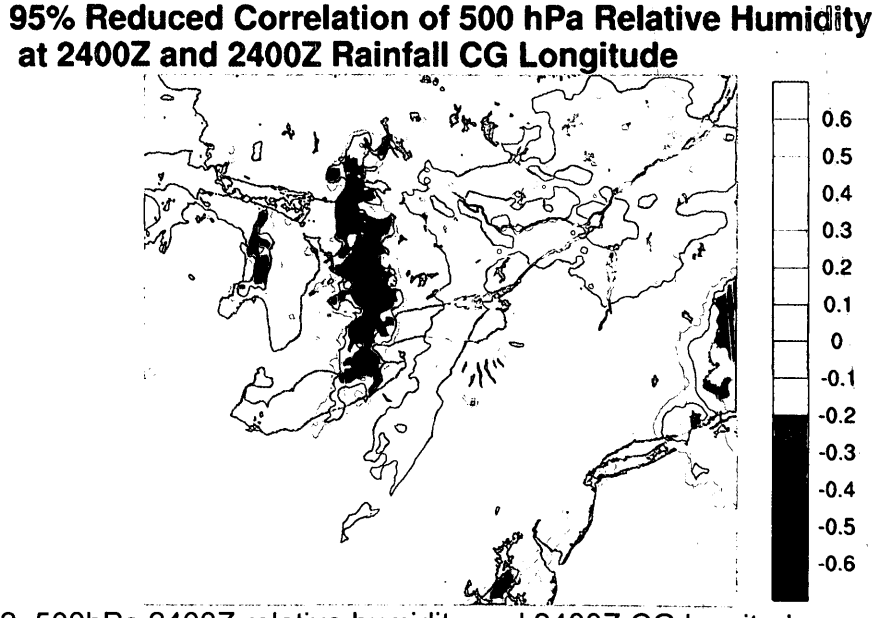


Figure 52: 500hPa 2400Z relative humidity and 2400Z CG longitude correlations

In Figures 53 to 60 we have similar plots of correlation of CG longitude with 850 hPa temperature and meridional velocity.

For the temperature plots, Figures 53 to 56, a situation similar to the relative humidity at 500 hPa is evident – bands of positive and negative correlation which appear to track with the squall line backward in time. In this case however, the correlation is much less pronounced and more subtle.

For the meridional velocity correlations, Figures 57 to 60, the relationship is more difficult to interpret, but again, a physical interpretation is not necessary. By taking the 95% lower bound of our correlations, we can be 95% sure that the correlations are statistically significant. Given that the meridional velocity correlations remain high (in excess of 0.6), even 12 hours beforehand, we can be sure that they should offer guidance for targeting efficient data assimilation activity.

