UCAR/RAL

Numerical Weather Prediction Model Sensitivity Analyses

Report Supporting MD&E per FY10 Technical Direction*

David Dowell, Frank McDonough, Bob Sharman, Matthias Steiner, Jenny Sun, Mei Xu

9/30/2010

"This research is in response to requirements and funding by the Federal Aviation Administration (FAA). The views expressed are those of the authors and do not necessarily represent the official policy or position of the FAA."

(*) Technical Direction Deliverables

10.5.19 Develop and refine techniques to assimilate radar radial velocity and reflectivity data through GSI and Rapid Refresh toward the HRRR

E2) Report on results from improved version of 13km/3km radar assimilation techniques for demonstration in FY10 exercise.

E3) Provide additional _report _on radar assimilation results for HRRR from winter 2007-08 case studies under the lead of ESRL with contributions from each organization.

10.5.24 Develop, test, and improve the 3-km WRF-based High-Resolution Rapid Refresh

E2- Collaborate with ESRL on analysis of convection-permitting forecast cases for 3-km ARW initialized with RUC-RR radar-initialized DFI grids. Draft and deliver summary of conclusions and results.

E3 - Deliver report summarizing all HRRR experimental results on sensitivity to physical parameterizations, initial conditions and assessment of HRRR results for key case studies from high impact weather days

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

1. INTRODUCTION

The FAA-supported model development and enhancement (MD&E) effort builds upon an ongoing, synergistic collaboration between the NOAA Earth System Research Laboratory (ESRL) Global Systems Division (GSD) and the NCAR Research Applications Laboratory (RAL). NCAR's expertise in numerical weather prediction (NWP) model development, data assimilation research, and weather analysis is utilized for carrying out a variety of sensitivity studies that are geared toward providing guidance for optimization of the NWP model configurations that are run operationally at NOAA/ESRL.

This year's focus on NCAR's side was expanded to take a broader look at the impact of particular model configurations on the detection and prediction of a variety of aviation hazards, including convective storms, turbulence, and in-flight icing. Four major aviation impact days were carefully selected so that the weather encountered on these days included convective storms, turbulence, and in-flight icing conditions. The purpose of that was to analyze identical days from different perspectives to document how changes in model configuration impact various aviation hazards. Most of the work in the past has been focused on convective storms, but optimization for convection may not necessarily yield a good configuration for turbulence and in-flight icing as well.

This report documents analyses carried out with a joint focus from convective storms, turbulence, and in-flight icing on four major impact days. The results discussed in the report provide an important first step toward optimization of the NWP model configurations that are run operationally at NOAA/ESRL, especially the 13 km WRF Rapid Refresh (WRF-RR) and the 3 km High-Resolution Rapid Refresh (HRRR) models. Clearly, these jointly focused, initial analyses need to be continued and expanded to achieve a solid basis from which to draw conclusive evidence on how to best configure the WRF-RR and HRRR models. Additional sensitivity studies are documented with a specific focus convective storm only.

Section 3 of the report documents the sensitivity analyses and the events selected. Sections 3, 4, and 5 discuss the various results obtained from a convective storms, turbulence, and in-flight icing perspective, respectively. An overall summary and outlook is provided in section 6. Some of the analyses are documented in an appendix to make the main body of the report more readable.

2. CASE SELECTION AND NUMERICAL EXPERIMENTATION

2.1 JOINT CASES FOR CONVECTION, TURBULENCE, AND IN-FLIGHT ICING

A majority of the analyses and forecast-sensitivity studies described in this report focus on the four cases listed in Table 2.1. These four cases were selected because they all included episodes of convection, turbulence, and in-flight icing that had a significant impacts on aviation. Thus, each case can be studied for multiple purposes. Three cases are during the cool season and feature large, strong baroclinic waves moving through the model domain. The fourth case (4 June 2008) is during the warm season, featuring widespread convection near a quasi-stationary front.

On 6 February 2008 (Figs. 2.1 and 2.2), an upper-level trough was moving eastward through the central US. In response, a surface low moved northeastward from Arkansas to Ohio. Near and ahead of an advancing cold front, convection was widespread from the Gulf coast to the Great Lakes. Turbulence was associated with both the baroclinic wave and associated convection in the central and eastern US and with another shortwave trough, moving eastward through the northern Rockies.

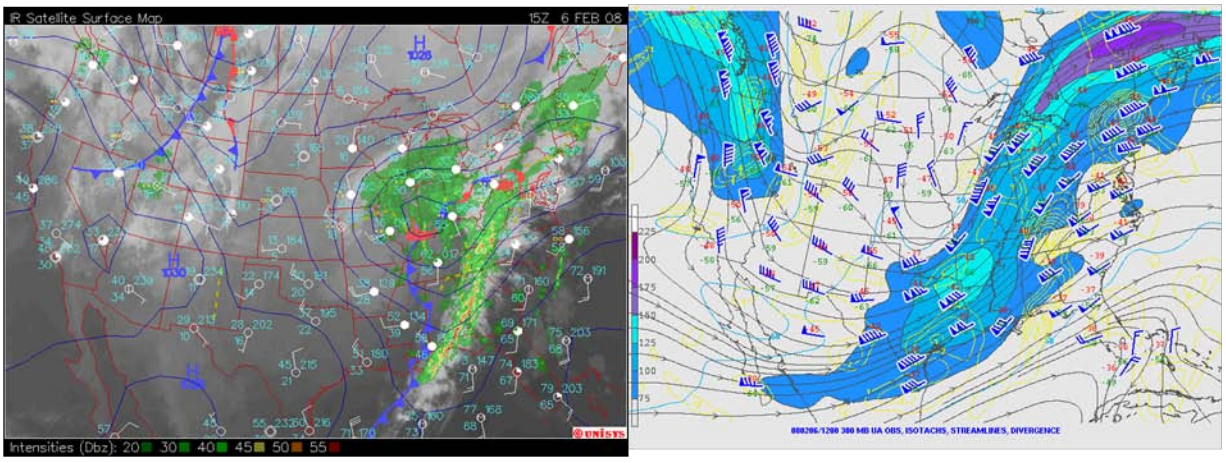


Figure 2.1, IR satellite, radar, and surface composite at 1500 UTC 6 February 2008 (left) and 300 mb upper-air analysis at 1200 UTC 6 February 2008 (right).

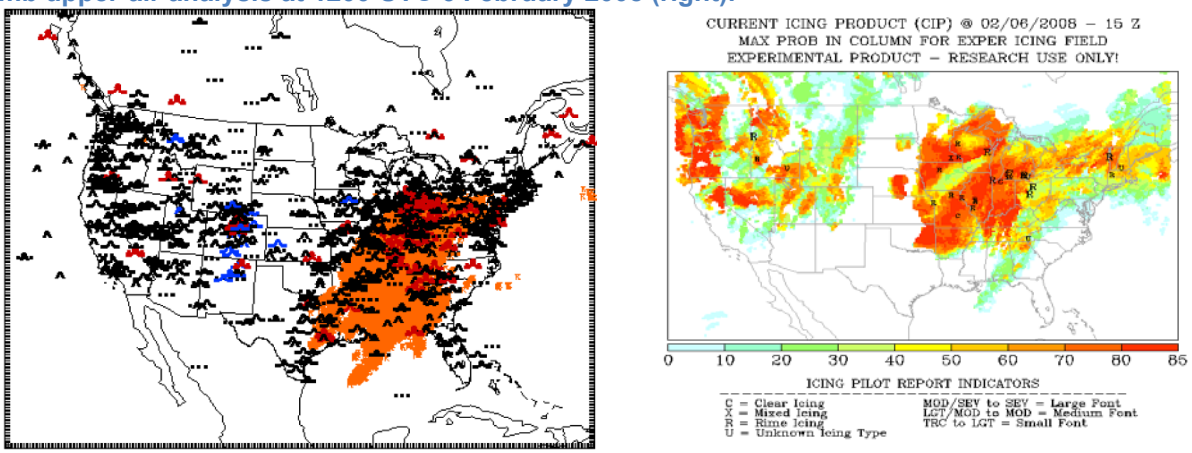


Figure 2.2, Pilot reports of turbulence (PIREPS) between 0000 UTC 6 February to 0000 UTC 7 February 2008(left), PIREPs for Icing and Current Icing Product (right)

Table 2.1, Summary of weather events and forecast-sensitivity experiments for joint cases for convection, turbulence, and in-flight icing studies.

Date	Model Initialization Time (UTC)	Phenomena of Interest	Forecast-Sensitivity Experiments
6 February 2008	1200	Convection: Line storms Southeast Icing: Warm front Midwest to Northeast Turbulence: Widespread turbulence associated with line thunderstorms over the SE; mountain wave turbulence over the Rockies mountain chain	Resolution: 3km (HRRR), 12km (RR) PBL: MYJ, MYNN, YSU
4 June 2008	1200	Convection: Severe isolated storms Central Plains; multiple MCSs Northeast Icing: Cold advection Pacific Northwest and northern/central Rockies Turbulence: Turbulence associated with storms over the central plains; mountain wave turbulence over much of the West	Resolution: 3km (HRRR), 12km (RR) PBL: MYJ, MYNN, YSU
9 October 2009	1200	Convection: Line and ordinary storms Southeast Icing: Various quadrants of central US synoptic cyclone Turbulence: Reported severe turbulence associated with convective line storms over Plains.	Resolution: 3km (HRRR), 12km (RR) PBL: MYJ, MYNN, YSU
9 December 2009	1200	Convection: Line storms Southeast Icing: Strong low over Midwest Turbulence: Severe turbulence associated with line storms over the Southeast, plus severe mountain wave turbulence over the Colorado Rockies.	Resolution: : 3km (HRRR), 12km (RR) PBL: MYJ, MYNN, YSU

On 4 June 2008 (Figs. 2.3 and 2.4), there was an upper-level trough over the western US and unseasonably strong westerly flow over much of country. A WSW-ENE oriented quasi-stationary surface front extended from the central high plains to the Great Lakes, the low-level atmosphere was very moist near and south of the front. Multiple rounds of significant convection occurred from the northeast US to the central high plains. Turbulence was associated with this convection and also with the trough in the southwest US.

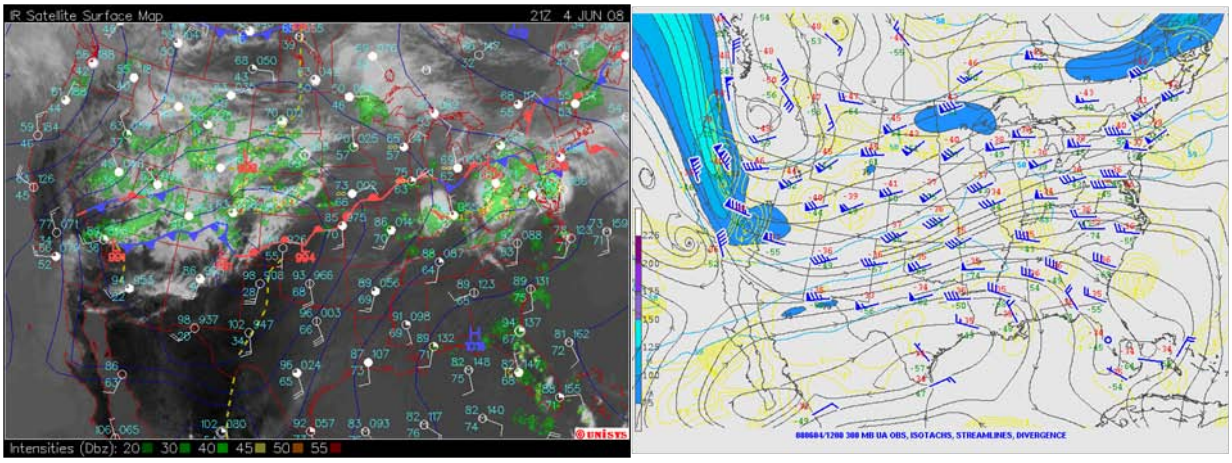


Figure 2.3, IR satellite, radar, and surface composite at 2100 UTC 4 June 2008 (left) and 300 mb upper-air analysis at 1200 UTC 4 June 2008 (right).

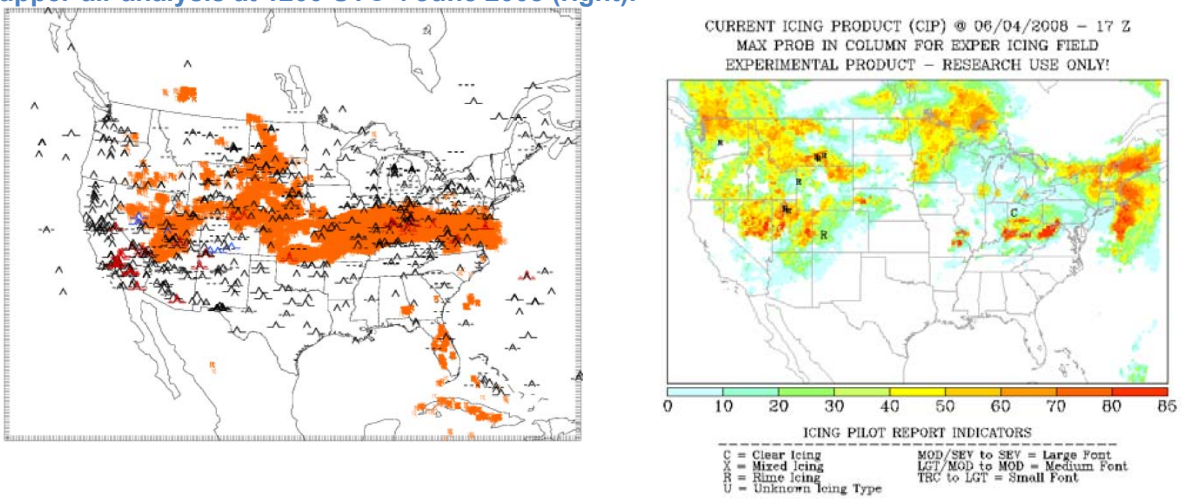


Figure 2.4, PIREPS 0000 UTC 4 June to 0000 UTC 5 June 2008 for turbulence (left); PIREPs for Icing and Current Icing Product (right)

On 9 October 2009 (Figs. 2.5 and 2.6), a strong jet aloft extended from the Southern Plains to Great Lakes. At the surface, a cold front that was roughly parallel to the jet advanced eastward. The front encountered moist low-level air in the southern US, and convective lines formed and persisted near the front. Reported turbulence was probably associated with the convection, the upper-level jet in the central US, and a shortwave trough moving southeastward through the northern Rockies.

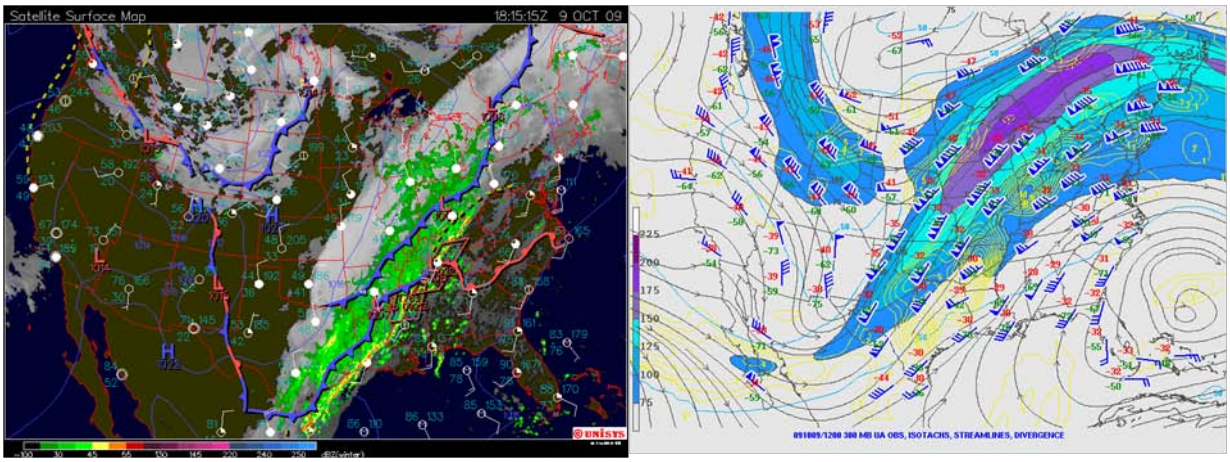


Figure 2.5, IR satellite, radar, and surface composite at 1815 UTC 9 October 2009 (left) and 300 mb upper-air analysis at 1200 UTC 9 October 2009 (right).

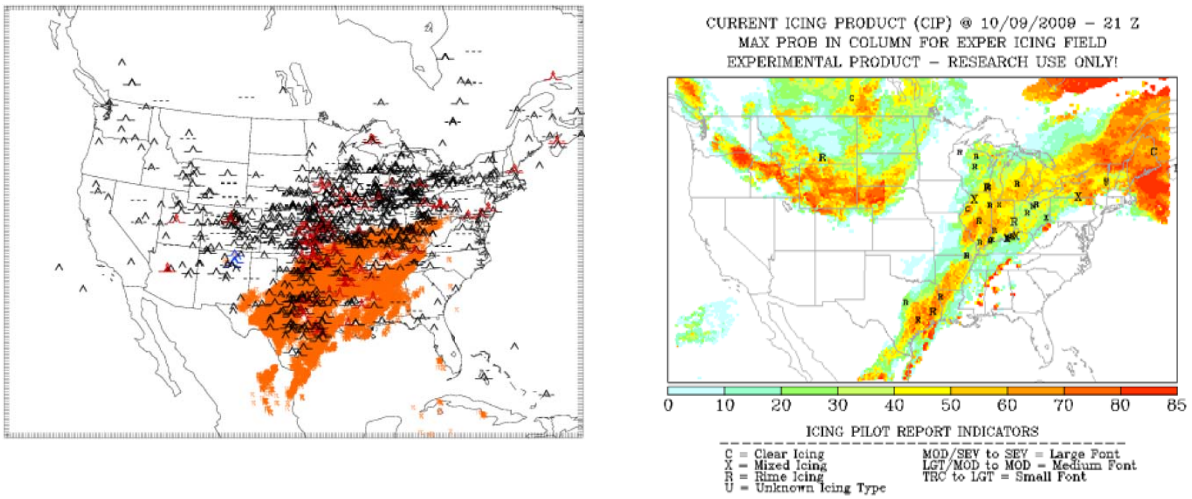


Figure 2.6, PIREPS 0000 UTC 9 October to 0000 UTC 10 October 2009 (left); PIREPs for Icing and Current Icing Product (right)

On 9 December 2009 (Figs. 2.7 and 2.8), there was a large, occluded cyclone over the Great Lakes. At the surface, warm, moist air was confined to the Southeast and mid-Atlantic regions, where most of the convection occurred. Strong westerly flow aloft was observed through the southern half of the US. Reported turbulence was associated with the convection in the Southeast, the advancing shortwave trough and precipitation in the Midwest and Northeast, and the strong westerly flow over Colorado.

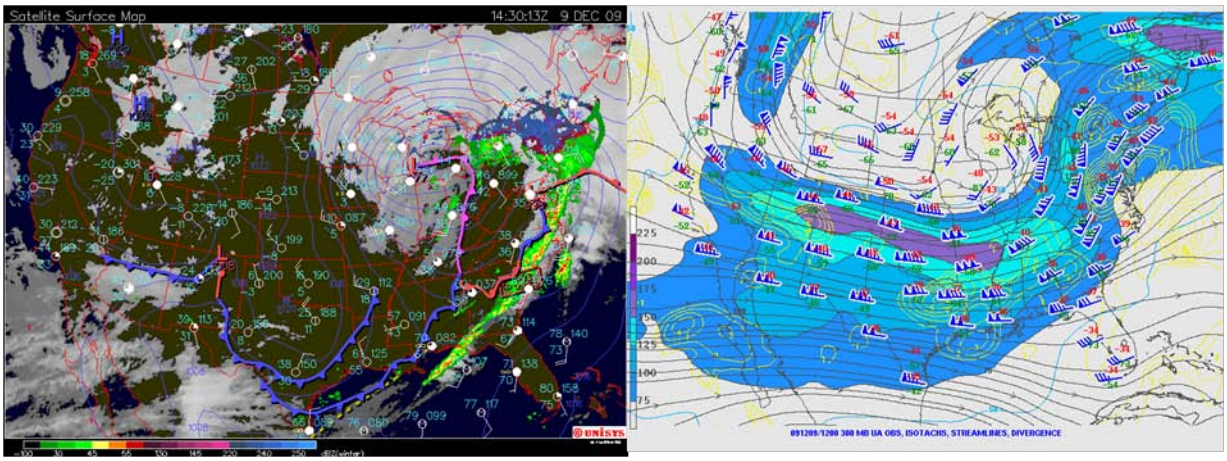


Figure 2.7, IR satellite, radar, and surface composite at 1430 UTC 9 December 2009 (left) and 300 mb upper-air analysis at 1200 UTC 9 December 2009 (right)

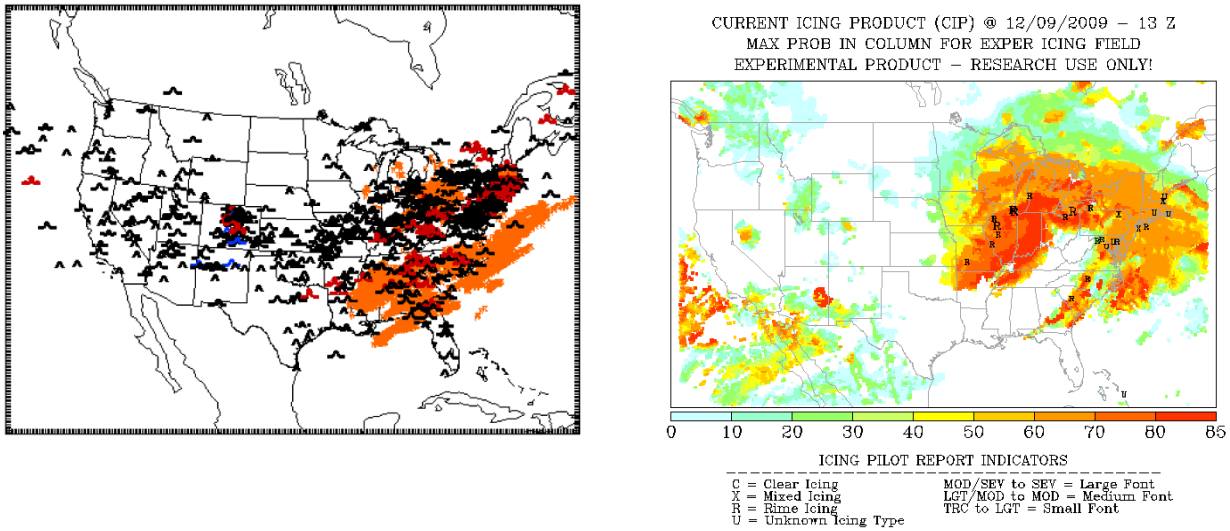


Figure 2.8, PIREPS 0000 UTC 9 December to 0000 UTC 10 December 2009 (left) and PIREPs for Icing and Current Icing Product (right).

For each of the four joint cases, baseline 12-h forecasts were produced with the WRF-ARW model, emulating both the NOAA/ESRL Rapid-Refresh (RR) model and High-Resolution Rapid-Refresh (HRRR) model. Parameters for these forecast are summarized in Table 2.2. The experiments here were configured as similarly as possible to the operational models run in real-time. Convection, turbulence, and icing aspects of these baseline forecasts are evaluated qualitatively and quantitatively in subsequent sections of this report.

The parallel RR-like and HRRR-like baseline simulations provide an opportunity to examine the influence of model resolution on the convection, turbulence, and icing forecasts. The former model configuration uses 13-km horizontal grid spacing and convective parameterization whereas the latter model configurations uses 3-km grid spacing and no convective parameterization (Table 2.2). Based on previous studies, we would expect significant differences in the forecasts, considering the 3-km model's ability to produce convective storms

explicitly and to resolve better orographic processes. Relative to the 13-km forecasts, the 3-km forecasts would be expected to have more realistic updraft strengths, explicit representation of how convective storms propagate (dynamically induced pressure gradients, cold pools, etc.), more realistic cloud and hydrometeor concentrations, significantly different precipitation rates, and improved representation of dynamical (e.g., turbulence) processes.

Table 2.2, WRF model configuration for baseline 3-km (HRRR-like) and 13-km (RR-like) forecasts.

Model Parameter	**"HRRR"**	**"RR"**
WRF-ARW model version	3.1.1 with NOAA/GSD code changes*	3.1.1 with NOAA/GSD code changes*
Horizontal grid	CONUS; Δx = 3 km 1680 $\times$ 1100 gridpoints	CONUS; Δx = 13 km 400 $\times$ 250 gridpoints
Vertical levels	51	51
Initialization	RUC	RUC
PBL scheme	MYJ	MYJ
Land-surface model	RUC	RUC
Microphysics scheme	Thompson	Thompson
Convective scheme	none	Grell-Devenyi
Radiation schemes	Dudhia (shortwave) RRTM (longwave)	Dudhia (shortwave) RRTM (longwave)
6th order horizontal diffusion coefficient	0.25	0.25

(*) For real-time HRRR forecasts in summer 2009, NOAA/GSD implemented minor changes to the WRF3.1.1 code involving the RUC land-surface model, diagnostic output fields, and digital filter initialization. This modified code was used for the forecasts described in this report.

In addition to resolution, forecast sensitivity to the choice of the planetary boundary layer (PBL) parameterization is also examined for the baseline forecasts (Table 2.1). With model parameters otherwise identical to those in the RR-like (13-km) baseline forecasts, additional forecasts have been produced with both the MYNN and YSU PBL parameterizations for the four joint cases. PBL parameterization is recognized as a significant challenge in quantitative precipitation forecasting and weather forecasting in general at the surface, and thus is an important topic for investigation here for model development and evaluation.

2.2 ADDITIONAL CASES FOR CONVECTION

In addition to the four joint cases (Table 2.1), five additional cases are used for convection-specific studies (Table 2.3). All are during the warm season, and the emphasis of the model evaluation is on predictions of specific regional convective events.

Table 2.3, Summary of weather events and forecast-sensitivity experiments for additional cases used for convection studies.

Date	Model Initialization Time (UTC)	Weather Phenomena	Forecast-Sensitivity Experiments
15 May 2009	1200	Ordinary storms Southeast; MCS Midwest	6th order horiz. diffusion: scheme 1, coeff. 0.12; scheme 2, coeff. 0.12; scheme 2, coeff. 0.25
6 June 2009	0000	Nebraska convective system	storm-scale initialization: DDFI, cloud analysis, both, neither
10 August 2009	1200	Line and cluster storms NY to OK	Horiz. grid: Δx = 3 km, 6 km, 9 km, 12 km 6th order horiz. diffusion coeff.: 0.12, 0.25, 0.50
12 August 2009	1200	Ordinary storms Southeast	Horiz. grid: Δx = 1 km, 3 km
4 August 2010	1000, 1300	Ordinary storms Southeast	6th order horiz. diffusion coeff.: 0.00, 0.12, 0.25

On 15 May 2009 (Fig. 2.9), a W-E oriented quasi-stationary surface front existed in the Midwest while a broad shortwave trough aloft approached from the northern Rockies. Convective lines and clusters developed near the front. Farther southeast, widespread ordinary convective cells formed in the moist, weak-flow regime.

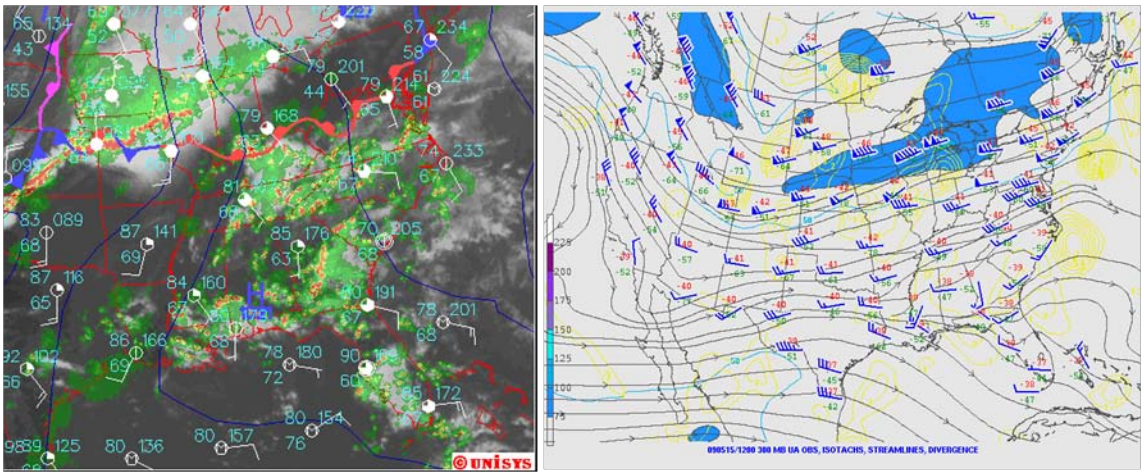


Figure 2.9, IR satellite, radar, and surface composite at 2100 UTC 15 May 2009 (left) and 300 mb upper-air analysis at 1200 UTC 15 May 2009 (right).

On 6 June 2009 (Fig. 2.10), there was a ridge aloft over the Plains and a shortwave trough in the West. Convective storms (supercell thunderstorms and mesoscale convective systems) formed on the central High Plains near a surface front.

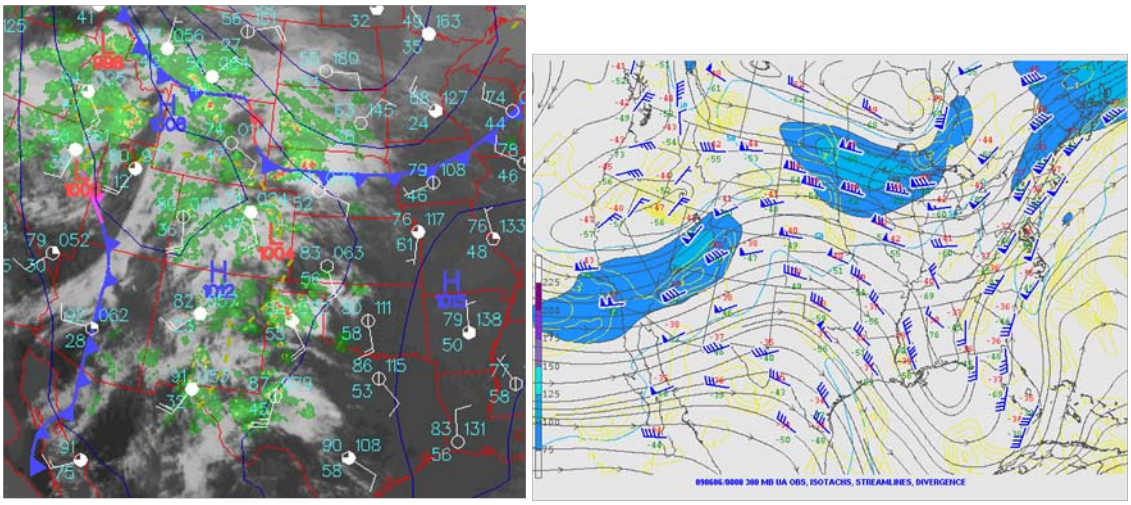


Figure 2.10, IR satellite, radar, and surface composite (left) and 300 mb upper-air analysis at 0000 UTC 6 June 2009 (right).

Three cases from August 2009 and 2010 (Figs. 2.11 to 2.13) feature widespread ordinary convective storms and a few convective clusters in the southeast US. In the 10 August 2009 and 4 August 2010 cases, these storms formed beneath upper-level ridges, far removed from

major surface boundaries. In contrast, in the 12 August 2009 case, the storms formed as a upper-level trough and surface front were approaching the region.

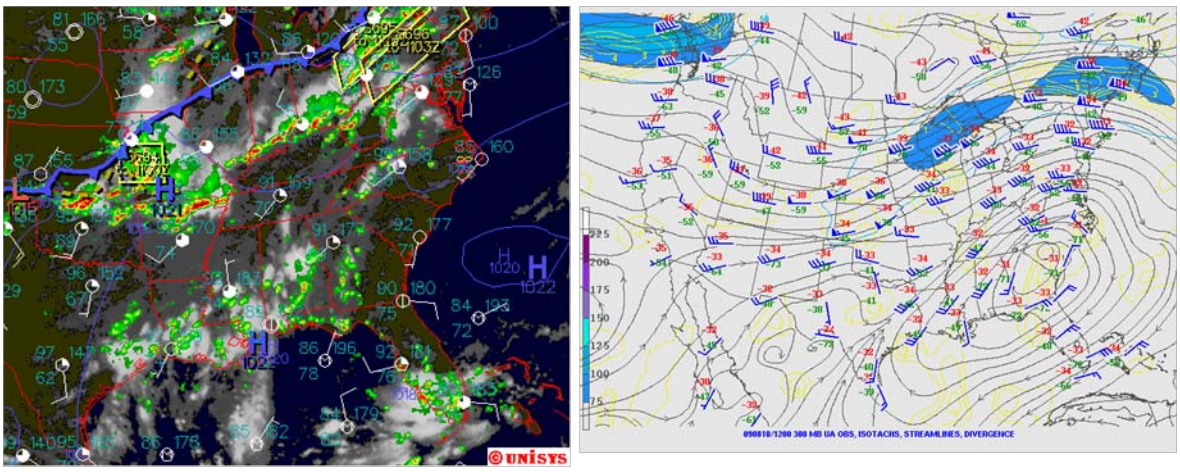


Figure 2.11, IR satellite, radar, and surface composite at 2130 UTC 10 August 2009 (left) and 300 mb upper-air analysis at 1200 UTC 10 August 2009 (right).

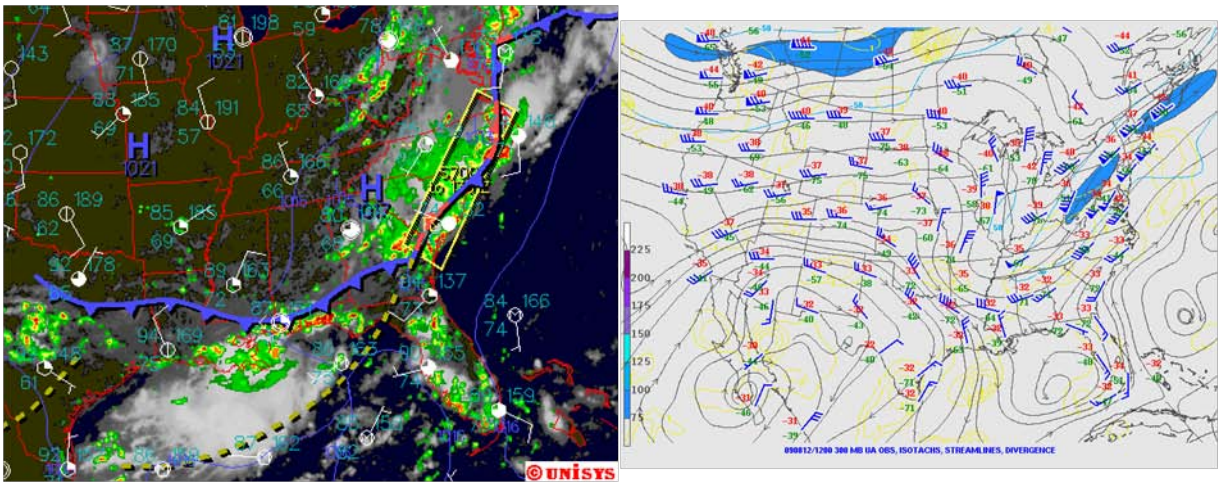


Figure 2.12, IR satellite, radar, and surface composite at 1930 UTC 12 August 2009 (left) and 300 mb upper-air analysis at 1200 UTC 12 August 2009 (right).

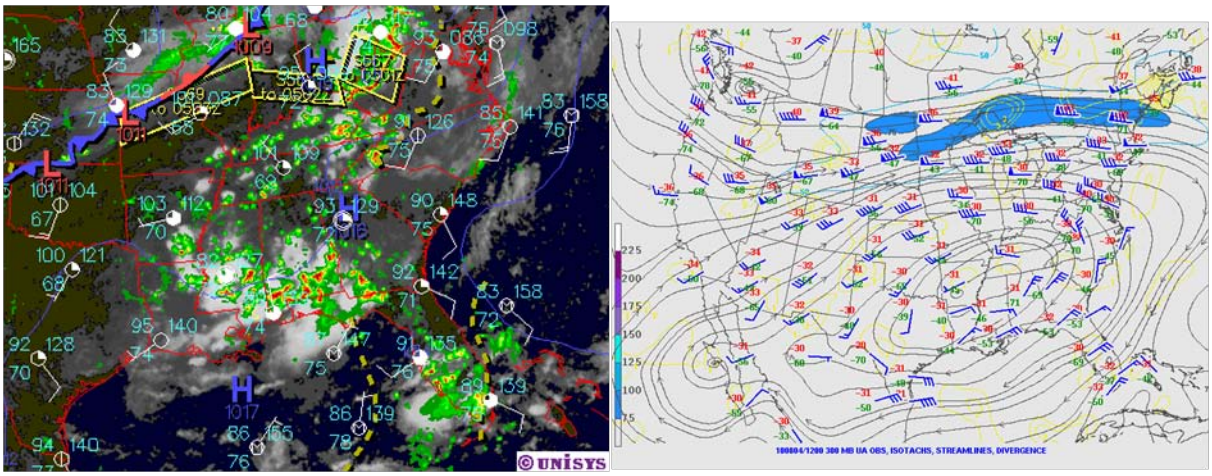


Figure 2.13, IR satellite, radar, and surface composite at 2130 UTC 4 August 2010 (left) and 300 mb upper-air analysis at 1200 UTC 4 August 2010 (right).

Last year's report (September 2009) illustrated that forecasts initialized with high-quality but different analyses are quite sensitive to these initial conditions, particularly for 0-8 h lead times. In those experiments, high-resolution WRF forecasts were initialized with NAM, GFS, and RUC analyses. These results motivate continued attention to making the best use of available observations in model initialization. Along these lines, the current report includes an investigation of model initialization with radar-reflectivity data.

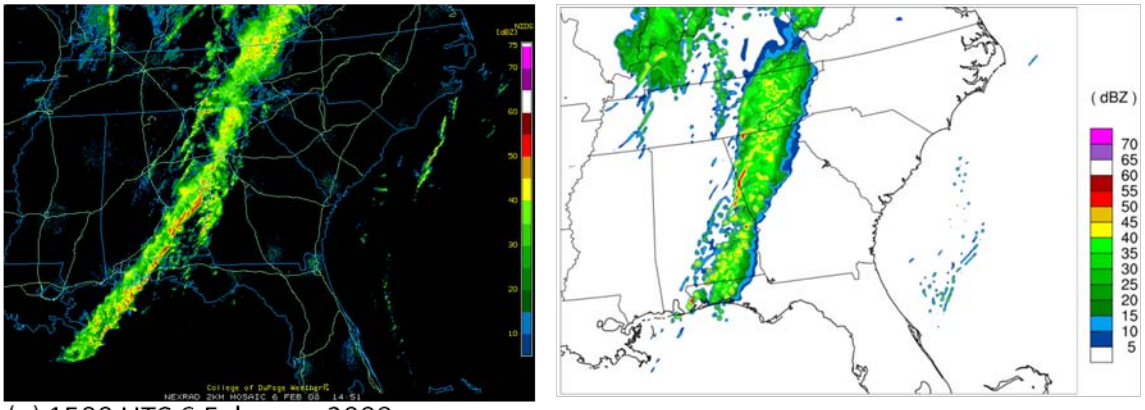
Other forecast-sensitivity studies with the five additional cases (Table 2.3) focus on horizontal diffusion (both the numerical scheme and the coefficient choice) and model horizontal grid spacing. These factors would be expected to affect potentially the structure of individual convective storms, the number of storms, and the representation of small-scale phenomena such as orographic processes.

3. ANALYSES FOCUSED ON CONVECTION

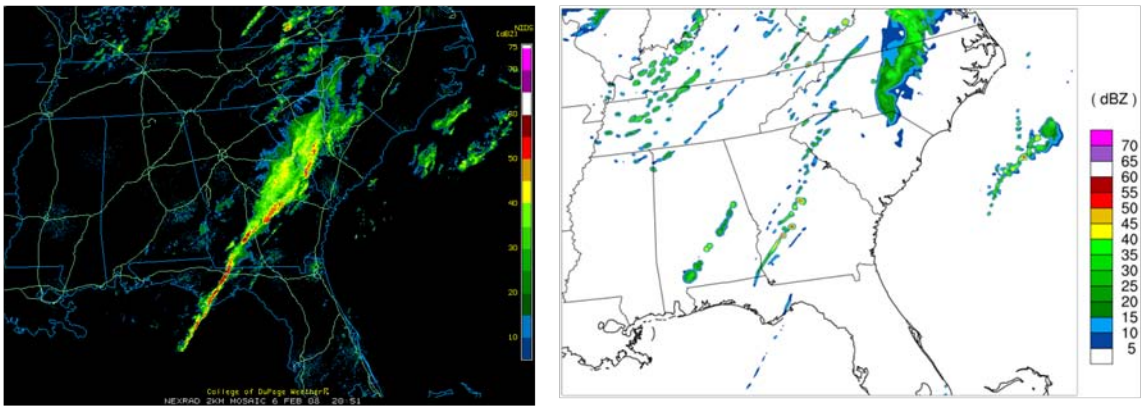
3.1 QUALITATIVE ASSESSMENT OF BASELINE AND SENSITIVITY RUNS

3.1.1 Qualitative Assessment of 3-km Baseline Runs

In this section, 3-km forecast performance is examined qualitatively for the most significant regional convective events on the four joint case days (Table 2.1). Appendix A.1 contains supplementary plots of observed and predicted rain accumulation on a national scale.



(a) 1500 UTC 6 February 2008

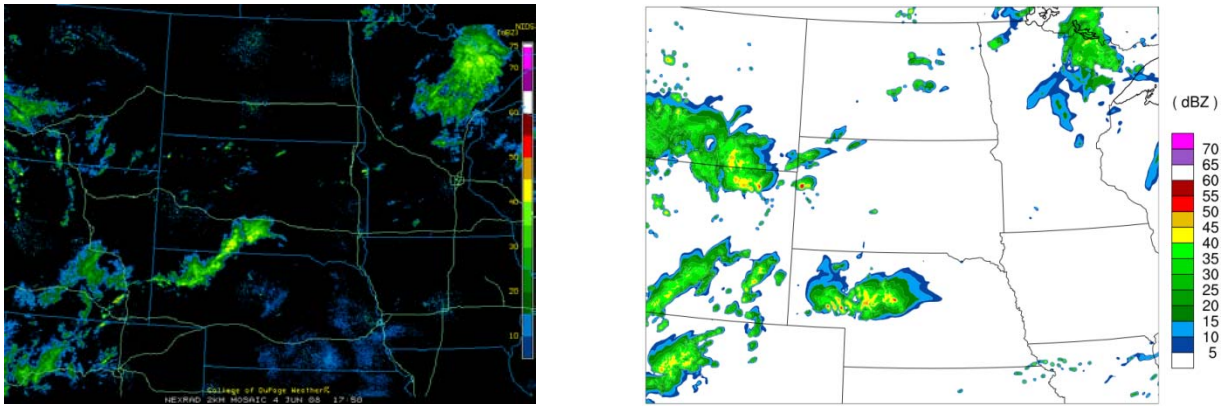


(b) 2100 UTC 6 February 2008

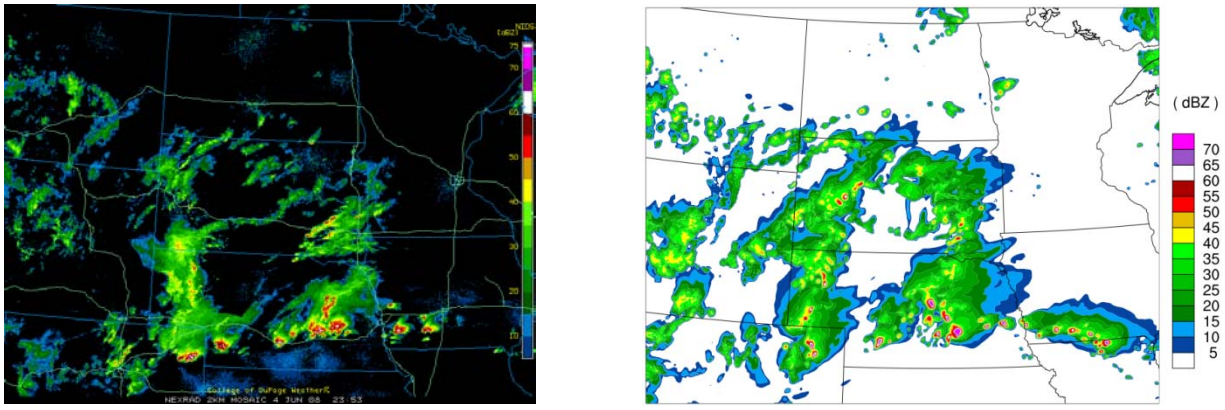
Figure 3.1.1, Observed (left) and forecast (right) composite reflectivity (dBZ) in the southeast US for the 6 February 2008 case.

On 6 February 2008, a long convective line from the Gulf of Mexico to Kentucky during the morning (Fig. 3.1.1.a) transitioned to a shorter line in the Southeast during the afternoon (Fig. 3.1.1.b). The baseline HRRR-like simulation was successful in predicting the position of the convective line throughout the day. This result is perhaps not surprising since strong synoptic-scale forcing (cf. Fig. 2.1), which is typically resolved well by the current generation of mesoscale numerical weather prediction models, likely played a significant role in determining convective storm locations. However, the baseline forecast tended to underpredict the amount

of convection. Both in the morning (Fig. 3.1.1.a) and afternoon (Fig. 3.1.1.b), the model underpredicted the north-south extent of the convective line and the concentration of high-reflectivity cells within the line.



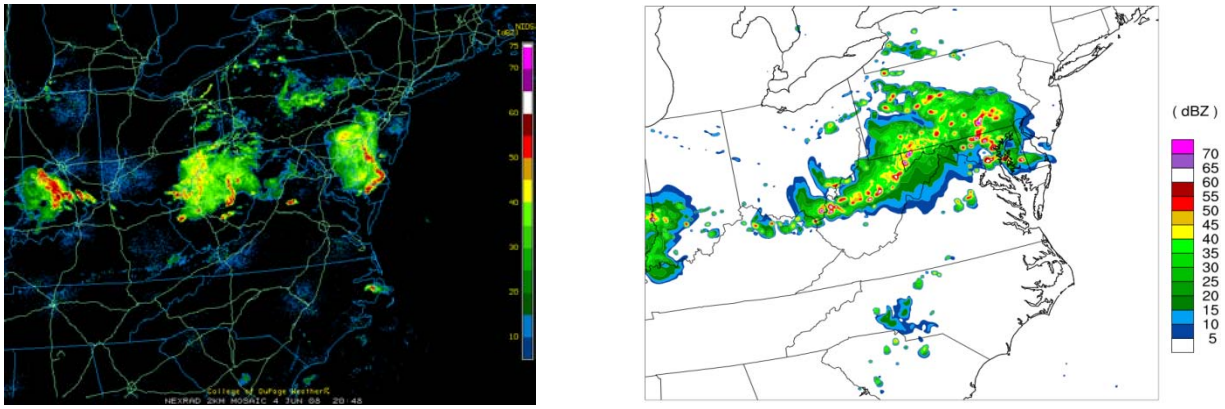
(a) 1800 UTC 4 June 2008



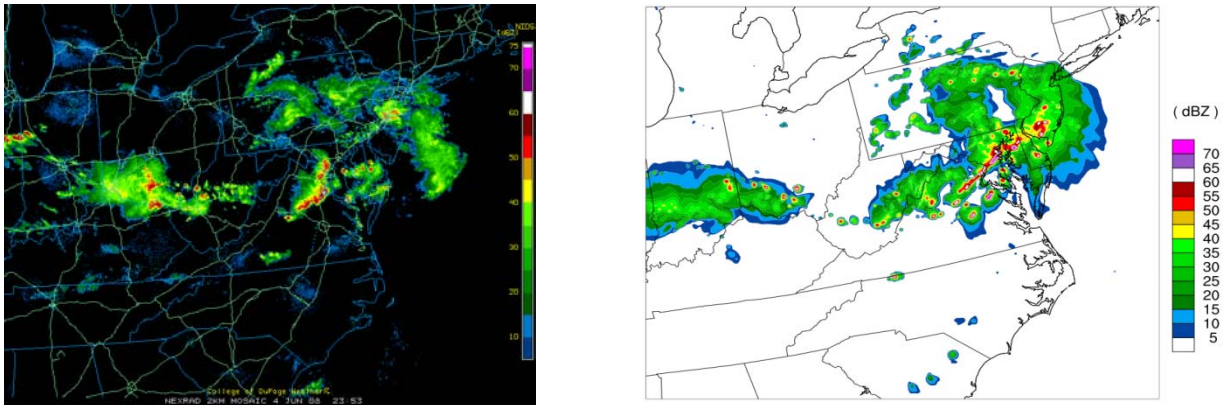
(b) 0000 UTC 5 June 2008

Figure 3.1.2, As in Fig. 3.1.1, except for the northern plains in the 4 June 2008 case.

On 4 June 2008, numerous severe thunderstorms formed in the Central Plains in the afternoon (Fig. 3.1.2) north of a quasi-stationary front (cf. Fig. 2.3). The 3-km baseline forecast produced strong convective storms 1-2 h too early in this region (Nebraska, northwest South Dakota, and northeast Wyoming in Fig. 3.1.2.a) but was otherwise quite accurate with the mesoscale aspects of the convection. In the 12-h forecast valid at 0000 UTC, the model depicted very realistic storm structures, and there was a good correspondence between predicted and observed storms locations, including the axis of most intense storms from northeast Colorado through southern Nebraska into southwest Iowa (Fig. 3.1.2.b).



