Responses to Anonymous Referee #1 (Page 1)

My co-authors and I wish acknowledge and thank Reviewer #1 for the time, energy, and effort applied in the detailed review of this manuscript. We do feel that a more narrow focus on microphysics and removal of the energy norm has improved upon the original manuscript and also address most if not all of the highlighted concerns.

Responses to General Comments:

1) "…my main issue with the paper which is whether we can evaluate microphysics schemes against analyses such as these in a useful way."

Both your comments and those of Reviewer #2 highlight this point. While we do believe that GFS analysis data can be useful for broader themes of our analysis (e.g., large-scale water vapor fields), its coarseness proves problematic was addressing specific microphysical-related questions. The revised manuscript now includes a new analysis making use of the Multi-Radar Multi-Sensor (MRMS) 3D volume data. These observation data, we argue, permit a more thorough investigation of smaller-scale impacts from the microphysics.

2) "Errors in the forecast are dominated by other causes, such as the initial analysis error, considering that these are initialized 72 hours ahead of the precipitation events. Perhaps initializing closer to the event would have given more accurate representations that could be compared with analyses."

In light of your suggestion and a similar comment from Reviewer #2, we shifted the model initialization time forward until 24 hours prior to cyclogenesis off the Mid-Atlantic United States and re-ran all 35 WRF model simulations. We believe that initializing 24 hours prior to cyclogenesis is ideal because it ensures each model simulation is sufficiently spun-up prior to the main cyclogenesis period and yet there are only minimal deviations (< 50 km) between WRF simulations and the GFS model analysis storm tracks.

3) "I especially am not convinced that the energy norm metric has been demonstrated to be useful."

We concur and agree that the energy norm, although useful, is not the most effective vehicle by which to evaluate microphysical-related simulation errors. Thus the energy norm would be more apt in a more general, bulk analysis of nor'easters where a focus on large-scale players are key. Due to our shift in model initialization time (see #2 above) and our shift to focus on microphysics (see #1 above), the energy norm analysis has been redacted from the revised manuscript.

4) "There are also aspects of the model set-up that I would criticize. It seems that the central 1.67 km domain is at the same position for all storms, and this means that some storms pass through it while other would miss it and only be resolved in the 5 km domain"

The WRF model domain positions were fixed for all nor'easter cases. This lead to a situation WRF-simulated nor'easters in cases 1 and 4 either missed or never fully entered the 1.667 km model grid (Domain 4) as the reviewer hypothesized. We have since increased the sizes of the 5 km and 1.167 (Domains 3 and 4, respectively) by 50%, shifted domain 3 southward, and tailored the location of domain 4 for all seven nor'easter events. To physically demonstrate these changes, Figure 1 shows our original and new WRF model configuration. All 35 model simulations were re-run and reanalyzed accordingly. As can be seen below, each model analysis track moves through the center of each respective domain 4.

Responses to Anonymous Referee #1 (Page 2)

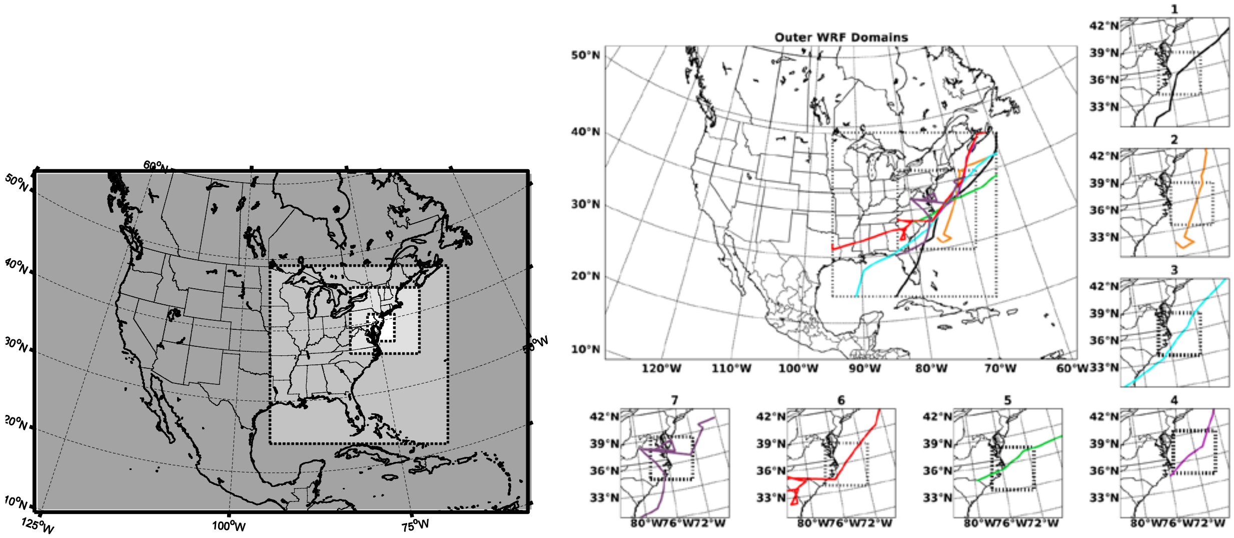

Fig. 1: Nested WRF configuration for the original manuscript (left) and the revised manuscript (right). The colored lines in the right panel show the GFS model analysis storm tracks for each of the seven cases.

Specific Comments:

1. line 141. What are the perturbations relative to, the GMA analysis? This is not stated.

All energy norm calculations are relative to the GFS model analysis. The energy norm section has been removed from the paper.

2. Section 3.2. It is not clear what area these results and Table 4 are for. It also seems that much of this would be in the 12 km domain where there is a cumulus scheme, and part is in domains 3 and 4 where there isn't.

Table 4 was originally based upon domain 2 (15 km domain). The revised manuscript keeps the same approach, but we use domain 3 (5 km grid spacing) instead because it is of similar resolution to the Stage IV precipitation product (4 km resolution), the cumulus parameterization is turned off, and we felt that domain 4 would be over too limited an area for comparison.

3. line 208. WRF's common heritage with GFS is implied. I don't think there is much common physics heritage except for some relationship in the land-surface scheme. What is meant here?

My assumption here was based upon that simulated storm tracks between GFS and WRF would be similar given WRF's common heritage in GFS. Similar tracks would, in theory, give a greater potential of similar forecasts. My comment about this heritage is no longer necessary and it has been removed from the revised manuscript.

Responses to Anonymous Referee #1 (Page 3)

4. Abstract does not mention that there are seven cases and five microphysics schemes and has nothing on the energy norm. It is not adequately describing the work carried out.

Given the significant changes to the manuscript in this revision, the abstract has been updated and overhauled to more aptly describe the work conducted.

5. line 234. What is meant by saturation heights?

Thank you for this asking this clarification. By saturation height, I am referring to the height at which each microphysical species reached its maximum value. This value however is part of the mixing ratio profile and I think distracts from the paper. I have elected to remove this term from the revised manuscript.

6. line 236. cloud water? This should probably be cloud droplet number concentration?

Thank you for finding this error. "Cloud water" has been changed to "cloud droplet number concentration" in the revised manuscript.

7. line 241-246. Without knowing where the freezing level is, it is difficult to follow this discussion. How much of the cloud water is supercooled?

Thank you for noting this challenge to understanding the microphysical species analysis section. To provide information on how much of the cloud water is super cooled, I have modified the composite mixing ratio diagrams with two dashed black lines which indicate both the 0°C and -40°C levels.

8. line 279. How does lack of a sedimentation term lead to low cloud ice? I thought sedimentation should reduce cloud ice extent and lifetime.

Thank you for the noting this logic error. A quick read into the literature found a cloud resolving model study addressing this very topic. Their findings do indeed show that the impact of the sedimentation in cloud ice is to increase its conversion rate to snow and graupel and thus decreasing the amount, extent, and lifetime of cloud ice hydrometeors. I have removed the erroneous comment from the revised manuscript.

Nomura, M., Tsuboki, K. and Shinoda, T., 2012. Impact of Sedimentation of Cloud Ice on Cloud-Top Height and Precipitation Intensity of Precipitation Systems Simulated by a Cloud-Resolving Model. *気象集誌. 第 2 輯*, *90*(5), pp.791-806.

9. line 282. 'assumed water saturation'. What assumption is made about water saturation in a purely ice process?

The original GCE6 scheme generated excess super cooled cloud water at temperature below -12°C where such droplets do not often occur. Therefore water saturation was extended down to much colder temperatures which allowed cloud ice to achieve supersaturation with respect to ice and made cloud ice to snow conversion rates

For further details please refer to page 2308 of the following reference:

Lang, S. E., Tao, W. -K., Zeng, X., *and* Li, Y.: Reducing the biases in simulated radar reflectivities from a bulk microphysics scheme: Tropical convective systems, J. Atmos. Sci., 68, 2306–2320, *2011*.

Responses to Anonymous Referee #1 (Page 4)

10. Figure 7 (vapor) would have been better presented as a difference from analysis. Nothing can be seen with this plot as it is.

Thank you for the suggestion. In new manuscript, this diagram (now Figure 2) has been updated to show the difference in water vapor.

11. Section 3.4. It is hard to interpret what is meant by lowest energy norms and the metrics in Table 5 in general. Also make clearer what is meant by model-relative and GMA-relative norms.

"GMA-relative" denotes diagnosing the simulated environment within a 600-km wide box centered on the GMA-indicated cyclone center in both GMA and each WRF simulation. "Model-relative" uses the same box, but centers it on the cyclone center determined from each individual model simulation. The energy norm analysis is no longer part of the manuscript.

12. As mentioned in the general comments, I do not think the energy norm statistics are adding anything useful to the paper. It would be better and more focused without this. There are so many factors that could make one simulation look temporarily better than another, related to timing and structure developments, that using such a high-level bulk measure as this conflates too many things to be useful in such an intercomparison.

While we do see some value in the energy norm results with respect to diagnosing which dynamical fields are responsible for observed error, we agree that in context of a microphysics- focused paper this metric is not sensitive enough to be of use. Pending the suggestion of both reviewers, this section has been redacted from the revised manuscript.

13. line 334. Regarding the low-level jet which case is being referred to? Can it really be inferred from the v component of the energy norm that this jet is the cause? This looks highly speculative.

We agree with the reviewer's viewpoint that the energy norm by itself could be considered speculative for Case 7. Our decision to not include a figure of 850-hPa winds (See Figure 2 below) in the original manuscript was made on the assumption that presence of the cyclone center, the small size of the model domain, and a bump in the u and v energy norm components at 850-hPa would be sufficient circumstantial evidence to support our claim without the need for an additional figure. In the revised manuscript, the energy norm section has been removed from the paper.

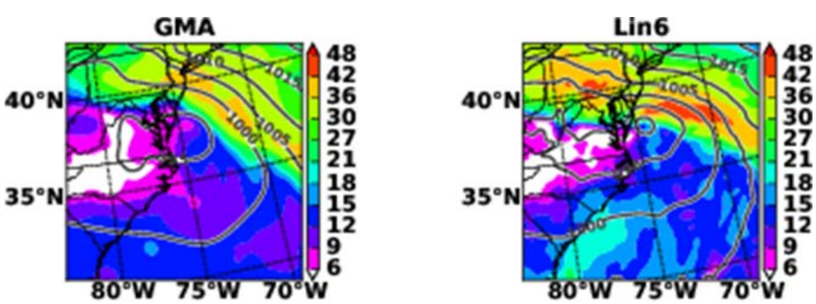

Fig. 2: 850-hPa wind speed (fills, m s-1) and sea-level pressure (contours, hPa) on 13 March 2010 at 18 UTC (Case 7).

Responses to Anonymous Referee #2 (Page 1)

My co-authors and I wish to thank Reviewer #2 for their time and consideration in reviewing this manuscript. Many comments are consistent with those of Reviewer #1 and have been incorporated into the revised manuscript.

General Comments

1) *I think that the spin-up time of 72 hours is too long for a simulation without any kind of assimilation. A test with a shorter spin up (12 hours) could be recommendable*

In light of your suggestion and a similar comment from Reviewer #1, we shifted the model initialization time forward until 24 hours prior to cyclogenesis off the Mid-Atlantic United States and re-ran all 35 WRF model simulations. We set our start time 24 hours beforehand because simulated radar reflectivity fields still appeared slightly "blooby" up through 9-10 hours. Starting the model simulations 24 hours before primary cyclogenesis allowed for full development of simulated radar reflectivity structures and WRF-GMA track differences tended to be modest (<50 km).

2) *"A microphysical comparison with observations could be useful because this topic is the main focus of the paper. Is it possible to retrieve data from radar or satellite platform"*

Thanks to your suggestion, we have given this revised paper more of a microphysics-style focus. I looked both into TRMM and CloudSat 2C-Ice products. TRMM offers a wide range radar observations but its orbital inclination is 35 degree (http://disc.sci.gsfc.nasa.gov/precipitation/additional/instruments/trmm_instr.html), which limits its usefulness when only half my analysis domains falls equatorward of 35°N. CloudSAT does provide profiles cloud ice, which my colleague used in a recent paper on global cloud species. It narrow swath range (see Figure 3) made getting a consistent "hit" on a nor'easter challenging.

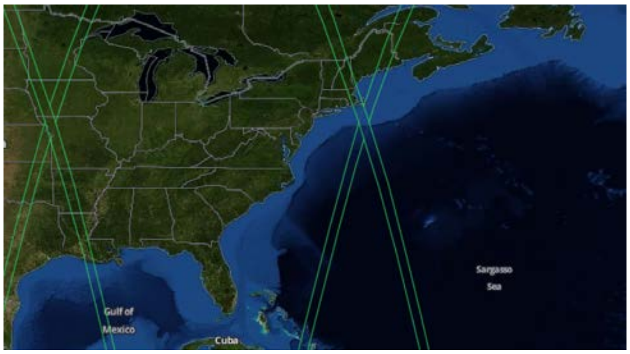

Fig. 3: CloudSAT orbital overpass sample from 2012.

I did find success with the Multi-Radar Multi-Sensor product from National Oceanagrahic and Atmospheric Association (NOAA), which provides hourly gridded 3D volume scans at 1-hour intervals (See Figure 4). Similar to StageIV, MRMS data only covers part of domain 4 in many of the seven cases, but the results thus far have been reasonable and useful.

Responses to Anonymous Referee #2 (Page 2)

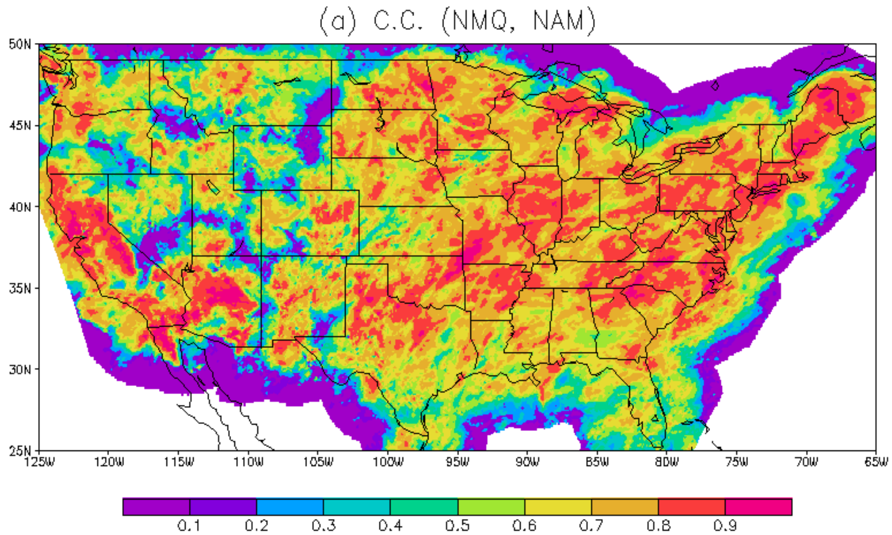

Fig. 4: MRMS coverage area (everywhere with colors).

Specific Comments:

1) Line 133: w is the mixing ratio of rain?

Although ‘w’ is often used in meteorology to denote mixing ratio, it represents vertical velocity in the energy norm equation. Instead, this formula uses ‘q’ to represent mixing ratio. With the removal of the energy norm from the paper’s results this particular comment is no longer valid.

2) Line 203: Not Fig. 4 but Fig. 5

Thank you for catching the typo. I have corrected the manuscript to refer to Fig. 5.

3) Figs. 5-6-7: insert letters in the panel to easy the reading of section 3.

