The Ability of Radar Data Assimilation to Improve Warm Season Heavy Rainfall Forecasts for use in Hydrologic Models

Ben A. Moser[1], William A. Gallus Jr.[1], and Ricardo Mantilla[2]
[1]Department of Geological and Atmospheric Sciences, Iowa State University
[2]Iowa Flood Center, University of Iowa

1. Introduction

Warm Season convective rainfall is one of the most poorly forecasted parameters in numerical models. Yet it is also one of the most important, because in many areas of the world including the Upper Midwest it often accounts for a large percentage of the annual rainfall. In addition, this rainfall often occurs with very high rates, which can lead to flooding if the duration of the event is sufficiently long. Unfortunately, quantitative precipitation forecasting (QPF) skill has traditionally been so poor that these forecasts are not used in hydrologic modeling for stream flow. Instead, stream flow forecasts are made using estimates of precipitation that has fallen, reducing the amount of lead-time for warnings from what could exist if forecasts were used. Thus a continued focus in the meteorological community has been on increasing the forecasting accuracy of warm season convective rainfall.

Several studies have shown the ability of radar data assimilation to improve precipitation forecasts, especially in the short-term (first 6 hours or so). For instance Macpherson (1999) along with Davolio and Buzzi (2004) found that assimilation of radar data via nudging techniques yielded improved precipitation forecasts in the first 6 hours. Other studies such as Sugimoto et al. (2009) have focused on the use of 3-dimensional variational (3DVAR) data assimilation techniques and have found that the assimilation of all possible Doppler radar data (radial velocity and reflectivity data) results in the best performance regarding short-range precipitation forecasting.

This study uses a 3DVAR analysis system that is developed within the Advanced Regional Prediction System (ARPS) model framework (Xue et al. 1995, 2000, 2001) in order to assimilate the radial velocity data and a cloud analysis procedure that is a component of both the ARPS 3DVAR system and the ARPS Data Analysis System (ADAS; Brewster 1996) to adjust the hydrometeor and cloud fields based off of reflectivity data. Simulations are then run using the Weather Research and Forecasting model (WRF). The simulated rainfall and post-processed radar data are analyzed in order to get an idea of the degree to which the accuracy of the convective rainfall simulations can be increased by adjusting the model initial conditions using radar data. The focus of this project is on quantifying the impact of radar data assimilation on QPF skill in Iowa.

2. Methodology

2.1 WRF

For the simulations in this project the Weather Research and Forecasting (WRF) version 3.3.1 ARW (Advanced Research WRF) core is used, which encompasses the ARW dynamics solver (Eulerian mass dynamics solver) along with the compatible components (e.g. physics schemes, dynamics options, etc.)

Corresponding author address: Ben A. Moser, 4112 Lincoln Swing St. #218, Ames, IA 50014. Email: bamoser@iastate.edu

of WRF with that solver. The external data used to create the initial and boundary conditions comes from the NAM (North American regional model) and is stored on the NCEP grid #218, which is a 12 km grid on a Lambert Conformal map projection. The domain size for the simulations is roughly 800 x 800 km centered over Iowa. A high-resolution 4 km spaced grid in the horizontal with 40 grid levels in the vertical is used. At a 4 km grid resolution it is assumed that deep convective clouds can be explicitly resolved to a good degree, so a cumulus parameterization scheme is not used in the model. The Mellor-Yamada-Janjic (MYJ; Janjic, 1994) planetary boundary layer (PBL) scheme was used with microphysics being parameterized with the Thompson scheme (Thompson et al. 2008). Two MCS events are simulated; both occurring in the early morning so the model is initialized at 00z (using 00z initialized NAM data) and run for a duration of 24 hours with boundary conditions being ingested every 3 hours. This study focuses on the first 12 hours of the model runs since radar data often doesn't result in much improvement after that.