(a) 2100 UTC 4 June 2008

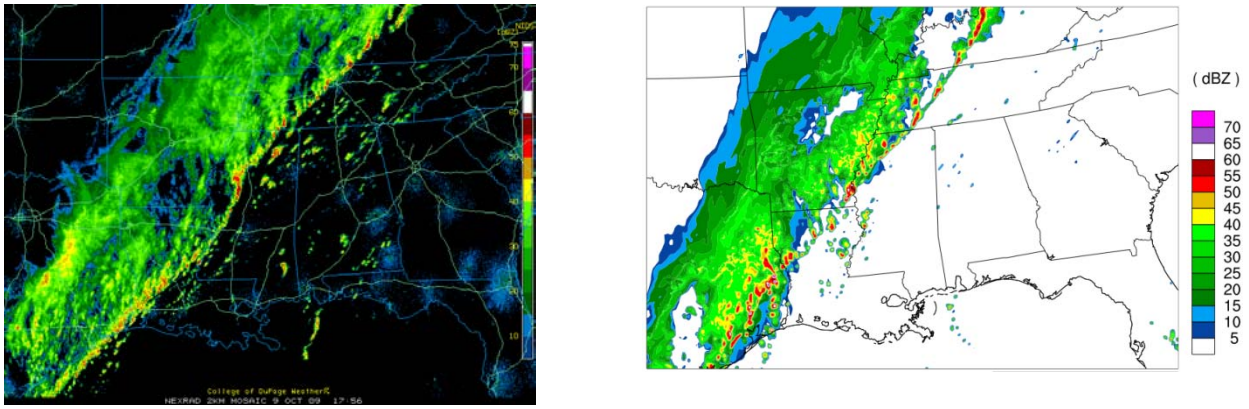


(b) 0000 UTC 5 June 2008

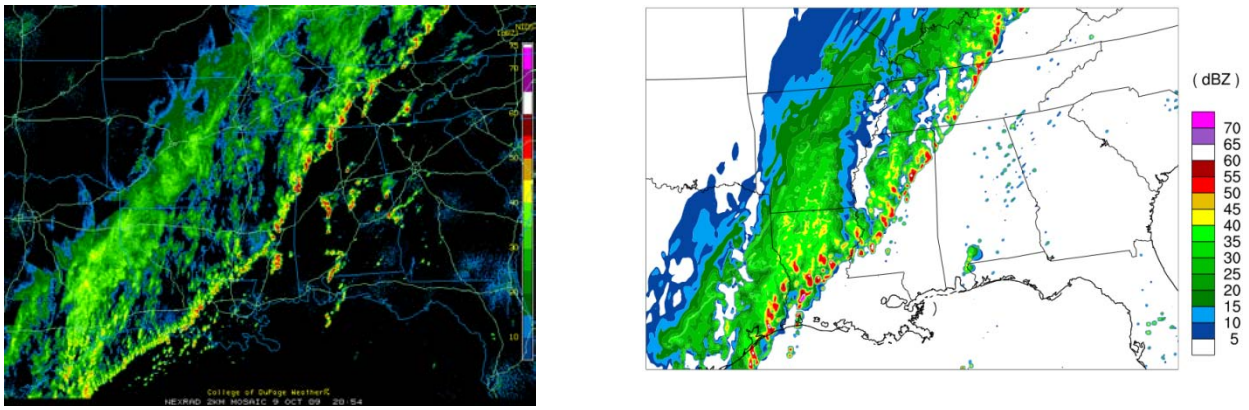
Figure 3.1.3, As in Fig. 3.1.1, except for the mid-Atlantic region in the 4 June 2008 case.

Strong convective storms also occurred in the northeast US on 4 June 2008 (Fig. 3.1.3). Near the surface front (cf. Fig. 2.3), multiple rounds of mesoscale convective systems and other convective storms moved eastward through the region. The timing and location of individual convective systems were not predicted accurately by the baseline 3-km forecast (Fig. 3.1.3). However, the model rather accurately predicted the swath of convective weather and the convective modes during the afternoon.

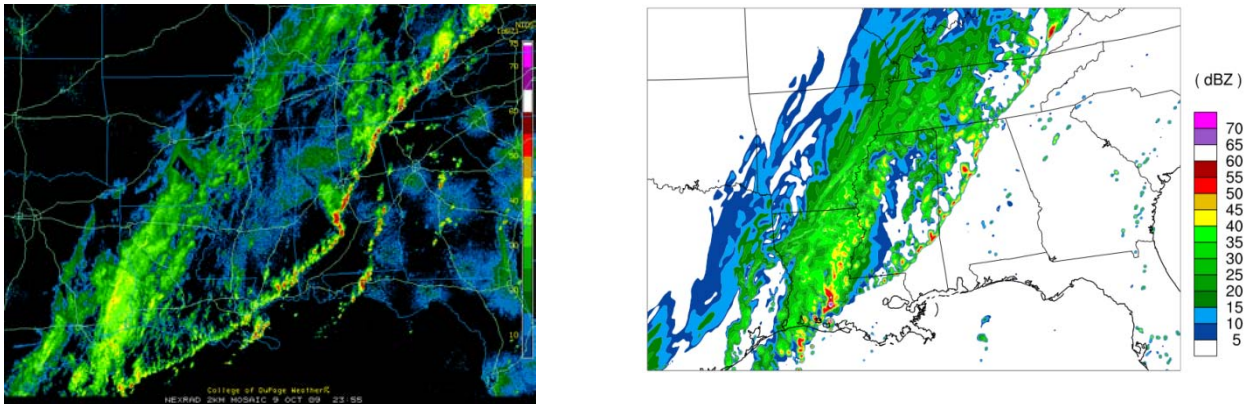
The 3-km baseline forecast for the southeast US on 9 October 2009 was generally quite good (Fig. 3.1.4). The model was accurate with the location of the main convective line, perhaps an indication of the enhanced predictability that comes with strong forcing on the well-resolved synoptic scale. The model was also quite accurate with predicting the overall character of precipitation pattern, including the thin, broken leading line and the trailing “stratiform” precipitation region. One negative aspect of the baseline forecast is that both the number and intensity of isolated cells in Alabama and Georgia, ahead of the main line, are underpredicted.



(a) 1800 UTC 9 October 2009



(b) 2100 UTC 9 October 2009

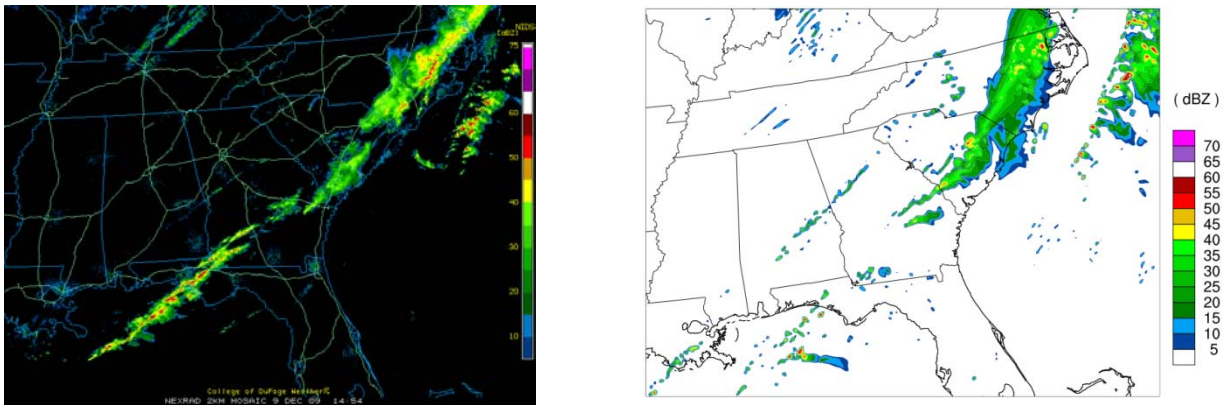


(c) 0000 UTC 10 October 2009

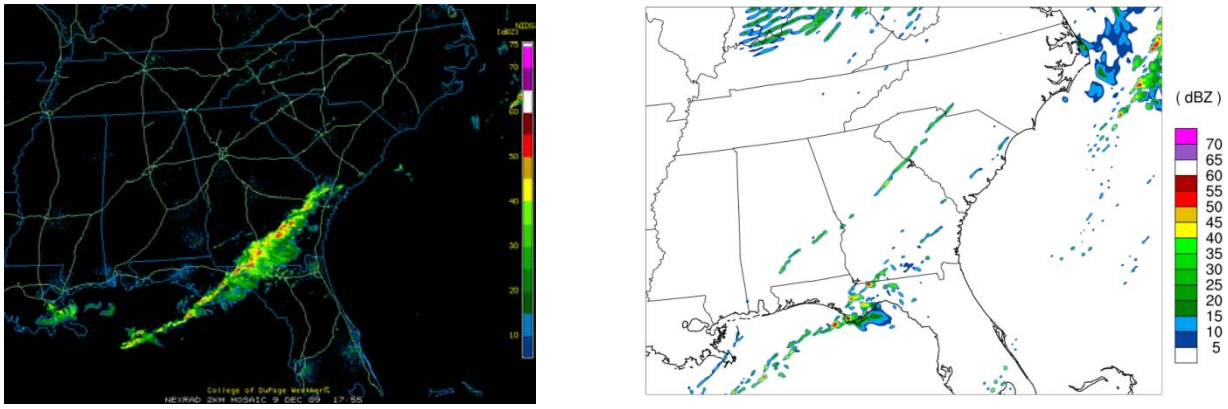
Figure 3.1.4, As in Fig. 3.1.1, except for the 9 October 2009 case.

The coverage and intensity of convective storms on 9 December 2009 (Fig. 3.1.5) was less than in the other three joint cases. The area of interest was again in the southeast US, where thin convective lines occurred, particularly in the morning. The 3-km baseline forecast predicted convective storms in the location of the observed convective line, but the predicted storms were

generally weaker and less numerous than was observed over Georgia, Florida, and the northeast Gulf of Mexico (Fig. 3.1.5).



(a) 1500 UTC 9 December 2009



(b) 1800 UTC 9 December 2009

Figure 3.1.5, As in Fig. 3.1.1, except for the Southeast in the 9 December 2009 case.

3.1.2 Forecast Sensitivity to Horizontal Grid Resolution

The parallel HRRR-like (3-km grid spacing) and RR-like (13-km grid spacing) forecasts for the four joint cases (Tables 2.1 and 2.2) provide the opportunity to identify qualitatively any advantages that the higher resolution forecasts provide in anticipating convective storm mode, location, and timing. The fields that are shown in this section are observed composite reflectivity, forecast composite reflectivity (3-km forecast only), forecast convective (parameterized) precipitation accumulation (13-km forecast only), and forecast maximum vertical velocity in each grid column. In the 13-km forecasts, (nearly) all of the convection is represented through the convective parameterization scheme rather than explicitly through model dynamics and precipitation microphysics. Thus, one can't compute reflectivity corresponding to the predicted convective precipitation, for purposes of comparison to the observed reflectivity. Instead, the field that is shown for the 13-km forecasts is the accumulated precipitation determined by the convective parameterization.

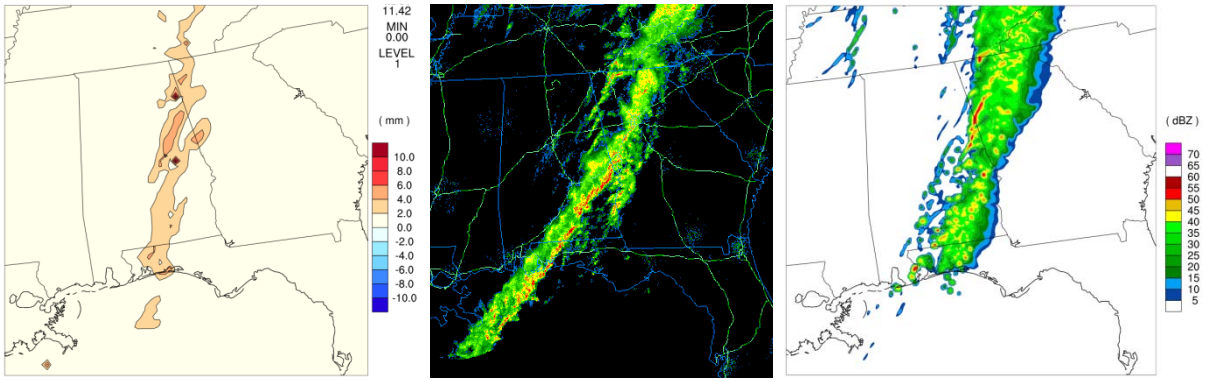


Figure 3.1.6, 13-km forecast 1-hr convective precipitation accumulation (mm; left), observed composite reflectivity (dBZ; center), and 3-km forecast composite reflectivity (dBZ; right) at 1500 UTC 6 February 2008.

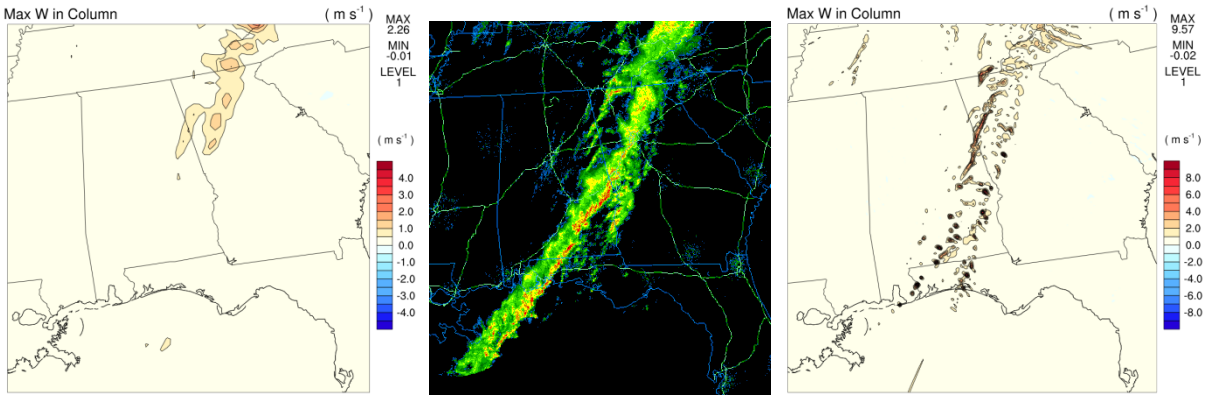


Figure 3.1.7, 13-km (left) and 3-km (right) forecast maximum vertical velocity in model column (m s^{-1}) at 1500 UTC 6 February 2008. Observed composite reflectivity (dBZ; center) is also shown.

For the comparison of the 3-km and 13-km forecasts, we focus mainly on regional convective events that had an impact on aviation. The convective line in the southeast US on 6 February 2008 (Figs. 3.1.6 and 3.1.7) was associated with ground stops due to thunderstorms at the

Atlanta, Georgia airport (ATL) between 1430 and 1500 UTC. Comparing the observed composite reflectivity to the forecast reflectivity / precipitation (Fig. 3.1.6), one notices more realism in the 3-km forecast in the overall convective system structure, the size of individual storms, and the number of storms. Although the 13-km forecast indicates a similar north-south extent of the convective line to that in the 3-km forecast, the 13-km forecast does not provide as much information about the nature of the convective line (that is, typical storm size, broken versus solid line, etc.). Furthermore, the convective line in the 13-km forecast lags behind both the observed and 3-km forecast line. One could reasonably speculate that the coarser-resolution forecast poorly represents process related to convective system propagation, such as the development of cold pools.

Vertical velocity forecasts are also much different in the 3-km and 13-km simulations. Since the 3-km model produces convective storms explicitly, it has more representative updraft strengths, sizes, and organization. In the 6 February 2008 case (Fig. 3.1.7), the vertical velocity pattern in the 13-km forecast does not provide a good indication of the north-south extent of the convective line, nor does it provide information on how individual convective storms are organized in the line.

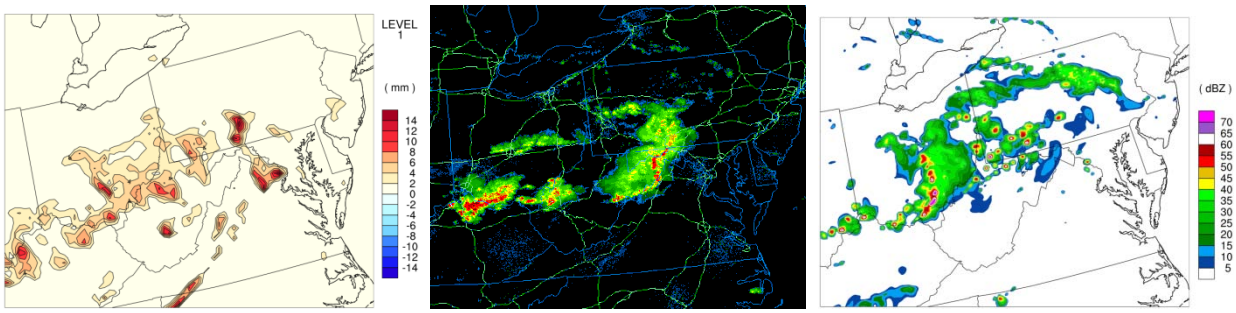


Figure 3.1.8, 13-km forecast 1-hr convective precipitation accumulation (mm; left), observed composite reflectivity (dBZ; center), and 3-km forecast composite reflectivity (dBZ; right) at 1800 UTC 4 June 2008.

Convective storms in the midwest and mid-Atlantic regions on 4 June 2008 were associated with ground stops at the Baltimore, Maryland (BWI) and Washington, DC (DCA and IAD) airports between 1800 and 0200 UTC. For this case, there were significant phase errors in the 3-km baseline forecast (Fig. 3.1.8). For example, the most significant north-south oriented convective line in the model at 1800 UTC was in Ohio, whereas it was in West Virginia, Maryland, and Pennsylvania in reality. Nevertheless, the 3-km forecast still provides useful information about convective mode (that is, the orientation of the most significant convective system; the typical storm size; the number of storms; and the mix of lines, clusters, and isolated cells) that the 13-km forecast does not provide. In addition, for this case, the 13-km forecast appeared to overpredict the areal coverage of convective storms (Fig. 3.1.8). This tendency is discussed more in section 3.2.1.

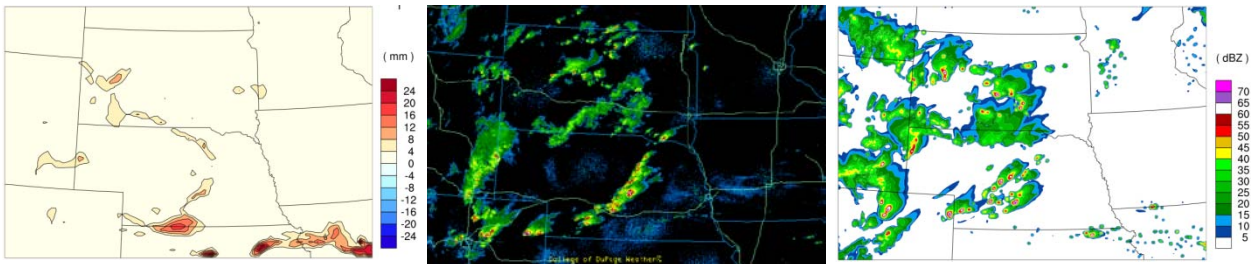


Figure 3.1.9, 13-km forecast 1-hr convective precipitation accumulation (mm; left), observed composite reflectivity (dBZ; center), and 3-km forecast composite reflectivity (dBZ; right) at 2100 UTC 4 June 2008.

Farther west on 4 June 2008, ground stops due to thunderstorms occurred at the Denver, Colorado (DEN) airport between 2030 and 2100 UTC. Strong storms were numerous in the central plains during the afternoon and evening (Fig. 3.1.9). The 13-km forecast has only weak indications of convection in Denver area, and it incorrectly emphasizes the region from northwestern KS to northern MO for the strongest convection. The 3-km forecast has strong storms in the Denver area; provides rather accurate locations of storm regions overall in CO, WY, NE, and SD; and realistically represents the convective mode (size, structure, etc.).

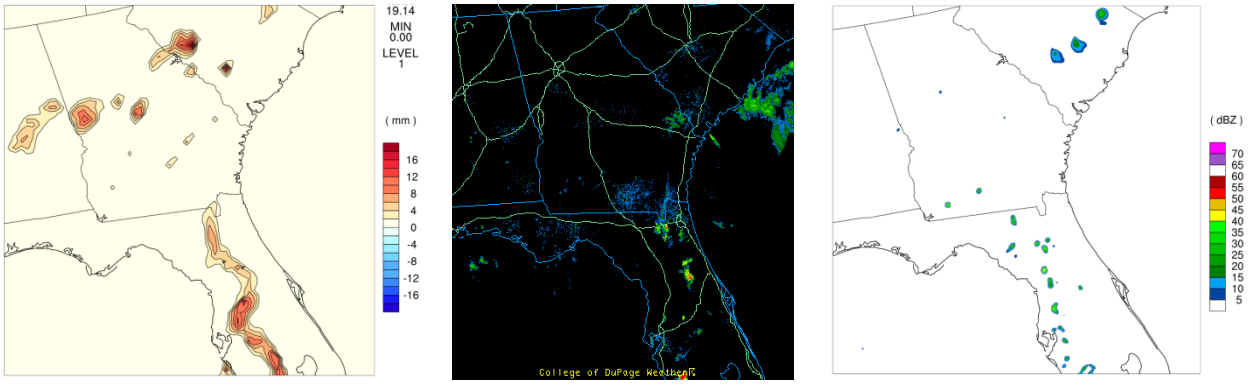


Figure 3.1.10, 13-km forecast 1-hr convective precipitation accumulation (mm; left), observed composite reflectivity (dBZ; center), and 3-km forecast composite reflectivity (dBZ; right) at 0000 UTC 5 June 2008.

A comparison of 13-km and 3-km forecasts for the southeast US for the same case (Fig. 3.1.10) demonstrates how “false alarms” (overprediction) in the 13-km model can sometimes be mitigated by using higher resolution instead. As discussed in section 3.2.1, the 13-km model tends to overforecast the areal extent of convective precipitation. In the 4-5 June 2008 case (Fig. 3.1.10), the convective precipitation pattern in the 13-km model suggests a line over Florida, which could have the potential to impact aviation significantly if it were a solid line of strong storms. The 3-km forecast provides better guidance about the observed pattern of scattered weak storms. Also, from Mississippi to South Carolina, the 13-km model has spurious storms, whereas the 3-km model more correctly indicates the absence of convective storms in this region.

3.2 QUANTITATIVE ASSESSMENT OF BASELINE AND SENSITIVITY RUNS

3.2.1 Precipitation Forecasting Skill

Objective precipitation verifications of the 3 km and 13 km simulations for the four joint cases (Table 2.1) are discussed in this section. To verify the precipitation forecasts, Stage IV estimates of hourly rainfall rate are first interpolated to the 3 km and 13 km model grid respectively. Skill scores are then calculated by comparing forecasts against observations at model grid points. Traditional Equitable Threat Score (ETS) (Fig. 3.2.1), bias score (Fig. 3.2.2), and Fractional Skill Score (FSS) (Fig. 3.2.3), of forecast 1-h precipitation accumulations are presented here.

It should be pointed out that, while the traditional skill scores are still widely used as indicators of forecast quality, they have limitations (especially for convection verification) and need to be interpreted with care (Hogan et al., 2010). The skill scores need to be considered in conjunction with examinations of the actual precipitation maps from the forecasts (Appendix A.1). From the precipitation verification of the 3 km and 13 km baseline forecasts (Figs. 3.2.1-3.2.3), the following points have been observed.

- Both the 3 km and 13 km forecasts have fairly good skills in the first hour, and do not experience significant spin-up problems. This is clearly a benefit of the RUC initialization. Both the 3 km and 13 km forecasts are initialized with RUC datasets that are updated every hour and enhanced with data assimilation aimed at short-term forecasting. Previous tests (e.g., September 2009 MD&E report) have shown that forecasts initialized with RUC datasets usually out-perform those initialized with NAM or GFS in 1-8 h forecasts.
- The 3 km forecasts are good with general location of convective areas or lines. The ETS skills for light precipitation are fairly good through 12 hours. However, for heavy precipitation, the skills decrease sharply after the first 4-6 hours. The skills for the 3 cool season cases are generally better than for the June case.
- The ETS scores for the 3 km and 13 km do not exhibit huge differences. However, the precipitation maps (Figs. A1.1-A1.4) as well as the bias scores (Fig. 3.2.2) show that the 13 km forecasts are biased toward significant overprediction and misrepresentation in many areas over the CONUS domain.
- Overall, both the 3 km and 13 km forecasts tend to overpredict precipitation areas as compared to Stage IV estimates (except for two of the 3 km forecasts in the first 2 hours for 2.5 mm/h threshold). However, the 13 km forecasts have much higher bias scores for both 0.1 mm/h and 2.5 mm/h thresholds (Fig. 3.2.2). The over-prediction by coarser resolution models that employ the Grell-Devenyi cumulus parameterization scheme (CPS) is also found in a case study for events on 10 August 2009 (Section 3.3.3).
- The lower ETS in 6-12 h forecasts for the 9 December 2009 case is an artifact. It is partly due to the storms having moved out of radar coverage, and due to missing radar data in the northeast. Therefore bias scores and FSS for the case are not shown in Figs. 3.2.2 and 3.2.3.

- While the equitable threat score measures the point-to-point matches of forecasts to observations at grid points, the fractional skill score measures the matches of fractional coverage in forecasts to fractional coverage in observations in the neighborhood around grid points (Roberts, 2005). It is a “fuzzy verification” that indicates the resemblance of the forecast to observations on a broader scale. For all four cases, the FSS for light precipitation is rather high in both 3 km and 13 km forecasts, in the range of 0.5-0.8 (Fig. 3.2.3; 13 km forecasts not shown). Again, for heavy precipitation, the skills decrease drastically after the first a few hours.

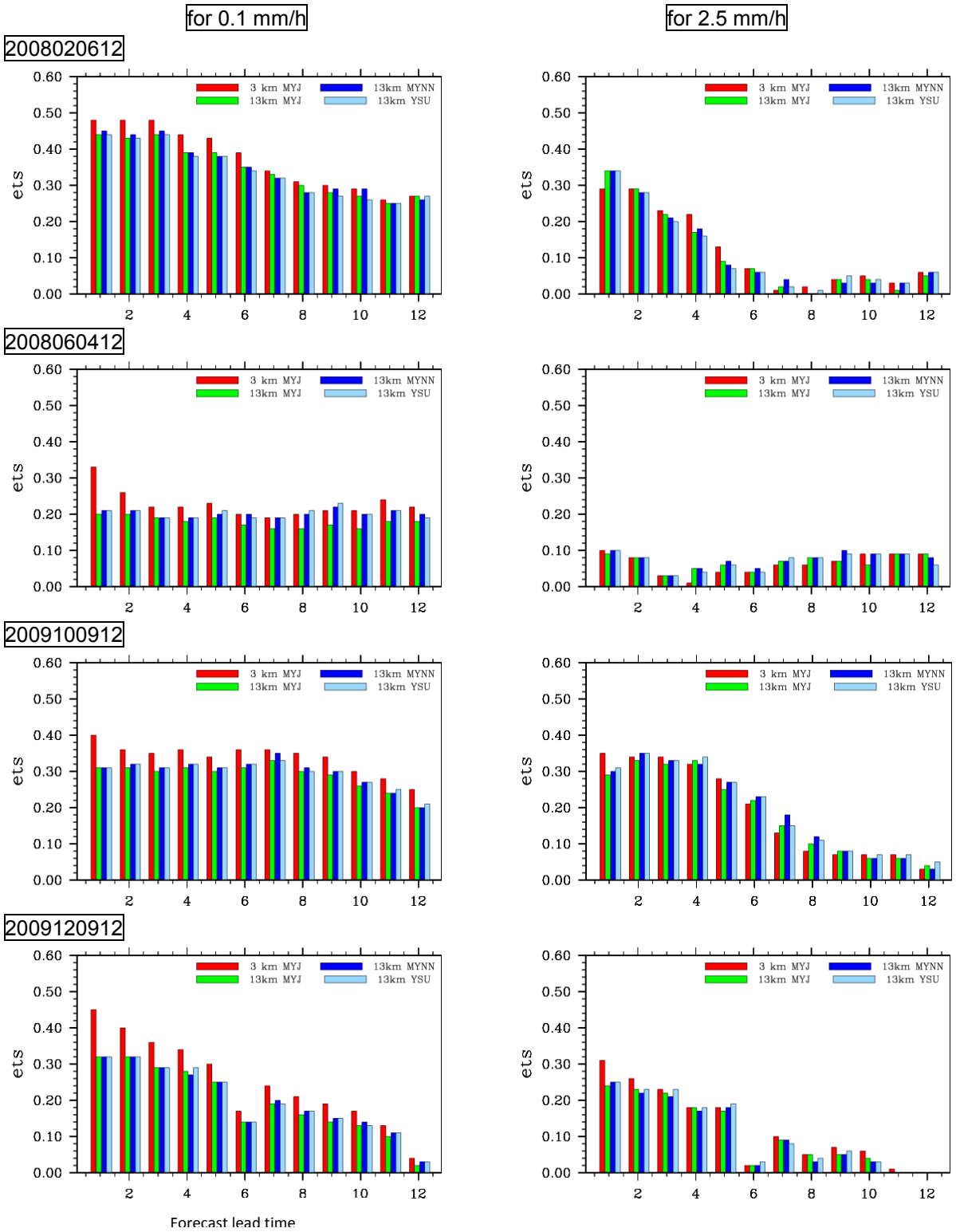


Figure 3.2.1, Traditional Equitable Threat Score (ETS) of 1-h precipitation forecasts from the 3 km (red) and 13 km (green) baseline runs, and the 13 km MYNN (blue) and YSU PBL (light blue) sensitivity runs, for the 4 primary cases.

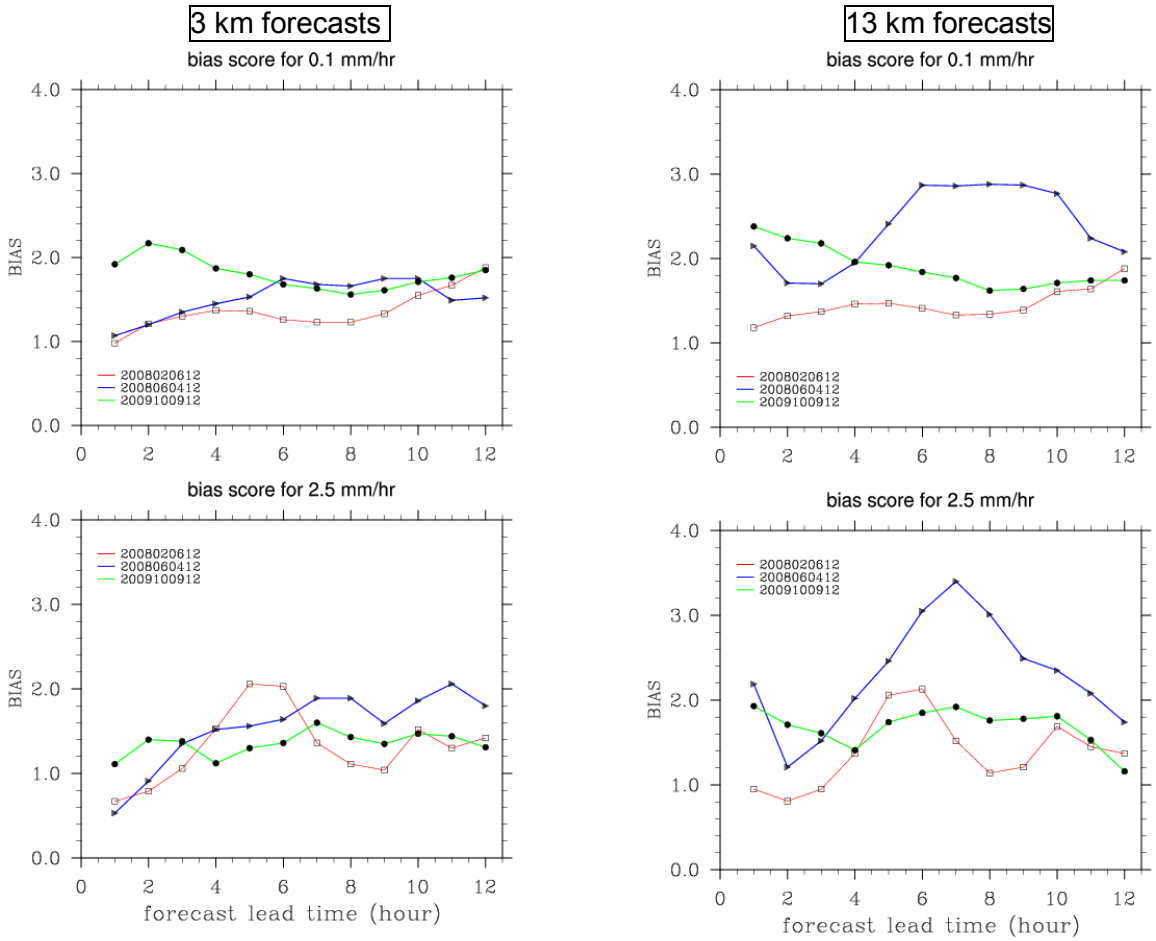


Figure 3.2.2, Bias score of the 3 km (left panels) and 13 km (right panels) forecasts. Shown are skills for three forecasts initialized at 2008020612 (red), 2008060412 (blue), and 2009100912 (green).

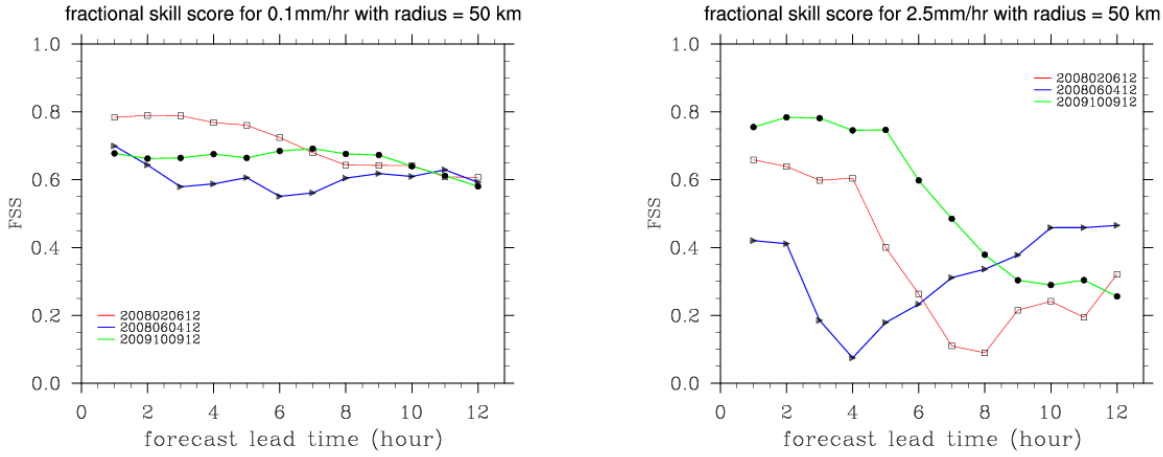


Figure 3.2.3, Fractional skill score of the 3 km forecasts for light precipitation (left panel) and heavy precipitation (right panel). Shown are skills for three forecasts initialized at 2008020612 (red), 2008060412 (blue), and 2009100912 (green).

3.2.2 Forecast Sensitivity to PBL Parameterization

For the four joint cases (Table 2.1), 13-km forecasts were produced for three different WRF planetary boundary layer (PBL) parameterizations: Mellor-Yamada-Janjic (MYJ), Mellor-Yamada-Nakanish-Niino (MYNN), and Yonsei University (YSU). The MYJ scheme is currently used by the real-time HRRR and RR models. The MYNN scheme is essentially a refinement of the MYN scheme and is being evaluated for possible implementation in real-time models such as the HRRR and RR models. The YSU is another commonly used scheme for WRF mesoscale and convective-scale forecasting.

Here, 3-, 6-, 9-, and 12-hour forecasts from the 13-km model with each of the three PBL parameterizations are verified with surface observations at 1500, 1800, 2100, and 0000 UTC, respectively. These surface METAR observations in the US, southern Canada, and northern Mexico were obtained from NOAA's Meteorological Assimilation Data Ingest System (MADIS). The forecast values are the WRF-diagnosed values at 2 m AGL (temperature and dewpoint) and 10 m AGL (*u* and *v* wind components). The verification fields are the root-mean-square (RMS) difference between the forecast and observations and the mean difference, which could be interpreted as model bias.

For the four joint cases, temperature forecast skills with the MYJ and MYNN schemes were not significantly different (Fig. 3.2.4). In contrast, the forecasts with the YSU scheme were noticeably different. Forecast surface temperatures with the YSU scheme were 1-4 K higher than those with the MYJ and MYNN schemes during the day (dashed and solid blue lines in Fig. 3.2.4). These differences became small roughly around the time of sunset. Furthermore, for 3 of the 4 cases, forecast surface temperatures with the YSU scheme were significantly higher (typically 1-2 K but up to 5 K) than surface-temperature observations during the daytime (long dashed blue lines in Fig. 3.2.4). Temperature forecasts with the YSU scheme also had the highest RMS errors for 3 of the 4 cases (dashed and solid red lines in Fig. 3.2.4).

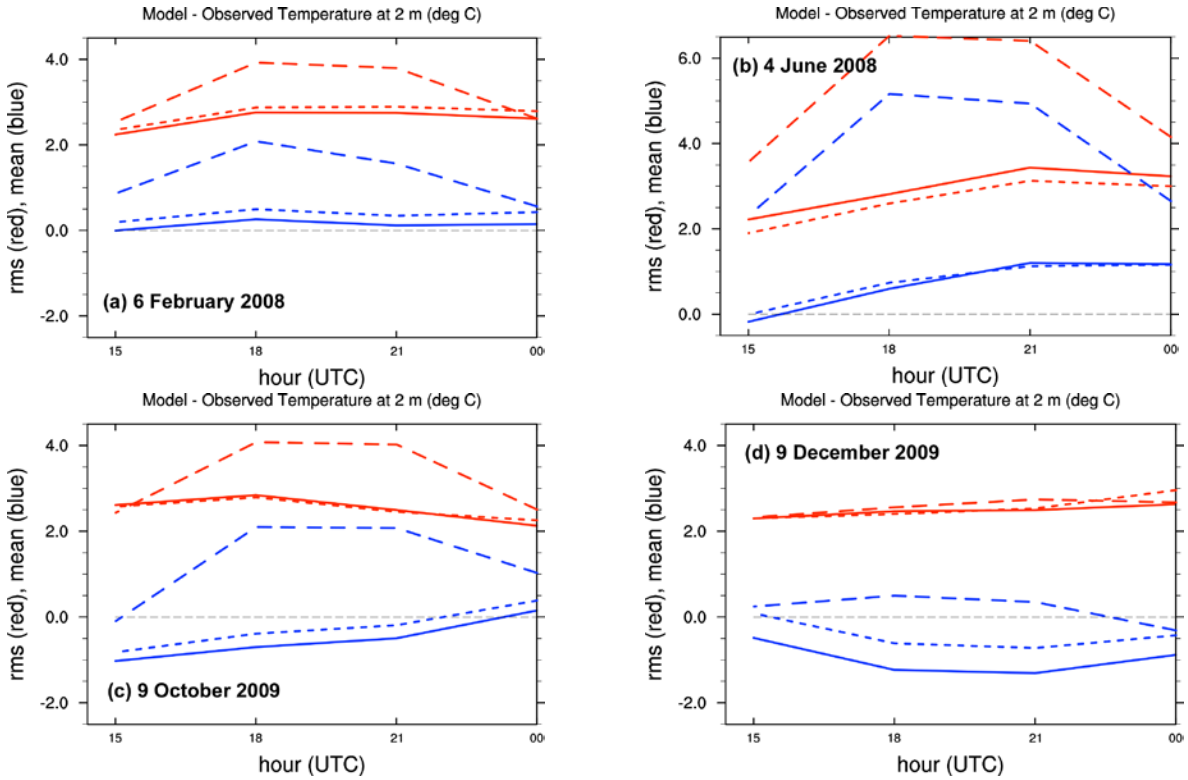


Figure 3.2.4, Root-mean-square (red) and mean (blue) temperature differences (K) between 13-km forecasts and surface observations. Verification statistics are shown for forecasts utilizing three different PBL parameterizations: MYJ (solid lines), MYNN (short dashes), and YSU (long dashes).

Point-wise illustrations of the differences between the YSU-based and MYJ-based surface-temperature forecasts indicate that these differences are situation dependent. Forecast temperatures with the YSU scheme are much higher than those with the MYJ scheme particularly during the day in locations where the YSU-based forecast indicates clear skies (Fig. 3.2.5). In at least two instances among the four joint cases, the YSU-based forecast had stratus that was neither observed nor predicted in the MYJ forecast (central plains in 6 February 2008 case, high plains in 9 December 2009 case) (Fig. 3.2.5). Thus, the YSU surface temperatures were significantly lower in these particular locations than the MYJ temperatures.

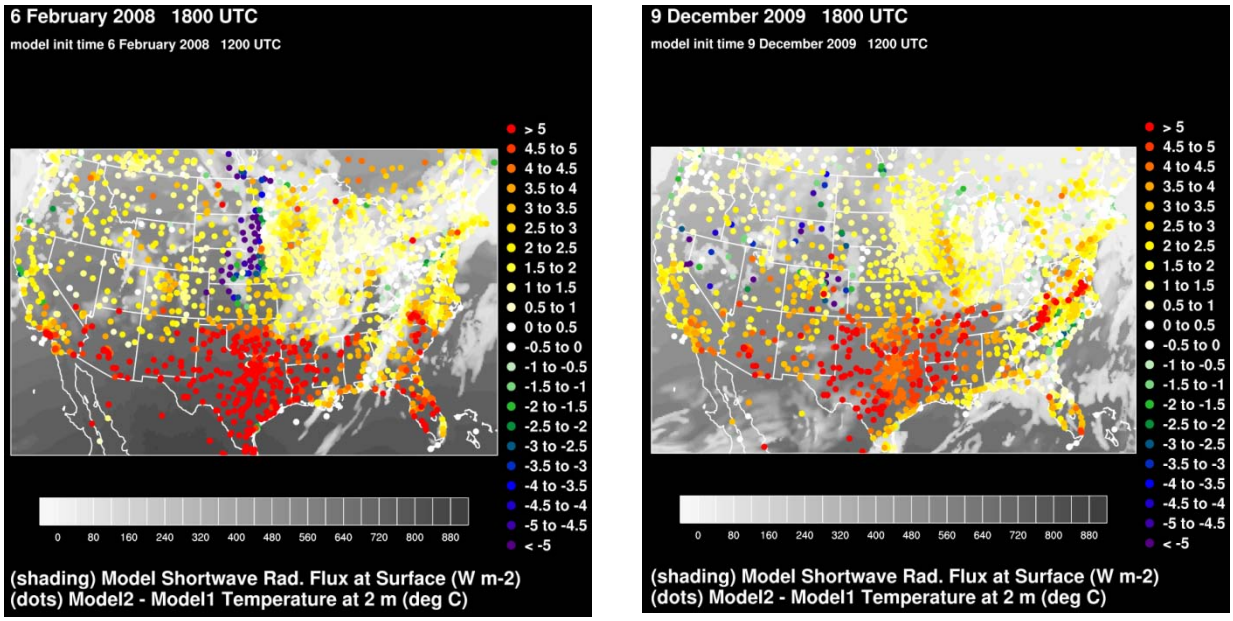


Figure 3.2.5, Temperature differences (colored dots; K) at surface-observation locations between 13-km forecasts with the YSU and MYJ PBL schemes at 1800 UTC 6 February 2008 (left) and 1800 UTC 9 December 2009 (right). Also shown is the downward shortwave radiation flux at the surface in the YSU forecasts (shading, in W m^{-2}).

The next series of verification plots is for dewpoint temperature at 2 m AGL (Figs. 3.2.6-3.2.8). Overall, dewpoints are lower in the YSU-based forecasts than in the MYJ-based forecasts, and these differences increase with time during the day (dashed and solid blue lines in Fig. 3.2.6). Relative to the MYJ-based forecasts, dewpoint RMS errors are less in both the YSU- and MYNN-based forecasts (dashed and solid red lines in Fig. 3.2.6).

If the verification is restricted to locations where the lower atmosphere has high magnitudes of water vapor, then the conclusions change somewhat. Here, we arbitrarily define these "moist" locations as locations where the observed dewpoint is $\geq 10°C$. The verification then reveals no clear advantage to any of the three individual PBL schemes, in terms of RMS errors (dashed and solid red lines in Fig. 3.2.7). At these "moist" sites, dewpoints are still lower in the YSU-based forecasts than in the MYJ- (and MYNN-) based forecasts (dashed and solid blue lines in Fig. 3.2.7).

Further examining point-specific forecasts, one notices a relationship between YSU – MYJ dewpoint differences (Fig. 3.2.8) and YSU – MYJ temperature differences (Fig. 3.2.5). Specifically, where it's warmer in the YSU-based forecast, it also tends to be drier.

Wind forecasts at 10 m AGL were also verified with observations, and plots for these verifications are shown in Figs. A1.5 - 1.6 in Appendix A.1. Forecast differences for the three different PBL schemes were not great. However, using the YSU or MYNN scheme did result in slightly lower surface-wind errors than using the MYJ scheme.

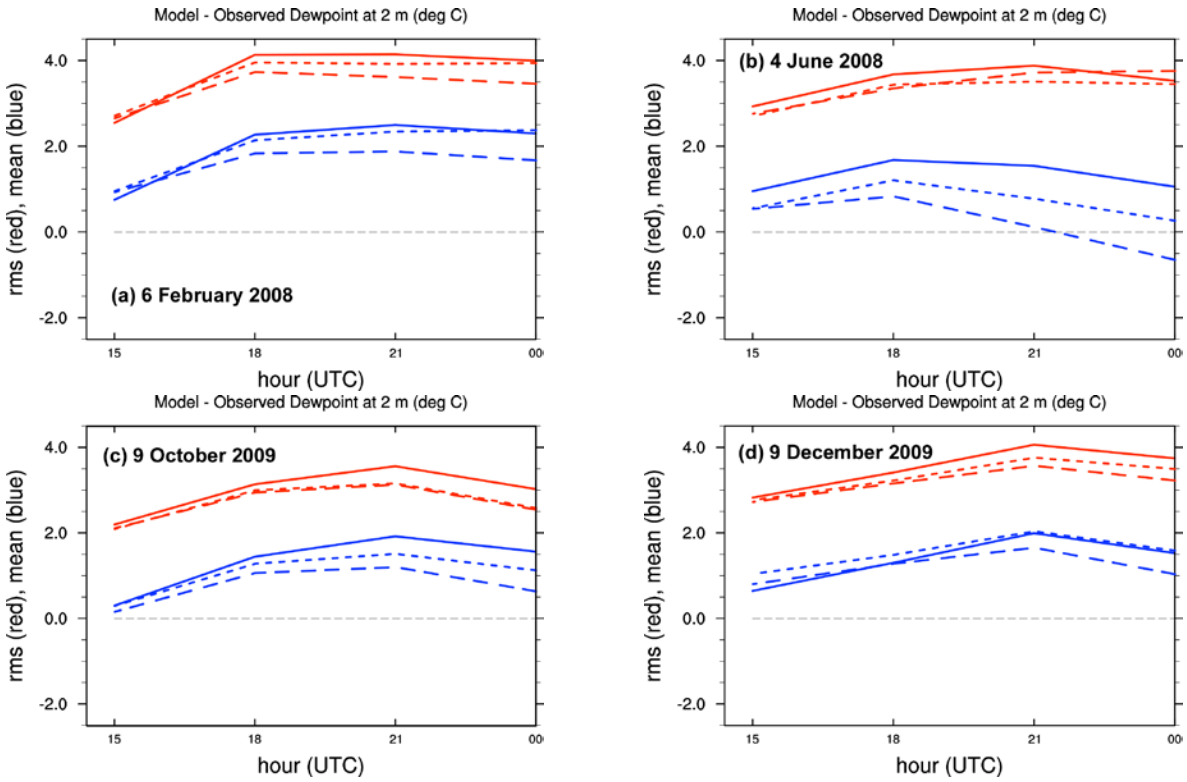


Figure 3.2.6, Root-mean-square (red) and mean (blue) dewpoint differences (K) between 13-km forecasts and surface observations. Verification statistics are shown for forecasts utilizing three different PBL parameterizations: MYJ (solid lines), MYNN (short dashes), and YSU (long dashes).

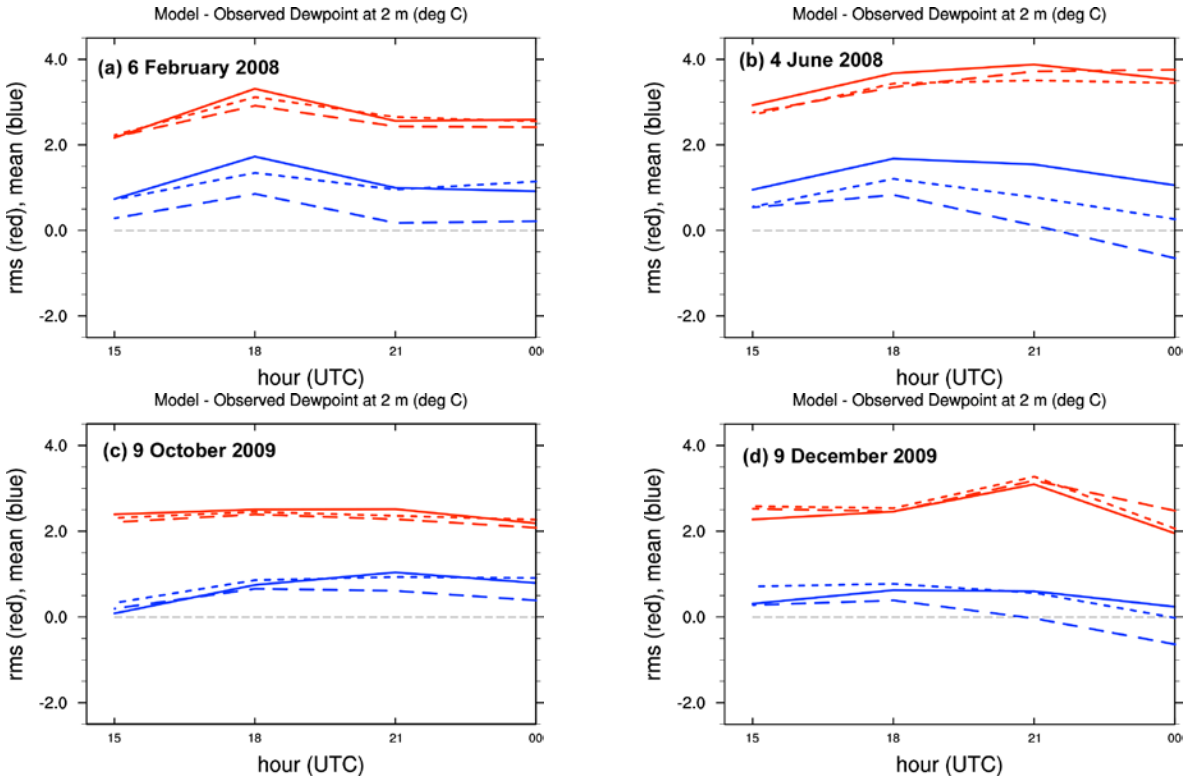


Figure 3.2.7, As in the previous figure, except the statistics are computed *only at locations where the observed dewpoint is ≥ 10°C.*

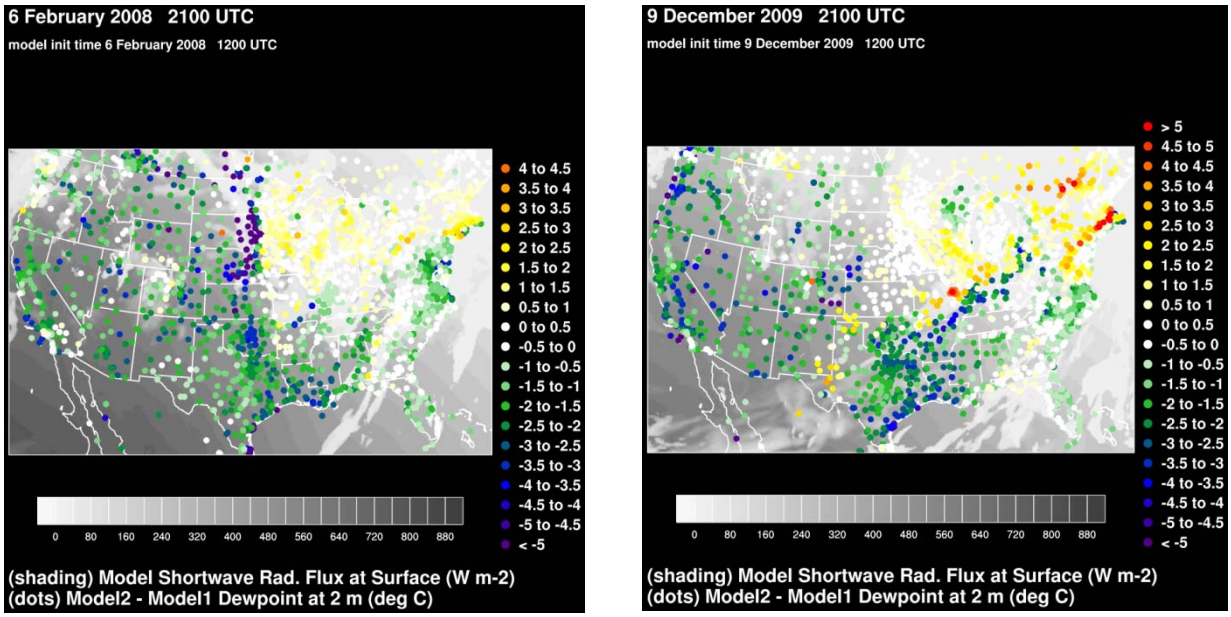


Figure 3.2.8, Dewpoint differences (colored dots; K) at surface-observation locations between 13-km forecasts with the YSU and MYJ PBL schemes at 2100 UTC 6 February 2008 (left) and 2100 UTC 9 December 2009 (right). Also shown is the downward shortwave radiation flux at the surface in the YSU forecast (shading, in W m^{-2}).