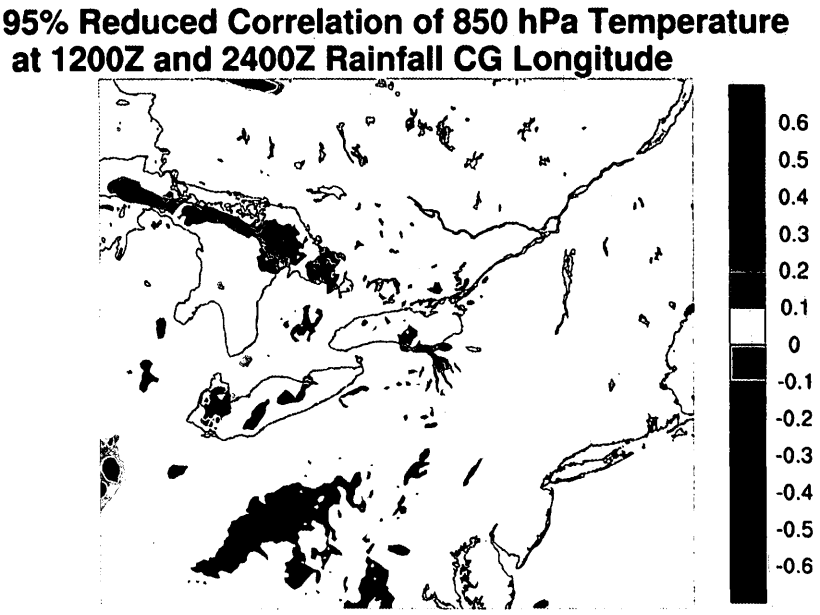


Figure 53: 850hPa 1200Z temperature and 2400Z CG longitude correlations

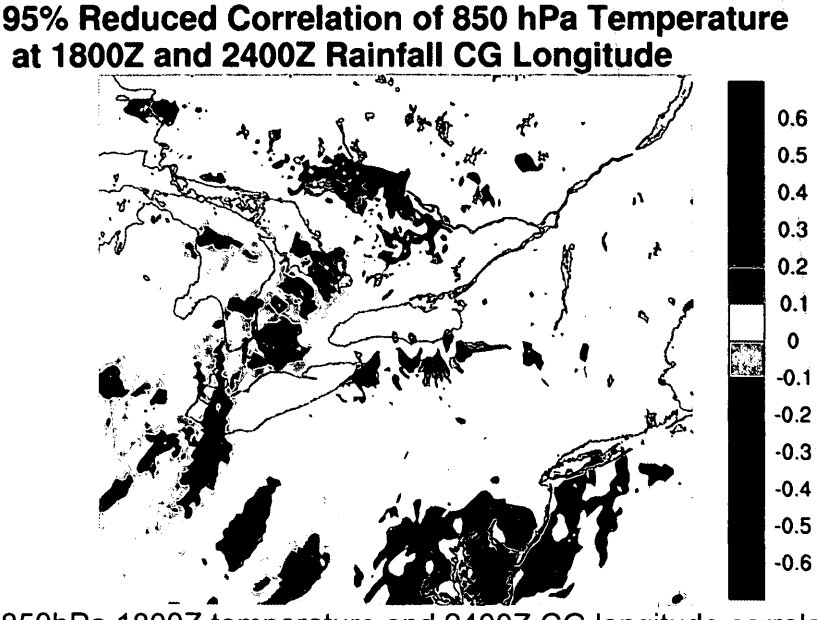


Figure 54: 850hPa 1800Z temperature and 2400Z CG longitude correlations

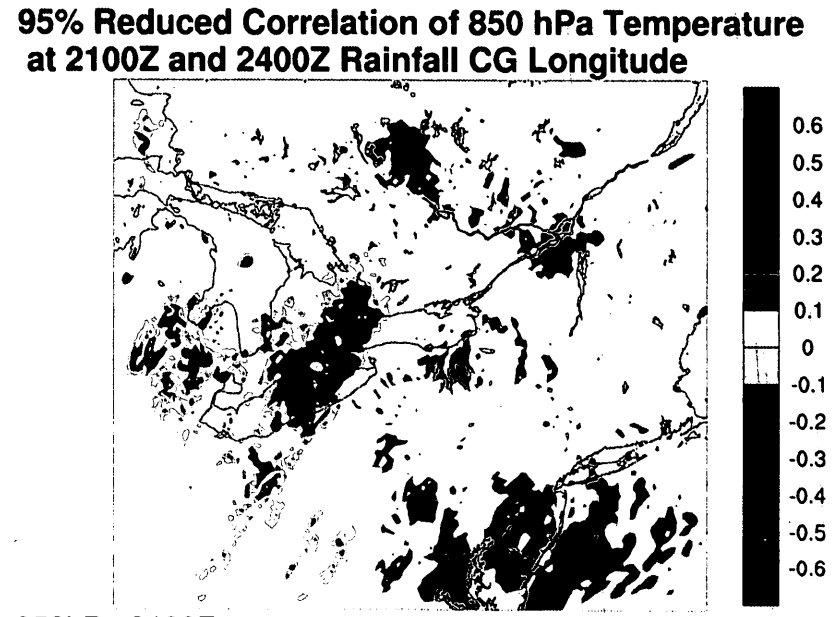


Figure 55: 850hPa 2100Z temperature and 2400Z CG longitude correlations

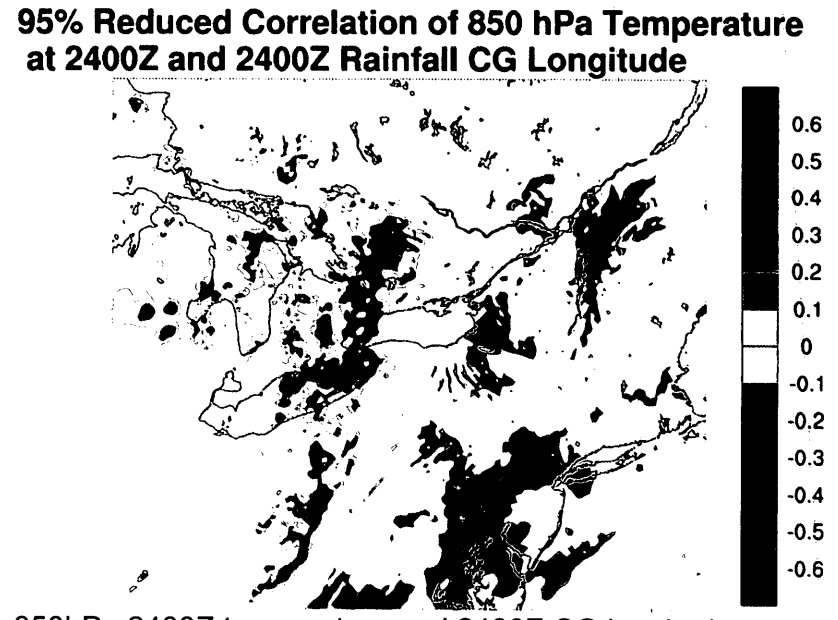


Figure 56: 850hPa 2400Z temperature and 2400Z CG longitude correlations

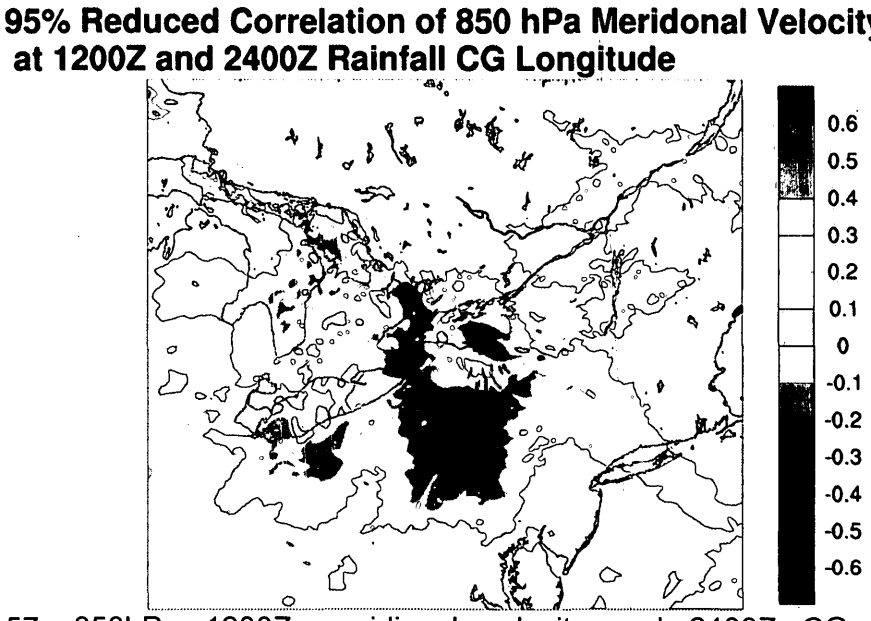


Figure 57: 850hPa 1200Z meridional velocity and 2400Z CG longitude correlations

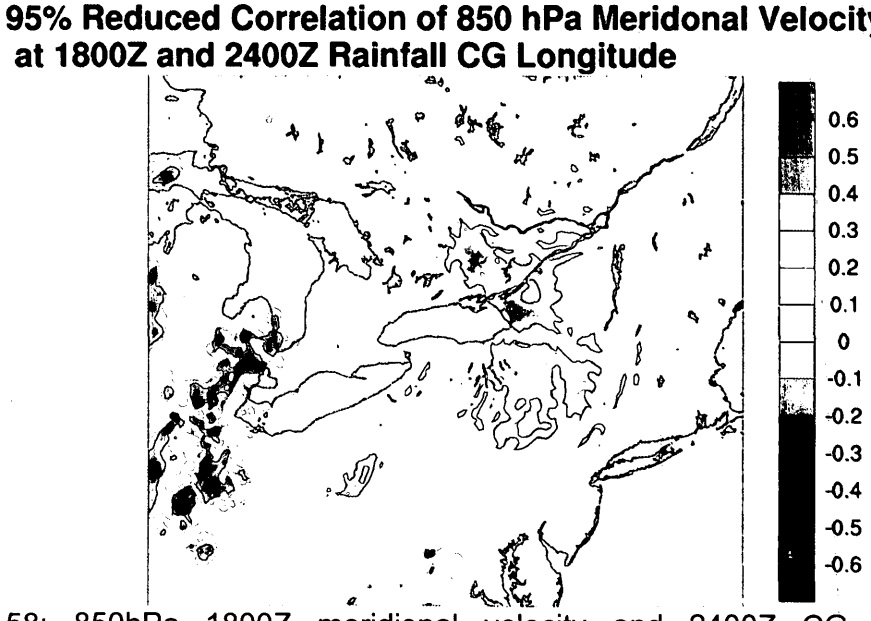


Figure 58: 850hPa 1800Z meridional velocity and 2400Z CG longitude correlations

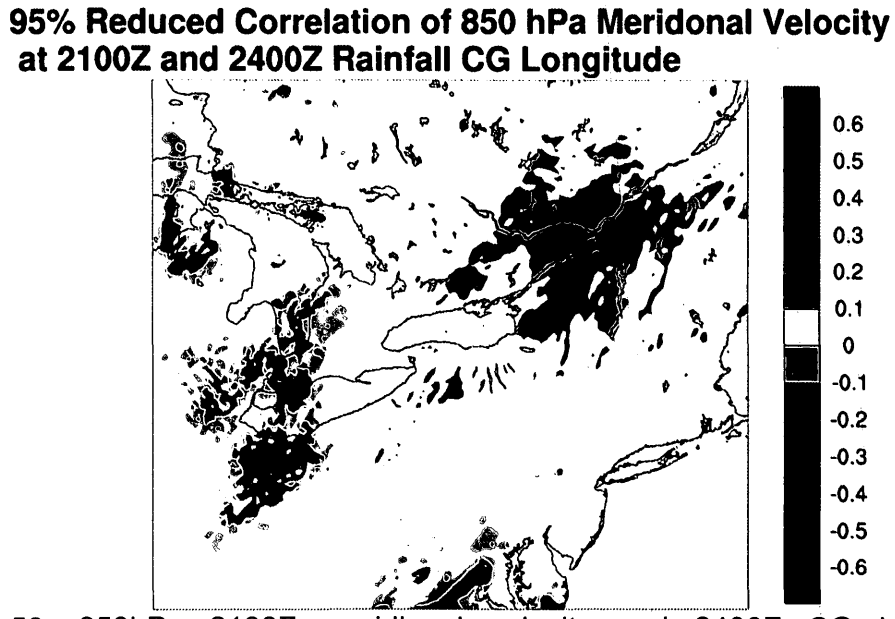


Figure 59: 850hPa 2100Z meridional velocity and 2400Z CG longitude correlations

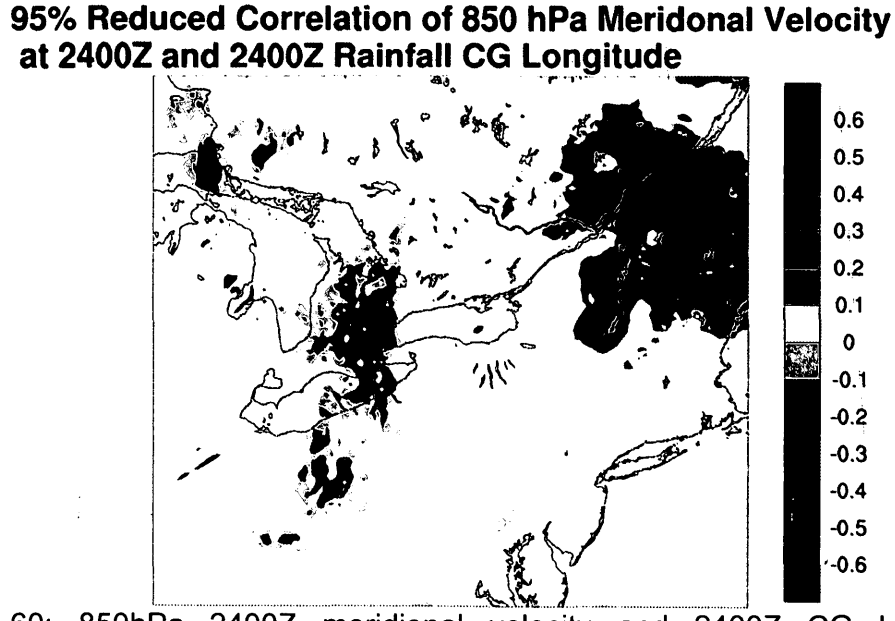


Figure 60: 850hPa 2400Z meridional velocity and 2400Z CG longitude correlations

3.3.2 Temperature Gradient 'Object-Based' Squall Line Correlations

We come now to our second sensitivity study, the correlation of input variables with the longitude of intersection of the object-based 'squall lines' with a reference latitude (the ensemble average mid-point latitude of these object-based 'squall lines'). Here however, as we noted in the Methodology, we are interested in the correlations not with basic variables distributed in time and space, but derived single value variables at 1200Z – namely pressure low latitude, longitude, strength and mean zonal wind flow. Again, as for the distributed correlation studies, a 95% lower bound was taken on the correlations. The results are as follows:

Variable	Correlation With Squall Line/Reference Latitude Intersection Longitude (95% Lower Error Bound, or 95% Upper Error Bound for Negative Correlations)
1200Z Pressure Low Longitude	0 (-0.006)[19]
1200Z Pressure Low Latitude	0 (-0.124)
1200Z Pressure Low Strength	0 (0.254)
1200Z 2nd Domain Mean Zonal Flow	0.294

Table 5: Object-based squall line correlations

19 The values for the first 3 correlations were taken to be zero, since an initial positive (negative) correlation became negative (positive) when the lower (upper) 95% correlation was calculated. See explanation in the Methodology section.

Clearly, these correlations are not as significant as those obtained for the CG longitude correlations above, except for the domain mean zonal flow, which is also admittedly quite low. This is probably because the 'output variable' chosen, the longitude of the intersection of the squall line with the reference latitude is a poor indicator of squall line position. Perhaps by choosing a larger lower threshold of temperature gradient to define our squall lines, we could use the object-based 'squall line' mid point longitude as our final variable, since we could be more sure in this case that the full squall line was being indicated by our squall line objects, rather than the more indirect method of using an intersection with a reference latitude. Nevertheless, given the poor correlations obtained, it is hard to say whether the correlations with these input variables are poor in actuality, or poor because of a poor choice of output variable.

3.4 Conclusions

We see then notably strong correlations between the 2400Z longitude of the rain weighted center of gravity (CG) and the low level (850 hPa) zonal and meridional velocities, and mid level (500 hPa) relative humidities even as far back as 1200Z hours. While not as strong, appreciable correlations with 850 hPa temperatures also exist. As noted in our Introduction, such strong correlations have been used to help target those variables, areas and times, for which data assimilation methods for improving forecast accuracy would be most effective.

The correlations derived from using the 2400Z longitude of object-based 'squall line' intersection with our reference latitude as our output variable, and 1200Z pressure low latitude, longitude, intensity and domain mean zonal velocity as our input variables, were much smaller. The results of this study are inconclusive as to whether these poor correlations are indeed indicative of poor of squall line zonal position correlation with our input variables, or rather just our poor choice of output variable to correlate with.

Nevertheless, the correlation studies as a whole have shown that there are indeed strong correlation between our squall line zonal position and many input

variables back to 1200Z. Thus, these results suggest that data assimilation strategies employed back at these and later times would be effective in reducing the biased zonal position of squall line predicted by the high and low resolution simulations of the first phase of this study and the ensemble mean distribution of the second phase.

4 Summarizing Conclusions and Future Work

We can see then that the low resolution (3km) simulations were just as successful as the high resolution simulation (1km) in capturing the basic dynamics and evolution of the storm system, modeling a squall line system as expected by observations. As was learned from our modeling, the basic driver for the storm was a low level jet situated adjacent to a moist, unstable air mass. This jet imparted very significant shear into the lower level flow and resulted in especially high helicity values along the squall line, which resulted in very high levels of vorticity when tilted by the updrafts simulated, leading to the genesis of supercells. Evidence of suspercell splitting along the squall line, and long-lived supercells were indeed witness at both resolutions. It would appear from our results that dynamically driven vertical pressure gradients resulting from supercell evolution were the primary lifting mechanism, although low level convergence may also have been a factor, and that the scale of this lifting mechanism was mesoscale in extent. Both the high and low resolution simulations resulted in a significant phase error of squall line position however.

A method of generating ensembles for research purposes, from a re-interpolation of lower resolution ensemble perturbations generated from the published

ensembles of forecast centers was proposed. It was found that the ensemble generated was indeed successful as it offered a sufficiently dispersive ensemble – one that after 24 hours was able to contain the true state of the atmosphere as a plausible member of the ensemble. Thus, despite concerns which have been raised regarding the usage of low resolution data for the generation of convective-scale ensembles, the ensemble generated by the method used in our study appears to have been successful.

The ensemble mean was found to give the same phase lagged position for the squall line position as the high and low resolution simulations, although peak values of quantities such as rainfall and vertical velocity along the squall line were unrealistically low.

There appear to be at least two scales of unpredictability at work in the system simulated – a small scale nonlinear scale related to low level temperature fluctuations that appears to saturate quickly after about 12 hours, and a larger, synoptic-scale scale which appears to saturate well after the 24 hour period of our simulation. Thus, information related to these larger scales would be expected to remain correlated with our ensembles' end states after 24 hours. This expected correlation is confirmed by our correlation results from the third

part of the study, which showed very significant correlations of squall line zonal position at 2400Z with model initial variables. Since the small scales have lost correlation with the model's end state by 2400Z (due to saturation), but the larger scale has retained significant correlations, we can surmise that the squall line position is driven by the activity at the larger scale, which would otherwise not correlate. Thus, the phase error may well be the result of some initial condition error, possibly an error in the background flow. This would explain why all three simulations – both the high and low resolution simulations, and the ensemble mean simulation, all showed the same phase error.

Again, our correlation studies show that significant correlations of squall line position remain with prior model variables at least 12 hour before. Thus, targeting areas for adding observations to improve upon the squall line position using data assimilation is certainly very feasible. Naturally, confirming this hypothesis remains a definite goal for future work.

Appendix A: WRF Control Files for High and Low Resolution Simulations

The following are the namelist input files for the high and low resolution models, as well as the eta levels used to define the vertical levels in all simulations.