While I will not dispute that Figs. 5-7 do attempt to show much data. In an earlier form of this paper, I actually tried putting letters into the panels, but these letters were difficult to place without blocking or interfering with the displayed data. I thank you for the suggestion, but I have decided to keep my “Microsoft Excel-like” approach to plot labelling.

Responses to Editor (Page 1)

Note to editor: Apologies for taking so long with these technical corrections, the end of the year is always a busy time and my co-authors were hard to pin down. While having a paper returned with "major revisions" is never a good feeling, we were glad to have the extra time because we were able to make your suggested corrections, do many other tweaks to the paper (tables, figures, wording, etc.), and even do a little more analysis. We hope the end product meets your satisfaction and that our efforts here (with luck) will be viewed favorably by the reviewers. A summary of the highlighted corrections and our response is below

Editor highlighted corrections (Corrections or changes are noted in bold):

1) *l158: ITS mixing ratio*

   Original: We exclude hail from our analysis because it is unique to GCE7 and it mixing ratio values are an order of magnitude smaller than other species.

   Modified: We exclude hail from our analysis because it is unique to GCE7 and **its** mixing ratio values are an order of magnitude smaller than other species.

2) *l168: will BE explained*

   Original: …is due to identifiable trends within the underlying assumptions made by BMPSs and will explained in more detail below

   Modified: …is due to identifiable trends within the underlying assumptions made by BMPSs and will **be** explained in more detail below

3) *l207: ITS*
   Original: GCE7 is in many ways at opposition to Lin6, where it simulations generate the most snow, …
   Modified: GCE7 is in many ways at opposition to Lin6, where **its** simulations generate the most snow, …

4) *l235: turned on*

   Original: WRF precipitation is generated from its microphysics and cumulus parameterization; the latter is turned for Domains 3 (5 km grid spacing) and 4 (1.667-km grid spacing)

   Modified: WRF precipitation is generated from its microphysics and cumulus parameterization; the latter is turned **on** for Domains 3 (5 km grid spacing) and 4 (1.667-km grid spacing)

5) l255: IN

   Original: As illustrated Figs. 6 and 7, all WRF simulations tended to generate similar coverage to Stage IV, …

   Modified: As illustrated **in** Figs. 6 and 7, all WRF simulations tended to generate similar coverage to Stage IV, …

Responses to Editor (Page 2)

6) *l272-273: ??*

No direct comment is given. I am assuming that the statement or phrasing here is confusing. In the revised manuscript, I no longer reference all the CFAD and CFAD-related figures in the first paragraph. Instead, I now describe the CFAD plot (Fig. 8) generally and then introduce the follow-on figures (Figs. 9-11) as needed. For reference the 3,000 m and 9,000 m heights were selected because they represented levels were the BMPSs varied (3,000 m) or where errors with the MRMS data product filtering (9,000 m) could be highlighted.

7) l305: ??

Similar to 6, I am assuming this line is confusing and too long. I have revised the sentence and boiled it down to more exact details.

Original: WRF-Stage IV accumulated precipitation comparisons reveal WRF demonstrate that although WRF generates precipitation fields of similar coverage to Stage IV precipitation intensities tended to be higher than observations and resulting in low to moderate (0.217–0.414) threat scores with WDM6 demonstrating marginally better forecast skill than its single-moment counterparts.

Modified: WRF-simulated precipitation fields exhibit similar coverage but trended towards higher precipitation amounts relative to Stage IV observations resulting in low-to-moderate threat scores (0.217–0.414).

8) l310:

Similar to 6 and 7, I am assuming this section of the manuscript was confusing or unclear. In the latest revision of the results, we were able to pull out more details about the Lin6 vs GCE7 comparison. By focusing more on GCE7 and Lin6 in our description, we think this helps make the conclusion here clearer to the reader.

Original: Finally, MRMS-based CFAD and CFAD scores show Lin6 and GCE7 to be notably better than GCE6, WSM6 and WDM6 in the lower troposphere, with GCE7 being the only BMPS scheme to produce the narrow core of maximum frequencies below10 dBZ due to its temperature and mixing ratio dependent aggregation and new snow map. Above 5,000 m GCE7 however becomes less skilled the combination of smaller hydrometers and entrainment reduced it cloud top height relative to other BMPSs.

Modified: Finally, MRMS-based contoured frequency with altitude diagrams (CFADs) and CFAD scores show Lin6 and GCE7 to perform the best in the lower half of the troposphere (below 6,300 m AMSL), where GCE7 most realistically reproduced the maximum frequency core between 5 and 15 dBZ due to its temperature and mixing ratio dependent aggregation and new snow size mapping. However, the overly large growth of graupel via its dry collection of snow suggests that Lin6 obtains high CFAD scores from a less realistic solution than GCE7. Above 6,300 m AMSL, model-simulated cloud tops are much more susceptible to entrainment and become more sporadic; this in conjunction with the non-precipitating echo filtering in the MRMS data makes evaluations less meaningful with increasing height.

Responses to Editor (Page 3)

9) Check for expressions like "cold temperature". The air is cold. A temperature is not cold, it is low!!

   - Thank you for pointing out this grammatical error. I have scanned through the paper and changed all “cold” temperatures to “low” temperatures. I along with the 4^{th} author also vetted the paper for any other similar logic errors and fixed them.

10) Unfortunately, it happens that scientists read only the abstract and conclusions. To help those readers, you may write the conclusion in a self-contained fashion. I simply mean, that you could again introduce the abbreviations you us, include a link to Table 2. You may also describe in a few more words what GMA, Stage IV precipitation and MRMS is. Is it model or measurement data?

    - Thank you for the advice and suggestions about the abstract and conclusions. While it would be ideal to think a reader would read the whole paper, it probably does not happen most of the time. I have adjusted both the abstract and conclusion to be more self-contained and descriptive as per your suggestion.

Influence of Bulk Microphysics Schemes upon Weather Research and Forecasting (WRF) Version 3.6.1 Nor'easter Simulations

Stephen D. Nicholls[1,2], Steven G. Decker[3], Wei-Kuo Tao[1], Stephen E. Lang[1,4], Jainn J. Shi[1,5], and Karen I. Mohr[1]

[1]*NASA-Goddard Space Flight Center, Greenbelt, 20716, United States of America*
[2]*Joint Center for Earth Systems Technology, Baltimore, NASA-Goddard Space Flight Center, Baltimore, 21250, United States of America*
3Department of Environmental Sciences, Rutgers, The State University of New Jersey, 08850, United States of America
[4]*Science Systems and Applications, Inc., Lanham, 20706, United States of America*
[5]*Goddard Earth Sciences Technology and Research, Morgan State University, 21251, United States of America*

Correspondence to: Stephen D. Nicholls (stephen.d.nicholls@nasa.gov)

Abstract. This study evaluated the impact of five, single- or double- moment bulk microphysics schemes (BMPS) on Weather Research and Forecasting (WRF, version 3.6.1) model simulations of seven, intense winter time cyclones events impacting the Mid-Atlantic United States. Five-day long WRF simulations were initialized roughly 24 hours prior to the onset of coastal cyclogenesis off the coast of North Carolina coastline. In all, 35 model simulations (5 BMPSs and seven cases) were run and their associated Validation efforts focus on microphysics-related storm properties (including hydrometeor mixing ratios, precipitation, and radar reflectivity) were evaluated against by comparing model output to model analysis and available gridded radar and ground-based precipitationrainfall products across 35 WRF model simulations (5 BMPSs and seven cases). Inter-BMPS Ccomparisons of column-integrated mixing ratios and mixing ratio profiles revealed little variability in non-frozen hydrometeor species due to their sharedcommon programming heritage, yet their assumptions concerningabout snow and graupel intercepts, ice supersaturation, snow and graupel density maps, and terminal velocities lead to considerable variability in both simulated frozen hydrometeor species and in turn radar reflectivitiesyies. WRF-simulated precipitation fields exhibit minor spatio-temporal variability amongst BMPSs, yet their spatial extent is largely conserved. Compared to ground-based precipitation data, WRF-simulations demonstrate low-to-moderate (0.217–0.414) threat scores and a rainfall distribution shifted toward higher values.WRF model simulations were found to produce similar precipitation coverage, but simulations favored excessively high precipitation amounts compared to observations and low to moderate (0.217–0.414) threat scores. Finally, an analysis of WRF and gridded radar reflectivity data via comparison of contoured frequency with altitude (CFAD) diagramsplots between WRF and gridded observed radar reflectivity fields yielded reveals notable variabilitytions amongstbetween BMPSs, where better performing schemes favored with schemes favoring lower graupel mixing ratios and better underlying and better aggregation assumptions compared more favorably to observations.

1 Introduction

Bulk microphysical parameterization schemes (BMPSs), within numerical modern weather prediction models (e.g., Weather Research and Forecasting model [WRF; Skamarock et al., 2008]), have become increasingly complex and computationally expensive. Presently, WRF Modern prognostic weather models, such as the Weather Research and Forecasting (WRF) model (Skamarock et al., 2008), offers BMPS options varyingranging from simplistic, warm rain physics (Kessler, 1969) to multi-phasecomplex, six-class, two-moment microphysics (Morrison et al., 2009). Microphysics and cumulus parameterizations drive cloud and precipitation processes within WRF and similar models numerical weather prediction models and which has consequences for directly or indirectly impacts radiation, moisture, aerosols, and other simulated meteorologicalsimulated processes. Tao et al. (2011) highlighted the importance of BMPSs in modelsCiting its importance, Tao et al. (2011) by summarizing detailed more than 36 published, microphysics-focused studies ranging focusing on from idealized simulations to, hurricanes to, or mid-latitude convection. More recently, the observation-based al studies of Stark (2012) and Ganetis and Colle (2015) investigated microphysical species variability within United States (U.S.) east coast winter-time cyclones (locally called "nor'easters") and have called for further investigation into how BMPSs impact these cyclones which motivates this nor'easter study.~~have called for further studies investigating how microphysical parameterizations impact simulations of these powerful cyclones~~.

A "nor'easter" is a large (~2000 km), mid-latitude cyclone occurring from October to April and is capable of bringing punishing winds, copious precipitation, and potential coastal flooding to the Northeastern U.S. (Kocin and Uccellini 2004; Jacobs et al., 2005; Ashton et al., 2008). This region is home to over 65 million people and produces 16 billion U.S. dollars of daily economic output (Morath, 2016). Given its high economic output, nor'easter-related damages and disruptions can be extreme. Just ten strong, December nor'easters, between 1980 and 2011, produced 29.3 billion U.S. dollars in associated damages (Smith and Katz, 2013). ~~BMPSs are key to accurate simulations of a nor'easter's precipitation and microphysical properties and will be the focus of this study.~~

Recent nor'easter studies are scarce given the extensive research efforts of~~in~~ the 1980s. These historical studies addressed key environmental ~~nor'easter~~ drivers including frontogenesis and baroclinicity (Bosart, 1981; Forbes et al., 1987; Stauffer and Warner, 1987), anticyclones (Uccelini and Kocin, 1987), latent heat release (Uccelini et al., 1987), and moisture transport by the low-level jet (Uccellini and Kocin, 1987; Mailhot and Chouinard, 1989). Despite extensive observational analyses, little ~~ess~~ attention has been given to role of BMPSs in mid-latitude winter cyclones. ~~has been provided to mid-latitude, winter cyclone simulations, especially those focused on BMPSs~~.

Reisner et al. (1998) ran several ~~several single and double-moment BMPS~~ Mesoscale Model Version 5 winter storm simulations, with multiple BMPS options, ~~of winter storms~~ that impacted~~ing~~ the Colorado Front Range during~~for~~ the Winter Icing and Storms Project. Double moment-based simulations produced more accurate simulations of supercooled water and ice mixing ratios than those originating from single-moment schemes. However, single~~-~~moment-based ~~-~~simulations vastly improved when the snow-size distribution intercepts were derived from a diagnostic equation rather than from a fixed value.

Wu and Pretty (2010) investigated how five, six-class BMPSs affected WRF simulations of four polar-low events (two over Japan, two over the Nordic Sea). Their simulations yielded nearly identical storm tracks, but notable cloud top temperature and precipitation errors. Overall, the WRF single-moment BMPS (Hong and Lim, 2006) produced marginally better cloud and precipitation process simulations than ~~compared to~~ those from other BMPSs. For warmer, tropical cyclones, Tao et al. (2011) investigated how four, six-class BMPSs impacted WRF simulations of Hurricane Katrina. They found BMPS choice minimally impacted storm track, yet sea-level pressure (SLP) varied up to 50 hPa.

Shi et al. (2010) evaluated several WRF single-moment BMPSs during a lake-effect snow event. Simulated radar reflectively and cloud top temperature validation revealed that WRF accurately simulated the onset, termination, cloud cover, and band extent of a lake-effect snow event, however snowfall totals at fixed points were less accurate due to interpolation of the mesoscale grid. Inter-BMPS simulation differences were small ~~They found BMPSs produced only minimal simulation differences~~ because low~~cold~~ temperatures and weak vertical velocities prevented graupel generation. Reeves and Dawson (2013) investigated WRF sensitivity to eight BMPSs during a December 2009 lake-effect snow event. Simulated ~~Their study found~~ precipitation rates and snowfall ~~-~~coverage were particularly sensitive to BMPSs because vertical velocities exceeded hydrometeor terminal fall speeds in ~~in~~ half of their simulations. Vertical velocity differences were attributed to varying BMPS frozen hydrometeor assumptions concerning snow density values, temperature-dependent snow-intercepts, and graupel generation terms.

~~Similar to previous studies, w~~We will evaluate WRF nor'easter~~winter storm~~ simulations and their sensitivity to six- and seven-class BMPSs with a focus ~~, but our primary focus will be~~ on microphysical properties and precipitation. The remainder of this paper is divided into three sections. Section 2 explains the methodology and analysis methods. Section 3 shows the results. Finally section 4 describes the conclusions, its implications, and prospects for future research.

2 Methods

2.1 Study design

We utilized WRF version 3.6.1 (hereafter W361) which solves a set of fully-compressible, non-hydrostatic, Eulerian equations in terrain-following coordinates (Skamarock et al., 2008). Figure 1 shows the four-domain WRF ~~model~~ grid configuration for this study with a 45-, 15-, 5-, and 1.667-km horizontal grid spacing, respectively. ~~used for this study~~. Additionally, this configuration includes ~~This grid also has~~ 61 vertical levels, a 50-hPa (~20 km) model top, two-way domain feedback, and ~~turns off~~ cumulus parametrization is turned off for Domains 3 and 4 which are convection permitting. Notably, the location of Domain 4 adjusts for each case (Fig. 1).~~in Domains 3 and 4. The fourth domain is convection-resolving and moves fior each simulation set (Fig. 1)~~. Global Forecasting System model operational analysis (GMA) data was used for WRF boundary conditions. The above model configuration (except for the 4th domain)~~This model configuration (except the 4th domain)~~ and ~~the below~~ parameterizations are derived from ~~identical to those in~~ Nicholls and Decker (2015). ~~and are consistent with past and present WRF model studies at NASA Goddard Space Flight Center (i.e., Shi et al., 2010; Tao et al. 2011)~~. Model parameterizations include:

- Longwave radiation: New Goddard Scheme (Chou and Suarez, 1999; Chou and Suarez, 2001)
- Shortwave radiation: New Goddard Scheme (Chou and Suarez, 1999)
- Surface layer: Eta similarity (Monin and Obukhov, 1954; Janjic, 2002)
- Land surface: NOAH (Chen and Dudhia, 2001)
- Boundary layer: Mellor-Yamada-Janjic (Mellor and Yamada 1982; Janjic 2002)
- Cumulus parameterization: Kain-Fritsch (Kain, 2004) ~~(Not applied to domains 3 and 4)~~

This study investigates the seven nor'easter cases described in Table 1 and shown in Fig. 1. These cases are identical to those in Nicholls and Decker (2015) and represent a small, diverse sample of nor'easter events of varying intensity and seasonal timing. ~~This study investigates the same, diverse, selectively chosen sample of seven nor'easter cases from Nicholls and Decker (2015) detailed in Table 1 and storm tracks are shown in Fig. 1. The seven, nor'easter cases in Table 1 include at least one event per month (October-March) and are sorted by month rather than chronological order.~~ In Table 1, the Northeast Snowfall Impact Scale (NESIS) value serves as proxy for storm severity (1 =~~is~~ notable, ~~and~~ 5 = extreme) and is based upon storm duration, ~~its value depends upon the~~ population impacted, area affected, and snowfall severity (Kocin and Uccellini, 2004). Early and late season storms (Cases 1, 2, and 7) did not have snow and thus lack~~do not have~~ a NESIS rating.