3.3 SENSITIVITY ANALYSES OF ADDITIONAL CASES

3.3.1 Forecast Sensitivity to Cloud Analysis and DDFI

The current operational HRRR uses the 13 km RUC analysis as initial conditions. The 13 km RUC is initialized by GSI that is a 3DVAR-based data assimilation system with cloud analysis (Hu et al. 2007) and diabatic digital filter initialization (DDFI, Weygandt et al. 2008). The DDFI is a technique that incorporates diabatic heating derived from radar reflectivity observations. The same data assimilation technique used in the 13 km RUC can be applied to initialize the 3 km HRRR directly by assimilating data at the 3 km resolution. For convenience, hereafter we will refer to the data assimilation technique that is applied to initialize the 3 km HRRR as HRRR DA. In this study, our objective is to test the robustness and performance of the HRRR DA in 0-12 hour convective initiation and forecast.

The convective case chosen for the study occurred on 5-6 June 2009 over the Rocky Mountain Front Range region. Scattered convective cells formed around 19 UTC on 5 June in the northwest part of Wyoming. The cells propagated eastward and organized into a convective system. Fig. 3.3.1 shows Stage 4 hourly rainfall at 01 and 03 UTC on 6 June respectively. We conducted four forecast experiments initialized at 00 UTC to evaluate the impact of DDFI and cloud analysis. Experiment CTRL is initialized by GSI without the use of DDFI and cloud analysis, experiment BOTH turns on both DDFI and the cloud analysis schemes, experiment CLOUD is the same as BOTH but without the DDFI scheme, and experiment DDFI is the same as BOTH but without the cloud analysis.

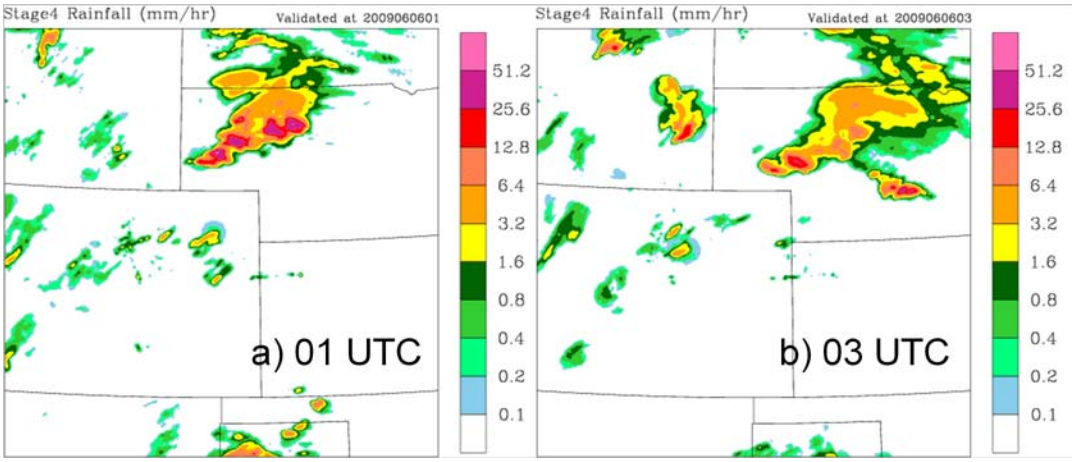


Figure 3.3.1, Stage 4 hourly precipitations at a) 01 UTC and b) 03 UTC of 6 June 2009.

Figs. 3.3.2 and 3.3.3 show 1-hour and 3-hour hourly rainfall forecasts from the four experiments. Comparing the results in Fig. 3.3.2 with the Stage 4 analysis (Fig. 3.3.1a), it can be seen that the experiment CTRL misses the major convective band in Nebraska while the experiment BOTH captures the band and some of the convective cells embedded in the band. The plots for CLOUD and DDFI indicate that the main contribution for the improved forecast in the experiment BOTH comes from the cloud analysis, although the DDFI plays a role in

strengthening the convective cells. Fig. 3.3.3 suggests that the impact of both cloud analysis and DDFI does not last very long -- the difference from the four experiments is not significant in the 3-hour forecasts.

More case studies are being performed to test the robustness and performance of HRRR DA for both storm initiation and forecasts. An emphasis will be on the study of convective initiation over the southeast region of the CoSPA domain. The operational HRRR that uses the 13 km RUC for initialization often fails to forecast these initiation events. Studies will also be conducted to investigate the reasons for the short-time impact of DDFI and cloud analysis. Techniques will be developed to prolong the impact. The future HRRR DA studies will be coordinated with other sensitivity studies.

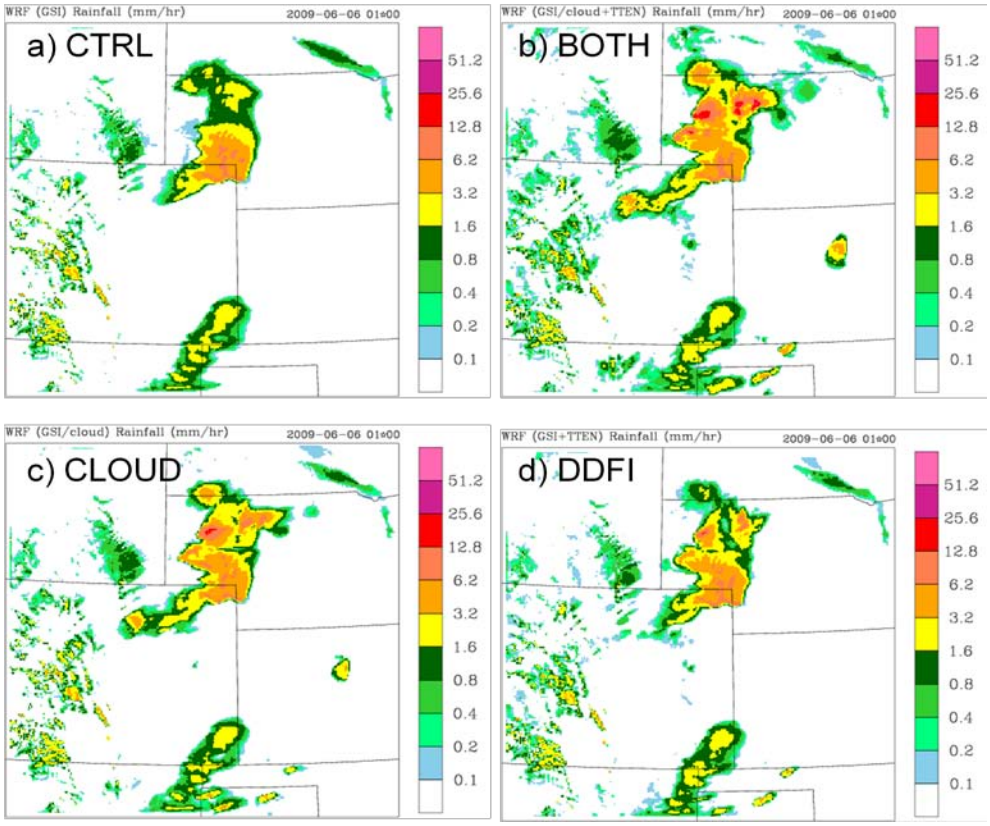


Figure 3.3.2, 1-hour rainfall forecast starting from 00 UTC of 6 June 2009 from the experiments CTRL, BOTH, CLOUD, and DDFI.

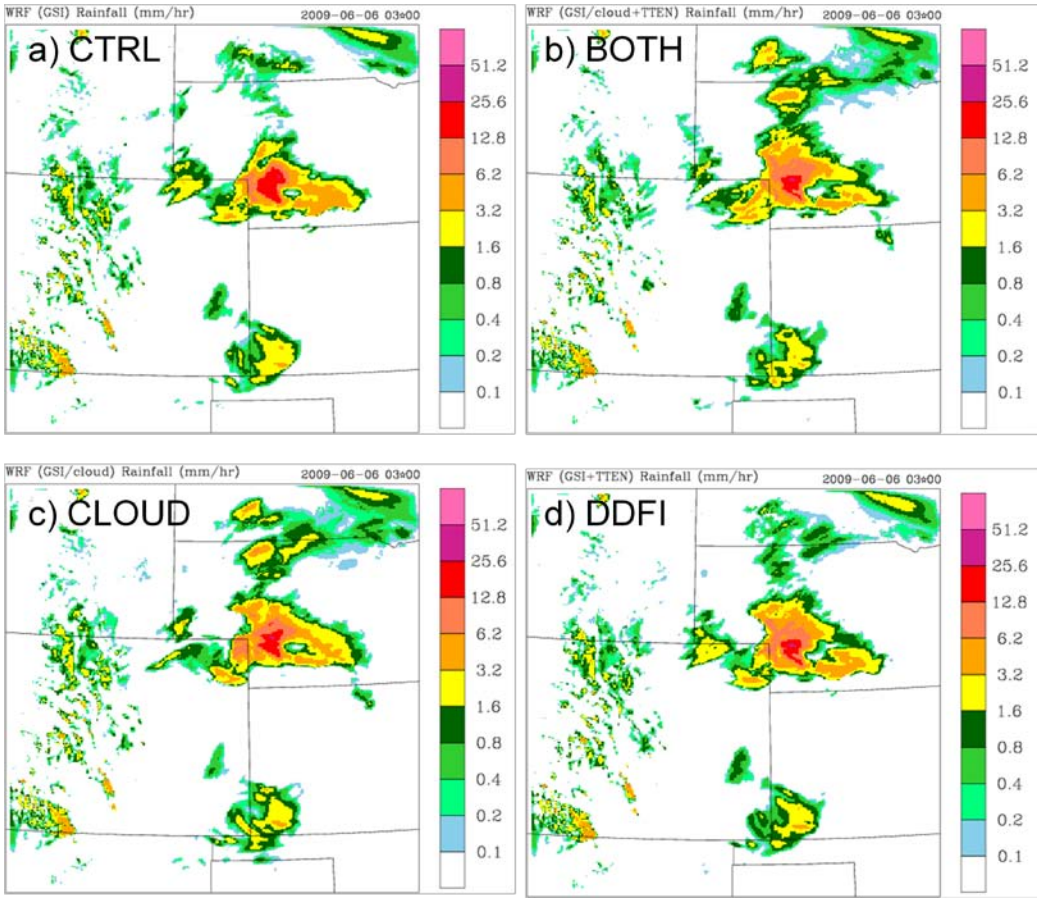


Figure 3.3.3, 3-hour rainfall forecast starting from 00 UTC of 6 June 2009 from the experiments CTRL, BOTH, CLOUD, and DDFI.

3.3.2 Forecast Sensitivity to Horizontal Diffusion

The WRF model includes a 6th-order horizontal numerical diffusion scheme for filtering out short-wave numerical noises on all variables (Skamarock et al., 2008; Knievel et al., 2007). Two choices of the scheme can be made, a simple diffusion (option = 1) and a positive definite one (option = 2). In addition, the horizontal diffusion coefficient is usually arbitrarily selected and its optimal value may be dependent on the phenomena of interest and other dynamics and physics options used in the model. During HRRR operations in spring 2009, the simple scheme (option 1) and the default value of 0.12 for the horizontal diffusion coefficient was used. The HRRR diffusion scheme was changed to option 2 and coefficient of 0.25 in summer 2009, and has remained so since then.

Sensitivity tests were conducted to find out the effect of changing horizontal diffusion options and coefficient value on HRRR forecasts of convection. Three cases are examined for events on May 15, 2009; August 10, 2009; and August 4, 2010. All three cases involve scattered convection in the Southeast and/or over the Appalachians, in addition to convective lines or clusters in other parts of the domain.

May 15, 2009:

During the late morning**,** coastal convection formed near the Georgia and South Carolina coasts, and gradually formed into a line parallel to the coastline. At 18 Z – 20 Z, the storms were well organized into a thin line of convection pushing inland, getting as far as 100-150 km from the coast (Fig. 3.3.4). The diffusion sensitivity experiment for the case was initially run with a WRFv3.0 based HRRR, and later repeated using WRFv3.1. Initialized at 12 Z, three sensitivity forecasts were conducted using the following options for horizontal diffusion:

(1) option 1, coefficient 0.12;

(2) option 2, coefficient 0.12; and

(3) option 2, coefficient 0.25.

All three forecasts predict scattered convection on the coasts and over a broad area inland, and all fail to predict the observed line formation and movement (Fig. 3.3.4). However, the amount and intensity of scattered convection are different among the forecasts. A significant reduction in the amount of scattered convection is seen when the diffusion option is changed from 1 to 2, and a smaller effect is seen when the coefficient is changed from 0.12 to 0.25.

Histograms for precipitation rate at 20 Z show that the number of grid points with precipitation rates across the range of 0.1 – 25.6 mm/hr is decreased by 1/3 when the diffusion option is changed from option 1 to 2 (Fig. 3.3.5). The domain average precipitation in the forecasts with option 2 is somewhat closer to that in Stage IV observations. When the coefficient is changed from 0.12 to 0.25, a smaller, further reduction is seen for most precipitation rate intervals. However, despite the effect on the domain mean precipitation, the timing/initiation of convection does not seem to be affected by the diffusion choices (Fig. 3.3.6). The experiment was repeated using a WRFv3.1 based HRRR, which showed similar results.

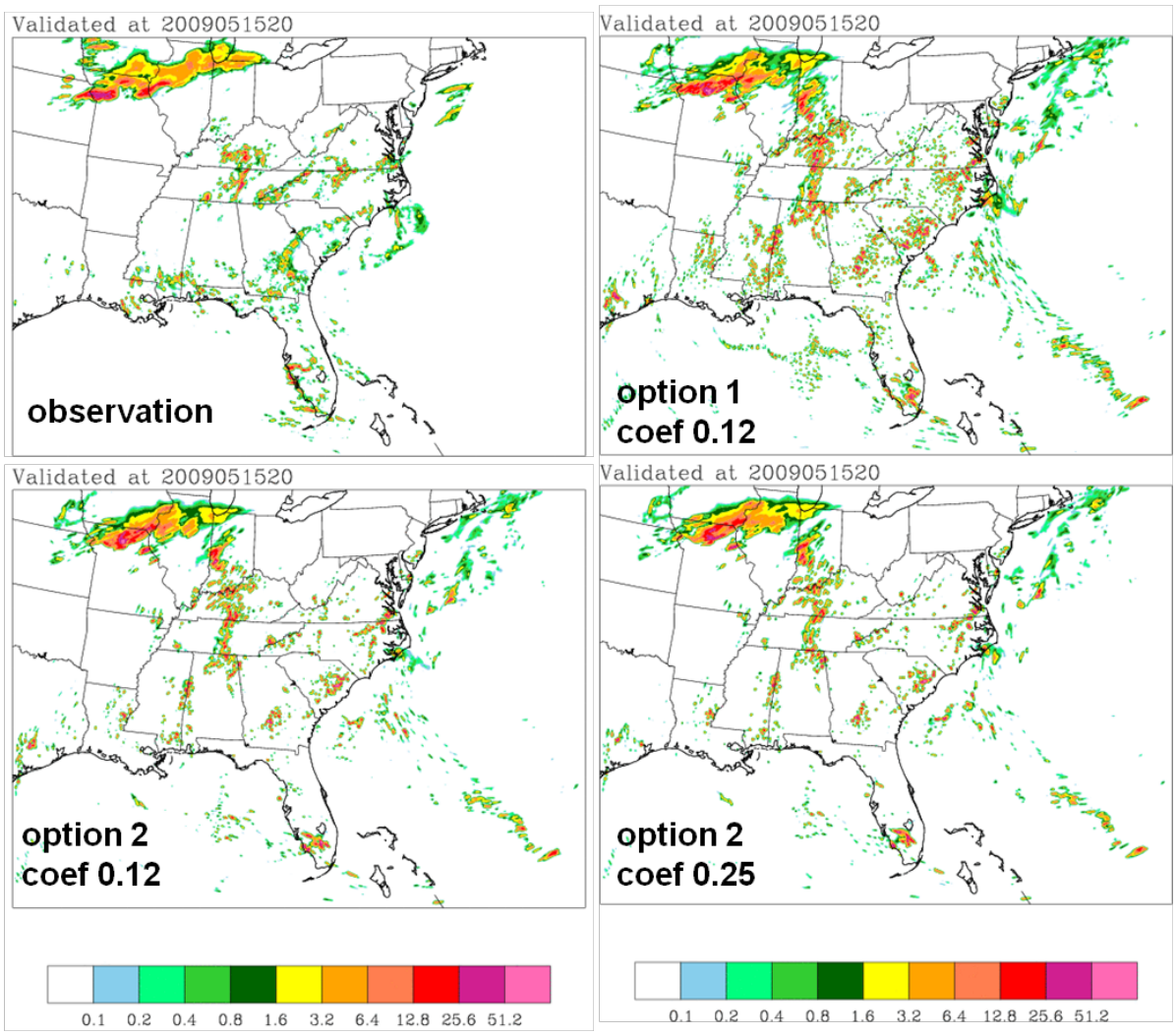


Figure 3.3.4, One-hour rainfall rate from Stage IV and 3 km HRRR (WRFv3.0) forecasts with 3 different horizontal diffusion settings. The forecasts are initialized at 12 Z and valid at 20 Z of May 15, 2009.

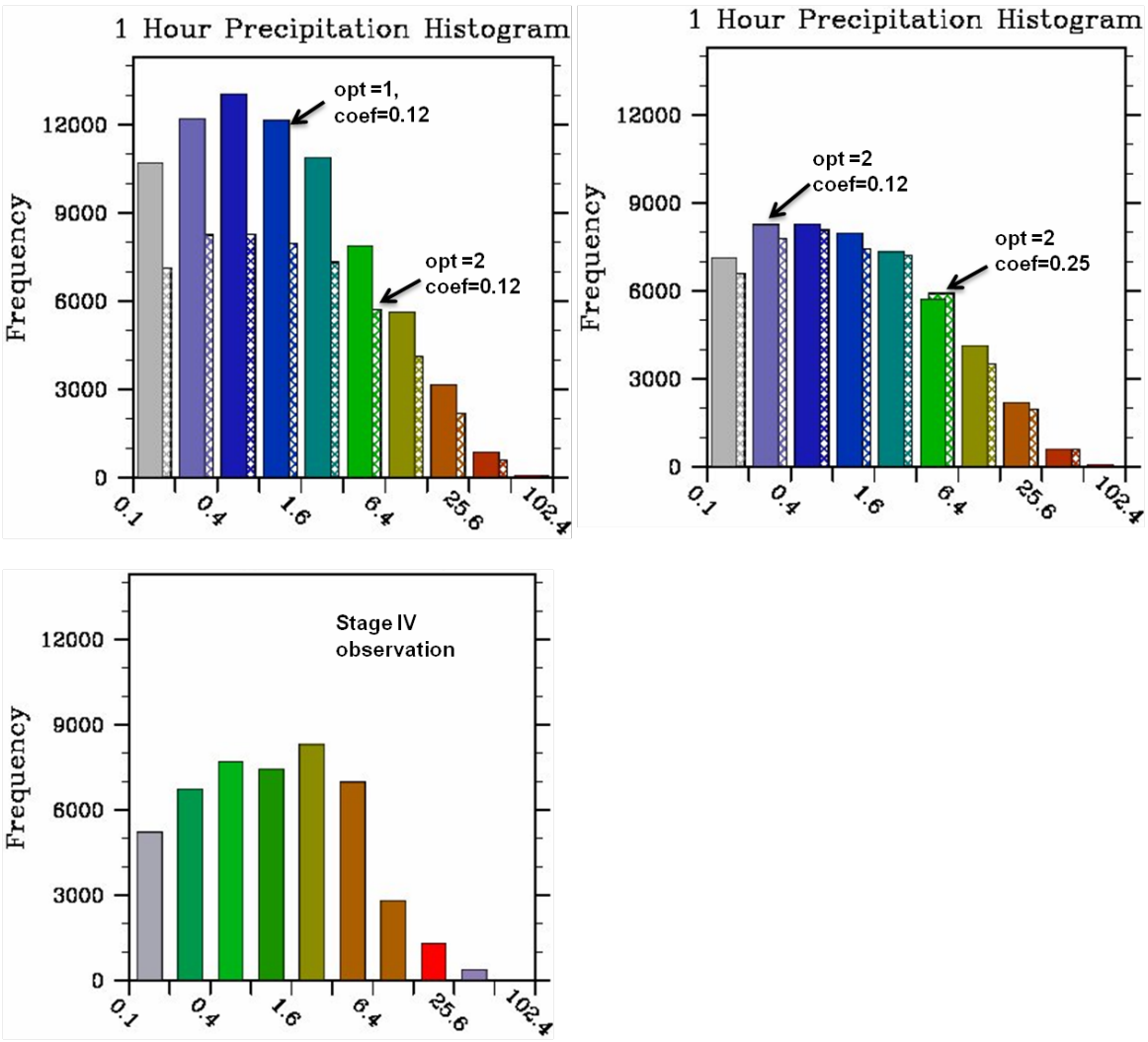


Figure 3.3.5, Histograms showing the number of grid points having precipitation rate in specified intervals. Upper: forecasts from 3 km HRRR with various horizontal diffusion settings. Lower: Stage IV observations. The forecasts are initialized at 12 Z and valid at 20 Z of May 15, 2009.

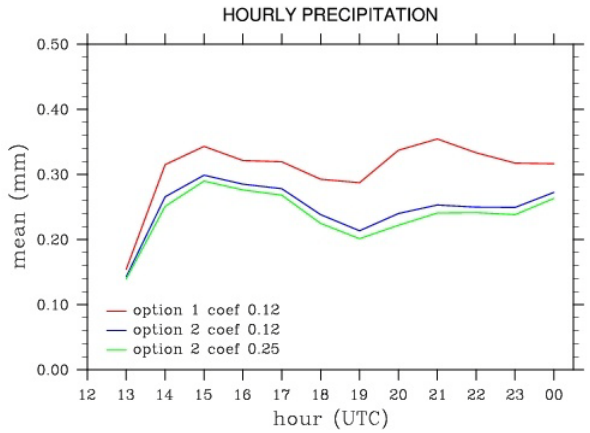


Figure 3.3.6, Hourly domain-mean precipitation rate (mm h^{-1}) in the 15 May 2009 forecasts with 3 horizontal diffusion settings.

August 10, 2009:

The events on the day include line storms in the Midwest to Northeast, clusters near the Kansas-Missouri-Arkansas borders and over West Virginia, and scattered convection in the Southeast (Fig. 3.3.7). The diffusion sensitivity tests were conducted with WRFv3.1.1 using the positive definite scheme (option 2) only, with three different diffusion coefficient values: 0.12, 0.25 and 0.5. The forecasts were initialized at 13 Z.

It is found that all forecasts largely missed the clusters in West Virginia. When a lower value is used for horizontal diffusion coefficient, scattered small cells with light precipitation appear in West Virginia and over the Appalachians, but none of them has the observed intensity. In general, when the diffusion coefficient is reduced (increased), the precipitation amount over the domain increases (decreases) (Fig. 3.3.8). On the other hand, all three forecasts under-predict the cluster storms and over-predict the line storms for this case.

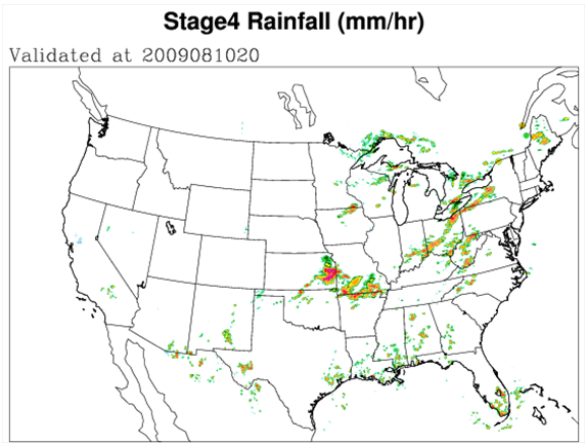


Figure 3.3.7, Stage IV hourly rainfall rate valid at 20 Z, August 10, 2009.

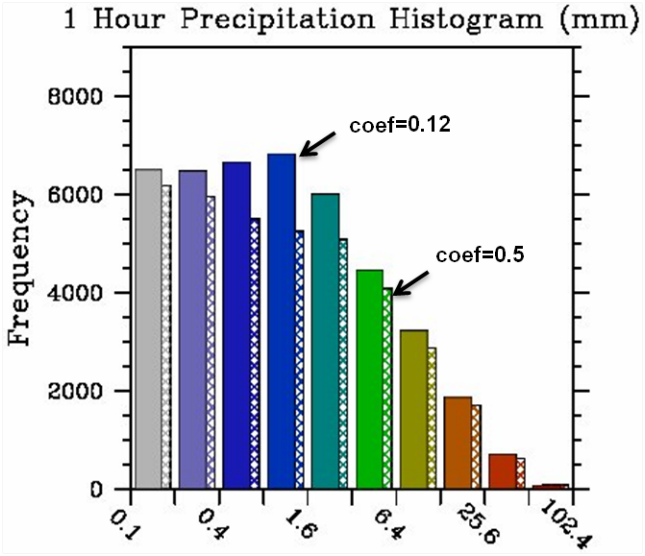


Figure 3.3.8, Number of grid points having 1-hour precipitation rate in specified intervals, for 3 km HRRR (WRFv3.1.1) forecasts valid at 20 Z, 10 August 2009.

August 4, 2010:

This case also involves afternoon (18-22 Z) ordinary convection in the Southeast, mainly on the Gulf coast extending northward to the Appalachians (Fig. 3.3.9). The diffusion sensitivity tests were run with GSD's modified WRFv3.1.1. Again only the positive definite scheme (option 2) is tested, with three different diffusion coefficient values: 0 (i.e. no 6th-order diffusion), 0.12 and 0.25.

Again, the HRRR forecasts for the case predicted some scattered convection in the region, but not the observed organization and evolution. The changes in the amount and intensity of the scattered convection with respect to diffusion coefficient are highly non-linear (Fig. 3.3.10). About 25% of the grid points with light to moderate precipitation rates are eliminated when the diffusion coefficient is changed from 0 to 0.12, and a much smaller effect is seen when the coefficient is changed from 0.12 to 0.25.

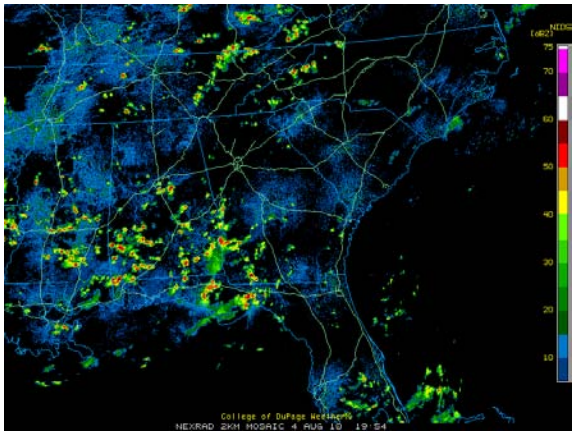


Figure 3.3.9, NEXRAD composite reflectivity (dBZ) at 20 Z on 4 August 2010.

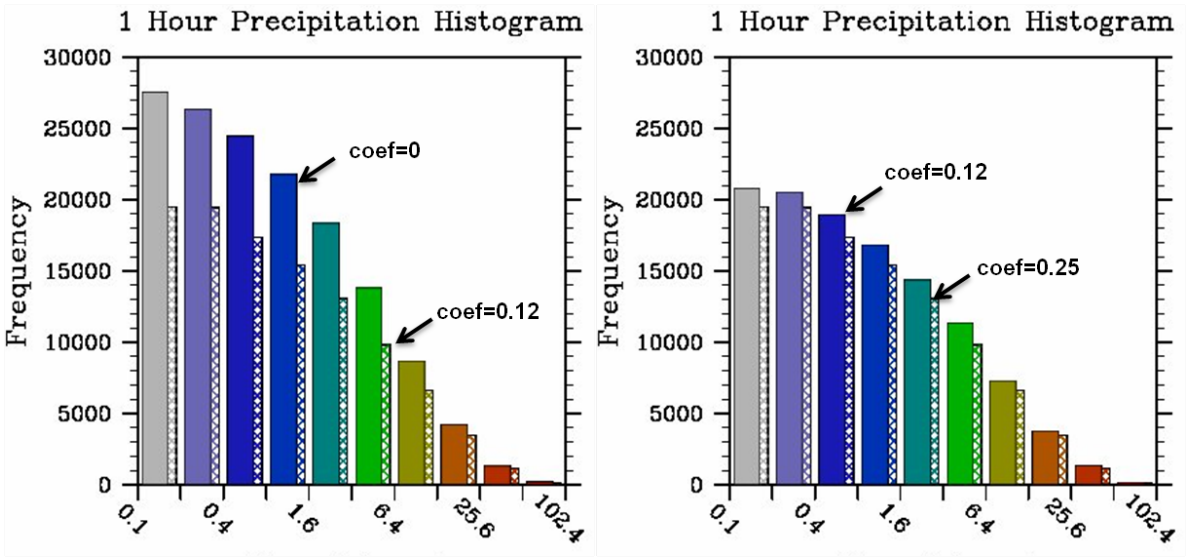


Figure 3.3.10, Number of grid points having precipitation rate in specified intervals, for 3 km HRRR (WRFv3.1.1) forecasts valid at 20 Z, 4 August, 2010. Left: 6th order diffusion coefficient = 0 and 0.25; Right: 6th order diffusion coefficient = 0.12 and 0.25.

3.3.3 Forecast Sensitivity to Horizontal Grid Resolution

In the case study for 10 August 2009, forecasts from model runs with 3, 6, 9 and 12 km grid were conducted and compared with each other. While the model with the 3 km grid uses HRRR configuration and resolves convection explicitly, the model with 6, 9 and 12 km grid have to employ a CPS (Grell-Devenyi ensemble scheme in this case).

Initialized at 12 Z with RUC datasets, the 3 km forecast correctly simulated a line of storms from the Midwest to Northeast, and cluster storms near the Kansas-Missouri-Arkansas borders in the afternoon (Fig. 3.3.11, see Fig 3.3.7 for Stage IV observation). The 3-km model underpredicted or missed the cluster storms over West Virginia and the scattered convection along the Gulf coast. The forecasts produced on 6 km, 9 km and 12 km grids are similar to each other but vastly different from either the 3 km run or Stage IV observations. Instead of line or cluster storms, the 6-12 km forecasts predicted multiple swaths of intense precipitation.

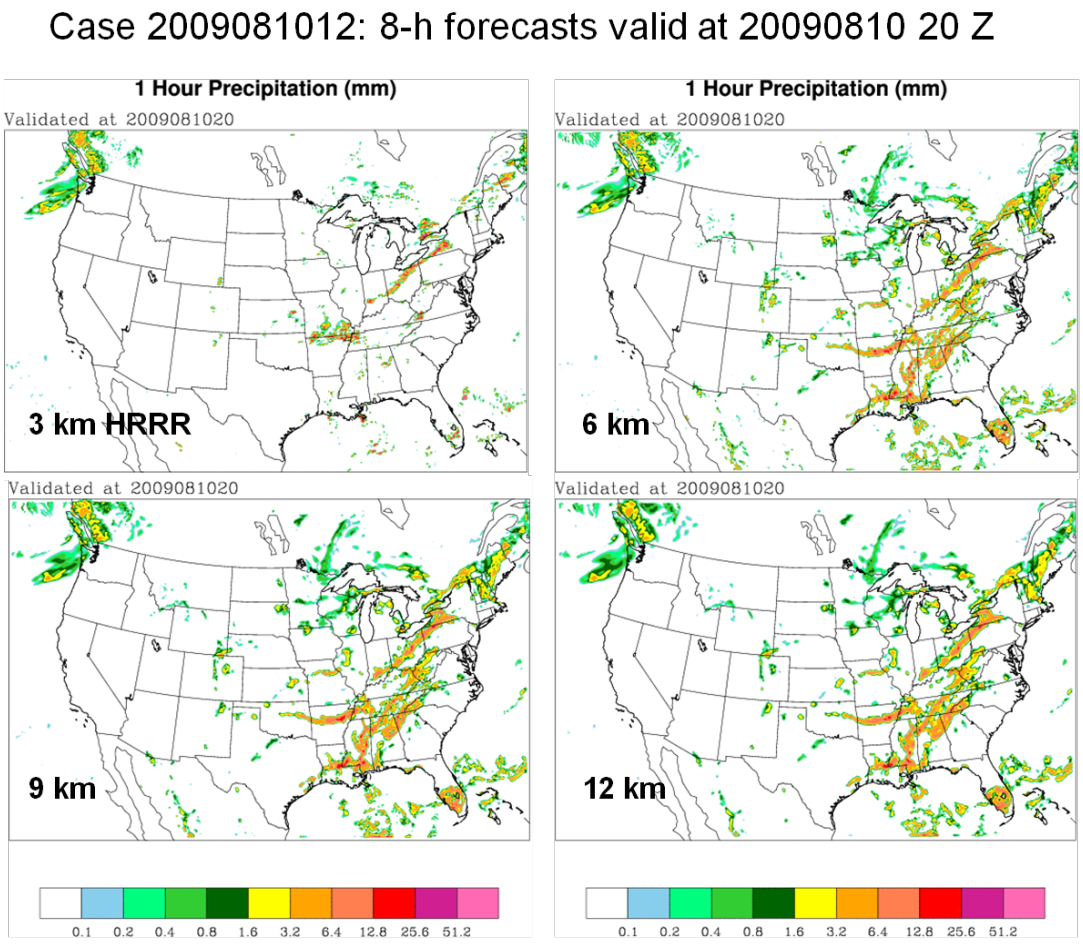
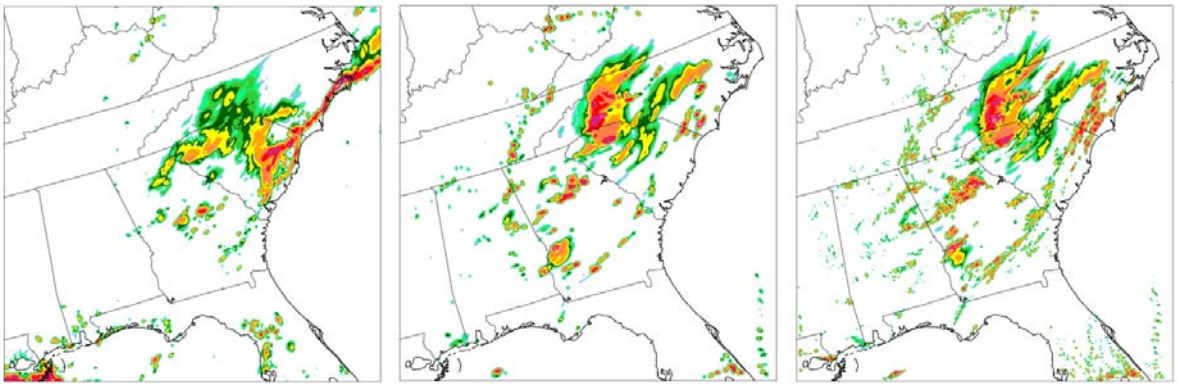


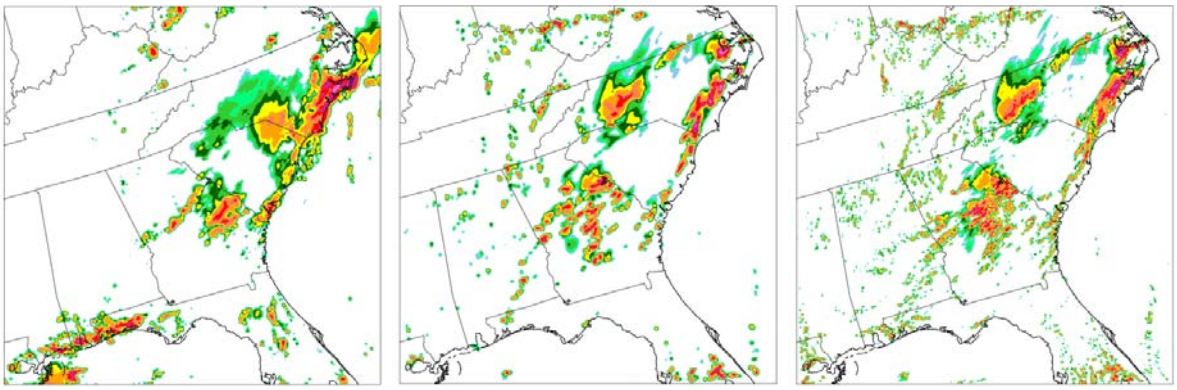
Figure 3.3.11, One-hour rainfall from 3 km HRRR and models with 6, 9, and 12 km horizontal grid spacing. The model was initialized at 12 Z of August 10, 2009 and the validation time shown is 20 Z on the same date.

With a final convection case, we continue the focus on the southeast US, investigating whether convective forecasts there would change significantly if very high model resolution were used.

For the 12 August 2009 case, we compare forecasts produced with the nominal HRRR horizontal grid spacing of 3 km to forecasts that are otherwise identical but employ a 1-km grid spacing (Figs. 3.3.12 - 3.3.17). This case featured a mesoscale convective system over the Carolinas, a convective cluster over Georgia, scattered ordinary storms, and convective lines near coastlines in North Carolina, South Carolina, the Florida Panhandle, and Mississippi (Fig. 3.3.12).



(a) 1800 UTC 12 August 2009 (6-h forecast)



(b) 2000 UTC 12 August 2009 (8-h forecast)

Figure 3.3.12, 1-h accumulated rainfall in stage 4 analysis (left), forecast with 3-km grid spacing (center), and forecast with 1-km grid spacing (right).

Although there were some differences in location, intensity, and structure, the 3-km forecast was successful in predicting a convective cluster in Georgia, a convective line in the eastern Carolinas, and a region of moderate precipitation farther west in the Carolinas (center panels in Fig. 3.3.12). On the broad scale, it appears these features are represented similarly in the forecast with 1-km grid spacing (right panels in Fig. 3.3.12). As one would expect with higher resolution, individual convective storms are typically smaller in the 1-km forecast. Furthermore,

the coverage of convective storms was increased slightly in the 1-km forecast, for example near the coastlines from North Carolina to Georgia, and also in the Florida Panhandle.

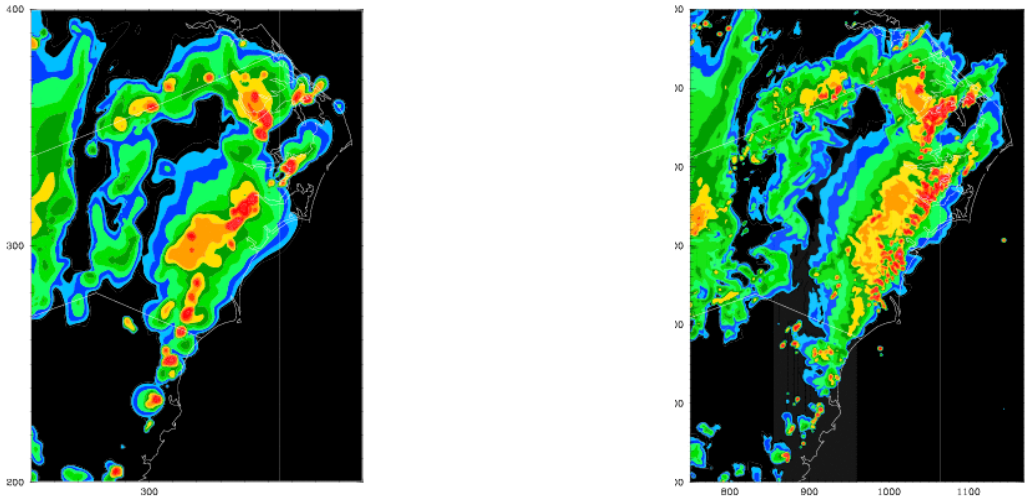


Figure 3.3.13, Composite reflectivity near the mid-Atlantic coast at 2000 UTC 12 August 2009 (8-h forecast) in the forecast with 3-km grid spacing (left) and forecast with 1-km grid spacing (right).

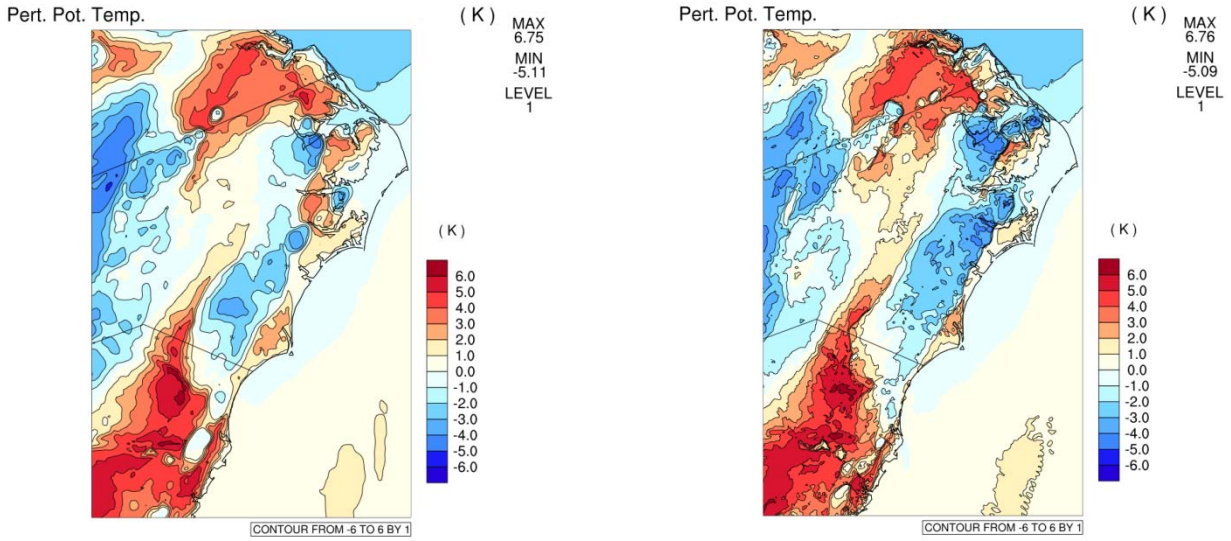


Figure 3.3.14, Perturbation potential temperature (K, relative to WRF reference temperature of 300 K) near the mid-Atlantic coast at 2000 UTC 12 August 2009 in the forecast with 3-km grid spacing (left) and forecast with 1-km grid spacing (right).

Figs. 3.3.13 - 3.3.16 focus on smaller regions in the forecast, to illustrate more notable differences in the representation of convective storms resulting from higher resolution. The first pair of figures highlights the convective system in North Carolina while the second pair highlights convection farther south, near the Georgia - South Carolina border. In both

examples, the 1-km grid spacing results in convective systems that mature more quickly than with 3-km grid spacing. Specifically, in the 1-km forecasts, the areal coverage of storms is greater, and the convective systems have more developed leading convective lines and trailing "stratiform" regions. At the surface, the cold pools are not stronger in terms of minimum temperature in the 1-km forecasts (right panels in Figs. 3.3.14 and 3.3.16) than in the 3-km forecasts (left panels in Figs. 3.3.14 and 3.3.16), but the cold pools in the 1-km forecasts are larger, consistent with the notion that the convective systems have matured more quickly. Also noteworthy in these forecasts is that the coverage of convective storms near the coast has been increased in the 1-km forecasts, suggesting that coastal circulations (sea breeze and land breeze) might be better represented in the higher-resolution forecasts.

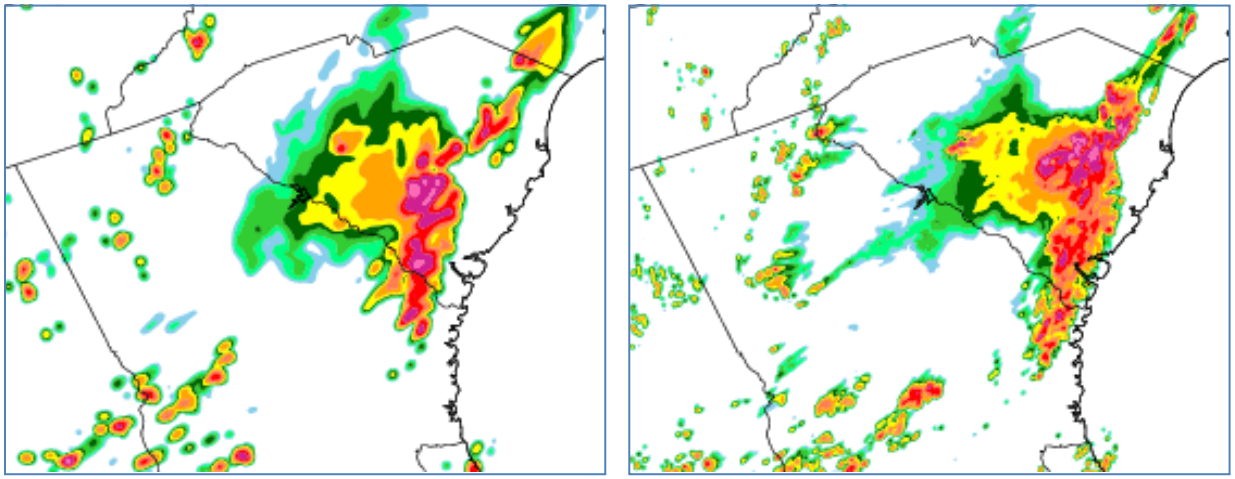


Figure 3.3.15, 1-h accumulated rainfall in the Georgia-South Carolina region at 0000 UTC 13 August 2009 (12-h forecast) in the forecast with 3-km grid spacing (left) and with 1-km grid spacing (right).

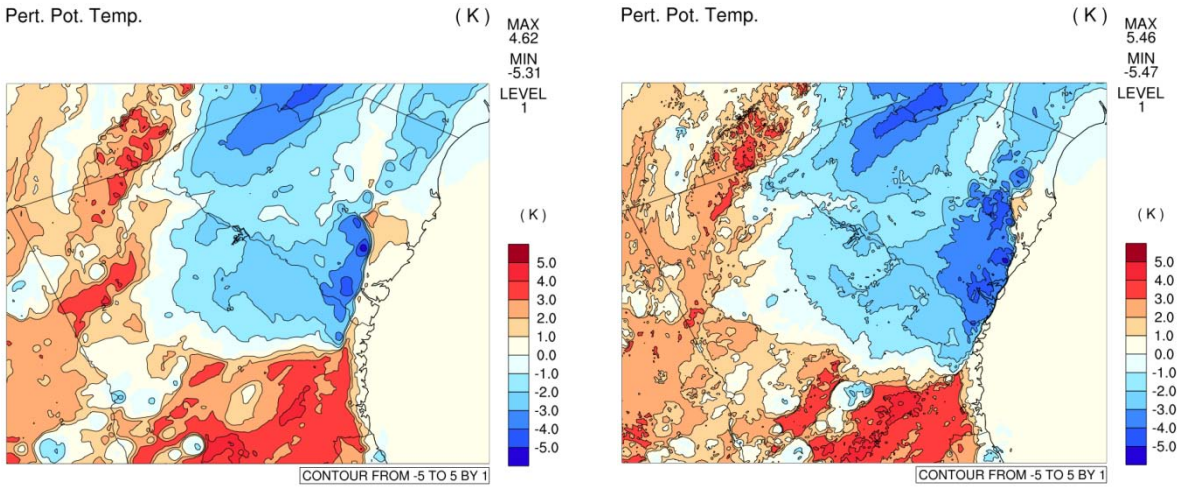


Figure 3.3.16, Perturbation potential temperature (K, relative to WRF reference temperature of 300 K) in the Georgia-South Carolina region at 0000 UTC 13 August 2009 in the forecast with 3-km grid spacing (left) and with 1-km grid spacing (right).

The qualitative impressions of the differences between the forecasts with 3-km and 1-km horizontal grid spacing are supported by domain-wide precipitation statistics (Fig. 3.3.17). The mean accumulated precipitation is about 5% greater in the 1-km forecast than in the 3-km forecast, consistent with the observation of more widespread storms in the higher-resolution forecast. Furthermore, the peak in the hourly accumulation rate occurs about 1 hour earlier in the 1-km forecast than in the 3-km forecast. This difference is consistent with the notion that convective storms and systems mature more quickly in the higher-resolution forecast.