2.2 ARPS 3DVAR

The 3DVAR Analysis System developed as a part of the Advanced Regional Prediction System (ARPS) along with a cloud analysis procedure that is a component of both ARPS 3DVAR and ADAS is used in this study to produce initial conditions (analysis) for the WRF model simulations. In order to carry out the data assimilation using the radial wind data, the ARPS 3DVAR uses an incremental form of cost function that includes the background, observation, and equation constraint terms (Hu et al. 2006). NAM Data at 00z (time of initialization) is interpolated to the ARPS gird and used as the background (or first-guess) field for the 3DVAR data assimilation procedure for both wind analysis and the cloud analysis. The cloud analysis is carried out using radar reflectivity data to construct three-dimensional cloud and precipitating fields. In order to perform the cloud analysis the reflectivity data are first remapped to the analysis grid using a local least squares procedure. The level II radar data comes from NEXRAD WSR-88D radar sites located within the model domain; all radar sites located within the model domain are used so that the areal data coverage provided by the radars is the best possible.

2.3 Probability Matched Method

Statistics are included in the results section for a probability-matched method (PM) that was used to produce a rainfall forecast for an 8 member mixed-physics model ensemble. Initial and lateral boundary conditions for the ensemble members come from the 12 km NAM (in 4 members) and the 40 km GFS (in the other 4 members). The Thompson, Ferrier and WSM-6 microphysics and the YSU and MYJ PBL parameterization schemes are varied also between the 8 members (Table 1).

Ensemble Members	
NAM-MYJ-Thompson	GFS-MYJ-Thompson
NAM-MYJ-Ferrier	GFS-MYJ-Ferrier
NAM-YSU-Thompson	GFS-YSU-Thompson
NAM-YSU-WSM6	GFS-YSU-WSM6

Table 1: Mixed physics ensemble members

The PM rainfall forecast was included in order to give an idea of how another method for improving rainfall forecasts would fare for these two cases. In order to implement the PM method it is assumed that the most accurate spatial representation of the rain field is given by the ensemble mean and that the best frequency distribution of rain rates is given by the ensemble of individual members QPFs (Ebert, 2001). Thus, the PM rainfall forecast is created by pooling the forecast rain rates for all members (and all grid points) and arranging them in order from largest to smallest and keeping every 8th value (since we have 8 members). Next, the rain rates in the simple ensemble mean are pooled and arranged in the

same order with the location of each value also being stored. The grid point with the highest rain rate in the simple ensemble mean rain field is then replaced with the highest value from the ensemble member rain rate distribution, and so on. This method is used since it preserves the mean and maximum rainfall intensity predicted by individual ensemble members and prevents the smearing or smoothing effect taking the simple mean has on the predicted rain area.

2.4 Analysis Methods

Model simulations were run for the June 22, 2010 and June 26, 2010 heavy rainfall cases that occurred in and around the state of Iowa. For each case there were two simulations, one where radar data was used to produce the initial conditions (analysis) and one where the initial conditions were produced by interpolating the NAM data to the WRF grid. The hourly simulated rainfall amounts and post-processed reflectivity were analyzed and compared between the different simulations, and then compared with observed rainfall and NEXRAD reflectivity for the two cases.

a. Equitable Threat Score (ETS) and Bias

ETS and bias were calculated for the control and radar adjusted model simulations along with the rainfall forecast that was produced using a probability matching technique for comparison. The ETS used in this study can be found in Schaefer (1990) and is given below (eq. 1) along with the bias equation (2). As stated in Schaefer (1990) "the score is

$$ETS = \frac{(\mathrm{CFA}-\mathrm{CHA})}{(\mathrm{F}+\mathrm{O}-\mathrm{CFA}-\mathrm{CHA})} \qquad (1)$$

$$bias = \frac{\mathrm{F}}{\mathrm{O}} \qquad (2)$$

the number of correctly forecasted points in excess of those that would verify by chance, divided by the number of cases when there was a threat that would not be foreseen by chance". A perfect forecast would result in an ETS score of 1 with lower values showing a less accurate forecast. Bias values much greater than 1 indicate that the model over predicted areal coverage while values less than 1 indicate the model under predicted areal coverage of rainfall greater than the threshold value.

b. Rainfall Characteristics

Along with the ETSs and bias, the number of grid points with rainfall above certain rainfall thresholds (areal coverage), along with the volume and rain rate for those points was also examined. These three measures help to better characterize the forecasted rainfall, since two runs may have a similar rain volume with one achieving it through lighter rain over a large area and the other through heavy rainfall over a smaller area.