Five-day, WRF model simulations for this study -were initialized 24 hours prior to the first precipitation impacts in the highly populated Mid-Atlantic region and prior to the onset of rapid, coastal cyclogenesis off the North Carolina coastline. This starting point provides sufficient time A 24 hour lead time provides sufficient time for WRF to establish fully develop mesoscale circulations and atmospheric vertical structure (Kleczek et al., 2014), and also to establish key surface baroclinic zones, and sensible and latent heat fluxes (Bosart, 1981; Uccelini and Kocin, 1987; Kuo et al., 1991; Mote et al., 1997; Kocin and Uccellini, 2004; Yao et al., 2008, Kleczek et al., 2014). We define the first precipitation impact time as the first 0.5 mm (~0.02 inch) precipitation reading from the New Jersey Weather and Climate Network (D. A. Robinson, pre-print, 2005) associated with a nor'easter event. A smaller threshold wasis not used to avoid capturing isolated showers occurring well ahead of the primary precipitation shield.

To investigate BMPS influence upon W361 nor'easter simulations, five BMPS are used (Table 2). These BMPSs include As shown in Table 2, the selected schemes include three, six-class, three-ice, single-moment schemes (-Lin [(Lin6; Lin et al., 1983; Rutledge and Hobbs, 1984]), Goddard Cumulus Ensemble [(GCE6; Tao et al., 1989; Lang et al., 2007], and WRF single moment [(WSM6; Hong and Lim 2006]), a seven-class, four-ice, single-moment Goddard Cumulus Ensemble scheme (GCE7; Lang et al. 2014), and finally, the a six-class, three-ice, WRF double-moment scheme (WRF double-moment, six class (WDM6; Lim and Hong 2010)). In total, 35 model simulations were completed (7 nor'easters x 5 BMPSs).For this study, we ran 35 W361 simulations covering five BMPS and seven nor'easter cases.

2.2 EvaluationVerification and analysis techniques

Model evaluation validation and analysis efforts involved comparing WRF model output to focused on comparisons of WRF to GMA, Stage IV precipitation (StIV; Fulton et al. 1998; Y. Lin and K.E. Mitchell, preprints, 2005), and Multi-Radar, Multi-Sensor (MRMS) 3D volume radar reflectivity (Zhang et al. 2016). GMA offers six-hourly, gridded dynamical fields, including water vapor, with global coverage. Stage IV is a six-hourly, 4-km resolution, gridded, combined radar and rain gauge precipitation product covering the United States and is derived from rain gauge and radar data. Finally, MRMS is two minute, 1.3-km resolution, gridded 3D volume radar mosaic product derived from S- and C-band radars covering the United States and Southern Canada (Zhang et al. 2016) and is the . MRMS serves as an the operational successor to the the better known National Mosaic and Multi-Sensor QPE (NMQ; Zhang et al. 2011) product radar mosaic products. Both Stage IV and MRMS, however are limited by the detection range of their surface-based assets. All cross comparisons between WRF and these evaluationvalidation data were were conducted at identical grid resolution.

Analysis of WRF model microphysical, precipitation, and simulated radar output was comprised of three main parts: precipitable mixing ratios and domain-averaged mixing ratio profiles, simulated precipitation, and simulated radar reflectivity. Precipitable mixing ratios are calculated for all six microphysical species (vapor, cloud ice, cloud water, snow, rain, and graupel) using the equation for precipitable water:

$$PMR = \frac{1}{\rho g}\int_{P_{top}}^{P_{sfc}} w\, dp \qquad (1)$$

In Eq. (1), *PMR* is the precipitable mixing ratio in mm, ρ is the density of water (1,000 kg m^{-3}); *g* is the gravitational constant (9.8 m s^{-2}); p_{sfc} is the surface pressure (Pa), p_{top} is the model top pressure (Pa); *w* is the mixing ratio (kg kg^{-1}); *dp* is the change in atmospheric pressure between model levels (Pa). We only evaluate water vapor PMR's in this study because all other GMA mixing ratio species are nonexistent Only water vapor can be validated because the other species are nonexistent in GMA and ground and space validation microphysical data are lacking, especially over the data-poor North Atlantic (Li et al., 2008; Lebsock and Su, 2014). Similarly, mixing ratio profiles will only be inter-compared amongst BMPSs because satellite-derived cloud ice profile products (e.g., CloudSat 2C-ICE; Deng et al. 2013) do not directly overpass, have a narrow scan width (1.3–1.7 km) and do not have direct overpass of Domain 4 during coastal cyclogenesis for any case. WRF-simulated precipitation fields and their distribution were evaluated against StIV and simulation error was quantified via qualitatively compared to Stage IV data and then evaluated with bias and threat score (critical success index; Wilks, 2011) values. Finally, contoured frequency with altitude diagrams (CFADs) were used to validate WRF-simulated radar reflectivity relative to MRMS similar to will validate WRF against observed MRMS data as in the similar radar validation efforts of Yuter and Houze (1995), Lang et al. (2011) and Lang et al. (2014). A CFAD offers the advantage of preserving frequency distribution information, yet is insensitive to spatio-temporal errors to both spatial and temporal mismatches. Additionally, CFAD-based scores were will also be calculated for each height level and with time at each height level and evaluated with time using Eq (2).

$$CS = 1 - \frac{\sum |PDF_m - PDF_o|_h}{200} \quad (2)$$

In (2), *CS* is the CFAD score and PDF_m and PDF_o (%) are the probability density functions (PDF) at constant height from WRF or the modeland MRMS-simulated and observed radar reflectivity, respectively. The CFAD score ranges between 0 (no PDF overlap) to 1 (identical PDFs).

3. Results

3.1 Hydrometeor species analysis

Figure 2 displays six classes (water vapor, cloud water, graupel, cloud ice, rain, and snow) of precipitable mixing ratios (mm) from each WRF simulation and GMA for six microphysics species (water vapor, cloud water, graupel, cloud ice, rain, and snow) and Fig. 3 shows corresponding simulated radar reflectivity (no MRMS on this date) at 4,000 m above mean sea level (AMSL) from Case 5, Domain 4 at 06 UTC February 2010. We chose this time and height because storm track errors are negligible, the cyclone is centralized within Domain 4, and mixing ratio profiles at this time (Fig. 4) show all hydrometeor species to coincide at 4,000 m AMSL and that snow and graupel mixing ratios approach their maximum values at this height. Figure 5, shows the Composite seven-case composite mixing ratios derived from hourly data during the residence time each nor'easter case in Domain 4 (24-30 hours). This composite illustrates that mixing ratio profiles profiles are averaged over the residence time of the nor'easter within Domain 4 (24-30 hours).largely preserve their shape, maximum mixing ratio heights, and mixing ratio tendencies (i.e., higher snow mixing ratios in GCE6 and GCE7), but hourly mixing ratio values themselves can vary up to 3.5x's higher (QRAIN; WDM6) at a given height than in the seven case composite (Fig. 5). Corresponding simulated radar reflectivity (dBZ) at 4,000 m is shown as Fig. 3. This case and time was selected for its negligible storm track error, centralized location in Domain 4, and expansive radar reflectivity coverage at 4,000 m where hydrometeor mixing ratios are high. Notably, MRMS are currently not available for this date. To supplement these data, Figs. 4 and 5 depict composite mixing ratios, temperature, and vertical velocity profiles for Case 5 (Fig. 4) and over all seven cases (Fig. 5) from Domain 4. Composite profiles are averaged over the residence time of the nor'easter within Domain 4 (24-30 hours). Figures 4 and 5 also contain two black dashed lines denoting the 0°C and -40°C heights, which denote the region where super-cooled water may occur. Although both the super-cooled water fraction and these temperature heights vary hourly, the latter demonstrates little to no inter-BMPS variability. To emphasize the fraction of supercooled water, two sets of dashed black lines are added to each panel in Figs. 4 and 5 to indicate the 0°C and -40°C heights from each model simulation. We exclude hail from our analysis because it is unique to GCE7 and it mixing ratio values are an order of magnitude smaller than other species.

Comparing Figs. 2 and 3 Comparing Figs. 2 and 4 to Fig. 3, reveals a strong correspondence between radar reflectivity signatures at 4,000 m AMSL and particular precipitable hydrometeor species structures, especially rain, graupel, and snow. and to a lesser extent cloud water. As seen in Fig. 4, all cloud water and rain above 3,500 m AMSL is super-cooled. Stronger nor'easter-related convection (reflectivity > 35 dBZ) in Fig. 3 best corresponds to precipitable rain and then graupel (Fig. 2) despite the near non-existence of the former at 4,000 m AMSL (Fig. 4). This apparent discrepancy suggests localized enhancement of rain mixing ratios where stronger vertical velocities near convection likely drive the freezing level higher than Fig. 4 indicates. Within the broader precipitation shield (20-35 dBZ), radar reflectivity patterns best correspond to precipitable snow and then precipitable graupel (Fig. 2) for all BMPSs except for Lin6 where this trend is reversed. Although Fig. 4 shows all five BMPS loosely agree on amount and height of maximum graupel at 4,000 m AMSL, Lin6 has little to any snow at this level which likely explains the trend reversal. Analysis of Fig. 4 reveals that cloud water at 4,000 m is super-cooled and graupel mixing ratios values are near their peak and given the corresponding precipitation mixing ration values in Fig. 2, these two species are well correlated with the strongest, convective reflectivity signatures (> 35 dBZ). Fig. 4 also reveals snow mixing ratio, except for Lin6 are also comparatively high at this level, yet precipitable snowfall values better correlate best with moderate reflectivity (20-35 dBZ) regions within the broader, more stratiform, precipitation shield. Notably, for Lin6, reduced snow mixing ratios are partially offset by an increase of graupel mixing ratio values within the precipitation shield. Inter-BMPS mixing ratio variability amongst BMPSs, both at this level and throughout the troposphere, is due to identifiable trends within the underlying assumptions made by BMPSs and will be explained in more detail below.

All evaluated BMPSs share a common heritage with the Lin schemein the Lin6 scheme (Note: Lin6 is a modified form of the original Lin scheme). Amongst the BMPSs, only WDM6 explicitly forecasts cloud condensation nuclei, rain, and cloud water number concentrations, the remaining schemes apply derivative equations for these quantities (Hong et al., 2010). Aside from the above, all five BMPS differ primarily in their treatment of frozen hydrometeors which is most evident from the nearly identical (exception: WDM6) rain mixing ratio profiles (Figs. 4 and 5) and precipitable water vapor (Fig. 2) and is a result consistent with Wu and Petty (2010). Comparing WSM6 to WDM6 reveals the second moment has little to no effect on precipitable rain coverage area (Fig. 2) yet, precipitable rain is enhanced (Fig.2) and rain mixing ratios drop sharply near the surface.

With the exception of the two-moment cloud water and rain and CCN-cloud droplet feedbacks in WDM6, the BMPSs differ primarily in how each addresses frozen hydrometeor species (cloud ice, graupel, and snow). Their common programming heritage is evident from the nearly identical (exception: WDM6) rain mixing ratio profiles (Figs. 4 and 5) and precipitable water vapor (Fig. 2) and is consistent with Wu and Petty (2010). WDM6, unlike single-moment BMPSs, explicitly forecasts CCN, rain and cloud droplet number concentrations and does not apply derivative equations (Hong et al., 2010). The forecasts result produce minimal changes to maximum mixing ratio height (Figs. 4 and 5) and precipitable rain coverage (Fig. 2), yet rain mixing ratios remain higher aloft and decrease sharply towards the surface unlike in single-moment simulations.

Similar to rain, precipitable cloud water extent (Fig. 2) and maximum cloud water height (Figs. 4 and 5) barely change, yet mixing ratio amounts (Figs. 2, 4, 5) did vary amongst the BMPSs. These cloud water mixing ratio differences are likely associated with both varying ice supersaturation allowances as described for the Goddard schemes by Chern et al. (2016) and for the WRF schemes by Hong et al. (2010) and assumed cloud water number concentrations (300 cm^{-3} for WSM6). Although WDM6 borrows much of its source code from WSM6, forecasts of cloud condensation nuclei and cloud water number concentrations alter inter-hydrometeor species interactions which in turn alter cloud water mixing ratios (Hong et al. 2010). The similarly between WSM6 and WDM6 in Figs. 2-4 indicate that forecasted cloud number concentrations for Case 5 are likely close to the 300 cm^{-3} value assumed by WSM6. For the other cases, cloud water mixing ratios did vary between WSM6 and WDM6 indicating that WDM6 cloud water number concentrations did stray from 300 cm^{-3} and therefore cause the apparent differences in composite cloud water mixing ratios (Fig. 5).

Similar to rain mixing ratios, cloud water mixing ratios exhibit little variability in either the precipitable cloud water extent (Fig. 6) or the maximum mixing ratio height and freezing level (Fig. 7), but maximum mixing ratio values vary even between single-moment BMPSs. Differing allowances in the amount of ice supersaturation between GCE7 (Chern et al. 2016) and WSM6 (Hong et al. 2010) are likely to account for the differences in the maximum cloud water mixing ratios. Although in WDM6 cloud water is double-moment, the maximum mixing ratios are only decreased slightly relative to WSM6. This result suggests that WDM6-forecasted cloud water number concentrations are likely close to prescribed 300 cm^{-3} number concentration assumed in WSM6 (Hong et al. 2010) and/or the larger-scale environment/forcing is a dominant factor as water supersaturation are negligible.

Amongst the BMPSs, Figuress. 2, 4, and 5 show that precipitable snow and snow mixing ratios vary considerably amongst the BMPSs with Lin6 and GCE6 having the smallest and largesthighest snow amounts of snow, respectively. Dudhia et al. (2008) and Tao et al. (2011) attribute the low snow mixing ratios in Lin6 to its high rates of dry collection of snow by graupel, its low snow size distribution intercept (decreased surface area), and its auto-conversion of snow to either graupel or hail at high mixing ratios. GCE6 turns off dry collection of snow and ice by graupel, greatly increasing the snow mixing ratios at the expense of graupel and reducing snow riming efficiency relative to Lin6 (Lang et al. 2007). Snow growth in GCE6 is further augmented by its assumption of water saturation for the vapor growth of cloud ice to snow (Reeves and Dawson, 2013; Lang et al. 2014). GCE7 addressed the vapor growth issueissue of GCE6 by and applied numerous other changes including the introducingtion and of a snow size and density mapping, snow breakup interactions, a relative humidity (RH)-based correction factor, and a new vertical-velocity-dependent ice super saturation assumption (Lang el al., 2007; Lang et al., 2011; Lang et al., 2014; Chern et al., 2016; Tao et al., 2016). Despite the reduced efficiency of vapor growth of cloud ice to snow due to stemming from the both the new RH correction factor and the ice super saturation adjustment, the new snow mapping and enhanced cloud ice--to--snow auto-conversion in GCE7 offset this potential reduction which kept and keep GCE snowfall mixing ratios higher than those in non-GCE BMPSs. Unlike Lin6, WSM6 and WDM6 assume that grid cell graupel and snow fall speeds are identical (Dudhia et al., 2008) and that ice nuclei concentration is a function of temperature (Hong et al., 2008). These two aspects, effectively eliminate the accretion of snow by graupel and increase snow mixing ratios at lowercolder temperatures (Dudhia et al., 2008; Hong et al., 2008). Figures 4 and 5 show the height of maximum snow mixing ratio height is roughly conserved in all non-Lin6 BMPSs. Lin6's assumption of non-uniform graupel and snow fall speeds and dry collection of snow- by graupel reduces snow mixing ratios in the middle troposphere and raises its maximum snow mixing ratio height.

Compared to snow, graupel mixning ratios are generally smaller -except for Lin6 where unrealistically high dry collection of snow by graupel dominates species growthfor non-Lin6 schemes due to Lin6's assumption of dry collection by snow dominates species growth which was proven unrealistic by (Stith et al. (2002). Graupel mixing ratios are lowest in GCE7 due to the net effect of its additions despite the inclusion of a new graupel size map. In particular, the combination of -is in many ways at opposition to Lin6, where it simulations generate the most snow, yet the least graupel. GCE7 includes graupel size mapping, but the combination of the new snow size mapping (decrease snow size aloft, -increases snow surface area, and enhances vapor growth), the addition of deposition conversion processes (graupel/hail particles experiencing deposition growth at lowercolder temperatures are converted to snow), and a reduction in super cooled droplets available for riming (cloud ice generation is augmented, see below) all favor snow growth at the expense of graupel (Lang et al. 2014; Chern et al., 2016; Tao et al., 2016). Consistent with Reeves and Dawson (2013), WSM6 and WDM6 graupel mixing ratios value are typically 30-50 % of their snow counterparts for WSM6 and WDM6.