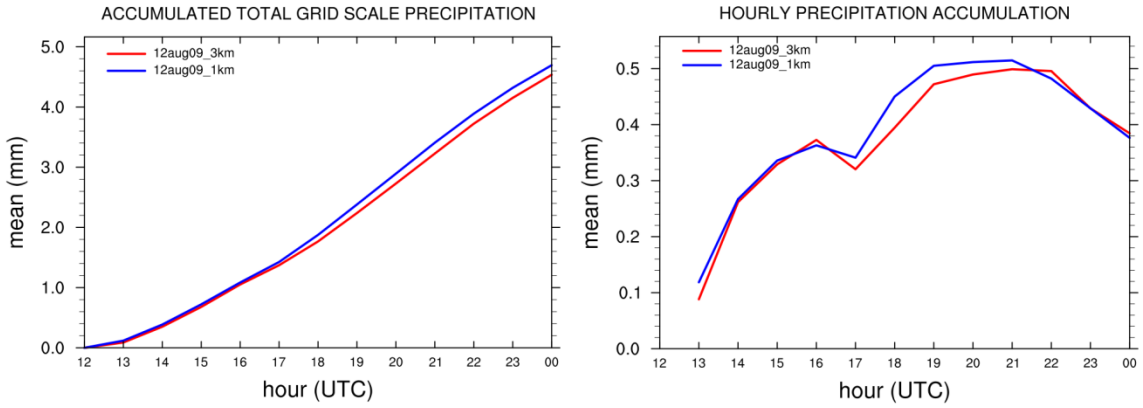


Figure 3.3.17, Hourly domain-mean accumulated precipitation (left; mm) and precipitation rate (right; mm h^{-1}) in the 12-13 August 2009 forecasts with 3-km grid spacing (red) and 1-km grid spacing (blue).

3.4 SUMMARY OF CONVECTION ANALYSES AND SUGGESTIONS FOR FUTURE WORK

Convective forecast-sensitivity experiments focused on the following model characteristics: resolution (horizontal grid spacing ranging from 13 to 1 km), PBL parameterization, horizontal diffusion (scheme and coefficient), and storm-scale initialization through radar-reflectivity data assimilation. As has been shown previously, the differences between forecasts employing 13-km (RR) and 3-km (HRRR) grid spacing were significant, largely due to the ability of the 3-km model to produce convective storms explicitly. The analyses discussed in this report are rather encouraging for the current HRRR configuration. For convective forecasting, 3-km horizontal grid spacing offers the following significant benefits over 13-km grid spacing:

Convective storm / system structures are more realistic.

Storm location and timing forecasts are more accurate.

Some of the "false alarms" (overprediction) in the 13-km forecasts are eliminated by the higher resolution.

Using even higher resolution than in the current HRRR could result in some forecast changes, as illustrated by a single case study here. In the example shown, convective systems matured more quickly in forecasts employing 1-km rather than 3-km horizontal grid spacing, and the storm coverage increased in a region -- the southeast US -- where underprediction is often a problem, as discussed below. Storm coverage near coastlines also increased, suggesting that land surface effects (land-sea contrasts and terrain) could be better represented in the higher-resolution forecasts. However, the differences between 1-km and 3-km forecasts were not nearly as great as differences between 3-km and 13-km forecasts. The 3-km grid spacing currently employed by the HRRR appears to be a cost effective grid spacing for convective forecasting.

During the summer 2009 and 2010 HRRR forecast evaluations, NCAR scientists on duty (SODs) have noted that the model often underpredicts the amount of convection in two situations: during the morning and in the southeast US, where the dominant mode is ordinary convective cells. Examples of underprediction in the Southeast were shown in this report: e.g., the underpredicted number and intensity of isolated cells ahead of the main line in Fig. 3.1.4, and the underpredicted line intensities in Figs. 3.1.1 and 3.1.5. Examinations of forecast sensitivity to the horizontal diffusion scheme and coefficient suggest that this diffusion could be one of several important factors affecting this underprediction. Differences between forecasts employing the WRF simple (option 1) and positive definite (option 2) diffusion schemes were significant, but the choice of the diffusion coefficient also affected the forecasts, with lower values leading to more precipitation. Quantitative verification of multiple cases, for the purposes of tuning the model to have an optimal diffusion coefficient, could be done in the future.

For the 13-km (RR) forecasts, selecting different options for the PBL parameterization (MYJ, MYNN, or YSU) in the WRF model did not affect precipitation-forecast skill significantly. However, forecasts of surface variables -- particularly temperature and dewpoint / water vapor -- were affected significantly. The MYNN scheme might have a slight edge in performance over the other two PBL schemes tested. In future work, it will be useful to determine if there would be significant forecast sensitivity to PBL parameterization for 3-km (HRRR) grid spacing as well.

Additional focus on the initial conditions is suggested. The September 2009 MD&E report demonstrated that sensitivities of 0-6 h convective forecasts to changes in initial conditions are much greater than sensitivities to model changes such as different parameterization choices. The current report demonstrated sensitivity of 1-h convective forecasts to how the model is initialized with radar-reflectivity data.

During the summer 2010 HRRR evaluation by NCAR SODs, it has been noted that the model often has difficulty maintaining mesoscale convective systems that exist at the model initialization time. Also, the model sometimes has difficulty predicting the secondary convective initiation and system regeneration that often occurs during the mid-late morning. Thus, in the future, it will be worthwhile to examine in detail how these systems are initialized through the radar and other data, and to develop methods for representing more balanced convective systems in the initial conditions.

4. ANALYSES FOCUSED ON TURBULENCE

The four (baseline) cases considered for detailed analysis were based in part on the large number of severe PIREPs for those days. To better appreciate this, for reference the 2008 and 2009 median and 95% percentile numbers of severe PIREPs (surface-65,000 ft) are

- 2008: median=13, 95% percentile=46
- 2009: median=11, 95% percentile=41

In contrast, the number of severe reports for the 4 cases studied in detail are

- 6 Feb 2008 = 85
- 4 June 2008 = 34
- 9 Oct 2009 = 103
- 9 Dec 2009 = 90

Thus, the number of severe reports in each case is much larger than even the 95% percentile. Plots showing the location of all PIREPs using standard symbology for the entire 24-hr period covering each of the four days are shown in Fig. 4.1 (and previously in Figs 2.2, 2.4, 2.6 and 2.8). Severe PIREPs are plotted in red and mountain-wave turbulence PIREPs are shown in blue. The locations of NLDN lightning flash data are shown in orange. As in other sections, here we present both qualitative and quantitative analyses of the four baseline cases. Sensitivity studies have not yet been completed.

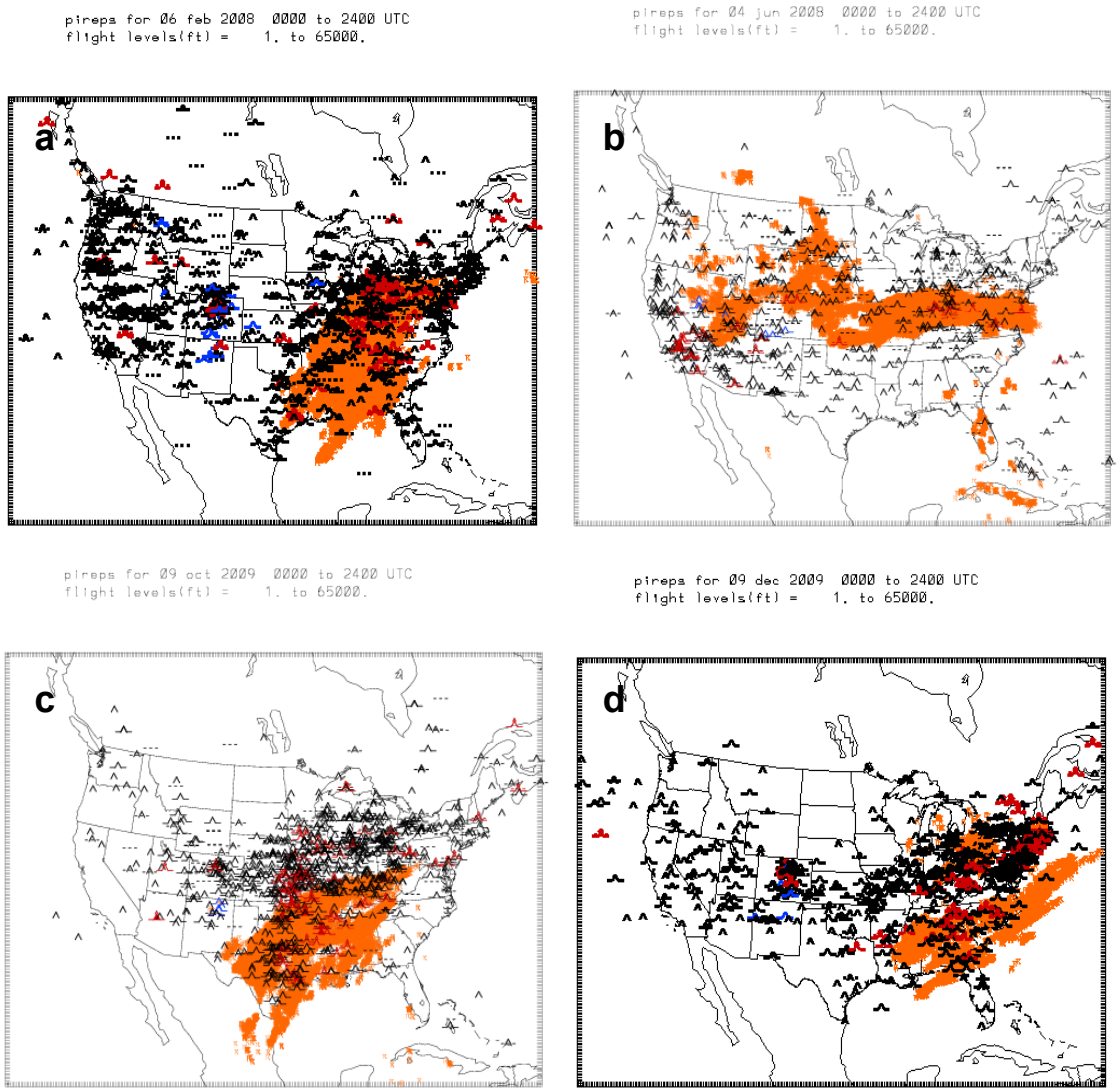


Figure 4.1, Pilot reports for 24 hour period, ground to 65,000 ft for a) Feb. 6, 2008, b) June 4, 2008, c) Oct. 9, 2009, and d) Dec. 9, 2009. Note turbulence was wide-spread on all four days.

4.1 QUALITATIVE COMPARISON OF BASELINE

Qualitative analyses consist of comparisons of PIREPs and insitu EDR data (Cornman et al. 1995) to selected turbulence diagnostics. The GTG combination (Sharman et al. 2006) is not included in the set since this would require a time-consuming calibration of diagnostic performance to each NWP model configuration. And, GTG performance depends on the performance of the individual diagnostics, such that the better the individual diagnostic discrimination performance, the better the GTG discrimination performance. The three diagnostics selected for comparisons are: 1) the Ellrod index, 2) EDR index, and 3) the MYJ subgrid-scale turbulence kinetic energy (SGS TKE) that comes directly out of the underlying

NMP model output (Janjić, 2002). The Ellrod index (Ellrod and Knapp 1992) was developed and has been used operationally by NOAA for turbulence forecasts for many years. The EDR (eddy dissipation rate) algorithm was developed at NCAR for GTG and is described in Frehlich and Sharman (2004). We have found this to be a very good general purpose diagnostic when used with a variety of NWP models. These comparisons have been performed at various flight levels within the computational domain, but only the flight level with the maximum number of reports are shown here. Also, a vertical profile of the Richardson's number (Ri) is shown for the grid point in each model closest to Denver, CO. Ri is a good overall indicator of which layers are susceptible to turbulence (small Ri), and in observations of upper level turbulence these patches tend to be pancake shaped, i.e., thin elongated patches with a vertical dimension much smaller the horizontal dimension (Vinnichenko et al. 1980). In all cases we show comparisons of the HRRR13 (13km RR-like WRF runs) to the HRRR3 (3km HRRR-like WRF runs) and to the RUC13 (13km operational RUC model runs). All results presented are for 6-hr forecasts, initialized at 12 UTC and valid at 18 UTC.

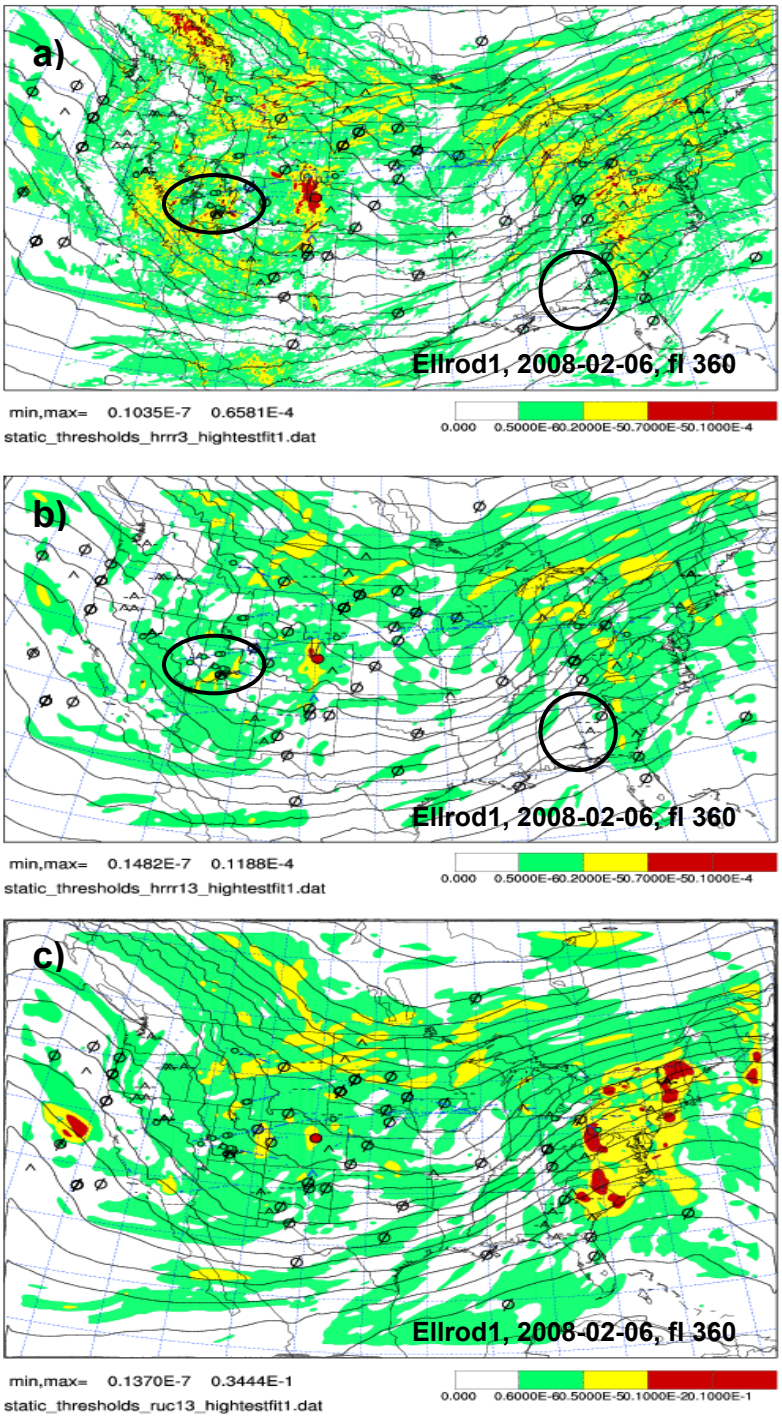


Figure 4.2, Ellrod index for 6 hour forecast initialized at 12 UTC for fl 360 on Feb 6, 2008 for model a) HRRR3, b) HRRR13, and c) RUC13. Light, moderate and severe turbulence is indicated by green, yellow and red, respectively. Colored circles indicate insitu reports. The large red dot marks location of Denver, CO.

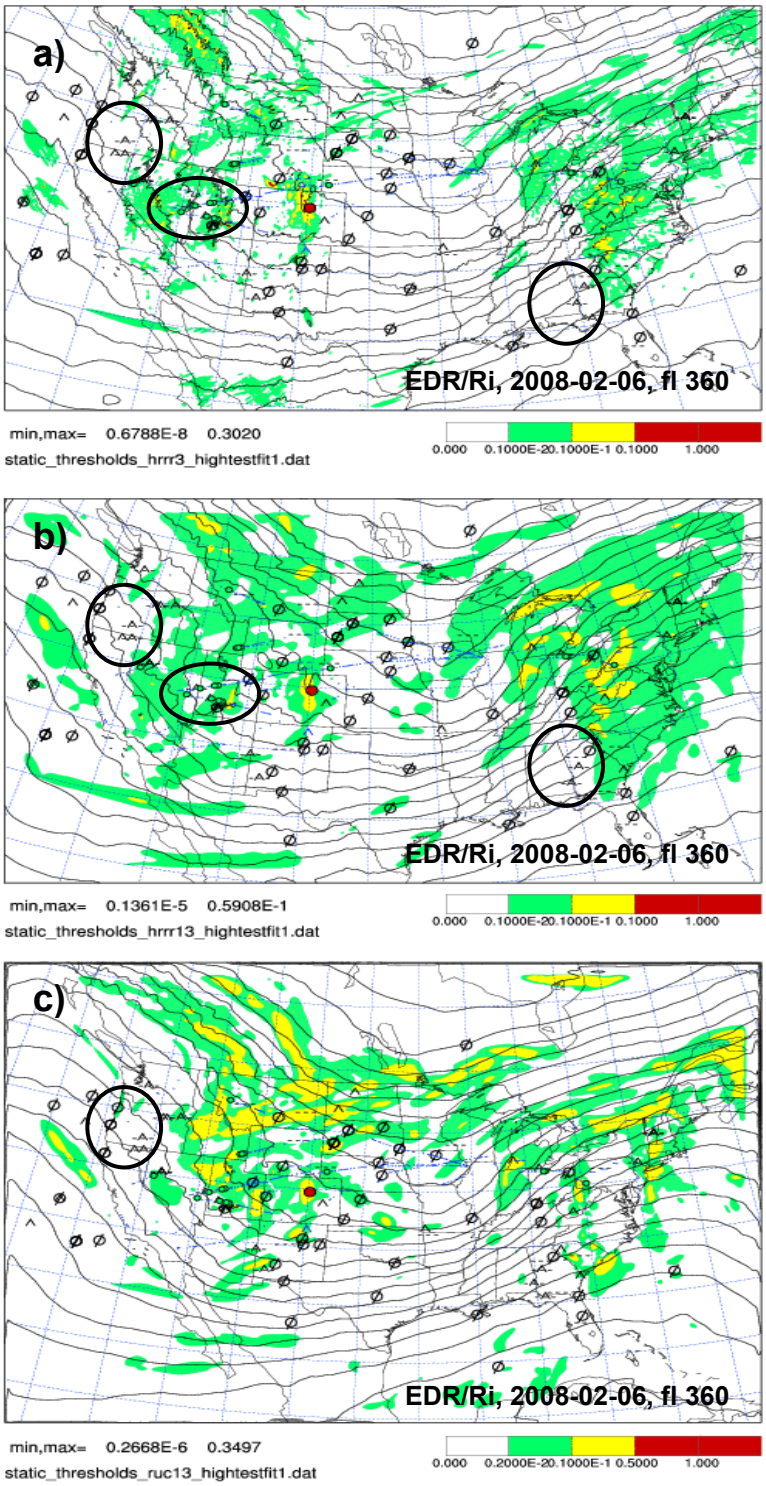


Figure 4.3, As in Fig. 4.2, except plots are for EDR/Ri index.

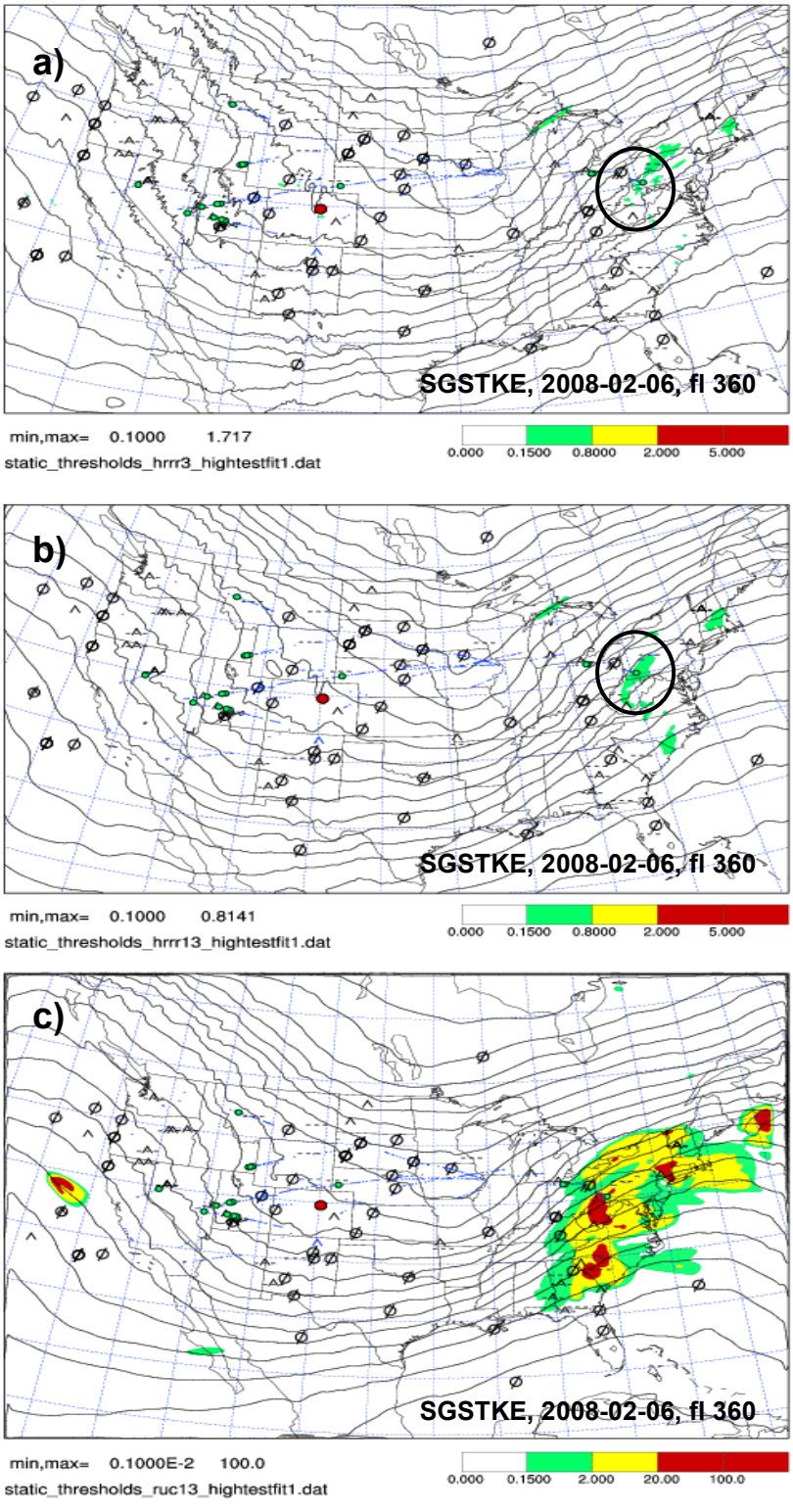


Figure 4.4, As in Fig. 4.2, except plots are for SGSTKE index.

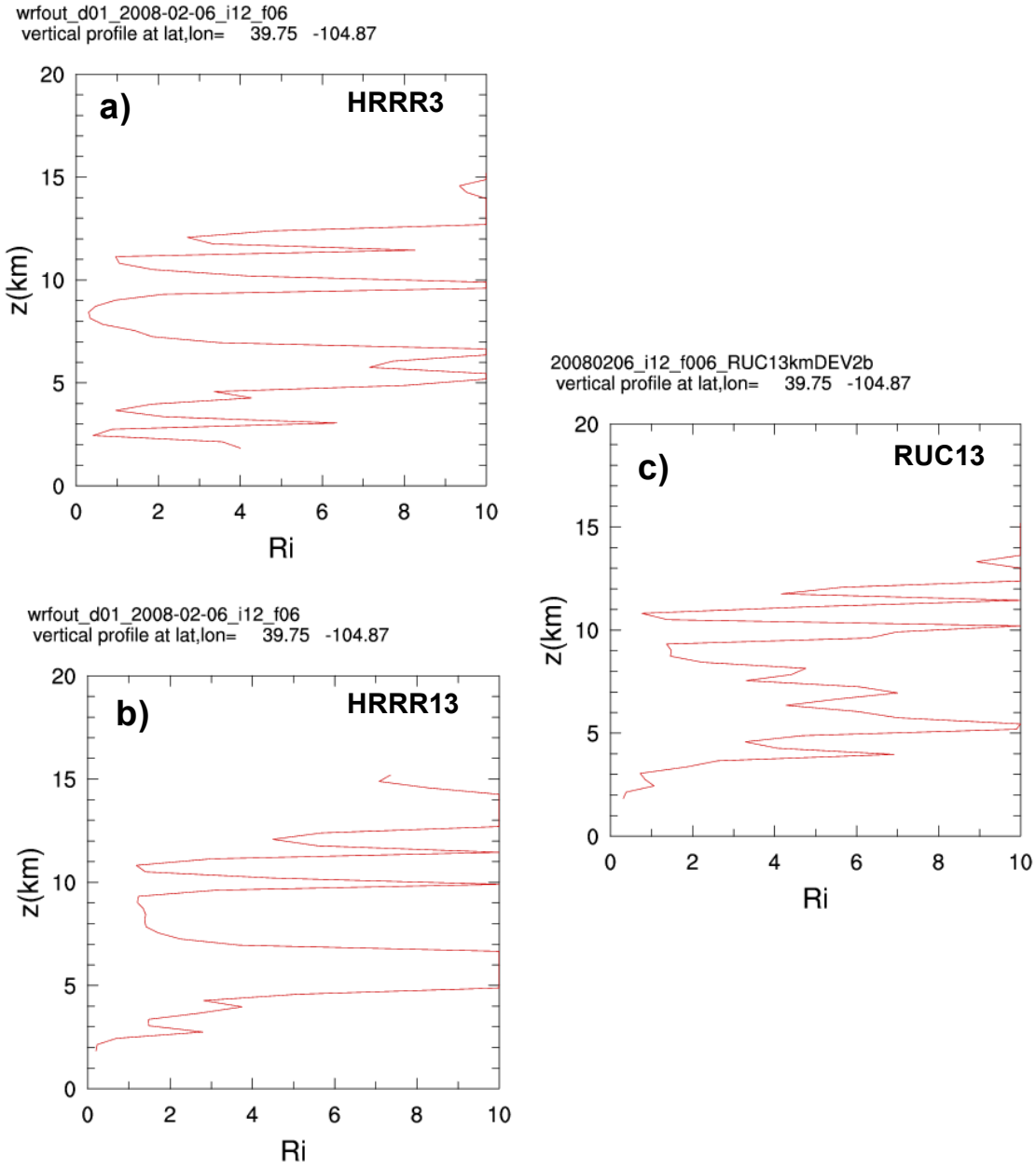


Figure 4.5, Vertical profiles of Richardson number for the grid point closed to Denver, CO for a) HRRR3, b) HRRR13, and c) RUC13 for Feb. 6, 2008.

4.1.1 Case 1 - 6 Feb 2008

Turbulence reports on Feb 6, 2008, 18 UTC, were wide-spread over the western U.S at FL360. A series of light to moderate reports extended from western Nevada and northern California through Oregon and Washington. Additionally, there was a cluster of light insitu reports over Utah, as well as light and light to moderate reports over New Mexico and Colorado. A scattering of light and light to moderate reports extended from Georgia up to Vermont.

Comparison of the 6 hour forecast Ellrod index at fl 360 for Feb 6, 2008 is shown in Fig. 4.2. Overall, for this altitude and index the areal extent of forecast turbulence was greater in the HRRR3 than in HRRR13. HRRR3 was better than HRRR13 at predicting the regions of turbulence captured by the insitu data over Utah and pireps over Georgia. Patches of severe turbulence predicted by HRRR3 over Colorado and RUC13 over the east coast were unsubstantiated. Regions of null pireps were better predicted by HRRR3 and HRRR13 than by RUC13. A number of null pireps were close to the boundary between null and light in HRRR3, suggesting that changes in threshold values might improve the forecast.

In contrast to the Ellrod index, the areal extent of turbulence as indicated by EDR/Ri is smaller for HRRR3 than HRRR13 (Fig. 4.3) for this same altitude. HRRR3 and HRRR13 partially capture the light turbulence reports in Utah. In all three models this particular index missed the turbulence clusters over Northern California/ Southern Oregon and western Georgia. The band of null reports over the middle of the country is well captured, especially by HRRR3.

The MYJ SGS TKE, based on sub-grid scale turbulence kinetic energy, predicts little or no turbulence across the U.S. for both HRRR3 and HRRR13 (Fig. 4.4). Only the light in situ report over western Pennsylvania is well represented. In contrast the RUC13 predicts significant patches of severe turbulence over the Eastern U.S where none was reported. For this index, none of the three models captured the reported turbulence over western U.S.

Fig. 4.5 shows vertical profiles of Richardson number at the model grid point closest to Denver, CO for all three models. While all three models show regions of low Ri around 9 and 11 km, RUC13 shows much more structure in Ri compared with HRRR3 and HRRR13.

4.1.2 Case 2 - 4 June 2008

June 4, 2008 was characterized by numerous turbulence reports throughout the Rocky Mountain region at 18 UTC, FL350. Both in situ and pilot reports were especially numerous over Wyoming, including a severe in situ report over southeastern Wyoming. Light and light to moderate reports also occurred throughout California. Only four pireps indicated turbulence over the eastern U.S. The three forecast indices for this time and date are shown in Figs. 4.6-4.8.

As with the previous case, both the severity and areal extend of turbulence as indicated by the Ellrod index was greater in HRRR3 than HRRR13. In general HRRR3 did a better job than HRRR13 of capturing turbulence over the Rocky Mountain region, and HRRR3 captured the severe turbulence in southern Wyoming (Fig. 4.6). The large swath of severe turbulence the HRRR3 Ellrod index predicted near the California – Nevada border was unsubstantiated. HRRR13 predicted no regions of severe turbulence at this altitude and index, whereas RUC13 had large patches over the eastern U.S. and a few small patches over the middle of the country, including southern Colorado.

The EDR/Ri index from the HRRR3 did match observations better than this index from HRRR13 (Fig. 4.7); note small patches of severe turbulence in southern Wyoming. EDR/Ri from the RUC13 model also missed the severe turbulence in Wyoming. All three models indicated stronger turbulence along the California-Nevada border than was reported.

The MYJ SGS TKE in both HRRR3 and HRRR13 indicated only a few very small patches of turbulence over the entire U.S, with moderate or greater limited to Nebraska (Fig. 4.8). RUC13 SGS TKE, in contrast, predicted several patches of severe turbulence over the Rocky Mountain region, as well as a few patches in the eastern U.S. and off the N.E. coast.

Vertical profiles of the Richardson number from the grid point closest to Denver, CO from all three models are shown in Fig. 4.9. The models consistently show minimum values of Ri just below 10 km. A second minimum around 13 km is more pronounced in RUC13 and HRRR13 than HRRR3.

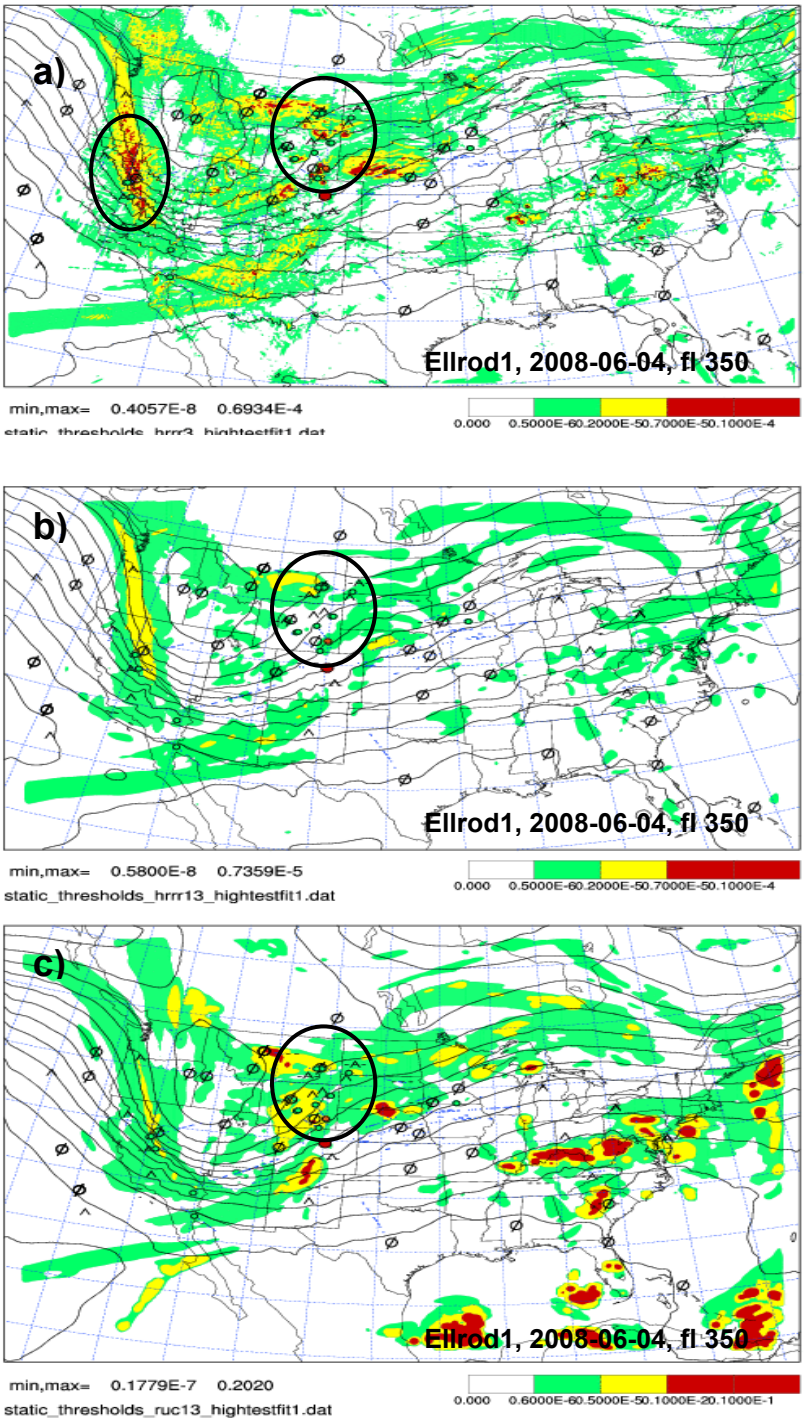


Figure 4.6, Ellrod index for 6 hour forecast initialized at 12 UTC for fl 350 on June 4, 2008 for model a) HRRR3, b) HRRR13, and c) RUC13. Light, moderate and severe turbulence is indicated by green, yellow and red, respectively. Colored circles indicate insitu reports. The large red dot marks location of Denver, CO.

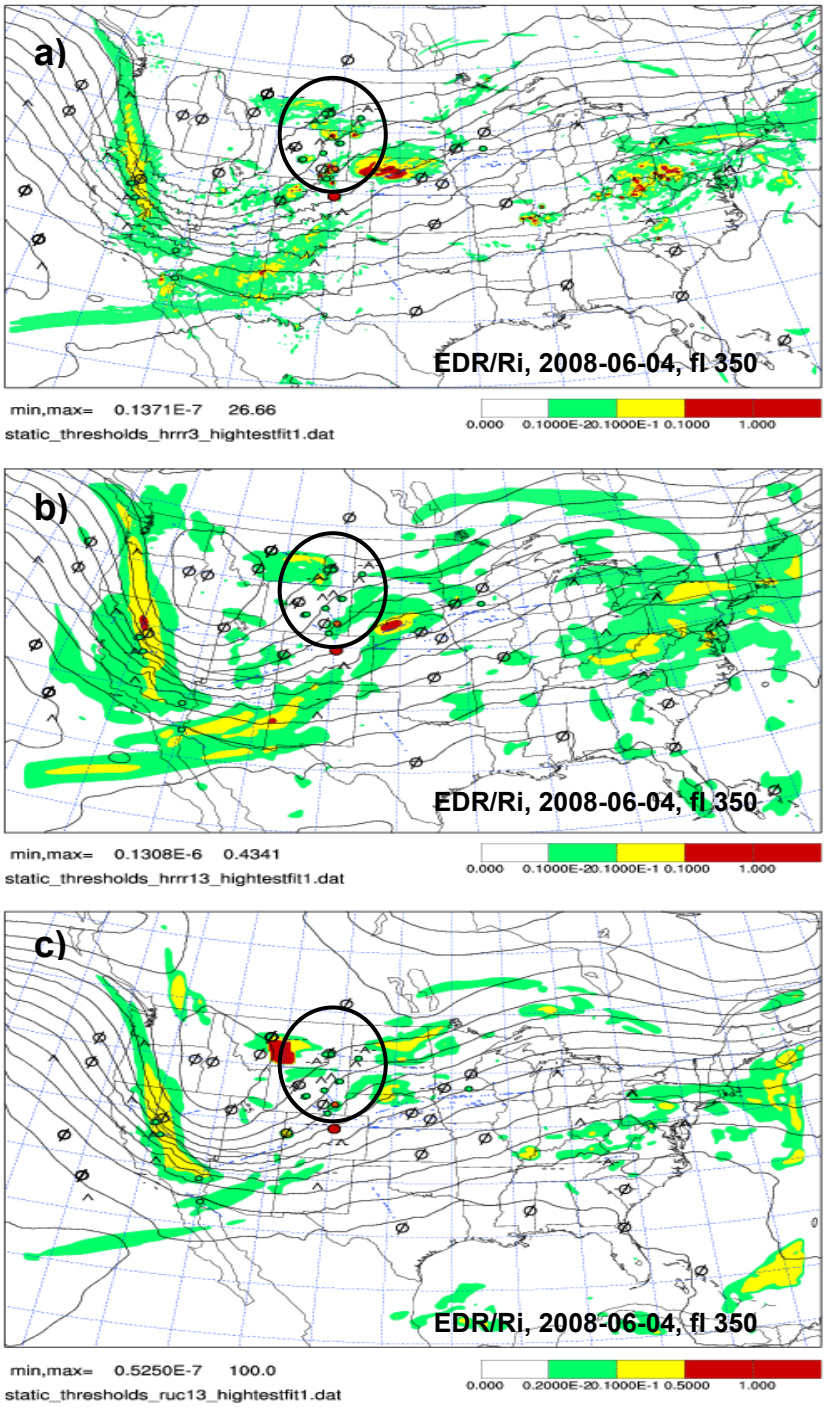


Figure 4.7, As in Fig. 4.6, except plots are for EDR/Ri index.

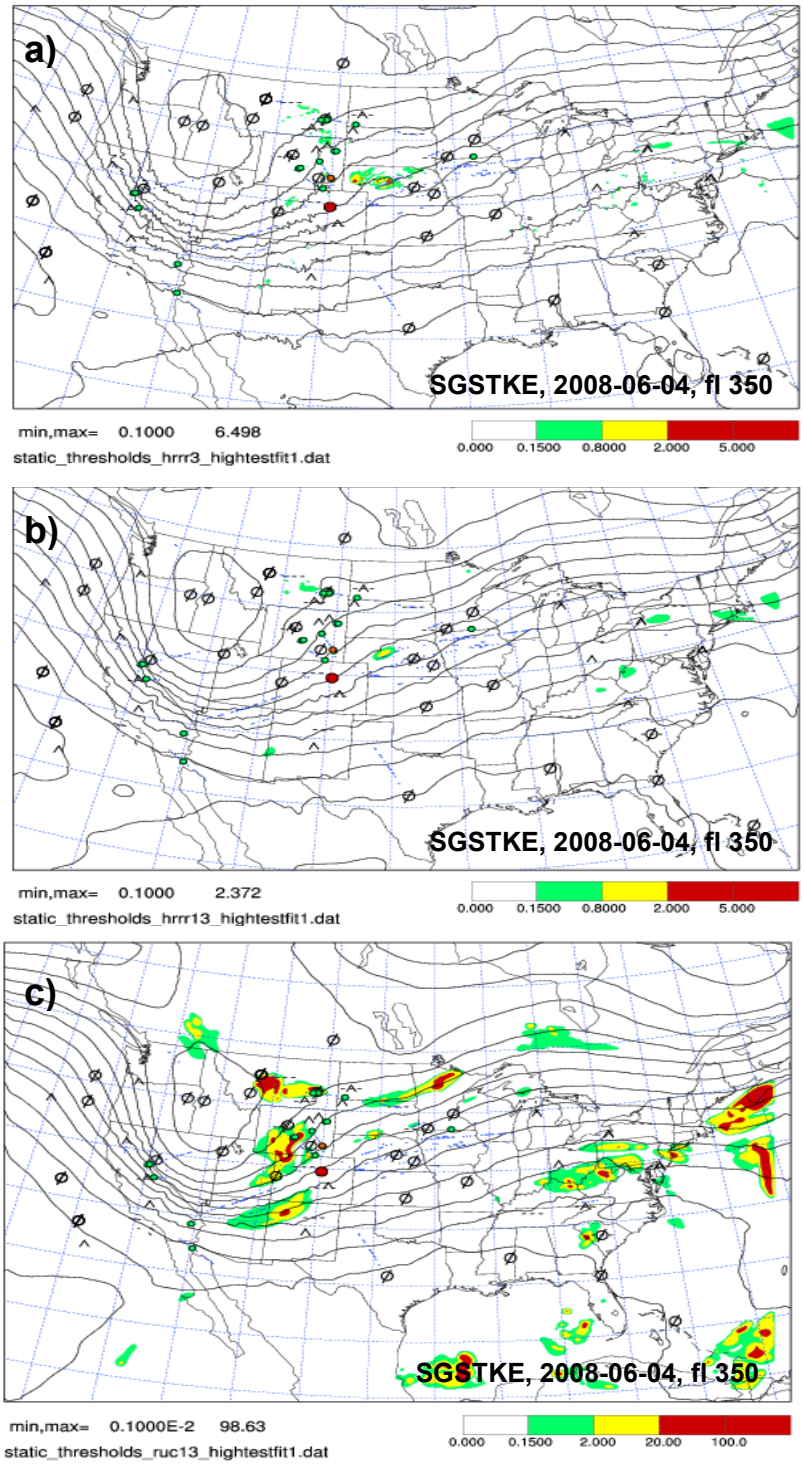


Figure 4.8, As in Fig. 4.6, except plots are for SGSTKE index.

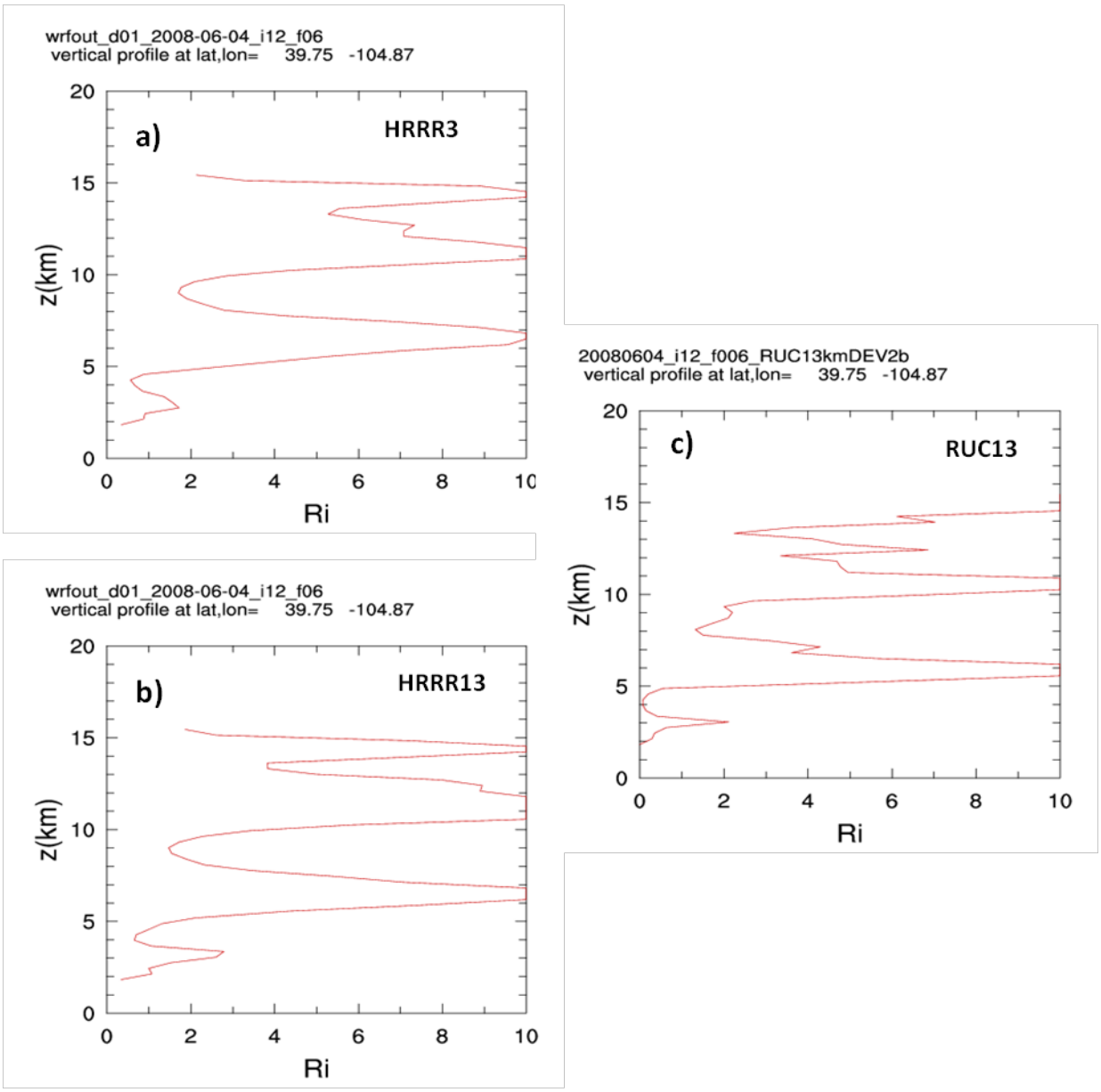


Figure 4.9, Vertical profiles of Richardson number for the grid point closed to Denver, CO for a) HRRR3, b) HRRR13, and c) RUC13 for June 4, 2008.

4.1.3 Case 3 - 9 Oct 2009

On Oct. 9, 2009 at FL290 and 18 UTC turbulence reports, including a number of severes, were primarily clustered along a swath running from eastern Kansas up through Wisconsin.

Both HRRR3 and HRRR13 were superior to RUC13 in predicting severe turbulence in this region using the Ellrod index (Fig. 4.10). For EDR/Ri, HRRR13 was the best model for predicting the extensive region of severe turbulence (Fig. 4.11).

In contrast to other cases discussed here, the MYJ SGS TKE in both HRRR3 and HRRR13 did a good job of capturing the cluster of severe pireps over eastern Kansas, as shown in Fig. 4.12. This was not true for SGS TKE from the RUC13.

The Richardson number profiles from Denver, CO show a pronounced minimum at 12 km in all 3 models (Fig. 4.13).

4.1.4 Case 4 - 9 Dec 2009

A number of light to moderate reports of turbulence extended along an east-west line from Colorado to Missouri at FL230 on Dec 9, 2009, 18UTC. Figs. 4.14 and 4.15 show this was captured by both the Ellrod and EDR/Ri indices in all three models. However, much of the large areal extent of turbulence outside of this region predicted by the Ellrod index in the HRRR3 was unconfirmed. All three models missed the moderate to severe report over Texas. The MYJ SGS TKE from all three models did a poor job of predicting turbulence on this day (Fig. 4.16).

Vertical profiles of the Richardson number from the grid point closest to Denver, CO from all three models are shown in Fig. 4.17. Although they all predict a minimum between 5 and 10 km, the region of low Ri is much deeper in RUC13.

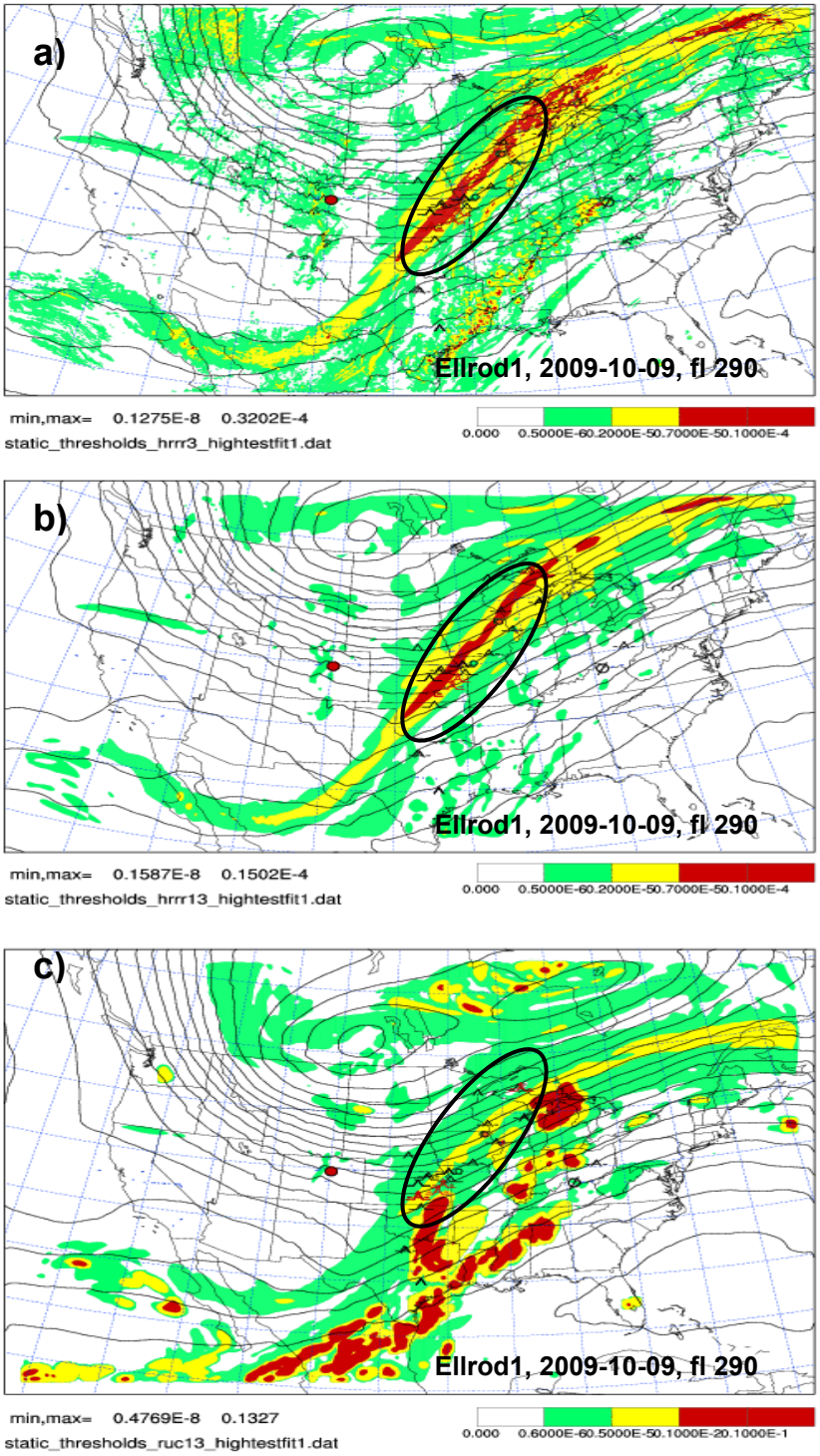


Figure 4.10, Ellrod index for 6 hour forecast initialized at 12 UTC for fl 290 on Oct. 9, 2009 for model a) HRRR3, b) HRRR13, and c) RUC13. Light, moderate and severe turbulence is indicated by green, yellow and red, respectively. Colored circles indicate insitu reports. The large red dot marks location of Denver, CO.