High Resolution Model (1km) namelist.input file:

```
&time_control

run_days                            = 0,
run_hours                           = 24,
run_minutes                         = 0,
run_seconds                         = 0,
start_year                          = 2009, 2009, 2009, 2009
start_month                         = 08,   08,   08,  08,
start_day                           = 20,   20,   20,  20,
start_hour                          = 00,   00,   00,  00,
start_minute                        = 00,   00,   00,  00,
start_second                        = 00,   00,   00,  00,
end_year                            = 2009, 2009, 2009, 2009,
end_month                           = 08,   08,   08,  08,
end_day                             = 21,   21,   21,  21,
end_hour                            = 00,   00,   00,  00,
end_minute                          = 00,   00,   00,  00,
end_second                          = 00,   00,   00,  00,
interval_seconds                    = 21600
input_from_file                     = .true.,.true.,.true.,.true.,
fine_input_stream                   = 0,     2,   2,   2,
history_interval                    = 60,  60,   60,  60,
frames_per_outfile                  = 1, 1, 1, 1,
restart                             = .false.,
restart_interval                    = 720,
io_form_history                     = 2
```

```
io_form_restart                     = 2
io_form_input                       = 2
io_form_boundary                    = 2
debug_level                         = 1000

/


&domains
time_step                           = 90,
time_step_fract_num                 = 0,
time_step_fract_den                 = 1,
max_dom                             = 3,
e_we                                = 193, 274, 421, 421,
e_sn                                = 129, 229, 391, 661,
e_vert                              = 28, 28,  28,  28,
p_top_requested                     = 20000,
num_metgrid_levels                  = 8,
num_metgrid_soil_levels             = 0,
dx                                  = 27000, 9000, 3000, 1000,
dy                                  = 27000, 9000, 3000, 1000,
grid_id                             = 1,     2,     3,    4,
parent_id                           = 1,     1,     2,    3,
i_parent_start                      = 1,     87,    92,   111,
j_parent_start                      = 1,     35,    60,    91,
parent_grid_ratio                   = 1,     3,     3,     3,
parent_time_step_ratio              = 1,     3,     3,     3,
feedback                            = 1,
smooth_option                       = 0

/


&physics
mp_physics                          = 10,   10,    10,   10,
ra_lw_physics                       = 1,     1,     1,    1,
```

```
ra_sw_physics                       = 1,     1,     1,    1,
radt                                = 27,    9,    3,     1,
sf_sfclay_physics                   = 1,     1,     1,    1,
sf_surface_physics                  = 1,     1,     1,    1,
bl_pbl_physics                      = 1,     1,     1,    1,
bldt                                = 0,     0,     0,    0,
cu_physics                          = 1,     1,     0,    0,
cudt                                = 5,     5,     5,    5,
isfflx                              = 1,
ifsnow                              = 0,
icloud                              = 1,
surface_input_source                = 1,

num_soil_layers                     = 5,
sf_urban_physics                    = 0,
maxiens                             = 1,
maxens                              = 3,
maxens2                             = 3,
maxens3                             = 16,
ensdim                              = 144,

/

&fdda
/

&dynamics
w_damping                           = 0,
diff_opt                            = 1,
km_opt                              = 4,
diff_6th_opt                        = 1,      1,      1,     1,
diff_6th_factor                     = 0.12,   0.12,   0.12,   0.12,
base_temp                           = 290.
damp_opt                            = 0,
zdamp                               = 5000.,  5000.,  5000.,  5000.,
dampcoef                            = 0.2,    0.2,    0.2,   0.2,
khdif                               = 0,      0,      0,    0,
```

```
kvdif                               = 0,      0,      0,     0,
non_hydrostatic                     = .true., .true., .true., .true.,
moist_adv_opt                       = 1,      1,      1,     1,
scalar_adv_opt                      = 1,      1,      1,     1,
/



&bdy_control
spec_bdy_width                      = 5,
spec_zone                           = 1,
relax_zone                          = 4,
specified                           = .true., .false.,.false.,.false.,
nested                              = .false., .true., .true.,.true.,

/



&grib2
/



&namelist_quilt
nio_tasks_per_group = 0,
nio_groups = 1,
/
```

Low Resolution Model (3km) namelist.input file:

```
&time_control
run_days                            = 0,
run_hours                           = 24,
run_minutes                         = 0,
run_seconds                         = 0,
start_year                          = 2009, 2009, 2009
start_month                         = 08,   08,   08
start_day                           = 20,   20,   20
start_hour                          = 00,   00,   00
start_minute                        = 00,   00,   00
start_second                        = 00,   00,   00
end_year                            = 2009, 2009, 2009
end_month                           = 08,   08,   08
end_day                             = 21,   21,   21
end_hour                            = 00,   00,   00
end_minute                          = 00,   00,   00
end_second                          = 00,   00,   00
interval_seconds                    = 21600
input_from_file                     = .true.,.true.,.true.
fine_input_stream                   = 0,     2,   2
history_interval                    = 60,  60,   60
frames_per_outfile                  = 1, 1, 1
restart                             = .false.,
restart_interval                    = 720,
io_form_history                     = 2
io_form_restart                     = 2
io_form_input                       = 2
io_form_boundary                    = 2
debug_level                         = 1000
/


&domains
time_step                           = 90,
time_step_fract_num                 = 0,
```

```
time_step_fract_den             = 1,
max_dom                         = 3,
e_we                            = 193, 274, 421
e_sn                            = 129, 229, 391
e_vert                          = 28, 28,  28
p_top_requested                 = 20000,
num_metgrid_levels              = 8,
num_metgrid_soil_levels         = 0,
dx                              = 27000, 9000, 3000
dy                              = 27000, 9000, 3000
grid_id                         = 1,     2,     3
parent_id                       = 1,     1,     2
i_parent_start                  = 1,     87,    92
j_parent_start                  = 1,     35,    60
parent_grid_ratio               = 1,     3,     3
parent_time_step_ratio          = 1,     3,     3
feedback                        = 1,
smooth_option                   = 0
/


&physics
mp_physics                      = 10,    10,    10
ra_lw_physics                   = 1,     1,     1
ra_sw_physics                   = 1,     1,     1
radt                            = 27,    9,     3
sf_sfclay_physics               = 1,     1,     1
sf_surface_physics              = 1,     1,     1
bl_pbl_physics                  = 1,     1,     1
bldt                            = 0,     0,     0
cu_physics                      = 1,     1,     0
cudt                            = 5,     5,     5
isfflx                          = 1,
ifsnow                          = 0,
icloud                          = 1,
surface_input_source            = 1,
num_soil_layers                 = 5,
```

```
sf_urban_physics                    = 0,
maxiens                             = 1,
maxens                              = 3,
maxens2                             = 3,
maxens3                             = 16,
ensdim                              = 144,
/

&fdda
/


&dynamics
w_damping                           = 0,
diff_opt                            = 1,
km_opt                              = 4,
diff_6th_opt                        = 1,      1,      1
diff_6th_factor                     = 0.12,   0.12,   0.12
base_temp                           = 290.
damp_opt                            = 0,
zdamp                               = 5000.,  5000.,  5000.
dampcoef                            = 0.2,    0.2,    0.2
khdif                               = 0,      0,      0
kvdif                               = 0,      0,      0
non_hydrostatic                     = .true., .true., .true.
moist_adv_opt                       = 1,      1,      1,
scalar_adv_opt                      = 1,      1,      1,
/

&bdy_control
spec_bdy_width                      = 5,
spec_zone                           = 1,
relax_zone                          = 4,
specified                           = .true., .false.,.false.
nested                              = .false., .true., .true.
/
```

```
&grib2
/

&namelist_quilt
nio_tasks_per_group = 0,
nio_groups = 1,
/
```

Eta Levels For WRF Simulations:

The following are the eta levels used to define the vertical levels in all WRF simulations (note that level numbers count from 0 for the first level):

1.000 , 0.990 , 0.978 , 0.964 , 0.946 , 0.922 , 0.894 , 0.860 , 0.817 , 0.766 ,
0.707 , 0.644 , 0.576 , 0.507 , 0.444 , 0.380 , 0.324 , 0.273 , 0.228 , 0.188 ,
0.152 , 0.121 , 0.093 , 0.069 , 0.048 , 0.029 , 0.014 , 0.000

Appendix B: Ensemble Object-Based Squall Lines at 2400Z

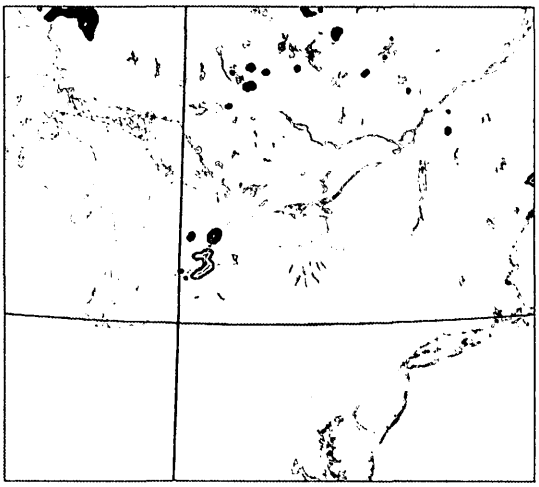
The following plots are of the 2400Z ensemble members where the 2m temperature gradient exceeded 50% of the domain maximum temperature gradient of each ensemble member. Typically, such areas fell along the squall line, although high temperature gradients were seen in other areas as well. Note that while these objects were called 'frontal objects' in the plots, such areas are also associated with pre-frontal phenomena such as squall lines as well as high temperature gradient areas such as cold fronts, so these term 'frontal object' is to be taken as only a informal term for all such high gradient 'objects'. Note also that the colors are not indicative of any specific physical properties of these 'frontal' objects – these objects were differentiated as separate contiguous objects by frontal numbers for analysis purposes in the database constructed from the data, and the colors are a just a by-product of the plotting algorithm of NCL – the language used for the data analysis and plotting of this study. Since the colors help differentiate the objects being studied, it was decided to leave them as plotted.

Lines were then drawn through those objects corresponding to the squall lines of the ensembles, and as described in the Methodology section of Chapter 3, the longitudes of the intersections of these lines with a 'reference' latitude was then

determined. This reference latitude was taken to be the average 'squall line' mid-point latitude among the ensemble members.

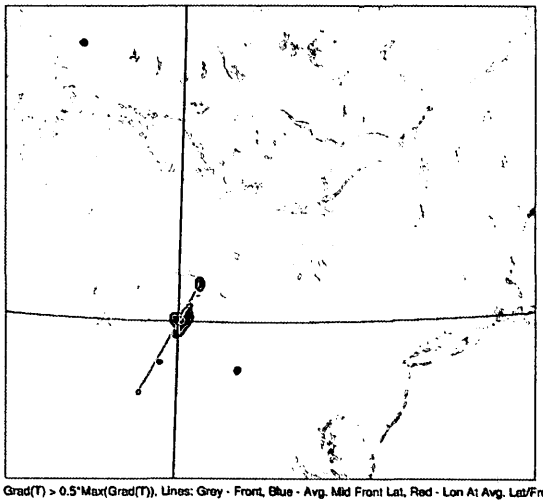
As can be seen from these plots, there was considerable variation in their positioning.