Although cloud ice mixing ratios are nearly an order of magnitude are up to ninety percent smaller than those for snow (GCE6), these mixing ratioscloud ice mixing ratios still vary greatly amongst the BMPSs as illustrated in Figs. 2, 4, and 5. Cloud ice mixing ratios are highest in GCE7 and lowest in Lin6. Wu and Petty (2010) similarly found low cloud ice mixing ratios in Lin6 simulations and ascribe it to dry collection by cloud ice by graupel and its fixed cloud-ice size distribution. Similar to Lin6, GCE6 uses a monodispersed cloud-ice size distribution (20 μm diameter), but assumes vapor growth of cloud ice to snow assuming under an assumption of water saturation conditions (yet supersaturated with respect ice) leading to higher cloud ice amounts, and but also increased cloud ice to snow conversion rates (Lang et al., 2011; Tao et al., 2016). GCE7 blunts this cloud ice--to--snow conversion ratesterm using a RH correction factor which is dependent upon ice supersaturation which is itself dependent up vertical velocity. Additionally, GCE7, also includes contact and immersion freezing terms (Lang et al., 2011), makes the cloud ice collection by snow efficiency a function of snow size (Lang et al., 2011; Lang et al., 2014), sets a maximum limit on cloud-ice particle size (Tao et al., 2016), makes ice nuclei concentrations follows the Cooper curve (Cooper, 1986;

Tao et al., 2016), and ~~it~~ allows cloud ice to persist in ice subsaturated conditions (i.e., RH for ice ≥ 70%) (Lang et al, 2011; Lang et al., 2014). Despite the increased cloud ice-to-snow auto conversion rates in GCE7 (Lang et al. 2014; Tao et al. 2016), precipitable cloud ice amounts nearly doubled relative to GCE6~~all the above changes nearly doubled cloud ice amounts in GCE7 than in GCE6~~ (See Fig. 2). Similar to GCE7, WSM6 ~~runs~~ generates larger cloud ice mixing ratios than Lin6, which Wu and Petty (2010) attribute to excess cloud glaciation at temperatures between 0°C and -20°C and its usage of fixed cloud ice size intercepts. Additionally, both WSM6 and WDM6 include ice sedimentation terms which promote smaller cloud ice amounts (Hong et al., 2008). Despite their varying assumptions, the maximum cloud ice heights~~amounts~~ for both Case 5 and overall (Figs. 4 and 5) are consistent between BMPSs.

3.2 Stage IV precipitation analysis

Excessive precipitation, whether frozen or not, is one of the most potentially crippling impacts of~~from~~ a nor'easter. ~~WRF precipitation is generated from its microphysics and cumulus parameterization; the latter is turned for Domains 3 (5 km grid spacing) and 4 (1.667 km grid spacing).~~ Figures 6 and 7 show Domain 3, ~~24-hour~~ accumulated precipitation, their difference from St~~age~~ IV, and the~~the~~ associated probability and cumulative distribution functions (PDF and CDF, respectively) ~~of precipitation~~ for Cases 5 and 7 based upon the 24-30 hour residence period of a nor'easter within Domain 4.~~. As for our composite microphysics plots, the data accumulation period only covers the nor'easter's residence time in Domain 4.~~ We focus our attention on Domain 3 for this analysis because most of Domain 4 resides close to or outside the StIV data boundaries.~~rather than Domain 4 because the latter is located near the boundary of the Stage IV dataset where its radar-based data tends to fade.~~ Cases 5 and 7 are chosen ~~are shown here~~ because of their ~~these cases have~~ near-shore tracks (Fig. 1) which affords good StIV data coverage ~~of their associated precipitation by Stage IV~~. Table 3 includes threat score and bias information from ~~or~~ all seven cases and their associated standard deviation statistics. Both threat score and model bias assume the same ~~a~~ 10 mm ~~precipitation accumulation~~ threshold value, which as seen in Figs. 6 and 7 is approximately the 25th percentile of accumulated precipitation on average.

Case 4 threat score and bias values (Table 3)~~Table 3 shows Case 4~~ are more than two standard deviations from the composite mean due to its non-coastal storm track (Fig. 1) and thus it is excluded from this analysis~~as a clear outlier where its low threat score and bias values deviate more than two standard deviation from the composite mean due to its non-coastal track (Fig. 1) and thus it will be excluded from this section of the analysis~~. The remaining six cases show WRF to have low-to-moderate forecast skill (Threat score: 0.217 [Lin6] – 0.414 [Lin6]) and to cover too large an area with precipitation values greater than 10 mm~~For the remaining six cases, Table 4 indicates low (0.217; Lin6, Case 2) to moderate (0.414; Lin6, Case 5) threat scores and a 10 mm precipitation contour spatial covers an area far exceeding Stage IV~~ (bias ~~range~~: 1.47 [Lin6, Case 7] – 4.05 [GCE7, Case 3]) relative to StIV. Inter-BMPS threat score and bias differences are an order or magnitude or less than the values from which they are derived. ~~Inter-BMPS barely varied with threat score and biases varying only up to an order of magnitude less than the threat and bias scores themselves.~~ Consistent with Hong et al. (2010), threat score and bias values from~~or~~ WSM6 are equal to or improved upon by WDM6 due to its inclusion of a cloud condensation nuclei ~~(CCN)~~ feedback. Overall, WDM6 shows marginally better precipitation forecast skill than other BMPSs (lowest threat score in four out of six cases and lowest mean threat score: 0.322), yet Lin6 is the least biased (lowest bias score in four of out of six cases and lowest mean bias: 2.55)generated marginally better simulated precipitation fields and has the lowest threat score in four out of six cases and it also has the lowest model mean (0.322), yet Lin6 was found to be the least bias in four out of six cases and it also has the lowest model mean (2.55).

PDF and CDF plots As illustrated from Figs. 6 and 7 show, all WRF to favor higher precipitation amounts and is consistent with the positive bias scores in Table 3. simulations tended to generate similar coverage to Stage IV, but its precipitation values tended to be smaller than for corresponding grid points in WRF resulting in low to moderate forecast skill and excessively heavy precipitation totals as illustrates in the PDF and CDF diagrams. Previous modelling studies of strong -convection by Ridout et al. (2005) and Dravitzki and McGregor (2011) found that both GFS and the Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS) produced too much light precipitation and too much heavy precipitation, which stands in contrast to our results, which show the opposite tendency. Unlike these two studies, our study region lacks nor'easters often track over the data spare North Atlantic, a region with no rain gauge data and is at is near or beyond- the operational range limits of S-band radars. These two issues could lead to an under bias in Stage IV precipitation data, especially near the data boundariesedges and suggests that WRF threat scores and biases are , which likely suggests that threat scores and biases are likely closer to observations than as Table 3 indicates shown. Marginal changes in accumulated precipitation PDFs and CDFs (<10 mm) between BMPS simulations and threat scores amongst BMPSs areis consistent with the investigation of simulatedion precipitation during warm-season precipitation events and a quasi-stationary front by (Fritsch and Carbone (, 2004); and Wang and Clark (2010), respectively..

Min (2015)

WDM6 has been reported to reduce light precipitation and increase moderate precipitation, reducing the systematic bias of WSM6 (Hong et al. 2010, Min et al 2015).

using simulated reflectivity products to compare model fields with radar has advantages over radarestimated precipitation fields because there is less uncertainty involved in the calculation of reflectivity from the model than precipitation from radar (Koch et al. 2005; Molthan and Colle 2012)

Among the six hydrometeors, q_{rain}, q_{snow}, and $q_{graupel}$ are used to calculate the reflectivity.

4-km resolution is needed to account for the complexity of the local topography and to compare directly with radar data

3.3 MRMS and radar reflectivity analysis

Figure 8 shows Domain 4, Case 4 ~~statistical~~ radar reflectivity CFADs ~~for Case 4, Domain 4~~ constructed over the ~~a~~ 24 hour residence time of the nor'easter within Domain 4 ~~period~~ (12 UTC 26–27 January 2015). Although not shown, ~~with~~ the 0°C and -40°C heights are at approximately 2~~3~~,000 and 8~~9~~,000 m AMSL, respectively.~~above mean sea level (not shown)~~. ~~Similar to the previous section, all CFAD and CFAD products are based only upon the 24-30 hour period a nor'easter resided within Domain 4.~~ We selected Case 4 because its radar data has been reprocessed with the latest MRMS algorithm, whereas the remaining cases used an older algorithm associated with NMQ and were still in the process of being updated.~~We selected Case 7 because its radar volume data from NMQ has been reprocessed with the latest algorithms associated with MRMS~~. The MRMS CFAD (Fig. 8) shows two, distinct frequency peaks between 2,300 – 5,000 m and 7,500 m – 11,000 m AMSL, that are not well matched in the models. To investigate these differences, Figs. 9 and 10 show radar reflectivity at 4,000 and 9,500 m AMSL on 18 UTC 26 January 2015. Finally, to evaluate model performance against MRMS, Fig. 11 shows a contoured plot of CFAD scores calculated hourly and at each height level.

CFADs in Fig. 8 depict a distinct bifurcation in the simulated CFADs above and below 6,000 m AMSL relative to MRMS. Below this level, model-based CFADs generally show a frequency swath that is broader than MRMS and overly favors the occurrence of stronger reflectivity values (exceptions: GCE7 and Lin6). Above this level, CFADs display a reflectivity frequency swath that is generally too narrow (exception: GCE6) and favors weaker reflectivity values relative to MRMS. Below the 0°C height level (2,000 m AMSL), all models over extend the reflectivity frequency range (5% frequency: Models = -15 – 32 dBZ; MRMS = -1 – 27 dBZ), yet only Lin6 and especially GCE7 correctly capture the maximum frequency core between 5 and 15 dBZ. Other schemes produce this core, but it is offset by 10 dBZ or more toward higher reflectivity values. Case 4 mixing ratio profiles (not shown) depict similar relationships amongst the hydrometer species as shown for Case 5 (Fig. 4), albeit with a comparably lower freezing level. Below the freezing level, both GCE7 and Lin6 have lower graupel mixing ratios values than other schemes. Given the earlier correspondence (Section 3.1) between graupel and stronger reflectivity values this does suggest a probable explanation for the better results of Lin6 and GCE7. At higher altitudes (2,000 – 6,000 m), both Lin6 and GCE7 maintain the best representation of the maximum frequency core with radar reflectivity values that are indeed lower than other schemes and closer to MRMS (Fig. 9 at 4,000 m AMSL). Above 6,000 m AMSL, WRF CFADs shift toward very low reflectivity values (< 0 dBZ) due to increased entrainment near the simulated cloud top. This shift if particularly pronounced in GCE7 where the combination of its new temperature and mixing ratio dependent aggregation rates and snow map produce smaller hydrometeors are lower temperatures and larger hydrometeors at higher temperatures. Near the top of the troposphere (> 7,500 m AMSL), CFADs values are based solely upon increasingly isolated reflectivity values (Fig. 10 at 9,500 m AMSL) which leads to the notable discrepancies in CFAD structure between the models and MRMS.

~~To supplement Fig. 8, MRMS and WRF simulated radar reflectivities are shown at 4,000 and 9,500 m above mean sea level on 18 UTC 26 January 2015 are shown as Figs. 9 and 10, respectively. These two heights were selected because they pass through the two MRMS dBZ frequency maxima shown in Fig. 8. Finally, Fig. 11 shows CFAD scores with height and time and their differences over the same time period as Fig. 8.~~

~~Figure 8 show a wider ranges of dBZ values (up to 40 dBZ) from WRF simulations than from MRMS (up to 27 dBZ) below the melting layer. Qualitatively, all model simulations below the melting layer have dBZ frequency ranges exceeding that of MRMS, yet only Lin6 and especially GCE7 correctly capture the core of maximum frequencies between 5-10 dBZ. All other schemes produce this same core, but at values over 10 dBZ. Figure 9 illustrates these radar reflectivity differences at the 4,000 m above sea level where radar reflectivity values from GCE6, WSM6, and WDM6 simulations are often 15 dBZ or more greater than MRMS. Between 3,000 and 6,000 m, only GCE7 produces a narrow core of maximum frequency values below 10 dBZ consistent with MRMS. Lang et al. (2014) attribute the narrow core to changes in aggregation which made it both temperature and mixing ratio dependent and to the new snow map. Together these changes favored the production of small hydrometeors at colder temperature and larger hydrometeors at warmer temperatures. Eventually above 6,500 m, all WRF CFADs collapse to very small radar reflectivities values (< 5 dBZ) whereas the core of dBZ frequencies increases in MRMS up through 11 km. As Fig. 10 shows, at 9,500 m in altitude radar reflectivity coverage has become spotty and quite sensitive to even small radar signatures.~~

Consistent with the above discussion, CFAD scores with height and time (Fig. 11) show Lin6 to qualitatively perform best overall, however, GCE7 simulations below 5,000 m AMSL typically attained even higher CFAD scores. Other BMPSs (~~as shown in~~ Fig. 8) typically favor unrealistically high~~er~~ distribution of reflectivity values and also ~~the~~ exhibit lower CFAD scores in the melting layer likely due to ~~which is likely associated with~~ higher graupel mixing ratios~~and cloud ice concentrations~~. Further aloft, aggregation of hydrometeors toward smaller sizes and entrainment likely cut off cloud tops in GCE7 more so than in other schemes and results in its lower CFAD scores above 6,000 m AMSL. The other six cases produce similar tendencies in their CFAD and CFAD scores as noted above for Case 4~~7~~, except cloud heights become higher and CFADs become wider with the introduction of stronger convection in~~with~~ early and late season events.

4 Conclusions

The role and impact of five bulk microphysics schemes (BMPSs; Table 2) upon seven, Weather Research and Forecasting model (WRF) ~~W361~~ winter time cyclone (“nor’easter”)nor’easter~~nor’easter~~ simulations (Table 1) are~~is~~ investigated and validated against GFS model analysis (GMA), Stage IV rain gauge and radar estimated precipitation, and the radar-derived, Multi-Radar, Multi-Sensor (MRMS) 3D volume radar reflectivity product ~~MRMS 3D volume reflectivity~~. Tested BMPSs include three~~four~~ single-moment, six class BMPSs (Lin6, GCE6, ~~GCE7,~~ and WSM6), one single-moment, seven class BMPS (GCE7), and one double-moment, six-class BMPS~~s~~ (WDM6). Simulated hydrometer mixing ratios show general similarities for non-frozen hydrometeor species (cloud water and rain) due to their common Lin6 heritage. However, frozen hydrometeor species (snow, graupel, cloud ice) demonstrate considerably larger variability amongst~~between~~ BMPSs. This variability results ~~Larger changes exist for frozen species due to~~ from different assumptions concerning ~~about~~ snow and graupel intercepts, degree of allowable ice supersaturation, snow and graupel density maps, and terminal velocities made by each BMPS. WRF-simulated precipitation fields exhibit similar coverage, but tended to favor higher precipitation amounts relative to Stage IV observations resulting in low-to-moderate threat scores (0.217–0.414). Inter-model differences were an order of magnitude or less than the threat score values, but WDM6 did demonstrate marginally better forecast skill overall. WRF-Stage IV accumulated precipitation comparisons reveal WRF demonstrate that although WRF generates precipitation fields of similar coverage to Stage IV precipitation intensities tended to be higher than observations and resulting in low to moderate (0.217–0.414) threat scores with WDM6 demonstrating marginally better forecast skill than its single-moment counterparts. Finally, MRMS-based contoured frequency with altitude diagrams (CFADs) and CFAD scores show Lin6 and GCE7 to be best in to be notably better than GCE6, WSM6 and WDM6 in the lower half of the –troposphere, where GCE7 most realistically reproduced the maximum frequency core between 5 and 15 dBZwith GCE7 being the only BMPS scheme to produce the narrow core of maximum frequencies below10 dBZ due to its temperature and mixing ratio dependent aggregation and new snow map. Above 65,000 m AMSL, model simulations approach or exceed their cloud tops where entrainment and hydrometeor sizes differences alter cloud top heights and reflectivity fields became increasingly spotty with height which made CFADs increasingly sensitive to individual reflectivity values. GCE7 however becomes less skilled the combination of smaller hydrometers and entrainment reduced it cloud top height relative to other BMPSs.