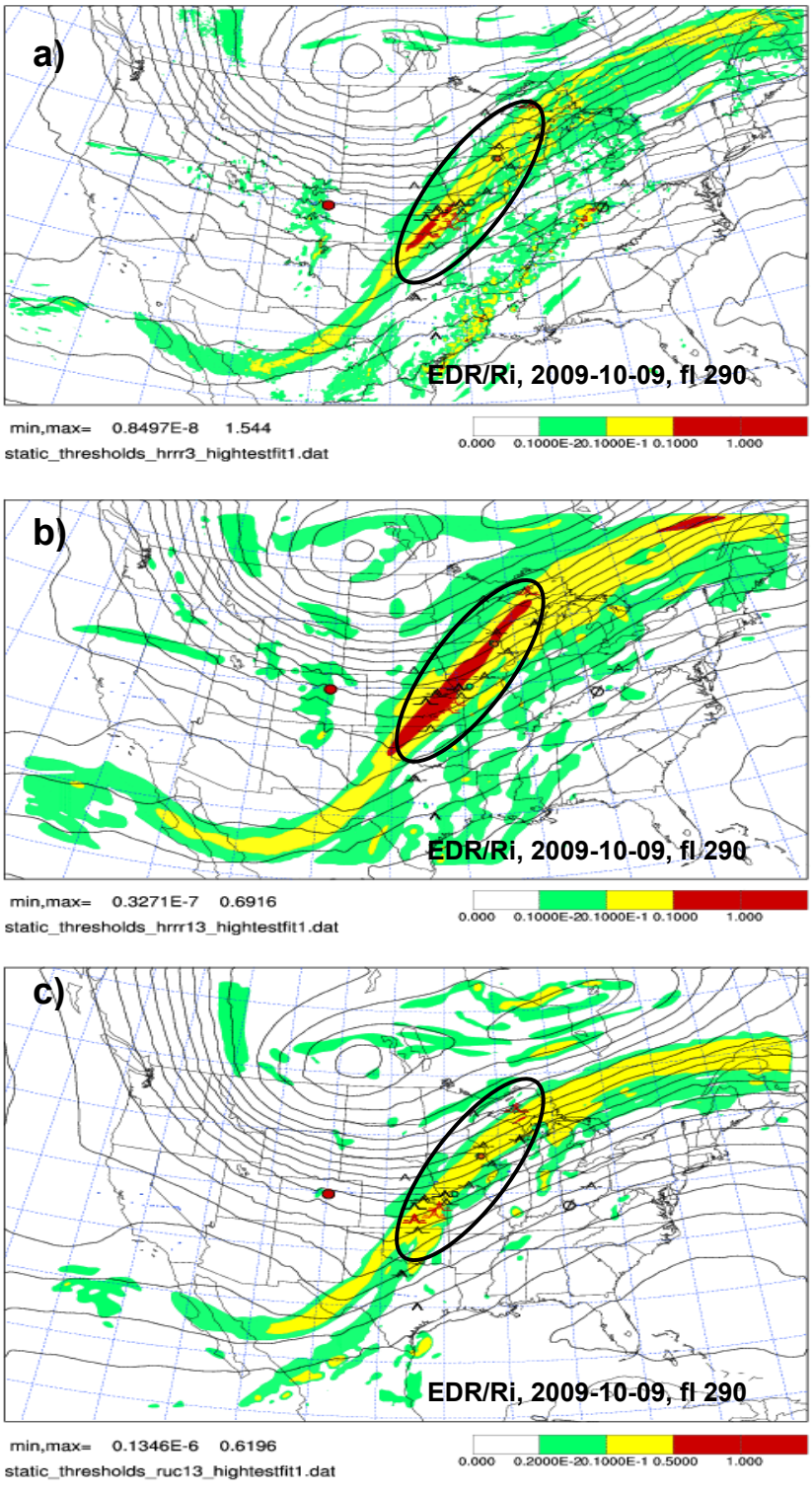


Figure 4.11, As in Fig. 4.10, except plots are for EDR/Ri index.

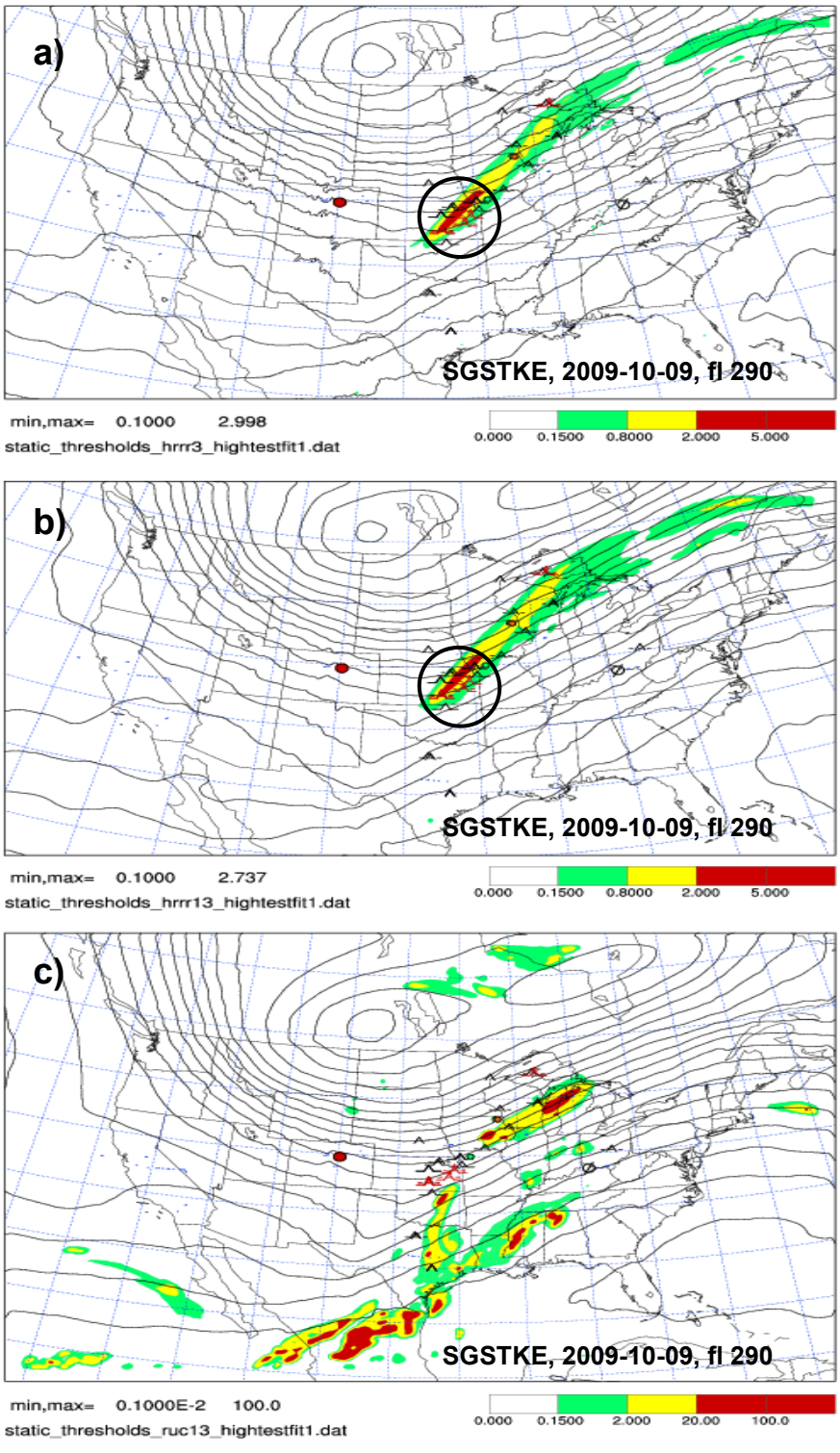


Figure 4.12, As in Fig. 4.10, except plots are for SGSTKE index.

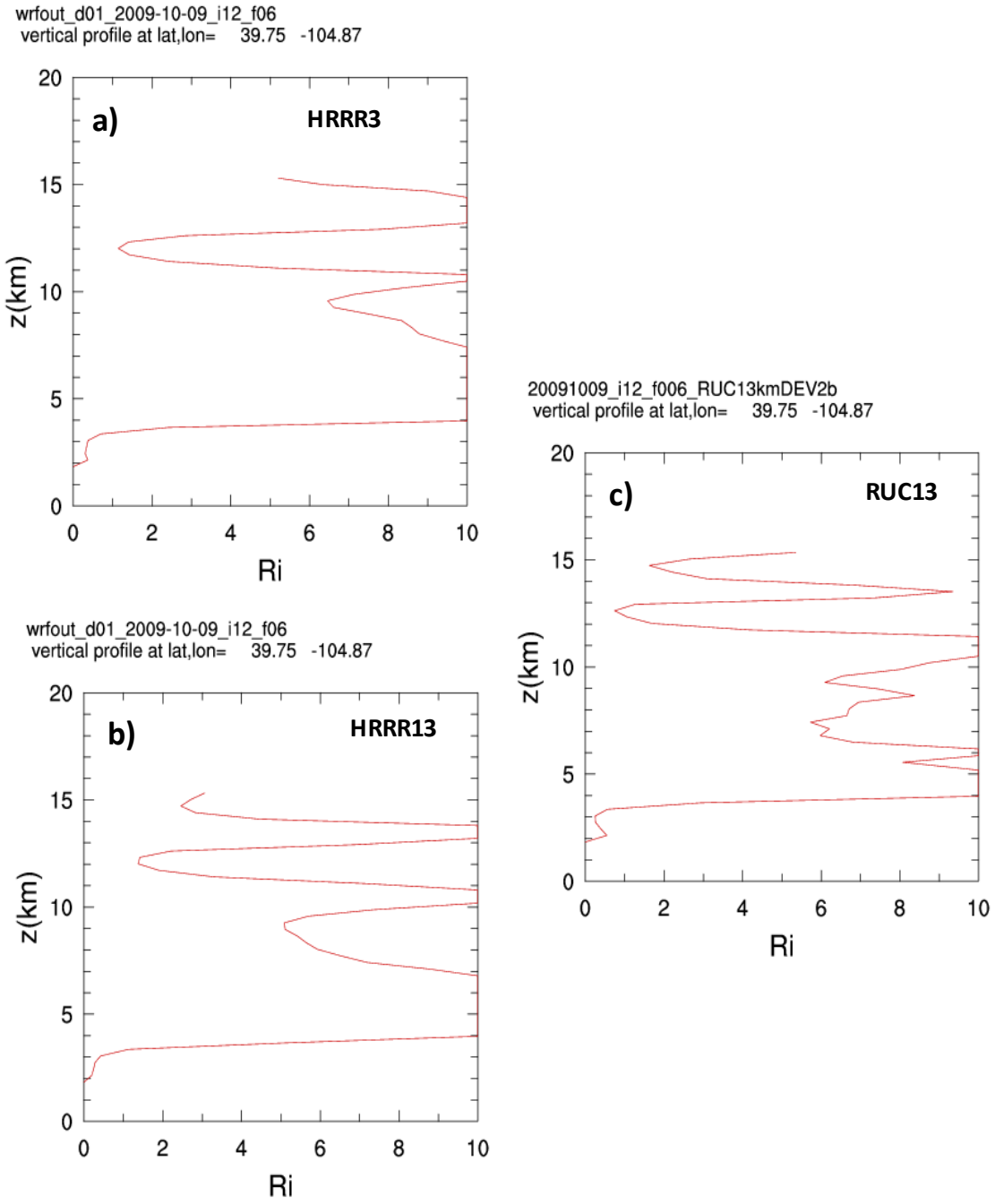


Figure 4.13, Vertical profiles of Richardson number for the grid point closed to Denver, CO for a) HRRR3, b) HRRR13, and c) RUC13 for Oct 9, 2009.

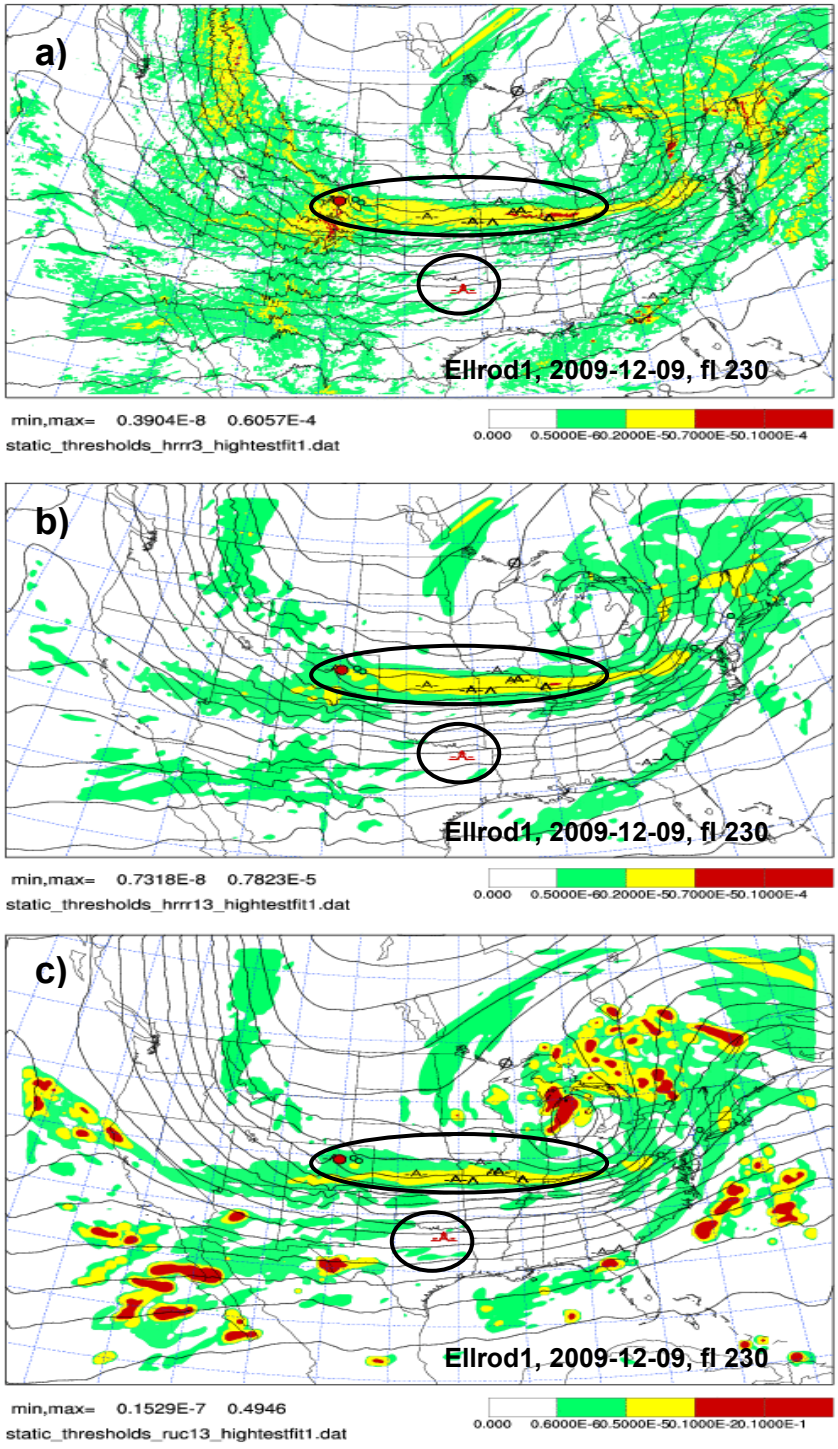


Figure 4.14, Ellrod index for 6 hour forecast initialized at 12 UTC for fl 230 on Dec. 9, 2009 for model a) HRRR3, b) HRRR13, and c) RUC13. Light, moderate and severe turbulence is indicated by green, yellow and red, respectively. Colored circles indicate in situ reports. The large red dot marks location of Denver, CO.

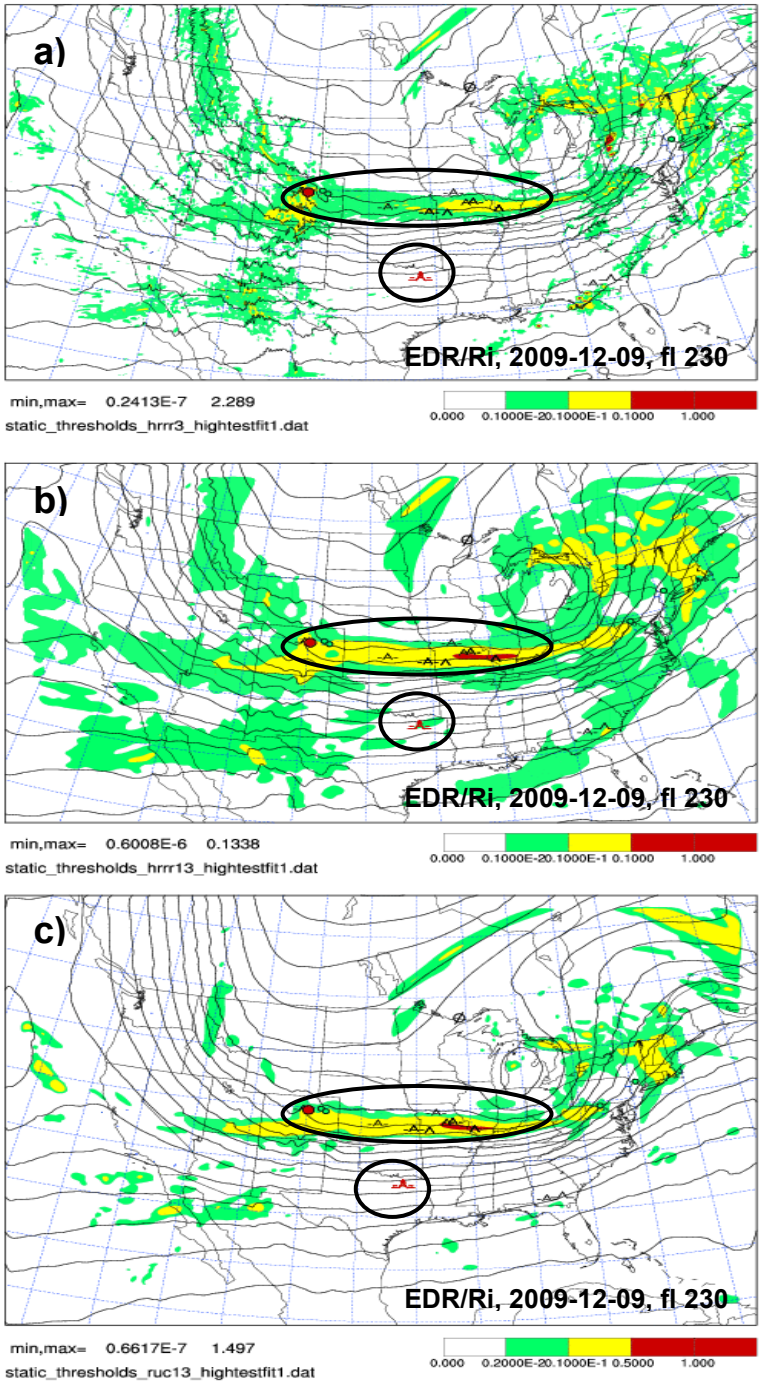


Figure 4.15, As in Fig. 4.14, except plots are for EDR/Ri index.

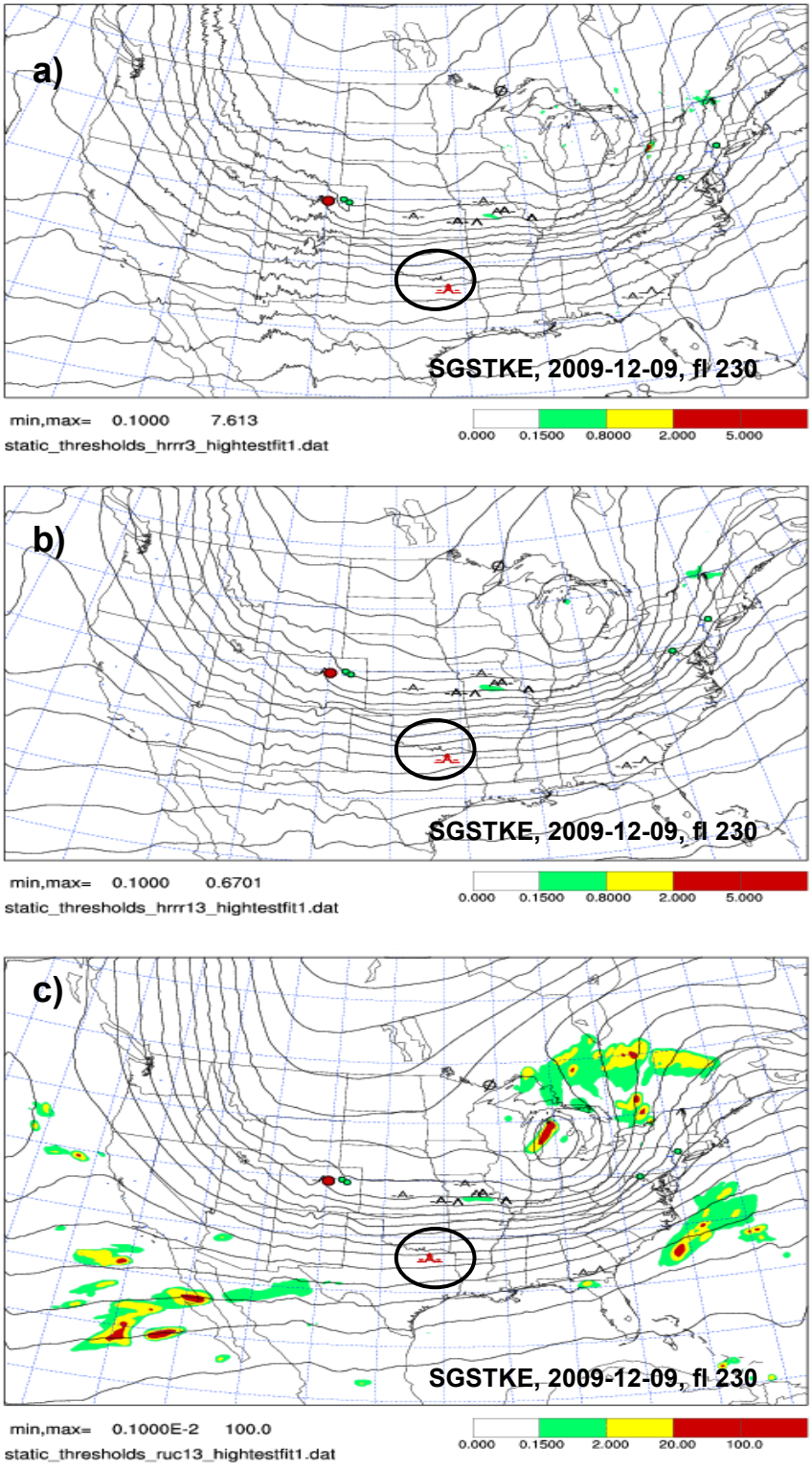


Figure 4.16, As in Fig. 4.14, except plots are for SGSTKE index.

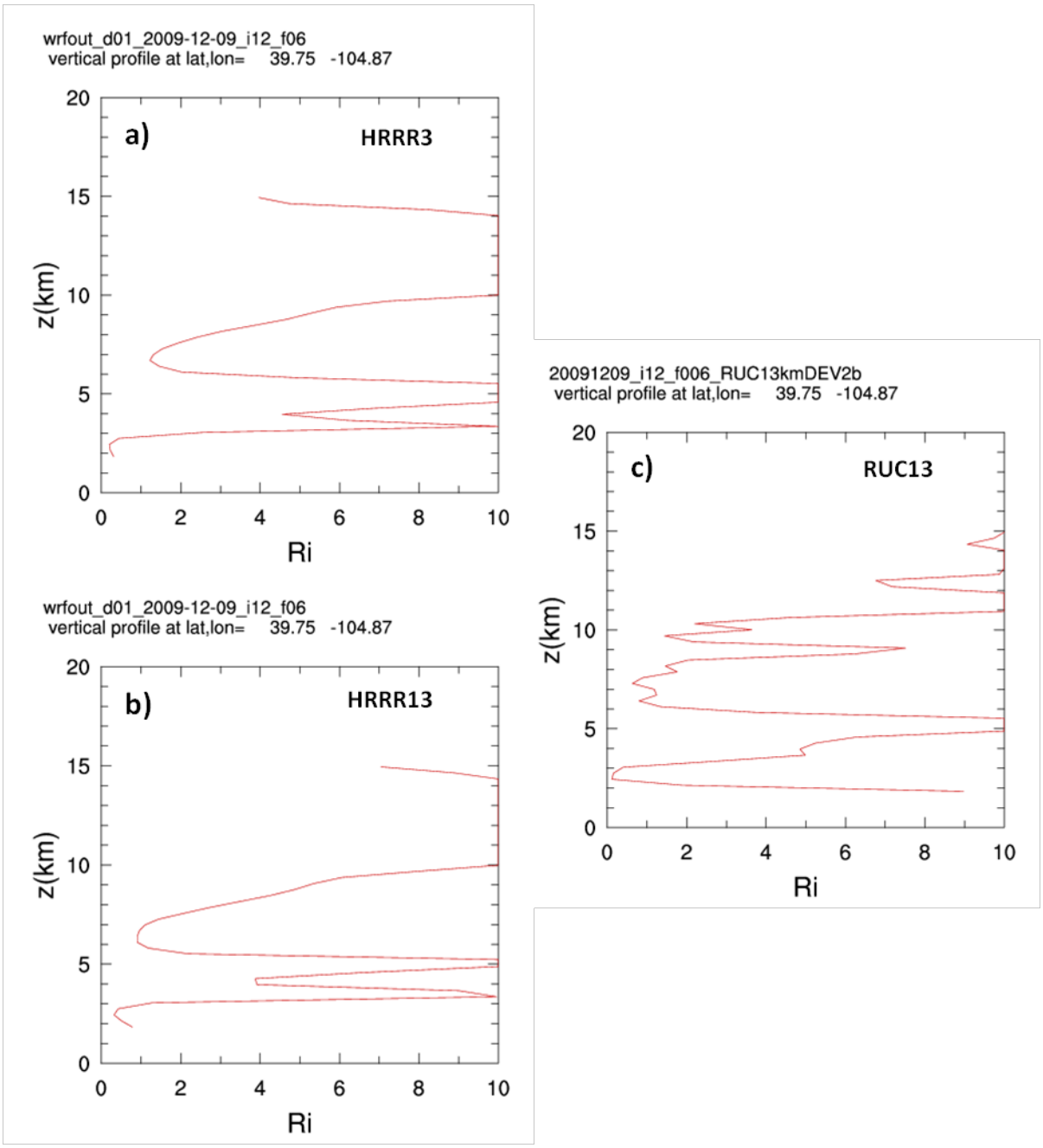


Figure 4.17, Vertical profiles of Richardson number for the grid point closed to Denver, CO for a) HRRR3, b) HRRR13, and c) RUC13 for Dec. 9, 2009.

4.2 QUANTITATIVE ASSESSMENT OF BASELINE RUNS

Quantitative analyses of the baseline cases are constructed by computing the discrimination capability of individual diagnostics. The GTG diagnostic is not included. Here the discrimination performance is based on a diagnostic's ability to discriminate between null or smooth events from moderate-or-greater (MOG) events. There are many measures of discrimination performance, most of which depend on a threshold, in this case, for MOG events. To avoid the difficulty of specifying a threshold, receiver operating characteristic or ROC curves are constructed. In this approach various threshold values are tried, and for each threshold probabilities of detection (POD) are computed. There are two PODs of interest, one is the POD of capturing the MOG events (PODY), the other is the POD of capturing the null (smooth) events (PODN). Curves obtained be varying the thresholds and tracing out the resultant PODY vs. PODN values are the ROC curves. The higher the curve in the PODY-PODN diagram (upper left corner) the better the overall discrimination performance. Equivalently, the larger the area under the ROC curves (AUC), the better the discrimination performance. See Sharman et al. (2006) for details.

However, the discrimination performance based on the AUC or any other metric does have some uncertainty in it because of the qualitative nature of the PIREP intensity and the uncertainty in the location of the PIREPs (average ~50 km) and even the in situ EDR data (average ~ 6 km). Note that both observation sources have a location uncertainty greater than the HRRR3 grid resolution and the PIREPs have an average location uncertainty greater than both HRRR13 and RUC13 grid resolutions. This causes some uncertainty in the placement of the ROC curves and therefore some uncertainty in the evaluation of the performance. Further only the four cases were available for discrimination capability assessment, and more cases would be necessary to better quantify the ROC curve values and the uncertainty in the values.

POD curves for 6-hr forecast performance based on the sum of the four cases for the HRRR13, HRRR3, and RUC13 underlying NWP models are shown in Fig. 4.18. Each curve in the diagram represents the performance of an individual diagnostic based on the roughly 6400 available observations (PIREPs and in situ); the curves in green indicate the EDR algorithm, and below those the curves in blue indicate the Ellrod diagnostic. The lower blue curve indicates the SGS TKE result. By eye there is no obvious difference in skill from the three different underlying NWP models.

Another way of assessing the differences is to compute the AUCs for individual diagnostics and these are tabulated in the table below. Here we only present the AUCs for the 10 best diagnostics and the SGS TKE diagnostic.

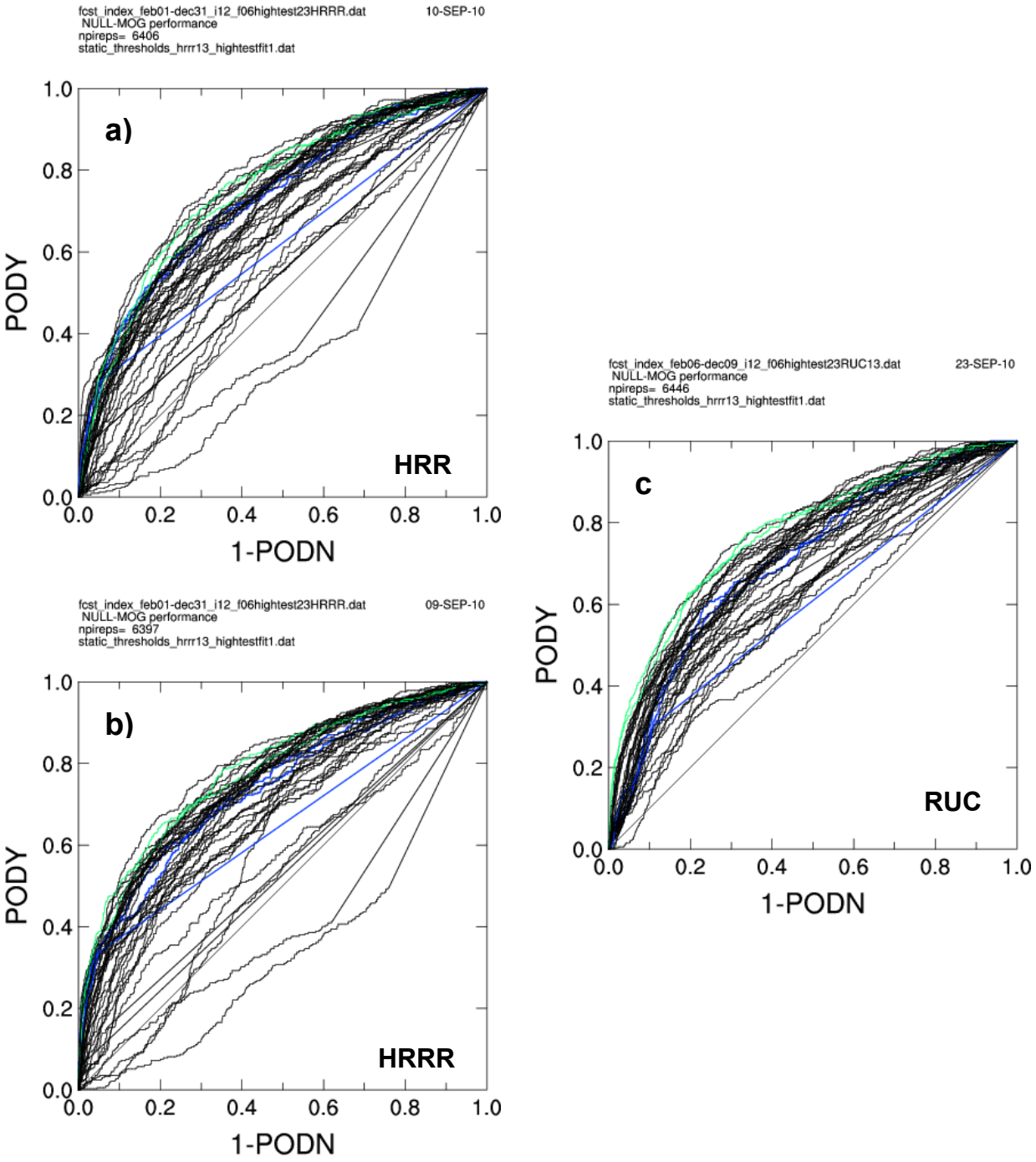


Figure 4.18, ROC curves for a) HRRR3, b) HRRR13, and c) RUC13.

Table 4.1 AUCs for the 10 best diagnostics and SGS TKE diagnostic.

Diagnostic	RUC13 AUC	HRRR13 AUC	HRRR3 AUC
EDR/Ri	0.782	0.790	0.762
Fθ/Ri	0.754	0.798	0.781
NCSU2/Ri	0.777	0.765	0.745
UBF/Ri	0.754	0.770	0.738
Iawind/Ri	0.741	0.751	0.731
CT2/Ri	0.731	0.762	0.698
Ellrod1	0.711	0.746	0.736
NGM2	0.705	0.761	0.745
DEF2/Ri	0.765	0.783	0.767
LHFK/Ri	0.741	0.755	0.729
SGSTKE	0.597	0.646	0.618

Again, there is some uncertainty in the areas given in the table, but some gross trends can be inferred.

1) The AUC for most of these diagnostics is 0.7-0.8. A perfect discriminator would have an AUC of 1.0.

2) For most of these diagnostics the HRRR3 performance is the worst of the three.

3) The single best performer is not the same for the three models, suggesting that a composite of diagnostics (GTG) would have to account for this.

4) The MYJ SGS TKE is a relatively poor performer, and if there were no post-processing system like GTG, this diagnostic, which comes directly from the NWP model output, would provide very poor performance.

4.3 SUMMARY OF TURBULENCE ANALYSES

Qualitative analyses of each of the four baseline cases was performed, as well as quantitative analyses based on statistical discrimination performance of the four cases combined. The overall impression based on these analyses indicate little significant advantage in the performance of the HRRR3 over either the HRRR13 or RUC13, in fact there is some evidence that the HRRR3 is the worst performer of the three. However, due to uncertainties in the observations, it is difficult to know this for sure.

Qualitatively, the synoptic patterns produced are very similar for the three models and therefore the overall patterns of predicted turbulence are more or less the same. The HRRR3 does seem to provide more fine scale detail horizontally, although it is not clear this finer scale detail is consistent with observations. Vertically, the fine scale detail provided by both the HRRR13 and HRRR3 is apparently not as well resolved as it is in the RUC13, indicating that more vertical levels in the HRRR might provide more realistic results.

Quantitatively, even accounting for uncertainties in the observation data, the MOG-NULL performance discrimination capability is not obviously different for any of the underlying NWP models.

Future work needs to better assess the quantitative performance of the HRRR13 and HRRR3. Sensitivity studies, some of which have already been initiated, need to be performed with regards to the model top, vertical resolution, and upper boundary condition used. Statistical evaluation methods that are more object-oriented would also reduce some of the uncertainty in the results.

5. ANALYSES FOCUSED ON IN-FLIGHT ICING

5.1 INTRODUCTION

In-flight icing occurs when supercooled (subfreezing) liquid water drops impact an aircraft while in-flight. Accurate numerical weather predictions of the presence and mass of supercooled liquid water (SLW) are of primary importance in correctly identifying icing conditions. Knowledge of the locations and times when SLW is present in the atmosphere are available primarily through pilot reports (PIREPs) that mention icing conditions. In addition to providing the locations and altitudes of the icing, the PIREPs include an icing severity (trace (1) – trace to light (2) – light (3) – light to moderate (4) - moderate (5) – moderate to severe (6) – heavy (7) – severe (8)). The accuracy of PIREP locations and times are of reasonable quality (Bernstein et al. 2005)).

Five configurations of the WRF were tested on 4 cases. A 3-km and 13-km baseline run, two 13-km boundary layer schemes (MYNN) and (YSU), along with a 13-km sensitivity (sen2) were all analyzed. Two winter cases February 6, 2008 and December 09, 2009 were run, along with a late spring case June 4, 2008, and a fall case October 9, 2009. The model was initialized at 1200 UTC and run through 12-hrs. The models produced 1 – 12-h forecasts of the thermodynamic and micro-structure of the atmosphere.

In this study we examine PIREPs, positive for icing, to test the five WRF model configuration's ability to predict icing conditions. We extracted the model output at the closest model location (within 0.02^{o} (lat/long) for the 3-km and 0.05^{o} for the 13-km) and altitudes (within in 300-m) of the icing PIREPs. Some PIREPs were only matched to the 3-km model since the closest 13-km gridpoint was more than 0.05^{o} from the reported icing location. The PIREP had to have occurred within one hour of the model valid time. In order to insure the model's microphysics were mature and had time to spin-up we only looked at output from the 6 and 9-hour forecasts. Here we present the results of the PIREP analysis. The case study results for several individual PIREPs are added in Appendix A.2.

5.2 ANALYSIS

A set of summary statistics was computed for each model configuration at the six and nine hour forecasts. These are defined as:

- **# grid points** – the number of three dimensional grid points matched to PIREPs
- **Pod SLW** – the percent of icing grid points at the PIREP locations that had any forecast SLW
- **Pod ice** – the percent of icing grid points that had only frozen condensate forecast
- **# cold grid points** – number of PIREPs matched to grid points that are colder than $-12^{o}C$
- **Pod Cold** – the percent of cold icing grid points that had any forecast SLW
- **Pod RH>90** – the percent of icing grid points that had forecast relative humidity > 90%
- **Pod W>0** – the percent of icing grid points that were associated with a forecast of rising air

- **Pod mog SLW** - the percent of icing grid points matched to moderate or greater icing severity (MOG) PIREP locations that had any forecast SLW
- **Pod mog ice** - the percent of mog icing grid points that had only frozen condensate predicted
- **# cold mog grid points** - number of mog PIREP matched grid points that are colder than -12^{o}C
- **Pod mog cold** - the percent of cold mog icing grid points that had any SLW predicted
- **Pod mog RH >90** – the percent of mog icing grid points that had a forecast of relative humidity > 90%
- **Pod mog W > 0** – the percent of mog icing grid points that were associated with forecast rising air

5.2.1 Case 1 (Feb 6, 2008)

The results of the summary statistics analysis for this case are presented in Table 5.1. Columns 2-6 present the WRF configurations from the 6-hour forecasts valid at 1800 UTC. Columns 7 – 10 present the 9-hr forecast valid at 2100 (the 13-km sensitivity run at 2100 UTC was not analyzed). At 1800 UTC the 3-km run had icing PIREPs mapped to 261 grid points. The majority of grid points were within a model predicted cloud (ice and/or SLW) but only 25% were correctly identified as containing SLW. The 13-km baseline, 13-km MYNN, 13-km YSU boundary layer schemes and 13-km sensitivity run also correctly identified the majority of icing grid points within cloud and were slightly better than the 3-km. But they also only had between 25 – 30% of the observed icing locations matched to model predicted SLW. None of the WRF configurations did well with the cold PIREPs with less than 5% of the observed icing within SLW conditions. The model-based relative humidities, which should be near 100% within icing conditions, were too low, especially the 3-km. This may be due to higher resolution ascent/descent vertical velocity couplets in the 3-km matched to the possible timing or spatial errors in the PIREPs. This would be more pronounced within the higher resolution WRF simulations.

Table 5.1, Summary statistics for the Feb 6, 2008 case. Columns 2-6 are for 1800 UTC (6 hour forecast) and 7-10 are for 2100 UTC (9 hour forecast).

	F6-3km	F6-13km	F6-13MYNN	F6-13YSU	F6-13sen2	F6-3km	F6-13km	F6-13MYNN	F6-13YSU
# grid points	261	138	138	138	138	292	156	156	157
Pod SLW	25.14	24.73	30.85	30.85	26.88	25.79	22.42	24.3	21.9
Pod ICE	34.85	43.01	36.17	37.23	39.78	34.21	47.66	44.86	48.57
# cold GPs	48	25	25	25	24	26	11	11	11
Pod Cold	4.16	4	4	4	0	3.84	0	0	0
POD RH>90	54.2	63.45	62.76	64.89	64.15	46.84	52.34	55.14	48.57
POD W>0	48.57	49.46	47.87	55.32	50.43	60.52	54.2	58.87	58.1
Pod mog SLW	33.33	31.58	30	50	42.11	19.23	15.6	18.75	15.6
Pod mog ICE	31.75	63.15	65	45	52.63	28.84	53.12	53.13	53.13
# cold mog GPs	13	4	4	4	4	14	8	8	8
Pod mog Cold	0	25	25	25	0	7.14	0	0	0
POD mog RH>90	50.79	84.21	80	80	84.21	40.38	50	46.87	40.62
POD mog W>0	69.84	63.16	65	85	63.16	61.53	59.37	56.25	46.88

The mass of supercooled liquid at the icing locations where the model correctly identified the icing clouds (i.e. supercooled liquid forecast at the icing PIREP location) was tabulated. A higher mass of WRF predicted SLW should correlate to more severe icing. The PIREP icing severity values are reported as 8 categories. All the PIREPs of matching severity for the case and 6 hour forecast time were grouped and the results presented in Figure 5.1. The 13-km had an increasing mass of SLW as the PIREP severity increases. The 3-km, with more gridpoints matched, had less mass than the 13-km at the 1, 3, 5 severity levels. The severity level 5 showed lower SLW mass than level 4 for the 3km model runs.

The distribution of supercooled liquid mass when it was correctly predicted for the 9 hour forecast is presented in figure 5.2. The 13-km had more supercooled liquid mass predicted for the moderate icing PIREPs than the 3-km and the mass increased as severity increased. The 3-km had an increasing median SLW mass between the trace and light icing (1-3) but no increase for the moderate PIREPs. Moderate icing usually required at least greater than 0.2 g kg^{-1}, which raises the question with regard to the level 5 matched data (.e.g, PIREP location accuracy, or incorrect microphysical phase).

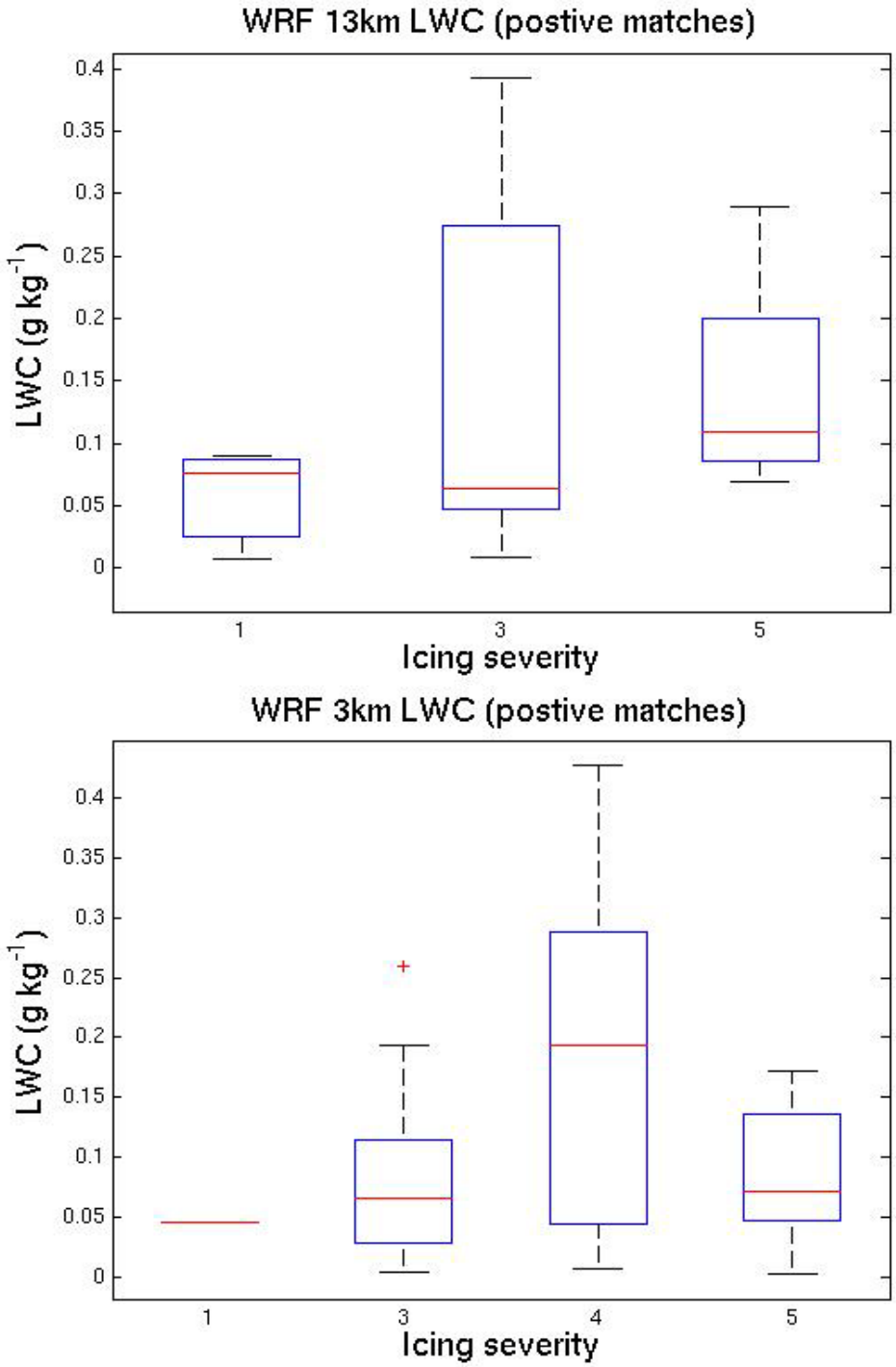


Figure 5.1, Summary of the mass of supercooled liquid at the locations where the model correctly identified icing conditions is shown above. The boxes summarize the PIREPs by reported icing severity (1 trace icing to 8 severe icing). 13-km (top) and 3-km (lower panel).

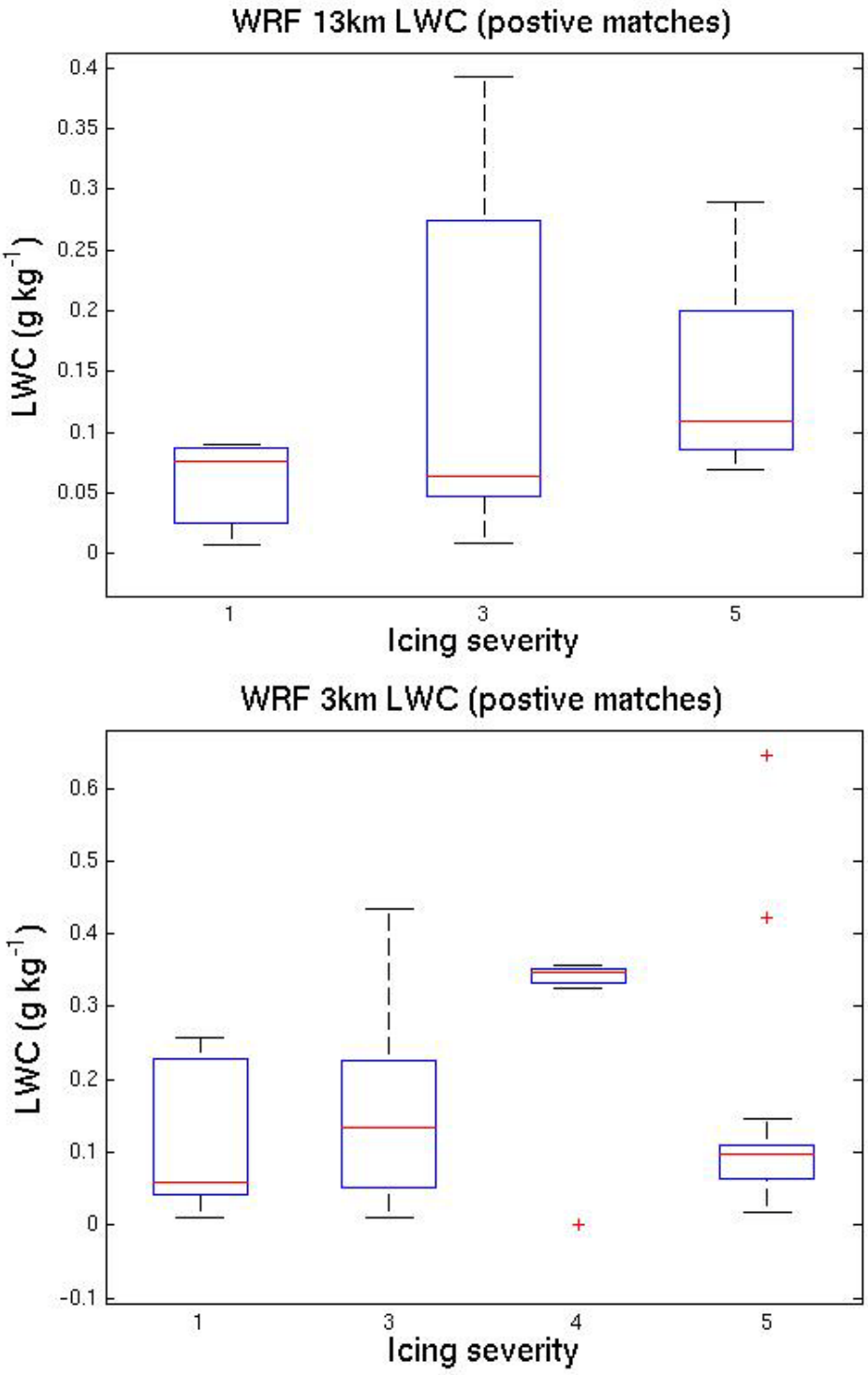


Figure 5.2, Summary of the mass of supercooled liquid at the locations were the model correctly identified icing conditions as a function of PIREP reported icing severity (1 trace icing to 8 severe icing). 13-km (top) and 3-km (lower panel).