Ensemble 1 Frontal Longitude



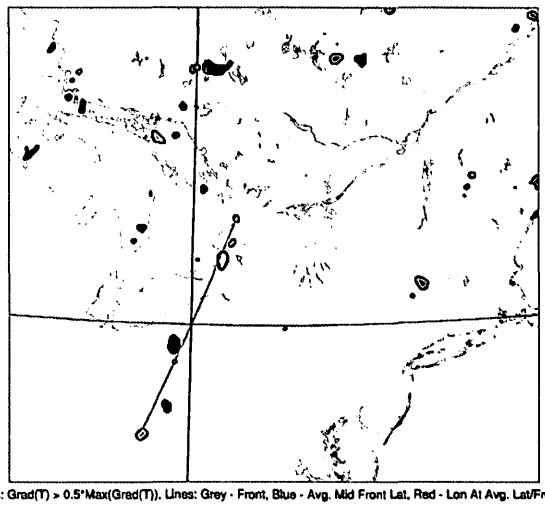
Frontal objects: Grad(T) > 0.5*Max(Grad(T)), Lines: Grey - Front, Blue - Avg. Mid Front Lat, Red - Lon At Avg. Lat/Front Intersection

Ensemble 2 Frontal Longitude



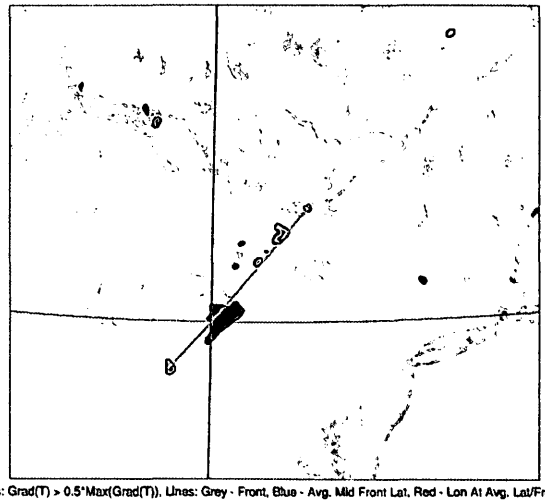
Frontal objects: Grad(T) > 0.5*Max(Grad(T)), Lines: Grey - Front, Blue - Avg. Mid Front Lat, Red - Lon At Avg. Lat/Front Intersection

Ensemble 3 Frontal Longitude



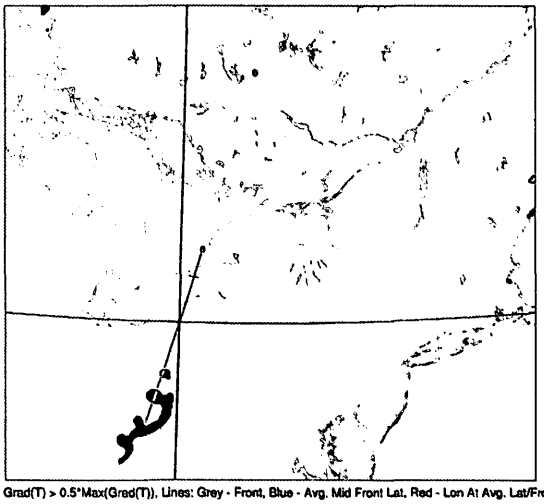
Frontal objects: Grad(T) > 0.5*Max(Grad(T)). Lines: Grey - Front, Blue - Avg. Mid Front Lat, Red - Lon At Avg. Lat/Front Intersection

Ensemble 4 Frontal Longitude



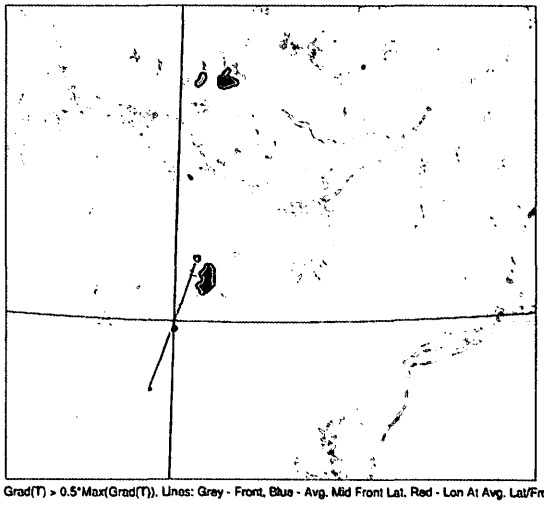
Frontal objects: Grad(T) > 0.5*Max(Grad(T)). Lines: Grey - Front, Blue - Avg. Mid Front Lat, Red - Lon At Avg. Lat/Front Intersection

Ensemble 5 Frontal Longitude



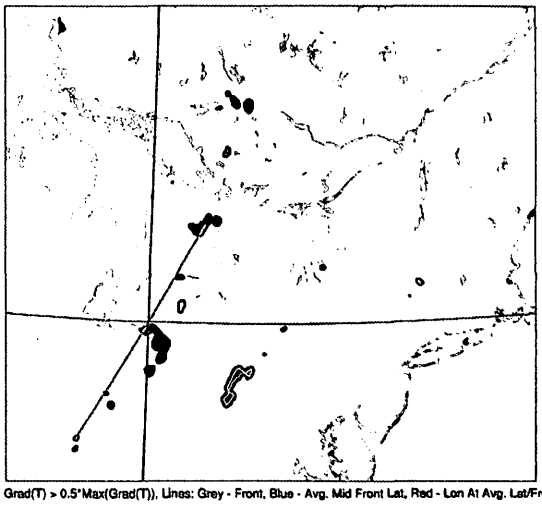
Frontal objects: Grad(T) > 0.5*Max(Grad(T)), Lines: Grey - Front, Blue - Avg. Mid Front Lat, Red - Lon At Avg. Lat/Front Intersection

Ensemble 6 Frontal Longitude



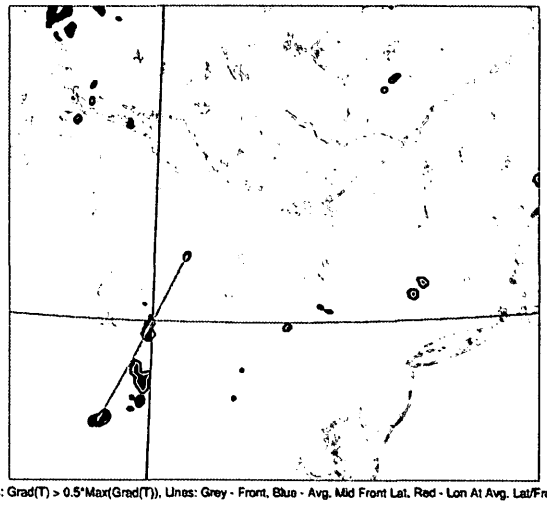
Frontal objects: Grad(T) > 0.5*Max(Grad(T)), Lines: Grey - Front, Blue - Avg. Mid Front Lat, Red - Lon At Avg. Lat/Front Intersection

Ensemble 7 Frontal Longitude



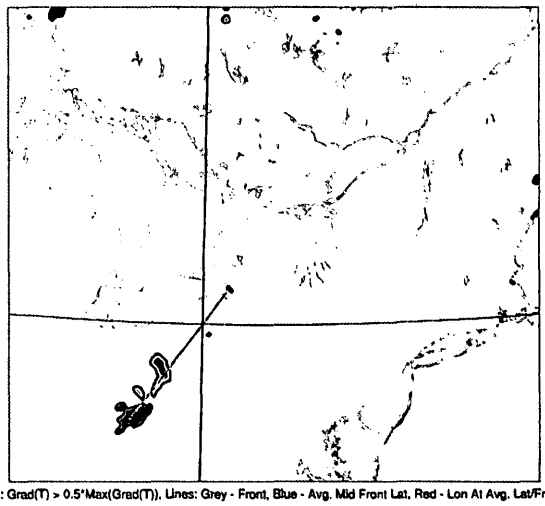
Frontal objects: Grad(T) > 0.5*Max(Grad(T)), Lines: Grey - Front, Blue - Avg. Mid Front Lat, Red - Lon At Avg. Lat/Front Intersection

Ensemble 8 Frontal Longitude



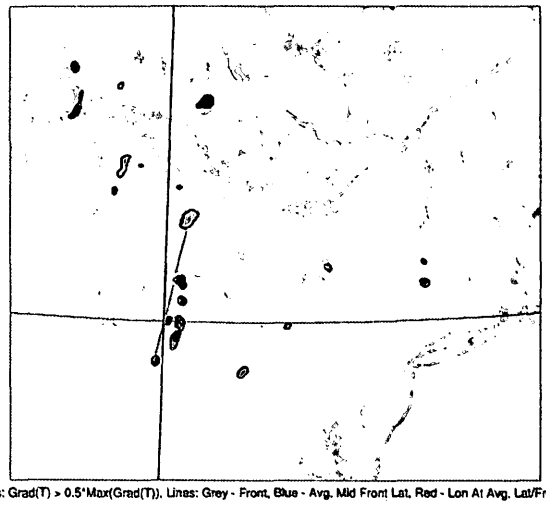
Frontal objects: Grad(T) > 0.5*Max(Grad(T)), Lines: Grey - Front, Blue - Avg. Mid Front Lat, Red - Lon At Avg. Lat/Front Intersection

Ensemble 9 Frontal Longitude



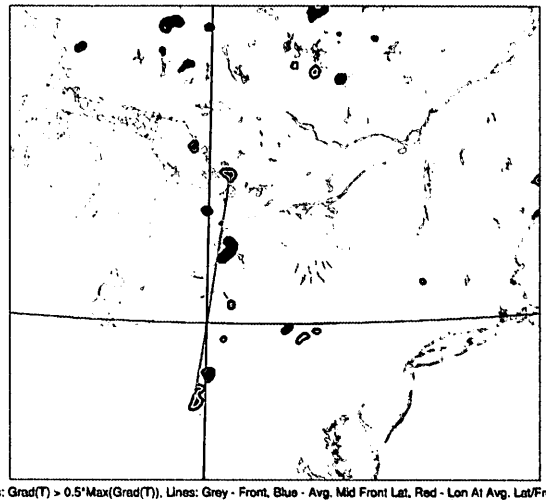
Frontal objects: Grad(T) > 0.5*Max(Grad(T)). Lines: Grey - Front, Blue - Avg. Mid Front Lat, Red - Lon At Avg. Lat/Front Intersection

Ensemble 10 Frontal Longitude



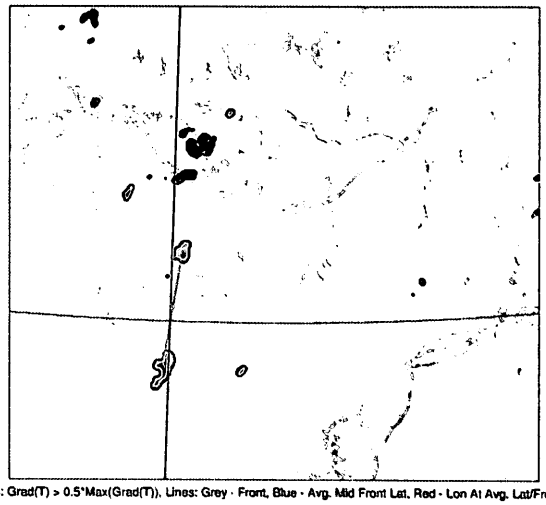
Frontal objects: Grad(T) > 0.5*Max(Grad(T)). Lines: Grey - Front, Blue - Avg. Mid Front Lat, Red - Lon At Avg. Lat/Front Intersection

Ensemble 11 Frontal Longitude



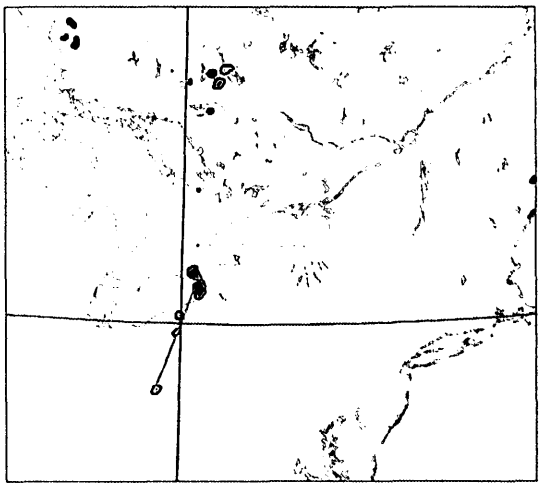
Frontal objects: Grad(T) > 0.5*Max(Grad(T)), Lines: Grey - Front, Blue - Avg. Mid Front Lat, Red - Lon At Avg. Lat/Front Intersection