The study has shown that although cloud microphysics lead to only -subtle in the large-scale environment, they do noticeably alter cloud microphysics do make small, but noticeable impacts in the microphysical and precipitation properties of a nor'easter. While no BMPS hasleads to consistently improved precipitation forecast skill, their underlying assumptions result in varying forecast skill of simulated radar reflectivity structures between individual BMPSs when compared to MRMS observations. do make notable change in the composition of radar reflectivity structure which itself can vary notably from observed radar reflectivity structures. Follow-on studies could investigate additional nor'easter cases or simulate other weather phenomena (polar lows, monsoon rainfall, drizzle, etc.). Results covering multiple phenomena may provide guidance to model users in their selection of BMPS for a given computational cost. Additionally, potential studies could specifically focus on -address key aspects of a nor'easter's structure (such as the low-level jet) or validation of model output against current and recently available satellite-based datasets from MODIS (Justice et al., 2008), CloudSat (Stephens et al., 2008), CERES, and GPM (Hou et al. 2014). Finally, other validation methods including object-oriented (Marzban and Sandgathe, 2006) or fuzzy verification (Ebert 2008) could be utilized.

5 Code availability

WRF version 3.6.1 is publically available for download from the WRF Users' Page (http://www2.mmm.ucar.edu/wrf/users/download/get_sources.html).

6 Data availability

GFS model analysis data boundary condition data can be obtained from the NASA's open access, NOMADS data server (ftp://nomads.ncdc.noaa.gov/GFS/Grid3/). Stage IV precipitation data is publically available from the

National Data and Software Facility at the University Center for Atmospheric Research (http://data.eol.ucar.edu/cgi-bin/codiac/fgr_form/id=21.093). Daily MRMS data is available from the National Severe Storms Laboratory (http://www.nssl.noaa.gov/projects/mrms/)

7 Author contributions

S. D. Nicholls designed and ran all ~~experimental~~ model simulations and prepared th~~is~~ise manuscript. S. G. Decker supervised S. D. Nicholls' research efforts, funded the research, and revised the manuscript. W. -K. Tao, S. E. Lang, and J. J. Shi brought their extensive knowledge and expertise on model microphysics which helped shape the project methodology and rationalize the results. Finally, K. I. Mohr helped to facilitate connections between the research team, supervised S. Nicholls' research, and was pivotal in revising the manuscript.

8 Acknowledgements

This research was supported by the Joint Center for Earth Systems Technology (JCET), the University of Maryland Baltimore County (UMBC), and in part by the New Jersey Agricultural Experiment Station. Resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center.

References

Ashton, A. D., Donnely, J. P., and Evans, R. L.: A discussion of the potential impacts of climate change on the shorelines of the Northeastern U.S.A. Mitig. Adapt. Strat. Glob. Change, 13, 719–743, 2008.

Bosart, L. F.: The Presidents' Day Snowstorm of 18–19 February 1979: A subsynoptic-scale event, Mon. Wea. Rev., 109, 1542–1566, 1981.

Chen, F., and Dudhia, J.: Coupling an advanced land-surface/ hydrology model with the Penn State/ NCAR MM5 modeling system. Part I: Model description and implementation, Mon. Wea. Rev., 129, 569–585, 2001.

Chern, J. -D., Tao, W. -K., Lang, S. E., Matsui, T., J. -L. F. Li, J. -L. F., Mohr, K. I.,, Skofronick-Jackson, G. M., and Peters-Lidard, C. D.: Performance of the Goddard multiscale modeling framework with Goddard ice microphysical schemes, J. Adv. Model. Earth Syst., 7, doi:10.1002/2015MS000469, 2016.

Chou, M. -D. and Suarez, M. J.: A solar radiation parameterization for atmospheric research studies. NASA Tech, Memo NASA/TM-1999-104606, 40 pp., 1999.

Chou, M. -D., and Suarez, M. J.: A thermal infrared radiation parameterization for atmospheric studies, NASA Tech. Rep. NASA/TM-1999-10466, vol. 19, 55 pp., 2001.

Deng M, G. G. Mace, Z. Wang, and R. P. Lawson: Evaluation of several A-Train ice cloud retrieval products with in situ measurements collected during the SPARTICUS campaign, J. Appl. Meteor. Climatol., 52, 1014–1030, 2013.

Dravitzki, S., and McGregor, J.: Predictability of heavy precipitation in the Waikato River Basin of New Zealand, Mon. Wea. Rev., 139, 2184–2197, 2011.

Dudhia, J., Hong, S. -Y., and Lim, K. -S.: A new method for representing mixed-phase particle fall speeds in bulk microphysics parameterizations, J. Meteor. Soc. Japan, 86A, 33–44, 2008.

Ebert, E. E.: Fuzzy verification of high-resolution gridded forecasts: A review and a proposed framework, Meteor. Applic., 15, 51-64, 2008.

Forbes, G. S., Thomson, D. W., and Anthes, R. A.: Synoptic and mesoscale aspects of an Appalachian ice storm associated with cold-air damming, Mon. Wea. Rev., 115, 564–591, 1987.

Fulton, R. A., J. P. Breidenbach, D.-J. Seo, D. A. Miller, and T. O'Bannon: The WSR-88D rainfall algorithm. Wea. Forecasting, 13, 377–395. 1998.

Fritsch, J. M., and Carbone, R. E.: Improving quantitative precipitation forecasts in the warm season: A USWRP research and development strategy, Bull. Amer. Meteor. Soc., 85**,** 955–965, 2004.

Ganetis, S. A. and Colle, B. A.: The thermodynamic and microphysical evolution of an intense snowband during the Northeast U.S. blizzard of 8–9 February 2013. Mon. Wea. Rev., 143, 4104-4125, 2015.

Hong, S -Y., and Lim, J. -O. .J.: The WRF single-moment 6-class microphysics scheme (WSM6), J. Korean Meteor. Soc., 42, 129-151, 2006.

Hong, S. -Y., Lim, K. -S. S., Lee, Y. -H., Ha, J. -C., Kim, H. -W., Ham, S. -J., and Dudhia, J.: Evaluation of the WRF double-moment 6-class microphysics scheme for precipitating convection, Adv. Meteor., 2010, doi:10.1155/2010/707253, 2010.

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K., and Iguchi, T.: The Global Precipitation Measurement Mission, Bull. Amer. Meteor. Soc., 95, 701–722, 2014.

Jacobs, N. A., Lackmann, G. M., and Raman, S.: The combined effects of Gulf Stream-induced baroclinicity and upper-level vorticity on U.S. East Coast extratropical cyclogenesis, Mon. Wea. Rev., 133, 2494–2501, 2005.

Janjic, Z. I.: Nonsingular implementation of the Mellor–Yamada level 2.5 scheme in the NCEP meso model, NCEP Office Note 437, 61 pp., 2002.

Justice, C. O. et al. (1998), The Moderate Resolution Imaging Spectroradiometer (MODIS): land remote sensing for global change research, IEEE Transactions on Geoscience and Remote Sensing, 36, 1228–1249, 1998.

Kain, J. S.: The Kain–Fritsch Convective Parameterization: An Update, J. Appl. Meteor., 43, 170–181, 2004.

Kessler, E.: On the distribution and continuity of water substance in atmospheric circulation, Meteor. Monogr., 32, Amer. Meteor. Soc., 84 pp, 1969.

Kleczek, M. A., G.-J. Steenveld, and A. A. M. Holtslag: Evaluation of the Weather Research and Forecasting Mesoscale Model for GABLS3: Impact of boundary-layer schemes, boundary conditions and spin-up, Boundary-Layer Meteorol, 152, 213–243, 2014.

Kocin, P. J. and Uccellini, L. W.: Northeast snowstorms. Vols. 1 and 2, Meteor. Monogr., No. 54., Amer. Met. Soc., 818 pp., 2004.

Kuo, Y. H., Low-Nam, S., and Reed, R. J.: Effects of surface energy fluxes during the early development and rapid intensification stages of seven explosive cyclones in the Western Atlantic. Mon. Wea. Rev., 119, 457–476, 1991.

Lang, S., Tao, W. -K., Cifelli, R., Olson, W., Halverson, J., Rutledge, S., and Simpson, J.: Improving simulations of convective system from TRMM LBA: Easterly and westerly regimes, J. Atmos. Sci., 64, 1141–1164, 2007.

Lang, S. E., Tao, W. -K., Zeng, X., and Li, Y.: Reducing the biases in simulated radar reflectivities from a bulk microphysics scheme: Tropical convective systems, J. Atmos. Sci., 68, 2306–2320, 2011.

Lang, S. E., Tao, W. -K., Chern, J. -D., Wu, D., and Li, X.: Benefits of a fourth ice class in the simulated radar reflectivities of convective systems using a bulk microphysics scheme, J. Atmos. Sci., 71, 3583–3612, doi:10.1175/JAS-D-13-0330.1, 2014.

Lebsock, M., and Su, H: Application of active spaceborne remote sensing for understanding biases between passive cloud water path retrievals, J. Geophys. Res. Atmos., 119, 8962–8979, doi:10.1002/2014JD021568, 2014.

Li, J. -L. F., Waliser, D., Woods, C., Teixeira, J., Bacmeister, J., Chern, J. -D.,, Shen, B. -W., Tompkins, A., Tao, W. -K.,, and Kohler, M.: Comparisons of satellites liquid water estimates to ECMWF and GMAO analyses, 20th century IPCC AR4 climate simulations, and GCM simulations, Geophys. Res. Lett., 35, L19710, doi:10.1029/2008GL035427, 2008.

Lim, K.-S. and Hong, S. -Y.: Development of an effective double-moment cloud microphysics scheme with prognostic cloud condensation nuclei (CCN) for weather and climate models, Mon. Wea. Rev., 138, 1587–1612, 2010.

Lin, Y. -L., Farley, R. D., and Orville, H. D.: Bulk parameterization of the snow field in a cloud model, J. Climate Appl. Meteor., 22, 1065–1092, 1983.

Mailhot, J. and Chouinard, C.: Numerical forecasts of explosive winter storms: Sensitivity experiments with a meso-scale model, Mon Wea. Rev., 117, 1311–1343, 1989.

Marzban C., and Sandgathe, S.: Cluster analysis for verification of precipitation fields, Wea. Forecasting, 21, 824–838, 2006.

Mellor, G. L., and Yamada, T.: Development of a turbulence closure model for geophysical fluid problems, Rev. Geophys. Space Phys., 20, 851–875, 1982.

Min, K.-H., S. Choo, D. Lee and G. Lee: Evaluation of WRF cloud microphysics schemes using radar observations. Weather and Forecasting, 30, 10.1175/WAF-D-14-00095.1, 2015.

Monin, A. S., and Obukhov, A. M.: Basic laws of turbulent mixing in the surface layer of the atmosphere. Tr. Akad. Nauk SSSR Geophiz. Inst., 24, 163–187, 1954.

Morath, E. (2016), Will a blizzard freeze U.S. economic growth for the third straight year, Wall Street Journal, 20 Jan. 2016.

Morrison, H., Thompson, G., and Tatarskii, V.: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes, Mon. Wea. Rev., 137, 991–1007, 2009.

Mote, T. L., Gamble, D. W., Underwood, S. J., Bentley, M. L.: Synoptic-scale features common to heavy snowstorms in the Southeast United States, Wea. Forecasting, 12, 5–23, 1997.

Nicholls, S. D. and Decker, S. G.: Impact of coupling an ocean model to WRF nor'easter simulations, Mon. Wea. Rev., 143, 4997–5016, 2015.

Reeves, H. D. and Dawson II, D. T.: The dependence of QPF on the choice of microphysical parameterization for lake-effect snowstorms, J. Appl. Meteor. Climatol., 52, 363–377, 2013.

Reisner, J. R., Rasmussen, R. M., and Bruintjes, R .T.: Explicit forecasting of supercooled liquid water in winter storms using the MM5 mesoscale model. Quar. J. Roy. Met. Soc., 124, 1071-1107, 1998.

Ridout, J. A., Y. Jin, and Liou, C. -S.: A cloud-base quasi-balance constraint for parameterized convection: Application to the Kain–Fritsch cumulus scheme, Mon. Wea. Rev., 133, 3315–3334, 2005.

Rutledge, S. A., and Hobbs, P. V.: The mesoscale and microscale structure and organization of clouds and precipitation in mid-latitude cyclones. XII: A diagnostic modeling study of precipitation development in narrow cloud-frontal rainbands. J. Atmos. Sci., 20, 2949–2972, 1984.

Shi, J. J. et al.: WRF simulations of the 20-22 January 2007 snow events of Eastern Canada: Comparison with in situ and satellite observations, J. Appl. Meteor. Climatol., 49, 2246–2266, 2010.

Skamarock, W.C., Klemp, J. P., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X. -Y., Wang, W., and Powers, J. G.: A description of the advanced research WRF version 3, NCAR Tech. Note NCAR/TN–475+STR, 125 pp., 2008.

Smith, A. B., and Katz, R. W.: US billion-dollar weather and climate disasters: Data sources, trends, accuracy and biases, Natural Hazards, 67, 387–410, 2013.

Stark, D.: Field observations and modeling of the microphysics within winter storms over Long Island, NY. M.S. thesis, School of Marine and Atmospheric Sciences, Stony Brook University, 132 pp., 2012.

Stauffer, D. R., and Warner, T. T.: A numerical study of Appalachian cold-air damming and coastal frontogenesis, Mon. Wea. Rev., 115, 799–821, 1987.

Stephens, G. L., et al.: CloudSat mission: Performance and early science after the first year of operation, J. Geophys. Res., 113, D00A18, doi:10.1029/2008JD009982, 2008.

Stith, J. L., Dye, J. E., Bansemer, A., Heymsfield, A. J., Grainger, C. A., Petersen, W. A, and Clfelli, R.: Microphysical observations of tropical clouds, J. Appl. Meteor., 41, 97–117, 2002.

Tao, W. -K., Simpson, J. and McCumber, M.: An ice-water saturation adjustment, Mon. Wea. Rev., 117, 231–235, 1989.

Tao, W. -K., Shi, J. J., Chen, S. S., Lang, S., Lin, P. -L., Hong, S. -Y., Peters-Lidard, C., and Hou, A.: The impact of microphysical schemes on hurricane intensity and track, Asia-Pacific J. Atmos. Sci., 47, 1–16, 2011.

Tao, W. -K., Wu, D., Lang, S., Chern, J. -D., Peters-Lidard, C., Fridlind, A., and Matsui, T.: High-resolution NU-WRF simulations of a deep convective-precipitation system during MC3E: Further improvements and comparisons between Goddard microphysics schemes and observations, J. Geophys. Res. Atmos., 121, 1278–1305, doi:10.1002/2015JD023986, 2016.

Uccellini, L. W. and Kocin, P. J.: The Interaction of jet streak circulations during heavy snow events along the east coast of the United States, Wea. Forecasting, 2, 289–308, 1987.

Wang, S.-Y., and Clark, A. J.: NAM Model forecasts of warm-season quasi-stationary frontal environments in the Central United States, Wea. Forecasting, 25**,** 1281–1292, 2010.

Wilks, D. S.: Statistical methods in the atmospheric sciences, third edition, Academic Press, Oxford, in press., 2011.

Wu, L., and Petty, G. W.: Intercomparison of bulk microphysics schemes in model simulations of polar lows, Mon. Wea. Rev., 138, 2211–2228, 2010.

Yao, Y., Pierre, W., Zhang, W., and Jiang, J.: Characteristics of atmosphere-ocean interactions along North Atlantic extratropical storm tracks, J. Geophys. Res., 113, doi:10.1029/2007JD008854, 2008.

Yuter, S. E., and R. A. Houze: Three-dimensional kinematic and microphysical evolution of Florida cumulonimbus part II: frequency distributions of vertical velocity, reflectivity, and differential reflectivity, Mon. Wea. Rev., 123, 1941–1963.

Zhang, J., K. Howard, C. Langston, S. Vasiloff, B. Kaney, A. Arthur, et al.: National mosaic and multi-sensor QPE (NMQ) system: description, results, and future plans. Bulletin of the American Meteorological Society, 92, 1321-1338, 2011.

Zhang, J., K. Howard, C. Langston, B. Kaney, B., Y. Qi, L. Tang, H. Grams, Y. Wang, S. Cocks, S. Martinaitis, and A. Arthur: Multi-radar multi-sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. Bull. Amer. Meteor. Soc., 97, 621–638, 2016.

Table 1. Nor'easter case list. The NESIS number is included for storm severity reference. Mean sea-level pressure (MSLP) indicates maximum cyclone intensity in GMA. The last two columns denote the first and last times for each model run. GMA storm tracks are displayed in Fig. 1.