5.2.2 Case 2 (June 4, 2008)

The results of the summary statistics analysis are presented in Table 5.2. Columns 2-6 present the WRF configurations from the 6-hour forecasts valid at 1800 UTC. Columns 7 – 10 present the 9-hr forecast valid at 2100 (the 13-km sensitivity run at 2100 UTC was not analyzed). Clouds forming in this summer case likely have smaller spatial and temporal scale. At 1800 UTC the 3-km run had icing PIREPs mapped to 75 grid points. Just under half of the grid points were within a model predicted cloud (ice and/or SLW) but only 8.8% were correctly identified as containing SLW. The 13-km baseline, 13-km MYNN, 13-km YSU boundary layer schemes and 13-km sensitivity run correctly identified the majority of icing grid points in a model cloud. But they only had between 7 – 15% of the observed icing within model predicted SLW. None of the WRF configurations did well with the cold PIREPs. The model-based relative humidities, which should be near 100% within icing conditions, were too low, especially the 3-km. By 2100 UTC the percentage of icing reports that matched model cloud (i.e., Pod SLW + Pod ICE) predicted by the various model configurations increased. For example, at 1800 UTC the 3-km is mapped to ~50% model clouds while at 2100 UTC it is matched to ~83% model clouds; the 13km MYNN is mapped to ~59% clouds at 1800 UTC and 100% clouds at 2100 UTC.

The distribution of supercooled liquid mass, when it was correctly predicted, at 1800 UTC on 4 June 2008 is presented in Figure 5.3. The 13-km sample size was low and only contained light icing. The median SLW value was 0.15 g kg^{-1}. The 3-km had a larger set of icing PIREPs correctly identified and the moderate group had a very large 0.8 g kg^{-1} median value. This mass is much larger than is typically found in the 13-km operational runs. By 2100 UTC the 13-km results (not shown) are very similar to the 1800 UTC run. The 3-km SLW mass values have returned to smaller than expected values (Fig. 5.4).

Table 5.2, Summary statistics for the Jun 4, 2008 case. Columns 2-6 are results from 1800 UTC and 7-10 are from 2100 UTC.

	J4-3km	J4-13km	J4-13MYNN	J4-13YSU	J4-13sen2	J4-3km	J4-13km	J4-13MYNN	J4-13YSU
# grid points	75	48	47	47	48	57	34	34	34
Pod SLW	8.82	14.81	14.81	7.4	14.81	31.42	0	6.25	6.25
Pod ICE	41.17	51.85	44.44	59.26	48.15	51.43	68.75	93.75	87.5
# cold GPs	20	16	17	17	16	17	13	13	13
Pod Cold	5	0	0	0	0	23.53	0	0	0
POD RH>90	20.58	44.44	44.44	44.44	44.44	42.86	6.25	18.85	18.5
POD W>0	52.94	48.15	48.15	62.96	48.15	45.71	75	87.5	100
Pod mog SLW	10	0	0	0	0	58.33	0	0	0
Pod mog ICE	60	80	60	80	70	25	83.33	100	83.33
# cold mog GPs	6	7	7	7	7	6	6	6	6
Pod mog Cold	16.67	0	0	0	0	66.67	0	0	0
POD mog RH>90	0	30	30	30	30	83.33	0	0	16.67
POD mog W>0	50	30	30	70	30	66.67	83.33	100	100

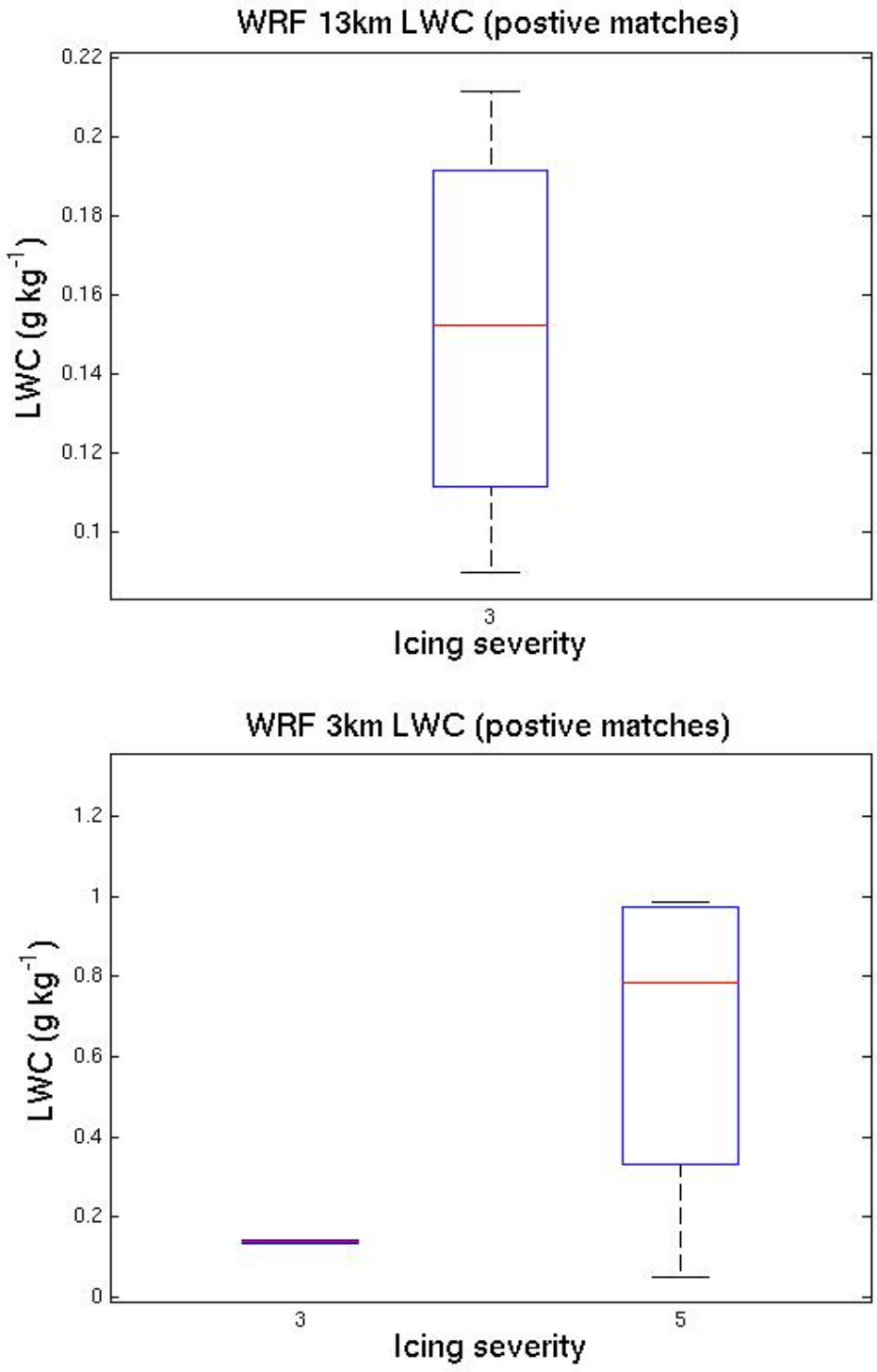


Figure 5.3, Summary of the mass of supercooled liquid at the locations were the model correctly identified icing conditions, as a function of PIREP reported icing severity (1 trace icing to 8 severe icing). 13-km (top) and 3-km (lower panel).

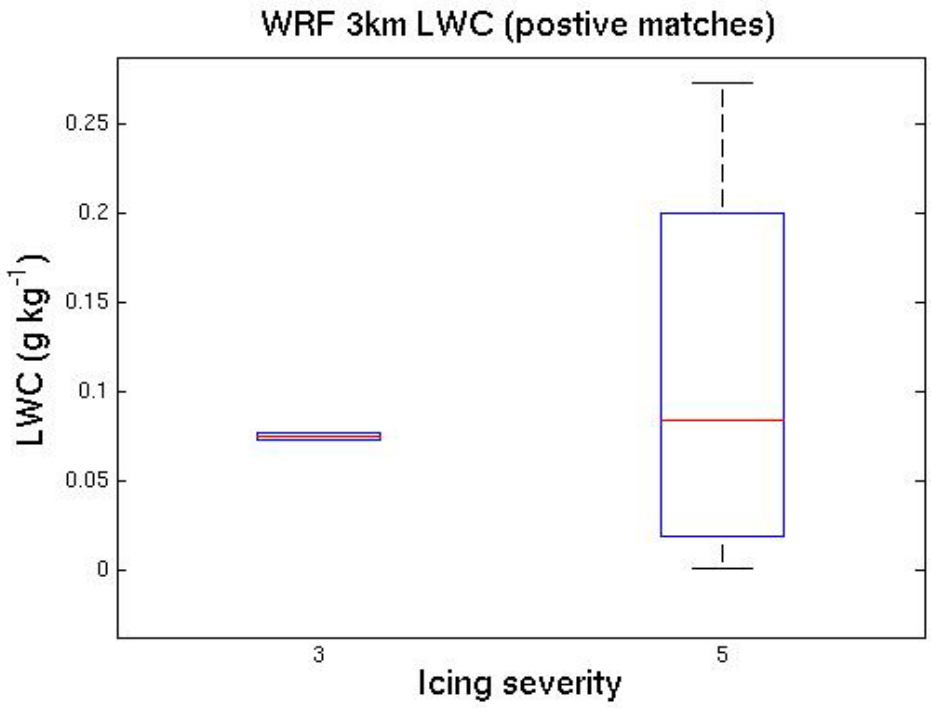


Figure 5.4, Summary of the mass of supercooled liquid at the locations were the model correctly identified icing conditions as a function of PIREP reported icing severity (1 trace icing to 8 severe icing).

5.2.3 Case 3 (Oct 9, 2009)

The results of the summary statistics analysis are presented in Table 5.3. Columns 2-6 present the WRF configurations from the 6-hour forecasts valid at 1800 UTC. Columns 7 – 10 present the 9-hr forecast valid at 2100 (the 13-km sensitivity run at 2100 UTC was not analyzed). Clouds forming in this fall case may exhibit hybrid characteristics containing both summer-like smaller spatial and temporal scales as well as winter-like larger scales. At 1800 UTC the 3-km run had icing PIREPs mapped to 265 grid points. Almost 75% of grid points were within a model predicted cloud (ice and/or SLW) but only 12.4% were correctly identified as containing SLW. The 13-km baseline, 13-km MYNN, 13-km YSU boundary layer schemes and 13-km sensitivity run correctly also identified the majority of icing grid points within a model cloud. They also only had between 6 – 11% of the observed icing within model predicted SLW. None of the WRF configurations did well with the cold PIREPs (the 13-km YSU did the best, 12.7%). The model-based relative humidities, which should be near 100% within icing conditions, were again too low but not as low as the Jun 4, 2008 case. By 2100 UTC all the model cloud (i.e., Pod SLW + Pod ICE) detections again improved significantly. The SLW POD predictions remained approximately the same. The 3-km and 13-km were similar in predicting SLW at the cold icing locations.

Table 5.3, Summary statistics for the Oct 9, 2008 case. Columns 2-6 are for 1800 UTC and 7-10 are for 2100 UTC.

	O9-3km	O9-13km	O9-13MYNN	O9-13YSU	O9-13sen2	O9-3km	O9-13km	O9-13MYNN	O9-13YSU
# grid points	265	159	158	158	159	277	137	138	139
Pod SLW	12.44	9.52	11.29	11.2	6.35	14.65	10.28	12.96	10.09
Pod ICE	51.67	57.14	52.42	56	63.49	65.95	67.29	65.74	69.72
# cold GPs	109	57	55	55	56	93	46	47	45
Pod Cold	8.26	7.02	10.91	12.72	1.79	10.75	6.52	4.26	4.44
POD RH>90	44.5	38.1	39.5	41.6	41.27	58.2	44.86	47.22	48.62
POD W>0	59.33	50.8	48.39	48.8	48.41	69.82	61.68	65.74	60.55
Pod mog SLW	18.75	26.67	26.67	26.67	33.33	16.67	21.74	13.04	21.74
Pod mog ICE	68.75	46.67	40	40	33.33	66.67	73.91	86.96	78.26
# cold mog GPs	11	3	3	3	3	26	15	15	14
Pod mog Cold	0	0	0	0	0	16.67	13.33	13.33	14.28
POD mog RH>90	62.5	40	46.67	46.67	53.33	58.33	43.48	43.47	47.83
POD mog W>0	62.5	46.67	53.33	53.33	46.67	81.25	56.52	65.22	69.57

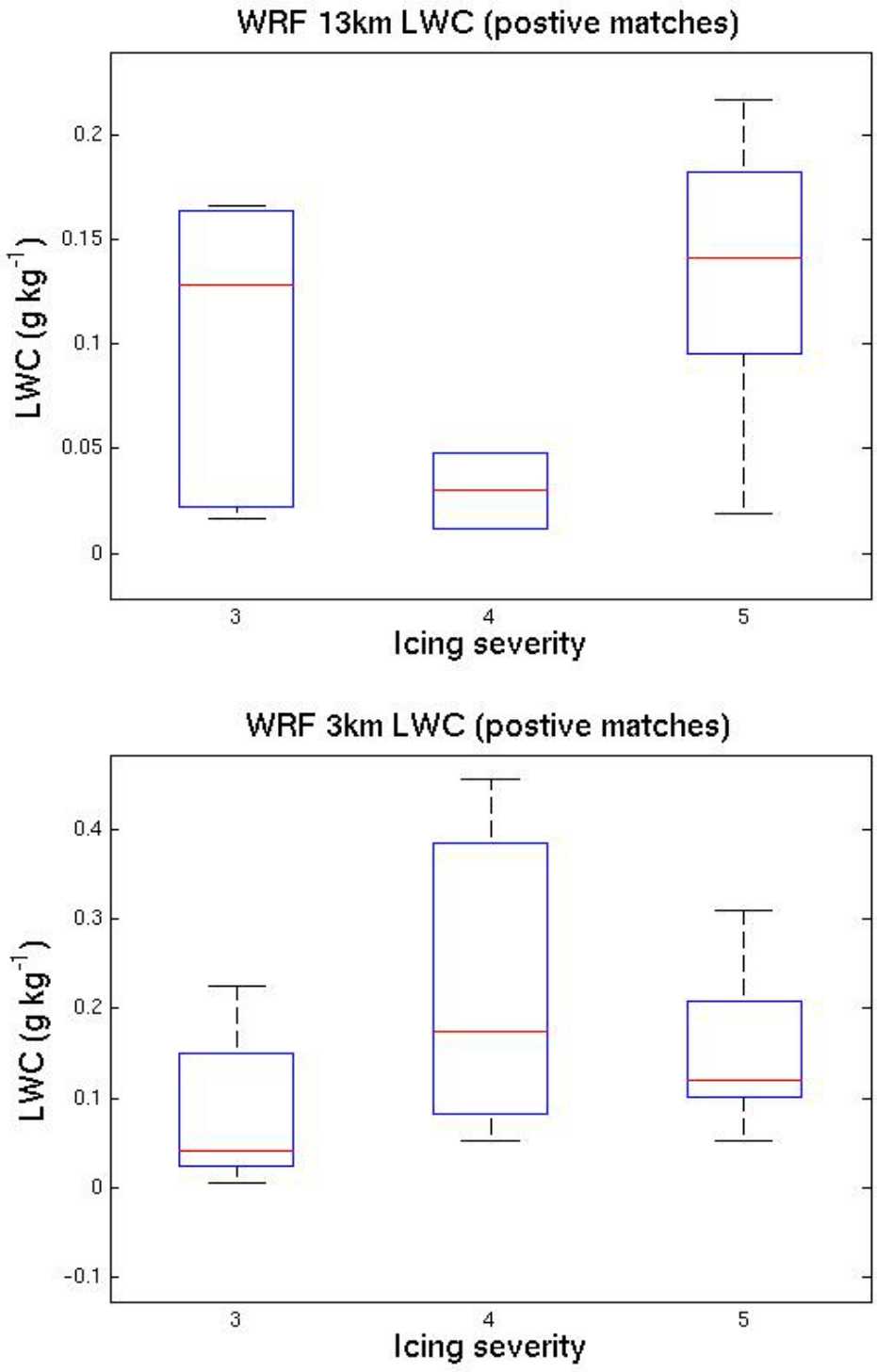


Figure 5.5, Summary of the mass of supercooled liquid at the locations were the model correctly identified icing conditions as a function of PIREP reported icing severity (1 trace icing to 8 severe icing). 13-km (top) and 3-km (lower panel).

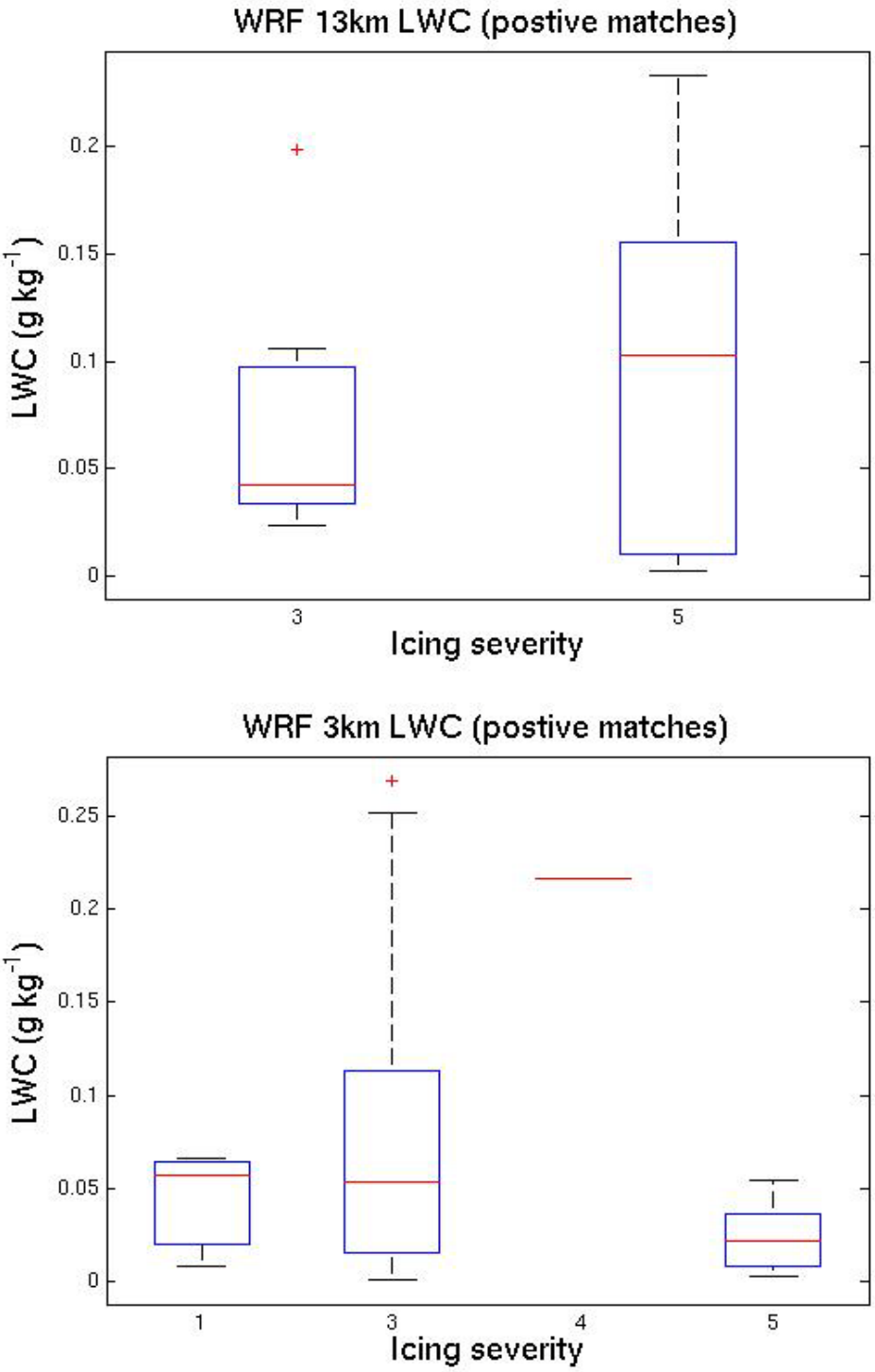


Figure 5.6, Summary of the mass of supercooled liquid at the locations were the model correctly identified icing conditions as a function of PIREP reported icing severity (1 trace icing to 8 severe icing). 13-km (top) and 3-km (lower panel).

The distribution of supercooled liquid mass, when it was correctly predicted, at 1800 UCT on 9 Oct 2009 is presented in figure 5.5. The 13-km SLW mass was slightly lower than the 3-km. The

median value for the 13-km WRF for moderate icing was 0.14 g kg^{-1}. The 3-km had more SLW in the light to moderate group (4) but the values were similar to the 13-km. By 2100 UTC the 13-km and 3-km SLW mass values have returned to smaller than what would be expected to produce moderate icing (Fig. 5.6).

5.2.4 Case 4 (Dec 9, 2009)

The results of the summary statistics analysis are presented in Table 5.4. Columns 2-6 present the WRF configurations from the 6-hour forecasts valid at 1800 UTC. Columns 7 – 10 present the 9-hr forecast valid at 2100 (the 13-km sensitivity run at 2100 UTC was not analyzed). Clouds forming in this winter case are expected to have a larger spatial and temporal scale than deep convective summer clouds. At 1800 UTC the 3-km run had icing PIREPs mapped to 157 grid points. Almost 80% of grid points (ice and/or SLW) were within a model predicted cloud but only 21.4% were correctly identified as containing SLW. The 13-km baseline, 13-km MYNN, 13-km YSU boundary layer schemes and 13-km sensitivity run correctly identified 87-93% of icing grid points within a model cloud. They also under-forecast SLW, with between 23 – 37% of the observed icing within model predicted SLW. Similar to the other cases, none of the WRF configurations did well with the cold PIREPs. The percent of model-based relative humidities greater than 90%(icing conditions should be near 100%) were better than the summer and fall cases, but again too low. By 2100 UTC all the model cloud detections were similar to the 1800 UTC runs.

The distribution of supercooled liquid mass, when it was correctly predicted, at 1800 UCT on 9 Dec 2009 is presented in figure 5.7. The 13-km and 3-km SLW mass was too low. The median value for the 13-km WRF for moderate icing was 0.07 g kg^{-1}, while the 3-km had 0.11 g kg^{-1}. By 2100 UTC both the 13-km and 3-km SLW mass values have returned to values that would be expected to produce light and moderate icing. The 3-km has a few outliers that exceed 0.5 g kg^{-1}. This may suggest the 3-km can produce more realistic cloud simulations when the cloud structure is well resolved.

Table 5.4, Summary statistics for the Dec 9, 2008 case. Columns 2-6 are for 1800 UTC and 7-10 are for 2100 UTC.

	D9-3km	D9-13km	D9-13MYNN	D9-13YSU	D9-13sen2	D9-3km	D9-13km	D9-13MYNN	D9-13YSU
# grid points	157	109	110	110	110	279	150	149	149
Pod SLW	21.48	37.5	28.39	23.46	35.8	27.27	25	29.5	26.43
Pod ICE	58.68	51.25	59.26	70.37	54.32	44.66	48.57	48.92	51.43
# cold GPs	38	20	19	20	20	86	69	71	74
Pod Cold	13.16	10	0	5	10	10.47	14.49	15.49	9.46
POD RH>90	52.89	62.5	62.96	59.26	65.43	64.03	57.86	60.43	59.29
POD W>0	52.89	77.5	79.01	74.07	75.31	54.15	64.29	74.82	66.43
Pod mog SLW	27.94	40.43	29.17	25	37.5	18.75	29.41	47.06	41.18
Pod mog ICE	52.94	51.06	60.41	72.92	56.25	37.5	23.53	17.65	17.65
# cold mog GPs	22	12	12	12	12	16	11	11	11
Pod mog Cold	9.09	40.42	0	0	8.33	0	0	18.18	9.09
POD mog RH>90	55.88	61.7	64.58	62.5	66.67	40.62	47.05	52.94	52.94
POD mog W>0	75	97.87	100	91.67	93.75	50	64.71	82.35	83.35

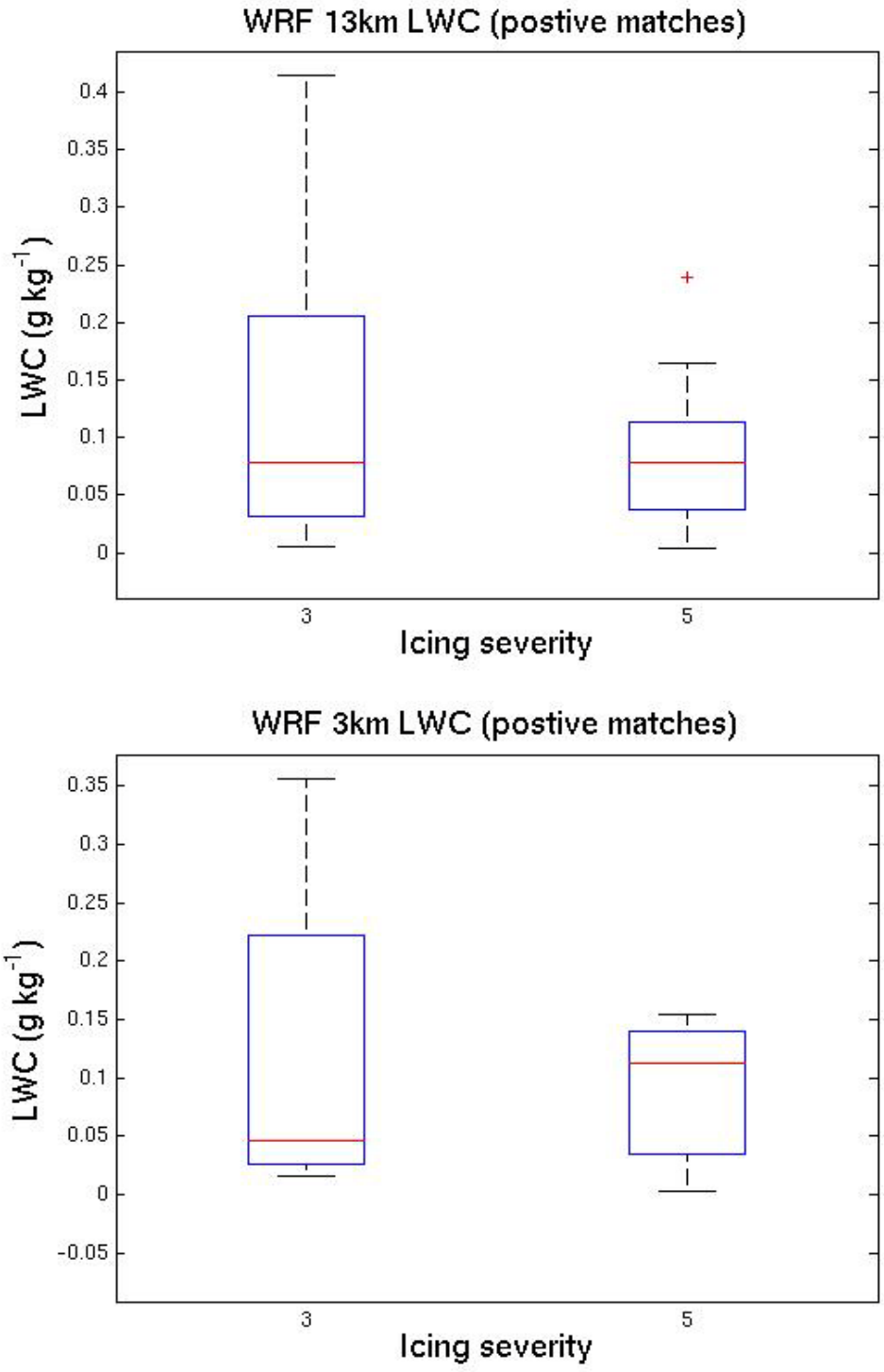


Figure 5.7, Summary of the mass of supercooled liquid at the locations were the model correctly identified icing conditions as a function of PIREP reported icing severity (1 trace icing to 8 severe icing). 13-km (top) and 3-km (lower panel).

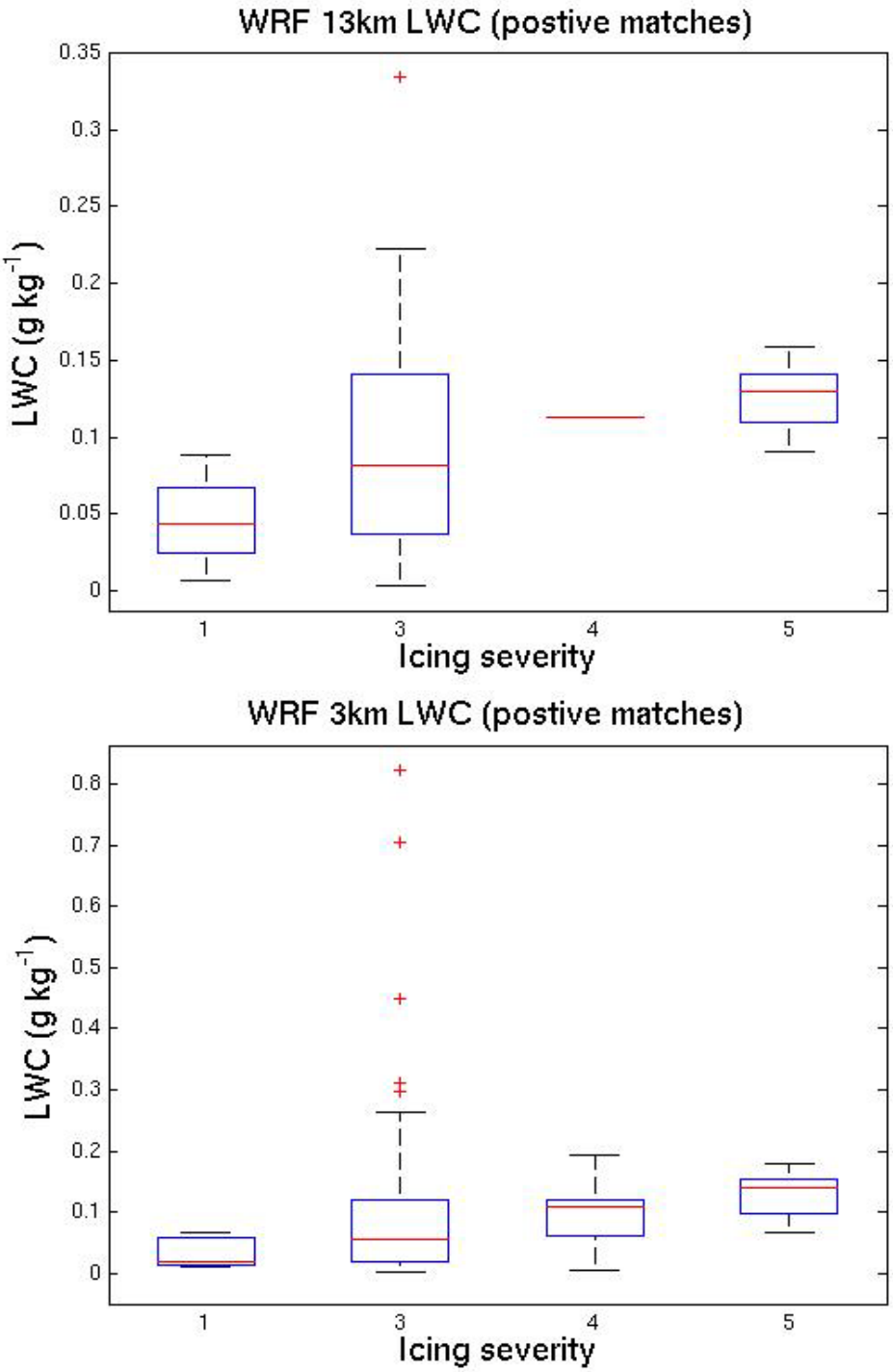


Figure 5.8, Summary of the mass of supercooled liquid at the locations were the model correctly identified icing conditions as a function of PIREP reported icing severity (1 trace icing to 8 severe icing). 13-km (top) and 3-km (lower panel).

5.3 MODEL RESOLUTION DIFFERENCES

Taking a broader look at the effects of model resolution differences, besides focusing simply on PIREP locations (as done in the discussion so far), we also extracted all model grid points that showed non-zero supercooled cloud water (SLW) within a 420-km by 600-km domain centered over central Illinois. This region had observed icing clouds on both February 6, 2008 and October 9, 2009. The cloud water, temperature, and vertical air motion data were extracted from the 3-km and 13-km models. Then only the subfreezing grid points that contain cloud water were tabulated, stratified into 0.1 g kg^{-1} bins, and the resulting distributions normalized by the total count of grid points belonging to each respective distribution.

Figures 5.9 and 5.10 show the normalized distributions of supercooled cloud water and vertical air motion for two valid times and two model resolutions. On February 6, 2008 the 3-km model runs exhibited substantially larger SLW mass than the 13-km data at 1800 UTC, but they were comparable at 2100 UTC. For both valid times analyzed on October 9, 2009 the higher resolution model exhibited a notably larger SLW mass. The larger mass of SLW in the 3-km model runs can be explained by a much wider distribution of vertical air motions with updrafts reaching 2 m s^{-1}.

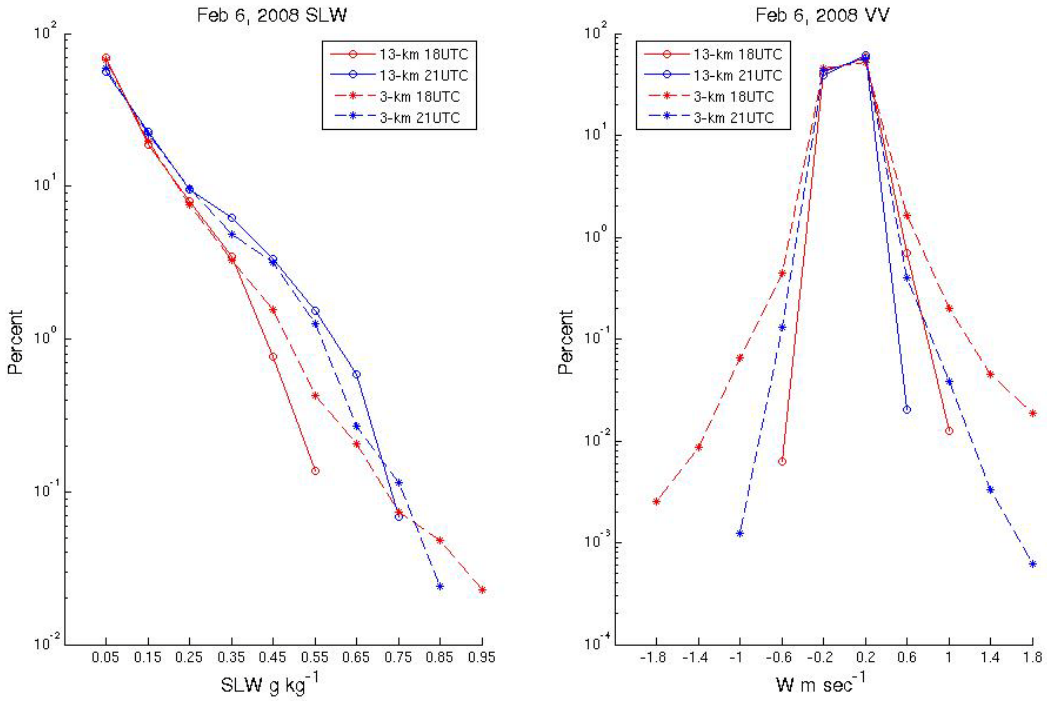


Figure 5.9, Normalized distributions of supercooled liquid water (SLW) mass are displayed in the left panel, while the right panel is for vertical air motions (W). Shown are the distributions for the 13 km (solid lines) and 3 km (dashed lines) model runs valid at 1800 UTC (6 hour forecast) and 2100 UTC (9 hour forecast), respectively, on February 6, 2008.

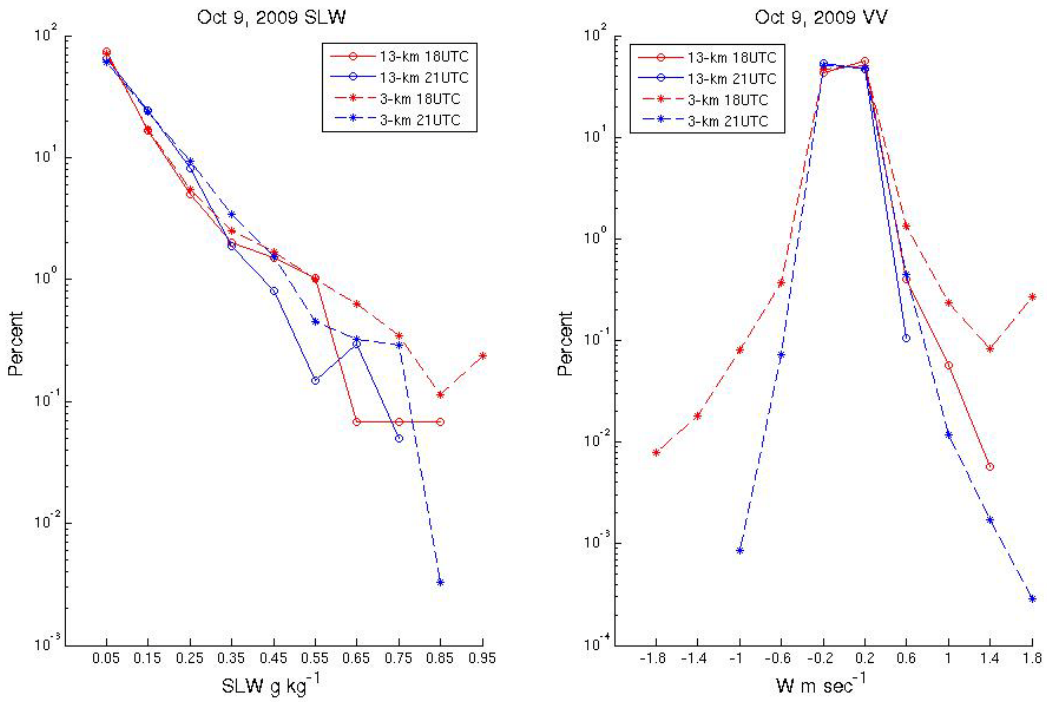


Figure 5.10, Normalized distributions of supercooled liquid water (SLW) mass are displayed in the left panel, while the right panel is for vertical air motions (W). Shown are the distributions for the 13 km (solid lines) and 3 km (dashed lines) model runs valid at 1800 UTC (6 hour forecast) and 2100 UTC (9 hour forecast), respectively, on October 9, 2009.

5.4 SUMMARY

Microphysical output from several configurations and resolutions of the WRF model simulations were reviewed. All of the model configurations did a reasonably good job identifying clouds, but struggled to differentiate ice phase from supercooled liquid phase, based on the evaluation against PIREPs. The mass of SLW was usually too low in all the model configurations. Moreover, the PIREP location errors may be more problematic in the 3-km WRF comparisons and might require alternative verification data, such as those obtained by research aircraft. Comparison of the distributions of SLW and vertical air motions over a large domain (central Illinois) revealed the expected outcome of somewhat larger SLW mass and wider vertical air motion distributions for the higher resolution model runs. However, these are preliminary studies that need to be substantiated using more extensive data sets and analyses.

6. SUMMARY OF FINDINGS AND OUTLOOK

Forecast-sensitivity experiments focused on the following model characteristics: resolution (horizontal grid spacing ranging from 13 to 1 km), PBL parameterization, horizontal diffusion (scheme and coefficient), and storm-scale initialization through radar-reflectivity data assimilation.

For Convective cases, as shown previously, the differences between forecasts employing 13-km (RR) and 3-km (HRRR) grid spacing were significant, largely due to the ability of the 3-km model to produce convective storms explicitly. The analyses discussed in this report are encouraging for the current HRRR configuration. For convective forecasting, 3-km horizontal grid spacing offers the following significant benefits over 13-km grid spacing:

- Convective storm / system structures are more realistic.
- Storm location and timing forecasts are more accurate.
- Some of the "false alarms" (overprediction) in the 13-km forecasts are eliminated by the higher resolution.

Focusing on turbulence, the synoptic patterns produced are very similar for the three model runs analyzed and therefore the overall patterns of predicted turbulence are more or less the same. The WRF 3km runs seem to provide more fine scale detail horizontally, although it is not clear this finer scale detail is consistent with observations. Vertically, the fine scale detail provided by both the 13km and 3km WRF is apparently not as well resolved as it is in the operational 13km RUC runs, indicating that more vertical levels in the WRF runs might provide more realistic results. Future work needs to better assess the quantitative performance of the 13km and 3km WRF mode. Sensitivity studies, some of which have already been initiated, need to be performed with regards to the model top, vertical resolution, and upper boundary condition used. Statistical evaluation methods that are more object-oriented would also reduce some of the uncertainty in the results.

In the case of in-flight Icing, microphysical output from several configurations and resolutions of the WRF model simulations were reviewed. All of the model configurations did a reasonably good job identifying clouds, but struggled to differentiate ice phase from supercooled liquid phase, based on the evaluation against PIREPs. The mass of SLW was usually too low in all the model configurations. Moreover, the PIREP location errors may be more problematic in the 3-km WRF comparisons and might require alternative verification data, such as those obtained by research aircraft. Comparison of the distributions of SLW and vertical air motions over a large domain (central Illinois) revealed the expected outcome of somewhat larger SLW mass and wider vertical air motion distributions for the higher resolution model runs.

These analyses focused on sensitivity impacts of numerical weather prediction model configurations on a variety of aviation hazards, and need to be extended. The analyses presented in this report are too limited to provide conclusive evidence and thus clear guidance for improving the 3km HRRR and 13 km WRF-RR model configurations

APPENDIX A.1 CONVECTION

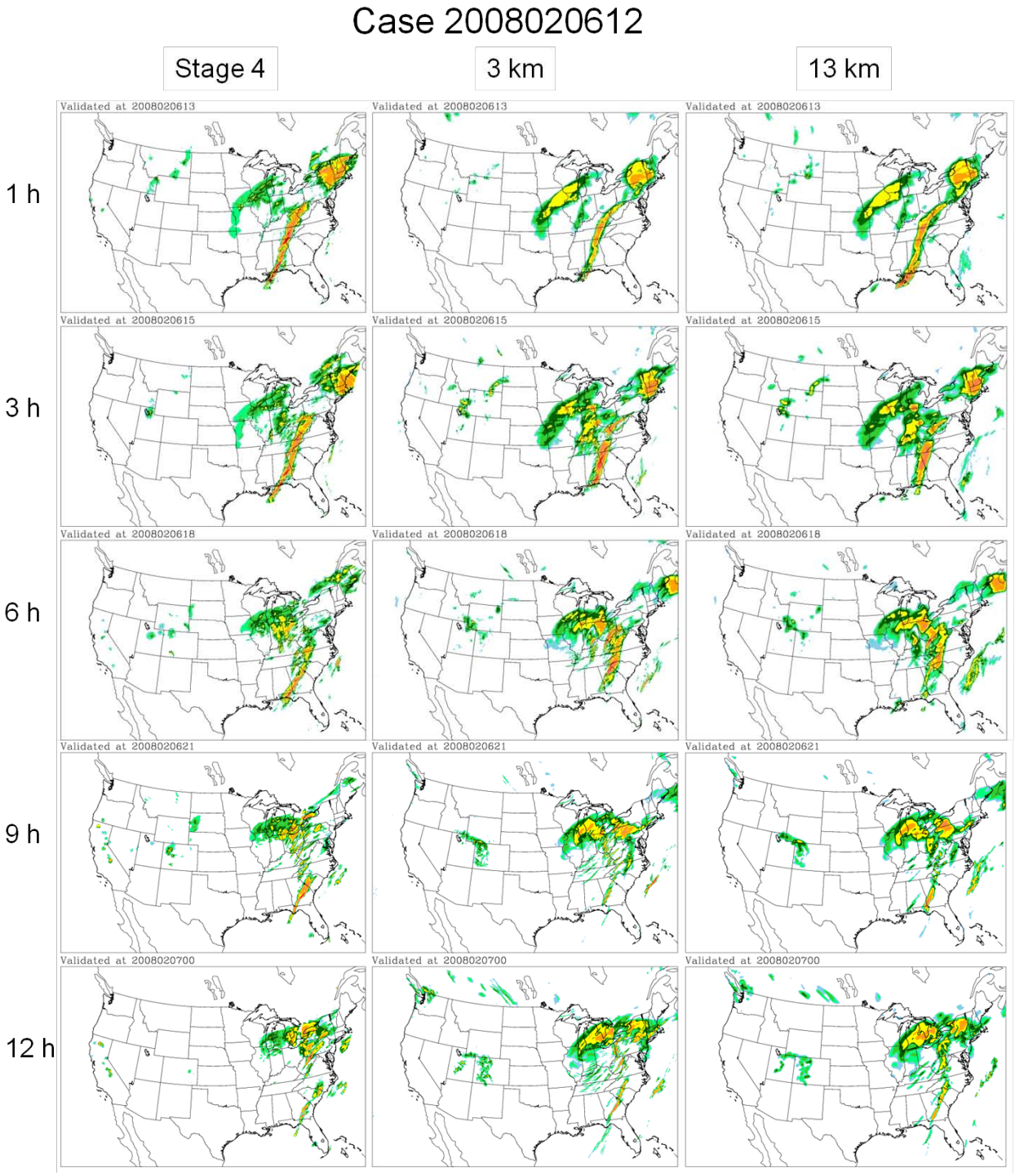


Fig. A1.1, One-hour rainfall rate from Stage IV and model forecasts on February, 6, 2008. The model was initialized at 12 Z and validation time shown are 13, 15, 18, 21 Z of 20080206, and 00Z of 20080207.

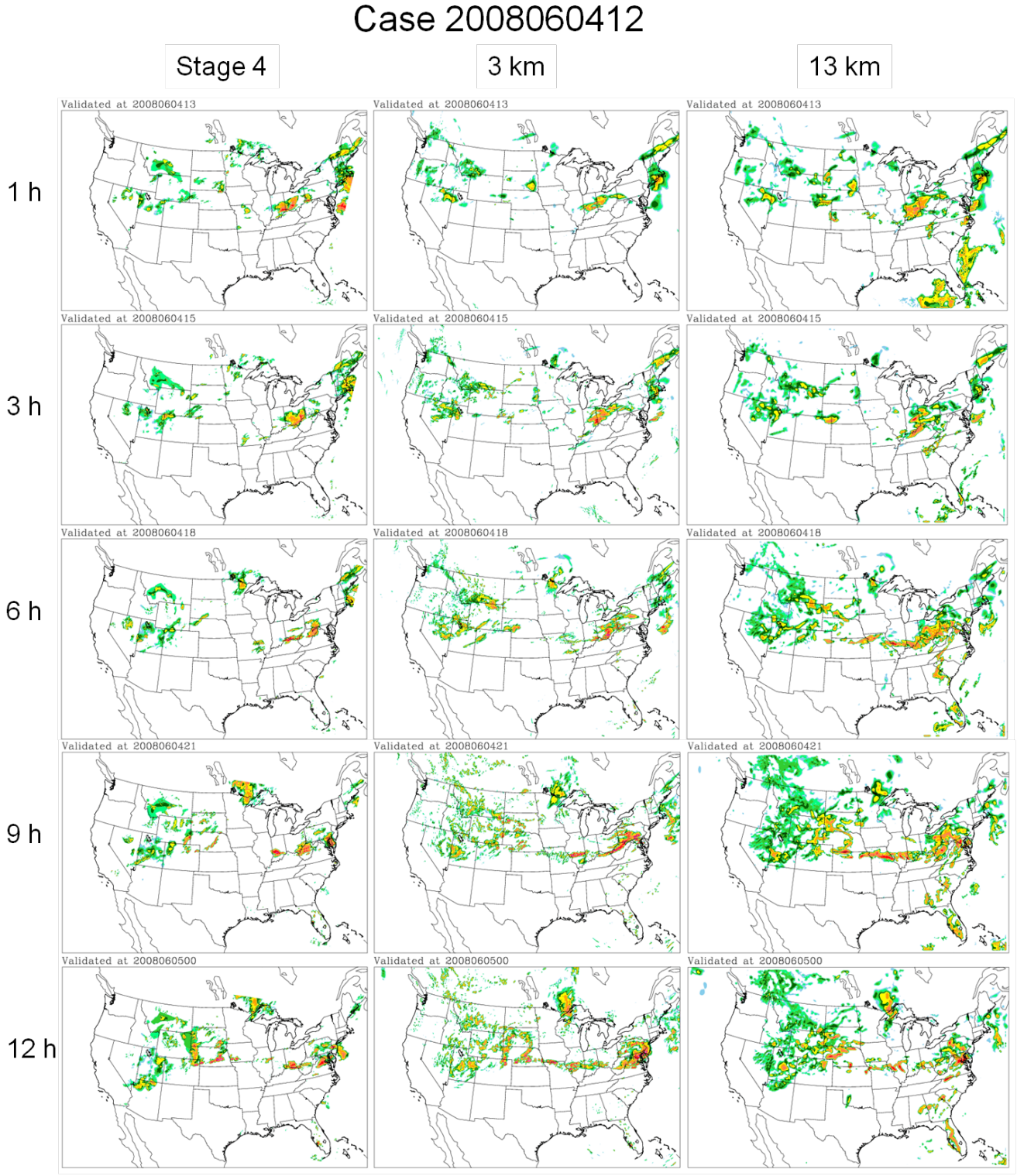


Fig. A1.2. One-hour rainfall rate from Stage IV and model forecasts on June, 4, 2008. The model was initialized at 12 Z and validation time shown are 13, 15, 18, 21 Z of 20080604, and 00Z of 20080605.