3. Results

3.1 June 22, 2010

The first heavy rainfall case to be examined occurred during the evening and overnight of June 21, 2010 into the morning hours of June 22 over parts of southern Iowa into northwestern Illinois. The focal point of this activity was initially a shortwave trough that moved across northern Iowa into southern Wisconsin and northern Illinois. As the activity with this was moving off to the east thunderstorms began to develop along the southern periphery of this activity near a stalled out frontal boundary. This development appears to be partially in response to a strengthening LLJ oriented perpendicular to the front, it was this activity that was responsible for the heavy rainfall that occurred over southern Iowa into northern Illinois.

From the reflectivity plots (Fig. 1) it is evident that the radar data assimilation worked well for the first few hours of the simulation in this case, as it correctly simulated thunderstorms along the frontal boundary in southern Iowa that weren't there in the cold-start run. However by

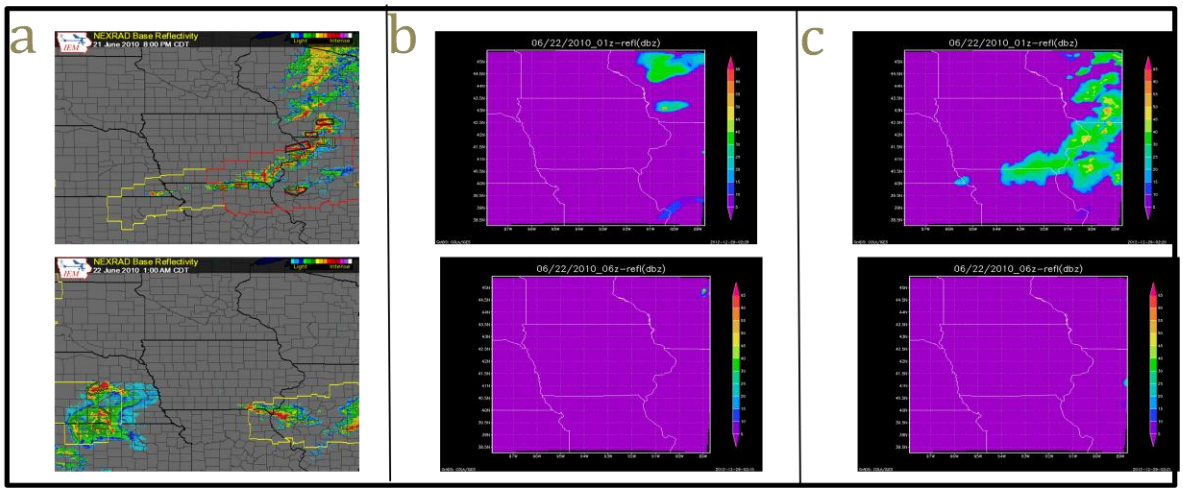

Fig. 1: Plots for June 22, 2010 of a) 2km NEXRAD reflectivity, b) cold start run, and c) radar assimilated run simulated reflectivity at 01z (top) and 06z (bottom)

06z these had diminished and moved off to the east leaving little activity in that area, while in reality as noted from the NEXRAD image at 06z (Fig. 1) storms continued to back-build and train over that area.