Ensemble 12 Frontal Longitude



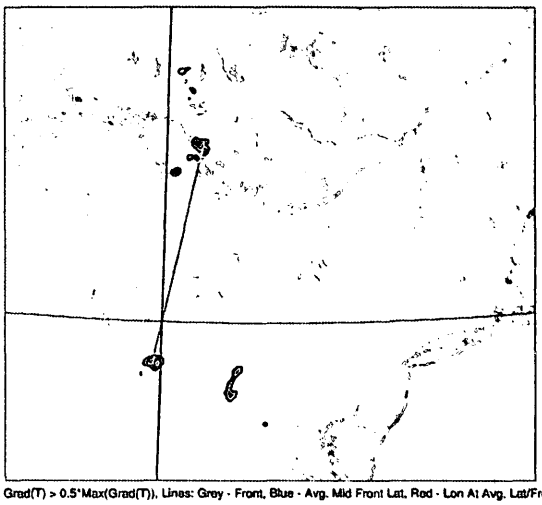
Frontal objects: Grad(T) > 0.5*Max(Grad(T)), Lines: Grey - Front, Blue - Avg. Mid Front Lat, Red - Lon At Avg. Lat/Front Intersection

Ensemble 13 Frontal Longitude



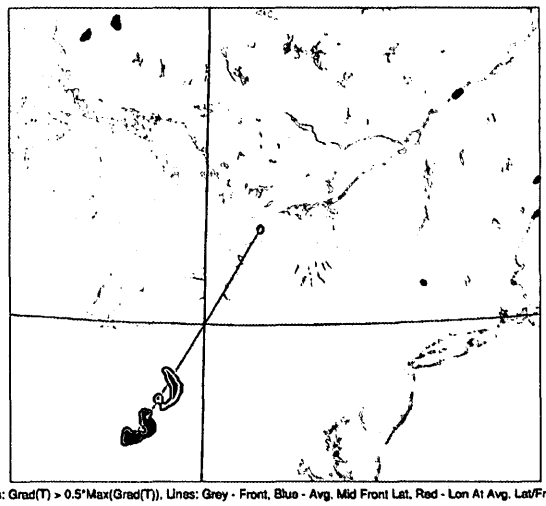
Frontal objects: Grad(T) > 0.5*Max(Grad(T)), Lines: Grey - Front, Blue - Avg. Mid Front Lat, Red - Lon At Avg. Lat/Front Intersection

Ensemble 14 Frontal Longitude



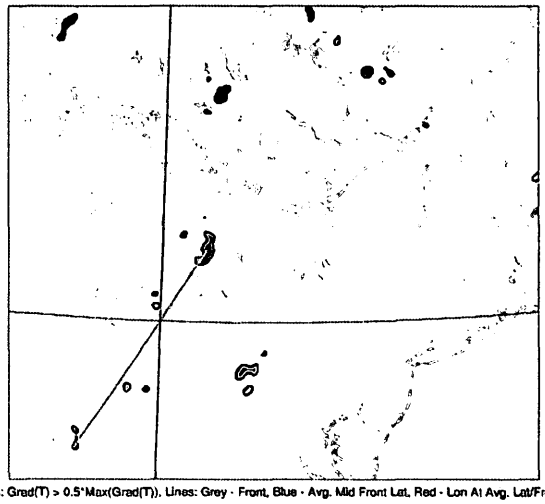
Frontal objects: Grad(T) > 0.5*Max(Grad(T)), Lines: Grey - Front, Blue - Avg. Mid Front Lat, Red - Lon At Avg. Lat/Front Intersection

Ensemble 15 Frontal Longitude



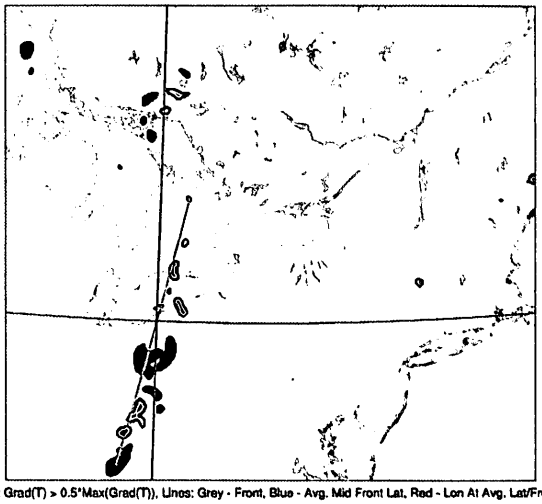
Frontal objects: Grad(T) > 0.5*Max(Grad(T)), Lines: Grey - Front, Blue - Avg. Mid Front Lat, Red - Lon At Avg. Lat/Front Intersection

Ensemble 16 Frontal Longitude



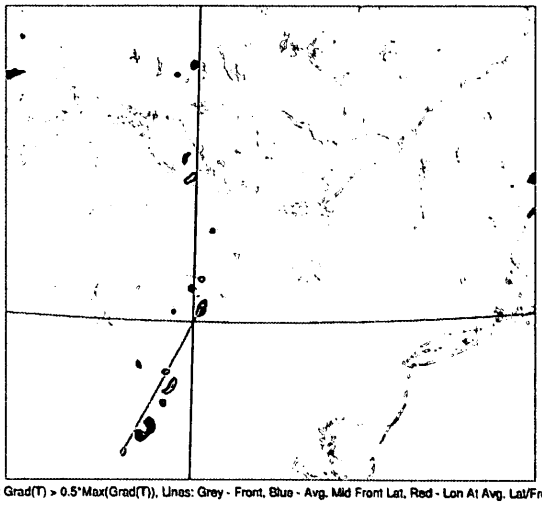
Frontal objects: Grad(T) > 0.5*Max(Grad(T)), Lines: Grey - Front, Blue - Avg. Mid Front Lat, Red - Lon At Avg. Lat/Front Intersection

Ensemble 17 Frontal Longitude



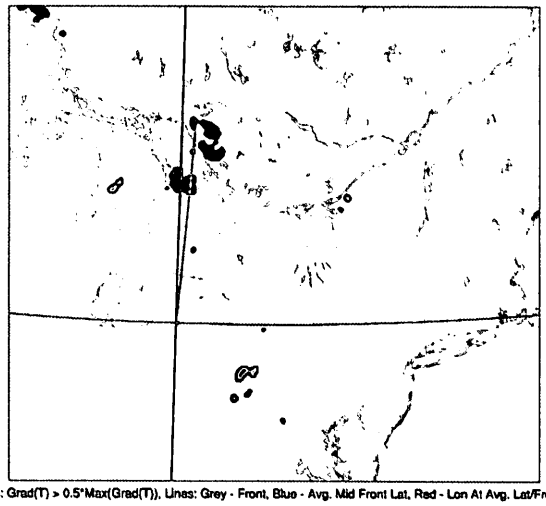
Frontal objects: Grad(T) > 0.5*Max(Grad(T)), Lines: Grey - Front, Blue - Avg. Mid Front Lat, Red - Lon At Avg. Lat/Front Intersection

Ensemble 18 Frontal Longitude



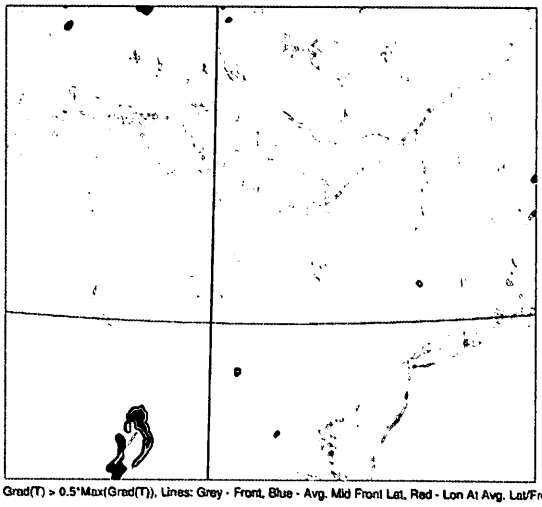
Frontal objects: Grad(T) > 0.5*Max(Grad(T)), Lines: Grey - Front, Blue - Avg. Mid Front Lat, Red - Lon At Avg. Lat/Front Intersection

Ensemble 19 Frontal Longitude



Frontal objects: Grad(T) > 0.5*Max(Grad(T)), Lines: Grey - Front, Blue - Avg. Mid Front Lat, Red - Lon At Avg. Lat/Front Intersection

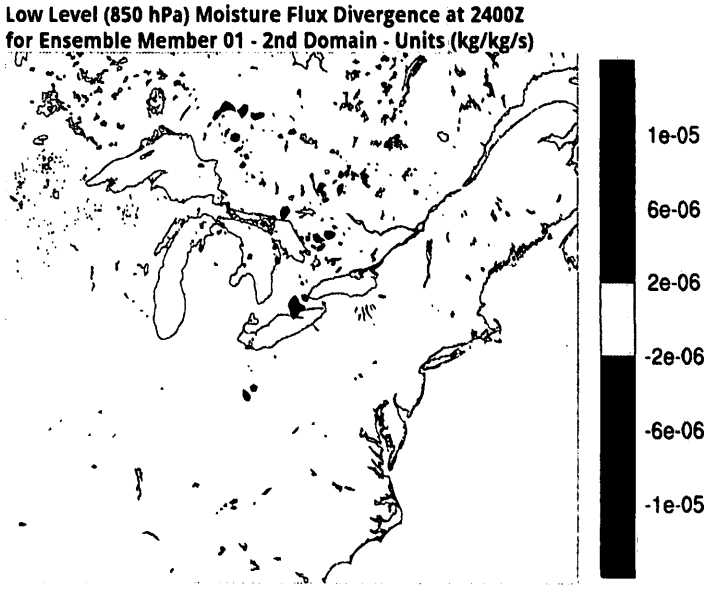
Ensemble 20 Frontal Longitude

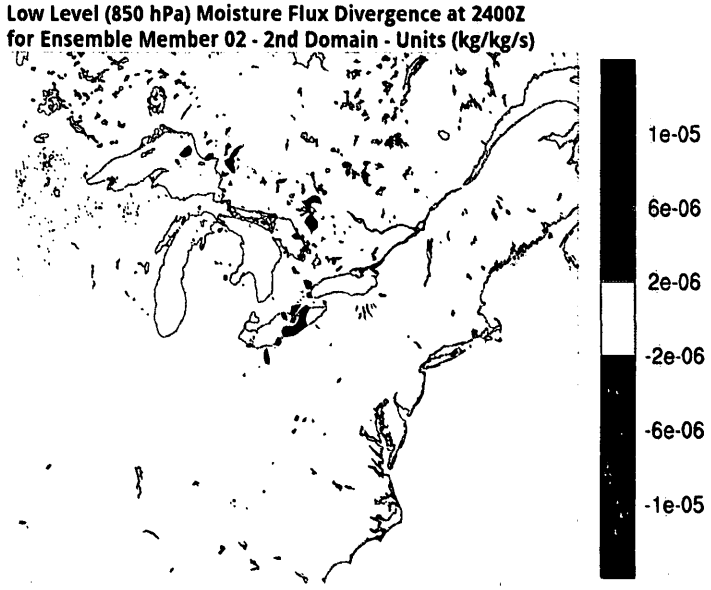


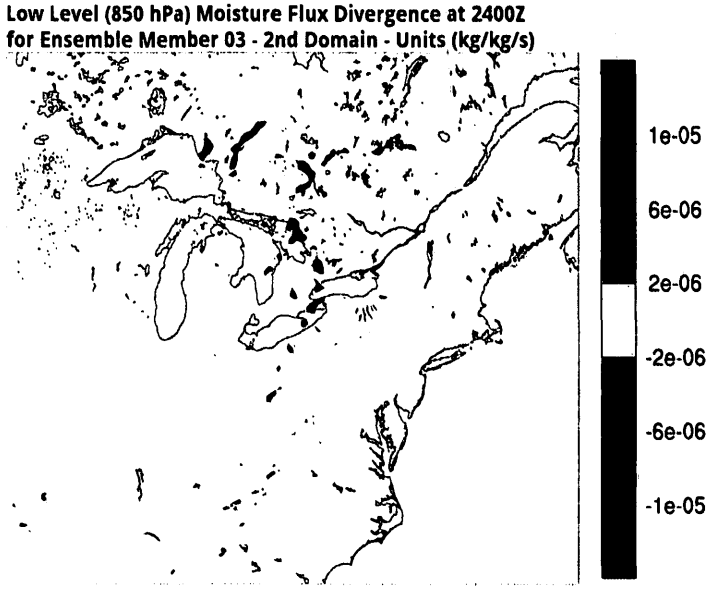
Frontal objects: Grad(T) > 0.5*Max(Grad(T)), Lines: Grey - Front, Blue - Avg. Mid Front Lat, Red - Lon At Avg. Lat/Front Intersection