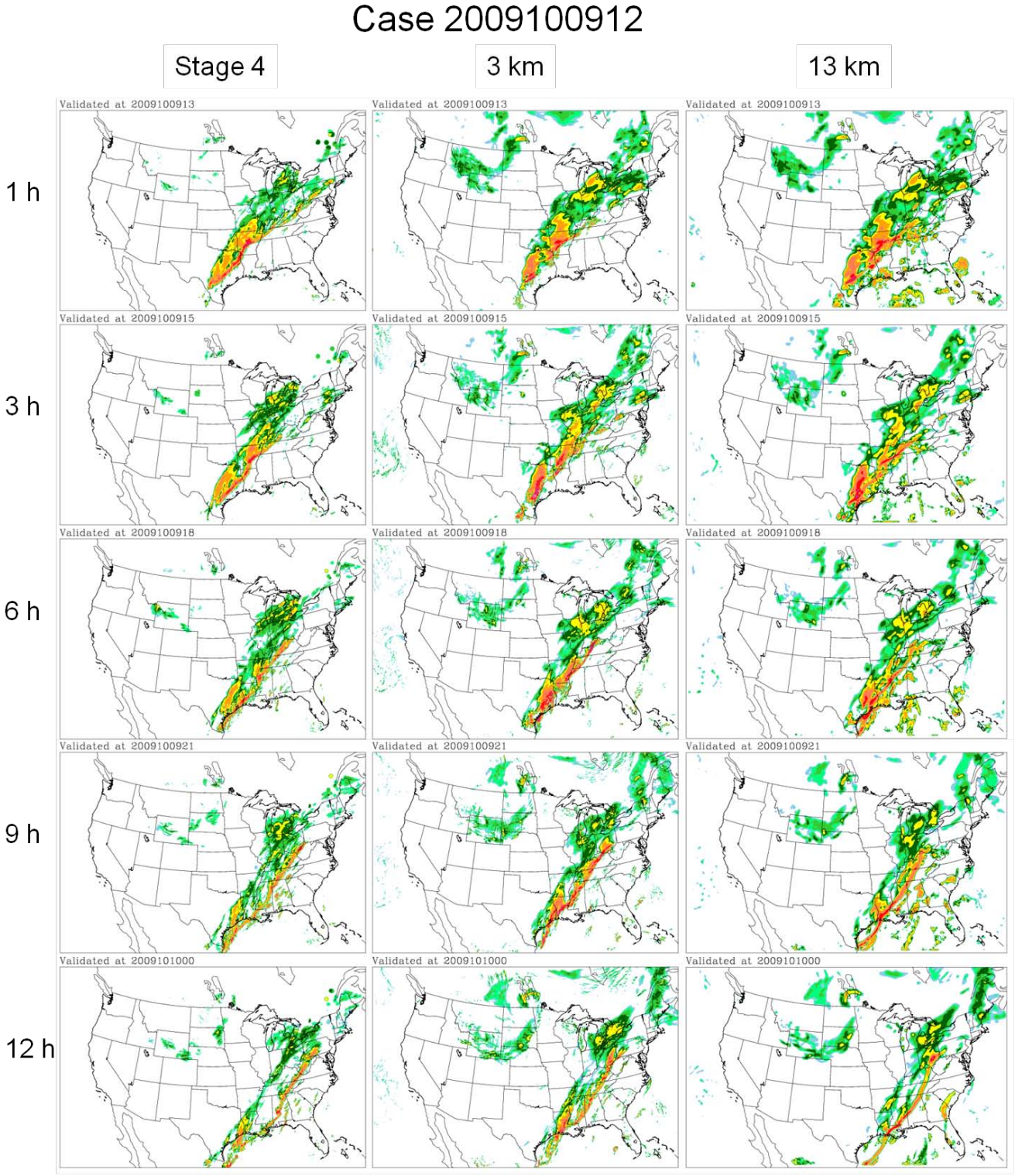


Fig. A1.3. One-hour rainfall rate from Stage IV and model forecasts on October, 9, 2009. The model was initialized at 12 Z and validation time shown are 13, 15, 18, 21 Z of 20091009, and 00Z of 20091010.

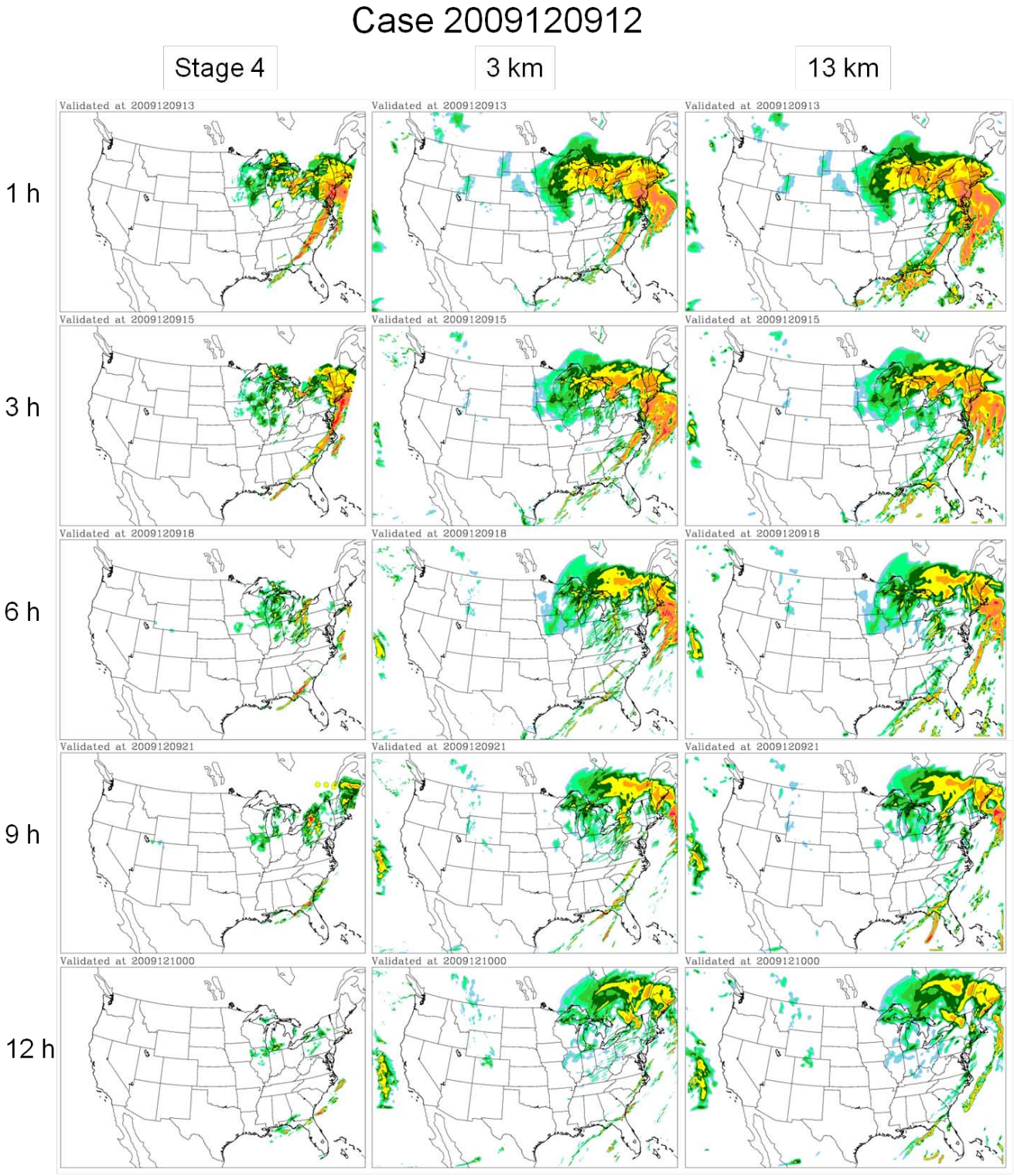


Fig. A1.4. One-hour rainfall rate from Stage IV and model forecasts on December, 9, 2009. The model was initialized at 12 Z and validation time shown are 13, 15, 18, 21 Z of 20091209, and 00Z of 20091210.

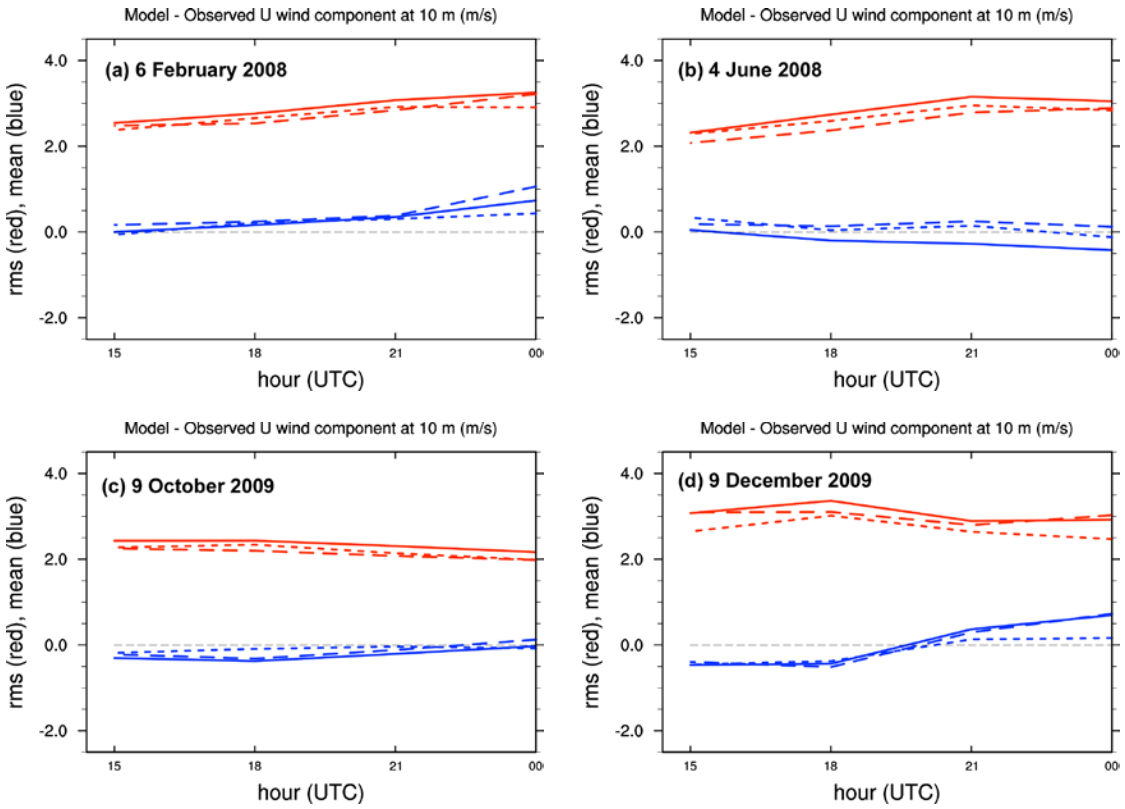


Figure A1.5. Root-mean-square (red) and mean (blue) *u* wind component differences (m s^{-1}) between 13-km forecasts and surface observations. Verification statistics are shown for forecasts utilizing three different PBL parameterizations: MYJ (solid lines), MYNN (short dashes), and YSU (long dashes).

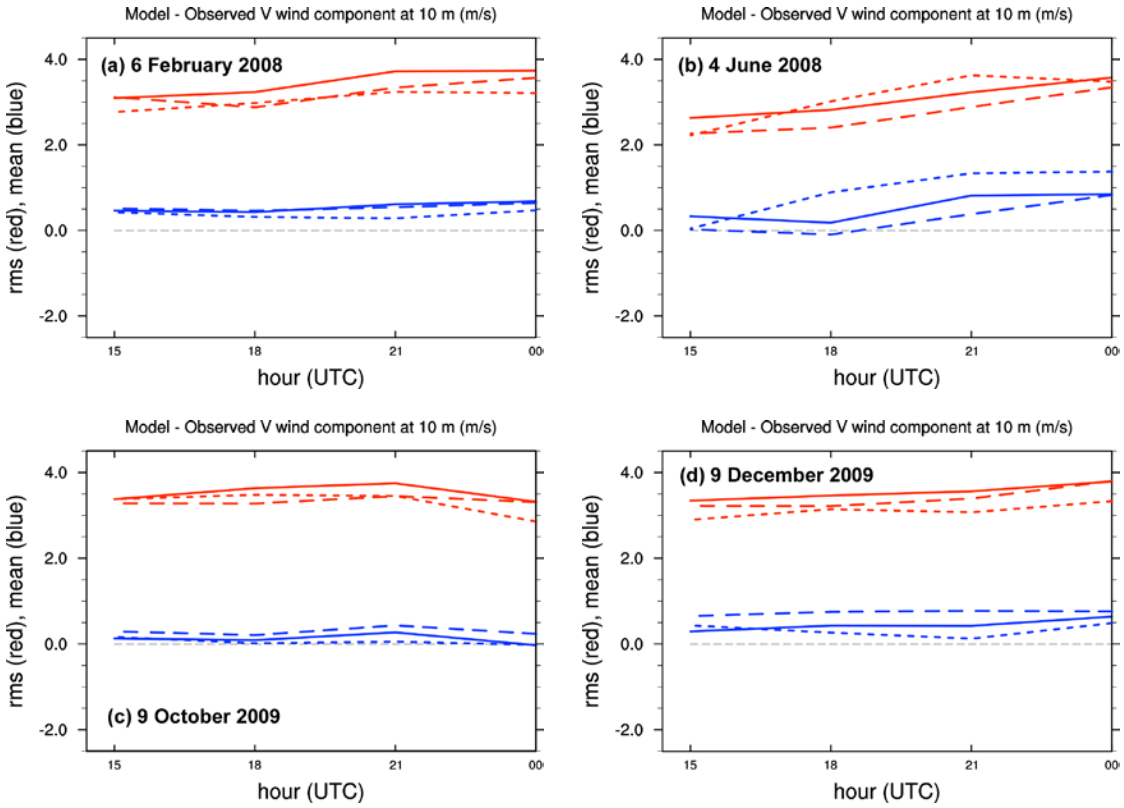


Figure A1.6. Root-mean-square (red) and mean (blue) *v* wind component differences (m s^{-1}) between 13-km forecasts and surface observations. Verification statistics are shown for forecasts utilizing three different PBL parameterizations: MYJ (solid lines), MYNN (short dashes), and YSU (long dashes).

APPENDIX A.2 IN-FLIGHT ICING

A.2.1 Feb. 6, 2008 21UTC – 9 hour forecast

PIREP 1. Negative icing was reported at ~10 km near Baltimore, MD. Both the 13-km and 3-km models correctly indicated that no supercooled liquid water was present. Both models had a very similar temperature and dew point temperature profile. The 3-km and 13-km WRF runs had similar condensate species forecast but the 3-km produced more than double the mass of ice condensate as the 13-km. The vertical velocity structure was significantly different at the base of the column with the 3-km having stronger lift at the level of maximum condensate (Fig. 1).

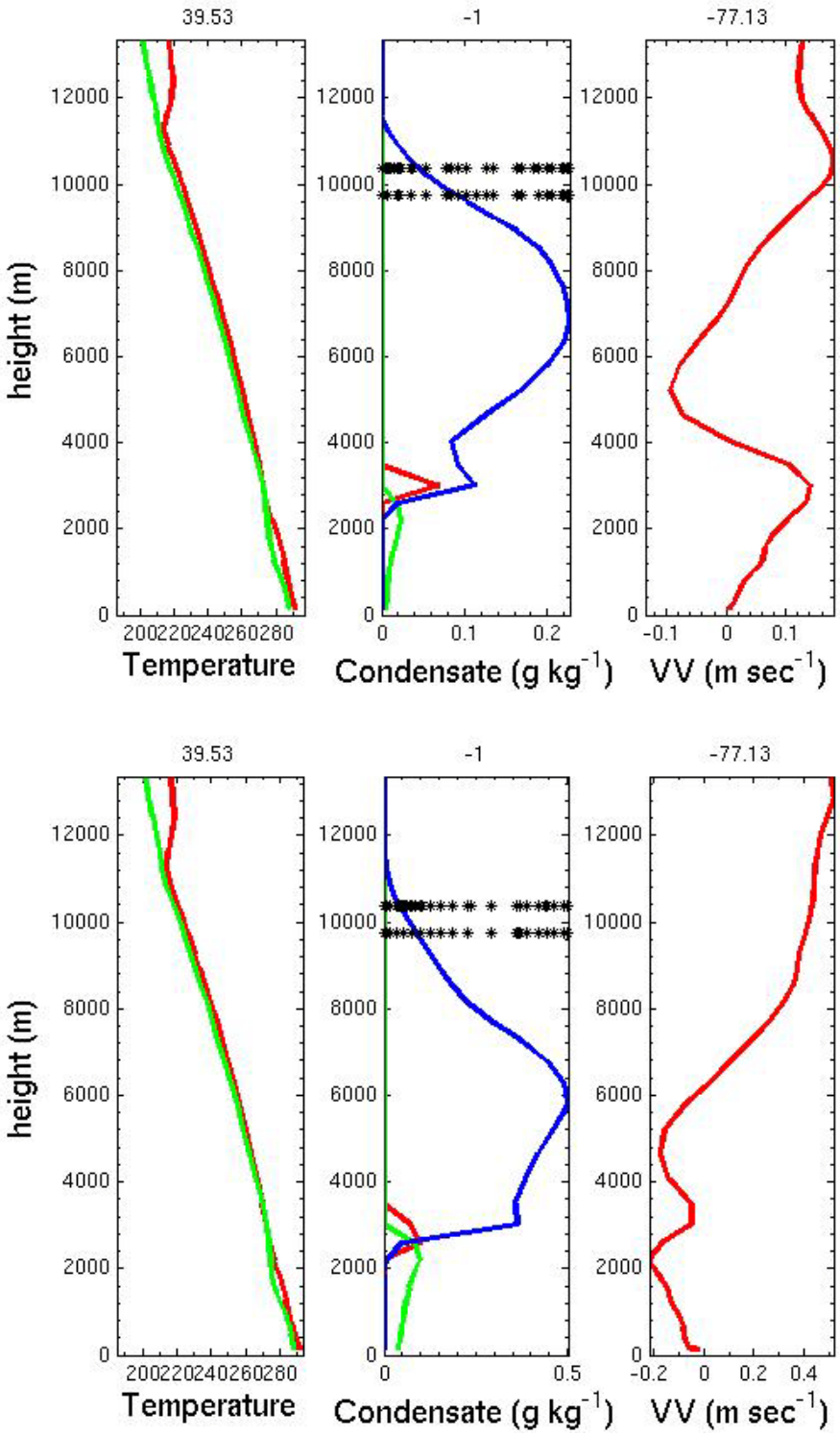


Figure 1. 13-km WRF run (top), 3-km WRF run bottom. Left panel; (T (red),Td (green)), middle panel; (cloud water (red), rain water (green), frozen condensate (blue)) black stars denote PIREP icing top and base, right panel; vertical velocity (red). Latitude, icing severity, and longitude shown along top of panels 1,2, and 3.

PIREP 2 (not shown). Light icing was reported at 2-km near St. Louis, MO. The temperature profiles from both the 13-km and 3-km were very similar. The condensate plots from both the 3 and 13-km model runs had no supercooled liquid predicted and a very low mass of ice condensate.

PIREP 3 (not shown). Light to moderate icing was reported over Lexington, KY between 2.5-km and 3.3-km. Both the 3-km and 13-km WRF runs were sub-saturated and neither had any condensate. The 3-km had weak ascent in the vertical velocity field while the 13-km vertical velocity was neutral.

PIREP 4 (not shown). Moderate clear icing was reported over southwestern GA between 5.5-km and 6.0-km. Both the 13-km and 3-km model runs were sub-saturated and neither correctly identified the icing conditions.

PIREP 5. Moderate mixed icing was reported over southern MI between 0.8-km and 1.5-km. Both models have a deep cloud at the icing location. The 3-km has up to 0.4 g kg^{-1} of ice condensate just above the icing level and about half as much supercooled liquid condensate. The 13-km has about half the ice condensate as the 3-km WRF and about the same amount of supercooled liquid water. Perhaps the increased ice condensate has removed too much liquid. At the reported icing altitudes the 3-km WRF has no supercooled liquid while the 13-km has almost 0.2 g m^{-3} (Fig 2).

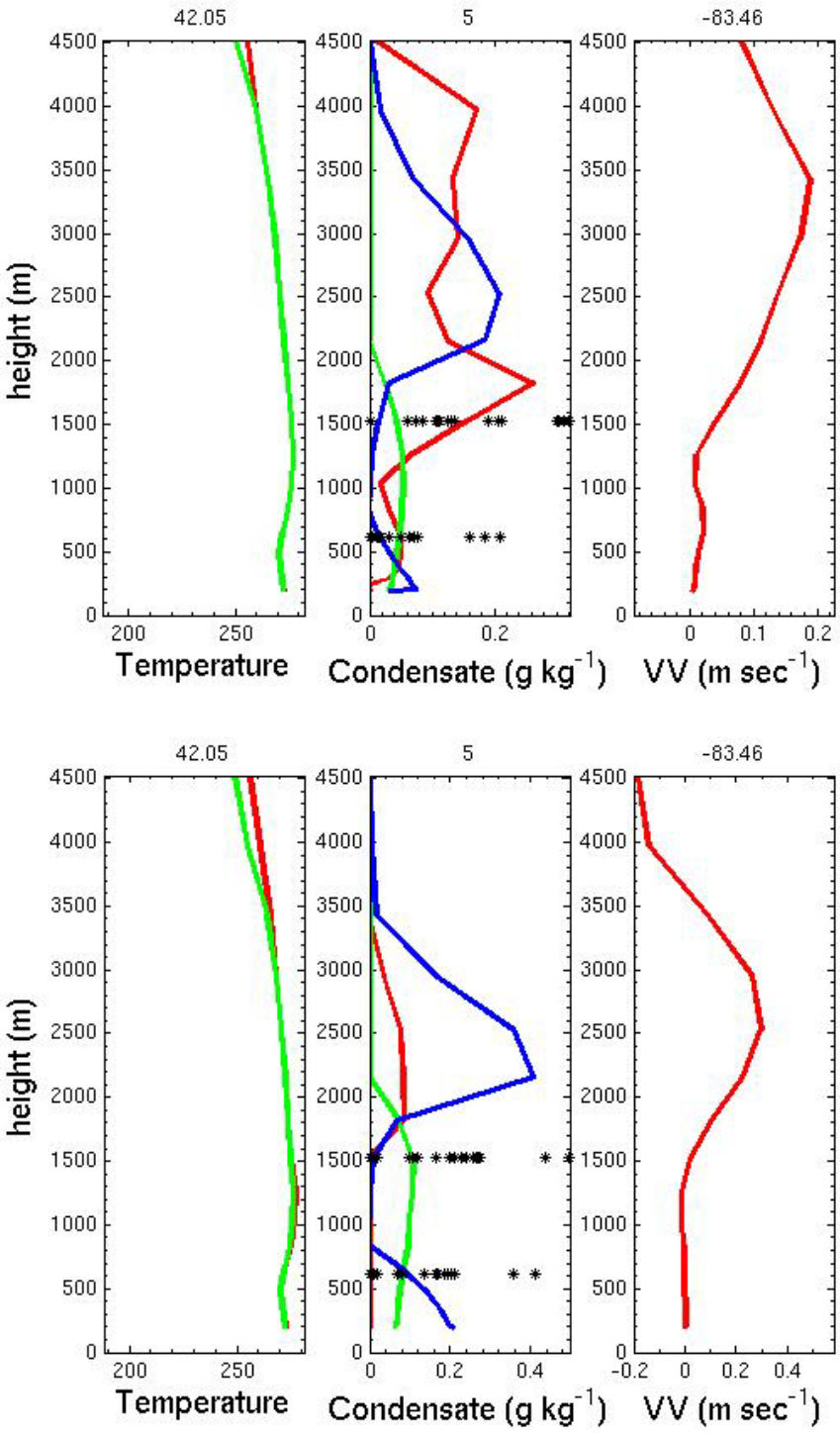


Figure 2. 13-km WRF run (top), 3-km WRF run bottom. Left panel; (T (red),Td (green)), middle panel; (cloud water (red), rain water (green), frozen condensate (blue)) black stars denote PIREP icing top and base, right panel; vertical velocity (red). Latitude, icing severity, and longitude shown along top of panels 1,2, and 3.

PIREP 6. Light rime icing was reported along the north shore of Lake Erie between 2.5 and 4-km. Neither model correctly identified the icing layer. The 13-km had up to 0.02 g kg^{-1} of ice condensate while the 3-km model had less that 0.01 g kg^{-1} of ice condensate. The 13-km model had neutral ascent through the icing layer whereas the 3-km WRF had weak ascending vertical velocity (Fig 3).

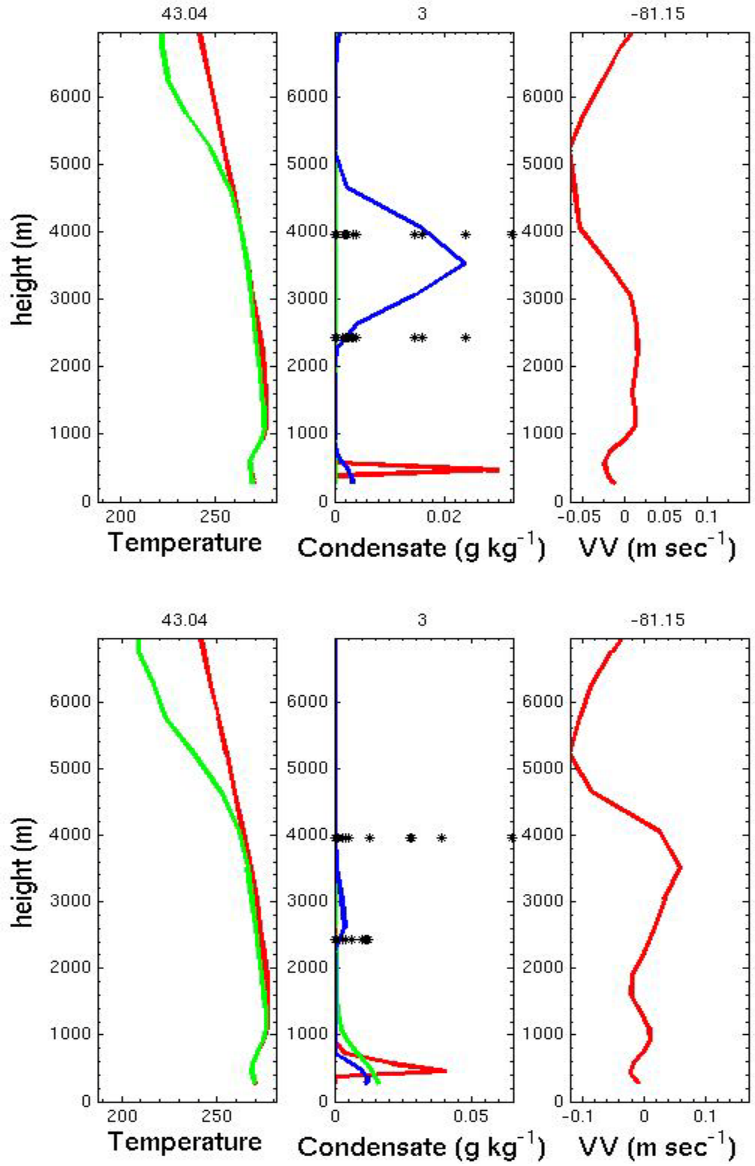


Figure 3. 13-km WRF run (top), 3-km WRF run bottom. Left panel; (T (red),Td (green)), middle panel; (cloud water (red), rain water (green), frozen condensate (blue)) black stars denote PIREP icing top and base, right panel; vertical velocity (red). Latitude, icing severity, and longitude shown along top of panels 1,2, and 3.

PIREP 7 (not shown). Numerous aircraft flying near Rochester, NY reported light icing between 0.6 and 1.8-km. Both the 3-km and 13-km model runs had a two-layer cloud. The upper cloud had more than 1.0 g kg^{-1} of ice condensate. The lower cloud layer contained the icing conditions. Both model runs correctly identified the icing cloud and had similar amounts of supercoled liquid (~0.3 g kg^{-1}).

PIREP 8. Moderate mixed icing was reported over Toronto, Canada between 4.5 and 5.5-km. Both models have a deep cloud layer predicted with notable differences. The 13-km has over 1.0 g kg^{-1} of ice phase at the icing altitudes and no supercooled liquid. The 3-km has significantly less ice condensate and a layer of supercooled liquid (0.1 g kg^{-1}). The vertical velocities are very strong supporting with the large condensate mass (Fig 4).

PIREP 9 (not shown). Moderate mixed icing was reported near Fort Wayne, IN between 1.5 and 2.5-km. Both model runs have boundary layer supercooled liquid cloud predicted with condensate mass of ~ 0.3 g kg^{-1}. The top of this SLW cloud just captures the base of the icing layer. The 13-km WRF has a small amount of ice phase though the rest of the icing layer whereas the 3-km is dryer and sub-saturated at this level.

PIREP 10. Light rime icing was reported between 1 and 2.5 km over Portland, OR. Neither model had any condensate predicted at this location and at these altitudes. The 3-km WRF did have fairly strong descent (0.4 m sec^{-1}) as compared to the 13-km WRF (0.04 m sec^{-1}) (Fig 5).

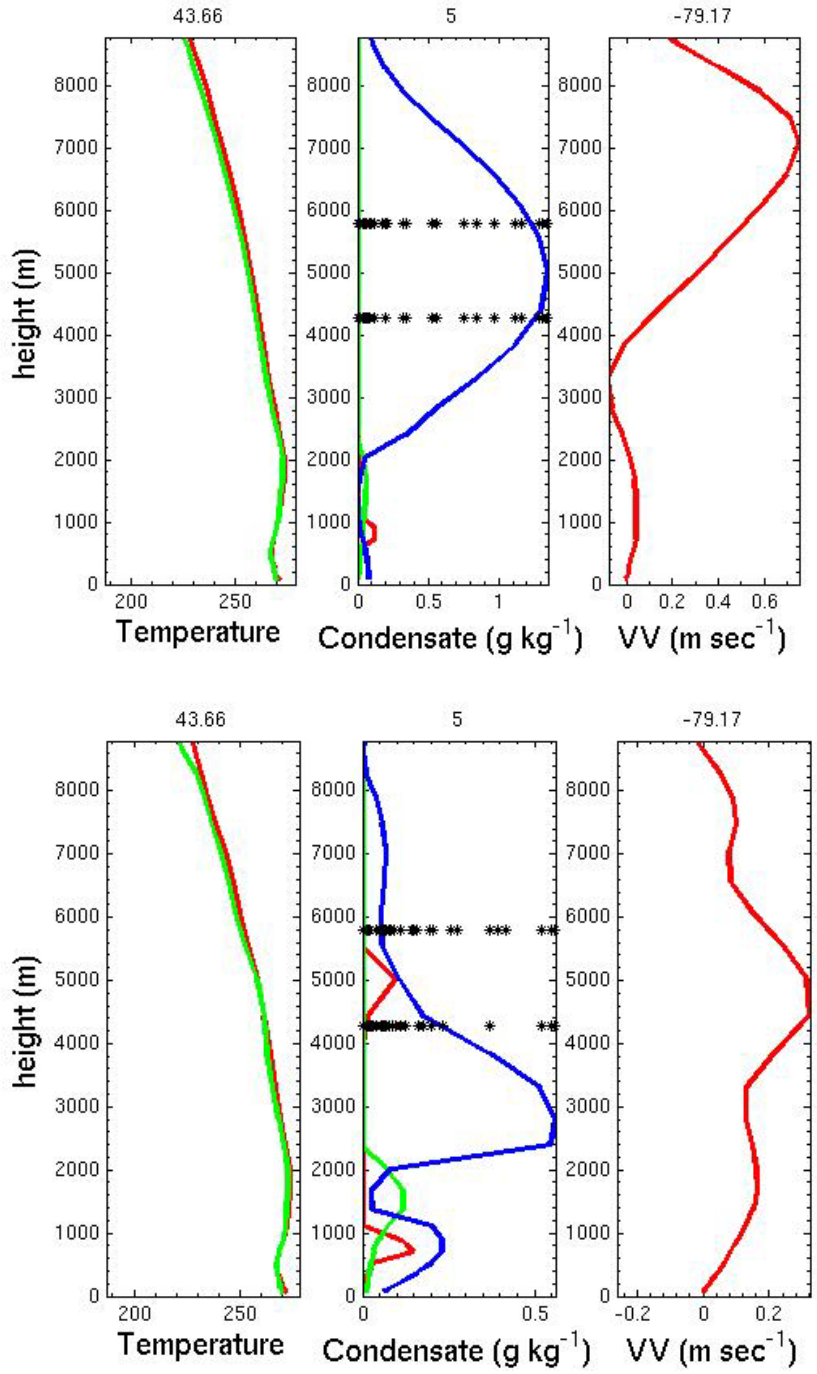


Figure 4. 13-km WRF run (top), 3-km WRF run bottom. Left panel; (T (red),Td (green)), middle panel; (cloud water (red), rain water (green), frozen condensate (blue)) black stars denote PIREP icing top and base, right panel; vertical velocity (red). Latitude, icing severity, and longitude shown along top of panels 1,2, and 3.

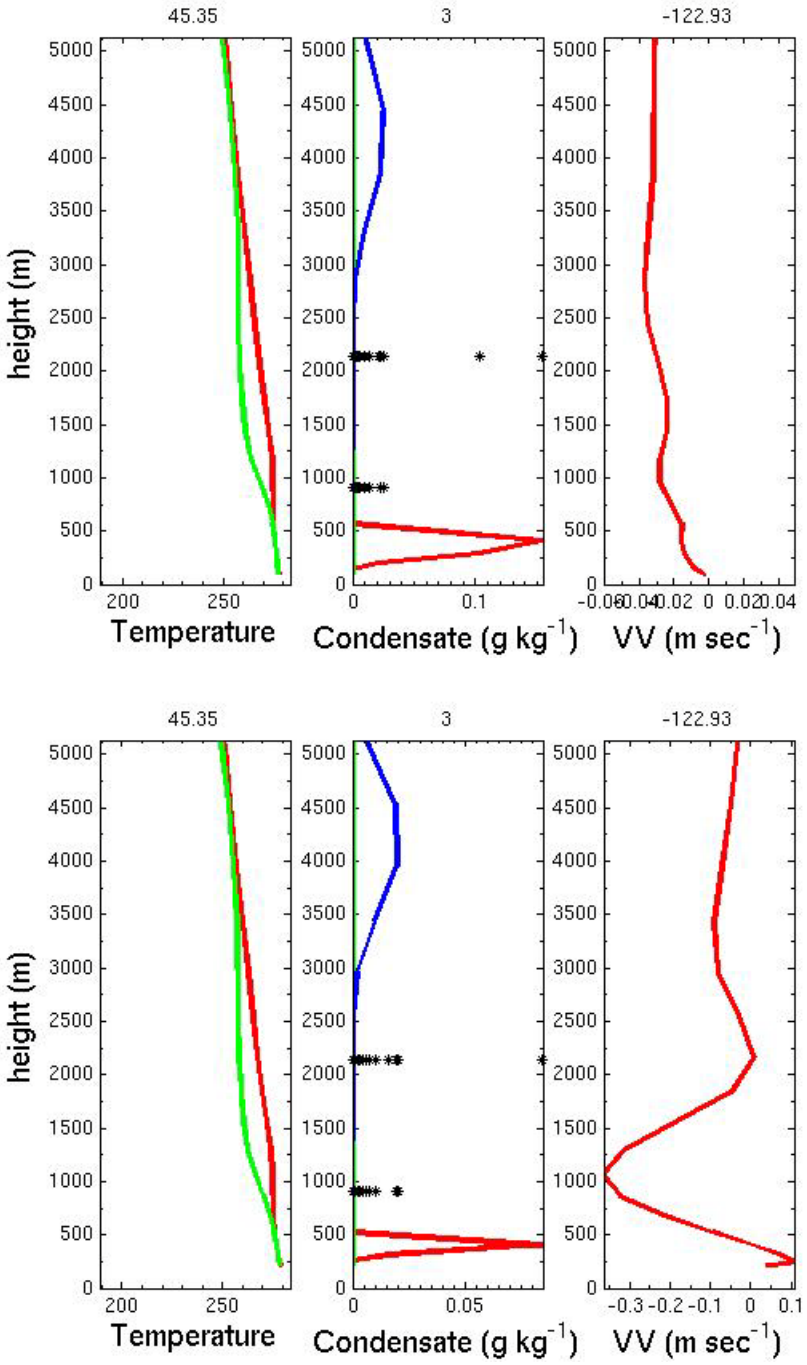


Figure 5. 13-km WRF run (top), 3-km WRF run bottom. Left panel; (T (red),Td (green)), middle panel; (cloud water (red), rain water (green), frozen condensate (blue)) black stars denote PIREP icing top and base, right panel; vertical velocity (red). Latitude, icing severity, and longitude shown along top of panels 1,2, and 3.

PIREP 11. Light rime icing was reported over western Nova Scotia, Canada between 3 and 3.5-km. Both models identified a low-condensate mass cloud layer at the icing altitudes. The 13-km WRF predicted all frozen condensate, while the 3-km correctly predicted 0.05 g kg^{-1} of supercooled liquid (Fig 6).

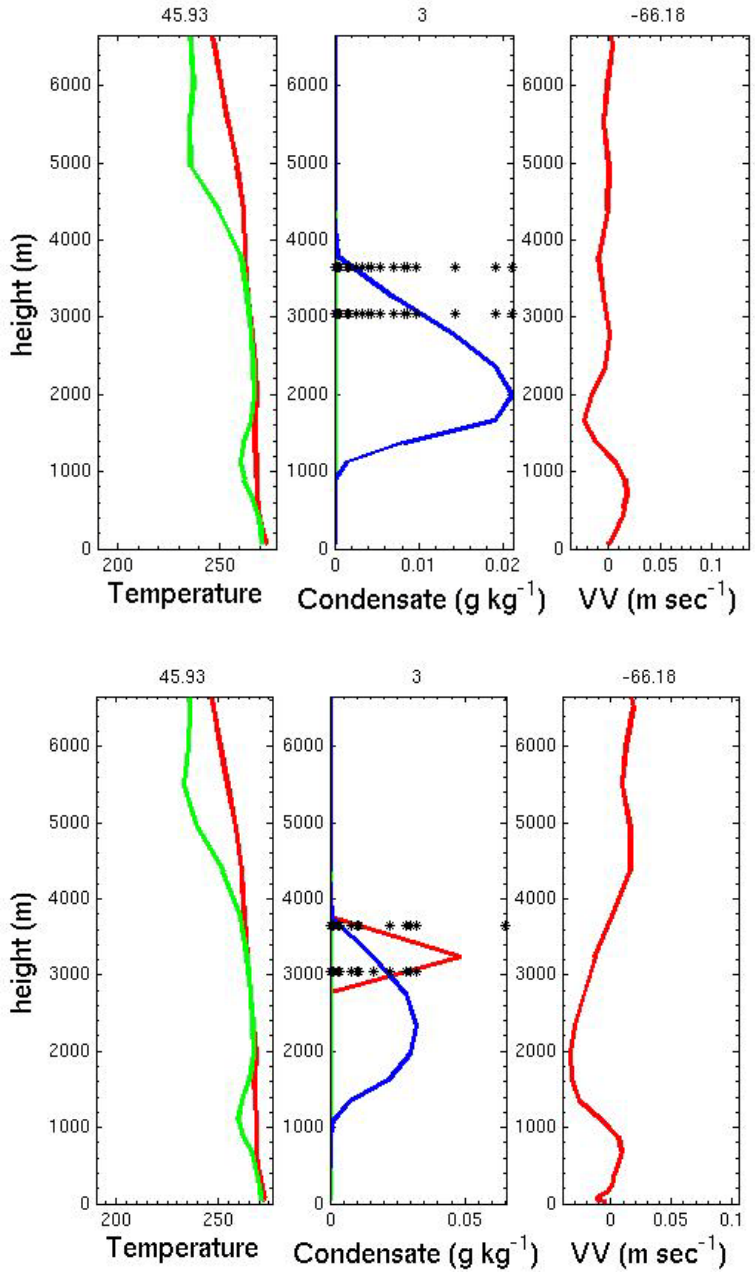


Figure 6. 13-km WRF run (top), 3-km WRF run bottom. Left panel; (T (red),Td (green)), middle panel; (cloud water (red), rain water (green), frozen condensate (blue)) black stars denote PIREP icing top and base, right panel; vertical velocity (red). Latitude, icing severity, and longitude shown along top of panels 1,2, and 3.

PIREP 12 (not shown). Moderate rime icing was reported below 2-km over St. Louis, MO. Both model runs did an excellent job predicting this icing cloud. The 3-km had over 0.2 g kg^{-1} of supercooled liquid and the 13-km slightly less than that amount.

PIREP 13. Light icing was reported over southern IL between 1 and 1.5-km at a warm temperature of -2C. Both models were quite similar and both correctly identified the icing cloud with up to 0.3 g kg^{-1} of supercooled liquid water (Fig 7).

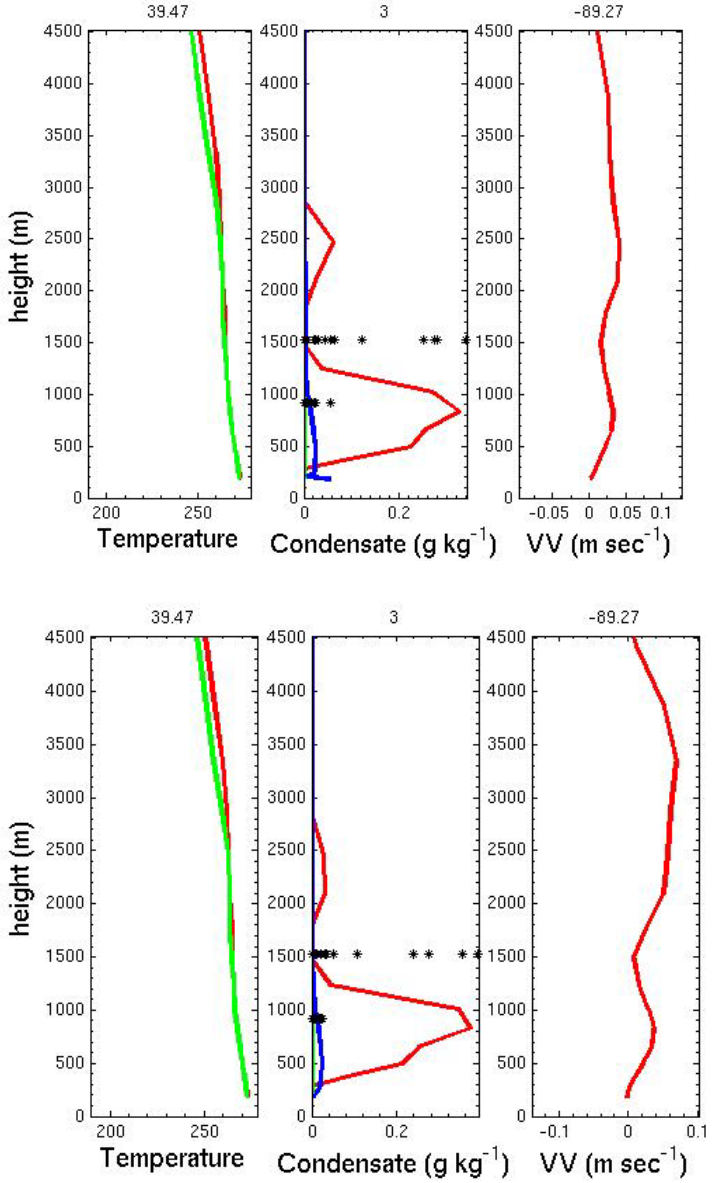


Figure 7. 13-km WRF run (top), 3-km WRF run bottom. Left panel; (T (red),Td (green)), middle panel; (cloud water (red), rain water (green), frozen condensate (blue)) black stars denote PIREP icing top and base, right panel; vertical velocity (red). Latitude, icing severity, and longitude shown along top of panels 1,2, and 3.

PIREP 14 (not shown). Light rime icing was reported below 2.5-km over eastern IN. Both model runs were very similar and had supercooled liquid forecast below 1.2-km. This is well below the reported icing top. The 13-km run had a very large 0.6 g kg^{-1} while the 3-km WRF had a more modest and more likely correct 0.2 g kg^{-1}.

PIREP 15. A trace of icing was reported near Decatur, IL between 0.7 and 1.5-km. Both model runs had a very similar cloud structure. Supercooled liquid was correctly forecast at the icing level. The mass was ~ 0.1 g kg^{-1} in both models (Fig 8).

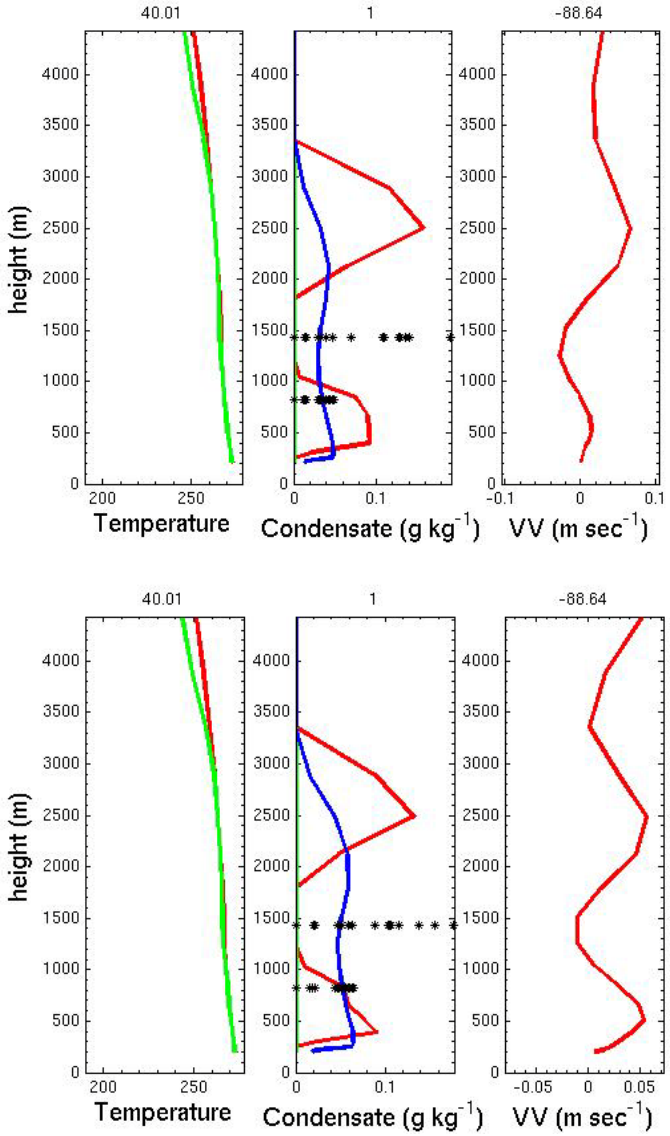


Figure 8. 13-km WRF run (top), 3-km WRF run bottom. Left panel; (T (red),Td (green)), middle panel; (cloud water (red), rain water (green), frozen condensate (blue)) black stars denote PIREP icing top and base, right panel; vertical velocity (red). Latitude, icing severity, and longitude shown along top of panels 1,2, and 3.

PIREP 16. Moderate mixing icing was reported west of Nova Scotia, Canada between 1.3 and 2.5-km. Neither model correctly identified liquid water at the icing level. Both had low (~ 0.02 g kg^{-1}) amounts of ice condensate. The 3-km WRF did predict a layer of supercooled liquid just above the icing levels (Fig 9).

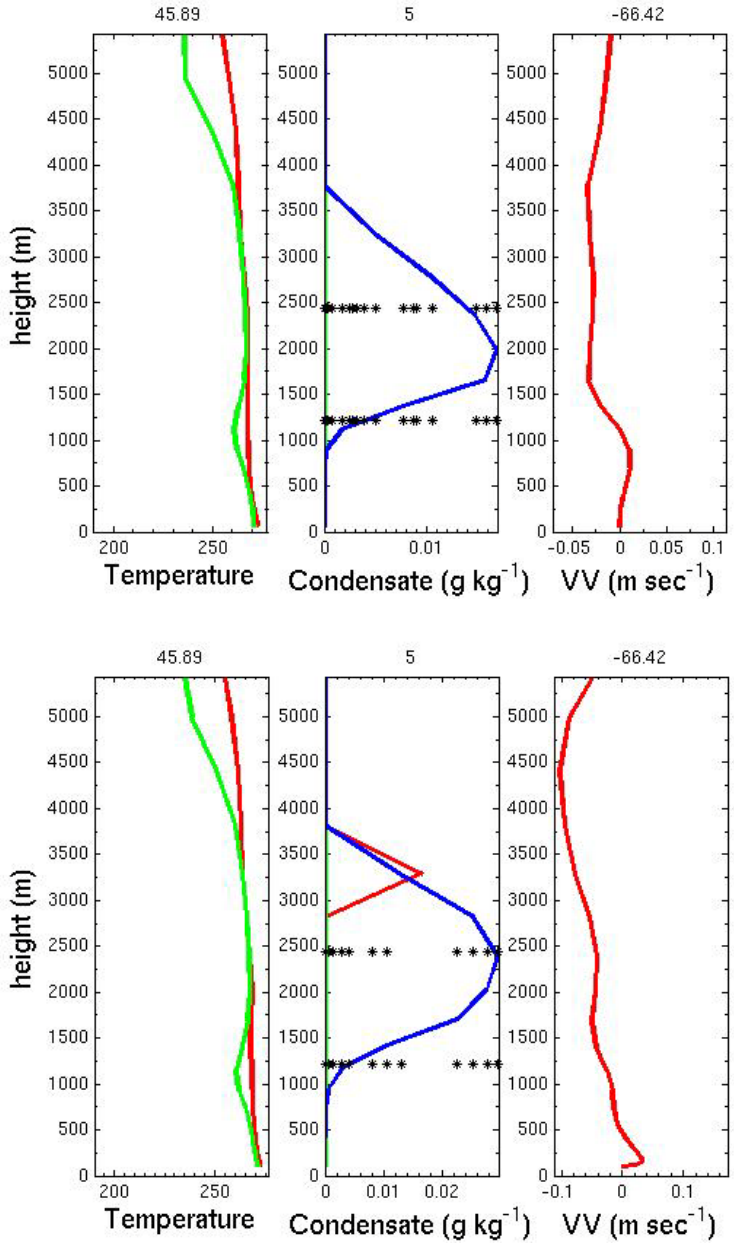


Figure 9. 13-km WRF run (top), 3-km WRF run bottom. Left panel; (T (red),Td (green)), middle panel; (cloud water (red), rain water (green), frozen condensate (blue)) black stars denote PIREP icing top and base, right panel; vertical velocity (red). Latitude, icing severity, and longitude shown along top of panels 1,2, and 3.