Comparing the characteristics between the simulated rainfall and the actual observed rainfall (Table 2) it is apparent that the radar assimilated run did a fairly good job of simulating the rainfall in the first 6 hours of the model run (00z-06z), although the areal extent and intensity of the rainfall for the points exceeding the 1 inch threshold was a bit overdone. In contrast the areal coverage for the .01 inch threshold was actually too small, this means that overall the simulated rainfall was more intense and occurred over a smaller area than the observed rainfall. The rainfall characteristics were not simulated well for the following 6 hour period (06z-12z) as very little rainfall was produced by the simulation.

June 22, 2010				
		Forecast Period (UTC)		
Threshold(in.)	**Parameters**	**00-06**	**06-12**	**00-12**
0.01	Observed areal coverage (points)	9246	10824	17590
	Observed rain rate (in.)	0.41	0.33	0.42
	Radar assimilated areal coverage (points)	6852	42	7156
	Radar assimilated rain rate (in.)	0.52	0.01	0.50
1	Observed areal coverage (points)	672	642	1543
	Observed rain rate (in.)	1.37	1.38	1.45
	Radar assimilated areal coverage (points)	1154	0	1155
	Radar assimilated rain rate (in.)	1.74	0.00	1.74
	Domain rain volume (in. X 16 km^2)			
0.01	Observed rain volume	3788.21	3530.77	7321.43
	Radar assimilated rain volume	3544.54	0.54	3565.65
1	Observed rain volume	923.96	884.91	2234.58
	Radar assimilated rain volume	2007.40	0.00	2011.08

Table 2: Observed and radar data assimilated run areal coverage, rain rate and rain volume for grid points that exceeded the specified rainfall thresholds

ETSs (Table 3) show that the radar data assimilated run displayed the best forecasting skill overall with it having much higher ETSs for the lighter thresholds than the PM forecast. However, the PM forecast did exhibit slightly higher ETSs for the higher thresholds, which is a bit concerning since the heavy rainfall areas are usually most important when it come to the potential of flooding. As would be expected from the reflectivity plots little to no forecasting skill was present after the first six hours with ETSs around zero for each of the thresholds. The bias values (Table 4) show the heavier rainfall was over-predicted, while the areal extent of the lighter rainfall thresholds was under-predicted. There was fairly low bias (values near 1) for the first 6 hour period (00z-06), but after that the forecast was so poor that bias values were near zero for each of the rainfall thresholds

ETS		June 22, 2010				
Method	Period (UTC)	Threshhold (in.)				
		0.01	0.1	0.5	1	1.5
Cold start	00-06	0.149	0.104	0.039	0.005	0
PM		0.245	0.222	0.211	0.174	0.13
Radar		0.551	0.395	0.183	0.121	0.073
Cold start	06-12	-0.001	0	0	0	0
PM		0.002	-0.002	0	0	0
Radar		0.001	0	0	0	0

Table 3: ETSs for a run that uses radar-data, a cold start run (no radar data), and a PM forecast for rainfall exceeding five different thresholds

bias	June 22, 2010			radar assimilated	
Period (UTC)	Threshhold (in.)				
	0.01	0.1	0.5	1	1.5
00-06	0.741	0.664	0.757	1.717	3.099
06-12	0.004	0	0	0	0
Avg.	0.3725	0.332	0.3785	0.8585	1.5495

Table 4: bias for the radar data assimilated run for rainfall exceeding five different thresholds

3.2 June 26, 2010

The second heavy rainfall case examined in this study occurred during the evening of June 25, 2010 through the early morning hours of June 26th, with the heaviest rainfall occurring over south-central Minnesota into north-central and north-western Iowa. At the surface a favorable synoptic set-up for precipitation was present in the area of southwestern Minnesota and eastern South Dakota where the convective system develop (not shown). This area is just north of a warm front that is extending eastward from a low pressure system with noticeably warmer and higher dew point air just to the south of the front (fairly strong boundary). Further south hot and humid conditions are found over the southern plains with temperatures well into the 90's and dew points in the 70's over many areas of Kansas, Oklahoma, and Texas. Also it should be noted that this set-up is very favorable for strong nocturnal LLJ development as there is a good chance the pressure gradient in the warm sector may be superimposed on the mesoscale pressure gradient associated with daytime differential heating due to the sloped terrain of the plains. In fact a strong southerly LLJ (low-level jet) did develop and played an important role in maintaining and allowing for back-building of this MCS to the south-west in the overnight hours.