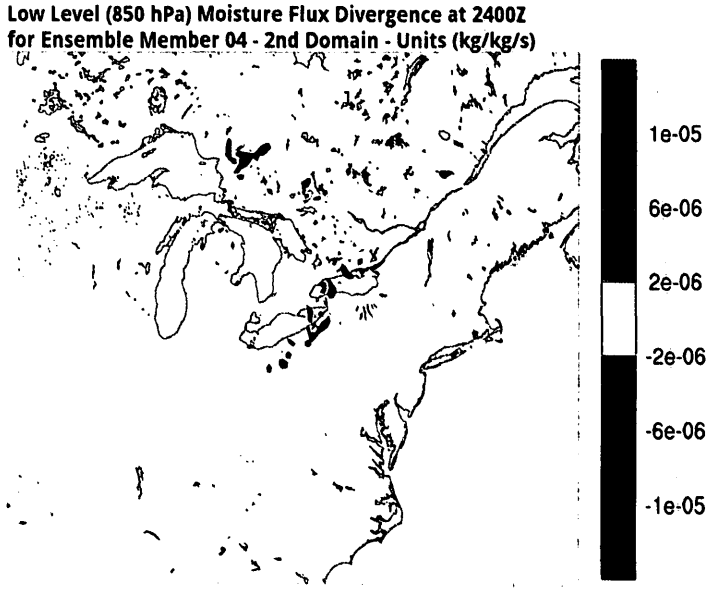
Appendix C: Moisture Flux Divergence Squall Lines at 2400Z

The following plots are of the 2400Z moisture flux divergence at the 850 hPa level for the 20 ensemble members. As was the case with the squall lines as estimated from the high 2m temperature gradient 'objects' of Appendix C, there is considerable variation in the squall line positions that can be estimated from the following figures. These estimated squall lines in most of the following figures can be subjectively delineated as the linearly extended regions adjacent to the Great Lakes which show the most highly concentrated regions of maximum moisture flux convergence and divergence. Note of course that some ensemble members showed more easily defined squall lines than others- for instance, the squall line is fairly evident for ensemble member 6, but not so well defined for ensemble member 1.

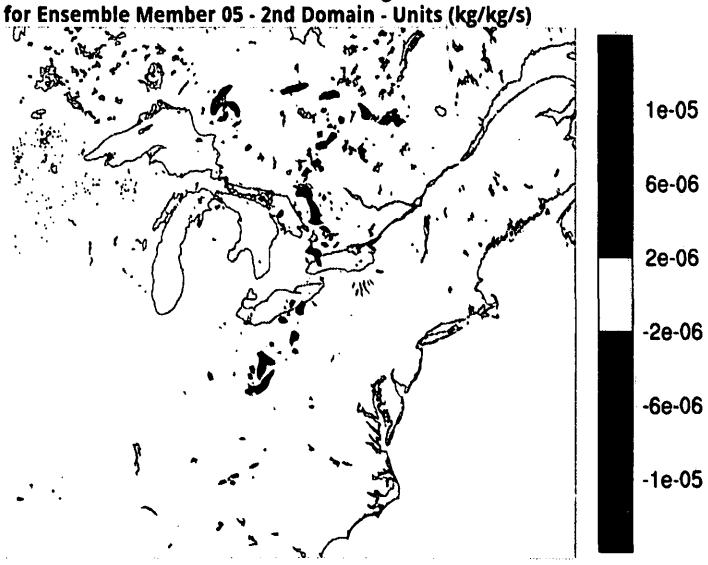

Low Level (850 hPa) Moisture Flux Divergence at 2400Z
for Ensemble Member 01 - 2nd Domain - Units (kg/kg/s)
1e-05
6e-06
2e-06
-2e-06
-6e-06
-1e-05


Low Level (850 hPa) Moisture Flux Divergence at 2400Z
for Ensemble Member 02 - 2nd Domain - Units (kg/kg/s)
1e-05
6e-06
2e-06
-2e-06
-6e-06
-1e-05

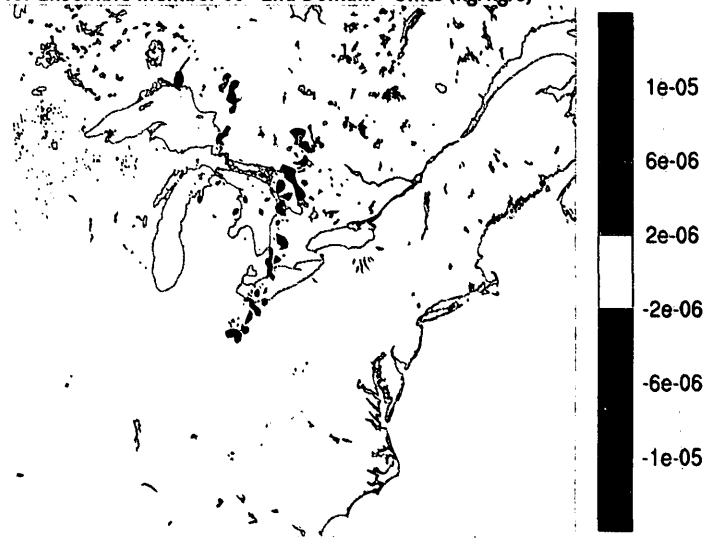

Low Level (850 hPa) Moisture Flux Divergence at 2400Z
for Ensemble Member 03 - 2nd Domain - Units (kg/kg/s)
1e-05
6e-06
2e-06
-2e-06
-6e-06
-1e-05


Low Level (850 hPa) Moisture Flux Divergence at 2400Z
for Ensemble Member 04 - 2nd Domain - Units (kg/kg/s)
1e-05
6e-06
2e-06
-2e-06
-6e-06
-1e-05

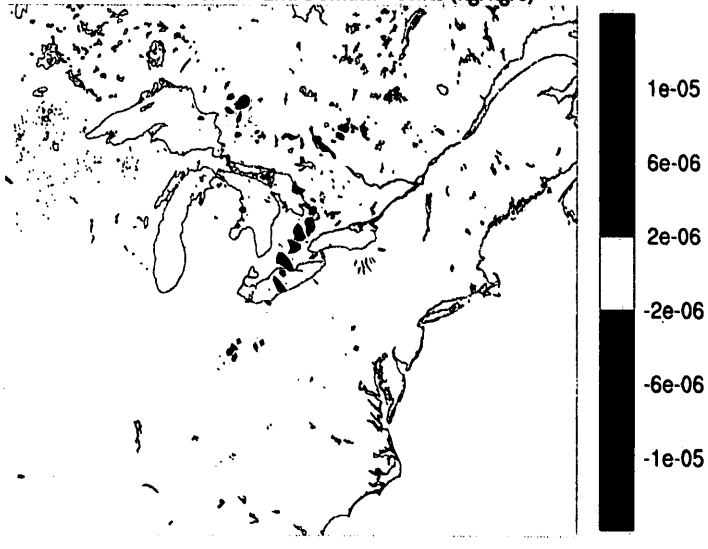
Low Level (850 hPa) Moisture Flux Divergence at 2400Z
for Ensemble Member 05 - 2nd Domain - Units (kg/kg/s)



Low Level (850 hPa) Moisture Flux Divergence at 2400Z
for Ensemble Member 06 - 2nd Domain - Units (kg/kg/s)

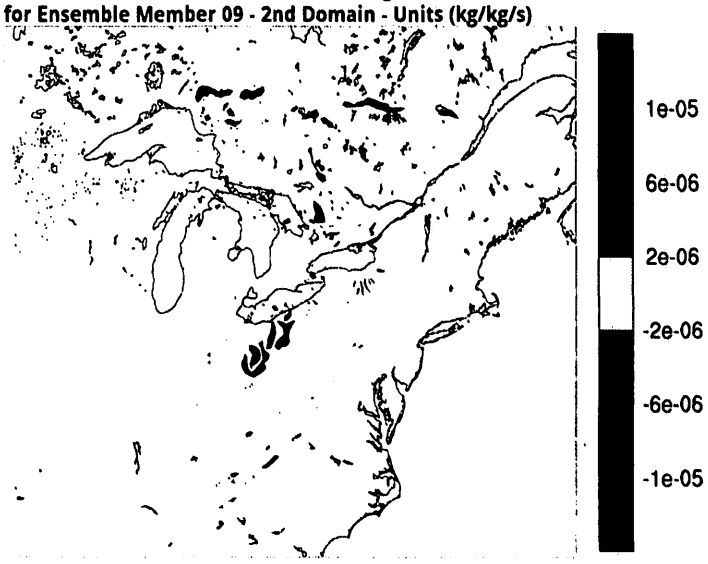


Low Level (850 hPa) Moisture Flux Divergence at 2400Z
for Ensemble Member 07 - 2nd Domain - Units (kg/kg/s)

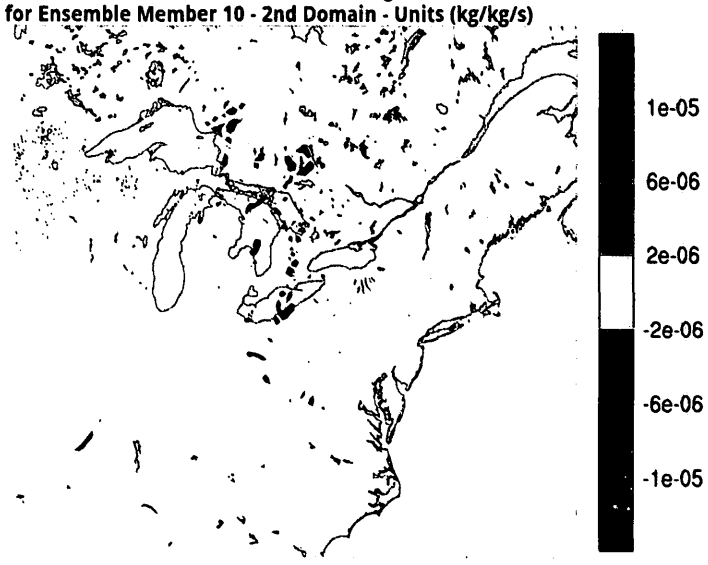


Low Level (850 hPa) Moisture Flux Divergence at 2400Z
for Ensemble Member 08 - 2nd Domain - Units (kg/kg/s)