PIREP 17 (not shown). Light clear icing was reported over Columbia, MO between 0.8 – 2.0-km. This was likely boundary layer icing. Both models had the top of the boundary layer too low and most of the icing layer did not have any supercooled liquid water present. The base of the icing layer had up to 0.2 g kg^{-1} of supercooled liquid predicted.

PIREP 18(not shown). Light clear icing was reported over southern MI near 3-km. Both models had a deep cloud forecast. The vertical distribution of condensate was quite different with the 3-km having more ice phase at higher altitudes while the 13-km had more ice at lower levels. Both models did have some supercooled liquid at the icing level with the 3-km having more than 13-km.

PIREP 19. (not shown). Light mixed icing was reported along the northeast shore of Lake Ontario between 1.2 and 2.8-km. Both model runs had an ice phase cloud predicted at the top of the icing layer with dry, sub-saturated air in the remainder of the icing layer.

PIREP 20. Negative icing was reported at Casper, WY at 3.5-km. Both model correctly identified the negative icing with an ice phase cloud at this location (Fig 10).

PIREP 21(not shown). Light rime icing was reported below 2.5-km over Flint, MI. Both models predicted a deep cloud with large ice phase condensate mass at and above the icing level. Both models also correctly identified supercooled liquid (~ 0.1 g kg^{-1}) at the icing levels. The vertical velocities from the 3-km were much larger.

PIREP 22. Light icing was reported at Bangor, ME between 1.5 and 3.0-km. Both model runs did a good job predicting the icing cloud (Fig 11).

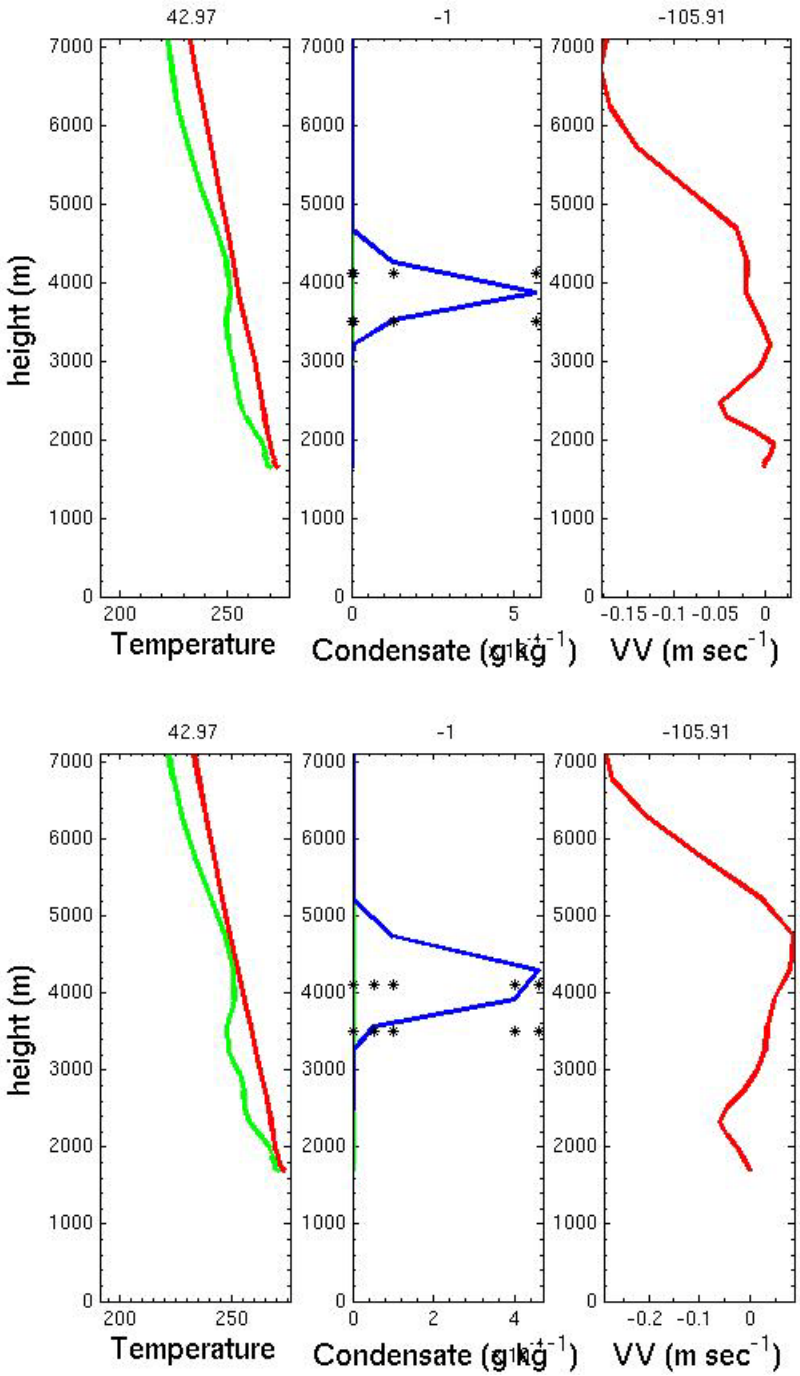


Figure 10. 13-km WRF run (top), 3-km WRF run bottom. Left panel; (T (red),Td (green)), middle panel; (cloud water (red), rain water (green), frozen condensate (blue)) black stars denote PIREP icing top and base, right panel; vertical velocity (red). Latitude, icing severity, and longitude shown along top of panels 1,2, and 3.

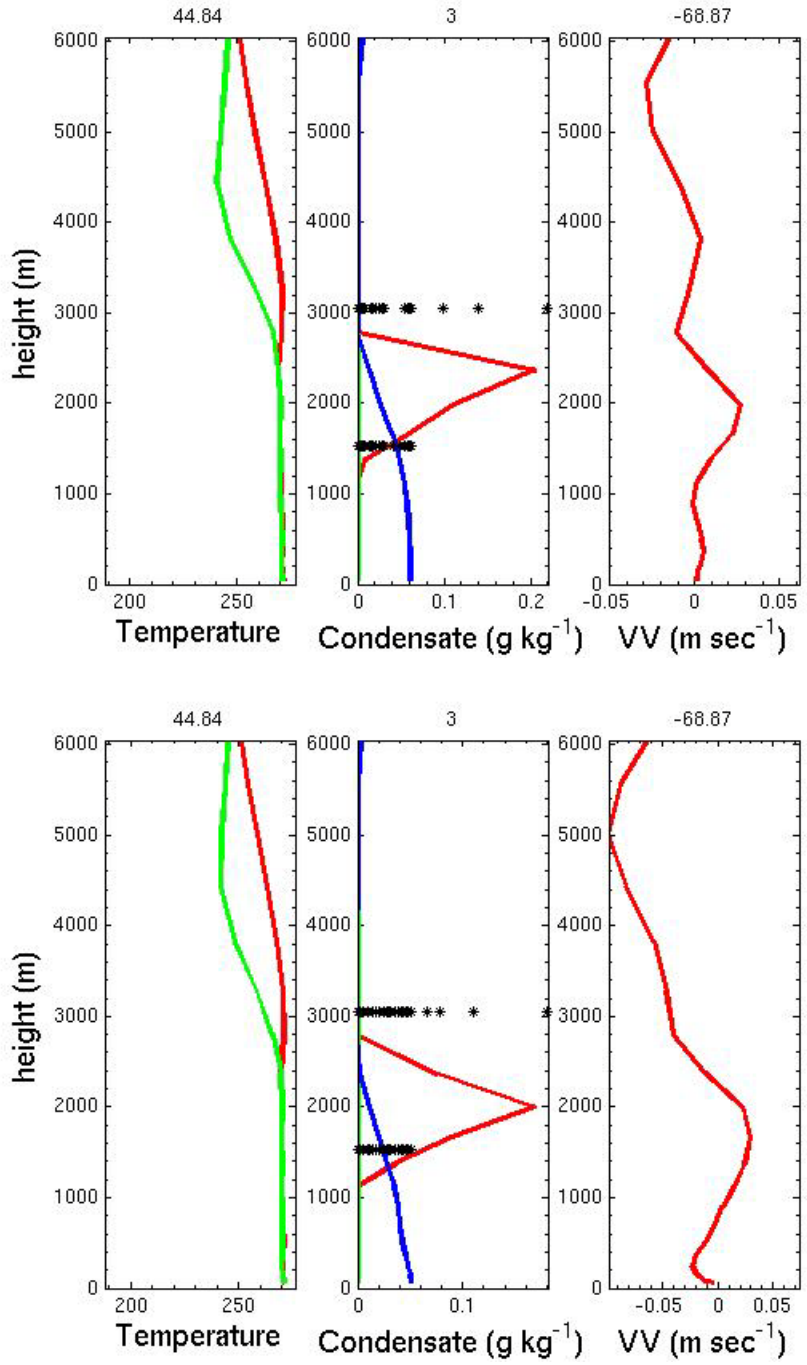


Figure 11. 13-km WRF run (top), 3-km WRF run bottom. Left panel; (T (red),Td (green)), middle panel; (cloud water (red), rain water (green), frozen condensate (blue)) black stars denote PIREP icing top and base, right panel; vertical velocity (red). Latitude, icing severity, and longitude shown along top of panels 1,2, and 3.

PIREP 23 (not shown). Light rime icing was reported over central, IL between 1.0 and 2.5-km. Both models had a deep cloud layer forecast with low mass of ice condensate (~ 0.02 g kg). Both models did predict a layer of supercooled liquid in the boundary layer but that was below the level were the icing was reported. It is unclear if the boundary layer depth was too shallow or if there was an icing layer that was misdiagnosed as ice phase.

A.2.2 June 4, 2008 2100 UTC

PIREP 1 – Moderate rime icing was reported near Las Vegas, NV between 5.6 – 6.2-km at -5C. Both the 3-km and 13-km WRF runs correctly identified cloudy conditions near the PIREP altitudes. The 13-km had minimal condensate ice and some SLW (0.06 g kg^{-1}), while the 3-km correctly had much more SLW predicted (0.5 g kg^{-1}). The vertical velocity in the 3-km was more than double (0.5 m sec^{-1}) the 13-km (Fig 1).

PIREP 2(not shown).. – Light rime icing was reported near Charleston, WV between 6.5- and 7-km at -14C. Both the 13-km and 3-km model runs correctly identified cloud at the icing altitudes but the both incorrectly had only frozen condensate predicted. The mass of ice phase in the 13-km model run was only 0.1 g kg^{-1}, while the 3-km had much more 2.0 g kg^{-1}. The vertical velocities were also much larger in the 3-km model.

PIREP 3- Light rime icing was reported over Larimie, WY between 5.5 – 6-km at -13C. Both models correctly identified clouds at the icing altitudes but both incorrectly only predicted ice phase clouds. The 13-km and 3-km WRF runs both had about 0.15 g kg^{-1} of ice phase at the icing altitudes. The 13-km had weak ascent through the icing layer, while the 3-km had descending air (Fig 2).

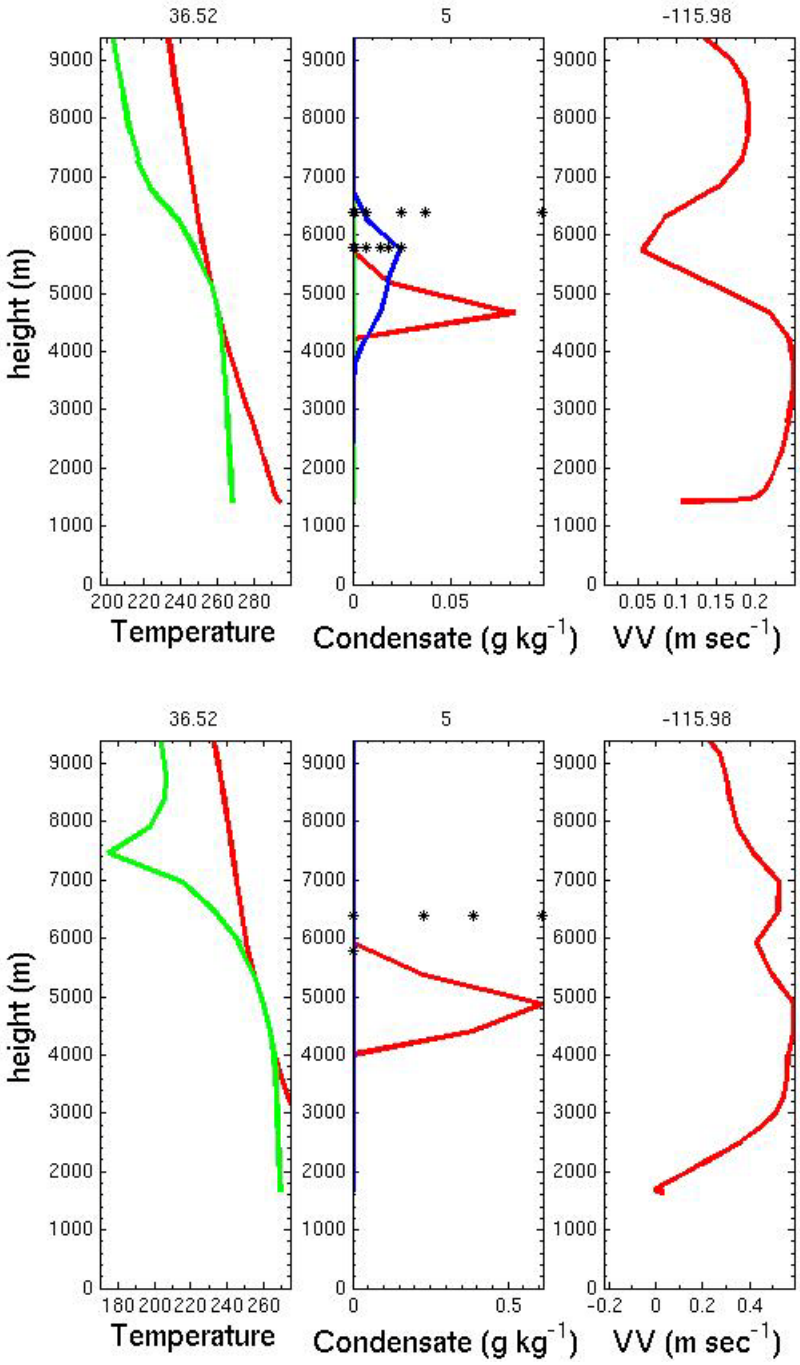


Figure 1. 13-km WRF run (top), 3-km WRF run bottom. Left panel; (T (red),Td (green)), middle panel; (cloud water (red), rain water (green), frozen condensate (blue)) black stars denote PIREP icing top and base, right panel; vertical velocity (red). Latitude, icing severity, and longitude shown along top of panels 1,2, and 3.

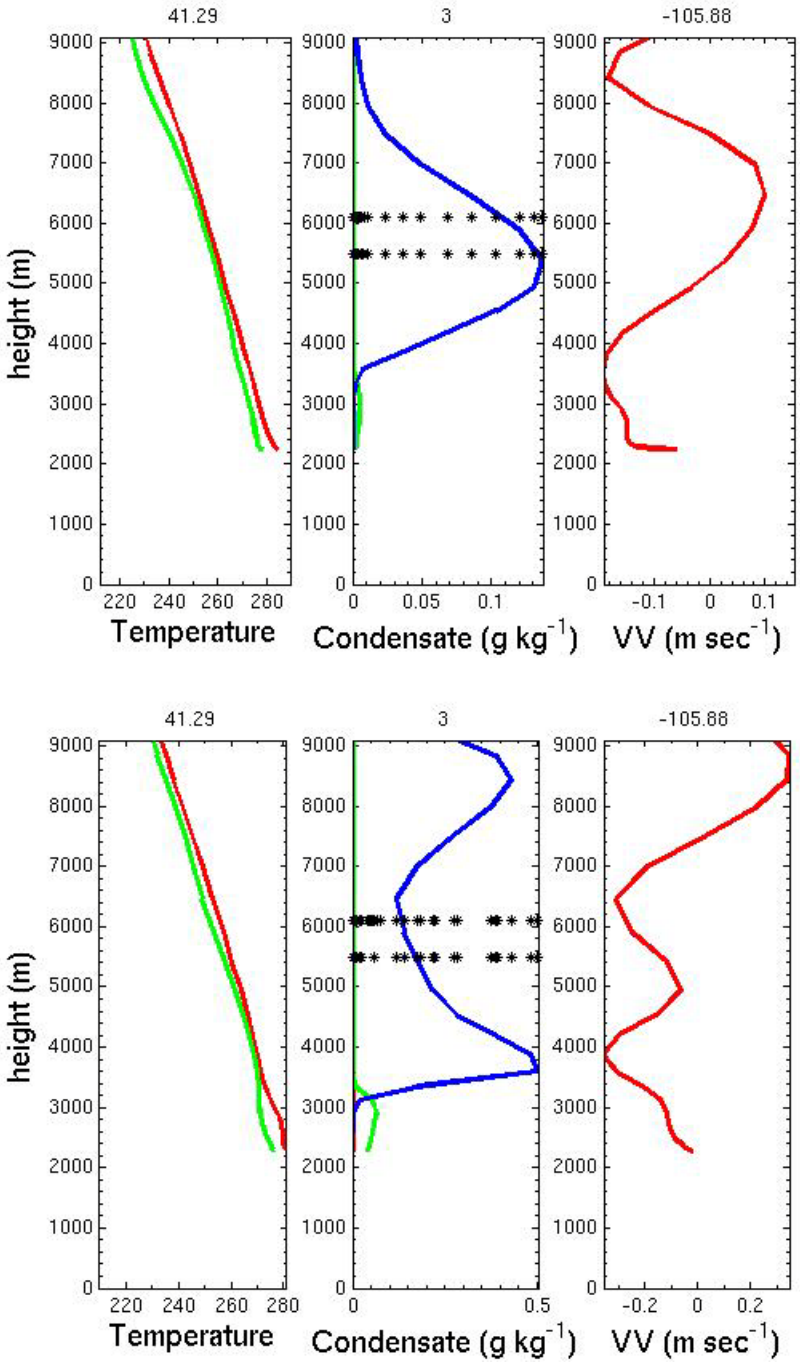


Figure 2. 13-km WRF run (top), 3-km WRF run bottom. Left panel; (T (red),Td (green)), middle panel; (cloud water (red), rain water (green), frozen condensate (blue)) black stars denote PIREP icing top and base, right panel; vertical velocity (red). Latitude, icing severity, and longitude shown along top of panels 1,2, and 3.

PIREP 4 – Moderate mixed icing was reported over the Rocky mountains west of Colorado Springs, CO between 5.9 – 6.2-km. Both models correctly identified clouds at the icing altitudes. The 13-km had up to 0.6 g kg^{-1} of ice phase, while the 3-km had more than 1.25 g kg^{-1} of ice phase and ~0.3 g kg^{-1} of supercooled liquid. The upward vertical velocities were also much stronger in the 3-km compared to the 13-km (Fig 3).

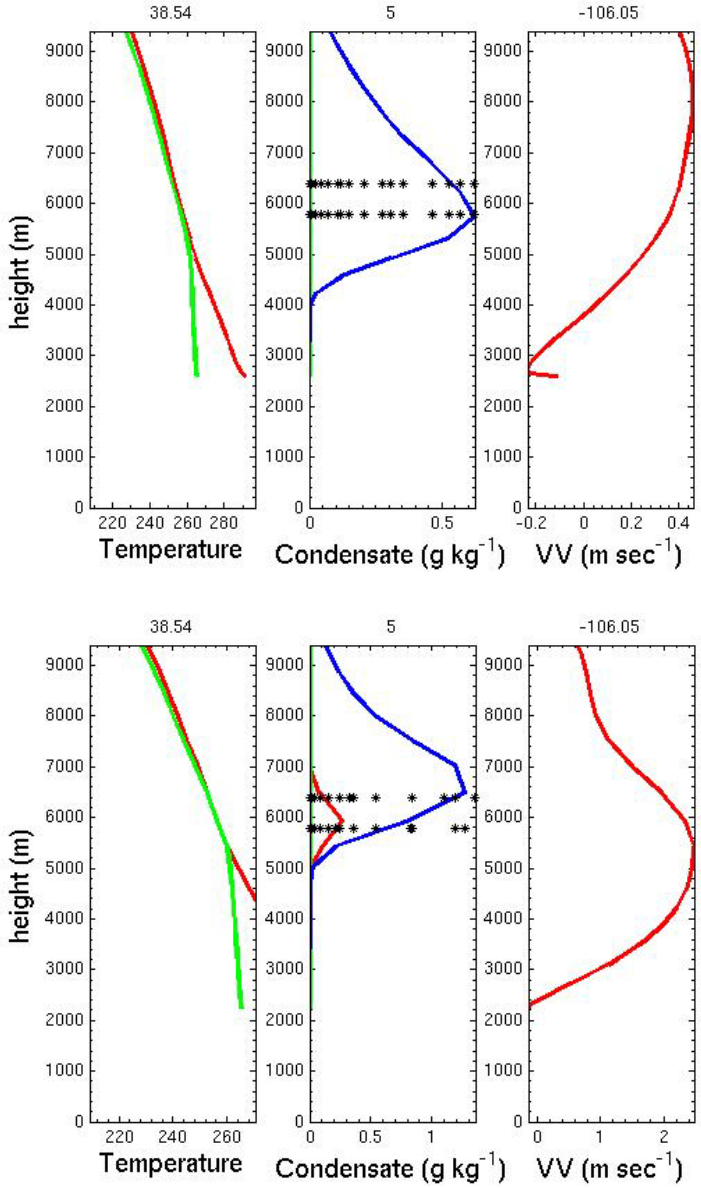


Figure 3. 13-km WRF run (top), 3-km WRF run bottom. Left panel; (T (red),Td (green)), middle panel; (cloud water (red), rain water (green), frozen condensate (blue)) black stars denote PIREP icing top and base, right panel; vertical velocity (red). Latitude, icing severity, and longitude shown along top of panels 1,2, and 3.

A.2.3 Dec 9, 2009 1800 UTC

PIREP 1 – Light mixed icing at 600-m was reported near Indianapolis IN. The 3-km WRF correctly identified the icing conditions with up to 0.3 g kg $^{-1}$ of supercooled liquid within the cloud layer. The 13-km WRF run also had supercooled liquid (up to 0.25 g kg^{-1}) predicted but the base of the supercooled cloud was a bit to high (Fig 1).

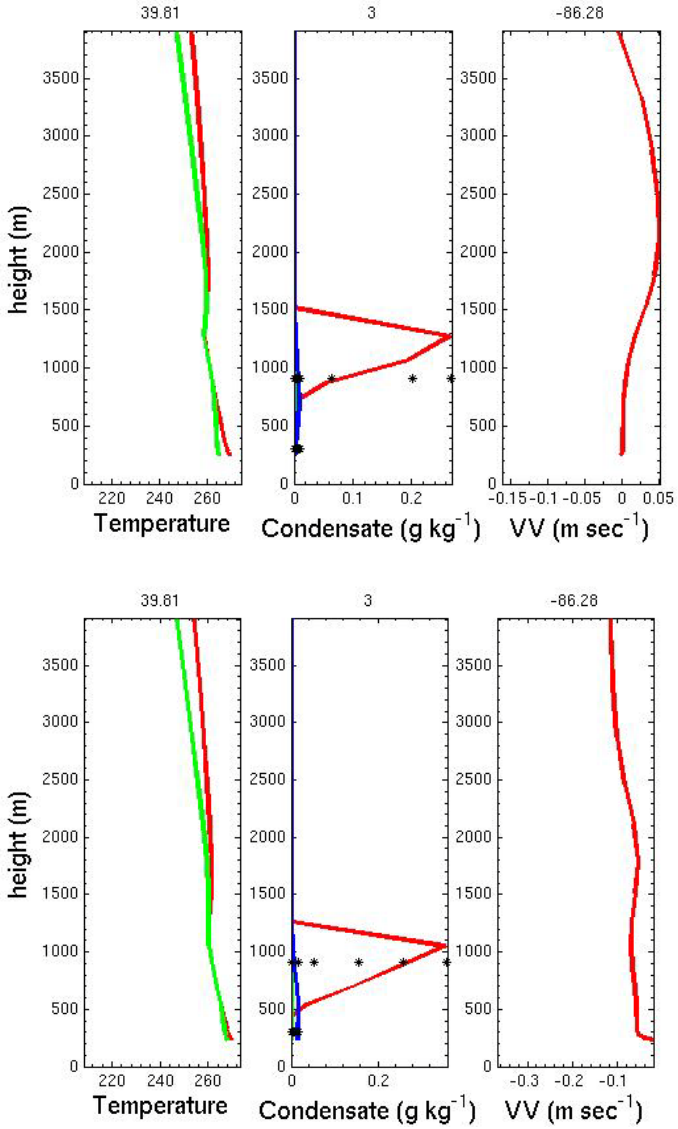


Figure 1. 13-km WRF run (top), 3-km WRF run(bottom). Left panel; (T (red),Td (green)), middle panel; (cloud water (red), rain water (green), frozen condensate (blue)) black stars denote PIREP icing top and base, right panel; vertical velocity (red). Latitude, icing severity, and longitude shown along top of panels 1, 2, and 3.

PIREP 2 – Moderate rime icing between 1700 and 3000 m was reported at Zanesville, OH. The model solutions between the 13-km and 3-km WRF runs were quite different. The 13-km had a deeper cloud layer (condensate top at 5-km) with up to 0.08 g kg^{-1} of ice phase condensate at the base of the icing layer. The 3-km WRF run had a condensate top at 2500-m, below the icing top, but it did have up to 0.06 g kg^{-1} of supercooled liquid at the base of the icing layer. The 3-km WRF had stronger upward vertical velocities through the icing layer (Fig 2).

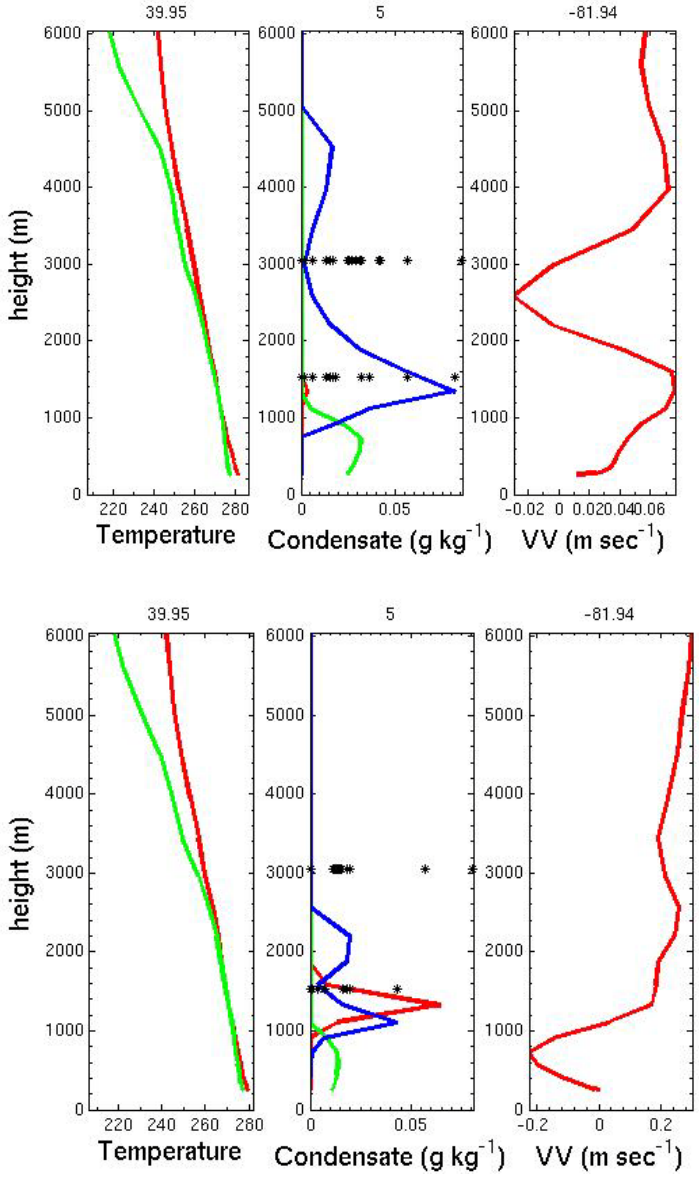


Figure 2. 13-km WRF run (top), 3-km WRF run (bottom). Left panel; (T (red),Td (green)), middle panel; (cloud water (red), rain water (green), frozen condensate (blue)) black stars denote PIREP icing top and base, right panel; vertical velocity (red). Latitude, icing severity, and longitude shown along top of panels 1, 2, and 3.

PIREP 3 – Heavy rime icing was reported at T = -22C between 4.25-km and 4.9-km over Jamestown, NY. Both the 13-km and 3-km models only had ice phase condensate predicted at the icing location and altitudes. The 13-km had more condensate mass than the 3-km WRF model run (Fig 3).

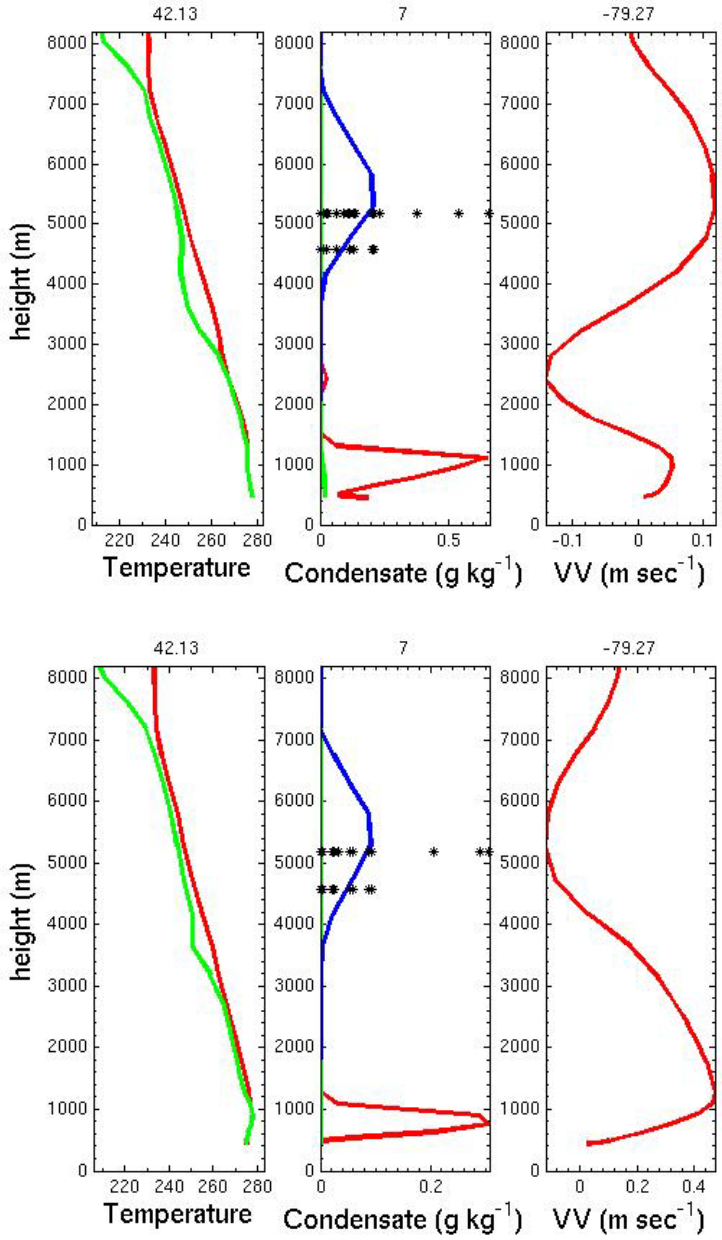


Figure 3. 13-km WRF run (top), 3-km WRF run (bottom). Left panel; (T (red),Td (green)), middle panel; (cloud water (red), rain water (green), frozen condensate (blue)) black stars denote PIREP icing top and base, right panel; vertical velocity (red). Latitude, icing severity, and longitude shown along top of panels 1, 2, and 3.

PIREP 4 – Moderate mixed icing was reported between 600 – 1200 m over South Bend IN. Both models had supercooled liquid predicted at the icing levels but the 3-km correctly had more than 3 times the mass of supercooled liquid as the 13-km (Fig 4).

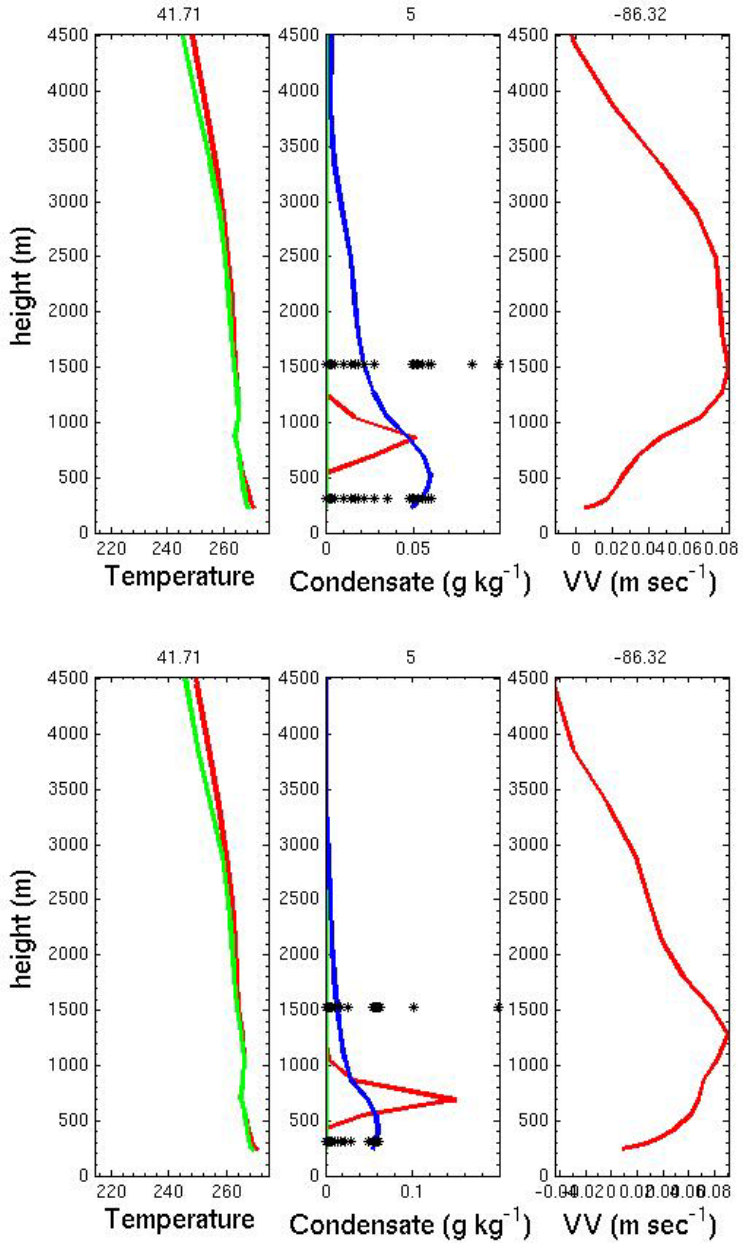


Figure 4. 13-km WRF run (top), 3-km WRF run(bottom). Left panel; (T (red),Td (green)), middle panel; (cloud water (red), rain water (green), frozen condensate (blue)) black stars denote PIREP icing top and base, right panel; vertical velocity (red). Latitude, icing severity, and longitude shown along top of panels 1, 2, and 3.

PIREP 5 (not shown)- Moderate rime icing was reported between 1000 and 2500 m over Joilet, IL. The model solutions were very similar. Both models had low amounts of ice phase at cloud top near 2500 m and both models had supercooled liquid water in the 0.1 – 0.2 g kg^{-1} range. Both models also had upward vertical velocities in the range of 15 – 20 cm sec^{-1}.

PIREP 6 – Light rime icing was reported near Dayton, OH between 1800 – 2200 m. The 3-km WRF correctly had the cloud top at 2500-m and had 0.001 g kg^{-1} of ice phase predicted through the icing layer. The 3-km model also had descending air through much of the cloud layer. The 13-km had the cloud top at 2000-m and correctly had up to 0.18 g kg^{-1} of supercooled liquid water within and below the reported icing altitudes. The 13-km WRF also had rising air where the 3-km had descending air (Fig. 5).

PIREP 7 – Light icing was reported near Grand Rapids, MI though the layer 500 – 4000 m. Both models had the condensate top at about 2800-m, well below the top of the icing layer. Both models also had ice phase predicted from the model cloud top to the surface. The 13-km had less total condensate but did have a thin layer of supercooled liquid water near the base of the icing layer while the 3-km only had ice phase and less total condensate. The vertical velocities were downward at the base of the layer and upwards near the top of the layer. The magnitudes of the vertical velocities were similar in both model configurations (Fig 6).

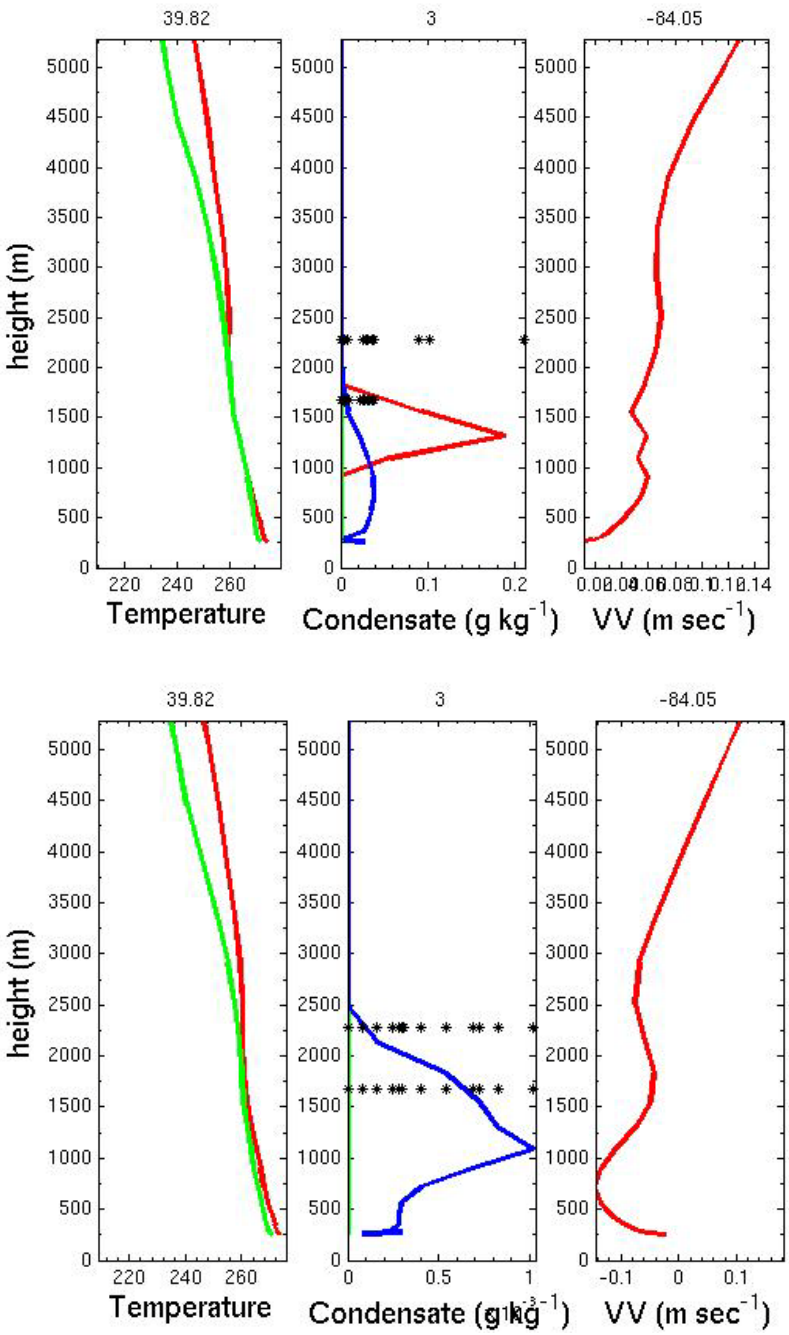


Figure 5. 13-km WRF run (top), 3-km WRF run (bottom). Left panel; (T (red),Td (green)), middle panel; (cloud water (red), rain water (green), frozen condensate (blue)) black stars denote PIREP icing top and base, right panel; vertical velocity (red). Latitude, icing severity, and longitude shown along top of panels 1, 2, and 3.

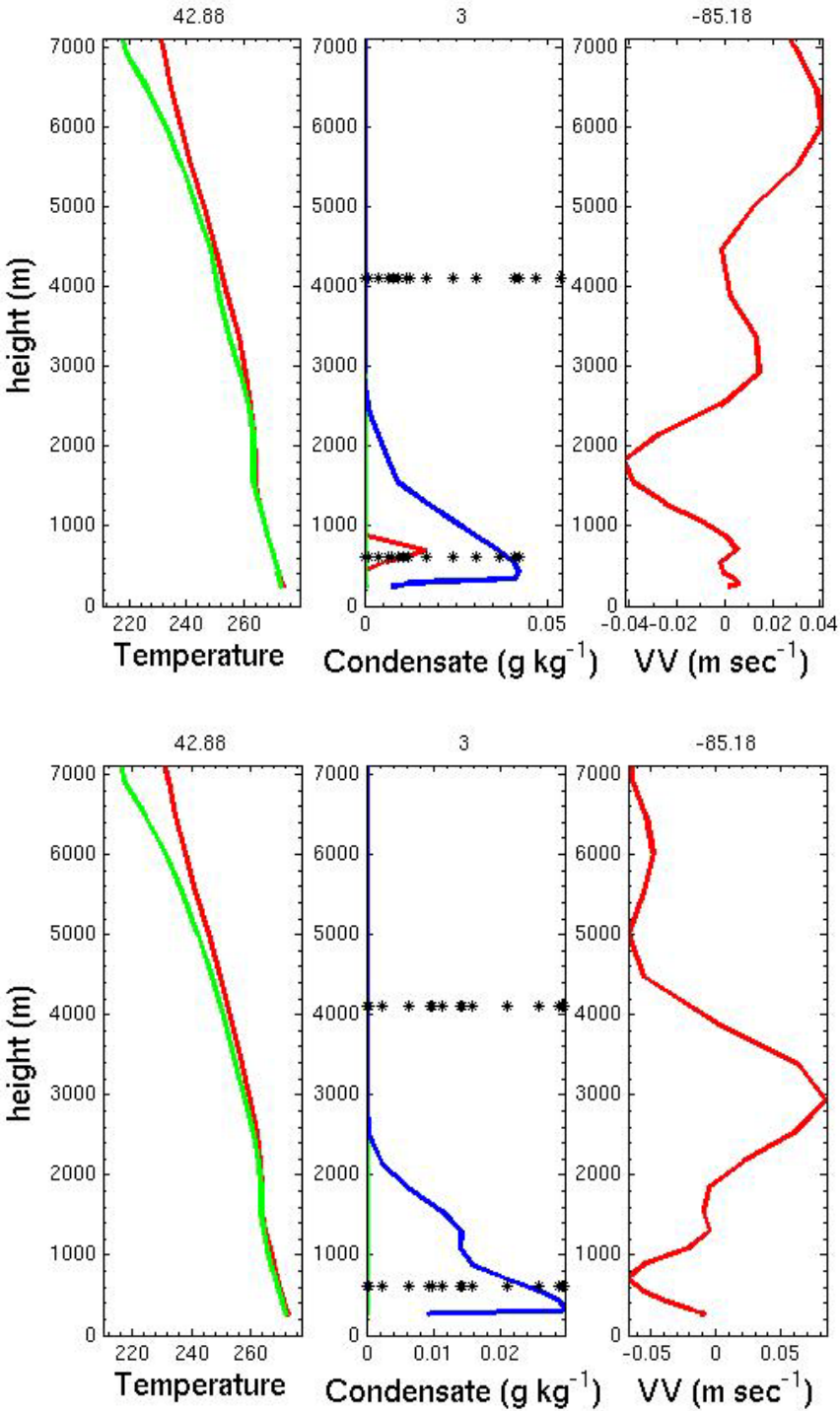


Figure 6. 13-km WRF run (top), 3-km WRF run (bottom). Left panel; (T (red),Td (green)), middle panel; (cloud water (red), rain water (green), frozen condensate (blue)) black stars denote PIREP icing top and base, right panel; vertical velocity (red). Latitude, icing severity, and longitude shown along top of panels 1, 2, and 3.

PIREP 8 – Moderate mixed icing was reported over Ottawa, Canada between 1500 – 4100 m. Both the 13-km and 3-km WRF runs correctly placed the PIREP within a model cloud, The 13-km only had ice condensate predicted (up to 0.4 g kg^{-1}) while the 3-km had both ice phase (up to 0.4 g kg^{-1}) and supercooled liquid (up to 0.15 g kg^{-1}) predicted (Fig 7).

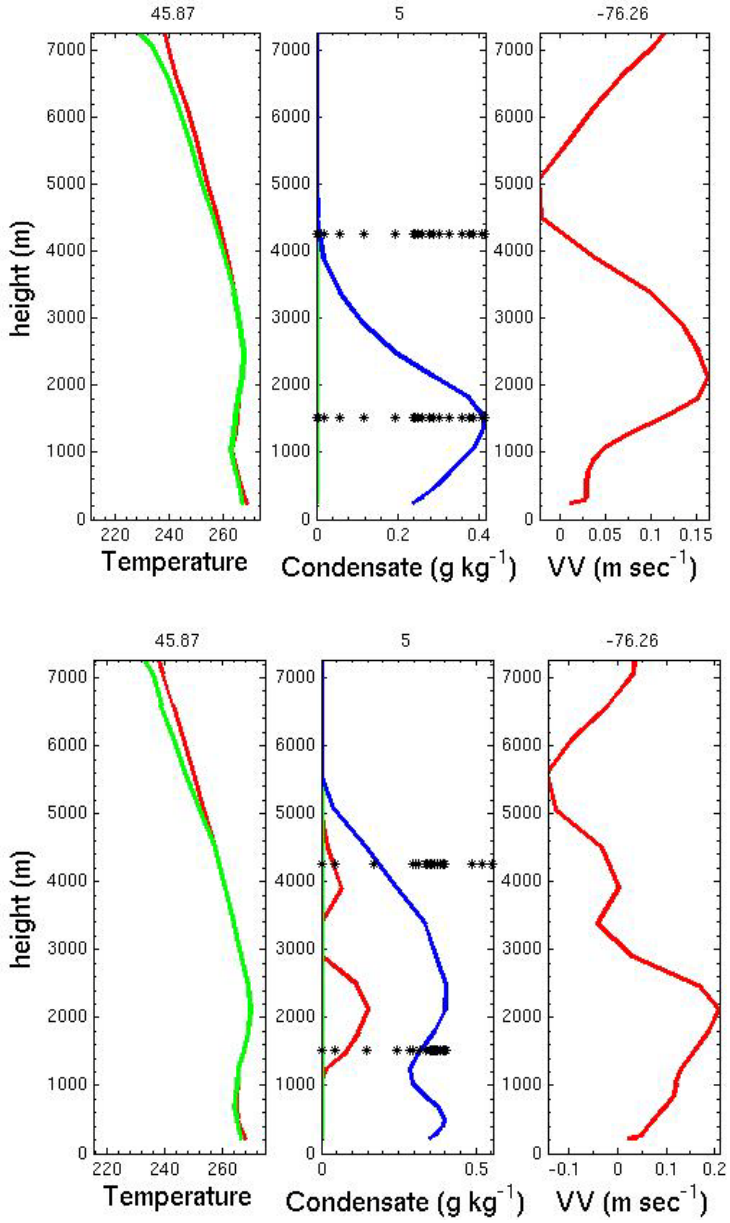


Figure 7. 13-km WRF run (top), 3-km WRF run (bottom). Left panel; (T (red),Td (green)), middle panel; (cloud water (red), rain water (green), frozen condensate (blue)) black stars denote PIREP icing top and base, right panel; vertical velocity (red). Latitude, icing severity, and longitude shown along top of panels 1, 2, and 3.

REFERENCES

Bernstein, B.C., F. McDonough, M.K. Politovich, B.G. Brown, T.P. Ratvasky, D.R. Miller, C.A. Wolff and G. Cunning, 2005: Current Icing Potential: Algorithm Description and Comparison with Aircraft Observations. J*. Appl. Meteor.,* Vol. 44**,** No. 7 (July), 969-986.

Cornman, L. B., C. S. Morse, and G. Cunning, 1995: Real-time estimation of atmospheric turbulence severity from in-situ aircraft measurements. *J. Aircraft*, **32,** 171-177.

Ellrod, G. P., and D. L. Knapp, 1992: An objective clear-air turbulence forecasting technique: Verification and operational use. *Wea. Forecasting*, **7**, 150-165.

Frehlich, R. and R. Sharman, 2004: Estimates of turbulence from numerical weather prediction model output with applications to turbulence diagnostics and data assimilation. *Mon. Wea. Rev*., **132**, 2308-2324.

Hogan, Robin J., Christopher A. T. Ferro, Ian T. Jolliffe, David B. Stephenson, 2010: Equitability Revisited: Why the “Equitable Threat Score” Is Not Equitable. Wea. Forecasting, 25, 710–726.

Hu, M., M. Xue, and K. Brewster, 2006: 3DVAR and cloud analysis with WSR-88D level-II data for the prediction of the Fort Worth, Texas, tonadic thunderstorm. Part II: impact of radial velocity analysis via 3DVAR. Mon. Wea. Rev., 134, 675-698.

Janjić, Z. I., 2002: Nonsingular Implementation of the Mellor–Yamada Level 2.5 Scheme in the NCEP Meso model, NCEP Office Note, No. 437, 61 pp.

Knievel, Jason C., George H. Bryan, Joshua P. Hacker, 2007: Explicit Numerical Diffusion in the WRF Model. Mon. Wea. Rev., 135, 3808–3824.

Roberts, N. M., 2005: An investigation of the ability of a storm scale configuration of the Met Office NWP model to predict flood-producing rainfall. Met Office Tech. Rep. 455, 80 pp.

Sharman, R., C. Tebaldi, G. Wiener, and J. Wolff, 2006: An integrated approach to mid- and upper-level turbulence forecasting. *Wea. Forecasting*, **21**, 268-287.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X.-Y. Huang, W. Wang, and J. G. Powers, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp.

Vinnichenko, N. K., N. Z. Pinus, S. M. Shmeter, and G. N. Shur, 1980: *Turbulence in the Free Atmosphere.* Plenum, 310 pp.

Weygandt, S. S., S. G. Benjamin, T. G. Smirnova, and J. M. Brown, 2008: Assimilation of radar reflectivity data using a diabatic digital filter within the Rapid Update Cycle. Preprints, 12th Conf. on IOAS-AOLS. January 2008, New Orleans, LA, AMS.