From the plots of NEXRAD, cold-start run, and radar data assimilated run reflectivity (Fig. 2) it is clear that the radar data assimilation resulted in a more intense MCS that was positioned further west than in the cold-start run. This further west and more intense MCS is in agreement with the actual NEXRAD reflectivity and STAGE IV observed rainfall (not shown). Unlike in the June 22nd case the radar data assimilation had a positive impact on the simulation through at least the first 12 hours with the convection building further west and remaining more intense than in the cold-start run through that time. Hourly rainfall amounts (not shown) showed that although the radar data assimilated run matched the location of the observed rainfall fairly well it produced too much rainfall in the heaviest precipitating regions.

Rainfall characteristics (Table 5) for this case show that the radar data assimilated

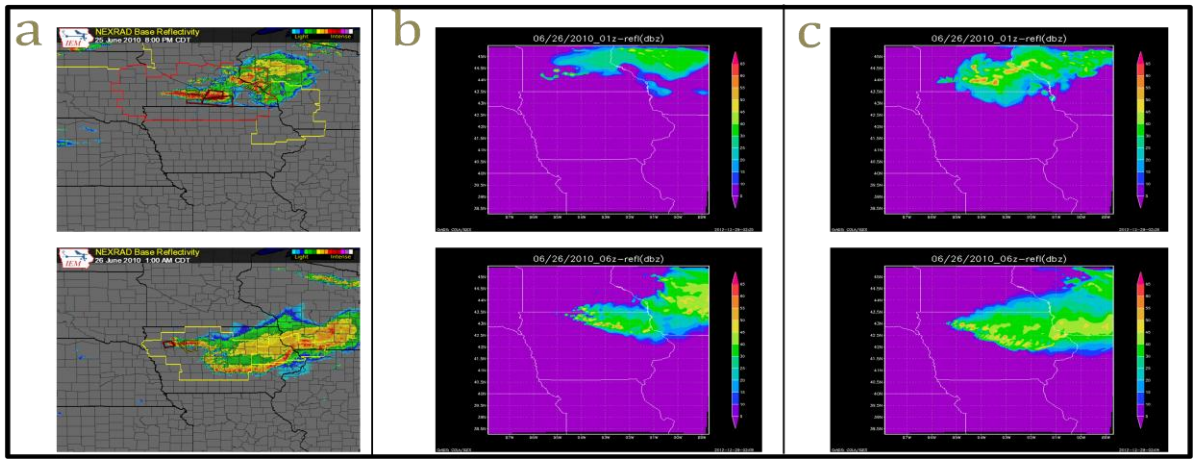

Fig. 2: Plots for June 26, 2010 of a) 2km NEXRAD reflectivity, b) cold start run, and c) radar assimilated run simulated reflectivity at 01z (top) and 06z (bottom)

was much more skillful with rainfall prediction over the first 12 hours compared to the cold-start run; however like in the June 22 case the simulation produces too large an area of rainfall exceeding the 1 inch threshold with the intensity of this rainfall also being a bit high. Unlike the June 22nd case the areal coverage for the rainfall exceeding .01 inch (which could be considered the entire precipitating region) for the radar data assimilated run was very close to that of the observed throughout the first 12 hours of the simulation, also the intensity only being a bit too high was fairly accurate as well through that time. The volume of the rainfall for all grid points exceeding the .01 inch threshold in the model domain was nearly 3,000 (in. x 16 km^2) too high through the first 6 hours of the simulation, then was fairly accurate for the following 6 hour period (06z-12z). Overall, the rainfall characteristics show that the radar data assimilated run gave a fairly accurate forecast through the first 12 hours, with only some minor discrepancies between the forecast and the observed rainfall.