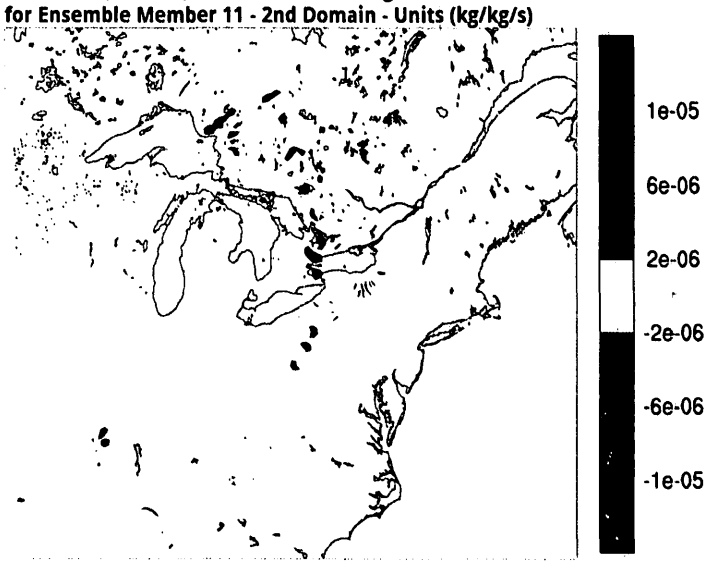
Low Level (850 hPa) Moisture Flux Divergence at 2400Z
for Ensemble Member 09 - 2nd Domain - Units (kg/kg/s)



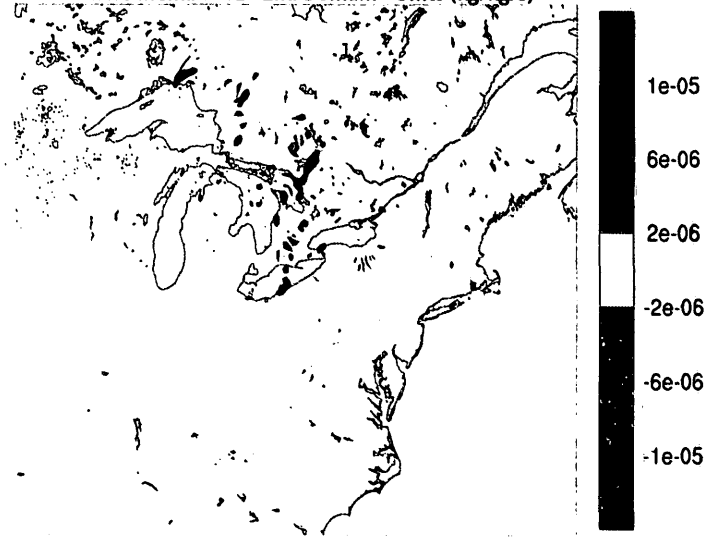
Low Level (850 hPa) Moisture Flux Divergence at 2400Z
for Ensemble Member 10 - 2nd Domain - Units (kg/kg/s)

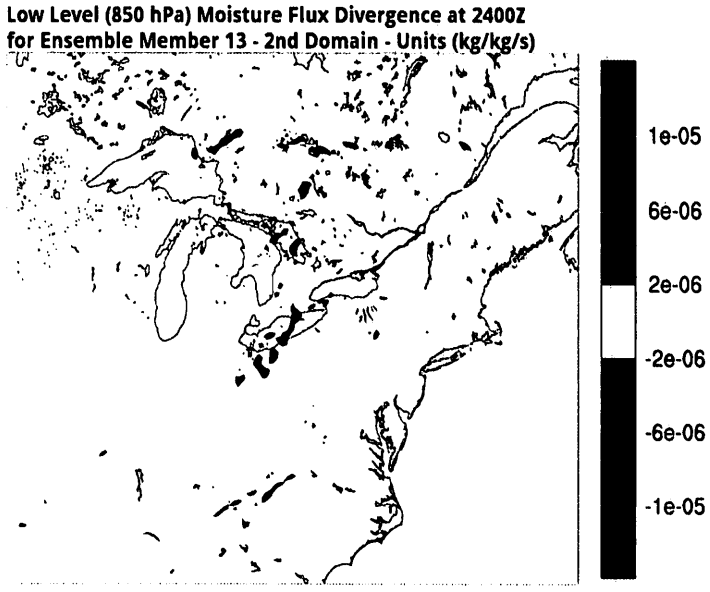


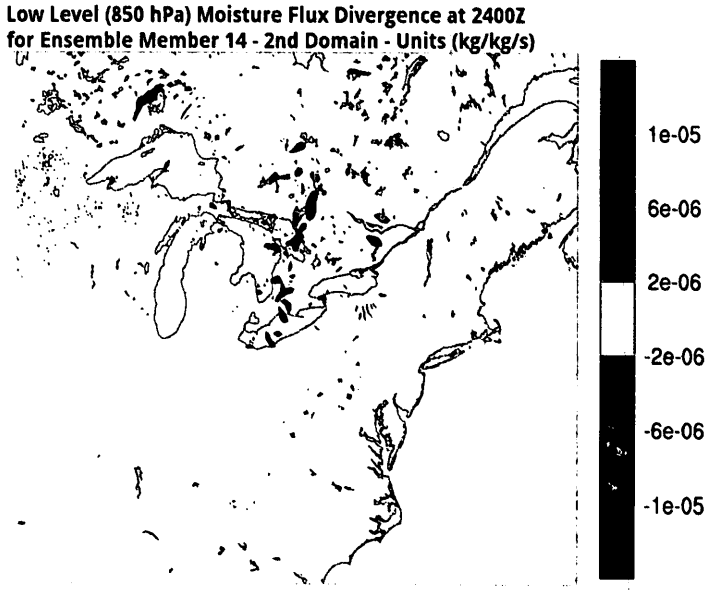
Low Level (850 hPa) Moisture Flux Divergence at 2400Z
for Ensemble Member 11 - 2nd Domain - Units (kg/kg/s)



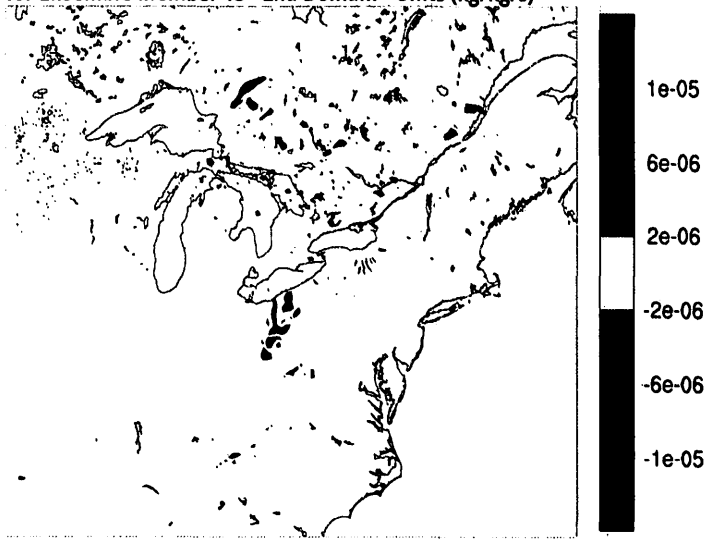
Low Level (850 hPa) Moisture Flux Divergence at 2400Z
for Ensemble Member 12 - 2nd Domain - Units (kg/kg/s)



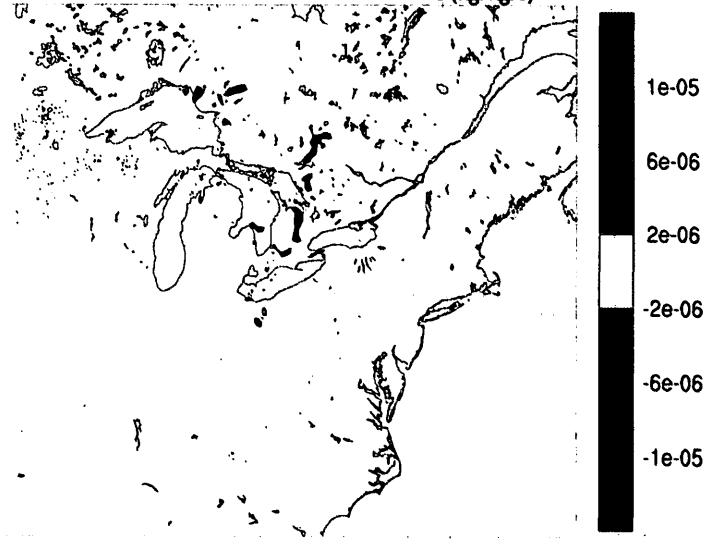

Low Level (850 hPa) Moisture Flux Divergence at 2400Z
for Ensemble Member 13 - 2nd Domain - Units (kg/kg/s)
1e-05
6e-06
2e-06
-2e-06
-6e-06
-1e-05


Low Level (850 hPa) Moisture Flux Divergence at 2400Z
for Ensemble Member 14 - 2nd Domain - Units (kg/kg/s)
1e-05
6e-06
2e-06
-2e-06
-6e-06
-1e-05

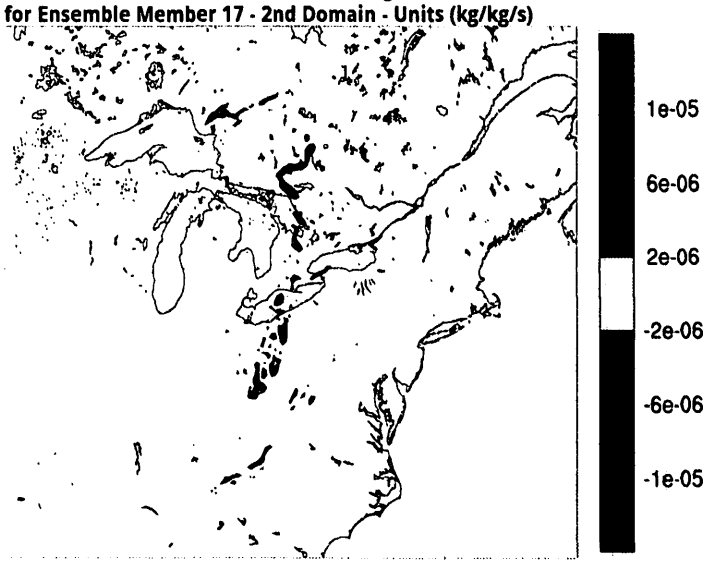
Low Level (850 hPa) Moisture Flux Divergence at 2400Z
for Ensemble Member 15 - 2nd Domain - Units (kg/kg/s)



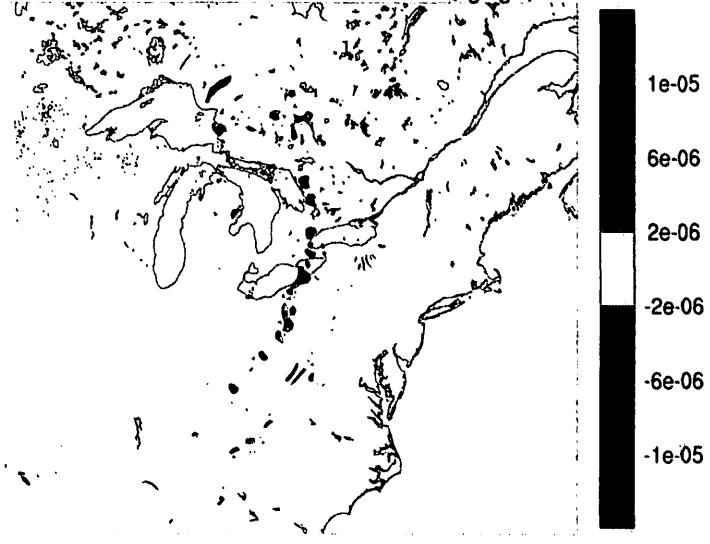
Low Level (850 hPa) Moisture Flux Divergence at 2400Z
for Ensemble Member 16 - 2nd Domain - Units (kg/kg/s)