Case Number	NESIS	MSLP (hPa)	Event Dates	Model Run Start Date	Model Run End Date
1	N/A	991.5	15–16 Oct 2009	10/15 00UTC	10/20 00UTC
2	N/A	989.5	07–09 Nov 2012	11/06 18UTC	11/11 18UTC
3	4.03	972.6	19–20 Dec 2009	12/18 18UTC	12/23 18UTC
4	2.62	980.5	26–28 Jan 2015	01/25 12UTC	01/30 12 UTC
5	4.38	979.7	05–07 Feb 2010	02/05 06UTC	02/10 06UTC
6	1.65	1005.5	02–03 Mar 2009	03/01 00UTC	03/06 00UTC
7	N/A	993.5	12–14 Mar 2010	03/11 18UTC	03/16 18UTC

Table 2. Applied bulk microphysics schemes and their characteristics. The below table indicates simulated mixing ratio species and number of moments. Mixing ratio species include: QV = water vapor, QC = cloud water, QH = hail, QI = cloud ice, QG = graupel, QR = rain, QS = snow.

Microphysics Scheme	QV	QC	QH	QI	QG	QR	QS	Moments	Citation
Lin6	X	X		X	X	X	X	1	Lin et al. (1983); Rutledge and Hobbs (1984)
GCE6	X	X		X	X	X	X	1	Tao et al. (1989); Lang et al. (2007)
GCE7	X	X	X	X	X	X	X	1	Lang et al. (2014)
WSM6	X	X		X	X	X	X	1	Hong and Lim (2006)
WDM6	X	X		X	X	X	X	2 (QC, QR)	Lim and Hong (2010)

Table 3. Stage IV-relative, accumulated precipitation threat scores and biases assuming a threshold value of 10 mm (25th percentile of 24 hour accumulated precipitation). Bolded value denote the model simulation with the threat score closest to 1 (perfect forecast) or a bias values closest to 1 (number of forecasted cells matches observations). The lower two panels indicate the number of standards deviations (stdev) each threat score and bias value deviates from the composite (all models + all cases) mean.

Domain 3

Threat Score	1	2	3	4	5	6	7	Mean	Mean w/o 4
Lin6	0.289	0.217	0.291	0.091	**0.414**	**0.304**	0.332	0.277	0.308
GCE6	0.286	**0.243**	0.320	0.091	0.406	0.291	0.356	0.285	0.317
GCE7	0.288	0.235	0.319	**0.096**	0.405	0.300	0.337	0.283	0.314
WSM6	**0.293**	0.237	0.315	0.093	0.404	0.292	0.356	0.284	0.316
WDM6	0.290	**0.243**	**0.329**	0.094	0.411	0.299	**0.357**	**0.289**	**0.322**

Bias	1	2	3	4	5	6	7	Mean	Mean w/o 4
Lin6	2.47	**3.53**	**2.72**	7.82	**2.22**	2.9	**1.47**	3.304	2.552
GCE6	**2.37**	3.88	2.85	8.09	2.26	2.93	1.64	3.431	2.655
GCE7	2.52	4.05	2.85	**7.75**	2.23	2.82	1.57	3.399	2.673
WSM6	2.47	3.75	2.86	8.13	2.26	2.93	1.62	3.431	2.648
WDM6	**2.37**	3.8	2.76	8.09	2.23	**2.82**	1.57	**3.377**	**2.592**

T. Score Stats:	All Stdev	0.094	All Mean	0.284			
Threat Score	1	2	3	4	5	6	7
Lin6	0.06	-0.71	0.08	-2.05	1.39	0.22	0.52
GCE6	0.03	-0.43	0.39	-2.05	1.31	0.08	0.77
GCE7	0.05	-0.52	0.38	-2.00	1.29	0.18	0.57
WSM6	0.10	-0.50	0.34	-2.03	1.28	0.09	0.77
WDM6	0.07	-0.43	0.48	-2.02	1.36	0.16	0.78

Bias Stats	All Stdev	2.007	All Mean	3.389			
Bias	1	2	3	4	5	6	7
Lin6	-0.46	0.07	-0.33	2.21	-0.58	-0.24	-0.96
GCE6	-0.51	0.24	-0.27	2.34	-0.56	-0.23	-0.87
GCE7	-0.43	0.33	-0.27	2.17	-0.58	-0.28	-0.91
WSM6	-0.46	0.18	-0.26	2.36	-0.56	-0.23	-0.88
WDM6	-0.51	0.21	-0.31	2.34	-0.58	-0.28	-0.91

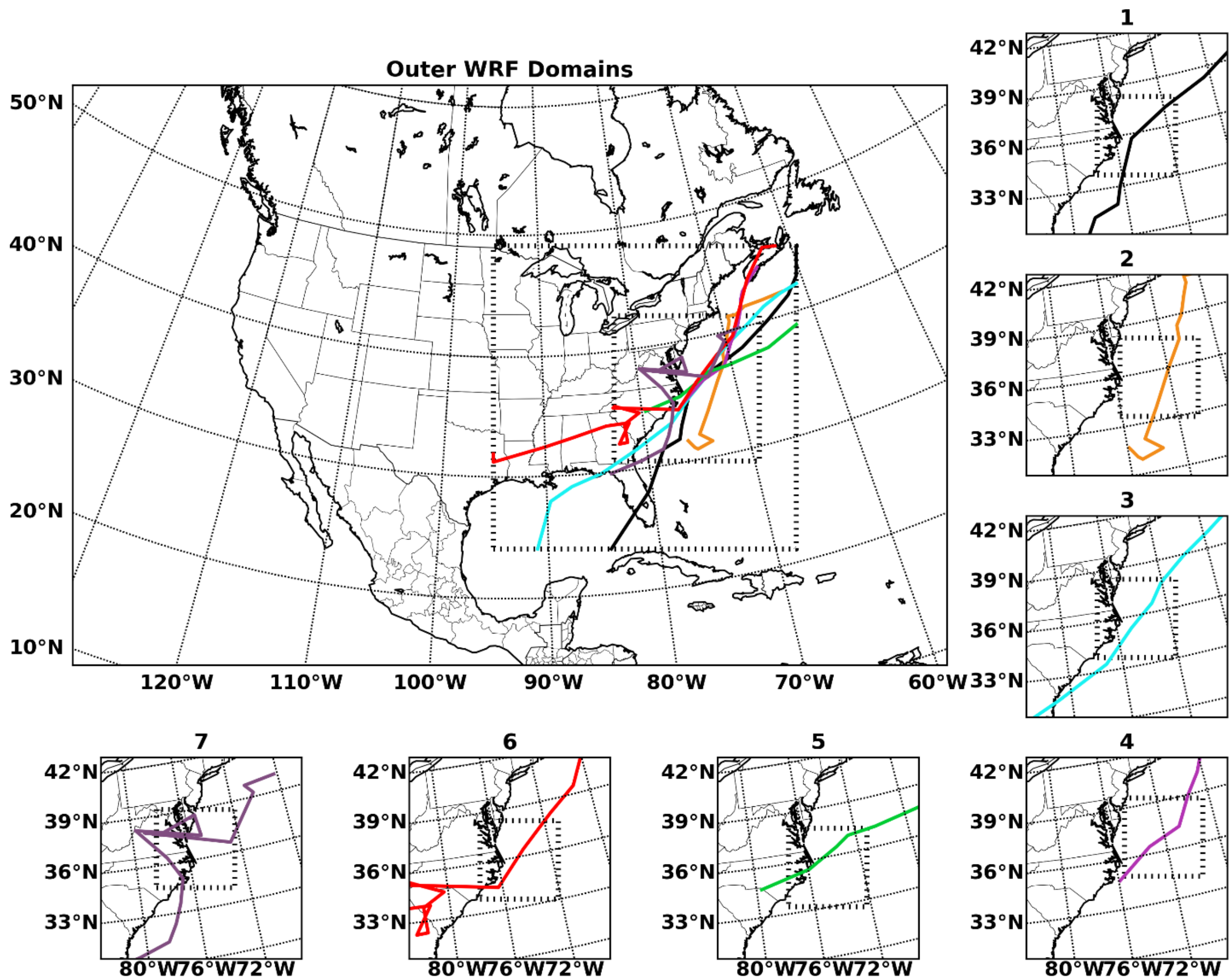

Figure 1. Nested WRF configuration used in simulations. The large panel shows the first 3 model domains (45-, 15-, 5- km grid spacing, respectively). The smaller panels show the location of domain 4 (1.667-km resolution) for each of the seven cases. The colored lines show the cyclone track as indicated by GMA for each nor'easter case.

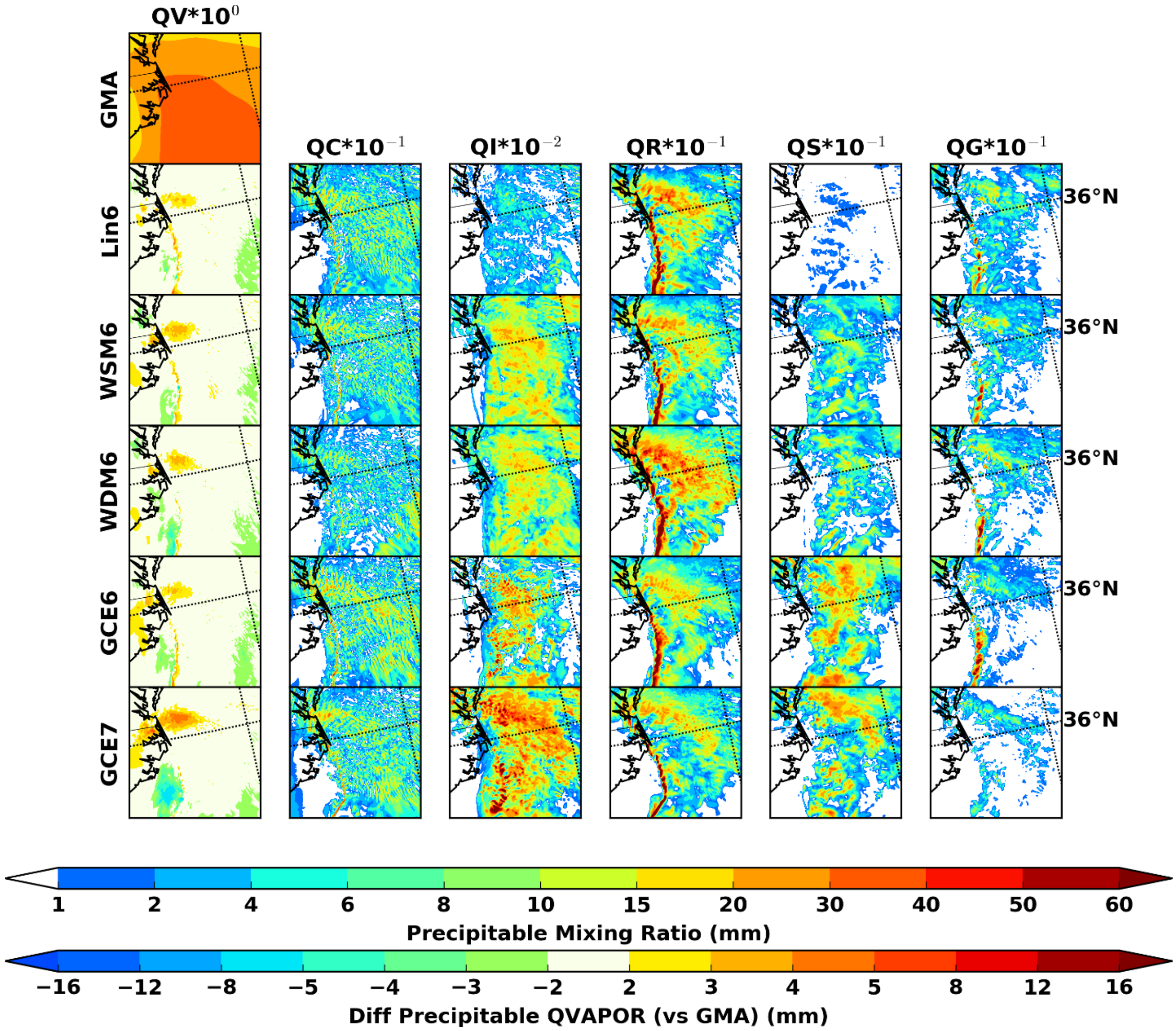

Figure 2. Domain 4 (1.667 km grid spacing), precipitable mixing ratios (mm) at 06 UTC 06 February 2010. Shown abbreviations for mixing ratios include: QV = water vapor, QC = cloud water, QG = graupel, QI = cloud ice, QR = rain, QS = snow.

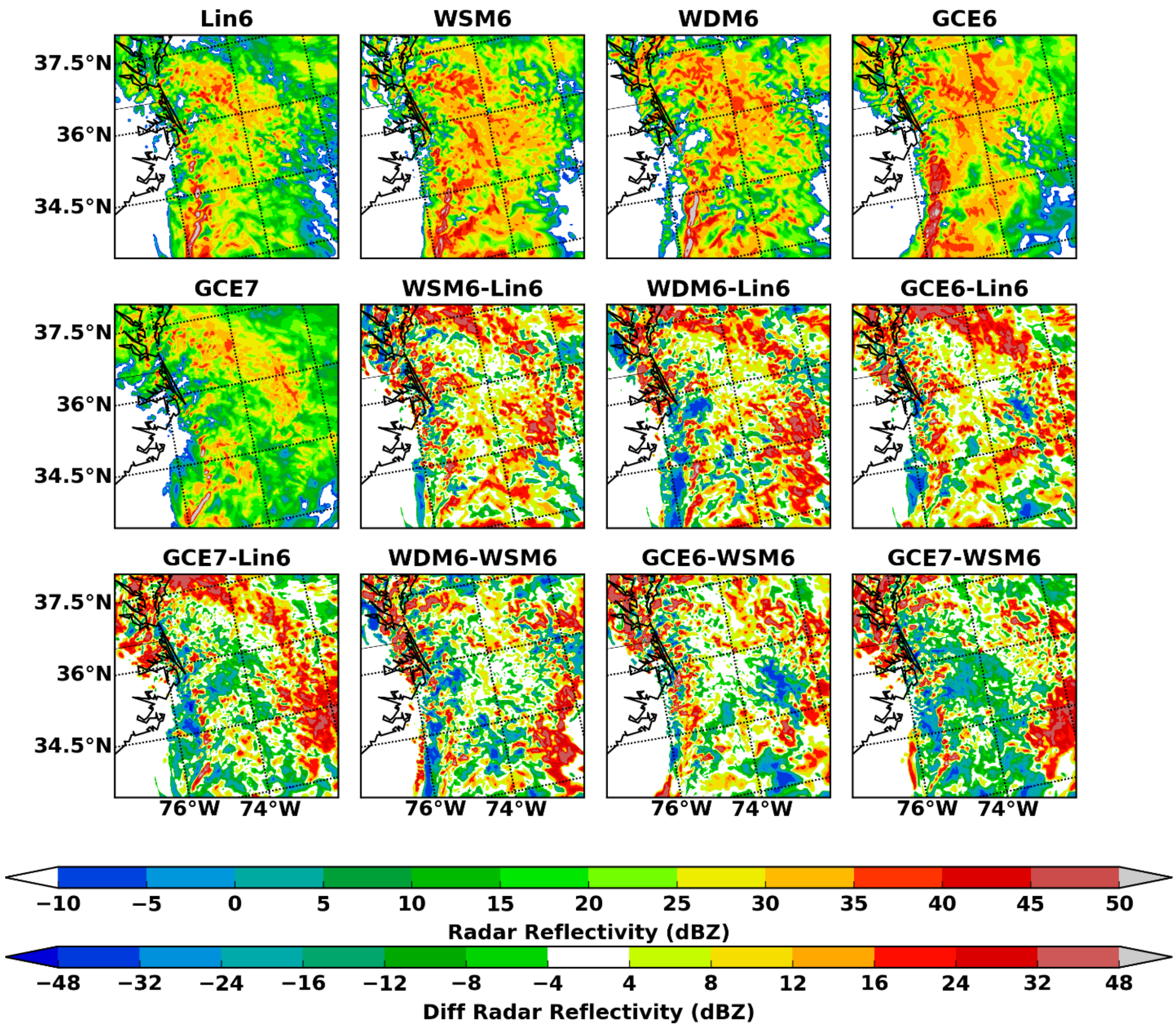

Figure 3. Simulated radar reflectivity (dBZ) at 4,000 m above mean sea level and their difference at the same time as Fig. 2.