June 26, 2010				
		Forecast Period (UTC)		
Threshold(in.)	**Parameters**	**00-06**	**12-Jun**	**00-12**
0.01	Observed areal coverage (points)	12903	10645	18775
	Observed rain rate (in.)	0.59	0.40	0.63
	Radar assimilated areal coverage (points)	12152	10352	17989
	Radar assimilated rain rate (in.)	0.84	0.46	0.84
1	Observed areal coverage (points)	2183	642	3286
	Observed rain rate (in.)	1.51	1.36	1.59
	Radar assimilated areal coverage (points)	3471	969	4546
	Radar assimilated rain rate (in.)	1.69	2.73	2.01
	Domain rain volume (in. X 16 km^2)			
0.01	Observed rain volume	7615.91	4226.26	11845.47
	Radar assimilated rain volume	10214.74	4799.99	15032.80
1	Observed rain volume	3295.76	875.09	5223.23
	Radar assimilated rain volume	5868.70	2644.84	9146.26

Table 5: Observed and radar data assimilated run areal coverage, rain rate and rain volume for grid points that exceeded the specified rainfall thresholds

ETSs (Table 6) show that the radar data assimilated run was more skillful than the PM forecast for the lighter thresholds, but as in the other case, the PM forecast was slightly better for the heaviest thresholds. The bias values (Table 7) were generally pretty good for the radar data assimilated run, with values near 1 for the lower rainfall thresholds (.01, .1, .5), but larger than 1 for the higher thresholds. The larger bias values make sense for the higher rainfall thresholds as it was already noted from the rainfall characteristics data that the radar data assimilated run over predicted the number of grid points exceeding 1 inch of rainfall.

ETS		June 26, 2010				
Method	Period (UTC)	Threshhold (in.)				
		0.01	0.1	0.5	1	1.5
Cold start	00-06	0.436	0.331	0.132	0.053	0.058
PM		0.516	0.412	0.193	0.168	0.182
Radar		0.736	0.658	0.37	0.327	0.15
Cold start	06-12	0.275	0.224	0.014	-0.011	-0.004
PM		0.447	0.407	.l63	0.143	0.159
Radar		0.433	0.477	0.225	0.053	0.034

Table 6: ETSs for a run that uses radar-data, a cold start run (no radar data), and a PM forecast for rainfall exceeding five different thresholds

bias	June 26, 2010				radar assimilated
Period (UTC)	Threshold (in.)				
	0.01	0.1	0.5	1	1.5
00-06	0.942	1.001	1.361	1.59	2.095
06-12	0.972	0.749	0.751	1.509	3.81
Avg.	0.957	0.875	1.056	1.5495	2.9525

Table 7: bias for the radar data assimilated run for rainfall exceeding five different thresholds

4. Conclusions

The main goal of this project was to quantify the impact of radar data assimilation on summertime heavy rainfall events in and around the state of Iowa. In the two events examined here the accuracy of placement and areal coverage of precipitating regions increased in the radar data assimilated runs over the cold-start runs. This was true for the regions of heavier rainfall (threshold greater than 1 inch) that are of importance for flooding, although it appears the heavy rainfall was actually a little too intense in the radar data assimilated runs. The rainfall volume over the model domain in both cases was also much closer to the observed for the radar data assimilated run compared to the cold start run. The improvements were evident throughout the first 12 hours of the simulation for the June 26th case. For the June 22nd case the radar data assimilation improved the accuracy initially, but not beyond 6 hours due to the poor NAM forecast (errors in placement and intensity of large-scale features noted) used for the first-guess field and boundary conditions. Thus, despite this improvement in the simulated rainfall the accuracy of the external forecast data (NAM) used is very important and as errors in it can greatly hinder the impact of the radar data assimilation on the model simulation. There may be ways to overcome the impact from poor external model data such as using a more intense method of radar data assimilation. For instance assimilating the radar data at more times than just at the model initialization (e.g. every 15 minutes over the first hour of the simulation) in order to further adjust the model conditions toward the radar data, could be a possibility.