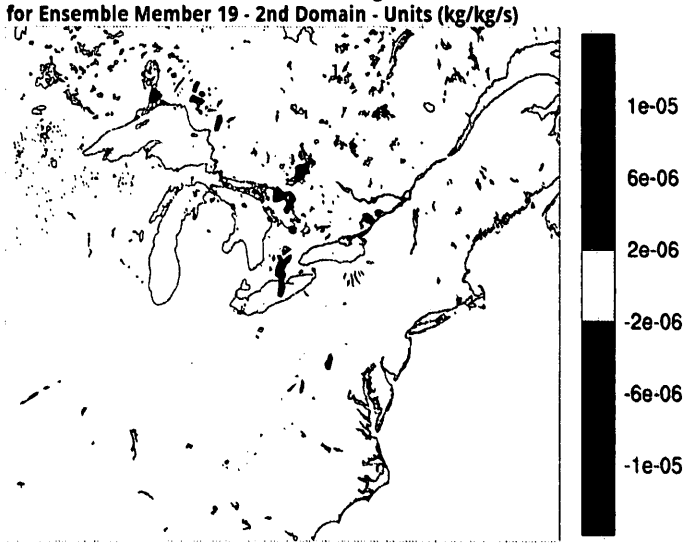
Low Level (850 hPa) Moisture Flux Divergence at 2400Z
for Ensemble Member 17 - 2nd Domain - Units (kg/kg/s)



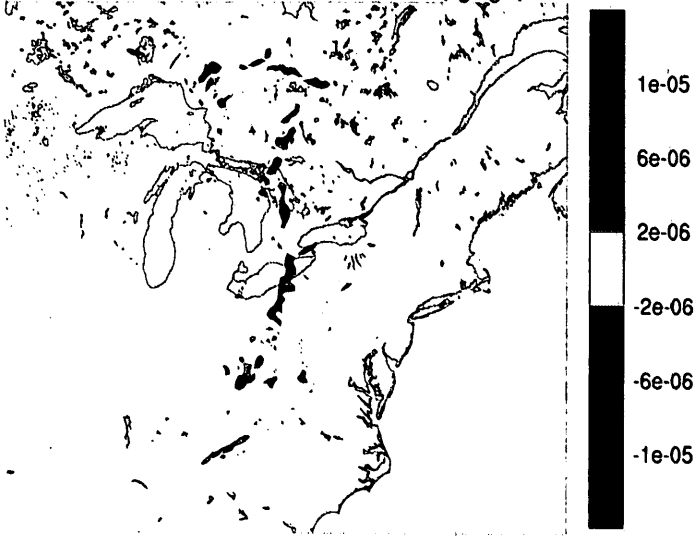
Low Level (850 hPa) Moisture Flux Divergence at 2400Z
for Ensemble Member 18 - 2nd Domain - Units (kg/kg/s)



Low Level (850 hPa) Moisture Flux Divergence at 2400Z
for Ensemble Member 19 - 2nd Domain - Units (kg/kg/s)



Low Level (850 hPa) Moisture Flux Divergence at 2400Z
for Ensemble Member 20 - 2nd Domain - Units (kg/kg/s)



Bibliography

Ancell, B., and G. Hakim, 2007: Comparing Adjoint and Ensemble Sensitivity Analysis with Applications to Observation Targeting, *Monthly Weather Review*, 135, 4117-4134

Banacos, P. C., and D. M. Schultz, 2005: The Use of Moisture Flux Convergence in Forecasting Convective Initiation: Historical and Operational Perspectives, *Weather and Forecasting*, 20, 351-366

Bishop, C. and D. Hodyss, 2007: Flow-Adaptive Moderation of Spurious Ensemble Correlations and Its Use In Ensemble-Based Data Assimilation , *Quarterly Journal of the Royal Meteorological Society*, 133, 2029-2044

Brooks, H. E., C. E. Doswell and R. Davies-Jones, 1993: Environmental Helicity and the Maintenance and Evolution of Low-Level Mesocyclones. *The Tornado: Its Structure, Dynamics, Predictions and Hazards*, Geophys. Monographs, No. 79, American Geophysical Union, 97-104

Bluestein, H. B., 1992: Principles of Kinematics and Dynamics. Vol. I, Synoptic–Dynamic
Meteorology in Midlatitudes. Oxford University Press, Oxford

Bluestein, H. B., 1992: Principles of Kinematics and Dynamics. Vol. I, Synoptic–Dynamic
Meteorology in Midlatitudes. Oxford University Press, Oxford

Davies-Jones, R. P., D. W. Burgess, and M. Forster, 1990: Test of helicity as a tornado forecast parameter. Preprints, *16th Conf. On Severe Local Storms*, Kananaskis Park, AB, Canada, Amer. Meteor. Soc., 588-592

Dudhia, J, 2011: State of Physics Parameterizations. *NWP Workshop on Model Physics with an Emphasis on Short-Range Prediction*, Camp Springs, Maryland, USA

Fankhauser, J. C., G. M. Barnes and M. A. LeMone, 1992: Structure of a Mid-Latitude Squall Line Formed in Strong Unidirectional Shear, *Monthly Weather Review*, 120, 237-260

Finley, Catherine A. Numerical Simulation of Intense Multi-Scale Vorticies Generated by Supercell Thunderstorms, Ph. D. Thesis, Colorado State University, 1998

Garcies, L., and V. Homar, 2009: Ensemble Sensitivities of the Real Atmosphere: Application to Mediteranian Intense Cyclones, *Tellus*, 61A, 394-406

Grimt, E. P. and C. F. Mass, 2007: Measuring the Spread-Error Relationship with a Probabilistic Approach: Stochastic Ensemble Results, *Monthly Weather Review*, 135, 203-221

Hamil, T. M. and S. J. Colluci, 1996: Verification of Eta–RSM Short-Range Ensemble Forecasts , *Monthly Weather Review*, 125, 1312-1327

Hamill, T.M. and C. Snyder, 2002: Using Improved Background- Error Covariances From an Ensemble Kalman Filter for Adaptive Observations, *Monthly Weather Review*, 130, 1552–1572.

Hanley, K. E., D. J. Kirshbaum, N. M. Roberts and G. Leoncini, 2011: Sensitivities of a Squall Line Over Central Europe In a Convective-Scale Ensemble, *Monthly Weather Review*, Preprint, DOI: 10.1175/MWR-D-12-00013.1

Hoskins, B. J., I. Draghici, and H. C. Davies, 1978: A New Look at the ω-Equation, *Quarterly Journal of the Royal Meteorological Society*, 111, 877-946

Hoskins, B. J. and M. A. Pedder, 1980: The Diagnosis of Mid-Latitude Synoptic Development, *Quarterly Journal of the Royal Meteorological Society*, 106, 707-719

Hohenegger, C. and C. Schar, 2007: Atmospheric Predictability at Synoptic Versus Cloud-Resolving Scales, *Bulletin of the American Meteorological Society*, 88, 1783-1793

Houtekamer, P. L., 1993: Global and Local Skill Forecasts, *Monthly Weather Review*, 121, 1834-1846

Junker, N. W., M. J. Brennan, F. Pereira, M. J. Bodner and R. H. Grum, 2009, Assessing the Potential for Rare Precipitation Events with Standardized Anomalies and Ensemble Guidance at the Hydrometeorological Prediction Center , *Bulletin of the American Meteorological Society*, 90, 445-453

Kalnay, E., 2003: Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, Cambridge, U. K.

Kong, F., K. Droegemeier and N. Hickmon, 2006: Multiresolution Ensemble Forecasts of an Observed Tornadic Thunderstorm System. Part I: Comparsion of Coarse- and Fine-Grid Experiments , *Monthly Weather Review*, 134, 807-833

Kong, F., K. Droegemeier and N. Hickmon, 2007: Multiresolution Ensemble Forecasts of an Observed Tornadic Thunderstorm System. Part II: Storm-Scale Experiments , Monthly *Weather Review*, 135, 759-782

Leoncini, G., R. S. Plant, S. L. Gray and P. A. Clark, 2010: Perturbation Growth at the Convective Scale for CSIP IOP18, *Quarterly Journal of the Royal Meteorological Society*, 136, 653-670

Lorenz, E., 1963: Deterministic Non-Periodic Flow, *Journal of the Atmospheric Sciences*, 20, 130-141

Lorenz, E., 1965: A Study of the Predictability of a 28-Variable Atmospheric Model, *Tellus*, 21, 321-333

Martin, J. E., 2006: Mid-latitude Atmospheric Dynamics: A First Course, J. Wiley and Sons, Chichester, U. K.

Markowski, P. and Y. Richardson, Mesoscale Meteorology in Midlatitudes, 2011: J. Wiley and Sonse, Chichester, U. K.

Peel, S. and L. Wilson, 2008: A Diagnostic Verification of the Precipitation Forecasts Produced by the Canadian Ensemble Prediction System, *Weather and Forecasting*, 23, 596-616

Shamrock, W. C., 2004: Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra, *Monthly Weather Review*, 134, 3019-3032

Shamrock, W. C., J. B. Dudhia, J. Gill, D. O. Barker, D. M. Barker, W. Wang, and J. G. Powers, 2005: A description of the Advanced Research WRF Version 2. NCAR Tech. Note TN-468 + STR., 88 pp.

Stensrud, D. J., H. E. Brooks, J. Du, S. Tracton, and E. Rogers, 1999: Using Ensembles for Short-Range Forecasting, *Monthly Weather Review*, 127, 433-446

Szunyogh, I., and Z. Toth, 2002: The Effect of Increases Horizontal Resolution on the NCEP Global Ensemble Mean Forecasts, *Monthly Weather Review*, 130, 1125-1133

Tang, Y., Z. Deng, X. Zhou and Y. Cheng, 2008: Interdecadal Variation of ENSO Predictability in Multiple Models, *Journal of Climate*, 21, 4811-4833

Torn, R. D., and G. J. Hakim, 2008: Ensemble-Based Sensitivity Analysis, *Monthly Weather Review*, 136, 663-677

Toth, Z. and E. Kalnay, 1993: Ensemble Forecasting at the NMC: The Generation of Perturbations, *Bulletin of the American Meteorological Society*, 74, 2317-2330

Tracton, S. and E. Kalna, 1993: Operational Ensemble Prediction at the National Meteorological Soceity: Practical Aspects, *NMC Notes: Weather and Forecasting*, 8, 379-398

Verret, R., L. Lefairvre, G. Pellerin, M. Klasa, F. Pithois, P. Houtekamer, and L. Wilson, 2002: Spread-Skill Relationship in the Canadian Ensemble Prediction System, 2002 AMS Symposium on Observations, Data Assimilation, and Probabilistic Prediction, Orlando, Florida, USA

Von Storch, H. and F. W. Zweirs, 1999, Statistical Analysis in Climate Research, Cambridge University Press, Cambridge, U. K.

Wallace, J. W. and P. V. Hobbs, 2006, Atmospheric Science, An Introductory Survey, Academic Press, London, U. K.

Wandishin, M. S., S. L. Mullen, D. J. Stensrud, and H. E. Brooks, 2001: Evaluation of a Short-Range Multimodel Ensemble System, *Monthly Weather Review*, 129, 729-747

Weisman, M. L. and R. Rotunno, 2000: The Use of Vertical Wind Shear Versus Helicity in Interpreting Supercell Dynamics. *Journal of the Atmospheric Sciences,* 57, 1452-1472

Whitaker, J. S. and A. F. Loughe, 1998: The Relationship between Ensemble Spread and Ensemble Mean Skill . *Monthly Weather Review*, 126, 3292-3302

Zhang, M. and D. L. Zhang, 2011: Sub-kilometer Simulation of a Torrential-Rain-Producing Mesoscale Convective System in East China. Part 1: Model Verification and Convective Organization. *Monthly Weather Review*, 140, 184-201