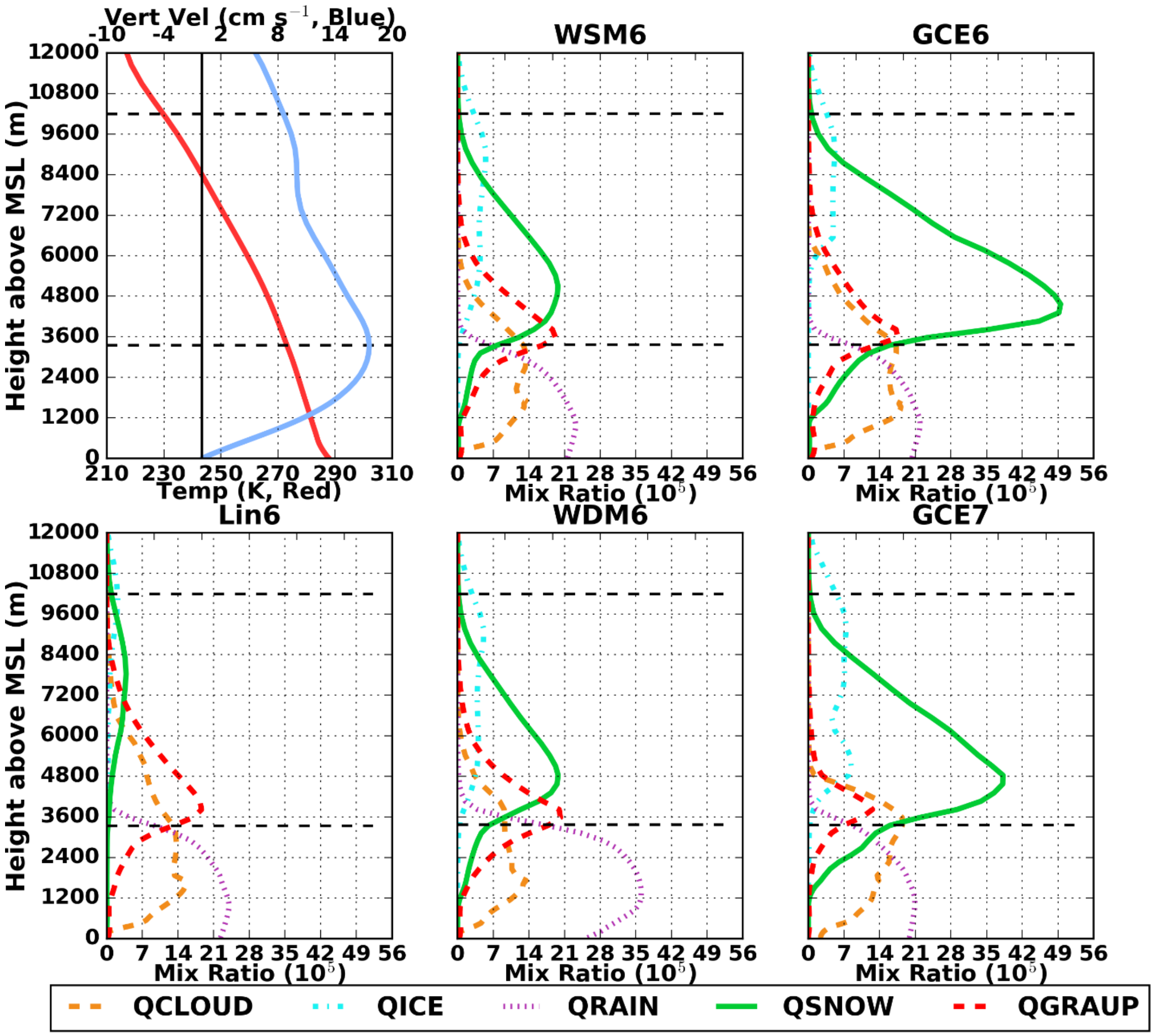

Figure 4. Domain 4-averaged (1.167-km grid spacing) mixing ratios (kg kg^{-1}), temperature (K), and vertical velocity (cm s^{-1}) at the same time as Fig. 2. . The black dashed lines denote the height above mean sea level (MSL) where the air temperature is 0°C or -40°C. The upper-left panel shows composited and model-averaged profiles of temperature (red line) and vertical velocity (blue). Mixing ratio species abbreviations are QCLOUD (cloud water), QGRAUP (graupel), QICE (cloud ice), QRAIN (rain), QSNOW (snow) and QHAIL (hail).

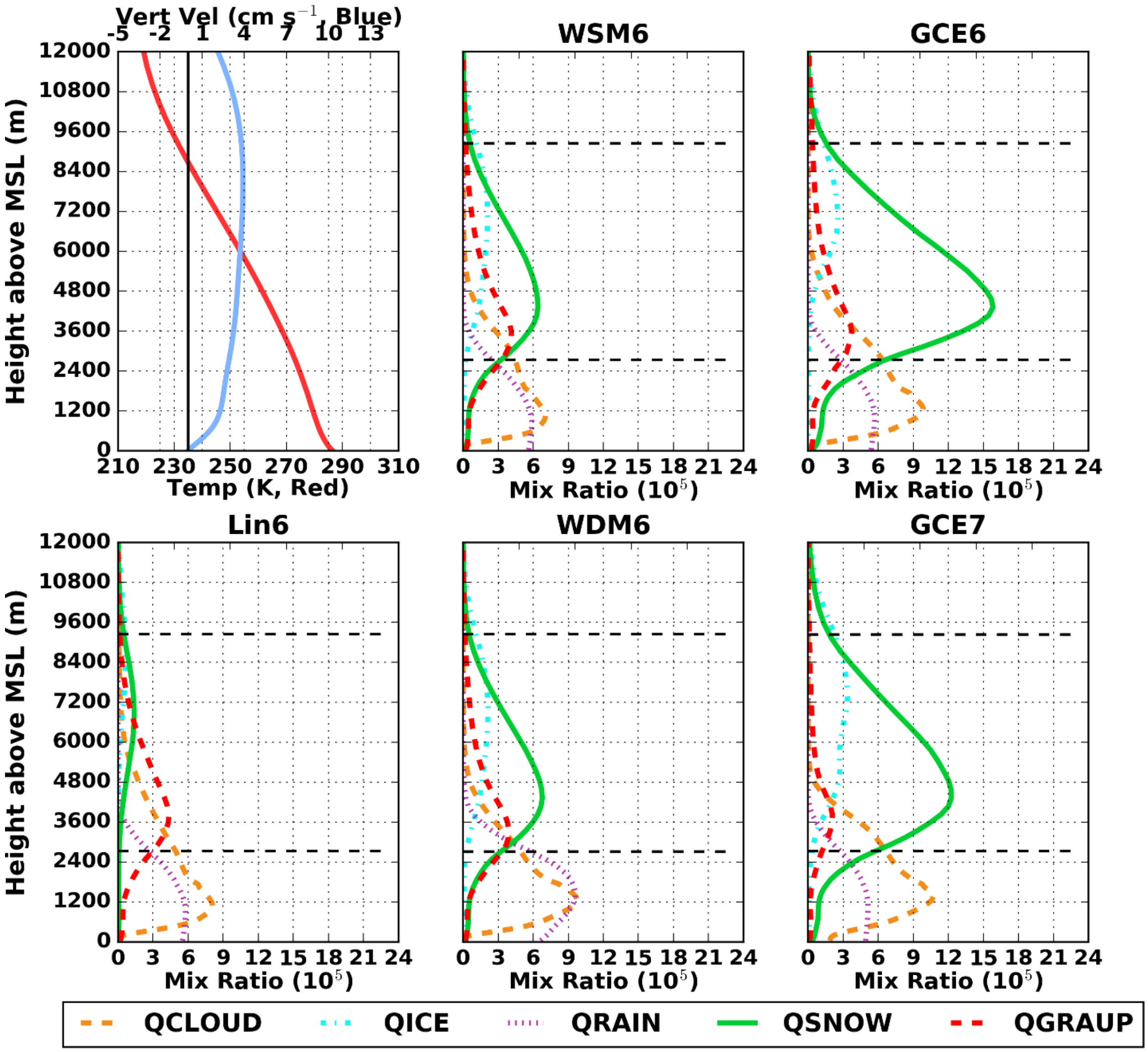

Figure 5. Domain 4-averaged (1.167-km grid spacing), composite mixing ratios (kg kg^{-1}), temperature (K), and vertical velocities (cm s^{-1}) composited over all seven nor'easter events. The black dashed lines denote the height above mean sea level (MSL) where the air temperature is 0°C or -40°C. The upper-left panel shows composited and model-averaged profiles of temperature (red line) and vertical velocity (blue). Mixing ratio species abbreviations are QCLOUD (cloud water), QGRAUP (graupel), QICE (cloud ice), QRAIN (rain), QSNOW (snow) and QHAIL (hail).

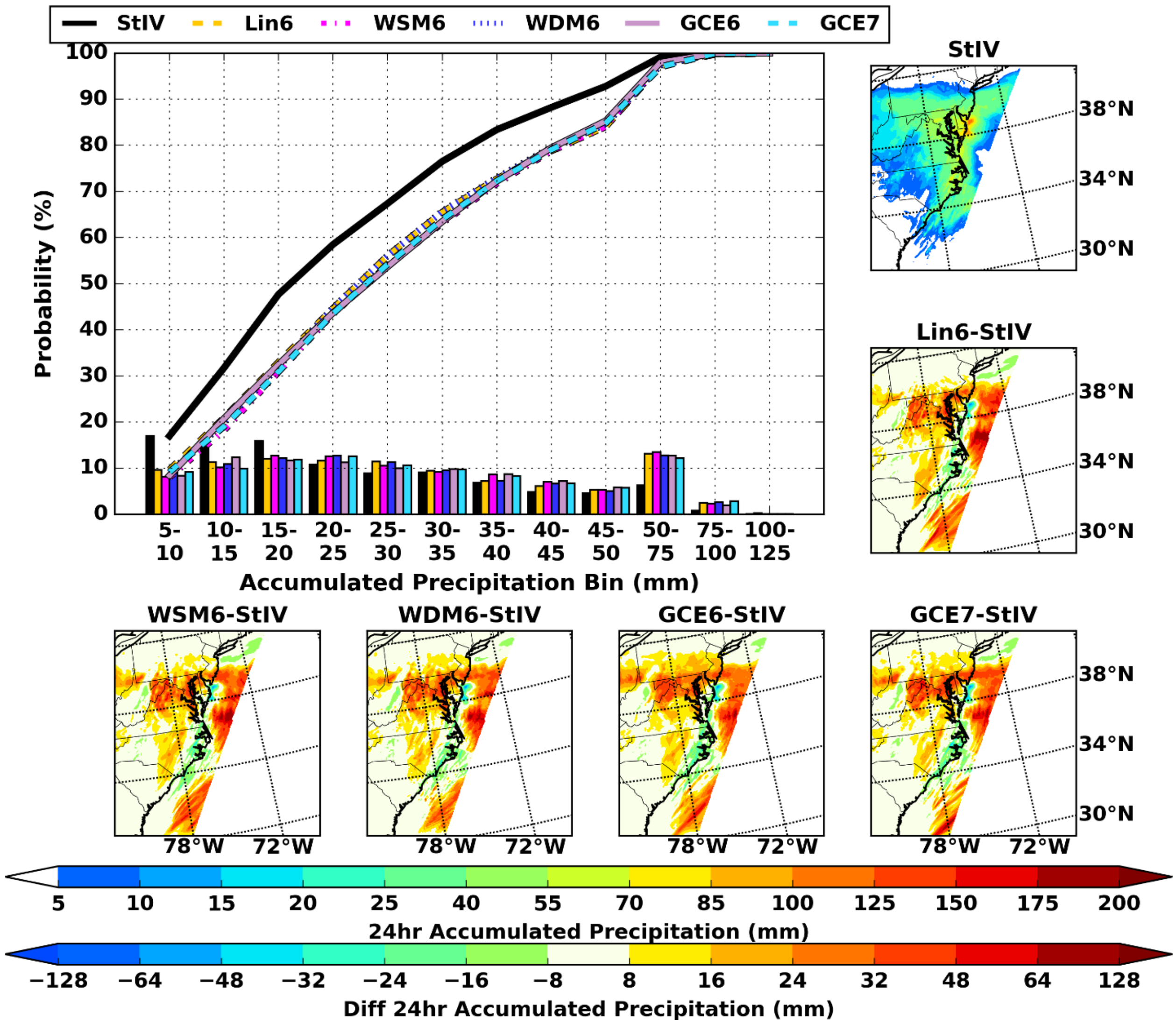

Figure 6. Case 5, 24-hour precipitation accumulation and their differences (mm, small panels) and corresponding probability density and cumulative distribution functions (big panel) of these same data derived from Stage IV and WRF model output. Accumulation period is from 00 UTC 06 February 2010 – 00 UTC 07 February 2010. Shown differences are model - Stage IV (StIV).

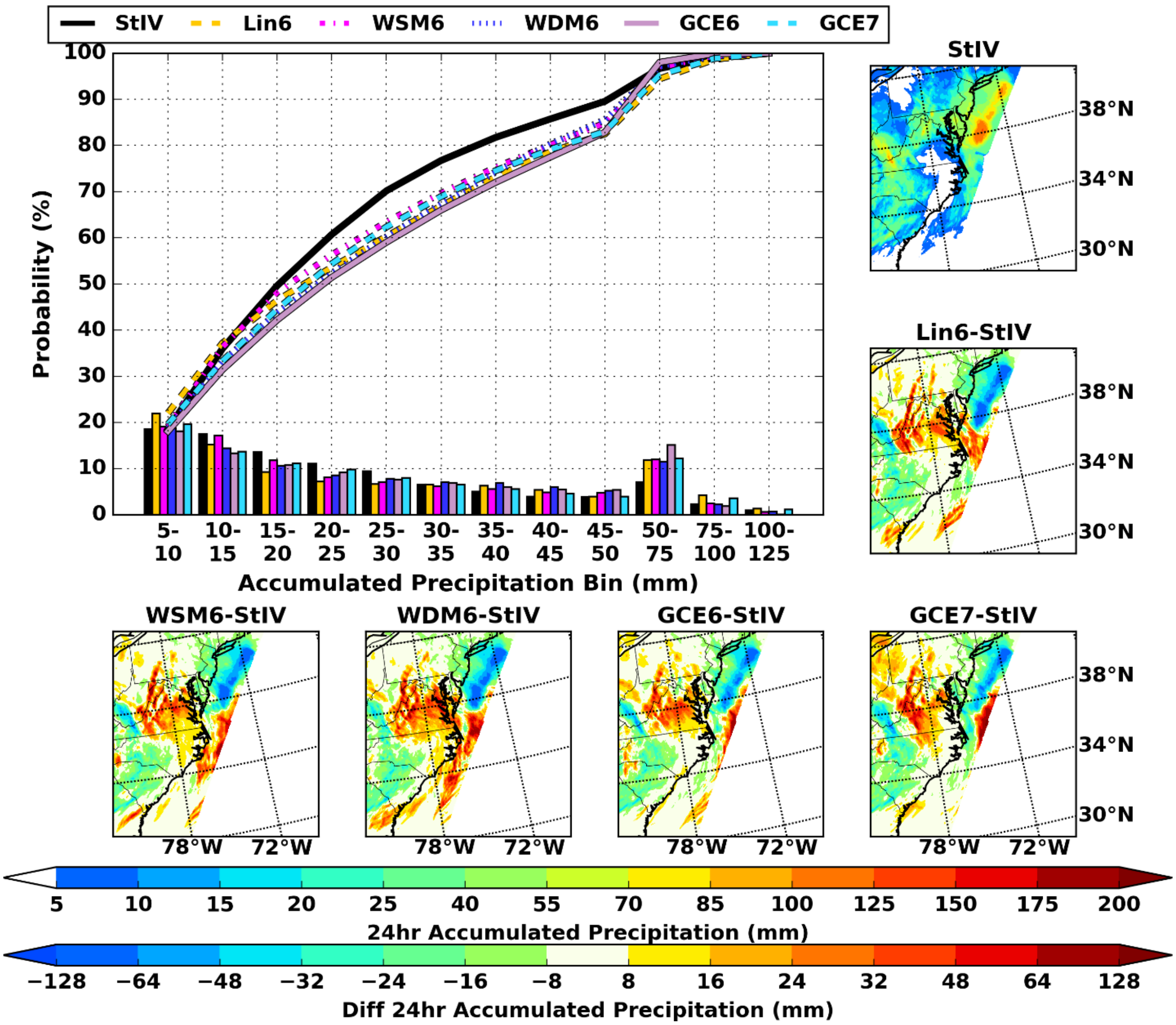

Figure 7. Case 7, 24-hour precipitation accumulation and their differences (mm, small panels) and corresponding probability density and cumulative distribution functions (big panel) of these same data derived from Stage IV and WRF model output. Accumulation period is from 18 UTC 12 March 2010 – 18 UTC 13 March 2010. Shown

differences are model - Stage IV (StIV).