Overall, analysis of the simulated rainfall shows that a significant improvement in QPF skill is achieved through radar data assimilation via ARPS 3DVAR, especially in the first 6 hours of the model simulation. The study demonstrates that the extent of the improvement along with how far out in time it might last is case dependent and likely a function of the accuracy of the external model data used as a first guess and for lateral boundary conditions. Future work will look into whether or not the improvement in QPF skill noted in this study is great enough to result in a statistically significant increase in the skill of a hydrology model's stream flow predictions.

5. Acknowledgments

This work was supported by the Iowa Flood Center. Dave Flory is thanked for his help in getting the WRF system up and running and for writing some of the scripts to extract the rainfall data from the WRF output. Kevin Thomas from CAPS at the University of Oklahoma is thanked for his help in getting the necessary ARPS programs running for this project.

6. References

Brewster, K., 1996: Application of a Bratseth analysis scheme including Doppler radar data. Preprints, *15th Conf. on Weather Analysis and Forecasting*, Norfolk, VA, Amer. Meteor. Soc., 92–95.

Ebert, E. E., 2001: Ability of a Poor Man's Ensemble to Predict the Probability and Distribution of Precipitation. *Mon. Wea. Rev.*, **129,** 2461-2480.

Davolio, S., and A. Buzzi, 2004: A nudging scheme for the assimilation of precipitation data into a mesoscale model. *Wea. and Forecasting*, **19,** 855-871.

Hu, M., M. Xue, J. Gao, K. Brewster, 2006: 3DVAR and Cloud Analysis with WSR-88D Level-II Data for the Prediction of the Fort Worth, Texas, Tornadic Thunderstorms. Part II: Impact of Radial Velocity Analysis via 3DVAR. *Mon. Wea. Rev.*, **134**, 699–721.

Janjic, Z. I., 1994: The Step-Mountain Eta Coordinate Model: Further Developments of the Convection, Viscous Sublayer, and Turbulence Closure Schemes. *Mon. Wea. Rev.*, **122**, 927–945.

Macpherson, B., 1999: Operational experience with assimilation of rainfall data in the Met Office Mesoscale model. *Met. Atm. Phys.*, **76,** 3-8.

Schaefer, J. T., 1990: The Critical Success Index as an Indicator of Warning Skill. *Wea. Forecasting*, **5,** 570-575.

Sugimoto, S., D. M. Barker, N. A. Crook, J. Sun, and Q. Xiao, 2009: An examination of WRF 3DVAR data assimilation on its capabilities in retrieving unobserved variables and forecasting precipitation through observing system simulation experiments. *Mon. Wea. Rev.*, **137,** 4011-4029.

Thompson, G., P. R. Field, R. M. Rasmussen, W. D. Hall, 2008: Explicit Forecasts of Winter Precipitation Using an Improved Bulk Microphysics Scheme. Part II: Implementation of a New Snow Parameterization. *Mon. Wea. Rev.,* **136**, 5095–5115.

Xue, M., K. K. Droegemeier, V. Wong, A. Shapiro, and K. Brewster, 1995: *ARPS Version 4.0 User's Guide*. 380pp.[Available online at http://www.caps.ou.edu/ ARPS.].

Xue, M., K. K. Droegemeier, and V. Wong, 2000: The Advanced Regional Prediction System (ARPS)—A multiscale nonhydrostatic atmospheric simulation and prediction tool. Part I: Model dynamics and verification. *Meteor. Atmos. Phys*, **75,** 161–193.

Xue, M., and Coauthors, 2001: The Advanced Regional Prediction System (ARPS)—A multi-scale nonhydrostatic atmospheric simulation and prediction tool. Part II: Model physics and applications. *Meteor. Atmos. Phys*, **76,** 143–166.