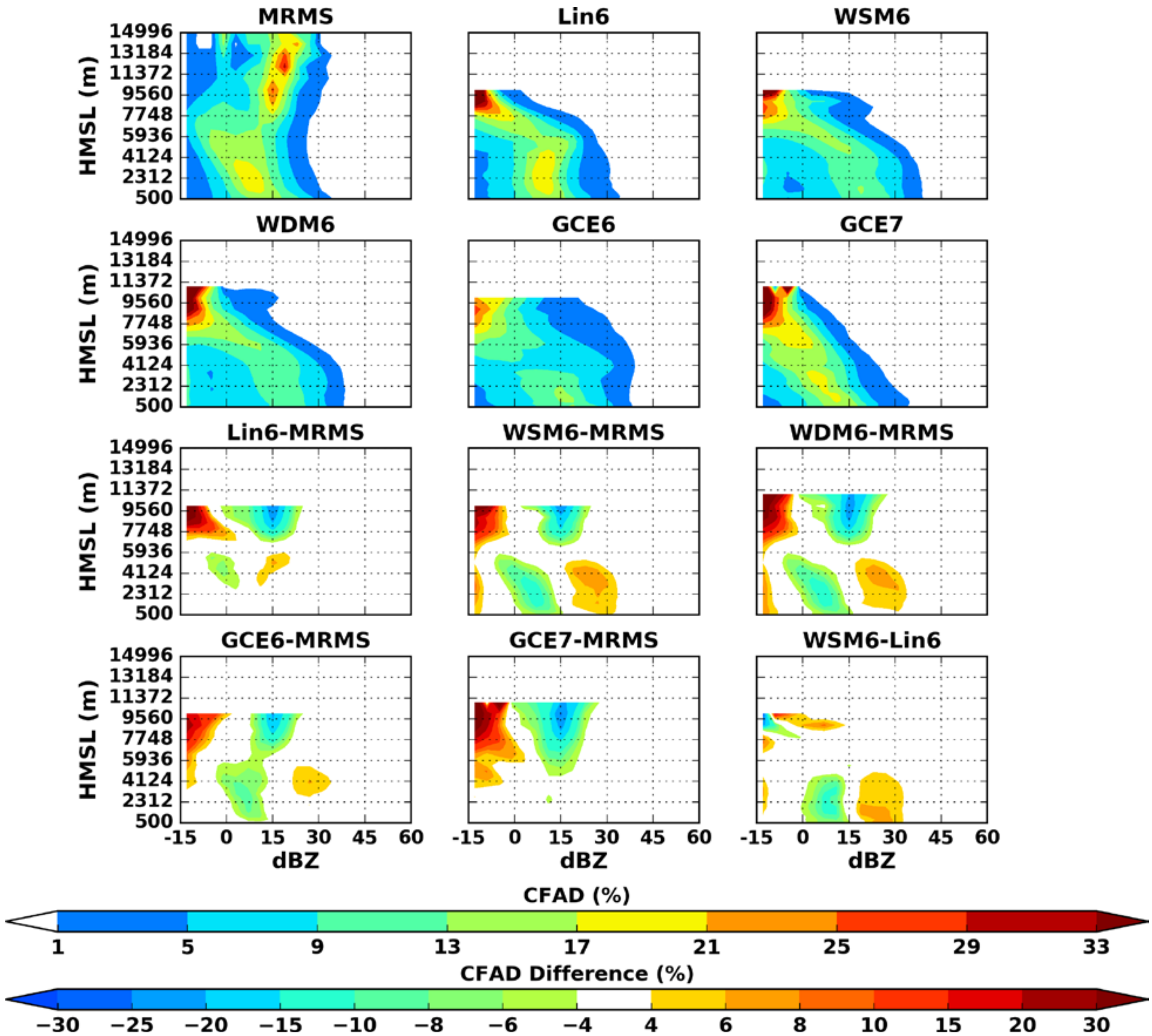

Figure 8. Domain 3 (5 km grid spacing), contoured frequency with altitude diagram (CFAD) of radar reflectivity and indicated differences from Case 4 (January 2015). Data accumulation period spans 12 UTC 26 January 2015 – 12 UTC 27 January 2015 during the transit of the nor'easter through Domain 4. The y-axis shows height above mean sea level (HMSL).

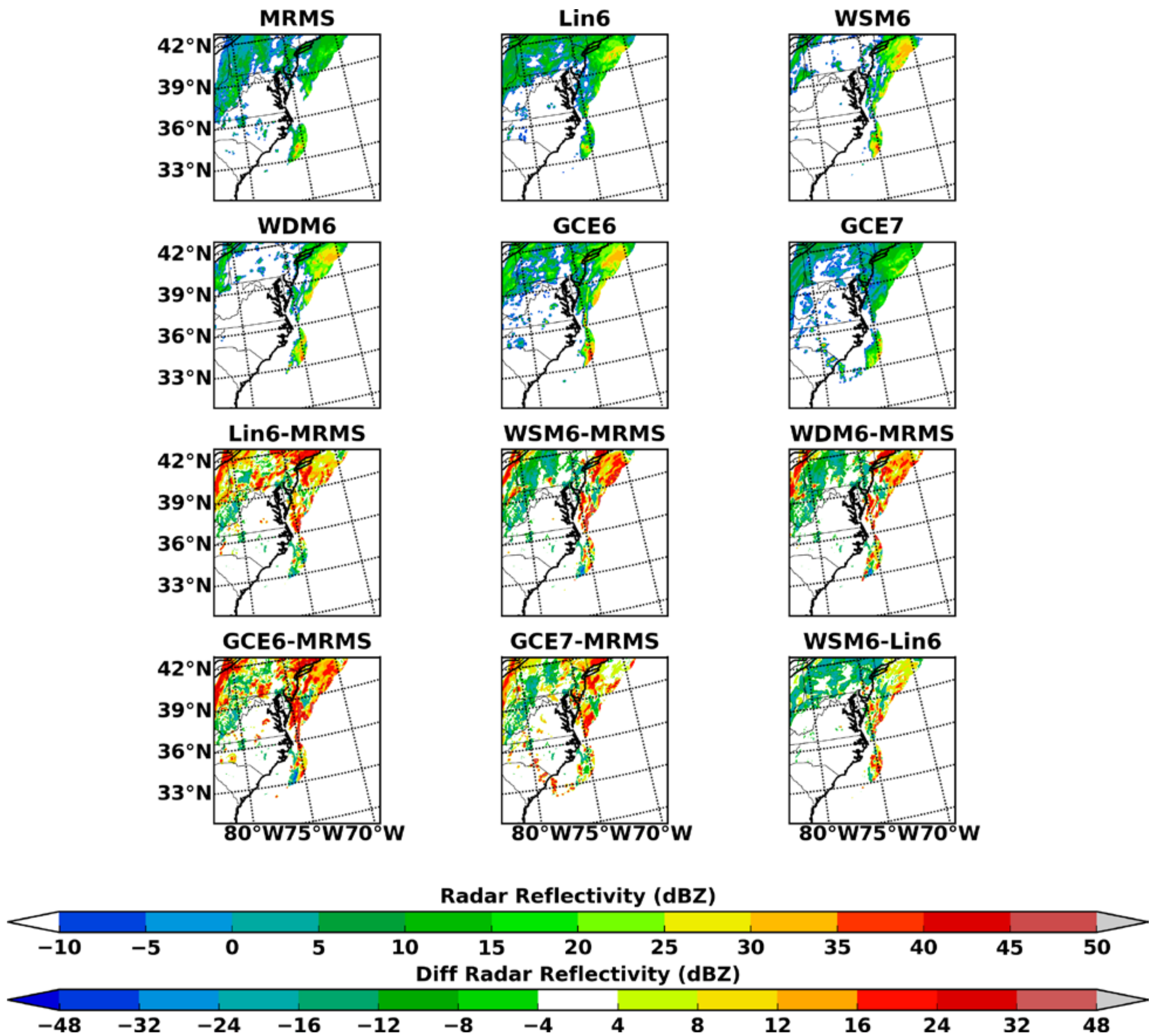

Figure 9. MRMS radar reflectivity and WRF simulated radar reflectivity (dBZ) at 3,000 m above sea level at 18 UTC 26 January 2015. Show radar reflectivity differences are as indicated.

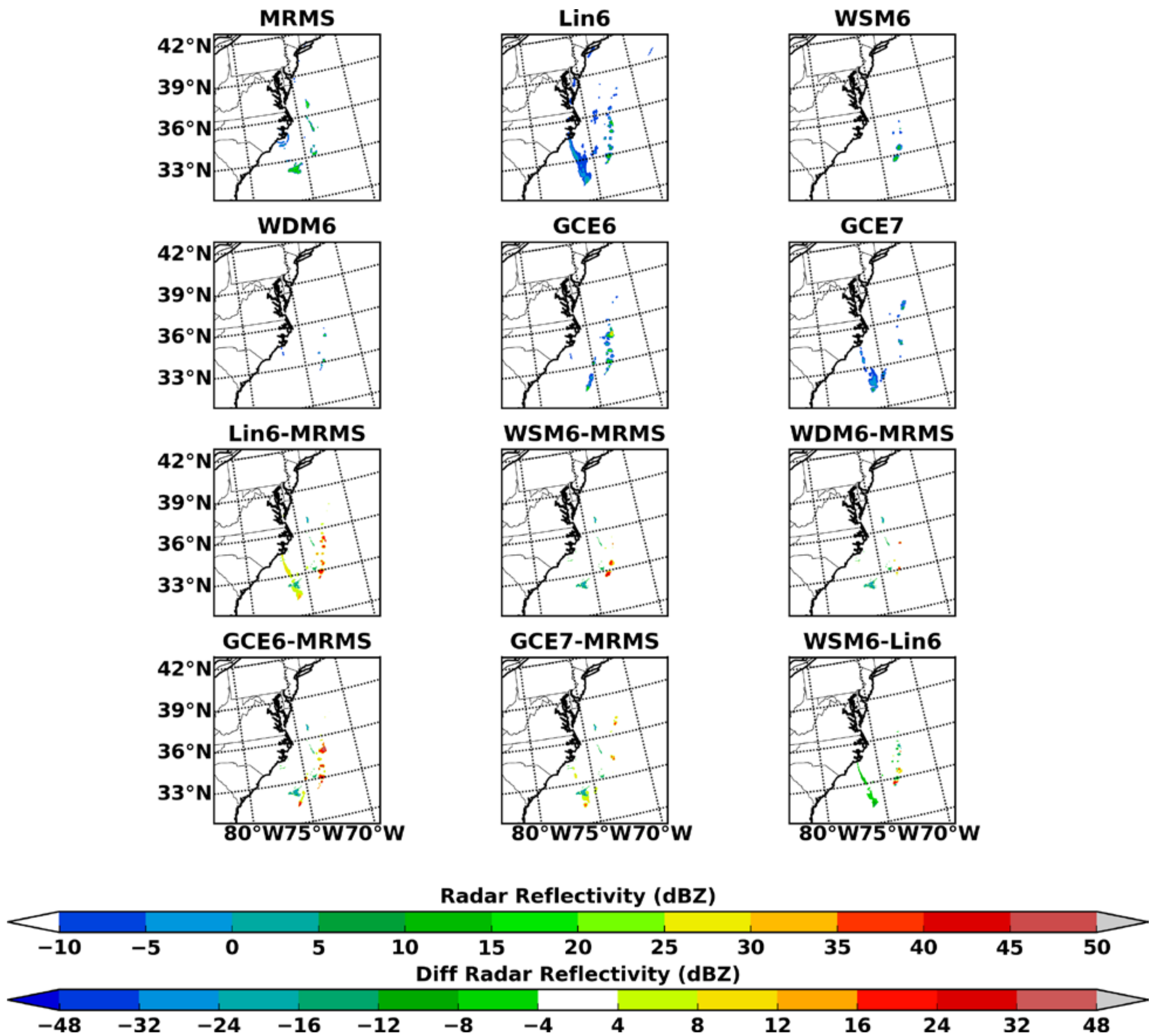

Figure 10. MRMS observed radar and WRF simulated radar reflectivity (dBZ) at 9,000 m above sea level at 18 UTC 26 January 2015. Show radar reflectivity differences are as indicated.

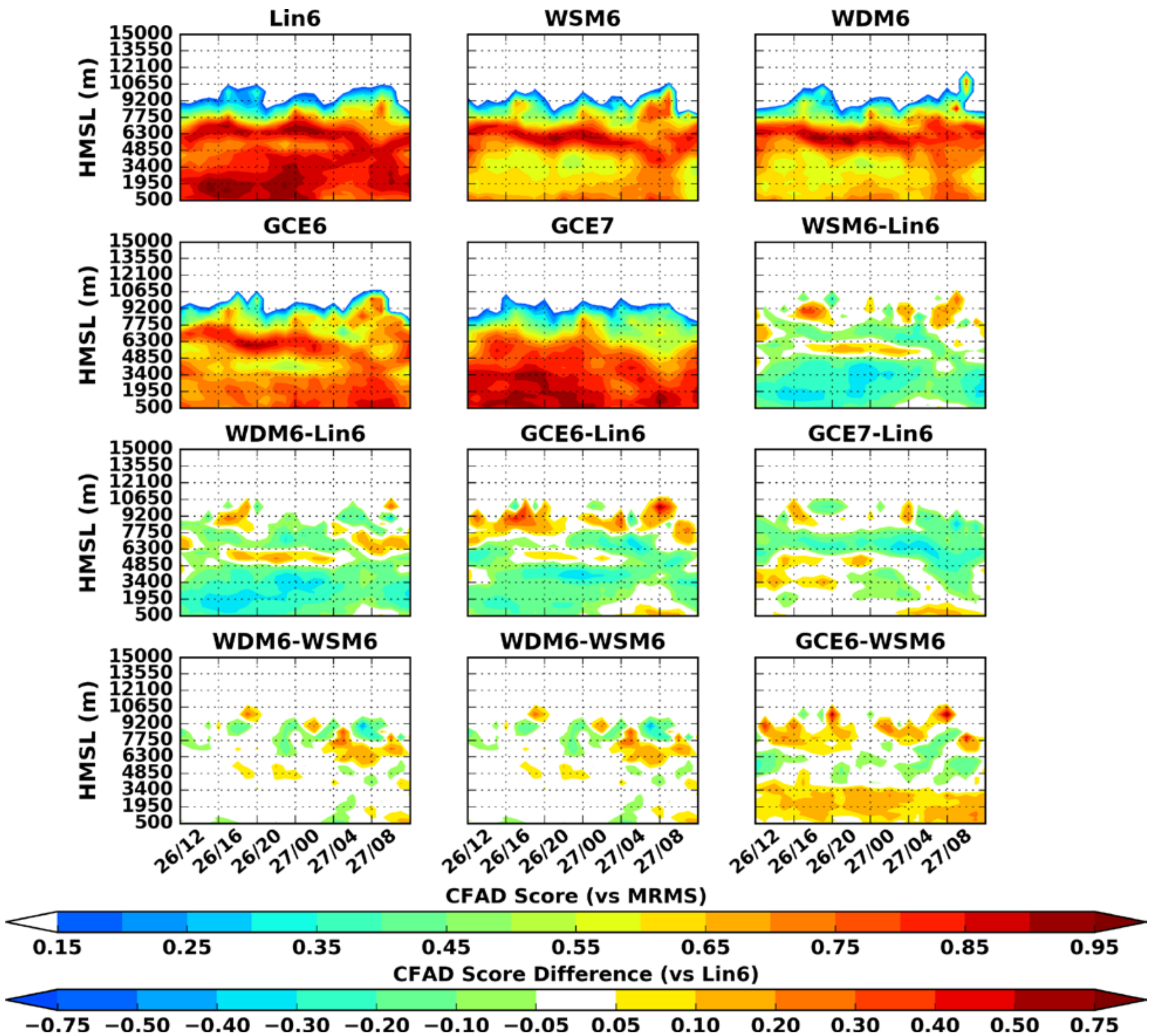

Figure 11. Domain 3, (5 km grid spacing), hourly CFAD scores (See Eq. 2) of radar reflectivity and indicated differences from Case 4 starting 12 UTC 26 January 2015 and ending on 12 UTC 27 January 2015. The time period corresponds to the same time period as in Figure 5. The y-axis shows height above mean sea level (m).